The effects of internal constraints on the configurations of chain molecules

Hue Sun Chan and Ken A. Dill

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

(Received 14 September 1989; accepted 16 November 1989)

We explore the three-dimensional configurations of chain molecules containing one or more self-contacts (constraints). We focus predominantly on the 1-, 2-, and 3-constraint ensembles. We take into account excluded volume by exhaustive computer enumeration of the conformational spaces of short chains on three-dimensional simple cubic lattices, and through use of the path integral approach of Edwards and Freed. We develop topological correlation functions to describe how the cyclization probability of one loop affects cyclization of another. There are two rather striking findings. (i) Considerable amounts of internal architecture (helices and antiparallel and parallel sheets) are predicted to arise in compact polymers due simply to steric restrictions. This appears to account for why there is so much internal organization in globular proteins. (ii) Several cyclization properties are remarkably ideal for chains which are relatively or highly compact in three dimensions. For example, in relatively compact molecules the correlation functions of loop pairs are well predicted by the random-flight model of Jacobson and Stockmayer; the number of configurations of maximally compact chains is predicted relatively well by the Flory theory of excluded volume, which is found to be better than the Huggins theory in three dimensions; and the probability of cyclization within globular chains is well predicted by the Bragg-Williams approximation.

I. INTRODUCTION

The properties of chain molecules are governed by the statistical mechanics of their ensembles of accessible conformations. Some conformations contain intrachain loops, or self-contacts of the chain. Such conformations are often of considerable importance, for example, in understanding ring-chain equilibria, the stabilization of globular proteins by disulfide bonds and other crosslinks, and conformational equilibria and phase transitions in double-stranded molecules such as DNA or triple-stranded molecules such as collagen.

Our purpose here is to explore the three-dimensional configurations of chain molecules which are constrained by the existence of one or more intrachain loops. We do so by two methods. First, we describe exhaustive simulations of the conformations of chains on three-dimensional simple cubic lattices. Second, we describe a first-order perturbation path-integral approach. The combination of methods is of value in identifying possible artifacts from either method alone.

The methods and results can be described most concisely if we first define some useful terminology. Connected neighbors are pairs of monomers which are adjacent along the sequence of the chain, residues \( i \) and \( i+1 \). Topological neighbors are pairs of monomers \( (i,j) \), \( i \neq j \), which are not adjacent in the sequence, \( j \neq i-1, i+1 \), but are adjacent in three-dimensional space. A pair of topological neighbors is also referred to as a contact.

A contact may be of two different types. A presumed contact is given in advance, specified a priori; an inferred contact is not. For example, consider the problem of exploring the configuration space of all conformations of a chain with topological neighbor pair \( (i,j) \). By the specification of the problem, the pair \( (i,j) \) is a presumed contact that must occur in all configurations. Some of the configurations which satisfy this constraint will also have other neighbor pairs; these are inferred contacts. These inferred contacts comprise the excluded volume interactions considered here for chains with a given single presumed contact \( (i,j) \).

The topology, i.e., pattern of intrachain loops or contacts, of a conformation can be displayed on a contact map, in which the contact \( (i,j) \) is represented by a dot at the \( i \)th row and \( j \)th column. An example of such representation is given in Fig. 1 (see also Ref. 8). The order of a contact \( k = |i-j| \) measures the loop size by the number of bonds along the chain between the two contacting residues \( i \) and \( j \). Contacts of the same order appear as dots on the same diagonal on the contact map.

II. THE LATTICE MODEL

First, we discuss the lattice model. Chains are configured on three-dimensional simple cubic lattices in which

![Diagram of contact map and chain geometry](https://example.com/diagram.png)

**FIG. 1.** Contact map. Dots indicate pairs of monomers which are spatially adjacent for the configuration shown on the left. For this conformation the topological contacts are \( \{(1,4), (1,6), (5,8)\} \).
each lattice site is occupied by no more than one monomer. Monomers are numbered sequentially, 1, 2, 3, ... , from one end of the chain. Contacts are identified when a pair of unconnected monomers occur in nearest-neighbor lattice sites. On the simple cubic lattice, as on the square lattice, only odd-order (k = odd number) contacts are possible. Even-order contacts are forbidden a priori, so their absence need not be represented explicitly on the contact map. Although this restriction could be circumvented by the use of other lattices, it is of little consequence for our present purposes. We study here only the simple cubic lattice.

Conformational freedom is affected by the presence of presumed contacts along the chain. One consequence is the appearance of inferred blocks, contacts that are forbidden by a given set of presumed contacts (see Fig. 2). This is an effect of excluded volume. Since the importance of excluded volume diminishes with increasing spatial dimensionality, for a given set of presumed contacts, fewer blocks are implied in three dimensions than in two dimensions. In fact, it can be proved by explicit construction that for a single presumed contact with order k > 3, there is no inferred block in the three-dimensional simple cubic lattice. In contrast, corresponding inferred blocks are always present on two-dimensional square lattices.

In the following analysis, the exact lattice enumerations are used to determine the number of conformations $Q$ under various constraints. In particular, $Q_0(N)$ is the total number of accessible conformations with $N$ bonds ($N + 1$ monomers), without any constraint except for the requirement of excluded volume. Due to the approximate exponential scaling of $Q_0(N)$ as a function of $N$,

$$Q_0(N) \sim N^{N \mu N},$$

(2.1)

only short chains can be exhaustively enumerated. The scaling (2.1) has been confirmed by exact enumerations$^{6-13}$ and is consistent with the prediction of renormalization group analysis.$^{7,12,13}$ $Q_0(N)$ for the simple cubic lattice has been computed by Sykes$^{10}$ for $N < 16$, based on his data, $Q_0(N)$ for $N = 17$ can be determined by extrapolation of Eq. (2.1), with the estimates $\gamma = 0.1667$ for any three-dimensional chains and $\mu = 4.682$ for simple cubic lattices. In the present enumeration of $Q$, the two ends of the chains are considered to be distinguishable, and rigid rotations and reflections are considered to be distinguishable.$^{14}$

![Figure 2: Implied blocks. The figure on the left shows all possible positions of monomer number 4 and 5 when the contact (0,3) is presumed. The blocks (squares) on the contact map on the right represent contacts which are impossible in any configuration, given the presumed contact which is represented as a circular dot. The dotted line indicates a possible (1,6) contact.](image-url)

### III. SINGLE SELF-CONTACTS

We first consider the configurations of chains constrained to have one presumed contact $(i,j)$. As in our two-dimensional study,$^4$ we define a "reduction factor" or cyclization probability $R(N;i,j)$ for a chain with $N$ bonds. If $Q(N;i,j)$ is the number of conformations constrained to have the $(i,j)$ contact pair, and $Q_0(N)$ is the total number of accessible conformations, the reduction factor due to the constraint of a single presumed contact is$^{15}$

$$R(N;i,j) = \frac{Q(N;i,j)}{Q_0(N)}.$$  

(3.1)

Thus $R(N;i,j)$ is the fractional restriction of conformational space due to the contact $(i,j)$.

The simplest theoretical treatment of this reduction factor is due to Jacobson and Stockmayer (JS).$^1$ They assumed random-flight statistics for the chains, valid in a theta solvent, but therefore excluded volume effects are ignored. In the JS model, the reduction factor is predicted to be

$$R(N;i,j) = (\Delta V) \left( \frac{3}{2 mr[i-j]} \right)^{3/2}, \quad N > |i-j|,$$

(3.2)

in three-dimensional space. The tolerance $\Delta V$ is an empirical parameter that characterizes the volume around a monomer within which another monomer must reside in order to be qualified as being in contact with the first. The JS theory, Eq. (3.2), predicts that the reduction factor depends only on the contact order $k = |i-j|$, according to the power law $k^{-3/2}$, and is independent of the chain length $N$ and the location of the contact along the chain sequence.

Exact enumerations on simple cubic lattices indicate that the dependence of $R$ on $k$ is significantly altered by effects of excluded volume. Specifically, there are end effects: reduction factors depend on the location of the contact along the chain. In other words, the cyclization probability of a loop depends on the two tail lengths $l_0$ and $l_0$ [Fig. 3(a)] attached to the two contacting monomers $(i,j)$ that define the loop. This feature can be clearly demonstrated by focusing on the dependence of reduction factors on one of the two tail lengths $l_0$. This is accomplished by considering the limit $R(\infty;i,j)$ of reduction factors of low-order contacts that are embedded in relatively long chains. This limit corresponds to the condition $N \gg k$. These factors are estimated by deducting limiting values of $R(N;i,j)$ for successively longer chains (increasing $N$). Figure 3(b) exhibits the $N \to \infty$ reduction factors $R(\infty;i_0, 1, l_0 + k + 1)$ of a single presumed contact of order $k = 3, 5, 7,$ and $9$. The results show that end effects are significant; the reduction factors decrease as the loop moves from the chain end $(l_0 = 0)$ towards the middle of the chain.

