THEORETICAL AND EXPERIMENTAL STUDIES OF CONFORMATIONS OF POLYPEPTIDES

HAROLD A. SCHERAGA

Department of Chemistry, Cornell University, Ithaca, New York 14850

Received October 28, 1970

Contents

I. Introduction

This is a report of the current status of our continuing efforts to understand the nature and magnitudes of the interactions which determine the stable conformations of polypeptides and proteins in solution. It will deal not only with theoretical studies, but also with some related experimental ones which have been carried out in order to examine the implications of various conformational energy calculations. Since both the theoretical[1-6] and experimental[7,8] aspects of this subject have been reviewed several times, we will assume that these reports are available to the reader, and hence will confine our discussion here primarily to our work of the last two years. However, for the sake of continuity and clarity, some of the earlier results will be mentioned briefly throughout this article.

Calculations are being carried out on the basis that stable conformations of polypeptides and proteins in solution correspond to local minima of a function (referred to, for brevity, as the "conformational energy") which is the sum of the potential energy for all intrapolypeptide interactions and the free energy for all interactions involving the solvent. The most stable conformation of a macromolecule is then the one which corresponds to that local minimum of the conformational energy surface which has the largest statistical weight. It is, therefore, of importance to consider the energetic and entropic terms which contribute to the statistical weight of an arbitrary conformation of a molecule. At the same time, it is necessary to consider experimental work which provides the basis for establishing the magnitudes of the various interactions and for testing the various predictions which result from the calculations. The theoretical and experimental aspects of this work will be presented in sections II and III, respectively.

II. Theoretical

A. THEORETICAL FOUNDATIONS

In a series of publications,[9-12] the theoretical foundations for computing stable conformations of macromolecules have been developed. The essential ideas emerging from these papers are the following. A macromolecule in solution is not a static entity. Aside from its translational and overall rotational motion, it undergoes internal vibrations[13-15] which arise from the stretching of bonds, the bending of bond angles, and the variations of dihedral angles for rotation around single bonds.[16] The nature of these internal motions depends on the magnitude of a "conformational energy" function, $F(Q)$, of the generalized coordinates Q; for a molecule *in vacuo*, $F(Q)$ is equal to $U(Q)$, the potential energy of all intrapolypeptide interactions, and for a molecule in solution, $F(Q)$ is given by the sum of $U(Q)$ and $V(Q)$; the latter term, being the free energy for all interactions involving the solvent, includes the

(1) H. A. Scheraga, *Advan. Phys. Org. Chem.*, **6**, 103 (1968).
(2) H. A. Scheraga, *Harvey Lect.*, **63**, 99 (1968).
(3) G. N. Ramachandran and V. Sasisekharan, *Advan. Protein Chem.*, **23**, 283 (1968).
(4) A. M. Liquori, Nobel Symposium 11 on "Symmetry and Function of Biological Systems at the Macromolecular Level," A. Engstrom and B. Strandberg, Ed., Almqvist and Wiksell, Stockholm, 1969, p 101; *Quart. Rev. Biophys.*, **2**, 65 (1969).
(5) (a) G. N. Ramachandran in ref 4, p 79; (b) C. M. Venkatachalam and G. N. Ramachandran, *Ann. Rev. Biochem.*, **38**, 45 (1969); (c) G. N. Ramachandran, *Int. J. Protein Res.*, **1**, 5 (1969).
(6) H. A. Scheraga in ref 4, p 43.
(7) G. D. Fasman in "Poly-α-Amino Acids," G. D. Fasman, Ed., Marcel Dekker, New York, N. Y., 1967, Chapter 11.
(8) M. Goodman, A. Verdini, N. Choi, and Y. Masuda, *Top. Stereochem.*, **5**, 69 (1970).
(9) S. Lifson and I. Oppenheim, *J. Chem. Phys.*, **33**, 109 (1960).
(10) N. Gō, M. Gō, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **59**, 1030 (1968).
(11) K. D. Gibson and H. A. Scheraga, *Physiol. Chem. Phys.*, **1**, 109 (1969).
(12) N. Gō and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969).
(13) E. B. Wilson, Jr., *ibid.*, **9**, 76 (1941).
(14) E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, "Molecular Vibrations," McGraw-Hill, New York, N. Y., 1955.
(15) N. Saito, K. Okano, S. Iwayanagi, and T. Hideshima, *Solid State Phys.*, **19**, 343 (1963).
(16) P. J. Flory, "Statistical Mechanics of Chain Molecules," Interscience, New York, N. Y., 1969.

entropy of solvation. The multidimensional surface, representing $F(Q)$ as a function of the conformation of the macromolecule, is a very complicated one, containing minima, maxima, saddle points, etc. Stable macroscopic conformations of the macromolecule in solution correspond to local minima of $F(Q)$, and the macromolecule undergoes stable oscillations around each of these minima. These oscillations contribute a conformational entropy; hence every macroscopic conformation corresponds to a collection of neighboring microscopic states [those located around each minimum of $F(Q)$], the statistical weight (or free energy) of each macroscopic conformation depending on both the magnitude of $F(Q)$ at the minimum and on the contribution from all the vibrational states in the neighborhood of the minimum.

It is assumed that the native conformation of a protein is that which corresponds to the local minimum of $F(Q)$ which has the *highest* statistical weight. As shown by several examples,[17-19] this need not be the global minimum of $F(Q)$; *i.e.*, a local minimum near the global one may have a higher statistical weight because of contributions from the librational motion.

The computation of stable conformations of polypeptides and proteins then involves three different problems. First, the geometry of the polypeptide chain must be known, and the functional form (and the values of the parameters) of $F(Q)$ must be determined. Second, $F(Q)$ must be minimized to locate the various local minima. Third, the statistical weights (which include the conformational entropy) of each minimum within a reasonable range of the global one, *i.e.*, within about 1 kcal/residue, must be calculated.

As far as the first problem is concerned, the conventions for the description of the polypeptide chain, geometrical data (bond lengths and bond angles), and procedures for transformation of coordinates were all discussed in a previous review.[1] Thus, it is possible to specify any conformation of a polypeptide chain in terms of Cartesian coordinates or internal coordinates in a fixed coordinate system. (It is important to note here that a new set of conventions has been adopted by an International Commission on Nomenclature,[20a] and replaces the one[20b] currently in use.)

The conformational energy, $F(Q)$, of an arbitrary conformation may then be expressed as a function of the coordinates of the macromolecule. Recent progress in the determination of appropriate energy functions is described in section II.B.

Considering next the second problem, several procedures[21-24] are available for the minimization of a function of many variables. These usually involve the evaluation of the first derivatives of $F(Q)$, with respect to all of the independent variables, and were discussed in an earlier review.[1] Since these procedures generally lead to stationary points[25] (which can be maxima, minima, or saddle points), the second derivatives must be examined[17] to determine which stationary points are indeed minima. The second derivatives also serve another important function, *viz.*, to provide the elements of a matrix **F** (defined below) required (together with a matrix **G**, also defined below) in the evaluation of the statistical weight of the macroscopic conformation corresponding to each minimum. Thus, there are two sources of entropy in the computations, the contributions from solvation [primarily in $V(Q)$] and from libration [in det **GF**].

While fairly efficient procedures are available for minimizing $F(Q)$, even for a molecule as large as lysozyme, a major problem remains, *viz.*, that arising from the existence of many local minima in the multidimensional $F(Q)$ surface. This problem has by no means been solved yet, but the following progress has been made. (1) Six different mathematical procedures have been developed to surmount energy barriers and have been applied to various peptide systems. All of them have given promising results in preliminary trials. These procedures are (a) minimization with an added extra variable,[25,26] (b) minimization of a constrained function,[25,26] (c) the deflation technique,[27] (d) method of slowest ascent,[28] (e) a statistical procedure,[29,30] and (f) a global search by elimination of high-energy regions (referred to as "spotting").[31] (2) A physicochemical procedure, based on the demonstrated importance of short-range interactions in proteins,[32-35] is being exploited to minimize the energies (and thereby determine the conformations) of short sequences[36] (seven to ten residues) in proteins. (3) A combination of procedures 1 and 2, using a perturbation technique to adjust a starting conformation by local changes in structure,[37,38] is being examined. Hopefully, some combination of several of the above procedures will solve the problem of multiple minima.

Turning now to the third problem, the calculation of the statistical weights is carried out in the following manner. For a molecule *in vacuo*, one should, in principle, minimize $U(Q)$, allowing for variation of *all* internal degrees of freedom. At the minimum point, Wilson's **G** and **F** matrix must be calculated to solve for the $(3n - 6)$ normal mode vibrational circular frequencies ω_i $(i = 1, 2, \ldots, 3n - 6)$, where n is the number of atoms in the molecule, the elements of the matrix **G** are the coefficients of the kinetic energy term in the canonical expression for the Hamiltonian (and are related to the effective moments of inertia for the internal motion of the molecule), and the elements of the matrix **F** are the second

