Supplemental Information

1. Appendix 1: Different definitions of cooperativity

The MWC model features several interesting features in its most general form: i) indirect regulation and ii) cooperativity. There are several ways of thinking about cooperativity. [1, 2] In this section we examine three of definitions of cooperativity and find that all of them yield the same two conclusions. First, the MWC molecule with only one binding site or with equal ligand binding affinities in the two states does not yield cooperative interactions. Second, all other MWC molecules exhibit signs of positive cooperativity.

1.1. Ratio of probabilities of ligand binding

A common definition of cooperativity is that the binding of one ligand should encourage or discourage the binding of the next. This definition excludes the one-site MWC molecule from exhibiting signs of cooperativity. Mathematically, this involves comparing the probability of the $(k+1)^{th}$ ligand binding given that k ligands are bound to the probability that the k^{th} ligand binds given that $k-1$ ligands are bound. [2] Label the sites of an n-site receptor as 1 through n. Let $P(k+1|k)$ denote the probability that a $(k+1)^{th}$ ligand binds to site $k+1$ given that sites 1 through k are bound. By definition, this conditional probability can be calculated from

$$P(k+1|k) = \frac{P(\text{sites 1 to k+1 bound})}{P(\text{sites 1 to k bound}) + P(\text{sites 1 to k+1 bound})}. \tag{1}$$

We claim that there is positive cooperativity when $P(k+1|k) > P(k|k-1)$, negative cooperativity when $P(k+1|k) < P(k|k-1)$, and no cooperativity in either direction if $P(k+1|k) = P(k|k-1)$.

To give an intuition as to why this definition works, we will start by analyzing an Adair-like [3] two-site receptor. The two binding sites have identical energies of ligand binding ε_b, but there is a direct energetic interaction $\Delta\varepsilon$ between the two ligand binding sites when both sites are bound. This latter constraint embodies the idea of direct regulation. Using the statistical mechanics formalism in Section 2 of the main text, one can show that

$$P(1|0) = \frac{P(\text{site one bound})}{P(\text{site one bound}) + P(\text{no sites bound})} = \frac{\frac{c}{c_0}e^{-\beta(\varepsilon_b-\mu_0)}}{1+\frac{c}{c_0}e^{-\beta(\varepsilon_b-\mu_0)}} \tag{2}$$

and

$$P(2|1) = \frac{P(\text{both sites bound})}{P(\text{both sites bound}) + P(\text{site one bound})} = \frac{e^{-\beta\Delta\varepsilon}\left(\frac{c}{c_0}e^{-\beta(\varepsilon_b-\mu_0)}\right)^2}{\frac{c}{c_0}e^{-\beta(\varepsilon_b-\mu_0)} + e^{-\beta\Delta\varepsilon}\left(\frac{c}{c_0}e^{-\beta(\varepsilon_b-\mu_0)}\right)^2} \tag{3}$$

Therefore, the probability that a second ligand binds relative to the probability that the first ligand binds is

$$\frac{P(2|1)}{P(1|0)} = e^{-\beta\Delta\varepsilon}, \tag{4}$$

which is greater than 1 if $\Delta\varepsilon < 0$ and less than 1 if $\Delta\varepsilon > 0$. This result agrees with our intuition that an energetically favorable interaction, $\Delta\varepsilon < 0$, between two bound ligands should yield positive cooperativity, encouraging the second ligand to bind. Similarly, an energetically unfavorable interaction, $\Delta\varepsilon > 0$, between two bound ligands should yield negative cooperativity, dissuading the second ligand from binding.

We can do a similar calculation for the n-site MWC molecule, and so we need to calculate

$$P(k+1|k) = \frac{P(\text{sites 1 to k+1 bound})}{P(\text{sites 1 to k+1 bound}) + P(\text{sites 1 to k bound})}. \tag{5}$$

As discussed in Section 2, the probability of a particular state is proportional to its weight, $e^{-\beta(E_{state} - n_{state}\mu)}$, where $\mu = \mu_0 + k_B T \ln \frac{c}{c_0}$ and n_{state} is the number of bound ligands. When all binding sites are identical,

$$E_{state} = \begin{cases} \varepsilon_A + n_{state}\ \varepsilon_b^{(A)} & \text{receptor is active} \\ \varepsilon_I + n_{state}\ \varepsilon_b^{(I)} & \text{receptor is inactive} \end{cases}. \tag{6}$$

The normalization constant Z is the sum of these weights for all of the states, which can be done analytically using the binomial series trick $(1+b)^n = \sum_{k=0}^{n} \binom{n}{k} b^k$,

$$\begin{aligned} Z &= e^{-\beta\varepsilon_A} \sum_{k=0}^{n} \binom{n}{k} \left(\frac{c}{c_0} e^{-\beta(\varepsilon_b^{(A)} - \mu_0)} \right)^k + e^{-\beta\varepsilon_I} \sum_{k=0}^{n} \binom{n}{k} \left(\frac{c}{c_0} e^{-\beta(\varepsilon_b^{(I)} - \mu_0)} \right)^k & (7) \\ &= e^{-\beta\varepsilon_A} \left(1 + \frac{c}{c_0} e^{-\beta(\varepsilon_b^{(A)} - \mu_0)} \right)^n + e^{-\beta\varepsilon_I} \left(1 + \frac{c}{c_0} e^{-\beta(\varepsilon_b^{(I)} - \mu_0)} \right)^n & (8) \end{aligned}$$

Thus the probability of sites 1 to k being bound is

$$P(\text{sites 1 to k bound}) = \frac{1}{Z} \left(e^{-\beta\varepsilon_A} \left(\frac{c}{c_0} e^{-\beta(\varepsilon_b^{(A)} - \mu_0)} \right)^k + e^{-\beta\varepsilon_I} \left(\frac{c}{c_0} e^{-\beta(\varepsilon_b^{(I)} - \mu_0)} \right)^k \right) \tag{9}$$

At this point, it is convenient to switch to thermodynamic notation, in which $K_A = c_0 e^{\beta(\varepsilon_b^{(A)} - \mu_0)}$, $K_I = c_0 e^{\beta(\varepsilon_b^{(I)} - \mu_0)}$, and $L = e^{-\beta(\varepsilon_I - \varepsilon_A)}$. This gives a normalization constant

$$Z = \left(1 + \frac{c}{K_A} \right)^n + L \left(1 + \frac{c}{K_I} \right)^n \tag{10}$$

and a probability of the state in which sites 1 to k are bound as

$$P(\text{sites 1 to k bound}) = \frac{1}{Z} \left(\left(\frac{c}{K_A} \right)^k + L \left(\frac{c}{K_I} \right)^k \right). \tag{11}$$

Therefore, we can calculate the conditional probability $P(k+1|k)$ and simplify, resulting in

$$\begin{aligned} P(k+1|k) &= \frac{P(\text{sites 1 to k+1 bound})}{P(\text{sites 1 to k+1 bound}) + P(\text{sites 1 to k bound})} & (12) \\ &= \frac{(\frac{c}{K_A})^{k+1} + L(\frac{c}{K_I})^{k+1}}{(\frac{c}{K_A})^{k+1} + L(\frac{c}{K_I})^{k+1} + (\frac{c}{K_A})^k + L(\frac{c}{K_I})^k} & (13) \\ &= \frac{c/K_A}{c/K_A + \frac{1 + L(K_A/K_I)^k}{1 + L(K_A/K_I)^{k+1}}}. & (14) \end{aligned}$$

Similarly, replacing k with $k-1$,

$$P(k|k-1) = \frac{c/K_A}{c/K_A + \frac{1+L(K_A/K_I)^{k-1}}{1+L(K_A/K_I)^k}}. \tag{15}$$

The ratio of the two is

$$\frac{P(k+1|k)}{P(k|k-1)} = \left(\frac{c}{K_A} + \frac{1+L(K_A/K_I)^{k-1}}{1+L(K_A/K_I)^k}\right) \Big/ \left(\frac{c}{K_A} + \frac{1+L(K_A/K_I)^{k}}{1+L(K_A/K_I)^{k+1}}\right). \tag{16}$$

If $K_A = K_I$, i.e. the inactive and active states of the receptor have similar affinity for the ligand, then this ratio is always 1, indicating no cooperativity. If $K_A \neq K_I$, then this ratio will be different from 1, regardless of k. If $K_A \neq K_I$, then one can show that

$$\frac{1+L(K_A/K_I)^{k-1}}{1+L(K_A/K_I)^k} > \frac{1+L(K_A/K_I)^{k}}{1+L(K_A/K_I)^{k+1}}, \tag{17}$$

implying that there is positive cooperativity.

