(9) Soda, K.; Wada, A. *Biophys. Chem.* **1984**, *20*, 185.
(10) Lee, W. I.; Schmitz, K. S.; Lin, S.-C.; Schurr, J. M. *Biopolymers* **1977**, *16*, 583.
(11) Schurr, J. M.; Schmitz, K. S. *Annu. Rev. Phys. Chem.* **1986**, *37*, 271.
(12) Fulmer, A. W.; Benbasat, J. A.; Bloomfield, V. A. *Biopolymers* **1981**, *20*, 1147.
(13) Fried, M.; Crothers, D. M. *Nucleic Acids Res.* **1981**, *9*, 6505.
(14) Axelrod, D.; Koppel, D. E.; Schlessinger, J. S.; Webb, W. W. *Biophys. J.* **1976**, *16*, 1055.
(15) Elson, E. L.; Magde, D. *Biopolymers* **1974**, *13*, 1.
(16) Magde, D.; Elson, E. L.; Webb, W. W. *Biopolymers* **1974**, *13*, 29.
(17) Phillies, G. D. J. *Biopolymers* **1975**, *14*, 499.
(18) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976; pp 268–270.
(19) Doi, M.; Shimada, T.; Okano, K. *J. Chem. Phys.* **1988**, *88*, 4070.
(20) Semenov, A. N. *J. Chem. Soc., Faraday Trans. 2* **1986**, *82*, 317.
(21) Schaefer, D. W.; Han, C. C. In *Dynamic Light Scattering: Applications of Photon Correlation Spectroscopy*; Pecora, R., Ed.; Plenum Press: New York, 1985; pp 181–243.
(22) Odijk, T. *Macromolecules* **1979**, *12*, 688.
(23) de Gennes, P. G.; Leger, L. *Annu. Rev. Phys. Chem.* **1982**, *33*, 49.
(24) Tirrell, M. *Rubber Chem. Technol.* **1984**, *57*, 523.
(25) Doi, M.; Edwards, S. F. *J. Chem. Soc., Faraday Trans. 2* **1978**, *74*, 560.
(26) Doi, M.; Edwards, S. F. *J. Chem. Soc., Faraday Trans. 2* **1978**, *74*, 918.
(27) Graves, D. E.; Yielding, L. W.; Watkins, C. L.; Yielding, K. L. *Biochim. Biophys. Acta* **1977**, *479*, 98.
(28) Scalettar, B. A.; Selvin, P. R.; Axelrod, D.; Hearst, J. E.; Klein, M. P. *Biophys. J.* **1988**, *53*, 215.
(29) Icenogle, R. D.; Elson, E. L. *Biopolymers* **1983**, *22*, 1919, and companion paper.
(30) Marchand, P.; Marmet, L. *Rev. Sci. Instrum.* **1983**, *54*, 1034.
(31) Savitzky, A.; Golay, J. E. *Anal. Chem.* **1964**, *36*, 1627.
(32) IMSL Inc. *User's Manual IMSL Library* **1984**, *4*, ZXSSQ-1.
(33) Packard, B. S.; Saxton, M. J.; Bissell, M. J.; Klein, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 449.
(34) Sorcher, S. M.; Klein, M. P. *Rev. Sci. Instrum.* **1980**, *51*, 98.
(35) Thomas, J. C.; Allison, S. A.; Schurr, J. M. *Biopolymers* **1980**, *19*, 1451.
(36) Koppel, D. E.; Axelrod, D.; Schlessinger, J.; Elson, E. L.; Webb, W. W. *Biophys. J.* **1976**, *16*, 1315.
(37) Andries, C.; Guedens, W.; Clauwaert, J. *Biophys. J.* **1983**, *43*, 345.
(38) Batchelor, G. K. *J. Fluid Mech.* **1976**, *74*, 1.
(39) Southwick, J. G.; McDonnell, M. E.; Jamieson, A. M.; Blackwell, J. *Macromolecules* **1979**, *12*, 305.
(40) Koene, R. S.; Mandel, M. *Macromolecules* **1983**, *16*, 220.
(41) Muller, G. In *Physical Optics of Dynamic Phenomena and Processes in Macromolecular Systems*; Sedlacek, B., Ed.; Walter de Gruyter & Co.: New York, 1985; pp 117–127.
(42) Zero, K. M.; Pecora, R. *Macromolecules* **1982**, *15*, 87.
(43) Kubota, K.; Chu, B. *Biopolymers* **1983**, *22*, 1461.
(44) Russo, P. S.; Karasz, F. E.; Langley, K. H. *J. Chem. Phys.* **1984**, *80*, 5312.
(45) Keep, G. T.; Pecora, R. *Macromolecules* **1988**, *21*, 817.
(46) Callis, P. R.; Davidson, N.; *Biopolymers* **1969**, *8*, 379.

Compact Polymers

Hue Sun Chan and Ken A. Dill*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143. *Received February 17, 1989; Revised Manuscript Received April 25, 1989*

ABSTRACT: The compact conformations of polymers are important because they are the principal configurations of the native states of globular proteins. We study the compact polymeric state by exhaustive computer enumeration of short-chain configurations on two-dimensional square lattices. The number of accessible configurations is found to depend on two factors: (i) configurational freedom and excluded volume, extensively studied in the past, and (ii) a "shape" entropy of the compact object. In addition, we study the probability of loop formation in compact polymers and topological correlations among pairs of loops. We find that there is an intrinsic steric driving force for loops to be configured into helices and antiparallel sheets in compact chains, similar to that we have recently found for more open chains. Also, we study in detail the few conformations that remain after the enormous reduction of configurational space upon increasing the compactness. Those few conformations are overwhelmingly dominated by secondary structures: helices, antiparallel and parallel sheets, and turns. This dominance increases with chain length. It is found to be exceedingly difficult to configure a compact chain with less than about 50% secondary structure. This suggests that the driving force for formation of secondary structures in proteins may be nonspecific steric interactions rather than hydrogen-bonding or other specific interactions.

1. Introduction

The compact conformations of a chain molecule comprise a very small but important subset of all the physically accessible conformations. Their importance derives from the fact that the native conformations of all globular proteins are compact. By "compact", we refer to those configurations of single-chain molecules which are tightly packed, i.e., fully contained within a volume of space (a box) with the minimal (or near minimal) surface/volume ratio. In contrast, by "open", we refer to the complete superset of all accessible conformations, including those that are compact. Relatively little attention has previously been directed toward the set of compact conformations of polymers. Considerably more effort in polymer science has focused on the more open conformations of chains, because they are far greater in number and because they are the predominant conformations of chains in solution or in the bulk. Our purpose in this paper is to explore in some detail the nature of the compact conformations and to show how they differ from the larger superset of all possible conformations. Our purpose here is served by exhaustive simulation of every possible conformation of short chains on two-dimensional square lattices. There are three advantages of studies in two dimensions for our present purposes. First, certain predictions can be compared with a significant literature dating back to the work of Orr[1] in 1947 on exhaustive lattice simulations of open conformations, most of which have been in two dimensions. Second, we can explore greater chain lengths for a given amount of computer time. Third, the surface/volume ratio, a principal determinant of the driving force for a protein to

* To whom inquiries about this paper should be addressed.

0024-9297/89/2222-4559$01.50/0

collapse,[2] for long chains in three dimensions is more closely approximated by short chains in two dimensions than by equally short chains in three dimensions. Also, it has been shown recently that a two-dimensional square-lattice model of short chains predicts certain general features of protein behavior:[3] copolymers of sequences of H (nonpolar) and P (other) residues, subject to excluded volume, and with increasing HH attraction energy, will collapse to a relatively small number (often one) of maximally compact conformations in which there is a core of H residues.

In addition, one principal motivation of the present study is to explore the hypothesis[4] that excluded-volume forces may be responsible for the formation of certain specific forms of internal chain organization in compact polymers. For open chains subject to excluded volume, we have recently found that the most probable configurations of chains containing two intrachain contacts are helices and antiparallel sheets, the principal forms of internal chain organization (secondary structures) in globular proteins.[4] Here we explore in addition the hypothesis that excluded volume may also be a principal driving force for the formation of secondary structures in more compact chains and thus perhaps a driving force for the formation of secondary structures in globular proteins. Although the present study is limited to two dimensions, other work[5] shows similar behavior in three dimensions.

2. The Model

We begin with a summary of some definitions and terminology of the model used in our analysis. For brevity, only the essentials are included here, since further details can be found elsewhere.[4] In the two-dimensional square lattice with coordination number $z = 4$, chains are comprised of a sequence of monomers or residues, each occupying one lattice site. Residues are numbered sequentially from one end of the chain, starting with 1. Coordinates of the ith residue are given by the vector $\mathbf{r}_i$; bond lengths are normalized to unity, such that $\mathbf{r}_i = (n, m)$, where n and m are integers. Bond angles are limited to ±90° and 180°. Excluded volume is taken into account by forbidding two different residues from occupying the same lattice site; i.e., $\mathbf{r}_i \neq \mathbf{r}_j$ for $i \neq j$. For a chain with $N + 1$ residues, the chain length is defined to be the number of bonds in the chain, which equals N.

Two residues i and j are taken to be in contact when they are nearest neighbors on the lattice, $|\mathbf{r}_i - \mathbf{r}_j| = 1$. There are two types of contacts: the *connected contacts* are those between residues i and $i + 1$, implied by the connectivity of the chain, whereas *topological contacts* are made between residues not adjacent in the sequence. A topological contact is represented as an ordered pair (i, j), $j \neq i - 1, i, i + 1$. The *order* k of a topological contact is the number of bonds along the chain between the two contacting residues, given by $k = |j - i| > 1$. A convenient way to represent contacts is in terms of the *contact map*, a two-dimensional matrix of size $(N + 1) \times (N + 1)$, in which a dot at the ith row and jth column denotes the topological contact (i, j). Clearly the contact map is symmetric under row–column interchange; hence, only the upper triangular half needs to be shown; see Figure 1. The shaded main diagonal corresponds to the connected contacts, which are always implicit and thus are not represented with dots. Any line parallel to the main diagonal is termed a *diagonal*. Contacts of the same order lie along the same diagonal. The contact map characterizes the *topology* of a chain conformation, in contrast to the list of coordinates that characterizes the *geometry* of a chain conformation.

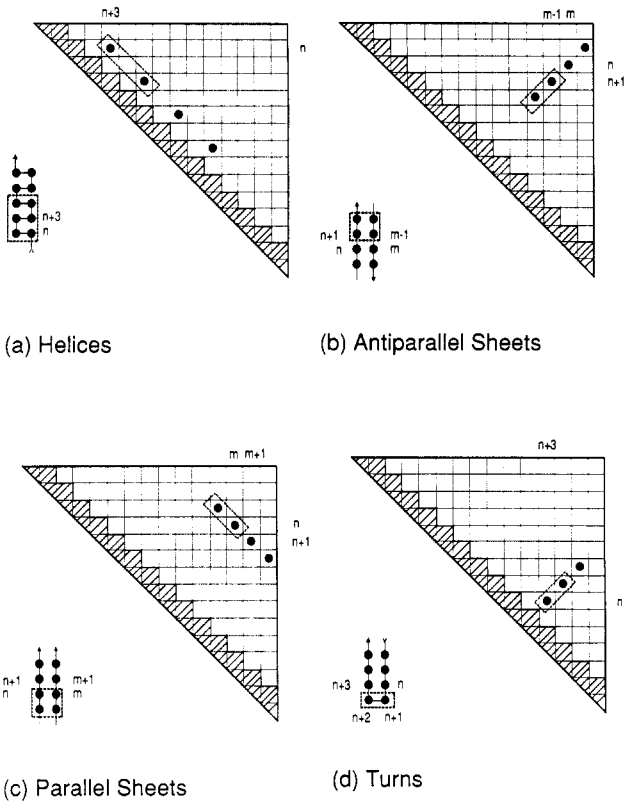

Figure 1. Secondary structures as contact patterns. The dotted boxes encircle minimal units that must be present to qualify as secondary structures. (a) Contacts for helices are along the order 3 diagonal. (b and d) Contacts for antiparallel sheets and turns form strings that are perpendicular to the main diagonal. (c) Contacts for parallel sheets form strings that are parallel to the main diagonal.

