(4) Abe, A.; Mark, J. E. *J. Am. Chem. Soc.* **1976**, *98*, 6468.
(5) Abe, A.; Hirano, T.; Tsuruta, T. *Macromolecules* **1979**, *12*, 1092.
(6) Riande, E. *J. Polym. Sci., Polym. Phys. Ed.* **1976**, *14*, 2231.
(7) Riande, E. *Makromol. Chem.* **1977**, *178*, 2001.
(8) Guggenheim, E. A. *Trans. Faraday Soc.* **1949**, *45*, 714.
(9) Smith, J. W. *Trans. Faraday Soc.* **1950**, *46*, 394.
(10) McClellan, A. L. "Tables of Experimental Dipole Moments"; W. H. Freeman: San Francisco, Calif., 1963; Vol. I. *Ibid.* Rahara Enterprises: El Cerrito, Calif., 1974; Vol. II.
(11) Leonard, W. J.; Jernigan, R. L.; Flory, P. J. *J. Chem. Phys.* **1965**, *43*, 2256.
(12) Riande, E.; Guzmán, J.; Welsh, W. J.; Mark, J. E., unpublished results.
(13) Flory, P. J. "Statistical Mechanics of Chain Molecules"; Wiley-Interscience: New York, 1969.
(14) Welsh, W. J.; Mark, J. E.; Riande, E. *Polym. J.* **1980**, *12*, 467.
(15) Nogami, N.; Sugeta, H.; Miyazawa, T. *Bull. Chem. Soc. Jpn.* **1975**, *48*, 3573.
(16) Sakakibara, M.; Matsuura, H.; Harada, I.; Shimanouchi, T. *Bull. Chem. Soc. Jpn.* **1977**, *50*, 111.
(17) In fact the value given for E_{σ_p} in ref 14 (0.7 kcal mol^{-1}) is somewhat larger than that reported in ref 2 (0.4 kcal mol^{-1}). In spite of this difference, the agreement between both values is still reasonable, considering the uncertainty in general in such calculated results.
(18) Flory, P. J. *Macromolecules* **1974**, *7*, 381.
(19) Riande, E.; Guzmán, J. *Macromolecules* **1979**, *12*, 952, 1117.

Evolution and Stability of Polypeptide Chain Conformation: A Simulation Study[1]

D. C. Rapaport[2a] and H. A. Scheraga*[2b]

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.
Received October 23, 1980

ABSTRACT: The evolution of the conformations of short polypeptide chains has been studied by using a variant of the Monte Carlo method. A detailed molecular model is used in which the atoms and bonds are represented in full; short- and long-range interactions, hydrogen bonding, and bond torsional energies are incorporated by means of the empirical ECEPP scheme. Partial and complete folding of the chains from initially open conformations was observed, and simulations begun in a helical state were observed to remain in the same conformation throughout the run. The computations made use of a high-speed array processor; this permitted simulation runs of lengths barely attainable on the fastest of large scientific computers currently available.

I. Introduction

It is well established that the amino acid sequence and the surrounding solvent dictate the unique structural form in which a globular protein exists in its native state in solution.[3,4] A considerable degree of effort has been invested in attempting to determine the nature of the interactions within the molecule that are responsible for this three-dimensional structure and for the folding pathways by which the molecule evolves from the denatured coiled state to the well-structured native state, but, despite limited success,[4] a great deal of work remains to be done.

A related, but somewhat more tractable problem involves the conformational properties of shorter polypeptide chains containing some 10 residues, as opposed to proteins which generally consist of more than 100 residues. The two problems are connected by the fact that the same structural units and interactions are present in each; the simplification results from the fact that, in any attempt to simulate these molecules numerically, the shorter the chain the fewer the number of degrees of freedom, and hence the greater the chance of sampling a significant proportion of the conformational space of the molecule.

In this paper we report on a series of numerical simulations of peptide chains using a Monte Carlo type approach. The chain is placed in a particular initial state and its conformation is allowed to evolve by means of small random changes of the internal coordinates; whether or not a particular change is accepted is determined by a Monte Carlo procedure which examines the energy differences involved. The simulation follows the conformation of the chain over a great many such changes, as many as 5×10^6. A variety of different kinds of behavior are found to occur; these will be detailed subsequently, but several of the runs lead to the formation of apparently stable right- or left-handed α-helices from initially extended chain conformations, whereas others did not succeed in producing persistent, recognizable structures.

The simulations described here are considerably more extensive than those of previous Monte Carlo work,[5-10] both in regard to the nature and extent of the sampling involved and also in the fact that the often drastic simplifications of the underlying model necessary to reduce the computational requirements in some of the earlier studies[5-8] have been obviated (see below). The key to the ability to carry out lengthy computations of the kind required in this study lies in the use of an array processor. This device will be described in greater detail below, but briefly, it is a small high-speed processor, supervised (in our case) by a minicomputer, which, if carefully programmed, can exhibit a performance comparable to, or even exceeding, most of the large-scale scientific computers currently in use. More significantly, this performance is achieved at a small fraction (less than 5%) of the cost of the large machines.

The earliest simulations[5] employed hard-sphere models; such systems lack the attractive forces which are essential for the appearance of ordered three-dimensional chain structure. A later Monte Carlo study[6] using a more detailed residue model failed to include the interactions between residues; effects arising from excluded volume and long-range interactions could not, therefore, be observed. Several subsequent studies[7-10] included all interresidue interactions. In some,[7-9] the sampling scheme permitted individual residue conformations corresponding only to low-energy states of the terminally blocked single residue; this approach excluded those low-energy conformations of the full polypeptide in which attractive long-range interactions compensate for any residues not in states com-

0024-9297/81/2214-1238$01.25/0

patible with this restriction. A later study[10] used a biased sampling scheme described earlier[6] in which the frequency of appearance of each residue in a particular conformation was weighted by using the appropriate terminally blocked single-residue energy, thereby favoring low-energy regions of the single-residue energy surface. While, in principle, this approach to Monte Carlo is correct, there is a distinct possibility that the ability to sample those low-energy regions of the *entire* polypeptide, which do not correspond to those in which each and every residue is in its own low (single-residue) energy state, will not be satisfactory. An additional restriction in each of these studies[5-10] is that the dihedral angles are confined to discrete sets of values. In the work described in this paper, the Monte Carlo algorithm avoids explicit reference to the properties of single residues; furthermore, the dihedral angles are allowed to vary over a continuous range of values.

Details of the calculations are described in sections II–IV; these include the chain structure and method for generating conformations (section II), the interaction potential (section III), and the Monte Carlo algorithm (section IV). An overview of the array processor, insofar as it bears on these calculations, and an outline of the implementation of the algorithm are given in section V. The results of the simulation, primarily graphical sequences illustrating chain evolution, are described in section VI. Finally, section VII provides a brief discussion of the limitations of the approach and an indication of possible extensions of the analysis.

II. Structure of the Polypeptide Chain

The model describing the individual amino acid residues from which the polypeptide chain is constructed assumes a rigid geometrical structure.[11] Thus, the bond lengths and bond angles for each residue are fixed, but complete freedom, subject only to the intra- and interresidue potentials (section III), is allowed for the bond dihedral angles. The bond lengths and bond angles are presumed independent both of the position of the residue in the chain and of its immediate neighbors.