1.2. Sigmoidal activity curves or binding curves

Cooperativity is related to the idea that the binding curve or activity curve are sigmoidal. [4] This graphical definition can only apply to plots in which the scale of concentrations is linear rather than logarithmic, since otherwise non-cooperative systems appear cooperative on log-log plots. (See Figure 1 of the main text.) A sigmoid curve $y(x)$ is often defined using a differential equation,

$$\frac{dy}{dx} = c_1 y\ (1 - c_2 y) \tag{18}$$

with a boundary condition, e.g. $y(0) = c_3$. The solution to this differential equation is

$$y(x) = \frac{1}{c_2 + \left(\frac{1}{c_3} - c_2\right) e^{-c_1 x}}, \tag{19}$$

which is the typical form of the sigmoidal function. In the context of cooperative binding curves, the importance of the sigmoid function is that its first derivative given in eqn. 18 increases with y until $y = \frac{1}{c_2}$ and then decreases for larger y. Equivalently, the sigmoid function's second derivative switches sign. The MWC activity curve given by

$$p_{active}(c) = \frac{\left(1+\frac{c}{K_A}\right)^n}{\left(1+\frac{c}{K_A}\right)^n + L\left(1+\frac{c}{K_I}\right)^n} \tag{20}$$

sometimes also has the property that the second derivative of $p_{active}(c)$ switches signs at some finite, non-negative concentration. The activity curve's first derivative (also known as the static gain G_0) is

$$\frac{dp_{active}}{dc} = \frac{n(K_I - K_A)}{K_A\ K_I} \left(\frac{\left(1+\frac{c}{K_A}\right)^{n-1}}{\left(1+\frac{c}{K_A}\right)^n + L\left(1+\frac{c}{K_I}\right)^n}\right) \left(\frac{L\left(1+\frac{c}{K_I}\right)^{n-1}}{\left(1+\frac{c}{K_A}\right)^n + L\left(1+\frac{c}{K_I}\right)^n}\right). \tag{21}$$

When $n = 1$, the static gain is monotonically decreasing with increasing $c \geq 0$. However, when $n > 1$, $G_0(c)$ increases and then decreases as c increases if $K_I > K_A$, or decreases and then increases as c increases if $K_A > K_I$. This is more easily observed if we rewrite eqn. 21 as

$$\frac{dp_{active}}{dc} = \frac{n(K_I - K_A)}{(c+K_A)(c+K_I)}\ p_{active}\left(1 - p_{active}\right). \tag{22}$$

Written in this way, the first derivative of the activity curve looks similar to the first derivative of the sigmoid function in eqn. 18 with $c_2 = 1$ and where c_1 is a function of x rather than a constant.

Some care must be taken in using graphical methods of inferring cooperativity. The plots of activity curves and binding curves in the main text are shown on a log-log scale, leading to activity curves that appear to be sigmoidal even when the system itself is non-cooperative. Conversely, a cooperative system might appear to have a non-sigmoidal activity curve or binding curve if it has a small transition point concentration.

1.3. Effective Hill coefficient

The last definition, which was presented in the main text, is perhaps the least intuitive. The effective Hill coefficient h_{eff} is mathematically defined as twice the slope of the "normalized" activity curve on a log-log plot at the transition point. The transition point is the concentration at which the activity is halfway between its minimum and maximum. [1] This definition leads to a rather complicated formula for h_{eff} as a function of MWC parameters, given in Figure 2 of the main text as a string of equations, but which we define here as

$$\frac{1}{2} = p_{active}^{norm}(c^*) = \frac{p_{active}(c^*) - p_{active}^{min}}{p_{active}^{max} - p_{active}^{min}} \tag{23}$$

and

$$h_{eff} = 2\frac{\partial \log p_{active}^{norm}}{\partial \log c}\Big|_{c=c^*} = 2c^*\frac{\partial \log p_{active}^{norm}}{\partial c}\Big|_{c=c^*} \tag{24}$$

This complicated definition is an analytical form for the Hill coefficient that will yield a good fit between the MWC activity curve and the Hill function. Recall that the Hill function has an activity curve given by

$$p_{active}^{Hill} = \frac{c^h}{c^h + K^h}. \tag{25}$$

In this equation, h is the Hill coefficient and K is a dissociation constant. If the Hill coefficient is positive, then a Hill coefficient $h > 1$ indicates positive cooperativity, a Hill coefficient of $h < 1$ indicates negative cooperativity, and a Hill coefficient of 1 indicates no cooperativity.

As stated earlier, the effective Hill coefficient provides a good fit between the activity curve of a MWC molecule and the Hill function. However, the Hill function always increases with ligand concentration c. If the MWC molecule activity curve decreases with ligand concentration c, then using eqn. 24 will yield a negative effective Hill coefficient, which amounts to switching the labels of the active and inactive states. Only the absolute value of the effective Hill coefficient is meaningful for determining whether or not there is positive cooperativity, negative cooperativity, or no cooperative interactions. Additionally, the activity curve of a MWC molecule does not necessarily have a minimum of 0 and a maximum of 1 as the ligand concentration varies from 0 to $\to \infty$. This is the reason for "normalizing" the MWC activity curve in eqn. 24.

As a sanity check, we show here that the effective Hill coefficient of the Hill function is the Hill coefficient, $h_{eff} = h$, and the transition point concentration is the dissociation constant, $c^* = K$. Unlike the MWC activity curve, p_{active}^{Hill} always has a minimum value of 0 at $c = 0$ and a maximum value of 1 as $c \to \infty$, and therefore, the "normalized" Hill function is the Hill function. The concentration at the transition point is given by

$$p_{active}^{Hill}(c^*) = \frac{(c^*)^h}{(c^*)^h + K^h} = \frac{1}{2} \to c^* = K. \tag{26}$$

From this simple formula for the concentration at the transition point, we can calculate an effective Hill

coefficient of

$$h_{eff}^{Hill} = 2\frac{\partial \log p_{active}^{Hill}}{\partial c}\Big|_{c=c^*} = 2\frac{c^*}{p_{active}^{Hill}(c^*)}\frac{\partial p_{active}^{Hill}}{\partial c}\Big|_{c=c^*} \quad (27)$$

$$= 2\frac{c^*}{p_{active}^{Hill}(c^*)}\left(\frac{h(c^*)^{h-1}}{(c^*)^h + K^h} - \frac{h(c^*)^{h-1}(c^*)^h}{((c^*)^h + K^h)^2}\right) \quad (28)$$

$$= 4h\left(\frac{(c^*)^h}{(c^*)^h + K^h}\right)\left(\frac{K^h}{(c^*)^h + K^h}\right) \quad (29)$$

$$= 4h(1/2)(1/2) = h. \quad (30)$$

In other words, the definition of the effective Hill coefficient and transition point concentration given in eqns. 24 and 23 are at least reasonable formulas in that when applied to the Hill function, they yield the Hill coefficient and the dissociation constant K.

It can be shown that the absolute value of the effective Hill coefficient of a MWC molecule with $K_A \neq K_I$ is always greater than 1, $|h_{eff}| > 1$, indicating positive cooperativity. Equality (and no cooperativity) holds only when $n = 1$ or $K_A = K_I$, which corresponds to the intuition that cooperativity requires multiple binding sites and states with different ligand binding affinities. No MWC parameters will cause the absolute value of the effective Hill coefficient to be less than 1, and this indicates that no MWC molecule without direct repulsive interactions between the ligands exhibits negative cooperativity.

2. Appendix 2: A short primer on information theory

In Section 4 of the main text, we introduced the definitions of mutual information and channel capacity as a general, principled way of quantifying how well an output tracks an input signal. In this appendix, we present a highly abridged sketch of some of the main ideas that lead to these definitions. Readers looking for comprehensive treatments of information theory should consult other textbooks or reviews. [5, 6, 7, 8, 9, 10, 11, 12, 13, 14] To quantify how much information is transmitted, we must define how much information is contained in the original signal.

A signal can be thought of as a message, i.e. a telephone message passed from one person to another. The telephone is called a "communication channel". These messages can be conceptualized in terms of probability distributions. Suppose someone has written three letters on a blank piece of paper. Without employing any contextual knowledge about the human language or what this person is doing, each of those letters has an equal probability of being any of the twenty-six letters in our alphabet. However, when we read the (smudged) word, we see that the first letter looks like an M or maybe an N, the second letter looks like a W, or perhaps a U or a V, and the third letter looks like a C, or maybe an O. After reading (i.e., observing a noisy signal) the probability distribution for each of these letters is quite different than before we read the word. That is, the probability of the first letter representing each of the letters in the alphabet is far from uniform; the first letter might have a probability $\frac{2}{3}$ of being an M and probability $\frac{1}{3}$ of being an N. The information gained by the act of reading has something to do with the difference between the probability distributions over these letters before and after reading.

