Combinatorial model of ligand-receptor binding

Mobolaji Williams

January 24, 2021

Abstract

We introduce a combinatorial model of ligand-receptor binding that allows us to quantitatively frame the question "How can ligands seek out and bind to their optimal receptor sites in a sea of other competing ligands and suboptimal receptor sites?" To answer the question, we first derive a formula to count the number of partial generalized derangements in a list; the result is an extension to a combinatorial result by Gillis and Even. We then compute the general partition function for the ligand-receptor system and derive the equilibrium expressions for the average number of bound ligands and the average number of optimally bound ligands. A visual model of squares assembling onto a grid allows us to easily identify fully optimal bound states. Equilibrium simulations of the system reveal its extremes to be one of two types, qualitatively distinguished by whether optimal ligand-receptor binding is the dominant form of binding at all temperatures and quantitatively distinguished by the relative values of two critical temperatures. One of those system types (termed "search-limited," as it was in previous work) does not exhibit kinetic traps and we thus infer that biomolecular systems where optimal ligand-receptor binding is functionally important are likely to be search-limited.

Keywords: Derangements, Laguerre Polynomials, Statistical Physics, Ligands and Receptors, Assembly

Contents

1 Introduction .................................................. 2
2 Partial Generalized Derangements .................. 5
3 General Partition Function .............................. 9
   3.1 Gendered Dimer-System Assembly ................ 11
4 Large \( N \) Limits of Special and General Cases .... 14
   4.1 Simple Binding Model .............................. 14
   4.2 Derangement Model ............................... 17
   4.3 General Case ...................................... 20
1 Introduction

The interaction between membrane receptors and extracellular ligands is the starting point for many cell-signaling pathways [SML+20]. Given the intricacy of these pathways, one might think that the initiating ligand-receptor interaction needs to be “highly specific” (i.e., one ligand type only binds to one receptor type). But work over the past two decades suggests the opposite: The specificity of the resulting processes requires such a precise code that only a combinatorial one, which makes use of various combinations of a finite number of inputs, can achieve it. This was found in the case of olfactory receptors [MHSB99] where different receptors recognized different combinations of ligands. Also, polypharmacology, a recent branch of drug design referring to creating ligands that act on multiple targets, has been found to be necessary for treating complex diseases such as
schizophrenia [RSK04]. Others have found that having many-to-many interactions between ligands and receptors promotes an increased diversity in range of responses for signaling pathways [SML+20].

All of these contexts for ligand-receptor interactions allow us to conceive of a stripped down model of the extracellular medium as one where receptors of various types and copy numbers exist on a cell surface surrounded by ligands of various types and copy numbers. Due to the many-to-many interactions between ligands and receptors (also called, “multi-specific” or “promiscuous” binding), in the most general case such a system exhibits bindings featuring all combinations of receptors and ligands. Still, to provide a reference point for the affinities, we can highlight, for each ligand type, a single binding between receptor and ligand that is the strongest for that ligand. We can term such interactions as “optimal” to distinguish them from other interactions.

Such a representation of the ligand-receptor system presents us with a question: How do the various binding affinities between specific receptors and ligands affect the global binding properties (e.g., total number of bound ligands, total number of optimally bound ligands, temperature at which optimal binding occurs, etc.) of the entire ligand-receptor system?

To answer this question, we have to consider both the combinatorial and kinetic subproblems of the system. The combinatorial subproblem for ligand-receptor binding concerns how ligands can arrange themselves so that each one attaches to its optimal receptor site, and the kinetic subproblem concerns how ligands can find these optimal receptor sites in the volume they occupy. The two subproblems together are also the prototypical definitions of a self-assembly process in which initially distanced units must come together and combine in the correct ordered configuration through a thermal system’s unforced evolution towards a free-energy minimum [Nel04, JLD10, PH15]. There are few analytically tractable model archetypes that can treat the respective influences of combinatorics and kinetics on such processes. This work aims to propose such an archetype.

There are some well known approaches to modeling ligand-receptor binding. Most well known is the law of mass action which has often been applied to ligand-receptor binding since it provides a coarse-grained framework to model how affinities affect bound concentrations (see chapter five of [RP08] for a summary). However, this approach does not take into account the competition between ligands that can place additional limits on the achievement of specific types of binding.

A combinatorial model of ligand-receptor binding was presented in chapter six of [PTKG12]. There the authors considered a collection of identical ligands in a grid-like space that contained a single receptor. The main combinatorial task in that analysis was determining the number of ways to arrange the ligands amongst the spatial grid for both bound and unbound configurations. From the answer, the authors computed the system partition function and the ligand-receptor binding concentration. However, the model did not consider different ligand species existing in the same environment and thus did not account for the combinatorial competition between various species in real systems.

In the present work, combinatorics is incorporated through the finiteness of the number of different particle species in the system and the resulting finiteness of the possible number of ligand-receptor interactions. Also, by considering multiple ligand-types each of different copy numbers and with distinct “promiscuous” (or multi-specific) binding affinities to a similarly diverse set of
1 Introduction

Figure 1: Example ligand-receptor system. The figure displays three different types of ligands and three corresponding types of receptors. All ligands can bind to all receptors, but each receptor has an optimal binding with a specific receptor type. The ligands not bound to any receptor sites are free and exist in the volume of the system. In this work we derive the conditions under which all ligands can bind to their optimal receptor sites.

receptor sites, the total possible set of combinatorial bindings better approaches that of a real system. The net consequence of these assumptions is to introduce combinatorial competition into the equilibrium statistical physics that define the system, making finite number effects particularly important in describing thermal properties.

The general system we consider is shown in Fig. 1. Say we have a system of ligands and receptors existing in the extracellular medium. The ligands come in multiple copies as do the receptors, and, as is consistent with the multi-specificity of some real ligand systems, we assume all ligands have the ability to bind to all receptors. However, we will also assume that each ligand type has a specific optimal binding with a particular receptor type. This latter assumption will provide us with an additional order parameter with which we can define our system.

There are some basic questions that a corresponding model for this system should be able to answer: How does the average number of bound ligands of various types depend on the system’s binding affinities? How does the average number of optimally bound ligands of various types depend on the system’s binding affinities? What thermal conditions define a system in which all ligands are optimally bound? For what properties of the affinities are such conditions even feasible? Are there ways we can categorize these systems so as to determine a priori from affinity properties what the expected binding behavior should be?

We answer these questions in the subsequent sections. In Section 2, we introduce the main combinatorial problem that underlies the binding of multiple ligand species to multiple receptor species. In Section 3, we use the solution to this problem to compute the partition function for the system and show that the result generalizes a special case derived in [Wil19]. In Section 4, we consider the large particle-number limit of the partition function for two limiting cases and the general case. The limiting cases help us build the intuition relevant to understanding the conditions that define fully optimal binding in the general case. In each case, we derive expressions for the average number of bound ligands of each species and the average number of optimally bound ligands supposing the relevant quantity is not trivially constrained by the case itself. In Section 5, we introduce an image
on a grid to give a visual handle on the various limiting cases of the system, and we simulate the
grid images to affirm that the analytical results accurately predict the average grid state at various
temperatures. In Section 6, we note that the temperature curves for the average number of bound
ligands and the average number of optimally bound ligands have distinct limiting behaviors contin-
gent on how model parameters vary with one another. We explore these distinct limiting behaviors
through simulations and argue that one limiting behavior is associated with kinetic traps. In Section
7, we return to the system that motivated our analyses and discuss the biophysical implications of
the results. In Section 8, we conclude by considering ways to extend the general model.

2 Partial Generalized Derangements

Our ultimate goal is to model the equilibrium thermodynamics of systems of the kind shown in Fig.
1. Achieving this amounts to computing a partition function and then using the partition function
to find order parameters, but first we need to solve the combinatorial problem at the heart of this
system.

We recall that the derangement of a list is a rearrangement of that list such no element is in its
original position. The formula for the number of derangements of a list with \( N \) unique elements
was first obtained by Pierre Mortmort and Nicholas Bernoulli in the early 18th century [dM13]:

\[
d_N = \sum_{j=0}^{N} \binom{N}{j} (-1)^j (N - j)!
\]

(1)

More than 200 years later, Gillis and Even derived the generalization to this result for the case
where elements occur with multiple copy number [EG76]. They showed that the number of ways
to completely derange an ordered list with \( n_1 \) elements of type 1, \( n_2 \) elements of type 2, \ldots, and \( n_R \)
elements of type \( R \) (where \( R \leq N \)) is

\[
G_n = \int_0^\infty dx \ e^{-x} \prod_{k=1}^{R} (-1)^{n_k} L_{n_k}(x),
\]

(2)

where \( n \equiv (n_1, n_2, \ldots, n_R) \) and \( L_n(x) \) is the Laguerre polynomial defined as

\[
L_n(x) = \sum_{j=0}^{n} \binom{n}{k} \frac{(-1)^j}{j!} x^j.
\]

(3)

We will call Gillis and Even’s result the “generalized derangement result.” For this work, we
want to obtain a further generalization to the generalized derangement result to the case where not
necessarily all elements of an initial list are included in a rearrangement. Finding this generalization
would allow us to model a system in which ligands can exist both on and off receptor sites.

The primary problem we need to solve is as follows:

We have \( n_1 \) elements of type 1, \( n_2 \) elements of type 2, \ldots, and \( n_R \) elements of type \( R \),
all of which are arranged in an initial list. All elements are then removed from the list.
2 Partial Generalized Derangements

What is the number of ways that we can choose and arrange \(k_1 \leq n_1\) elements of type \(1\), \(k_2 \leq n_2\) elements of type \(2\), ..., and \(k_R \leq n_R\) elements of type \(R\) such that none of the elements has the same position as it has in the original list?

We call the answer to this question the “partial generalized derangement result,” given that we are considering derangements of partial collections of the total set of elements with repeats. The resulting quantity will be termed \(B_{n,k}\) where \(n = (n_1, n_2, \ldots, n_R)\) and \(k = (k_1, k_2, \ldots, k_R)\), and we will obtain an explicit expression for it by reasoning according to the principle of inclusion and exclusion.

To apply the principle of inclusion and exclusion in the desired case, it is helpful to first review it in the simpler case of Eq.(1). With the summation index \(j\) denoting the number of elements that are fixed in their original positions, the factor \(\binom{N}{j}\) is the number of ways to choose \(j\) fixed elements out of \(N\) possible elements. The factor \((-1)^j\) is the common principle of inclusion and exclusion factor that leads sets of “correct position” elements to be alternately subtracted from and added to the first term of \(N!\) which is a count of all permutations. The factor \((N - j)!\) counts the number of ways to arrange the remaining elements given that \(j\) are fixed in their original positions. The end result after summing over all \(j\) is a count of only permutations that do not include any elements in their original positions.

Thus, there are three essential factors in the summand of Eq.(1): The counting of the number of ways to arrange elements in their original position; the principle of inclusion and exclusion factor \((-1)^j\) for each such original-position element; and the factor that counts the number of ways to arrange the remaining elements.

We can define analogous factors for \(B_{n,k}\) and use them to write a summation expression for the quantity. The result is

\[
B_{n,k} = \sum_{j_1=0}^{k_1} \cdots \sum_{j_R=0}^{k_R} \binom{n_1}{j_1} \cdots \binom{n_R}{j_R} (-1)^{j_1 + \cdots + j_R} \frac{(n_1 - j_1 + \cdots + n_R - j_R)!}{(k_1 - j_1 + \cdots + k_R - j_R)!} \\
\times \binom{k_1 - j_1 + \cdots + k_R - j_R}{(k_1 - j_1)!(k_R - j_R)!}.
\] (4)

To understand Eq.(4) we consider how each factor in the summand contributes to the final expression. The factor \(\binom{n_i}{j_i}\), for \(i = 1, \ldots, R\), is the number of ways to fill \(j_i\) out of the \(n_i\) positions of type \(i\) with their original elements. The factor \((-1)^{j_1 + \cdots + j_R}\) is the net principle of inclusion and exclusion factor for the \(j_i\) elements of type \(i\) (for \(i\) running from 1 to \(R\)) that are in their original positions. After fixing these positions with their original elements, there are now \(n_1 - j_1 + \cdots + n_R - j_R\) possible positions which we must fill with \(k_1 - j_1 + \cdots + k_R - j_R\) elements. The number of ways to choose which of these remaining positions to fill is represented by a binomial factor. The last factor \(\binom{k_1 - j_1 + \cdots + k_R - j_R}{(k_1 - j_1)!(k_R - j_R)!}\) is the number of ways to permute the \(k_1 - j_1 + \cdots + k_R - j_R\) elements amongst the chosen positions divided by factors to correct for the fact that elements of the same type are identical.

To affirm correctness, we can perform some sanity checks on Eq.(4) to ensure that this result is consistent with related ones.
Figure 2: Microstate of partial generalized derangements. In a partial generalized derangement, elements occur in multiple copies and partially occupy deranged positions in a list. The figure shows a microstate for the system with $R = 3$, $n_1 = 6$, $n_2 = 3$, and $n_3 = 5$ with squares, triangles, and circles associated with 1, 2, and 3 respectively. There are $k_1 = 3$, $k_2 = 2$, and $k_3 = 2$ elements of the various types in contact with the lattice. Some of the elements on the lattice are in deranged positions and some are in correct positions. Taking $m = (m_1, m_2, m_3)$ to define the vector counting the number of elements of each type in correct positions, we have $m_1 = 2$, $m_2 = 1$, and $m_3 = 1$. Given these correct positions, the microstate in this figure contributes to the count for $B_{n-m,k-m}$.

For what follows, it will be most useful to express Eq.(4) as an integral expression. To do so, we introduce the generalized Laguerre polynomial:

$$L_n^{(\alpha)}(x) = \sum_{j=0}^{n} \binom{n+\alpha}{n-j} \frac{(-1)^j}{j!} x^j.$$  \hspace{1cm} (5)

Using Eq.(5) and the definition of the Gamma function, we find that Eq.(4) can be written as

$$B_{n,k} = \frac{1}{(\sum_i \alpha_i)!} \int_0^\infty dx \, e^{-x} \prod_{i=1}^{R} (-1)^{k_i} x^{\alpha_i} L_k^{(\alpha_i)}(x),$$ \hspace{1cm} (6)

where we defined

$$\alpha_i \equiv n_i - k_i$$ \hspace{1cm} (7)

For the first sanity check, we expect that Eq.(6) should reduce to Eq.(2) when we take $k_j = n_j$ for all $j$. Namely when we are considering the full (rather than a partial) set of elements, the partial generalized derangement result should reduce to the generalized derangement result. Imposing this equality on Eq.(6) and noting that $L_j \equiv L_j^{(0)}$, we indeed find that Eq.(2) is reproduced.

One can show (as was done in the appendix of [Wil19]) that if we have $R$ different elements each of which is associated with a particular site out of $R$ lattice sites, then the number of ways select $K \leq R$ elements to arrange amongst the lattice sites such that none is in its associated site is

$$b_{R,K} = \sum_{J=0}^{R} (-1)^J \binom{R}{J} \binom{R-J}{K-J}^2 (K-J)!.$$ \hspace{1cm} (8)
2 Partial Generalized Derangements

Thus, we should be able to show that Eq.(6) reduces to Eq.(8) under the right conditions. In particular if we take \( n = (1, 1, \ldots, 1) \equiv n_0 \) (i.e., we have \( R \) unique elements, each of a single copy-number), then the vector \( k \) in Eq.(6) can only have elements of 1 or 0, and thus \( k \) defines a particular subset of the total set of elements. \( B_{n_0, k} \) then represents the number of ways to completely derange a particular collection of unique elements where the collection is defined by the vector \( k \). In order to find the total number of ways to completely derange \( K \) total elements (i.e., what is represented in Eq.(8)), we need to sum \( B_{n_0, k} \) over all possible values of \( k \) such that \( \sum_j k_j = K \). Thus the consistency check we must make is

\[
\sum_{k_1=0}^{1} \cdots \sum_{k_R=0}^{1} B_{n_0, k} \delta(K, k_1 + \cdots + k_R) = b_{R, K}. \tag{9}
\]

It takes more work to demonstrate Eq.(9) (see Appendix B.1), but doing so affirms that Eq.(6) is consistent with its simpler manifestations.

As a final consistency check, we note that there should be a summation condition for the total number of ways to order \( N \) unique elements is also the number of ways to select \( m \) fixed elements and derange the rest summed over all possible values of \( m \). It is straightforward to check that Eq.(1) satisfies Eq.(10).

Towards finding an analogous summation condition for \( B_{n, k} \), we note that \( \left( \begin{array}{c} n_1 \\ m_1 \end{array} \right) \cdots \left( \begin{array}{c} n_R \\ m_R \end{array} \right) B_{n-m, k-m} \) is the number of ways to choose \( m_i \) out of \( n_i \) positions (for \( i = 1, \ldots, R \)) to contain their original elements while the remaining \( k_i - m_i \) elements are completely deranged with respect to the \( n_i - m_i \) remaining original positions of type \( j \). If we sum this quantity over all possible values of \( m_i \), as in

\[
I_{n, k} \equiv \sum_{m_1=0}^{n_1} \cdots \sum_{m_R=0}^{n_R} \left( \begin{array}{c} n_1 \\ m_1 \end{array} \right) \cdots \left( \begin{array}{c} n_R \\ m_R \end{array} \right) B_{n-m, k-m}. \tag{11}
\]

we should obtain the number of ways to arrange (and not necessarily derange) \( k_i \leq n_i \) elements for \( i = 1, \ldots, R \) across a total of \( n_1 + \cdots + n_R \) lattice sites.

