Biophysical Chemistry 8 (1978) 327–339

COOPERATIVE AND NON-COOPERATIVE BINDING OF LARGE LIGANDS TO A FINITE ONE-DIMENSIONAL LATTICE. A MODEL FOR LIGAND-OLIGONUCLEOTIDE INTERACTIONS

Irving R. EPSTEIN *
Max-Planck-Institut fur biophysikalische Chemie, D-3400 Gottingen, W. Germany

Received 13 April 1978; revised manuscript received 25 May 1978

A combinatorial approach is employed to calculate exact expressions for the extent of binding to a finite one dimensional lattice of ligands which cover more than one lattice site. The binding may be either cooperative or non-cooperative. It is found that the assumption of an effectively infinite lattice is generally a good one, except with relatively low concentrations of strongly cooperative ligands. An approach to analyzing experimental data is suggested which makes explicit use of the lattice length dependence of binding to extract more information about the binding parameters than can be obtained using the infinite lattice approximation. It is shown that irreversible binding cannot be viewed as a limiting case of reversible binding. The reasons for this difference are discussed, and expressions for the extent of irreversible binding are derived.

1. Introduction

A large number of biochemically significant reactions involve the binding of ligands to a molecule which constitutes an essentially one-dimensional lattice. In many of these reactions, e.g., when the lattice is an oligo- or polynucleotide and the ligand is a basic polypeptide, the binding may be primarily electrostatic, and hence non-specific, so that the lattice may be regarded as homogeneous. A further feature of such processes is that the ligands are often *large*, in the sense that when a particular lattice site** is covered by a bound ligand, one or more contiguous sites are then unavailable for the binding of further ligand molecules. The binding characteristics of such large ligands are quite different from those of small ligands which occupy only a single site.

The differences between large and small ligands have been thoroughly discussed by McGhee and von Hippel [1], who developed a theory to describe the extent of binding, either non-cooperative or cooperative, of large ligands to an infinite one dimensional lattice. Schwarz [2] has recently presented a more general discussion of such systems as well as considering more complex combinations of binding processes.

Recent experimental studies of the binding of E. coli ribosomal protein S1 [3] and of T4 gene 32-protein [4] to relatively short (~2–20 bases) oligonucleotides have provided important insights into the possible *in vivo* functioning of these proteins. With the improvement of techniques [5] for the preparation of pure, well characterized oligonucleotide sequences, it would seem desirable to have a clear theoretical understanding of large ligand binding to lattices in which the infinite lattice approximation may be expected to break down. Also, a more precise characterization of when a lattice may be taken to be "effectively infinite" should be of some interest.

Latt and Sober [6] have analyzed the non-cooperative binding of large ligands to a finite lattice using a combinatorial theory of ligand binding. While com-

* Permanent address: Department of Chemistry, Brandeis University, Waltham, Mass. 02154, USA.

** We shall use the term *site* throughout this paper to refer to a single element of the lattice. A collection of n adjacent sites, required in order for an n-site ligand to bind, will be referred to as a binding region or simply, a *region*.

binatorial methods are of little use in dealing with infinite lattices, they are ideally suited for finite lattices, where the conditional probability approach of McGhee and von Hippel [1] or other approaches based upon doublet closure [8] become unwieldy, owing to the necessity of treating each lattice site separately. The alternative and powerful method of sequence generating functions [9] also becomes inapplicable when the effects of a finite lattice must be taken into account.

In this paper, we use a combinatorial approach to derive exact expressions for the cooperative as well as non-cooperative binding of large ligands to a finite lattice of arbitrary length. We then look at how rapidly the results approach the infinite lattice limit as the lattice length is increased. We indicate how our equations might be used to extract equilibrium constants, ligand lengths and cooperativity parameters from experimental binding data. Situations in which more than one type of ligand may bind or, alternatively, in which the same ligand may bind in two or more different ways to the lattice are also considered.

Finally, we investigate the *irreversible* binding of large ligands or the equivalent process of the irreversible transformation of a lattice in blocks comprising several adjacent sites. By extending an approach first employed by Flory [10], we obtain results which show that the extent of irreversible binding is *not* the limit of the extent of reversible binding as the binding constant and/or the free ligand concentration approaches infinity. This apparently paradoxical result is interpreted in kinetic terms. The differences between long and short ligand binding extend, of course, to kinetic as well as to thermodynamic properties, and we shall treat the kinetics of long ligand-lattice interactions elsewhere.

2. Reversible binding

2.1. Definitions and fundamental relations

We shall consider the binding of a ligand of length $n \geq 1$ to a lattice which contains M sites. In order for binding to occur, there must be n adjacent empty sites (a binding region) available on the lattice. For the sake of definiteness, we number the lattice sites 1 to M from left to right and also take the binding of an individual ligand to proceed in this direction. Then, binding a ligand *at site q* (i.e., in the region which "starts at" site q, and includes sites $q, q+1, \ldots, q+n-1$) implies that ligands may no longer be bound at sites $q-n+1$ through $q+n-1$, a total of $2n-1$ sites. It is this exclusive or "multiple contact" [2] binding which, for $n > 1$, gives rise to the special nature of large ligand-lattice interactions. We note that the above description does not necessarily imply that the ligand *binds* to n distinct monomers in the lattice, but only that by virtue of binding, steric, or other factors it requires and renders unavailable for further binding n consecutive sites. These assumptions imply that the rightmost $n-1$ positions of the lattice may not be utilized as starting points for ligand binding, though they may, of course, be occupied by ligands which start binding at points further from the end of the lattice. Further comments on end effects will be found in sections 2.5 and 2.6.

