Binding Polynomials

We've looked at three cases of ligand binding so far:

The single set of independent sites (ss[i]s)

$$\bar{\upsilon} = \frac{nK[L]}{1 + K[L]} \tag{1}$$

Multiple sets of independent sites (ms[i]s, or m[i]ss)

$$\bar{\upsilon} = \sum_m \frac{n_i K_i [L]}{1 + K_i [L]} \tag{2}$$

'All or none', or 'two-state' cooperativity

$$\bar{\upsilon} = \frac{nK[L]^n}{1 + K[L]^n} \tag{3}$$

(1) & (3) are the limiting cases of no cooperative interactions, and very strong cooperative interactions, respectively, for a system with one type of binding site.

Q: What are the 'units' of K in Eq. (3)? (Why is 'units' placed in quotations.) What is the relationship to K in Eq. (1)?

Equations 1 to 3 spell out the statistical distributions of protein microstates, with specific reference to the extent of ligand binding (and assumptions about the nature of the binding). *In general*, the numerator of $\bar{\upsilon}$ accounts for the number of moles of bound ligand, and the denominator accounts for the number moles of protein. The general case with *n* binding sites is written as

$$\bar{\upsilon} = \frac{[PL] + 2[PL_2] + 3[PL_3] \cdots + n[PL_n]}{[P] + [PL] + [PL_2] + [PL_3] \cdots + [PL_n]} \tag{4}$$

Have there been assumptions made about whether the sites interact, or whether there is one or more class of sites? (The answer to both is *no*, hence its generality.) *Also, without loss of generality,* substitutions can be made for the PL_i (where *i* ranges from 0 to *n*).

$$\bar{\upsilon} = \frac{K_1[P][L] + 2K_1K_2[P][L]^2 + 3K_1K_2K_3[P][L]^3 \cdots + nK_1K_2K_3 \cdots K_n[P][L]^n}{[P] + K_1[P][L] + K_1K_2[P][L]^2 + K_1K_2K_3[P][L]^3 \cdots + K_1K_2K_3 \cdots K_n[P][L]^n} \tag{5}$$

From here, relationships among binding constants can be assumed and then exploited in Equation 5 to reach Equations 1 to 3. Some of the subtleties in Equation 5 are, in part, a matter of definition. K_1 through K_n are 'macroscopic' association constants; the relationship to the underlying intrinsic constant(s) will change according to the number of sets (or classes) of sites, and the nature of the interaction among sites, *i.e.* are there interactions, or not. In all cases [*P*] can be factored out to give:

$$\overline{\upsilon} = \frac{K_1[L] + 2K_1K_2[L]^2 + 3K_1K_2K_3[L]^3 \cdots + nK_1K_2K_3 \cdots K_{n-2}K_{n-1}K_n[L]^n}{1 + K_1[L] + K_1K_2[L]^2 + K_1K_2K_3[L]^3 \cdots + K_1K_2K_3 \cdots K_{n-2}K_{n-1}K_n[L]^n} \tag{6}$$

Factoring out [*P*] implicitly acknowledges it ([*P*]) as a 'reference state' with respect to statistical and energetic considerations, more on this below.

Q: What are the expressions for K_1 in terms of the intrinsic association constants, if (i) there is a single set of independent sites, (ii) there are n sets of independent sites, (iii) there is one set of strongly (all or none) interacting sites?

SSS. The relationship between the macroscopic association constants and a single intrinsic association constant (*K*), simply reflects the statistics of ligand binding through a single class of sites. The individual steps of the binding process, catalogued according to the number of ligands bound, are given by

$$\begin{aligned}
P + L &\rightleftharpoons PL & K_1 &= [PL]/[P][L] \\
PL + L &\rightleftharpoons PL_2 & K_2 &= [PL_2]/[PL][L] \\
PL_2 + L &\rightleftharpoons PL_3 & K_3 &= [PL_3]/[PL_2][L] \\
&\vdots & &\vdots \\
PL_{n-1} + L &\rightleftharpoons PL_n & K_n &= [PL_n]/[PL_{n-1}][L]
\end{aligned} \tag{7}$$

The K_i (macroscopic constants) are related to *K* (the intrinsic constant) by

$$K_i = \frac{\Omega_{n,i}}{\Omega_{n,i-1}} K \tag{8}$$

where $\Omega_{n,i}$ and $\Omega_{n,i-1}$ represent the number of ways of assorting i and $i-1$ ligand molecules, respectively, on n independent, equivalent binding sites.

$$\Omega_{n,i} = \frac{n!}{(n-i)!i!} \tag{9}$$

Equation 9 is recognizable as the equation for coefficients of the binomial expansion. For small values of *n*, *e.g.* 4, the validity of the Equations 8 and 9 in generating the

relationships between the K_i and K can be validated by enumerating the different microscopic forms of PL_i and PL_{i-1} diagrammatically.

