1970

A Statistical-Mechanical Model for the Binding of Flexible Ligands to Proteins

Nora Dawn Laiken


PRO BONO HUMANI GENERIS
The Rockefeller University
1901

A STATISTICAL-MECHANICAL MODEL FOR THE BINDING OF FLEXIBLE LIGANDS TO PROTEINS

A thesis submitted to the Faculty of The Rockefeller University

in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

by

Nora Dawn Laiken, B.S.

Approved for publication.

George Némethy

Assistant professor

1 April 1970

The Rockefeller University

New York

ACKNOWLEDGEMENTS

There are several people whose assistance has been especially important to me in the course of preparing this dissertation. First and foremost, I would like to thank my research advisor, Dr. George Némethy, for his patience and encouragement as well as for the valuable guidance he has provided. In addition, I would like to thank the members of my Faculty Committee, Dr. E.G.D. Cohen, Dr. Mark Kac, Dr. Gerald Manning, and Dr. Gertrude Perlmann, for their interest in my work almost from its beginning stages. Gratitude also goes to Dr. Fred Dodge, for his generosity in making available to me the 360-91 computer at IBM-Watson laboratories, without which very few of my statistical-mechanical formulae could have been transformed into binding isotherms.

I also would like to thank the staff of Graphic Services, and in particular Miss Elsa Rivera, for their skillful preparation of the illustrations, and the Physics Department, for lending me a typewriter which could cope with even the more unconventional equations.

ABSTRACT

In this dissertation, a theoretical model is presented for the binding of flexible ligands to proteins. The model explicitly accounts for the ability of these chain-like molecules to bind in a large number of configurations (including many in which not all segments are in contact with the protein) and therefore is quite different from the theory of multiple equilibria which commonly is used to analyze such interactions. The latter assumes that the ligands bind rigidly to point binding sites, neglecting the internal degrees of freedom of the bound molecules. Comparisons of binding data calculated using the present model with those obtained experimentally indicate that this model, rather than the theory of multiple equilibria, is the appropriate theoretical model for the interactions between flexible ligands and proteins (e.g. the nonspecific binding of substituted alkanes to proteins).

Formally, the model is similar to the statistical-mechanical theories of polymer adsorption at a surface, since the protein surface is represented by a lattice, each site of which binds a single ligand segment. However, the lattice chosen has several features which make it especially suitable for the treatment of protein-ligand interactions. For example, the lattice has been folded over a closed surface by introducing a small number of irregularities in the lattice pattern, and chemically different sites (e.g. sites with ionic, polar,

and nonpolar character) in various arrangements may be included. The binding sites may cover the entire protein surface or only a small part of it. The number of configurations of the bound ligands is determined by counting the possible arrangements on the lattice, regarding the placement of each ligand on the lattice as a Markov process. For chemically homogeneous binding surfaces, an alternative form of the model is possible which does not utilize a lattice for the enumeration of ligand configurations.

The model allows a calculation of parameters which characterize the average configuration of each bound ligand as well as the equilibrium properties of protein-ligand systems, such as the relationship between the free ligand concentration and the average number of ligands bound per protein molecule(which may be compared to an experimental binding isotherm). Furthermore, the dependence of these quantities on the specific characteristics of the ligand and the protein surface may be determined. Among the most important of these results is that flexible ligands bind in configurations with a larger number of desorbed segments as the lattice sites on the protein become saturated. This variation in the average configuration of the bound ligands produces a change in the intrinsic affinity of the protein for the ligand, so that the Scatchard and double-reciprocal plots of binding data never can be expected to be linear for flexible molecules, even when all of the binding sites are chemically identical. For chemically homogeneous systems, the theory of multiple equilibria predicts linearity for such plots since it assumes

an equal intrinsic affinity of the protein for all ligands, an assumption which is not applicable to flexible molecules. The model predicts binding data quite similar to those observed experimentally. In fact, it is found that only the present model which accounts for the flexibility of the bound ligands can explain the experimental observation that the strength of binding increases with the chain length of the ligand. Furthermore, the model allows certain correlations to be made between the chemical nature of the binding region and the general features of experimental binding data. In particular, the binding data for several substituted alkanes (long-chain anions) indicate that the binding region must include a large number of nonpolar sites (which interact with the hydrocarbon segments of the ligand) as well as ionic sites (which interact with the ionic head group). Finally, the model may be used to illustrate the limited nature of the conclusions which may be drawn from class analyses of binding data according to the theory of multiple equilibria.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Complexes between proteins and small molecules, or ligands, are important in many biological systems. Perhaps the most familiar examples are the highly specific complexes formed between enzymes and their substrates. In addition, many proteins form complexes of a somewhat less specific nature with a variety of small molecules, including the alkanes and their substituted derivatives. In this dissertation, a theoretical model for the interaction of these chain-like molecules with proteins will be presented. It will be seen that the model predicts binding data which are quite similar to those observed experimentally, and more importantly, accounts for several general features of binding data which could not be explained previously or which were interpreted incorrectly.

Protein-ligand interactions commonly have been analyzed using the standard formulae governing multiple equilibria (Klotz, 1953), in which the bound ligands are grouped into classes according to the strength with which they are bound. In the simplest possible case, all of the ligands bind independently and with equal intrinsic binding constants. If a maximum of M ligands is bound, each with an intrinsic association constant k_{int}, then the average number of ligands bound per protein molecule, $\bar{N}$, bears a simple relationship to the free ligand concentra-

tion, c (Klotz, 1953)

$$\bar{N} = \frac{Mk_{int}c}{1 + k_{int}c} \qquad (I.1)$$

Furthermore, the equilibrium constants for the binding of successive ligands differ only by a statistical factor which accounts for the different ways of placing the ligands on the protein

$$k_N = \frac{M-N+1}{N} k_{int} \qquad (I.2)$$

where k_N is the equilibrium constant for the binding of the Nth ligand.

According to equation I.1, the binding isotherms, expressed as $\bar{N}$ vs. log c, will have a symmetrical, sigmoidal shape for single-class binding. The binding equation may be written in two alternative forms

$$\frac{\bar{N}}{c} = k_{int}M - k_{int}\bar{N} \qquad (I.3)$$

$$\frac{1}{\bar{N}} = \frac{1}{Mk_{int}c} + \frac{1}{M} \qquad (I.4)$$

Thus, plots of $\bar{N}/c$ vs. $\bar{N}$ and $1/\bar{N}$ vs. $1/c$ (the so-called Scatchard and double-reciprocal plots, respectively) will be linear when all ligands bind as one class. Examples of single-class binding data are illustrated in Figure 1.

In experimental studies of protein-ligand interactions, symmetrical binding isotherms (and linear Scatchard and double-reciprocal plots) rarely are obtained (e.g. Figure 32). Such behavior most commonly is explained by assuming that the bound ligands belong to several classes

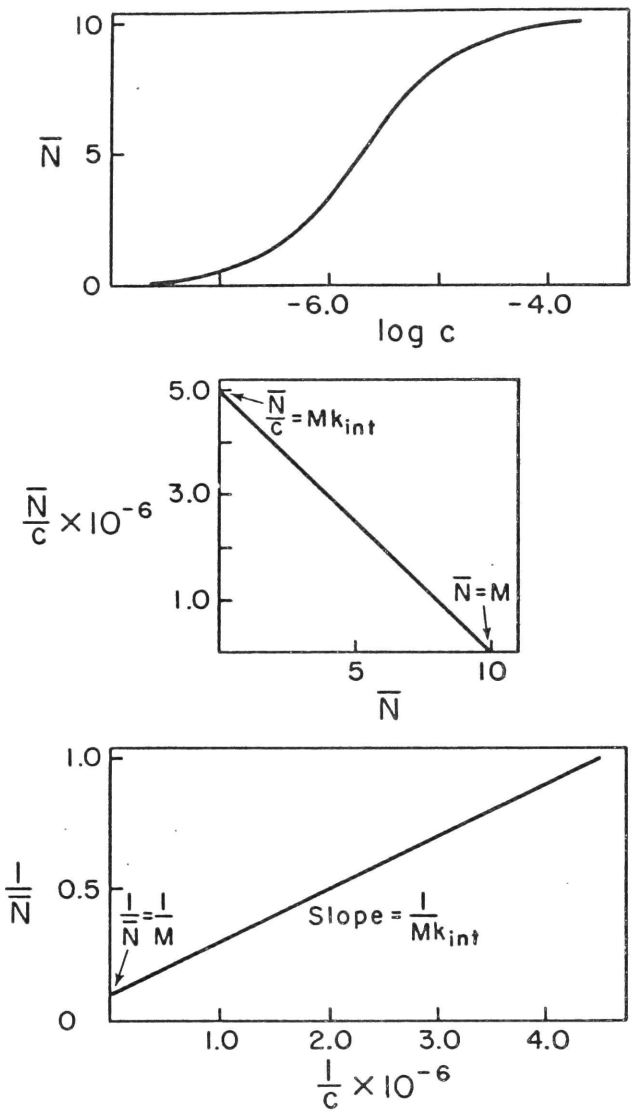


Figure 1. Graphical presentation of single-class binding data. Top: binding isotherm; center: Scatchard plot; bottom: double-reciprocal plot. All curves were generated for a hypothetical protein with ten binding sites (M=10), each with an intrinsic association constant of 5×10^5 ($k_{int} = 5 \times 10^5$). $\bar{N}$ is the average number of ligands bound per protein molecule and c is the concentration of unbound ligand. M and k_{int} may be determined from the x- and y-intercepts of the Scatchard plot, or from the y-intercept and the slope of the double-reciprocal plot.

(Scatchard et al., 1950)

$$\bar{N} = \sum_{i=1}^{n_c} \frac{M_i k_{int}^{(i)} c}{1 + k_{int}^{(i)} c} \qquad (I.5)$$

where n_c is the number of classes.

There are two important reasons, however, why the theory of multiple equilibria is not suitable to describe the interactions of flexible, chain-like molecules with proteins. First of all, the binding site in the theory of multiple equilibria is defined as a point to which a ligand binds in an all-or-none manner. Although this may accurately represent the binding of metallic ions and other rigid, monofunctional ligands, a chain-like molecule is composed of many functional units and, because of its flexibility, not all units have to be in contact with the protein simultaneously. Thus, for flexible molecules it is more realistic to define a binding site as an area on the protein which serves as a point of attachment for a single functional group, or segment, of the ligand in question. Secondly, in the theory of multiple equilibria the internal degrees of freedom of the bound ligands are ignored, so that all ligands in a given class are assumed to bind in the same configuration. Again, this is a reasonable assumption for the metallic ions and other rigid ligands which can bind in only one configuration (Figure 2a), but for flexible molecules it implies that all ligands belonging to a given class are bound in an identical, rigid configuration. Actually, such chain-like molecules can assume a large number of configurations when bound (Figure 2b), including many in which not all segments are in contact with the protein (con-

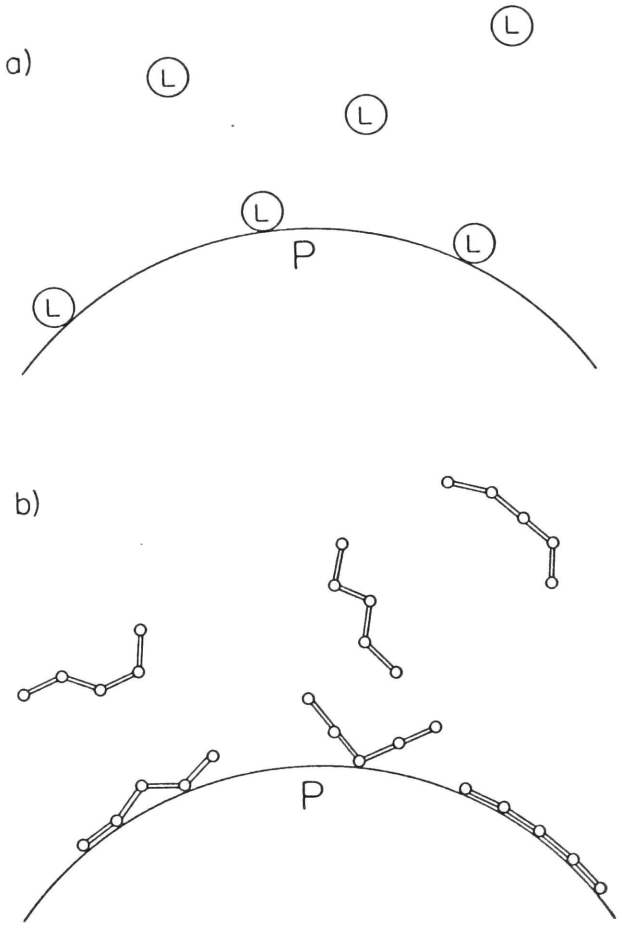


Figure 2. Schematic illustration of the binding of rigid and flexible ligands to a protein. a) Small ions and other rigid ligands (L) bind to a protein (P) with no changes in configuration. b) Flexible ligands may assume a large number of configurations both in solution and in the bound state.

figurations with desorbed loops and free ends). Therefore, a new model for multiple equilibria is necessary if the binding of flexible ligands is to be considered. The model developed in this dissertation fulfills this need, since it specifically accounts for the flexibility of the bound ligands. Although the model has certain formal similarities to the statistical-mechanical theories of flexible polymer adsorption at a surface, several features have been included which make the model especially suitable for the description of protein-ligand interactions.

CHAPTER II

DESCRIPTION OF THE MODEL

Limitations of Earlier Models

Several statistical-mechanical models of polymer adsorption at a surface have been constructed (Silberberg, 1962; Roe, 1965a, 1965b. Further references are included in these papers.). In these theories, the central problem is to enumerate all possible configurations that a flexible polymer might assume in the vicinity of a surface when at least one of its segments is adsorbed to the surface. The enumeration of configurations usually is accomplished with the aid of a lattice model. Neglecting the self-excluded volume of the polymer, there is a one-to-one correspondence between random walk paths on the lattice and polymer chain configurations.

However, the currently available models for polymer adsorption cannot be applied to the description of protein-ligand interactions since each suffers from one or more of the following limitations. First of all, infinite, planar surface lattices are utilized. The surface of a protein is neither infinite nor flat, as most proteins have a fairly compact shape. Secondly, the surface sites commonly are regarded as equivalent in their ability to adsorb polymer segments, and the polymer chain is said to be composed of identical segments. In contrast, a protein surface has many different types of binding sites (due to the chemical diversity of its

constituent amino acids), and many flexible ligands contain segments which differ chemically (e.g. the substituted alkanes). A third limitation of most lattice treatments is that they are formulated only for the limiting case of infinite or very long chains, while most ligands employed in protein binding studies are relatively short (fewer than 20 carbon atoms for the alkanes and their derivatives). Finally, most models emphasize the properties of a single polymer chain at a surface, and the binding of additional molecules, if treated at all, is assumed to occur in the absence of competition for surface sites among adjacent molecules. Experimentally, protein-ligand interactions are studied over a range of concentrations, throughout much of which several ligands are bound to each protein. Due to the finite size of the protein surface, the independence of successively bound ligands cannot be assumed automatically.

Choice of Model

The placement of a finite, planar lattice upon a closed surface such as a cube or a sphere requires the introduction of a small number of irregularities in the lattice pattern. Variations in either the coordination number of certain sites or the spacing between sites, or both, are possible. As an example of an alteration in coordination number, a square lattice may be placed upon a cube by introducing tri-coordinated sites at the eight corners (Figure 3a). The placement of a square lattice on a globe to demarcate longitude and latitude provides an example in

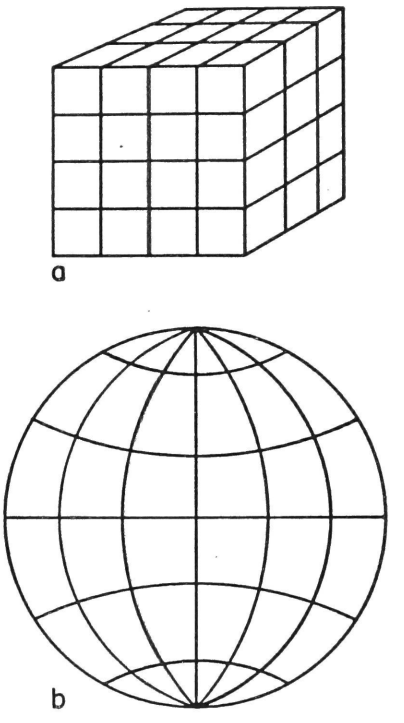


Figure 3. The placement of planar lattices on closed surfaces. a) A square lattice may be placed upon a cube by introducing tri-coordinated sites at the eight corners. b) A square lattice may be placed upon a sphere by altering the spacing between adjacent lattice sites and changing the coordination number of the two polar sites.

which irregularities have been introduced in both the spacing between adjacent lattice sites and the coordination number of the (two polar) sites (Figure 3b).

In choosing a lattice to represent a protein surface, irregularities in coordination number are to be preferred to large irregularities in the spacing between sites, since the bond lengths between adjacent segments of most flexible ligands are constant. Therefore, although a sphere is a fairly good model for globular proteins, the square lattice of Figure 3b is not suitable. A planar, triangular (hexa-coordinated) lattice can be placed upon a sphere by introducing a small number of penta-coordinated sites at regular intervals as illustrated in Figure 4(Caspar and Klug, 1962), so that the lattice is characterized by two coordination numbers: $z_1=6$ for the hexa-coordinated sites and $z_2=5$ for the penta-coordinated sites. This lattice has been selected for the present model.[1]

The lattice of Figure 4 actually represents the outward projection of the sites on an icosahedron (Marks, 1960), the regular polyhedron with 20 equilateral triangular faces and 12 vertices (Figure 5). If each face is subdivided into D^2 equilateral triangles (producing a division of each edge into D segments), a lattice is formed which contains $10(D^2-1)$ hexa-coordinated sites and 12 penta-coordinated sites, the latter located at the vertices. To treat the case in which only a restricted area of the

---

[1]The choice of a triangular lattice is reasonable in terms of the geometry of the ligands to be considered (Chapter V).

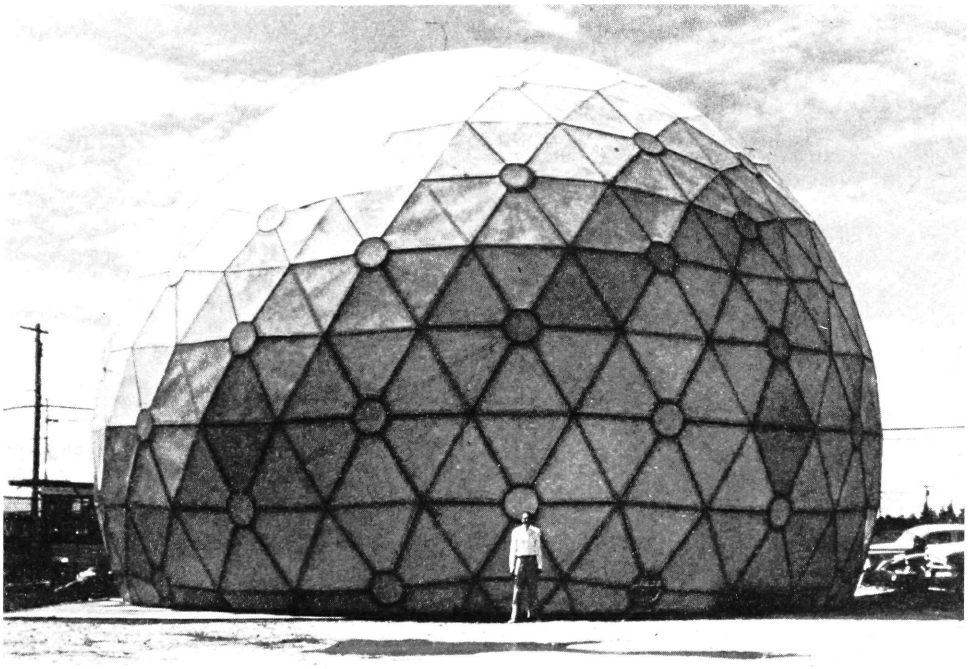


Figure 4. The placement of a planar, triangular lattice on a sphere by introducing a small number of penta-coordinated sites at regular intervals (from Caspar and Klug, 1962).

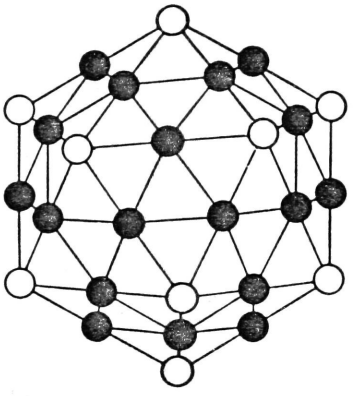


Figure 5. An icosahedron, the regular polyhedron with 20 equilateral triangular faces and 12 vertices (marked in white). Each triangular face of this icosahedron has been subdivided into four equilateral triangles ($D^2=4$), thereby creating additional sites on the lattice (marked in black). The sites at the 12 vertices are penta-coordinated (z_2 sites), while all other sites are hexa-coordinated (z_1 sites).

protein (i.e. a patch) is capable of binding ligands, a single triangular face of the icosahedron, or part of it, may be used (Figure 6). Such a single face can be placed on a surface of any shape and need not cover 1/20th of the surface area, so that the model no longer is restricted to spherical proteins when just a patch of sites is used.[1]

In addition to their role in adjusting the triangular lattice to a spherical surface, the z_2 sites serve as reference points on the surface. Within each triangular face, each z_1 site may be classified according to its positional relationship to the z_2 sites, providing a convenient labeling system which allows each type of chemical grouping found on a protein to be represented by a different class of sites. The actual classifications used will be described in Chapter III.

The enumeration of ligand configurations requires that a lattice describe the solution surrounding the protein surface as well as the surface itself, since configurations with loops and free ends are possible. The lattice in the immediate vicinity of the protein surface is constructed of concentric layers of lattice points with patterns identical to the pattern of sites on the surface. Therefore, the classification established for the surface sites may be retained in each layer of sites in the surrounding solution. When a ligand segment is on a site in a given layer above the surface, it is assumed that the subsequent segment can be placed not only on one of the z_1 (or z_2) adjacent

---

[1]The projection of more than one face also is possible if the faces are non-adjacent to each other and it is assumed that the ligands bind independently to each.

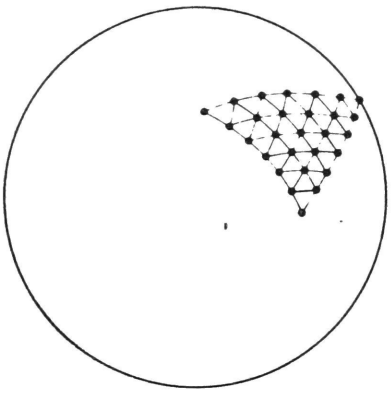


Figure 6. The projection of a single triangular face of an icosahedron onto a sphere, to give a patch of binding sites.

sites in that layer, but on any of the z_1 (or z_2) sites in the layers above and below as well. Thus, the solution sites in the vicinity of the protein surface are characterized by coordination numbers $3z_1$ and $3z_2$.[1]

A simplified, but somewhat more limited statistical-mechanical model for protein-ligand interactions which does not require a lattice for the enumeration of configurations also may be formulated. Appendix A includes a brief discussion of this model and some of the results predicted by it.

---

[1]The use of concentric layers of lattice points which retain the pattern of sites on the surface implies that the spacing between lattice points within a given layer will increase slightly as additional layers are added. Some objections might be raised, therefore, if long chains were considered, some of whose segments might extend many layers above the surface while other segments remained attached to the protein. However, as mentioned above, the ligands used in protein binding studies are short so that the approximations involved in the choice of solution lattice are not serious.

CHAPTER III

STATISTICAL-MECHANICAL TREATMENT

Basic Formulae

The theory is formulated using the grand canonical ensemble, in which the individual protein molecules, each with n_0 surface lattice sites, are the constituent systems. The solution in which the protein molecules are immersed serves as a reservoir of both heat at temperature T and ligands with chemical potential μ. Thus, each system is characterized by the thermodynamic variables μ, n_0, T.[1] The solutions considered are sufficiently dilute in protein molecules to allow the assumption of independence of systems.

The grand canonical partition function, $\mathcal{Z}(\mu, n_0, T)$, has the form

$$\mathcal{Z}(\mu, n_0, T) = \sum_{N=0}^{M} \mathcal{Q}_x(N, n_0, T)\, e^{\mu N/kT} \qquad \text{(III.1)}$$

where $\mathcal{Q}_x(N, n_0, T)$ is the canonical partition function for the system when N ligands, each with x segments, are bound. M is the maximum number of ligands which can bind to the protein and is defined by[2]

$$M = [n_0/x] \qquad \text{(III.2)}$$

---

[1]The thermodynamic variable n_0 is analogous to the volume of the system which is encountered most commonly as the extensive variable in applications which utilize the grand canonical ensemble.

[2]Throughout this and all subsequent discussions, the square bracket notation will be used to denote the integral part of the enclosed expression.

The terms $\mathcal{Q}_x(N,n_0,T)$ are of the form

$$\mathcal{Q}_x(N,n_0,T) = \frac{\prod_{i=0}^{N} q_x(i,n_0,T)}{N!} \qquad (III.3)$$

with

$$q_x(0,n_0,T) = 1 \qquad (III.4)$$

where $q_x(i,n_0,T)$ is the canonical partition function for the placement of the ith ligand on the protein surface.

To compare the predictions of the present theory with experiment, the most important thermodynamic quantity which can be calculated from $\Xi(\mu,n_0,T)$ is $\bar{N}$, the average number of ligands bound per protein molecule

$$\bar{N}(\mu,n_0,T) = \frac{\sum_{N=0}^{M} N\mathcal{Q}_x(N,n_0,T)\, e^{\mu N/kT}}{\Xi(\mu,n_0,T)}$$

$$= kT \frac{\partial \ln \Xi}{\partial \mu} \qquad (III.5)$$

The equilibrium constant for the binding of the Nth ligand may be calculated from the terms of $\Xi(\mu,n_0,T)$

$$k_N = \frac{\dfrac{\mathcal{Q}_x(N,n_0,T)}{V}}{\dfrac{\mathcal{Q}_x(N-1,n_0,T)}{V}\,\dfrac{q_{x,u}(V)}{V}}$$

$$= V \frac{\mathcal{Q}_x(N,n_0,T)}{\mathcal{Q}_x(N-1,n_0,T)\, q_{x,u}(V)} \qquad (III.6)$$

where $q_{x,u}(V)$ is the partition function for the unbound ligand in a volume V of solution.

The Binding of a Single Ligand

In this section, equations will be developed which enumerate the possible configurations of a single bound ligand, leading to an expression for $q_x(1,n_0,T)$. Except for a correction eliminating immediate reversals, the self-excluded volume of the ligand will be neglected. A general method of correcting for excluded volume and some of the results of such a calculation are described in Appendix B.

A given configuration of a bound ligand consists of a series of stretches of adsorbed segments separated by loops of desorbed segments. A state of adsorption of the ligand (Roe, 1965a), S , is defined by the set of parameters

$$S = \{d_0 , s_1 , d_1 , s_2 , d_2 \ldots\ldots s_{\ell-1} , d_{\ell-1} , s_\ell , d_0' \} \qquad (III.7)$$

where s_i and d_i are the number of segments in the ith adsorbed stretch and desorbed loop, respectively, and ℓ is the number of adsorbed stretches. The parameters d_0 and d_0' represent the number of segments in the free chain ends; if the segment at an end is in the adsorbed state, d_0 and/or d_0' is assigned the value zero. Since at least one segment of the ligand must be adsorbed, ℓ has the range of values

$$1 \leq \ell \leq [(x+1)/2] \qquad (III.8)$$

In addition, the following restriction must hold

$$x = d_0 + \sum_{\rho=1}^{\ell-1} (s_\rho + d_\rho) + s_\ell + d_0' \qquad (III.9)$$

Any state of adsorption is allowed provided that equations III.8 and III.9 are satisfied. It should be noted that many possible configura-

tions are included in a single state of adsorption. Not only could a ligand with a given S be placed on different parts of the lattice, but different arrangements of its segments are possible once the first segment has been placed.

$q_x(1,n_0,T)$ is calculated by assigning a statistical weight, $t(S)$, to each state of adsorption

$$q_x(1,n_0,T) = \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} t(S) \qquad (III.10)$$

where the summation over $\{S\}$ refers to all possible states of adsorption for a given ℓ consistent with equation III.9.

The Totally Adsorbed Ligand In order to write general expressions for the weights $t(S)$, a method for enumerating the configurations of a bound ligand must be developed. The counting procedure used may be illustrated most clearly through a discussion of the totally adsorbed ligand, for which $\ell=1$ and

$$S = \{0,x,0\} \qquad (III.11)$$

To obtain an expression for $t(0,x,0)$, the random walk of the ligand on the lattice is regarded as a Markov process, each step of which corresponds to the placement of a ligand segment on a lattice site. The states of the Markov process are defined in terms of the classification of lattice sites presented below. Thus, the ith state of the Markov process is defined as the placement of a ligand segment on a class i site. The

relevant stochastic matrix is[1]

$$\underline{W} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \cdots\cdots\cdots & w_{1C} \\ w_{21} & w_{22} & w_{23} \cdots\cdots\cdots & w_{2C} \\ & \cdots\cdots\cdots\cdots & & \\ & \cdots\cdots\cdots\cdots & & \\ & \cdots\cdots\cdots\cdots & & \\ w_{C1} & w_{C2} & w_{C3} \cdots\cdots\cdots & w_{CC} \end{pmatrix} \qquad \text{(III.12)}$$

whose elements w_{ij} are transition coordination numbers, representing the number of ways in which a ligand segment can be placed on a class j site given that the preceding segment is on a class i site. If a given lattice has C classes of sites, the transition coordination numbers must satisfy the relations[2]

$$\begin{aligned} &\sum_{j=1}^{C} w_{ij} = z_1 = 6 \qquad \text{for } i=2,3,\ldots.C \\ &\sum_{j=1}^{C} w_{ij} = z_2 = 5 \qquad \text{for } i=1 \end{aligned} \qquad \text{(III.13)}$$

if the z_2 sites are said to form class 1.