Mathematically, the information contained in a probability distribution is exactly the negative of the entropy of a probability distribution. Readers may be familiar with the concept of entropy from thermodynamics as a proxy for the amount of disorder in a system of particles. Similarly, in the parlance of information theory, entropy can be thought of as the uncertainty of the value of a sample from a given probability distribution. For instance, if we toss a unweighted coin, it will have a half-half chance of landing on heads and tails. We are uncertain as to the outcome of the toss. However, if we toss a weighted coin, we can be fairly certain that it will land on (say) heads. The outcome of the toss of an unweighted coin has more uncertainty and more entropy than the toss of a weighted coin, and therefore the outcome of the toss of the unweighted coin conveys more information.

This intuition can be quantified as follows. Let X be a random variable that can take values $\{x_1, ..., x_n\}$ with probabilities $\{p(x_1), ..., p(x_n)\}$. We ask that any function $H(X)$ that purports to quantify entropy should satisfy some of our intuitions about what uncertainty means, namely that: uniform probability distributions should have more entropy than any other probability distribution; that two probability distributions that are "nearby" should have similar entropy; and that the entropy of a probability distribution should be the same no matter how we group the outcomes. More detailed description of these axioms can be found in many other texts. [6, 7] From these axioms alone, we can derive a function for the entropy up to a multiplicative constant as

$$H(X) = -\sum_x p(x) \log_2 p(x). \tag{31}$$

Interested readers are highly encouraged to consult one of many detailed texts on the subject, e.g. [6]. It is possible to multiply $H(X)$ in eqn. 31, and the new $H(X)$ will still satisfy the axioms listed above. Changing the multiplicative constant is equivalent to changing the base of the logarithm. Two standard choices are 2, so that entropy is in bits, and e, as in statistical mechanics. This definition of entropy is easily extended to definitions of "joint entropy" and "conditional entropy". If we have a joint probability distribution of two random variables, $p(x, y)$, its "joint entropy" is

$$H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y). \tag{32}$$

The "conditional entropy" $H(X|Y)$ is perhaps less intuitive than the entropy or the joint entropy; the conditional entropy $H(X|Y)$ is the expected number of bits required to describe X given knowledge of Y, namely,

$$H(X|Y) \quad = \quad \sum_y p(y) \left(-\sum_x p(x|y) \log_2 p(x|y) \right) = -\sum_{x,y} p(y)p(x|y) \log_2 p(x|y). \tag{33}$$

When X and Y vary indepdently ($p(x, y) = p(x)p(y)$) the conditional entropy is maximal, $H(X|Y) = H(X)$. On the other hand, if X and Y are tightly correlated, then $H(X|Y)$ is small.

With the definition of the entropy of a random variable X in hand, we can define the mutual information between a random variable X and Y as the information gained about X by measuring Y. Specifically, this implies

$$I(X; Y) = H(X) - H(X|Y). \tag{34}$$

The mutual information can also be thought of as the amount of shared information between random variables X and Y. If X and Y are independent random variables, then they share no information, and accordingly $I(X; Y) = 0$. If Y tracks X perfectly, then the conditional entropy $H(X|Y) = 0$, giving a maximal mutual information $I(X; Y) = H(X)$. The channel capacity is the maximum value[1] of $I(X; Y)$ attainable by varying $p(x)$, i.e.

$$I_{opt}(X; Y) = \max_{p(x)} I(X; Y). \tag{35}$$

According to the noisy channel coding theorem, the channel capacity is also the maximum possible rate of error-free information transmission. [6, 7] In this sense, the channel capacity is a single-number characterization of the information transmission limits of the system. [6, 7] For instance, in Section 4 of the main text, the channel capacity $I_{opt}(c; N_{open})$ is the maximal amount of shared information between

[1] The channel capacity is more correctly defined using a supremum rather than a maximum, $I_{opt}(X; Y) = \sup_{p(x)} I(X; Y)$. [6, 7]

the ligand concentration c and the number of open ion channels N_{open}. The probability distribution $p^*(c)$ that maximizes the mutual information $I(c; N_{open})$ has noise that is "well-matched" to the noise in the communication channel, $p(N_{open}|c)$.

We are interested in rewriting eqn. 34 in terms of the probability distributions $p(x)$ and $p(y|x)$, since we often have access to the conditional probability distribution $p(y|x)$. Using the definition of the conditional distribution, $p(x|y)p(y) = p(x, y)$, the conditional entropy takes the form

$$H(X|Y) = -\sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(y)}, \tag{36}$$

and similarly we can use the definition of the marginal distribution, $p(x) = \sum_y p(x, y)$, to rewrite the entropy of X in eqn. 32 as

$$H(X) = -\sum_x (\sum_y p(x, y)) \log_2 (\sum_y p(x, y)) = -\sum_{x,y} p(x, y) \log_2 (\sum_y p(x, y)). \tag{37}$$

These two expressions enable us to rewrite the mutual information $I(X; Y)$ given in eqn. 34 in terms of the probability distribution $p(x, y)$,

$$I(X; Y) = -\sum_{x,y} p(x, y) \log_2 p(x) + \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(y)} = \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}. \tag{38}$$

If X and Y are continuous random variables, then the sums are easily changed to integrals and $p(x, y)$ becomes a probability density rather than a list of probabilities, e.g.

$$I(X; Y) = \int \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dxdy. \tag{39}$$

The conditional probability distribution $p(y|x)$ is related to the joint probability distribution $p(x, y)$ using $p(x, y) = p(x)p(y|x)$; substituting this expression into the above expression for mutual information $I(X; Y)$ gives

$$I(X; Y) = \int \int p(x)p(y|x) \log_2 \frac{p(y|x)}{p(y)} dxdy. \tag{40}$$

There are many other equivalent expressions for the mutual information, e.g. shown in Theorem 2.4.1 of [6].

3. Appendix 3: Mutual Information and Channel Capacity for MWC Molecules

In Section 4.1 of the main text, we presented a sketch of the derivation of the channel capacity of a number of N independent MWC molecules, where the input is ligand concentration and the output is the number of active MWC molecules. Here, we present that derivation in more mathematical detail for interested readers. A pedagogical derivation with even more mathematical detail is presented in Section 6.4 of Bialek's excellent new biophysics textbook. [14]

If the joint probability distribution of c and N_{open} is $p(c, N_{open})$, then following the procedure outlined in eqn. 38, the mutual information between the input ligand concentration and the output number of open channels is defined as

$$I(c; N_{open}) = \int \sum_{N_{open}=0}^{N} p(c, N_{open}) \log_2 \frac{p(c, N_{open})}{p(c)p(N_{open})} dc. \tag{41}$$

Using the definition of the joint probability as $p(c, N_{open}) = p(N_{open}|c)p(c)$ we can rewrite the mutual information in eqn. 41 as

$$I(c; N_{open}) = \int \sum_{N_{open}=0}^{N} p(N_{open}|c)p(c) \log_2 \frac{p(N_{open}|c)p(c)}{p(c)p(N_{open})} dc. \tag{42}$$

By exploiting the rule $log(A/B) = log(A) - log(B)$, we can now rewrite this as

$$\begin{aligned} I(c; N_{open}) &= \int p(c) \left(\sum_{N_{open}=0}^{N} p(N_{open}|c) \log_2 p(N_{open}|c) \right) dc \\ &\quad - \int p(c) \left(\sum_{N_{open}=0}^{N} p(N_{open}|c) \log_2 p(N_{open}) \right) dc \end{aligned} \tag{43}$$

The N independent MWC molecules can be conceptualized as N independent binary random variables, as each of these receptors is open with a probability $p_{open}(c)$. Therefore, the conditional probability distribution is a binomial distribution,

$$p(N_{open}|c) = \binom{N}{N_{open}} (p_{open}(c))^{N_{open}} (1 - p_{open}(c))^{N - N_{open}}. \tag{44}$$