Calculating this quantity another way, we note that (including filled and empty sites) we are technically trying to order a total of \( n_1 + \cdots + n_R \) sites: There are \( k_1 + \cdots + k_R \) filled sites and \( n_1 - k_1 + \cdots + n_R - k_R \) empty sites. Consequently there are \( (n_1 + \cdots + n_R)! \) ways to order the total collection. Given that the filled-site elements occur in multiple copies, we must correct for equivalent orderings by dividing this count by \( k_j! \) for each element type. Also, since the empty sites act as an extra ”type” of element, we must also divide the count by \( (n_1 - k_1 + \cdots + n_R - k_R)! \), the number of ways to reorder these empty sites. Thus we should find

\[
I_{n, k} = \frac{(n_1 + \cdots + n_R)!}{k_1! \cdots k_R!(n_1 - k_1 + \cdots + n_R - k_R)!}. \tag{12}
\]
In Appendix B.2, we show that Eq.(11) produces Eq.(12).

With our combinatorial expression found and consistency affirmed, we can now work towards building the partition function for the system.

3 General Partition Function

We recall that our objective is to study the equilibrium thermodynamics of the physical system depicted in Fig. 1. The system is one where a fixed set of ligands can exist as bound or unbound to a collection of receptors. When a ligand is bound to a receptor, it can be bound either to an optimal receptor or to a suboptimal receptor. To study the thermodynamics of such a system, we needed to compute a combinatorial factor that counts the number of ways ligands can be bound to receptor sites where some of these bindings are suboptimal. Having computed this quantity in Sec. 2, we can now use what we found to calculate the partition function.

Say that we have $R$ different types of ligands. A ligand type is labeled as $i$ for $i = 1, \ldots, R$. The ligand of type $i$ has $n_i$ copies in the system, and each ligand can either be bound to a receptor or is free to move in the space surrounding the receptor sites. There are $n_1 + \cdots + n_R$ receptors to which each ligand can bind to any one of them, but for each ligand of type $i$ there are $n_i$ receptors to which it binds most strongly with a binding energy of $-\Delta_i$ (with $\Delta_i \geq 0$) relative to its binding energy for other receptors. We call such bindings “optimal” or “correct.”

When a ligand of type $i$ is not bound to a receptor, it has the single-particle partition function $Q_i^F$. When a ligand of type $i$ is bound to a receptor, but not to its optimal receptor, it has the single-particle partition function $Q_i^B$. When a ligand of type $i$ is bound to its optimal receptor, it has the single particle partition function $Q_i^{F_{iR}}$ for a system at temperature $T$.

We can now express the system’s partition function as a summation over the possible number of bound and optimally-bound ligands of each type. We have the expression

$$Z_n = \sum_k \sum_m B_{n-m,k-m} \prod_{i=1}^{R} \left( \frac{n_i}{m_i} \right) e^{\beta m_i \Delta_i} (Q_i^B)^{k_i} (Q_i^{F_{iR}})^{n_i-k_i} (n_i-k_i)!,$$  \hspace{1cm} (13)

where we defined

$$\sum_k \equiv \prod_{i=1}^{R} \sum_{k_i=0}^{n_i} \quad \sum_m \equiv \prod_{i=1}^{R} \sum_{m_i=0}^{k_i}.$$

In the summands of Eq.(13), $k_i \leq n_i$ represents the total number of ligands of type $i$ that are bound to receptors, and $m_i \leq k_i$ represents the total number of ligands of type $i$ that are optimally bound to receptors.

The factor $B_{n-m,k-m}$ (with $B_{n,k}$ defined in Eq.(4)) represents the number of ways to select and arrange $k_i - m_i$ ligands of type $i$ (for $i = 1, \ldots, R$) across a total set of $n_i - m_i + \cdots + n_R - m_R$ receptors such that no ligand of type $i$ is bound to one of its $n_i - m_i$ optimal receptors. The factors $\left( \frac{n_1}{m_1} \right) \cdots \left( \frac{n_R}{m_R} \right)$ count the number of ways to choose $m_i$ receptors from the $n_i$ possible receptors for $i = 1, \ldots, R$ to be occupied by optimal-binding partner ligands. The exponent $e^{\beta \Delta_1 + \cdots + \beta \Delta_R}$ is the combined Boltzmann factor for all optimally-bound ligands. The factor $(Q_i^B)^{k_i} \cdots (Q_R^B)^{k_R}$ is the combined Boltzmann factor for all bound ligands. The factor $(Q_i^{F_{iR}})^{n_i-k_i}/(n_i-k_i)! \cdots (Q_R^{F_{iR}})^{n_R-k_R}/(n_R-k_R)!$
3 General Partition Function

is the combined Boltzmann factor for all unbound ligands. As is the case for free particle partition functions, we divide \((Q^F_i)^{n_i-k_i}\) by a factorial to correct for equivalent permutations of particle positions in space.

For notational simplicity, we will define some additional constants. We define

\[
c_n \equiv \prod_{i=1}^{R} (Q^F_i)^{n_i}, \quad \delta_i \equiv e^{\beta \Delta_i}, \quad \gamma_i = \frac{Q^B_i}{Q^F_i},
\]

thus giving us

\[
Z_n = c_n \sum_k \sum_m B_{n-m,k-m} \prod_{i=1}^{R} \left( \frac{n_i}{m_i} \right) \frac{1}{(n_i - k_i)!} \delta_i^{m_i} \gamma_i^{k_i}.
\]

Given that pre-factors do not affect physical predictions in canonical partition functions, Eq.(16) reveals that it is only the ratios of our single-particle partition functions that are thermodynamically relevant. This result makes sense given that only free-energy differences (i.e., logarithms of partition function ratios) should affect the physics of a system. Thus, without loss of generality, we can impose \(Q^F_i = 1\) for all \(i\) under the assumption that the thermal dependence of each \(Q^F_i\) can be absorbed into a redefinition of \(\gamma_i\) and \(\delta_i\) with no change in the physical implications of Eq.(16). With this imposition we have \(c_n = 1\).

Moving forward, we recognize that the partition function becomes more analytically useful to us if we can replace the discrete summation with an integral\(^1\). To do so we make use of the integral form of \(B_{n,k}\) in Eq.(6) and a few Laguerre polynomial identities. After some work (see Appendix C), we obtain

\[
Z_n(\delta, \gamma) = \frac{1}{2\pi i} \oint_{\Gamma} \frac{dz}{z} \int_0^\infty dx \exp \left[ \mathcal{F}_n(z, x; \delta, \gamma) \right],
\]

where \(\Gamma\) is a closed contour about the origin in the complex plane and

\[
\mathcal{F}_n(z, x; \delta, \gamma) = z - x + \sum_{j=1}^{R} \ln \left[ (\gamma_j (\delta_j - 1))^{n_j} L_{n_j} \left( \frac{x (z \gamma_j + 1)}{z \gamma_j (1 - \delta_j)} \right) \right],
\]

with \(L_n(x)\) the \(n\)th Laguerre polynomial. Eq.(17) provides the starting point for our thermal equilibrium analysis. But first we derive expressions for the order parameters written in terms of this partition function.

From Eq.(16), we can derive expressions for the two main observables of the system. The average number of bound ligands and the average number of optimally-bound ligands are, respectively,

\[
\langle k \rangle = \sum_{i=1}^{R} (k_i) = \sum_{i=1}^{R} \gamma_i \frac{\partial}{\partial \gamma_i} \ln Z_n, \quad \langle m \rangle = \sum_{i=1}^{R} \langle m_i \rangle = \sum_{i=1}^{R} \delta_i \frac{\partial}{\partial \delta_i} \ln Z_n.
\]

We can use the second equation in Eq.(19) to write an alternative expression for \(\langle m \rangle\). For the

---

\(^1\)Integrals, unlike discrete summations, are more amenable to the methods of analysis.
3.1 Gendered Dimer-System Assembly

Figure 3: System for gendered dimer assembly: The general system studied in this work for the case of $n_i = 1$. For this case, there is one copy of each particle species and each particle has a one-to-one correspondence with an optimal binding site. Also, the binding affinities and optimal-binding affinity advantages for each particle varies with the particle species. Consequently, this case is a slight generalization of the case of dimer assembly with fixed binding sites treated in [Wil19].

function $f_n(x; q) = (q - 1)^n L_n(x/(1-q))$ we can use the identity $uL^m_n(u) = mL_m(u) - mL_{m-1}(u)$ to prove $\partial_q f_n(x; q) = nf_n(x; q)$. We can then show

$$\frac{\partial}{\partial \delta_i} \left[ (\gamma_i(\delta_i - 1))^n_i L_n_i \left( \frac{x(z\gamma_i + 1)}{z\gamma_i(1-\delta_i)} \right) \right] = n_i \gamma_i(\delta_i - 1)^{n_i-1} L_{n_i-1} \left( \frac{x(z\gamma_i + 1)}{z\gamma_i(1-\delta_i)} \right).$$ (20)

Thus from Eq.(19), we have

$$\langle m \rangle = \sum_{j=1}^{K} n_j \gamma_j \frac{Z_{n_j}(\delta, \gamma)}{Z_n(\delta, \gamma)},$$ (21)

where $n_j$ is $n$ with 1 subtracted from the $j$th component: $n_j = (n_1, \ldots, n_j - 1, \ldots, n_K)$. There is no analogous simplified expression for $\langle k \rangle$.

As is common for partition functions written as integrals, approximating the partition function by the maximum (or, in the case of complex values, stationary) value of its integrand allows us to derive more tractable expressions for the equilibrium conditions. Before we pursue these conditions, we will show how Eq.(17) is a generalization of a result established in a previous paper, by using a special case of the former to derive the equilibrium equations of the latter. The purpose of establishing such a generalization is to extrapolate some of the physical results explored in the paper to this more complex case.

3.1 Gendered Dimer-System Assembly

In the appendix of [Wil19], we considered “gendered dimer assembly.” We recall that this refers to a system where there are two types of particles and where a particle of one type can only bind to a particle of the other type. By fixing the positions of all particles of one type, we were able to apply the results to the case of particles binding to a lattice of possible sites (e.g., ligands binding to receptors). We assumed each particle type had a single copy and that all particles had the same binding affinities to the lattice and the same optimal-binding affinities to their correct sites.

In this section we use the general expression Eq.(17) to derive a slight generalization of these past results. In [Wil19] we assumed a global binding affinity and optimal-binding affinity for all types
of particles, but here we will assume that particles’ binding affinities and optimal-binding affinities vary according to particle type. In effect, in Eq.(17) we will take \( n_i = 1 \) for all \( i \) but retain the index dependence of \( \delta_i \) and \( \gamma_i \). Since gendered dimer assembly (with one of the “genders” fixed in space) is a more specific case of the ligand-receptor binding system considered in this work, we should find that the equilibrium equations derived for this special case of Eq.(17) match those found for the gendered dimer system in [Wil19].

First, imposing the condition \( n_i = 1 \) for all \( i \) on Eq.(17), we find the partition function

\[
Z_R(\delta, \gamma) = \frac{1}{2\pi i} \oint \frac{dz}{z} \int_0^{\infty} dx \exp [F_R(z, x; \delta, \gamma)], \tag{22}
\]

where

\[
F_R(z, x; \delta, \gamma) = z - x + \sum_{i=1}^{R} \ln \left( \gamma_j (\delta_j - 1 + x) + \frac{x}{z} \right). \tag{23}
\]

The corresponding average number of bound particles and average number of optimally bound particles for a particle of type \( j \) are the same as what is given in Eq.(19):

\[
\langle k_j \rangle = \gamma_j \frac{\partial}{\partial \gamma_j} \ln Z_R, \quad \langle m_j \rangle = \delta_j \frac{\partial}{\partial \delta_j} \ln Z_R. \tag{24}
\]

Applying the large \( N \) integral approximation (specifically \( R \gg 1 \) in this case) to Eq.(22) yields

\[
Z_R(\delta, \gamma) \simeq \frac{1}{(2\pi \det H)^{1/2}} \exp \left[ F_R(\bar{z}, \bar{x}; \delta, \gamma) \right], \tag{25}
\]

where \( H \) is the hessian matrix with second-order derivative components \( H_{\alpha,\beta} = \frac{\partial^2 F_R}{\partial \alpha \partial \beta} \bigg|_{z=\bar{z}, x=\bar{x}} \) \((\alpha, \beta \in \{x, z\})\) and \( \bar{z} \) and \( \bar{x} \) are defined by the conditions

\[
0 = \frac{\partial_x F_R(z, x; \delta, \gamma)}{z = \bar{z}, x = \bar{x}}; \quad 0 = \frac{\partial_z F_R(z, x; \delta, \gamma)}{z = \bar{z}, x = \bar{x}}. \tag{26}
\]

In order to compute equilibrium conditions for \( \langle k_j \rangle \) and \( \langle m_j \rangle \) from Eq.(25) we first need the conditions for \( \bar{x} \) and \( \bar{z}^2 \). Using Eq.(23) and Eq.(26) to find these conditions, we have, from \( \partial_x F_R = 0 \) and \( \partial_z F_R = 0 \), respectively,

\[
1 = \sum_{j=1}^{R} \frac{\bar{z} \gamma_j + 1}{\bar{x} + \bar{z} \gamma_j (\delta_j - 1 + \bar{x})}, \quad \bar{z} = \sum_{j=1}^{R} \frac{1}{\bar{x} + \bar{z} \gamma_j (\delta_j - 1 + \bar{x})}. \tag{27}
\]

Next, computing \( \langle k_j \rangle \) and \( \langle m_j \rangle \), we have

\[
\langle k_j \rangle = \gamma_j \frac{\partial}{\partial \gamma_j} F_R = \frac{\bar{z} \gamma_j (\delta_j - 1 + \bar{x})}{\bar{z} \gamma_j (\delta_j - 1 + \bar{x}) + \bar{x}}. \tag{28}
\]
3.1 Gendered Dimer-System Assembly

\[
\langle m_j \rangle = \delta_j \frac{\partial}{\partial \delta_j} F_R = \frac{\bar{\varepsilon} \gamma_j \delta_j}{\bar{\varepsilon} \gamma_j (\delta_j - 1 + \bar{x}) + \bar{x}} \tag{29}
\]

where, in applying Eq.(24) to Eq.(25), we neglected the exponential pre-factor in the latter since it is subleading in the \( R \gg 1 \) limit.

Using Eq.(27) to eliminate the \( \bar{x} \) and \( \bar{\varepsilon} \) from Eq.(28) and Eq.(29) (see Appendix D), we find the coupled equilibrium conditions

\[
\sum_{j=1}^{R} \frac{1}{\gamma_j} \left( \langle k_j \rangle - \langle m_j \rangle (1 - \delta_j^{-1}) \right) = \left( N - \langle k \rangle \right)^2 \tag{30}
\]

\[
\sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1} = \frac{\langle k \rangle - \langle m \rangle + \sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1}}{N - \langle m \rangle + \sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1}}. \tag{31}
\]

Eq.(30) and Eq.(31) define how the average number of bound and optimally-bound particles for each species \( j \) vary with one another and with the parameters for binding affinity \( \gamma_j \) and optimal-binding affinity \( \delta_j \). For practical purposes, when trying to solve this system of equations it is necessary to first solve Eq.(27) and then insert the obtained values of \( \bar{x} \) and \( \bar{\varepsilon} \) into Eq.(28) and Eq.(29) to find \( \langle k_j \rangle \) and \( \langle m_j \rangle \). But Eq.(30) and Eq.(31) do provide an affirming pathway to more familiar results. If we take \( \delta_j = \delta \) and \( \gamma_j = \gamma \) for all \( j \), and note that \( \langle k \rangle = \sum_{j=1}^{R} \langle k_j \rangle \) (and similarly for \( \langle m_j \rangle \)), we find

\[
\frac{1}{\gamma} \left( \langle k \rangle - \langle m \rangle (1 - \delta^{-1}) \right) = \left( N - \langle k \rangle \right)^2, \quad \langle m \rangle \delta^{-1} = \frac{\langle k \rangle - \langle m \rangle (1 - \delta^{-1})}{N - \langle m \rangle (1 - \delta^{-1})}. \tag{32}
\]

where we took the total number of particles (or, equivalently, the total number of binding sites) to be \( N \equiv R \). The results in Eq.(32) are the very same ones we found in [Wil19] for the gendered dimer system. By imposing the condition \( \langle k \rangle = \langle m \rangle \) on the second equation in Eq.(32), we can show that

\[
\langle k \rangle = \langle m \rangle = \frac{N - 1}{1 - \delta^{-1}}, \tag{33}
\]

suggesting that the condition \( \langle k \rangle = \langle m \rangle \) only occurs when essentially all the particles are bound to their optimal binding sites. Inserting this value for \( \langle k \rangle \) and \( \langle m \rangle \) into the second equation of Eq.(32) yields the thermal condition under which this fully optimal binding configuration occurs. We find

\[
N - 1 = \gamma \delta \frac{(1 - N \delta^{-1})^2}{1 - \delta^{-1}}. \tag{34}
\]

In [Wil19], we used Eq.(34) to infer the existence of generally two types of binding systems with quite different relationships between \( \langle k \rangle \) and \( \langle m \rangle \). When \( \delta \gg \gamma > 1 \), Eq.(34) became \( \gamma \delta \simeq N \) and we had a “search-limited” system in which optimal binding was primarily limited by the ability of particles to find their optimal binding site in the surrounding volume; when \( \gamma \gg \delta > 1 \), Eq.(34) became \( \delta \simeq N \) and we had a “combinatorics-limited” system in which optimal binding was primarily limited by the ability of particles to avoid the combinatorial sea of suboptimal ones.