The number of ways of placing the first ligand segment on the lattice is given by a vector of initial configuration numbers

$$\underline{a}^{(1)} = \left(a_1^{(1)} \quad a_2^{(1)} \quad a_3^{(1)} \quad \cdots\cdots\cdots \quad a_C^{(1)}\right) \qquad \text{(III.14)}$$

---

[1]Throughout this and all subsequent discussions, underlined upper case letters will be used to represent matrices, while underlined lower case letters will represent vectors. All vector products represent scalar products.

[2]When binding to a patch of sites is considered, somewhat different relations will be necessary for values of i representing sites on the edge of the patch.

whose elements $a_i^{(1)}$ represent the number of class i sites available to the first segment of the ligand. In the vector

$$\underline{a}^{(n)} = \underline{a}^{(1)} \underline{W}^{n-1} \qquad \text{(III.15)}$$

each element $a_i^{(n)}$ represents the number of ligand configurations which terminate with the nth segment on a class i site. Thus, the number of totally adsorbed configurations of a chain of x segments, $\nu(x)$, is

$$\nu(x) = \underline{a}^{(x)} \underline{e} = \underline{a}^{(1)} \underline{W}^{x-1} \underline{e} \qquad \text{(III.16)}$$

where $\underline{e}$ is a column vector of dimension C, all of whose elements are unity.

To determine the elements of $\underline{a}^{(1)}$ and $\underline{W}$, the classification of sites must be specified. Superficially, it might appear that a satisfactory classification could be obtained simply by grouping the sites according to their distance in lattice steps, r , from the nearest z_2 site. Although such a classification is perhaps the simplest one possible, rigorously it cannot provide the basis for defining the states of a Markov process. For example, of the sites at r=2 in Figure 7, the site labeled g has three neighbors at r=3, while the sites labeled h have only two neighbors at r=3. Thus, if all sites at r=2 were treated as one class, the value of w_{23} would have to be some sort of average of the values of three (for sites such as g) and two (for sites such as h). This averaging, or lumping process, is not allowed because it would

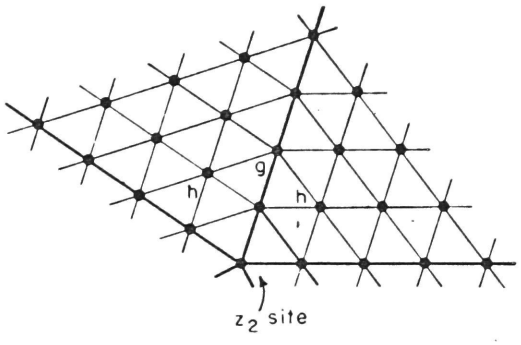


Figure 7. Detail of surface lattice in the vicinity of a z_2 site. The lattice sites cannot be classified according to their distance in lattice steps, r , from the nearest z_2 site. For example, the sites labeled g and h are located at r=2, but the g site has three neighbors at r=3, while the h sites have only two neighbors at r=3.

destroy the Markovian character of the ligand random walk.[1] Therefore, the more refined classifications illustrated in Figure 8 must be used.[2] Actually, the larger number of classes for a given lattice size is an advantage, increasing the number of ways in which different types of sites may be arranged. Furthermore, the classifications of Figure 8 also are applicable to the single face analogues of the lattices, although some modifications in the elements of $\underline{a}^{(1)}$ and $\underline{W}$ are necessary (Figure 9).

The classifications chosen possess a high degree of regularity, since each face not only has a three-fold axis of symmetry about its center but appears 20 times in covering the surface as well. The latter source of regularity is eliminated when only a single patch of sites is used. Furthermore, it will be shown that the precise arrangement of different types of sites has little effect on the results and the conclusions which can be drawn from them (Chapter V).

$t(0,x,0)$ may be calculated using the above enumeration procedure if each configuration is weighted with Boltzmann factors for the site-segment contacts formed. Thus, the elements $a_j^{(1)}$ and w_{ij} must be multiplied by

$$f_j^{(n)} = e^{-\epsilon_j^{(n)}/kT} \qquad (III.17)$$

---

[1] Under certain conditions, lumping of states in Markov processes is allowed (Kemeny and Snell; 1960).

[2] Calculations performed using the lumped states, however, show that the resulting errors in $\nu(x)$ appear only in the fourth significant figure for ligands as long as 30 segments and are considerably smaller for shorter ligands.

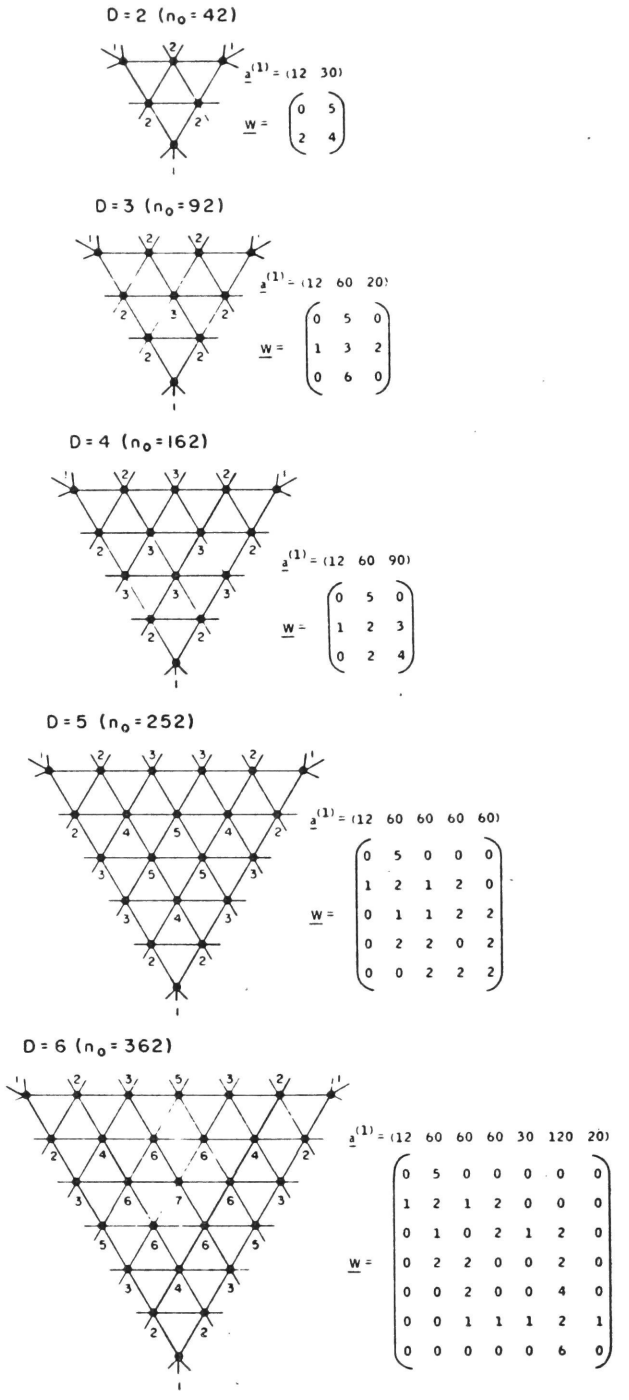
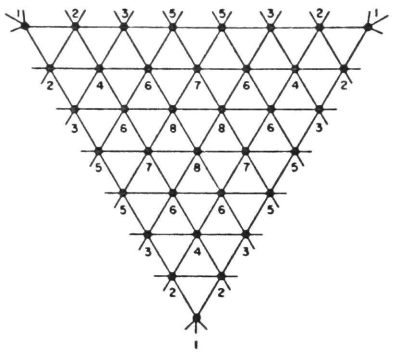


Figure 8. Classification of lattice sites. The initial configuration vector, $\underline{a}^{(1)}$, and stochastic matrix, $\underline{W}$, for each lattice also are presented. D is the number of lattice steps between adjacent z_2 sites and n_0 is the total number of sites on the lattice.

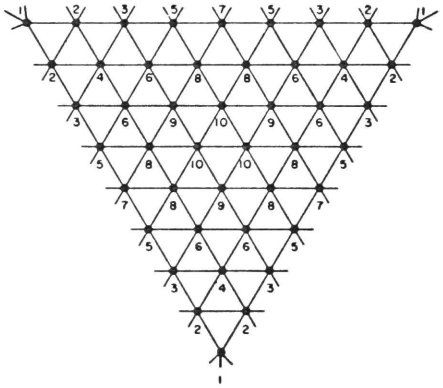
D = 7 (n_o = 492)



$$\underline{a}^{(1)} = (12\ \ 60\ \ 60\ \ 60\ \ 60\ \ 120\ \ 60\ \ 60)$$

$$\underline{W} = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 2 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 & 2 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 \end{pmatrix}$$

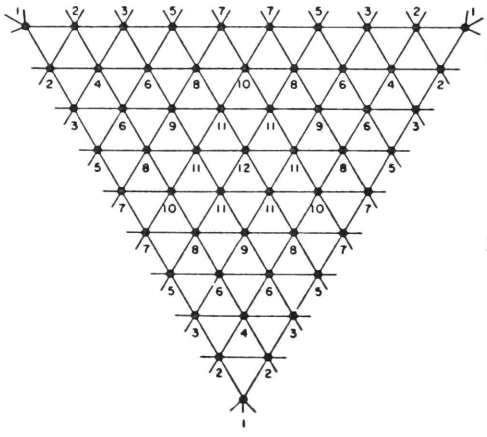
D = 8 (n_o = 642)



$$\underline{a}^{(1)} = (12\ \ 60\ \ 60\ \ 60\ \ 60\ \ 120\ \ 30\ \ 120\ \ 60\ \ 60)$$

$$\underline{W} = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 \end{pmatrix}$$

D = 9 (n_o = 812)



$$\underline{a}^{(1)} = (12\ \ 60\ \ 60\ \ 60\ \ 60\ \ 120\ \ 60\ \ 120\ \ 60\ \ 60\ \ 120\ \ 20)$$

$$\underline{W} = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 \end{pmatrix}$$

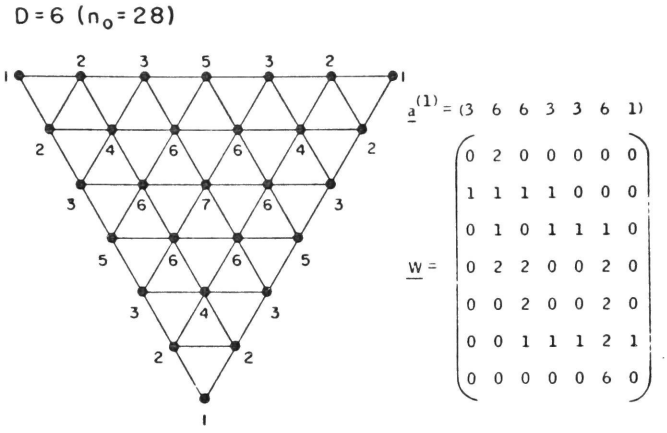


Figure 9. Single face of the lattice D=6. Although the classification of sites is the same as that for the corresponding lattice which covers a complete surface (Figure 8), the initial configuration vector and stochastic matrix must be modified. A somewhat less angular patch may be obtained if the class 1 (corner) sites are omitted to give a patch with n_0=25, which requires the deletion of the first element of $\underline{a}^{(1)}$ and the first row and column of $\underline{W}$.

where $\epsilon_j^{(n)}$ is the energy of interaction between the nth segment of the ligand and a class j site.[1] For the case in which all segments of the ligand are identical and only one type of lattice site is included, the same Boltzmann factor is used for all elements

$$f = e^{-\epsilon/kT} \qquad \text{(III.18)}$$

Furthermore, in the absence of a detailed correction for self-excluded volume (Appendix B), all elements w_{ij} in matrices describing segments following the second must be multiplied by a correction factor to exclude immediate reversals of the ligand

$$\frac{z_i - 1}{z_i} \qquad \text{(III.19)}$$

where z_i represents the coordination number of a class i site.[2]

For notational clarity, $\underline{a}^{(1)}$ and $\underline{W}$ will refer to the initial configuration vector and stochastic matrix, respectively, when only an enumeration of configurations is desired (e.g. equation III.16). When Boltzmann factors and/or immediate reversal corrections are used, as in the following calculation of t(0,x,0), $\underline{a}_s^{(1)}$ and $\underline{W}_s$ will be used. Thus,

$$t(0,x,0) = \underline{a}_s^{(1)} \underline{W}_s^{x-1} \underline{e} \qquad \text{(III.20)}$$

General Equations for t(S) Based on the principles discussed above, it is possible to write expressions for the weights t(S) for states of adsorption containing desorbed loops and free ends. The total number of config-

---

[1]The zero of energy corresponds to an infinite separation of the segment and site in question, surrounded by the appropriate solvent.

[2]Any rigorous correction would depend on j as well as i. Equation III.19 represents an average over all j.

urations of a loop or free end may be equated to the product of its vertical and horizontal configurations if it is assumed that the vertical and horizontal motion of the desorbed segments can be treated independently. Such an assumption may be made for the present model because the lattice points in each layer above the surface are labeled in a manner identical to those on the surface, allowing the classification of sites presented above to be retained as the basis for defining the states of the Markov process.[1]

If H(d) and $H_0(d_0)$ represent the number of vertical configurations of a loop of d segments and a free end of d_0 segments, respectively, then

$$t(S) = \underline{a}_d^{(1)} \underline{W}_d^{d_0-1} H_0(d_0) \prod_{\rho=1}^{\ell-1} \left(\underline{W}_s^{s_\rho} \underline{W}_d^{d_\rho} H(d_\rho) \right) \underline{W}_s^{s_\ell} \underline{W}_d^{d_0'} H_0(d_0') \underline{e} \qquad \text{(III.21)}$$

for the chain with free ends and

$$t(S) = \underline{a}_s^{(1)} \underline{W}_s^{s_1-1} \prod_{\rho=1}^{\ell-1} \left(\underline{W}_s^{s_\rho} \underline{W}_d^{d_\rho} H(d_\rho) \right) \underline{W}_s^{s_\ell} \underline{e} \qquad \text{(III.22)}$$

when $d_0=d_0'=0$. $\underline{a}_s^{(1)}$ and $\underline{W}_s$ are the vector and matrix described above for adsorbed segments, while $\underline{a}_d^{(1)}$ and $\underline{W}_d$ are the vector and matrix for desorbed segments. To form the latter from the elements of $\underline{a}^{(1)}$ and $\underline{W}$, the only modifications required are immediate reversal correction factors for the elements of matrices representing segments following the second of a desorbed loop or free end

$$\frac{3z_i - 1}{3z_i} \qquad \text{(III.23)}$$

---

[1] On more typical lattices, the assumption of independence is a good one only for small extensions of a long chain (Flory, 1953).

Expressions for H and H_0 The parameters H(d) and $H_0(d_0)$ may be calculated in two ways. The first method is of interest primarily because it treats the vertical motion of the ligand as a Markov process and thus is more similar to the methods presented above for the horizontal motion of the ligand. The second method, however, gives closed expressions for H(d) and $H_0(d_0)$. For a loop of d segments, the first and dth segments must lie in the layer h=1, while the intermediate d-2 segments may be placed in any vertical layer $h \geq 1$. One possible vertical configuration of a loop of ten segments is illustrated in Figure 10. Only the most proximal segment of a free end is constrained to the layer h=1.

Markov Method The vertical motion of the ligand may be treated as a Markov process in which the ith state is defined as the placement of a ligand segment in the vertical layer h=i. Thus, the elements y_{ij} of the stochastic matrix Y represent the number of ways of placing a segment in the layer j given that the preceding segment is in layer i. Since steps to adjacent layers or within the same layer are the only ones possible, Y is a tri-diagonal matrix of the form

$$\underline{Y} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 1 & 1 & 0 & \cdots \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \end{pmatrix} \qquad \text{(III.24)}$$

The first segment of a loop or free end always is in the layer h=1.

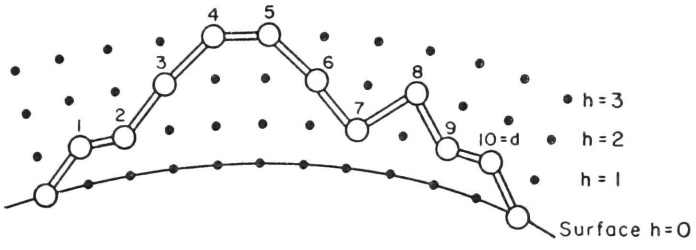


Figure 10. A possible vertical configuration of a loop of 10 (=d) segments. The first and dth segments are constrained to the first layer of lattice points above the surface (h=1), while the intermediate segments may be placed in any layer above the surface.

Hence, the initial configuration vector $\underline{b}^{(1)}$ is

$$\underline{b}^{(1)} = \left(1 \quad 0 \quad 0 \quad 0 \quad \ldots\ldots\ldots\ldots\ldots\right) \qquad \text{(III.25)}$$

Theoretically, $\underline{b}^{(1)}$ and $\underline{Y}$ have infinite dimensions, although for a loop or free end of d segments a d-element vector and a matrix of order d are sufficient since h cannot exceed d. The number of vertical configurations of a loop or free end which result in the placement of the dth segment on a site in layer i is given by the ith element of the vector

$$\underline{b}^{(d)} = \underline{b}^{(1)}\underline{Y}^{d-1} \qquad \text{(III.26)}$$

Since the last segment of a loop must lie in the layer h=1, the total number of vertical configurations of a loop of d segments, H(d), may be equated to $b_1^{(d)}$

$$H(d) = b_1^{(d)} = \underline{b}^{(d)}\underline{e}' = \underline{b}^{(1)}\underline{Y}^{d-1}\underline{e}' \qquad \text{(III.27)}$$

where $\underline{e}'$ is a column vector whose first element is unity and all other elements are zero. On the other hand, the final segment of a free end may lie in any vertical layer above the surface. Hence, the total number of vertical configurations of a free end of d_0 segments, $H_0(d_0)$, may be equated to the sum of all elements of $\underline{b}^{(d_0)}$

$$H_0(d_0) = \underline{b}^{(d_0)}\underline{e} = \underline{b}^{(1)}\underline{Y}^{d_0-1}\underline{e} \qquad \text{(III.28)}$$

Combinatorial Method Closed expressions for H(d) and $H_0(d_0)$ may be obtained by modifying some combinatorial theorems for one-dimensional random walks(Feller, 1957) in which the outcome of each step is in one of two possible directions (from level h to level h+1 or h-1). In the present model, three possible outcomes exist for each step, as

discussed above. However, any vertical configuration of a loop can be visualized as a one-dimensional random walk interrupted by steps which retain a constant value of h. Thus, for the loop illustrated in Figure 10, steps 2 and 3 (the placement of segments 3 and 4) and 5 through 8 (the placement of segments 6 through 9) perform a one-dimensional walk, while steps 1, 4, and 9 remain at positions of constant h. For a loop of d segments (d-1 steps), any number of steps 2j may be involved in the one-dimensional walk, subject to the restriction

$$0 \leq j \leq [(d-1)/2] \tag{III.29}$$

which insures that segment d of the loop can return to the layer h=1. In addition, all segments of the loop must remain in layers above the surface. The total number of configurations for such a one-dimensional walk is (Feller, 1957)

$$\frac{1}{j+1}\,\frac{(2j)!}{j!\,j!} \tag{III.30}$$

Because the remaining (d-1-2j) steps at constant h can be placed anywhere along the loop, the total number of vertical configurations of the entire loop for a given j is

$$\frac{(d-1)!}{(2j)!\,(d-1-2j)!}\,\frac{1}{j+1}\,\frac{(2j)!}{j!\,j!} = \frac{1}{j+1}\,\frac{(d-1)!}{j!\,j!\,(d-1-2j)!} \tag{III.31}$$

Hence

$$H(d) = \sum_{j=0}^{[\frac{d-1}{2}]} \frac{1}{j+1}\,\frac{(d-1)!}{j!\,j!\,(d-1-2j)!} \tag{III.32}$$

Similarly, a free end of d_0 segments (d_0-1 steps) may be visualized as an interrupted one-dimensional walk. Any number of steps m between zero and d_0-1 may be involved in the one-dimensional walk as long as all

segments remain above the surface. The number of configurations for such a one-dimensional walk of m steps is (Feller, 1957)

$$\frac{m!}{[\frac{m}{2}]!\,[\frac{m+1}{2}]!} \tag{III.33}$$

These m steps may be distributed in any manner among the remaining (d_0-1-m) constant h steps of the free end. Hence, for a fixed m, the total number of vertical configurations is

$$\frac{(d_0-1)!}{(d_0-1-m)!\,m!}\;\frac{m!}{[\frac{m}{2}]!\,[\frac{m+1}{2}]!} = \frac{(d_0-1)!}{(d_0-1-m)!\,[\frac{m}{2}]!\,[\frac{m+1}{2}]!} \tag{III.34}$$

Therefore

$$H_0(d_0) = \sum_{m=0}^{d_0-1} \frac{(d_0-1)!}{[\frac{m}{2}]!\,[\frac{m+1}{2}]!\,(d_0-1-m)!} \tag{III.35}$$

Evaluation of $q_x(1,n_0,T)$ Equations III.21 and III.22 may be combined into a single, more general one with the aid of the Kronecker delta

$$t(S) = \left(\underline{a}_s^{(1)}\underline{W}_s^{-1}\delta_{d_0 0} + (1-\delta_{d_0 0})\,\underline{a}_d^{(1)}\underline{W}_d^{d_0-1}H_0(d_0)\right) \times \prod_{\rho=1}^{\ell-1}\left(\underline{W}_s^{s_\rho}\underline{W}_d^{d_\rho}H(d_\rho)\right)\,\underline{W}_s^{s_\ell}\left(\delta_{d_0' 0} + (1-\delta_{d_0' 0})\,\underline{W}_d^{d_0'}H_0(d_0')\right)\,\underline{e} \tag{III.36}$$

so that equation III.10 becomes

$$q_x(1,n_0,T) = \sum_{\ell=1}^{[\frac{x+1}{2}]}\sum_{\{S\}}\underline{u}_0(d_0)\prod_{\rho=1}^{\ell-1}\left(\underline{V}(s_\rho)\,\underline{U}(d_\rho)\right)\,\underline{V}(s_\ell)\,\underline{u}_0'(d_0') \tag{III.37}$$

where

$$\underline{u}_0(d_0) = \underline{a}_s^{(1)}\underline{W}_s^{-1}\delta_{d_0 0} + \underline{a}_d^{(1)}\underline{W}_d^{d_0-1}H_0(d_0)\,(1-\delta_{d_0 0}) \tag{III.38}$$

$$\underline{V}(s_\rho) = \underline{W}_s^{s_\rho} \tag{III.39}$$

$$\underline{U}(d_\rho) = \underline{W}_d^{d_\rho} H(d_\rho) \qquad \text{(III.40)}$$

$$\underline{u}_0'(d_0') = \left(\delta_{d_0'0} + (1-\delta_{d_0'0})\, \underline{W}_d^{d_0'} H_0(d_0')\right) \underline{e} \qquad \text{(III.41)}$$

Equation III.37 cannot be written in closed form. However, an exact evaluation is possible for relatively small values of x, such as those of the ligands commonly used in protein binding studies. As a first step in this evaluation, the generating function of $q_x(1,n_0,T)$ is formed

$$\Gamma(\zeta) = \sum_{x=1}^{\infty} q_x(1,n_0,T)\,\zeta^x$$

$$= \sum_{x=1}^{\infty}\zeta^x \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} \underline{u}_0(d_0) \prod_{\rho=1}^{\ell-1}\left(\underline{V}(s_\rho)\,\underline{U}(d_\rho)\right) \underline{V}(s_\ell)\,\underline{u}_0'(d_0') \qquad \text{(III.42)}$$

Following the procedure of Lifson (1964), the first two summations are rearranged and the x factors of ζ^x distributed among the terms of $q_x(1,n_0,T)$ to give

$$\Gamma(\zeta) = \sum_{\ell=1}^{\infty} \sum_{x=2\ell-1}^{\infty} \sum_{\{S\}} \underline{u}_0(d_0)\zeta^{d_0} \prod_{\rho=1}^{\ell-1}\left(\underline{V}(s_\rho)\zeta^{s_\rho}\,\underline{U}(d_\rho)\zeta^{d_\rho}\right)$$

$$\times\ \underline{V}(s_\ell)\zeta^{s_\ell}\,\underline{u}_0'(d_0')\zeta^{d_0'} \qquad \text{(III.43)}$$

When summing at fixed ℓ over all possible values of $x \geq (2\ell-1)$, restriction III.9 on $\{S\}$ no longer holds. Thus

$$\Gamma(\zeta) = \sum_{\ell=1}^{\infty} \sum_{d_0=0}^{\infty} \underline{u}_0(d_0)\zeta^{d_0} \prod_{\rho=1}^{\ell-1}\left(\sum_{s_\rho=1}^{\infty} \underline{V}(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} \underline{U}(d_\rho)\zeta^{d_\rho}\right)$$

$$\times \sum_{s_\ell=1}^{\infty} \underline{V}(s_\ell)\zeta^{s_\ell} \sum_{d_0'=0}^{\infty} \underline{u}_0'(d_0')\zeta^{d_0'} \qquad \text{(III.44)}$$

$q_x(1,n_0,T)$ will be the coefficient of ζ^x in this power series (cf. equation III.42). Using the following convolution formula for the product of two power series

$$\sum_{i=1}^{\infty} a_i\zeta^i \sum_{i=1}^{\infty} b_i\zeta^i = \sum_{i=1}^{\infty}\left(\sum_{j=1}^{i} a_j b_{i+1-j}\right)\zeta^{i+1} \qquad \text{(III.45)}$$

it is possible to compute this coefficient for a given x from the known functions $\underline{u}_0$, $\underline{V}$, $\underline{U}$, and $\underline{u}_0'$.

Average Properties Any thermodynamic average derivable from $q_x(1,n_0,T)$ may be evaluated in a similar fashion. The averages of interest are those associated with the configuration of the bound ligand, which will be referred to as configurational parameters. For example, the average number of ligand segments adsorbed to the surface may be equated to the coefficient of ζ^x in the following power series, divided by $q_x(1,n_0,T)$

$$\sum_{\ell=1}^{\infty} \sum_{d_0=0}^{\infty} \underline{u}_0(d_0)\zeta^{d_0} \Bigg[\sum_{s_1=1}^{\infty} s_1 \underline{V}(s_1)\zeta^{s_1} \sum_{d_1=1}^{\infty} \underline{U}(d_1)\zeta^{d_1} \times \prod_{\rho=2}^{\ell-1} \left(\sum_{s_\rho=1}^{\infty} \underline{V}(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} \underline{U}(d_\rho)\zeta^{d_\rho} \right) \sum_{s_\ell=1}^{\infty} \underline{V}(s_\ell)\zeta^{s_\ell}$$
$$+ \sum_{s_1=1}^{\infty} \underline{V}(s_1)\zeta^{s_1} \sum_{d_1=1}^{\infty} \underline{U}(d_1)\zeta^{d_1} \sum_{s_2=1}^{\infty} s_2 \underline{V}(s_2)\zeta^{s_2} \sum_{d_2=1}^{\infty} \underline{U}(d_2)\zeta^{d_2} \times \prod_{\rho=3}^{\ell-1} \left(\sum_{s_\rho=1}^{\infty} \underline{V}(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} \underline{U}(d_\rho)\zeta^{d_\rho} \right) \sum_{s_\ell=1}^{\infty} \underline{V}(s_\ell)\zeta^{s_\ell} + \cdots$$
$$\cdots + \prod_{\rho=1}^{\ell-1} \left(\sum_{s_\rho=1}^{\infty} \underline{V}(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} \underline{U}(d_\rho)\zeta^{d_\rho} \right) \sum_{s_\ell=1}^{\infty} s_\ell \underline{V}(s_\ell)\zeta^{s_\ell} \Bigg] \times \sum_{d_0'=0}^{\infty} \underline{u}_0'(d_0')\zeta^{d_0'} \qquad \text{(III.46)}$$

The average number of segments adsorbed commonly is divided by x to give the average fraction of segments adsorbed, $\bar{\theta}$.

