bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A general model of multivalent binding with ligands of heterotypic subunits and multiple surface receptors

Zhixin Cyrillus Tan [1] and Aaron S. Meyer *[1,2,3,4]

[1] *Bioinformatics Interdepartmental Program, University of California, Los Angeles*
[2] *Department of Bioengineering, University of California, Los Angeles*
[3] *Jonsson Comprehensive Cancer Center, University of California, Los Angeles*
[4] *Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles*

Abstract

Multivalent cell surface receptor binding is a ubiquitous biological phenomenon with functional and therapeutic significance. Predicting the amount of ligand binding for a cell remains an important question in computational biology as it can provide great insight into cell-to-cell communication and rational drug design toward specific targets. In this study, we extend a mechanistic, two-step multivalent binding model to account for multiple ligands and receptors, optionally allowing heterogeneous complexes. We derive the macroscopic predictions for both specifically arranged and randomly assorted complexes, and demonstrate how this model enables large-scale predictions on mixture binding and the binding space of a ligand. This model provides an elegant and computationally efficient framework for analyzing multivalent binding.

Keywords General Binding Model, Multivalent Binding, Combinatorics

*Department of Bioengineering, UCLA, Los Angeles, CA 90095; ameyer@asmlab.org

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 Introduction

Binding to extracellular ligands is among the most fundamental and universal activities of a cell. Many important biological activities, and cell-to-cell communication in particular, are based on recognizing extracellular molecules via specific surface receptors. For example, multivalent ligands are common extracellular factors in the immune system [8], and many computational models have been applied to study IgE-FcεRI [5], MHC-T cell receptor [7], and IgG-FcγR interaction [16].

In this study, we extend a simple two-step, multivalent binding model to cases involving multiple receptors and ligand subunits [1, 2, 3, 12, 7]. By harnessing the power of combinatorics via applying the multinomial theorem and focusing on macrostates, we can predict the amount of binding for each ligand and receptor at the equilibrium state. Our model provides both generality and computational efficiency, allowing large-scale predictions such as characterizing synergism of using a mixture of ligands and depicting the binding space of a compound. The compactness and elegance of the formulae enable both analytical and numerical analyses. We expect this binding model will be widely applicable to many biological contexts.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2 Preliminaries

2.1 Vector and matrix notation

In this work, we denote a vector in boldface letter and its entry in the same letter but with subscript and not in boldface, e.g. $\mathbf{C} = [C_1, C_2, \ldots, C_n]$. The sum of elements for a vector is denoted as $|\mathbf{C}| = \sum_{i=1}^{n} C_i$.

For any matrix (A_{ij}) of size $m \times n$, we denote the vector formed by its i-th row as $\mathbf{A_{i\bullet}} = [A_{i1}, A_{i2}, \cdots, A_{in}]$, and the vector formed by its j-th column as $\mathbf{A_{\bullet j}} = [A_{1j}, A_{2j}, \cdots, A_{mj}]$. The row sums of matrix (A_{ij}), therefore, can be written as $|\mathbf{A_{1\bullet}}|$, $|\mathbf{A_{2\bullet}}|$, $\cdots$, $|\mathbf{A_{m\bullet}}|$, and column sums $|\mathbf{A_{\bullet 1}}|$, $|\mathbf{A_{\bullet 2}}|$, $\cdots$, $|\mathbf{A_{\bullet n}}|$.

In this work, multinomial coefficients such as n choose $k_1, k_2, \cdots, k_n$ will be written as

$$\binom{n}{\mathbf{k}} = \binom{n}{k_1 \quad k_2 \quad \cdots \quad k_n} = \frac{n!}{k_1! k_2! \cdots k_n!}.$$

The implicit assumption here is that $|\mathbf{k}| = n$, and each $k_i \in \mathbb{N}$.

2.2 Some useful theorems in combinatorics

From the binomial theorem, we know that

$$\sum_{i=0}^{f} \binom{f}{i} \Phi^i = (1+\Phi)^f.$$

Differentiating both sides by Φ, we get

$$\sum_{i=0}^{f} i \binom{f}{i} \Phi^i = f\Phi(1+\Phi)^{f-1}. \tag{1}$$

We can derive similar property from the multinomial theorem. Assume the elements of a nonnegative integer vector $\mathbf{q}$ add up to f, or $|\mathbf{q}| = f$. Given another nonnegative vector $\boldsymbol{\varphi}$ with sum of elements $|\boldsymbol{\varphi}|$, we have

$$\sum_{|\mathbf{q}|=f} \binom{f}{\mathbf{q}} \prod_i {\varphi_i}^{q_i} = |\boldsymbol{\varphi}|^f.$$

Differentiate both sides by φ_m where φ_m can be any entry of $\boldsymbol{\varphi}$, and rearrange, we have

$$\sum_{|\mathbf{q}|=f} \binom{f}{\mathbf{q}} q_m \prod_i {\varphi_i}^{q_i} = \varphi_m f |\boldsymbol{\varphi}|^{f-1}. \tag{2}$$

We can multiply two different multinomial theorem equations together, too. Let $\mathbf{u}$ and $\mathbf{v}$ are two nonnegative integer vectors, $\mathbf{a}$ and $\mathbf{b}$ are two nonnegative vectors, and $|\mathbf{u}| = m$, $|\mathbf{v}| = n$, we have

$$\sum_{\substack{|\mathbf{u}|=m \\ |\mathbf{v}|=n}} \binom{m}{\mathbf{u}} \binom{n}{\mathbf{v}} \prod_i a_i^{u_i} \prod_j b_j^{v_j} = |\mathbf{a}|^m |\mathbf{b}|^n.$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Throughout this paper, we consolidate multiple summation symbols into one. In this case, we use $\sum_{|\mathbf{u}|=m,|\mathbf{v}|=n}$ as a shorthand for $\sum_{|\mathbf{u}|=m}\sum_{|\mathbf{v}|=n}$. From Eq.(2), we can derive the sum of a linear combination of two exponents from each multinomial term as

$$\sum_{\substack{|\mathbf{u}|=m\\|\mathbf{v}|=n}} \binom{m}{\mathbf{u}}\binom{n}{\mathbf{v}} \underbrace{(k_1u_p+k_2v_q)}_{\text{linear combination}} \prod_i a_i^{u_i}\prod_j b_j^{v_j} = k_1a_pm|\mathbf{a}|^{m-1}|\mathbf{b}|^n + k_2|\mathbf{a}|^m b_q n|\mathbf{b}|^{n-1}$$
$$= \left[\frac{k_1ma_p}{|\mathbf{a}|}+\frac{k_2nb_q}{|\mathbf{b}|}\right]|\mathbf{a}|^m|\mathbf{b}|^n,$$

where k_1 and k_2 are constants.

We can extend this to the product of N multinomial equations. Let $\mathbf{q_1},\cdots,\mathbf{q_N}$ be N nonnegative integer vectors, each with $|\mathbf{q_i}|=\theta_i$, and $\boldsymbol{\psi_1},\cdots,\boldsymbol{\psi_N}$ be N nonnegative vectors. Then, the sum of any linear combination of exponent terms $\sum_r k_rq_{s_rt_r}$, where k_r's are constants and each $q_{s_rt_r}$ is the t_r-th element of $\mathbf{q_{s_r}}$, can be calculated as

$$\sum_{\substack{|\mathbf{q_1}|=\theta_1\\ \cdots\\ |\mathbf{q_N}|=\theta_N}} \left(\sum_r k_rq_{s_rt_r}\right)\prod_{i=1}^N\binom{\theta_i}{\mathbf{q_i}}\left(\prod_j \psi_{ij}^{q_{ij}}\right) = \left[\sum_r \frac{k_r\theta_{s_r}\psi_{s_rt_r}}{|\boldsymbol{\psi_{s_r}}|}\right]\prod_{i=1}^N|\boldsymbol{\psi_i}|^{\theta_i}. \quad (3)$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3 Model setup

