APPENDIX I:

THE IMPORTANCE OF BINDING CONSTANTS: THEORY, MODELS, AND DISCUSSION

A.I.0 INTRODUCTION

The non-covalent binding of one or more ligands to a single macromolecule carries enormous consequences in the biological world: the binding of oxygen to hemoglobin enables oxygen to be transported from mammalian lungs to tissue organs; non-covalent interactions of DNA and RNA allow life to be passed on from generation to generation; the mechanism of many pharmaceutics rely heavily on their uptake by targeted receptors; the mode of action of enzymes are based strictly on non-covalent interactions; the list goes on. Indeed, it has been stated that there are as many intracellular reactions in which binding plays an integral part as there are total bond breaking/reforming chemical reactions. It should become readily apparent, then, that to understand the inner workings of life itself--insofar as treating life as a complex series of chemical reactions and conversions--one must pay great attention to the importance of non-covalent interactions.

A large part of our existing knowledge of non-covalent binding is based on the measurement of equilibrium constants. Equilibrium constants afford the scientist information on the mechanism of the chemical process involved. In general, the basic process can be distilled down to the association, or binding, of one or more ligands to a host molecule. Despite multiple reports otherwise,[†] for our present purposes we will assume that only *one* ligand may occupy each receptor (or host) site. A scheme for this event may be represented by:

$$nG + H \rightleftharpoons HG_n \qquad (1)$$

where *n* signifies the maximum number of binding sites on substrate *H* and *G* denotes the ligands that are free to bind to each site.

[†] For example, Bryant et al. have reported crystal structures of BPP34C10 + 2 ammonium salts, where one host macromolecule is threaded by two ligands. [Bryant, W. S.; Guzei, I. A.; Rheingold, A. L.; Merola, J. S; Gibson, H. W. *J. Org. Chem.* **1998**, 63, 7634.]

It is important to keep in mind that this is a reversible process, and fortunately so for us at that. Given initial reaction conditions, a number of host receptor sites may or may not be occupied by the individual specific guest molecules in question; a second set of reaction conditions may reverse the initial interaction. For example, the chemical environment in mammalian lungs favors the uptake of four molecules of O_2 onto each hemoglobin (Hb) strand. In order to utilize hemoglobin for transport, there must exist a set of conditions outside of the lungs that favors the release of all four molecules of O_2. The body does this by utilizing the partial pressure of oxygen in the Hb solution. Until the Hb solution reaches peripheral tissue, the partial pressure of oxygen remains constant and relatively large. At the perhipheral capillaries, the partial pressure decreases substaintially, favoring the release of O_2 [Voet, D.; Voet, J. *Biochemistry*. John Wiley & Sons: New York, **1990**, 1st Ed., 33]. This reversibility allows for very efficient use and re-use of biochemical entities. Given a set of initial conditions, this efficiency can be quantified using aforementioned association constants. One such constant frequently reported by biochemists in multi-site binding systems is called the dissociation constant K_d.

Much as the name suggests, dissociation constants directly measure how well host molecules bind ligands. The derivation of K_d is straightforward and is in fact the inverse of the equilibrium association constant K_a, $K_a = 1 / K_d$. Whereas chemistry literature favors the use of K_a, the employment of dissociation constants abounds in biochemistry and biology journals and will be presented below according to convention.

At equilibrium, ligand molecules can either be found in the free state, *G*, or in the bound state, *HG* (Equation 1). Given dilute conditions, which holds for most natural systems as well as the experiments presented in sections I-III, the assumption that concentration approximates chemical activity holds true. Thus:

$$K_d = \frac{[H][G]^n}{[HG_n]} \tag{2}$$

Although this equation may appear very undemanding upon first glance, the inclusion of n (the number of host receptor sites) adds a great deal of complexity to the issue: numerous cases exist when adjacent receptor sites influence their nearest neighbor's ability to bind other ligand molecules. It will therefore be intuitive to present individual cases in a sequential manner, starting with the simplest system before concluding with neighboring interactions in multi-site binding.