The coefficient of ζ^x in the power series

$$\sum_{\ell=1}^{\infty} \ell \sum_{d_0=0}^{\infty} \underline{u}_0(d_0)\zeta^{d_0} \prod_{\rho=1}^{\ell-1} \left(\sum_{s_\rho=1}^{\infty} \underline{V}(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} \underline{U}(d_\rho)\zeta^{d_\rho} \right) \times \sum_{s_\ell=1}^{\infty} \underline{V}(s_\ell)\zeta^{s_\ell} \sum_{d_0'=0}^{\infty} \underline{u}_0'(d_0')\zeta^{d_0'} \qquad \text{(III.47)}$$

when divided by $q_x(1,n_0,T)$, represents the average number of adsorbed

stretches, $\bar{\ell}$. The number-average lengths of the desorbed ends, $\bar{d}_0$ and $\bar{d}_{0'}$, are calculated from the appropriate coefficients in the following two power series, respectively[1]

$$\sum_{\ell=1}^{\infty} \sum_{d_0=0}^{\infty} d_0 \underline{u}_0(d_0)\zeta^{d_0} \prod_{\rho=1}^{\ell-1} \left(\sum_{s_\rho=1}^{\infty} \underline{V}(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} \underline{U}(d_\rho)\zeta^{d_\rho} \right) \times \sum_{s_\ell=1}^{\infty} \underline{V}(s_\ell)\zeta^{s_\ell} \sum_{d_{0'}=0}^{\infty} \underline{u}_{0'}(d_{0'})\zeta^{d_{0'}} \qquad \text{(III.48)}$$

$$\sum_{\ell=1}^{\infty} \sum_{d_0=0}^{\infty} \underline{u}_0(d_0)\zeta^{d_0} \prod_{\rho=1}^{\ell-1} \left(\sum_{s_\rho=1}^{\infty} \underline{V}(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} \underline{U}(d_\rho)\zeta^{d_\rho} \right) \times \sum_{s_\ell=1}^{\infty} \underline{V}(s_\ell)\zeta^{s_\ell} \sum_{d_{0'}=0}^{\infty} d_{0'}\, \underline{u}_{0'}(d_{0'})\zeta^{d_{0'}} \qquad \text{(III.49)}$$

Other configurational parameters of the ligand may be calculated directly from the above set. Thus, the number-average length of an adsorbed stretch, $\bar{s}$, and the number-average length of a desorbed loop, $\bar{d}$, are given by

$$\bar{s} = \frac{x\bar{\theta}}{\bar{\ell}} \qquad \text{(III.50)}$$

$$\bar{d} = \frac{x - x\bar{\theta} - \bar{d}_0 - \bar{d}_{0'}}{\bar{\ell} - 1} \qquad \text{(III.51)}$$

The Binding of Additional Ligands

The expressions for $q_x(2,n_0,T)$ through $q_x(M,n_0,T)$ are identical in form to equation III.10. However, since fewer surface sites are available for binding as additional ligands are placed on the surface, the elements of $\underline{a}_s^{(1)}$ and $\underline{W}_s$ must be decreased. On the lattice sites

[1]For ligands composed of identical segments, the Markov process is reversible (Kemeny and Snell, 1960), so that $d_0=d_{0'}$.

above the surface, a smaller fraction of configurations is excluded by multiple-site occupancy as a result of the many layers of lattice points available to the segments of loops and free ends. Therefore, no reductions in the elements of $\underline{a}_d^{(1)}$ and $\underline{W}_d$ are made. Thus, for N>1

$$q_x(N,n_0,T) = \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} t_{N-1}(S) \qquad \text{(III.52)}$$

with

$$t_{N-1}(S) = \left({}_{N-1}\underline{a}_s^{(1)} \, {}_{N-1}\underline{W}_s^{-1} \delta_{d_0 0} + (1-\delta_{d_0 0}) \underline{a}_d^{(1)} \underline{W}_d^{d_0-1} H_0(d_0) \right) \times \prod_{\rho=1}^{\ell-1} \left({}_{N-1}\underline{W}_s^{s_\rho} \underline{W}_d^{d_\rho} H(d_\rho) \right) {}_{N-1}\underline{W}_s^{s_\ell} \times \left(\delta_{d_0' 0} + (1-\delta_{d_0' 0}) \underline{W}_d^{d_0'} H_0(d_0') \right) \underline{e} \qquad \text{(III.53)}$$

The subscript N-1 indicates that corrections have been made for the N-1 ligands previously bound to the surface.

The correction factors for the elements of $\underline{a}_s^{(1)}$ and $\underline{W}_s$ may be calculated once the number of occupied sites of each class has been determined. If $\bar{n}_i(N-1)$ is the average number of occupied class i sites when N-1 ligands have been bound (equation III.60), then

$${}_{N-1}a_i^{(1)} = a_i^{(1)} - \bar{n}_i(N-1) \qquad \text{(III.54)}$$

For the elements of $\underline{W}_s$, the corrections must account for the fact that each ligand already bound to the surface occupies contiguous lattice sites (Flory, 1953). Therefore, if any of the z_j sites adjacent to a class j site previously has been established as vacant, the probability that the class j site is vacant is somewhat greater than the

average fraction of vacant j sites, $\bar{\xi}_j(N-1)$, which may be equated to the ratio of the average number of vacant j sites to the total number of j sites

$$\bar{\xi}_j(N-1) = \frac{a_j^{(1)} - \bar{n}_j(N-1)}{a_j^{(1)}} \qquad \text{(III.55)}$$

The desired correction factor is $\xi_j(N-1)$, the probability that a class j site is vacant when the vacancy of an adjacent site already has been established. Thus, $\xi_j(N-1)$ is the ratio of the number of vacant j sites to the total number of j sites which are compatible with the previously confirmed vacancy of a given adjacent site. If ϕ_j is the fraction of the $\bar{n}_j(N-1)$ occupied j sites which have a neighboring ligand segment on an adjacent site, then (Guggenheim, 1944)

$$\xi_j(N-1) = \frac{a_j^{(1)} - \bar{n}_j(N-1)}{a_j^{(1)} - \phi_j \bar{n}_j(N-1)} \qquad \text{(III.56)}$$

ϕ_j may be equated to the ratio of the average number of neighboring segments, per segment, to the coordination number of the site

$$\phi_j = \frac{\frac{2(x-2)+2}{x}}{z_j} = \frac{2(x-1)}{xz_j} \qquad \text{(III.57)}$$

Hence

$$\xi_j(N-1) = \frac{a_j^{(1)} - \bar{n}_j(N-1)}{a_j^{(1)} - \frac{2(x-1)}{xz_j}\bar{n}_j(N-1)} \qquad \text{(III.58)}$$

and

$$_{N-1}w_{ij} = w_{ij}\xi_j(N-1) \qquad \text{(III.59)}$$

Calculation of $\bar{n}_j(N-1)$ If, for a given j, the jth element of $_{N-2}\underline{a}_s^{(1)}$ and the elements of the jth column of $_{N-2}\underline{W}_s$ are multiplied by the dummy variables $a_j=1$ and $\omega_j=1$, respectively, then $\bar{n}_j(N-1)$ may be calculated from the relation[1]

$$\bar{n}_j(N-1) = \bar{n}_j(N-2) + \frac{\partial \ln q_x(N-1,n_0,T)}{\partial \omega_j} + \frac{\partial \ln q_x(N-1,n_0,T)}{\partial a_j}$$

$$= \bar{n}_j(N-2) + \frac{1}{q_x(N-1,n_0,T)} \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} \Big(\delta_{d_0 0}\ {}_{N-2}\underline{a}_s^{(1)} \begin{bmatrix} \underline{E}_C & \underline{O}_C \end{bmatrix} \times {}_{N-2}\hat{\underline{W}}_s^{-1} + (1-\delta_{d_0 0})\, \underline{a}_d^{(1)} \begin{bmatrix} \underline{E}_C & \underline{O}_C \end{bmatrix} \hat{\underline{W}}_d^{d_0-1} H_0(d_0) \Big)$$
$$\times \prod_{\rho=1}^{\ell-1} \Big({}_{N-2}\hat{\underline{W}}_s^{s_\rho}\ \hat{\underline{W}}_d^{d_\rho} H(d_\rho) \Big)\ {}_{N-2}\hat{\underline{W}}_s^{s_\ell}$$
$$\times \Big(\delta_{d_0' 0} + (1-\delta_{d_0' 0})\, \hat{\underline{W}}_d^{d_0'} H_0(d_0') \Big) \begin{bmatrix} \underline{O}_C \\ \underline{E}_C \end{bmatrix} \underline{e}$$

$$+ \frac{1}{q_x(N-1,n_0,T)} \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} {}_{N-2}\hat{\underline{a}}_s^{(1)}\ {}_{N-2}\underline{W}_s^{-1}$$
$$\times \prod_{\rho=1}^{\ell-1} \Big({}_{N-2}\underline{W}_s^{s_\rho}\ \underline{W}_d^{d_\rho} H(d_\rho) \Big)\ {}_{N-2}\underline{W}_s^{s_\ell}$$
$$\times \Big(\delta_{d_0' 0} + (1-\delta_{d_0' 0})\, \underline{W}_d^{d_0'} H_0(d_0') \Big)\ \underline{e} \qquad \text{(III.60)}$$

where $\underline{E}_C$ and $\underline{O}_C$ are, respectively, the identity and null matrices of order C and

$$ {}_{N-2}\hat{\underline{W}}_s = \begin{pmatrix} {}_{N-2}\underline{W}_s & {}_{N-2}\underline{W}'_s \\ \underline{O}_C & {}_{N-2}\underline{W}_s \end{pmatrix} \qquad \text{(III.61)}$$

---

[1]The use of the augmented matrix for differentiation of a matrix product is a technique adopted from Flory (Flory and Miller, 1966).

$$\hat{\underline{W}}_d = \begin{pmatrix} \underline{W}_d & \underline{O}_C \\ \underline{O}_C & \underline{W}_d \end{pmatrix} \qquad \text{(III.62)}$$

$$\hat{N-2\underline{a}_s}^{(1)} = \frac{\partial\, N-2\underline{a}_s^{(1)}}{\partial\, a_j} \qquad \text{(III.63)}$$

with

$$N-2\underline{W}'_s = \frac{\partial\, N-2\underline{W}_s}{\partial\, \omega_j} \qquad \text{(III.64)}$$

Equation III.60 may be evaluated by forming the appropriate generating function and following the procedure described above for the evaluation of $q_x(1,n_0,T)$.

The Unbound Ligand

In order to calculate the equilibrium properties of protein-ligand systems (equations III.5 and III.6), a value for $q_{x,u}(V)$, the partition function for the unbound ligand, must be specified. The latter may be equated to the number of configurations of a free ligand in solution. In the calculations described in the subsequent Chapters, an approximate enumeration was obtained by counting ligand configurations on a lattice constructed by stacking planar, triangular lattices, so that each site in the bulk solution is characterized by coordination number $3z_1$. This lattice was selected because of its relationship to the lattice on which the configurations of the desorbed loops and free ends of bound ligands were counted. Hence

$$q_{x,u}(V) = \frac{V}{v_s}\, 3z_1 (3z_1 - 1)^{x-2} \qquad \text{(III.65)}$$

where V is the volume of the solution and v_s is the volume of a lattice site. Although this choice of enumeration procedure is somewhat arbitrary, the use of a different counting method simply would require the multiplication of all equilibrium constants by a constant factor and the shift of all binding isotherms along the log c axis by a constant increment. Thus, the general conclusions of the present model are not affected by the value of $q_{x,u}(V)$.

The chemical potential of the ligand, μ, is

$$\mu = \mu_0 + kT \ln c \qquad \text{(III.66)}$$

where c is the concentration of free ligand and μ_0 is the standard chemical potential. The latter may be equated to the chemical potential of a solution in which c=1. Neglecting ligand-ligand interactions,[1] the partition function for such a solution, $\mathcal{Q}^0_{x,u}(N_u, V)$, is[2]

$$\mathcal{Q}^0_{x,u}(N_u, V) = \frac{\left(q_{x,u}(V)\right)^{N_u}}{N_u!} \qquad \text{(III.67)}$$

where N_u is the number of unbound ligands in a volume V of solution

---

[1] Ligand-ligand interactions may be neglected since experimental binding studies are performed in solutions very dilute in ligand. This is an experimental limitation due to two considerations: (i) many ligands dimerize (carboxylic acids) or form micelles (alkyl sulfates and sulfonates) as the concentration is raised; (ii) partial or complete unfolding of the protein often is induced by high ligand concentrations (this effect is particularly characteristic of the alkyl sulfates). As discussed in Chapter V, the inability to study binding at high ligand concentrations can be a serious hindrance in the analysis of binding data.

[2] Ligand-solvent interactions do not have to be included in the calculation of μ_0 due to the choice of zero for the site-segment interaction energies.

when c=1. Thus

$$\mu_0 = -kT \frac{\partial \ln Q^0_{x,u}(N_u, V)}{\partial N_u}$$

$$= -kT \ln \frac{q_{x,u}(V)}{N_u} \qquad \text{(III.68)}$$

A Model for Rigid Binding

To better evaluate the importance of the flexibility of the bound ligand in quantitative discussions of protein-ligand interactions, a model in which segmented ligands are flexible in solution but bind in a single, rigid configuration also has been constructed. This model for rigid binding is essentially a statistical-mechanical formulation of the theory of multiple equilibria (Klotz, 1953) and is entirely equivalent to the latter. However, a statistical-mechanical treatment allows a more direct comparison of the results of this commonly used theory with those of the present model.

The basic statistical-mechanical formulae are identical to those of the model for flexible binding (equations III.1 to III.6). However, the expressions for the canonical partition functions for the individual ligands (cf. equation III.37) may be simplified considerably, since the internal degrees of freedom of the bound ligands are neglected and the ligands bind with a unique orientation to non-overlapping regions of the protein surface. For the case in which the M ligands are assumed to bind as one class (equation I.1), the canonical partition function for the binding of the Nth ligand to the protein is

$$q_x(N, n_0, T) = (M-N+1)\, e^{-E/kT} \qquad \text{(III.69)}$$

where (M-N+1) is a combinatorial factor giving the number of ways of placing the Nth ligand on the protein. E is the energy of interaction when the entire ligand binds to the surface, in contrast to the flexible binding model in which the energies of individual site-segment interactions appear in the partition function (equation III.17). Since the unbound ligand is flexible, it may be described by the equations of the previous section.

If the bound ligands belong to more than one class (equation I.5), the grand canonical partition function for the system (cf. equation III.1) is written as the product of the grand canonical partition functions for the individual classes. The latter are formulated in the manner just described for single-class binding.

CHAPTER IV

RESULTS FOR HOMOGENEOUS SYSTEMS

The statistical-mechanical theory developed in Chapter III allows a calculation of two types of thermodynamic averages. First of all, parameters characterizing the configuration of the Nth bound ligand, the so-called configurational parameters, may be obtained from the canonical partition function $q_x(N,n_0,T)$ (equations III.46 to III.51). Secondly, from the grand canonical partition function $\Xi(\mu,n_0,T)$, the equilibrium properties of the system may be calculated (equations III.5 to III.6). Although quantities in the former category cannot be measured experimentally, their variation as N ranges from one to M provides a more detailed understanding of the equilibrium properties.

The properties of homogeneous systems, in which ligands with identical segments bind to surfaces with only one type of site, are presented in this Chapter. Although such systems are not representative of those encountered in actual binding studies, many features of the model can be illustrated most clearly in discussions of these simpler systems.[1] Heterogeneous systems, in which ligands with non-identical segments

[1] If some simplifying assumptions are made regarding the immediate reversal correction factors, equation III.37, when applied to homogeneous systems, can be factored into contributions due to the vertical and horizontal motion of the ligand, allowing the model to be extended to the limit of very long chains. A further discussion is presented in Appendix C.

bind to surfaces composed of many types of sites, are considered in Chapter V.

Configurational Parameters

When the ligands x=8 and x=4 bind to the lattice n_0=162 (D=4; see Figure 8), the configurational parameters vary with xN (the number of segments belonging to bound chains) as illustrated in Figures 11 to 15.[1] Data at three values of ϵ are included. As N increases, the reduction in the number of vacant surface sites results in larger contributions to the partition function from configurations with fewer segments adsorbed to the surface. Thus, the average fraction of segments adsorbed, $\bar{\theta}$, and the average length of an adsorbed stretch, $\bar{s}$, decrease as additional ligands are bound (Figures 11 and 13). Through intermediate values of xN, the smallest decrease in $\bar{\theta}$ is observed for ϵ=-4kT, since the energy of the site-segment interaction dominates the free energy change per segment upon binding, overriding the possible entropy gain from the formation of loops and free ends. The resulting reduction in the number of vacant surface sites is so large that a very sharp drop in $\bar{\theta}$ occurs as xN approaches n_0. The drop is somewhat larger for the shorter ligand, since a larger change in $\bar{\theta}$ results from the desorption of a single segment. As $|\epsilon|$ decreases, $\bar{\theta}$ drops with increasing xN at a more constant rate, since energy and entropy effects are balanced more closely. At these lower values of $|\epsilon|$, the drop is somewhat smal-

[1] It should be emphasized that the configurational parameters characterize the Nth ligand being bound and are not averages of ligands one through N.

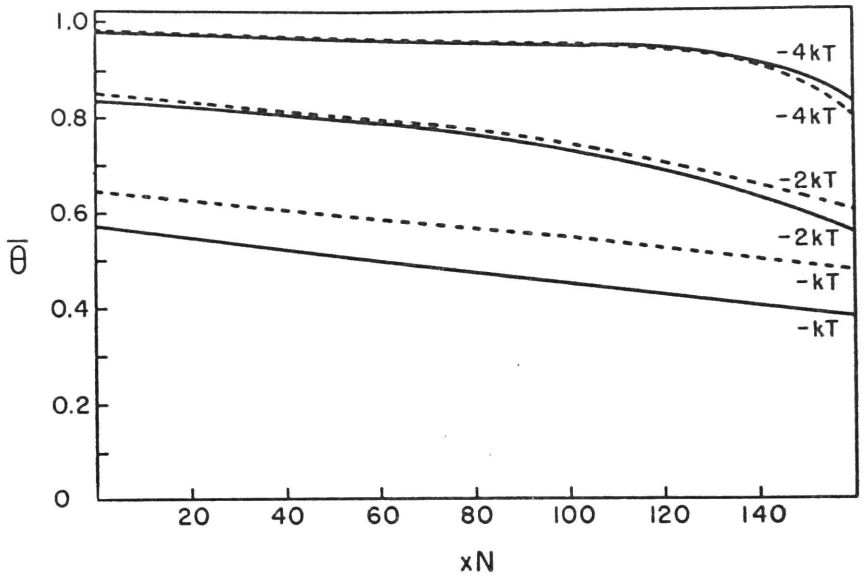


Figure 11. The variation in $\bar{\theta}$, the average fraction of ligand segments adsorbed, with xN, the number of segments belonging to bound chains, for the ligands x=8 (——) and x=4 (- - -). All curves were obtained with the lattice n_0=162. The value of ϵ is indicated by each curve.

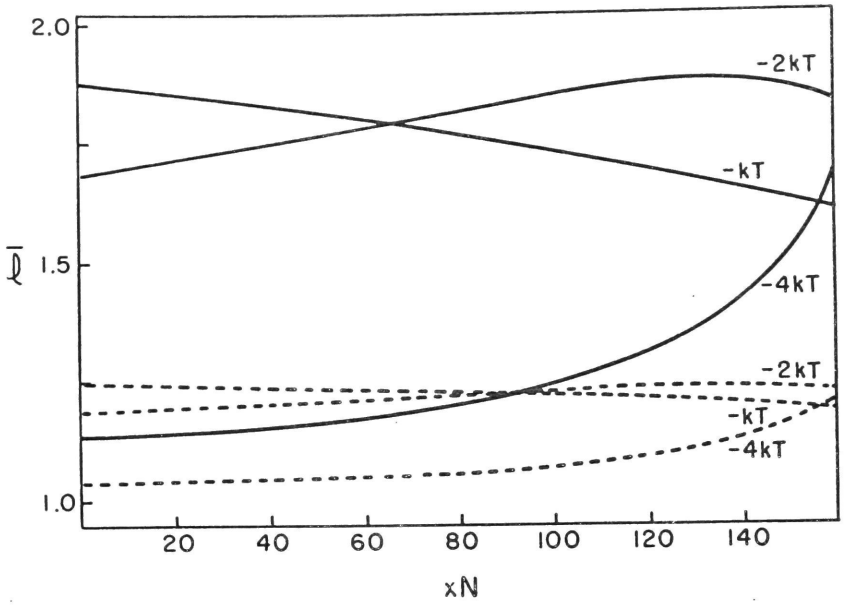


Figure 12. The variation in $\bar{\ell}$, the average number of adsorbed stretches, with xN, the number of segments belonging to bound chains, for the ligands x=8 (——) and x=4 (- - -). The curves were obtained with the lattice n_0=162, with values of ϵ as indicated.

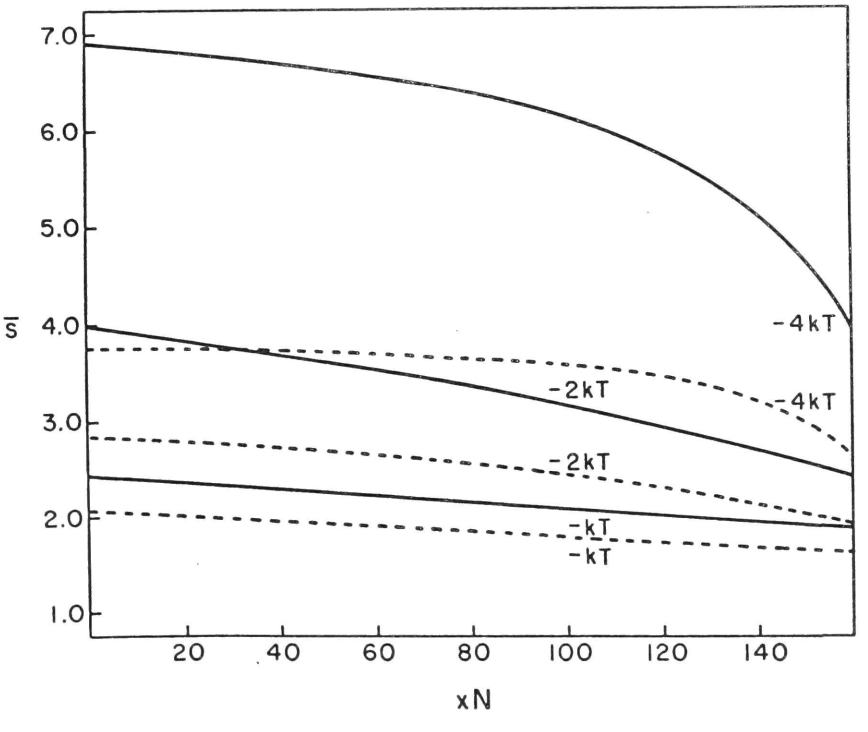


Figure 13. The variation in $\bar{s}$, the average length of an adsorbed stretch, with xN, the number of segments belonging to bound chains, for the ligands x=8 (——) and x=4 (- - -). The curves were obtained with the lattice n_0=162, with values of ϵ as indicated.

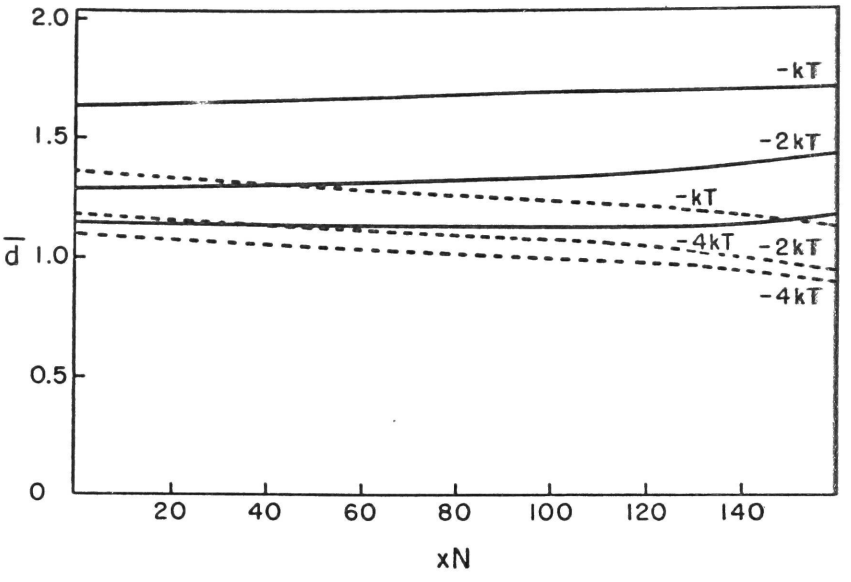


Figure 14. The variation in $\bar{d}$, the average length of a desorbed loop, with xN, the number of segments belonging to bound chains, for the ligands x=8 (——) and x=4 (- - -). The curves were obtained with the lattice n_0=162, with values of ϵ as indicated.

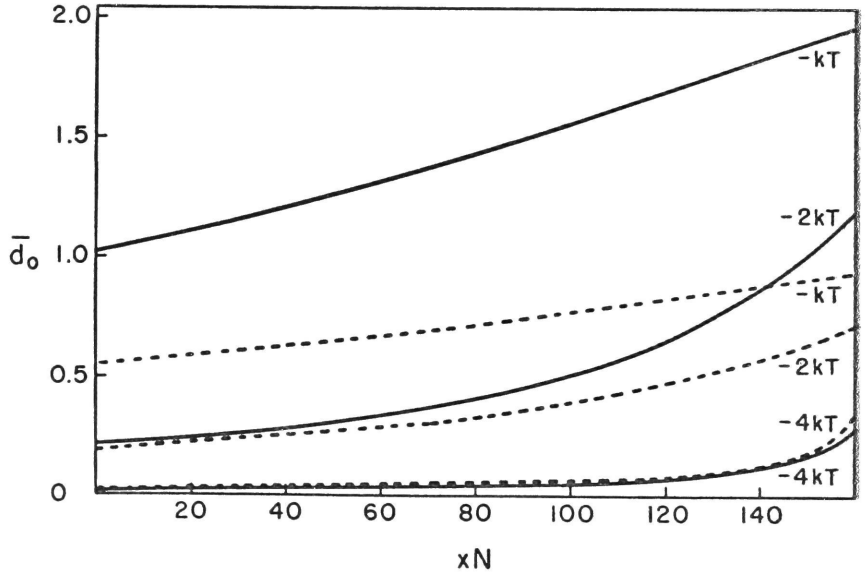


Figure 15. The variation in $\bar{d}_0$, the average length of a single desorbed end, with xN, the number of segments belonging to bound chains, for the ligands x=8 (——) and x=4 (- - -). The curves were obtained with the lattice n_0=162, with values of ϵ as indicated.

ler for the shorter ligand, since it is easier to find smaller clusters of contiguous vacant sites. $\bar{\theta}$ is larger for x=4 at the smaller values of $|\epsilon|$, since the shorter ligand can gain less entropy by forming loops (there are fewer ways of distributing a loop along a shorter ligand).

Additional desorbed segments can appear in either loops or free ends. For x=8, the average length of a single desorbed end, $\bar{d}_0$, increases markedly with xN, while the average length of a loop, $\bar{d}$, remains almost constant (Figures 14 and 15). However, at ϵ=-4kT, the total number of segments in loops does increase, since the average number of loops, $\bar{\ell}$ -1, increases (Figure 12). Increases in the number of loops rather than in $\bar{d}$ are favored because a larger number of configurations is gained by increases in ℓ (through the additional states of adsorption available to the ligand) than in d. At ϵ=-kT, configurations with $d_0 > 1$ predominate. Therefore, the additional desorbed segments appear almost exclusively in free ends, since desorbed segments in free ends exceeding one segment in length have a larger number of configurations than loops of an equivalent length. In addition, some of the ends lengthen at the expense of loops, causing $\bar{\ell}$ to decrease as xN approaches n_0. The graph of $\bar{\ell}$ vs. xN for ϵ=-2kT exhibits a maximum, since $\bar{\ell}$ increases until $\bar{d}_0$ becomes sufficiently large to favor free end formation instead of additional loops.

For x=4, where the maximum number of loops is one and the number of states of adsorption with a single loop is limited by the small value of x, the increase in desorbed segments with xN occurs almost totally in the

free ends. The increases in $\bar{\ell}$ at ϵ=-4kT and ϵ=-2kT (low xN) are less marked than those for x=8, since such rises are possible only with a decrease in $\bar{d}$ due to the short length of the chain. As ϵ=-kT and ϵ=-2kT (high xN), decreases in $\bar{\ell}$ also are observed for x=4, but are not as sharp as those for x=8 since the $\bar{d}_0$ values are smaller.

The effect of energy on the configurational parameters for the ligand x=8 is summarized in Table I, which shows these parameters for N=1.

Table I

Dependence of Configurational Parameters on ϵ [1]

ϵ	$\bar{\theta}$	$\bar{\ell}$	$\bar{s}$	$\bar{d}$	$\bar{d}_0$
-kT	0.561	1.87	2.41	1.64	1.05
-2kT	0.831	1.70	3.91	1.28	0.23
-4kT	0.975	1.14	6.86	1.15	0.02

[1]The values shown are for the ligand x=8 binding to the lattice n_0=162 when N=1.

The present model allows a treatment of binding to a patch of sites as well as a lattice which covers a complete surface. Figure 16 illustrates the configurational parameters for a patch of 25 sites (corresponding to a single face of the lattice D=6, omitting the corner sites; see Figure 9) for the ligand x=4 at three values of ϵ. The trends observed are consistent with those for the complete surface and require no further explanation.