These end effects are consequences of excluded volume interactions, based on the same physical principles as in two dimensions. Excluded volume effects are less important in higher spatial dimensions; the end effects observed here for three-dimensional chains are less prominent than those in two dimensions. For example, the ratios between reduction factors in Fig. 3(b) at chain end and midchain are 1.29, 1.44, 1.54, and 1.58 for $k = 3, 5, 7,$ and $9$, respectively, for three-dimensional chains, whereas the corresponding ratios on
two-dimensional square lattices are 1.60, 2.25, 2.79, and 3.13, respectively.\textsuperscript{8}

Next we consider the power-law dependence of \( R(N; i, j) \) on the contact order \( k = |i - j| \). It is well known\textsuperscript{8,16-18} that for large \( k \), the reduction factor varies approximately as \( k^{-\nu} \). The exponent \( \nu \) is universal, i.e., it is independent of the choice of lattice.\textsuperscript{19} For three-dimensional chains subject to excluded volume, earlier work\textsuperscript{16-18} and our present observations show that the exponent \( \nu \) is somewhat larger than the random-flight JS prediction of 3/2.

In the determination of the scaling exponent \( \nu \) from exact lattice simulations, some authors\textsuperscript{17,18} cast the power-law dependence of the reduction factor \( R(k; 1, k + 1) \) as \((k + 1)^{-\nu} \) instead of \( k^{-\nu} \). The length \( k + 1 \) corresponds to the number of bonds in the ring formed by joining the contacting monomers 1 and \( k + 1 \). Using either \( k \) or \( k + 1 \) in the power-law scaling should give the identical exponent \( \nu \) in the \( k \to \infty \) asymptotic regime. However, the two conventions give slightly different estimates for \( \nu \) for short chains. In order to facilitate better comparison with earlier lattice simulations, we plot the reduction factors \( R(N; i_0 + 1, i_0 + k + 1) \) as functions of \( (k + 1) \) in Fig. 4(b), and base our estimates of \( \nu \) on the \( (k + 1)^{-\nu} \) power law.

For ring polymers, \( k = N \), Wall, Hiller, and Atchison\textsuperscript{20} were the first to use Monte Carlo sampling on the lattice to obtain an estimate of \( \nu \approx 2 \) for large \( k \) in both two and three dimensions. Subsequently, exact lattice enumerations by Hiley and Sykes\textsuperscript{17} and Martin, Sykes, and Hioe\textsuperscript{18} sharpened the estimations to \( \nu \approx 11/6 \) in two dimensions and \( \nu \approx 23/12 \) in three dimensions. Our enumerations confirm this power-law dependence of cyclization probability.\textsuperscript{21} In addition, we have also studied cases in which the presumed contact \((i, j)\) is not of the two chain ends, but may occur anywhere within the polymer. Our results show that the exponent \( \nu \) depends both on \( k \) and on the location of the cycle in the chain. This is illustrated in Fig. 4(a): (1) is a ring polymer, corresponding to the case considered in Refs. 16-18, (2) has a loop at one end of the chain, connected to a long tail, (3) is a loop at the middle of a long chain, connected to two long tails. It is clear from Fig. 4(b) that the scaling exponents \( \nu_1 \), \( \nu_2 \) and \( \nu_3 \) for case (1), (2), and (3) in Fig. 4(a) are different, with \( \nu_1 < \nu_2 < \nu_3 \). The best estimates from our exact enumeration data are \( \nu_1 \approx 1.99 \), \( \nu_2 \approx 2.18 \), and \( \nu_3 \approx 2.42 \). The existence of multiple scaling exponents is another consequence of excluded volume, resulting from interactions between the loop and the tails, and is also observed in two dimensions.\textsuperscript{8}

Wall, Hiller, and Atchison\textsuperscript{20} were the first to consider the distinction between case (1) and (2). They noted that for a loop of a specific size, a loop of type (1) is more probable to form than a type (2) loop. Nevertheless they concluded that the exponents \( \nu_1 \) and \( \nu_2 \) are approximately equal for large \( k \), and the difference in probability for the two cases is attributed to the different overall multiplicative constants. However, more accurate subsequent studies by other authors suggested a small difference between the exponents \( \nu_1 \) and \( \nu_2 \), as reviewed by Redner.\textsuperscript{22} From an examination of exact enumeration data on the two-dimensional triangular lattice and the three-dimensional fcc lattice, Redner\textsuperscript{22} also provided the estimate of \( \nu_1 = 2.95 \pm 0.20 \) and

\[ \frac{R}{k+1} \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]

\[ \left( \begin{array}{c}
\nu_1 \\
\nu_2 \\
\nu_3
\end{array} \right) \]
\( N + 1 = 14 \) monomers. The exact values for \( Q(13,i,j) = Q_{0}(13)R(13;i,j) \) are given in Table I. Although both plots indicate that the smallest loops are the most strongly favored—the deepest free energy trough is adjacent to the main diagonal, and is shown as the darkest region—there are significant differences. In particular, the bending of the contours near the two ends of the chains in Fig. 5 is a manifestation of excluded-volume end effects, that loops near chain ends are favored. A further consequence of this is that the least-favored loops, shown as the lightest regions in these plots, occur around (2,13) in this example, instead of at (1,14) as predicted by the random-flight JS theory. The location of the least-favored contacts are important in protein engineering: if these contacts are formed by disulfides or other crosslinks, they would lead to maximum stabilization of the folded state of globular proteins.² ⁴

IV. THE PATH-INTEGRAL MODEL

We now turn to the path-integral approach ² ⁷ for chains with a single presumed contact. In this approach, the total number of accessible conformations is represented by the partition function

\[
Q_{0}(N_{0}) = \int \mathcal{D} \mathbf{e} e^{-H(N_{0},\mathbf{v}_{0})}
\]

(4.1)

for chains with length \( N_{0} \), where the Boltzmann factor

\[
H(N_{0},\mathbf{v}_{0}) = \frac{1}{2} \int_{0}^{N_{0}} d\tau \left| \frac{d\mathbf{c}(\tau)}{d\tau} \right|^{2} + \frac{v_{0}}{2} \int_{0}^{N_{0}} d\tau
\]

\[
\times \int_{|\tau - \tau'| > \rho} d\tau' \delta[\mathbf{c}(\tau) - \mathbf{c}(\tau')]
\]

(4.2)

consists of two terms: the first accounts for the random-flight statistics, while the second is a simple delta-function two-body repulsion that approximates the intrachain excluded volume effects. The quantity \( v_{0} > 0 \) in the second term is an empirical parameter for the strength of the excluded volume interaction. In Eqs. (4.1) and (4.2), \( \tau \) and \( \tau' \) are contour lengths along the chain. The vector \( \mathbf{c}(\tau) \) specifies the chain conformations, through the relation

\[
\mathbf{c}(\tau) = \sqrt{d} \mathbf{r}(\tau) = \sqrt{3} \mathbf{r}(\tau) \quad (d = \text{spatial dimension} = 3),
\]

where \( \mathbf{r}(\tau) \) is the spatial vector of the chain at contour length.

### Table I. Number of conformations \( Q(N_{0},i,j) \) consistent with each given \( (i,j) \) contact tabulated here for all contacts \( (i,j) \) along chains with \( N + 1 = 14 \) monomers configured on simple cubic lattices. The data is plotted as contours in Fig. 5.