3.1 Parameters and notations

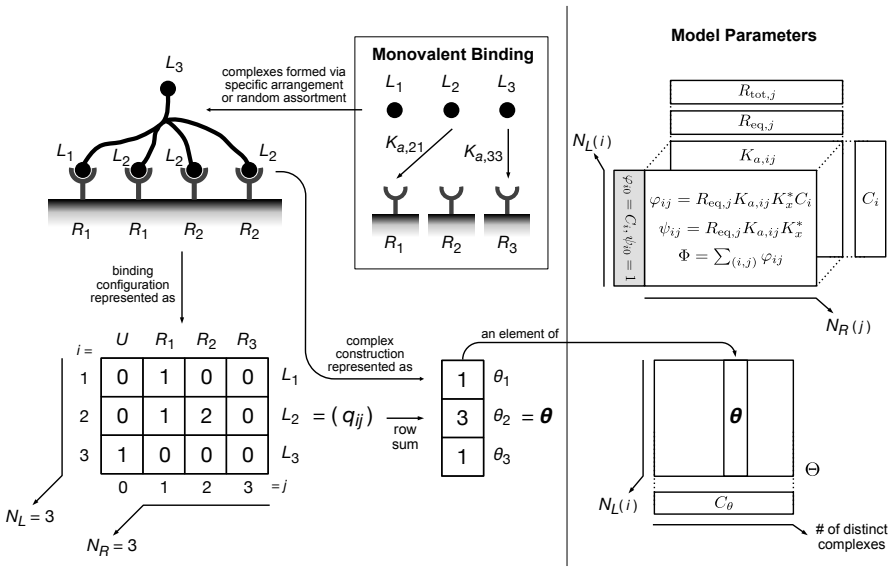

Figure 1: General setup of the model. In this study, we investigate the binding behavior of complexes formed by monomer ligands in either specific arrangement or random assortment. We propose that the binding configuration between a complex and several receptors on a cell can be described as a matrix (q_{ij}). The construction of a complex can be written as a vector $\boldsymbol{\theta}$. The figure shows the dimensions of the model's parameters: C_i, the monomer compositions, are in a vector of N_L; $R_{\mathrm{tot},j}$ and $R_{\mathrm{eq},j}$, the receptor expression and equilibrium level are in vectors of N_R; the binding affinities, $K_{a,ij}$, are in a matrix of $N_L \times N_R$; φ_{ij} and ψ_{ij} are in the matrices of $N_L \times (N_R + 1)$. Θ is a set of all possible $\boldsymbol{\theta}$'s, with $C_{\boldsymbol{\theta}}$ as their compositions. Each $\boldsymbol{\theta}$ is a vector of N_L, and $C_{\boldsymbol{\theta}}$ should be in a vector of the same size as Θ.

In this study, we investigate the binding between multivalent ligand complexes and a cell expressing various surface receptors. As shown in Figure 1, we consider N_L types of distinct monomer ligands, namely L_1, L_2, ..., L_{N_L}, and N_R types of distinct receptors expressed on a cell, namely R_1, R_2, ..., R_{N_R}. The monovalent binding association constant between L_i and R_j is defined as $K_{a,ij}$. A ligand complex consists of one or several monomer ligands, and each of them can bind to a receptor independently. Its construction can be described by a vector $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_{N_L}]$, where each entry θ_i represents how many L_i this complex contains. The sum of elements of vector $\boldsymbol{\theta}$, $|\boldsymbol{\theta}|$, is f, the valency of this complex.

The binding configuration at equilibrium between an individual complex and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a cell expressing various receptors can be described as a matrix (q_{ij}) with N_L rows and $(N_R + 1)$ columns. For example, the complex bound as shown on the top left corner in Figure 1 can be described as the matrix below it. Entry q_{ij} represents the number of L_i to R_j binding, and q_{i0}, the entry on the 0-th column, is the number of unbound L_i on that complex in this configuration. This matrix can be unrolled into a vector form $\mathbf{q} = [q_{10}, q_{11}, ..., q_{1N_R}, q_{20}, ..., q_{2N_R}, q_{30}, ..., q_{N_L N_R}]$ of length $N_L(N_R + 1)$. Note that this binding configuration matrix (q_{ij}) only records how many L_i-to-R_j pairs are formed, regardless of which exact ligand on the complex binds. For example, in Figure 1, swapping the two L_2's binding to R_2's will give us the same configuration matrix. Therefore, we will need to account for this combinatorial factor when applying the law of mass action.

We know from the conservation of mass that for this complex, $\theta_i = q_{i0} + q_{i1} + q_{i2} + \cdots + q_{iN_R} = |\mathbf{q_{i\bullet}}|$ must hold for all i. Mathematically, vector $\boldsymbol{\theta}$ is the row sums of matrix (q_{ij}). The corresponding $\boldsymbol{\theta}$ of a binding configuration $\mathbf{q}$, $\boldsymbol{\theta}(\mathbf{q})$ which is written in the format of a function, can be determined by this relationship. Also, the sum of elements in $\mathbf{q}$, $|\mathbf{q}| = f$, the valency.

The concentration of complexes in the solution is L_0 (not to be confused with L_i, the name of ligands, when $i = 1, 2, \cdots, N_L$). The number of ligand complexes in the solution is usually much greater than that of the receptors and so it is a common practice to assume binding does not deplete the ligand concentration.

On the receptor side, $R_{\text{tot},i}$ is the total number of R_i expressed on the cell surface. This usually can be measured experimentally. $R_{\text{eq},i}$ is the number of unbound R_i on a cell at the equilibrium state during the ligand complex-receptor interaction, and it needs to be calculated from $R_{\text{tot},i}$ as we will explain later.

The binding of a ligand complex, a large molecule, is complicated. To simplify the matter, we will need to make some key thermodynamic assumptions. In this model, we make two assumptions on the binding dynamics:

1. The initial binding between a free (unbound) complex and a surface receptor R_j has the same affinity (association constant, $K_{a,ij}$) as the monomer ligand L_i;

2. In order for the detailed balance to hold, the affinity constant of any subsequent binding event on the surface of a cell after the initial interaction must be proportional to their corresponding monovalent affinity. We assume the subsequent binding affinity in multivalent interactions between L_i and R_j to be $K_x^* K_{a,ij}$.

K_x^* is a term coined as the crosslinking constant. It captures the difference between free and multivalent ligand-receptor binding, including but not limited to steric effects and local receptor clustering [4]. In practice this term is often fit to apply this model to a specific biological context.

We create two more variables that will help to simplify our equations throughout this work. For all i in $\{1, 2, ..., N_L\}$, we define $\psi_{ij} = R_{\text{eq},j} K_{a,ij} K_x^*$ and $\varphi_{ij} = R_{\text{eq},j} K_{a,ij} K_x^* C_i$ where $j = \{1, 2, ..., N_R\}$, and we define $\psi_{i0} = 1$, $\varphi_{i0} = C_i$. Therefore, $\varphi_{ij} = \psi_{ij} C_i$ holds for all i and j. Then we define the sum of this new

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

matrix (φ_{ij}) as $\sum_{i=1}^{N_L}\sum_{j=1}^{N_R}\varphi_{ij}=\Phi$, and $\sum_{i=1}^{N_L}\sum_{j=0}^{N_R}\varphi_{ij}=\Phi+\sum_{i=1}^{N_L}C_i=1+\Phi$. The rationale of these definitions will become clear in future sections.

3.2 The amount of a specific binding configuration

Now we will derive the amount of complexes bound with the configuration described as $\mathbf{q}$ on a cell at equilibrium, $v_{\mathbf{q}}$.

Within the definitions of our model, we know that the composition of any complex can be described by a vector $\boldsymbol{\theta}$ of length N_L, where each entry θ_i represents the number of monomer L_i this complex consists of. We can enumerate all possible binding configuration of $\boldsymbol{\theta}$ complex by filling the matrix (q_{ij}) with any nonnegative integer values so long as its row sums equal $\boldsymbol{\theta}$. Conversely, we can imply the complex composition given any binding configuration $\mathbf{q}$ by finding its row sums, $\boldsymbol{\theta}(\mathbf{q})$. For a certain configuration $\mathbf{q}$, its $\boldsymbol{\theta}(\mathbf{q})$ is determined and has concentration $L_0C_{\boldsymbol{\theta}(\mathbf{q})}$. If the corresponding complex $\boldsymbol{\theta}(\mathbf{q})$ does not exist in the solution, $C_{\boldsymbol{\theta}(\mathbf{q})}=0$. Since we assumed that binding will not deplete the ambient concentration of any $\boldsymbol{\theta}(\mathbf{q})$, it will remain $L_0C_{\boldsymbol{\theta}(\mathbf{q})}$ at equilibrium.