The chains considered here are homopolymers, each consisting of a sequence of L-alanine or glycine residues. The backbone is terminated at the α-amino end by an acetyl ($COCH_3$) group and at the α-carboxyl end by a methylamide ($NHCH_3$) group. The structural parameters of both residues and end groups are given in ref 11. The variable dihedral angles for a particular residue i are ϕ_i for the $N-C^\alpha$ bond, ψ_i for the $C^\alpha-C'$ bond, and ω_i for the peptide $C'-N$ bond; for alanine there is also a single side-chain dihedral angle χ_i^1 about the $C^\alpha-C^\beta$ bond.

Once the values of the variable dihedral angles have been specified, the determination of the Cartesian atomic coordinates for that conformation follows a general procedure applicable to all kinds of residues, including those with branched side chains (preliminary studies involving more complicated residues have been carried out but are not reported here). We assume that the polypeptide chain has been constructed from the available residue data and a set of reference atomic coordinates generated for (say) all the dihedral angles set equal to zero.

The generation of coordinates for a particular chain state proceeds by means of an iterative scheme starting with the atoms at the N terminus and finishing at the C terminus. To start the process, all atoms which are not affected by a rotation about the first variable bond are placed in position using the reference coordinates. The next step is to determine the range of atoms affected by the rotation about the first variable bond; the rotation matrix for this group of atoms is evaluated, and the appropriate rotation and translation operations are applied to the reference coordinates to place the atoms in position. This procedure is repeated—each group of atoms whose locations are affected by the variation of a particular backbone dihedral angle but not its successor are identified, the rotation matrix for that variation is computed and premultiplied by a product of the rotation matrices for all the preceding bonds, and the resulting matrix is used to perform the rotation. Side chains are treated in a similar fashion, although it is necessary to record the state of the rotation matrix product immediately prior to branching away from the backbone and to restore its value prior to processing further backbone atoms. An extension of this procedure treats further branches within side chains.

The actual construction of a rotation matrix is a straightforward exercise in trigonometry. Consider the bond joining atoms k and $k+1$ about which a clockwise rotation θ (as viewed from k to $k+1$) is to occur. Assume that in the reference state of the chain the direction cosines of the bond are c_1, c_2, c_3. If $\kappa = \cos\theta$ and $\sigma = \sin\theta$, then the total rotation matrix $\mathbf{R}$ for the atoms affected by rotation θ is related to the preceding rotation matrix $\mathbf{R}'$ (used for atoms not affected by θ) by $\mathbf{R} = \mathbf{R}' \times \mathbf{B}$, where $\mathbf{B}$ is the matrix

$$\mathbf{B} = \begin{bmatrix} \kappa' c_1^2 + \kappa & \kappa' c_1 c_2 - \sigma c_3 & \kappa' c_1 c_3 + \sigma c_2 \\ \kappa' c_1 c_2 + \sigma c_3 & \kappa' c_2^2 + \kappa & \kappa' c_2 c_3 - \sigma c_1 \\ \kappa' c_1 c_3 - \sigma c_2 & \kappa' c_2 c_3 + \sigma c_1 & \kappa' c_3^2 + \kappa \end{bmatrix} \quad (1)$$

with $\kappa' = 1 - \kappa$. The initial value of the cumulative rotation matrix is of course the unit matrix.

III. Computation of the Interaction Energy

The interactions between pairs of atoms within the molecule are computed by means of the ECEPP (empirical conformational energy program for peptides) scheme.[11] Only the variable portion of the total interaction energy is computed; those pairs of atoms whose separations are invariant under a change of dihedral angle are excluded from the scheme. ECEPP has been employed in a variety of studies of oligopeptides and proteins;[4] to date the computational approach has involved energy minimization, and this is the first attempt to use it within a Monte Carlo framework. Briefly, ECEPP represents an attempt to summarize both experimental and theoretical results for peptides in a unified manner; the formulas thus obtained are assumed to be applicable to polypeptides and proteins without further modification.

The total interaction energy E of a specified molecular conformation is constructed as a sum of several components: an electrostatic term E_{el}, a term covering the interaction between nonbonded pairs of atoms E_{nb}, a term E_{hb} representing the energy of hydrogen-bonded pairs, and finally a bond torsional term E_{tor}. It should be emphasized that this decomposition is for convenience of parameterization only and is not to be taken too literally.

The detailed forms of the individual components are as follows: The electrostatic term is

$$E_{el} = \sum_{\langle ij \rangle} q_i q_j / D r_{ij} \quad (2)$$

where the sum is over all pairs of atoms $i...j$ whose separation depends on the values of one or more *variable* dihedral angles, r_{ij} denotes the corresponding interatomic distance, q_i and q_j are the partial atomic charges derived from CNDO/2 calculations,[11] and D (=2) is an effective dielectric constant. The nonbonded contribution is based on the Lennard-Jones 6–12 potential

$$E_{nb} = \sum_{\langle ij \rangle} f_{ij} A_{t_i t_j} / r_{ij}^{12} - B_{t_i t_j} / r_{ij}^{6} \quad (3)$$

Here, the sum is over all pairs not designated as hydrogen bonded (see eq 4) which are separated by at least one variable dihedral angle. The indices t_i and t_j denote atom type (a combination of both atomic number and type of bonding), the coefficients A and B depend on the types of both members of the atom pair, and the factor f_{ij} has the value 0.5 if only one variable dihedral angle separates the pair and unity otherwise. Those pairs of atoms which form hydrogen bonds (certain H···O and H···N pairs) interact by means of a modified Lennard-Jones interaction

$$E_{hb} = \sum_{\langle ij \rangle} A'_{t_i t_j}/r_{ij}^{12} - B'_{t_i t_j}/r_{ij}^{10} \quad (4)$$

where the coefficients A' and B' are again pair dependent. The torsional term is a sum of contributions from each of the peptide and variable side-chain bonds

$$E_{tor} = (1/2)\sum_k U_k\{1 + (-1)^{l_k} \cos(n_k\theta_k)\} \quad (5)$$

where U_k is the barrier height, n_k (=2, 3) is the symmetry factor, and l_k (=0, 1) allows for the barrier maximum to be located at dihedral angle values $\theta_k = 0$ or 180°. All the parameters in eq 2–5 are to be found in ref 11.

IV. Monte Carlo Algorithm

The general features of Monte Carlo simulation are too well-reviewed to be repeated here.[12] Our application of the method to the polypeptide problem does, however, differ in one important respect from some earlier work on the subject.[5,6,9,10] Whereas the previous studies involved the generation of a set of *independent* chains, we have chosen to generate sequences of *conformationally correlated* chains, each chain being derived from its predecessor by a small variation in the dihedral angles. Thus, in a sense, the sequence of conformations can be taken to represent the conformational evolution of the molecule. Clearly, no true dynamics are involved in the Monte Carlo approach, but it is reasonable to expect that the stochastic evolution of conformation produced by Monte Carlo sampling should resemble, at least qualitatively, the evolution of a polypeptide in solution under the influence of the essentially random interactions with the solvent molecules.

