This homework consists of 3 problems exploring the concept of equilibrium steady states in bioregulation. Problem 1 gets you some hands-on idea about how equilibrium enables models made up of physically measurable quantities. Then, we venture into discussions about the idea of equilibrium.

Equilibrium is a very powerful concept, yet it is often not fully appreciated in our understanding of biological systems. On one hand, biologists may out-right say there exists non-equilibrium driving forces everywhere in biology. This does not prevent the applicability of equilibrium at all, since although biology is driven, it is *sparsely driven*, so most parts of biology have to be at equilibrium. And we don't need everything in life to be equilibrium for this concept to be (very) useful. This is highlighted in Problem 3, where equilibrium is used to understand binding reactions, and this is compared with other models we have used so far, i.e. rate equations and chemical master equations.

On the other hand, biophysicists, in response to the existence of driving forces in biology, over-emphasize the importance of non-equilibrium concepts. This is discussed in Problem 2, where the point "it takes a lot of driving to be not like equilibrium" is made. What is worse is that, as a result of this, when indeed nonequilibrium behaviors become important, there is so much energy spent that theoretical tools of nonequilibrium calculating how much energy is spent for some cycle fluxes etc become useless. System behavior become almost deterministic, or as non-equilibrium as you want, and what becomes important instead, is the question of functional design. What mechanisms, or how to wire parts together, so that a certain function is achieved? The question about energy spending become rather secondary.

Above is the perspective I would like to convey through this homework's exploration of equilibrium. Hope you enjoy it.

## 1 Equilibrium builds models from measurable quantities – allosteric regulation

One power of equilibrium models is that the parameters are often quantities that can be individually measured by separate experiments. This resolves a problematic issue of kinetic models we often use, such as chemical reaction networks, where kinetic rate constants such as catalysis rates, on rates and off rates are often hard to determine independently. As a result, often when kinetic models are used to explain data, the data is simply fit to the model with all parameters varied. But this greatly reduces the explanatory power of the kinetic model, since many models with enough parameters can fit the same set of data. Therefore, it is arguable that reducing the number of parameters left to fit to data can greatly enhance the explanatory power of a model. Let us have a taste of this by looking at the models of gene regulation via allostery, as presented in [1]. An allosteric model of gene expression by the Lac operon is constructed. This model consists of two parts. On one hand, the LacI repressor is regulated by inducers via allostery, so it has both an activated form and an inactive form, denoted $R_A$ and $R_I$, respectively. These quantities are related to the total concentration of the repressor via the probability that a repressor is in activated form, $p_A(c)$, by $R_A = p_A(c)R$, where $R$ is the total concentration of repressors, and $c$ is the inducer concentration. Note that an assumption is made here that either we can directly tune $c$, or that $c \gg R$ so that the inducer concentration does not change significantly upon binding with repressors. These assumptions are needed so that each repressor molecules can be considered independent and identically distributed, such that the probability of one repressor can simply multiply the total concentration of repressors to represent the whole population of repressors.

On the other hand, the transcriptional activity is determined by the probability that RNAP is bound to the promoter, $p_{\text{bound}}(R_A, R_I)$, which changes with the concentration of the repressor in activated form $R_A$, and that of the repressor in inactive form $R_I$. Note that similar to the consideration for $p_A(c)$ above, an assumption that $R_A$ $R_I$, and the concentration of RNAP are much larger than the concentration of promoter is made here.

Combine $p_{\text{bound}}(R_A, R_I)$ with $p_A(c)$ and $R_A = p_A(c)R$, we can obtain $p_{\text{bound}}(c)$, the transcriptional activity as a function of inducer concentration $c$. This result then serves as a prediction for experimental observations, with parameters inside that can be separately determined.

We follow [1] to derive the formula for $p_{\text{bound}}(c)$, and then investigate the formula's predictive power over data.