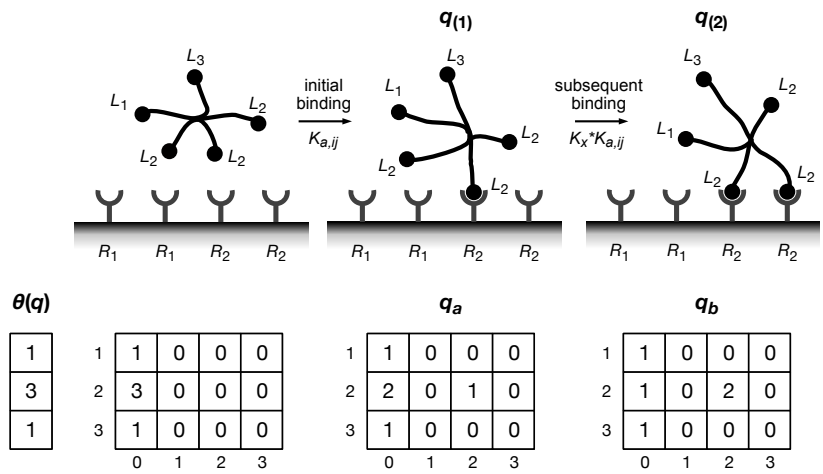

Figure 2: A scheme of cell-complex binding step by step. We assume the initial binding event has the same affinity as monomer binding, $K_{a,ij}$, while subsequent binding has an association constant scaled by K_x^*, the crosslinking constant. Each binding configuration scheme above can be described by the $\mathbf{q}$ right below, if we ignore the statistical factors. $\boldsymbol{\theta}(\mathbf{q})$ is the structure of the complex and can be derived from $\mathbf{q}$.

Initial binding We start with the initial binding reaction of a complex, L_i-to-R_j. As shown in Figure 2, the reactants of this reaction are the free complexes and the free receptors R_j (in this case R_2), and the product are L_i-to-R_j (in this

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

case L_2-R_2) monovalently bound complexes $\mathbf{q_{(1)}}$. We denote the concentration of this new complex as $v_{\mathbf{q_{(1)}}}$. The concentration of free complexes is $L_0 C_{\boldsymbol{\theta}(\mathbf{q_{(1)}})}$. By the assumption of the model, the equilibrium constant for the reaction is $K_{a,ij}$. Therefore, we have

$$v_{\mathbf{q_{(1)}}} = L_0 C_{\boldsymbol{\theta}(\mathbf{q_{(1)}})} R_{\mathrm{eq},j} K_{a,ij}.$$

While the binding configuration of $\mathbf{q_{(1)}}$ can be described by $\mathbf{q_a}$, the total amount of complexes that bind as described as $\mathbf{q_a}$ may not be the same as $v_{\mathbf{q_{(1)}}}$, since $\mathbf{q_a}$ does not consider the number of ways this binding L_i can be chosen. An equivalent explanation is that, $\mathbf{q_{(1)}}$ is only one possible microstate to achieve the $\mathbf{q_a}$ configuration, and we need to count how many microstates are possible for $\mathbf{q_a}$. Accounting for this statistical factor, we have

$$v_{\mathbf{q_a}} = v_{\mathbf{q_{(1)}}} \binom{\theta_i}{1} = v_{\mathbf{q_{(1)}}} \binom{\theta_i}{\mathbf{q_{a,i\bullet}}},$$

$$v_{\mathbf{q_a}} = L_0 C_{\boldsymbol{\theta}(\mathbf{q_a})} R_{\mathrm{eq},j} K_{a,ij} \binom{\theta_i}{\mathbf{q_{a,i\bullet}}},$$

since $\boldsymbol{\theta}(\mathbf{q_{(1)}}) = \boldsymbol{\theta}(\mathbf{q_a})$. $\mathbf{q_{a,i\bullet}}$ is a vector formed by the i-th row of $\mathbf{q_a}$. For example, in Figure 2, $\mathbf{q_{a,2\bullet}} = [2, 0, 1, 0]$. Conceptually, $\binom{\theta_i}{\mathbf{q_{a,i\bullet}}}$ can be understood as the number of ways to split θ_i L_i's into q_{i0} of unbound units, q_{i1} of R_1-bound, q_{i2} of R_2-bound, ..., and q_{iN_R} of R_{N_R}-bound. Here, only q_{i0} and q_{ij} will be nonzero, with $q_{i0} = \theta_i - 1$ and $q_{ij} = 1$, so it is effectively the same as $\binom{\theta_i}{1}$. However, the multinomial coefficient expression can be generalized into more complicated cases.

Subsequent binding For a subsequent binding between L_i and R_j (i and j are not necessarily the same as in initial binding), we have the reactants as a bound complex, $\mathbf{q_{(1)}}$, and a free receptor R_j (in the case shown by Figure 2, R_2), while the product is another bound complex, $\mathbf{q_{(2)}}$. The equilibrium constant is $K_x^* K_{a,ij}$, then

$$v_{\mathbf{q_{(2)}}} = v_{\mathbf{q_{(1)}}} R_{\mathrm{eq},j} K_x^* K_{a,ij}.$$

To account for the statistical factors for $v_{\mathbf{q_b}}$, we have $v_{\mathbf{q_b}} = v_{\mathbf{q_{(2)}}} \binom{\theta_i}{\mathbf{q_{b,i\bullet}}}$. For example, in Figure 2, $\mathbf{q_{b,2\bullet}} = [1, 0, 2, 0]$. Putting these together, we have

$$v_{\mathbf{q_b}} = v_{\mathbf{q_a}} R_{\mathrm{eq},j} K_x^* K_{a,ij} \frac{\binom{\theta_i}{\mathbf{q_{b,i\bullet}}}}{\binom{\theta_i}{\mathbf{q_{a,i\bullet}}}}.$$

By recursion, we can solve $v_{\mathbf{q}}$ for any $\mathbf{q}$ from these equations. It is

$$v_{\mathbf{q}} = \frac{L_0 C_{\boldsymbol{\theta}(\mathbf{q})}}{K_x^*} \prod_{(i,j)=(1,1)}^{(N_L,N_R)} (R_{\mathrm{eq},j} K_x^* K_{a,ij})^{q_{ij}} \prod_{i=1}^{N_L} \binom{\theta_i}{\mathbf{q_{i\bullet}}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$= \frac{L_0 C_{\boldsymbol{\theta}(\mathbf{q})}}{K_x^*} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} \psi_{ij}^{q_{ij}} \prod_{i=1}^{N_L} \binom{\theta_i}{\mathbf{q_{i\bullet}}}$$

if we define $\psi_{ij} = R_{\mathrm{eq},j} K_x^* K_{a,ij}$ for $j = 1, 2, \cdots, N_R$ and $\psi_{i0} = 1$ for all i. $\prod_{(i,j)=(1,0)}^{(N_L,N_R)}$ is a shorthand for $\prod_{i=1}^{N_L} \prod_{j=0}^{N_R}$. In the next section, we will use this formula repeatedly.

Notice that this equation is not suitable for calculating the concentration of unbound $\mathbf{q}$, when every nonzero values are on its 0-th column. The concentration of unbound ligands should always be $L_0 C_{\boldsymbol{\theta}(\mathbf{q})}$. However, for algebraic convenience, we allow such definition and will name it $v_{0,\mathrm{eq}}$ which equals $L_0 C_{\boldsymbol{\theta}(\mathbf{q})}/K_x^*$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4 Macroscopic equilibrium predictions

From here we will investigate the macroscopic properties of binding, such as the total amount of ligand bound and receptor bound on a cell surface at equilibrium. We consider two different ways complexes in the solution to be formed. First, complexes can be formed in a specific arrangement. In this case, the structure and exact concentration for each kind of complex are designed and known. Alternatively, we can set a fixed valency f for all complexes given the known proportion of each ligand monomer. Through random assortment, any combination of f monomer ligands can form a complex, and their concentration will follow a multinomial distribution. We will explore these two cases separately.

4.1 Complexes formed in a specific arrangement

When complexes are specifically arranged, the structure and proportion of each kind are well-defined. To formulate this mathematically, we assume that we have various kinds of complexes, and each of them can be described by a vector $\boldsymbol{\theta}$ of length N_L, with each entry θ_i as the number of L_i in this complex. The valency of each complex may be different, and for complex $\boldsymbol{\theta}$ its valency is $|\boldsymbol{\theta}|$. The proportion of $\boldsymbol{\theta}$ among all complexes is defined as $C_{\boldsymbol{\theta}}$, and the concentration of each $\boldsymbol{\theta}$ complex will be $L_0 C_{\boldsymbol{\theta}}$. For example, if we create a mixture of 20% of bivalent L_1 and 80% of bispecific $L_1 - L_2$, then $\boldsymbol{\theta_1} = [2, 0]$, $\boldsymbol{\theta_2} = [1, 1]$, $C_{\boldsymbol{\theta_1}} = 20\%$, and $C_{\boldsymbol{\theta_2}} = 80\%$. If the mixture solution has a total concentration of 10 nM, then the concentration of $\boldsymbol{\theta_1}$ is 2 nM, and the concentration of $\boldsymbol{\theta_2}$ is 8 nM.