As we increased the temperature in search-limited systems, the value of \( \langle m \rangle \) remained close to
4 Large $N$ Limits of Special and General Cases

the value of $\langle k \rangle$ thus indicating that such systems could have partial binding to sites but with all such bindings being optimal. Conversely, in combinatorics-limited systems, increasing the temperature led to the value of $\langle m \rangle$ being much lower than the value of $\langle k \rangle$ indicating that when particles were bound, such binding was likely suboptimal. With some heuristic arguments, we suggested that biophysical systems are more likely to be of the search-limited type, but such an inference was limited by the simplicity of our model.

In this work, we want to extend the analysis in this simpler case to one where there are multiple particle types of various copy number and various binding and optimal-binding affinities. For this general case, the objective is to find a condition (if only approximate) akin to Eq.(34) that will allow us to distinguish various binding behaviors in the system and thus tell us if our previous combinatorics-limited and search-limited framings still apply. Due to its incorporation of multiple-copy number and type-dependent binding affinities, this more general case will be more biophysically relevant and could thus serve as a firmer basis for categorizing biophysical systems as one of the two types.

But before we consider this most general case, we consider two more specific cases to build the intuition and techniques for how binding and combinatorics affect ligand-receptor systems.

4 Large $N$ Limits of Special and General Cases

In studying the system modeled by Eq.(22), we will first work through two special cases that establish the intuition and methods we will apply to the most general case. The first special case is that of $\delta_i = 1$ for all $i$. This is the case of ligands binding to receptor sites with no binding preference for the site but with each receptor type having distinct binding affinities to the set of receptors. The second special case is that of $\gamma_i \to \infty$ corresponding to a system where ligands can only exist as attached to a receptor and where the various microstates consist of derangements of the ligands amongst the set of receptors. This latter case is a generalization of the permutation glass studied in [Wil18] to the case of repeated components. With these two cases established, we will then consider the general case with no prior assumptions on the values of $\delta_i$ and $\gamma_i$.

4.1 Simple Binding Model

One simplification of the most general scenario associated with Eq.(17) is to have each ligand type have the same binding affinity regardless of to which receptor it binds. This simplification amounts to taking $\delta_i = 1$ (or $\Delta_i = 0$ by Eq.(15)). Phrased differently, this condition implies that, for a single type, the optimally and suboptimally bound ligand partition functions are equal, and thus there is no thermal advantage for a ligand to be bound to any particular receptor; from the perspective a single type of ligand, all receptors are thermodynamically identical. However, the ligands of different types are not thermodynamically identical to each other. Since $\gamma_i$ is not presumed to be the same for all $i$, each ligand type has a different binding affinity to an arbitrary receptor. Thus, our system contains many different ligands in the presence of many identical receptors.

The thermal implications of this condition can be found by computing the partition function.
4.1 Simple Binding Model

Figure 4: Two limiting cases for the general model. In (b), we have the case of $\delta_i = 1$. For this case, all receptors are equivalent and a distinct ligand species (denoted with a distinct color or border in the figure) of type $i$ has a binding affinity of $\gamma_i$. In (c), we have the case where $\gamma_i \to \infty$. For this case, all ligands are bound to the lattice, and the system microstates consist of the various ways to permute a list with repeated elements. The top figure shows the “correct microstate” (where each ligand is in its optimal binding site), and the bottom figure shows a “deranged microstate” (where each ligand is in a suboptimal binding site). The most general case of ligand-receptor binding is a combination of these two cases.

Taking $\delta_i \to 1$ in Eq.(17) we can define the new partition function

$$W_n(\gamma) \equiv \lim_{\delta_i \to 1} \frac{Z_n(\delta, \gamma)}{\prod_i n_i!}$$

where we divide by the multinomial coefficient $(\sum_i n_i)!/\prod_i n_i!$ to account for the fact that, in contrast to the assumptions that underly Eq.(17), we assume here that all $\sum_i n_i$ receptors are thermodynamically identical. Using the limit identity $\lim_{\alpha \to 0} \alpha^m L_m \left( \frac{x}{\alpha} \right) = (-1)^m x^m / m!$, the definition Eq.(35) gives us

$$W_n(\gamma) = \frac{1}{2\pi i} \oint \frac{dz}{z} e^z \prod_{i=1}^R \left( \gamma_i + \frac{1}{z} \right)^{n_i}.$$  

(36)

We could also have obtained Eq.(36) without use of the limiting case by starting from the expression

$$W_n(\gamma) = \prod_i \frac{n_i!}{(\sum_i n_i)!} \sum_k I_{n,k} \prod_{j=1}^R \frac{\gamma_i^{k_j}}{(n_i - k_i)!} = \sum_k \frac{1}{(\sum_i n_i - k_i)!} \prod_{j=1}^R \binom{n_j}{k_j} \gamma_j^{k_j},$$

(37)

where each sum over $k_j$ runs from 0 to $n_j$, and $I_{n,k}$ (defined in Eq.(12)) is the number of ways to arrange $k_i$ objects of type $i$, for $i = 1, \ldots, R$ amongst $n_1 + n_2 + \cdots + n_R$ positions. We again divide by the multinomial pre-factor $(\sum_i n_i)!/\prod_i n_i!$ since these positions are all equivalent.

Using Eq.(19) and Eq.(37), we find that the average number of bound ligands in the system can be written as

$$\langle k \rangle = \sum_{j=1}^R n_j \gamma_j \frac{W_{n_j}(\gamma)}{W_n(\gamma)},$$

(38)

where $n_j$ is $n$ with 1 subtracted from the $j$th component: $n_j = (n_1, \ldots, n_j - 1, \ldots, n_R)$. Eq.(38)
provides us with a reliable means for computing the order parameter presuming we have a reliable
means for computing the partition function. For our use case, we will use the large $N$ saddle-point
approximation as the basis for this latter computation.

First, defining

$$A_n(z; \gamma) \equiv z + \sum_{j=1}^{R} n_j \ln \left( \gamma_j + \frac{1}{z} \right),$$

and then applying the saddle point approximation to $W_n$ defined in Eq.(36), we obtain the approximate partition function

$$W_n(\gamma) = \frac{1}{2\pi i} \oint \frac{dz}{z} \exp \left[ A_n(z; \gamma) \right] \simeq \frac{1}{(2\pi \bar{z}^2 A''_n(\bar{z}; \gamma))^{1/2}} \exp \left[ A_n(\bar{z}; \gamma) \right],$$

where $\bar{z}$ is defined by the constraint $0 = \partial_z A_n(z; \gamma)|_{z=\bar{z}} \equiv A'_n(z; \gamma)|_{z=\bar{z}}$. Eq.(40) is an approximation, but we henceforth use an equality symbol for notational sparsity. Using Eq.(39) to compute $\bar{z}$ given its constraint definition, we find

$$\bar{z} = \sum_{j=1}^{R} \frac{n_j}{\gamma_j \bar{z} + 1}.$$  

Next, we can identify the $j$th term in the sum Eq.(38) with the average number of bound ligands of type $j$: $\langle k_j \rangle$. With this identification and Eq.(41), we see that $\langle k_j \rangle$ can be approximated as

$$\langle k_j \rangle = \frac{n_j \gamma_j \bar{z}}{\gamma_j \bar{z} + 1}.$$  

Eq.(42) and Eq.(41) provide a means for approximately determining the number of bound ligands of a specific type. Given the list of binding affinities $\gamma = (\gamma_1, \ldots, \gamma_R)$ for our ligands, we can numerically solve Eq.(41) and insert the result into Eq.(42). However, there is a different perspective on these results that connects them to the way binding is typically represented in chemistry. Using our system of equations to solve for $\bar{z}$ in two ways, we have

$$\bar{z} = N - \langle k \rangle, \quad \bar{z}^2 = \sum_{j=1}^{R} \langle k_j \rangle \gamma_j^{-1},$$

where $N = \sum_{j=1}^{R} n_j$. The first equation is found by subtracting Eq.(42) from $n_j$ and summing over $j$. The second equation is found by dividing Eq.(42) by $\gamma_j$ and summing over $j$. Eliminating $\bar{z}$ from Eq.(43), we then have

$$(N - \langle k \rangle)^2 = \sum_{j=1}^{R} \langle k_j \rangle \gamma_j^{-1},$$

which is the law of mass action for a system of ligands with a list of binding affinities $\gamma = (\gamma_1, \ldots, \gamma_R)$. We could have anticipated this result: When $\delta_i = 1$, the combinatorics of various binding configurations becomes irrelevant since all re-orderings of the same set of bound ligands are thermally equivalent, and thus the equilibrium properties should be governable by averages constrained by
4.2 Derangement Model

The order of our approximation does not allow for $\langle k \rangle = N$ (i.e., the case where all ligands are bound), but it does allow for $\langle k \rangle = N - 1$, a state in which essentially all ligands are attached to a receptor. With our definition of $\bar{z}$ in Eq.(43) this almost-complete binding state is defined by $\bar{z} = 1$. Taking $N \gg 1$ (as is the case in our large $N$ approximation), the thermal condition that defines this almost-completely bound state is

$$1 = \sum_{j=1}^{R} n_j \gamma_j^{-1} + O(\gamma^{-2}).$$

(45)

Given a temperature-dependence for $\gamma_j$, we can numerically solve Eq.(45) for the temperature at which essentially all of the ligands are bound to receptor sites. We will do so in Sec. 5.1 when we simulate this system.

4.2 Derangement Model

Another simplification of the most general scenario associated with Eq.(17) is to consider it as purely a combinatorial one in which the only available microstates are those consisting of permutations of ligand positions amongst the receptor sites. This can only occur if the ligand binding affinity is so large that the bound-ligand partition function is infinitely larger that the corresponding unbound-ligand partition function. Quantitatively, by Eq.(15), this amounts to taking $\gamma_i \to \infty$. In such a case all ligands are bound to a receptor and thermal fluctuations only lead the ligands to switching receptors.

Taking the partition function Eq.(17) to the $\gamma_i \to \infty$ limit and dividing out the thermodynamically irrelevant factors, we can define

$$X_n(\delta) \equiv \lim_{\gamma_i \to \infty} \frac{1}{\prod_{i=1}^{R} \gamma_i^{n_i}} Z_n(\delta, \gamma),$$

(46)

Computing the limit gives us

$$X_n(\delta) = \int_0^{\infty} dx \, e^{-x} \prod_{i=1}^{R} (\delta_i - 1)^{n_i} L_{n_i} \left( \frac{x}{1-\delta_i} \right),$$

(47)

where $L_n(x)$ is the $n$th Laguerre polynomial. In defining Eq.(46), we note that the $\gamma_i$ are thermodynamically irrelevant for a system consisting only of bound ligands and thus their factors can be divided out of the partition function.

We could have derived Eq.(46) without a limiting case by recognizing that the various microstates of this system are "derangements" of a list. Specifically, we could have written $X_n$ as a summation over these derangements:

$$X_n(\delta) = \sum_{m} G_{n-m} \prod_{j=1}^{R} \left( \frac{n_j}{m_j} \right) \delta_j^{m_j},$$

(48)
4 Large $N$ Limits of Special and General Cases

where $G_n$ is defined in Eq. (2).

Also, Eq. (46) is an "elements with repeats" generalization of the permutation glass considered in [Wil18]. If we take $n_i = 1$ for all $i$ in the product in Eq. (46), we find

$$X_n(\delta)|_{n_i=1} = \int_0^\infty dx \, e^{-x} \prod_{i=1}^{N} (\delta_i - 1 + x),$$

(49)

which, with $\delta_i = e^{\beta \Delta_i}$, is identical to the partition function derived in [Wil18]. In that work, we derived necessary but not sufficient conditions for the system to settle into the "completely correct" microstate, which in our case corresponds to all ligands being bound in their optimal receptors. Here we attempt to derive analogous conditions for this more general case.

Using Eq. (19) and Eq. (48), we find that the average number of ligands bound to their optimal receptors is

$$\langle m \rangle = \sum_{j=1}^{R} n_j \delta_j \frac{X_{n_j}(\delta)}{X_n(\delta)},$$

(50)

where $n_j$ is $n$ with 1 subtracted from the $j$th component: $n_j = (n_1, \ldots, n_j - 1, \ldots, n_R)$. Eq. (50) tells us that if we have a consistent means for computing the partition function $X_n(\delta)$, we can calculate the order parameter with little extra work. For this system, the consistent means we have for computing the partition function is the large $N$ approximation.

Defining

$$F_n(x; \delta) \equiv -\sum_{j=1}^{R} \ln \left[ (\delta_j - 1)^{n_j} L_{n_j} \left( \frac{x}{1 - \delta_j} \right) \right].$$

(51)

And applying Laplace’s method to $X_n$ defined in Eq. (47), we have the approximation

$$X_n(\delta) = \int_0^\infty dx \, \exp \left[ -F_n(x; \{\delta_i\}) \right] \approx \left( \frac{2\pi}{F_n''(\bar{x}; \delta)} \right)^{1/2} \exp \left[ -F_n(\bar{x}; \delta) \right]$$

(52)

where $\bar{x}$ is defined by the constraint

$$0 = \partial_x F_n(x; \{\delta_i\})|_{x=\bar{x}} \equiv F_n'(x; \{\delta_i\})|_{x=\bar{x}}.$$  

Eq. (52) is an approximation, but we henceforth use an equality symbol for notational sparsity. Applying this constraint to Eq. (51) and using the recursive Laguerre identity $uL_n'(u) = n(L_n(u) - L_{n-1}(u))$, we find that $\bar{x}$ is defined implicitly by

$$\bar{x} = \sum_{j=1}^{R} n_j \left( 1 - \frac{L_{n_j-1}(\bar{\sigma}_j)}{L_{n_j}(\bar{\sigma}_j)} \right); \quad \bar{\sigma}_j \equiv \frac{\bar{x}}{1 - \delta_j},$$

(53)

Using Eq. (50) with Eq. (52) (and the fact that we are in the $N \gg 1$ limit), we see that the average number of optimal bindings is

$$\langle m \rangle = \sum_{j=1}^{R} \frac{n_j \delta_j}{\delta_j - 1} \frac{L_{n_j-1}(\bar{\sigma}_j)}{L_{n_j}(\bar{\sigma}_j)}.$$  

(54)

In the original paper, we defined our microstates in terms of energy penalties rather than energy benefits so the partition function here differs from the original by a multiplicative constant.
4.2 Derangement Model

With Eq. (54), we can determine whether a microstate consisting of all ligands bound to their optimal receptors can be achieved in this system. Finding the condition that makes such a microstate possible would require us to look at the low temperature behavior of the system, but let’s momentarily go in the opposite direction and ask what is the behavior of Eq. (54) when \( T \to \infty \)? First, as \( T \to \infty \), \( \delta_i \) goes to 1. Given the definition of the Laguerre polynomial, we can derive the limit
\[
\lim_{\alpha \to 0} \alpha^m L_m(x/\alpha) = (-1)^m x^m / m!.
\]
Using these limits, we find for \( T \to \infty \) that
\[
\langle m \rangle \to \sum_{j=1}^R n_j^2 / \bar{x}
\]
and \( \bar{x} \to \sum_{j=1}^R n_j \). Therefore,
\[
\lim_{T \to \infty} \langle m \rangle = \frac{\sum_{j=1}^R n_j^2}{\sum_{j=1}^R n_j}.
\] (55)

A naive entropic argument might lead us to presume that there would be no optimal bindings at infinite temperature (i.e., a physical regime where the energy advantage of optimal bindings is irrelevant) since the macrostate of completely deranged ligands (i.e., \( \langle m \rangle \simeq 0 \)) would supposedly have the largest number of microstates and thus the largest entropy. However, Eq. (55) suggests that the macrostate for fully deranged bindings is in fact not entropically favored at infinite temperature, and thus that such a macrostate takes up less configuration space than seemingly more ordered and constrained macrostates. Indeed, when you have various types of ligands each of which occurs in large numbers, then it becomes more constraining to require no ligand to be in its optimal binding site than it is to have some ligands be optimally bound.

We can get a better sense of the meaning Eq. (55) by writing it in terms of a variance rather than a second moment. Defining \( \bar{n} = \sum_{j=1}^R n_j / R \) and \( \bar{n}^2 = \sum_{j=1}^R n_j^2 / R \) we have
\[
\lim_{T \to \infty} \langle m \rangle = \bar{n} \left( 1 + \frac{\sigma_n^2}{\bar{n}^2} \right),
\] (56)

The quantity \( \bar{n} \) is the average particle-number across all types of ligands and \( \sigma_n^2 = \bar{n}^2 - \bar{n}^2 \) is the variance in particle-number. What Eq. (56) shows is that the greater the variability in the number of ligand-copies of each type, the larger the lower limit on the number of optimally bound ligands. With more copy-number variability, it becomes more likely that at least some ligands, just from random assorting, will be bound to their optimal receptors.