1. Follow the argument in Figure 2A of [1] to show that Equation 1 (copied below) holds.

$$p_{\text{bound}} = \frac{\frac{P}{N_{NS}} e^{-\varepsilon_P}}{1 + \frac{R_A}{N_{NS}} e^{-\Delta\varepsilon_{RA}} + \frac{R_I}{N_{NS}} e^{-\Delta\varepsilon_{RI}} + \frac{P}{N_{NS}} e^{-\varepsilon_P}} \tag{1}$$

2. Follow the arguments below Equation 2 about the physical estimates on $P$, $N_{NS}$, and $\Delta\varepsilon_P$, to simplify the expression for $p_{\text{bound}}$ above, and obtain Equation 3 for fold-change.

3. Follow the arguments in Figure 2B to derive Equation 4 for $p_A(c)$. Then combine this with Equation 3 to derive Equation 5 for fold change as a function of $c$, reproduced below.

$$\text{foldchange} = \frac{p_{\text{bound}}(R>0)}{p_{\text{bound}}(R=0)} = \left(1 + \frac{(1+c/K_A)^n}{(1+c/K_A)^n + e^{-\varepsilon_{AI}}(1+c/K_I)^n} \frac{R}{N_{NS}} e^{-\varepsilon_{RA}}\right)^{-1}. \tag{2}$$

4. Read the section "Experimental Design", as well as "Inferring Allosteric Parameters from Previous Data". Try to describe, in your own words, what are the difficulties involved in estimating the parameters of the fold-change formula above, and what efforts did the authors make to estimate the parameters independently from the data, so that they can achieve physical prediction, rather than just fitting to data?

## 2 It takes a lot of driving to not look like equilibrium

Equilibrium behaviors should be considered the "default" since it does not require incessant energy spending. Therefore, from a bio-design perspective where our goal is to build a certain function in cells, if we can do so via equilibrium mechanisms, it is much more preferable than using non-equilibrium mechanisms. And if non-equilibrium mechanisms can not be avoided, it should be minimized. Similarly, from a bio-science perspective where our goal is to infer what mechanisms govern a behavior we want to understand, we should always try hard to build equilibrium models to explain it, since equilibrium models are much easier to analyze. Furthermore, if both an equilibrium and a non-equilibrium model explain the behavior, the equilibrium one is more likely to hold since it has less complexity and less cost to the cell.

The view above is supported by the fact that albeit many systems indeed contain non-equilibrium driving here and there, their behaviors are simply equivalent to equilibrium systems. In other words, although there are certain driving forces in these systems, if we only look at the overall behavior of the system, (e.g. input-output,) then it looks just like an equilibrium system! We may give a slogan to this phenomena, that "it takes a lot of driving to not look like equilibrium." We investigate this phenomena in this problem.

### 2.1 On-off switches and chains are always equilibrium-like

Certain networks (chemical reactions or Markov chains), just because of the network topology, their steady states are always equilibrium ones, for all positive reaction rate constants.

The precise condition is called the Wegscheider condition (see Wikipedia page for example), which basically states that (1) the reaction network is reversible, i.e. for every reaction in the network, its reverse reaction is also in the network, and (2) there are no cycle fluxes, so the product of rate constants through each cycle in one direction is

the same as that in the other direction. Therefore, if a network's topology automatically forces the reaction rate constants to always satisfy the no-cycle-flux condition, then all steady states are equilibrium.

We consider one simple example here.

1. Consider an on-off switch where we just have two states 1 and 2, with transition rates $k_{12}$ from 2 to 1, and $k_{21}$ from 1 to 2. This corresponds to the following rate equation:

$$X_1 \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} X_2. \tag{3}$$

   Show that the steady state $(x_1, x_2)$ satisfies detailed-balance condition

$$k_{12}x_2 = k_{21}x_1. \tag{4}$$

   So the steady state is an equilibrium steady state in the sense of detailed balance. As a result, we can define the energy of $x_1$ as $\varepsilon_1$, and the energy of $x_2$ as $\varepsilon_2$, and their energy difference $\Delta\varepsilon_{12} = \varepsilon_1 - \varepsilon_2$ is related to the rate constants by $\frac{x_1}{x_2} = \frac{k_{12}}{k_{21}} = e^{-\Delta\varepsilon_{12}}$.