<table>
<thead>
<tr>
<th>( i )</th>
<th>( j = 4 )</th>
<th>( j = 5 )</th>
<th>( j = 6 )</th>
<th>( j = 7 )</th>
<th>( j = 8 )</th>
<th>( j = 9 )</th>
<th>( j = 10 )</th>
<th>( j = 11 )</th>
<th>( j = 12 )</th>
<th>( j = 13 )</th>
<th>( j = 14 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>126 890 448</td>
<td>54 018 888</td>
<td>27 855 816</td>
<td>17 261 952</td>
<td>12 259 848</td>
<td>12 673 920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>101 305 008</td>
<td>40 109 688</td>
<td>20 043 936</td>
<td>12 443 472</td>
<td>9 517 608</td>
<td>12 259 848</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>98 334 026</td>
<td>38 565 360</td>
<td>18 927 984</td>
<td>11 948 664</td>
<td>12 043 472</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>98 615 688</td>
<td>37 426 920</td>
<td>18 454 200</td>
<td>18 927 984</td>
<td>12 259 848</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>97 682 712</td>
<td>37 406 976</td>
<td>20 043 936</td>
<td>17 261 952</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>98 615 688</td>
<td>38 565 360</td>
<td>27 855 816</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>98 334 026</td>
<td>40 109 688</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>98 334 026</td>
<td>101 305 008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>101 305 008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>126 890 448</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[ f(N; l_0 + 1, l_0 + k + 1) = 4 \left[ \sqrt{N} - \sqrt{N - k} + \sqrt{k} \cos^{-1} \sqrt{\frac{k}{k + 4l_0}} \right]. \] (4.4)

The path-integral reduction factor Eqs. (4.3) and (4.4) through \( O(v_0) \) predicts remarkably well the end effects observed in the exact lattice enumerations. Figure 8 is the contour plot of \(-\ln R(N; l_0 + 1, l_0 + k + 1) + \text{const}\) taken from Eqs. (4.3) and (4.4) for chains with \( N + 1 = 14 \) residues, with \( v_0 = 0.41 \). A comparison with Fig. 5 shows that the path-integral method reproduces well the lattice results. The successful reproduction of features observed in exact lattice enumerations by the path-integral model, such as the enhancement of cyclization probability near chain ends, and the existence of a peak near, but not at, the upper right corner of the contact map, indicates that these features are not artifacts of a particular lattice. The fact that these features are predicted by these two very different approaches implies that they follow from general principles of excluded volume interactions.

FIG. 8. Topological energy surface for a single presumed contact—path-integral results through first order in \( v_0 \). Contours are given in 0.3\( kT \) steps.
We now focus on the variation of the power-law dependence of reduction factors on the location of the loop along the chain. Of special interest are the three cases depicted in Fig. 4(a). Since these are extreme cases, they can be used to obtain bounds for intermediate situations. The function in Eq. (4.4) can easily be computed for these special cases by requiring \( l_0 = l'_0 = 0 \) for (1), \( l_0 = 0, l'_0 \to \infty \) for (2), and \( l_0, l'_0 \to \infty \) for (3). The result is \( f = 4\sqrt{k}, f = 2\pi_0/k \), and \( f = 4\pi_0/k \), respectively. Substituting these expressions for \( f \) into Eq. (4.3), we find a monotonic decay in cyclization probability with increasing contact order \( k \) in all three cases, the rate of decline increasing in the order \( (1) < (2) < (3) \). This result is consistent with the exact enumeration data presented in Fig. 4(b), which shows the order \( v_1 < v_2 < v_3 \) for the scaling exponents.

Quantitative prediction, however, cannot be expected from this path integral perturbation method without renormalization because the validity of the perturbation procedure requires the quantity \( v_0 N^{-d/2} \) to be small\(^7 \) \( (v_0 N^{-d/2} \ll 1) \), where \( N \) is the chain length and \( d = 4 - d \). This implies that for all spatial dimensions \( d < 4 \), \( v_0 \) is restricted to small values if the chain is long (\( N \) large). The success of the above comparison between the three-dimensional path-integral model and the results of lattice chains with full excluded volume interaction is due to the shortness of the chains, for which a reasonably large value of \( v_0 \) is permissible. For long chains, however, \( v_0 \) has to be small, in accordance to the above condition \( v_0 N \ll 1 \) for \( d = 3, \epsilon = 1 \). Though a small \( v_0 \) is often adequate for describing polymers near the theta point, it is not sufficient to account for the full excluded volume effect.

The renormalization group analysis\(^7,12,13 \) is able to circumvent this difficulty for long chains by allowing the spatial dimension \( d \) to become a continuous variable, and by considering expansions in powers of \( \epsilon \). This procedure allows \( \epsilon \) to be arbitrarily small if \( N \) is large, thus preserving the condition for the validity of perturbation theory even when \( v_0 \) is not small. According to the procedure of Freed and coworkers,\(^7 \) the scaling exponent \( v_i \) for the reduction factors can be determined for chains with full excluded volume (i.e., at the nontrivial fixed point) by obtaining the coefficient of the \( \ln a \) term in the equations of the diagrams in Fig. 7, which are pertinent to the function \( f \) in four dimensions (\( d = 4 \)). These terms have "logarithmic divergences," since \( \ln a \) diverges as the cutoff vanishes \((\alpha \to 0)\). We have performed such analysis for cases (1), (2), and (3) of Fig. 4(a). Since the renormalization group analysis based on the path integral (4.1) and (4.2) is well documented,\(^7 \) the details are omitted here. Our calculations give the long-chain scaling exponents through first order in \( \epsilon \),

\[
\begin{align*}
v_1 &= 2 - \frac{\epsilon}{8} + O(\epsilon^2), \\
v_2 &= 2 + O(\epsilon^2), \\
v_3 &= 2 + \frac{\epsilon}{4} + O(\epsilon^2).
\end{align*}
\]

Substituting \( d = 3, \epsilon = 1 \) into the above equations yields \( v_1 \approx 1.88, v_2 \approx 2.0, \) and \( v_3 \approx 2.25 \). Although Eq. (4.5) is only accurate to \( O(\epsilon) \), the predicted scaling exponents are in approximate agreement with our estimates from exact enumerations of short chains. Furthermore, Eq. (4.5) demonstrates that the different slopes observed in the last section for cases (1), (2), and (3), with \( v_1 < v_2 < v_3 \), are not lattice artifacts, but are universal features that follow from general principles of excluded volume interactions.

Our simple \( O(\epsilon) \) analysis is consistent with a previous study by des Cloizeaux\(^{24} \) through \( O(\epsilon^3) \). Instead of the path integral (4.1) and (4.2), des Cloizeaux used an earlier and more complicated formalism, which exploits a correspondence between the polymer excluded volume problem and the \( n \to 0 \) limit of \( O(n) \) field theory, originally discovered by de Gennes.\(^{13} \) Combining the results of both authors\(^{13,24} \) gives the scaling exponents of cases (1), (2), and (3) through \( O(\epsilon^3) \),

\[
\begin{align*}
v_1 &= 2 - \frac{\epsilon}{8} + \frac{27\epsilon^2}{256} + O(\epsilon^3), \\
v_2 &= 2 + \frac{\epsilon^2}{16} + O(\epsilon^3), \\
v_3 &= 2 + \frac{\epsilon}{4} - \frac{15\epsilon^2}{128} + O(\epsilon^3).
\end{align*}
\]

These improved formulas give \( v_1 \approx 1.980, v_2 \approx 2.125, \) and \( v_3 \approx 2.133 \) for chains configured in three dimensions, again confirmed by our lattice observation \( v_1 < v_2 < v_3 \). The values for \( v_1 \) and \( v_2 \) compare well with the exact lattice enumeration estimates of the last section, but there is considerable discrepancy between the renormalization group prediction of \( v_3 \) and our lattice estimate. This may be due to the limitation to short chains in exact lattice enumerations, whereas power-law scaling is strictly valid only for very long chains.

Some aspects of these predictions have recently been experimentally tested. Tsunashima and Kurata\(^{25} \) performed light scattering measurements on long polymers in good solvents. They found that the probability of forming an internal loop is indeed lower than that of forming an end-to-end ring of the same size, and deduced that \( v_1 \approx 2.22 \). Their experimental value compares favorably with the theoretical treatment of des Cloizeaux,\(^{24} \) who utilized the explicit terms through \( O(\epsilon^3) \) calculated by perturbative renormalization group analysis [Eq. (4.6)] to obtain the estimation of \( v_3 \approx 2.18 \) by an approximate summation of the asymptotic series for \( v_3 \).

V. MULTIPLE CONSTRAINTS

In the preceding sections, we have considered how the configurations of a chain molecule are affected by the existence of a single given constraint, namely that monomers \((i, j)\) are topological neighbors. In the following sections, we consider how the configurations of a chain molecule are affected by more than a single internal constraint, i.e., when there are multiple topological neighbors \((i_1, j_1), (i_2, j_2), \ldots\). There are several reasons for interest in this problem. First, the most elementary building block of "secondary structures" in globular proteins, i.e., helices and parallel and antiparallel sheets, is a set of two such contact pairs. This is
evident from Fig. 9, in which the lattice models of secondary structures are given. The adequacy of these as models for real protein secondary structures has been discussed elsewhere. This study of multiple constraints therefore permits exploration of internal architecture in highly constrained polymer molecules. Second, many of the problems for which a single loop formation is important as an idealization, multiple loop formation is generally the reality. For example, globular proteins stabilized by crosslinks are generally complicated by more than one. One of the questions of interest in those cases is: can multiple loops be generally considered independent, and the free energy contributions to protein stabilization considered additive? Third, chains in poor solvents are seldom adequately described by first-order perturbation theory. It is therefore of value to study the nature of effects of multiple constraints on the configurations of chain molecules.