In its most frequently used form, the Monte Carlo procedure leads to a Boltzmann distribution of states once equilibration is complete; thus, states of low free energy are favored (in contrast to energy minimization which searches for an energy minimum rather than a free energy one), and if, for example, a helix corresponds to such a state, its appearance and relative stability ought to be observed during the simulation. In this approach to Monte Carlo, however, and, for that matter, also in molecular dynamics simulation, there is a strong possibility of chain entanglement; given sufficient computation time, the chain may succeed in extricating itself, but this is a problem of principle not readily overcome.

A simulation run typically begins with the molecule in an ordered (e.g., helical or planar zigzag) or a randomly generated conformation. A single step in the simulation consists of taking the current state Υ, modifying it in some way to generate a new trial state Υ', and performing an appropriate test to decide whether the next state should be Υ' or Υ. A simulation run may consist of as many as several million such steps.

Assume that Υ corresponds to the dihedral angle set $\{\theta_i\}$, with E being the total interaction energy. The trial state Υ' is constructed by changing the *entire set* of dihedral angles to $\{\theta_i'\}$, where each $\theta_i' = \theta_i + \delta_i$ and δ_i is a uniformly distributed random value in a specified range $(-\Delta, +\Delta)$. The atomic coordinates for Υ' are computed and the energy E' is evaluated. The Monte Carlo test then amounts to accepting Υ' in place of Υ with probability $p = \min[1, \exp(E - E')/kT]$, where k is the Boltzmann constant and T the absolute temperature (here T = 298 K); if Υ' is rejected, then the previous Υ is retained for the next step in the simulation.

The quantity Δ, the limit of angular change during a single step, is a free parameter whose value should be chosen to optimize the efficiency of the simulation. There are, however, two conflicting requirements for Δ; its value must be sufficiently small to maintain a degree of continuity between successive molecular states, while at the same time it must be large enough to permit significant evolution over a limited number of steps. The actual criterion used in selecting a value for Δ, which turns out to satisfy the above requirements, is that the root-mean-square change in the dihedral angles over a specified number of Monte Carlo steps should be as large as possible. This is equivalent to maximizing the "diffusion" rate in configuration space, an ideal goal in any Monte Carlo simulation. The value chosen was $\Delta = 2°$; this resulted in a root-mean-square change in dihedral angle of between 10 and 40° over 200 steps, with a corresponding acceptance ratio ranging from 10 to 70%, depending on the degree of packing of the molecule.

During the course of each simulation run, the energy, end-to-end distance, and dihedral angles were monitored and recorded at regular intervals for subsequent analysis. The monitoring made it possible to determine, albeit subjectively, which of the runs should be allowed to continue for long periods, and which not. Typically, those in which the onset of an ordered state was apparent were allowed to continue, though clearly there is no guarantee that a terminated run would not have evolved into such a state if permitted to continue. The availability of only finite computing resources forces the adoption of this pragmatic approach. A summary of the longer runs made to date is included in section VI.

V. Array Processor Implementation

In recent years, an answer to the problem of achieving high-speed floating-point computations at low cost has appeared in the form of the array processor (AP). A number of machines of this kind exist, the most widely used probably being the AP-120B produced by Floating Point Systems Inc. This machine is intended to be run as a slave attached to a host computer, in our case a Prime minicomputer. An impression of the performance capability of this AP, whose cost is that of a typical medium-scale minicomputer system, can be gained from the fact that it is able to run simulations—both Monte Carlo and molecular dynamics—of interacting many-body systems at speeds several times faster than large scientific computers typified by the IBM 370/168.

The task of software development for the AP can be approached at several levels, with a remarkable variation in efficiency being observed across the range. The program to be run on the AP must be developed on the host computer using specialized software tools; the programming language may be Fortran or a special AP assembly language. Once developed, it is combined with a program running on the host whose primary task is to supervise the AP operation and to ensure the correct transfers of programs and data between the machines. Fortran AP programming permits a short development cycle; the disadvantages of this approach are that the full capability of the AP is not realized (typically the performance is only 10–20% of a carefully written assembly language program) and, furthermore, there are also limitations on the size of program which can be stored in the AP (the Monte Carlo

program, for example, would exceed the available memory).

At the other extreme lies the assembly language program, with the most heavily used code segments carefully optimized to achieve maximal hardware utilization. This can be a relatively complex task, for the AP architecture relies heavily on both parallel operation and pipelining to achieve its high throughput. The implication is that the software is responsible for ensuring that the various pipelines (namely, the independent floating-point addition and multiplication units and the fetching or storing of data) are used effectively and that data are supplied to or removed from these pipelines at the correct moments. The theoretical upper limit to the performance is 12 million floating-point operations per second, together with as many other operations (integer arithmetic, memory accesses, tests for branches, etc.) as can be fitted into each 0.167-μs machine instruction. For a full description of the capability of the machine, the original literature should be consulted.[13-15]

Though not originally intended for simulations of many-body systems—the machine was designed for signal-processing applications—it turns outs to be extremely well suited for calculations of this kind if one is prepared to ignore a single caveat, namely, potential problems of software development. Various successful applications have been surveyed recently;[14] on our machine, in addition to the Monte Carlo work described here, software has been developed for use in energy minimization computations,[15] and a large-scale molecular dynamics simulation of water has been implemented.[16]

In order to gain an impression of the amount of computation involved in a typical polypeptide Monte Carlo run of 10^6 or more steps, consider the following facts: A typical chain studied consisted of 8 alanine residues with end groups, for a total of 92 atoms; thus, there are some 4000 pairs of atoms whose mutual interactions must be computed. Each such computation involves a calculation of the separation r_{ij}, itself the square root of a sum of squares, followed by the calculation of various negative powers of r_{ij} and several other additions and multiplications required to evaluate a single term in eq 2-4. For the full run, a calculation of this kind must be repeated some 4×10^9 times. Only a carefully designed algorithm coded in assembly language is capable of handling a computation of this magnitude while using only a moderate amount of computer time.

The major part of the computational effort is the evaluation of the energy contributions defined by eq 2-4. Relative to this, the other portions of the calculation, viz., the generation of coordinates and the associated matrix algebra, the evaluation of eq 5, and the Monte Carlo decision process, are very modest in their computational requirements. In evaluating the energy sums of eq 2-4, we follow the general lines used in the energy minimization approach,[15] though several changes of detail occurred. Thus, the sum in eq 3 is split into independent parts; for a given atom i, the majority of atoms j with which it is paired form a consecutively numbered sequence, and for each such pair $f_{ij} = 1$. Such portions of the sum (or the analogous term of eq 4) together with the corresponding contributions from the sum of eq 2 are evaluated by a carefully designed program loop; the remaining terms, far fewer in number, are processed separately. The optimized central loop utilizes the parallel processing and pipelining capability of the AP to the extent that five pairs of atoms are being processed at any one time, and individual contributions to the energy sum are generated at a rate exceeding one per 3 μs.