2. We could add a driving force on top. For example, we could imagine coupling the $X_1$ to $X_2$ transition with a non-equilibrium driving force, such as ATP hydrolysis, so that $X_1$ to $X_2$'s transition rate is larger. Let $q_{21} = k_{21} + \Delta k_{21}$ be the new transition rate under driving. Let us also denote $q_{12} = k_{12}$ as the new transition rate from 2 to 1, although it has not changed.

   Show that the new steady state also satisfies detailed balance condition,

$$q_{12}x_2' = q_{21}x_1', \tag{5}$$

   where $x_1'$ and $x_2'$ are the new steady states. As a result, we can again define an energy $\varepsilon_1'$ for state 1 and $\varepsilon_2'$ for state 2, so their energy difference $\Delta\varepsilon_{12}' = \varepsilon_1' - \varepsilon_2'$ is related to the new rate constants by

$$\frac{x_1'}{x_2'} = \frac{q_{12}}{q_{21}} = e^{-\Delta\varepsilon_{12}'}. \tag{6}$$

   Note, however, $\Delta\varepsilon_{12}'$ is not the same as $\Delta\varepsilon_{12}$ anymore. So although we can still define an effective energy, and consider the steady state as an effectively equilibrium steady state, it is no longer a physical equilibrium where the equilibrium energy is uniquely defined.

3. The same idea that all steady states are effectively equilibrium steady states extends to networks beyond a simple on-off switch. This holds true for a chain, for example, including infinitely long ones, as long as there are no loops in the chain. Consider a production-degradation chain for example, where we have states $X_0$, $X_1$, $X_2$, ..., with a rate $k_i^+$ for transitioning from $X_i$ to $X_{i+1}$, and a rate $k_i^-$ for transitioning from $x_i$ to $x_{i-1}$. Show that its steady state distribution always satisfy detailed balance condition, i.e. $k_i^+ x_i = k_{i+1}^- x_{i+1}$, $i = 0, 1, 2, \ldots$.

## 2.2 Non-equilibrium enables non-monotonic gene expression under strong driving

The work [2] give a great study looking into the phenomena about the generality of equilibrium behaviors. Considering non-equilibrium driving over gene expression modeled as a 4-state transition, the authors looked at behaviors that only non-equilibrium can have.

1. Look at Figure 1 of [2]. This is the model we consider. The promoter is assumed to have 4 states, $S$, the empty genome state, $X$, the state with transcription factor $X$ bound, $P$, the state with RNAP bound, and $XP$, the state with both transcription factor $X$ and RNAP bound. The input is of course the concentration of the transcription factor $X$. The output, or activity, of gene expression, is considered a non-negative combination over the probability of each state, $\langle r\rangle = \sum_i r_i p_i$, where $i \in \{S, X, P, XP\}$ and $r_i$ is the activity of state $i$, and $p_i$ is the probability to be in state $i$.

   Consider the system's equilibrium steady state. Let us show that the equilibrium behavior is always of the following form, which is Equation 2 in the paper.

$$\langle r\rangle^{\text{equ}} = \frac{A' + B'X}{C' + D'X}, \tag{7}$$

   where $A', B', C', D'$ are positive parameters independent of $X$. This form eliminates the possibility of non-monotonic behaviors. Always, $\langle r\rangle^{\text{equ}}$ gradually increases or decreases with $X$.

   Denote $\rho_i = \frac{p_i}{p_S}$, for $i \in \{X, P, XP\}$. Show that

$$\langle r\rangle^{\text{equ}} = \frac{r_S + r_P\rho_P + r_X\rho_X + r_{XP}\rho_{XP}}{1 + \rho_P + \rho_X + \rho_{XP}}. \tag{8}$$

2. Show that, at equilibrium,

$$\rho_P = \frac{k_{PS}P}{k_{SP}}, \quad \rho_X = \frac{k_{SX}X}{k_{XS}}, \quad \rho_{XP} = \frac{k_{X,XP}P}{k_{XP,X}}\frac{k_{SX}X}{k_{XS}} = \frac{k_{P,XP}X}{k_{XP,P}}\frac{k_{PS}P}{k_{SP}}. \tag{9}$$

   where the rate constant notation is that $k_{ij}$ is going from state $i$ to $j$, following the convention in the paper.