One of the principal virtues of the contact-map representation is its direct representation of secondary structures, which are defined by their topologies rather than by their geometries. The principal secondary structures observed in globular proteins—helices, parallel and antiparallel sheets, and turns—are all represented by very simple regular patterns on the contact map; see Figure 1. The patterns representing secondary structures are strings of dots.[6] Clearly, at least two contacts are needed to form a string; accordingly, the dotted boxes in parts a–c of Figure 1 encircle "minimal units" that must be present in order to be qualified as helices or sheets. Each of these minimal units consists of two dots on the contact map. In terms of chain geometry, the minimal units for helices have six residues, while those for sheets have four. There is no upper limit to the length of the strings of contacts for secondary structures, since secondary structures can be extended indefinitely. Hence, the number of residues participating in a helix may be 6, 8, 10, ..., while the corresponding numbers of residues in a sheet are 4, 6, 8, 10, ... As indicated by the dotted box in Figure 1d, there are always two residues at a turn, and they must be connected to an antiparallel sheet.

It is worthwhile to comment on the distinction between topological and geometric descriptions of secondary structures. In principle and in practice, helices can be readily represented either by their geometries or topologies. The topological representation entails specification of a particular pattern of spatial-neighboring, but nonconnected, residues along the chain. The geometric representation, on the other hand, entails specification of a particular region of acceptable ϕ–ψ angles among the connected backbone neighboring monomers. Either the

geometric or topological representation can be suitably used to identify helices. However, it is clearly much less sensible to represent sheets by specifying the geometries, the ϕ–ψ angles among connected neighboring residues, for the following reasons. Although ϕ–ψ angles can distinguish sheets from helices, they cannot so readily distinguish antiparallel and parallel sheets. This distinction is immediately clear, however, in the topological representation. Moreover, for long sheets of either type, specification in terms of local bond angles would require extreme precision and in so doing could not capture the essence of small twists and bends. It is our convinction therefore that what is meant by "sheet" must be more akin to the "train-track" nature of organization of strands of chain rather than some small class of bond angles among connected residues. Hence, for our present purposes, we adopt the view that secondary structures are defined in terms of spatially localized units of intrachain contacts rather than by precise details of the geometries of bond angles among connected neighbors.

It follows that lattice models can be expected to adequately represent secondary structures, even though they simplify the representation of bond angle geometries into a small number of discrete states. Even within geometric representations of secondary structures, quite wide latitude is generally taken in defining broad regions of ϕ–ψ angles of peptide backbones for classifying secondary structures.[7-10] The discretized conformations of lattice models probably do no worse than this. Moreover, topological properties are relatively independent of lattice type and spatial dimensionality. The intrinsic α-helical contact pattern represented by $(i, i + 3)$, for example, can be readily represented on several different types of lattices in two or three dimensions, including the two-dimensional square lattices studied here. In three-dimensional models, we obtain predictions that are qualitatively similar to those presented below for the two-dimensional square lattice.[5]

3. Enumeration of the Chain Conformations

In this work, we carry out exhaustive explorations of every accessible conformation of a polymer chain molecule on the two-dimensional square lattice, as a function of the chain "compactness". In later sections, we study the properties of these conformations; in the present section, we simply enumerate how many conformations there are at each compactness, as a function of the chain length. We first define variables and then summarize related past efforts on exhaustive explorations of chain conformations on lattices. We then devote considerable attention to the enumeration of conformations that are maximally compact.

A fundamental measure of conformational freedom is the total number of conformations, or states, that are accessible. For open chains with $N + 1$ residues, modeled here as unconstrained self-avoiding walks of N steps, this is given by $\Omega_0(N)$, which is the number of all such walks. Among these, some conformations will involve a certain number of topological self-contacts; hence, $\Omega_0(N)$ may be written as the sum

$$\Omega_0(N) = \sum_{t=0}^{t_{max}} \Omega^{(t)}(N) \tag{3.1}$$

where $\Omega^{(t)}(N)$ is the number of conformations with exactly t contacts and t_{max} is the maximum number of topological contacts that can be present in chains with $N + 1$ residues. $\Omega^{(t)}(N)$ measures conformational freedom as a function of the number of intrachain contacts. The compactness of the molecule is proportional to t, as described in more detail below, and is clearly related to the density of the molecule, the number of chain segments per unit volume.

In the two-dimensional square lattice, the *perimeter* of an isolated lattice site is 4; hence, the *maximum perimeter*, $P_{max}(N)$, of a chain that occupies $N + 1$ sites is $2(N + 2)$, when there are no intrachain contacts among the $N + 1$ residues. In general, the number of topological contacts t is related to the perimeter $P(N)$ of such chains by the equation

$$t = \frac{1}{2}[P_{max}(N) - P(N)] = N + 2 - \frac{P(N)}{2} \tag{3.2}$$

since formation of each contact decreases $P(N)$ by two units. The perimeter P serves as a two-dimensional analogue of the surface area exposed to solvent for a three-dimensional chain molecule in solution. When a lattice chain becomes more compact by forming more topological contacts, its perimeter necessarily decreases. Hence, the maximum number of topological contacts, t_{max}, may be computed from eq 3.2 by setting P equal to the minimum, or *compact perimeter* P_c. The only shapes that have minimum perimeter-to-area ratio on the square lattice are either squares or rectangles in which the lengths of the two sides differ by only one unit. Therefore for a chain of $N + 1$ residues confined within such shapes, it can easily be deduced that

$$P_c(N) = 2(2m + 1) \qquad \text{for } m^2 < N + 1 \le m(m + 1)$$

$$P_c(N) = 4(m + 1) \qquad \text{for } m(m + 1) < N + 1 \le (m + 1)^2 \tag{3.3a}$$

hence

$$t_{max} = N + 1 - 2m \qquad \text{for } m^2 < N + 1 \le m(m + 1)$$

$$t_{max} = N - 2m \qquad \text{for } m(m + 1) < N + 1 \le (m + 1)^2 \tag{3.3b}$$

where m is a positive integer.

In this paper, $\Omega_0(N)$, $\Omega^{(t)}(N)$, and related quantities are computed by exhaustive enumeration of every conformation that does not violate excluded-volume contraints. Exhaustive simulation of short lattice chains has been a major theoretical tool in polymer science.[11] Orr[1] was the first to adopt the method in 1947 by considering $N \le 8$ for square lattices and $N \le 6$ for simple cubic lattices. With the invention of high-speed computers, a systematic effort to reach longer chain lengths was pioneered by Domb and his collaborators[12-15] in the beginning of the 1960s. That work is now recognized as the basis for most for the modern developments in polymer theoretical physics, including scaling law methods[16] and the modern path integral and renormalization theories.[17,18] Inspiration from these analytic theories in turn channeled most of the more recent exhaustive simulation work into the determination of various scaling exponents and connective constants. A recent review of all the available data in two-dimensional lattices can be found in the work by Guttmann.[19] To our knowledge, the longest chain length on a square lattice for which $\Omega_0(N)$ has been successfully enumerated is $N = 25$, reported by Rapaport[20] in a study of the end-to-end distances of polymers.

Systematic exhaustive simulations have also been employed to address polymer problems involving intrachain contacts. The first was due to Orr,[1] who enumerated $\Omega^{(t)}(N)$ in a treatment of polymers in solution. Values for all possible t were provided by Orr. Fisher and Hiley[21] extended this work and provided $\Omega^{(0)}(N)$ for $N \le 14$ on square lattices, $N \le 12$ on triangular lattices, and $N \le 10$ on simple cubic lattices. Recently, Ishinabe and Chikahisa[22] have been able to enumerate conformations for longer chains and have obtained values for $\Omega^{(0)}(N)$ for N

Table I
Number of Conformations $\Omega^{(t)}(N)$ on the Square Lattice as a Function of the Number of Contacts, t, and Chain Length, N

		$\Omega^{(t)}(N)$									
N	$\Omega_0(N)$ (total)	$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$	$t = 8$	$t = 9$
3	5	4	1								
4	13	9	4								
5	36	21	11	4							
6	98	50	32	16							
7	272	118	92	43	19						
8	740	281	254	134	66	5					
9	2034	666	672	425	173	98					
10	5513	1584	1778	1229	576	298	48				
11	15037	3743	4622	3450	1944	803	444	31			
12	40617	8877	11938	9625	5718	2830	1262	367			
13	110188	20934	30442	26467	16736	9538	3722	1989	360		
14	296806	49522	77396	71570	48452	28297	13650	5655	2122	142	
15	802075	116579	194896	191814	138446	84607	45564	18733	8662	2705	69

≤ 22 on square lattices and $N \leq 20$ on tetrahedral lattices. In both of the recent works cited above, much attention was devoted to chains with absolutely no self-contacts, the $t = 0$ case of polymers in "super-solvents". By comparison, much less attention was devoted to the compact configurations. These states have been of little interest for polymer solution theories: according to Fisher and Hiley,[21] calculations for the conformational freedom of a *single* chain with a high probability of self-contact formation are "not very significant since polymer molecules in solution then tend to attract one another so that coagulation and eventually phase separation set in". However, for proteins, these conformations are of fundamental importance.

We have determined[4] $\Omega_0(N)$ on square lattices for $N \leq 17$ and $\Omega^{(t)}(N)$, $0 \leq t \leq t_{max}$, for $N \leq 15$. Values for $\Omega^{(t)}(N)$ are listed in Table I. In our enumeration, the two ends of a chain are considered to be distinct, and only conformations that are not related by translation, rigid rotation, or reflection are counted. With appropriate transformations to take care of minor differences in definitions,[23] our results for $\Omega^{(t)}(N)$ agree with those of Orr.[1] Values for $\Omega^{(0)}(N)$ in Table I also agree with those of Fisher and Hiley[21] except for a slight discrepancy[24] for $N = 14$. Also our calculations for $\Omega^{(0)}(N)$ are in total agreement with those of Ishinabe and Chikahisa.[22]

The probability that a chain will form t contacts is dependent upon its chain length N. The maximum possible number of contacts, t_{max}, also depends on N; see eq 3.2 and 3.3. We therefore define the *compactness*, ρ, as the ratio of the number of topological contacts of a conformation *relative* to the maximum number of contacts attainable for a given chain length.

$$\rho \equiv \frac{t}{t_{max}} \qquad 0 \leq \rho \leq 1 \tag{3.4}$$

Figure 2 is a histogram showing the distribution of the number of accessible conformations as a function of ρ for $N = 15$. From the data of Table I, we can compute the average compactness

$$\langle \rho(N) \rangle \equiv \sum_{t=0}^{t_{max}} \frac{t}{t_{max}} \frac{\Omega^{(t)}(N)}{\Omega_0(N)} \tag{3.5}$$

which characterizes the tendency of a chain of length N to form contacts. Values for $\langle \rho(N) \rangle$ computed for $3 \leq N \leq 15$ show a very gradual decreasing trend with increasing N, with some slight oscillations, ranging from 0.24 to 0.30; see the inset of Figure 2. A related quantitity $q \equiv \sum_t (t/N)\Omega^{(t)}(N)/\Omega_0(N)$ has been computed by Ishinabe and Chikahisa[22] for $N \leq 22$, and the limit $q \to 0.16$ was extrapolated for $N \to \infty$. Since $\langle \rho(N) \rangle = q(N/t_{max}) > q$ for finite N and $\langle \rho(N) \rangle \to q$ as $N \to \infty$, we estimate that

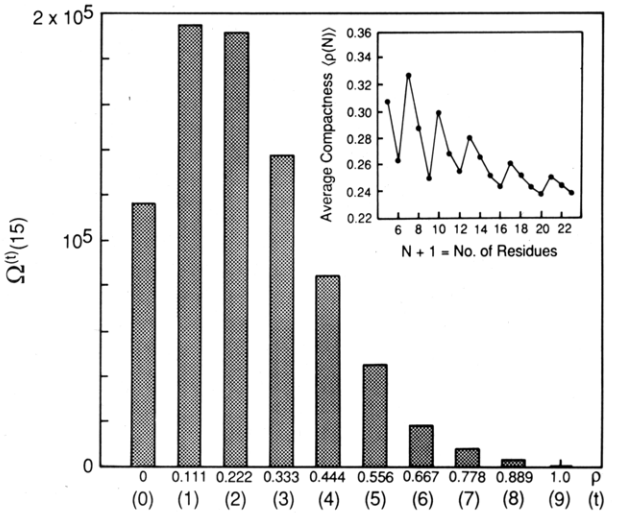

Figure 2. Number of conformations as a function of compactness ρ, for $N = 15$, $\Omega_0(15) = 802\,075$. The number of topological contacts t are shown in parentheses below ρ. The inset shows the variation of average compactness $\langle \rho(N) \rangle$ as a function of chain length. Data for $N + 1 \geq 17$ is from Ishinabe and Chikahisa.[22]

$\langle \rho(N) \rangle$ is bounded between 0.16 and 0.24 for intermediate to very long chain lengths.