We could now calculate the channel capacity numerically by substituting eqn. 44 above into the expression for mutual information in eqn. 43, and then finding an optimal $p^*(c)$ using a numerical approximation scheme. However, when N is large, we can invoke a "small-noise approximation" and approximate the discrete binomial distribution in eqn. 44 by a continuous function. This allows us to derive an analytic expression for the channel capacity. Recall from the text that the mean of this binomial distribution is

$$\bar{N}_{open}(c) = N p_{open}(c) \tag{45}$$

and its variance is

$$\sigma^2_{N_{open}}(c) = N p_{open}(c)(1 - p_{open}(c)). \tag{46}$$

When N is very large, the conditional probability distribution $p(N_{open}|c)$ in eqn. 44 is highly peaked about its mean and looks Gaussian. We wish to rewrite the binomial distribution for $p(N_{open}|c)$ in a more convenient form as a continuous normal distribution, and this normal approximation should have the same mean and variance as the binomial distribution. The approximating normal distribution is therefore

$$p(N_{open}|c) \simeq \frac{1}{\sqrt{2\pi\sigma^2_{N_{open}}}} e^{-(N_{open} - \bar{N}_{open})^2 / 2\sigma^2_{N_{open}}}, \tag{47}$$

where N_{open} runs from $-\infty$ to ∞. As long as $\sigma_{N_{open}} << \bar{N}_{open}$, since then there is vanishingly small probability of N_{open} attaining a negative value or a value greater than N. This is precisely what it means to make the "small noise" approximation. This small-noise approximation is often used in other contexts to simplify formulae, e.g. polymer physics. [3] To make the dependence upon key variables such as the concentration of ligand clear, we rewrite this in the more cumbersome but equivalent form using eqns. 45 and 46 as

$$p(N_{open}|c) = \frac{\exp(-(N_{open} - N p_{open}(c))^2 / 2N p_{open}(c)(1 - p_{open}(c)))}{\sqrt{2\pi N p_{open}(c)(1 - p_{open}(c))}}. \tag{48}$$

With that approximation in hand, we can now simplify the expression for the mutual information in equation 43. Specifically, the first parenthetical in eqn. 43 can be approximated as

$$\begin{aligned}\sum_{N_{open}=0}^{N} p(N_{open}|c)\log_2 p(N_{open}|c) &\approx \int_{-\infty}^{\infty} \frac{e^{-(N_{open}-\bar{N}_{open})^2/2\sigma^2_{N_{open}}}}{\sqrt{2\pi\sigma^2_{N_{open}}}} \log_2 \frac{e^{-(N_{open}-\bar{N}_{open})^2/2\sigma^2_{N_{open}}}}{\sqrt{2\pi\sigma^2_{N_{open}}}} dN_{open} \\ &= -\frac{1}{2}\log_2(2\pi e\sigma^2_{N_{open}}(c)), \end{aligned} \tag{49}$$

where we have used standard rules about Gaussian integrals. [6] This result leaves us poised for substitution into the first term of eqn. 43. To evaluate the second parenthetical term in equation 43, we have

$$\sum_{N_{open}=0}^{N} p(N_{open}|c)\log_2 p(N_{open}) \approx \int_{-\infty}^{\infty} p(N_{open}|c)\log_2 p(N_{open}) dN_{open}. \tag{50}$$

We can approximate this integral using a Taylor expansion of the integrand, expanding $\log_2 p(N_{open})$ about $\bar{N}_{open}$,

$$\begin{aligned}\log_2 p(N_{open}) &= \log_2 p(\bar{N}_{open}) + \frac{\partial \log_2 p(N_{open})}{\partial N_{open}}\Big|_{\bar{N}_{open}} (N_{open}-\bar{N}_{open}) \\ &\quad + \frac{1}{2}\frac{\partial^2 \log_2 p(N_{open})}{\partial N^2_{open}}\Big|_{\bar{N}_{open}} (N_{open}-\bar{N}_{open})^2 \end{aligned} \tag{51}$$

We then integrate this Taylor expansion according to eqn. 50,

$$\begin{aligned}\int_{-\infty}^{\infty} p(N_{open}|c)\log_2 p(N_{open}) dN_{open} &\approx \int_{-\infty}^{\infty} p(N_{open}|c)\ \Big(\log_2(p(\bar{N}_{open})) \\ &\quad + \frac{\partial \log_2 p(N_{open})}{\partial N_{open}}\Big|_{\bar{N}_{open}} (N_{open}-\bar{N}_{open}) \\ &\quad + \frac{1}{2}\frac{\partial^2 \log_2 p(N_{open})}{\partial N^2_{open}}\Big|_{\bar{N}_{open}} (N_{open}-\bar{N}_{open})^2\Big)\ dN_{open}. \end{aligned} \tag{52}$$

The coefficients in this Taylor expansion are independent of N_{open}, and as such, we can evaluate the right hand side of eqn. 52. The zeroth order term in the Taylor expansion for $\log_2 p(N_{open})$ integrates to

$$\int_{-\infty}^{\infty} p(N_{open}|c)\log_2 p(\bar{N}_{open}) dN_{open} = \log_2 p(\bar{N}_{open}) \int_{-\infty}^{\infty} p(N_{open}|c) dN_{open} = \log_2 p(\bar{N}_{open}) \tag{53}$$

using the condition that probability distributions are normalized. The first order term in the Taylor expansion for $\log_2 p(N_{open})$ integrates to

$$\begin{aligned}\int_{-\infty}^{\infty} p(N_{open}|c)\ \frac{\partial \log_2 p(N_{open})}{\partial N_{open}}\Big|_{\bar{N}_{open}} (N_{open}-\bar{N}_{open}) dN_{open} &= \frac{\partial \log_2 p(N_{open})}{\partial N_{open}}\Big|_{\bar{N}_{open}} \times \\ &\quad \int_{-\infty}^{\infty} (N_{open}-\bar{N}_{open}) p(N_{open}|c) dN_{open} \\ &= 0, \end{aligned} \tag{54}$$

since the mean $\bar{N}_{open}$ is defined by $\int_{-\infty}^{\infty}(N_{open}-\bar{N}_{open})p(N_{open}|c)dN_{open} = 0$. (See eqn. 47.) Finally, the

second order term in the Taylor expansion evaluates to

$$\begin{aligned}\frac{1}{2}\int_{-\infty}^{\infty} p(N_{open}|c)\ \frac{\partial^2 \log_2 p(N_{open})}{\partial N_{open}^2}\Big|_{\bar{N}_{open}}(N_{open}-\bar{N}_{open})dN_{open} &= \frac{1}{2}\frac{\partial^2 \log_2 p(N_{open})}{\partial N_{open}^2}\Big|_{\bar{N}_{open}} \times \\ &\quad \int_{-\infty}^{\infty}(N_{open}-\bar{N}_{open})^2 p(N_{open}|c)dN_{open} \\ &= \frac{1}{2}\frac{\partial^2 \log_2 p(N_{open})}{\partial N_{open}^2}\Big|_{\bar{N}_{open}}\sigma^2_{N_{open}}, \qquad (55)\end{aligned}$$

using the definition of variance, $\int_{-\infty}^{\infty}(N_{open}-\bar{N}_{open})^2 p(N_{open}|c)dN_{open} = \sigma_{N_{open}}(\bar{N}_{open})^2$. Therefore, eqn. 52 simplifies to

$$\int_{-\infty}^{\infty} p(N_{open}|c)\log_2 p(N_{open})dN_{open} \approx \log_2 p(\bar{N}_{open}) + \frac{1}{2}\frac{\partial^2 \log_2 p(N_{open})}{\partial N_{open}^2}\Big|_{\bar{N}_{open}}\sigma^2_{N_{open}}. \qquad (56)$$

using eqns. 53-55. Hence, using the simplifications achieved in equations 49 and 56, we are now ready to compute the mutual information itself by plugging back into equation 43 resulting in

$$I(c;N_{open}) \approx -\int p(c)\left(\log_2\sqrt{2\pi e\sigma^2_{N_{open}}(c)} + \log_2 p(\bar{N}_{open})\right)dc. \qquad (57)$$

We have exploited a second consequence of the small noise approximation by ignoring the term proportional to $\sigma^2_{N_{open}}$ in eqn. 56; although $\sigma^2_{N_{open}} \propto N$ using eqn. 46, its coefficient $\frac{\partial^2 \log_2 p(N_{open})}{\partial N_{open}^2}\Big|_{\bar{N}_{open}} \propto \frac{1}{N^2}$. Together, these proportionality relationships suggest that $\frac{\partial^2 \log_2 p(N_{open})}{\partial N_{open}^2}\Big|_{\bar{N}_{open}}\sigma^2_{N_{open}} \propto \frac{1}{N}$, indicating that this second order term is dwarfed by $\log_2\sigma^2_{N_{open}} \propto \log_2 N$ in the large N, small noise limit. Finally, since $\bar{N}_{open}$ is a deterministic function of c, we can relate their probability distributions exactly using a change of variables formula,

$$p(c)dc = p(\bar{N}_{open})d\bar{N}_{open}. \qquad (58)$$

This change of variables formula can be loosely thought of as a statement about the conservation of the probability. Eqn. 58 permits us to rewrite the integral above as

$$I(c;N_{open}) \simeq -\int p(\bar{N}_{open})\left(\log_2\sqrt{2\pi e\sigma^2_{N_{open}}(c)} + \log_2 p(\bar{N}_{open})\right)d\bar{N}_{open}. \qquad (59)$$

Even with the small noise approximation, a determination of the mutual information still requires detailed knowledge of the empirical distribution of ligand concentration or (equivalently) a detailed knowledge of the mean number of open channels, $p(\bar{N}_{open})$. However, we now employ a second major assumption, which is that the biological system chooses a distribution $p(c)$ and thus a distribution $p(\bar{N}_{open})$ that maximizes the mutual information. Alternatively, we are calculating the channel capacity $I_{opt}(c;N_{open})$ of the ensemble of receptors, which is its mutual information at maximal information transmission.