3. Insert out expression for $\rho_i$ above into $\langle r\rangle^{\text{equ}}$, and show that it can be written as Eqn (7).

   Argue that, this means the gene expression's response to changing transcription factor concentration can only be monotonically increasing or decreasing.

4. Now we consider non-equilibrium behavior. The general solution is rather hard, as is almost always with non-equilibrium, even for this case of just 4 states forming a square. The solution method in this paper uses the Matrix Tree Theorem, where the nonequilibrium steady state probabilities are expressed in terms of trees that are subgraphs of the 4-state square state transition graph.

   (If the reader is interested, a simple case to consider is the extremely nonequilibrium case where only the clockwise rates are nonzero, with the addition of rate $k_{P,XP}X$ from $P$ to $XP$, and solve the steady state by looking at the null space of the state transition matrix. This is one of the simplest cases that yields non-equilibrium behavior. In fact, if only clockwise rates are nonzero, even though this is an nonequilibrium extreme, its behavior is just like equilibrium...)

   The reader is referred to the original paper for derivation of the general case. The result is Equation 1 of the paper:

$$\langle r\rangle = \frac{A + BX + CX^2}{D + EX + FX^2}, \tag{10}$$

   where $A, B, C, D, E, F$ are positive constants. So this allows non-monotonic behavior.

   Look at Figure 5A of the paper, where the dependence of the non-monotonicity to the driving force is depicted. How much driving force is needed to exhibit non-monotonic behavior that simply cannot be

explained by equilibrium models? Is this driving force realistic for this biological setting? (Note that X is transcription factor, and the driving force is measured from the cycling flux produced, which could be a significant underestimate of chemical fuel burned when driving the cycle.) Does this depend on the direction of driving (i.e. increasing or decreasing a rate constant)?

What would be your conclusion overall about whether it is easy or hard to produce behaviors that look distinctively different from equilibrium ones?

## 3 Stochastic, discrete, and statistical mechanics of binding reactions

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

$$E + S \underset{k^-}{\overset{k^+}{\rightleftharpoons}} C \tag{11}$$

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 $(N_E, N_S, N_C)$, which are the number of free enzymes, free substrates, and complexes.

### 3.1 Deterministic steady state solution

Let us solve for the deterministic solution first.

1. Show that the deterministic rate equation of the binding reaction is

$$\frac{d}{dt}x_C = k^+ x_E x_S - k^- x_C, \tag{12}$$

where $(x_E, x_S, x_C)$ are continuous variables representing concentrations for free enzyme $E$, free substrate $S$, and the complex $C$. Here the association rate constant $k^+$ has unit per nM per second, and $k^-$ has unit per second.

2. Show that the steady state equation is

$$x_C = \frac{x_E x_S}{K}, \tag{13}$$

where $K = \frac{k^-}{k^+}$.

3. Define $q_E = x_E + x_C$ as the total concentration of $E$, and $q_S = x_S + x_C$ as the total concentration of $S$. Show that, to solve for steady state $x_C$ in terms of $(q_E, q_S, K)$, we can derive the following equation from the steady state condition.

$$x_C K = (q_E - x_C)(q_S - x_C). \tag{14}$$

4. Solve the above to obtain

$$x_C = \frac{1}{2}\left((q_E + q_S + K) - \sqrt{(q_E + q_S + K)^2 - 4q_E q_S}\right). \tag{15}$$

### 3.2 Chemical master equation of simple binding

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

For a given initial condition, the total number of enzymes $N_{\text{tot}}^E = N_E + N_C$ and substrates $N_S^{\text{tot}} = N_S + N_C$ are fixed, since they are not changed by the reactions. Define $N_C^{\max} = \min\{N_E^{\text{tot}}, N_S^{\text{tot}}\}$. Because the number of complexes is always less than the total of enzyme and the total of substrates, i.e. $N_C \le N_C^{\max}$, the states that $N_C$ can jump to are $\{0, 1, \dots, N_C^{\max}\}$. The states that $N_E$ and $N_S$ can jump to can be obtained from the fact that $N_E = N_E^{\text{tot}} - N_C$. For example, if $N_S^{\text{tot}} \ge N_E^{\text{tot}}$, then $N_S$ can reach states $\{N_S^{\text{tot}} - N_E^{\text{tot}}, N_S^{\text{tot}} - N_E^{\text{tot}} + 1, \dots, N_S^{\text{tot}}\}$.