VI. CORRELATION BETWEEN TWO CONTACTS: THE LATTICE MODEL

We begin by considering the simplest multiple constraint problem, namely that of two loops in the chain. The correlation between the cyclization probabilities of the two loops \((i_1,j_1)\) and \((i_2,j_2)\) can be defined, as it has been for chains in two dimensions, by the correlation function

\[
g_{k_1,k_2}(L) = \frac{R(N;i_1,j_1;i_2,j_2)}{R(N;i_1,j_1)R(N;i_2,j_2)}, \tag{6.1}
\]

where \(k_1 = |i_1 - j_1|\) and \(k_2 = |i_2 - j_2|\) are the contact orders of the two contacts respectively. Without loss of generality, we require \(i_1 < j_1, i_2 < j_2,\) and \(k_1 < k_2\) in the above definition, and the separation \(L\) along the chain between the two contacts is given by \(L = j_2 - j_1\). The definitions for \(L\) are shown graphically in Fig. 10, and the three possible topologies for chains with two contacts are shown in Figs. 10(b) and 10(c).

In Eq. (6.1), \(R(N;i_1,j_1)\) and \(R(N;i_2,j_2)\) are the single-contact reduction factors as defined in Eq. (3.1) for chains of length \(N\). \(R(N;i_1,j_1;i_2,j_2)\) is the reduction factor for the pair of presumed contacts. The correlation \(g\) is the ratio of the actual number of conformations which satisfy the two presumed contacts to the number of conformations if the two contacts were independent. Hence it is a measure of the degree to which one loop hinders or enhances the formation of another: enhancement is indicated when \(g > 1\), hindrance \(g < 1\), and \(g = 1\) implies independence. The correlation function \(g_{k_1,k_2}\) is predominantly a function of \(L\), as indicated in Eq. (6.1), even though in general it is also a function of the two tail lengths \(l_0\) and \(l_0'\) (see Fig. 10); i.e., \(g_{k_1,k_2} = g_{k_1,k_2}(L;l_0,l_0')\). The dependence on \(l_0\) and \(l_0'\) is neglected here because it is relatively weak, particularly if the two contacts are near the middle of a long chain \((l_0,l_0' \rightarrow \infty)\).

As in the one-loop case, we begin with the simplest theoretical treatment of the two loop problem. We first show the extension of Jacobson–Stockmayer (JS) random-flight theory, which neglects effects of excluded volume. For a pair of contacts of different orders \((k_2 > k_1)\) and nonnegative separation \((L > 0)\), the random-flight correlation is given by

![FIG. 9. Helices and sheets are represented as simple dot patterns on the contact map. (a) Four best representations of helices on the simple cubic lattice; all are cylindrical repeating patterns involving \((i+3)\) contacts. (b)–(d) Sheets and turns. Sheets are defined by at least six coplanar monomers (3 contacts). Dotted line circumscribe the fundamental units.](http://jcp.aip.org/jcp/copyright.jsp)
for chains configured in three dimensions. Here the labels I, II, and III represent different intervals of \( L \), corresponding to the three different topologies shown in Figs. 10(b) and 10(c). The superscript \((0)\) in \( g \) indicates that this is the random-flight result, corresponding to the zeroth order term in the path-integral treatment (next section). Correlation functions for negative \( L \) are obtained by symmetry through the relation

\[
g^{(0)}_{k_1,k_2}(L) = g^{(0)}_{k_2,k_1}(k_2 - k_1 - |L|), \quad L < 0,
\]

and intervals for negative \( L \) are labeled by applying the definition (6.2) to the quantity \( k_2 - k_1 + |L| \). Explicitly, for \( L < 0 \),

\[
g^{(0)}_{k_1,k_2}(L) = \begin{cases} 
(1 - [(k_1 + L)^2/k_1k_2])^{-3/2} & \text{for } 0 < -L < k_1 \quad \text{(region II)} \\
1 & \text{for } -L > k_1 \quad \text{(region III)}
\end{cases}
\]

and region I is absent in the range of negative \( L \). When the two contact orders are equal \((k_1 = k_2 = k)\), region I is absent altogether, and

\[
g^{(0)}_{k,k}(L) = \begin{cases} 
(1 - [(k - |L|)/k]^{-2} & \text{for } 0 < |L| < k \quad \text{(region II)} \\
1 & \text{for } |L| > k \quad \text{(region III)}
\end{cases}
\]

It is not surprising that the random-flight correlation \( g^{(0)} \) is completely independent of the two tail lengths \( l_0 \) and \( l_0' \), since there is no excluded volume interaction between the tails and the two loops in random-flight chains. The general behavior of \( g^{(0)} \) is shown in Figs. 11(a) and 11(b). Because of the symmetry relation (6.3), only the sketch for \( L > 0 \) needs to be shown. The division of the contact map by one presumed contact into regions I, II, and III is also shown in Fig. 11(c). The general behavior of \( g^{(0)} \) in three dimensions is almost identical to that in two dimensions. The only difference is that the plateau in region I is higher, and consequently the slope in region II is steeper, in three dimensions than in two dimensions. This is due simply to the fact that the absolute value of the exponent in Eqs. (6.2) and (6.5) is 3/2 (in 3D) instead of 1 (in 2D).

As expected, the correlation functions deviate from the JS random-flight predictions when excluded volume effects...
are taken into account. Typical results of the exhaustive lattice enumerations are shown in Fig. 12. In these plots, there are variations in region II (see Fig. 11) that are not predicted by JS. In particular, the g = 0 value at L = 1 in the first plot corresponds to an implied block, also not predicted by JS (see also Fig. 2). The plots also indicate that two loops are independent (g = 1) only when the separation L is substantially larger than the value $k_2$ (at the boundary between regions II and III) predicted by JS. Moreover, the lattice results always exhibit a local minimum at $L = k_2$, a feature also observed in two dimensions. This can be explained in terms of excluded-volume "repulsion" between the two loops.  

In spite of these differences, the JS model is successful in providing qualitatively correct predictions, particularly the enhancement when one loop is completely within another (region I). The discrepancy between the JS model and the excluded-volume lattice simulations is smaller in three dimensions than in two dimensions, simply because of the diminished importance of excluded volume effects in higher spatial dimensions. One striking difference between two and three dimensions is the absence of implied blocks ($g = 0$) for all contacts with order $k > 5$. Implied blocks are restricted to region II. The absence of blocks in three dimensions leads to the reduced amplitudes of variation of the lattice correlation functions compared to that in two dimensions. In fact, as shown in Fig. 12, the general trend of the JS prediction follows approximately the three-dimensional lattice results in most cases.

The most striking conclusion from these excluded-volume correlation shown in Fig. 12 is that certain internal architectures are preferred within the two-constraint ensemble. For example, the maximum at $L = 2$ of $g_{3,3}(L)$ corresponds to a helical configuration, of type (ii), (iii), and (iv) (see Fig. 9). The maximum at $L = 1$ of $g_{5,5}(L)$ includes the type (i) helix in Fig. 9; and all the peaks in $g_{3,2}(L)$, $g_{5,2}(L)$, and $g_{7,2}(L)$ are contact patterns that correspond to antiparallel sheets.

Two-contact correlation can also be studied by employing topological free energy surfaces, constructed as follows. Suppose contact $(i_1, j_1)$ is given, indicated by the circular dot on the contact maps in Figs. 13 and 14. Then the conditional probability of forming a second contact $(i_2, j_2)$ in chains of length $N$ subject to the presumed contact is simply $R(N; i_1, j_1; i_2, j_2) / R(N; i_1, j_1)$. The natural logarithm of this quantity is a dimensionless free energy, in units of $kT$, and is plotted as contours for all possible positions $(i_2, j_2)$ on the contact map. The contours then represent the relative probabilities of any second contact for the given first contact. In this representation, the two contact "interaction" or correlation is depicted by a potential surface, analogous to the field description of particle-particle interactions in classical physics. The preference for the formation of internal architectures is shown quite clearly in this representation.