A.I.1 ONE-SITE BINDING

In the simplest case, where only one binding site exists per host molecule, $n = 1$. It is thus possible to establish the fraction of total binding sites occupied, θ, as

$$\theta = \frac{[G]_{bound}}{[H]_{total}} = \frac{[HG]}{[H]+[HG]} \tag{3}$$

Determining θ is often a trivial practice in spectrophotometric manipulation. In the experiments presented in sections I-III, a specified concentration of guest molecules $[G]_0$ is added to a constant and set concentration of host moieties $[H]_0$. In the cases in which a slow exchange regime is present (see individual sections) observation and integration of the resultant ^{1}H NMR or UV absorbance spectra permits one to readily establish θ.[‡] Given θ, we can then solve Equation 2 for $[HG]$ and substitute the answer into Equation 3 to obtain

$$\theta = \frac{[G]}{K_d + [G]} \tag{4}$$

[‡] Establishing θ by NMR in systems which exhibit fast exchange is much more complex and is discussed in Sections II and III.

This theoretical treatment of one-site binding suggests that a plot of θ versus $[G]$[‡] at a constant concentration of $[H]_{total}$ will yield a hyperbola whose midpoint will surrender K_d. Such a non-linear binding curve is known as the Langmuir isotherm, or the "direct" plot. As it will be instructive to demonstrate graphically various binding models, a sample direct plot is shown in Figure 1a. Here, the association constant was arbitrarily chosen to be 1000 M^{-1} and the system assumed to exhibit the full range of complexation (θ = 0.0% to 99.9%). The respective guest concentrations were thus determined accordingly at regularly spaced intervals.

From the experimental chemist's point of view, the non-linearity of the direct plot is cumbersome in that it is necessary to observe data over an indiscriminate range of [G] in order to establish the midpoint of the hyperbola. Further manipulation of Equation 4 yields two linear forms that are commonly encountered in binding studies. Simply taking the reciprocal of Equation 4 yields the "double-reciprocal" plot:

$$\frac{1}{\theta} = K_d\left(\frac{1}{[G]}\right)+1 \qquad (5)$$

In enzyme kinetic studies, a plot of $1 / \theta$ versus $1 / [G]$ has been coined a Lineweaver-Burk plot [Lineweaver, H.; Burk, D. *J. Am. Chem. Soc.* **1934**, *56*, 658]. For our purposes, the double-reciprocal plot will be called the Benesi-Hildebrand binding curve, as it has commonly been referred to in spectroscopic studies of supramolecular complexes [Benesi, H.; Hildebrand, J. H. *J. Am. Chem. Soc.* **1949**, *71*, 2703]. Figure 1b shows such a plot, utilizing the same data points found in the direct binding curve. Note that the slope of the Benesi-Hildebrand plot readily gives K_d.

‡ [G], which refers to the concentration of uncomplexed guest moieties, is often represented by $[G]_{uc}$ or $[G]_{eq}$. In this manuscrip, $[G]_{uc}$ is preferred and is used interchangeable with [G] and $[G]_{eq}$.

A second conversion of the hyperbola of Equation 3 to a straight line is called the x-reciprocal plot:

$$\frac{\theta}{[G]} = \frac{1}{K_d} - \frac{\theta}{K_d} \tag{6}$$

A plot of θ / [G] versus θ yields a straight line whose slope is -1 / K_d. In enzyme kinetic studies, the x-reciprocal manipulation has been coined the Eadie plot [Eadie, G. S. *J. Biol. Chem.* **1942**, *146*, 85]; in supramolecular binding complexes, this curve is called a Scatchard plot [Scatchard, *G. Ann. N.Y. Acad. Sci.* **1949**, *51*, 660]. An example can be found in Figure 1c.

A.I.1.1 COMPARISON OF THREE BINDING MODELS

As one might expect, the use of linear plots is preferred over non-linear plots if for no other reason than for simplicity's sake. More precisely, the two major advantages of the linear plots over the direct-plot are: 1) the scatter of points about linearity afford a qualitative measurement about the system's binding capability and 2) K_d can be readily obtained simply by calculating the slope of the linear plots. Although both linear forms are derived from the same equation, the derivation of each results in two models that are experimentally different.