We further conceptualize that Θ is a set of all existing $\boldsymbol{\theta}$'s. By this setting, we should have $\sum_{\boldsymbol{\theta} \in \Theta} C_{\boldsymbol{\theta}} = 1$. These complexes will bind in various configurations which can all be described as a $\mathbf{q}$. We define Q as a set of all possible $\mathbf{q}$'s, and we borrow the notation $\mathbf{q} \subseteq \boldsymbol{\theta}$ to indicate any binding configuration $\mathbf{q}$ that can be achieved by complex $\boldsymbol{\theta}$. This is equivalent to $|\mathbf{q_{i\bullet}}| = \theta_i$ for all i, or $\boldsymbol{\theta}$ is the row sum of (q_{ij}).

Solve the amount of free receptors A remaining problem in the model setup is that in practice we can only experimentally measure the total amount of receptor of each kind expressed by a cell, $R_{\text{tot},j}$, while the amount of free receptors at equilibrium, $R_{\text{eq},j}$, though being used extensively in the model derivation, is unknown. To find $R_{\text{eq},j}$, we first need to derive the amount of bound receptors of each kind, $R_{\text{bound},j}$, then use conservation of mass to solve $R_{\text{eq},j}$ numerically.

To calculate the amount of bound ligand $R_{\text{bound},n}$, we can simply add up all entries at the n-th column for every $\mathbf{q}$'s:

$$R_{\text{bound},n} = \sum_{\mathbf{q} \in Q} |\mathbf{q_{\bullet n}}| v_{\mathbf{q}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$
\begin{aligned}
&= \sum_{\mathbf{q}\in Q} |\mathbf{q_{\bullet n}}| \frac{L_0 C_{\boldsymbol{\theta}(\mathbf{q})}}{K_x^*} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} \psi_{ij}^{q_{ij}} \prod_{i=1}^{N_L} \binom{\theta_i}{\mathbf{q_{i\bullet}}} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \sum_{\mathbf{q}\subseteq\boldsymbol{\theta}} |\mathbf{q_{\bullet n}}| \prod_{(i,j)=(1,0)}^{(N_L,N_R)} \psi_{ij}^{q_{ij}} \prod_{i=1}^{N_L} \binom{\theta_i}{\mathbf{q_{i\bullet}}} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[\frac{\psi_{1n}}{|\boldsymbol{\psi_{1\bullet}}|}\theta_1 + \cdots + \frac{\psi_{N_L n}}{|\boldsymbol{\psi_{N_L\bullet}}|}\theta_{N_L} \right] \prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[\sum_{i=1}^{N_L} \frac{\psi_{in}}{|\boldsymbol{\psi_{i\bullet}}|}\theta_i \right] \prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i},
\end{aligned}
$$

where $|\mathbf{q_{\bullet n}}| = \sum_{m=1}^{N_L} q_{mn}$, and $|\boldsymbol{\psi_{i\bullet}}| = \sum_{j=0}^{N_R} \psi_{ij}$.
By the conservation of mass, we have

$$
\begin{aligned}
R_{\mathrm{tot},n} &= R_{\mathrm{eq},n} + R_{\mathrm{bound},n} \\
&= R_{\mathrm{eq},n} + \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[\sum_{i=1}^{N_L} \frac{\psi_{in}}{|\boldsymbol{\psi_{i\bullet}}|}\theta_i \right] \prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i}.
\end{aligned}
$$

In this equation, $R_{\mathrm{tot},n}$ are known, and any $\psi_{i\bullet}$ is a function of every $R_{\mathrm{eq},j}$, $j = 1, 2, \cdots, N_R$, so all $R_{\mathrm{eq},j}$ need to be solved together. This system of equations usually does not have a closed form and must be solved numerically. When implementing, we suggest taking the logarithm of both sides of these equations so the exponents can be eliminated and the range is restricted to positive numbers.

As a side note, the total amount of bound receptors regardless of which kind is

$$
\begin{aligned}
R_{\mathrm{bound}} &= \sum_{n=1}^{N_R} R_{\mathrm{bound},n} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \sum_{n=1}^{N_R} \left[\frac{\psi_{1n}}{|\boldsymbol{\psi_{1\bullet}}|}\theta_1 + \cdots + \frac{\psi_{N_L n}}{|\boldsymbol{\psi_{N_L\bullet}}|}\theta_{N_L} \right] \prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[(1 - \frac{\psi_{10}}{|\boldsymbol{\psi_{1\bullet}}|})\theta_1 + \cdots + (1 - \frac{\psi_{N_L n}}{|\boldsymbol{\psi_{N_L\bullet}}|})\theta_{N_L} \right] \prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[|\boldsymbol{\theta}| - \sum_{i=1}^{N_L} \frac{\theta_i}{|\boldsymbol{\psi_{i\bullet}}|} \right] \prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i}.
\end{aligned}
$$

The amount of bound ligand complexes Our model makes many macroscopic predictions readily accessible. For example, the amount of ligand bound

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

at equilibrium is a useful quantity when measuring the overall quantity of tagged ligand. To compute this number, we can add up all $v_{\mathbf{q}}$ except the $\mathbf{q}$'s that only have nonzero values on the 0-th column, $v_{0,\text{eq}}$. Consequently, the model prediction of bound ligand at equilibrium is

$$\begin{aligned}
L_{\text{bound}} &= \sum_{\mathbf{q}\in Q} v_{\mathbf{q}} - v_{0,\text{eq}} \\
&= \sum_{\mathbf{q}\in Q} \frac{L_0 C_{\boldsymbol{\theta}(\mathbf{q})}}{K_x^*} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} {\psi_{ij}}^{q_{ij}} \prod_{i=1}^{N_L} \binom{\theta_i}{\mathbf{q_{i\bullet}}} - \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[\sum_{\mathbf{q}\subseteq\boldsymbol{\theta}} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} {\psi_{ij}}^{q_{ij}} \prod_{i=1}^{N_L} \binom{\theta_i}{\mathbf{q_{i\bullet}}} - 1 \right] \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[|\boldsymbol{\psi_{1\bullet}}|^{\theta_1} |\boldsymbol{\psi_{2\bullet}}|^{\theta_2} \dots |\boldsymbol{\psi_{N_L\bullet}}|^{\theta_{N_L}} - 1 \right] \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \left[\prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i} - 1 \right]
\end{aligned}$$

when $|\boldsymbol{\psi_{i\bullet}}| = \sum_{j=0}^{N_R} \psi_{ij}$, and the predicted amount of bound complex $\boldsymbol{\theta}$ (complex of each kind) is

$$L_{\text{bound},\boldsymbol{\theta}} = \frac{L_0 C_{\boldsymbol{\theta}}}{K_x^*} \left[\prod_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|^{\theta_i} - 1 \right]$$

The amount of fully bound ligands In multivalent complexes like bispecific antibodies, drug activity may require that all subunits be bound to their respective targets [13]. The predicted amount of ligand bound f-valently can be calculated as

$$\begin{aligned}
v_{f,\text{eq}} &= \sum_{\boldsymbol{\theta}\in\Theta} \sum_{q_{10},\dots,q_{N_L 0}=0} \frac{L_0 C_{\boldsymbol{\theta}}}{K_x^*} \prod_{(i,j)=(1,1)}^{(N_L,N_R)} \psi_{ij}^{q_{ij}} \binom{\theta_1}{\mathbf{q^*_{1\bullet}}} \cdots \binom{\theta_{N_L}}{\mathbf{q^*_{N_L\bullet}}} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \prod_{i=1}^{N_L} \left(\sum_{j=1}^{N_R} \psi_{ij} \right)^{\theta_i} \\
&= \frac{L_0}{K_x^*} \sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} \prod_{i=1}^{N_L} \left(|\boldsymbol{\psi_{i\bullet}}| - 1 \right)^{\theta_i},
\end{aligned}$$

with $\mathbf{q^*_{i\bullet}} = (q_{i1}, \dots, q_{iN_R})$, the $\mathbf{q_{i\bullet}}$ vector without q_{i0}. In this equation, the multinomial coefficient $\binom{\theta_i}{\mathbf{q^*_{i\bullet}}}$ describes the number of ways one can allocate θ_i

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

receptors to any position in the i-th row of the (q_{ij}) matrix except the 0-th row which stands for unbound.