Because of the constraints that $N_S^{\text{tot}}$ and $N_E^{\text{tot}}$ are constant, we can simplify the state of the system by considering just one variable, $N_C$. Then the other two variables can be obtained from $N_S = N_S^{\text{tot}} - N_C$, and $N_E = N_E^{\text{tot}} - N_E$. So the system becomes a one-dimensional stochastic jump process on the state space $N_C \in \{0, 1, \dots, N_C^{\max}\}$. Transition $N_C \to N_C + 1$ has rate $k^+(N_E^{\text{tot}} - N_C)(N_S^{\text{tot}} - N_C)$, and transition $N_C \to N_C - 1$ has rate $k^- N_C$.

Now, we write down the dynamics of this jump chain and solve for the steady state distribution of $N_C$.

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

$$\begin{aligned}\frac{d}{dt}p(N_C,t) =& k_d^-(N_C+1)p(N_C+1,t) + k_d^+(N_E^{\text{tot}} - N_C + 1)(N_S^{\text{tot}} - N_C + 1)p(N_C-1,t) \\ &- \Big(k_d^- N_C + k_d^+(N_E^{\text{tot}} - N_C)(N_S^{\text{tot}} - N_C)\Big)p(N_C,t),\end{aligned} \tag{16}$$

   when $N_C \in \{0, \dots, N_C^{\max}\}$. Here $p(N_C, t)$ is assumed zero always for $N_C < 0$ or $N_C > N_C^{\max}$. Here $k_d^+ = k^+/V$ is the reaction rate constant for discrete changes of molecule counts, which is the continuous reaction rate constant $k^+$ divided by the reaction volume $V$ (assumed constant.) Therefore, $k_d^+$ has the unit of per molecule per second. The discrete change reaction rate constant for dissociation is $k_d^-$, which has the unit of per second. This is the same unit as the continuous version, so $k_d^- = k_c^-$.

2. 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:

$$k_d^+(N_E^{\text{tot}} - N_C)(N_S^{\text{tot}} - N_C)p(N_C) = k_d^-(N_C+1)p(N_C+1), \qquad N_C = 0, \dots, N_C^{\max} - 1. \tag{17}$$

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

3. Using the above detailed balanced condition, show that this implies the steady state distribution satisfies

$$p(N_C) \propto \frac{K_d^{-N_C}}{N_E! N_S! N_C!}, \qquad N_E = N_E^{\text{tot}} - N_C,\ N_S = N_S^{\text{tot}} - N_C,\ K_d = \frac{k_d^-}{k_d^+} = \frac{k^-}{k^+}V = KV. \tag{18}$$

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

4. We can connect this steady state distribution $p(N_C)$ to the deterministic steady state solution $x_C = \frac{x_E x_S}{K}$, where $(x_E, x_S, x_C)$ are steady state concentrations for free enzyme $E$, free substrate $S$, and the complex $C$.