The Benesi-Hildebrand plot transforms the otherwise equally spaced data points of the direct plot in such a way that the data points become inhomogeneous: points at high [G] are grouped tightly while points at low [G] become widely spaced apart. This is readily apparent upon observation of Figure 1, where the same data used to construct the direct plot is manipulated to construct the two linear models. The major advantage of the Benesi-Hildebrand plot is that the variables θ and [G] remain independent on the

Figure 1. Representative Binding Curves for Various Models Manipulating the Same Data Set (K_a = 1000 M^{-1}) for Each Plot. a) Direct Plot, b) Benesi-Hildebrand Plot, c) Scatchard Plot.

Figure 1a: Direct Plot

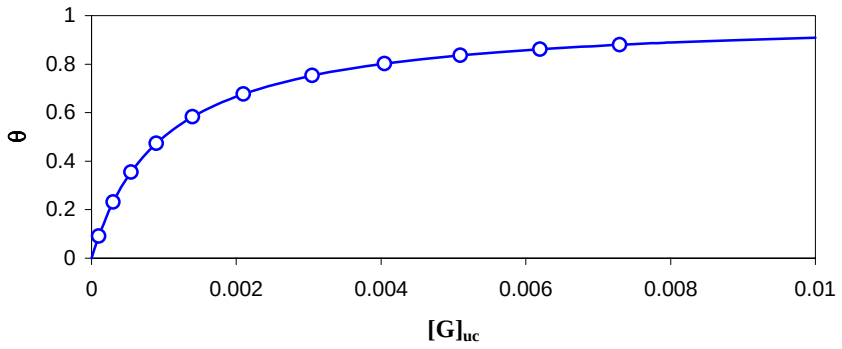


Figure 1b: Benesi-Hildebrand Plot

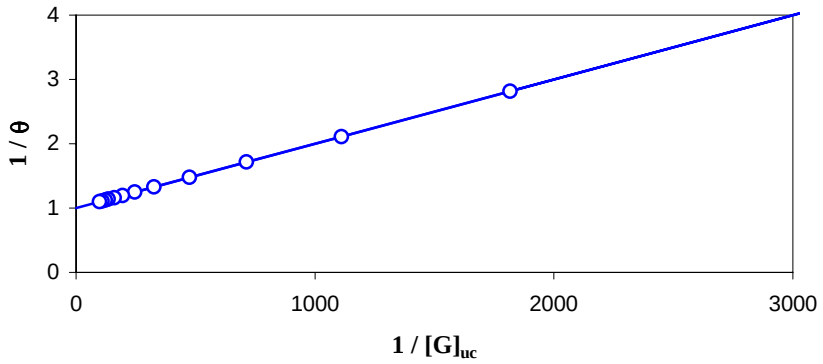
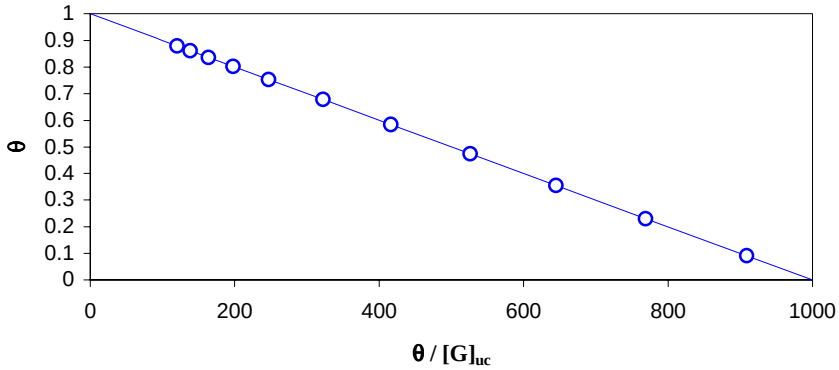


Figure 1c: Scatchard Plot



abscissa and ordinate, whereas other graphical representations allow the variables to become mixed.