(17) K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **63**, 242 (1969).
(18) M. Gō, N. Gō, and H. A. Scheraga, *J. Chem. Phys.*, **52**, 2060 (1970).
(19) N. Gō, P. N. Lewis, and H. A. Scheraga, *Macromolecules*, **3**, 628 (1970).
(20) (a) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970); (b) J. T. Edsall, P. J. Flory, J. C. Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, *Biopolymers*, **4**, 121, 1149 (1966); *J. Biol. Chem.*, **241**, 1004, 4167 (1966); *J. Mol. Biol.*, **15**, 339 (1966); **20**, 589 (1966).
(21) W. C. Davidon, AEC Research and Development Report, ANL-5990, 1959.
(22) R. Fletcher and M. J. D. Powell, *Comput. J.*, **6**, 163 (1963).
(23) R. Fletcher and C. M. Reeves, *ibid.*, **7**, 149 (1964).
(24) J. D. Pearson, *ibid.*, **12**, 171 (1969).
(25) K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **63**, 9 (1969).
(26) K. D. Gibson and H. A. Scheraga, *Comput. Biomed. Res.*, **3**, 375 (1970).
(27) G. M. Crippen and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **64**, 42 (1969).
(28) G. M. Crippen and H. A. Scheraga, *Arch. Biochem. Biophys.*, submitted for publication.
(29) See addendum to ref 30.
(30) F. A. Momany, G. Vanderkooi, R. W. Tuttle, and H. A. Scheraga, *Biochemistry*, **8**, 744 (1969).
(31) G. M. Crippen and H. A. Scheraga, *Arch. Biochem. Biophys.*, submitted for publication.
(32) D. Kotelchuck and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **61**, 1163 (1968).
(33) D. Kotelchuck and H. A. Scheraga, *ibid.*, **62**, 14 (1969).
(34) D. Kotelchuck, M. Dygert, and H. A. Scheraga, *ibid.*, **63**, 615 (1969).
(35) P. N. Lewis, N. Gō, M. Gō, D. Kotelchuck, and H. A. Scheraga, *ibid.*, **65**, 810 (1970).
(36) D. Kotelchuck and H. A. Scheraga, unpublished work.
(37) N. Gō and H. A. Scheraga, *Macromolecules*, **3**, 178 (1970).
(38) N. Gō and H. A. Scheraga, *ibid.*, **3**, 188 (1970).

derivatives of the conformational energy function $F(Q)$ at the minimum point. Then, the (quantum-mechanical) statistical weight is given[11,12] by

$$Z = \prod_{i=1}^{3n-6} [2 \sinh \frac{1}{2}(\beta\hbar\omega_i)]^{-1} \exp(-\beta U_{\min}) \quad (1)$$

where $\beta = 1/kT$ and $U_{\min}$ is the value of $U(Q)$ at the minimum point. However, in practice, eq 1 cannot be used for a macromolecule because the computation of the ω_i's for such a large molecule is too formidable a problem at the present time. Moreover, eq 1 cannot be used for a molecule in solution because $F(Q)$ contains an entropy term (from solvation) which cannot be treated by the quantum-mechanical procedures involved in the normal mode analysis to obtain the ω_i's. Therefore, one must use an approximate treatment of the motion of of a macromolecule (both *in vacuo* and in a solvent) to compute the statistical weight; also, it produces a good approximation for eq 1, when applied to a molecule *in vacuo*.

In the approximate treatment, cognizance is taken of the fact that the frequencies for bond stretching and bond angle bending ("hard" variables) are generally much higher than those for variations of dihedral angles for rotation around single bonds ("soft" variables), and hence that the two classes of internal vibrations can be decoupled.[11,12] When solvent is present, the motions of the solvent molecules couple with the internal motions of the polymer. However, the low and high frequency modes of the whole system (including the solvent molecules) can still be decoupled. The dihedral angles are regarded as the only independent variables; their variation is governed by $F(Q) = U(Q) + V(Q)$ and can be treated by classical mechanics. The bond lengths and bond angles (hard variables) are then either functions of the instantaneous values of the soft variables [approximation A or B, depending on whether the zero-point energies of the motions of the hard variables are included (A) or not (B)], or considered to be constant (approximation C). The instantaneous values of the bond lengths and bond angles, determined by the instantaneous values of the soft variables in approximations A and B, correspond physically to the average values of the high-frequency (quantum-mechanical) motions of the hard variables. The zero-point energies $\epsilon(Q)$ of these motions are not affected much by solvent and, as indicated above, constitute a part of the conformational energy (approximation A). However, since the dependence of $\epsilon(Q)$ on the dihedral angles is not large, they can be neglected as an approximation (approximation B). For a molecule *in vacuo* or in solution, whose motion is that described above, the statistical weight is given by[10,12]

$$Z = \left(\frac{1}{2\pi\hbar}\right)^m \int \left[\left(\frac{2\pi}{\beta}\right)^m \frac{1}{\det \mathbf{G}}\right]^{1/2} \exp(-\beta F)\, dQ \quad (2)$$

where m is the number of soft variables in the molecule, the matrix $\mathbf{G}$ is the one for the molecule in which only the dihedral angles are independent variables,[39] and $F = F_0(Q) + \epsilon(Q)$ for approximation A, $F = F_0(Q)$ for approximation B, and $F = F_1(Q)$ for approximation C, with $F_0(Q)$ and $F_1(Q)$ being the conformational energy function in which the values of the hard variables are functions of the instantaneous values of the soft variables, or constants, respectively. *If* $F(Q)$ can be approximated by a multidimensional parabola in the neighborhood of each minimum (*the assumption of small conformational fluctuations*), the free energy of the macroscopic conformation corresponding to a local minimum of $F(Q)$ is given, using eq 2, by

$$F = F_0(Q_0) + (1/2)RT \ln [\det \mathbf{GF}]_{Q=Q_0} \quad (3)$$

for approximation B, and by

$$F = F_1(Q_1) + (1/2)RT \ln [\det \mathbf{GF}]_{Q=Q_1} \quad (4)$$

for approximation C. In approximation A, the zero-point energies $\epsilon(Q_0)$ of the motion of the hard variables are added to the right-hand side of eq 3. In eq 3 and 4, Q_0 and Q_1 are minimum points of $F_0(Q)$ and $F_1(Q)$, respectively. Since terms, which are independent of conformation, have been dropped from eq 3 and 4, these equations can be used only to consider *differences* in free energies for different conformations. The last term on the right-hand side of eq 3 and 4 is the contribution to the statistical weight from the librational degrees of freedom of the molecule. Approximation A, the most accurate of the three approximations, leads to the following result, which is indeed a good approximation to eq 1, when applied to a molecule *in vacuo*.[12]

$$Z = \prod_{i=1}^{m} (\beta\hbar\omega_i)^{-1} \exp\left\{-\beta\left[U_{\min} + (1/2)\sum_{i=m+1}^{3n-6} \hbar\omega_i\right]\right\} \quad (5)$$

Even though it is incorrect to apply eq 1 to molecules in solution, it is legitimate to apply approximation A (with the zero-point energies of the hard variables computed by a normal mode treatment of the hard variables) in principle to molecules in solution because the motions of the hard variables are not affected much by the presence of the solvent, *i.e.*, the term $V(Q)$ in $F(Q)$ does not affect the quantum-mechanical motion of the hard variables Q' much. However, the practical calculation of the zero-point energies is a formidable one for large molecules, as already stated. Also, since the zero-point energies do not depend very much on conformation, they can be neglected (approximation B). Therefore, approximation A is more of theoretical rather than of practical interest.

For actual calculations, the following recipe, consisting of three steps, has been proposed.[12] (1) Start with approximation C, *i.e.*, assume that the bond lengths and bond angles are constant, and construct a function $F_1(Q)$ by taking into account all intramolecular interactions (including solvent effects for a molecule in solution). Using a suitable minimization procedure, locate the minimum point(s) Q_1 in the function $F_1(Q)$. (2) Then, adopt approximation B by allowing all bond lengths and bond angles to be variable, and construct a conformational energy function $F(Q,Q')$. In this process, a set of force constants for variation of the hard variables Q' is required.[40] The minimization of the function $F(Q,Q')$, *starting* from the point (Q_1,Q_1') (Q_1' being the fixed value of the hard variables in step 1), will lead to the point (Q_0, Q_0'). (3) At the point (Q_0, Q_0'), evaluate $F_0(Q_0)$ and $(\det \mathbf{GF})_{Q=Q_0}$. This gives us the value of the free energy of a macroscopic conformation in approximation B. Unless there is serious steric hindrance in the conformation obtained in step 1,

(39) The matrix G of ref 10 is inverse to the matrix G of ref 12, the latter being the same as in eq 2 above. There are several typographical errors in ref 10 (there should be a factor of $1/2$ in eq 8 and 10, a factor of 2 in eq 9, and π should be replaced by 2π in eq 11).

(40) See, for example, the force constants adopted by S. Lifson and A. Warshel, *J. Chem. Phys.*, **49**, 5116 (1968), and by A. Warshel, M. Levitt, and S. Lifson, *J. Mol. Spectrosc.*, **33**, 84 (1970).

approximation C would be expected to be a fairly good one (*i.e.*, the deviation of the bond lengths and bond angles from standard values would be expected to be small, and therefore need not be considered). In this case, step 2 can be skipped, and $F_1(Q_1)$ and $[\det \mathbf{GF}]_{Q=Q_1}$ can be used to obtain the free energy of the macroscopic conformation in approximation C. When all three steps are followed, it is probably sufficient to use $(1/2)RT \ln [\det \mathbf{GF}]_{Q=Q_1}$, the contribution to the statistical weight from the librational degrees of freedom of the molecule treated as a rigid one, together with $F_0(Q_0)$ in step 3, rather than $(1/2)RT \ln [\det \mathbf{GF}]_{Q=Q_0}$, because the conformational dependence of the librational contribution is much smaller than that of the conformational energy; in other words, the main difference between the quantities F of eq 3 and 4 is the one between $F_0(Q_0)$ and $F_1(Q_1)$. The procedures for computing det **G** and det **F** for rigid molecules are discussed elsewhere[11,12,18] and sketched briefly below. They have been applied[19,38] to cyclo(Gly_3Pro_2) (see section II.H for the application of approximation C to this cyclic pentapeptide); the extension of this procedure to flexible molecules is now in progress.[41] The procedure for computing det **G** has also been applied to a polyethylene-type polymer,[12] and those for computing det **F** have been applied to deca-L-alanine[17] (see section II.D), and to polyglycine and poly-L-alanine[18] (see section II.F). The calculations for cyclo(Gly_3Pro_2) are the most complete ones so far carried out, in the sense that, heretofore, the conformational calculations have been done either by completely neglecting the conformational entropy, or by taking it into account only partially (by computing only det **F** *or* det **G** but not det **GF**); the calculation can be made even more complete (a) by allowing the molecule to be flexible,[41] and (b) by introducing the effects of solvation for isolated molecules in solution (see section II.B), or (b′) by introducing the effects of crystallization for molecules in the crystalline form (see section II.E).