Table I
Summary of Runs[a]

run no.	chain	initial	final	no. of steps
1	8 Ala	RH helix	unchanged	1.0×10^6
2	8 Ala	planar zigzag	part RH helix	4.0×10^6
3	8 Ala	planar zigzag	part RH helix	4.7×10^6
4	8 Ala	random	part LH helix	3.5×10^6
5	8 Ala	random	part LH helix	4.0×10^6
6	8 Ala	planar zigzag	RH helix	5.0×10^6
7	12 Ala	RH helix	unchanged	1.3×10^6
8	12 Ala	RH helix	unchanged	1.0×10^6
9	12 Ala	LH helix	unchanged	0.7×10^6
10	12 Ala	planar C	part C	0.6×10^6
11	12 Ala	planar zigzag	part C	4.2×10^6
12	8 Gly	planar zigzag	mixed	4.6×10^6
13	12 Gly	RH helix	unchanged	1.1×10^6
14	12 Gly	LH helix	part LH helix	1.5×10^6

[a] The initial conformations are to be interpreted as follows: RH helix: $\phi = -70°$, $\psi = -50°$, $\omega = 180°$, $\chi^1 = 60°$ (if present); planar zigzag: all dihedral angles = 180°; random: all dihedral angles random except $\omega = 180°$; LH helix: $\phi = 60°$, $\psi = 60°$, $\omega = 180°$; planar C: $\phi = -80°$, $\psi = 80°$, $\omega = 180°$, $\chi^1 = 60°$. In the description of final conformations, the term "part" is used to indicate that a significant proportion of residues lies in the stated region.

Unlike the earlier energy calculations,[15] which made considerable use of the host computer resources in addition to the AP, the Monte Carlo simulations are carried out entirely within the AP. The amount of data which must be transferred between the host and AP during the run is therefore minimal, with the result that the simulation is able to proceed at a rate of some 2×10^5 steps/h (for the 8-alanine chain mentioned earlier) with little or no impact on the other users of the time-shared host minicomputer. A 1 000 000-step run requires a mere 5 h, hardly a heavy request for a minicomputer system. In order not to prevent other users from using the AP for short periods while a run was in progress, a checkpointing scheme was implemented which enabled the current state of the simulation to be preserved on disk, and the simulation continued when the AP became available again.

As part of the energy minimization study[15] comparisons were made of the relative speeds of the AP and a large scientific computer, the IBM 370/168. These benchmarks indicated a speed increase of between 5 and 11 in favor of the AP; the ratio depends on whether the IBM version of the program was coded in Fortran or assembly language. A similar speed ratio also applies to the Monte Carlo computations.

VI. Results

Simulation runs were carried out for homopolypeptide chains of 8 and 12 L-alanine or glycine residues, with end groups as indicated in section II. The runs were of various lengths, and a full breakdown of the chain types, numbers of Monte Carlo steps, and initial and final conformations are summarized in Table I. It should be stressed that each step involved an attempt to alter *all* the dihedral angles in the molecule—backbone, side chains (for alanine), and end groups.

Early in the calculations it became clear that, for the lengths of runs studied here, chains of 8 residues are capable of exhibiting a great variety of conformational forms, whereas 12-residue chains evolve in a more limited fashion. This is not surprising, but it does imply that chains beyond a certain size cannot be studied by this kind of approach which imposes no a priori restriction on the dihedral angle ranges. Some short simulation runs for 16-residue chains were not very encouraging in this respect and tended to

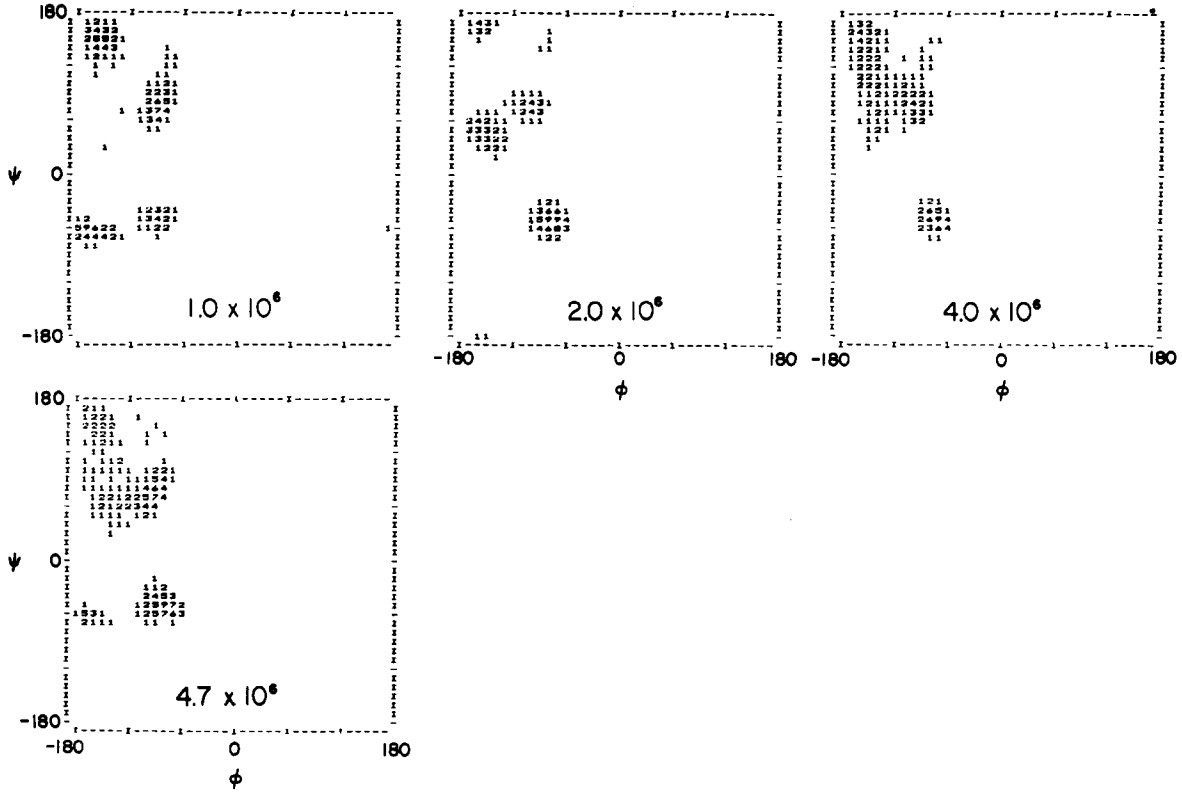

Figure 1. Chain of 8 L-alanines (no. 3 in Table I) evolving from an initially planar zigzag to a partial right-handed helical conformation.

confirm this conclusion.