In fact, the predicted amount of any specific-valently bound ligands can be derived in such manner. For example, the amount of ligands that bind monovalently can be calculated as

$$\begin{aligned} v_{1,\text{eq}} &= \sum_{\boldsymbol{\theta}\in\Theta} \frac{L_0 C_{\boldsymbol{\theta}}}{K_x^*} \sum_{i=1}^{N_L}\sum_{j=1}^{N_R} \psi_{ij}^{q_{ij}} \binom{\theta_i}{1} \\ &= \sum_{\boldsymbol{\theta}\in\Theta} \frac{L_0 C_{\boldsymbol{\theta}}}{K_x^*} \sum_{i=1}^{N_L} |\boldsymbol{\psi_{i\bullet}}|\theta_i. \end{aligned}$$

This can be used for estimating the amount of multimerized ligands, $L_{\text{multi}} = L_{\text{bound}} - v_{1,\text{eq}}$, and multimerized receptors, $R_{\text{multi}} = R_{\text{bound}} - v_{1,\text{eq}}$.

4.2 Complexes formed through random assortment

Another common mode of forming multivalent complexes in biology, such as in the formation of antibody-antigen complexes [16], is engagement of monomer units to a common scaffold. Instead of resulting in a specific arrangement, we provide binding compounds of a fixed valency f and a litany of monomer ligands, and complexes can form through random assortment. The concentration of these complexes, therefore, will follow a multinomial distribution.

To formulate this mathematically, we denote the proportion of L_i as C_i, and $\sum_{i=1}^{N_L} C_i = 1$. For example, we have 40% L_1 and 60% L_2 in the solution to form dimers ($f = 2$), then $C_1 = 40\%$, $C_2 = 60\%$. Assume complex formation follows a binomial distribution, there will be 16% bivalent L_1, 36% bivalent L_2, and 48% $L_1 - L_2$ complex. When a complex is randomly assembled from the monomer ligands, the probability of such complex formed as described by $\boldsymbol{\theta}$ is

$$C_{\boldsymbol{\theta}} = \binom{f}{\boldsymbol{\theta}} C_1^{\theta_1} C_2^{\theta_2} \ldots C_{N_L}^{\theta_{N_L}} = \binom{f}{\boldsymbol{\theta}} \prod_{i=1}^{N_L} C_i^{\theta_i}$$

Since $\sum_{i=1}^{N_L} C_i = 1$, we know that

$$\sum_{\boldsymbol{\theta}\in\Theta} C_{\boldsymbol{\theta}} = \sum_{\boldsymbol{\theta}\in\Theta} \binom{f}{\boldsymbol{\theta}} \prod_{i=1}^{N_L} C_i^{\theta_i} = (C_1 + C_2 + \ldots + C_{N_L})^f = 1$$

Plugging this relationship between $C_{\boldsymbol{\theta}}$ and C_i into the equation for the amount of a specific binding configuration derived in the previous section, we have

$$v_{\mathbf{q}} = \frac{L_0}{K_x^*} \binom{f}{\boldsymbol{\theta}(\mathbf{q})} \prod_{i=1}^{N_L} C_i^{\theta_i(\mathbf{q})} \prod_{(i,j)=(1,1)}^{(N_L,N_R)} \psi_{ij}^{q_{ij}} \prod_{i=1}^{N_L} \binom{\theta_i(\mathbf{q})}{\mathbf{q_{i\bullet}}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$
\begin{aligned}
&= \frac{L_0}{K_x^*}\binom{f}{\mathbf{q}} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} C_i^{q_{ij}} \prod_{(i,j)=(1,1)}^{(N_L,N_R)} \psi_{ij}^{q_{ij}} \\
&= \frac{L_0}{K_x^*}\binom{f}{\mathbf{q}} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} \varphi_{ij}^{q_{ij}},
\end{aligned}
$$

where $\varphi_{ij} = R_{\mathrm{eq},j} K_{a,ij} K_x^* C_i$ and $\varphi_{i0} = C_i$.

Solve the amount of free receptors Like in the specific arrangement case, we still need to solve $R_{\mathrm{eq},n}$ numerically from $R_{\mathrm{tot},n}$. We first derive the amount of bound receptors of each kind at equilibrium as

$$
\begin{aligned}
R_{\mathrm{bound},n} &= \sum_{\mathbf{q}\in Q} |\mathbf{q}_{\bullet\mathbf{n}}| v_{\mathbf{q}} \\
&= \sum_{\mathbf{q}\in Q} |\mathbf{q}_{\bullet\mathbf{n}}| \binom{f}{\mathbf{q}} \frac{L_0}{K_x^*} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} \varphi_{ij}^{q_{ij}} \\
&= \frac{L_0}{K_x^*} \sum_{\mathbf{q}\in Q} \sum_{m=1}^{N_L} q_{mn} \binom{f}{\mathbf{q}} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} \varphi_{ij}^{q_{ij}} \\
&= \frac{L_0}{K_x^*} \sum_{m=1}^{N_L} \varphi_{mn} f \left[\sum_{(i,j)=(1,0)}^{(N_L,N_R)} \varphi_{ij} \right]^{f-1} \\
&= \frac{L_0 f}{K_x^*} |\boldsymbol{\varphi}_{\bullet\boldsymbol{n}}| (1+\Phi)^{f-1}.
\end{aligned}
$$

Then by the conservation of mass, we have the equation to numerically solve for $R_{\mathrm{eq},n}$:

$$
\begin{aligned}
R_{\mathrm{tot},n} &= R_{\mathrm{eq},n} + R_{\mathrm{bound},n} \\
&= R_{\mathrm{eq},n} + \frac{L_0 f}{K_x^*} |\boldsymbol{\varphi}_{\bullet\boldsymbol{n}}| (1+\Phi)^{f-1}.
\end{aligned}
$$

Again, since Φ is a function of every $R_{\mathrm{eq},n}$, all $R_{\mathrm{eq},n}$ need to be solved together.

The amount of k-valently bound complexes For randomly assorted complexes, we first derive the amount of ligands that bind k-valently. As we will show, it has a nice expression that can used to calculate many other quantities conveniently. First, let's break $\mathbf{q}$ into two separate vectors, $\mathbf{q} = (\mathbf{q}_{\bullet\mathbf{0}}, \mathbf{q}_{\bullet\mathbf{x}})$. We define the vector formed by the 0-th column of $\mathbf{q}$ which stand for unbound as $\mathbf{q}_{\bullet\mathbf{0}}$, and the one formed by the other elements as $\mathbf{q}_{\bullet\mathbf{x}}$. By the model setup, we know $|\mathbf{q}| = f$, $|\mathbf{q}_{\bullet\mathbf{x}}| = k$, and $|\mathbf{q}_{\bullet\mathbf{0}}| = f - k$. We then have

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$
\begin{aligned}
v_{k,\mathrm{eq}} &= \sum_{\substack{|\mathbf{q}_{\bullet\mathbf{x}}|=k \\ |\mathbf{q}_{\bullet\mathbf{0}}|=f-k}} v_{\mathbf{q}} = \sum_{\substack{|\mathbf{q}_{\bullet\mathbf{x}}|=k \\ |\mathbf{q}_{\bullet\mathbf{0}}|=f-k}} \binom{f}{\mathbf{q}_{\bullet\mathbf{x}} \quad \mathbf{q}_{\bullet\mathbf{0}}} \frac{L_0}{K_x^*} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} {\varphi_{ij}}^{q_{ij}} \\
&= \sum_{\substack{|\mathbf{q}_{\bullet\mathbf{x}}|=k \\ |\mathbf{q}_{\bullet\mathbf{0}}|=f-k}} \binom{f}{k} \binom{k}{\mathbf{q}_{\bullet\mathbf{x}}} \binom{f-k}{\mathbf{q}_{\bullet\mathbf{0}}} \frac{L_0}{K_x^*} \prod_{(i,j)=(1,1)}^{(N_L,N_R)} {\varphi_{ij}}^{q_{ij}} \prod_{i=1}^{i=N_L} {C_i}^{q_{i0}} \\
&= \frac{L_0}{K_x^*} \binom{f}{k} \left[\sum_{|\mathbf{q}_{\bullet\mathbf{x}}|=k} \binom{k}{\mathbf{q}_{\bullet\mathbf{x}}} \prod_{(i,j)=(1,1)}^{(N_L,N_R)} {\varphi_{ij}}^{q_{ij}} \right] \left[\sum_{|\mathbf{q}_{\bullet\mathbf{0}}|=f-k} \binom{f-k}{\mathbf{q}_{\bullet\mathbf{0}}} \prod_{i=1}^{i=N_L} {C_i}^{q_{i0}} \right] \\
&= \frac{L_0}{K_x^*} \binom{f}{k} \left[\sum_{(i,j)=(1,1)}^{(N_L,N_R)} {\varphi_{ij}}^{q_{ij}} \right]^k \left[\sum_{i=1}^{i=N_L} C_i \right]^{f-k} = \frac{L_0}{K_x^*} \binom{f}{i} \Phi^k.
\end{aligned}
$$