A better patch-total surface comparison is made in Figure 17, which illustrates the configurational parameters for the ligand x=6 binding to a patch and a complete surface with an identical number of

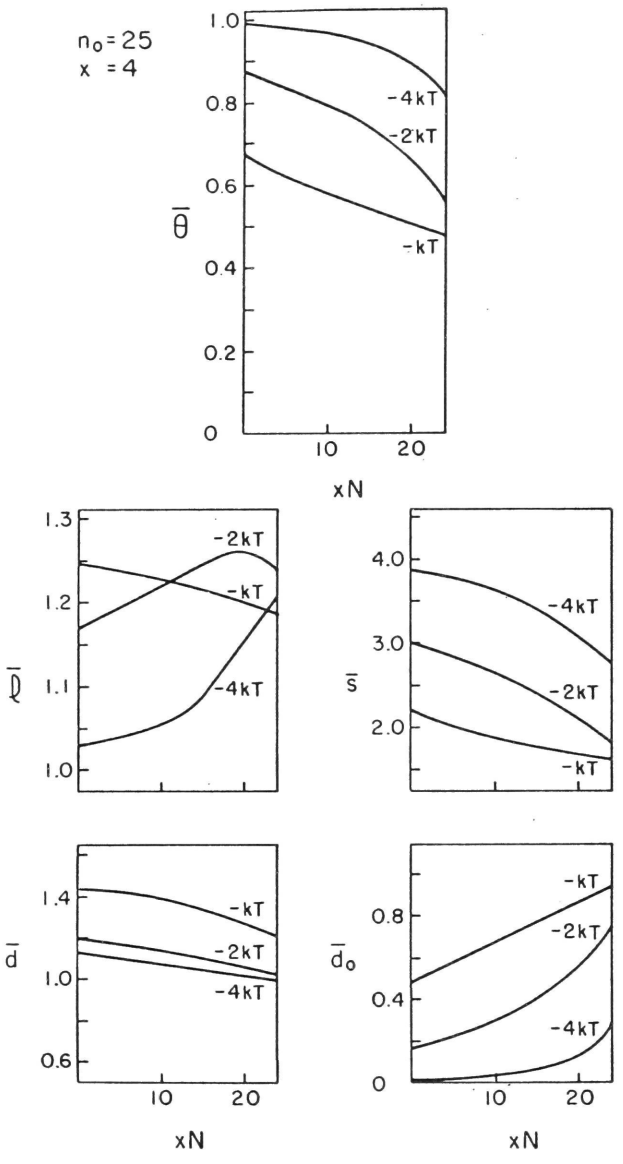


Figure 16. Configurational parameters for the ligand x=4 binding to a patch of sites (n_0=25), with values of ϵ as indicated. The curves are similar to those in Figures 11 to 15 for binding to a complete surface.

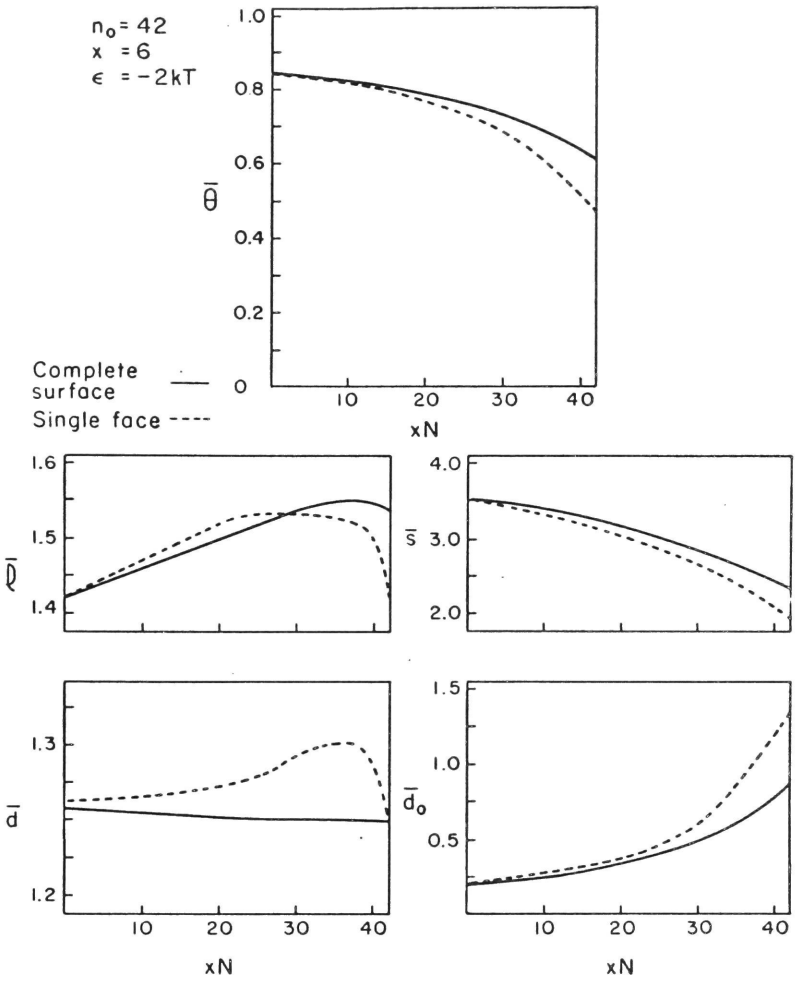


Figure 17. Configurational parameters for the ligand x=6 binding to a complete surface (——) and a patch (- - -) of identical size (n_0=42). All curves were calculated with ϵ=-2kT. The presence of edges on the patch becomes evident as xN increases.

sites (n_0=42, corresponding to a single face of the lattice D=8, omitting the corner sites, and the complete surface D=2, respectively; see Figure 8). At low xN, the average properties are nearly indistinguishable for the two cases. However, the presence of boundaries on the patch becomes evident as additional ligands are bound. Thus, larger decreases in $\bar{\theta}$, $\bar{s}$, and $\bar{\ell}$ and a sharper increase in $\bar{d}_0$ are observed for the patch. An interesting feature is the presence of a maximum in the $\bar{d}$ vs. xN curve for the patch. Because of the reduced coordination number of the boundary sites, the first ligands bind preferentially in configurations which occupy the interior sites. The ligands N=5 and N=6 bind by forming a large loop over the occupied class 9 and class 10 sites, causing an increase in $\bar{d}$, but such a loop no longer is possible for the ligand N=7 since the sites of classes 4, 6, and 8 no longer are available. Thus, $\bar{d}$ drops after N=6 and the desorbed segments appear as free ends.

Equilibrium Properties

Figure 18 illustrates binding isotherms[1] for the ligands x=8 and x=4 binding to the lattice n_0=162 at three values of ϵ (cf. Figures 11 to 15). The position of the curve shifts to lower c values as ϵ becomes more

[1]An advantage of such a plot for the representation of binding data is that the conclusions drawn from it are unaffected by the method of calculating μ_0, the standard chemical potential of the ligand, from the partition function for the unbound ligand (equation III.68), since a change in μ_0 simply would shift the positions of all isotherms along the log c axis by a constant amount. In anticipation of applications of the model, the value of v_s, the volume of a site, used to calculate μ_0 is that appropriate for hydrocarbon segments (Chapter V).

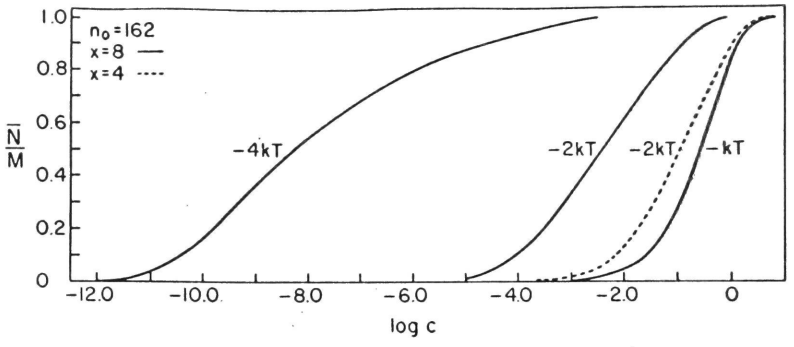


Figure 18. Isotherms for the ligands x=8 (——) and x=4 (- - -) binding to the lattice n_0=162, with values of ϵ as indicated. The variations in the position and shape of the isotherms with ϵ and x are discussed in the text.

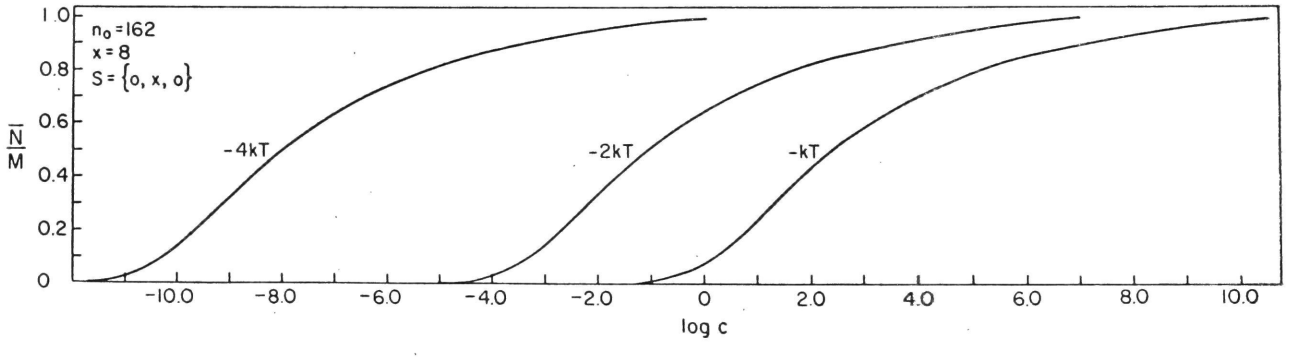


Figure 19. Isotherms for totally adsorbed ligands. The curves were obtained for the ligand x=8 binding to the lattice n_0=162 (cf. Figure 18). However, the ligands were constrained to bind with all segments adsorbed to the surface ($S=\{0,x,0\}$). The value of ϵ is indicated by each curve.

negative since the energy gained upon binding is better able to compensate for the resulting loss of translational and configurational entropy. Such a correlation between binding energy and curve position, in which the more negative energy values are associated with binding at lower values of c, has been stressed by other workers in discussions of enzyme-substrate complexes (Koshland et al., 1966).

The isotherms of Figure 18 are not symmetrical about their midpoints. Broadening is observed as saturation is approached since the decrease in the number of vacant surface sites reduces the number of configurations available to additional ligands, making their binding more difficult. This flattening is most marked for ϵ=-4kT, because the high $\bar{\theta}$ values result in a more rapid depletion of the vacant surface sites by the first ligands bound. Deviations of binding curves from symmetry also have been noted in studies of enzyme-substrate complexes (Koshland et al., 1966). Cooperativity is said to occur when the first substrate molecules bound facilitate the binding of additional substrate molecules, causing the binding curves to steepen. On the other hand, broadening of enzyme-substrate binding curves is said to result from negative cooperativity, when the first substrate molecules bound make the binding of additional substrate molecules more difficult. Thus, the flattening of the isotherms in the present model may be referred to as negative cooperativity because of its formal similarity to the broadening of enzyme-substrate binding curves. However, whereas the negative cooperativity observed in enzyme-substrate interactions

generally is attributed to conformational changes in the protein (Koshland et al., 1966), the form of negative cooperativity found in the present model has a totally different origin and might be characterized as a type of ligand-ligand interaction, differing from a direct energetic interaction (such as the electrostatic effects in multiple ionization (Klotz, 1953)).

The fraction of the maximum number of ligands bound, $\bar{N}/M$, has been used as the ordinate in Figure 18 to establish the effect of chain length on the binding isotherm. The isotherm shifts to lower c values as x is increased, since the energy of the site-segment interaction dominates the free energy change per segment upon binding. The isotherm for x=4 is somewhat steeper than that for x=8 since it is easier to find smaller clusters of contiguous vacant surface sites. Thus, the negative cooperativity responsible for flattening is less for the shorter ligand.

As a further illustration of the negative cooperativity of the present model, isotherms have been obtained for the case in which all segments of the bound ligands are adsorbed to surface sites, no loops or free ends being permitted ($S = \{0,x,0\}$). As illustrated in Figure 19, the curves are equally broad at all values of ϵ since $\bar{\theta}=1.0$ in each case, resulting in a maximum amount of negative cooperativity.

When ligand binding to a patch of sites is studied as a function of ϵ, the variations in curve position and shape are identical to those described for a lattice which covers a complete surface (Figure 20; cf. Figure 16). Isotherms for a patch and a complete surface with an identical number of sites are compared in Figure 21. As a result of the

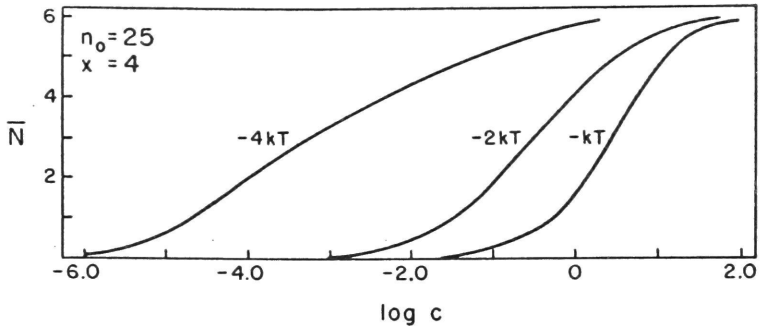


Figure 20. Isotherms for the ligand x=4 binding to a patch of 25 sites, with values of ϵ as indicated. The variations in curve position and shape with ϵ are similar to those observed for the lattice which covers a complete surface (Figure 18).

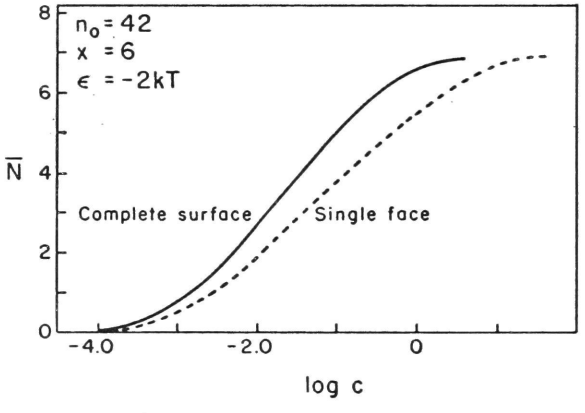


Figure 21. Isotherms for the ligand x=6 binding to a complete surface (——) and a patch (- - -) of identical size (n_0=42). Both curves were calculated with ϵ=-2kT. Due to edge effects, binding to the patch is somewhat more difficult than binding to the complete surface.

edge effects mentioned in the discussion of configurational parameters (Figure 17), binding to the patch is somewhat more difficult.

As discussed in Chapter I, when the commonly used theory of multiple equilibria (Klotz, 1953) forms the theoretical framework for analyzing protein-ligand interactions, flexible ligands are assumed to bind in a rigid configuration. Using the model for rigid binding discussed in Chapter III, a comparison of the isotherms for single-class binding in the theory of multiple equilibria with those for homogeneous systems in the present model is possible. For example, the isotherms illustrated in Figure 22 are obtained when the systems of Figure 18 are studied under the conditions of rigid binding. The resulting curves are symmetrical about their midpoints, as expected for single-class binding (Figure 1), and there is no variation in curve shape with energy. The shift in position as the energy is varied is in the direction observed in the model for flexible binding. However, in contrast to the latter, the entropy loss per segment dominates the free energy change per segment at $\epsilon = -2kT$, and the position of the isotherm shifts to <u>higher</u> values of c when the chain length is increased. With rigid binding, values of $|\epsilon| \gg 2kT$ are necessary to produce the opposite chain length dependence, since the bound ligand lacks configurational entropy, giving a much larger entropy loss per segment.

In the rigid binding model, as in the theory of multiple equilibria, when all ligands bind as one class the equilibrium constants for the formation of successive protein-ligand complexes differ only by a statistical factor which accounts for the different ways of placing the ligands

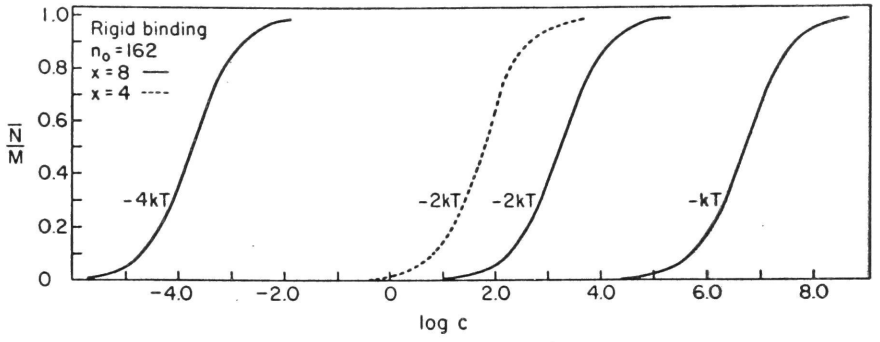


Figure 22. Isotherms for rigid binding. The curves for the ligand x=8 (——) were obtained with M=20 and E=8ϵ, while those for x=4 (- - -) were obtained with M=40 and E=4ϵ, allowing a comparison with the isotherms of Figure 18. Although the variation in curve position with ϵ is similar to that found in the model for flexible binding, the variation in position with x is opposite to that found in the flexible binding model. The curve shape is not affected when either x or ϵ is varied, in contrast to the flexible binding model.

on the protein. Thus, each constant can be written as the product of a statistical factor and an intrinsic association constant, the same for all N (equation I.2). For such systems, plots of $\bar{N}/c$ vs. $\bar{N}$ or $1/\bar{N}$ vs. $1/c$ will be linear (Figure 1). A very significant feature of the present model for flexible binding is that the above two plots fail to be linear, even for homogeneous systems (Figures 23 and 24). Due to the neglect of the internal degrees of freedom of the bound ligands in the rigid binding model, single-class binding not only implies that the surface is chemically homogeneous, but that all ligands bind in an identical configuration as well. In the present model, in which the flexibility of the bound ligands is included, any given ligand can bind in many different configurations. The number of possible configurations decreases as additional ligands are bound, so that the ligands do not bind in an identical fashion even though all of the binding sites are chemically the same. Each ligand binds with a characteristic equilibrium constant which reflects its unique manner of binding as well as simple statistical considerations. Hence, the conventional plots for multiple equilibria are not linear. The asymmetry of the isotherms (Figures 18 to 21) also results from this failure of the ligands to bind in an identical configuration.

Figure 25 illustrates the variation in the equilibrium constants as successive ligands are bound (equation III.6), as calculated with the present model and the model for rigid binding.

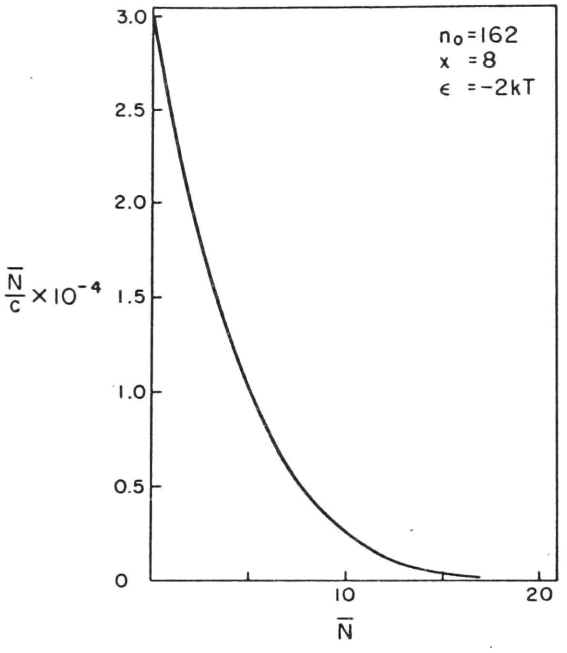


Figure 23. A Scatchard plot of binding data calculated by the present model (cf. Figure 18). The plot is not linear even though the system is homogeneous.

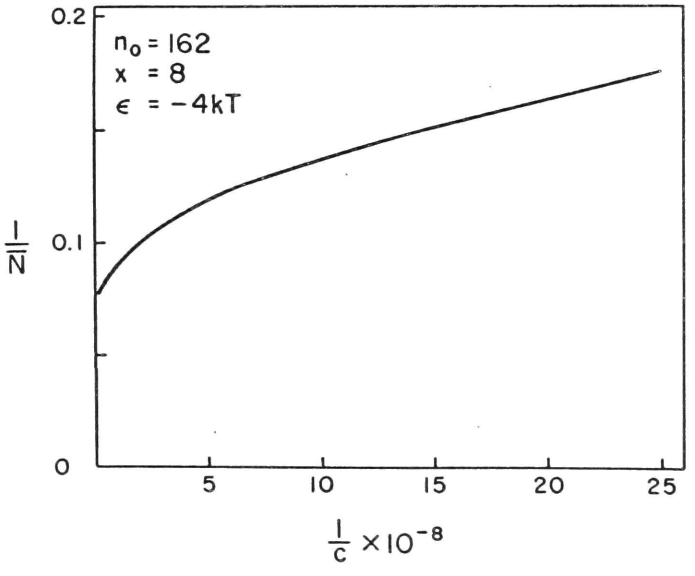


Figure 24. A double-reciprocal plot of binding data calculated by the present model (cf. Figure 18). The plot is not linear even though the system is homogeneous.

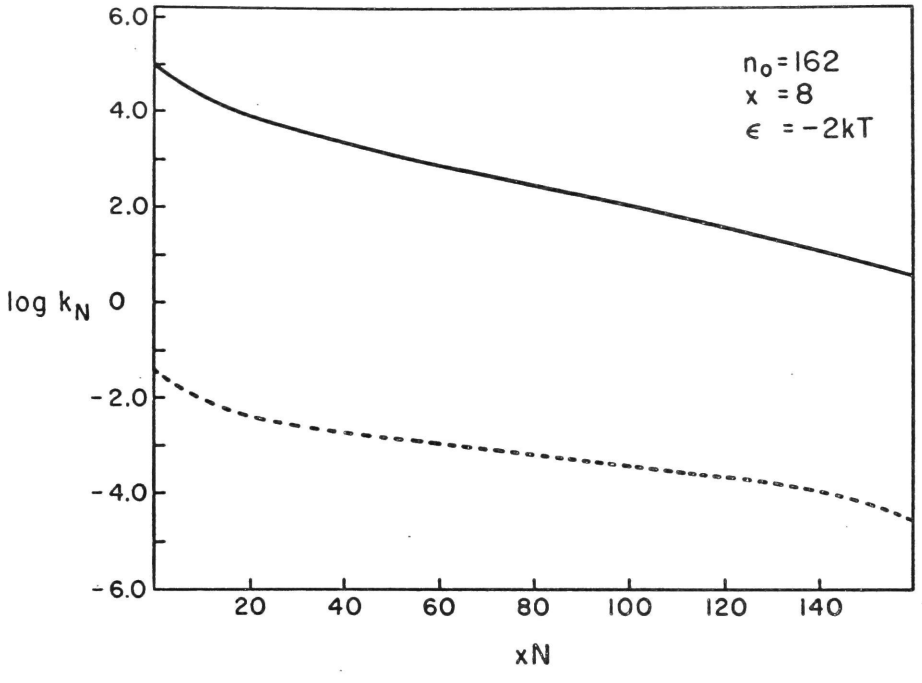


Figure 25. The variation in $\log k_N$, the logarithm of the equilibrium constant for the binding of the Nth ligand, with xN, the number of segments belonging to bound chains. The results calculated with the present model for flexible binding (——) may be compared with those obtained with the rigid binding model (---), in which the equilibrium constants differ only by a statistical factor.

CHAPTER V

RESULTS FOR HETEROGENEOUS SYSTEMS

Physical Parameters

Appropriate values for several physical parameters must be obtained before heterogeneous systems (in particular, the interactions between substituted alkanes and proteins) can be treated with the present model.

Definition of the Ligand Segment The segment of the ligand, as defined for the purposes of enumerating configurations, corresponds to a step in the random walk of the ligand chain. This so-called statistical segment (Flory, 1953) need not correspond to the chemical segment of the ligand on a one-to-one basis (for the hydrocarbon portion of the substituted alkanes, the latter would be defined as a methylene group ($-CH_2-$)). In fact, a one-to-one correspondence between statistical and chemical segments would be somewhat unrealistic given the triangular lattice of the present model, since adjacent carbon-carbon bonds always form a tetrahedral angle (109.5°). Thus, it is more realistic to equate two chemical segments to a single statistical segment, which allows the ligand to be accommodated quite easily on a triangular lattice. This choice of statistical segment has the additional advantage of reducing the size of the systems, thereby simplifying the calculations.[1] This

[1]The naturally-occurring substituted alkanes, as well as those most frequently used experimentally, have an even number of carbon atoms.

definition of the statistical segment for the hydrocarbon portion of the ligand suggests that the (non-carbon) atoms of the ionic or polar head group be regarded as a single segment.

Volume of a Site The volume of a site, v_s, may be equated to the volume of the statistical segment. From the known bond lengths, bond angles, and van der Waals radii characterizing the atoms of a hydrocarbon chain (Scheraga, 1968), an estimated value of 48 $\mathring{A}^3$ has been calculated,[1] equivalent to a sphere of radius (r_s) 2.25 $\mathring{A}$. A very similar estimate for the $-CH_2-CH_2-$ volume, 42 $\mathring{A}^3$, has been obtained from Courtauld's models (Poland and Scheraga, 1965).[2]

Surface Sites The precise chemical grouping to which the term site is to be applied must correspond in size to the previously defined statistical segment and the resulting value of v_s. Thus, the backbone portion of each amino acid is divided into two sites, corresponding to the carbonyl group ($>C=O$) and the α-carbon-amide group ($-NHC^{\alpha}H-$), respectively. A nonpolar (hydrocarbon)site is defined as two methylene groups ($-CH_2-CH_2-$) to correspond directly to a hydrocarbon segment. Most polar and ionic groups on amino acid side chains ($-COO^-$, $-NH_3^+$, $-CH_2OH$, etc.) constitute a single site. Other assignments can be

---

[1]The volumes of the head groups of interest do not differ sufficiently to warrant the use of an additional parameter.

[2]It should be emphasized that v_s appears only in the calculation of $q_{x,u}(V)$, the partition function for the unbound ligand, and hence μ_0, the standard chemical potential of the ligand. Thus, none of the conclusions of the present model is affected by the value of v_s.

made in a manner consistent with these examples.

An estimate of the number of surface sites on a protein of a given size is needed. Two sites are necessary to represent the backbone of each amino acid, and the side chain volumes cited by Waugh (1954) suggest an average of two sites per amino acid side chain. Thus, a protein with n amino acids contains a total of 4n sites. If a protein is not too different from spherical, an estimate of the number of the 4n sites which lie on the surface can be obtained by comparing the volume of a shell of thickness $2r_s$ to the total volume, $4nv_s$. For a small protein such as myoglobin (with $n \approx 150$), approximately 325 sites would appear on the surface, while a larger protein such as hemoglobin or serum albumin (with $n \approx 575$) would require approximately 900 surface sites. In many applications, however, a patch of sites is used to represent the binding region, and no fixed relationship between the size of the patch and the size of the protein is necessary.

Site-Segment Interaction Energies No precise values for the energies of interaction between various ligand segments and surface sites are available, either from theoretical or empirical considerations. However, estimates for these energies can be obtained with the aid of methods currently being developed for calculating protein conformations from known amino acid sequences (Scheraga, 1968). The segment and site are assumed to contact each other such that their constituent atoms are separated by a distance equal to the sum of their van der Waals radii, at which the potential energy of interaction between the atoms is

a minimum. Furthermore, it is assumed that standard bond length and bond angle values apply, and that no energy contributions from torsion about single bonds result when forming a given contact. The interaction energy (equation III.17) is equated to a sum of contributions from nonbonded (van der Waals) interactions, electrostatic interactions, and desolvation

$$\epsilon_j^{(n)} = {}_{NB}\epsilon_j^{(n)} + {}_{E}\epsilon_j^{(n)} + {}_{S}\epsilon_j^{(n)} \qquad (V.1)$$

where the subscripts NB, E, and S refer to the contribution from nonbonded interactions, electrostatic interactions, and desolvation, respectively.