Little is to be gained from a detailed presentation of data from all the runs; we have therefore selected several typical runs and propose to discuss the key aspects of the behavior of each. The results will be presented in the form of two-dimensional histograms which measure the joint distribution of the backbone dihedral angles ϕ (N–C$^\alpha$) and ψ (C$^\alpha$–C′). The histograms in fact record the frequency of appearance of pairs of values (ϕ,ψ) for *each* of the residues in the chain, excluding the terminal groups; only a single histogram is maintained for the entire chain; separate results for individual residues are not tabulated. Each histogram spans a sequence of 10^5 Monte Carlo steps; every 200 steps the current (ϕ,ψ) pairs are used to update the histogram. The quantization interval is 10°; thus the histogram is based on a 36 × 36 grid. Once accumulation of data for a histogram is complete, the values are scaled to integers in the range 0–9, and it is these results which are used in the figures which follow. The zero entries have been suppressed for clarity, and the step number included with each histogram is the number of the final contributing step.

Figure 1 shows four characteristic histograms from a 4.7 × 10^6 step run of an 8-alanine residue chain. The run started with the chain in an extended, planar zigzag conformation, with all dihedral angles set to 180°. After 2 × 10^6 steps almost half the residues tend to lie in the right-handed α-helical region, namely, $(\phi,\psi) \approx (-70°,-50°)$, but this ratio has dropped to about 30% by the time the run ended. Examination of the final structure (not shown) revealed that the central 4–5 residues lay in an essentially helical conformation.

The larger set of histograms in Figure 2 illustrates the evolution of a separate 8-alanine chain, again starting from a planar conformation, but using a different random number sequence to generate the conformations (section IV). After only 0.5 × 10^6 steps a significant proportion of helical residues are present. Subsequent histograms show

Table II
Dihedral Angles (Deg) Corresponding to the RH Helix Shown in Figure 3

residue	ϕ	ψ	ω	χ^1
$COCH_3$		−60.5	−170.4	
Ala	−79.5	69.3	−170.9	−79.9
Ala	−94.2	−10.6	171.4	−166.1
Ala	−64.8	−26.7	−174.9	61.0
Ala	−81.2	−19.0	177.0	−163.0
Ala	−88.6	−40.4	−179.8	−59.7
Ala	−64.4	−44.7	−172.5	49.0
Ala	−68.9	−53.2	−168.7	−62.4
Ala	−74.8	−46.4	174.4	−171.6
$NHCH_3$	−145.4			

that the preferred states move away from the helix region of the (ϕ,ψ) map [see Figure 1 of ref 15 for a characterization of the various (ϕ,ψ) regions] but somewhere between 2.7 × 10^6 and 2.8 × 10^6 steps there is a rapid shift back to the right-handed α-helical state which is complete after 3 × 10^6 steps. The molecule remains in this region of (ϕ,ψ) space for the rest of the 5 × 10^6 step run. Two perspective views of the final state of the molecule are shown in Figure 3; the helical form is quite apparent, and while some of the hydrogen bonds may be somewhat stretched, their overall orientation is consistent with that expected for a helical conformation. Table II lists the values of all the dihedral angles in this final state.

A corresponding run for a glycine chain is shown in Figure 4. The initial state is once again planar, and while there is a considerable variation in the structure over the course of 4.6 × 10^6 steps, there is no hint of a stable structure emerging. In fact the (ϕ,ψ) distribution tends to reflect the shape of some of the low-energy regions of the single-residue energy map (ref 17, Figure 3).

Figure 5 shows the case of an 8-alanine chain started in a randomly generated state, with the exception of the peptide bond dihedral angles (ω), which were initially set to 180°. On examination, the initial state was not found

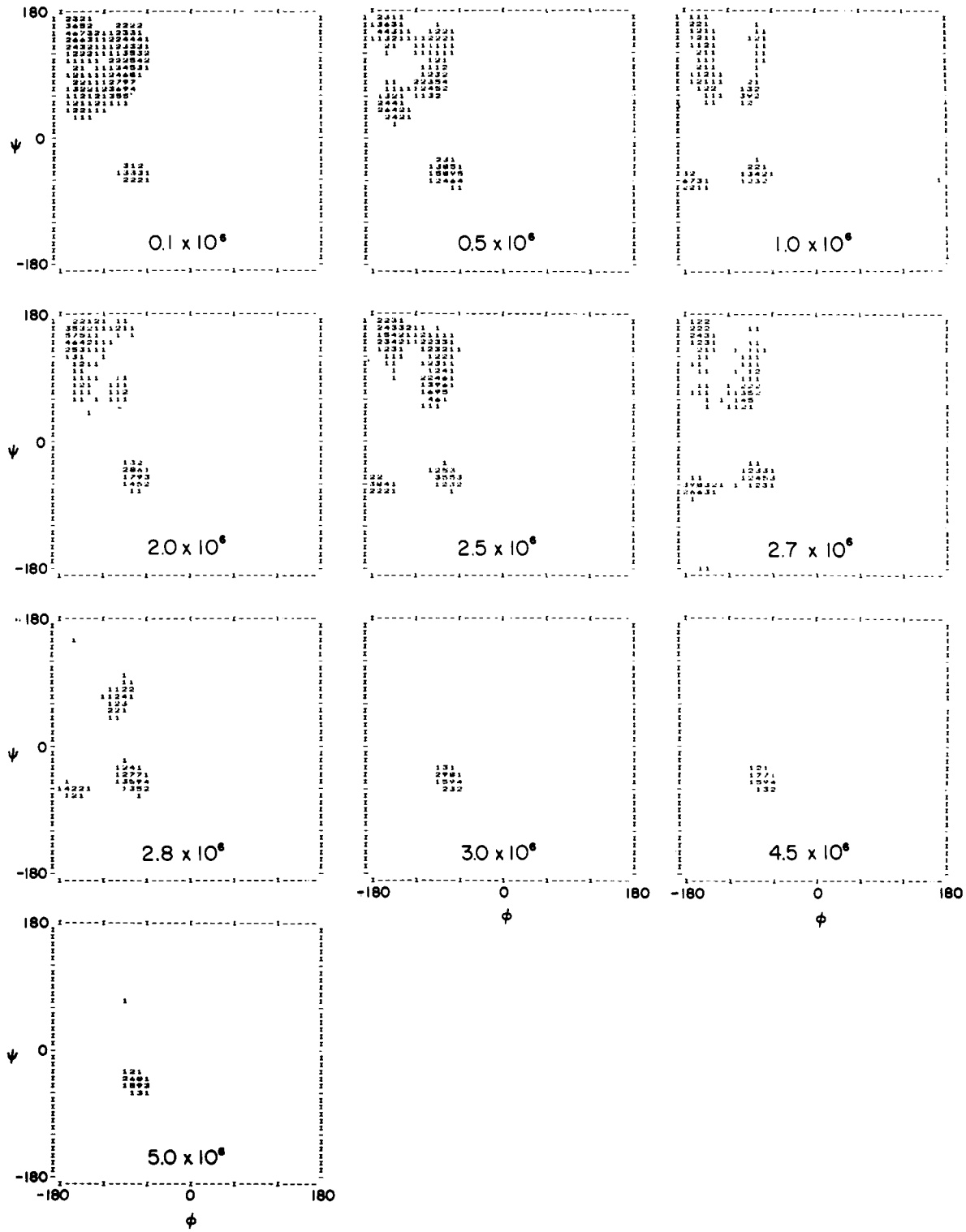

Figure 2. Chain of 8 L-alanines (no. 6) evolving from a planar state to a stable right-handed helix. The final stage of helix formation between 2.7×10^6 and 3×10^6 steps should be noted.

to show any preference toward the left-handed α-helical region for which $(\phi,\psi) \approx (60°,60°)$; nevertheless, the molecule rapidly evolves into states where over half the residues lie in this region. The left-handed helix does in fact correspond to a relative low-energy state for the single residue.[17]

An example of a 12-alanine chain appears in Figure 6. While a certain fraction of the residues lie in the right-handed helical region throughout the run, there is a pronounced preference for other regions of (ϕ,ψ) space, in particular those corresponding to low-energy states of the single residue (ref 17, Figure 2).