Figures 13 and 14 show two such topological free energy surfaces for chains with $N + 1 = 14$ monomers on simple cubic lattices. In Fig. 13, the presumed contact is of order 5 at (5,10). In Fig. 14, the presumed contact is of order 7 at (4,11). The most important features of these free energy surfaces are the locations of the deepest minima, which identify the most probable second loop given a specific presumed first loop. From both plots, it is clear that the most probable second loop is that which is in the antiparallel sheet configuration relative to the first loop. In Fig. 13, the most probable second contact is that which fills in the turn of the antiparallel sheet, (6,9); it is approximately 2.7$kT$ more favorable than the least favored contact. In Fig. 14, the most favored second contact is one that continues the antiparallel sheet, at

**FIG. 12.** Correlation of two presumed contacts: The first three plots are correlations between contacts of the same order. The last three plots are correlations between contacts of different orders. Square dots represent lattice simulations. Continuous lines without dots show random-flight predictions $g^{(2)}$. 

numbers of constraints. Since similar results are obtained on
two-dimensional square lattices, it appears that these
conclusions are not artifacts of a particular lattice. Hence
simply excluded volume and conformational freedom are
significant driving forces for the formation of "secondary"
and "tertiary" structures observed in globular proteins,
which are chains with the maximum or near-maximum
number of self-contacts.

Figures 15 and 16 are the random-flight predictions for
comparison with Figs. 13 and 14, respectively. As discussed
above in conjunction with correlation functions in Fig. 12,
the discrepancies between the excluded-volume and ran-
dom-flight results are smaller in three dimensions than in
two dimensions. It is clear from Figs. 15 and 16 that the
random-flight theory is successful in predicting that con-
tacts required for antiparallel sheet formation are among the
most favored contacts, but random-flight theory fails to sin-
gle out the antiparallel sheet contact as the most favored
contact in each contact map. Specifically, in Fig. 15, contacts
(5,8) and (7,10) have the same formation probability as the
antiparallel sheet contact (6,9). Similarly in Fig. 16, con-
tacts (4,9) and (6,11) also have the same formation proba-
bility as the antiparallel sheet contact (5,10). This failure is
related to the constant value of $g^{(0)}$ in region I predicted by
random-flight theory (see Fig. 11). Lattice chains with ex-
cluded volume, on the other hand, always exhibit a peak in g
at $(k_{1} - k_{2})/2$ in the middle of region I (see Fig. 12), a
feature not reproduced by random-flight considerations. We
shall demonstrate in the next section, through use of the
path-integral approach, that the existence of these peaks is
indeed a general consequence of excluded volume.

There is one notable distinction between lattice simul-
ation contact maps on the one hand, and random-flight and
path-integral contact maps on the other. One unphysical
artifact of the random-flight and path-integral models is that
they allow formation of arbitrarily small loops with arbitrar-
VII. CORRELATION BETWEEN TWO CONTACTS: THE PATH- INTEGRAL MODEL

We now apply the path-integral method to the treatment of chains with two presumed contacts. As noted in the last section, there are three different ways in which two contacts can be formed along the chain (Fig. 10). These three basic topologies (labeled I, II, and III) are shown in Fig. 17. The corresponding definitions for these variables and their interrelationships are given in Table II for two contacts of order $k_1$ and $k_2$.

Following the same procedure of diagrammatic expansion as in Ref. 8, the $O(v_0)$ term $h(N;i_1;i_2;j_2)$ in the three-dimensional $(d = 3)$ two-contact reduction factor

$$R(N;i_1;i_2;j_2) = R^{(0)}(i_1;i_2;j_2) \times \left[ 1 - \frac{v_0}{(2\pi)^{3/2}} h(N;i_1;i_2;j_2) + O(v_0^2) \right]$$

is computed from the contributions of 45 Feynman-type diagrams, reproduced here in Fig. 18. The reduction factor (7.1) takes three different forms for the three basic structures that correspond to regions I–III in Fig. 17. In three dimensions, the zeroth-order two-contact reduction factor follows from random-flight statistics,\textsuperscript{8,27}

$$R^{(0)}(i_1;i_2;j_2;j_3) = \begin{cases} \left[(\Delta l)^2[3/(2\pi)]^3[l_2(l_1 + l_3)]^{-3/2} \right. & \text{in region I} \\ \left[(\Delta l)^2[3/(2\pi)]^3q^{-3/2} \right. & \text{in region II} \\ \left[(\Delta l)^2[3/(2\pi)]^3[k_1k_3]^{-3/2} \right. & \text{in region III} \end{cases}$$

where

$$q \equiv l_1l_2 + l_4l_3 + l_1l_4.$$  

The first order correction $h(N;i_1;i_2;j_2)$ is computed by collecting contributions from three separate groups of 15 diagrams in the three regions,

$$h(N;i_1;i_2;j_2) = \begin{cases} h_1(N;i_1;i_2;j_2) & \text{in region I} \\ h_2(N;i_1;i_2;j_2) & \text{in region II} \\ h_III(N;i_1;i_2;j_2) & \text{in region III} \end{cases}$$

Applying diagrammatic rules in three dimensions yields

$$h_1(N;i_1;i_2;j_2) = 4\pi \sqrt{\frac{l_2(l_1 + l_3)}{l_1 + l_2 + l_3}} + \sqrt{N} - \sqrt{l_6 + l_6^0} + \sqrt{l_1 + l_3} \left( \cos^{-1} \sqrt{\frac{l_1 + l_3}{l_1 + l_2 + 4l_6}} \right)$$

$$+ \cos^{-1} \sqrt{\frac{l_1 + l_3}{l_1 + l_2 + 4l_6^0}} + \sqrt{l_2} \left( 2\sin^{-1} \sqrt{\frac{l_2(l_1 + l_3)}{l_2(l_1 + l_3) + 4l_6^0l_3}} \right) - \sin^{-1} \sqrt{\frac{l_2(l_1 + l_3)}{l_2 + 4l_6^0}(l_1 + l_3 + 4l_6l_3)} \right]$$

$$(7.5a)$$
\[ h_{II}(N; i_1, j_1; i_2, j_2) = 4 \left[ \sqrt{N} + \frac{\sqrt{l_0 + \sqrt{l_0 + l_1}}}{q} + \frac{\sqrt{l_0 + \sqrt{l_0 + l_1}}}{q} - \sqrt{\frac{l_0}{q} + \sqrt{l_0 + l_1}} - \sqrt{l_0 + \frac{l_1}{q} + \sqrt{l_0 + l_1} - 4 \sqrt{\frac{l_0 l_1}{q}} \right] + \left[ \frac{2}{l_2 + l_3} \sin^{-1} \left( \frac{l_2 l_3 - l_1 (l_2 + l_3)}{\sqrt{q + 4l_0 (l_2 + l_3)}} \right) + \sin^{-1} \left( \frac{l_2 l_3 - l_1 (l_2 + l_3)}{\sqrt{q + 4l_0 (l_2 + l_3)}} \right) \right] \]

\[ - 2 \sin^{-1} \left( \frac{l_2 l_3 - l_1 (l_2 + l_3)}{\sqrt{q + 4l_0 (l_2 + l_3)}} \right) \cos^{-1} \left( 1 - \frac{8l_1 l_2 l_3 (l_2 + l_3)}{q^2} \right) + \left( \frac{l_2 l_3}{l_2 + l_3} \right) \left( 2 \pi - \cos^{-1} \left( 1 - \frac{8l_1 l_2 l_3 (l_2 + l_3)}{q^2} \right) \right) \right] + \{i_1 \leftrightarrow i_2 \} + \{i_1 \leftrightarrow i_3 \}, \quad (7.5b) \]

\[ h_{III}(N; i_1, j_1; i_2, j_2) = 4 \left[ \sqrt{N} - \frac{\sqrt{l_0 - l_0 + l_1}}{q} + \sqrt{k_1 \left( \cos^{-1} \left( \frac{k_1}{k_1 + 4l_0} \right) + \cos^{-1} \left( \frac{k_1}{k_1 + 4l_0} \right) \right) \right] \]

\[ + \sqrt{k_2 \left( \cos^{-1} \left( \frac{k_2}{k_2 + 4l_0} \right) + \cos^{-1} \left( \frac{k_2}{k_2 + 4l_0} \right) \right) \right] \]

\[ + \frac{k_1 k_2}{k_1 + k_2 + 4l_0} \cos^{-1} \left( 1 - \frac{2k_1 k_2}{(k_1 + 4l_0)(k_2 + 4l_0)} \right) \right], \quad (7.5c) \]

where the notation \( \{i_1 \leftrightarrow i_2 \} \) and \( \{i_1 \leftrightarrow i_3 \} \) in \( h_{II} \) denotes terms that result from interchanging \( i_1, i_2 \) and \( i_1, i_3 \) in the preceding expressions enclosed in curly brackets. The ranges for the functions \( \sin^{-1} \) and \( \cos^{-1} \) are \([-\pi/2, \pi/2]\) and \([0, \pi]\), respectively, and the \( \theta \) function in \( h_{II} \) is defined as

\[ \theta(x) = \begin{cases} 
1 & \text{for } x > 0 \\
0 & \text{for } x < 0. \\
1/2 & \text{for } x = 0 \end{cases} \quad (7.6) \]