To calculate the nonbonded contribution, ${}_{NB}\epsilon_j^{(n)}$, the Lennard-Jones "6-12" potential is used (Scheraga, 1968)

$${}_{NB}\epsilon_j^{(n)} = \sum_{i,k} \frac{a_{ik}}{r^{12}} - \frac{\beta_{ik}}{r^{6}} \qquad (V.2)$$

where the summation is over all extended atoms[1] of the site and segment in question, a_{ik} and β_{ik} are the coefficients for repulsion and attraction, respectively, and r is the distance between extended atoms i and k and is equated to r_{ik}, the sum of their van der Waals radii. The coefficient β_{ik} is calculated from the Slater-Kirkwood equation (London, 1937; Pitzer, 1959)

$$\beta_{ik} = \frac{\Delta_i \Delta_k}{\sqrt{(\frac{\Delta_i}{\Lambda_i})} + \sqrt{(\frac{\Delta_k}{\Lambda_k})}} \frac{3}{2}\hbar \frac{q_e}{\sqrt{m_e}} \qquad (V.3)$$

[1]For simplicity, hydrogen atoms are not considered individually. Instead, they are regarded as part of an extended atom, such as $-CH_2-$ or -NH- (Gibson and Scheraga, 1967).

where q_e is the electronic charge, m_e is the electronic mass, $\hbar$ is Planck's constant, Δ_i and Δ_k are the atomic polarizabilities of the extended atoms i and k, respectively, and Λ_i and Λ_k are the effective number of outer-shell electrons for the extended atoms in question. Values for the parameters Δ and Λ have been obtained from Gibson and Scheraga (1967). The coefficients a_{ik} are evaluated by solving the equations

$$\left(\frac{\partial \left(\frac{a_{ik}}{r^{12}} - \frac{\beta_{ik}}{r^{6}} \right)}{\partial r} \right)_{r=r_{ik}} = 0 \qquad (V.4)$$

The electrostatic contribution, $E_j^{\epsilon(n)}$, is a sum of Coulomb interactions between all pairs of extended atoms of the site and segment (Hirschfelder et al., 1954). For a charged atom i and a polar atom k (charge-dipole interaction), the contribution is

$$-\frac{1}{3kT} \frac{q_i^2 \mu_k^2}{r_{ik}^4 D} \qquad (V.5)$$

where q_i is the charge on atom i, μ_k is the dipole moment of atom k, and D is the local dielectric constant of the medium, taken as 3.5 in the vicinity of a protein in aqueous solution (Brant and Flory, 1965). When the interaction is between two polar atoms (dipole-dipole interaction)

$$-\frac{2}{3kT} \frac{\mu_i^2 \mu_k^2}{r_{ik}^6 D} \qquad (V.6)$$

The contribution for a charged atom i and a nonpolar atom k (charge-induced dipole interaction) is

$$-\frac{q_i^2 \Delta_k}{2r_{ik}^4 D} \qquad (V.7)$$

Electrostatic interactions between a polar and a nonpolar atom (dipole-induced dipole interactions) are sufficiently weak to be neglected. Similarly, dipole-dipole, charge-induced dipole, and dipole-induced dipole interactions may be neglected in the calculation of charge-dipole interactions (equation V. 5), and dipole-induced dipole interactions may be neglected in the calculation of dipole-dipole interactions (equation V. 6) and charge-induced dipole interactions (equation V. 7). Estimates for electrostatic contributions due to charge-charge interactions have been obtained from Arvidsson (1965).

The desolvation contribution, $S\epsilon_j^{(n)}$, is equated to the sum of the energies of desolvating the extended atoms involved in the interaction. The contribution from a given atom consists of a desolvation energy per water molecule removed multiplied by the number of water molecules displaced when the contact is established. Estimates for the latter may be obtained by dividing the volumes of the approaching extended atoms (Bondi, 1964) by 30 $\mathring{A}^3$, the volume of a water molecule (Gibson and Scheraga, 1967). Given the polar character of a water molecule, the desolvation energies for charged and polar groups may be calculated as charge-dipole and dipole-dipole interactions, respectively (equations V. 5 and V. 6). Estimates of nonpolar group desolvation energies may be obtained from the thermodynamic properties of aqueous solutions of the hydrocarbons (Némethy and Scheraga, 1962; Gibson and Scheraga,

1967).

Selected site-segment interaction energies are presented in Table II. The values listed include those which appear in the calculations discussed in the subsequent sections of this Chapter. Because the carboxylic acids are among the most commonly encountered flexible ligands, they will be emphasized in the results presented below. The ϵ-amino group of lysine has been selected as the ionic site to which the carboxyl segment binds, based on the evidence of Teresi (1950) and Spector et al. (1969). Electrostatic interactions between these sites have been neglected. All calculations have been performed at T=25°C.

Table II

Site - Segment Interaction Energies

Site	ϵ (cal/mole)[1]	
	Carboxyl Segment	Hydrocarbon Segment
$-NH_3^+$	-5000	$\approx \infty$
$> C=O$	-2700	-150
$-NHC^{\alpha}H-$	-1500	-450
$-CH_2OH$	-2000	-1100
$-CH_2-CH_2-$	$\approx \infty$	-1300
$-COO^-$	$\approx \infty$	$\approx \infty$

[1]Interactions between a segment and site of like charge or between hydrocarbon groups and ionic groups are so unfavorable that the energies have been assigned the value ∞, giving a Boltzmann factor of zero for these contacts.

Configurational Parameters[1]

Many of the general properties of heterogeneous systems can be illustrated using a lattice with just two types of sites. A surface with 12 ionic sites ($-NH_3^+$) and 150 nonpolar sites ($-CH_2-CH_2-$) can be constructed from the lattice $n_0=162$ by assigning ionic groups to the class 1 sites and nonpolar groups to the sites of classes 2 and 3 (see Figure 8). Figure 26 illustrates the variations in $\bar{\theta}$ and $\bar{d}_0$ with xN for the carboxylic acids x=9 and x=5. The energy of the carboxyl segment-amino group interaction is so favorable the the first 12 ligands bind almost exclusively in configurations with their carboxyl segments adsorbed to ionic sites. Thus, a sharp increase in $\bar{d}_0$ and a resulting drop in $\bar{\theta}$ occur at N=12 since configurations with $d_0 \geqslant 1$ are the only ones available thereafter. The more gradual changes in $\bar{\theta}$ and $\bar{d}_0$ reflect the reduced availability of vacant surface sites, as observed in homogeneous systems (Figures 11 and 15).

As illustrated in Figure 27, similar variations in $\bar{\theta}$ and $\bar{d}_0$ occur when binding to a patch of ionic and nonpolar sites is considered ($n_0=25$; cf. Figure 16). The results with one ionic site (the single class 7 site) may be compared with those obtained with three ionic sites (the class 4 sites). The sharp changes in $\bar{\theta}$ and $\bar{d}_0$ when N exceeds the number of ionic sites are somewhat larger for the latter case, since fewer vacant nonpolar

---

[1]Attention will be focused on two parameters, $\bar{\theta}$ and $\bar{d}_0$ (the latter referring only to the carboxyl end of the ligand), which are especially sensitive to the effects of heterogeneity.

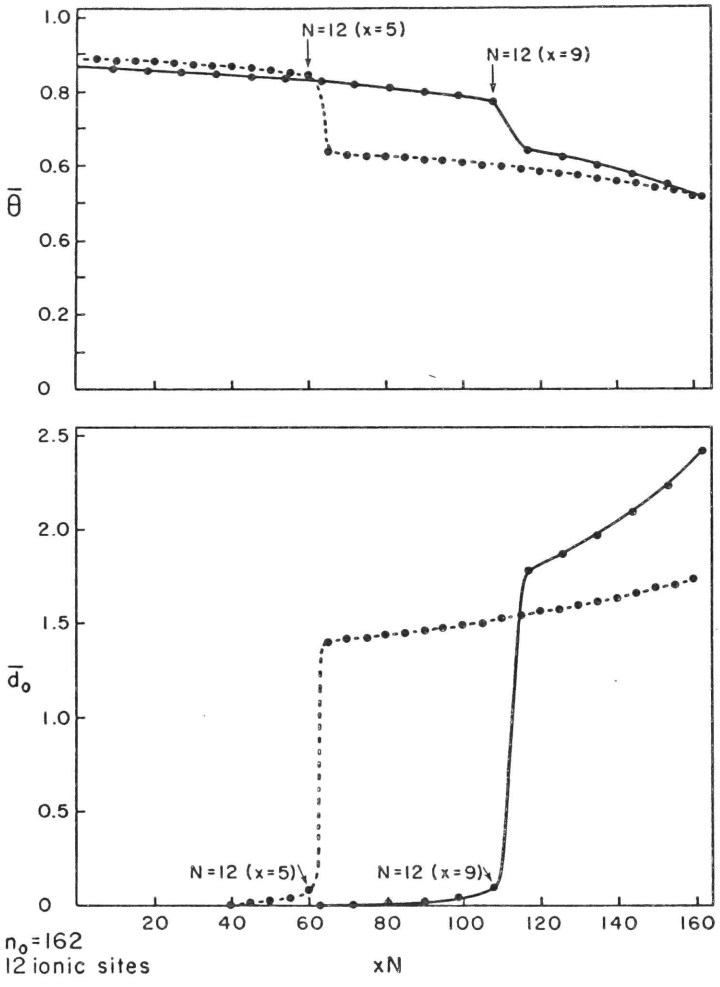


Figure 26. Variations in $\bar{\theta}$, the average fraction of ligand segments adsorbed, and $\bar{d}_o$, the average length of the free carboxyl end, with xN, the number of segments belonging to bound chains, for the carboxylic acids x=9 (●—●) and x=5 (●- -●). The curves were calculated for the lattice n_o=162, with 12 ionic sites and 150 nonpolar sites. A marked change in the average configuration of the ligands occurs at N=12, corresponding to the saturation of the ionic sites.

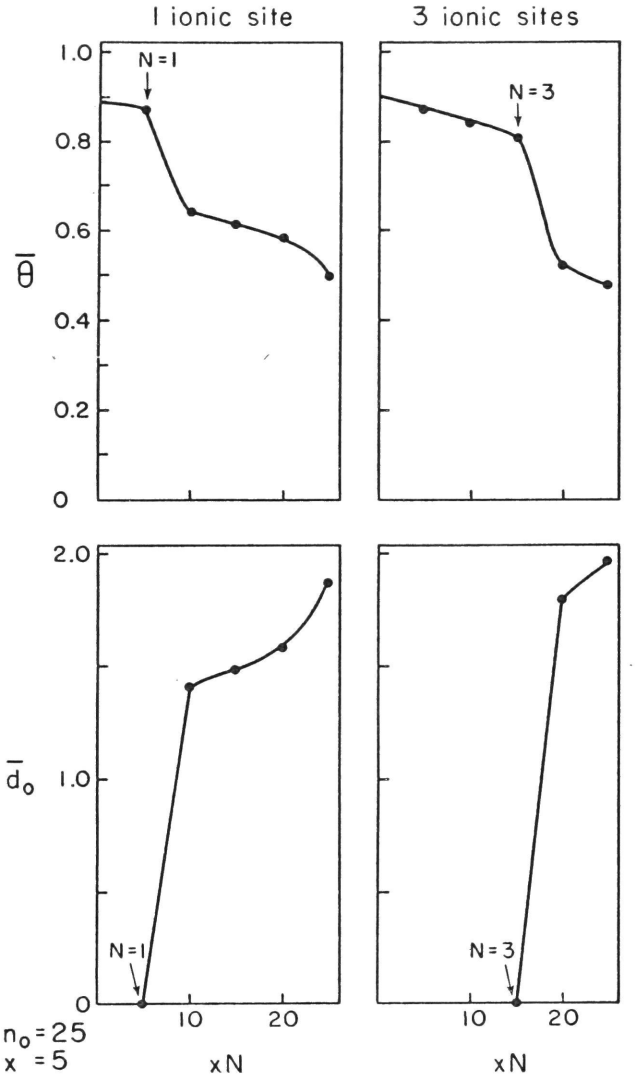


Figure 27. Variations in the configurational parameters $\bar{\theta}$ and $\bar{d}_0$ for the carboxylic acid x=5 binding to a patch of 25 sites. The curves on the left were obtained with one ionic site and 24 nonpolar sites, while those on the right were obtained with three ionic sites and 22 nonpolar sites.

sites remain after the ionic sites are occupied.

The specific position of the ionic sites is relatively unimportant in determining the behavior of the parameters $\bar{\theta}$ and $\bar{d}_0$, as shown in Figure 28, which illustrates binding to a patch of 42 sites (cf. Figure 17) with three ionic sites located either near the center of the patch (the class 9 sites) or near its corners (the class 4 sites).[1] Figure 28 also illustrates the effect of introducing additional types of sites, such as polar sites representing the chemical groups of the peptide backbone. Thus, in the central arrangement, $-NHC^{\alpha}H-$ groups have been placed on the class 8 sites and $>C=O$ groups on the sites of classes 7 and 10, while $-NHC^{\alpha}H-$ groups and $>C=O$ groups have been placed on class 3 and 2 sites, respectively, in the corner arrangement, for a total of 12 polar sites in each case.[2] With such additional heterogeneity, the changes in the configurational parameters are not as sharp as those for the surfaces with just two types of sites, since the polar sites remain as possible binding sites for the carboxyl segment after the ionic sites are filled.

Considering the lattices with polar sites in more detail, the rise in $\bar{d}_0$ and drop in $\bar{\theta}$ are largest for the corner arrangement, since the configurations containing carboxyl segment-polar site interactions make a smaller contribution to the partition function due to the reduced coordi-

---

[1]When referring to these two arrangements of ionic sites, the terms central arrangement and corner arrangement will be used.

[2]The polar sites have been placed near the ionic sites because such an arrangement is most consistent with current concepts of protein structure.

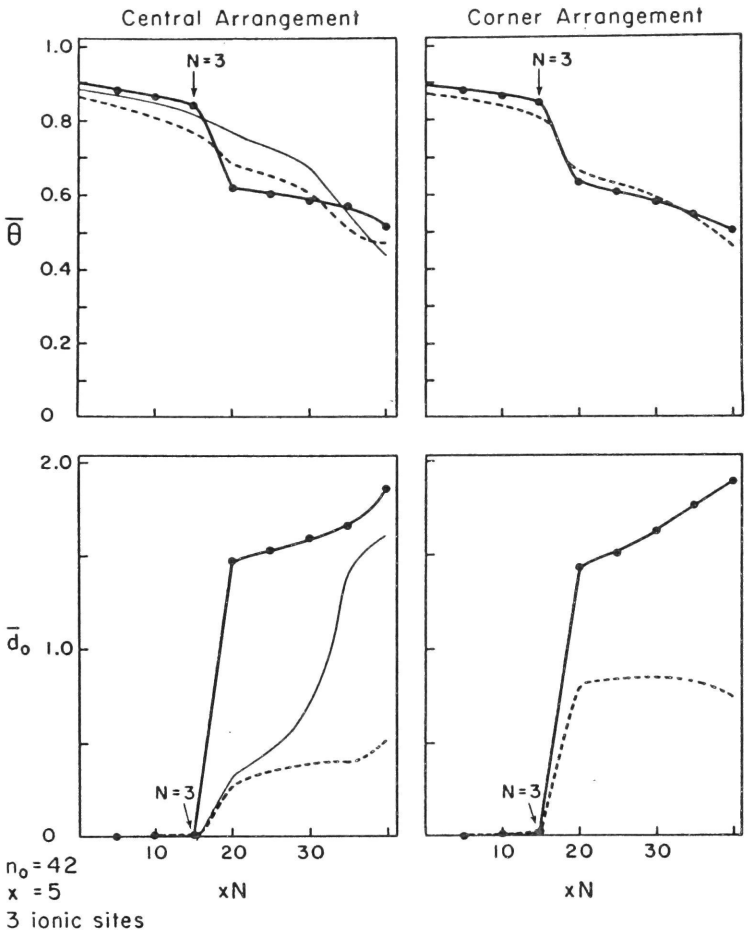


Figure 28. Variations in the configurational parameters $\bar{\theta}$ and $\bar{d}_0$ for the carboxylic acid x=5 binding to a patch of 42 sites, with 39 nonpolar sites and three ionic sites (●—●). The results obtained with the ionic sites located near the center of the patch (central arrangement) may be compared with those obtained with the ionic sites located near the corners (corner arrangement). The effects of introducing 12 polar sites in the vicinity of the ionic sites in each arrangement are indicated (- - -). In the central arrangement, a patch with six polar sites and three ionic sites also has been studied (——).

nation number of the polar sites in this arrangement. The $\bar{n}_j(N-1)$ values indicate that those carboxyl segments which are adsorbed after N=3 for the corner arrangement bind preferentially to $-NHC^{\alpha}H-$ sites, since the greater inaccessibility of the $>C=O$ sites (in this particular arrangement) overrides their energetic advantage (Table II). But after N=6, the availability of surface sites has been curtailed considerably and the possible energy gain for binding a carboxyl segment to the relatively vacant $>C=O$ sites becomes more important, causing a decrease in $\bar{d}_0$ at high values of N. An accompanying rise in $\bar{\theta}$ does not occur, however, due to the scarcity of vacant nonpolar sites.

For the central arrangement, the ligands after N=3 bind with their carboxyl segments attached primarily to the $>C=O$ sites, since the latter not only are favored energetically over the $-NHC^{\alpha}H-$ sites, but are occupied to a lesser extent by previously bound hydrocarbon segments as well (for the latter, the $-NHC^{\alpha}H-$ sites are favored energetically over the $>C=O$ sites; see Table II). The accelerated rise in $\bar{d}_0$ after N=6 results from the reduced accessibility of the $>C=O$ sites.

Figure 28 also illustrates configurational parameters for the case in which the central arrangement includes only <u>six</u> polar sites ($-NHC^{\alpha}H-$ groups on the class 10 sites and $>C=O$ groups on the class 7 sites). On this lattice, as with the central arrangement with 12 polar sites, the carboxyl segments bind preferentially to the $>C=O$ sites after N=3. The rather marked change in the parameters at N=6 reflects the saturation of the three $>C=O$ sites; binding of the carboxyl segment to $-NHC^{\alpha}H-$ sites

does not occur thereafter due to the reduced accessibility of these sites.

Equilibrium Properties

Isotherms for the carboxylic acids x=9 and x=5 binding to the lattice n_0=162 with 12 ionic sites (cf. Figure 26) are presented in Figure 29. The curves might be described as biphasic, since an intermediate plateau occurs at $\bar{N}$=12 (corresponding to the saturation of the ionic sites) followed by another rise to the final plateau at $\bar{N}$=M (representing the binding of ligands with free carboxyl ends). The second phase occurs at higher ligand concentrations in the absence of the favorable energy contribution from the carboxyl segment-amino group interaction.

The first phase of the isotherm is broader for the ligand x=9 since the 12 ionic sites are saturated at xN=108, while the corresponding number is only 60 for x=5 so that less broadening due to negative cooperativity results. A similar correlation between chain length and isotherm broadness is found in the second phase of the curves. For the ligand x=5, the second phase is even steeper than the curve expected for the binding of four hydrocarbon segments to the available number of nonpolar sites. Due to the entropy gain possible when $d_0 > 1$ ($H_0(n) > 1$ for $n > 1$), the prior formation of a free end by the carboxyl segment favors the lengthening of that end. As a result, the ligands bound subsequently to N=12 have lower $\bar{\theta}$ values than would be anticipated given the hydrocarbon segment-nonpolar site interaction energy ($\approx$-2kT at 25°C; see Figure 18), steepening the isotherm. No such effect is possible for x=9,

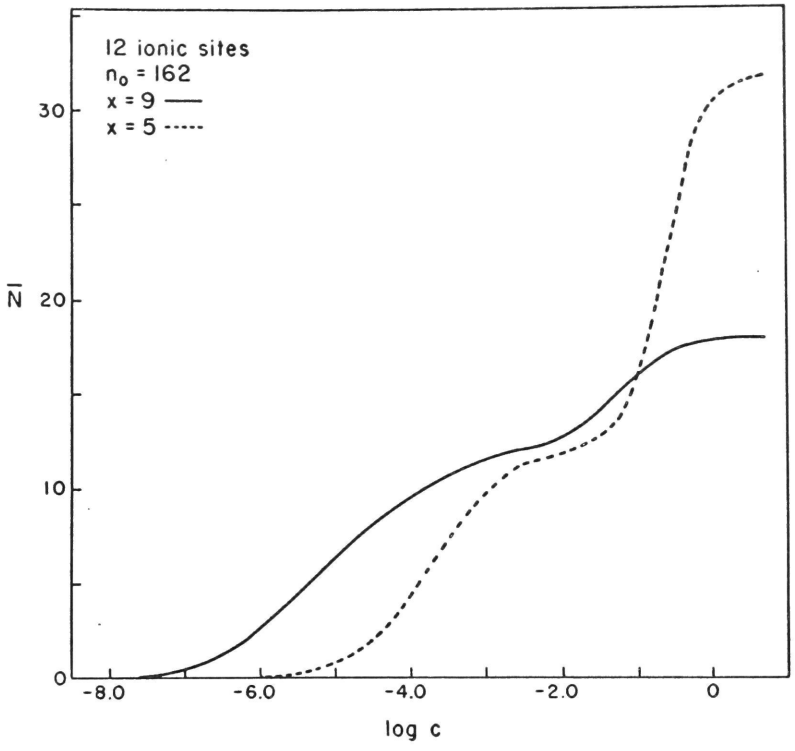


Figure 29. Isotherms for the carboxylic acids x=9 (——) and x=5 (- - -) binding to the lattice n_0=162 with 12 ionic sites. The marked change in the average configuration of the ligands when the ionic sites are saturated (Figure 26) produces an intermediate plateau in the curve at $\bar{N}$=12, giving a biphasic isotherm.

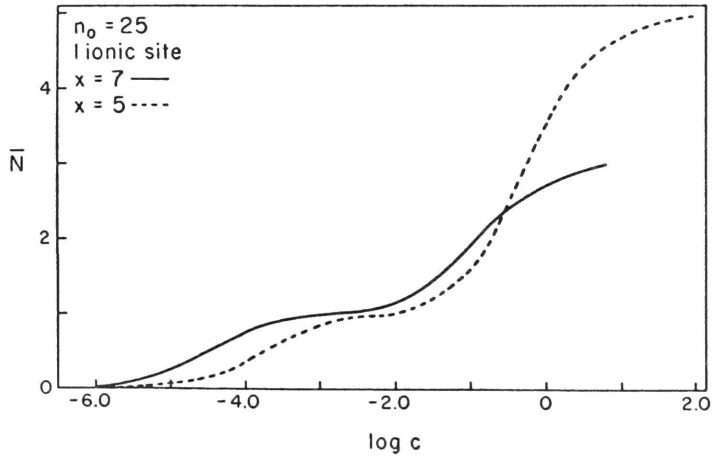


Figure 30. Isotherms for the carboxylic acids x=7 (——) and x=5 (- - -) binding to a patch of 25 sites with one ionic site. An intermediate plateau is observed at $\bar{N}$=1.

for at the extent of nonpolar site occupancy at N=12, hydrocarbon segments would form a free end even if the unbound carboxyl segment were not present.

Isotherms for a patch of 25 sites with one ionic site (cf. Figure 27) are shown in Figure 30. As for the lattice which covers a complete surface, the isotherms are biphasic with an intermediate plateau at a value of $\bar{N}$ corresponding to the saturation of the ionic sites, and comparable effects of chain length on isotherm position and shape are observed.

Figure 31 illustrates binding to a patch of 42 sites with three ionic sites in two different arrangements (cf. Figure 28). The location of the ionic sites hardly affects either the position or the shape of the isotherms. The introduction of 12 polar sites in each arrangement shifts the first phase of the curve to higher values of c, since a smaller amount of energy is gained when ligands bind in configurations with some hydrocarbon segments adsorbed to polar sites (such configurations will occur since the polar sites are adjacent to the ionic sites). On the other hand, the presence of polar sites shifts the second phase of the isotherms to lower values of c, because of the energy gained from carboxyl segment-polar site interactions (without the polar sites, the carboxyl segment would be unbound). In both phases, the shifts are larger for the central arrangement, in which most of the polar sites are hexa-coordinated so that configurations which include them make larger contributions to the partition function. As might be expected, the isotherm for the central arrangement with only six polar sites (cf. Figure 28) would fall between the two illustrated for the central arrangement.

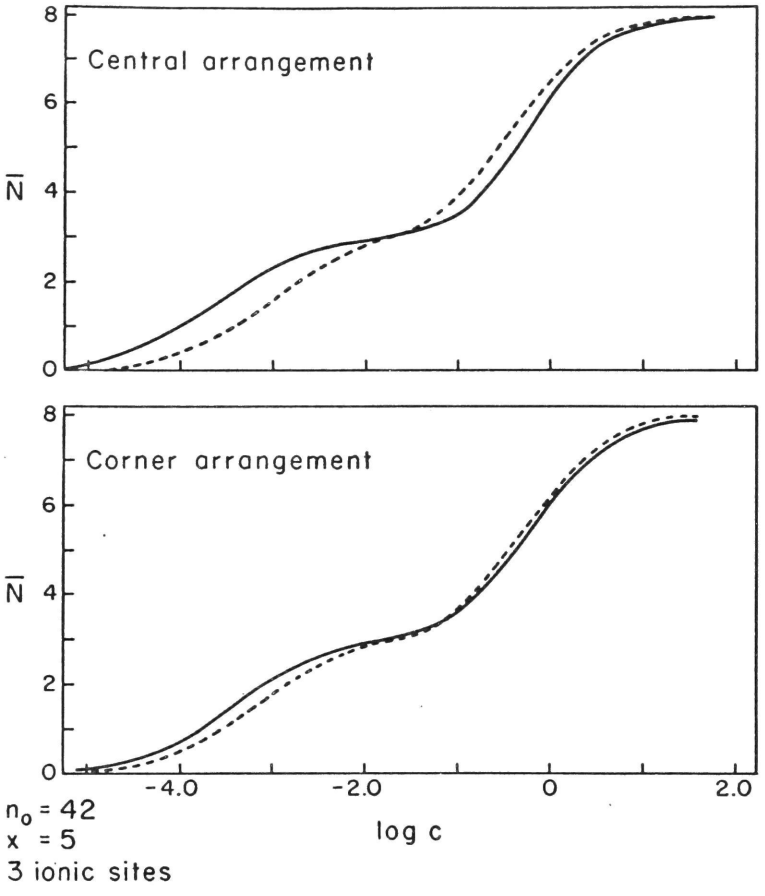


Figure 31. Isotherms for the carboxylic acid x=5 binding to a patch of 42 sites with three ionic sites (——). In the upper figure, the ionic sites are located near the center of the patch (central arrangement), while in the lower figure they are located near the corners (corner arrangement). The effects of introducing 12 polar sites in the vicinity of the ionic sites also are indicated (- - -).

Thus, the general effect of introducing polar sites is to decrease the difference in position between the two phases of the isotherm, shortening the intermediate plateau region. Further studies have shown that if a sufficiently large number of polar sites is introduced (for example, if less than one-half of the lattice sites are nonpolar sites), the intermediate plateau disappears almost entirely, and the isotherm appears to be a simple, sigmoidal curve, similar to those obtained for homogeneous systems. Thus, if a sigmoidal binding isotherm is observed in an experiment, this in itself does not establish that the binding region is chemically homogeneous. Conversely, the presence of a marked intermediate plateau may be taken as evidence that a fairly large number of nonpolar sites is present in the vicinity of the ionic sites.

Comparison with Experimental Data

Experimental data for the binding of hexadecanoic (palmitic) acid to serum albumin (Goodman, 1958) are presented in Figure 32. The results are typical of those observed in other binding studies using carboxylic acids, alkyl sulfates, and other long-chain flexible ligands (Karush and Sonenberg, 1949; Teresi and Luck, 1952: Ray et al., 1966; Reynolds et al., 1967, 1968; Spector et al., 1969). If the theory of multiple equilibria (Klotz, 1953) is used as the theoretical framework for analyzing protein-ligand interactions, the nonlinearity of the Scatchard plot indicates that the ligands do not bind as a single class (Chapter I). In the example under discussion (Goodman, 1958), smooth curves that fit the data points

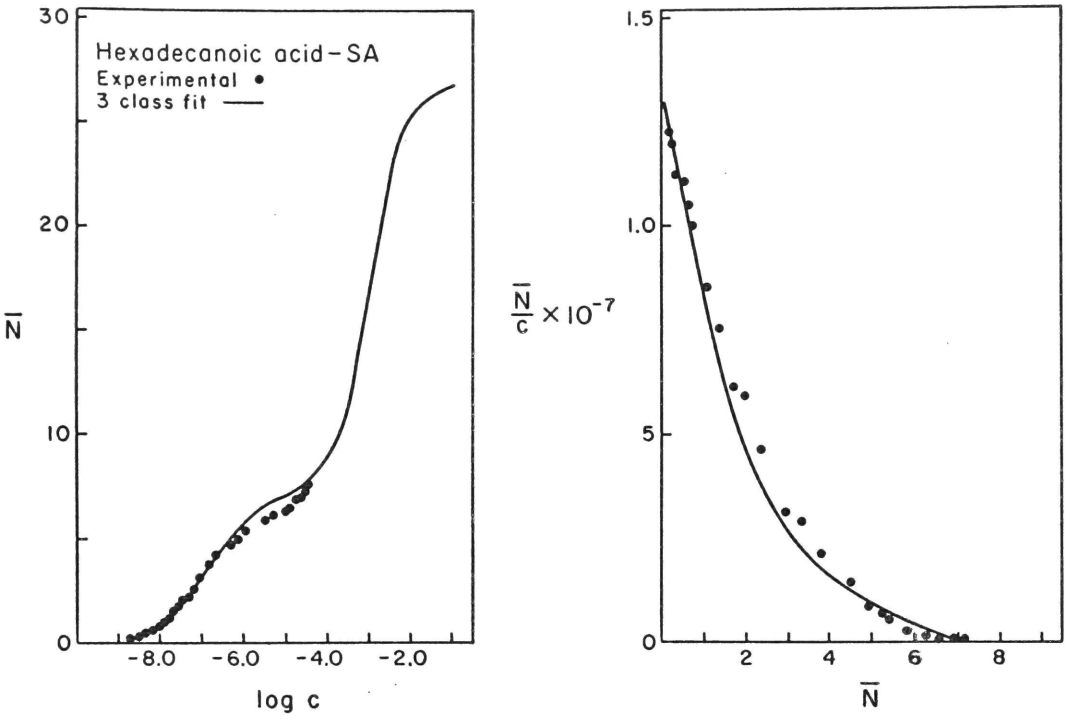


Figure 32. Experimental binding isotherm (left) and Scatchard plot (right) for the interaction of hexadecanoic (palmitic) acid with serum albumin (SA) (Goodman, 1958). The points correspond to actual experimental observations; the smooth curves represent a three-class fit of the data according to the theory of multiple equilibria (equation I.5). The data are typical of those observed in other binding studies using carboxylic acids, alkyl sulfates, and other long-chain flexible ligands (references are listed in the text).

quite well were obtained by assuming that the bound ligands belong to three classes (in accordance with equation I.5). As discussed below, the significance of such an analysis has certain limitations, especially since it is impossible experimentally to achieve saturation of the protein with ligand. Nevertheless, it is clear that the general features of the experimental points as well as the curve fitted to them are similar to those observed for heterogeneous systems in the present model. Most important, the isotherm is biphasic. The position of the intermediate plateau suggests that six ionic sites are present.