We define a *binding constant* K which characterizes the binding of a ligand to an *isolated* binding region, i.e., to a group of n unoccupied sites bordered at each end by at least one further unoccupied site. If we let u signify an unoccupied site, (L_n) a free and (l_n) a bound ligand molecule, then K is the equilibrium constant for the association reaction

$$\ldots \underbrace{\mathrm{uu} \ldots \mathrm{uu}}_{(n+2)} \ldots + (L_n) \rightleftharpoons \ldots \mathrm{u}\,(l_n)\,\mathrm{u} \ldots \qquad (1)$$

We also define a *cooperativity parameter* ω such that the equilibrium constant for binding a ligand to a *singly contiguous region*, i.e., for the reaction

$$\ldots (l_n)\underbrace{\mathrm{u} \ldots \mathrm{uu}}_{(n+1)} \ldots + (L_n) \rightleftharpoons \ldots (l_n)(l_n)\mathrm{u} \ldots \qquad (2a)$$

or

$$\ldots \underbrace{\mathrm{uu} \ldots \mathrm{u}}_{(n+1)}\,(l_n) \ldots + (L_n) \rightleftharpoons \mathrm{u}(l_n)(l_n) \ldots \qquad (2b)$$

is $K\omega$. Consistency requires that the equilibrium constant for binding to a *doubly contiguous region*, i.e., one bordered on both sides by a bound ligand, be $K\omega^2$. Values of ω greater than 1 define a cooperatively interacting ligand, $\omega = 1$ implies non-cooperativity, and $\omega < 1$ corresponds to anticooperativity. We as-

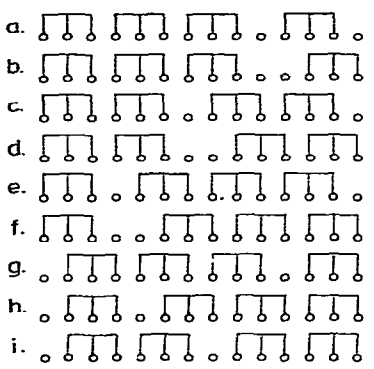

Fig. 1. The nine distinct configurations with $M = 14$, $n = 3$, $k = 4, j = 2$.

sume that binding constants are unaffected by the states of any sites other than those of the immediate nearest neighbors.

We further define

$$g = [M/n] \tag{3}$$

where $[M/n]$ signifies the greatest integer less than or equal to M/n, as the maximum number of ligands which may bind to the lattice. Let

$$r = M - ng \tag{4}$$

be the number of lattice sites remaining when the lattice is fully covered, $0 \leqslant r \leqslant n - 1$. Finally, let the free ligand concentration be denoted by L.

Now consider the various possible arrangements of ligands on the lattice. Any given configuration may be characterized by the number of ligands bound, k, and by the number of *adjacencies*, j. The number of adjacencies is defined as the number of times, starting from the left end of the lattice, that we encounter a ligand with its right end adjacent to the left end of another bound ligand. Clearly, $j \leqslant k - 1$. Fig. 1 illustrates the nine possible configurations with $k = 4$, $j = 2$ for a 3-site ligand binding to a 14-site lattice. Any equilibrium property of the ligand-lattice system should be calculable as an appropriately weighted average over all possible configurations. The number of configurations, however, grows exponentially with M. Since all binding (as opposed, for example, to kinetic) properties must be the same for configurations, such as those in fig. 1, with equal values of k and j, the problem may be rendered tractable by considering separately only *sets* of configurations with distinct k and j values. The task then becomes that of calculating the statistical weight of each set (k, j).

It is clear from our assumption of non-specific binding and from our definitions above, that the relative probability of any single configuration in the set (k, j) is just $(KL)^k \omega^j$, where we assign unit relative probability to the reference state of a naked lattice $(k, j) = (0, 0)$. What we now need to know is the number of configurations in each set (k, j). Let us denote this quantity as $P_M(k, j)$, i.e., $P_M(k, j)$ is the number of distinct ways that k n-site ligands may bind to an M-site lattice with j adjacencies. If the value of n is not obvious from the context, we may include it as a superscript on $P_M^n(k, j)$.

If the $P_M(k, j)$ are available, then we may express, for example, the average number of bound ligands as

$$\bar{k} = \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} k P_M(k, j) (KL)^k \omega^j \right\} \times \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} P_M(k, j) (KL)^k \omega^j \right\}^{-1}. \tag{5}$$

We shall be particularly interested in the extent of binding or the fraction of occupied sites θ, defined by

$$\theta = n\bar{k}/M. \tag{6}$$

Another quantity of potential experimental interest [1,11,12] is the average length of *runs*, where a run is defined as the number of consecutive sites covered by bound ligand. For example, in fig. 1, configurations a, b, e, f, g and h have run lengths of 9 and 3, while configurations c, d, and i each has two runs of length 6. The average run length for every configuration in fig. 1 is thus 6. The number of runs in any configuration is just $k - j$. Therefore if we define an average number of adjacencies $\bar{j}$ by

$$\bar{j} = \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} j P_M(k, j) (KL)^k \omega^j \right\} \times \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} P_M(k, j) (KL)^k \omega^j \right\}^{-1},$$

the average run length $\bar{l}$ is given by

$$\bar{l} = n\bar{k}/(\bar{k} - \bar{j}) = M\theta/(\bar{k} - \bar{j}). \tag{8}$$

Other quantities such as the probability of occurrence of any given number of bound ligands, adjacencies or runs, may be similarly calculated.

Thus, if the $P_M(k,j)$ are found, all quantities of interest with respect to binding may be written as ratios of polynomials in the free ligand concentration, where the coefficients depend upon the binding constant and the cooperativity parameter, and the degree of the polynomials is the maximum number of ligands which may be bound. We now turn to the determination of the $P_M(k,j)$.

2.2. Calculation of the $P_M(k,j)$

We wish to determine the number of distinct ways of distributing k n-site ligands on an M-site lattice with precisely j adjacencies. We shall obtain this quantity in two different ways, first by a purely combinatorial analysis which yields a multiple sum, and then by an approach based on the notion of building up the lattice from smaller lattices, from which we derive recursion relations.