The Scatchard plot also transforms the evenly spaced data points of the direct plot in a non-uniform sequence; however it provides a more optimal weighting than does the Benesi-Hildebrand plot [Marshall, A. G. *Biophysical Chemistry: Principles, Techniques, and Applications*. John Wiley & Sons: New York, NY, **1978**, 74]. The Scatchard plot is also useful in establishing multi-site binding events. In addition, the Scatchard plot is usually favored in that the independent axis is θ, thereby readily affording the experimenter information on the range of complexation he/she is looking at. This issue was once at the forefront of describing how well the three models predicted actual binding behavior and will be discussed in the following section.

Arguments as to which graphical representation should be used abound in the literature [Conors, K. A. *Binding Constants: The Measurement of Molecular Complex Stability*. John Wiley & Sons: New York, NY, **1987**, 68]. Common practice dictates experimental data be plotted in several forms and a conclusion consistent with each be drawn.

A.I.2 APPROPRIATE COMPLEXATION BINDING RANGES

It should be evident upon examination of the direct plot (Figure 1a, Equation 4) that if one were given only very low values of the independent variable [*G*], it would be difficult to determine K_a. Initial inspection might lead one to the incorrect observation that binding occurs in a linear fashion; only when one continually loads up the host molecules does the experimenter observe the curvature typical of real one-site binding systems. At the other extreme, it becomes increasingly difficult to follow the behavior of thoroughly loaded host molecules, in which greater than 85% of the binding sites are occupied, due to solubility limitations, non-ideal binding behavior, and other experimental difficulties [Connors, K. A. *Binding Constants: The Measurement of Molecular Complex Stability*. John Wiley & Sons: New York, NY, **1987**, pp. 46-86]. These observations become important because, as Person asserted, it is only possible to

establish that chemical equilibrium exists for systems in which association constants are low ($K_a < 10^4\ M^{-1}$) when binding is observed into the nonlinear regime of the curve [Person, W. B. *J. Am. Chem. Soc* **1965**, *87*, 167]. Person went on to generalize that for 1:1 binding, the guest concentration should be extended to at least the value of $10K_d$. Deranleau built upon this analysis to determine statistically the relative error of K_d in one site binding systems and found the approximate relative error, σ / K, to be:

$$\frac{\sigma}{K} = \frac{\sigma}{\theta(1-\theta)} \tag{7}$$

Figure 2 plots shows a graphical representation of this error in plotting θ versus $1 / (\theta(1\text{-}\theta))$. The sharp increase in error at either extreme significantly limits the effective observable binding range; the relative error in K_d is minimized when θ falls in the range of 0.20 to 0.80.

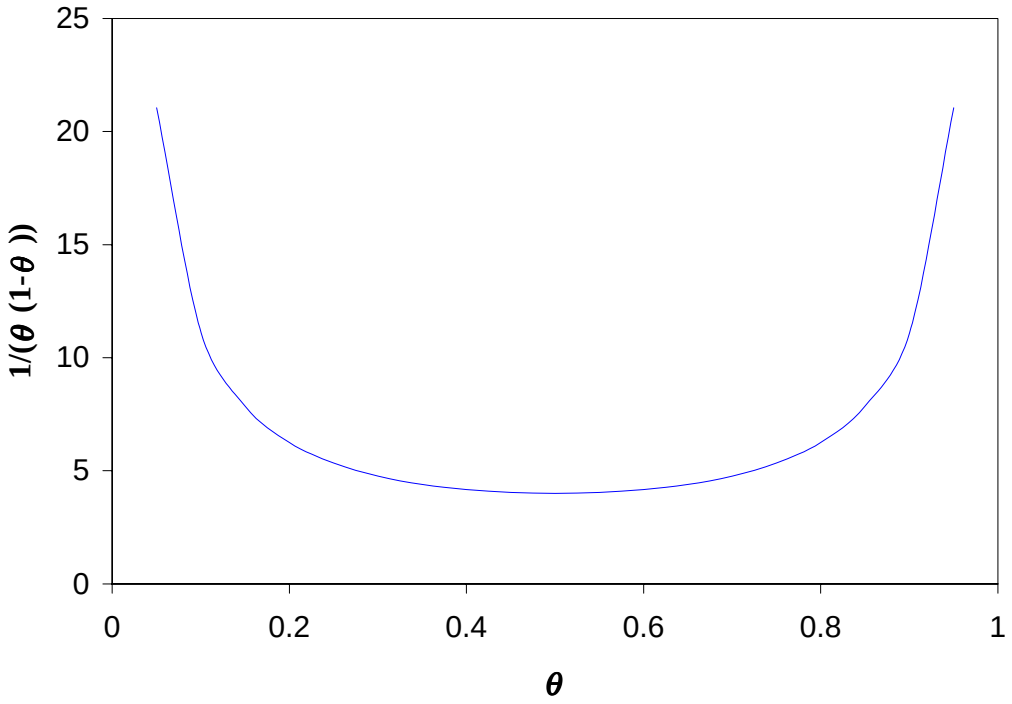
Utilizing the work of Person and Deranleau, Weber demonstrated that given prior knowledge of binding constants, it is possible to determine the proper host and guest concentrations that will enable the experimenter to observe binding in the 20 to 80% loaded range [Weber, G. *Molecular Biophysics*. Pullman, B.; Weissbluth, M., Ed. Academic Press: New York, NY, **1965**, 369-367]. Weber defined the probability of binding as the ratio of the concentration of the complex, [HG] divided by the concentration of the minor component added, either $[H]_0$ or $[G]_0$:

$$p = \frac{[HG]}{[G]_0} \quad \text{if } [H]_0 \geq [G]_0 \tag{8}$$

$$p = \frac{[HG]}{[H]_0} \quad \text{if } [G]_0 \geq [H]_0 \tag{9}$$

For systems that exhibit low equilibrium constants, the host concentration is always the minor component. In these systems, Weber established experimentally that to extend binding into its upper limit, a concentration of guest molecules equal to about ten times K_d was sufficient. Coupled with Person's observation that the lower binding limit occurs when the minor component's concentration is equal to about one tenth of K_d, Weber generalized the following: in order to observe binding within the 20 to 80% regime, the host concentration should remain constant at a value equal to one tenth of the dissociation constant while the guest should be added at concentrations varying between $0.1K_d$ and $10K_d$.

Figure 2. Approximate Relative Error in One-site Binding.



A.I.3 MULTI-SITE BINDING

Although the above models do a good job of describing the binding behavior of the pseudorotaxanes described in section II, they are not well suited for most biological applications, which tend to have more than one binding site. The introduction of multiple binding sites on an individual host molecule carries enormous consequences, particularly when the effectiveness of the binding sites is dependent upon each other. To be independent, individual receptors must not influence binding at any other site on the macromolecule. If interaction among sites does exist, allosteric effects are said to be present. In such cases, the binding behavior of a host changes as a direct consequence of initial interaction with a guest molecule. In typical allosteric interactions, the host's affinity for subsequent guest molecules increases; cases also exist where the host's affinity for bound molecules decreases upon interaction with the initial guest.

Looking first at the case where all binding sites are identical and independent, we can return to Equation 2 to establish the average number of guest molecules bound per host macromolecule as:

$$\theta = \frac{n[\mathrm{G}]}{\mathrm{K_d} + [\mathrm{G}]} \tag{10}$$

Equation 10 is nearly identical to Equation 4. Ergo, the integrity of the one-site binding models is preserved upon extension to identical and independent multi-site binding. Table 1 lists the equivalent models described for one-site binding. The dissociation constant extracted from each of these models is an average over all binding sites and is often referred to as a macroscopic dissociation constant.

It is possible to determine the stepwise binding constants of each individual host receptor as guest molecules bind successively to it. If we let *k* represent the microscopic association constant of binding to a single site, which is constant if all sites are truly identical and independent, we can then define a stepwise binding constant as K_{Gi}, which will vary statistically from receptor to receptor based on the number of sites previously occupied. In other words, given a macromolecular host containing ten receptor sites, it

will be much easier for the first guest added to find a binding site than it will be for each successive guest added. The first guest has *ten* sites from which to choose; the tenth guest has only *one* site on which to bind.