Now, we consider the opposite temperature limit under the frame of a specific question: At what temperature are all of the ligands bound to their optimal receptors? To answer this, we first use Eq. (53) and Eq. (54) to obtain the identity
\[
\bar{x} = N - \sum_{j=1}^R \langle m_j \rangle (1 - \delta_j^{-1}),
\] (57)

where \( N = \sum_{j=1}^R n_j \) and \( \langle m_j \rangle \) are the elements of the sum in Eq. (54). When all ligands are bound to their optimal receptors, we have \( \langle m_j \rangle = n_j \). Thus, at this desired critical temperature, we have the condition
\[
\bar{x} = \sum_{j=1}^R n_j \delta_j^{-1}.
\] (58)

19
4 Large $N$ Limits of Special and General Cases

To move forward, we will make two assumptions whose consistency we will check at the end of the calculation. First, we assume that the desired temperature is sufficiently low that $\delta_i \gg 1$ for all $i$. Second, we assume that $\delta_i \gg \bar{x}$. With these assumptions, we find $L_{n_j-1} (\bar{\sigma}_j) / L_{n_j} (\bar{\sigma}_j) = 1 - \bar{x} \delta_j^{-1} + O(\delta_j^{-2})$. Expanding Eq.(54) to first order in $\delta_j^{-1}$ yields

$$\langle m \rangle = N + (1 - \bar{x}) \sum_{j=1}^{R} n_j \delta_j^{-1} + O(\delta^{-2}),$$

where $O(\delta^{-2}) \equiv \sum_{j=1}^{R} O(\delta_j^{-2})$ and $O(\delta_j^{-2})$ represents terms of order $\delta_j^{-2}$. At the temperature at which all ligands are optimally bound, we have $\langle m \rangle = N$. Thus, Eq.(59) implies $\bar{x} = 1 + O(\delta_j^{-2})$ and by Eq.(58) we obtain the final condition

$$1 = \sum_{j=1}^{R} n_j \delta_j^{-1} + O(\delta^{-2}).$$

(60)

Given the temperature dependence for $\delta_j$, we can numerically solve Eq.(60) for the temperature at which all of the ligands are optimally bound to receptor sites. We will do so in Sec. 5.2 when we simulate this system. To check consistency with our two initial assumptions (i.e., $\delta_j \gg 1$ and $\delta_j \gg \bar{x}$), we note that, for the large particle-number limit, Eq.(60) implies $\delta_j > n_j \gg 1 = \bar{x} + O(\delta_j^{-2})$.

4.3 General Case

Having explored various limiting cases, we are now ready for the full case. In this section, our objective is two fold: First, determine the equations for both the average number of bound and optimally-bound ligands of each type; second, use these equations to determine the thermal conditions that define the system settling into the microstate in which each ligand is bound to its optimal receptor (i.e., the fully optimally bound). To get to either objective, we first need to approximate the partition function and compute the standard observables (Eq.(19)) according to this approximation. We will apply methods similar to those applied to the limiting cases to analyze this general case.

Applying the saddle point approximation to Eq.(17), we have

$$Z_n(\delta, \gamma) = \frac{1}{2\pi i} \oint \frac{dz}{z} \int_0^\infty dx \exp \left[ F_n(z, x; \delta, \gamma) \right] \approx \frac{1}{(\bar{z}^2 \det H)^{1/2}} \exp \left[ F_n(\bar{z}, \bar{x}; \delta, \gamma) \right]$$

(61)

where

$$F_n(\bar{z}, \bar{x}; \delta, \gamma) = \bar{z} - \bar{x} + \sum_{j=1}^{R} \ln \left( \gamma_j (\delta_j - 1) \right)^{n_j} L_{n_j} \left( \bar{\sigma}_j \right); \quad \bar{\sigma}_j = 1 - \bar{\sigma}_j \left( 1 + \frac{1}{\bar{z} \gamma_j} \right).$$

(62)

The quantity $H$ is the complex hessian matrix of $F_n$

$$H \equiv \partial_{\alpha, \beta} F_n(z, x; \delta, \gamma) \big|_{x, z = \bar{x}, \bar{z}}.$$  

(63)
where the variables $\alpha$ and $\beta$ can be $x$ or $z$. The critical points $\bar{z}$ and $\bar{x}$ are defined by the conditions

$$0 = \partial_z F_n(z, x; \delta, \gamma) \bigg|_{z=x=\bar{z}, \bar{z}}, \quad 0 = \partial_x F_n(z, x; \delta, \gamma) \bigg|_{x=\bar{x}, \bar{z}}. \quad (64)$$

Applying the critical point conditions to Eq.(62), we can derive

$$\bar{z} = \sum_{j=1}^{R} \frac{n_j}{\bar{z} \gamma_j + 1} \left(1 - \frac{L_{n_j-1} (\bar{\phi}_j)}{L_{n_j} (\phi_j)}\right), \quad \bar{x} = \sum_{j=1}^{R} n_j \left(1 - \frac{L_{n_j-1} (\bar{\phi}_j)}{L_{n_j} (\phi_j)}\right). \quad (65)$$

The associated values of $\langle k \rangle$ and $\langle m \rangle$ can then be found by applying Eq.(19) to the approximated partition function Eq.(61) while ignoring the subleading pre-factor. We obtain

$$\langle k_j \rangle = \frac{n_j}{\bar{z} \gamma_j + 1} \left(\bar{z} \gamma_j + \frac{L_{n_j-1} (\bar{\phi}_j)}{L_{n_j} (\phi_j)}\right), \quad \langle m_j \rangle = \frac{n_j \delta_j}{\delta_j - 1} \frac{L_{n_j-1} (\bar{\phi}_j)}{L_{n_j} (\phi_j)}. \quad (66)$$

With Eq.(66), we can use the solutions for $\bar{x}$ and $\bar{z}$ determined from Eq.(65) to find the average number of bound and correctly bound ligands of each type, thus fulfilling the first objective.

For the second objective of determining the thermal conditions for fully optimal ligand-receptor binding, we first use Eq.(65) and Eq.(66) together to obtain two equations relating the four quantities:

$$\bar{z} = N - \langle k \rangle, \quad \bar{x} = N - \sum_{j=1}^{R} \langle m_j \rangle (1 - \delta_j^{-1}), \quad (67)$$

where we took $N = \sum_{j=1}^{R} n_j$ and $\langle k \rangle = \sum_j \langle k_j \rangle$.

In Sec. 3.1, we showed that in the gendered dimer assembly system, the correct assembly condition yielded the result $\langle k \rangle = \langle m \rangle = (N - 1)/(1 - \delta^{-1}) = N - 1 + O(\delta^{-1})$. With this result, we were then able to find the thermal condition that defined the fully optimized state. We want to do something similar for this more general case.

We start by defining the state of fully optimal ligand-receptor binding in a way analogous to the definition in Eq.(33): We assert that the system is in the state where all ligands are optimally bound to receptors when $\langle k \rangle$ and $\langle m \rangle$ satisfy

$$\langle k \rangle = \langle m \rangle = N - 1 + O(\delta^{-1}), \quad (68)$$

where $O(\delta^{-1}) = \sum_{j=1}^{R} O(\delta_j^{-1})$ and $O(\delta^{-1})$ represents terms of order $\delta_j^{-1}$. To find the thermal condition that defines fully optimal binding, we need to find the thermal condition that is consistent with Eq.(68). Applying Eq.(68) to Eq.(67) and Eq.(66), we find

$$1 = \sum_{j=1}^{R} n_j \delta_j^{-1} (1 + \gamma_j^{-1}) + O(\delta^{-2}), \quad (69)$$

where $O(\delta^{-2}) = \sum_{j=1}^{R} O(\delta_j^{-2})$ (see Appendix E). Eq.(69) is the general thermal condition that defines fully optimal binding. Given the temperature dependences of $\delta_j$ and $\gamma_j$, we can use Eq.(69)
5 Simulation Comparison

Consider a two-dimensional grid of square lattice sites each of which can be filled with various colored squares. The colored squares represent the ligands of the system with a specific color defining a ligand type, and the lattice sites are the receptors. Each color-type binds optimally to a particular collection of lattice sites. For pictorial convenience we can arrange the optimal lattice sites for each particle type such that a figure is created. This way it is obvious whether our system is in the fully optimally bound configuration. We depict this system in Fig. 5.

As a clarifying point, the model we developed for ligands binding to receptors applies equally well to a one-dimensional chain as to an $n$-dimensional grid as long as both are finite. This is because coordinates on a finite grid can be mapped one-to-one to a finite list such that a fixed collection of objects exploring various positions in the multi-dimensional grid is equivalent to those objects being placed in various orderings in a list.

In what follows, we use this graphical lattice model to present simulation results for the two limiting cases and the general case discussed in Sec. 4.
5.1 Simple Binding Model Simulation \((\delta_i = 1)\)

In this section, we affirm the theoretical results in Sec. 4.1 by simulating a simplification of the grid system in Fig. 5 at various temperatures. The simplification is to assume particles have no "optimal" position on the grid (i.e., \(\delta_i = 1\)) and thus a single particle has the same binding affinity to every site on the grid. The system was simulated using the Metropolis Hastings algorithm in which the microstates transitioned into one another dependent on free energy differences of the form

\[
\beta E = \sum_j k_j \ln \gamma_j,
\]

where \(k_j\) is the number of bound particles of type \(j\) in the microstate and \(\gamma_j\) is the associated binding affinity. We allowed for two types of microstate transitions: particle binding to the grid and particle dissociation from the grid. As is typical for Metropolis Hastings algorithms, to fully determine the transition probabilities we also had to incorporate the difference in probabilities of selecting the particles for the forward and reverse transitions between microstates (See Appendix A for a description of a more general simulation system and Supplementary Code in Sec. 11 for associated code).

To incorporate an explicit temperature dependence into the system, we set \(\gamma_j = (\beta E_V)^{3/2} e^{\beta E_j}\) where \(\beta = 1/k_B T\). The quantity \(E_V\) represents the volume-based energy of a free particle in the system (e.g., \(E_V \equiv \hbar^2/2\pi m V^{2/3}\) for an ideal gas particle of mass \(m\)) and thus \((\beta E_V)^{3/2}\) represents the ratio between the kinetic partition functions for free and bound particles. The quantity \(E_j\) rep-
5 Simulation Comparison

represents the binding energy of a particle of type \( j \). For simplicity, we did not assume a \( j \) dependence for \( E_V \).

Taking \( k_B T_{\text{bind}} = \beta_{\text{bind}}^{-1} \) to be the critical temperature at which the complete (or, more precisely, "almost-complete") binding state is achieved, Eq.(45) thus became

\[
1 = (\beta E_V)^{-3/2} \sum_{j=1}^{R} n_j e^{-\beta_{\text{bind}} E_j} + O(e^{-2\beta_{\text{bind}} E_j}),
\]

(S70)

Solving Eq.(70) for \( k_B T_{\text{bind}} \) gives us the temperature at which the thermal advantage of each particle binding to the grid (at any site) is large enough to overcome the entropic disadvantage of the particle existing statically in the grid rather than freely in the volume. In a sense, Eq.(70) defines the thermal condition under which all particles are able to search for and successfully find the grid in the space they occupy. This "searching" is encoded by the product of the \( V \) and \( n_j \) factors in the equation: As \( V \) (defined in \( E_V \)) and \( n_j \) increase, the volume in which a particle must search increases and the number of particles doing the searching increases, respectively. Both increases make it more difficult for the system to settle into a state in which all particles are bound: Increasing volume increases the space in which particles must search for the grid; increasing the number of particles increases the number of units that need to conduct this search successfully. Thus increasing either of these values makes settling into the complete binding state more difficult, that is, unless the temperature is lowered sufficiently so that the binding energy is strong enough to overcome the entropic disadvantage of having the particles exist freely. Thus it is only below \( k_B T_{\text{bind}} \), that the searching entropy succumbs to the energy advantage and the system settles into its full binding state. We employ this spatial search metaphor to distinguish this system from one grounded in a combinatorial search of possible states. We discuss this latter system in the next section.

To simulate the system, we chose numerical values for all parameters. For simplicity, we took all energy parameters in the system to be dimensionless. The values of \( E_j \) defining \( \gamma_j \) were sampled from a Gaussian distribution \( N(\mu_E, \sigma_E^2) \) with mean \( \mu_E = 6.0 \) and variance \( \sigma_E = 2.0 \). The value of \( E_V \) was set to \( E_V = 10^{-3} \). The values of \( n_j \) were determined directly from Fig. 5: Inspecting the count of squares for each of the \( R = 8 \) colors and taking each color to be a particle type, we have \( n = (9, 9, 10, 5, 7, 6, 3, 51) \).

In Fig. 6a, we show the simulated grid at various equilibrium temperatures. The particles are colored squares where particle-type is distinguished by color. Particles not bound to the grid are not shown. The values of \( k_B T \) are dimensionless because we are taking the energy parameters of the system to be dimensionless. In the (i) image of Fig. 6a, we see that all particles are bound although they are not in their "correct positions" as defined by the fully optimally bound state in Fig. 5. This is of course because, with \( \delta_i = 1 \), there is no thermal advantage to being in such entropically limited positions. As the temperature increases, fewer particles occupy the grid which confirms our intuition that the system melts at higher temperatures.

In Fig. 6b, we plot the theoretical temperature-dependence of \( \langle k \rangle = \sum_{j=1}^{R} k_j \) against the simulated temperature-dependence. We mark the points in the curve that are associated with the grid depictions in Fig. 6a. The temperature computed from Eq.(70) is denoted as \( k_B T_{\text{bind}} \). We see excellent agreement between the simulation results and the theoretical results. Moreover, the predicted
5.2 Derangement Model Simulation ($\gamma_i \to \infty$)

In this section, we affirm the theoretical results in Sec. 4.2 by simulating a simplification of the grid system in Fig. 5 at various temperatures. The simplification is to consider the system for the case in which all particles remained on the grid (i.e., $\gamma_j \to \infty$) and where state transitions are confined to particles exchanging positions within one another. The system was simulated using the Metropolis Hastings algorithm where microstates transitioned into one another contingent on free energy differences of the form $\beta E = \sum_j m_j \ln \delta_j$, where $m_j$ is the number of optimally bound particles of type $j$ in the microstate and $\delta_j$ is the additional binding affinity factor for optimal binding. We allowed for only one type of transition: single-step permutations of particle positions (See Appendix A for a description of a more general simulation system and Supplementary Code in Sec. 11 for associated code).

To incorporate temperature into the system, we returned to our original expression for $\delta_j$ in Eq.(19): $\delta_j = e^{\beta \Delta_j}$. We recall that $\Delta_j$ is the energy-advantage an optimal binding has over any other binding for a ligand of type $j$. The associated critical temperature at which all particles were optimally bound was defined as $k_B T_{\text{derang}}$. Taking $k_B T_{\text{derang}} = \beta_{\text{derang}}^{-1}$, Eq.(60) yields

$$1 = \sum_{j=1}^R n_j e^{-\beta_{\text{derang}} \Delta_j} + O(e^{-2\beta_{\text{derang}} \Delta}).$$

Solving Eq.(71) for $k_B T_{\text{derang}}$ gives us the temperature at which the thermal advantage of each particle settling into its optimal site is large enough to overcome the entropic disadvantage of choosing that site in the space of all other combinatorial possibilities. In a way, Eq.(71) defines a limit on how influential combinatorics can while still being able to achieve this optimal state, where the influence of “combinatorics” is encoded by $\mathbf{n} = (n_1, n_2, \ldots, n_R)$: As $n_j$ increases, the number of possible combinatorial states in the system increases and thus it becomes more difficult for a ligand to thermally select the optimal site in a sea of suboptimal ones, unless the temperature is lowered to diminish how much the entropy influences the free energy. It is only at or below $k_B T_{\text{derang}}$ that combinatorial entropy succumbs to the energy-advantage of optimal sites, and the system settles into its fully optimal state.

To simulate the system, we chose numerical values for all parameters. The values of $\Delta_j$ were sampled from a Gaussian distribution $N(\mu_\Delta, \sigma_\Delta^2)$ with mean $\mu_\Delta = 4.0$ and variance $\sigma_\Delta = 2.0$; energy parameters were taken to be dimensionless. The values of $n_j$ were determined directly from Fig. 5: Taking each particle to represent a particle type, we have $\mathbf{n} = (9, 9, 10, 5, 7, 6, 3, 51)$.

In Fig. 7a, we show the simulated grid at various equilibrium temperatures. The particles are colored squares where particle-type is distinguished by color. The values of $k_B T$ are dimensionless because we are taking the energy parameters of the system to be dimensionless. In the (i) image of Fig. 7a, we see that all particles are bound in their “correct positions” as defined by the fully
5 Simulation Comparison

Figure 7: Grid snapshots and theory vs. simulation for derangement model Sec. 4.2: We used $\delta_j = e^{\beta \Delta_j}$ where $\Delta_j$ was sampled from a normal distribution $N(\mu, \sigma^2)$ with $\mu = 4.0$ and $\sigma = 2.0$. All energy parameters were taken to be dimensionless. In (a) we see snapshots of the simulated system at various equilibrium temperatures. As temperature increases, the system becomes more "deranged," meaning particles are less likely to assume their optimal binding sites. In (b) we plot a theory vs. simulation comparison for $\langle m \rangle$ as a function of temperature, and we mark the binding temperature $k_B T_{\text{derang}}$ computed from Eq.(71). The "Exact" values of $\langle m \rangle$ were computed from Eq.(50). The "Large $N$" values of $\langle m \rangle$ were computed from Eq.(54). The "Simulation" values of $\langle m \rangle$ were computed from the results of a Metropolis Hastings algorithm where each point is the average of the result of five simulations. We see that Eq.(71) indeed identifies the temperature at which the system is in its fully optimally bound state and that both Eq.(50) and Eq.(54) accurately model the simulation. (See Supplementary Code in 11 for link to code repository used to produce this figure.)

As the temperature increases, the particles become increasingly "deranged" from their correct positions, though we note that even at high temperatures some particles (in particular the ones with large $n_j$) do maintain many of their correct positions. This latter result is consistent with the discussion surrounding Eq.(56).