Consider the situation with 4 sites. In Equation 10, the formulae for the K_i, expressed in terms of K (*via* Equation 8) have been substituted into Equation 5

$$\bar{\upsilon} = \frac{(4K)[L] + 2(4K)\left(\frac{6}{4}K\right)[L]^2 + 3(4K)\left(\frac{6}{4}K\right)\left(\frac{4}{6}K\right)[L]^3 + 4(4K)\left(\frac{6}{4}K\right)\left(\frac{4}{6}K\right)\left(\frac{1}{4}K\right)[L]^4}{1 + (4K)[L] + (4K)\left(\frac{6}{4}K\right)[L]^2 + (4K)\left(\frac{6}{4}K\right)\left(\frac{4}{6}K\right)[L]^3 + (4K)\left(\frac{6}{4}K\right)\left(\frac{4}{6}K\right)\left(\frac{1}{4}K\right)[L]^4} \quad (10)$$

The terms in the numerator are collected in such as fashion to accentuate the number of sites (4).

$$\bar{\upsilon} = \frac{4(K)[L] + 4\left(3K^2\right)[L]^2 + 4\left(3K^3\right)[L]^3 + 4\left(K^4\right)[L]^4}{1 + 4K[L] + 6K^2[L]^2 + 4K^3[L]^3 + K^4[L]^4} \quad (11)$$

Factoring out $4K[L]$ in the numerator produces a 3rd order polynomial. The denominator is a 4th order polynomial.

$$\bar{\upsilon} = \frac{4K[L]\left\{1 + 3K[L] + 3K^2[L]^2 + K^3[L]^3\right\}}{1 + 4K[L] + 6K^2[L]^2 + 4K^3[L]^3 + K^4[L]^4} \quad (12)$$

Both of these are of the form $(1 + K[L])^n$,

$$\bar{\upsilon} = \frac{4K[L](1 + K[L])^3}{(1 + K[L])^4} = \frac{4K[L]}{(1 + K[L])} \quad (13)$$

Generalizing to n sites gives Equation 1.

The Partition Function, *Q*. The denominator of Equation 6 is equal to the total protein concentration divided by the concentration of free protein is $[P]_{tot}/[P]$. This quantity has meaning in statistical mechanics; it is an example of a partition function, Q.

$$Q = 1 + K_1[L] + K_1K_2[L]^2 + K_1K_2K_3[L]^3 \cdots + K_1K_2K_3 \cdots K_{n-2}K_{n-1}K_n[L]^n \quad (14)$$

Q is the sum of the number of ways a given energy state may be formed. $[P]$, represented by the leading term on the right hand side of Equation 14, is the 'reference state' and has a relative energy equal to 1, and a statistical factor equal to 1. The partition function for the equivalent sites (sss) model with 4 sites is

$$Q = 1 + 4K[L] + 6K^2[L]^2 + 4K^3[L]^3 + K^4[L]^4 \quad (15)$$

When written in a slightly different way, as Equation 16,

$$Q = 1K^0[L]^0 + 4K^1[L]^1 + 6K^2[L]^2 + 4K^3[L]^3 + 1K^4[L]^4 \quad (16)$$

shows explicitly the contributions from statistics (the lead numerical coefficients) and terms that represent the energies of the different protein forms with one or more ligand molecules bound ($K^i[L]^i$). These are relative energy terms, relative to the energy of the reference state [*P*], and are also know as 'statistical weights' (they weight the statistical contribution). *Q*, for the general case for *n* equivalent binding sites is

$$Q = \sum_{i=0}^{n} \Omega_{n,i} K^i [L]^i \quad (17)$$

Each term in the summation of Equation 17 is the product between the number of distinguishable microscopic arrangements ($\Omega_{n,i}$) and the statistical weight, $K^i[L]^i$. ($K[L]$ is sometimes referred to as the reduced ligand concentration.) The $K^i[L]^i$ represent ratios of protein forms, *e.g.* $[PL_i]/[P]$, in a Boltzmann-like manner, $[PL_i]/[P] = \exp[i\Delta g/k_B T]$, where $\Delta g = \Delta g^0 + k_B T \ln[L]$. The progressive binding of ligand incrementally adds to the statistical weight. Note that for the independent sites model, the energy increment is the same for each ligand molecule that binds to the protein:

i	statistical weight	Ratio	Δg
0	$K^0[L]^0$	$[P]/[P]$	0
1	$K^1[L]^1$	$[PL]/[P]$	Δg
2	$K^2[L]^2$	$[PL_2]/[P]$	$2\Delta g$
3	$K^3[L]^3$	$[PL_3]/[P]$	$3\Delta g$
.	.	.	.
.	.	.	.
n	$K^n[L]^n$	$[PL_n]/[P]$	$n\Delta g$

Finally, it is interesting to note that

$$\bar{\upsilon} = \frac{(K[L])}{Q}\frac{dQ}{d(K[L])} = \frac{d\ln Q}{d\ln(K[L])} \quad (18)$$

Discussions of the relationship between experimental observables and the partition function, of which Equation 18 is an example, are found in statistical thermodynamics and biophysics textbooks.