Although three different functions represent regions I–III in the path-integral prediction of the reduction factor \( (7.1) \), it is straightforward to check that there is no discontinuity at any boundary among these three regions. The three-dimensional results \( (7.5) \) are simpler than the corresponding results in two dimensions in that all integrals can be performed analytically and the final results \( (7.5) \) are completely expressible in terms of elementary functions, while in two dimensions some integrals cannot be so expressed.\(^8\)

We now compare the path-integral results, Eqs. \((7.1)-(7.5)\), with the lattice enumerations for prediction of the conditional probabilities of formation of a second contact given that the first contact has been formed. Figure 19 shows the topological free energy surface for a chain with \( N + 1 = 14 \) monomers and an order 7 presumed contact at \((4,11)\); three different excluded-volume interaction

![FIG. 17. Basic structures for chains with two presumed contacts. Definitions for the contour lengths of various chain sections are given in Table II.](image1)

![FIG. 18. The Feynman-type diagrams that contribute to the first-order correction of two-contact reduction factors.](image2)
\[ g_{k_1,k_2}(L;l_0,l'_0) = g^{(0)}_{k_1,k_2}(L) \left[ \frac{1 - \frac{v_0}{(2\pi)^{3/2}}} {2N!i_1i_2j_1j_2} \left\{ f(Ni_1i_2j_1j_2) \right\} + O(v_0^2) \right]. \]

(7.7)

where the contact orders \( k_1 \equiv i_1 - i_0 \) and \( k_2 \equiv j_2 - i_2 \), and the separation \( L \equiv j_1 - i_1 \) are defined in Fig. 17 and Table II. \( g^{(0)} \) is the zeroth-order (random-flight) correlation given in Eqs. (6.2)-(6.5), and is independent of the two tail lengths \( l_0 \) and \( l'_0 \). The correlation \( g \) for chains with excluded volume depends on \( l_0 \) and \( l'_0 \), but the dependence is weak, especially for two contacts located in the middle of long chains \( (l_0,l'_0 \to \infty) \). Figure 20 shows the correlation functions predicted by path-integral theory (7.7) compared with the lattice enumerations. Representative results of \( g_{5,5}, g_{5,6}, \) and \( g_{6,7} \) are shown. In these plots, the path-integral calculations are carried out for contacts located in the middle of sufficiently long chains \( (N = 300) \), so that the dependence on \( l_0 \) and \( l'_0 \) is negligible and \( g_{k_1,k_2} \) may be approximated as a function of \( L \) alone. The corresponding lattice results are obtained from enumerations of chains with lengths as long as computationally feasible: \( N + 1 = 17 \) residues. A value of \( v_0 = 0.2 - 0.3 \) gives a reasonable reproduction of the general profile of the lattice simulations. The path-integral correlation functions always reach a local minimum at \( L = k_2 \) at the boundary between region II and III, and a local maximum at \( L = (k_2 - k_1)/2 \) at the middle of region I, which are features that also appear in lattice simulations, and in two dimensions. More detailed analysis of the path-integral correlation (7.7) demonstrates that the existence of a \( L = k_2 \) minimum and a \( L = (k_2 - k_1)/2 \) maximum are not limited to these few specific examples; instead, they are universally valid for all contact orders. For example, in the infinite tail length limit \( (l_0,l'_0 \to \infty) \), the path-integral correlation in region III \( (L > k_2) \) is given by

\[
\lim_{l_0,l'_0 \to \infty} g_{k_1,k_2}(L;l_0,l'_0) = 1 - \frac{4v_0}{(2\pi)^{3/2}} \sqrt{\frac{k_1 k_2}{k_1 + k_2 + 4l}} \times \cos^{-1} \left[ 1 - \frac{2k_1 k_2}{(k_1 + 4l)(k_2 + 4l)} \right] + O(v_0^2),
\]

(7.8)

where \( l = L - k > 0 \). The expression (7.8) approaches unity as \( L \to \infty \), and decreases with decreasing \( L \), reaching the minimum

\[
1 - v_0 \sqrt{\frac{2k_1 k_2}{\pi (k_1 + k_2)}} + O(v_0^2)
\]

at \( L = k_2 \). It can also be demonstrated by differentiating the infinite-tail-length path-integral correlation that there is always a local maximum at \( L = (k_2 - k_1)/2 \) at the middle of region I. Hence these features appear to be general properties of excluded volume, and are not restricted to a particular

---

TABLE II. Definitions of the separation \( L \) between two contacts and the contour lengths of various chain sections for the three basic structures in Fig. 17.

<table>
<thead>
<tr>
<th>Region I</th>
<th>Region II</th>
<th>Region III</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0 &lt; L &lt; k_2 - k_1 )</td>
<td>( k_1 - k_0 &lt; L &lt; k_2 )</td>
<td>( L &gt; k_2 )</td>
</tr>
<tr>
<td>( l_0 = i_0 \to i_0 )</td>
<td>( i_0 = i_1 \to i_2 )</td>
<td>( i_0 = i_1 \to i_2 )</td>
</tr>
<tr>
<td>( l'_0 = N - j_0 \to j_0 )</td>
<td>( N - j_1 \to j_0 )</td>
<td>( N - j_0 \to j_0 )</td>
</tr>
<tr>
<td>( l_1 = j_0 - i_0 \to i_1 - L )</td>
<td>( j_1 - j_0 \to k_1 - k_2 + L )</td>
<td>( k_1 - j_0 \to k_1 - i_1 )</td>
</tr>
<tr>
<td>( l_2 = j_0 - j_1 \to j_1 - L )</td>
<td>( j_1 - j_0 \to k_1 - L )</td>
<td>( k_1 - j_0 \to k_1 - i_1 )</td>
</tr>
</tbody>
</table>

---

lattice, nor a consequence of shortness of chains. From this behavior of the correlation function \( g \), we infer that: (i) Two contact pairs that share a common contacting monomer, such as \((3,8)\) and \((8,11)\), is a disfavored conformation. This follows from the existence of a minimum at \( L = k_2 \), and can be described as a physical "repulsion" between two close loops. \(^8\) (ii) The antiparallel sheet structure, corresponding to \( L = (k_2 - k_1)/2 \) in region I, is favored over other contacts in region I only when \( v_0 > 0 \). This shows clearly the importance of excluded-volume interactions in antiparallel sheet formation.

The principal weakness of the path integral method is its limited ability to account for short-ranged excluded volume. \(^7\) Consequently, there are quantitative discrepancies between the path-integral and lattice models. For example, the path integral fails to predict some of the variations in region II which are present in the lattice correlations. This discrepancy also occurs, and is larger, in two dimensions. \(^8\)

**VIII. MULTIPLE CHAIN CONTACTS AND COMPACT POLYMERS**

In this section, we consider chains that are constrained by more than two given self-contacts. First, we consider the case of three presumed contacts. Then we consider the properties of chain configurations as a function of the number of internal chain contacts, all the way up to the maximum number possible. Finally, we consider certain properties of chains in the maximally compact ensemble.

We study the three-contact problem by considering chains with two fixed presumed contacts. Subject to these two fixed constraints, the formation probability of any third contact is represented by a topological free energy surface, constructed by plotting the contours of the three-contact reduction factor \( R(N; i_1, j_1; i_2, j_2; i_3, j_3) \) with fixed \((i_2, j_2)\) and \((i_3, j_3)\) on the contact map. Figure 21 shows such a free energy surface with two order 3 presumed contacts \((5,8)\) and \((7,10)\) in a helical pattern. The contours show that the most favored contacts are those associated with extension of the helix: the order 3 contacts \((3,6)\) and \((9,12)\) extend a type (iii) or (iv) helix, while the order 5 contacts \((4,9)\) and \((6,11)\) form part of the type (ii) helical pattern, see Fig. 9.