The amount of total bound ligands and receptors Many macroscopic properties can be derived from $v_{k,\mathrm{eq}}$. For example, the amount of total bound ligands is simply the sum of ligands bound monovalently to fully, and can be simplified to

$$
\begin{aligned}
L_{\mathrm{bound}} &= \sum_{k=1}^{f} v_{k,\mathrm{eq}} = \sum_{k=0}^{f} v_{k,\mathrm{eq}} - v_{0,\mathrm{eq}} \\
&= \sum_{k=0}^{f} \frac{L_0}{K_x^*} \binom{f}{k} \Phi^k - \frac{L_0}{K_x^*} \binom{f}{0} \Phi^0 \\
&= \frac{L_0}{K_x^*} \left[(1+\Phi)^f - 1 \right].
\end{aligned}
$$

Similarly, the total bound receptors should be

$$
\begin{aligned}
R_{\mathrm{bound}} &= \sum_{k=1}^{f} k v_{k,\mathrm{eq}} = \sum_{k=1}^{f} k \frac{L_0}{K_x^*} \binom{f}{k} \Phi^k \\
&= \frac{L_0}{K_x^*} f \Phi (1+\Phi)^{f-1}.
\end{aligned}
$$

As we show here, these quantities all have elegant closed form solutions, and they are only dependent on Φ, a single value that incorporate all information about receptor amounts, monomer ligand compositions, and binding affinities.

The number of cross-linked receptors In some biological contexts such as T cell receptor-MHC [7] or antibody-Fc receptor [16] interactions, signal

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

transduction is driven by receptor cross-linking due to multivalent binding. The amount of total cross-linked receptors can be derived from $v_{k,\mathrm{eq}}$ as

$$\begin{aligned} R_{\mathrm{multi}} &= \sum_{k=2}^{f} k v_{k,\mathrm{eq}} = R_{\mathrm{bound}} - v_{1,\mathrm{eq}} \\ &= \frac{L_0}{K_x^*} f\Phi(1+\Phi)^{f-1} - \frac{L_0}{K_x^*}\binom{f}{1}\Phi = \frac{L_0}{K_x^*} f\Phi\left[(1+\Phi)^{f-1} - 1\right]. \end{aligned}$$

To find the number of crosslinked receptors of a specific kind, R_n, requires extra consideration. Similar to how $v_{k,\mathrm{eq}}$ was found, we break break $\mathbf{q}$ into three separate vectors, $\mathbf{q} = (\mathbf{q_{\bullet 0}}, \mathbf{q_{\bullet n}}, \mathbf{q_{\bullet x}})$. $\mathbf{q_{\bullet 0}}$ is the vector formed by the 0-th column of $\mathbf{q}$, $\mathbf{q_{\bullet n}}$ is the vector formed by the n-th column of $\mathbf{q}$, and $\mathbf{q_{\bullet x}}$ contains all others. If we assume that a complex is s-valently bound, then $|\mathbf{q_{\bullet 0}}| = f - s$. We further assume that $|\mathbf{q_{\bullet n}}| = t$, then $|\mathbf{q_{\bullet x}}| = s - t$. By this setup, we have

$$\begin{aligned} R_{\mathrm{multi},n} &= \sum_{s=2}^{f}\sum_{t=0}^{s} t \sum_{\substack{|\mathbf{q_{\bullet x}}|=s-t \\ |\mathbf{q_{\bullet n}}|=t \\ |\mathbf{q_{\bullet 0}}|=f-s}} v_{\mathbf{q}} = \sum_{s=2}^{f}\sum_{t=0}^{s} t \sum_{\substack{|\mathbf{q_{\bullet x}}|=s-t \\ |\mathbf{q_{\bullet n}}|=t \\ |\mathbf{q_{\bullet 0}}|=f-s}} \binom{f}{\mathbf{q}} \frac{L_0}{K_x^*} \prod_{(i,j)=(1,0)}^{(N_L,N_R)} {\varphi_{ij}}^{q_{ij}} \\ &= \sum_{s=2}^{f}\sum_{t=0}^{s} \frac{tL_0}{K_x^*}\binom{f}{s-t\quad t\quad f-s} \left[\sum_{|\mathbf{q_{\bullet x}}|=s-t} \binom{s-t}{\mathbf{q_{\bullet x}}} \prod_{\substack{(i,j)=(1,0) \\ j\neq n}}^{(N_L,N_R)} {\varphi_{ij}}^{q_{ij}}\right] \\ &\left[\sum_{|\mathbf{q_{\bullet n}}|=t} \binom{t}{\mathbf{q_{\bullet n}}} \prod_{i=1}^{N_L} {\varphi_{in}}^{q_{in}}\right]\left[\sum_{|\mathbf{q_{\bullet 0}}|=f-s} \binom{f-s}{\mathbf{q_{\bullet 0}}} \prod_{i=1}^{N_L} {C_i}^{q_{i0}}\right] \\ &= \sum_{s=2}^{f}\sum_{t=0}^{s} \frac{tL_0}{K_x^*}\binom{f}{s-t\quad t\quad f-s} (\Phi - |\boldsymbol{\varphi_{\bullet n}}|)^{s-t} |\boldsymbol{\varphi_{\bullet n}}|^t \left(\sum_{i=1}^{N_L} C_i\right)^{f-s} \\ &= \sum_{s=2}^{f} \frac{L_0}{K_x^*}\left[\sum_{t=0}^{s} t\binom{s}{t}\left(\frac{|\boldsymbol{\varphi_{\bullet n}}|}{\Phi - |\boldsymbol{\varphi_{\bullet n}}|}\right)^t\right]\binom{f}{s}(\Phi - |\boldsymbol{\varphi_{\bullet n}}|)^s \\ &= \sum_{s=2}^{f} \frac{L_0}{K_x^*} s\left(\frac{|\boldsymbol{\varphi_{\bullet n}}|}{\Phi - |\boldsymbol{\varphi_{\bullet n}}|}\right)\left(\frac{\Phi}{\Phi - |\boldsymbol{\varphi_{\bullet n}}|}\right)^{s-1}\binom{f}{s}(\Phi - |\boldsymbol{\varphi_{\bullet n}}|)^s \\ &= \frac{L_0}{K_x^*}\left[\sum_{s=2}^{f} s\binom{f}{s}\Phi^s\right]\frac{|\boldsymbol{\varphi_{\bullet n}}|}{\Phi} \\ &= \frac{L_0}{K_x^*}\frac{|\boldsymbol{\varphi_{\bullet n}}|}{\Phi}\left[f\Phi(1+\Phi)^{f-1} - f\Phi\right] \\ &= \frac{L_0 f}{K_x^*}|\boldsymbol{\varphi_{\bullet n}}|[(1+\Phi)^{f-1} - 1]. \end{aligned}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

This formula can useful when investigating the role of each receptor in a pathway that requires multimerized binding.

Of course, the macroscopic predictions provided in this section cannot exhaust many biological quantities one may wish to study, but with the ideas we have demonstrated here, the readers can derive their own formulae as needed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

5 Application examples

In previous sections, we have shown how all macroscopic predictions made in this work can be written in closed form formulae. Therefore, many computational methods such as auto-differentiation and sensitivity analysis can be easily applied. These analyses will bring great insights into the complex behavior of multivalent binding. Here, we provide two examples to demonstrate the advantage of large-scale predictions made possible by this model.