In Fig. 7b, we plot the theoretical temperature-dependence of $\langle m \rangle = \sum_{j=1}^{R} \langle m_j \rangle$ against the simulated temperature-dependence. In particular we compare the simulations to the "Exact" theoretical prediction defined in Eq.(50) and the "Large $N$" theoretical prediction defined in Eq.(54). We mark the points in the curve that are associated with the grid depictions in Fig. 7a. The temperature computed from Eq.(71) is denoted as $k_B T_{\text{derang}}$. We see excellent agreement between the simulation results and the theoretical results. Moreover, the predicted temperature computed from Eq.(71) accords with the results of the simulation. Inspecting (ii) in Fig. 6a, as we expect, beyond the critical temperature the grid begins to show deranged particle states.

Having explored the two limiting cases of the general model, we are now prepared to consider the fully general case. We will proceed as we did in these two example sections: Starting with a theoretical analysis stemming from an approximation and then finally simulating our results. The objective is to obtain a condition similar to Eq.(70) and Eq.(71) that implicitly defines the temperature at which the fully optimally bound state is achieved. We turn to this task in the next section.
In this section, we simulate the system outlined in Sec. 4.3 for the lattice grid depicted in Fig. 5.

To incorporate temperature into the system we took $\delta_j = e^{\beta \Delta_j}$ and $\gamma_j = (\beta E_V)^{3/2}e^{\beta E_j}$, where $\Delta_j$ is the energy advantage for the particle of type $j$ binding to its optimal lattice site. The quantity $E_V$ represents the volume-based energy of a free particle in the system (e.g., $E_V \equiv h^2/2\pi mV^2/3$ for an ideal gas particle of mass $m$) and thus $(\beta E_V)^{3/2}$ represents the ratio between the kinetic partition functions for free and bound particles. The quantity $E_j$ is the “base-binding energy” of the particle of type $j$ to any position on the lattice. For example, if each particle of type $j$ only had a non-zero binding affinity for its optimal site, we would have $E_j = 0$ and $\Delta_j > 0$ for all $j$. With these thermal dependences, and taking $k_B T_{\text{crit}} = \beta_{\text{crit}}^{-1}$ to be the critical temperature at which the system achieves the fully optimally bound state, Eq.(69) becomes

$$1 = \sum_{j=1}^{R} n_j e^{-\beta_{\text{crit}} \Delta_j} \left( 1 + (\beta_{\text{crit}} E_V)^{-3/2} e^{-\beta_{\text{crit}} E_j} \right) + O(e^{-2\beta_{\text{crit}} \Delta_j}).$$

Comparing Eq.(72) to Eq.(70) and Eq.(71), it appears that the first is a combination of the latter two. Moreover, given our “search” and “combinatorics” interpretation of Eq.(70) and Eq.(71) respectively, it appears that Eq.(72) embodies aspects of both limits contingent on various relative values of the parameters. In the next section, we will explore these relationships further.

To simulate the system, we chose numerical values for all parameters. The values of $\Delta_j$ were sampled from a Gaussian distribution $N(\mu_\Delta, \sigma_\Delta^2)$ with mean $\mu_\Delta = 3.0$ and variance $\sigma_\Delta = 1.0$. The values of $E_j$ were sampled from a Gaussian distribution $N(\mu_E, \sigma_E^2)$ with mean $\mu_E = 12.0$ and variance $\sigma_E = 4.0$. The value of $E_V$ was set to $E_V = 10^{-3}$. The values of $n_j$ were determined directly from Fig. 5: Specifically, taking each color to represent a particle type, we had $n = (9, 9, 10, 5, 7, 6, 3, 51)$

In Fig. 8, we display the results of simulated thermodynamics for our grid assembly system for this general case. For such equilibrium simulations, we can choose whatever state transitions we like as long as detailed balance is satisfied, that is, as long as the ratios between the forward and reverse transitions are equivalent to the ratios of the Boltzmann factors between the final and initial states [Kra06]. Therefore when simulating the equilibrium behavior of the system, we chose state transitions which led to an efficient-in-time exploration of the state space even if such transitions were unphysical. We included three state-transitions: particle binding (i.e., ligand association to a receptor), particle unbinding (i.e., ligand dissociation from a receptor), and binding permutation (i.e., bound ligands switching receptor sites). The binding permutation transition does not occur in real biomolecular systems, but it was useful for our simulations since it ensured that the system did not remain trapped in non-equilibrated states at low temperature. Theoretically with particle binding and unbinding alone, the system should always find its true equilibrium eventually, but consistently finding such an equilibrium on the finite time scales of realistic simulations is difficult. Thus the principal effect of including the binding permutation transition is to reduce the time needed to simulate these systems.

In Fig. 8a, we show the simulated grid at various equilibrium temperatures. In the (i) image of
5 Simulation Comparison

Figure 8: Grid snapshots and theory vs. simulation for general ligand-receptor binding from Sec. 4.3: We used $\delta_j = e^{\beta \Delta_j}$ and $\gamma_j = (\beta E_V)^{3/2} e^{\beta E_j}$ where $\Delta_j$ was sampled from a normal distribution $N(\mu_\Delta, \sigma_\Delta^2)$ with $\mu_\Delta = 3.0$ and $\sigma_\Delta = 1.0$, $E_V = 10^{-3}$, and $E_j$ was sampled from a normal distribution $N(\mu_E, \sigma_E^2)$ with $\mu_E = 14.0$ and $\sigma_E = 2.0$. All energy parameters were taken to be dimensionless.

In (a), we have snapshots of the simulated system at various temperatures. Particles not bound to the grid are not shown. In (b) we plot a theory vs. simulation comparison for both $\langle k \rangle$ and $\langle m \rangle$ as functions of temperature. The critical temperature computed from Eq.(72) is denoted by $k_B T_{\text{crit}}$.

The theoretical values of $\langle k \rangle$ and $\langle m \rangle$ were computed from Eq.(66) and summed over $j$. The “Simulation” values of $\langle k \rangle$ and $\langle m \rangle$ were computed from the results of a Metropolis Hastings algorithm where each point is the average of the result of five simulations. We see that Eq.(72) identifies the temperature at which the system is in the fully optimally bound state and that the theory curves well match the simulation results. Thus the validity of the large-number limit solutions is affirmed in this case. We note that the $\langle m \rangle$ curve falls away faster than the $\langle k \rangle$ curve which is a consequence of the specific parameter choices we made. In Sec. 6 we explore how different choices could lead to different behaviors. (See Supplementary Code in 11 for link to code repository used to produce this figure.)

Fig. 8a, we see that all particles are bound in their “optimal positions” as defined by the fully optimally bound state in Fig. 5. As the temperature increases, fewer particles occupy the grid and fewer of the particles which occupy the grid are in their optimal binding sites which is consistent with our intuition that the system should lose both binding and combinatorial order as the temperature increases.

In Fig. 8b, we plot the theoretical temperature-dependences of $\langle k \rangle = \sum_{j=1}^R \langle k_j \rangle$ and $\langle m \rangle = \sum_{j=1}^R \langle m_j \rangle$ against their simulated temperature-dependences. We mark the points in the curve that are associated with the grid depictions in Fig. 8a. The temperature computed from Eq.(72) (i.e., the temperature at which fully optimal binding is achieved) is denoted as $k_B T_{\text{crit}}$. We see excellent agreement between the simulation results and the theoretical results. Moreover, the computed critical temperature is in accordance with the results of the simulations. Inspecting (ii) in Fig. 8a (which is at a temperature above the critical temperature) the system is no longer in its fully optimally bound configuration, as we should expect.

The $\langle m \rangle$ and $\langle k \rangle$ curves depicted in Fig. 8b represent only one type of relationship between the
temperature dependences of total binding and optimal binding. In this case, we see that as we heat the system above the critical temperature, the average number of optimally bound particles falls more quickly than does the average number of total bound particles. Therefore, slightly above the critical temperature we have a grid-image such as that depicted in (ii) of Fig. 8a: particles are mostly bound but not all of them are in their optimal sites.

This thermal relationship between $\langle k \rangle$ and $\langle m \rangle$ was pre-determined by our parameter choices for $\Delta_j$, $E_j$, and $E_V$ as was the temperature computed from Eq.(72). In the discussion following Eq.(34) for the gendered dimer assembly model, we noted how such parameter choices could lead us to categorize the extremes of that assembly system as one of two types. In the next section, we will attempt to do something similar with this general ligand-receptor system.

6 Search and Combinatorics Limited Systems

In [Wil19], we found that systems of dimer assembly could often be characterized as either search-limited or combinatorics-limited contingent on the relationships between the binding parameters. Given that the system we are currently studying is a generalization of a version of gendered dimer assembly (as shown in Sec. 3.1), we can naturally ask if the ligand-receptor binding system exhibits similar divisions.

Consider the theoretical condition defining the microstate where all ligands are optimally bound to a receptor site (rewritten here from Eq.(69)):

$$1 = \sum_{j=1}^{R} n_j \delta_j^{-1} (1 + \gamma_j^{-1}) + O(\gamma^{-2}) + O(\delta^{-2}).$$

To obtain Eq.(73), we took $\delta_j \gg 1$. Therefore, all of our system-limits will necessarily be defined with the base assumption that the optimal ligand-receptor binding affinity is large. But even with this base assumption, we can still find two important limits that give us different approximations for the critical temperature.

Assume first that $\gamma_j \gg 1$ for all $j$ at this critical temperature. Then $1/\gamma_j$ is sub-dominant in the second term in the parentheses of Eq.(73), and we have the approximation

$$1 = \sum_{j=1}^{R} n_j \delta_j^{-1} + O(\gamma^{-1}) + O(\delta^{-2}).$$

[Combinatorics Limiting Condition] (74)

Recalling that $\gamma_j$ is the base-binding affinity of a type $j$ ligand to any receptor, we know that if $\gamma_j \gg 1$ for all $j$, then all ligands have a sufficiently strong binding to any receptor that ligand-receptor binding will occur even at low temperature. Thus, whether the system consists entirely of optimal bindings is primarily determined by whether $\delta_j$ is strong enough to bias such optimal bindings over their combinatorial disadvantage. We call Eq.(74) a “combinatorics-limiting condition” (akin to Eq.(60)) for fully optimal binding since the achievement of fully optimal binding is limited by the influence of the combinatorial properties of the system.

On the other hand, assume that $\gamma_j \ll 1$ (with $\gamma_j \delta_j \gg 1$) for all $j$ at the critical temperature. Then
6 Search and Combinatorics Limited Systems

$1/\gamma_j$ is dominant in the second term in the parentheses for Eq. (69) and we have the approximation

$$1 = \sum_{j=1}^{R} n_j \delta_j^{-1} \gamma_j^{-1} + O(\gamma^0) + O(\delta^{-2}). \quad \text{[Search Limiting Condition]} \tag{75}$$

In this case, the product $\delta_j \gamma_j$ represents the net-binding affinity for an unbound ligand to not merely bind to any receptor but to specifically bind to its optimal receptor. A small value of $\gamma_j$ means that unbound ligands are not attracted to suboptimal receptors, and thus without the additional optimal-binding affinity factor $\delta_j$, ligands would generally not bind at all. In particular, since $\gamma_j \ll 1$ but $\delta_j \gamma_j \gg 1$ the optimal binding affinity is already strong enough to overcome the combinatorial disadvantage of such bindings. Thus achieving the fully optimally bound state is not limited by the influence of combinatorics. Instead, we call Eq. (75) a “search-limiting condition” (akin to Eq. (45)) to highlight the fact that for this case, achieving the fully optimally bound state is primarily limited by ligands’ abilities to search for their optimal receptor sites in the volume they occupy.

Given temperature dependences for $\delta_j$ and $\gamma_j$, we can compute the temperatures corresponding to the approximations Eq. (74) and Eq. (75). We term these temperatures, respectively, $T_{\text{comb}}$ and $T_{\text{search}}$. Using the temperature dependences which yielded Eq. (72), we find Eq. (74) and Eq. (75) become, respectively,

$$1 = \sum_{j=1}^{R} n_j e^{-\beta_{\text{comb}} \Delta_j}, \quad 1 = (\beta_{\text{search}} E V)^{-3/2} \sum_{j=1}^{R} n_j e^{-\beta_{\text{search}} (\Delta_j + E_j)} \tag{76}$$

where $k_B T_{\text{search}} = \beta_{\text{search}}^{-1}$ and similarly for $k_B T_{\text{comb}}$, and we dropped sub-leading terms for notational simplicity. We can use the equations in Eq. (76) to establish upper bounds on the true critical temperature $T_{\text{crit}}$: From the temperature dependences in each equation and comparing Eq. (72), it is straightforward to show

$$T_{\text{crit}} < T_{\text{comb}}, \quad T_{\text{search}}. \tag{77}$$

Why is it important to define these limiting cases at all and then compute temperatures from them? Because the relative values of the temperatures are associated with qualitatively different behaviors for the $\langle k \rangle$ and $\langle m \rangle$ curves, and thus computing these temperatures can immediately provide us with a sense of how the difference $\langle k \rangle - \langle m \rangle$ varies with temperature.

We show this in Fig. 9. In each figure, we plot simulation vs theory curves for $\langle k \rangle$ and $\langle m \rangle$ (akin to that displayed in Fig. 8b), for various parameter distributions of $\gamma_i$ and $\delta_i$. We see that depending on these distributions, $k_B T_{\text{comb}}$ and $k_B T_{\text{search}}$ have different relative values and these relative values can in turn be used to infer properties of the relationship between $\langle k \rangle$ and $\langle m \rangle$. Specifically, if $T_{\text{comb}} < T_{\text{search}}$, then above the critical temperature, the difference $\langle k \rangle - \langle m \rangle$ grows quickly indicating that a significant fraction of ligands can be bound suboptimally to the set of receptors (Fig. 9a). Conversely if $T_{\text{comb}} > T_{\text{search}}$ then above the critical temperature, the difference $\langle k \rangle - \langle m \rangle$ is small, indicating that even when the system consists of only partially bound ligands, most of these ligands are attached to their optimal receptor sites (Fig. 9c).

We can also use these temperatures to approximately define when a system is either search-
Figure 9: Search-limited, indeterminate, and combinatorics-limited systems: We took $\delta_j = e^{\beta \Delta_j}$ and $\gamma_j = (\beta E_V)^{3/2} e^{\beta E_j}$ where $\Delta_j$ was sampled from a normal distribution $N(\mu_\Delta, \sigma_\Delta^2)$, $E_V = 10^{-3}$ and $E_j$ was sampled from a normal distribution $N(\mu_E, \sigma_E^2)$. In (a) with $(\mu_\Delta, \sigma_\Delta) = (4.75, 2.0)$ and $(\mu_E, \sigma_E) = (16.0, 3.0)$, we have a combinatorics-limited system; In (b) with $(\mu_\Delta, \sigma_\Delta) = (6.75, 2.0)$ and $(\mu_E, \sigma_E) = (10.75, 3.0)$, we have a system of indeterminate or in-between type; In (c) with $(\mu_\Delta, \sigma_\Delta) = (7.7501, 2.0)$ and $(\mu_E, \sigma_E) = (3.0, 1.0)$, we have a search-limited system. In the combinatorics-limited system there is a significant difference between $\langle k \rangle$ and $\langle m \rangle$ above the critical temperature, indicating that such systems can have a significant fraction of suboptimally bound ligands. In the search-limited system there is little difference between $\langle k \rangle$ and $\langle m \rangle$, indicating that even when ligands in such systems are partially bound, the ligands are primarily bound optimally. (See Supplementary Code in 11 for link to code repository used to produce this figure.)

Thus, similar to what was found for dimer system self-assembly [Wil19], we have found that we can categorize the ligand-receptor system as constrained by two extremes: A search-limited extreme and a combinatorics-limited extreme. Given the results of the special cases considered in Sec. 4.1 and Sec. 4.2, we could also rename the search-limited condition and combinatorics-limited condition as binding-limited conditions and a derangement-limited conditions, respectively, the binding-limited condition defining whether particles can go from free-space to attaching to their preferred site on the grid, and the derangement-limited condition defining whether the particles attach to their preferred site relative to other sites.
Figure 10: Simulations in time of the $k_B T = 0.5$ values of $m/N$ for the systems in Fig. 9 allowing for unphysical (top-row) and physical (bottom-row) types of transitions. Each simulation began with all particles unbound from the grid. The black dashed line at $m/N = 1$ represents the expected equilibrium value that all simulations should approach (as inferred from the plots in Fig. 9). For (a), (b), and (c), we used all the same state transitions used to simulate Fig. 9, namely particle binding, particle unbinding, and particle permutation. For (d), (e), and (f), we only used the physically relevant state transitions of particle binding and unbinding. Plots (a) and (d) have the combinatorics-limited parameters used for Fig. 9a. Plots (b) and (e) have the indeterminate system parameters used for Fig. 9b. Plots (c) and (f) have the search-limited parameters used for Fig. 9c. We see that while all system types approach their expected equilibrium values when using the unphysical (but efficient-in-time) particle permutation transition, only the search-limited system reproduces the equilibrium estimate on the finite time scale of simulations when only particle binding and unbinding are allowed transitions. This suggests that, for optimal ligand-receptor binding, only search-limited systems are able to avoid kinetic traps in real non-equilibrium situations. (See Supplementary Code in 11 for link to code repository used to produce this figure.)

sion is permitted as long as forward and backward transition ratios equate to Boltzmann factor ratios), but if we want to understand realistic non-equilibrium behavior—such as how a system of initially free ligands binds over time to a collection of receptors—we can only use the physical transitions of binding and unbinding. Limiting our transition choices in this way reveals another difference between combinatorics and search-limited systems: When only using physically realistic state transitions, combinatorics-limited systems were more likely to get trapped in non-equilibrated metastable states than were search-limited systems. In particular, when we only allowed for particle binding and unbinding transitions in systems where $\gamma_j \gg 1$ and the temperature satisfied $T < T_{\text{crit}}$ (thus indicating the system should satisfy $\langle k \rangle = \langle m \rangle \approx N - 1$), the simulated system did not always find the “true equilibrium” of fully optimal binding even if analytical predictions suggested it should.