2.2.1. Combinatorial analysis

If we are to have k bound ligands with j adjacencies, we must have exactly $k-j$ distinct runs on the lattice. In order to keep the runs separated, each of the $k-j-1$ leftmost runs must be terminated at its righthand end by an "attached" vacant site. The remaining vacant sites may then be arbitrarily distributed on the lattice along with the runs and their attached sites. The number of such unattached vacancies is just

$$N_u = M - nk - (k-j-1) \tag{8}$$

Therefore, if we consider the $k-j$ runs as one class and the N_u unattached vacant sites as a second class of independent element, the number of possible configurations is just

$$N_c = \frac{(N_u + k - j)!}{N_u!(k-j)!}. \tag{9}$$

In obtaining eq. (9) we have treated the $k-j$ runs as being identical, ignoring the fact that they will in general be of different lengths. To calculate $P_M(k,j)$, we must multiply N_c by the number of different ways in which the run lengths may be arranged, i.e., by the number of different ways in which the integer k may be partitioned into $k-j$ positive integers. In the example of fig. 1, we have $k=4, j=2$, and the three possible partitions are $3+1$, $2+2$, and $1+3$. The partition $3+1$ corresponds, for example, to the structures a, b, and g in which a run of 3 ligands is followed by a run of 1 ligand. Each of the other partitions also corresponds to $N_c = 3$ configurations.

The number of distinct partitions of the integer k into $k-j$ positive integers is given by

$$N_p = \frac{(k-1)!}{j!(k-j-1)!}. \tag{10}$$

Combining eqs. (8)–(10), we finally obtain

$$P_M(k,j) = N_c N_p = \frac{(M-nk+1)!(k-1)!}{(M-nk-k+j+1)!(k-j)!j!(k-j-1)!} \tag{11}$$

In the case of non-cooperative binding, $\omega = 1$, the number of adjacencies j is irrelevant for most purposes. The k ligands and the $M-nk$ unoccupied sites are now all independent units and the expression (11) becomes

$$P_M(k) = \frac{(M-nk+k)!}{(M-nk)!k!}. \tag{12}$$

On suppressing the index j and the sum over j in eq. (5), we obtain the equation given by Latt and Sober [6], with the generalization that M need not be an integral multiple of n.

2.2.2. Recursion relations

The analysis leading to recursion relations for the $P_M(k,j)$ is somewhat more direct. We ask in how many ways we may obtain a configuration which contributes to $P_M(k,j)$ by adding on either unoccupied sites or bound ligands to configurations of smaller lattices. Without loss of generality, we may take this addition to occur at the right hand end of the lattice, and we partition $P_M(k,j)$ into two components

$$P_M(k,j) = P_M(k,j)^+ + P_M(k,j)^-, \tag{13}$$

where the plus superscript indicates that the last (right-most) site of the lattice is occupied and the minus that it is empty. (In fig. 1, $P_{14}(4,2)^+ = 6$, $P_{14}(4,2)^- = 3$.)

On reflection it becomes clear that there are just

three ways in which shorter lattices can give rise to configurations which contribute to $P_M(k,j)$:

a) We add an unoccupied site to an $(M-1)$ site lattice with a (k,j) configuration. Thus $P_{M-1}(k,j)^+$ and $P_{M-1}(k,j)^-$ contribute to $P_M(k,j)^-$.

b) We add a bound ligand to an $(M-n)$ site lattice with a $(k-1,j)$ configuration and the last site unoccupied. Thus $P_{M-n}(k-1,j)^-$ contributes to $P_M(k,j)^+$.

c) We add a bound ligand to an $(M-n)$ site lattice with a $(k-1,j-1)$ configuration and the last site occupied. Thus $P_{M-n}(k-1,j-1)^+$ contributes to $P_M(k,j)^+$.

Possibilities a)–c) are disjoint and inclusive. We therefore have

$$P_M(k,j)^+ = P_{M-n}(k,j)^+ + P_{M-n}(k-1,j-1)^+, \quad (14)$$

$$P_M(k,j)^- = P_{M-1}(k,j)^- + P_{M-1}(k,j)^+. \quad (15)$$

Combined with eq. (13) and with the conditions

$$\begin{aligned} P_M(0,0)^+ &= 0, \\ P_M(0,0)^- &= 1, \quad \text{if } M \geqslant 0 \\ P_M(k,j) &= 0, \quad \text{if } j \text{ or } k < 0 \\ &\quad \text{or} \quad \text{if } j \geqslant k \\ &\quad \text{or} \quad \text{if } nk \geqslant M \end{aligned} \quad (16)$$

eqs. (14) and (15) suffice to determine all the $P_M(k,j)$. We note that since the matrices $P_M(k,j)$ are independent of the binding and cooperativity parameters, once a ligand length and maximal lattice length are chosen, the $P_M(k,j)$ may be constructed once and then stored for any future calculations with other parameter values.

2.3. Dependence of θ on lattice length

The assumption of an infinite lattice is, of course, always an approximation. Nevertheless, the equations given by McGhee and von Hippel [1] and by Schwarz [2] make the infinite lattice approximation an attractive one for the analysis of experimental data. It is therefore desirable to know how long the lattice must be in order for the infinite lattice approximation to hold to a desired degree of accuracy. In order to answer this question, as well as to study the nature of the approach of the extent of binding θ to its infinite lattice limit θ_∞, we have carried out a number of calculations of θ as a function of lattice length M for various combinations of the parameters n, KL, and ω.

As previously suggested [1], $\theta(M)$ generally approaches θ_∞ from below as the lattice length is increased, though for anticooperative ligands in relatively high concentrations (e.g., $n=2, KL=10$, $\omega=0.2$), the approach can be from above. For noncooperative and anticooperative ligands, θ tends to approach its infinite lattice limit monotonically with increasing M, after some initial fluctuations, but for cooperative ligands, $\theta(M)$ exhibits pronounced oscillations of period n about an envelope which increases toward the asymptotic limit θ_∞. This behavior results from the tendency of highly cooperative ligands to cluster into a single run which efficiently covers the lattice, leaving only the r (eq. 4) sites which cannot be filled. We therefore expect, and find, that the oscillations will have maxima for $r=0$ (i.e., M an integral multiple of n) and minima for $r=n-1$, and that they will be most pronounced for large values of ω. We show in fig. 2 examples of the typical monotonic increase of $\theta(M)$ for ligands without extreme positive or negative cooperativity, the anomalous decreasing θ function for an anticooperative ligand, and the characteristic oscillations of a highly cooperative binding process.