In order to evaluate $(1/2)RT \ln [\det \mathbf{GF}]_{Q=Q_1}$, det **G** and det **F** are obtained as follows. For obtaining det **G**, one first obtains Wilson's **G** matrix for the molecule treated as a flexible one (variable bond lengths and bond angles), using eq 20–22 of ref 12; if there are n atoms in the molecule, the **G** matrix is $(3n - 6) \times (3n - 6)$. The elements of the **G** matrix are obtained, with the aid of a computer, by calculating Wilson's **s** vectors for the given Cartesian coordinates and connectivity (chemical bonds) of the atoms. Then the condition of rigidity is introduced (corresponding to the freezing of bond lengths and bond angles) by using eq 29 of ref 12 to obtain an $m \times m$ **G** matrix whose elements are functions of the hard variables; the computation of these elements requires the inversion (performed numerically) of a $(3n - 6 - m) \times (3n - 6 - m)$ submatrix (corresponding to the $3n - 6 - m$ hard variables) of the original $(3n - 6) \times (3n - 6)$ matrix, and the multiplication by $m \times (3n - 6 - m)$ and $(3n - 6 - m) \times m$ submatrices. To obtain det **F**, $F(Q)$ is expressed in terms of the m soft variables (dihedral angles). The conformational energy $F(Q)$ is given as a sum of various energy terms. A second derivative of $F(Q)$, *i.e.*, an element of the matrix **F**, is given as a sum of second derivatives of these energy terms. Since some of the terms in $F(Q)$ are explicit functions of interatomic distance rather than of dihedral angle, it is necessary to express the interatomic distance as a function of the dihedral angles. A procedure for computing the first and second derivatives of $F(Q)$ with respect to the dihedral angles is given in the Appendix of ref 18. See also ref 11 and 19 for calculations of the magnitudes of the average fluctuation in the conformation of a protein.

While the above procedure (making use of eq 3 or 4) is applicable to a protein, where the conformational fluctuations can be expected to be small, it is not applicable to random coils for which large conformational fluctuations may be expected. In order to calculate the statistical weight of a polymer in the random coil state, the integration in eq 2 has to be carried out over the *entire* Q space, rather than simply over the neighborhood of local minima. To perform this integration practically, the following assumptions and simplification of the integrand of eq 2 have to be made. First, only approximation C is used. Second, long-range interactions are assumed to contribute negligibly to eq 2. This is equivalent to the assumption that the polymer is at the θ-point.[16,42] Thus, $F(Q)$ in eq 2 is separated into its short-range and long-range terms, and only the short-range ones (*i.e.*, those energy terms depending on either only one or two consecutive dihedral angles) are retained. However, it has not yet been found possible to use the same procedure for computing ln det **G**. For the present, we have made the interim assumption that ln det **G** in eq 2 can be treated as a constant for conformations with large fluctuations. After these assumptions and simplifications are made, the integration of eq 2 can be performed numerically for any polymer by replacing the integration over the range of $[0, 2\pi]$ for any dihedral angle by a summation over N points in this range, where $N = 36$ is usually large enough.[12] However, if the molecule is a homopolymer, a matrix method[10,18] (involving the calculation of eigenvalues) can be used to perform the integration of eq 2; this technique has been applied to the calculation of the parameters s and σ for the helix–coil transition in polyglycine and poly-L-alanine[10,18] (see section II.F).

B. CONFORMATIONAL ENERGIES

In using empirical functions[1] for $U(Q)$ and $V(Q)$, the contributions to $U(Q)$ include internal rotation potentials, nonbonded interactions, dipole and charge interactions, hydrogen bond potentials, potential for torsion about the peptide bond, and potentials for bond angle bending and bond stretching. The influence of solvent is included in the dielectric constant and in $V(Q)$ (see below); the term $V(Q)$ takes account of the liberation of bound water as parts of the chain approach each other when the dihedral angles change,[43] and leads to a tendency for a polypeptide chain to fold up so that, on the average, most of its polar side-chain groups are on the outside of the molecule in contact with water and most of the nonpolar ones are in the interior. The contribution from hydrophobic bonding appears in both $U(Q)$ and $V(Q)$. For structures of rigid geometry containing rings (*e.g.*, cyclic peptides, or loops formed by disulfide bonds), two procedures have been used to assure the formation of the ring. In the first, an artificial loop-closing potential (which vanishes when the loop is properly closed) is introduced to bring the ends of the loop together.[1,44,45] In the second, the ring is closed by determining

(41) N. Gō, G. C. C. Niu, and H. A. Scheraga, work in progress.

(42) P. J. Flory, "Principles of Polymer Chemistry," Cornell University Press, Ithaca, N. Y., 1953, Chapter 14.

(43) K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **58**, 420 (1967).

(44) K. D. Gibson and H. A. Scheraga, *ibid.*, **58**, 1317 (1967).

(45) R. A. Scott, G. Vanderkooi, R. Tuttle, P. Shames, and H. A. Scheraga, *ibid.*, **58**, 2204 (1967).

the range of values of the dihedral angles which allow ring closure; in such a ring of n dihedral angles, six dihedral angles become functions of the remaining $n - 6$ which can be varied independently (within the limited range which keeps the ring closed) to change the conformation of the ring.[37] If nonrigid geometry is allowed, the potentials for bond stretching and bond angle bending accomplish loop closure, in addition to their other roles (see, *e.g.*, ref 41 and 46). Since the empirical functions have been discussed in earlier reviews,[1-6] we confine our consideration here to subsequent developments.

One problem to which more attention has recently been paid is the role of water in influencing conformation. It appears that we can distinguish three types of solvent effect,[18,47] *viz.*, (1) specific interactions of one or more solvent molecules with the backbone NH and CO (and side-chain polar) groups of the residues in which these groups are not themselves hydrogen bonded, (2) effect of the solvent molecules in which they are regarded as a bulk medium with a dielectric constant other than unity, and (3) additional interaction energies between atoms, or groups of atoms, in a polypeptide, arising from the presence of nearby solvent molecules (*e.g.*, hydrophobic bonding). Effects of type 1 have been considered for polyglycine and poly-L-alanine in water[47] (see section II.F), wherein the free energy of binding, b, was taken as

$$b = -2RT \ln (1 + Ka) \tag{6}$$

where a is the activity of water and K is the equilibrium constant for binding water molecules (assumed to be the same for NH and CO groups). Expressing Ka as

$$Ka = \exp[-(\Delta H_B - T\Delta S_B)/RT] \tag{7}$$

the enthalpy and entropy changes, respectively, due to water binding were found (by adjusting theoretical melting curves to experimental data for poly-L-alanine in water; see section II.F) to be $\Delta H_B = -2.74$ kcal/mol and $\Delta S_B = -5.78$ eu. Effects of type 2 have been taken into account by varying the dielectric constant, D; in particular, D has been assigned[47] either the value 1.0 or 4.0 for interactions between charged groups (in water) separated by up to about 6 Å, and infinity (to simulate the high dielectric constant of water) at larger distances (see section II.F). While effects of type 3 are small for polyglycine and poly-L-alanine,[47] they are not so for longer nonpolar groups. [See an earlier review[1] for a discussion of the assignment of hydration free energies (in the term $V(Q)$ which is really a *free* energy) to nonpolar, as well as polar, groups.]

Efforts have also been devoted recently toward improving and strengthening the physical basis of the empirical methods which are being applied to conformational energy calculations on polypeptides, and toward refining the parameters of the empirical energy functions. Molecular orbital calculations[48-50] have been carried out for several series of homologous compounds to provide a semiquantitative understanding of the physical basis of the empirical methods, and the structures of crystals of small molecules have been computed[6,51] to refine the empirical energy parameters.

The extended Hückel theory (EHT) of Hoffmann[52] and the complete neglect of differential overlap theory (CNDO/2) of Pople and Segal[53] have been used for calculations on amides,[48] esters and related compounds,[49] and hydrogen-bonded complexes of amides.[50] The calculations yielded the net partial charges on each atom and thus the molecular dipole moments, the energies for internal rotation about bonds, the electronic singlet-state energy levels, and some trends in the electronic spectra. These properties were obtained for different conformations of the compounds studied, and thus give some insight into the relations between electronic structure and conformational changes. Initial results[54] indicate that there is a similarity in the appearance of the contours of the ϕ-ψ maps[20b] of glycyl and alanyl residues computed by the EHT method and by the use of empirical energy functions. While these methods, like the perturbative configuration interaction using localized orbitals theory (PCILO) recently applied by Maigret, *et al.*,[55] are only approximations to more exact procedures, they give valid information on *trends* in various *series* of related compounds. Furthermore, while the molecular orbital calculations provide only approximate information about charge distributions and other parameters, these can then be refined in the calculations on crystals of small molecules.

The calculations on several compounds[48] related to formamide and acetamide indicated that the partial charges appear to be the same no matter whether the amide group is in the cis- or trans-planar conformation, but vary as the amide group departs from planarity. The reasonableness of the values of the partial charges is attested to by the fact that the calculated (CNDO/2) and experimental dipole moments are in good agreement. The computed (EHT) barrier to rotation about the peptide bond (24 kcal/mol) and the energy difference between the cis and trans forms (3 kcal/mol) are in qualitative agreement with experimental values. The calculated preference for the trans form arises primarily from the destabilization of the cis conformation of *N*-methylacetamide because of the steric repulsion of the cis methyl groups. The barriers to internal rotation of the single methyl groups in *N*-methylformamide and in acetamide remain small (less than 1 kcal/mol), with a *small* variation as the amide group departs significantly from planarity; also, the symmetries of these internal rotational energy functions correspond to those in the empirical intrinsic potential functions of Scott and Scheraga[56] rather than to those of Maigret, *et al.*[55] It was also found from the calculations that the (n,π^*) and (π,π^*) amide transitions are shifted toward the red as the group departs significantly from planarity; the larger calculated shift for the (π,π^*) transition, suggesting that the two bands might cross at some angle of twist, remains to be tested experimentally.