To demonstrate that the present model for flexible binding better represents experimental data than the theory of multiple equilibria in which the ligands bind rigidly, experimental isotherms for the binding of hexadecanoic and dodecanoic (lauric) acids to serum albumin (Goodman, 1958) are compared with those calculated using the present model and the rigid binding model for the lattice n_0=162 with 12 ionic sites[1] (Figure 33). Only the present model can explain the experimental observation that the strength of binding increases with the chain length of the ligand. In contrast, the rigid binding model predicts a decrease in the strength of binding with increasing chain length. Furthermore, with the specific set of physical parameters used, the isotherms for rigid binding occur at concentrations several orders of magnitude above those obtained

[1]This system corresponds to two-class binding in the rigid model. The first class contains 12 members which form one carboxyl segment-ionic site contact and x-1 hydrocarbon segment-nonpolar site contacts, while the second class contains M-12 members which bind with x-1 hydrocarbon segment-nonpolar site interactions only.

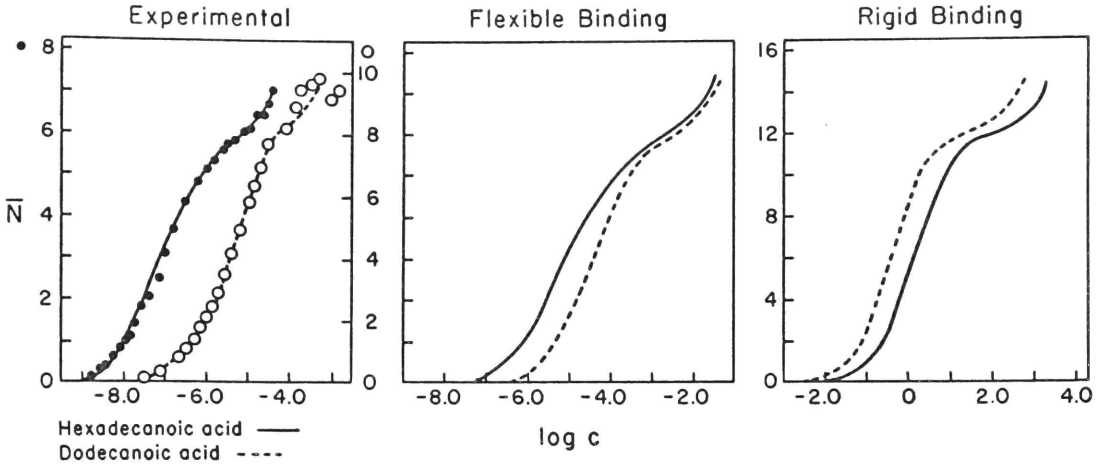


Figure 33. Comparison of experimental isotherms (left) for the binding of hexadecanoic (palmitic) and dodecanoic (lauric) acids to serum albumin (Goodman, 1958) with the isotherms calculated using the present model for flexible binding (center) and the rigid binding model (right). To facilitate the comparison of the curves, the ordinates have been chosen so that all intermediate plateaus occur at the same vertical level. Only the model for flexible binding predicts the experimental observation that the strength of binding increases as the chain length of the ligand increases. Furthermore, the isotherms for flexible binding occur at concentrations much closer to those observed experimentally.

experimentally, whereas the positions of the isotherms of the present model are much closer to those actually observed.

A quantitative demonstration of the chain length dependence is presented in Figure 34, in which ΔF_1^o, the change in the standard Helmholtz free energy for the binding of the first ligand (calculated as $-RT \ln k_1$) is plotted as a function of the number of carbon atoms in the ligand. In the present model for flexible binding, ΔF_1^o becomes more negative as the chain length of the ligand is increased, as observed experimentally for both the carboxylic acids and the alkyl sulfates (Karush and Sonenberg, 1949; Teresi and Luck, 1952; Goodman, 1958; Ray et al., 1966; Reynolds et al., 1967, 1968; Spector et al., 1969). This trend is not affected by the choice of lattice. In contrast, if it is assumed that the ligands bind rigidly, as in the theory of multiple equilibria, ΔF_1^o becomes more positive with increasing chain length.

The slope of the experimental ΔF_1^o vs. chain length curve is somewhat steeper than that of the curves calculated with the present model. Although the calculated slope could be made more negative simply by choosing a more negative energy contribution per segment, an unrealistically large value for the nonpolar site-hydrocarbon segment interaction energy would be required to correct the observed discrepancy. A more plausible explanation for the difference in slopes is that all configurations of the unbound ligand have been equally weighted. Actually, there will be some preference for compactly folded configurations which minimize unfavorable water-hydrocarbon contacts, giving a smaller value

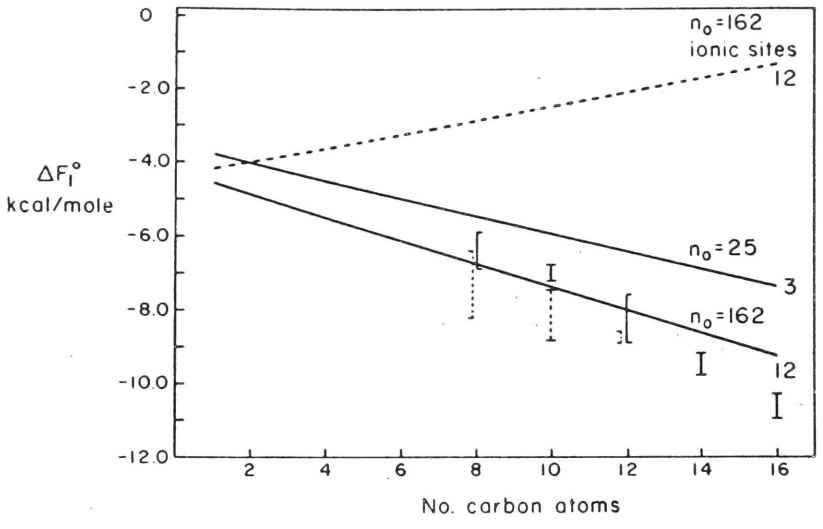


Figure 34. The variation in ΔF_1^o, the change in free energy for the binding of the first ligand, with the number of carbon atoms in the ligand. In the present model for flexible binding, ΔF_1^o decreases with increasing chain length (——), as observed experimentally for both the carboxylic acids (I) and the alkyl sulfates (I). The bars indicate the ranges of the experimental values reported in the literature (references are listed in the text). The trend of the present model is not altered by the value of n_0 or the number of ionic sites on the lattice. In contrast, in the rigid binding model ΔF_1^o increases with increasing chain length (---).

for $q_{x,u}(V)$, the partition function for the unbound ligand, than calculated. Correcting the latter would shift ΔF_1^o to more negative values, since a smaller entropy loss would result upon binding, and because $q_{x,u}(V)$ increases exponentially with the chain length (equation III.65), the shift would be larger for the longer ligands. Folding of hydrocarbon chains in aqueous solution actually has been observed for the carboxylic acids and alkyl sulfates in ionic conductance studies (Elworthy, 1963).

That the experimental isotherms are biphasic is a significant result if binding data are to be used to provide clues about protein structure, since an intermediate plateau will not be observed unless the binding region includes a fairly large number of nonpolar sites as well as ionic sites (see discussion following Figure 31). The importance of the hydrocarbon chain in producing strong binding of the substituted alkanes frequently has been mentioned in experimental studies, although the actual nature of the interaction of the hydrocarbon segments with the protein was not specified (Karush and Sonenberg, 1949; Teresi and Luck, 1952; Goodman, 1958; Ray et al., 1966; Reynolds et al., 1967, 1968; Spector et al., 1969). However, recent nuclear magnetic resonance (NMR) studies of the binding of long-chain anions to proteins by Rosenberg et al. (1969) provide direct evidence that the hydrocarbon segments of the ligand do interact with nonpolar groups on the protein, since the resonance peaks of the ligand protons shift upon binding in a direction indicating transfer to a more hydrophobic environment. The NMR studies of Gillberg-LaForce and Forsén (1970) confirm the observation made using more conventional

methods that binding becomes stronger with increasing chain length (Figure 34). In addition, the nature of the ionic head group of the ligand (carboxyl, sulfate, or sulfonate) was shown to have little effect on the strength of binding.

The isotherms calculated with the present model can be used to reproduce experimental data. To duplicate an experimental biphasic isotherm, a lattice with a number of ionic sites corresponding to the position of the intermediate plateau and a fairly large number of nonpolar sites must be selected. The results are less sensitive to the precise arrangement of sites on the surface, the presence of a small number of polar sites, and the size of the lattice. Of course, the calculation can be no more accurate than the physical parameters with which it is performed. For example, the site-segment interaction energies can be regarded only as approximate. In addition, there are uncertainties in the value of μ_0, the standard chemical potential of the ligand, due to the arbitrary assumptions involved in its calculation. Furthermore, the comparison of any results with experiment is complicated by scatter in much of the data and the impossibility of observing saturation of the protein with ligand. In spite of these difficulties, however, experimental isotherms can be matched quite closely with those of the present model. For example, the data of Goodman (1958) for the binding of dodecanoic acid to serum albumin (cf. Figure 33) can be fit by assuming that the protein surface consists of eight independent binding regions, each of which includes a single ionic site and 24 nonpolar sites

(Figure 35).[1] The only adjustment necessary to obtain this fit was less than a ten per-cent increase in μ_0, which causes a small shift in the horizontal position of the curve without changing any of its features. Such an increase in μ_0 would, in fact, be required if folding reduces the number of configurations of the unbound ligand, as discussed above.

Limitations of Class Analyses of Binding Data

The isotherms of the present model may be used to demonstrate potential errors that may arise from an uncritical use of class analyses in applications of the theory of multiple equilibria. In such analyses, binding data are fit with those values of M_j and $k_{int}^{(j)}$ for which the function Ψ is a minimum (cf. equation I.5)

$$\Psi = \sum_{i=1}^{p} \left(\bar{N}_i - \sum_{j=1}^{n_c} \frac{M_j k_{int}^{(j)} c_i}{1 + k_{int}^{(j)} c_i} \right)^2 \qquad (V.8)$$

where $\bar{N}_i$ and c_i correspond to a selected point of the theoretical isotherm and p is the number of points used. Using a nonlinear least squares algorithm (Laiken and Printz, 1970), the results in Table III were obtained for the binding of the carboxylic acids x=9, x=7, and x=5 to the lattice n_0=162 with 12 ionic sites. As indicated by the low values of Ψ, the theoretical isotherms can be fit quite well with two classes (the fit for the ligand x=5 is illustrated in the top graph of

---

[1] It also would be possible to fit the data using a single binding region with eight ionic sites or several binding regions which differ in size. However, the present choice is the simplest one possible, and in the absence of precise information about the arrangement of sites on the protein surface, additional refinements hardly are warranted.

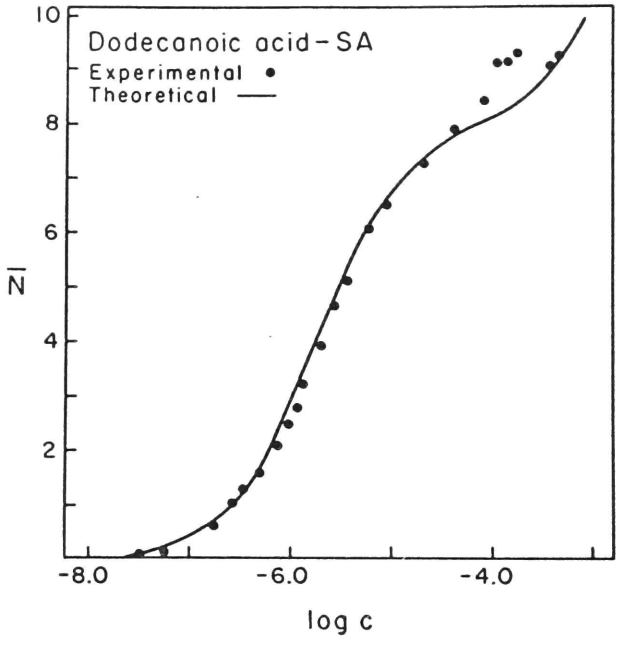


Figure 35. Fit of experimental data using the calculated isotherms of the present model. The experimental points represent the data of Goodman (1958) for the binding of dodecanoic (lauric) acid to serum albumin (SA) (cf. Figure 33). The data are matched quite well by an isotherm calculated with the present model (——), assuming that the protein contains eight independent patches of binding sites, each of which includes a single ionic site and 24 nonpolar sites.

Table III

Class Analyses of Theoretical Isotherms[1]

	Complete Isotherm			Incomplete Isotherm
	x=9	x=7	x=5	x=5
Theoretical[2] M	18	23	32	32
p	34	26	28	23
Two-Class Fit				
Ψ	12.2	3.07	8.17	1.96
M_1	10	11	11	12
M_2	7	12	23	66
$k_{int}^{(1)} \times 10^{-4}$	20.9	3.66	.712	.585
$k_{int}^{(2)}$	42.3	12.8	3.84	.530
Three-Class Fit				
Ψ	0.26	0.48	8.16	0.58
M_1	6	5	6	9
M_2	6	7	4	2
M_3	7	12	23	60
$k_{int}^{(1)} \times 10^{-4}$	79.2	13.1	.715	.698
$k_{int}^{(2)} \times 10^{-4}$	2.17	1.19	.708	.232
$k_{int}^{(3)}$	20.0	11.1	3.84	.580

[1]All values of M_j have been rounded to the nearest integer.

[2]Calculated using equation III.2.

Figure 36). Furthermore, the values of M_j giving the best fit correspond approximately to the number of ligands bound in each phase of the isotherm. A somewhat better fit is possible with three classes, simply because of the additional statistical degrees of freedom. The extra class forms from a division of the high-energy class (class 1) into two new classes and does not represent another type of binding. When experimental data have been analyzed using a three-class fit (Goodman, 1958), the sum of the first two classes has been found to correspond approximately to the position of the intermediate plateau. The latter most likely represents the number of ionic sites in the binding region, and the division of the high-energy class into two classes appears to have no physical significance.

However, the above analyses of calculated isotherms were performed on complete isotherms in which the plateau of the second phase of binding actually was observed. Experimental data usually are available only for that short portion of the second phase immediately following the intermediate plateau region (Figures 32 and 33). Therefore, it is of interest to consider the analysis of an incomplete isotherm. The latter can be formed from the calculated curves of the present model by cutting off a large portion of the second phase (bottom graph of Figure 36). As shown in Table III, a two-class fit of this truncated isotherm gives classes containing 12 and 66 members. In other words, although the size of the first class still corresponds to the number of ligands bound in the first phase of the isotherm, the size of the second class bears

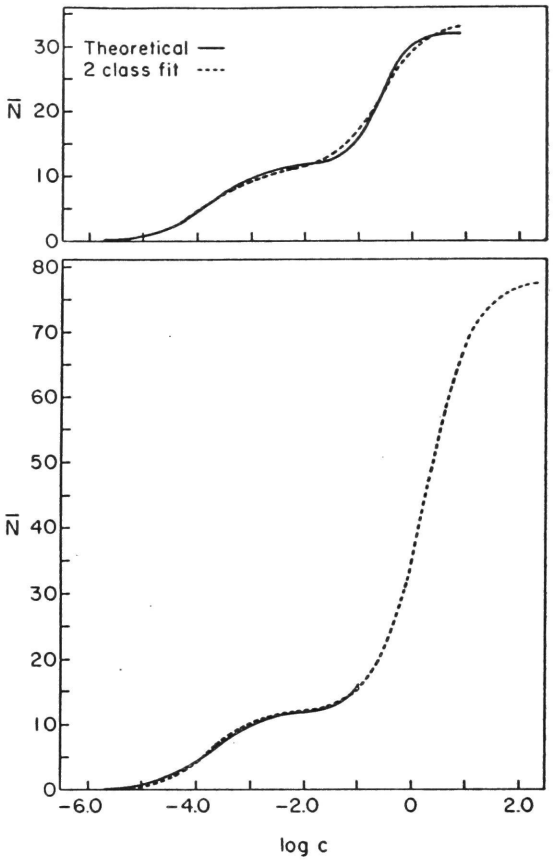


Figure 36. Class analyses of isotherms calculated with the present model. In the upper figure, the theoretical curve (——) was obtained with the carboxylic acid x=5 binding to the lattice n_0=162 with 12 ionic sites (cf. Figure 29). A two-class fit according to the theory of multiple equilibria (equation I. 5) matches the theoretical curve quite well (- - -). In the lower figure, the theoretical curve has been truncated to produce an incomplete isotherm, such as those observed experimentally (Figures 32 and 33). Although a two-class fit matches the incomplete isotherm quite well, its match of the missing portion of the curve is strikingly inaccurate.

no relationship whatsoever to the number of ligands actually bound in the second phase.[1] A class which consists of a large number of members, representing the binding of ligands in the second phase, has been postulated by several workers from analyses of such incomplete isotherms (Teresi and Luck, 1952; Goodman, 1958; Spector et al., 1969) and is probably a very misleading estimate of the actual number of ligands which could bind in the second phase. Therefore, the only physically significant information which can be obtained from a class analysis of incomplete biphasic isotherms is the number of ionic sites, and this quantity could, in many cases, be obtained simply from the position of the intermediate plateau.

---

[1]Since the analysis of the first phase of an incomplete isotherm retains its above-mentioned correspondence to the number of ionic sites in the binding region, the use of the k_{int} values for the high-energy class obtained by such analyses to calculate experimental $\Delta F\rho$ values for Figure 34 is justified.

APPENDIX A

AN ALTERNATIVE MODEL[1]

As an alternative model for the binding of flexible ligands to proteins, the protein surface may be regarded as a continuum, so that the adsorption of ligand segments no longer is restricted to the binding sites provided by lattice points. Thus, the protein is visualized as a sticky sphere. To be consistent with this continuous nature of the binding surface, the possible arrangements of the ligand segments are less restricted as well.

Statistical-Mechanical Treatment

The basic statistical-mechanical formulae of the sticky sphere model are identical to equations III.1 to III.6, except for the replacement of the extensive variable n_0 by A, the area of the surface

$$A = n_0 \pi r_s^2 \qquad (A.1)$$

where r_s is the radius of the ligand segment. Furthermore, the state of adsorption, S , may be retained as the basis for classifying ligand configurations (equations III.7 to III.9). However, a new set of statistical weights, t(S) , is needed to calculate the canonical partition functions for the individual ligands (according to equation III.10).

---

[1]The author would like to thank Dr. E.G.D. Cohen for suggesting the approach used in this Appendix.

The Totally Adsorbed Ligand The first segment of the ligand N=1 may be placed anywhere on the surface. Hence, its contribution to t(0,x,0) is[1]

$$a_s^{(1)} = \left(\frac{2\pi m_L kT}{h^2}\right) A f \qquad (A.2)$$

where m_L is the mass of the ligand and f is the Boltzmann factor for the interaction between the surface and the first segment. If each segment is a sphere of radius r_s, the center of the second segment may be placed anywhere along the equator of a sphere of radius $2r_s$ whose center coincides with the center of the first segment (Figure 37). The center of each remaining segment has only two-thirds of this equatorial arc available, due to the presence of the previous ligand segment (Figure 37). In addition, each segment is allowed to move away from the equator by a small angle $\pm\psi$, which would correspond to a vibrational motion (the value of ψ is discussed below). Thus, the contribution of the second segment to t(0,x,0) is

$$w_s = \left(\frac{2\pi m_s kT}{h^2}\right) 2\pi (2r_s)(2r_s 2\psi) f$$

$$= \left(\frac{2\pi m_s kT}{h^2}\right) 16\pi r_s^2 \psi f \qquad (A.3)$$

while that for each remaining segment is

$$w_s = \left(\frac{2\pi m_s kT}{h^2}\right) \frac{2}{3} 16\pi r_s^2 \psi f \qquad (A.4)$$

---

[1]As in the lattice model, $a^{(1)}$ will represent the contribution of the first segment to the partition function and w will represent the contribution of each remaining segment (both are scalars in the present discussion). Subscripts s and d will be used to represent adsorbed and desorbed segments, respectively, as in the lattice model.

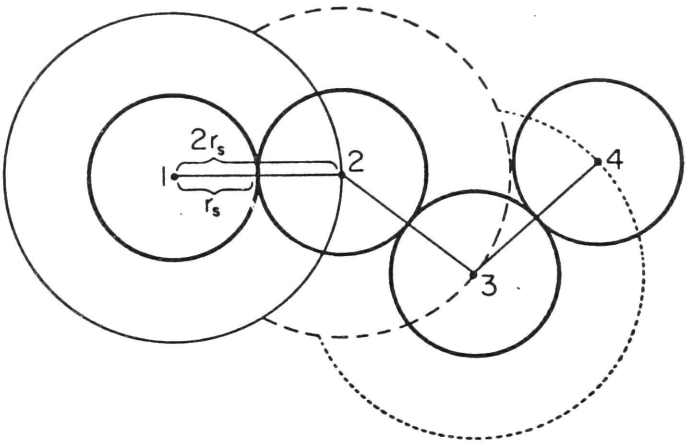


Figure 37. The placement of successive adsorbed ligand segments in the sticky sphere model. Once the first segment (1) has been placed on the surface, the center of the second segment (2) may be placed anywhere along the equator of a sphere of radius $2r_s$ whose center coincides with the center of the first segment (——). The center of the third segment (3) may be placed anywhere along the equator of a sphere of radius $2r_s$ whose center coincides with the center of the second segment, excluding that portion of the equator occupied by the first segment (- - -). Similarly, the placement of the center of the fourth segment (4) is along the equator of a sphere of radius $2r_s$ whose center coincides with the center of the third segment, excluding that portion occupied by the second segment (• • •).

where m_s is the mass of a ligand segment and f is calculated using the appropriate energy of interaction.[1] Using the above contributions, t(0,x,0) has the form

$$t(0,x,0) = a_s^{(1)} w_s^{x-1} \qquad (A.5)$$

In contrast to the lattice model, t(0,x,0) depends on the mass of the ligand and its segments and contains a temperature dependence in addition to that included in the Boltzmann factor. This is because the vibration of the ligand segments on the lattice sites has been neglected.

General Equations for t(S) The contribution of a desorbed first segment is

$$a_d^{(1)} = \left(\frac{2\pi m_L kT}{h^2}\right) A \qquad (A.6)$$

The second segment of a free end and the first and second segments of a loop may be placed anywhere on the surface of a sphere of radius $2r_s$, while the available area is reduced by a factor of one-quarter for the third and subsequent segments of a loop or free end (corresponding to the surface area of a spherical segment of height r_s). Thus

$$w_d = \left(\frac{2\pi m_s kT}{h^2}\right) 16\pi r_s^2 \qquad (A.7)$$

for the second segment of a free end and the first and second segments of a loop and

$$w_d = \left(\frac{2\pi m_s kT}{h^2}\right) 12\pi r_s^2 \qquad (A.8)$$

---

[1] Equations A.3 and A.4 give a value of 0.667 for the immediate reversal correction factor, in contrast to the value 0.833 for the hexa-coordinated sites in the lattice model (equation III.19).

for the third and subsequent segments.[1] However, the contribution of a loop of d segments cannot be equated to

$$w_d^d \qquad (A.9)$$

since there would be no way of assuring that the dth segment of the loop is located such that the (d+1)th segment can be placed on the surface and that all segments of the loop remain above the surface. In fact, with the continuum model there is no simple way of calculating the exact contribution to the partition function of a loop which includes these restrictions. In the lattice model, the ratio of the number of vertical configurations of a loop of d segments (equation III.32) to the total number of vertical configurations available to d segments, 3^d, gives the fraction of the total number of vertical configurations of d segments which meet the requirements for loop formation. It is assumed that this ratio gives a reasonable estimate of the fraction of the configurations included in equation A.9 which fulfill the above requirements. Similarly, the ratio of the number of vertical configurations of a free end of d_0 segments in the lattice model (equation III.35) to the total number of vertical configurations of d_0 segments, 3^{d_0}, is used as an estimate of the fraction of allowed configurations for a free end.

The value of ψ (equations A.3 and A.4) was chosen arbitrarily such that

$$\frac{w_s}{w_d} = \frac{f}{3} \qquad (A.10)$$

---

[1] Equations A.7 and A.8 give an immediate reversal correction factor of 0.750 for desorbed segments, in contrast to 0.944 for desorbed segments on hexa-coordinated sites in the lattice model (equation III.23).

to be comparable to the ratio of the coordination numbers of the surface and solution points in the lattice model. This gives a value of 19.1° for ψ.

With the above contributions, one may write

$$t(S) = a_d^{(1)} w_d^{d_0-1} \frac{H_0(d_0)}{3^{d_0}} \prod_{\rho=1}^{\ell-1} \left(w_s^{s_\rho} w_d^{d_\rho} \frac{H(d_\rho)}{3^{d_\rho}} \right) w_s^{s_\ell} w_d^{d_0'} \frac{H_0(d_0')}{3^{d_0'}} \qquad (A.11)$$

for the chain with free ends and

$$t(S) = a_s^{(1)} w_s^{-1} \prod_{\rho=1}^{\ell-1} \left(w_s^{s_\rho} w_d^{d_\rho} \frac{H(d_\rho)}{3^{d_\rho}} \right) w_s^{s_\ell} \qquad (A.12)$$

when $d_0=d_0'=0$. Combining these two equations into a more general one

$$t(S) = \left(a_s^{(1)} w_s^{-1} \delta_{d_0 0} + (1-\delta_{d_0 0}) a_d^{(1)} w_d^{d_0-1} \frac{H_0(d_0)}{3^{d_0}} \right) \times \prod_{\rho=1}^{\ell-1} \left(w_s^{s_\rho} w_d^{d_\rho} \frac{H(d_\rho)}{3^{d_\rho}} \right) w_s^{s_\ell} \left(\delta_{d_0' 0} + (1-\delta_{d_0' 0}) w_d^{d_0'} \frac{H_0(d_0')}{3^{d_0'}} \right) \qquad (A.13)$$

This general expression for t(S) may be used in the evaluation of $q_x(1,A,T)$ and its associated thermodynamic averages following the procedures described in conjunction with equations III.42 to III.51.

The Binding of Additional Ligands For ligands bound subsequently to the first, the expressions for $a_s^{(1)}$ and w_s must be modified since a smaller area of the surface is available for binding due to the presence of ligands previously bound. As in the lattice model, no modifications are introduced in the terms describing the desorbed segments, so that $a_d^{(1)}$ and w_d retain the forms presented above (equations A.6 to A.8). If $\bar{n}(N-1)$ is the average number of segments adsorbed to the surface, in ligands one through N-1, then (cf. equation III.54)

$$_{N-1}a_s^{(1)} = \left(\frac{2\pi m_L kT}{h^2}\right)\left(A - \bar{n}(N-1)\pi r_s^2\right) f \qquad (A.14)$$

and, as an approximation which assumes that the occupied areas are distributed uniformly over the surface (cf. equation III.59)

$$_{N-1}w_s = w_s \left(\frac{A - \bar{n}(N-1)\pi r_s^2}{A}\right) \qquad (A.15)$$

The quantity $\bar{n}(N-1)$ is calculated from the configurational parameter $\bar{\theta}$, the average fraction of ligand segments adsorbed, for ligands one through N-1. If $\bar{\theta}(i)$ is the average fraction of segments adsorbed for the ith ligand, then

$$\bar{n}(N-1) = \sum_{i=1}^{N-1} x\bar{\theta}(i) \qquad (A.16)$$

The Unbound Ligand The partition function for the unbound ligand is (cf. equation III.65)

$$q_{x,u}(V) = \left(\frac{2\pi m_L kT}{h^2}\right)^{3/2} V w_d^{x-1} \qquad (A.17)$$

To correct for immediate reversals, equation A.8 must be used for segments subsequent to the second. Equation III.68 may be used to calculate μ_0, the standard chemical potential of the ligand.

Results and Discussion

Before presenting some of the results calculated using the sticky sphere model, it is important to mention some of its limitations, especially in regard to its applicability to protein-ligand interactions. First of all, the model is designed for systems in which all parts of the surface are equivalent (homogeneous systems). If patches differing in their chemical characteristics were placed on the surface, it would

be impossible to specify what types of areas were available to a segment once the previous segment had been bound, since there is no way of calculating the occupancies of the various regions. Secondly, there is no way of introducing edge effects, so that a sticky patch, analogous to the single face (Figure 6), cannot be treated. Furthermore, a lattice probably is a better representation of a protein surface than the continuum of the sticky sphere model, since ligand segments interact with discreet groups of atoms on proteins. Also, ligand segments do not have the configurational freedom implied by the continuum model due to bond angle restrictions.[1]

The practical advantage of the sticky sphere model is its simplicity. In particular, due to the absence of matrices the calculations are less involved than those of the lattice model.

As shown in Figure 38 (cf. Figure 18), the isotherms of the sticky sphere model for homogeneous systems are quite similar to those of the lattice model. However, the former are shifted by about one and one-half units in the positive direction on the log c axis, because the ratio of the partition function for the unbound ligand to that for the bound ligands is larger in the sticky sphere model.