These contacts are favored by approximately \(2.5kT\) compared to the least favored possible contact. Due to end effects, order 3 contacts at chain ends, \((1,4)\) and \((11,14)\), are also favored by approximately the same free energy. Hence given two contacts, the most favored third contact in this case is the one which continues formation of helices and sheets. Thus the entropic forces that favor the growth of secondary structures from 1-constraint chains are also important for 2-constraint chains.

As a chain acquires more intrachain contacts, it neces-

---

**TABLE III.** Number of conformations \( Q^{(t)}(N) \) on the simple cubic lattice as a function of number of contacts \( t \) and chain length \( N \).

<table>
<thead>
<tr>
<th>( N )</th>
<th>( t = 0 )</th>
<th>( t = 1 )</th>
<th>( t = 2 )</th>
<th>( t = 3 )</th>
<th>( t = 4 )</th>
<th>( t = 5 )</th>
<th>( t = 6 )</th>
<th>( t = 7 )</th>
<th>( t = 8 )</th>
<th>( t = 9 )</th>
<th>( t = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>150</td>
<td>126</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>726</td>
<td>534</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3,534</td>
<td>2,214</td>
<td>1,032</td>
<td>288</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>16,926</td>
<td>9,246</td>
<td>5,376</td>
<td>2,112</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>81,390</td>
<td>38,142</td>
<td>26,688</td>
<td>11,928</td>
<td>4,488</td>
<td>0</td>
<td>144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>387,966</td>
<td>157,974</td>
<td>128,880</td>
<td>66,192</td>
<td>29,424</td>
<td>4,632</td>
<td>864</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1,853,886</td>
<td>649,086</td>
<td>605,664</td>
<td>353,544</td>
<td>168,504</td>
<td>65,472</td>
<td>7,680</td>
<td>3,936</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8,809,878</td>
<td>2,675,022</td>
<td>2,802,576</td>
<td>1,817,208</td>
<td>956,832</td>
<td>412,416</td>
<td>116,928</td>
<td>24,288</td>
<td>4,608</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>41,934,150</td>
<td>10,966,470</td>
<td>12,725,136</td>
<td>9,092,592</td>
<td>5,269,920</td>
<td>2,435,256</td>
<td>1,059,120</td>
<td>257,880</td>
<td>94,272</td>
<td>0</td>
<td>3,504</td>
</tr>
<tr>
<td>12</td>
<td>198,842,742</td>
<td>45,054,630</td>
<td>57,525,552</td>
<td>44,547,912</td>
<td>27,853,080</td>
<td>16,693,312</td>
<td>6,441,456</td>
<td>2,418,264</td>
<td>651,648</td>
<td>161,856</td>
<td>19,104</td>
</tr>
<tr>
<td>13</td>
<td>943,974,510</td>
<td>184,400,910</td>
<td>256,574,352</td>
<td>214,532,136</td>
<td>143,817,120</td>
<td>79,878,144</td>
<td>38,486,304</td>
<td>17,723,064</td>
<td>6,191,616</td>
<td>1,957,344</td>
<td>289,392</td>
</tr>
</tbody>
</table>
three-dimensional simple cubic lattice, the maximum number of topological contacts \( t_{\text{max}} \) is determined for chains with \( N \) bonds (\( N+1 \) monomers) as follows. First it is easy to see that the lattice chain has a surface area \( 2(2N+3) \) if it does not make any intrachain contacts. Then for every contact made, the total surface area is reduced by 2. Therefore, if \( A_{\text{min}}(N) \) is the minimum surface area attainable by a chain with \( N \) bonds (\( N+1 \) monomers),

\[
t_{\text{max}} = \frac{1}{2} \left[ 2(2N+3) - A_{\text{min}}(N) \right] = 2N+3 - \frac{A_{\text{min}}(N)}{2} .
\]

(8.3)

On the simple cubic lattice,

\[
A_{\text{min}}(N) = \begin{cases} 
6m^2 + P_{\text{min}}(N-m^3) & \text{for } m^2 < N+1 < m^2(m+1) \\
2(3m^2 + 2) + P_{\text{min}}(N-m^2(m+1)) & \text{for } m^2(m+1) < N+1 < m^2(m+1)^2 \\
2(m+1)(3m+1)+P_{\text{min}}(N-m^2(m+1)^2) & \text{for } m(m+1)^2 < N+1 < m(m+1)^3 
\end{cases}
\]

(8.4)

where \( m \) is a positive integer. \( P_{\text{min}}(N') \) is the minimum perimeter of a chain with \( N' \) bond (\( N'+1 \) monomers) on two-dimensional square lattices, given by

\[
P_{\text{min}}(N') = \begin{cases} 
2(2m'+1) & \text{for } m^2 < N'+1 < m(m'+1) \\
4(m'+1) & \text{for } m(m'+1) < N'+1 < (m'+1)^2 
\end{cases}
\]

(8.5)

The maximum number of contacts \( t_{\text{max}} \) is then obtained by substituting Eq. (8.4) into Eq. (8.3).

It is generally supposed that the number of maximally compact conformations \( Q_c(N) \equiv Q_{(t_{\text{max}})}(N) \) increases approximately exponentially with \( N \), i.e., \( Q_c(N) \sim e^{\kappa N} \). The connective constant \( \kappa \) for compact chains describes the conformational freedom per chain segment in compact configurations. Since the total number of all accessible configurations also has a near-exponential behavior [see Eq. (2.1)], the quantity \( \kappa \) is important inasmuch as \( \ln(\kappa/\mu) \) is the entropy loss per segment due to compactness, relative to the freedom of open configurations. In a recent study in two dimensions, \( \kappa \) was discovered that the behavior of the number of \( \rho = 1 \) compact chains depends not only on the chain length \( N' \), but also on the number of overall shapes \( S \) that are possible. It is the number of compact conformations per shape that exhibits a general exponential trend, with oscillations related to the "articulation" of the shapes. Similar behavior in three dimensions is expected, such that the number of maximally compact conformations per shape \( Q_c(N)/S \) oscillates between two bounding exponentials

\[
c_1 \kappa^{N+1} < Q_c(N)/S < c_2 \kappa^{N+1}
\]

(8.6)

with the same connective constant \( \kappa \). As in two dimensions, \( c_1 \) and \( c_2 \) are determined by considering a special class of number of monomers \( N+1 \) referred to as "magic numbers." On three-dimensional simple cubic lattices, the magic numbers are \( N = 1 \Rightarrow 8, 12, 18, 27, 36, 48, 64, \ldots \), etc., corresponding to overall shapes of perfect cubes (for
example, 27 = 3 × 3 × 3), or near-perfect cubes, i.e., square-based cylinders wherein the length of one side of the base differs from the height of the cylinder by one unit (for example, 48 = 3 × 4 × 4). These shapes are useful in determining the constants in the upper bound \( c_2 \kappa^{N + 1} \) of Eq. (4.6) because they are least articulated, and hence allow maximum conformational freedom for compact chains. In addition to the data given in Table III, we have also enumerated \( Q_c(N) \) for \( N + 1 = 18 \) and 27. Both \( Q_c(N) \) and the shape-averaged \( Q_c(N)/S \) for the magic numbers \( N + 1 = 8, 12, 18, \) and 27 are listed in Table IV. Note that \( S = 1 \) for the perfect cubes, but \( S = 3 \) for the near-perfect cubes because each near-perfect cube has three possible orientations in space.