When small amide molecules, like formamide and *N*-methylacetamide, form binary and higher complexes (involving hydrogen bonding and other intermolecular interactions), there is a redistribution of charges (and even charge transfer

(46) M. Bixon and S. Lifson, *Tetrahedron*, 23, 769 (1967).
(47) M. Gō, N. Gō, and H. A. Scheraga, *J. Chem. Phys.*, submitted.
(48) J. F. Yan, F. A. Momany, R. Hoffmann, and H. A. Scheraga, *J. Phys. Chem.*, 74, 420 (1970).
(49) J. F. Yan, F. A. Momany, and H. A. Scheraga, *J. Amer. Chem. Soc.*, 92, 1109 (1970).
(50) F. A. Momany, R. F. McGuire, J. F. Yan, and H. A. Scheraga, *J. Phys. Chem.*, 74, 2424 (1970).
(51) F. A. Momany, G. Vanderkooi, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, 61, 429 (1968).
(52) R. Hoffmann, *J. Chem. Phys.*, 39, 1397 (1963); 40, 2474, 2480, 2745 (1964).
(53) J. A. Pople and G. A. Segal, *ibid.*, 44, 3289 (1966).
(54) R. Hoffmann and A. Imamura, *Biopolymers*, 7, 207 (1969).
(55) B. Maigret, B. Pullman, and M. Dreyfus, *J. Theor. Biol.*, 26, 321 (1970).
(56) R. A. Scott and H. A. Scheraga, *J. Chem. Phys.*, 45, 2091 (1966).

from one monomer to the other) as the monomers approach each other (in various relative orientations).[50] In a linear hydrogen-bonded dimer of formamide, the electronic interactions distort the oxygen line-pair orbitals in such a way that they appear to follow the H(N) atom; hence a stable conformation (the angle between the C=O and linear O···H—N bonds) is adopted which reflects the complete molecular electronic interaction, rather than being determined by the direction implied by pure sp^2 hybridization. Similarly, in the linear hydrogen-bonded dimer of *N*-methylacetamide, there is a significant departure from a direction implied by pure sp^2 hybridization, the specific orientation of the two monomers in the complex again being dominated by the total molecular environment rather than by strong directional orbital properties of the nonbonded monomers. The formation of the hydrogen bond influences not only the charge distribution but also, as a result, the dipole moment and the barrier to rotation about the peptide bond.

In the formamide linear hydrogen-bonded trimer,[50] the energy per hydrogen bond is lower than that found for the equivalent linear dimer. This indicates that linear hydrogen-bond formation is a cooperative process; *i.e.*, the formation of a linear trimer is more favorable than the formation of two linear dimers. It is of interest that qualitative arguments had been advanced by Frank and Wen,[57] and quantitative ones by Hoyland and Kier,[58] for the existence of such cooperativity in another hydrogen-bonded system, *viz.*, liquid H_2O. These results imply that the formation of extended hydrogen-bonded structures in a polypeptide should be favored. Also, the charge distribution in the amide groups of a large hydrogen-bonded polypeptide structure (*e.g.*, the α helix) would be expected to differ from that of, say, a non-hydrogen-bonded dipeptide. Similar molecular orbital calculations on hydrogen-bonded amide complexes have also been carried out by Murthy, *et al.*[59]

The calculations also indicated that other complexes (*e.g.*, formamide planar cyclic and parallel-plane dimers) should also be stable.[50] These various complexes appear in the crystal, and, together with the requirements for packing, lead to hydrogen-bond distances which would be expected to (and do) differ from those computed for *isolated* complexes.

In all of the calculations on hydrogen-bonded complexes,[50] the hydrogen-bond strength seemed to correlate better with the magnitude of the resultant charge-separation of the molecules making up the complex rather than with the charge-transfer which does occur. These calculations (and similar ones on related systems, involving, for example, the effect of hydrogen bonding to water molecules) are being used to formulate an empirical potential function for the formation of a hydrogen bond.[60] Hopefully, this function (and its refinement from calculations of crystal structures) will be an improvement over the Lippincott–Schroeder functions[61-63] now in use.[1]

Since the ester group occurs in many polyamino acids of interest (*e.g.*, benzyl glutamates and aspartates), similar molecular orbital calculations were also carried out for simple esters and related compounds.[49] Aside from information gained about charge distributions, dipole moments, and barriers to internal rotation about bonds in side chains containing ester groups, new insight was gained about intrinsic torsional potentials. In contrast to earlier procedures,[1] in which internal rotational energies were separated into intrinsic, nonbonded, and electrostatic contributions, it now appears that the intrinsic torsional potential already *includes* the electrostatic contribution for polar groups which are very close together (specifically, when they are involved in 1,4-type interactions, *i.e.*, when the intereacting atoms are separated by three bonds); of course, the 1,4-type nonbonded empirical energy still must be added in separately.

Similar molecular orbital calculations are currently in progress[60] for all naturally occurring amino acids, and for dipeptides formed from some of these, in order to provide a better phyical understanding of the empirical energy functions (partial charges, barriers to internal rotation, etc.) currently in use[1] in computations of stable conformations of such molecules. All of these results will then be refined by calculations of crystal structures of small molecules. Maigret, *et al.*,[55] are carrying out similar calculations with the PCILO method.

The components of the function $U(Q)$ should also play a role in determining the most stable conformations of dry crystals of small molecules. At the present time, it tentatively appears that separate sets of functions may be required for the short distances involved in intramolecular interactions and for the longer ones involved in intermolecular interactions. However, this is still an unsettled question. In an earlier review,[6] the computed lattice constants of 21 crystals were reported. In these calculations separate sets of functions were used for intra- and intermolecular interactions. An alternative procedure, involving the application of conditions of static equilibrium, was applied to crystalline benzene.[6,51] In the latter calculations, no partial charges were assigned to the carbon and hydrogen atoms, and an anomaly was observed in the absence of an attractive term for C···C nonbonded interactions. On the basis of the molecular orbital calculations mentioned above, it appears that even carbon and hydrogen atoms in *aliphatic* hydrocarbons may have some delocalization of their electronic distribution. Therefore, the calculations of crystal structures are being repeated, also with the inclusion of thermodynamic and spectroscopic data,[60,64] starting with information acquired from the molecular orbital calculations. The aforementioned anomaly can be resolved by slight adjustment of the C–H bond length and by the inclusion of partial charges on all the atoms.[60] It remains to be seen whether a universal empirical energy function, applicable to intra- and intermolecular interactions in all crystals, will emerge. In an alternative approach, the energy surface was represented as a polynomial, and the parameters were adjusted to match experimental data for the formic acid dimer;[65] the resulting energy function was then applied to the acetic acid dimer. The results of conformational energy calculations on polypeptides reported here, and in earlier reviews, were obtained with earlier "best" estimates of the parameters of the *empirical* conformational energy function on the basis of

(57) H. S. Frank and W. Y. Wen, *Discuss. Faraday Soc.*, 24, 133 (1957).
(58) J. R. Hoyland and L. B. Kier, *Theor. Chim. Acta*, 15, 1 (1969).
(59) A. S. N. Murthy, K. G. Rao, and C. N. R. Rao, *J. Amer. Chem. Soc.*, 92, 3544 (1970).
(60) R. F. McGuire, F. A. Momany, and H. A. Scheraga, unpublished work.
(61) E. R. Lippincott and R. Schroeder, *J. Chem. Phys.*, 23, 1099 (1955).
(62) R. Schroeder and E. R. Lippincott, *J. Phys. Chem.*, 61, 921 (1957).
(63) W. G. Moulton and R. A. Kromhout, *J. Chem. Phys.*, 25, 34 (1956).
(64) A. Warshel and S. Lifson, *ibid.*, 53, 582 (1970).
(65) W. P. Minicozzi and D. F. Bradley, *J. Comput. Phys.*, 4, 118 (1969).

various physicochemical data on small molecules. Also, except where noted otherwise in the results described here, only the term $U(Q)$ of $F(Q)$ has been computed.

C. DIPEPTIDES

The earlier calculations of energy contours on a ϕ–ψ map[20b] for "dipeptides" of the type Gly-Gly and Gly-Ala, and the application of these results to the calculation of the end-to-end distances of random coils,[66–68] has been discussed previously.[1] Further studies of single residues, also using empirical energy functions, have been carried out by Ponnuswamy and Sasisekharan.[69] Recently, in connection with the application of the deflation technique to the problem of multiple minima in the energy surface, a more complete contour map (shown in Figure 1) of the "dipeptide" glycyl-L-alanine was computed.[27] Aside from the fact that many of the numerous stationary points were found by the deflation technique, a deep minimum was computed for the region near $(\phi, \psi) \backsim (240°, 120°)$, which had previously[70,71] been thought to be disallowed for a residue with a β-carbon atom in its side chain. It is of interest in this regard to note that there is *one* nonglycine residue in lysozyme, *viz.*, Phe 38, whose dihedral angles[72] (292°, 153°) lie near the (240°, 120°) minimum. Thus, this minimum (found by the deflation technique[27]), while not the global one, is nevertheless in an accessible energy region, and may account for the fact that Phe 38 in lysozyme has this conformation.