The stepwise association constant can be determined as follows:

$$K_{Gi} = \frac{\#\text{ of unoccupied sites on host}}{\#\text{ of occupied sites}} k \qquad (11)$$

$$= \frac{n-i+1}{i} k$$

Table 2 lists relative stepwise binding constants in the absence of allosteric, or cooperative, effects. Strictly a statistical effect, these stepwise constants decrease as *i* increases.

Table 1. Extrapolation of Equation 10 to Multi-Site Binding Models.

Model	Equation	Plot	Graphical Extractions
Direct Plot	$\theta = \frac{n[G]}{K_d + [G]}$	θ versus [G]	$n = \lim_{[G]\to\infty} \theta$ $K_d = [G]$ for $\theta = n/2$
Benesi-Hildebrand	$\frac{1}{\theta} = \frac{K_d}{n[G]} + \frac{1}{n}$	$1/\theta$ versus $1/[G]$	$n = 1/$(y-intercept) $K_d = -1/$(x-intercept)
Scatchard	$\frac{\theta}{[G]} = \frac{n}{K_d} - \frac{\theta}{K_d}$	$\theta/[G]$ versus θ	n = x-intercept $K_d = \left(\frac{\text{y-intercept}}{n}\right)^{-1}$

Returning to the discussion on identical and independent multi-site binding graphical representations, it is important to note that a deviation from linearity in the Scatchard plot (and to a lesser extent in the Benesi-Hildebrand plot) offers useful information on the nature of binding by the macroscopic host. A curved Scatchard plot suggests that the binding sites are not identical and independent. Cooperative effects for identical and dependent binding sites are discussed below; for examples and treatment of cases in which different types of receptors are available to an individual host macromolecule, the reader is referred to *Physical Biochemistry: Applications to Biochemistry and Molecular Biology* [Freifelder, D. *Physical Biochemistry: Applications to Biochemistry and Molecular Biology*. W. H. Freeman and Co.: San Francisco, CA. **1982**, 660-661].

Table 2. Statistical Influences on K_{Gi} for Multi-Site Hosts with *n* Identical, Independent Binding Receptors. K_{Gi}'s reported as relative values.

n	K_{11}	K_{12}	K_{13}	K_{14}	K_{15}	K_{16}	K_{17}	K_{18}	K_{19}	K_{110}
1	1									
2	2	1/2								
3	3	1	1/3							
4	4	1 1/2	2/3	1/4						
5	5	2	1	1/2	1/5					
6	6	2 1/2	1 1/3	3/4	2/5	1/6				
7	7	3	1 2/3	1	3/5	1/3	1/7			
8	8	3 1/2	2	1 1/4	4/5	1/2	2/7	1/8		
9	9	4	2 1/3	1 1/2	1	2/3	3/7	1/4	1/9	
10	10	4 1/2	2 2/3	1 3/4	1 1/5	5/6	4/7	3/8	2/9	1/10

A.I.3.1 ALLOSTERIC INTERACTIONS

In cases where the one-site binding models discussed above deviate from linearity, it becomes important to investigate the supramolecular system utilizing models developed to describe multi-site binding. A particularly useful multi-site model is known as the Hill plot, the derivation of which is described below.

Making use of Equations 1 and 2, and defining ρ as the number of guest molecules *G* bound per host molecule *H*:

$$\rho = \frac{n[HG_n]}{[HG_n]+[H]} \tag{12}$$

we can solve Equation 2 for [HG_n] and substitute into Equation 12 to obtain:

$$\rho = \frac{n[G]^n}{[G]_n + K_d} \tag{13}$$

Rewritten:

$$\frac{\rho}{n-\rho} = \frac{[G]^n}{K_d} \tag{14}$$

This equation can be simplified in noting that the fraction of sites bound, θ, is simply the number of sites occupied, ρ, divided by the number of sites available, n (see Equation 3).

$$\theta = \frac{\rho}{n} \tag{15}$$

Thus, Equation 14 becomes

$$\frac{\theta}{1-\theta} = \frac{[G]^n}{K_d} \tag{16}$$

which is known as the Hill equation.