As shown in Table I and Figure 2, the number of conformations is maximal at small ρ and decreases rapidly as ρ approaches unity. For $\rho = 1$, the number of maximally compact conformations

$$\Omega_c(N) \equiv \Omega^{(t_{max})}(N) \tag{3.6}$$

constitutes only a very small fraction of the full conformational space. For instance $\Omega_c/\Omega_0 = 69/802075 = 8.6 \times 10^{-5}$ for $N = 15$. Throughout the rest of this paper, "compact" will refer to conformations for which $\rho = 1$. Most real proteins are well represented as maximally compact, or nearly so.[25]

Due to the reduced conformational freedom in the compact states, longer chain lengths are more computationally accessible than for the open conformations. We have enumerated compact square-lattice chains with lengths $N + 1 \leq 30$ and $N + 1 = 36$. The numbers of compact conformations, $\Omega_c(N)$, are given in the second column of Table II.

It is generally believed that Ω_c has an approximate exponential dependence on N; i.e. $\Omega_c(N) \sim \kappa^N$, where $\kappa \geq 1$ is known as the connective constant.[19] Since the dominant functional dependence for open chains is also approximately exponential,[14,26,27] $\Omega_0(N) \sim N^\gamma \mu^N$, with a different connective constant $\mu > \kappa$, then $\ln(\kappa/\mu)$ represents the entropy loss *per segment* (residue) due to compactness,

Table II
Number of Compact (ρ = 1) Conformations $\Omega_c(N)$ on the Square Lattice[a]

N (N + 1)	$\Omega_c(N)$	S	N (N + 1)	$\Omega_c(N)$	S
3 (4)	1	1	17 (18)	1673	30
4 (5)	4	8	18 (19)	544	8
5 (6)	4	2	19 (20)	503	2
6 (7)	16	22	20 (21)	11226	187
7 (8)	19	6	21 (22)	11584	68
8 (9)	5	1	22 (23)	4577	22
9 (10)	98	30	23 (24)	3997	6
10 (11)	48	8	24 (25)	1081	1
11 (12)	31	2	25 (26)	100750	238
12 (13)	367	68	26 (27)	52594	88
13 (14)	360	22	27 (28)	45238	30
14 (15)	142	6	28 (29)	16294	8
15 (16)	69	1	29 (30)	13498	2
16 (17)	1890	88	35 (36)	57337	1

[a] N is the number of bonds, N + 1 is the number of residues, and S is the number of compact shapes.

relative to the freedom of open configurations.

Considerable effort has been devoted to the determination of κ. According to the Flory approximation,[28] $\kappa \simeq (z - 1)/e$, where z is the coordination number and e = 2.71828 ... is the base of the natural logarithm. Substituting z = 4, this gives κ = 1.1036 for the square lattice. By taking into account the conditionality resulting from the vacancy of the adjoining site reserved for occupation by the preceding segment, the more refined Huggin approximation[29,30] leads to $\kappa \simeq (z - 1)/a$, where $a = (1 - 2/z)^{(1-z/2)}$; hence, κ = 1.5 for the square lattice. Utilizing earlier results of Chang[31] and Miller[32] as bounds, together with exact enumeration results, Orr[1] estimated in 1947 that $\kappa \simeq 1.4$ for the square lattice and $\kappa \simeq 1.9$ for the simple cubic lattice.

Since these early efforts, Kasteleyn[33] was able to solve exactly by analytic methods the enumeration problem on two-dimensional oriented lattices. In contrast to the regular square lattice on which the bond angles ±90° and 180° are allowed at each step except for excluded volume, he employed the so-called "Manhattan" lattice rules, which restrict bond angles in a manner similar to one-way traffic intersections. The connective constant κ = 1.3385 ... was computed exactly for such a lattice. Gordon, Kapadia, and Malakis[34] pointed out that this exact value is a lower bound for the value of κ that would be obtained for more realistic models of chain conformations on the square lattice. On the other hand, Domb[35] pointed out that an upper bound for κ can be obtained from the exact "square ice" value 1.539 due to Lieb.[36] Gujrati and Goldstein[37] combined these two observations and state $1.338 \leq \kappa \leq 1.539$ as rigorous bounds. More recently, Schmalz, Hite, and Klein[38] estimated κ by counting compact closed loops (Hamiltonian circuits) using transfer matrix methods. They gave an estimate $\kappa \simeq 1.472$ and a lower bound $\kappa \geq 1.3904$, the latter of which is an improvement over the Manhattan lattice lower bound of 1.3385.

We now turn to our own simulations. From a casual inspection, it is clear that the values of $\Omega_c(N)$ in Table II do not follow any approximate exponential trend. Instead, as N increases, $\Omega_c(N)$ oscillates with ever-increasing amplitudes. This apparent anomaly can be resolved, however, by realizing that, in each of the earlier calculations that predicted $\Omega_c(N) \sim \kappa^N$, it is assumed, either explicitly or implicitly, that only *one* overall shape is allowed for the compact chain conformations. However, isolated compact chains can adopt many different shapes. For certain chain lengths, *many* possible overall shapes are consistent with the minimum perimeter condition (eq 3.3a). Therefore,

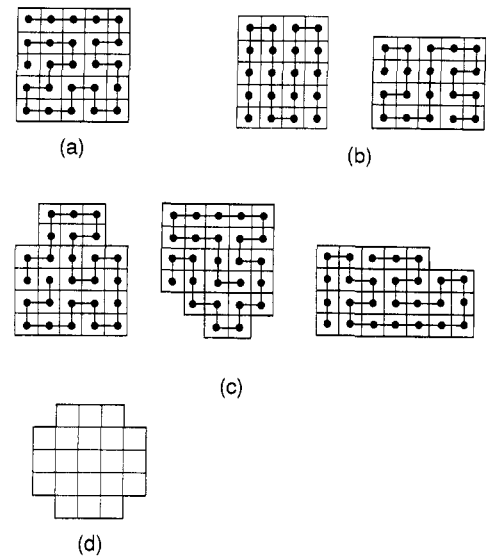

Figure 3. Compact shapes on the square lattice with minimum perimeter determined by the number of sites N + 1. (a) Only one shape is possible for $N + 1 = (m + 1)^2$. (b) Exactly two shapes are possible for $N + 1 = m\ (m + 1)$. (c) Three among a total of 238 possible shapes for N + 1 = 26. An example compact chain conformation is shown for each shape in (a)–(c). (d) No continuous chain with N + 1 = 21 residues can be fit into this 21-site compact shape.

in the earlier efforts cited above, other authors have considered, in essence, the number of compact conformations *per shape*, or the shape average, $\Omega_c(N)/S$, where S is the total number of compact shapes for chain length, N. As will be shown below, $\Omega_c(N)/S$ more closely approximates an exponential function of N than does $\Omega_c(N)$.

When the number of residues is a perfect square, $N + 1 = (m + 1)^2$, then the only maximally compact shape is a square; see Figure 3a. When N + 1 is a "near-perfect square", $N + 1 = m(m + 1)$, the compact conformations can take two possible shapes, as shown in Figure 3b, which are related by a rigid rotation of 90°. These two categories of numbers of residues (N + 1) = 2, 4, 6, 9, 12, 16, 20, 25, 30, 36, 42, 49, ... are special because essentially only one overall shape on the square lattice is accessible to their compact conformations. For convenience, these numbers of residues will be referred to below as "magic numbers". Obviously these magic numbers apply only to the square lattice and are lattice-dependent. Different sets of magic numbers will be found for other lattices. In general, for chain lengths not equal to the magic numbers, there will be a much larger number of geometric shapes consistent with the requirement of maximum compactness. For example, 238 shapes are possible for N + 1 = 26, a few of which are depicted in Figure 3c. Values for S are listed in the third column of Table II. The conventions[39] here for S are as follows: (i) All a priori shapes consistent with eq 3.3 are counted by S, no matter whether a chain can be suitably configured to fill it (see Figure 3d). (ii) Nonidentical rotations and reflections of shapes are counted as distinct. As an illustration, S = 1 for $N + 1 = (m+1)^2$ while S = 2 for $N + 1 = m(m+1)$; see parts a and b of Figure 3.

The logarithm of the *shape-averaged* number of compact conformations $\Omega_c(N)/S$ is plotted against $N + 1 \leq 30$ in Figure 4. The data shows an overall exponential increase, with oscillations bounded approximately by two dotted lines of the same slope $\kappa \simeq 1.41$, such that

$$c_1\kappa^N \leq \Omega_c(N)/S \leq c_2\kappa^N \tag{3.7}$$

where $c_1 \simeq 0.0787$ and $c_2 \simeq 0.318$. The global increasing

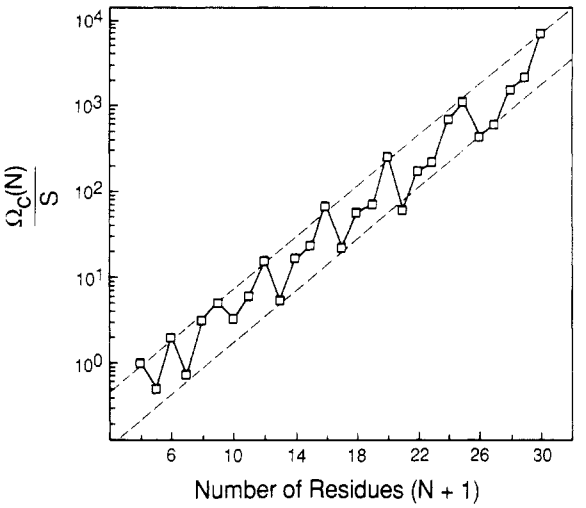

Figure 4. Shape-average number of compact conformations $\Omega_c(N)/S$ vs number of residues $N + 1$. The general trend is an exponential increase. Oscillations are due to articulateness of the shapes; see text for details.

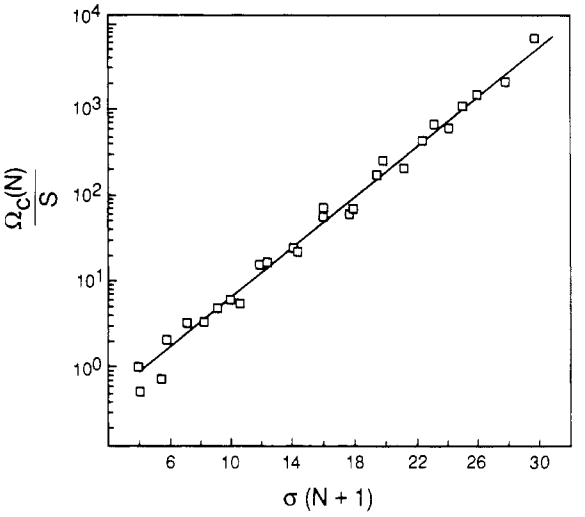

Figure 5. Shape-average number of compact conformations $\Omega_c(N)/S$ vs the parameter $\sigma(N + 1)$. The straight line is the best fit, $\Omega_c(N)/S = c\kappa^{\sigma(N+1)}$, with $c \simeq 0.226$ and $\kappa \simeq 1.40$.

trend (κ^N dependence) is attributable to the intrinsic conformational freedom associated with each chain segment, which leads to a general exponential increase of $\Omega_c(N)/S$ with chain length N. The oscillations, on the other hand, can be accounted for as corrections due to shape. Thus the number of conformations, and therefore the entropy, is comprised of two factors: (i) a conformational entropy per chain segment and (ii) a shape entropy for the compact object. It is clear that all values on the upper dotted line in Figure 4 ($c_2 \simeq 0.318$) correspond to the magic numbers; the values on the lower dotted line ($c_1 \simeq 0.0787$) represent chains for which the number of residues equals *1 plus* a magic number. Intermediate cases fall between the two dotted lines. The magic number result $\Omega_c(35)/S = \Omega_c(35) = 57\,337$ for $N + 1 = 36$ is not plotted in Figure 4 because data for $31 \leq N + 1 \leq 35$ are not yet available; nevertheless, it is easy to verify that $c_2\kappa^N$ gives a good estimation of this value to within 7%. Shapes for magic-numbered chains are square and near-square rectangles, but shapes for chains containing a number of residues between magic numbers are in general more articulated, as shown by examples in Figure 3c. The most articulated shapes occur when the number of residues equals 1 plus a magic number, since they have the maximum freedom to take different shapes. The "articulateness" of a compact shape may be quantitated as the number of corners on the perimeter. Since squares and rectangles with four corners are the only shapes possible for magic-numbered chains, according to this standard, they have the minimum articulateness of 4; by comparison, the average number of corners for $N + 1 = 26$ compact chains is 7.67.