Also recently, in connection with the work described in section II.H, conformational energy calculations were carried out for the two dipeptides Pro-Pro and Pro-Gly, both of whose conformations depend on only one angle, ψ, the dihedral angle for rotation about the central C^α–C′ bond. In these calculations, cognizance was taken of the puckering of the proline ring;[73] thus, two conformations [designated Pro(I) and Pro(II)] are available to this ring. The conformational energies of these two dipeptides are shown in Figures 2 and 3 for all possible combinations of the Pro(I) and Pro(II) geometry. It can be seen that the difference between Pro(I) and Pro(II), *i.e.*, the variation in the position of the C^γ atom, significantly affects the low-energy conformation of Pro-Pro in the range between $\psi \backsim 95°$ and $\backsim 175°$, but not that of Pro-Gly. The close contacts which are responsible for these energy diagrams are described elsewhere.[38] The minimum in the α-helical region (which lies in this range of ψ) is allowed for Pro-Pro only if the ring is puckered. This accounts for the fact that Schimmel and Flory[74] (who assumed that the pyrrolidine ring is planar) concluded that this range of ψ is disallowed, while Leach, *et al.*[70] (who assumed that the pyrrolidine ring is puckered), found that the range of ψ between 130 and 150° is allowed. Hopfinger and Walton[75] have also considered the four types of proline dipeptides, resulting from the puckering of the proline ring, as well as longer oligomers of proline (including poly-L-proline; see also Schimmel and Flory[74]).

As indicated in section II.B, molecular orbital calculations[55,60] are also being carried out for dipeptides. The ϕ–ψ maps[20a] for these molecules have been computed and the minimum-energy regions compared with experimental results. These calculations are also being used[60] as a basis for further improvement of empirical energy functions.

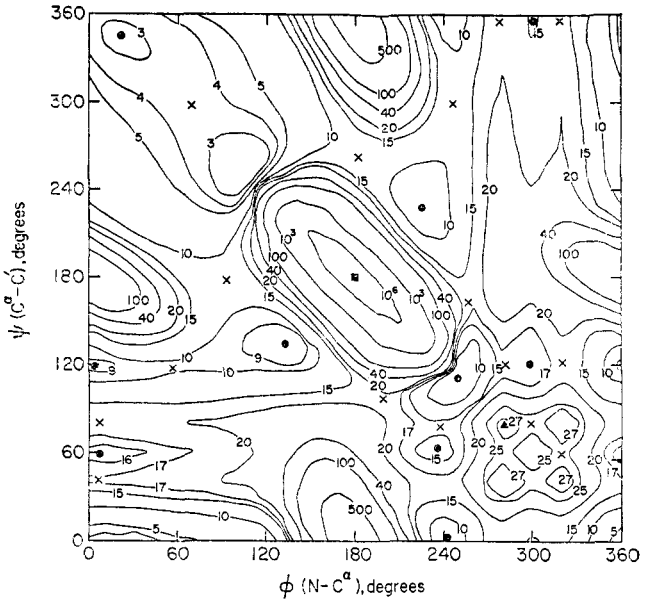

Figure 1. Energy contour map of the glycyl-L-alanine "dipeptide."[27] The dihedral angles[20b] ω, ω', and χ were held constant at $\omega = \omega' = 0°$ and $\chi = -60.3°$. The energy units are kcal/mol, with the zero of energy chosen so that the global minimum [at $(\phi, \psi) = (105°, 252°)$] has an energy of 1.4 kcal/mol. The points marked ●, ×, ■, and ▲ are minima, saddle points, singularities, and maxima, respectively, and were located by the deflation technique.[27]

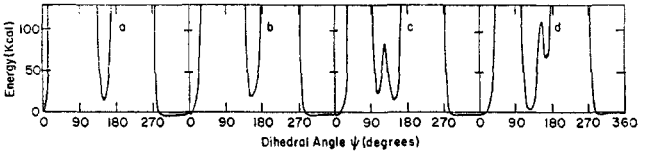

Figure 2. Conformational energy[38] in "dipeptide" (a) Pro(I)-Pro(I), (b) Pro(I)-Pro(II), (c) Pro(II)-Pro(I), and (d) Pro(II)-Pro(II).

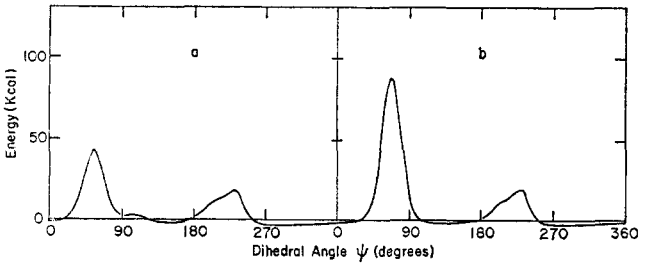

Figure 3. Conformational energy[38] in "dipeptide" (a) Pro(I)-Gly and (b) Pro(II)-Gly.

(66) D. A. Brant and P. J. Flory, *J. Amer. Chem. Soc.*, **87**, 2791 (1965).
(67) D. A. Brant, W. G. Miller, and P. J. Flory, *J. Mol. Biol.*, **23**, 47 (1967).
(68) W. Miller, D. A. Brant, and P. J. Flory, *ibid.*, **23**, 67 (1967).
(69) P. K. Ponnuswamy and V. Sasisekharan, *Int. J. Protein Res.*, **2**, 37, 47 (1970).
(70) S. J. Leach, G. Nemethy, and H. A. Scheraga, *Biopolymers*, **4**, 369 (1966).
(71) G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.*, 7, 95 (1963).
(72) C. C. F. Blake, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Proc. Roy. Soc., Ser. B*, **167**, 365 (1967).
(73) Y. C. Leung and R. E. Marsh, *Acta Cryst.*, **11**, 17 (1958).
(74) P. R. Schimmel and P. J. Flory, *Proc. Nat. Acad. Sci. U. S.*, **58**, 52 (1967).
(75) A. J. Hopfinger and A. G. Walton, *J. Macromol. Sci., Phys.*, **3**, 171 (1969).

D. POLYAMINO ACID α HELICES

Some of the earlier results[1,2,6] on helical structures of polyamino acids include the following: (1) the prediction[76] of the existence of the α_{II} helix, which was subsequently confirmed[77] by its discovery at the C termini of several α_I helices in lysozyme and myoglobin; (2) the prediction[78] that the right-handed α-helical form of poly-L-valine is sterically capable of forming, and the subsequent experimental verification[79] of this fact; (3) the calculation, in agreement with experiment, of the correct screw sense of the α helix in approximately 25 different homopolyamino acids.[78,80] In these calculations, the dipole–dipole interaction between side-chain ester groups and backbone amide groups was identified as an important feature determining helix sense in a series of poly-L-aspartates and poly-L-glutamates, thereby accounting for the differences in helix sense among some of these polymers. Also, a right-handed helix sense was deduced for poly-L-tyrosine, not only on the basis of conformational energy calculations,[78] but also from a calculation of the optical rotatory properties[81] of this polyamino acid. Additional results are (4) the calculation of the preferred conformations of the side chains of α-helical structures[78,80] [experiments are now being carried out (see section III.F) to test these predictions], and (5) the demonstration of the important influence of the side chain on the computed[82] rotational strength of the n,π^* transition, which is observed in circular dichroism and optical rotatory dispersion experiments on polyamino acids.

Recently, calculations were carried out[49] to determine the effect of a chlorine atom (in the ortho and meta positions of the benzene ring) on the helix sense of poly-β-benzyl-L-aspartate. Similar calculations had already been carried out[80] for the *p*-chlorobenzyl esters of poly-L-aspartic acid and poly-L-glutamic acid; the calculations for the former were in agreement with experiment[83–86] (*viz.*, that this polymer forms a right-handed α helix whereas the parent poly-β-benzyl-L-aspartate forms a left-handed one[87–89]), and those for the latter were subsequently verified by experiment[90] (*viz.*, that this polymer and its parent[89] poly-γ-benzyl-L-glutamate form a right-handed α helix). The recent computations[49] indicated that the ortho and meta chlorobenzyl-L-aspartates should form left-handed α helices. These predictions were subsequently verified by experiment[91] (see section III.E). Here again, dipole–dipole interactions were found to play an important role in determining the helix sense. For example, Figure 4 shows that, for the meta derivative, the interaction between the C–Cl and the nearest backbone amide dipole favors the left-handed conformation.[49] In the left-handed form, the C–Cl dipole is nearly antiparallel to the nearest backbone amide dipole whereas, in the right-handed form, these dipoles are parallel.