Fortunately, it also is possible to compare the sticky sphere model with a specific case of the lattice model for heterogeneous systems. One can consider the binding of carboxylic acids (or other substituted alkanes)

---

[1]As mentioned in Chapter V, the hexa-coordinated lattice of the lattice model is a fairly realistic one for chains of hydrocarbon segments with the statistical segment equated to two chemical segments.

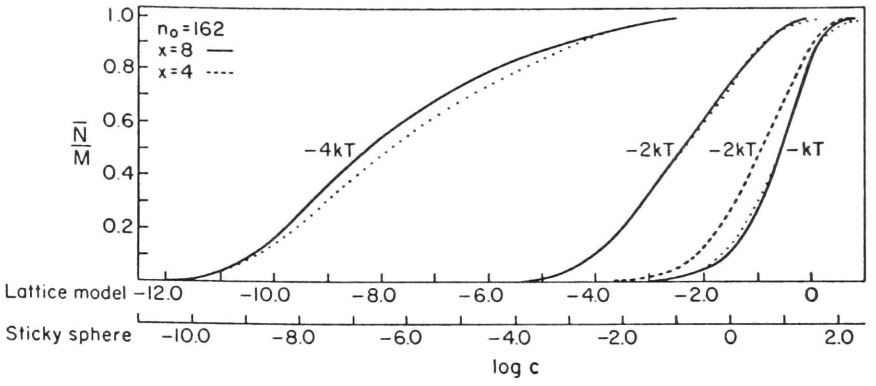


Figure 38. Comparison of isotherms for homogeneous systems calculated by the sticky sphere model with those obtained using the lattice model (Figure 18). The isotherms of the sticky sphere model (• • •) can be superimposed almost exactly on those obtained with the lattice model by shifting the former one and one-half units in the negative direction along the log c axis, as indicated. The sticky sphere isotherm at ϵ=-2kT for the ligand x=4 coincides too closely with that calculated using the lattice model to be distinguished from the latter.

to a protein with isolated areas of an ionic character distributed over a nonpolar surface, so that a comparison with the heterogeneous systems with two types of lattice sites on a complete surface is possible (e.g. Figure 29). The areas of an ionic character can bind only the carboxyl segment of the ligand, while the rest of the surface can bind only hydrocarbon segments. Furthermore, once the carboxyl segment is bound to an ionic area, it is assumed that a hydrocarbon segment will not be placed next to another ionic area, which is a good assumption only if the ligand is shorter than the distance between neighboring ionic areas. As shown in Figure 39, similar results again are obtained with the two models, with the curves of the sticky sphere model shifted by one and one-half units in the positive direction along the log c axis. Of course, with the sticky sphere model, different arrangements of the ionic areas cannot be tested, nor can additional regions with different chemical characteristics (e.g. polar regions) be introduced. However, it has been shown that the general features of the isotherms for heterogeneous systems are not affected significantly by the arrangement of the ionic sites or the introduction of a small number of polar sites. Thus, these two limitations of the sticky sphere model appear relatively unimportant for a discussion of the binding of carboxylic acids or other substituted alkanes to proteins, but this could not have been known without the more detailed results possible with the lattice model. A more serious limitation of the sticky sphere model is that a sticky patch cannot be treated, since binding in actual protein-ligand systems may be restricted to a single region or regions of the surface.

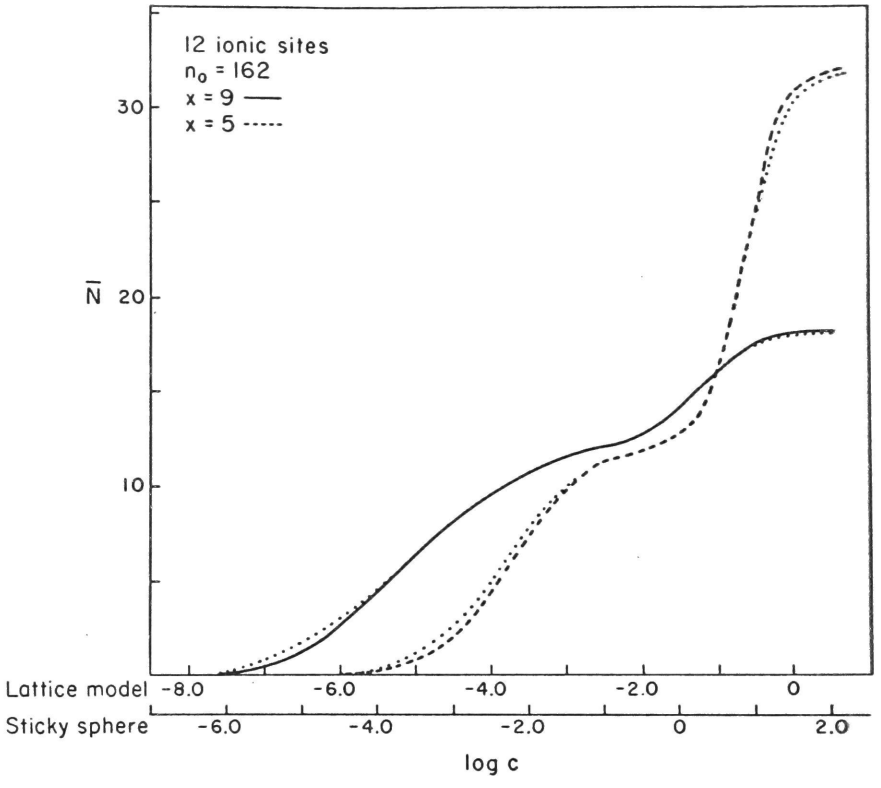


Figure 39. Comparison of isotherms for a specific heterogeneous system calculated by the sticky sphere model with those obtained using the lattice model (Figure 29). The isotherms of the sticky sphere model (• • •) can be superimposed almost exactly on those obtained with the lattice model by shifting the former one and one-half units in the negative direction along the log c axis, as indicated.

APPENDIX B

EXCLUDED VOLUME CORRECTIONS

Theoretical Considerations

In an exact count of the configurations of a ligand on a lattice, no lattice site can be occupied by more than one ligand segment. This Appendix describes a method which provides a more detailed correction for excluded volume than the average immediate reversal correction factors described in Chapter III (equations III.19 and III.23). The subject of excluded volume has received much attention in theoretical studies of polymer configuration (reviewed by Domb, 1969). However, the mathematical problems involved in any such treatment are formidable, and no general analytical methods have been found for the enumeration of polymer configurations on a lattice with a complete correction for excluded volume. Thus, the corrections described in this Appendix apply only to the totally adsorbed ligand ($S=\{0,x,0\}$). Although the mathematics are simplified greatly by considering this two-dimensional case, physically the excluded volume problem is a more serious one in two dimensions.

The first correction necessary is the elimination of immediate reversals of the ligand. For example, when the product

$$w_{ij}w_{ji} \qquad (B.1)$$

appears in any term in $\nu(x)$ (equation III.16), a correction must be applied

since some of the ligand configurations which go from a class i site to a class j site and back to a class i site will produce an immediate reversal.[1] A second-order correction eliminates the triangles formed when products such as

$$w_{ij}w_{jk}w_{ki} \tag{B.2}$$

appear in $\nu(x)$, while the third-order correction eliminates diamonds[2]

$$w_{ij}w_{jk}w_{kj}w_{ji} \tag{B.3}$$

Such corrections could be extended to any order, the nth-order correction eliminating those configurations which form an (n+1)-sided polygon on the lattice with no prior reversals or intersections. This approach resembles calculations by other workers of the number of polygonal closures of n steps on a variety of lattices, using both exact enumeration (Fisher and Sykes, 1959) and statistical, or Monte Carlo methods (reviewed by Gans, 1965).

An outline of the method used for the first-order corrections now will be described, followed by a discussion of some of the results.[3] The necessary corrections may be accomplished most simply if, after the kth segment of the ligand has been placed, a correction matrix $\underline{C}^{(k)}$ is subtracted before the configurations of the (k+1)th segment are gener-

---

[1]The necessary correction must be more precise than the average immediate reversal correction factor given by equation III.19, which depends on j alone.

[2]The immediate reversals produced by $w_{jk}w_{kj}$ in equation B.3 already are eliminated in the first-order correction.

[3]Limitations on available computer facilities have prevented an extension of the calculations to include higher-order corrections.

ated, so that

$$\nu_C(x) = \underline{a}^{(1)}\Big(\Big(\cdots\Big(\underline{W}^2 - \underline{C}^{(3)}\Big)\underline{W} - \underline{C}^{(4)}\Big)\underline{W} - \underline{C}^{(5)}\Big)\underline{W} \cdots\cdots$$
$$\cdots\cdots - \underline{C}^{(x-1)}\Big)\underline{W} - \underline{C}^{(x)}\Big)\underline{e}$$
$$= \underline{a}^{(1)}\left(\underline{W}^{x-1} - \sum_{i=0}^{x-3} \underline{C}^{(i+3)}\,\underline{W}^{x-3-i}\right)\underline{e} \qquad \text{(B.4)}$$

where the subscript C indicates that $\nu(x)$ has been corrected for immediate reversals. A computer program has been developed which generates the correction matrices $\underline{C}^{(k)}$. The elements of $\underline{C}^{(k)}$, $c_{ij}^{(k)}$, represent the number of configurations excluded by immediate reversals due to the placement of the kth segment when the first segment of the ligand is on a class i site and the kth segment is on a class j site. For example, if the first segment of the ligand is on a class i site, the third segment can produce an immediate reversal only by returning to a class i site, and there are z_i such excluded configurations for a given site. Hence

$$\underline{C}^{(3)} = \begin{pmatrix} 5 & 0 & 0 & 0 & \cdots\cdots \\ 0 & 6 & 0 & 0 & \cdots\cdots \\ 0 & 0 & 6 & 0 & \cdots\cdots \\ & & \cdots\cdots & & \\ & & \cdots\cdots & & \end{pmatrix} \qquad \text{(B.5)}$$

As a further example, if the first segment of the ligand is on a class i site and immediate reversals for segments one through three already have been eliminated, the number of immediate reversals produced by the placement of the fourth segment on a class j site is

$$c_{ij}^{(4)} = w_{ij}(z_j - 1) \qquad \text{(B.6)}$$

Results and Discussion

Wall et al. (1954, 1955), in their Monte Carlo studies of polymer configurations with corrections for excluded volume, found that after several segments of a polymer had been placed on a lattice, the fraction of the total number of configurations generated which had no multiple-site occupancies decreased with the addition of polymer segments according to a simple exponential law

$$\phi_x = e^{-\lambda x} \tag{B.7}$$

where ϕ_x represents this fraction for a polymer of x segments and λ is the so-called attrition coefficient. Equation B.7 was found to apply to a variety of two- and three-dimensional lattices, and values of λ were determined from the slopes of $\log \phi_x$ vs. x graphs (Wall et al., 1954, 1955). This exponential law is equivalent to the statement that the ratio

$$\mu = \frac{\nu_C(x)}{\nu_C(x-1)} \tag{B.8}$$

tends to a constant value as x approaches infinity (Fisher and Sykes, 1959). This ratio is related to the attrition coefficient by

$$\mu = e^{-\lambda(z-1)} \tag{B.9}$$

where z is the coordination number of the lattice. Hammersley and Morton (1954) have given a mathematical justification for the existence of such a limiting ratio. The values of μ for a hexa-coordinated lattice obtained by the Monte Carlo methods of Wall et al. (1955) and the exact enumeration procedure of Fisher and Sykes (1959) are compared with the corresponding results obtained using the excluded volume corrections described in this Appendix in Table IV. In the latter calculations, μ was found to

reach the constant value in the Table after only a few segments of the ligand were placed on the lattice. Even though the excluded volume calculations discussed in this Appendix only eliminate immediate reversals, it is clear that a serious overestimate of the number of allowed configurations is not obtained for a short ligand. Thus, the first-order correction for excluded volume, the elimination of immediate reversals, must be the most important one. The values of μ presented here for the exact treatment of immediate reversals described in this Appendix are almost identical to those obtained using the average immediate reversal correction factors given by equation III.19, and each approaches the value 5.0 characteristic of a hexa-coordinated lattice as n_0 is increased. Furthermore, the values of μ obtained by Wall et al. (1954, 1955) and Fisher and Sykes (1959) for a variety of lattices indicate that the discrepancy between the first-order correction and the complete correction is even smaller in three dimensions than in two dimensions. Therefore, the use of only an average immediate reversal correction factor to correct for excluded volume in the calculations of the present model is not a serious deficiency.

Table IV

Values of μ

Exact Enumeration[1]	4.152
Monte Carlo[2]	4.20
Present Model:	
n_0=92	4.884
n_0=162	4.936
n_0=252	4.960
n_0=362	4.972

[1]From Fisher and Sykes (1959).

[2]From Wall et al. (1955).

APPENDIX C

FACTORIZATION OF $q_x(1,n_0,T)$ IN HOMOGENEOUS SYSTEMS

Equation III.36 for t(S) may be simplified considerably in homogeneous systems in which the Boltzmann factors are identical for all vector and matrix elements. If the immediate reversal correction factors for all sites are assumed to be those characteristic of the z_1 sites, then

$$\underline{a}_s^{(1)} = \underline{a}^{(1)} f \qquad \underline{a}_d^{(1)} = \underline{a}^{(1)} \tag{C.1}$$

for adsorbed and desorbed first segments

$$\underline{W}_s = \underline{W} f \qquad \underline{W}_d = \underline{W} \tag{C.2}$$

for subsequent segments which require no immediate reversal correction factors

$$\underline{W}_s = \underline{W} f g_s \qquad \underline{W}_d = \underline{W} g_d \tag{C.3}$$

for the remaining segments which require these factors, where g_s and g_d are the immediate reversal correction factors for adsorbed and desorbed segments, respectively, given by equations III.19 and III.23 with $z_1=6$. Thus, t(S) assumes the form[1]

$$t(S) = \underline{a}^{(1)} \underline{W}^{x-1} \underline{e} \left(\delta_{d_0 0} + (1-\delta_{d_0 0})\, g_d^{-2}\, g_d^{d_0} H_0(d_0)\right) \prod_{\rho=1}^{\ell-1} \left(g_s^{-2}\, g_s^{s_\rho}\, f^{s_\rho}\, g_d^{d_\rho - 2}\, H(d_\rho)\right)$$
$$\times\ g_s^{-2}\, g_s^{s_\ell}\, f^{s_\ell} \left(\delta_{d_0' 0} + (1-\delta_{d_0' 0})\, g_d^{-2}\, g_d^{d_0'}\, H_0(d_0')\right) \tag{C.4}$$

---

[1] Rigorously, equation C.4 cannot be used for all S. For example, when $d_0=1$, no immediate reversal correction factor should be used, rather than g_d^{-1} as implied by that equation. For the calculations in this Appendix, however, this discrepancy is not of sufficient magnitude to warrant the more exact treatment used in the calculations described in Chapter III.

and

$$q_x(1,n_0,T) = \underline{a}^{(1)}\underline{W}^{x-1}\underline{e} \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} u_0(d_0) \prod_{\rho=1}^{\ell-1}\Big(v(s_\rho)\,u(d_\rho)\Big)\, v(s_\ell)\, u_0(d_0') \tag{C.5}$$

with[1]

$$u_0(d_0) = \delta_{d_0 0} + (1-\delta_{d_0 0})\, g_d^{-2} g_d^{d_0} H_0(d_0) \tag{C.6}$$

$$v(s_\rho) = g_s^{-2} g_s^{s_\rho} f^{s_\rho} \tag{C.7}$$

$$u(d_\rho) = g_d^{d_\rho - 2} H(d_\rho) \tag{C.8}$$

$$u_0(d_0') = \delta_{d_0' 0} + (1-\delta_{d_0' 0})\, g_d^{-2} g_d^{d_0'} H_0(d_0') \tag{C.9}$$

Thus, $q_x(1,n_0,T)$ can be written as a product of the horizontal and vertical contributions[2] $q_{x,H}$ and $q_{x,V}$

$$q_x(1,n_0,T) = q_{x,H}\, q_{x,V} \tag{C.10}$$

where

$$q_{x,H} = \underline{a}^{(1)}\underline{W}^{x-1}\underline{e} = \nu(x) \tag{C.11}$$

and

$$q_{x,V} = \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} u_0(d_0) \prod_{\rho=1}^{\ell-1}\Big(v(s_\rho)\,u(d_\rho)\Big)\, v(s_\ell)\, u_0(d_0') \tag{C.12}$$

With the partition function so simplified, the present model can be generalized to consider systems similar to those found in more typical models of polymer-surface interactions (Silberberg, 1962; Roe, 1965a, 1965b). In particular, the vertical partition function can be evaluated for infinite or very long chains, so that some of the limiting proper-

---

[1]The immediate reversal correction factors are written as indicated in equations C.4 to C.9 to facilitate the calculation of average properties (e.g. equations C.49 and C.73 to C.74).

[2]For convenience, the immediate reversal correction factors are included in the vertical contribution to the partition function, $q_{x,V}$.

ties of the model can be considered.

Horizontal Partition Function

Through a study of the horizontal partition function, one may evaluate the effects of the irregularities in the triangular lattice (z_2 sites) on the number of configurations available to the ligand. A useful reference model is the Flory-Huggins theory of polymer configurations on a liquid lattice (Flory, 1953), according to which the number of ways of placing a chain of x segments on a uniform triangular lattice of n_0 sites is[1]

$$\nu_F(x) = n_0 z_1^{x-1} \qquad (C.13)$$

where the subscript F denotes the Flory model.[2] If 12 randomly distributed z_2 sites are included among the n_0 sites of the Flory lattice, equation C.13 becomes

$$\nu_{F,p}(x) = n_0 \left(z_1 p + z_2 (1-p) \right)^{x-1} \qquad (C.14)$$

where p is the fraction of z_1 sites on the lattice

$$p = \frac{n_0 - 12}{n_0} \qquad (C.15)$$

Table V includes a comparison of $\nu(x)$ $(=q_{x,H})$, $\nu_F(x)$, and $\nu_{F,p}(x)$ for fixed values of x and n_0. $\nu_F(x)$ is larger than both $\nu(x)$ and $\nu_{F,p}(x)$, since it is increased by a factor of $z_1=6$ with the placement of each seg-

---

[1]The original form of equation C.13 is $n_0 z_1 (z_1-1)^{x-2}$ to correct for immediate reversals. However, since the immediate reversal correction factors in the present model have been included in the vertical partition function, the original Flory equation has been written without such a correction.

[2]Equation C.13 strictly is suitable only for very large lattices on which edge effects can be neglected.

Table V

Effect of z_2 Sites on Ligand Configurations

n_0	p	x	$\nu(x)$	$\nu_F(x)$	$\nu_{F,p}(x)$
92	0.869	5	1.10×10^5	1.19×10^5	1.09×10^5
		10	7.80×10^8	9.27×10^8	7.61×10^8
		20	3.90×10^{16}	5.61×10^{16}	3.69×10^{16}
		30	1.95×10^{24}	3.39×10^{24}	1.75×10^{24}
162	0.926	5	2.01×10^5	2.10×10^5	2.00×10^5
		10	1.48×10^9	1.63×10^9	1.46×10^9
		20	8.09×10^{16}	9.87×10^{16}	7.79×10^{16}
		30	4.41×10^{24}	5.70×10^{24}	4.16×10^{24}
252	0.952	5	3.18×10^5	3.27×10^5	3.16×10^5
		10	2.39×10^9	2.54×10^9	2.36×10^9
		20	1.35×10^{17}	1.54×10^{17}	1.32×10^{17}
		30	7.67×10^{24}	9.28×10^{24}	7.37×10^{24}
362	0.967	5	4.60×10^5	4.69×10^5	4.59×10^5
		10	3.50×10^9	3.65×10^9	3.47×10^9
		20	2.02×10^{17}	2.21×10^{17}	1.99×10^{17}
		30	1.17×10^{25}	1.33×10^{25}	1.14×10^{25}

ment. The randomly placed z_2 sites on the modified Flory lattice reduce the rate of increase of $\nu_{F,p}(x)$ to 5+p, while the lattice of the present model represents an intermediate case, since $\nu(x)$ is increased by a factor of $z_1=6$ ($= \sum_{j=1}^{C} w_{ij}$ for i= 2, 3,.... C) for the placement of all segments except those which follow a z_2 site and therefore contribute w_{12} to the terms of $\nu(x)$. The differences between $\nu(x)$, $\nu_F(x)$, and $\nu_{F,p}(x)$ are magnified as x increases due to the exponential dependence of these quantities on the coordination number of the lattice. As n_0 increases and p approaches unity, the presence of the z_2 sites is less evi-

dent and a decrease in the difference between these three quantities is seen.

Vertical Partition Function

Three methods of evaluating the vertical partition function will be discussed in this section. Although the first two apply only to infinite or very long chains, the approximations which may be introduced in these limiting cases result in closed expressions for $q_{x,V}$. The first is the method of sequence-generating functions (Lifson, 1964), which utilizes the assumption that the special properties of the ends of the ligand may be neglected for infinite chains (i.e. all desorbed segments may be treated as loops). The second, the method of steepest descents (Darwin and Fowler, 1922; Schrödinger, 1964), can be used to evaluate $q_{x,V}$ for x large but finite, when the effects of the chain ends cannot be neglected. The third method is applicable only to ligands with fewer than about 30 segments, for which an exact evaluation of $q_{x,V}$ is possible using the convolution procedure described in Chapter III (equations III.42 to III.45). By performing calculations with this exact method, the effects of the approximations used in the other procedures can be evaluated.

In all three methods, one begins by forming the generating function for $q_{x,V}$

$$\Gamma(\zeta) = \sum_{x=1}^{\infty} q_{x,V} \zeta^x$$

$$= \sum_{x=1}^{\infty} \zeta^x \sum_{\ell=1}^{[\frac{x+1}{2}]} \sum_{\{S\}} u_0(d_0) \prod_{\rho=1}^{\ell-1} \left(v(s_\rho)\, u(d_\rho) \right) v(s_\ell)\, u_0(d_0') \qquad (C.16)$$

Rearranging the summation indices (cf. equations III.43 to III.44)

$$\Gamma(\zeta) = \sum_{\ell=1}^{\infty} \sum_{d_0=0}^{\infty} u_0(d_0)\zeta^{d_0} \prod_{\rho=1}^{\ell-1} \left(\sum_{s_\rho=1}^{\infty} v(s_\rho)\zeta^{s_\rho} \sum_{d_\rho=1}^{\infty} u(d_\rho)\zeta^{d_\rho} \right)$$
$$\times \sum_{s_\ell=1}^{\infty} v(s_\ell)\zeta^{s_\ell} \sum_{d_0'=0}^{\infty} u_0(d_0')\zeta^{d_0'} \qquad (C.17)$$

The method of sequence-generating functions and the steepest descents procedure also utilize the generating functions

$$V(\zeta) = \sum_{s_\rho=1}^{\infty} v(s_\rho)\zeta^{s_\rho} = \sum_{s_\rho=1}^{\infty} g_s^{-2} g_s^{s_\rho} f^{s_\rho} \zeta^{s_\rho} \qquad (C.18)$$

$$U(\zeta) = \sum_{d_\rho=1}^{\infty} u(d_\rho)\zeta^{d_\rho} = \sum_{d_\rho=1}^{\infty} g_d^{d_\rho - 2} H(d_\rho)\zeta^{d_\rho} \qquad (C.19)$$

$$U_0(\zeta) = \sum_{d_0=0}^{\infty} u_0(d_0)\zeta^{d_0} = \sum_{d_0'=0}^{\infty} u_0(d_0')\zeta^{d_0'}$$
$$= 1 + \sum_{d_0=1}^{\infty} g_d^{-2} g_d^{d_0} H_0(d_0)\zeta^{d_0} \qquad (C.20)$$

so that equation C.17 may be written as

$$\Gamma(\zeta) = \sum_{\ell=1}^{\infty} U_0(\zeta) \prod_{\rho=1}^{\ell-1} \Big(V(\zeta)\, U(\zeta) \Big)\, V(\zeta)\, U_0(\zeta)$$
$$= \frac{U_0^2(\zeta)\, V(\zeta)}{1 - V(\zeta)\, U(\zeta)} \qquad (C.21)$$

Closed expressions for equations C.18 to C.20 are necessary for the two methods which use these generating functions. Thus, for $V(\zeta)$

$$V(\zeta) = \sum_{s_\rho=1}^{\infty} g_s^{-2} g_s^{s_\rho} f^{s_\rho} \zeta^{s_\rho} = g_s^{-2} \frac{f g_s \zeta}{1 - f g_s \zeta} \qquad (C.22)$$

Using the generating function for H(d)[1]

$$\sum_{d-1=0}^{\infty} H(d)\, \lambda^{d-1} = \frac{1}{2\lambda^2} \left(1 - \lambda - \sqrt{(1 - 2\lambda - 3\lambda^2)} \right) \qquad (C.23)$$

$U(\zeta)$ assumes the form

---

[1]This expression was brought to the author's attention by Dr. John Riordan, who also provided valuable assistance in the derivation of equation C.36.

$$U(\zeta) = \sum_{d_\rho=1}^{\infty} g_d^{d_\rho-2} H(d_\rho)\zeta^{d_\rho} = \zeta g_d^{-1} \sum_{d_\rho-1=0}^{\infty} H(d_\rho)\,(g_d\zeta)^{d_\rho-1}$$

$$= \frac{1 - g_d\zeta - \sqrt{(1-2g_d\zeta-3g_d^2\zeta^2)}}{2g_d^3\zeta} \qquad (C.24)$$

A generating function for $H_0(d_0)$ is required to write a closed expression for $U_0(\zeta)$. Beginning with equation III.35

$$H_0(d_0) = \sum_{i=0}^{d_0-1} \frac{(d_0-1)!}{[\frac{i}{2}]!\,[\frac{i+1}{2}]!\,(d_0-1-i)!} = \sum_{i=0}^{d_0-1} \frac{(d_0-1)!}{(d_0-1-i)!\,i!} \frac{i!}{[\frac{i}{2}]!\,[\frac{i+1}{2}]!} \qquad (C.25)$$

or, if the common notation for binomial coefficients is introduced

$$H_0(d_0) = \sum_{i=0}^{d_0-1} \binom{d_0-1}{i} h_i \qquad (C.26)$$

where

$$h_i = \frac{i!}{[\frac{i}{2}]!\,[\frac{i+1}{2}]!} \qquad (C.27)$$

The terms of equation C.26 may be grouped into even and odd parts

$$H_0(d_0) = H_0^e(d_0) + H_0^o(d_0)$$

$$= \sum_{i=0}^{[\frac{d_0-1}{2}]} \binom{d_0-1}{2i} h_{2i} + \sum_{i=0}^{[\frac{d_0-2}{2}]} \binom{d_0-1}{2i+1} h_{2i+1} \qquad (C.28)$$

where the superscripts e and o denote even and odd, respectively.

Splitting the binomial coefficients[1]

$$H_0(d_0) = \sum_{i=0}^{[\frac{d_0-1}{2}]} \left(\binom{d_0-2}{2i} + \binom{d_0-2}{2i-1} \right) h_{2i} + \sum_{i=0}^{[\frac{d_0-2}{2}]} \left(\binom{d_0-2}{2i+1} + \binom{d_0-2}{2i} \right) h_{2i+1}$$

$$= H_0^e(d_0-1) + \sum_{i=0}^{[\frac{d_0-1}{2}]} \binom{d_0-2}{2i-1} h_{2i} + H_0^o(d_0-1) + \sum_{i=0}^{[\frac{d_0-2}{2}]} \binom{d_0-2}{2i} h_{2i+1} \qquad (C.29)$$

---

[1] It is possible to substitute $H_0^e(d_0-1)$ and $H_0^o(d_0-1)$ into the final line of equation C.29 since a binomial coefficient $\binom{a}{b}$ is zero by convention when b is a negative integer (Moran, 1968).