Transforming Equation 15 by taking the log of both sides, we arrive at the Hill plot:

$$\log\left(\frac{\theta}{1-\theta}\right) = n_H \log[G] - \log K_d \tag{17}$$

A plot of log [θ / (1-θ)] versus log [G] will yield a straight line with slope n_H, termed the Hill coefficient, and intercept –log K_d. The Hill coefficient is a qualitative measure of the degree of cooperativity and it is experimentally less than the actual number of binding sites present, n—an important distinction which will be discussed shortly. If $n_H > 1$, the system is said to be positively cooperative: the addition of successive guest molecules is facilitated by prior loading. Returning to the hemoglobin example in which four molecules of O_2 are bound per strand of hemoglobin, the Hill coefficient has been found to be equal to 2.8 [Freifelder, D. *Physical Biochemistry: Applications to Biochemistry and Molecular Biology*. W. H. Freeman and Co.: San

Francisco, CA. **1982**, 660-661]. Cooperativity suggests that once the first O_2 molecule binds, the remaining three bind so quickly it is as if all four have attached to hemoglobin at once. If $n_H < 1$, the system is said to be anti-cooperative: the addition of each successive guest molecules is hindered by prior loading. In other words, it becomes increasingly more difficult to add guest molecules to anti-cooperative systems once the first ligand is bound.

Because Equation 17 assumes $n_H = n$, the Hill equation does not describe the real situation precisely. To be sure, when a Hill plot is constructed over a wide range of $[G]_{uc}$, a curve similar to that found in Figure 3 is observed. At the extremes of $[G]_{uc}$, the continuity of the Hill plot is broken. More specifically, the slope at either end is approximately one. This observation can readily be explained by the following: when the ligand concentration is either very low or very high, cooperativity does not exist. It seems reasonable to suggest that in the low concentration case, individual guest ligands are more likely to find non-coordinated host molecules than they are to occupy successive sites on a pre-coordinated host: single-site binding occurs. At the other extreme, every receptor site but one is filled. Thus, at high concentration, the Hill plot again represents single-site binding. It falls directly from this line of reasoning that Hill plots are more adept at describing cooperativity in molecules that exhibit a large number of receptor sites (see Figures 3 and 4).

The discerning reader will have made the observation that when $n_H = 1$, identical and independent multi-site binding exists. Indeed, one of the major advantages of the Hill plot is that it can be used to clearly indicate when cooperativity is *not* present. On the flip side, the Hill plot can *only* be used to establish the presence of cooperativity if the total number of binding sites is known. Again, Hill plots say nothing quantitative about binding constants; they merely provide an indicator of allosteric effects.

Figure 3. A Hypothetical Hill Plot for Cooperative Binding of Four Molecules of O_2 to a Single Hemoglobin Strand. The First and Fourth Additions Account for 50% of Overall Binding.

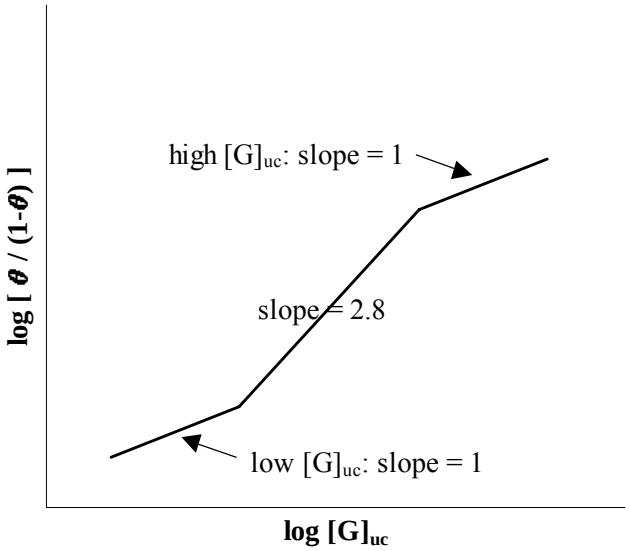


Figure 4. A Hypothetical Hill Plot for Cooperative Binding of 16 Guest Molecules to a Single Host Bearing 16 Receptor Sites, Displaying the Same Hill Coefficient as Hemoglobin. The First and Sixteenth Additions Account for 12.5% of Overall Binding.