All of the calculations discussed in this section were based on the assumption of regularity, *i.e.*, that the pair of values of ϕ and ψ was the same in every residue. Recently this restriction has been removed and nonregular helices have been treated[18] (see section II.F); also, conformations which depart significantly from helical structures have been considered, as illustrated by the following computations on deca-L-alanine.[25]

Actually, the computations[25] on deca-L-alanine were carried out for a different purpose, *viz.*, to demonstrate how one can pass successively through lower energy minima in the multidimensional energy surface. In these computations, the function $F(Q)$ was used; *i.e.*, the influence of the hydration term $V(Q)$ was included. Three arbitrary starting conformations (the α helix, the β conformation, and the one having $\phi = 60°$, $\psi = 180°$) were selected and, allowing each dihedral angle to vary independently, two of the procedures[25,26] mentioned in section II.A (designed to locate successively lower energy minima) were applied. Table I indicates, first of all, that it is possible to move from minimum to minimum, of decreasing energy, by the procedures applied. Secondly, while a conventional minimization procedure would confine the conformation to that near a nonregular α helix,[18] which was satisfactory for the purpose of the work described in ref 18 (see section II.F), the ability to move out of local minima enabled the nonregular α helix (minimum no. 1 of Table I, with energy of 17.71 kcal/mol) to depart from the helical conformation to a less regular one (minimum no. 14 of Table I, with energy of −20.36 kcal/mol). Using the second derivatives of $F(Q)$, it was possible to demonstrate that the stationary points of Table I are indeed minima[17] and to compute a part [$(1/2)R$ ln det **F**] of the contribution of the librational entropy to the free energy {but not the other part [$(1/2)R$ ln det **G**], the calculation of which requires the procedure described in section II.B}. From an examination of the resulting free energies (last column of Table I), it can be seen that the order of the minima in the $F(Q)$ surface can sometimes be inverted when statistical weights are computed (*cf.* minima no. 14 and 21); this effect is also demonstrated in section II.H, where the complete term [$(1/2)R$ ln det **GF**] is calculated for a cyclic pentapeptide. This shows that the structure of highest statistical weight need not correspond to the *global* minimum of the $F(Q)$ surface. An additional example of the importance of the librational entropy [even though only $(1/2)R$ ln det **F** was computed] is its effect in producing a change in helix sense in poly-L-alanine as the temperature is raised[18] (see section II.F). Another point of interest in the data of Table I is that, among the *very limited* range of conformations examined, extended conformations seem to have the highest statistical weight. This result is in agreement with observations of Ingwall,

(76) S. J. Leach, G. Nemethy, and H. A. Scheraga, *Biopolymers*, **4**, 887 (1966).
(77) G. Nemethy, D. C. Phillips, S. J. Leach, and H. A. Scheraga, *Nature*, **214**, 363 (1967).
(78) T. Ooi, R. A. Scott, G. Vanderkooi, and H. A. Scheraga, *J. Chem. Phys.*, **46**, 4410 (1967).
(79) R. F. Epand and H. A. Scheraga, *Biopolymers*, **6**, 1551 (1968).
(80) J. F. Yan, G. Vanderkooi, and H. A. Scheraga, *J. Chem. Phys.*, **49**, 2713 (1968).
(81) A. K. Chen and R. W. Woody, *J. Amer. Chem. Soc.*, **93**, 29 (1971).
(82) J. N. Vournakis, J. F. Yan, and H. A. Scheraga, *Biopolymers*, **6**, 1531 (1968).
(83) M. Hashimoto and J. Aritomi, *Bull. Chem. Soc. Jap.*, **39**, 2707 (1966).
(84) M. Hashimoto, *ibid.*, **39**, 2713 (1966).
(85) M. Hashimoto, S. Arakawa, and K. Nakamura, Preprints International Symposium on Macromolecular Chemistry, Tokyo-Kyoto, Sept 1966, p IX-12.
(86) M. Hashimoto and S. Arakawa, *Bull. Chem. Soc. Jap.*, **40**, 1698 (1967).
(87) R. H. Karlson, K. S. Norland, G. D. Fasman, and E. R. Blout, *J. Amer. Chem. Soc.*, **82**, 2268 (1960).
(88) E. M. Bradbury, A. R. Downie, A. Elliott, and W. E. Hanby, *Proc. Roy. Soc., Ser. A*, **259**, 110 (1960).
(89) G. D. Fasman in ref 7, p 515.
(90) E. H. Erenrich, R. H. Andreatta, and H. A. Scheraga, *Biopolymers*, **7**, 805 (1969).
(91) E. H. Erenrich, R. H. Andreatta, and H. A. Scheraga, *J. Amer. Chem. Soc.*, **92**, 1116 (1970).

Table I

Free Energies of Minima of Deca-L-alanine[17]

Starting point[a] *for minimization*	*Minimum no.*	$F_1(Q_1)$[b]	*(1/2)RT ln [det* $F]_{Q-Q_1}$[b]	F[b]
α helix	1	17.71	31.44	49.25
(ϕ = 120°,	2	17.67	31.90	49.57
ψ = 130°)	3	13.75	33.02	46.77
	4	8.36	32.16	40.52
	5	8.11	31.20	39.31
	6	7.84	32.50	40.34
	7	7.63	31.32	38.95
	8	5.63	30.32	35.95
	9	−0.94	27.89	26.95
	10	−1.75	28.78	27.03
	11	−8.25	27.62	19.37
	12	−12.34	28.47	16.13
	13	−16.47	30.73	14.26
	14	−20.36	25.27	4.91
β conformation	15	−25.37	20.41	−4.96
(ϕ = 60°,	16	−25.44	17.11	−8.33
ψ = 240°)	17	−25.53	16.51	−9.02
	18	−27.52	17.35	−10.17
Other conformation	19	−5.10	27.75	22.65
(ϕ = 60°,	20	−17.11	28.84	11.73
ψ = 180°)	21	−22.29	28.94	6.61

[a] Starting point for the sequential application of Davidon's minimization technique[21] and one of the methods[25,26] of searching for lower energy conformations. [b] Energies are in kcal/mol; free energies are calculated at 25°. See eq 4 and related discussion for definitions of these symbols.

et al.,[92] on the optical rotatory dispersion of a sandwich copolymer of deca-L-alanine and DL-lysine, which gave no indication of α-helical structure in aqueous solution (see section III.B), although the results of Ingwall, *et al.*,[92] may have included a possible effect of electrostatic repulsion between the DL-lysine end blocks; on the other hand, poly-L-alanine blocks of larger chain length *did* adopt the α-helical conformation.[92]

E. INTERMOLECULAR INTERACTIONS

While the results discussed in section II.D pertain to single-chain polymers, in which only intramolecular interactions are involved, recent attempts have been made to treat multiple-stranded structures, and also packed structures in crystals of polyamino acids. Calculations have been carried out for assemblies of α-helical structures, both in the presence and absence of water, multiple-stranded polyglycine, collagen models, and crystalline β and ω structures.

Parry and Suzuki[93] found that the intramolecular energies of single-stranded α helices and coiled coil (pitch length = 186 Å and radius = 5.5 Å) conformations of poly-L-alanine are very similar. However, when the energies of small (parallel and antiparallel) assemblies of coiled coil ropes were compared,[94] it was found that the coiled coil form is more stable because of intermolecular interactions. Parry[95] has recently made use of the coiled coil conformation, together with a

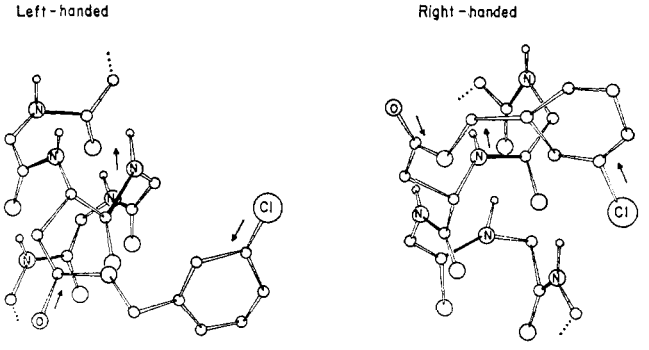

Figure 4. Orientation[49] of the side chains of the left- and right-handed α helices of poly(*m*-chlorobenzyl-L-aspartate). The solid arrows represent the directions of the C–Cl, ester, and amide dipoles, respectively.

repeating distribution of nonpolar residues, to account for some of the properties of α-fibrous proteins.

Assemblies of α helices, formed from a single chain folded in hair-pin fashion in a polymer such as poly-L-alanine, can be considerably more stable than a straight α helix[92,96] because of hydrophobic bonding between methyl side chains of adjacent helices. Taking the hydration term $V(Q)$ into account in the computation of $F(Q)$, it has been found[97] that two antiparallel α helices, whose methyl groups are enmeshed in a gear-like manner, form a very stable structure. A considerable increase in stability is achieved by surrounding a single central α helix with as many as six α helices, with hydrophobic bonding among them. These structures, comprised of interacting α helices, may be regarded as incipient forms of one of the compact structures found in globular proteins.

Ramachandran, *et al.*,[98] have examined the hydrogen bonding in polyglycine II, a multiple-stranded collagen-like helix, and Venkatachalam and Ramachandran[5b] have made similar observations for polyglycine I, an assembly of polypeptide chains in the extended β conformation. Although no conformational energy calculations were carried out, it was concluded that intermolecular interactions stabilize these two structures (to approximately the same extent), compared to the single-stranded polyglycine chain, for which the right- and left-handed α-helical conformations are the most stable.[56]

Hopfinger and Walton[99-101] have carried out calculations on several polypeptides which serve as models for collagen. For poly-L-hydroxyproline,[99] a minimum-energy packed structure, in agreement with experimental results, was obtained; it involves strong intermolecular C=O···H—O hydrogen bonding, with the steric energy making the dominant contribution in the determination of the packed structure. The packed arrangements of poly(Gly-Pro-Gly) also form in response primarily to intermolecular interactions.[100] In the case of various polymers of the Gly-Pro-X and Gly-X-Pro type,[101] some form stable collagen-like triple helices involving intermolecular interactions in proline-sparse regions and in-

(92) R. T. Ingwall, H. A. Scheraga, N. Lotan, A. Berger, and E. Katchalski, *Biopolymers*, 6, 331 (1968).
(93) D. A. D. Parry and E. Suzuki, *ibid.*, 7, 189 (1969).
(94) D. A. D. Parry and E. Suzuki, *ibid.*, 7, 199 (1969).
(95) D. A. D. Parry, *J. Theor. Biol.*, 26, 429 (1970).
(96) D. C. Poland and H. A. Scheraga, *Biopolymers*, 3, 305, 335 (1965).
(97) D. N. Silverman and H. A. Scheraga, unpublished results.
(98) G. N. Ramachandran, C. Ramakrishnan, and C. M. Venkatachalam in "Conformation of Bioplymers," G. N. Ramachandran, Ed., Academic Press, London, 1967, p 429.
(99) A. J. Hopfinger and A. G. Walton, *J. Macromol. Sci., Phys.*, 3, 195 (1969).
(100) A. J. Hopfinger and A. G. Walton, *ibid.*, 4, 185 (1970).
(101) A. J. Hopfinger and A. G. Walton, *Biopolymers*, 9, 29 (1970).