The final example, appearing in Figure 7, illustrates a somewhat different aspect of the simulation, viz., the problem of disrupting what is presumably a relatively stable structure. This simulation is of a chain of 12 glycine residues initially arranged as a right-handed α-helix. After a run of 1.1×10^6 steps there is still no indication that the helix is about to rupture.

It is clear from the behavior seen during these simulations that long-range interactions play a crucial role in stabilizing the observed right- or left-handed helical

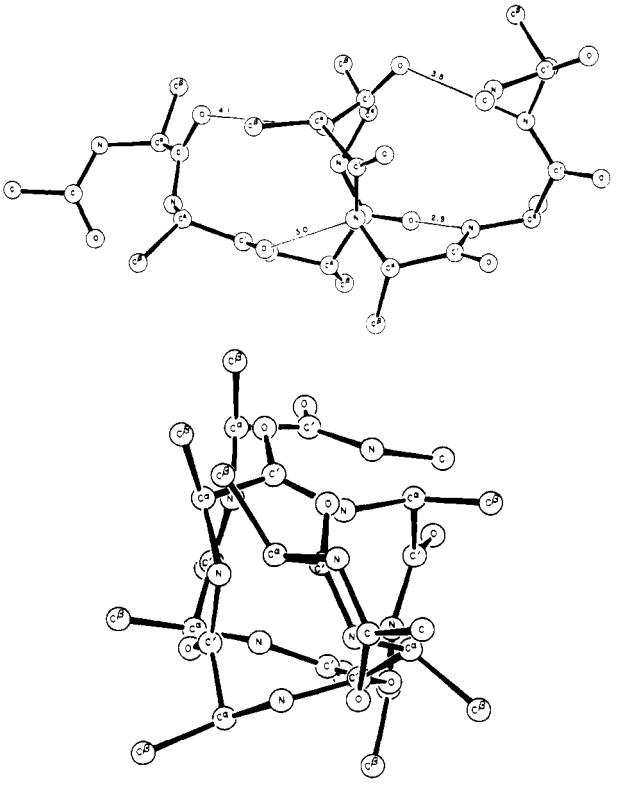

Figure 3. Lateral and end views of the final state of the 8-L-alanine chain of Figure 2. The hydrogen atoms have been omitted for clarity. The corresponding dihedral angles are listed in Table II. The positions and lengths (in Å) of the interior hydrogen bonds are indicated.

structures. It is equally clear, however, that the energetic preferences of single and immediately adjacent residues—the short-range interactions—are strongly felt in the distributions of dihedral angles of the polypeptides. This latter observation in no sense justifies the sampling schemes used in previous Monte Carlo studies which either restricted[7-9] or weighted[10] the individual residue conformations using single-residue energies; the failure to treat short- and long-range interactions on an equal footing may well destroy the delicate balance of interactions which helps determine polypeptide and protein structure.

VII. Discussion

An important consideration in attempting to decide on ways of extending the scope of these simulations is that some 200 h of AP time were expended on the work reported here. If we recall the great speed of the AP (see section V) it is clear that computations whose aim is the systematic exploration of the evolution of the structures of polypeptide molecules from various starting conformations, or the study of the change in helix stability with increasing chain length—to cite two examples—probably lie beyond the limits of present-day computers (at reasonable cost). In view of the fact that the time scales over which polypeptide and protein folding occurs are several orders of magnitude longer than the time span covered by a typical computer simulation (a molecular dynamics study is unlikely to exceed 10^{-10} s, and the Monte Carlo approach is of roughly equivalent efficiency, even though a time variable is not explicitly present), there is an incompatibility that even a considerable improvement in computer technology may be unable to bridge.

There is yet another problem, namely, that the solvent has been ignored. While the effective dielectric constant in eq 2 has been set to account for a uniform solvent background, it is only by explicit inclusion of the solvent molecules in the simulation that a true picture of the behavior can be reached. A calculation of this kind is far more time-consuming than those involving the isolated chains analyzed here. A highly idealized system of this kind, consisting of a polymer constructed of linked hard spheres immersed in a hard-sphere solvent has indeed been

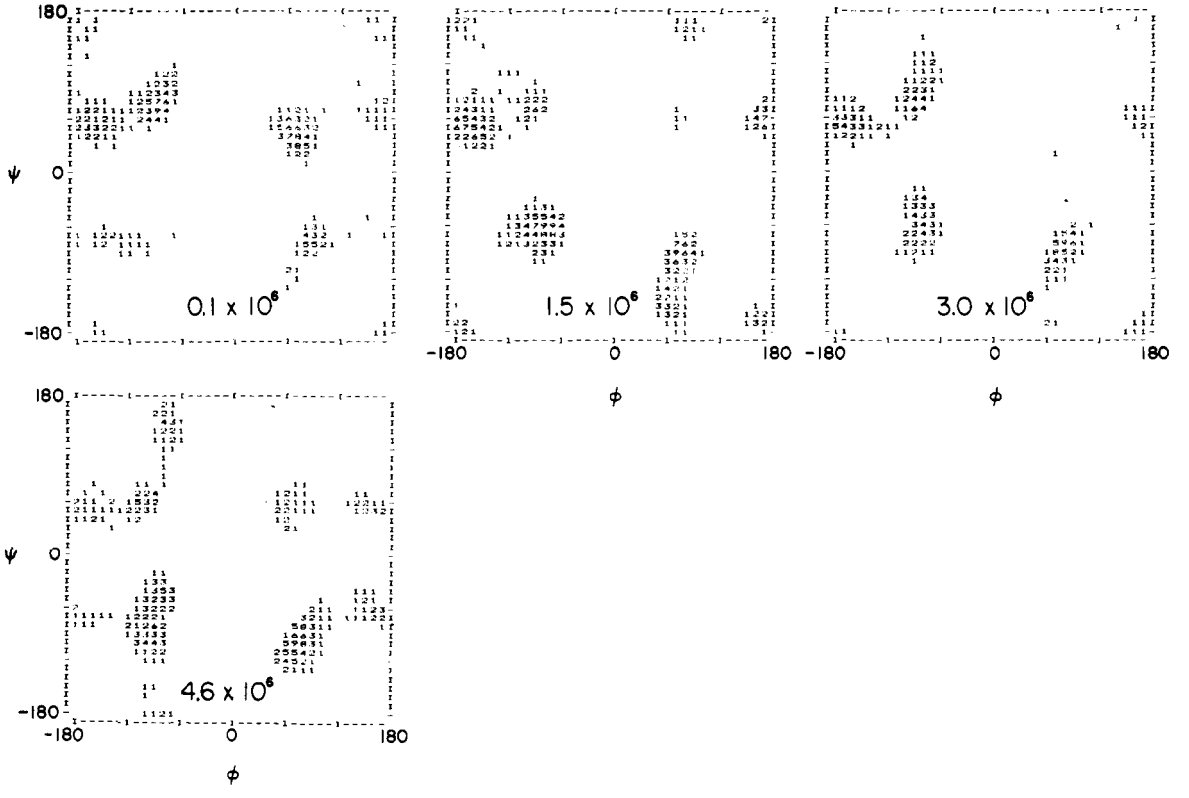

Figure 4. Chain of 8 glycines (no. 12) evolving through various mixed conformations.