The more articulated is the shape of the "box" into which the chain is "fit", the fewer are the configurations that can fill it precisely. In some cases of extreme articulation, there is no possible way that a chain can configure itself into that shape; for example, see Figure 3d, for $N + 1 = 21$. This observation explains the drop in $\Omega_c(N)/S$ when one residue is added to a magic-numbered chain: despite the intrinsic gain in conformational freedom due to an increased chain length, the shape effects due to increased articulateness dominate, resulting in a decrease in $\Omega_c(N)/S$. The local maxima of $\Omega_c(N)$ in Table II are those with $N + 1$ equal to 1 plus a magic number, coinciding with the local minima of $\Omega_c(N)/S$. In other words, the local maxima in $\Omega_c(N)$ are caused by shape multiplicity, *in spite of* the fact that the corresponding shape-averaged number of conformations $\Omega_c(N)/S$ is at their local minima. Equation 3.7 best summarizes our findings here: the general exponential trend is verified with a connective constant $\kappa \simeq 1.41$, which is very close to that of Orr[1] and is consistent with the recent lower bound proposed by Schmalz, Hite, and Klein.[38] The principal correction is due to boundary effects: more articulated shapes are shown to be more restrictive to conformational freedom.

In some earlier works, such as that of Kasteleyn,[33] boundary effects were avoided by adopting periodic boundary conditions, which effectively configures the chain on the surface of a torus. More recently, the importance of boundary effects of free edges on $\Omega_c(N)$ has been pointed out by Gordon, Kapadia, and Malakis.[34] Numerical results were provided by Malakis[40] for Manhattan lattices, but only for boundaries of rectangular shapes.

In the more general case of interest here, of the many possible compact shapes that can realize the P_c condition (eq 3.3a), it would be of value to have a quantity to correct for articulateness. To characterize the variation within the class of $\rho = 1$ compact chains, we define the parameter

$$\sigma \equiv \frac{N + 1}{[P_c/4]^2} \tag{3.8}$$

which is the ratio of the actual area occupied by chain residues to the area of a *hypothetical* square with the *actual* perimeter P_c. It is straightforward to show that

$$\left[\frac{2m}{2m + 1}\right]^2 < \sigma \leq 1 \qquad \text{for } m^2 < N + 1 \leq (m + 1)^2 \tag{3.9}$$

and $\sigma = 1$ only if $N + 1 = (m + 1)^2$. As such, the deviation of σ from unity may be used to measure the deviation of a compact shape from a perfect square. This measure is consistent with the consideration of articulateness above: the σ values for the $N + 1 = m(m + 1)$ magic numbers are $4N/(4N + 1)$, which is very close to 1 for large N with an $O(1/N)$ deviation. The most articulated compact chains with $N + 1$ equal to 1 plus a magic number have a corresponding deviation of $O(1/\sqrt{N})$. We also found empirically that the exponential relation

$$\Omega_c(N)/S \simeq c\kappa^{\sigma(N+1)} \tag{3.10}$$

where $c \simeq 0.226$ and $\kappa \simeq 1.40$ seems to hold approximately. This is shown graphically in Figure 5, in which

the oscillations of Figure 4 are dramatically suppressed by the use of this quantity σ.

σ may be considered as a refined measurement of compactness *within* the class of $\rho = 1$ compact chains. The closer the values of σ are to unity, the more "well-packed" are the compact chains, since they have a smaller perimeter to area ratio.

In this paper, we study all accessible $\rho = 1$ compact chains. We include data from all chain lengths $N + 1 \leq 30$ and $N + 1 = 36$. This not only provides us with the chain-length dependence of various physical properties but also allows us to probe the differential packing effects within the class of $\rho = 1$ compact chains, since different values of σ are sampled by chains of different lengths N. Thus a wider spectrum of data is obtained compared to studies that only consider magic-numbered chains.

4. Compact Chains with One Presumed Self-Contact

Our purpose in the present secton is to study cyclization, or loop formation, in compact polymers. Cyclization is an important property, which has previously been studied extensively in open chains. Using random-flight theory, Jacobson and Stockmayer first showed that the probability of spatial adjacency between monomers (residues) i and j is a diminishing function of their separation along the chain.[41] That theory has been augmented to take into account effects of local chain stiffness[42] and excluded volume.[4,26,43] In the present section, we explore the same question, but applied to only the compact conformations. The principal question addressed here is the following: how much is the compact conformational space restricted by the constraint that residues i and j are adjacent? We refer to this constraint (i, j) as a *presumed* contact, implying that it is specified a priori. This terminology is to distinguish that contact from the many other contacts in any particular compact conformation which arise simply as degrees of freedom of the system, not specified in advance. The effect of presuming a single contact pair (i, j) is measured by the reducton factor $R(N; i, j)$, defined to be the ratio[4]

$$R(N; i, j) \equiv \frac{\Omega(N; i, j)}{\Omega_0(N)} \tag{4.1}$$

where $\Omega(N; i, j)$ is the number of conformations that have the contact pair (i, j) and $\Omega_0(N)$ is the total number of accessible conformations.

We are interested in the cyclization probability as a function of compactness. We define the following quantities

$$\omega^{(t)}(N; i, j) \equiv \sum_{t'=t}^{t_{max}} \Omega^{(t')}(N; i, j) \tag{4.2a}$$

$$\omega_0^{(t)}(N) \equiv \sum_{t'=t}^{t_{max}} \Omega^{(t')}(N) \tag{4.2b}$$

$$R^{(t)}(N; i, j) \equiv \frac{\omega^{(t)}(N; i, j)}{\omega_0^{(t)}(N)} \tag{4.2c}$$

as a function of the compactness parameter, t. In eq 4.2a $\Omega^{(t')}(N; i, j)$ is the number of conformations with a total of t' contacts that also satisfy the condition that the pair of residues (i, j) are in contact, and $\Omega^{(t')}(N)$ in eq 4.2b is defined in eq 3.1. In these equations, the superscript (t) in $\omega^{(t)}$, $\omega_0^{(t)}$, and $R^{(t)}$ denotes the *minimum* number of contacts in the chains under consideration. Hence $R^{(t)}(N; i, j)$ measures the effect of presuming the contact pair (i, j) in the collection of chains that have *at least* t contacts.

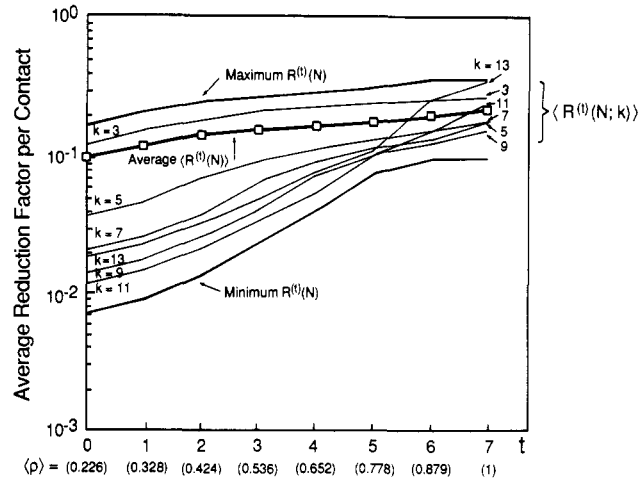

Figure 6. Average reduction factor per contact for chains with $N + 1 = 14$ residues, as a function of increasing average compactness $\langle\rho\rangle$. On the horizontal axis, t refers to the *minimum* number of contacts in the chains.

In these collections, the average chain compactness $\langle\rho\rangle$, over all chains with number of contacts $t' \geq t$ ($t' \leq t_{max}$), increases with t, while the size of the conformational space contracts with increasing t, since the conformational space of $R^{(t+1)}$ is a *subspace* of the conformational space of $R^{(t)}$. For $t = 0$, $R^{(0)}$ equals the open-chain reduction factor R in eq 4.1; on the other hand, when $t = t_{max}$, $R^{(t_{max})}$ becomes the *compact* reduction factor R_c that only takes into account the compact ($\rho = 1$) conformations, defined as

$$R_c(N; i, j) \equiv \frac{\Omega_c(N; i, j)}{\Omega_c(N)} = R^{(t_{max})}(N; i, j) \tag{4.3}$$

where $\Omega_c(N; i, j) = \Omega^{(t_{max})}(N; i, j)$ is the number of compact conformations that have the (i, j) contact. $R^{(t)}$ therefore interpolates between the open and compact chains. Its behavior as a function of t is useful for studing how the formation of specific contacts are affected by the overall packing density of the chain.

The general restricting effects of contacts on conformational freedom may be characterized by the *average reduction factor* $\langle R^{(t)}(N)\rangle$ *per contact*, defined as

$$\langle R^{(t)}(N)\rangle \equiv \frac{\sum_{i<j} R^{(t)}(N; i, j)\omega^{(t)}(N; i, j)}{\sum_{i<j}\omega^{(t)}(N; i, j)} \tag{4.4}$$

In the above equation, the numerator is a weighted sum of the reduction factor $R^{(t)}$ over all possible contact pairs (i, j). The weight $\omega^{(t)}(N; i, j)$ is the number of configurations that have the (i, j) contact and compactness greater than or equal to t/t_{max}. The denominator

$$\sum_{i<j}\omega^{(t)}(N; i, j) = \sum_{t'=t}^{t_{max}} t'\Omega^{(t')}(N) \tag{4.5}$$

is simply the total number of *contacts* present in the same conformational space.

The average reduction factor $\langle R^{(t)}(N)\rangle$ is computed for chains with $N + 1 = 14$ residues, and the data are shown as squares in Figure 6. $\langle R^{(t)}(N)\rangle$ gradually increases with chain compactness toward $\rho = 1$. Physically, this implies that the a priori constraint (i, j) becomes less restrictive relative to all other constraints, as the number of total constraints increases. Although the constraint (i, j) is less restrictive of compact chains than of open chains, nevertheless its restriction on conformational freedom is still significant. For the short chain example $N + 1 = 14$, the average reduction factor for $\rho = 1$ compact chains when

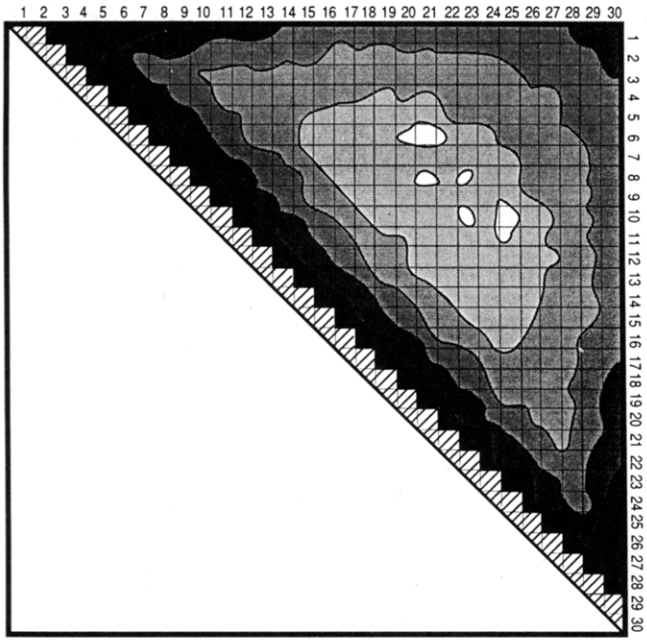

Figure 7. Free energy contour plot for the entropy of contact formation in compact square-lattice chains with 30 residues. Contours are given in $0.4kT$ steps.

$t = t_{max}$ is $\langle R_c(13)\rangle = 0.224$, compared with $\langle R^{(0)}(13)\rangle = \langle R(13)\rangle = 0.098$ for open chains. The average reduction factor is only increased by a factor of 2.30 by packing. As a function of chain length, our calculations show that both $\langle R(N)\rangle$ and $\langle R_c(N)\rangle$ tend to decrease gradually as N increases. For example, $\langle R_c(N)\rangle$ for $N + 1 = 20$, 25, 30, and 36 are 0.177, 0.155, 0.138, and 0.127, respectively, indicating that an average contact is more restrictive in longer than in shorter chains.

The average reduction factor $\langle R^{(t)}(N; k)\rangle$ of contacts (loops) of order (size) k, lying along the kth diagonal on the contact map, can be computed by restricting the summations over i and j in eq 4.4 to $j - i = k$. $\langle R^{(t)}(N; k)\rangle$ for $N + 1 = 14$ are plotted in Figure 6. The two thicker curves that envelop the $k = 3$–13 thinner curves represent the minimum and maximum reduction factor found in each conformational space labeled by t. As the chains increase compactness, $\langle R^{(t)}(N; k)\rangle$ exhibits more dramatic changes than the smooth behavior of $\langle R^{(t)}(N)\rangle$. The differences in average reduction factors for different contact orders become smaller, accompanied by a narrowing of range between the minimum and maximum reduction factor. At maximum compactness $\rho = 1$, $\langle R^{(t_{max})}(N; k)\rangle$ for all k are comparable, resulting mainly from the rapid increase for large k upon packing. The highest order $k = 13$ actually overtakes the lowest order $k = 3$ as the most favored contact order at $\rho = 1$.