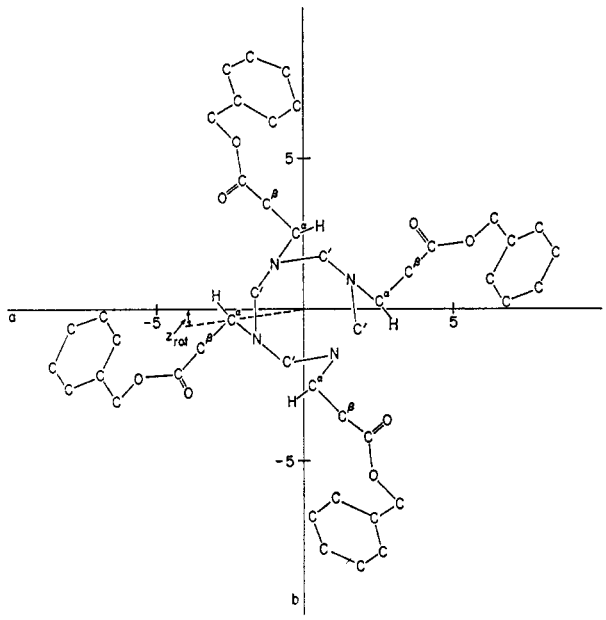

Figure 5. An a–b projection[103] of one complete turn in the minimum-energy conformation for the left-handed ω helix (four residues) of poly-β-benzyl-L-aspartate. The helix axis is perpendicular to the plane of the paper. (The scale from 0 to 5 is in Å.) z_{rot} is a right-handed rotation about the helix axis when looking from the N to the C terminus, and is 5.4° in this figure.

tramolecular interactions in proline-rich regions. Since effects, such as antiparallel chain folding and the influence of solvent, were not taken into account, it is not yet possible to assess how well these models conform to a collagen structure. Calculations along these lines are in progress in this laboratory.[102]

Recently a technique has been used to study the packing of both small molecules and also polypeptides in crystals.[103] The application of the method to small molecules, for the purpose of refining the empirical energy parameters, has been reviewed earlier[6] and referred to briefly in section II.B. The procedure has also been used to calculate the conformation and crystal (*i.e.*, intermolecular) packing energies of the β structure of poly-L-alanine and the ω helix of poly-β-benzyl-L-aspartate. The structure is influenced by the internal torsional energies, as well as by both the intramolecular and intermolecular nonbonded, hydrogen-bonded, and electrostatic energy contributions. In these calculations the energy was minimized both with respect to the intramolecular variables (*i.e.*, the backbone and side-chain dihedral angles) and also variations in the intermolecular orientations and crystal packing of the homopolymers. In the case of poly-L-alanine, the lowest energy was found for a β structure, involving a statistical fluctuation, among several types of packing of the sheets, as proposed by Arnott, *et al.*,[104] from their X-ray diffraction studies. The calculations indicate that the observed packing is caused primarily by intermolecular hydrogen bonds and the nonbonded side-chain interactions, with very little influence from the intramolecular interactions. In the case of poly-β-benzyl-L-aspartate, a left-handed ω helix exists in stretched fibers and films,[105] whereas a left-handed α helix exists in solution.[89] The lowest energy structure, which is compatible with the X-ray data of Bradbury, *et al.*,[105] is shown in Figures 5 and 6, which illustrate the conformation of the single chain and the packing arrangement, respectively. In the lowest energy structure, the amide group departs slightly from planarity, ω (the dihedral angle for rotation about the peptide bond[20b]) being −3.0°. For this polyamino acid, the major contribution to the total interaction energy arises from the intramolecular interactions, although the observed conformation and packing result from contributions from both the intermolecular and intramolecular interactions. The technique used in these calculations is applicable, not only to small molecules[6] and homopolymers, but also to heteropolymers and proteins (*e.g.*, to gramicidin S, lysozyme, etc.). Thus, calculations, such as those described in section II.H for single molecules, can be extended to crystal structures for comparison with X-ray diffraction data. At the present stage of the calculations, it is necessary to know in advance at least the symmetry and number of molecules per unit cell of the crystal, which is the minimal information necessary from X-ray diffraction studies.

F. HELIX–COIL TRANSITIONS IN HOMOPOLYMERS

A homopolyamino acid can undergo a thermally induced helix–coil transition. In the current theories (*e.g.*, those of Zimm and Bragg[106] or Lifson and Roig[107]), the behavior of the system is essentially a balance between the low energy and entropy of the helix and the high energy and entropy of the coil, the transition curves being characterized by two phenomenological parameters σ and s. If u_i and v_j are the statistical weights of sequences of i coil states and j helical states, respectively, then[10,18]

$$s = \lim_{j\to\infty}\left(\frac{v_{j+1}}{u_{j+1}}\right)\Big/\left(\frac{v_j}{u_j}\right) \tag{8}$$

$$\sigma = \sigma' s^2 \text{ with } \sigma' = \lim_{j\to\infty}(1/s^j)(v_j/u_j) \tag{9}$$

The statistical weights u_i and v_j are computed from the conformational energies of single residues and helical sequences, respectively. Such calculations have been carried out for polyglycine and poly-L-alanine in the absence[18] and presence[47] of water (also with the omission of the factor det **G**).

In the absence of water (except for the inclusion of a type-2 solvent effect, as discussed in section II.B), u_i was computed by taking into account only short-range contributions to $F(Q)$ and carrying out the integration of eq 2 by a matrix method[10,18] (involving the calculation of eigenvalues). In computing v_j, the condition of regularity was not imposed,[18] and the results of Figure 7 were obtained. It can be seen that, for a long chain, the central portion does remain regular with departures from regularity near the ends; this diffuseness extends over about five residues at each end. As in the case of deca-L-alanine,[17,25] the entropy factor in v_j was obtained from the second derivative [of $U(Q)$ in this case]. It was shown[18] that the calculated values of σ are not affected by the binding of solvent molecules to free NH and CO groups,[18] and hence

(102) M. Miller and H. A. Scheraga, work in progress.
(103) R. F. McGuire, G. Vanderkooi, F. A. Momany, R. T. Ingwall, G. M. Crippen, N. Lotan, R. W. Tuttle, K. L. Kashuba, and H. A. Scheraga, *Macromolecules*, **4**, 112 (1971).
(104) S. Arnott, S. D. Dover, and A. Elliott, *J. Mol. Biol.*, **30**, 201 (1967).
(105) E. M. Bradbury, L. Brown, A. R. Downie, A. Elliott, R. D. B. Fraser, and W. E. Hanby, *ibid.*, **5**, 230 (1962).
(106) B. H. Zimm and J. K. Bragg, *J. Chem. Phys.*, **31**, 526 (1959).
(107) S. Lifson and A. Roig, *ibid.*, **34**, 1963 (1961).

that the resulting values of σ may be compared with experimental ones. The results[18] are quite sensitive to the energy parameters chosen, but values of σ in the range of experimental ones were obtained for these polymers. Further, it was found that, in the absence of water, poly-L-alanine transforms from the right- to the left-handed α helix, before melting to the coil, as the temperature is raised,[18] because of the inclusion of the entropy term $(1/2)R$ ln det **F**. This may account for the observation[108-110] of such thermally induced changes in helix sense of other polymers *in organic solvents*, although the effect may also possibly be due to thermal (intra- and intermolecular) vibrations which manifest themselves as an increase in the effective radii of the various atoms, especially hydrogen atoms.[111]

The influence of water was taken into account[47] by inclusion of the three solvent effects referred to in section II.B. As pointed out in this section, the type-1 solvent effect was taken into account by inclusion of the free energy of water binding to NH and CO groups of the coil residues (see eq 6), with ΔH_B and ΔS_B being obtained by fitting the theoretical melting curves (obtained by calculation in the absence of water[18]) to experimental data[92] for poly-L-alanine in water. The type-2 solvent effect was included by using a low dielectric constant for short-range electrostatic interactions and cutting off the interactions beyond about 6 Å. Type-3 solvent effects were shown to be small for polyglycine and poly-L-alanine. While poly-L-alanine was found to melt in the temperature range between 0 and 100°, polyglycine was found to be in the random-coil form over this whole range. The use of the cutoff of the type-2 solvent effect beyond 6 Å improved the agreement between the calculated and experimental values[92] of σ for poly-L-alanine. The type-1 solvent effect altered the values of s (computed for polymers in the absence of water[18]) from those corresponding to abnormally high melting temperatures to those with melting temperatures in agreement with experimental data.[92] The enthalpy and entropy contributions (both of which are temperature dependent) to σ and s were resolved into their short-range and long-range components. The greater stability of the poly-L-alanine α helix compared to the polyglycine α helix in water arises from side-chain-to-backbone interactions in poly-L-alanine; the dominance of such near-neighbor side-chain-to-backbone interactions in determining the conformation of a residue in a polypeptide chain will be considered again in section II.J. In the presence of water, the left-handed α helix of poly-L-alanine is not stable; hence, in water, this polymer exhibits a conformational transition from the right-handed α helix to the random coil form when the temperature is raised.[47,92] It remains to be seen whether sufficiently accurate experimental data can be obtained to test the prediction that σ is a temperature-dependent quantity; some preliminary experimental data are available[112] to indicate that σ may indeed be temperature dependent.

As a result of the calculations discussed here, it has been possible to formulate a simple model of the helix–coil transition,[47,118] which is applicable to both homopolymers and copolymers. This model is based on the demonstrated dominance of short-range interactions.