The index of the first summation in equation C.29 may be changed from i to j=i-1

$$H_0(d_0) = H_0^e(d_0-1) + \sum_{j=0}^{[\frac{d_0-1}{2}]-1} \binom{d_0-2}{2j+1} h_{2j+2} + H_0^o(d_0-1) + \sum_{i=0}^{[\frac{d_0-2}{2}]} \binom{d_0-2}{2i} h_{2i+1} \tag{C.30}$$

With the substitutions

$$h_{2j+2} = 2h_{2j+1} \tag{C.31}$$

and

$$h_{2i+1} = \frac{2i+1}{i+1} h_{2i} = (2 - \frac{1}{i+1}) h_{2i} \tag{C.32}$$

equation C.30 becomes

$$H_0(d_0) = H_0^e(d_0-1) + H_0^o(d_0-1) + 2 \sum_{i=0}^{[\frac{d_0-1}{2}]-1} \binom{d_0-2}{2i+1} h_{2i+1} + \sum_{i=0}^{[\frac{d_0-2}{2}]} \binom{d_0-2}{2i} (2 - \frac{1}{i+1}) h_{2i} \tag{C.33}$$

where j has been replaced by i. Upon further simplification

$$H_0(d_0) = 3\left(H_0^e(d_0-1) + H_0^o(d_0-1)\right) - \sum_{i=0}^{[\frac{d_0-2}{2}]} \binom{d_0-2}{2i} \frac{h_{2i}}{i+1}$$

$$= 3\, H_0(d_0-1) - \sum_{i=0}^{[\frac{d_0-2}{2}]} \binom{d_0-2}{2i} \frac{h_{2i}}{i+1} \tag{C.34}$$

The remaining summation is $H(d_0-1)$, since (equation III.32)

$$H(d) = \sum_{i=0}^{[\frac{d-1}{2}]} \frac{1}{i+1} \frac{(d-1)!}{i!\, i!\, (d-1-2i)!} = \sum_{i=0}^{[\frac{d-1}{2}]} \frac{1}{i+1} \frac{(d-1)!}{(2i)!\, (d-1-2i)!} \frac{(2i)!}{i!\, i!}$$

$$= \sum_{i=0}^{[\frac{d-1}{2}]} \binom{d-1}{2i} \frac{h_{2i}}{i+1} \tag{C.35}$$

Hence

$$H_0(d_0) = 3\, H_0(d_0-1) - H(d_0-1) \tag{C.36}$$

or

$$H_0(d_0) = 3^{d_0-1} - \sum_{i=1}^{d_0-1} 3^{d_0-1-i} H(i) \tag{C.37}$$

since $H_0(1)=1$. Therefore, the generating function for $H_0(d_0)$ is

$$\sum_{d_0-1=0}^{\infty} H_0(d_0)\lambda^{d_0-1} = \sum_{d_0-1=0}^{\infty} 3^{d_0-1}\lambda^{d_0-1} - \sum_{d_0-1=0}^{\infty}\sum_{i=0}^{d_0-1} 3^{d_0-1-i} H(i)\lambda^{d_0-1}$$

$$= \frac{1}{1-3\lambda} - \sum_{i=0}^{\infty}\sum_{d_0-1=i}^{\infty} 3^{d_0-1-i} H(i)\lambda^{d_0-1}$$

$$= \frac{1}{1-3\lambda}\left(1 - \sum_{i=0}^{\infty} H(i)\lambda^{i}\right) = \frac{1}{1-3\lambda}\left(1 - \sum_{i=1}^{\infty} H(i)\lambda^{i}\right)$$

$$= \frac{1}{2\lambda(1-3\lambda)}\left(\sqrt{(1-2\lambda-3\lambda^2)} - (1-3\lambda)\right) \qquad (C.38)$$

and

$$U_0(\zeta) = 1 + \sum_{d_0=1}^{\infty} g_d^{-2} g_d^{d_0} H_0(d_0)\zeta^{d_0} = 1 + \zeta g_d^{-2} g_d \sum_{d_0-1=0}^{\infty} H_0(d_0)\,(g_d\zeta)^{d_0-1}$$

$$= 1 + \frac{g_d^{-2}\left(\sqrt{(1-2g_d\zeta-3g_d^2\zeta^2)} - (1-3g_d\zeta)\right)}{2(1-3g_d\zeta)} \qquad (C.39)$$

Method of Sequence-Generating Functions This method is based on the assumption that the Helmholtz free energy of a ligand, F_x, is linearly related to the number of segments, x, in the ligand. This assumption should be a good one in the limit as x approaches infinity, so that any special effects due to the ends of the ligand may be neglected. F_x will increase linearly with x if $q_{x,V}$ has the form

$$q_{x,V} = q_1^x \qquad (C.40)$$

so that

$$F_x = -kT \ln q_{x,V} = -xkT \ln q_1 \qquad (C.41)$$

The quantity q_1 may be called the residue partition function since it represents the contribution of one residue to $q_{x,V}$.

If $q_{x,V}$ has the form indicated in equation C.40

$$\Gamma(\zeta) = \sum_{x=1}^{\infty} q_{x,V}\zeta^x = \sum_{x=1}^{\infty} q_1^x \zeta^x \qquad (C.42)$$

This series converges only on the interval

$$0 < \zeta < \zeta_1 = q_1^{-1} \qquad (C.43)$$

and is a positive, monotonically increasing function of ζ on this interval.

Therefore, $1/\Gamma(\zeta)$ is a positive, monotonically decreasing function of ζ on the interval C.43, approaching zero when $\zeta=\zeta_1$, and ζ_1 is the smallest positive root of the equation[1]

$$1/\Gamma(\zeta) = 0 \tag{C.44}$$

or

$$1 - V(\zeta)U(\zeta) = 0 \tag{C.45}$$

so that

$$\zeta_1 = \frac{-\frac{g_d}{g_s^{-2}}\left(1+\frac{g_d}{g_{sf}}-\frac{2g_d^3}{g_s^{-2}g_{sf}}\right)+\left(\frac{g_d}{(g_s^{-2})^2}\left(g_d+\frac{4g_s^{-2}}{g_{sf}}-\frac{2g_d^2}{g_{sf}}-\frac{3g_d^3}{(g_{sf})^2}\right)\right)^{1/2}}{2\left(1-\frac{g_d^2}{g_s^{-2}}+\frac{g_d^4}{(g_s^{-2})^2}\right)} \tag{C.46}$$

With ζ_1 so determined, $q_{x,V}$ may be evaluated

$$q_{x,V} = q_1^x = \zeta_1^{-x} \tag{C.47}$$

Configurational parameters for the ligand also may be calculated. Thus, the average fraction of segments adsorbed, $\bar{\theta}$, is given by

$$\bar{\theta} = \frac{1}{x}\frac{\partial \ln q_{x,V}}{\partial \ln f} = -\frac{\partial \ln \zeta_1}{\partial \ln f} \tag{C.48}$$

and the average number of adsorbed stretches, $\bar{\ell}$, is[2]

$$\bar{\ell} = \frac{\partial \ln q_{x,V}}{\partial \ln g_s^{-2}} = -x\frac{\partial \ln \zeta_1}{\partial \ln g_s^{-2}} \tag{C.49}$$

From $\bar{\theta}$ and $\bar{\ell}$, the average length of an adsorbed stretch, $\bar{s}$, and the average length of a desorbed loop, $\bar{d}$, may be determined

$$\bar{s} = \frac{x\bar{\theta}}{\bar{\ell}} \tag{C.50}$$

$$\bar{d} = \frac{x(1-\bar{\theta})}{\bar{\ell}} \tag{C.51}$$

---

[1] $U_0(\zeta)$ does not appear in equation C.45 since the effects of the chain ends have been neglected when $q_{x,V}$ is written in the form C.40.

[2] The quantity $\bar{\ell}/x$ is used for infinite chains and provides the most meaningful comparison with the results for shorter ligands.

The Method of Steepest Descents For x large but finite so that the special properties of the ends of the ligand cannot be neglected, $q_{x,V}$ may be evaluated by the method of steepest descents. According to Cauchy's residue theorem (Sokolnikoff and Redheffer, 1966)

$$q_{x,V} = \frac{1}{2\pi i} \oint \Gamma(\zeta)\, \zeta^{-(x+1)}\, d\zeta \qquad (C.52)$$

where the integration is performed counterclockwise along any closed contour about the origin and within the circle of convergence of $\Gamma(\zeta)$. (equation C.43). $\Gamma(\zeta)$ is a monotonically increasing function of ζ between zero and ζ_1 on the real axis(as noted in the discussion of the method of sequence-generating functions), while $\zeta^{-(x+1)}$ is a monotonically decreasing function of ζ on that interval. Therefore, the integrand will exhibit one and only one minimum on the real, positive axis in the interval C.43, at a point ζ_0. The contour of integration is chosen to pass through ζ_0 in a direction giving a large decrease in the absolute value of the integrand (the path of steepest descents), so that the major contribution to the integral comes from the point $\zeta=\zeta_0$ and its neighborhood.

To select the appropriate contour, equation C.52 is rewritten as

$$q_{x,V} = \frac{1}{2\pi i} \oint e^{x\Phi(\zeta)}\, d\zeta \qquad (C.53)$$

where

$$\Phi(\zeta) = \frac{1}{x} \ln \frac{\Gamma(\zeta)}{\zeta^{x+1}} \qquad (C.54)$$

Expanding $\Phi(\zeta)$ in a Taylor's series about ζ_0

$$\Phi(\zeta) = \Phi_0 + \frac{1}{2!}\Phi_0^{(2)}(\zeta-\zeta_0)^2 + \frac{1}{3!}\Phi_0^{(3)}(\zeta-\zeta_0)^3 + \frac{1}{4!}\Phi_0^{(4)}(\zeta-\zeta_0)^4 + \frac{1}{5!}\Phi_0^{(5)}(\zeta-\zeta_0)^5 + \ldots\ldots \qquad (C.55)$$

where the superscript (j) represents the jth derivative with respect to ζ and the subscript 0 indicates evaluation at the point ζ_0.[1] If α is the angle between the vector $\zeta-\zeta_0$ and the real axis, the quantity $\zeta-\zeta_0$ may be written in the form

$$\zeta - \zeta_0 = re^{i\alpha} \tag{C.56}$$

while the derivatives $\Phi_0^{(j)}$ can be expressed as

$$\Phi_0^{(j)} = \rho_j e^{i\beta_j} \tag{C.57}$$

Hence

$$\Phi(\zeta) = \Phi_0 + \frac{1}{2!}\rho_2 r^2 e^{2i\alpha} e^{i\beta_2} + \frac{1}{3!}\rho_3 r^3 e^{3i\alpha} e^{i\beta_3} + \frac{1}{4!}\rho_4 r^4 e^{4i\alpha} e^{i\beta_4}$$

$$+ \frac{1}{5!}\rho_5 r^5 e^{5i\alpha} e^{i\beta_5} + \ldots\ldots \tag{C.58}$$

$\Phi(\zeta)$ can be split into real and imaginary parts

$$\Phi(\zeta) = A(\zeta) + iB(\zeta) \tag{C.59}$$

where

$$A(\zeta) = A_0 + \frac{1}{2!}\rho_2 r^2 \cos(\beta_2+2\alpha) + \frac{1}{3!}\rho_3 r^3 \cos(\beta_3+3\alpha) + \frac{1}{4!}\rho_4 r^4 \cos(\beta_4+4\alpha)$$

$$+ \frac{1}{5!}\rho_5 r^5 \cos(\beta_5+5\alpha) + \ldots\ldots \tag{C.60}$$

$$B(\zeta) = B_0 + \frac{1}{2!}\rho_2 r^2 \sin(\beta_2+2\alpha) + \frac{1}{3!}\rho_3 r^3 \sin(\beta_3+3\alpha) + \frac{1}{4!}\rho_4 r^4 \sin(\beta_4+4\alpha)$$

$$+ \frac{1}{5!}\rho_5 r^5 \sin(\beta_5+5\alpha) + \ldots\ldots \tag{C.61}$$

From these equations, it is evident that by moving off the real axis in a direction such that $\cos(\beta_2+2\alpha)=-1$, a large decrease in $A(\zeta)$, the real part of $\Phi(\zeta)$, will result.[2] Therefore, the path for which $\cos(\beta_2+2\alpha)=-1$ is the path of steepest descents. Since ζ_0 is on the real axis, all derivatives $\Phi_0^{(j)}$ will be real, so that $\beta_2=\beta_3=\beta_4=\beta_5=\ldots=0^\circ$. Therefore, $\alpha=\frac{\pi}{2}$

---

[1] The quantity $\Phi_0^{(1)}$ is zero by the definition of ζ_0.

[2] It is the real part of $\Phi(\zeta)$ which determines the behavior of the absolute value of the integrand of equation C.52.

and a circle of radius ζ_0 may be chosen as the contour of integration.

On this path, in the vicinity of ζ_0

$$A(\zeta) = A_0 - \frac{\Phi_0^{(2)} r^2}{2!} + \frac{\Phi_0^{(4)} r^4}{4!} - \cdots\cdots \tag{C.62}$$

$$B(\zeta) = B_0 - \frac{\Phi_0^{(3)} r^3}{3!} + \frac{\Phi_0^{(5)} r^5}{5!} - \cdots\cdots \tag{C.63}$$

so that

$$\Phi(\zeta) = \Phi_0 - \frac{\Phi_0^{(2)} r^2}{2!} - \frac{i\Phi_0^{(3)} r^3}{3!} + \frac{\Phi_0^{(4)} r^4}{4!} + \frac{i\Phi_0^{(5)} r^5}{5!} - \cdots\cdots \tag{C.64}$$

Hence[1,2]

$$q_{x,V} \cong \frac{1}{2\pi i} \oint e^{x\Phi_0 - \frac{x\Phi_0^{(2)} r^2}{2!} - \frac{ix\Phi_0^{(3)} r^3}{3!} + \frac{x\Phi_0^{(4)} r^4}{4!} + \frac{ix\Phi_0^{(5)} r^5}{5!}} \, d\zeta \tag{C.65}$$

Performing the change of variable $\zeta \to w$, where

$$w = \sqrt{\left(\frac{\Phi_0^{(2)} x}{2}\right)}\, r \tag{C.66}$$

so that

$$d\zeta = \sqrt{\left(\frac{2}{\Phi_0^{(2)} x}\right)}\, e^{i\alpha} dw = i\sqrt{\left(\frac{2}{\Phi_0^{(2)} x}\right)}\, dw \tag{C.67}$$

equation C.65 becomes

$$q_{x,V} \cong \frac{e^{x\Phi_0}}{2\pi} \sqrt{\left(\frac{2}{\Phi_0^{(2)} x}\right)} \int_{-\infty}^{\infty} e^{-w^2} e^{\left(- \frac{i\Phi_0^{(3)} w^3}{3!\, \frac{\Phi_0^{(2)}}{2} \sqrt{\left(\frac{\Phi_0^{(2)} x}{2}\right)}} + \frac{\Phi_0^{(4)} w^4}{4!\, \frac{\Phi_0^{(2)}}{2} \left(\frac{\Phi_0^{(2)} x}{2}\right)} + \frac{i\Phi_0^{(5)} w^5}{5!\, \frac{\Phi_0^{(2)}}{2} \left(\frac{\Phi_0^{(2)} x}{2}\right)^{3/2}} \right)} dw$$

$$\cong \frac{e^{x\Phi_0}}{2\pi} \sqrt{\left(\frac{2}{\Phi_0^{(2)} x}\right)} \int_{-\infty}^{\infty} e^{-w^2} \left\{ 1 - \frac{i\Phi_0^{(3)} w^3}{3!\, \frac{\Phi_0^{(2)}}{2} \sqrt{\left(\frac{\Phi_0^{(2)} x}{2}\right)}} + \frac{\Phi_0^{(4)} w^4}{4!\, \frac{\Phi_0^{(2)}}{2} \left(\frac{\Phi_0^{(2)} x}{2}\right)} + \frac{i\Phi_0^{(5)} w^5}{5!\, \frac{\Phi_0^{(2)}}{2} \left(\frac{\Phi_0^{(2)} x}{2}\right)^{3/2}} - \frac{1}{2!} \left(\frac{\Phi_0^{(3)} w^3}{3!\, \frac{\Phi_0^{(2)}}{2} \sqrt{\left(\frac{\Phi_0^{(2)} x}{2}\right)}} \right)^2 \right\} dw \tag{C.68}$$

---

[1] Any number of terms of the Taylor's series could be used without changing the evaluation procedure.

[2] Other maxima on this contour would be possible only if all exponents of ζ in $\Gamma(\zeta)$ shared a common factor other than unity.

where terms up to w^6 only have been included. The use of $-\infty$ to ∞ as integration limits introduces no significant error since the maximum contribution to the integrand is in the vicinity of ζ_0. Grouping terms with like powers of x

$$q_{x,V} \simeq \frac{e^{x\Phi_0}}{2\pi}\sqrt{\left(\frac{2}{\Phi_0^{(2)}x}\right)}\left\{\int_{-\infty}^{\infty} e^{-w^2}dw - \frac{1}{\sqrt{x}}\frac{i\Phi_0^{(3)}}{3\Phi_0^{(2)}\sqrt{\left(\frac{\Phi_0^{(2)}}{2}\right)}}\int_{-\infty}^{\infty} w^3 e^{-w^2}dw\right.$$
$$+\frac{1}{x}\left(\frac{\Phi_0^{(4)}}{6(\Phi_0^{(2)})^2}\int_{-\infty}^{\infty} w^4 e^{-w^2}dw - \frac{(\Phi_0^{(3)})^2}{9(\Phi_0^{(2)})^3}\int_{-\infty}^{\infty} w^6 e^{-w^2}dw\right)$$
$$\left.+\frac{1}{x^{3/2}}\frac{i\Phi_0^{(5)}}{30(\Phi_0^{(2)})^2\sqrt{\left(\frac{\Phi_0^{(2)}}{2}\right)}}\int_{-\infty}^{\infty} w^5 e^{-w^2}dw\right\} \qquad (C.69)$$

The oddness of the second and fifth integrands causes these terms to vanish, so that

$$q_{x,V} \simeq \frac{e^{x\Phi_0}}{2\pi}\left\{\frac{1}{\sqrt{x}}\sqrt{\left(\frac{2\pi}{\Phi_0^{(2)}}\right)} + \frac{1}{x^{3/2}}\sqrt{\left(\frac{2\pi}{\Phi_0^{(2)}}\right)}\left(\frac{\Phi_0^{(4)}}{8(\Phi_0^{(2)})^2} - \frac{5(\Phi_0^{(3)})^2}{24(\Phi_0^{(2)})^3}\right)\right\}$$
$$\simeq \frac{e^{x\Phi_0}}{2\pi}\frac{1}{\sqrt{x}}\sqrt{\left(\frac{2\pi}{\Phi_0^{(2)}}\right)}\left\{1 + \frac{1}{x}\frac{1}{(\Phi_0^{(2)})^2}\left(\frac{\Phi_0^{(4)}}{8} - \frac{5(\Phi_0^{(3)})^2}{24\Phi_0^{(2)}}\right)\right\} \qquad (C.70)$$

and as a first approximation for large x

$$q_{x,V} \simeq \frac{e^{x\Phi_0}}{2\pi}\frac{1}{\sqrt{x}}\sqrt{\left(\frac{2\pi}{\Phi_0^{(2)}}\right)} \qquad (C.71)$$

It will be shown below that this first approximation is adequate for most purposes.

Configurational parameters may be obtained by differentiation of equation C. 71 (or C. 70, if more precision is required)

$$\bar{\theta} = \frac{1}{x}\frac{\partial \ln q_{x,V}}{\partial \ln f} = \frac{1}{x}\frac{\partial \ln q_{x,V}}{\partial \zeta_0}\frac{\partial \zeta_0}{\partial \ln f} \qquad (C.72)$$

$$\bar{\ell} = \frac{\partial \ln q_{x,V}}{\partial \ln g_s^{-2}} = \frac{\partial \ln q_{x,V}}{\partial \zeta_0}\frac{\partial \zeta_0}{\partial \ln g_s^{-2}} \qquad (C.73)$$

$$\bar{d}_0 = \frac{1}{2} \frac{\partial \ln q_{x,V}}{\partial \ln g_{\bar{d}}^{-2}} = \frac{1}{2} \frac{\partial \ln q_{x,V}}{\partial \zeta_0} \frac{\partial \zeta_0}{\partial \ln g_{\bar{d}}^{-2}} \qquad (C.74)$$

Equations III.50 and III.51 may be used to evaluate $\bar{s}$ and $\bar{d}$, respectively.[1]

Convolution Procedure The convolution procedure has been discussed in Chapter III and therefore needs no further description. One evaluates $q_{x,V}$ by an actual determination of the coefficient of ζ^x in the generating function C.17, using the convolution formula III.45. Configurational parameters are obtained as described in equations III.46 to III.51.

Results and Discussion A comparison of the values for $q_{x,V}$ (on a per residue basis) obtained by the above three methods is presented in Table VI. The residue partition function decreases with increasing chain length, gradually approaching the value calculated for infinite chains by the method of sequence-generating functions. In the latter, the special properties of the free ends have been neglected entirely. Additional configurations are possible when free ends are included, causing an increase in the residue partition function. This increase is largest for shorter chains whose ends form a larger fraction of the total number of segments. The results obtained by the method of steepest descents are not changed markedly by the use of equation C.70 rather than C.71, and the disparity between these two approximations is less for the longer ligands, as expected. The close agreement be-

[1]The factor of one-half appears in equation C.74 because there are two ends per ligand.

Table VI

Vertical Partition Function[1]

ϵ	x	Seq-Gen. Funct.	Steepest Desc. Approx. 1st[2]	Steepest Desc. Approx. 2nd[3]	Convoln. Proc.
-kT	∞	1.2980			
	5000		1.2981	1.2981	
	1000		1.2985	1.2984	
	100		1.3029	1.3022	
	50		1.3077	1.3065	
	30		1.3135	1.3121	1.3123
	20		1.3120	1.3189	1.3193
	10		1.3327	1.3362	1.3375
-2kT	∞	2.0192			
	5000		2.0193	2.0193	
	1000		2.0196	2.0195	
	100		2.0229	2.0222	
	50		2.0264	2.0251	
	30		2.0308	2.0291	2.0292
	20		2.0356	2.0340	2.0342
	10		2.0473	2.0483	2.0492

[1]Values listed represent $\frac{1}{x} \ln q_{x,V}$ calculated by the methods indicated.

[2]Calculated from equation C.71.

[3]Calculated from equation C.70.

tween the results of the steepest descents procedure and the exact, convolution method (even for the shortest ligands for which the residue partition function was calculated by both procedures) testifies to the validity of the approximations made in the former method. The agreement improves as ϵ becomes more negative, since the absolute value of the integrand of equation C. 52 drops off more sharply upon moving away from the point ζ_0.

The configurational parameter of most interest is $\bar{\theta}$, the average fraction of ligand segments adsorbed, because its variation with the Boltzmann factor, f, is formally similar to the helix-coil transition in biopolymers (reviewed by Poland and Scheraga, 1970)[1] and order-disorder phenomena in general. Thus, a plot of $\bar{\theta}$ vs. ln$f$ exhibits a "transition," as shown in Figure 40.[2] Curves are presented for both an infinite chain and a chain of 20 segments (calculated by the method of sequence-generating functions and the convolution procedure, respectively). The "transition" for the infinite chain is considerably sharper than that for the ligand of 20 segments, since states of intermediate $\bar{\theta}$ are more important when the special properties of the ends are included.[3] A similar dependence on the chain length has been found

---

[1] In the helix-coil transition, $\bar{\theta}$ corresponds to the average fraction of segments in the helical state.

[2] In experimental and theoretical studies of helix-coil transitions, variations in energy correspond to changes in temperature or solvent.

[3] A sharp cut-off in the $\bar{\theta}$ vs. ln$f$ curve for infinite chains is observed at ln$f \cong 0.3$, since real roots ζ_1 for equation C. 45 are possible only when ln$f$ is greater than this critical value.

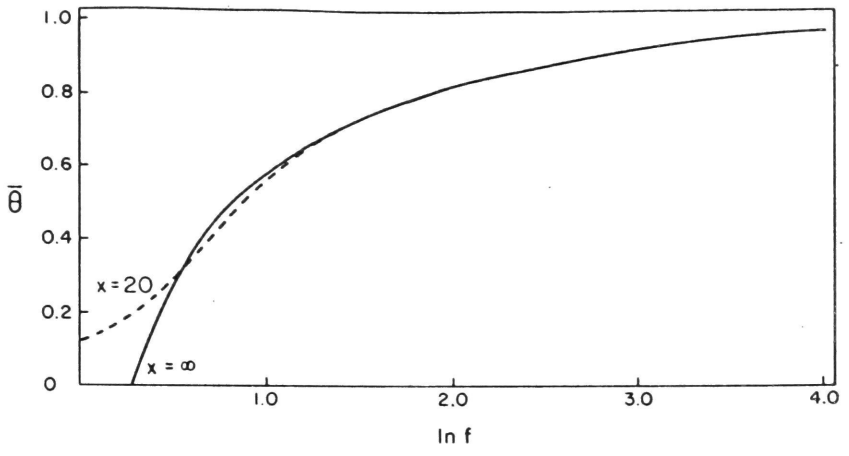


Figure 40. The variation in $\bar{\theta}$, the average fraction of ligand segments adsorbed, with ln f, the logarithm of the Boltzmann factor, for the ligands $x=\infty$(——) and x=20 (- - -), calculated by the method of sequence-generating functions and the convolution procedure, respectively. The quantity ln f is directly proportional to the energy of the site-segment interaction.

both theoretically and experimentally for helix-coil transitions in biopolymers. In the latter, however, the transitions are in general much sharper than those observed here, because the formation of a helical region is a cooperative process.[1]

[1]For example, in the polynucleotides, favorable interactions between adjacent helical segments are possible for helical sequences which exceed one residue in length.

BIBLIOGRAPHY

Arvidsson, E. O. (1965), Studies on Small Molecule-Protein Interactions with a Note on the Use of Tracers in Transport Systems, Lund, Studentlitteratur.

Bondi, A. (1964), J. Phys. Chem. 68, 441.

Brant, D.A., and Flory, P. J. (1965), J. Am. Chem. Soc. 87, 2791.

Caspar, D. L. D., and Klug, A. (1962), Cold Spring Harbor Symp. Quant. Biol. 27, 1.

Darwin. C.G., and Fowler, R.H. (1922), Phil. Mag. 44, 450.

Domb, C. (1969), Adv. Chem. Phys. 15, 229.

Elworthy, P.H. (1963), J. Chem. Soc., 388.

Feller, W. (1957), An Introduction to Probability Theory and its Applications, Volume I, New York, N. Y., John Wiley and Sons, Inc.

Fisher, M. E., and Sykes, M.F. (1959), Phys. Rev. 114, 45.

Flory, P. J. (1953), Principles of Polymer Chemistry, Ithaca, N. Y., Cornell University Press.

Flory, P. J., and Miller, W.G. (1966), J. Mol. Biol. 15, 284.

Gans, P. J. (1965), J. Chem. Phys. 42, 4159.

Gibson, K. D., and Scheraga, H. A. (1967), Proc. Natl. Acad. Sci. 58, 420.

Gillberg-LaForce, G., and Forsén, S. (1970), Bioch. Biophys. Res. Comm. 38, 137.

Goodman, D.S. (1958), J. Am. Chem. Soc. 80, 3892.

Guggenheim, E.A. (1944), Proc. Roy. Soc. London 183A, 203.

Hammersley, J.M., and Morton, K.W. (1954), J. Roy. Stat. Soc. B16, 23.

Hirschfelder, J.O., Curtiss, C.F., and Bird, R.B. (1954), Molecular Theory of Gases and Liquids, New York, N.Y., John Wiley and Sons, Inc.

Karush, F., and Sonenberg, M. (1949), J. Am. Chem. Soc. 71, 1369.

Kemeny, J.G., and Snell, J.L. (1960), Finite Markov Chains, Princeton, N.J., D. Van Nostrand Company, Inc.

Klotz, I.M. (1953), in The Proteins, IB, Neurath, H., and Bailey, K., Eds., New York, N.Y., Academic Press, Inc.

Koshland, D.E., Jr., Némethy, G., and Filmer, D. (1966), Biochemistry 5, 365.

Laiken, S., and Printz, M. (1970), Biochemistry 9, 1547.

Lifson, S. (1964), J. Chem. Phys. 40, 3705.

London, F. (1937), Trans. Faraday Soc. 33, 8.

Marks, R.W. (1960), The Dymaxion World of Buckminster Fuller, New York, N.Y., Reinhold Publishing Corporation.

Moran, P. (1968), An Introduction to Probability Theory, Oxford, Clarendon Press.

Némethy, G., and Scheraga, H.A. (1962), J. Chem. Phys. 36, 3401.

Pitzer, K.S. (1959), Adv. Chem. Phys. 2, 59.

Poland, D.C., and Scheraga, H.A. (1965), J. Phys. Chem. 69, 2431.

Poland, D.C., and Scheraga, H.A. (1970), Theory of Helix-Coil Transitions in Biopolymers, New York, N.Y., Academic Press, Inc.

Ray, A., Reynolds, J.A., Polet, H., and Steinhardt, J. (1966), Biochemistry 5, 2606.

Reynolds, J., Herbert, S., Polet, H., and Steinhardt, J. (1967), Biochemistry 6, 937.

Reynolds, J., Herbert, S., and Steinhardt, J. (1968), Biochemistry 7, 1357.

Roe, R.-J. (1965a), Proc. Natl. Acad. Sci. 53, 50.

Roe, R.-J. (1965b), J. Chem. Phys. 43, 1591.

Rosenberg, R.M., Crespi, H.L., and Katz, J.J. (1969), Biochim. Biophys. Acta 175, 31.

Scatchard, G., Scheinberg, I.H., and Armstrong, S.H., Jr. (1950), J. Am. Chem. Soc. 72, 535.

Scheraga, H.A. (1968), Adv. Phys. Org. Chem. 6, 103.

Schrödinger, E. (1964), Statistical Thermodynamics, Cambridge, Cambridge University Press.

Silberberg, A. (1962), J. Phys. Chem. 66, 1872, 1884.

Sokolnikoff, I.S., and Redheffer, R.M. (1966), Mathematics of Physics and Modern Engineering, New York, N.Y., McGraw-Hill Book Company.

Spector, A.A., John, K., and Fletcher, J.E. (1969), J. Lipid Res. 10, 56.

Teresi, J.D. (1950), J. Am. Chem. Soc. 72, 3972.

Teresi, J.D., and Luck, J.M. (1952), J. Biol. Chem. 194, 823.

Wall, F.T., Hiller, L.A., Jr., and Atchison, W.F. (1955), J. Chem. Phys. 23, 913.

Wall, F.T., Hiller, L.A., Jr., and Wheeler, D.J. (1954), J. Chem. Phys. 22, 1036.

Waugh, D.F. (1954), Adv. Prot. Chem. 9, 325.

End