In Fig. 10, we again simulated the systems whose equilibrium properties are depicted in Fig.
except in this case we only tracked the evolution of the number of optimally bound particles, denoted \( m \), over the course of the simulation. In each case the system started from a state consisting entirely of free particles and then it evolved according to its transition properties. The black dashed line at \( m/N = 1 \) represents the approximate predicted equilibrium value of \( \langle m \rangle/N \) for the associated system type at \( k_B T = 0.5 \) (as obtained from the plots in Fig. 9). The colored continuous lines are various simulations of the system at the given temperature. For Fig. 10a, Fig. 10b, and Fig. 10c, we used the same state-transitions used to simulate Fig. 9: Particle binding, particle unbinding, and particle permutation. For Fig. 10d, Fig. 10e, and Fig. 10f, we only used the physically relevant state transitions of particle binding and unbinding. Plots Fig. 10a and Fig. 10d have the combinatorics-limited parameters used for Fig. 9a. Plots Fig. 10b and Fig. 10e have the indeterminate system parameters used for Fig. 9b. Plots Fig. 10c and Fig. 10f have the search-limited parameters used for Fig. 9c. We see that while all system types approach their “correct” equilibrium value when using the unphysical (but efficient-in-time) particle-permutation transition, only the search-limited system reproduces the equilibrium result when only particle binding and unbinding are allowed transition steps.

The physical explanation for this behavior is simple. If we begin in a state of free ligands in a system satisfying \( \gamma_j \gg 1 \), then the ligands have sufficiently strong binding to all receptors to bind to any one of them and not necessarily to their optimal receptors. Once such ligands are bound, it is unlikely they will dissociate from these receptors and then bind to their true optimal receptor because \( \gamma_j \) is so large that dissociation after binding is unlikely. This situation exists in distinction to that of a search-limited system where \( \gamma_j \ll 1 \) and thus where dissociation from suboptimal receptors is thermodynamically feasible and ligands only bind strongly (and largely irreversibly) to their optimal binding sites.

Therefore, combinatorics-limited and search-limited are not merely convenient equilibrium distinctions between systems. They also typify distinctions in realistic non-equilibrium behavior. Theoretically, for infinite time, all systems should reach their true equilibrium regardless of whatever transitions they manifest. However, achieving such a true equilibrium in finite time is made difficult when kinetic traps exist in the system. In contrast to combinatorics-limited systems, search-limited systems have virtually no kinetic-traps since strong binding only occurs when receptors bind to their optimal receptor sites which are associated with the true equilibrium. Thus, in cases where ligands need to specifically bind to certain receptors in finite time, the systems will necessarily need to be search-limited to ensure for rapid optimal binding.

This “prediction” is necessarily a soft one because we have yet to define at a parameter-level the distinction between the two principal system types. In [Wil19] we derived necessary but insufficient conditions typifying the distinction between combinatorics and search-limited systems for dimer assembly, but in this work the index-dependence of our biophysical parameters make the analogous such conditions difficult to derive.
7 Biophysical Implications

The theoretical investigations of the previous sections were motivated by the biophysics of ligand-receptor binding. Now we will consider whether our results can help us better understand this starting point.

For the system of ligand-receptor binding depicted in Fig. 1, the results Eq. (66) afford us the ability to predict $\langle k_j \rangle$ and $\langle m_j \rangle$ for various ligand species given $\gamma_j$, $\delta_j$, and $n_j$. Moreover, with Eq. (73) (and temperature dependences for the affinities), we can determine the temperature at which the system settles into the fully optimally bound configuration. Such a prediction would first require finding values for $\gamma_j$, $\delta_j$, and $n_j$. Given that $\gamma_j$ and $\delta_j$ are “effective” model parameters representing, respectively, a ligand of type $j$’s binding affinity to a non-optimal site and the binding affinity advantage to an optimal site, these values would have to be approximated from available data on ligand-receptor interactions. A simple manifestation of this approximation would amount to taking $\gamma_j$ to be the average of the binding affinities for a particular ligand $j$ to all receptors (besides the optimal one) in a system and $\delta_j$ to be the additional binding affinity factor to the ligand’s optimal receptor (i.e., $\gamma_j \delta_j$ would be ligand $j$’s absolute binding affinity to this optimal receptor). However, ligand-receptor affinities, though useful theoretical quantities for modeling, are notoriously difficult to calculate in practice [JAVH20] so obtaining accurate predictions of $k_B T_{\text{crit}}$ might be similarly challenging.

Still, we can use the theoretical properties of Eq. (73) to make qualitative statements about the properties of optimally bound systems. First, we recognize that the terms within the sum must each be less than unity. Therefore as one of the factors of a single term increases, the other factors must decrease in order to keep the total product less than unity. Thus for a system with a large value of $n_j$ for ligand $j$, we need a correspondingly large $\delta_j$ in order for the thermal constraint condition to be soluble. Conceptually, this implies that the more copies we have of a ligand-type, then the more strongly that ligand must bind to its optimal receptor in order for the entire system’s fully optimally bound configuration to be achievable.

This result is also true for the total number of ligands in the system: The larger the number of total ligands in the system, the greater the average optimal binding affinities of the ligands must be in order for the system to settle into its fully optimally bound configuration at constant temperature. We can see this by applying Jensen’s inequality to the first equation in Eq. (76) and using Eq. (77). Doing so we obtain

$$k_B T_{\text{crit}} < \frac{1}{\ln N} \sum_{j=1}^{R} n_j \Delta_j,$$

where we recall that $N \equiv \sum_{j=1}^{R} n_j$. Eq. (80) sets an upper limit on the the temperature at which all ligands settle into their optimal receptors. Since $n_j/N$ is the fraction of elements of type $j$ and $\Delta_j$ is the associated binding energy benefit for being in an optimal site, Eq. (80) shows that the critical temperature is bounded above by the weighted average of energy benefits, $\bar{\Delta} \equiv \sum_{j=1}^{R} n_j \Delta_j/N$. Moreover, if we want this bound to remain the same as we increase $N$, the average optimal binding energy must increase in tandem. Thus the more ligands we have in the system the greater we expect the average binding affinity to be, presuming the energetically optimal binding configuration is a
desired state in the system.

One could likely have guessed a result of the form in Eq. (80). Namely, it makes sense that the critical temperature should be of the same order as (if not outright bounded above by) an average binding energy in the system. However, what is perhaps surprising is that the limiting temperature scales as \( 1/ \ln N \) meaning that as the total number of ligands in the system increases, the temperature at which the low energy system is accessible decreases in tandem but does so logarithmically. This scaling seems archetypal for combinatorial statistical physics systems [Wil18, Wil19].

We could also invert this inequality and use it to establish an upper-limit on the number of ligands in the system. Assuming we know \( k_B T_{\text{crit}} \), we can impose an upper limit on \( N \) as

\[
N < e^{\beta_{\text{crit}} \bar{\Delta}},
\]

which is another way to interpret the limiting relationship between total number of ligands and average optimal binding energy. As \( \bar{\Delta} \) increases, so too does the limit on \( N \). Thus systems with larger average optimal binding energies can also admit more ligands and still achieve the fully optimally bound configuration at nearly the same temperatures. The main implication is that cells with more proteins should also have larger average binding affinities for those proteins. For example, human cells and prokaryotic cells, which differ in numbers of proteins by a few order of magnitude [Mil13], should also differ in the average optimal binding energies for such proteins. In particular if the compared collection of cells are typically found in the similar thermal environments (i.e., exist at the same \( k_B T \)) and the particle number was stringently bound by Eq. (81), we would expect the relative average values of the binding energies for human and prokaryotic ligands should be the same as the order of magnitude that differentiates their relative number of proteins.

However, the limiting effect of Eq. (81) is only relevant if the associated system is specifically combinatorics-limited, i.e., if Eq. (74) well approximates the general thermal condition Eq. (73). The results in Sec. 6.1 suggest that real biophysical systems (where optimal ligand-receptor binding is functionally important) are likely search-limited rather than combinatorics-limited since it is only in the former that fully optimal binding can be achieved through physical transitions on finite time scales. We recall that search-limited systems are those for which the energy advantage for optimal binding is sufficiently high that the only limiting factor to ligands finding their optimal site is their ability to “search” for these sites in the constituent volume. More formally, search-limited systems are roughly defined as those for which \( T_{\text{search}} \) (computed from Eq. (75)) provides a good approximation for \( T_{\text{crit}} \) (computed from Eq. (73))

Given that real ligand-receptor systems are likely search-limited, they would also likely easily satisfy the combinatorics condition Eq. (74) at their typical temperature. Therefore, the derived inequality Eq. (81) would not be a strong limit on the number of particles. Instead, a stronger limit would be found by using the search-limiting condition Eq. (75) to find a bound on particle number. Finding such a bound first requires us to determine a form for \( \gamma_j \) that is a convex function of temperature, and such a \( \gamma_j \) depends on our exact model of how free ligands and bound ligands exist in the system of interest. But as a toy-case, we can also use our grid-assembly expression for \( \gamma_j \) to
8 Discussion

get a sense of the general form of the limit. Using the grid-assembly expression for $\gamma_j$ amounts to using the second equation in Eq.(76) to define the critical temperature. The Jensen-inequality based derivation that shows how this equation limits $N$ is similar to that which leads to Eq.(81). Ultimately, we find

$$N < (\beta_{\text{crit}} E_V) \frac{3}{2} e^{\beta_{\text{crit}}(\Delta + \bar{E})},$$

where $\bar{E} \equiv \sum_{j=1}^{R} n_j E_j / N$. Thus the implication is the same as that which follows Eq.(81): Ligand-receptor systems which have a large number of ligands and where optimal binding is physically important, should also have large average binding energies for those ligands.

It is possible that the critical temperature $k_B T_{\text{crit}}$ is not a desirable temperature in a real physical system. Eq.(73) gives the thermal condition under which fully optimal ligand-receptor binding can be achieved, but there is not a universal reason for why such a bound configuration would be functionally optimal in all biomolecular systems. Perhaps the microstate where all ligands are in their optimal receptor sites is too rigid to be biophysically useful, and thus it is preferred if the system exists in a partially-bound state where most (but not all) of the bound ligands are in their optimal configuration. Thus the value in this formalism might not exist in its ability to predict the temperature at which the system is in a fully bound state, but rather in its power to predict and categorize binding properties above this temperature. For this value, the system-type distinction is the major biophysical utility of this formalism: Ligand-receptor systems are inherently combinatorial, but it seems that if nature were to evolve such systems so as to avoid the kinetic traps of suboptimal bindings, it would engineer ligands to have sufficiently high binding affinities to their optimal-receptors as to easily achieve the combinatorics-limit Eq.(74) at in vivo temperatures.

8 Discussion

We began this work with the goal of using combinatorics to model how the competition between distinct ligands affects the ligands’ equilibrium binding properties to receptors. Before developing a physical model of such binding, we needed to solve a modified version of a well known problem on derangements. After solving this problem and using it to compute a general partition function, we were able to derive implicit formulas for the average number of bound ligands and the average number of optimally bound ligands for each type (Eq.(66)) and for the condition under which all ligands bind to their optimal receptors (Eq.(73)).

With the observables defined in Eq.(66), we can compute the equilibrium binding properties of our ligand-receptor system at any temperature (provided we have thermal dependences for $\gamma_j$ and $\delta_j$ for all $j$). But there are also softer biophysical predictions from this model. Eq.(81) and Eq.(82) can be seen as such predictions, labeled as “soft” because rather than explicitly predicting that some observable has a value, they predict that, in order for a system condition to be satisfied, an observable cannot exceed a certain value. The main implication of these inequalities is that the larger the binding energy advantage for optimal contacts, the more particles the system can have and still be

\footnote{We note that the expression for $\delta_j = e^{\beta \Delta_j}$ is also associated with the toy-model, but since $\delta_j$ is defined as a ratio of binding affinities, a purely exponential representation of $\delta_j$ is arguably general. Taking $\gamma_j = (\beta E_V)^{3/2} e^{-\beta E_j}$ was primarily for analytic convenience and real systems would have a different thermal dependence for $\gamma_j$.}
capable of achieving the fully optimally bound state. Thus, systems with more ligands should also have larger optimal binding energies supposing such bindings have functional importance. Note that Eq. (82) was derived for the simple model where free-ligands were taken to be point particles, but an analogous inequality could be obtained for more physically informed values of \( \gamma_j \).

For the task of quantitatively modeling optimal ligand-receptor binding, there are many limitations to the introduced model.

First, we are only partly exploring the combinatorics of our system, a combinatorics that can be characterized by derangements from a pre-defined "correct" or optimal configuration. A more general and flexible model would not \textit{a priori} define such a configuration and would instead have an interaction matrix defining how various ligand types interact with various receptor types, and would then use this matrix to determine the optimal matchings between ligands and receptors. However, we avoided using such a matrix since doing so would limit the solubility of the partition function.

Second, again for solubility, we assumed that the system contained the same number of receptor sites as ligands. This assumption was convenient for the framework of derangements, but does not match a realistic biophysical scenario wherein the number of ligands and receptors are not necessarily equal.

Third, we didn’t discuss how the spatial organization of receptors affects the propensity of ligands to bind to them. We instead assumed all receptors were on equal footing in terms of spatial accessibility, and all binding variances could be encoded into the collection of parameters \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_R) \) and \( \delta = (\delta_1, \delta_2, \ldots, \delta_R) \). How spatial organization affects binding could likely be better modeled through a matrix-based interaction scheme where spatially occluded receptors have reduced binding affinities.

All of these more realistic additions would complicate the combinatorial simplicity of what we have presented here, so much so that an entirely new model would likely have to be erected in its place. As in all modeling, building such a new model would require us to balance the increase in relevance from an incorporation of more realistic properties with a concomitant loss in transparency from the new model’s relative insolubility.

Finally, the model we developed does not answer the implicit question posed in the introduction. In Sec. 1, we first motivated our initial steps towards the general model presented in Sec. 3 by noting that many ligand-receptor systems exhibit many-to-many interactions in which each type of ligand can bind to many receptors and vice versa. This fact led us to build a model in which even when ligands interacted optimally with a few receptors, these ligands still had the potential to interact with sub-optimal receptors. However, the results of Sec. 6.1 suggest that if optimal ligand-receptor binding (where each ligand binds to its energetically optimal receptor) is functionally important in the system, then that binding has to be highly specific to combat its combinatorial disadvantage. So although our model was motivated by multi-specific ligand binding, the model’s results primarily pertain to highly specific ligand binding and the thermal conditions needed to achieve such binding. Therefore left over from the introduction is the question of how does one quantitatively model the ligand-receptor binding underlying the combinatorial codes that make multi-specific systems so necessary in pharmacology and real biomolecular environments. Building such a model would
likely require us to make use of an interaction matrix since it is only through such a matrix that one can encode for different sets of ligands interacting optimally with different sets of receptors.

As an ending remark, we note that although Sec. 4.2 explored a limiting case to our more general model, the model introduced there can also stand alone as a probabilistic model of derangements. Say that we begin with a state space of permutations of a list with repeated elements (e.g., all of the ways to permute the characters and spaces in a sentence). We assume that there is a single permutation where the elements are said to be in their correct order. Let the parameter $w_i$ be proportional to the probability that an element of type $i$ is in its correct position. Then, the partition function we previously computed in Eq. (47) becomes the weighted sum of states

$$X_n(w) = \int_0^\infty dx \, e^{-x} \prod_{k=1}^R (w_k - 1)^{n_k} L_{n_k} \left( \frac{x}{1-w_k} \right).$$

Eq. (83) is essential to the computation of two quantities: First, the probability that $\ell_i$ elements of type $i$, for $i = 1, \ldots, R$, are in their correct positions; Second, the average number of elements that are in their correct positions. Respectively, these quantities are

$$P_\ell = \frac{G_n - \ell}{X_n(w)} \prod_{k=1}^R \binom{n_k}{\ell_k} w_k^{\ell_k}, \quad \langle \ell_{\text{tot}} \rangle = \sum_{j=1}^R n_j w_j \frac{X_{n_j}(w)}{X_n(w)},$$

where $G_n$ is the generalized derangement formula Eq. (2), and $n_j$ is $n$ with 1 subtracted from the $j$th component: $n_j = (n_1, \ldots, n_j - 1, \ldots, n_R)$. Aside from deriving them directly, we can obtain the expressions in Eq. (84) by translating the Boltzmann factor expressions in Sec. 4.2 into the language of probability weights by taking $\delta_j \rightarrow w_j$. Analogous to Eq. (80), using Jensen’s inequality, we find that a necessary, but not sufficient, condition for the average $\langle \ell_{\text{tot}} \rangle$ to be equal to $N$ (i.e., for all elements to be in their correct order) is

$$1 \leq \sum_{j=1}^R n_j \ln w_j \frac{N}{N \ln N}.$$ 

This probabilistic model of derangements would be relevant in a computing context where one is interested in the probability of various derangements that differ from a given sequence by a fixed number of elements.