The question of when the infinite lattice approximation becomes satisfactory depends strongly upon the parameters of the binding process. In table 1, we give values of g_{25} and g_{10}, which are defined as those values of g (eq. 3), the maximum number of ligands which may be bound, such that for the given values of n, KL and ω, $\theta(M)$ is within 25% of θ_∞ if M/n exceeds g_{25} and within 10% if $M/n > g_{10}$. In general, we find that the infinite lattice approximation should be satisfactory in most experiments even for fairly short lattices. There does appear to be one major exception, however. When we have relatively low concentrations of a strongly cooperative ligand ($KL \ll 1$, $\omega \gg 1$, $KL\omega \approx 1$), the approach to the infinite lattice limit becomes quite slow. For a 5-site ligand, for example, with $\omega = 10^2$ and $KL = 10^{-2}$, even a 75-site lattice gave only 72% of the infinite lattice value of θ, while no other combination of parameters investigated for this value of n gave a g_{25} value greater than 6. Experimentalists who wish to employ the infinite lattice expressions in analyzing

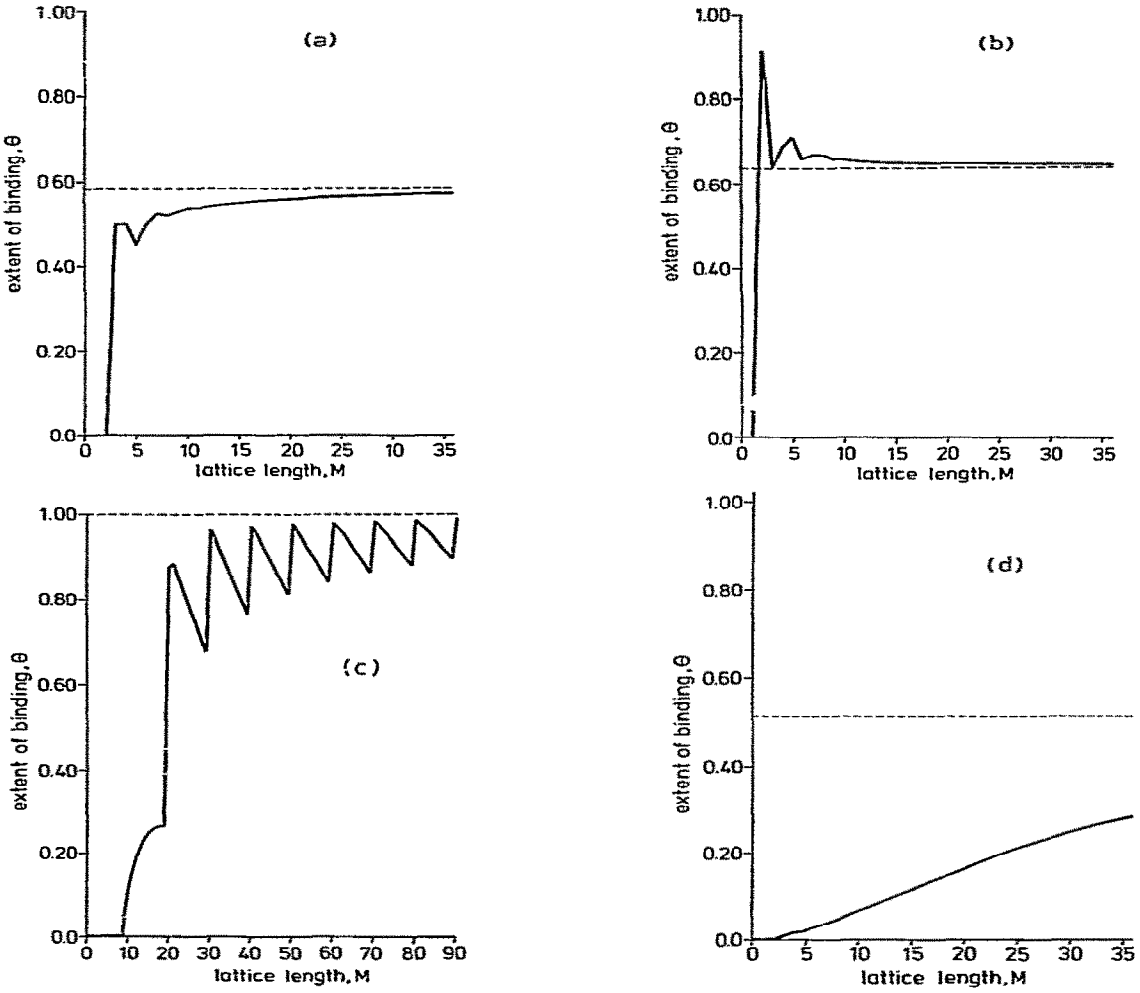

Fig. 2. Dependence of the extent of binding, θ, on the lattice length M. a) A typical case with weak or no cooperativity, $n = 3$, $KL = 1.0$, $\omega = 1.0$; b) Anticooperative binding, $n = 2$, $KL = 10$, $\omega = 0.2$. c) Strongly cooperative binding showing oscillatory approach of $\theta(M)$ to θ_∞, $n = 10$, $KL = 0.1$, $\omega = 10^3$; d) Slow approach of $\theta(M)$ to θ_∞ for low concentrations of strongly cooperative ligands, $n = 3$, $KL = 10^{-2}$, $\omega = 10^2$. In all cases, dashed line indicates value of θ_∞.

binding to oligomers should be careful to choose ligand concentrations which avoid this regime. This anomalously slow approach to θ_∞ appears to be associated with a tendency of the bound ligands to cluster together in the center of the lattice and to avoid occupying the end sites.

Along with their derivation of the equations for the infinite lattice limit θ_∞, McGhee and von Hippel [1] also present expressions for estimating θ for finite lattices from θ_∞ in the limits $\theta \to 0$ and $\theta \to 1$. We find that if $\theta_\infty \approx 0.75$, then the McGhee-von Hippel approximation

$$\theta(M) \approx \frac{ng}{M}\theta_\infty \quad \text{for } \theta \to 1 \tag{17}$$

Table 1
Minimum lattice lengths required for 25% and 10% accuracy of infinite lattice approximation to θ.