5.1 Mixture binding prediction

Leveraging the synergistic effect among two or more drugs is of great interest in pharmaceutical development. A challenge in investigating synergy is to identify its underlying source. Most biological pathways follow a similar pattern: when the drug binds to certain surface receptors of a cell, a downstream pathway in the cell is initiated, leading to some actions. Therefore in general, synergism can come from either the initial binding events themselves or downstream processes. Binding-level synergy means that merely using a combination of ligands boosts the amount of binding to the important receptors and thus intensifies the overall effect. Downstream effect synergy indicates that the benefit of using mixtures arises from other cellular regulatory mechanisms two ligands can bring about. The binding model we introduced can help to investigate this issue by offering accurate predictions for the binding of multivalent complex mixtures.

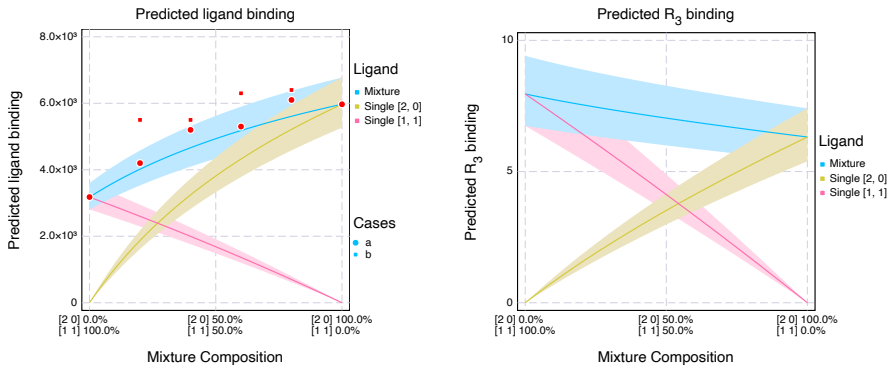

Figure 3: Prediction on mixture binding of $\boldsymbol{\theta_1} = [2, 0]$ and $\boldsymbol{\theta_2} = [1, 1]$. The left panel shows the predicted total ligand binding, while the right shows the amount of bound R_3 at equilibrium. Shaded areas are simulated confidence interval by varying the receptor levels up and down by 10%. The red dots on the left panel are simulated experimental results. In case a (red circles), since most data points are inside the confidence interval, we can assume the measurement error can explain these variations. In case b (red squares), however, the synergism of these complexes are beyond the binding level.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In Figure 3, we provide an example of mixture binding predictions. We investigate a mixture of two types of ligand complexes, bivalent L_1 ($\boldsymbol{\theta_1} = [2, 0]$) and bispecific $L_1 - L_2$ ($\boldsymbol{\theta_2} = [1, 1]$). The crosslinking constant is set to be $K_x^* = 10^{-12}$, similar to previous results [16]. We predict the amount of binding of this mixture to a cell expressing three types of receptors, with $\mathbf{R_{tot}} = [2.5 \times 10^4, 3 \times 10^4, 2 \times 10^3]\,\mathrm{cell}^{-1}$. The affinity constants of L_1 to these three receptors are $\boldsymbol{K_{a,1\bullet}} = [1 \times 10^8, 1 \times 10^5, 6 \times 10^5]\,\mathrm{M}^{-1}$, and of L_2, $\boldsymbol{K_{a,2\bullet}} = [3 \times 10^5, 1 \times 10^7, 1 \times 10^6]\,\mathrm{M}^{-1}$. Figure 3 shows the predicted ligand bound (left panel) and R_3 bound (right panel) for only $\boldsymbol{\theta_1}$ or $\boldsymbol{\theta_2}$ with L_0 from 0 to 1 nM, and their mixtures in every possible composition with total concentration $L_0 = 1\,\mathrm{nM}$ (from $C_{\boldsymbol{\theta_1}} = 0$ and $C_{\boldsymbol{\theta_2}} = 1$ to $C_{\boldsymbol{\theta_1}} = 1$ and $C_{\boldsymbol{\theta_2}} = 0$).

Mixture binding prediction can help us identify the source of synergy. To connect model predictions to experimental measurements, ligand binding might be measured by fluorescently-tagged ligands, while the number of bound receptors of a specific type might associate with an indirect measurement such as cellular response. After making a series of measurements for different compositions of mixtures, we can fit the 100% of one complex cases (numbers on the two ends on the plot) first and then compare the mixture measurements to the predictions. Determining whether the downstream effect contributes to the observed synergy (or antagonism) can be framed as a hypothesis testing problem:

$\boldsymbol{H_0}$: *The synergism of the mixture can be explained solely by binding.*

The uncertainty of mixture binding prediction comes from measurement errors of receptor abundance and binding affinities. Usually, the receptor expression of a cell population has an empirical distribution which can be measured. The confidence interval in Figure 3 is drawn with the assumption that receptor expression fluctuates up and down for 10%, similar to the confidence interval of a log-normal distribution. Also, due to the measurement technique, the binding affinities may be over- or underestimated [14]. The confidence interval of mixture prediction can be determined by the model with all these considered, and a p-value can be even derived.

If most mixture measurements fall within the confidence interval of the predictions (such as case a annotated by the red circles in Figure 3, left panel), the synergy will very likely come from binding only. However, if the measurements are obviously beyond the confidence interval (case b, the red squares), it is reasonable to suspect a synergistic (or antagonistic) effect beyond binding alone. Because of the binding model's flexibility, this method can also be extended to a mixture of more than two compounds.

5.2 Binding space of a ligand

When a dose of ligands (drug, hormone, cytokine, etc.) is released into the circulation system of an individual due to either physiological responses or exogenous administration, the compounds will spread and bind to many cell populations to varying extents. An essential question in pharmacology is how

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

much a compound will bind to their intended target populations compared to off-target ones. This question is important for understanding basic biology as well as developing new therapeutics. For example, hormones and cytokines are important signaling molecules, and having a quantitative prediction of on- and off-target binding can help us understand their mechanism greatly. For drug development, binding prediction can guide optimization to improve specificity toward the intended targets [18]. A cell population can be defined by the protein they express, especially their surface receptors. Therefore, given the parameters of the dose and the receptor profile of a cell population, our model can make all the predictions discussed previously.

From the perspective of this binding model, there is nothing special about one specific cell population. If the local concentration is constant everywhere, our model can map any cell with a certain receptor expression to the amount of binding induced by this dose. If the biological activity of this compound on a cell is related to the quantity of binding to a certain ligand or receptor, the effect of this dose can be written as a function f, with

$$\mathbf{R_{tot}} \in \mathbb{R}_+^{N_R} \mapsto f(\mathbf{R_{tot}}) \in \mathbb{R}_+,$$

where $\mathbf{R_{tot}}$ is a vector of nonnegative entries that describes the cell's expression of N_R receptors, and $f(\mathbf{R_{tot}})$ is the amount of binding. Here, we define the binding behavior of this dose (or any compound) as its binding space.

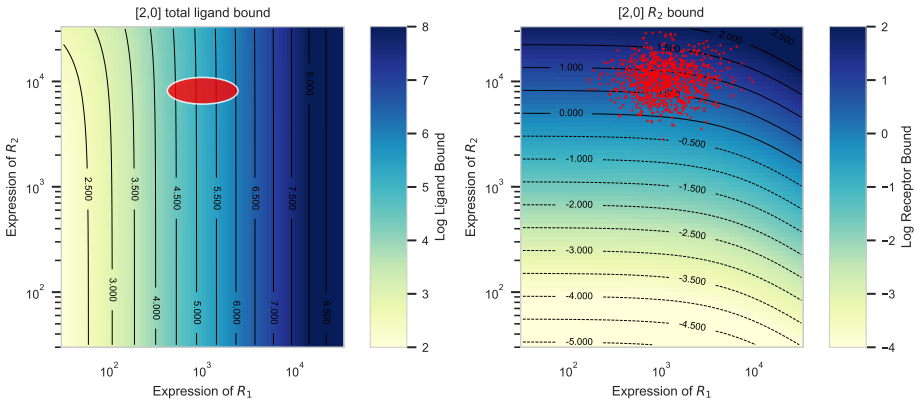

Figure 4: The binding space of 1 nM $\boldsymbol{\theta} = [2, 0]$. The left panel shows the amount of total ligand bound, while the right panel shows receptor R_2 bound predictions. The x- and y-axis show the expression of R_1 and R_2, while the expression of R_3 is a constant, 2.0×10^3 cell^{-1}, and not shown. Any cell population can be drawn on the binding space. For example, the red ellipse on the left panel represents a cell population with receptor expression at about $\mathbf{R_{tot}} = [1.0, 10.0, 2.0] \times 10^3$ cell^{-1}. We can alternatively project points of experimental single cell expression data onto a binding space, as shown on the right panel.