9 Conclusion

We introduced a combinatorial model of ligand-receptor interactions. The model advances the subject of ligand-receptor modeling in two directions: First, it concretely frames the question of how a finite collection of ligands can compete for a finite collection of receptors. Such a question is important because ligands in real biomolecular systems always exist in crowded environments with other types of ligands, and, with the multi-specificity of ligand-receptor binding, such crowded environments ordinarily exhibit ligands struggling to bind to their optimal sites in a sea of suboptimal ones. Second, to better organize the counting of the system’s microstates, the model introduces a combi-
natorial problem (and its solution) as the foundational framework for such systems. This approach to building biomolecular models by beginning with combinatorial questions appears to be generalizable to similar contexts since biomolecular systems often contain finite numbers of particles where particles can only interact in precise ways that can often be defined by some combinatorial set.

10 Acknowledgements

The author thanks Rostam Razban for helpful comments on the work.

11 Supplementary Code

The code used to generate all figures is found in the repository https://github.com/mowillia/LigandReceptor.

A Simulations

To simulate our system and create the plots in Fig. 6b, Fig. 7b, Fig. 8b, and Fig. 9 we implemented a Metropolis Hastings algorithm [Kra06]. The code for recreating these figures is linked to in Sec. 11. Here we review the salient parts of the implementation.

For these simulations, we needed to define a microstate, the probability of transitions between microstates, and the types of transitions between microstates.

A.1 Microstate Definition

A microstate of our system was defined by two lists: one representing the collection of unbound particles, and the other representing particles bound to their various binding sites. The particles themselves were denoted by unique strings and came in multiple copies according to the system parameters. For example, a system with $R = 3$ types of particles with $n_1 = 2$, $n_2 = 3$, and $n_3 = 1$ could have a microstate defined by unbound_particles = [A_2, A_2, A_3] and bound_particles = [A_1, -, A_2, -, A_1, -] where “-” in the bound list stands for an empty binding site.

Since the number of optimally bound particles was an important observable for the system, we also needed to define the optimal binding configuration for the microstates. Such an optimal configuration was chosen at the start of the simulation and was defined as a microstate with no unbound particles and all the bound particles in a particular order. For example, using the previous example, we might define the optimal binding configuration as optimal_bound_config = [A_1, A_1, A_2, A_2, A_2, A_3], in which case the number of optimally bound particles of each type in bound_particles = [A_1, -, A_2, -, A_1, -] is $m_1 = 1$, $m_2 = 1$, and $m_3 = 0$. The number of bound particles of each type is $k_1 = 2$, $k_2 = 1$, and $k_3 = 0$. We note that the order of the elements in unbound_particles is not physically important, but, since the number of optimally bound particles is an important observable, the order of the elements in bound_particles is physically important.
A Simulations

For these simulations, the temperature normalized energy of a microstate with \( k_i \) bound particles of type \( i \) and \( m_i \) optimally bound particles of type \( i \) was defined as

\[
\beta E(k, m) = \sum_{i=1}^{R} (m_i \ln \delta_i + k_i \ln \gamma_i),
\]

(86)

where \( k = (k_1, k_2, \ldots, k_R) \), \( m = (m_1, m_2, \ldots, m_R) \), \( \gamma_i \) is the binding affinity, and and \( \delta_i \) is the optimal-binding affinity.

A.2 Transition Probability

For transitioning between microstates, we allowed for three different transition types: Particle binding to a site; particle unbinding from a site; permutation of two particles in two different binding sites. Particle binding and unbinding both occur in real physical systems, but permutation of particle positions is unphysical. This latter transition type was included to ensure an efficient-in-time sampling of the state space. For simulations of equilibrium systems it is valid to include physically unrealistic transition types as long as the associated transition probabilities obey detailed balance.

At each time step, we first randomly selected one of the three transition types with equal probability for each type, then randomly selected the final proposed microstate given the initial microstate, and finally computed the probability that said proposal was accepted. By the Metropolis Hastings algorithm [Kra06], the probability that the transition is accepted is given by

\[
\text{acceptance prob}(\text{init} \rightarrow \text{fin}) = \min \left\{ 1, e^{-\beta(E_{\text{fin}} - E_{\text{init}})} \frac{\pi(\text{fin} \rightarrow \text{init})}{\pi(\text{init} \rightarrow \text{fin})} \right\},
\]

(87)

where \( \beta \) is inverse temperature, \( E_{\text{init}} \) is the energy of the initial microstate state, and \( E_{\text{fin}} \) is the energy of the final microstate, with energy defined in Eq.(86). The quantity \( \pi(\text{init} \rightarrow \text{fin}) \) is the probability of randomly proposing the final microstate state given the initial microstate state and \( \pi(\text{fin} \rightarrow \text{init}) \) is defined similarly. The ratio \( \pi(\text{fin} \rightarrow \text{init})/\pi(\text{init} \rightarrow \text{fin}) \) varied for each transition type.

A.3 Transition Types and Examples

Below we give examples of the three types of transitions along with the value of the ratio \( \pi(\text{fin} \rightarrow \text{init})/\pi(\text{init} \rightarrow \text{fin}) \) in each case. In the following, \( N_f \) and \( N_b \) represent the number of free particles and the number of bound particles, respectively, before the transition.

- **Particle Binding to Site**: One particle was randomly chosen from the unbound_particles list and placed in a randomly chosen empty site in the bound_particles list.

\[
\pi(\text{fin} \rightarrow \text{init})/\pi(\text{init} \rightarrow \text{fin}) = (N_b + 1)^{-1}/(N_f^{-1} \times N_f) = N^2_f/(N_b + 1).
\]

Example:

unbound_particles = [A2, A2, A3] and bound_particles = [A1, -, A2, -, A1, -]

(transitions to)

unbound_particles = [A2, A3] and bound_particles = [A1, A2, A2, -, A1, -];
Transition weight: $\pi(\text{fin} \to \text{init})/\pi(\text{init} \to \text{fin}) = 9/4$

- **Particle Unbinding from Site**: One particle was randomly chosen from the `bound_particles` list and placed in the `unbound_particles` list. $\pi(\text{fin} \to \text{init})/\pi(\text{init} \to \text{fin}) = ((N_f + 1)^{-1} \times (N_i + 1)^{-1})/N_b/(N_f + 1)^2$.

  Example:
  
  `unbound_particles = [A_2, A_2, A_3]` and `bound_particles = [A_1, -, A_2, -, A_1, -]`
  (transitions to)
  
  `unbound_particles = [A_2, A_2, A_3, A_2]` and `bound_particles = [A_1, -, -, -, A_1, -];`
  Transition weight: $\pi(\text{fin} \to \text{init})/\pi(\text{init} \to \text{fin}) = 3/16$

- **Particle Permutation**: Two particles were randomly selected from the `bound_particles` list, and their positions in the list were switched. $\pi(\text{fin} \to \text{init})/\pi(\text{init} \to \text{fin}) = 1$.

  Example:
  
  `unbound_particles = [A_2, A_2, A_3]` and `bound_particles = [A_1, -, A_2, -, A_1, -]`
  (transitions to)
  
  `unbound_particles = [A_2, A_2, A_3]` and `bound_particles = [A_2, -, A_1, -, A_1, -];`
  Transition weight: $\pi(\text{fin} \to \text{init})/\pi(\text{init} \to \text{fin}) = 1$

For impossible transitions (e.g., particle binding when there are no free particles) the probability for accepting the transition was set to zero. At each temperature, the simulation was run for anywhere from 10,000 to 30,000 time steps depending on observed convergence, and the final 2.5% of the time steps were used to compute ensemble averages of $\langle k \rangle$ and $\langle m \rangle$. These simulations were repeated five times, and each point in Fig. 6b, Fig. 7b, Fig. 8b, and Fig. 9 represents the average $\langle k \rangle$ and $\langle m \rangle$ over these five runs.

## B Consistency Checks for $B_{n,k}$

In this section, we affirm the various consistency checks for $B_{n,k}$. The first check ensures that $B_{n,k}$ has the proper limiting case when $n$ has components with value 1. The second check ensures that $B_{n-m,k-m}$ has the correct normalization when summed over all possible values of $m$.

### B.1 Checking Eq. (9)

We want to check that

$$
\sum_{k_1=0}^{1} \cdots \sum_{k_R=0}^{1} B_{n_0,k}\delta(K, k_1 + \cdots + k_R) = \sum_{j=0}^{R} (-1)^j \binom{R}{j} \binom{R-j}{K-j}^2 (K-j)!,
$$

(88)
B Consistency Checks for $B_{n,k}$

where $n_0 = (1,\ldots,1)$. We start from Eq.(4) (reproduced here for convenience):

$$B_{n,k} = \sum_{j_1=0}^{k_1} \cdots \sum_{j_R=0}^{k_R} \left( \begin{array}{c} n_1 \\ j_1 \end{array} \right) \cdots \left( \begin{array}{c} n_R \\ j_R \end{array} \right) (1)^{j_1+\cdots+j_R} \left( \begin{array}{c} n_1 - j_1 + \cdots + n_R - j_R \\ k_1 - j_1 + \cdots + k_R - j_R \end{array} \right) \times \frac{(k_1 - j_1 + \cdots + k_R - j_R)!}{(k_1 - j_1)! \cdots (k_R - j_R)!}.$$  

(89)

Given $j_i \leq k_i \leq n_i$, if $n_i = 1$ then we must have $(k_i - j_i)! = 1$. We also have \( \left( \begin{array}{c} n_j \\ k_j \end{array} \right) = \left( \begin{array}{c} 1 \\ k_j \end{array} \right) = 1 \). Thus Eq.(4) becomes

$$B_{n_0,k} = \sum_{j_1=0}^{k_1} \cdots \sum_{j_R=0}^{k_R} (1)^{j_1+\cdots+j_R} \left( \begin{array}{c} R - j_1 - \cdots - j_R \\ K - j_1 - \cdots - j_R \end{array} \right) \times (K - j_1 - \cdots - j_R)!,$$

(90)

where we defined $K \equiv \sum_{i=1}^{R} k_i$. With the identity,

$$1 = \sum_{J=0}^{K} \delta(J,j_1+\cdots+j_R),$$

(91)

we can introduce a Kronecker delta and obtain

$$B_{n_0,k} = \sum_{J=0}^{K} \sum_{j_1=0}^{k_1} \cdots \sum_{j_R=0}^{k_R} (1)^{j_1+\cdots+j_R} \left( \begin{array}{c} R - j_1 - \cdots - j_R \\ K - j_1 - \cdots - j_R \end{array} \right) \times (K - j_1 - \cdots - j_R)! \delta(J,j_1+\cdots+j_R).$$

(92)

Isolating the summation over the Kronecker delta yields

$$\sum_{j_1=0}^{k_1} \cdots \sum_{j_R=0}^{k_R} \delta(J,j_1+\cdots+j_R) = \frac{1}{2\pi i} \oint_{|z|=1} \frac{dz}{z^{J+1}} \prod_{i=1}^{k_i} \sum_{j_i=0}^{R} z^{j_i}$$

$$= \frac{1}{2\pi i} \oint_{|z|=1} \frac{dz}{z^{J+1}} (1+z)^{k_1+\cdots+k_R}$$

$$= \left( \begin{array}{c} K \\ J \end{array} \right).$$

(93)

Thus, we obtain

$$B_{n_0,k} = \sum_{J=0}^{K} (-1)^J \left( \begin{array}{c} R - J \\ K - J \end{array} \right) (K - J)! \left( \begin{array}{c} K \\ J \end{array} \right).$$

(94)
Performing the final summation in Eq. (88), we obtain
\[
\sum_{k_1=0}^{1} \cdots \sum_{k_R=0}^{1} B_{n_0,k} \delta(K, k_1 + \cdots + k_R) = \sum_{j=0}^{K} (-1)^j \binom{R-J}{K-J} (K-J)! \binom{K}{j} \times \sum_{k_1=0}^{1} \cdots \sum_{k_R=0}^{1} \delta(K, k_1 + \cdots + k_R)
\]
\[
= \sum_{j=0}^{K} (-1)^j \binom{R-J}{K-J} (K-J)! \binom{K}{j} \binom{R}{J}. \tag{95}
\]
and with the identity
\[
\binom{K}{J} \binom{R}{J} = \binom{R}{J} \binom{R-J}{K-J}, \tag{96}
\]
we have
\[
\sum_{k_1=0}^{1} \cdots \sum_{k_R=0}^{1} B_{n_0,k} \delta(K, k_1 + \cdots + k_R) = \sum_{j=0}^{K} (-1)^j \binom{R-J}{K-J}^2 (K-J)! \binom{R}{J}. \tag{97}
\]
as expected.

Why does this result make sense? When \( n = (1, \ldots, 1) \equiv n_0 \), the vector \( k \) can only have elements of 1 or 0. Thus \( B_{n_0,k} \) represents the number of ways to completely derange a particular collection of \( k \) objects out of \( R \) unique objects. In order to find \( b_{R,K} \), the number of ways to completely derange \( K \) objects, we need to count and sum the number of derangements for all collections of objects. Therefore, to obtain \( b_{R,K} \) we need to sum \( B_{n_0,k} \) over all possible values of \( k \) such that \( \sum_j k_j = K \).

**B.2 Checking equivalence between Eq.(11) and Eq.(12)**

We want to find a reduced form for
\[
I_{n,k} = \sum_{m_1=0}^{k_1} \cdots \sum_{m_R=0}^{k_R} \left( \binom{n_1}{m_1} \cdots \binom{n_R}{m_R} \right) B_{n-m,k-m}. \tag{98}
\]
The expression for \( B_{n,k} \) is
\[
B_{n,k} = \frac{1}{(\sum_i \alpha_i)!} \int_0^{\infty} dx e^{-x} \prod_{i=1}^{R} (-1)^{k_i} x^{\alpha_i} L^{(\alpha_i)}_{k_i}(x), \tag{99}
\]
Noting that \( \alpha_i = n_i - k_i = (n_i - m_i) - (k_i - m_i) \), from Eq.(98) and Eq.(99), we find
\[
I_{n,k} = \frac{1}{(\sum_i \alpha_i)!} \int_0^{\infty} dx e^{-x} x^{\sum_i \alpha_i} \prod_{i=1}^{R} \sum_{m_i=0}^{k_i} \binom{n_i}{m_i} (-1)^{k_i-m_i} L^{(\alpha_i)}_{k_i-m_i}(x). \tag{100}
\]
Next, we make the change of variables
\[
q_i \equiv k_i - m_i. \tag{101}
\]
B Consistency Checks for $B_{n,k}$

We then have

$$I_{n,k} = \frac{1}{(\sum_i \alpha_i)!} \int_0^\infty dx \ e^{-x \sum_i \alpha_i} \prod_{i=1}^R \sum_{q_i=0}^{k_i} (k_i + \alpha_i) (-1)^q \ L^{(\alpha_i)}_{q_i}(x). \quad (102)$$

To simplify this result we need to find an identity for

$$\sum_{q=0}^k \frac{(k + \alpha)}{(k - q)} U^q L^{(\alpha)}_q(x). \quad (103)$$

Expanding the generalized Laguerre polynomial $L^{(\alpha)}_q(x)$ according to its definition, we obtain

$$\sum_{q=0}^k \frac{(k + \alpha)}{(k - q)} U^q L^{(\alpha)}_q(x) = \sum_{q=0}^k \frac{(k + \alpha)}{(k - q)} U^q \sum_{i=0}^q \frac{(q + \alpha)}{i!} (-1)^i x^i. \quad (104)$$

More identity wrangling gives us

$$\frac{(k + \alpha)}{(k - q)} \frac{(q + \alpha)}{(q - i)} = \frac{(k + \alpha)}{(k - i)} \frac{(k - i)}{(q - i)}, \quad (105)$$

and thus Eq. (104) becomes

$$\sum_{q=0}^k \frac{(k + \alpha)}{(k - q)} U^q L^{(\alpha)}_q(x) = \sum_{i=0}^k \frac{(k + \alpha)}{(k - i)} \frac{(-1)^i}{i!} x^i \sum_{q=i}^k U^q \frac{(k - i)}{(q - i)}$$

$$= \sum_{i=0}^k \frac{(k + \alpha)}{(k - i)} \frac{(-1)^i}{i!} x^i U^i \sum_{q=i}^k \frac{(k - i)}{(q - i)}$$

$$= \sum_{i=0}^k \frac{(k + \alpha)}{(k - i)} \frac{(-1)^i}{i!} x^i U^i (1 + U)^{k-i}, \quad (106)$$

which, by the definition of the Laguerre polynomial yields

$$\sum_{q=0}^k \frac{(k + \alpha)}{(k - q)} U^q L^{(\alpha)}_q(x) = (1 + U)^k L^{(\alpha)}_k \left( \frac{xU}{1+U} \right). \quad (107)$$

Also, from the definition of the Laguerre polynomial, we can show

$$\lim_{\lambda \to 0} \lambda^k L^{(\alpha)}_k \left( \frac{x}{\lambda} \right) = (-1)^k \frac{x^k}{k!}. \quad (108)$$

Therefore, with Eq. (107) and Eq. (108), we have

$$\sum_{q=0}^k \frac{(k + \alpha)}{(k - q)} (-1)^q L^{(\alpha)}_q(x) = \lim_{U \to -1} (1 + U)^k L^{(\alpha)}_k \left( \frac{xU}{1+U} \right) = \frac{x^k}{k!}. \quad (109)$$
Inserting this result into Eq.(102) gives us

\[ I_{n,k} = \frac{1}{(\sum \alpha_i)!} \int_0^\infty dx \frac{e^{-x} x^{\alpha_1 k_1 + \cdots + \alpha_R k_R}}{k_1! \cdots k_R!} \]

which, with the Gamma function definition, yields

\[ I_{n,k} = \frac{(n_1 + \cdots + n_R)!}{(n_1 - k_1 + \cdots + n_R - k_R)! k_1! \cdots k_R!}, \tag{110} \]

Why does this result make sense? From one perspective, the quantity \( \binom{n_i}{m_i} \cdots \binom{n_R}{m_R} B_{n-m,k-m} \) is the number of ways to choose \( m_j \) out of \( n_j \) positions to contain their correct elements while the remaining \( k_j - m_j \) elements are completely deranged with respect to their \( n_j - m_j \) remaining correct positions, for \( j = 1, \ldots, R \). If we sum over all possible values of \( m_j \) we should obtain the number of ways to arrange (i.e., not only derange) \( k_j \leq n_j \) objects for \( j = 1, \ldots, R \) across a total of \( n_1 + \cdots + n_R \) lattice sites.