The optimal $p(\bar{N}_{open})$ is found using variational calculus, which is loosely speaking the calculus of functions of functions. [15, 16] The basic idea is that the mutual information $I(c;N_{open})$ is a function of the probability distribution $p(\bar{N}_{open})$ according to eqn. 59. To make this dependency clear, we can write $I[p(\bar{N}_{open})]$. Our goal is to find a probability distribution $p^*(\bar{N}_{open})$ such that a slightly different probability distribution, $p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})$, is guaranteed to have a lower mutual information,

$$I[p^*(\bar{N}_{open})] \geq I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})] \qquad (60)$$

for $|\varepsilon| << 1$ and all "allowable" functions $h(\bar{N}_{open})$. An allowable function $h(\bar{N}_{open})$ is any function such that both $p^*(\bar{N}_{open})$ and $p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})$ are normalized probability distributions, i.e.

$$1 = \int p^*(\bar{N}_{open})d\bar{N}_{open} \qquad (61)$$

and

$$1 = \int \left(p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})\right) d\bar{N}_{open}. \tag{62}$$

Subtracting these two equations gives us a restriction on $h(\bar{N}_{open})$, namely,

$$0 = \int h(\bar{N}_{open}) d\bar{N}_{open}. \tag{63}$$

When trying to find a local maximum of a function of a scalar variable in calculus, we typically take its derivative and set it to zero. To connect to this intuition, we rewrite eqn. 60 in terms of a derivative of the mutual information,

$$\lim_{\varepsilon \to 0} \frac{\partial}{\partial \varepsilon} \left(I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})] - I[p^*(\bar{N}_{open})] \right) = 0, \tag{64}$$

for any function $h(\bar{N}_{open})$ that satisfies the normalization condition in eqn. 63. A function that satisfies eqn. 64 will not necessarily satisfy eqn. 60, but a function that satisfies eqn. 60 will satisfy eqn. 64. Along this line of reasoning, we try to find an expression for $I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})] - I[p^*(\bar{N}_{open})]$ in the limit of infinitesimally small ε. We start by tackling the term $I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})]$,

$$\begin{aligned} I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})] &= -\int \left(p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})\right) \Big(\log_2 \sqrt{2\pi e \sigma^2_{N_{open}}(\bar{N}_{open})} \\ &\quad + \log_2 \left(p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})\right) \Big) d\bar{N}_{open}, \end{aligned} \tag{65}$$

from eqn. 59. For small ε, using the rule $\log(A+B) = \log\left(A \times (1 + \frac{B}{A})\right) = \log A + \log(1 + \frac{B}{A})$, and applied to eqn. 65, this yields

$$\log_2 \left(p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})\right) = \log_2 p^*(\bar{N}_{open}) + \log_2 \left(1 + \varepsilon \frac{h(\bar{N}_{open})}{p^*(\bar{N}_{open})}\right). \tag{66}$$

Using the Taylor approximation $\log(1+x) \simeq x$ for small x,

$$\log_2 \left(p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})\right) = \log_2 p^*(\bar{N}_{open}) + \frac{\varepsilon h(\bar{N}_{open})}{p^*(\bar{N}_{open}) \log 2} + O(\varepsilon^2). \tag{67}$$

Substituting this expression into eqn. 65 and expanding to first order in ε yields

$$\begin{aligned} I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})] &= -\int \left(p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})\right) \Big(\log_2 \sqrt{2\pi e \sigma^2_{N_{open}}(\bar{N}_{open})} \\ &\quad + \log_2 p^*(\bar{N}_{open}) + \varepsilon \frac{h(\bar{N}_{open})}{p^*(\bar{N}_{open}) \log 2} \Big) d\bar{N}_{open} + O(\varepsilon) \qquad (68) \\ &= -\int p^*(\bar{N}_{open}) \left(\log_2 \sqrt{2\pi e \sigma^2_{N_{open}}(\bar{N}_{open})} + \log_2 p^*(\bar{N}_{open}) \right) d\bar{N}_{open} \\ &\quad -\varepsilon \int h(\bar{N}_{open}) \Big(\log_2 \sqrt{2\pi e \sigma^2_{N_{open}}(\bar{N}_{open})} + \frac{1}{\log 2} \\ &\quad + \log_2 p^*(\bar{N}_{open}) \Big) d\bar{N}_{open} + O(\varepsilon^2). \qquad (69) \end{aligned}$$

We fortuitously recognize the first term in this expression as

$$I[p^*(\bar{N}_{open})] = -\int p^*(\bar{N}_{open}) \left(\log_2 \sqrt{2\pi e \sigma^2_{N_{open}}(\bar{N}_{open})} + \log_2 p^*(\bar{N}_{open}) \right) d\bar{N}_{open}, \tag{70}$$

and substituting this into eqn 69 gives us

$$I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})] - I[p^*(\bar{N}_{open})] = -\varepsilon \int h(\bar{N}_{open}) \Big(\log_2 \sqrt{2\pi e \sigma_{N_{open}}(\bar{N}_{open})^2} + \frac{1}{\log 2} + \log_2 p^*(\bar{N}_{open}) \Big) d\bar{N}_{open} + O(\varepsilon^2). \quad (71)$$

Based on eqn. 64, we wish to find $p^*(\bar{N}_{open})$ such that the expression

$$\lim_{\varepsilon \to 0} \big(I[p^*(\bar{N}_{open}) + \varepsilon h(\bar{N}_{open})] - I[p^*(\bar{N}_{open})] \big) = -\int h(\bar{N}_{open}) \Big(\log_2 \sqrt{2\pi e \sigma_{N_{open}}(\bar{N}_{open})^2} + \frac{1}{\log 2} + \log_2 p^*(\bar{N}_{open}) \Big) d\bar{N}_{open} \quad (72)$$

is 0 for any $h(\bar{N}_{open})$ that satisfies the normalization condition in eqn. 63. This will happen if the integrand $\log_2 \sqrt{2\pi e \sigma_{N_{open}}(\bar{N}_{open})^2} + \frac{1}{\log 2} + \log_2 p^*(\bar{N}_{open})$ is a constant for all $\bar{N}_{open}$, due to the normalization condition on $h(\bar{N}_{open})$ in eqn. 63. Therefore,

$$p^*(\bar{N}_{open}) \propto \frac{1}{\sigma_{N_{open}}(\bar{N}_{open})}. \quad (73)$$

With appropriate normalization of $p^*(\bar{N}_{open})$ given in eqn. 61, the form of $p(\bar{N}_{open})$ that maximizes the mutual information is of the form

$$p^*(\bar{N}_{open}) = \frac{1}{Z} \frac{1}{\sigma_{N_{open}}(\bar{N}_{open})} \quad (74)$$

where Z is a normalization constant,

$$Z = \int_{\bar{N}_{open}^{min}}^{\bar{N}_{open}^{max}} \frac{d\bar{N}_{open}}{\sigma_{N_{open}}(\bar{N}_{open})}. \quad (75)$$

The variance $\sigma_{N_{open}} = \sqrt{N p_{open}(1 - p_{open})}$ depends on $\bar{N}_{open}$ in that the $\bar{N}_{open} = N p_{open}$, giving

$$\sigma_{N_{open}}(\bar{N}_{open}) = \sqrt{\bar{N}_{open} \left(1 - \frac{\bar{N}_{open}}{N} \right)}. \quad (76)$$

Finally, using the optimal probability distribution of ligand concentration in eqn. 74 in place of $p(\bar{N}_{open})$ in eqn. 59, we find that the channel capacity is

$$I_{opt}(c; N_{open}) = \max_{p(c)} I(c; N_{open}) \quad (77)$$

$$= -\int p^*(\bar{N}_{open}) \left(\log_2 \left(\sqrt{2\pi e} \sigma_{N_{open}} \right) + \log_2 \frac{1}{Z} \frac{1}{\sigma_{N_{open}}} \right) d\bar{N}_{open}. \quad (78)$$