This phenomenon is easy to understand physically. When the chains are more open, most of the contacts are concentrated at the lower orders close to the main diagonal on the contact map,[4] because they are less restrictive of conformational freedom. However, when chains become more compact by forming increasing numbers of contacts, the lower order contacts become saturated and higher order contacts must be formed. Consequently higher order contacts are enhanced by packing.

A more convenient global representation of the contact pattern in compact chains is the topological contact free energy surface introduced in studies of open chains.[4] Here the *compact* reduction factor $R_c(N; i, j)$ of eq 4.3 is computed for every (i, j). The *contours* for the quantity [$-\ln R_c(N; i, j)$ + constant] are then plotted on the contact map.

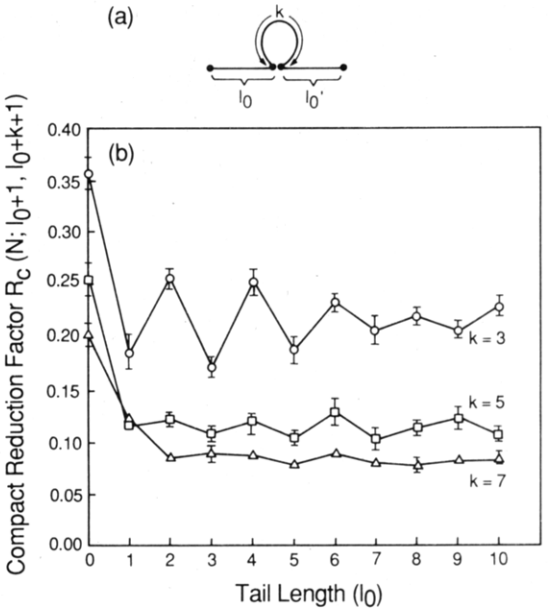

Figure 8. Reduction factors in compact chains vs tail length l_0 defined in diagram (a). Contacts are favored at chain ends.

This quantity is a dimensionless free energy, in units of kT, due only to chain conformational entropy, for relative formation probabilities of the different contacts.

Figure 7 shows such a topological free energy plot for compact chains with $N + 1 = 30$ residues. As is also observed for open chains,[4] contours curve near the vertical and horizontal boundaries of the contact map, showing that there are end effects. Contact formation is more favored at the chain ends. The least favored contacts (lightest colored) are located in the middle of the contact map. In this example with $N + 1 = 30$, the least favored contacts are those formed between residues that are approximately 7 segments from the two chain ends.

For open chains, loops at either chain end are favored relative to loops internal in the chain, for two reasons. First, there is less volume excluded at the ends to interfere with their configurational freedom. Second, end segments can have one more topological neighboring residue than segments at midchain, because end segments have only one connected neighbor, whereas midchain segments have two connected neighbors. The least favored contacts in open chains are those between segments near the two chain ends, since contacts between those segments would cause the greatest restriction of conformational space. For compact chains also, the end effects arise from these same factors. However, since higher order contacts are more favored in compact chains than in open chains, there are two differences. First, the least favored contacts in compact chains are for pairs of chain segments somewhat closer together in the sequence than for open chains (i.e, they are found on a lower order diagonal). Second, the overall free energy landscape is much flatter for compact chains than for open chains (see also Figure 6).

Figure 8 shows in somewhat more detail the end effects on the probability of cyclization of small loops, in compact polymers. Exact compact reduction factors $R_c(N; i, j)$ are computed for contact orders $k = |j - i| = 3$, 5, and 7 at all possible locations along the chain, for various chain lengths $N < 30$. $R_c(N; i, j)$ is expected to depend on the two tail lengths $l_0 = i - 1$ and $l_0' = N - j - 1$, shown in Figure 8a. In Figure 8b, the *average* of $R_c(N; l_0 + 1, l_0 + k + 1)$ over different chain lengths, N, are plotted as a function of l_0. It is appropriate that only chain lengths that are substantially longer than the contact orders are included in

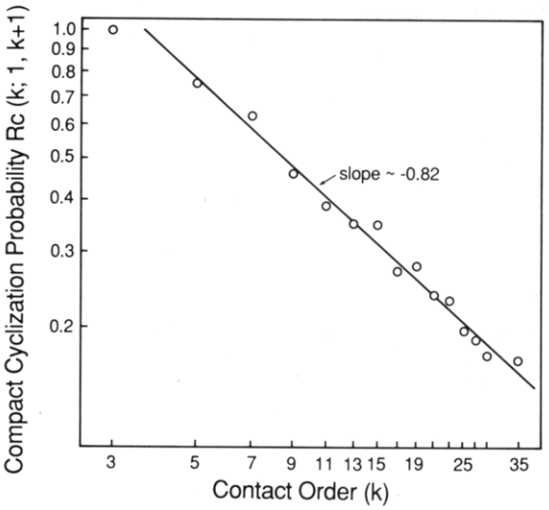

Figure 9. Power law dependence of the compact cyclization probability.

the averaging. Data in Figure 8b are obtained from averaging over $14 < N < 30$ for $k = 3$, $15 < N < 30$ for $k = 5$, and $16 < N < 30$ for $k = 7$. The variation in R_c for different N is recorded by the 1 standard deviation error bars in the plot. These variations are in general small; most of them are well below 10% of R_c. We believe that they are caused by the packing constraint, $\rho = 1$. By contrast, the corresponding variations are much smaller for open chains, in which case[4] limiting N-independent values of R are rapidly reached for $N \geq 16$.

Figure 8b confirms that there are end effects in compact chains, as in open chains.[4] For each of the three contact orders, $k = 3$, 5, and 7, R_c is highest at $l_0 = 0$, implying that contact (loop) formation is more preferred at the chain ends than in the middle. The large-amplitude even–odd oscillation of the $k = 3$ and $k = 5$ curves near $l_0 = 0$ is a feature peculiar to compact chains, not observed previously in open chains.[4] This phenomenon is a packing effect, since a loop with a tail chain comprised of an even number of residues is easier to pack into a compact conformation than one with an odd-numbered tail chain. On the topological free energy plot of Figure 7, these oscillations are manifested by the wavy features of some contours near the main diagonal.

The compact reduction factors R_c differ from their open-chain counterparts R in one important respect: R_c decreases much slower than R as the contact order k increases. In fact, at $l_0 = 1$, R_c values for $k = 5$ and $k = 7$ are approximately equal. At middle-chain positions, the ratio between R_c of $k = 3$ and $k = 5$ is approximately $0.22/0.12 = 1.83$ and between $k = 5$ and $k = 7$ is approximately $0.12/0.08 = 1.50$. Both numbers are much smaller than the corresponding ratios of 4.42 and 2.27 for open chains,[4] hence the flattened landscapes in compact polymers as noted above.

Next we consider the power law dependence of the compact reduction factor $R_c(k; 1, k + 1)$, the cyclization probability that the chain forms a ring with no tails by making contact between the two ends, $k = N$. The data are plotted in Figure 9, which show that for sufficiently large k, this probability has an approximate $k^{-\nu}$ dependence. Our best estimate from the data points with $k \geq 11$ gives $\nu \simeq 0.82$, which is about half of the corresponding exponent $\nu \simeq 1.63$ for open chains.[4] Hence, for chains of the same length, cyclization probabilities are relatively higher in compact chains than in open chains. Indeed, the probability is still reasonably high even for compact chains with intermediate lengths; for example, $R_c(k; 1, k + 1)$ equals $2310/13498 = 0.171$ for $N + 1 = k + 1 = 30$ and $9648/57337 = 0.168$ for $N + 1 = k + 1 = 36$. In summary, the two ends are quite likely to be near each other if the whole chain is confined to a small space. This may account for the fact that the two ends of protein molecules are often observed to be close to each other in the native state.[44]

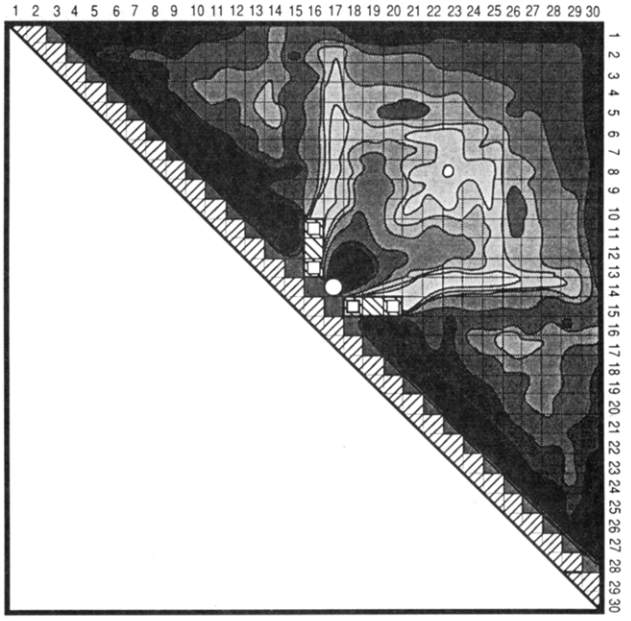

Figure 10. Contour plot ($0.4kT$ steps) of inference potential surface for compact chains with 30 residues and a single presumed contact at (14, 17). Most favored second contacts occur at helix and antiparallel sheet positions.

Our value of the exponent ν may be compared with a recent study of the mean-square end-to-end distance R_N^2 for compact chains by Ishinabe and Chikahisa.[22] Since the ring formation probability in two dimensions is approximately proportional to R_N^{-2}, their exponent $\nu_c \simeq 0.43$ ($2\nu_c \simeq 0.86$) in the scaling relation $R_N^2 \sim N^{2\nu_c}$ is equivalent to an exponent $\nu \simeq 0.86$ for the ring-formation probability. In view of the fact that not all data points lie exactly on the best fit line, their result is consistent with our value of $\nu \simeq 0.82$.

5. Correlation among Contacts in Compact Chains

In the preceding section, we considered the occurrence of a single specified loop, or intrachain contact, within the ensemble of compact polymer conformations. In the present section, we consider pairs of loops, two contacts (i_1, j_1) and (i_2, j_2) within the same chain. The reason this is of interest is because a pair of loops, represented by a set of two dots on the contact map, is the most elementary building block for secondary structures: helices, parallel and antiparallel sheets, and turns. We have recently shown, through development of a theory for "topological" pair correlation functions for interactions among the loops, that the most probable conformations of open chains containing two intrachain contacts are helices and antiparallel sheets.[4] Our aim in the present section is to extend the application of this topological pair correlation function approach to the subset of compact conformations of polymers.

If two loops in a chain are near each other in the sequence, then their cyclization probabilities may not be independent. The interdependence can be defined as follows.[4] Suppose contact (i_1, j_1) is presumed to be given, indicated by the hollow circles on the contact maps in Figures 10 and 11. Then if $R_c(N; i_1, j_1; i_2, j_2)$ is the reduction factor for simultaneously presuming two contacts

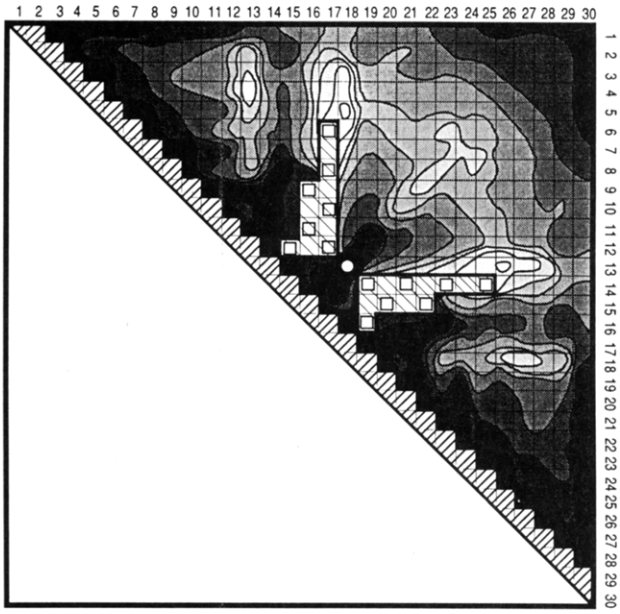

Figure 11. Contour plot ($0.4kT$ steps) of inference potential surface for compact chains with 30 residues and a single presumed contact at (13, 18). For this order 5 presumed first contact, the most favored second contacts occur at antiparallel sheet position (14, 17).