N	KL	ω	g_{25} [a]	g_{10} [a]	θ_∞
2	0.01	0.2	2	5	0.019
		1.0	2	5	0.019
		10.0	3	7	0.023
		100.0	>25 [b]	>55 [b]	0.508
		1000.0	4	7	1.000
	0.1	0.2	1	3	0.139
		1.0	2	4	0.155
		10.0	7	17	0.523
		100.0	3	7	0.997
		1000.0	2	5	1.000
	1.0	0.2	1	1	0.420
		1.0	1	2	0.558
		10.0	3	6	0.970
		100.0	2	5	0.999
		1000.0	2	5	1.000
	10.0	0.2	1	3	0.634
		1.0	2	3	0.844
		10.0	2	5	0.994
		100.0	2	5	1.000
		1000.0	2	5	1.000
5	0.01	0.2	3	8	0.045
		1.0	3	8	0.046
		10.0	4	9	0.054
		100.0	>18 [b]	>43 [b]	0.520
		1000.0	4	10	0.999
	0.1	0.2	2	5	0.253
		1.0	3	6	0.273
		10.0	6	12	0.550
		100.0	4	10	0.991
		1000.0	3	8	1.000
	1.0	0.2	1	3	0.545
		1.0	1	4	0.619
		10.0	3	7	0.934
		100.0	3	8	0.998
		1000.0	3	8	1.000
	10.0	0.2	1	3	0.708
		1.0	2	3	0.810
		10.0	3	7	0.980
		100.0	3	8	0.999
		1000.0	3	8	1.000
10	0.01	0.2	4	9	0.083
		1.0	4	9	0.085
		10.0	4	11	0.098
		100.0	>13 [b]	>33 [b]	0.532
		1000.0	4	12	0.999
	0.1	0.2	2	6	0.356
		1.0	3	6	0.373
		10.0	4	10	0.582
		100.0	4	11	0.983
		1000.0	3	9	1.000

Table 1 (*continued*)

N	KL	ω	g_{25} [a]	g_{10} [a]	θ_∞
	1.0	0.2	2	3	0.620
		1.0	2	4	0.664
		10.0	3	6	0.901
		100.0	3	9	0.996
		1000.0	3	9	1.000
	10.0	0.2	2	4	0.752
		1.0	2	3	0.808
		10.0	3	7	0.964
		100.0	3	9	0.998
		1000.0	3	9	1.000

[a] See text for definition.
[b] Estimated.

is accurate to better than 10% for essentially all values of $M(g > 2)$. The low binding limit

$$\theta(M) \approx \frac{M - n + 1}{M}\,\theta_\infty \qquad \text{for } \theta \to 0 \tag{18}$$

appears to hold whenever $\theta_\infty \lesssim 0.15$.

2.4. Analysis of experimental data

The above results show that for a large number of experiments, analysis using the infinite lattice approximation should suffice. This is particularly reassuring, since experiments with relatively long lattices generally employ an indeterminate mixture of lattice lengths. We should emphasize again, however, that for strongly cooperative ligands binding to moderately long lattices, care should be taken to avoid making measurements at concentrations such that $L \approx (K\omega)^{-1}$ if the infinite lattice expressions are to hold. Of course, in such cases one can always employ the finite lattice results derived here.

One problem in analyzing data within the infinite lattice approximation has recently been noted by Schwarz [2]. If one measures the extent of binding as a function of the free ligand concentration and attempts to extract all the relevant binding parameters, a serious difficulty arises. The quantities n and ω have countervailing effects on the binding [1], and analysis of Scatchard plots and expressions for their slopes and intercepts at the high and low saturation extremes shows [2] that only the ratio n/ω, but not the individual parameters, may be extracted with any accuracy from typical experimental data. We may ask whether the additional information provided by measurements on finite lattices of known length might enable us to infer accurate, independent values of n and ω.

Suppose that one is able to measure θ as a function of L for several different (and known) values of the lattice length M. In particular, suppose that these measurements may be carried out at rather low values of θ, i.e., for $KL \ll 1, 1/\omega$. Under these conditions, the influence of the cooperativity parameter ω is slight; that is, we need consider only the terms with $k = 0$ and $k = 1$ ($j = 0$) in eq. (5). Using eq. (11) we find that

$$P_M(0,0) = 1, \qquad P_M(1,0) = M - n + 1 \tag{19,20}$$

If we now differentiate eq. (6) with respect to L and let KL be small, we obtain the limiting value s_0 for the slope of a plot of θ versus L as both quantities approach zero:

$$s_0 = \lim_{\substack{KL \to 0 \\ \theta \to 0}} \frac{d\theta}{dL} = \frac{nK}{M}\,\frac{P_M(1,0)}{P_M(0,0)} = \frac{nK(M - n + 1)}{M}. \tag{21}$$

If s_0 values are obtained for several different lattice lengths, then a plot of s_0 versus $1/M$ has slope $nK(1 - n)$ and intercept nK, so that n and K may be determined. Having found n, one then calculates the coefficients $P_M^n(k,j)$. Using eqs. (5), (6) and (11), ω and, if desired, an improved value of K may be com-

puted by a least-squares fit or other procedure from the measured values of θ as a function of L.

We also note that in the limit of a saturated lattice ($KL \gg 1/\omega$), $d\theta/dL$ approaches zero and θ approaches ng/M, which is also independent of ω. Measurement of this limiting value at several preferably small and consecutive values of M will also yield the value of n, but this limit may be experimentally more difficult to reach.

2.5. End effects

In treating binding to a finite lattice, it is necessary to make some assumption about the nature of binding at the ends of the lattice. The effects of this choice disappear as the length of the lattice is increased, but they may be significant for finite lattices. We have simply assumed that no binding is possible unless the ligand has a full region of n sites available at the end of the lattice. An alternative and more general assumption is that ligands can bind to fewer than n sites at the end, but with some reduced binding constant K_e. One might even permit binding to $n-1$, $n-2$, ..., 1 end sites with each type of binding having its own distinct binding constant. While such an approach may be somewhat closer to reality, and does not render the combinatorial analysis totally intractable, it does introduce one or more additional parameters, which would probably be difficult, if not impossible, to get at experimentally. Perhaps a more appropriate procedure in some cases would be to allow for the possibility of imperfect binding or "wobble", i.e., binding as if the ligand had fewer than n sites, anywhere on the lattice including the ends. Such a situation would be equivalent to an experiment involving two or more different ligands and would be susceptible to the analysis given in the following section.