Let us look at the mode of the steady state distribution, i.e. $N_C$ that achieves the maximum of $p(N_C)$, in the limit that molecule numbers $N_E, N_S, N_C$ are large, so that we can consider the variables as continuous. Use Stirling's formula that $\log N! \approx N \log N - N$ when $N$ is large to show that when $N_E, N_S, N_C$ are large,

$$\log p(N_C) \approx \text{const} - N_C \log K_d - N_E \log N_E + N_E - N_S \log N_S + N_S - N_C \log N_C + N_C, \tag{19}$$

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

$$\frac{d}{dN_C} \log p(N_C) = \log \frac{(N_C^{\text{tot}} - N_C)(N_S^{\text{tot}} - N_C)}{N_C K_d}. \tag{20}$$

Conclude that the maximum of $p(N_C)$ is achieved at

$$N_C^* = \frac{(N_E^{\text{tot}} - N_E)(N_S^{\text{tot}} - N_C)}{K_d} = \frac{N_E N_S}{K_d}, \tag{21}$$

which is the deterministic steady state solution. Indeed, using the fact that the discrete dissociation constant relates to the continuous dissociation constant by $K_d = \frac{k^-}{k^+} V = KV$, and $x_E = \frac{N_E}{V}$, we have

$$x_C^* = \frac{(q_E - x_E)(q_S - q_C)}{K} = \frac{x_E x_S}{K}. \tag{22}$$

### 3.3 Equilibrium distribution from statistical physics

We can equivalently derive the equilibrium distribution via statistical mechanical arguments. This gives a direct derivation of the equilibrium distribution, independent of the chemical reaction network and the rate constants, and offers an energy interpretation of all the parameters.

1. Consider a lattice model, where there are $\Omega$ lattice sites, and each enzyme molecule $E$ or a substrate molecule $S$ occupies one lattice site. The enzyme and substrate molecules can be present in the form of a free enzyme $E$, a free substrate $S$, or a complex $C$ combining one enzyme and one substrate. Let $(N_E, N_S, N_C)$ denote the number of free enzyme, free substrate, and complexes, and $N_E^{\text{tot}}$, $N_S^{\text{tot}}$ as the total number of enzyme and substrate molecules.

   Argue that the system state with $N_C$ molecules of complexes has energy

   $$N_C \varepsilon_b + (N_S^{\text{tot}} - N_C)\varepsilon_f^S + (N_E^{\text{tot}} - N_C)\varepsilon_f^E, \tag{23}$$

   where $\varepsilon_b$ is the energy of enzyme-substrate binding, $\varepsilon_f^S$ is the energy of a free substrate, and $\varepsilon_f^E$ is the energy of a free enzyme.

   Then argue that the multiplicity of this state is approximately the following, assuming $\Omega$ is large:

   $$\frac{\Omega^{N_E+N_S+N_C}}{N_E! N_S! N_C!} = \Omega^{N_E^{\text{tot}}+N_S^{\text{tot}}} \frac{\Omega^{-N_C}}{(N_E^{\text{tot}} - N_C)!(N_S^{\text{tot}} - N_C)! N_C!}. \tag{24}$$

2. Put the energy and the multiplicity together to show that the equilibrium probability in state $N_C$ is proportional to

   $$p(N_C) \propto \frac{(\Omega e^{-\Delta\varepsilon})^{N_C}}{N_E! N_S! N_C!} \tag{25}$$

   where $\Delta\varepsilon = \varepsilon_b - \varepsilon_f^S - \varepsilon_f^E$.

   Compare this with the steady state distribution we derived in the previous case. How is the binding constant $K$ related to energies?

### Binding networks' equilibrium distributions are always Poisson-like (Further references)

In the previous calculation, we saw that the steady state distribution of $(N_E, N_S, N_C)$ of the binding reaction satisfies that it is Poisson-like, i.e. it is a Poisson distribution under the stoichiometric constraint that $N_E + N_C = N_E^{\text{tot}}$ and $N_S + N_C = N_S^{\text{tot}}$. This result turns out to hold in general for binding networks. See [3] for example, especially Theorem 4.5 from an earlier work by Whittle in 1986.

The specific reasoning is the following. A binding network is reversible, in the sense that for every chemical reaction in the network, this reaction's reverse reaction is also in the network. It is a known result that (Theorem 4.5 in [3] mentioned above) for a reversible chemical reaction network, if its rate constants satisfies that its deterministic rate equation has a detailed balance steady state, i.e. an equilibrium steady state in our words, then its chemical master equation has an equilibrium steady state distribution $\pi$. Here a distribution is equilibrium, or detailed balanced, is in the sense of a reversible distribution of Markov chains. In other words, for any two states $x$ and $y$, denote $q_{xy}$ as the rate of transiting from $y$ to $x$ in this Markov chain, then $q_{xy}\pi_y = q_{yx}\pi_x$.