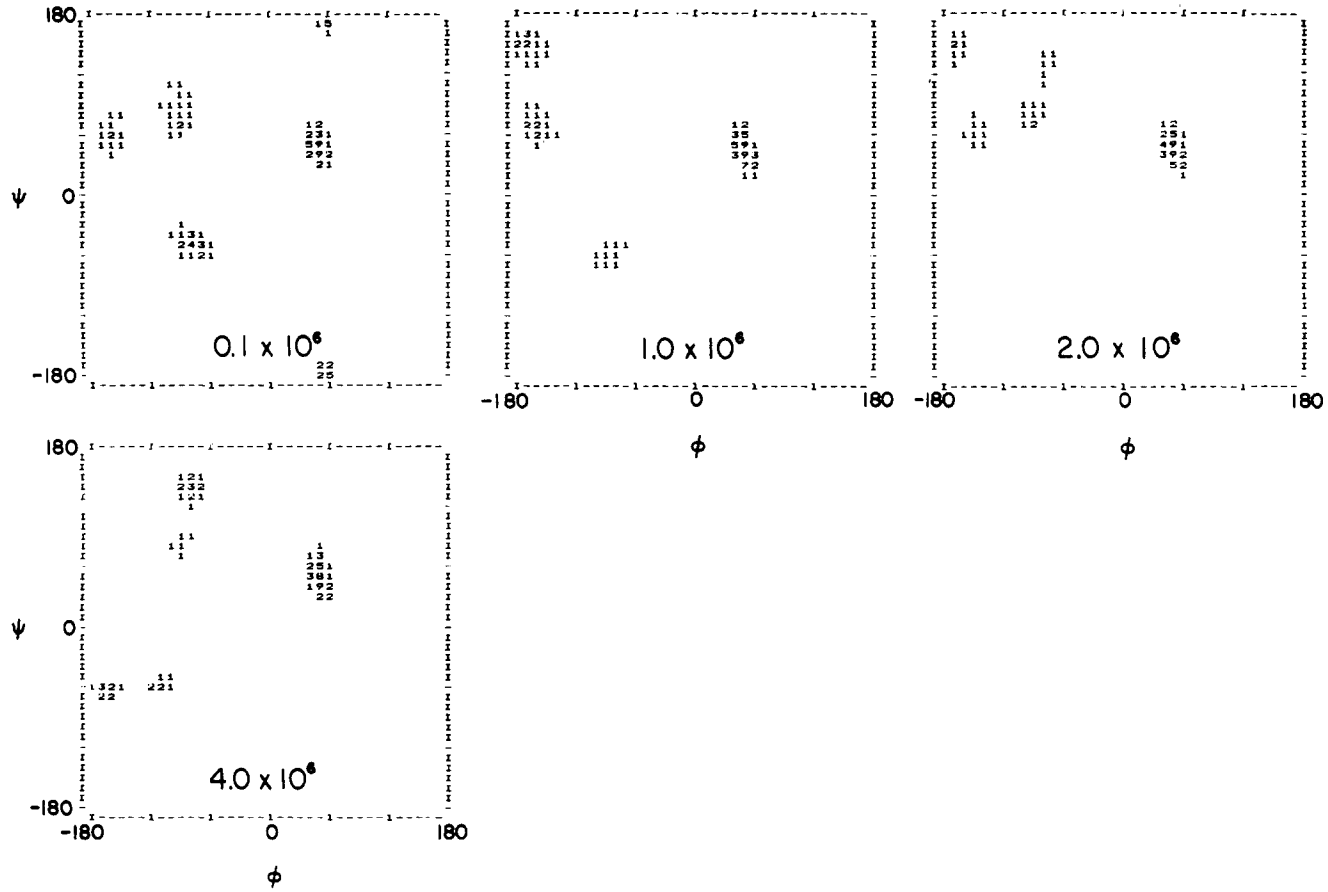

Figure 5. Chain of 8 L-alanines (no. 5) evolving from a random to a partial left-handed helical conformation.

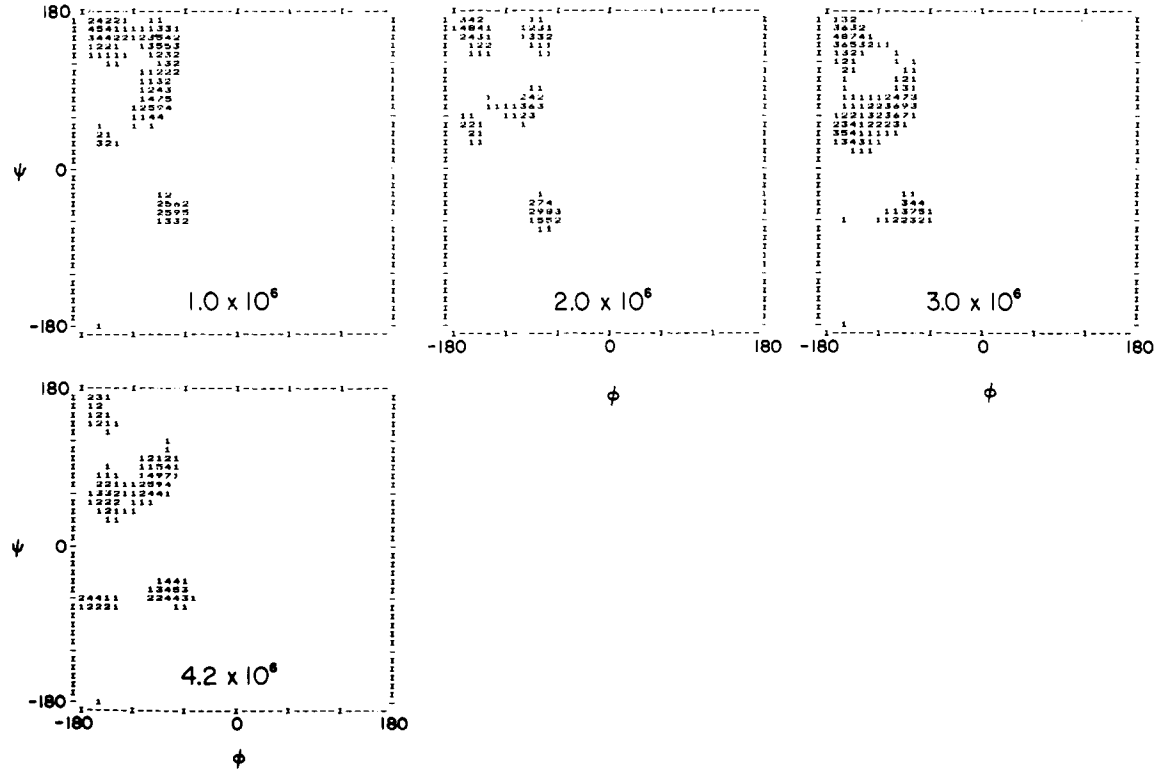

Figure 6. Chain of 12 L-alanines (no. 11).