In deriving our recursion relations, eqs. (14)–(16), we divided $P_M(k,j)$ into two components (eq. 13) according to whether or not a particular configuration had the last lattice site occupied. We can use this information to calculate the probability, as a function of lattice length, that the end position is occupied. If we call this probability R_M, we have (cf. eqs. 5, 11, 13)

$$R_M = \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} P_M^{+}(k,j)(KL)^k \omega^j \right\} \times \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} P_M(k,j)(KL)^k \omega^j \right\}^{-1} \tag{22}$$

One might expect that R_M should approach θ_∞ as $M \to \infty$. This is not the case. R_M does approach a limiting value R_∞, but in general we find that

$$\theta_\infty/n < R_\infty < \theta_\infty. \tag{23}$$

To account for the somewhat surprising result of eq. (23) we must first recognize that the reason that binding properties approach limiting values as $M \to \infty$ is not that the end positions begin to resemble the middle, but that the contribution made by the two end positions becomes negligible with respect to that of the $M-2$ interior sites. *The end positions are always different.* In effect, the growth of the lattice occurs in the middle. Eq. (23) can now be explained by observing that a site in the interior of the lattice may be occupied by ligands which "start" binding at any of n sites (the site in question and its $n-1$ iefthand neighbors), while the last site is occupied only if a ligand binds n sites from the end. On this basis R_∞ should be roughly θ_∞/n. However, there is also a factor which tends to favor binding at the lattice ends. Binding at interior sites may be blocked by ligands bound at any of the $n-1$ positions on either side. For the last allowed binding site however, ligands cannot bind to its right, so the probability of its being available is increased by nearly a factor of 2. We thus expect, ignoring cooperativity effects, $R \approx 2\theta/n$, which is in agreement with eq. (23) and with our results for $\omega = 1$.

The only exceptions we find to eq. (23) are in cases of very high cooperativity and relatively low KL, in which the few ligands bound tend to cluster in the lattice interior, and R falls below θ/n. These cases also correspond to the very slow convergence of θ to θ_∞ with increasing lattice length (e.g., fig. 1d).

Whether the special nature of the end sites, even in very long lattices, may have an effect on the behavior of polynucleotide lattices which have particular functional sites located near one end, remains to be seen.

2.6. More than one type of binding

In many cases of experimental interest one wishes to be able to predict the extent of binding of each of

two or more different ligands to a lattice. Such cases include actual competition experiments in which a dye molecule or simple ion is present in addition to a polypeptide ligand [13], situations in which aggregation of ligands before binding may be significant, and studies in which partial binding of ligands, as discussed in the previous section, may play a role.

While the methods of section 2.2 may be extended in a straightforward, if tedious, way to the binding of an arbitrary number of ligands all of which have cooperative interactions with one another, it is clear that the proliferation of parameters in such an approach will render the results of minimal use to the experimentalist. We limit ourselves, therefore, to consideration to two relatively simple situations: the binding of several noncooperative ligands, and a competition experiment between an n-site cooperative ligand and a short non-cooperative ligand (e.g. a sodium ion [6]).

If there are q noncooperative ligands each with length n_i, binding constant K_i and free ligand concentration L_i, we may extend the analysis of sections 2.1–2.2 directly. We can define the maximum number of bound ligands of type i as

$$g_i = [M/n_i], \quad i = 1, 2, ..., q, \tag{24}$$

and the extent of binding of each ligand by

$$\theta_i = n_i \bar{k}_i / M, \quad i = 1, 2, ..., q, \tag{25}$$

where

$$\bar{k}_i = \left\{ \sum_{k_1=1}^{g_1} \sum_{k_2=1}^{g_2} \cdots \sum_{k_q=1}^{g_q} k_i P_M(k_1, k_2, ..., k_q) \times \prod_{f=1}^{q} (K_f L_f)^{k_f} \right\} \times \left\{ \sum_{k_1=1}^{g_1} \sum_{k_2=1}^{g_2} \cdots \sum_{k_q=1}^{g_q} P_M(k_1, k_2, ..., k_q) \times \prod_{f=1}^{q} (K_f L_f)^{k_f} \right\}^{-1} \tag{26}$$

To obtain the statistical weights $P_M(k_1, k_2 \ldots k_q)$, i.e., the number of configurations containing k_i ligands of type i, $i = 1, 2, ..., q$, we observe that each ligand and each of the $M - \Sigma n_i k_i$ vacant sites represents an independent unit. The analog of eq. (12) is thus

$$P_M(k_1, k_2, ..., k_j) = \left\{ \left(M - \sum_{i=1}^{q} (n_i - 1) k_i \right)! \right\} \times \left\{ \left(M - \sum_{i=1}^{q} n_i k_i \right)! \prod_{i=1}^{q} k_i! \right\}^{-1} \tag{27}$$

Now consider the binding of an n-site ligand with binding and cooperativity parameters K and ω and concentration L in the presence of a small ligand with binding constant K_s, concentration L_s, and no cooperative interactions either with itself or with the larger ligand. We define g and θ for the large ligand as in eqs. (3) and (6). Let θ_s be the extent of binding of the small ligand, and let $P_M(k, j, t)$ be the number of configurations containing k large ligands with j adjacencies and t small ligands. We have

$$\theta_s = \bar{t}/M \tag{28}$$

and

$$\bar{k} = \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} \sum_{t=0}^{M} k P_M(k, j, t) (KL)^k (K_s L_s)^t \omega^j \right\} \times \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} \sum_{t=0}^{M} P_M(k, j, t) (KL)^k (K_s L_s)^t \omega^j \right\}^{-1}, \tag{29}$$

$$\bar{t} = \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} \sum_{t=0}^{M} t P_M(k, j, t) (KL)^k (K_s L_s)^t \omega^j \right\} \times \left\{ \sum_{k=0}^{g} \sum_{j=0}^{k-1} \sum_{t=0}^{M} P_M(k, j, t) (KL)^k (K_s L_s)^t \omega^j \right\}^{-1}. \tag{30}$$

Since the small ligands are non-cooperative and occupy only one site, they are equivalent, from the point of view of counting independent units, to vacant sites. Thus the analysis of section 2.2.1 still holds, with the number of vacant sites being reduced by the number of small ligands bound. The analysis of the distribution of the runs of large ligands is totally unaffected by the presence of the small ligands (their effect is felt only in the $(K_s L_s)^t$ terms in eqs. (29)–(30)), so we obtain

$$P_M(k, j, t) = \frac{(M - nk + 1)!(k - 1)!}{t!(M - nk - k + j + 1 - t)!(k - j)!j!(k - j - 1)!} \tag{31}$$

From eqs. (28)–(31) a complete analysis of such a competition experiment may be carried out.