Figure 22 shows the plot of the logarithm of magic-numbered \( Q_c(N)/S \) as a function of the number of monomers \( N + 1 \). The simulations are well characterized by

\[
Q_c(N)/S \approx c_2 \kappa^{N + 1} \quad \text{for} \quad N + 1 = \text{magic number}
\]

(8.7)

for which our best estimate is \( c_2 \approx 1.655 \) and \( \kappa \approx 1.736 \). Our estimate value for \( \kappa \) for simple cubic lattices is within the bounds

\[
1.55 < \kappa < 2.22
\]

(8.8)

given by Orr\(^{29}\) in 1947, who obtained Eq. (8.8) by utilizing even earlier results of Chang\(^{32}\) and Miller.\(^{33}\) However, our lattice estimate \( \kappa \approx 1.736 \) is smaller than the value \( \kappa \approx 1.90 \) derived by Orr, who simply took the arithmetic mean of the lower and upper bounds in Eq. (8.8). The \( \kappa \) from our lattice enumeration is also smaller than that predicted by both the Flory\(^{34}\) and Huggins\(^{35}\) approximations, which give \( \kappa = (z - 1)/\epsilon = 1.8394 \) and \( \kappa = (z - 1)(1 - 2/\epsilon)^{1/2} - 1 = 2.222 \), respectively, for simple cubic lattices with coordination number \( z = 6 \). It is interesting to note that the Flory approximation has been criticized\(^{36}\) for underestimating \( \kappa \) in two dimensions. (See also Ref. 37). However, the above comparison shows that in fact

**Table IV. Number of magic-numbered compact (\( \rho = 1 \)) conformations \( Q_c(N) \) on three-dimensional simple cubic lattices. \( N + 1 \) is the number of residues and \( S \) is the number of compact shapes.**

<table>
<thead>
<tr>
<th>( N + 1 )</th>
<th>( Q_c(N) )</th>
<th>( S )</th>
<th>( Q_c(N)/S )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>144</td>
<td>1</td>
<td>144</td>
</tr>
<tr>
<td>12</td>
<td>3504</td>
<td>3</td>
<td>1168</td>
</tr>
<tr>
<td>18</td>
<td>101280</td>
<td>3</td>
<td>33760</td>
</tr>
<tr>
<td>27</td>
<td>4960608</td>
<td>1</td>
<td>4960608</td>
</tr>
</tbody>
</table>

**FIG. 23. Topological free energy surface of contact formation in compact chains with 27 monomers on simple cubic lattices. The surface is remarkably flat except for end effects. Contours are given in 0.3\( k_B \)T steps.**

**FIG. 24. Topological free energy surface given an order 3 presumed contact in maximally compact simple cubic lattice chains with 27 monomers.**
both Flory and Huggins approximations overestimate $\kappa$ in three dimensions, at least for simple cubic lattices, indicating that the severity of the packing constraints in three dimensions is underestimated by these approximation schemes. Moreover, it is clear that the simpler Flory approximation is better than the Huggins approximation in three dimensions.

Cyclization is nearly ideal in compact chains. Figure 23 shows the contact formation probability for any pair $(i,j)$ in the ensemble of maximally compact chain of length $N + 1 = 27$ monomers on the three-dimensional simple cubic lattice. The most striking feature of the surface is its remarkable flatness. Although flattening of the topological free energy surface by compactness has also been observed in two dimensions,\textsuperscript{28} the surface in three dimensions is flatter than the corresponding surface in two dimensions. In Fig. 23, the only significant variation from the near-flat area that covers most of the contact map are contacts along the lowest order ($k = 3$) diagonal and those at chain ends. Conformational complexity imposed by chain compactness and connectivity notwithstanding, this exact lattice result demonstrates clearly that the Bragg–Williams mean-field approximation is quite good for computing pairwise segment probabilities in the ensemble of compact chains.

Finally, correlations between two loops can be computed for the ensemble of maximally compact chains, as they have been computed above for the full ensemble. The topological free energy surfaces with one presumed contact of order $k = 3, 5, 7$ in compact cubic-lattice chains with 27 monomers are shown in Figs. 24, 25 and 26, respectively. Since these contacts are not located exactly at the middle of the chain, these surfaces do not have mirror symmetry along the perpendicular bisection of the main diagonal. Nevertheless, these surfaces clearly demonstrate that the most favored contacts are those corresponding to secondary structural patterns. For example, the contacts (10, 13) and (14, 17) in Fig. 24 correspond to helices, while the contact (12, 15) in Fig. 25 and (11, 16) in Fig. 26 correspond to antiparallel sheets. These contacts are favored by 1.2–1.8$kT$ when compared to the least probable contacts, showing that secondary structural motifs are favored by conformational freedom and excluded volume in maximally compact chains, just as they are in the more complete ensemble of all conformations.

**IX. CONCLUSIONS**

We have studied the accessible conformations of chain molecules in three dimensions through the use of exhaustive simulations on simple cubic lattices, and through the use of the path integral method. In particular, we have considered chains constrained by one given intrachain contact between monomers $(i,j)$ or constrained by two, or three, or the maximum possible number of intrachain contacts. For a single such contact, the present results give the first-order excluded-volume correction to the Jacobson–Stockmayer theory, and are in agreement with the first and second order perturbation predictions of the des Cloizeaux–de Gennes approach. We present a method of topological correlation functions and topological free energy surfaces with which we can explore the “interactions” among loops, the degree to which one intrachain contact favors or hinders another. The principal result is that certain features of internal architecture are favored within the two-constraint ensemble of flexible chain molecules, simply by conformational freedom and excluded volume. These features are helices and antiparallel sheets, the principal components of globular proteins. We find that the same architectures are favored in multiple-constraint ensembles. Hence individual polymer molecules may be driven to the formation of internal architectures by any monomer–monomer attraction which causes multiple self-contacts. These internal architectures may be observable for isolated polymer molecules in very poor solvents. We find that loop correlations for chains with two presumed contacts can be well predicted by the Jacobson–Stockmayer theory.
In highly constrained ensembles, such as that of chains which are maximally compact, we find a remarkable ideality of conformational properties: (i) pairwise contacts occur with a frequency largely independent of their separation along the chain, and (ii) the Flory approximation is better than the Huggins approximation at predicting the number of configurations of maximally compact polymers in three dimensions.

ACKNOWLEDGMENTS

We thank the NIH, the URI Program of DARPA, and the Pew Scholars Program in the Biomedical Sciences for research support.

3. C. R. Cantor and P. R. Schimmel, Biophysical Chemistry (Freeman, San Francisco, 1980).
14. Note that the counting convention of $Q$ is different from that of the quantity $\Omega$ we used in an earlier work (Ref. 8). If $\Omega_0^c$ and $\Omega_0^c$ are the total number of accessible conformations enumerated according to the convention that only conformations that are not related by rigid rotations and reflections are considered to be distinguishable, then $\Omega_0^c$ on the cubic lattice is given by $Q_0 = 48\Omega_0^c - 24\Omega_0^c - 18$. Here $\Omega_0^c$ is equivalent to the quantity $\Omega_0$ used in Ref. 8.
15. Note that the conformational enumeration convention of $Q$'s used in the present work is different from the $\Omega$'s used in Ref. 8. However, for reasonably long chains, $Q$ is approximately proportional to $\Omega$, hence reduction factors calculated from ratios of $Q$'s only differ negligibly from those calculated from ratios of $\Omega$. In fact, the two definitions give identical reduction factors in the $N \to \infty$ limit, see Ref. 14. $Q$ is used here because the enumeration convention of $Q$ is one that is consistent with the path integral approach—conformations related by rigid rotations and reflections (other than the identity) are considered to be distinguishable in the path integral (4.1) and (4.2).
19. The notation $v$ for the exponent in $v^{-k}$ or $(k + 1)^{-1}$ is chosen here for simplicity. The $v$'s in this paper should not be confused with the exponent in the average square end-to-end distance, $R_s^2 \sim N^{v}$, a notation commonly employed in the polymer literature. In terms of the $v$ for $R_s^2$, which is estimated to be approximately 0.588 in 3D (J. L. Guillel and J. Zinn-Justin, Phys. Rev. Lett. 39, 95 (1977)), the exponent in the power-law dependence of ring formation probabilities on loop size $k$ for $k + 1$ is sometimes expressed as $(d + \theta)v$, where $d$ is the spatial dimension (Refs. 22 and 24).
21. Hiley and Sykes (Ref. 17) provide values of $Q(k, k + 1)$ for $k + 1 < 14$ on the simple cubic lattice, as coefficient of $x^k$ in their $U(x)$. Our enumeration is in complete agreement with theirs for $k + 1 < 14$, in addition, we obtain the $k + 1 = 16$ value for $Q(k, k + 1)$, which is equal to 218 904 768. Note that the quantity $p_{k+1}$, in Rushbrooke and Eve (Ref. 16) represents the number of distinct polygons, thus it is related to $Q$ by $Q(k, k + 1) = 2(k + 1)p_{k+1}$. Our data is also in complete agreement with Rushbrooke and Eve, who provided $p_{k+1}$ for $k + 1 < 14$.
23. If the $(k + 1)^{-1}$ convention is used instead of the $k^{-1}$ in Ref. 8, then the best value for $v_1$ in two dimensions is estimated to be 3.17, which is consistent with Redner's estimate. See also Ref. 21 of Ref. 8.
30. Orr does not distinguish the two ends of a chain in his conformational enumeration. Hence his coefficient of $\eta$ in $G(\eta)$ is equal to the quantity $Q^{1/2}/\eta$ in the present work.