Two Models of Finite Cooperativity

Cooperativity is rarely adequately described by an 'all or none' scenario, so other models have been developed. Two widely used models are the 'concerted' and 'sequential' models.

1. The Monod-Wyman-Changeux (MWC), or concerted, model was published in 1965. A central assumption is that cooperative proteins consist of an oligomeric cluster of protomers (subunits), in which all the protomers are in the same conformation, although the protein can undergo a transition between (at least) two conformational states in a *concerted* fashion. The different conformations have different ligand binding affinities. The details of the assumptions are found in the paper that described the model: Monod, Wyman & Changeux. 1965. *J. Mol. Biol.* **12,** 88-118. (Posted on the course website.) Also, see Cantor & Schimmel, Vol. III, Chapter 17 – the full reference for this book is on the course website.

The two assumptions of the MWC model (i) that two conformations of the macromolecule are assumed, and (ii) that a difference in binding affinity exists between these two forms, are combined in a way that generates positive cooperativity.

The relaxed (*R*) and tense (*T*) forms of the macromolecule, each engage in ligand binding equilibria. (Here the ligand is represented by the letter '*F*'.)

$$\begin{array}{ll} R + F \rightleftharpoons RF & T + F \rightleftharpoons TF \\ RF + F \rightleftharpoons RF_2 & TF + F \rightleftharpoons TF_2 \\ RF_2 + F \rightleftharpoons RF_3 & TF_2 + F \rightleftharpoons TF_3 \\ \quad . & \quad . \\ \quad . & \quad . \\ \quad . & \quad . \\ RF_{n-1} + F \rightleftharpoons RF_n & TF_{n-1} + F \rightleftharpoons TF_n \end{array} \qquad (19)$$

An equilibrium, which exists between the *R* and *T* forms in the absence of ligand, is given by

$$R \rightleftharpoons T \qquad L_0 = [T]/[R] \qquad (20)$$

where L_0 is generally greater than one, which is implies that the *T* form (the low affinity form) is more stable than the *R* form in the absence of ligand. The relationship between ligand binding constants to the *T* and *R* forms, which are characterized by the intrinsic association constants K_T and K_R, respectively, is given by

$$c = K_T/K_R \qquad (21)$$

By definition the *T* form binds ligand less tightly than the *R* form, $c < 1$. The average degree of saturation takes on a specific form, which is illustrated by a protein with four ligand-binding sites

$$\bar{\upsilon} = \frac{[RF] + 2[RF_2] + 3[RF_3] + 4[RF_4] \; + \; [TF] + 2[TF_2] + 3[TF_3] + 4[TF_4]}{[R] + [RF] + [RF_2] + [RF_3] + [RF_4] \; + \; [T] + [TF] + [TF_2] + [TF_3] + [TF_4]} \qquad (22)$$

Factoring the polynomials in the same manner, as had been done for the single sites model, gives

$$\bar{\upsilon} = \frac{4K_R[R][F](1+K_R[F])^3 \;+\; 4K_T[T][F](1+K_T[F])^3}{[R](1+K_R[F])^4 \;+\; [T](1+K_T[F])^4} \tag{23}$$

Application of the definitions for c ($K_T = cK_R$) and L_0 ($[T] = L_0[R]$) allows $[T]$ and K_T to be factored out

$$\bar{\upsilon} = \frac{4K_R[F](1+K_R[F])^3 \;+\; 4cL_0K_R[F](1+cK_R[F])^3}{(1+K_R[F])^4 \;+\; L_0(1+cK_R[F])^4} \tag{24}$$

Generalization to the protein that has n binding sites, gives

$$\bar{\upsilon} = \frac{nK_R[F](1+K_R[F])^{n-1} \;+\; ncK_RL_0[F](1+cK_R[F])^{n-1}}{(1+K_R[F])^n \;+\; L_0(1+cK_R[F])^n} \tag{25}$$

What values of c and L_0 must take on to reduce to either (i) the single-set-of-sites model, or (ii) the all-or-none model?

2. The Koshland-Nemethy-Filmer (KNF) Sequential Model generates cooperativity by allowing the intrinsic constants change in a way that reflects the changing probability of a protein subunit to undergo the ligand-dependent transition from a low to a high affinity state, which is influenced by ligand binding to one (or more) neighboring subunits. A two-site model suffices to illustrate the point

$$\bar{\upsilon} = \frac{2K_1[F] + 2K_1K_2[F]^2}{1 + 2K_1[F] + K_1K_2[F]^2} \tag{26}$$

where K_1 and K_2 are intrinsic binding constants. In the sequential model, $K_1 < K_2$ indicates positive cooperativity, $K_1 > K_2$ indicates negative cooperativity, and $K_1 = K_2$ indicates independent sites. The exact nature of the binding polynomial and the interaction terms depends on the symmetry of the macromolecule. See the original reference for details. (Koshland, Nemethy and Filmer. 1966. *Biochemistry*. 5:365-385; posted at the course website)