3. Irreversible binding

If we let KL become very large in eq. (5), we obtain the limiting value

$$\lim_{KL\to\infty} \bar{k} = g \tag{32}$$

which, on insertion in eq. (6) and using eq. (4), yields

$$\lim_{KL\to\infty} \theta = ng/M = 1 - r/M. \tag{33}$$

Thus, it would appear that for irreversible binding, $K \to \infty$ and $\theta(M)$ should equal unity whenever M is an integral multiple of n and should approach one as the lattice grows longer. That is, the ligands should fill the lattice in the most efficient possible manner.

However, Flory [10] showed nearly 40 years ago that in the irreversible intramolecular reaction between neighboring substituents of vinyl polymers, a process equivalent to the irreversible binding of a 2-site ligand, the fraction of reacted sites, i.e., θ, should approach not 1, but $1 - e^{-2} = 0.865$. This prediction is, in fact, consistent with the experimental results [14]. We therefore wish to reassess the applicability of eq. (33) to the irreversible binding of long ligands.

On consideration, it is clear that eq. (33) cannot be correct for irreversible binding, since it implies, for example, that when $M = gn$ all the ligands bind in a single run without any gaps. For irreversible binding, except in the limit of infinite cooperativity, this situation is extremely improbable, since if two ligands ever bind with fewer than n sites between them, those sites must remain forever isolated and vacant. If the binding is reversible, even if only slightly, then the ligands can rearrange themselves to bring about the efficiency of binding which eq. (33) implies. In the limit of a very small rate constant for dissociation from the lattice, we obtain eq. (33), but the necessary rearrangement of the ligands may take a very long time. These kinetic considerations thus imply that the treatment of sections 2.1 and 2.2, which implicitly assumes this rearrangement of the ligands, must be modified to deal with irreversible binding.

3.1. Recursion relations

In order to treat irreversible binding properly, we adopt an approach similar to that of Flory [10]. Let us define S_M as the average number of vacant lattice sites at saturation when an n-site ligand binds irreversibly with cooperativity parameter ω to an M-site lattice *. It is obvious that

$$S_0 = 0, \qquad S_n = 0$$

$$S_M = M, \qquad \text{if } 1 \leqslant M \leqslant n - 1. \tag{34}$$

Let us also define T_M as the average number of vacant lattice sites when our ligand binds irreversibly to an M-site lattice which is bordered by a bound ligand at *one* of its ends. Thus T_M should be greater than, less than or equal to S_M according to whether the ligand binds cooperatively, anti-cooperatively or noncooperatively. Finally we define U_M to be the corresponding quantity for a lattice bordered at *both* ends by bound ligands. Schematic diagrams of the lattices involved are shown in fig. 3.

Now consider binding the first ligand to our naked M site lattice. Suppose the ligand binds starting at site l. This binding divides the lattice into two permanently separated sublattices of length $l - 1$ and $M - n - (l - 1)$. Each of these lattices is singly bordered (see fig. 3). If we take the number of lattice sites remaining after each of the two sublattices has reached saturation, add these together, and then average this result over all possible points of first attack l, we will obtain the desired quantity S_N. Starting from the naked lattice, the first ligand is equally likely to bind at any of the $M - n + 1$ sites $l = 1, 2, \ldots, M - n + 1$. On letting l range from 1 to $M - n + 1$, we find that each of the sublattice lengths $l - 1$ and $(M - n - (l - 1))$ assumes values from 0 to $M - n$. Since the sublattices are singly bordered, they have by definition T_{l-1} and $T_{M-n-(l-1)}$ sites remaining unoccupied at saturation. We thus have the recursion relation

$$S_M = \frac{2}{M - n + 1} \sum_{l=0}^{M-n} T_l. \tag{35}$$

* The cooperativity parameter ω is defined here as the ratio of the forward (association) rate constants for reactions (1) and (2), i.e., we consider the dissociation rate constants to be equal and negligibly small.

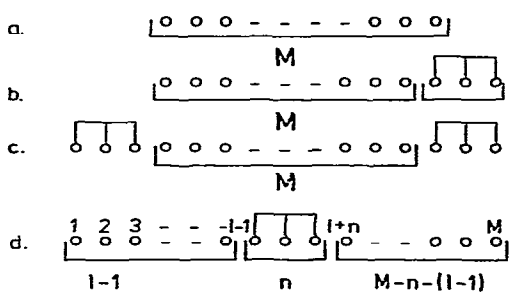

Fig. 3. The irreversible binding of an n-site ligand to an M-site lattice. (a) The naked lattice used to calculate S_M. (b) The singly bordered lattice used to calculate T_M. (c) The doubly bordered lattice used to calculate U_M. (d) The ligand binds at position l creating two singly bordered lattices of lengths $l-1$ and $M-n-(l-1)$.