On the other hand, if we are trying to calculate the number of ways to arrange \( k_j \) objects of type \( j \) for \( j = 1, \ldots, R \) across \( n_1 + \cdots + n_R \) lattice sites, we can use the language of multinomials. Say that the objects represent “filled” sites on the lattice and the spaces between objects are “empty” sites. Then there are \( k_1 + \cdots + k_R \) filled sites (of which \( k_i \) are identical for each \( i \)) and \( n_1 - k_1 + \cdots + n_R - k_R \) empty sites. Finding the number of ways to arrange the objects amongst the \( n_1 + \cdots + n_R \) sites is equivalent to finding the total number of ways to order this collection of filled and empty sites while correcting for equivalent orderings due to reordering the positions of the same type of site. Including both filled and empty sites there is a total of \( n_1 + \cdots + n_R \) sites amongst which we have \( n_1 - k_1 + \cdots + n_R - k_R \) “copies” of empty sites, \( k_1 \) copies of filled sites of type 1, \( k_2 \) copies of filled sites of type 2, \ldots and \( k_R \) copies of filled sites of type \( R \). Counting the number of ways to order the \( n_1 + \cdots + n_R \) sites and correcting for the equivalent reorderings arising from the multiple copies of various types of sites leads to the multionomial

\[ \frac{(n_1 + \cdots + n_R)!}{(n_1 - k_1 + \cdots + n_R - k_R)! k_1! \cdots k_R!}, \tag{112} \]

which we have shown is equivalent to the result written in terms of a summation over \( B_{n-m,k-m} \).

\section{Deriving General Partition Function}

In this section, we derive Eq.(17) from Eq.(16). First we note from the definition of \( B_{n,k} \) in Eq.(6) that

\[ B_{n-m,k-m} = \frac{1}{(\sum \alpha_j)!} \int_0^\infty dx e^{-x} x^{\sum \alpha_j} \prod_{j=1}^R (-1)^{k_j - m_j} f_{k_j - m_j}(x), \tag{113} \]
C Deriving General Partition Function

where \( \alpha_j = n_j - k_j \). Thus the summation over \( m \) becomes

\[
\sum_m B_{n-m,k-m} \prod_{j=1}^R \binom{n_j}{m_j} \delta^{m_j}_j = \frac{1}{(\sum_j \alpha_j)!} \int_0^\infty dx \, e^{-x} x^{\sum_j \alpha_j} \times \prod_{j=1}^R \sum_{m_j=0}^{k_j} \binom{n_j}{m_j} \delta^{m_j}_j (-1)^{k_j-m_j} L^{(\alpha_j)}_{k_j-m_j}(x),
\]

where we changed variables from \( m_j \) to \( q_j = k_j - m_j \) in the final line. Using the Laguerre polynomial identity (derived in Eq. (106))

\[
\sum_{q=0}^k \binom{k + \alpha}{k - q} U^q L^{(\alpha)}_q(X) = (1 + U)^k L^{(\alpha)}_k \left( \frac{XU}{1+U} \right), \tag{115}
\]

we then have

\[
\sum_{m_j=0}^{k_j} \binom{k_j + \alpha_j}{k_j - q_j} \delta^{k_j-q_j}_j (-1)^{q_j} L^{(\alpha_j)}_{q_j}(x) = (\delta_j - 1)^{k_j} L^{(\alpha_j)}_{k_j} \left( \frac{x}{1-\delta_j} \right). \tag{116}
\]

Therefore, the summation over \( m \) becomes

\[
\sum_m B_{n-m,k-m} \prod_{j=1}^R \binom{n_j}{m_j} \delta^{m_j}_j = \frac{1}{(\sum_j \alpha_j)!} \int_0^\infty dx \, e^{-x} \prod_{j=1}^R x^{\alpha_j} (\delta_j - 1)^{k_j} L^{(\alpha_j)}_{k_j} \left( \frac{x}{1-\delta_j} \right)
\equiv C_{n,k} \tag{117}
\]

For the summation over \( k \), we use the contour integral identity for inverse factorial

\[
\frac{1}{(\sum_j \alpha_j)!} = \frac{1}{2\pi i} \oint_{C^{\infty}} \frac{dz}{z^{\sum_j \alpha_j + 1} e^z}, \tag{118}
\]

where \( C \) is a closed contour about the origin, to obtain

\[
\sum_k C_{n,k} \prod_{j=1}^R \binom{n_j}{k_j} L^{(\alpha)}_{k_j} = \frac{1}{2\pi i} \oint_C \frac{dz}{z^{\sum_j \alpha_j + 1} e^z} \int_0^\infty dx \, e^{-x} \prod_{j=1}^R \sum_{k_j=0}^{n_j} (x)_{k_j}^{\alpha_j} \gamma_j^{k_j} (\delta_j - 1)^{k_j} L^{(\alpha_j)}_{k_j} \left( \frac{x}{1-\delta_j} \right)
\]

\[
= \frac{1}{2\pi i} \oint_C \frac{dz}{z^{\sum_j \alpha_j + 1} e^z} \int_0^\infty dx \, e^{-x} \prod_{j=1}^R \sum_{k_j=0}^{n_j} \frac{1}{(n_j-k_j)!} \left( \frac{x}{z^{\gamma_j} (\delta_j - 1)} \right)^{n_j-k_j} \times L^{(n_j-k_j)}_{k_j} \left( \frac{x}{1-\delta_j} \right). \tag{119}
\]
Using the identity
\[ \sum_{k=0}^{n} \frac{Y^{n-k}}{(n-k)!} L_k^{(n-k)}(X) = L_n(X-Y), \]  
(120)
(which can be proved from the definition of the generalized Laguerre polynomial Eq.(5) with the identification \( L_n(X) \equiv L_n^{(0)}(X) \)), we find that the final partition function is
\[ Z_n = \frac{1}{2\pi i} \oint_{\Gamma} dz \int_{0}^{\infty} dx e^{z-x} \prod_{j=1}^{R} (\gamma_j (\delta_j - 1))^{n_j} L_{n_j} \left( \frac{x(z\gamma_j + 1)}{z\gamma_j (1-\delta_j)} \right). \]  
(121)

\section{D Gendered Dimer Assembly Equilibrium Conditions}

Here we derive the equilibrium conditions Eq.(30) and Eq.(31) from the large \( N \) approximation conditions Eq.(27) and the observable definitions Eq.(28) and Eq.(29). First, we rewrite some of our previous results for convenience. The conditions defining \( \bar{x} \) and \( \bar{z} \) are
\[ 1 = \sum_{j=1}^{R} \frac{\bar{z}\gamma_j + 1}{\bar{x} + \bar{z}\gamma_j (\delta_j - 1 + \bar{x})}, \quad \bar{z} = \bar{x} \sum_{j=1}^{R} \frac{1}{\bar{x} + \bar{z}\gamma_j (\delta_j - 1 + \bar{x})}. \]  
(122)
and the expressions for \( \langle k_j \rangle \) and \( \langle m_j \rangle \) are
\[ \langle k_j \rangle = \frac{\bar{z}\gamma_j (\delta_j - 1 + \bar{x})}{\bar{z}\gamma_j (\delta_j - 1 + \bar{x}) + \bar{x}}, \]  
(123)
\[ \langle m_j \rangle = \frac{\bar{z}\gamma_j \delta_j}{\bar{z}\gamma_j (\delta_j - 1 + \bar{x}) + \bar{x}}. \]  
(124)
Next, we seek to eliminate the \( \bar{z} \) and \( \bar{x} \) dependence from these observables. From Eq.(122), Eq.(123), and Eq.(124), we can show
\[ \bar{z} = N - \langle k \rangle, \quad \bar{x} = N - \sum_{j=1}^{R} \langle m_j \rangle (1 - \delta_j^{-1}), \]  
(125)
where we used \( \langle k \rangle = \sum_{j=1}^{R} \langle k_j \rangle, \) and the total particle number \( N \) is equal to the number of particle types \( R \) when there is one copy per particle. With Eq.(123) and Eq.(124), we can also show
\[ \langle k_j \rangle - \langle m_j \rangle (1 - \delta_j^{-1}) = \frac{\bar{z}\gamma_j \bar{x}}{\bar{z}\gamma_j (\delta_j - 1 + \bar{x}) + \bar{x}}, \]  
(126)
and with the second equation in Eq.(122) and the first equation in Eq.(125), we obtain
\[ \sum_{j=1}^{R} \frac{1}{\gamma_j} \left( \langle k_j \rangle - \langle m_j \rangle (1 - \delta_j^{-1}) \right) = \left( N - \langle k \rangle \right)^2, \]  
(127)
which is our first equilibrium condition.

47
E Derivation of Eq. (69)

From both equations in Eq. (122), we have

$$\sum_{j=1}^{R} \frac{\bar{z} \gamma_j}{\bar{x} + \bar{z} \gamma_j (\bar{\delta}_j - 1 + \bar{x})} = 1 - \sum_{j=1}^{R} \frac{1}{\bar{x} + \bar{z} \gamma_j (\bar{\delta}_j - 1 + \bar{x})} = 1 - \bar{z}/\bar{x}. \quad (128)$$

Using Eq. (128) in Eq. (123) and Eq. (124), we can show

$$\langle k \rangle - \langle m \rangle + \sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1} = \sum_{j=1}^{R} \left( \langle k_j \rangle - \langle m_j \rangle (1 - \delta_j^{-1}) \right) = \bar{x} - \bar{z} \quad (129)$$

$$\sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1} = 1 - \bar{z}/\bar{x}. \quad (130)$$

With Eq. (129) and Eq. (130) and the second equation in Eq. (125), we then obtain

$$\sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1} = \frac{\langle k \rangle - \langle m \rangle + \sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1}}{N - \langle m \rangle + \sum_{j=1}^{R} \langle m_j \rangle \delta_j^{-1}}, \quad (131)$$

which is our second equilibrium condition.

E Derivation of Eq. (69)

In this section, we derive Eq. (69), the general thermal condition for the fully optimally bound state. This state is defined as

$$\langle k \rangle = \langle m \rangle = N - 1 + O(\delta^{-1}). \quad (132)$$

To derive the associated thermal condition, we will need some of our previous results. In particular, we need the average number of optimally bound ligands of type \( j \)

$$\langle m_j \rangle = \frac{n_j \delta_j}{\bar{\delta}_j - 1} \frac{L_{n_j - 1}(\bar{\delta}_j)}{L_{n_j}(\bar{\delta}_j)}, \quad (133)$$

and the equations relating \( \bar{z} \) and \( \bar{x} \) to our observables

$$\bar{z} = N - \langle k \rangle, \quad \bar{x} = N - \sum_{i=1}^{R} \langle m_i \rangle (1 - \delta_i^{-1}). \quad (134)$$

From Eq. (132), we can infer that for the fully optimally bound state we have

$$\langle m_j \rangle = n_j - O(R^{-1}) + O(\delta_j^{-1}), \quad (135)$$
where we are explicitly referencing the \(O(R^{-1})\) terms in order to satisfy \(\sum_{j=1}^{R} O(R^{-1}) = 1\) required of Eq.(132). Using Eq.(135) in the second equation in Eq.(134) yields

\[
\bar{x} = 1 + \sum_{j=1}^{R} n_j \delta_j^{-1} + O(\delta^{-1}),
\]

(136)

where we subsumed terms of order \(O(R^{-1}\delta^{-1})\) into \(O(\delta^{-1})\). For the fully optimally bound configuration, ligands need to greatly favor their optimal receptors in order to bind only to such receptors. Thus for the fully optimally bound state, we can assume that the optimal-binding affinity for each particle-type is much greater than 1: \(\delta_j \gg 1\). From this assumption we can take \(\bar{\phi}_j\) defined as

\[
\bar{\phi}_j = \frac{\bar{x}}{1 - \delta_j} \left( 1 + \frac{1}{\bar{z}\gamma_j} \right)
\]

(137)
to satisfy \(|\bar{\phi}_j| \ll 1\). We will check this latter assumption at the end of the calculation. From here, we find

\[
\langle m \rangle = \sum_{j=1}^{R} \frac{n_j}{1 - \delta_j^{-1}} \frac{L_{n_j-1}(\bar{\phi}_j)}{L_{n_j}(\bar{\phi}_j)} = 1 + \bar{\phi}_j + O(\delta_j^2) = 1 - \bar{x}\delta_j^{-1} \left( 1 + \frac{1}{\bar{z}\gamma_j} \right) + O(\delta_j^{-2}).
\]

(138)

Inserting Eq.(138) into Eq.(133), we obtain

\[
\langle m \rangle = N - 1 + \bar{x} - \bar{x} \sum_{j=1}^{R} n_j \delta_j^{-1} \left( 1 + \frac{1}{\bar{z}\gamma_j} \right) + O(\delta^{-1}).
\]

(139)

Using Eq.(136), we can write this result as

\[
\langle m \rangle = N - 1 + \bar{x} - \bar{x} \sum_{j=1}^{R} n_j \delta_j^{-1} \left( 1 + \frac{1}{\bar{z}\gamma_j} \right) + O(\delta^{-1}).
\]

(140)

Therefore, we see that Eq.(132) is reproduced if

\[
1 = \sum_{j=1}^{R} n_j \delta_j^{-1} \left( 1 + \frac{1}{\bar{z}\gamma_j} \right).
\]

(141)

With \(\langle k \rangle = N - 1 + O(\delta^{-1})\) and \(\bar{\varepsilon} = N - \langle k \rangle\), we find \(\bar{\varepsilon} = 1 + O(\delta^{-1})\), thus giving us the final thermal condition

\[
1 = \sum_{j=1}^{R} \frac{n_j}{\delta_j} \left( 1 + \frac{1}{\gamma_j} \right) + O(\delta^{-2}).
\]

(142)

Now to check the assumption of \(|\bar{\phi}_j| \ll 1\) for consistency. From Eq.(142), we can conclude that \(\sum_{j=1}^{R} n_j \delta_j^{-1} < 1\) and that, in the large number limit (i.e., \(n_j \gg 1\)), \(1 \gg \delta_j^{-1}(1 + 1/\gamma_j)\). Thus \(\bar{x}\) defined in Eq.(136) is \(O(1)\), and \(\bar{\phi}_j = \bar{x}(1 + 1/\bar{z}\gamma_j)/(1 - \delta_j) \simeq -\bar{x}\delta_j^{-1}(1 + 1/\gamma_j) + O(\delta_j^{-2})\) can indeed be taken to satisfy \(|\bar{\phi}_j| \ll 1\).
References

[dM13] Pierre Rémond de Montmort. *Essay d’analyse sur les jeux de hazard...* J. Quillau, 1713.

[EG76] Shimon Even and J Gillis. Derangements and laguerre polynomials. *Mathematical Proceedings of the Cambridge Philosophical Society*, 79:135–143, 1976.

[JAVH20] Inga Jarmoskaite, Ishraq AlSadhan, Pavanapuresan P Vaidyanathan, and Daniel Herschlag. How to measure and evaluate binding affinities. *Elife*, 9:e57264, 2020.

[JLD10] Iain G Johnston, Ard A Louis, and Jonathan PK Doye. Modelling the self-assembly of virus capsids. *Journal of Physics: Condensed Matter*, 22(10):104101, 2010.

[Kra06] Werner Krauth. *Statistical mechanics: algorithms and computations*, volume 13. OUP Oxford, 2006.

[MHSB99] Bettina Malnic, Junzo Hirono, Takaaki Sato, and Linda B Buck. Combinatorial receptor codes for odors. *Cell*, 96(5):713–723, 1999.

[Mil13] Ron Milo. What is the total number of protein molecules per cell volume? a call to rethink some published values. *BioEssays*, 35(12):1050–1055, 2013.

[Nel04] Philip Nelson. *Biological physics: Energy, information life*, chapter 8, pages 315–332. WH Freeman New York, 2004.

[PH15] Jason D Perlmutter and Michael F Hagan. Mechanisms of virus assembly. *Annual review of physical chemistry*, 66:217–239, 2015.

[PTKG12] Rob Phillips, Julie Theriot, Jane Kondev, and Hernan Garcia. *Physical biology of the cell*. Garland Science, 2012.

[RP08] Valerica Raicu and Aurel Popescu. *Integrated molecular and cellular biophysics*, volume 10. Springer, 2008.

[RSK04] Bryan L Roth, Douglas J Sheffler, and Wesley K Kroeze. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nature reviews Drug discovery*, 3(4):353–359, 2004.

[SML+20] Christina J Su, Arvind Murugan, James M Linton, Akshay Yeluri, Justin Bois, Heidi Klumpe, Yaron E Antebi, and Michael B Elowitz. Ligand-receptor promiscuity enables cellular addressing. *bioRxiv*, 2020.

[Wil18] Mobolaji Williams. Permutation glass. *Physical Review E*, 97(1):012139, 2018.

[Wil19] Mobolaji Williams. Self-assembly of a dimer system. *Physical Review E*, 99(4):042133, 2019.