studied by molecular dynamics techniques;[18] there too it was found that the bulk of the computational effort went into dealing with the large number of solvent molecules rather than with the polymer molecule itself. A more realistic model has been assumed for both solute and solvent in a molecular dynamics study of a terminally blocked single residue in water;[19] however, the very short chain length does not permit the simulation of the development of helical or other ordered structures. An alternative approach to the solvent problem[20] employs the

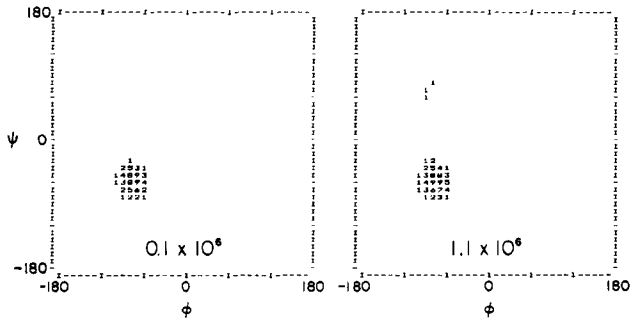

Figure 7. Chain of 12 glycines (no. 13) started and remaining in a right-handed helical state.

Monte Carlo method to simulate the arrangement of water molecules around fixed conformations of a blocked residue.

To strike a more optimistic note, the results of the simulation are encouraging. For the first time, it has been shown that an empirical potential function such as that of ECEPP contains the correct ingredients to fold polypeptides into helices. While the use of such a detailed model in simulations of protein molecules would seem to be beyond the limits of feasibility, it ought to be possible to use simulations of the kind described here to simplify the polypeptide model in ways which do not alter its structural preferences significantly. This simplified, but still physically meaningful model would then be the subject of further simulation, with appropriate solvent effects, in an attempt to emulate the protein-folding phenomenon.

References and Notes

(1) (a) This work was supported by research grants from the National Science Foundation (PCM79-20279, PCM77-09104, and PCM79-18336) and from the National Institute of General Medical Sciences, National Institutes of Health, U.S. Public Health Service (GM-14312 and GM-25138). (b) A preliminary account of this work was presented at the "International Symposium on the Statistical Mechanics of Phase Transitions in Polymers", Cleveland, 1980. *Ferroelectrics* **1980**, *30*, 159.
(2) (a) On leave of absence from the Physics Department, Bar-Ilan University, Ramat-Gan, Israel. (b) To whom requests for reprints should be addressed.
(3) Anfinsen, C. B. *Science* **1973**, *181*, 223.
(4) Némethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.
(5) Warvari, H. E.; Knaell, K. K.; Scott, R. A. *J. Chem. Phys.* **1971**, *55*, 2020.
(6) Premilat, S.; Hermans, J. *J. Chem. Phys.* **1973**, *59*, 2602.
(7) Tanaka, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 3802.
(8) Tanaka, S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1320.
(9) Premilat, S.; Maigret, B. *J. Chem. Phys.* **1977**, *66*, 3418.
(10) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842.
(11) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.
(12) Valleau, J. P.; Whittington, S. G. In *Mod. Theor. Chem.* **1977**, *5*, 137. Valleau, J. P.; Torrie, G. M. *Ibid.* **1977**, *5*, 169.
(13) Whittmayer, W. R. *Comput. Des.* **1978**, *17* (3), 93.
(14) Ostlund, N. S. "Attached Scientific Processors for Chemical Computations: A Report to the Chemistry Community"; NRCC Report, Lawrence Berkeley Laboratory, LBL-10409, UC-32, 1980.
(15) Pottle, C.; Pottle, M. S.; Tuttle, R. W.; Kinch, R. J.; Scheraga, H. A. *J. Comp. Chem.* **1980**, *1*, 46.
(16) Rapaport, D. C.; Scheraga, H. A. *Chem. Phys. Lett.* **1981**, *78*, 491.
(17) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.
(18) Rapaport, D. C. *J. Chem. Phys.* **1979**, *71*, 3299.
(19) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913.
(20) Hagler, A. T.; Osguthorpe, D. J.; Robson, B. *Science* **1980**, *208*, 599.

Photoresponsive Polymers. On the Dynamics of Conformational Changes of Polyamides with Backbone Azobenzene Groups

Masahiro Irie† and Wolfram Schnabel*

Hahn-Meitner-Institut für Kernforschung Berlin GmbH, Bereich Strahlenchemie, D-1000 Berlin 39, Federal Republic of Germany. *Received April 6, 1981*

ABSTRACT: Continuous irradiations and flash photolysis experiments have been carried out with *N,N*-dimethylacetamide and *N,N*-dimethylformamide solutions of a polyamide having azobenzene groups in the polymer backbone and with a low molecular weight model compound. The continuous irradiations revealed that trans → cis isomerization occurs upon irradiation with UV light (350–410 nm) whereas cis → trans isomerization can be induced by visible light ($\lambda > 470$ nm). The kinetics of cis → trans isomerization were studied by time-resolved optical absorption measurements. It turned out that this process occurred to about 90% during the 20-ns flash (530-nm light). The remainder was completed within 100 ns after the flash, indicating the occurrence of a relatively slow relaxation of chain segments. The conformational change of the total macromolecule, subsequent to cis → trans isomerization, was studied by time-resolved light scattering measurements. Relaxation times covering the range from 0.47 (neat DMF) to 1.1 ms (DMF/ethanol, volume ratio 3:2) were found, the difference reflecting the influence of solvent quality.

Introduction

Conformational changes of linear macromolecules in solution have attracted the interest of many researchers in the past.[1] Their work was mostly devoted to polyelectrolytes, and usually equilibrium properties were measured. Comparatively little is known about the dynamics of conformational transitions. In the studies reported so far, stopped-flow or temperature jump techniques in conjunction with optical rotation,[2] circular dichroism,[3] and light scattering[4] measurements have been employed. Usually, the time resolution was not better than 10^{-3} s.

Optical absorption or emission proved to be insensitive for the observation of conformational changes.[5]

In previous papers[6] one of us (M.I.) has reported that linear polyamides with azobenzene groups in the backbone undergo conformational changes upon irradiation with

† On leave from The Institute of Scientific and Industrial Research, Osaka University, Suita, Osaka, Japan.

0024-9297/81/2214-1246$01.25/0