Since the number of sites remaining is independent of where the ligand binds for $M \leqslant 2n-1$, we have

$$T_M = S_M \quad \text{for } 0 \leqslant M \leqslant 2n-1. \tag{36}$$

We need, however, a way to determine the T_M for $M \geqslant 2n$. We apply the above procedure now to a singly bordered naked lattice of length M. The first ligand can bind at site 1 with relative probability ω (since it will be adjacent to the bordering ligand), leaving a singly bordered lattice of length $M-n$. Alternatively it can bind, with relative probability 1 to any of the $M-n$ sites $l = 2, 3, \ldots, M-n+1$ yielding a doubly bordered lattice of length $M-n-(l-1)$. As before, we can calculate T_M as the appropriately weighted sum over the ligands remaining after the singly bordered (contribution $T_{M-n-(l-1)}$) and doubly bordered (contribution U_{l-1}) sublattices have reacted. We obtain

$$T_M = \left\{\omega T_{M-n} + \sum_{l=1}^{M-n-1} T_l + \sum_{l=1}^{M-n} U_l\right\} \Big/ (M-n+\omega). \tag{37}$$

Again it is clear that

$$U_M = S_M \quad \text{for } 0 \leqslant M \leqslant 2n-1. \tag{38}$$

All that remains now is to derive a recursion relation for U_M if $M \geqslant 2n$. Fortunately, a lattice can be no more than doubly bordered, so we find that the first ligand binding to a doubly bordered naked M-site lattice yields either a doubly bordered lattice of length $M-n$ with relative probability ω if it binds at either end ($l = 1$ or $l = M-n+1$) or two doubly bordered lattices of lengths $l-1$ and $M-n-(l-1)$ with relative probability 1 if it binds at an interior position $l = 2, 3, \ldots, M-n$. We thus obtain our final relation

$$U_M = \left\{2\left(\omega U_{M-n} + \sum_{l=1}^{M-n-1} U_l\right)\right\} \Big/ (M-n-1+2\omega). \tag{39}$$

Eqs. (34–39) constitute a sufficient set for calculating all the S_M. The extent of binding θ is then given simply by

$$\theta = 1 - S_M/M. \tag{40}$$

In the case of non-cooperative binding, $\omega = 1$, we have $U_M = T_M = S_M$ for all M and the set of equations reduces to eqs. (34–35) with T_M replaced by S_M. For noncooperative binding with $n = 2$, the recursion relation may be transformed into a difference equation [10] which is soluble and which yields

$$\lim_{M\to\infty} S_M = M/e^2. \tag{41}$$

which is the result referred to at the beginning of this section.

3.2. Comparison with reversible binding

On carrying out the calculation of θ as a function of n, ω and M for irreversible binding, we note several significant differences from our results for reversible binding. For a given degree of cooperativity and fixed lattice length, we find that θ decreases with ligand size n. In contrast, table 1 shows that longer ligands bind *more* efficiently than shorter ones if all other parameters are held fixed in the reversible binding process. We attribute this difference to the fact that for irreversible binding longer ligands are less able to "get out of each other's way" than are smaller ones, so that they tend to bind inefficiently with relatively large gaps, which would tend to disappear if dissociation and rebinding were permitted.

Surprisingly, only for rather strongly anticooperative ligands does the extent of irreversible binding

exceed that of reversible binding with only moderately high concentrations. For example, with $\omega = 1$, θ for irreversible binding is found to be roughly equal to that for reversible binding with $KL \approx 10$ over a wide range of M and n values. With $\omega = 5$, irreversible and reversible binding give comparable θ values when $KL \approx 1$. As cooperativity increases, it becomes increasingly more important for the ligands to be able to enter and leave the lattice in order that the "efficient" configurations which contain long runs get the high statistical weight they are "entitled to".

4. Conclusions

Our study of large ligand binding to finite lattices has revealed several pieces of potentially useful and possibly surprising information. The approximation of an infinite lattice should be an accurate one in most cases even for quite short lattices. However, in cases where strong cooperativity is likely, a concentration regime may be encountered in which this approximation breaks down. Use of analyses based on a finite lattice combined with experiments performed on lattices of different lengths may provide more information about the binding parameters than the conventional analysis based on the infinite lattice approximation. Irreversible binding has been seen, on essentially kinetic grounds, to be fundamentally different, and far less effective in covering the lattice, than reversible binding. Depending upon future experimental developments, it may prove useful to consider more sophisticated models involving end effects, inhomogeneous lattices, cooperativity of non-nearest neighbors, etc. We feel strongly that the combinatorial approach employed here can be of considerable utility in attacking a wide variety of such problems.

Acknowledgements

We thank the Dreyfus Foundation for a Teacher-Scholar Award and the National Science Foundation for a Science Faculty Professional Development Grant. We are grateful to Prof. Manfred Eigen for his kind hospitality during our stay in Göttingen and to Dr. Dietmar Pörschke and Dr. Thomas Jovin for enlightening discussions of these and related problems.

References

[1] J.D. McGhee and P.H. von Hippel, J. Mol. Biol. 86 (1974) 469.
[2] G. Schwarz, Biophys. Chem. 6 (1977) 65.
[3] D.E. Draper and P.H. von Hippel, ICN-UCLA Sym. Mol. Cell. Biol. 5 (1976) 421.
[4] R.C. Kelly, D.E. Jensen and P.H. von Hippel, J. Biol. Chem. 251 (1976) 7240.
[5] H. Gould and H.R. Matthews, Separation methods for nucleic acids and oligonucleotides (North Holland, Amsterdam, 1976).
[6] S.A. Latt and H.A. Sober, Biochemistry 6 (1967) 3293.
[7] A.S. Zasedatelev, G.V. Gurskii and M.V. Volkenshtein, Mol. Biol. 5 (1971) 245.
[8] G. Schwarz, Ber. Bunsenges. Physik. Chem. 75 (1971) 40.
[9] S. Lifson, J. Chem. Phys. 40 (1964) 3705.
[10] P.J. Flory, J. Amer. Chem. Soc. 61 (1939) 1518.
[11] R.F. Boyer, J. Phys. Colloid Chem. 51 (1947) 80.
[12] H. Delius, N.J. Mantell and B. Alberts, J. Mol. Biol. 67 (1972) 341.
[13] T.M. Jovin and G. Striker, in: Molecular biology, biochemistry and biophysics, Vol. 24, eds. I. Pecht and R. Rigler (Springer-Verlag, Berlin, 1977), pp. 245-281.
[14] C.S. Marvel and C.L. Levesque, J. Amer. Chem. Soc. 60 (1938) 280.