In Figure 4, we plot the binding space of a bivalent L_1 ligand $\boldsymbol{\theta} = [2, 0]$ with

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

concentration 1 nM. The binding affinities are the same as described in the last subsection. In this binding space, we consider three receptors, R_1, R_2, and R_3. We plot how the amount of binding relates to the cell expression profile, $\mathbf{R_{tot}}$. Here, the amount of R_1 and R_2 varies with the two axes, while R_3 is held constant at $2.0 \times 10^3\,\text{cell}^{-1}$. Then we use colors and contour lines to show the amount of binding. From these two plots, we can see that although both ligand binding and R_2 binding increase with more receptors, ligand binding is more sensitive to R_1 amounts, and R_2 binding R_2 amounts. To consider any specific cell population, we only need to determine where its expression profile falls on the plot and read the predictions from the contour line. For example, on the left panel, the red cell population will have about $e^{5.2} = 181$ bound ligands per cell. The number of contour lines a population ride on can also show intrapopulation variation. In this case, we expect the variation in ligand binding to fall between $e^{4.3} = 74$ and $e^{6.0} = 403$.

The binding space can provide ample information about the compound. It is an intrinsic property of a ligand given its concentration and other ligand it mixes with, independent of any specific cell. The biological process of drug diffusion to a certain cell is analogous to sampling a point from this binding space. Its gradient indicates in which direction the binding level increases the fastest, as well as to which receptor it is more sensitive. An inactive antagonist that introduces binding competition with the ligand can distort its binding space, and we can visualize it by the change of shape in the contour lines. This plot can also intuitively demonstrate intrapopulation binding variance and interpopulation cell specificity of the compound. With the development of high-throughput single-cell methods such as flow cytometry, the expression profiles of a collection of cells can be identified en masse, and we can overlap their results onto a binding space plot (as in Figure 4, right panel). This shows the promise of applying our model to single-cell data. Although we can only visualize two receptors in a plot, binding space applies to any N_R types of receptors. Theoretically, the concept of the binding space of a ligand is only complete when all relevant surface receptors are considered.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

6 Discussion

In this work, we propose a mechanistic multivalent binding model that accounts for the interaction among multiple receptors and a mixture of ligand complexes formed by binding monomers. We first derive the amount of ligand of a specific binding configuration at equilibrium through the law of mass action. Using this formula, we make macroscopic predictions by applying the multinomial theorem strategically. Our predictions cover cases where complexes are formed by specific arrangement or random assortment. Finally, we provide two practical examples of how this model can help with biological research.

Compared with many previous approaches, this model has several clear advantages. First of all, it is extremely efficient, and it is capable of handling a large number of receptors, ligands, and complexes types. This allows the model to make large-scale predictions easily, enabling mixture synergy analysis and binding space calculations. The mathematical elegance of the model welcomes analytical studies and incorporating it into more complicated frameworks.

The assumptions made in this model may compromise its accuracy in some cases. For example, the steric effects of a multivalent ligand can be more complicated and context-dependent. Our setup has a single crosslinking constant, K_x^*, to reflect the multivalency effect. In practice, this model works well in predicting experimental binding results [18, 15]. Some other computational approaches investigate the steric effect more meticulously, but inevitably introduce considerable added complexity [4]. When the actual situation is not known, our model can serve as an adequate starting point.

Although this model is very general purpose, it mainly focuses on the binding dynamics on a cell surface, similar to the previous work on which it is based [1, 2, 3]. For intracellular ligands discordant with the multivalent velcro shape shown in Figure 2, this model may be less suitable. For example, some previous works focus scaffold proteins in the cell signaling system for quantitative analysis [9], and various computational models different from ours have been developed [11, 6, 10].

Surface receptor binding is a universal event in biology. A prevalent question calls for a general enough solution. The model we present in this work can be successfully applied to many contexts, including predicting Fc-FcγR interaction [16] and fitting epithelial cell adhesion molecule binding data [18, 15]. With the arise of multispecific drugs in the recent decade [17], we expect this model to apply even more widely, exhibit its full competence and facilitate both basic scientific research and new therapy development.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Declaration of interest This work was supported by NIH U01-AI-148119 to A.S.M. The authors declare no competing financial interests.

Author contributions Z.C.T.: Methodology, Writing – original draft; A.S.M.: Funding acquisition, Writing – review & editing.

References

[1] Alan S Perelson and Charles DeLisi. "Receptor clustering on a cell surface. I. Theory of receptor cross-linking by ligands bearing two chemically identical functional groups". In: *Mathematical Biosciences* 48.1-2 (1980), pp. 71–110.

[2] Alan S. Perelson. "Receptor clustering on a cell surface. II. theory of receptor cross-linking by ligands bearing two chemically distinct functional groups". In: *Mathematical Biosciences* 49.1 (1980), pp. 87–110.

[3] Alan S Perelson. "Receptor clustering on a cell surface. III. Theory of receptor cross-linking by multivalent ligands: description by ligand states". In: *Mathematical Biosciences* 53.1-2 (1981), pp. 1–39.

[4] William S Hlavacek, Richard G Posner, and Alan S Perelson. "Steric effects on multivalent ligand-receptor binding: exclusion of ligand sites by bound cell surface receptors". In: *Biophysical journal* 76.6 (1999), pp. 3031–3043.

[5] William S Hlavacek et al. "Quantifying aggregation of IgE-FcεRI by multivalent antigen". In: *Biophysical journal* 76.5 (1999), pp. 2421–2431.

[6] Andre Levchenko, Jehoshua Bruck, and Paul W Sternberg. "Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties". In: *Proceedings of the National Academy of Sciences* 97.11 (2000), pp. 5818–5823.

[7] Jennifer D Stone, Jennifer R Cochran, and Lawrence J Stern. "T-cell activation by soluble MHC oligomers can be described by a two-parameter binding model". In: *Biophysical Journal* 81.5 (2001), pp. 2547–2557.

[8] Jodi M Paar et al. "Bivalent ligands with rigid double-stranded DNA spacers reveal structural constraints on signaling by FcεRI". In: *The Journal of Immunology* 169.2 (2002), pp. 856–864.

[9] Stephen A Chapman and Anand R Asthagiri. "Quantitative effect of scaffold abundance on signal propagation". In: *Molecular systems biology* 5.1 (2009), p. 313.

[10] Yinghao Wu et al. "Transforming binding affinities from three dimensions to two with application to cadherin clustering". In: *Nature* 475.7357 (2011), pp. 510–513.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.10.434776; this version posted March 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[11] Jin Yang and William S Hlavacek. "Scaffold-mediated nucleation of protein signaling complexes: elementary principles". In: *Mathematical biosciences* 232.2 (2011), pp. 164–173.

[12] Catherine A Macken and Alan S Perelson. *Branching processes applied to cell surface aggregation phenomena*. Vol. 58. Springer Science & Business Media, 2013.

[13] Emily C Piccione et al. "A bispecific antibody targeting CD47 and CD20 selectively binds and eliminates dual antigen expressing lymphoma cells". In: *MAbs*. Vol. 7. 5. Taylor & Francis. 2015, pp. 946–956.

[14] SA Hunter and JR Cochran. "Cell-binding assays for determining the affinity of protein–protein interactions: technologies and considerations". In: *Methods in enzymology* 580 (2016), pp. 21–44.

[15] Clifford M Csizmar et al. "Multivalent ligand binding to cell membrane antigens: defining the interplay of affinity, valency, and expression density". In: *Journal of the American Chemical Society* 141.1 (2018), pp. 251–261.

[16] Ryan A Robinett et al. "Dissecting FcγR regulation through a multivalent binding model". In: *Cell systems* 7.1 (2018), pp. 41–48.

[17] Raymond J Deshaies. "Multispecific drugs herald a new era of biopharmaceutical innovation". In: *Nature* 580.7803 (2020), pp. 329–338.

[18] Zhixin Cyrillus Tan, Brian Orcutt-Jahns, and Aaron S Meyer. "A quantitative view of strategies to engineer cell-selective ligand binding". In: *bioRxiv* (2020).