As a result of this, if $\boldsymbol{x} \in \mathbf{R}^n_{\geq 0}$ is the deterministic equilibrium steady state of the system, then the equilibrium steady state distribution is always of the form

$$\pi(\boldsymbol{N}) = \sum_{\Gamma} \alpha_\Gamma \pi_\Gamma(\boldsymbol{N}), \qquad \pi_\Gamma(\boldsymbol{N}) = \begin{cases} M_\Gamma \prod_{i=1}^n \frac{x_i^{N_i}}{N_i!}, & \boldsymbol{N} \in \Gamma \\ 0, & \boldsymbol{N} \notin \Gamma, \end{cases} \tag{26}$$

where $\boldsymbol{N} \in \mathbb{Z}^n_{\geq 0}$ is the count vector of the molecular species, $\alpha_\Gamma \geq 0, \sum_\Gamma \alpha_\Gamma = 1$, and the sum is over the closed, irreducible subsets $\Gamma$ of the state space. Here $M_\Gamma$ is a normalizing constant.

For our example of one binding reaction $E + S \rightleftharpoons C$, each $(N_E^{\text{tot}}, N_S^{\text{tot}})$ defines one $\Gamma$. So using the formula, the equilibrium distribution is

$$\pi(N_E, N_S, N_C) \propto \frac{x_E^{N_E} x_S^{N_S} x_C^{N_C}}{N_E! N_S! N_C!},$$

if $(N_E, N_S, N_C)$ is in the irreducible subset $\Gamma$ defined by $N_E^{\text{tot}}$ and $N_S^{\text{tot}}$. Now, since $K = x_E x_S / x_C$, and $N_E = N_E^{\text{tot}} - N_C$, we can write the numerator as $x_E^{N_E^{\text{tot}}} x_S^{N_S^{\text{tot}}} K^{-N_C}$, which can be simplified to just $K^{-N_C}$ since the other factors can be absorbed into the proportional sign. So we recover the formula we derived in this problem.

Knowing the exact steady state distribution of a detailed balance network is very powerful. In [4], it is shown that chemical reaction networks can be designed to have equilibrium steady state distributions that have exhibit an arbitrary 2D pattern, where the two axis are the counts of two species, $X$ and $Y$, and each pixel's grayscale corresponds to the probability that the count of $(X, Y)$ is $(N_X, N_Y)$. It is funny that their work showed the power of their construction by exhibiting a pattern like a photo of Darth Vader from Star Wars, but they withdrew those due to potential copyright issues from Disney...

## References

[1] M. Razo-Mejia, S. L. Barnes, N. M. Belliveau, G. Chure, T. Einav, M. Lewis, and R. Phillips, "Tuning transcriptional regulation through signaling: A predictive theory of allosteric induction," *Cell Systems*, vol. 6, no. 4, pp. 456–469.e10, Apr 2018. [Online]. Available: https://doi.org/10.1016/j.cels.2018.02.004

[2] S. D. Mahdavi, G. L. Salmon, P. Daghlian, H. G. Garcia, and R. Phillips, "Flexibility and sensitivity in gene regulation out of equilibrium," *Proceedings of the National Academy of Sciences*, vol. 121, no. 46, p. e2411395121, 2024. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.2411395121

[3] D. F. Anderson, G. Craciun, and T. G. Kurtz, "Product-form stationary distributions for deficiency zero chemical reaction networks," *Bulletin of Mathematical Biology*, vol. 72, no. 8, pp. 1947–1970, Nov 2010. [Online]. Available: https://doi.org/10.1007/s11538-010-9517-4

[4] D. Cappelletti, A. Ortiz-Muñoz, D. F. Anderson, and E. Winfree, "Stochastic chemical reaction networks for robustly approximating arbitrary probability distributions," *Theoretical Computer Science*, vol. 801, pp. 64–95, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030439751930502X