(i_1, j_1) and (i_2, j_2) in compact chains of length N, the *conditional* probability of forming the second contact (i_2, j_2) subject to the presumed first contact is simply

$$\frac{R_c(N; i_1, j_1; i_2, j_2)}{R_c(N; i_1, j_1)} \tag{5.1}$$

The logarithm of this quantity is plotted as contours for all possible positions (i_2, j_2) on the contact map. The contours then represent the relative probabilities of forming any second contact for a given first contact. In this representation, the two-contact "interaction" is represented by this potential surface, analogous to the field description of particle–particle interactions in classical physics.

Figures 10 and 11 show two such potential surfaces for chains with $N + 1 = 30$ residues. In Figure 10, the presumed contact is of order 3 at (14, 17). In Figure 11, the presumed contact is of order 5 at (13, 18). The 30-residue compact chains are taken here as a typical example representative of intermediate-length chains. We have verified by exhaustive enumerations of various chain lengths that potential surfaces of comparable lengths have features similar to those shown in Figures 10 and 11. The subtlety and complexity, which arises simply from excluded volume in the compact state, is quite remarkable.

The most important features of these free energy surfaces are the locations of the deepest minima. The deepest minima identify the most probable second loop, given a specific presumed first loop. It is clear from Figure 10 that given the smallest possible presumed loop at $(i_1, j_1) = (14, 17)$, the two most probable configurations in the compact conformational space are the antiparallel sheet (13, 18) and the helix (12, 15) and (16, 19). Those configurations are favored by $3.2kT$ and $2.8kT$, respectively, relative to the least favored contacts. Likewise the same conclusion is drawn from Figure 11. For the presumed contact of order 5 at (13, 18), the most probable second contact is that which "fills in" the turn for the antiparallel sheet (14, 17), $4.0kT$ more favorable than the least favored contact. The next most favorable contacts are those that may be associated with helices: (11, 14) or (17, 20) or extension of the antiparallel sheet (12, 19), all of which are equally favored by $3.2kT$ relative to the worst contacts. It is also clear that the end effects are as strongly favored as the second most probable conformation in each case. These results are obtained through exhaustive search of every conformation accessible to the compact chain.

Similar results have been observed for open chains.[4] In that case also, given one presumed contact of order 3 or 5, the most probable second contact results from "zipping up" the helix or antiparallel sheet from that point. Hence there is a driving force for secondary structure formation when only two intrachain contacts have formed, $t = 2$. As shown above, the same driving force exists at $t = t_{max}$, i.e., in the compact state. In the next section, it is shown that this driving force persists for the formation of every additional contact for all densities increasing from $t = 2$ to $t = t_{max}$. Thus steric forces act to drive the formation of secondary structures.

Certain other features of these free energy surfaces are also similar for compact and open chains. The hollow squares on some sites of the contact maps in Figures 10 and 11 represent "implied blocks", contacts which could not be formed for *any* chain configuration, open or compact, for the given presumed contact. Clearly, therefore, these lines of blocks will appear identically for open or compact chains. Similarly, the extensions of these lines of implied blocks are disfavored conformations, for either open or compact conformations (see columns 16–18 or rows 12–14 in Figures 10 and 11). For open conformations, the origins of these disfavored conformations in competing effects of excluded volume have been described elsewhere.[4]

The interdependences of cyclization of two loops within a chain can be described in terms of a correlation function $g_{c;k_1,k_2}$ for two contacts of orders $k_1 = |j_1 - i_1|$ and $k_2 = |j_2 - i_2|$

$$g_{c;k_1,k_2}(L) = \frac{R_c(N; i_1, j_1; i_2, j_2)}{R_c(N; i_1, j_1)\, R_c(N; i_2, j_2)} \tag{5.2}$$

defined here for compact chains. This definition for correlation function differs from that given previously[4] only in that the subscript c indicates that it applies to the compact conformations. In eq (5.2), R_c refers to compact reduction factors defined above, and L is the separation between the two contacts under consideration, measured as the number of bonds from the starting point on the larger loop to the corresponding point on the smaller loop.[4] g_c is the ratio of the actual number of compact conformations that satisfy the two presumed contacts to the number of compact 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 the other: enhancement is indicated when $g_c > 1$, hindrance is indicated when $g_c < 1$, and $g_c = 1$ implies independence.

Shown in Figure 12 is the compact correlation function $g_{c;k_1,k_2}$ calculated for $k_1 = k_2 = 3$ and $k_1 = 3$, $k_2 = 5$ and $N + 1 = 30$ residues. Depending on the exact location of the two contacts along the chain, there is considerable variation even when the separation L between the two contacts remains constant. This is an effect of the compact packing. In open chains with the two contacts well embedded in the middle of long chains, such variations are negligible.[4] The *averages* of $g_{c;k_1,k_2}$ over all possible positions along the chain are shown as square boxes, while the variations are represented by 1 standard deviation error bars. The thinner continuous lines correspond to the open chain (all ρ) results obtained earlier,[4] reproduced here for comparison.

Secondary structure enhancement is apparent from Figure 12, variations notwithstanding. Indeed, the average

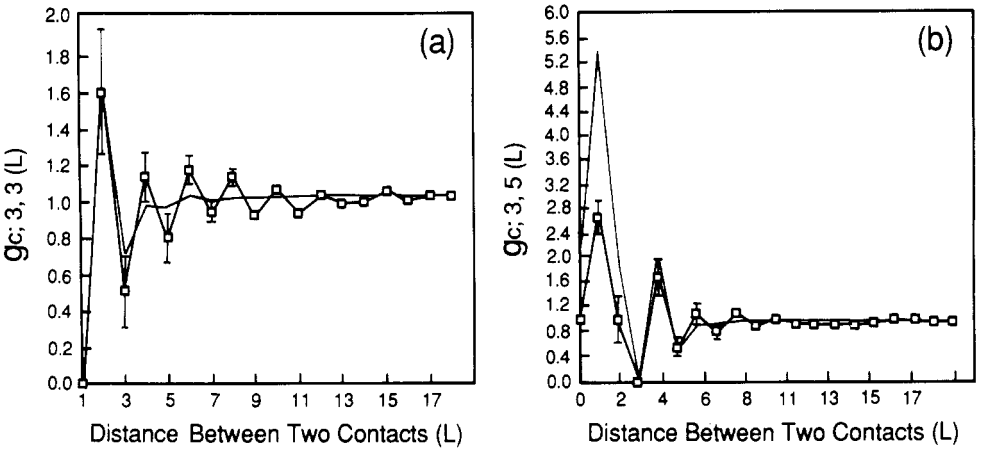

Figure 12. Two-contact correlation $g_{c;k_1,k_2}(L)$ for compact chains. (a) Correlation between two order 3 contacts, $k_1 = k_2 = 3$. (b) Correlation between an order 3 and an order 5 contact, $k_1 = 3$ and $k_2 = 5$. The variations along the chain are represented by 1 standard deviation error bars. Error bars that lie within the hollow squares are not shown. In both cases, the peak positions indicate that certain secondary structures are favored.

correlation g_c attains its maximum at $L = 2$ in Figure 12a, which is the helix position, and at $L = 1$ in Figure 12b, which is the antiparallel sheet position. The open chain g also has a maximum at these positions.[4]

In addition to the secondary structure peaks, these correlation function plots also show much similarity for compact and open chains, indicating that the local steric restrictions are similar. The following are observed in both cases: (i) contacts are essentially independent ($g, g_c \simeq 1$) when the separation L is large compared to the contact orders; and (ii) there is always a local minimum when L equals the size of the larger of the two contacts,[4] which corresponds to $L = 3$ in Figure 12a and $L = 5$ in Figure 12b.

Two differences between the open and compact cases should be noted however. First, the even–odd oscillation in Figure 12a is probably due to the same packing effect discussed above in conjunction with Figure 8. Second, notice in Figure 12b that the correlation function, g_c, is smaller at the peak value of compact than at those for open chains. Although the absolute probability of forming the antiparallel sheet is much higher in the compact chains (0.074) than in the open chains (0.014), recall that g_c is only a ratio, relative to the independent loops (see eq 5.2). In this case, it is simply that the single loops are also significantly enhanced by compactness.

6. Secondary Structures Driven by Packing

In the previous section, we found that steric constraints favor the association of two loops into helical or antiparallel sheet configurations in compact chains, in agreement with similar results for open chains.[4] In the present section, we ask the broader question of the joint distribution of *all* the possible loops within the molecule, including now also parallel sheets and turns. In addition, we study the distribution of secondary structural elements as a function of the compactness and length of the molecule and of the position within the chain.

When the definition introduced in section 2 is used, the number of residues participating in various secondary structures are computed for each of the 802 075 accessible conformations with $N + 1 = 16$ residues. Figure 13 shows the fractional participation, the total number of residues participating in secondary structures divided by $N + 1$, as a function of compactness, $\rho = t/t_{max}$. The principal conclusion from Figure 13 is that the fraction of residues in secondary structures increases with compactness and is remarkably high in compact molecules. Clearly, since

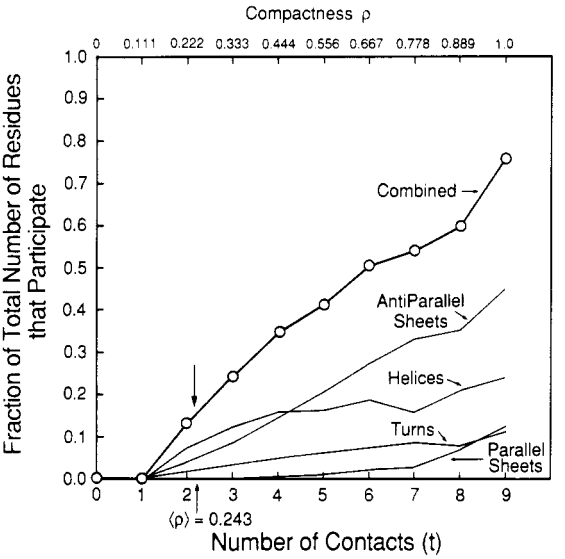

Figure 13. Secondary structures in chains with 16 residues. All secondary structure types are enhanced by compactness. The arrows indicate the average compactness $\langle\rho\rangle$ of the global conformational space and the corresponding combined participation rate of 0.153.

any secondary structure requires for its definition a minimum of $t = 2$ intrachain contacts, then no secondary structure can occur for $t = 0, 1$. Hence in the limit as $\rho \to 0$, for polymers in "super-solvents" or which are highly charged, the amount of secondary structure should be negligible. For maximally compact ($\rho = 1$) molecules, however, the fractional participation in secondary structures reaches 76.1% for the case shown. It is noteworthy that if the global conformational space of open chains (all possible ρ) is considered as a whole, the combined participation rate is about 15.3% for $N + 1 = 16$. Due to the near linearity of the combined participation curve, this rate is given approximately by the curve at $\rho = \langle\rho\rangle = 0.243$, indicated by an arrow in Figure 13. It is interesting that proteins that are unfolded by "weak" denaturing agents, such as temperature or pH/salt in some cases, and have high-density unfolded states are then also predicted to have much secondary structure. This is in accord with experimental evidence.[45,46]

All four kinds of secondary structure are enhanced by compactness. In the present model, antiparallel sheets overtake helices at intermediate compactness as the most prominent type of structure. This phenomenon is

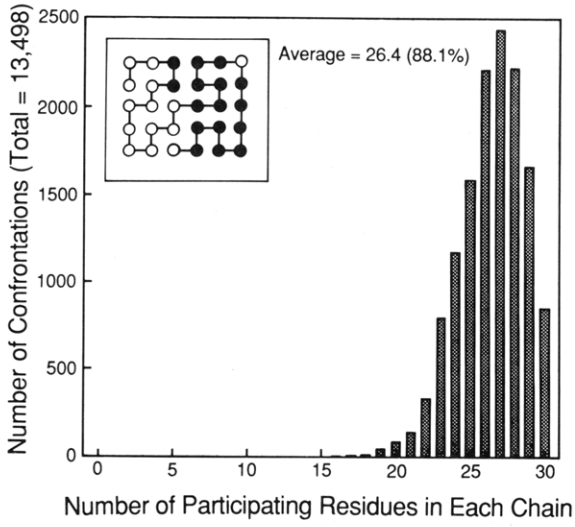

Figure 14. Histogram showing how secondary structure is distributed throughout the compact conformational space, for chains with 30 residues. There is a lower cutoff at 16; no compact chains with 30 residues can configure without at least 16 residues participating in secondary structures. The inset shows an example of one conformation having this minimal participation. Residues participating in secondary structures are filled dots; others are represented as hollow circles.

straightforward to understand in terms of packing enhancement of higher order contacts. Helices are confined to the order 3 diagonal, but sheets are not subject to this limitation. At low compactness, the order 3 diagonal is most favored. However, beyond a certain compactness, the order 3 diagonal will be saturated and higher order contacts will have to be formed. Whereas Figure 13 shows the behavior of the fractional number of *residues*, we have verified separately that the fractional number of *contacts* participating in sheets and helices has the same crossover at intermediate compactness.