Using the familiar rules about logarithm products, $\log(AB) = \log A + \log B$, this simplifies to

$$I_{opt}(c; N_{open}) = -\int p^*(\bar{N}_{open}) \log_2 \sqrt{\frac{2\pi e}{Z}} d\bar{N}_{open}. \quad (79)$$

Finally, since $p^*(\bar{N}_{open})$ is a normalized probability distribution such that $\int p^*(\bar{N}_{open}) d\bar{N}_{open} = 1$, and since $\log_2 \sqrt{\frac{2\pi e}{Z}}$ is a constant,

$$I_{opt}(c; N_{open}) = \log_2 \frac{Z}{\sqrt{2\pi e}}. \quad (80)$$

Now all that remains is to actually calculate $I_{opt}(c; N_{open})$ by calculating Z in eqn. 75 using the variance in eqn. 46. Using a change of variables $\bar{N}_{open} = Np_{open}$, we find that Z is the product of a term that depends on the number of receptors N and the parameters that govern the binding curve of a single receptor,

$$Z = \int_{Np_{open}^{min}}^{Np_{open}^{max}} \frac{d(Np_{open})}{\sqrt{Np_{open}(1-p_{open})}} = \int_{p_{open}^{min}}^{p_{open}^{max}} \frac{N dp_{open}}{\sqrt{N}\sqrt{p_{open}(1-p_{open})}} \tag{81}$$

$$= \sqrt{N} \int_{p_{open}^{min}}^{p_{open}^{max}} \frac{dp_{open}}{\sqrt{p_{open}(1-p_{open})}}. \tag{82}$$

Here p_{open}^{min} and p_{open}^{max} are exactly the probabilities of a single receptor being in the active state where there is no ligand and when there is a saturating amount of ligand, with formulas given in Figure 2 and Section 4 of the main text. The integral in eqn. 82 can be evaluated using trigonometric substitution. Explicitly, the indefinite integral $\int \frac{dp}{\sqrt{p(1-p)}}$ can be evaluated by setting another variable, $p = \sin^2 u$, giving

$$\int \frac{dp}{\sqrt{p(1-p)}} = \int \frac{d(\sin^2 u)}{\sqrt{\sin^2 u \ (1-\sin^2 u)}} = \int \frac{2 \sin u \cos u \ du}{\sin u \cos u} = 2u = 2 \sin^{-1} \sqrt{p}. \tag{83}$$

Therefore, eqn. 82 is

$$Z = 2\sqrt{N} \left(\sin^{-1} \sqrt{p_{open}^{max}} - \sin^{-1} \sqrt{p_{open}^{min}} \right). \tag{84}$$

This yields the equation for channel capacity that appeared in Section 4 of the main text, namely,

$$I_{opt} = \log_2 \frac{Z}{\sqrt{2\pi e}} = \log_2 \left(\sin^{-1} \sqrt{p_{open}^{max}} - \sin^{-1} \sqrt{p_{open}^{min}} \right) + \log_2 \sqrt{\frac{2N}{\pi e}}. \tag{85}$$

Increasing p_{open}^{max} or decreasing p_{open}^{min} will increase both dynamic range ($p_{open}^{max} - p_{open}^{min}$) and channel capacity I_{opt} in eqn. 85 above.

4. Appendix 4: Constructing transition matrices for a dynamical MWC model

In Section 5.1 of the main text, we presented a transition matrix for a dynamical model of a two-site MWC molecule, but we only derived its first row. For completeness, we derive here the elements of the other rows in this transition matrix. We then give a transition matrix for the n-site receptor.

4.1. Transition matrix for the two-site MWC molecule

As described in Section 5.1 of the main text, the dynamical model of a two-site MWC molecule describes the time evolution of the state vector

$$x = \begin{pmatrix} [O_2] \\ [O_1] \\ [O_0] \\ [C_0] \\ [C_1] \\ [C_2] \end{pmatrix} \tag{86}$$

by the first-order ordinary differential equation

$$\frac{dx}{dt} = Mx, \tag{87}$$

where M is a transition matrix. Here we use a different state vector than that used in Section 5.1, although they only differ by a normalization constant $[R]$, the total concentration of receptor. The elements of M can be determined by listing the elementary chemical reactions that lead to a change in the state vector, writing down the corresponding rate equations, and comparing those rate equations to eqn. 87. Non-elementary reactions, i.e. reactions that can be broken into elementary chemical reactions, are implicitly described by eqn. 87 since they can be described in terms of two or more elementary reactions.

In Section 5.1 of the main text, we derived the first row of M by considering the reactions $O_1 + L \leftarrow O_2$ and $O_2 \rightarrow O_1 + L$, written succinctly as $O_1 + L \leftrightharpoons O_2$, as the two elementary chemical reactions that affect the concentration of O_2. We reproduce that here for completeness. Using the kinetic rates defined in Section 5.1 of the main text, by the Law of Mass Action,

$$\frac{d[O_2]}{dt} = f_O[L][O_1] - 2b_O[O_2]. \tag{88}$$

The first row of eqn. 87 reads

$$\frac{d[O_2]}{dt} = M_{11}[O_2] + M_{12}[O_1] + M_{13}[O_0] + M_{14}[C_0] + M_{15}[C_1] + M_{16}[C_2]. \tag{89}$$

Comparing eqn. 88 and eqn. 89 yields

$$M_{11} = -2b_O,\ M_{12} = f_O c,\ M_{13} = M_{14} = M_{15} = M_{16} = 0. \tag{90}$$

Determining the second row of the matrix M proceeds similarly. The four reactions that we consider are $O_0 + L \leftrightharpoons O_1$ and $O_1 + L \leftrightharpoons O_2$, giving

$$\frac{d[O_1]}{dt} = 2f_O[L][O_0] - b_O[O_1] - f_O[L][O_1] + 2b_O[O_2], \tag{91}$$

where the factors of 2 denote that there are two sites that a ligand can bind to or unbind from on the receptor. The second row of eqn. 87 reads

$$\frac{d[O_1]}{dt} = M_{21}[O_2] + M_{22}[O_1] + M_{23}[O_0] + M_{24}[C_0] + M_{25}[C_1] + M_{26}[C_2]. \tag{92}$$

Comparing eqns. 91 and 92 yields

$$M_{21} = 2b_O,\ M_{22} = -f_O c - b_O,\ M_{23} = 2f_O c,\ M_{24} = M_{25} = M_{26} = 0. \tag{93}$$

There are four reactions to consider for O_0: $O_0 \leftrightharpoons C_0$ and $O_0 + L \leftrightharpoons O_1$. In the main text, we denoted the forward rate kinetic rate of the latter reaction as f_T and the backwards kinetic rate as b_T, giving

$$\frac{d[O_0]}{dt} = -2f_O[L][O_0] + b_O[O_1] - f_T[O_0] + b_T[C_0]. \tag{94}$$

The third row of eqn. 87 reads

$$\frac{d[O_0]}{dt} = M_{31}[O_2] + M_{32}[O_1] + M_{33}[O_0] + M_{34}[C_0] + M_{35}[C_1] + M_{36}[C_2]. \tag{95}$$

Comparing eqns. 94 and 95 yields

$$M_{32} = b_O,\ M_{33} = -2f_O c - f_T,\ M_{34} = b_T,\ M_{31} = M_{35} = M_{36} = 0. \tag{96}$$

There are four reactions that affect the concentration of C_0, $O_0 \leftrightharpoons C_0$ and $C_0 + L \leftrightharpoons C_1$, giving

$$\frac{d[C_0]}{dt} = -2f_C[L][C_0] + b_C[C_1] - b_T[C_0] + f_T[O_0]. \tag{97}$$

The third row of eqn. 87 reads

$$\frac{d[C_0]}{dt} = M_{41}[O_2] + M_{42}[O_1] + M_{43}[O_0] + M_{44}[C_0] + M_{45}[C_1] + M_{46}[C_2]. \tag{98}$$

Comparing eqns. 97 and 98 yields

$$M_{43} = f_T,\ M_{44} = -2f_C c - b_T,\ M_{45} = b_C,\ M_{41} = M_{42} = M_{46} = 0. \tag{99}$$

The four reactions that we account for as affecting $[C_1]$ are $C_0 + L \leftrightharpoons C_1$ and $C_1 + L \leftrightharpoons C_2$, giving

$$\frac{d[C_1]}{dt} = 2f_C[L][C_0] - b_C[C_1] - f_C[L][C_1] + 2b_C[C_2]. \tag{100}$$

The second row of eqn. 87 reads

$$\frac{d[C_1]}{dt} = M_{51}[O_2] + M_{52}[O_1] + M_{53}[O_0] + M_{54}[C_0] + M_{55}[C_1] + M_{56}[C_2]. \tag{101}$$

Comparing eqns. 100 and 101 yields

$$M_{54} = 2b_C,\ M_{55} = -f_C c - b_C,\ M_{56} = 2f_C c,\ M_{51} = M_{52} = M_{53} = 0. \tag{102}$$

Finally, considering the reaction $C_1 + L \leftrightharpoons C_2$ leads to

$$\frac{d[C_2]}{dt} = f_C[L][C_1] - 2b_C[C_2]. \tag{103}$$

The sixth row of eqn. 87 reads

$$\frac{d[C_2]}{dt} = M_{61}[O_2] + M_{62}[O_1] + M_{63}[O_0] + M_{64}[C_0] + M_{65}[C_1] + M_{66}[C_2]. \tag{104}$$

Comparing eqns. 103 and 104 yields

$$M_{65} = f_C c,\ M_{66} = 2b_C,\ M_{61} = M_{62} = M_{63} = M_{64} = 0. \tag{105}$$

4.2. Transition matrix for the n-site MWC molecule

Consider a MWC molecule with n identical binding sites, and denote the receptor's two states as either O (open/active) or C (closed/inactive). Each ligand-receptor can be denoted as O_k or C_k, where k gives the number of bound ligands, and O and C denote the state of the receptor. A dynamical MWC model describes the time evolution of a vector x,

$$x = \begin{pmatrix} [O_n] \\ \vdots \\ [O_0] \\ [C_0] \\ \vdots \\ [C_n] \end{pmatrix}, \tag{106}$$

using a first-order differential equation,

$$\frac{dx}{dt} = Mx. \tag{107}$$

The elements of the $2(n+1) \times 2(n+1)$ matrix M can be determined in the same way that the elements of M were determined for the two-site receptor. For the n-site MWC molecule we consider the following elementary reactions:

- $O_k + L \leftrightharpoons O_{k+1}$, with forward kinetic rate f_O and backward kinetic rate b_O related by $K_o = \frac{b_O}{f_O}$,
- $O_0 \leftrightharpoons C_0$, with forward kinetic rate f_T and backward kinetic rate b_T related by $K = \frac{f_T}{b_T}$, and
- $C_k + L \leftrightharpoons C_{k+1}$, with forward kinetic rate f_C and backward kinetic rate b_C related by $K_c = \frac{b_C}{f_C}$.

Here we list the corresponding rate equations without proof as

$$\frac{d[O_k]}{dt} = \begin{cases} -kb_O[O_k] + (n-k+1)f_O[L][O_{k-1}] - (n-k)f_O[L][O_k] + (k+1)b_O[O_{k+1}] & 1 \le k \le n-1 \\ -f_O[L][O_0] + nb_O[O_1] - f_T[O_0] + b_T[C_0] & k = 0 \\ f_O[L][O_{n-1}] - nb_O[O_n] & k = n \end{cases} \tag{108}$$

and

$$\frac{d[C_k]}{dt} = \begin{cases} -kb_C[C_k] + (n-k+1)f_C[L][C_{k-1}] - (n-k)f_C[L][C_k] + (k+1)b_C[C_{k+1}] & 1 \le k \le n-1 \\ -f_C[L][C_0] + nb_C[C_1] - b_T[C_0] + f_T[O_0] & k = 0 \\ f_C[L][C_{n-1}] - nb_C[C_n] & k = n \end{cases} . \tag{109}$$

We can explicitly the elements of M in terms of the function $\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ as

$$M_{ij} = \begin{cases} f_O c\ \delta_{i,j-1} - nb_O\ \delta_{i,j} & i = 1 \\ -((n-i+1)b_O + (i-1)f_O c)\ \delta_{i,j} + if_O c\ \delta_{i,j-1} + (n-i+2)b_O\ \delta_{i,j-1} & i = 2, ..., n \\ -(f_O c + f_T)\,\delta_{i,j} + b_T\ \delta_{i,j+1} + nb_O\delta_{i,j-1} & i = n+1 \\ -(f_C c + b_T)\,\delta_{i,j} + f_T\ \delta_{i,j-1} + nb_C c\ \delta_{i,j+1} & i = n+2 \\ -((i-n-2)b_C + (2n+2-i)f_C c)\ \delta_{i,j} + (2n+3-i)f_C c\ \delta_{i,j-1} + (i-n-1)b_C\ \delta_{i,j+1} & i = n+3, ..., 2n+1 \\ f_C c\ \delta_{i,j-1} - nb_C\ \delta_{i,j} & i = 2n+2 \end{cases} . \tag{110}$$

5. Appendix 5: Response functions of MWC molecules

In Section 5.2 of the main text, we presented a sketch of the derivation of the frequency response function of an MWC molecule to changing ligand concentration. Here, we present that derivation for the general n-site MWC molecule and in more mathematical detail for interested readers. Many of the manipulations presented here are standard manipulations found in textbooks, e.g. [17].

Consider a MWC molecule with n identical binding sites, and denote the receptor's two states as either O (open/active) or C (closed/inactive). Each ligand-receptor can be denoted as O_k or C_k, where k gives the number of bound ligands, and O and C denote the state of the receptor. A dynamical MWC model describes the time evolution of a vector x,

$$x = \begin{pmatrix} P(O_n, t) \\ \vdots \\ P(O_0, t) \\ P(C_0, t) \\ \vdots \\ P(C_n, t) \end{pmatrix} . \tag{111}$$

Recall from Section 5.1 of the main text that the list of probabilities of each of the ligand-receptor configurations for an MWC molecule evolves in time as

$$\frac{dx}{dt} = M(c)x, \tag{112}$$

where $M(c)$ is a transition matrix that depends on the ligand concentration c. For examples of transition matrices, see Appendix 4 above. Recall also from Section 5.1 of the main text that the probability of the receptor being in the open state is related to x using the projection matrix P_R using

$$p_{open} = P_R\ x. \tag{113}$$

We wish to relate $p_{open}(t)$ to changes in ligand concentration, $c(t)$. As described in Section 5.1 of the main text, we will find the relationship using the frequency response function

$$G(\omega) = \frac{\mathcal{F}(p_{open})}{\mathcal{F}(c)}, \tag{114}$$

where $\mathcal{F}(p_{open})$ is the Fourier transform of p_{open},

$$\mathcal{F}(p_{open}) = \int_{-\infty}^{\infty} p_{open}(t)\ e^{-i\omega t}\ dt, \tag{115}$$

and $\mathcal{F}(c)$ is the Fourier transform of c,

$$\mathcal{F}(c) = \int_{-\infty}^{\infty} c(t)\ e^{-i\omega t}\ dt. \tag{116}$$

Some write these formulas for the Fourier transform with an additional factor of $\frac{1}{2\pi}$ or $\frac{1}{\sqrt{2\pi}}$, [17, 15] but when we take the ratio of two Fourier transforms to find $G(\omega)$, these multiplicative factors cancel. Calculating the frequency response function $G(\omega)$ implicitly requires that a sinusoidal oscillation $c(t) = A_0 + A_1 \sin \omega t$ will result in a sinusoidal oscillation of $p_{open}(t) = B_0 + B_1 \sin(\omega t + \phi)$. This assumption does not hold in the case of the MWC molecule responding to changes in ligand concentration; according to eqn. 112, changes in ligand concentration are coupled in a nonlinear way to changes in the state vector x. However, it holds approximately when the variations in ligand concentration are very small. We will work in the limit that variations in ligand concentration are very small, which will allow us to "linearize" the system. [17] Explicitly , if c_0 is the time-averaged ligand concentration, then we assume that $|\Delta c(t)| << c_0$ where

$$c(t) = c_0 + \Delta c(t). \tag{117}$$

Similarly, we linearize all other variables. The state vector $x(t)$ is perturbed slightly from its equilibrium state $x_{eq}(c_0) = x_0$,

$$x(t) = x_0 + \Delta x(t), \tag{118}$$

where x_0 is defined by $M(c_0)x_0 = 0$. The probability of the receptor existing in the open state is also perturbed slightly from equilibrium $p_{open}(c_0) = P_R x_0 = p_0$,

$$p_{open}(t) = p_0 + \Delta p_{open}(t). \tag{119}$$

Finally, the transition matrix M given in Appendix 4 is linear in $c(t)$, but even if it were not, we could Taylor approximate $M(c)$ about $c = c_0$ using

$$M(c) \simeq M(c_0) + \frac{\partial M}{\partial c}\Big|_{c_0} (c - c_0) = M_0 + M_1 \Delta c(t). \tag{120}$$