Figure 14 shows the distribution of the number of conformations that each have a particular number of secondary structure participating residues for all compact 30-residue chains. Figure 14 shows the striking result that there is *no* compact conformation with less than 53% secondary structure for this chain length. The inset in Figure 14 shows a randomly chosen example conformation with the *minimum* participation rate. The same general result holds for other chain lengths, though compact chains with non-magic numbers of residues tend to have broader distributions than those with magic numbers, and a few conformations may be able to configure without forming any secondary structures; see Figure 15. This is not surprising because non-magic-numbered compact chains are not as well-packed as magic-numbered chains and therefore are subject to less packing constraint. An overwhelming majority of conformations accessible to compact chains are those with a high fraction of secondary structures. Compactness is simply inconsistent with most other arrangements of the chain. In Figure 14, quite a few (847 conformations, 6.3%) compact chains in fact have 100% participation rates. Figure 16 depicts a few examples of how this is achieved.

The high proportion of secondary structure in compact $\rho = 1$ chains is not an artifact due to the shortness of the chains. Quite the contrary, longer compact chains have even larger fractions of residues participating in secondary structures. For instance, compact chains with 30 residues have a participation rate of 88.1%, whereas compact chains with 36 residues have an even higher participation rate of 90.1%. Figure 17 shows the result of our calculations for all compact chains with the number of residues $N + 1$ ranging from 14 to 30, demonstrating that the high participation rate is a universal phenomenon for two-dimensional compact chains. There are oscillatory variations with chain length. Secondary structure participation is

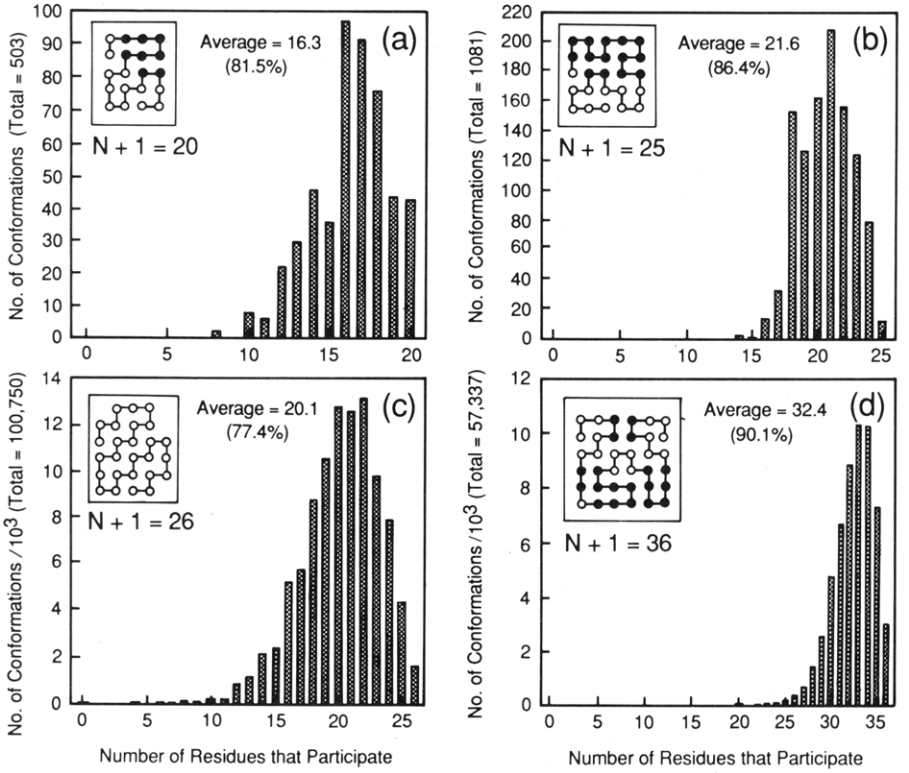

Figure 15. Same as Figure 14, for different chain lengths: (a) 20, (b) 25, (c) 26, and (d) 36 residues. Insets show examples of conformations with minimal participation. Residues participating in secondary structures are filled dots. Note that non-magic-numbered chains (c) have a broader distribution.

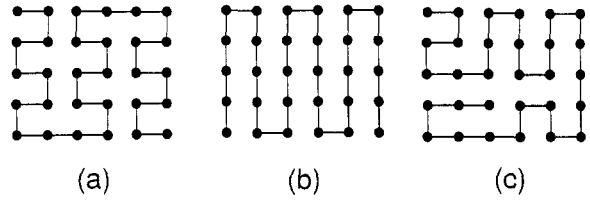

Figure 16. Three examples of compact chains of 30 residues with 100% participation in secondary structure: (a) all helices; (b) all antiparallel sheets and turns; (c) an example that contains all four types of secondary structure.

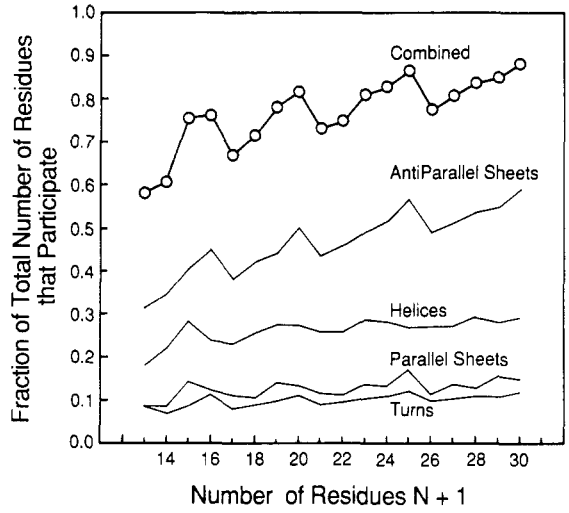

Figure 17. Secondary structure participation in compact chains. Participation rates increase with chain length: magic-numbered chains have more secondary structure than others.

largest for chain lengths with $N + 1$ equal to the magic numbers (defined in section 3) and is smallest when the number of residues is one unit longer than a magic number. Since magic-numbered chains are differentially more well-packed within the class of compact chains (see section 3), then this measure of increased "packing" also correlates with increased secondary structure.

Figure 17 suggests that the secondary structure participation rate will be even higher for longer chains. In order to eliminate the magic-number-related oscillations and obtain a better extrapolation, we plot these participation rates of compact chains in Figure 18 as a function of the perimeter/area ratio $P_c/(N + 1) = 4/[\sigma(N + 1)]^{1/2}$ (see eq 3.8), which tends to zero for chains of infinite length. From the general trend exhibited in Figure 18, it is not unreasonable to conjecture that the combined participation rate in two dimensions may asymptotically approach unity as $N \rightarrow \infty$; i.e., the domination of secondary structure motifs may be complete at infinite chain length. Clearly, conformations with lower than unity participation rates are always possible for finite N, but the extrapolation indicates that those conformations form a diminishing fraction of all compact chains as the chain length, N, tends to infinity.

The distributions of various secondary structure types along the chain are shown in Figure 19 for compact chains with 30 residues. Aside from end effects, the distributions are quite uniform. Without the incorporation of any specific interactions, secondary structure is predicted to be distributed uniformly along the chain. The end effects are also easy to understand: (i) The decrease in helix probability toward chain ends is a consequence of the fact that helices can form in only one direction at chain ends, but two directions are available in the middle of the chain. (ii) There are no turns at chain ends simply because at least two residues have to be attached to each of the two

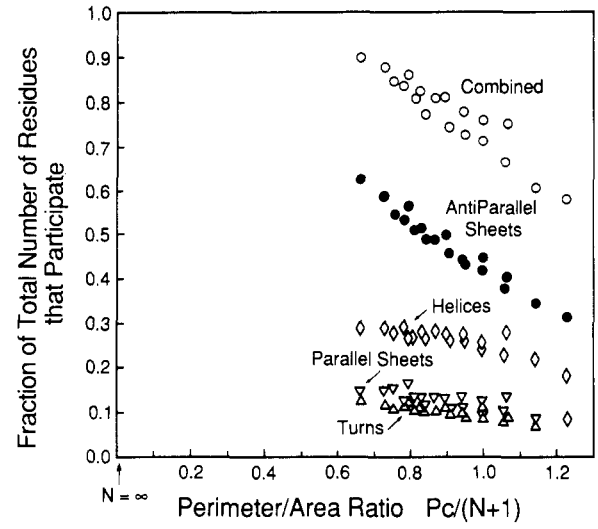

Figure 18. Participation in secondary structure as a function of the perimeter/area ratio in two-dimensional compact square-lattice chains. Data from chain lengths $N + 1 = 13$–30 and $N + 1 = 36$ are included. It is conjectured that the combined participation rate tends to unity as $P_c/(N + 1) \rightarrow 0$ or $N \rightarrow \infty$.

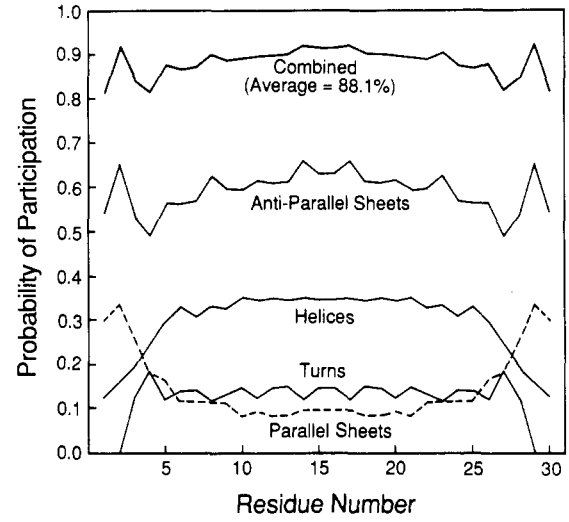

Figure 19. Distribution of various types of secondary structure along the sequence for compact chains with 30 residues.

residues comprising the turn. (iii) Parallel sheets are favored at chain ends because of the following. For two chain segments running in parallel, a chain segment of length considerably longer than the parallel segments is required to join the two parallel segments into a continuous chain. Clearly such connecting chain segments are more readily available if one of the parallel segments is located near a chain end.

To explore how sensitively our conclusions of secondary structure enhancement depend on the lattice definitions of secondary structure adopted in section 2, we have repeated all the above calculations with a modified and more restrictive definition. In this alternate approach, the definition for helices and turns (parts a and d of Figure 1) remained unchanged, but the minimal units of sheets (parts b and c of Figure 1) are modified from 2 to 3 contacts or from 4 to 6 residues. One motivation for experimenting with this alternate definition is that it eliminates the possibility of one residue participating in both a helix and a sheet, which is allowed by the definition of section 2 but impossible in real protein molecules. Obviously, reduced participation rates for sheets will result from this modification. It will be shown below however that this altered definition has only a minor quantitative effect on

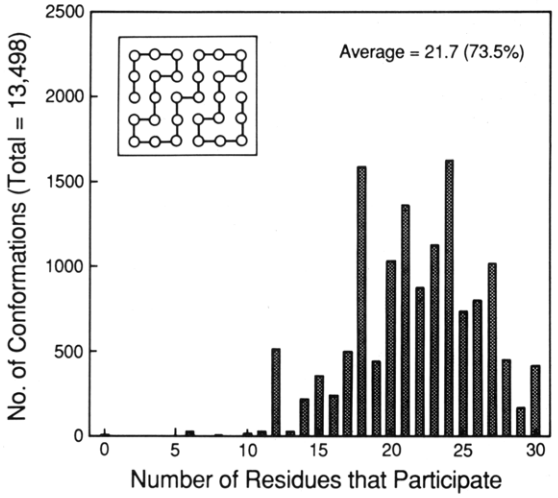

Figure 20. Histogram of the number of residues participating in secondary structure according to the more restrictive definition. Only one of the 13 498 compact chains with 30 residues configures without forming any secondary structure. This conformation is shown in the inset.

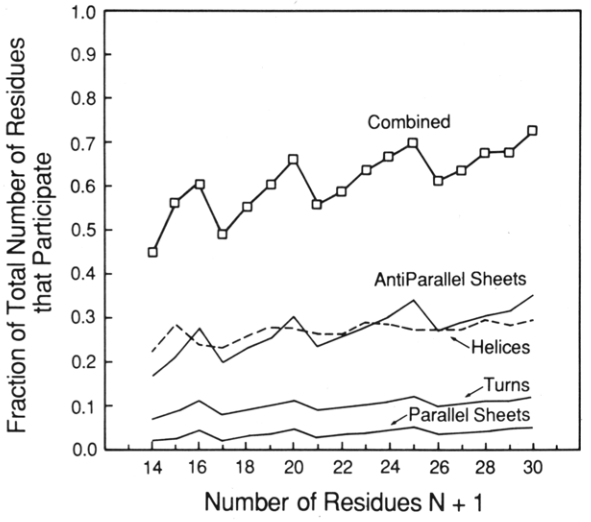

Figure 21. Secondary structure participation in compact chains according to the more restrictive definition.

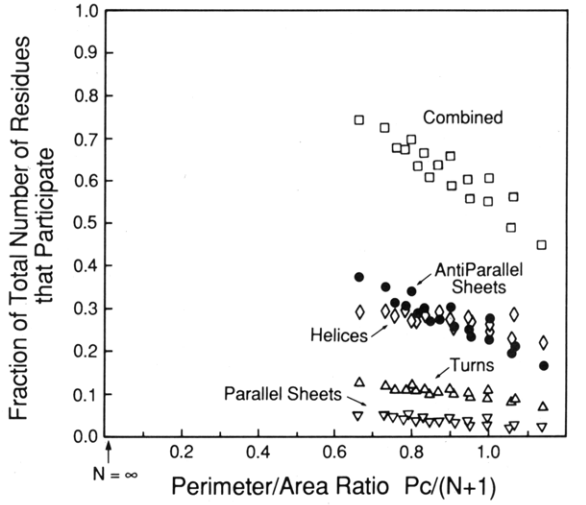

Figure 22. Participation in secondary structure as a function of the perimeter/area ratio under the more restrictive definition. Data from chain lengths $N + 1 = 14$–30 and $N + 1 = 36$ are included.

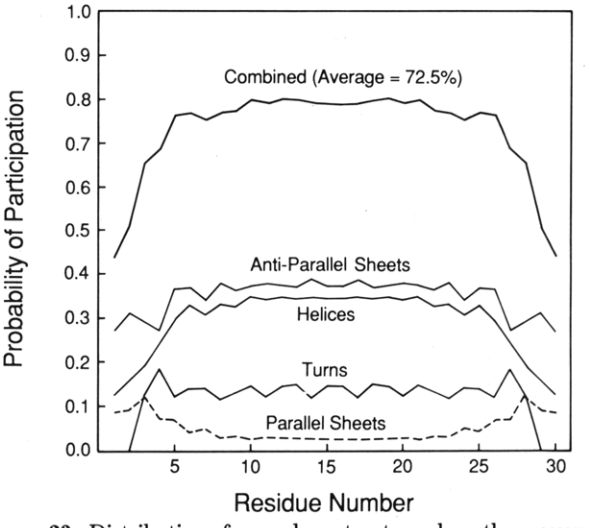

Figure 23. Distribution of secondary structure along the sequence of compact chains with 30 residues, according to the more restrictive definition.

the predictions above. The combined participation rates remain high: 72.5% and 74.4% for compact chains with 30 and 36 residues, respectively, for example. These calculations using modified definitions of secondary structures serve the function of "control experiments". They show that reasonable variations in the criteria for cataloging secondary structures have little effect on the conclusions drawn from this model.

Figures 20–23 are equivalent to Figures 14 and 17–19, calculated using the alternate secondary structure definition instead of that given in section 2. It is clear that many qualitative features remained unchanged. Under the more restrictive secondary structure definition, a broader distribution of the number of participating residues is shown in the histogram of Figure 20. In contrast to Figure 14, there is *exactly one* conformation that is able to configure without forming any alternately defined secondary structure. This conformation is shown in the inset of Figure 20. It is clear that even that particular structure has some order: an "antiparallel sheet" along the diagonal of the square lattice (see also Figure 15c). Nevertheless, this is excluded from our present definitions of secondary structure. Indeed, this example illustrates that some form of ordered contact pattern is almost unavoidable in two-dimensional compact chains. Figures 21 and 22 show that the participation rate remains high (over 60%) for compact chains of intermediate lengths ($N + 1 \geq 22$) and continues to increase for longer chains.

Similar calculations have been carried out in three dimensions.[5] The amount of secondary structure driven by packing forces is only slightly smaller on a three-dimensional cubic lattice. There is greater freedom of definition of secondary structures on the cubic lattice than on the square lattice,[47] however, but even so there is a minimum of about 35–40% secondary structure observed in compact chains in three dimensions.

Hence, our principal conclusion of packing enhancement of secondary structure is robust. It is valid under two different definitions of secondary structures in two dimensions and is valid for chains configured in both two and three dimensions. The basic principle that emerges here is that locally compact and nonarticulated substructures such as the secondary structures are favored in compact chains simply because they least obstruct the rest of the chain to configure into a compact conformation.

Acknowledgment. We thank K. F. Lau, J. Naghizadeh, and D. Yee for helpful discussion. We thank the NIH, the URI Program of DARPA, and the Pew Scholars Pro-

gram in the Biomedical Sciences for support of this work.

References and Notes

(1) Orr, W. J. C. *Trans. Faraday Soc.* **1947**, *43*, 12.
(2) Dill, K. A. *Biochemistry*, **1985**, *24*, 1501.
(3) Lau, K. F.; Dill, K. A. A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins. *Macromolecules* **1989**, *22*, 3986.
(4) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1989**, *90*, 492.
(5) Chan, H. S.; Dill, K. A., submitted for publication.
(6) It can be shown by explicit construction that in a two-dimensional square lattice a contact (dot on the contact map) *cannot* take part in more than one type of secondary structure.
(7) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211, 222.
(8) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(9) Kabsch, W.; Sander, C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1075.
(10) Pullman, B.; Pullman, A. *Adv. Protein Chem.* **1974**, *28*, 347.
(11) Barber, M. N.; Ninham, B. W. *Random and Restricted Walks; Theory and Applications*; Gordon & Breach: New York, 1970; and references therein.
(12) Domb, C. *Adv. Phys.* **1960**, *9*, 149.
(13) Sykes, M. F. *J. Math. Phys.* **1961**, *2*, 52.
(14) Domb, C.; Sykes, M. F. *J. Math. Phys.* **1961**, *2*, 63.
(15) Domb, C. *J. Chem. Phys.* **1963**, *38*, 2957.
(16) de Gennes, P-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, New York, 1979.
(17) Edwards, S. F. *Proc. Phys. Soc. London* **1965**, *85*, 613.
(18) Freed, K. F. *Renormalization Group Theory of Macromolecules*; Wiley: New York, 1987; and references therein.
(19) Guttmann, A. J. *J. Phys.* **1984**, *A17*, 455.
(20) Rapaport, D. C. *J. Phys.* **1985**, *A18*, L39; L201.
(21) Fisher, M. E.; Hiley, B. J. *J. Chem. Phys.* **1961**, *34*, 1253. In this reference, Fisher and Hiley gave $t_{max} = N(z-2)/2$, which is an approximation; the full effect of excluded volume has not been taken into account. The deviation of their approximation to the exact value given by eq 3.3b of the text is of $O(\sqrt{N})$.
(22) Ishinabe, T.; Chikahisa, Y. *J. Chem. Phys.* **1986**, *85*, 1009.
(23) The number of sites n in Orr's notation is equivalent to $N+1$ in our notation. He counted rigid rotations and reflections of conformations as distinct but considered two ends of the chain as identical. Therefore, his $G(0)$ equals our $2[2\Omega^{(0)}-1]$, and the coefficient of η^t in his $G(\eta)$ equals our $4\Omega^{(t)}$ for $t > 0$. The n of Fisher and Hiley and of Ishinabe and Chikahisa is equivalent to our N. In their convention, rigid rotations and reflections of conformations are counted as distinct, and two ends of the chain are also distinguished. Hence, their $c_n(0)$ equals our $4[2\Omega^{(0)}(N)-1]$ for $n = N$.
(24) Fisher and Hiley[21] gave $c_{14}^{(0)} = 396\,204$, but our result for $\Omega^{(0)}(14)$ is 49 522; according to the translation rules given in the last footnote, our result implies that $c_{14}(0)$ should be 396 172.
(25) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(26) Hiley, B. J.; Sykes, M. F. *J. Chem. Phys.* **1961**, *34*, 1531.
(27) Guttmann, A. J.; Ninham, B. W.; Thompson, C. J. *Phys. Rev.* **1968**, *172*, 554.
(28) Flory, P. J. *J. Chem. Phys.* **1942**, *10*, 51.
(29) Huggins, M. L. *J. Phys. Chem.* **1942**, *46*, 151.
(30) Huggins, M. L. *Ann. N.Y. Acad. Sci.* **1942**, *4*, 1.
(31) Chang, T. S. *Proc. R. Soc. London, A* **1939**, *A169*, 512.
(32) Miller, A. R. *Proc. Cambridge Philos. Soc.* **1942**, *38*, 109; **1943**, *39*, 54.
(33) Kasteleyn, P. W. *Physica* **1963**, *29*, 1329.
(34) Gordon, M.; Kapadia, P.; Malakis, A. *J. Phys.* **1976**, *A9*, 751.
(35) Domb, C. *Polymer* **1974**, *15*, 259.
(36) Lieb, E. H. *Phys. Rev. Lett.* **1967**, *18*, 692, 1046.
(37) Gujrati, P. D.; Goldstein, M. *J. Chem. Phys.* **1981**, *74*, 2596.
(38) Schmalz, T. G.; Hite, G. E.; Klein, D. J. *J. Phys.* **1984**, *A17*, 445.
(39) Our compact shapes constitute a subset of *compact lattice animals* commonly employed in lattice statistical mechanics. See, for example: Privman, V.; Forgacs, G. *J. Phys.* **1987**, *A20*, L543.
(40) Malakis, A. *Physica* **1976**, *84A*, 256.
(41) Jacobson, H.; Stockmayer, W. H. *J. Chem. Phys.* **1950**, *18*, 1600.
(42) Flory, P. J.; Semlyen, J. A. *J. Am. Chem. Soc.* **1966**, *88*, 3209.
(43) Martin, J. L.; Sykes, M. F.; Hioe, F. T. *J. Chem. Phys.* **1967**, *46*, 3478.
(44) Thornton, J. *Nature* **1988**, *335*, 10.
(45) Goto, Y.; Fink, A. L. *Biochemistry* **1989**, *28*, 945.
(46) Dill, K. A.; Alonso, D. O. V.; Hutchinson, K. *Biochemistry* **1989**, *28*, 5439.
(47) The greater freedom of definition of lattice secondary structures on the cubic lattice is due to higher spatial dimension and the increased coordination number ($z = 6$ instead of $z = 4$ for the square lattice). Although further increase in coordination number should lead to increased accuracy in representation of secondary structures and could therefore affect somewhat the predicted amounts of secondary structure, nevertheless the amount of secondary structure should not become diminishingly smaller with increased coordination number since there is a minimum resolution, or fraction of conformational space, required for its definition. Inasmuch as z for proteins has been estimated to be approximately 3.8,[2] then the present model should not grossly misrepresent that fraction.

Microphase Structure of Block Ionomers. 1. Study of Molded Styrene–4-Vinylpyridinium ABA Blocks by SAXS and SANS

Jean-Pierre Gouin,† Claudine E. Williams,‡ and Adi Eisenberg*,†

Department of Chemistry, McGill University, Montréal, Québec, Canada H3A 2K6, and Laboratoire pour l'utilisation du rayonnement électromagnétique (LURE), CNRS-MEN-CEA, Université Paris-Sud, 91405 Orsay Cedex, France.
Received November 23, 1988; Revised Manuscript Received April 11, 1989

ABSTRACT: The bulk morphology of 4-vinylpyridinium-*b*-styrene-*b*-4-vinylpyridinium ABA block ionomers with short ionic blocks has been studied by small-angle X-ray scattering (SAXS) as a function of the length of both A and B blocks. A broad asymmetric intensity maximum is observed for each sample and is interpreted as arising from correlation between phase-separated ionic domains on a paracrystalline lattice. The length of the ionic segments is found to have a much stronger effect on the average characteristic distance than that of the nonionic ones. Comparison of spherical, cylindrical, or lamellar morphologies of the ionic domains points to a pseudocubic arrangement of elongated or flat large multiplets. The morphology of the nonionic precursors of some of the above materials has been studied by small-angle neutron scattering (SANS). The shapes of the scattering patterns were found to be generally similar to those of the corresponding ionic materials except for peak positions and relative widths.

Introduction

Ionic aggregation in ion-containing polymers, specifically ionomers, has been the subject of extensive interest in the last 20 years.[1-7] Ion pairs tend to aggregate in media of low dielectric constant, and the presence of these ionic aggregates modifies the properties of polymers profoundly; thus, it is of continuing interest to elucidate the relationship between the structure of the aggregates and the physical properties of the materials. Most studies have

† McGill University.
‡ Université Paris-Sud.

0024-9297/89/2222-4573$01.50/0