Substituting eqns. 117-119 into eqns. 112 and 113 gives us a linearized system for which we can calculate the frequency response of Δp_{open} to Δc. Noting that

$$\frac{dx}{dt} = \frac{d}{dt}(x_0 + \Delta x) = \frac{d\Delta x}{dt}, \tag{121}$$

eqn. 112 becomes

$$\frac{d\Delta x}{dt} = (M_0 + M_1\Delta c)(x_0 + \Delta x) \simeq M_0\Delta x + M_1\Delta c, \tag{122}$$

using $M_0 x_0 = 0$ and ignoring the second-order, nonlinear term $\Delta c\Delta x$. Noting that $p_0 = P_R x_0$, eqn. 113 becomes

$$\Delta p_{open} = P_R\Delta x. \tag{123}$$

Solving eqns. 122 and 123 is a much simpler proposition than solving eqns. 112 and 113, precisely because the system is now linear. The solution to a more general version of eqns. 122 and 123 is actually a standard formula in various textbooks on dynamical systems, e.g. [17]. A general approach to solving such equations is to apply the Fourier transform (or Laplace or z-transform) to each of them. The Fourier transform of eqn. 122 is

$$\mathcal{F}\left(\frac{d\Delta x}{dt}\right) = \mathcal{F}(M_0\Delta x) + \mathcal{F}(M_1\Delta c). \tag{124}$$

In the Fourier domain, time derivatives become multiplication by $i\omega$ according to the product rule:

$$\mathcal{F}\left(\frac{d\Delta x}{dt}\right) \quad = \quad \int_{-\infty}^{\infty} \frac{d\Delta x}{dt}\, e^{-i\omega t} dt = \Delta x\, e^{-i\omega t}\Big|_{-\infty}^{\infty} + i\omega \int_{-\infty}^{\infty} \Delta x\, e^{-i\omega t} dt. \tag{125}$$

We assume that as $|t| \to \infty$, $\Delta x \to 0$ and recognize the second term on the right hand side of this equation as $i\omega\mathcal{F}(\Delta x)$, leading to

$$\mathcal{F}\left(\frac{d\Delta x}{dt}\right) \quad = \quad i\omega\mathcal{F}(\Delta x). \tag{126}$$

Additionally, $\mathcal{F}(ax) = a\mathcal{F}(x)$ if a is independent of time; therefore, as long as the kinetic rates in matrices M_0 and M_1 are time-independent,

$$\mathcal{F}(M_0\Delta x) = M_0\mathcal{F}(\Delta x) \tag{127}$$

and

$$\mathcal{F}(M_1\Delta c) = M_1\mathcal{F}(\Delta c). \tag{128}$$

Eqns. 126-128 simplify eqn. 124 to

$$i\omega\mathcal{F}(\Delta x) = M_0\mathcal{F}(\Delta x) + M_1\mathcal{F}(\Delta c). \tag{129}$$

Moving both $\mathcal{F}(\Delta x)$ terms to the left-hand side gives

$$(i\omega - M_0)\mathcal{F}(\Delta x) = M_1 x_0\mathcal{F}(\Delta c) \tag{130}$$

This allows us to solve for the Fourier transform of the changes in the state vector by multiplying the left and right hand side by the inverse of $i\omega - M_0$.

$$\mathcal{F}(\Delta x) = (i\omega - M_0)^{-1} M_1 x_0\mathcal{F}(\Delta c) \tag{131}$$

From this, we can find the Fourier transform of Δp_{open} as

$$\mathcal{F}(\Delta p_{open}) = P_R\mathcal{F}(\Delta x) = P_R(i\omega - M_0)^{-1} M_1 x_0\mathcal{F}(\Delta c), \tag{132}$$

and from this, the frequency response function of this ligand-gated ion channel,

$$G(\omega) = \frac{\mathcal{F}(\Delta p_{open})}{\mathcal{F}(\Delta c)} = P_R(i\omega - M_0)^{-1} M_1 x_0. \tag{133}$$

Earlier, we blithely took the inverse of $i\omega - M_0$. This inverse exists as long as the determinant of $i\omega - M_0$ is nonzero, or equivalently, as long as the product of the eigenvalues of $i\omega - M_0$ is nonzero. [18] Luckily, the eigenvectors of M_0 are also eigenvectors of $i\omega - M_0$, which makes it easy to calculate the conditions under which $i\omega - M_0$ does not have an inverse. Let $u_1, ..., u_n$ be the eigenvectors of M_0 with corresponding (real) eigenvalues $\lambda_1, ..., \lambda_n$. The absolute values of these eigenvalues are the inverse of the internal time constants of the receptor's dynamics, and one of these eigenvalues is guaranteed to be 0 with eigenvector x_0. Note that

$$(i\omega - M_0)u_k = i\omega u_k - M_0 u_k = i\omega u_k - \lambda_k u_k = (i\omega - \lambda_k)u_k. \tag{134}$$

Hence by definition, u_k is an eigenvector of $i\omega - M_0$ with eigenvalue $i\omega - \lambda_k$. The only way the product of these eigenvalues can be zero is if one of the eigenvalues is 0 itself, which happens only when $\omega = 0$. If we then write the diagonalized form of $M_0 = PDP^{-1}$ where $D = \text{diag}(\lambda_k)$, $i\omega - M_0 = P\ \text{diag}(i\omega - \lambda_k)\ P^{-1}$, and $(i\omega - M_0)^{-1} = P^{-1}\ \text{diag}(\frac{1}{i\omega - \lambda_k})\ P$, we find that the frequency response is

$$\begin{aligned} G(\omega) &= \frac{\mathcal{F}(\Delta p_{open})}{\mathcal{F}(\Delta c)} = P_R P^{-1} \text{diag}(\frac{1}{i\omega + \lambda_k}) M_1 x_0 & (135) \\ &= \sum_k a_k(c_0) G_k(\omega), & (136) \end{aligned}$$

where

$$G_k(\omega) = \frac{1}{i\omega + \lambda_k} = \frac{1}{i\omega + \omega_{cutoff,k}}, \tag{137}$$

and $a_k(c_0)$ are linear weighting coefficients that are frequency-independent. These filters in Eqn. 137 are low-pass first-order frequency filters with cutoff frequencies at the inverse time constant of the MWC molecule's internal dynamics,

$$|G_k(\omega)| = \frac{1}{\sqrt{\omega^2 + \omega_{cutoff,k}^2}}, \tag{138}$$

as given in Section 5.1 of the main text.

References

1. T. L. Hill. *Cooperativity theory in biochemistry : steady-state and equilibrium systems.* Springer-Verlag, New York, 1985.

2. A. Ben-Naim. *Cooperativity and Regulation in Biochemical Processes.* Kluwer Academic/Plenum Publishers, New York, 2001.

3. R. Phillips, J. Kondev, J. Theriot, and H. G. Garcia. *Physical biology of the cell, 2nd Edition.* Garland Science, New York, 2013. (Illustrated by N. Orme).

4. J. Monod, J. Wyman, and J. P. Changeux. On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol*, 12:88–118, 1965.

5. C. E. Shannon. A mathematical theory of computation. *The Bell System Technical Journal*, 27, 1948.

6. T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, 2006.

7. David J. C. Mackay. *Information Theory, Inference, and Leaning Algorithms*. Cambridge University Press, 2003.

8. D. J. Depew and B. H. Weber. *Entropy, Information, and Evolution: New Perspectives on Physical and Biological Evolution*. Bradford Books/MIT Press, 1988.

9. F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA, 1997.

10. G. Tkačik and A. M. Walczak. Information transmission in genetic regulatory networks: a review. *J. Phys.: Condens. Matter*, 23:153102, 2011.

11. C. Adami. Information theory in molecular biology. *Physics of Life Reviews.*, 1:3–22, 2004.

12. A. Rhee, R. Cheong, and A. Levchenko. The application of information theory to biochemical signaling systems. *Phys. Biol.*, 9:045011–045021, 2012.

13. A. Borst and F. E. Theunissen. Information theory and neural coding. *Nature Neurosci.*, 2:947–957, 1999.

14. W. Bialek. *Biophysics: Searching for Principles*. Princeton University Press, 2012.

15. M. L. Boas. *Mathematical Concepts in the Physical Sciences*. John Wiley and Sons Inc., New York, NY, 2005.

16. Hans Sagan. *Introduction to the Calculus of Variations*. Dover Publications, Inc., 1992.

17. K. J. Astrom and R. M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ, 2008.

18. Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 2 edition, 2012.