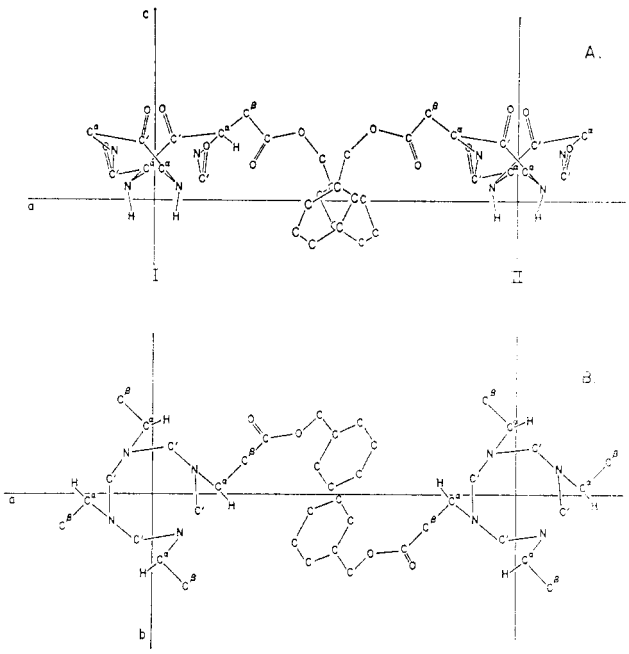

Figure 6. *a–c* (A) and *a–b* (B) projections[103] of two packed ω helices (I and II) in their minimum-energy conformation for poly-β-benzyl-L-aspartate, showing the arrangement of the side chains and the overlapping packing. For clarity, only the backbone atoms of one complete helix turn, plus one side chain in each chain, are shown for each ω helix.

G. HELIX–COIL TRANSITIONS IN COPOLYMERS

The calculations described in section II.F for a homopolymer can also be carried out for the helix–coil transition in a copolymer having a specific sequence of, say, two kinds of units A and B. Since such calculations are, at present, in their initial stages, we confine our discussion here to the *phenomenological* aspects of the helix–coil transition in *random* copolymers, *i.e.*, to the theoretical and experimental approach to the determination of the phenomenological parameters σ_A and s_A (for each kind of unit A in the random copolymer), and to the applicability of the experimental values of σ_A and s_A to obtain information about the folding of a protein chain; it is expected that these experimental values of σ_A and s_A in copolymers ultimately can be obtained by calculation from molecular quantities, as described for homopolymers in section II.F. The use of values of σ_A and s_A for an A-type residue in a protein (as described below) is based on the result (already indicated in section II.F and developed further in section II.J) that the conformation of an amino acid residue in a synthetic homopolymer or copolymer, or in a protein, is determined mainly by short-range interactions between a side chain and those backbone atoms nearest to it.

Accepting the dominance of short-range interactions, σ_A and s_A for any residue A can be determined, in principle, from helix–coil transition experiments on homopolymers of A in

(108) E. M. Bradbury, B. G. Carpenter, and H. Goldman, *Biopolymers*, 6, 837 (1968).
(109) E. M. Bradbury, B. G. Carpenter, and R. M. Stephens, *ibid.*, 6, 905 (1968).
(110) C. Toniolo, M. L. Falxa, and M. Goodman, *ibid.*, 6, 1579 (1968).
(111) N. Lotan, F. A. Momany, J. F. Yan, G. Vanderkooi, and H. A. Scheraga, *ibid.*, 8, 21 (1969).
(112) N. Gō, M. Gō, V. S. Ananthanarayanan, P. H. Von Dreele, and H. A. Scheraga, work in progress.

(113) N. Gō, P. N. Lewis, M. Gō, and H. A. Scheraga, work in progress.

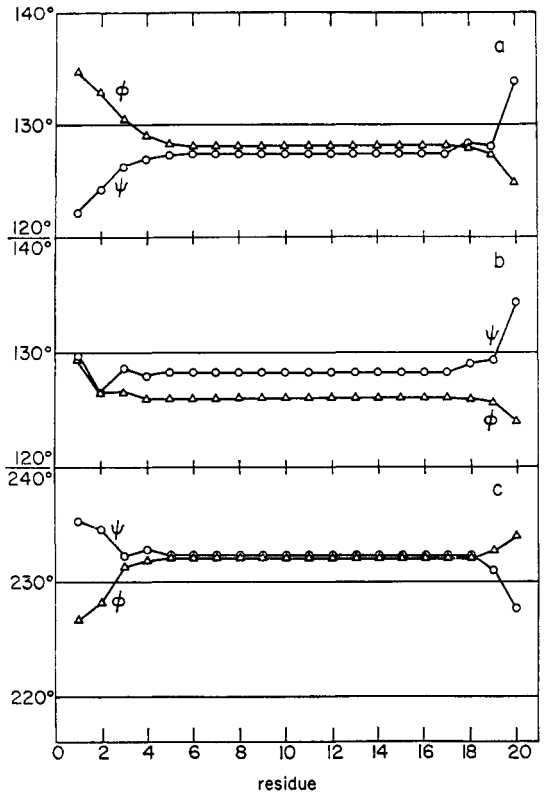

Figure 7. Variations of dihedral angles with position in the chain for (a) α helix of polyglycine, (b) the right-, and (c) the left-handed α helices of poly-L-alanine.[18]

say, water. However, as will be discussed in section III.A, there are experimental limitations to such an approach, and it is therefore necessary to resort to another procedure. For this purpose, random copolymers of two types of residues A and B are used in the melting experiments. Such copolymers are still devoid of the compact folding (involving long-range interactions) characteristic of globular proteins, and hence the copolymer chain (like the homopolymer) conforms to the one-dimensional Ising model. In the random copolymer, the assumption of the dominance of short-range interactions implies that the state of a given residue depends on whether it is A or B, but is independent of whether its neighbors are A or B; however, its statistical weight does depend on the conformational state of its neighbors, according to the dictates of the nearest-neighbor Ising model. The rationale for using random copolymers of A and B units to obtain experimental values of, say, σ_A and s_A, when σ_B and s_B are known, is discussed in sections III.A and III.D. Here we simply note that the phenomenological theory for the helix–coil transition in such random copolymers has been developed,[114,115] so that the melting curve for the copolymer can be related to the parameters for the melting of the corresponding homopolymers. If the sequences of A and B units in the copolymer are not very long, the sections of A and B units do not behave independently. Instead, the melting behavior of any section of the chain depends primarily on whether it is richer in A or B, even if the corresponding homopolymers of these amino acids have significantly different melting points. Hence, we may regard σ_A and s_A for all naturally occurring amino acids as determinable from experiments of the type described in section III.D.

Assuming then that values of σ_A and s_A for all naturally occurring amino acids are known from experiments on random copolymers, as discussed in section III.D, and that short-range interactions are also the most important ones (although not exclusively so) in determining the conformation of each residue in a protein, we may consider the use of these parameters to obtain information about the folding of a protein chain which has a given amino acid sequence. The native state of a globular protein involves specific long-range interactions which cannot be treated by the simple Ising model used in, say, the Zimm–Bragg[106] and Lifson–Roig[107] theories for homopolymers. However, above the denaturation temperature, the protein is devoid of tertiary structure and hence, presumably, of long-range interactions other than excluded volume effects, and the polypeptide chain conforms to the one-dimensional Ising model. Thus, we can compute[85] the probability that any given residue of the denatured chain will be in the helical or coil conformation, respectively. The statistical weight or partition function, Z, and the probability, $P_H(i)$, that the ith amino acid (of type A) in a chain of N residues is in the helical conformation are given by

$$Z = (0,1)\left[\prod_{j=1}^{N} \mathbf{W}_A(j)\right]\begin{pmatrix}1\\1\end{pmatrix} \tag{10}$$

and

$$P_H(i) = (0,1)\left[\prod_{j=1}^{i-1} \mathbf{W}_A(j)\right]\frac{\partial \mathbf{W}_A(i)}{\partial \ln s_A(i)}\left[\prod_{j=i+1}^{N} \mathbf{W}_A(j)\right]\begin{pmatrix}1\\1\end{pmatrix}\Big/Z \tag{11}$$

respectively, where $\mathbf{W}_A(j)$ is the matrix of statistical weights for the jth residue which is of amino acid type A, *viz.*

$$\mathbf{W}_A(j) = \begin{bmatrix} s_A(j) & 1 \\ \sigma_A(j)s_A(j) & 1 \end{bmatrix} \tag{12}$$

$s_A(j)$ is the statistical weight assigned to this residue when it is in a helical conformation and preceded by a residue in the helical conformation, and $\sigma_A(j)s_A(j)$ is the statistical weight assigned to this residue when it is in a helical conformation and preceded by a residue in the coil conformation. The σ of eq 12 is the same as σ' of eq 9, but we omit the prime here for convenience; the use of eq 12 implies that the statistical weight of a helical sequence is $\sigma' s^j$. The average helix content, θ_H, is given by

$$\theta_H = \frac{1}{N}\sum_{i=1}^{N} P_H(i) \tag{13}$$

By using experimental values of σ_A and s_A, determined as described in section III.D, it is possible to compute values of $P_H(i)$ and θ_H for proteins above the denaturation temperature. The implications of such calculations for the nucleation and folding of protein chains is discussed in sections II.J and II.K.

H. CYCLIC POLYPEPTIDES

By allowing all dihedral angles to vary independently, as in the case of nonregular homopolymer helices,[18,25] we are no longer restricted to the regular structures of homopolymers. Thus,

(114) G. W. Lehman and J. P. McTague, *J. Chem. Phys.*, **49**, 3170 (1968).
(115) D. Poland and H. A. Scheraga, *Biopolymers*, **7**, 887 (1969).

calculations can be carried out for the copolymers discussed in section II.G, and for polypeptides of arbitrary (but specific) sequence, such as natural polypeptides or proteins. Earlier work on the naturally occurring cyclic polypeptides oxytocin,[44] vasopressin,[44] and gramicidin S,[45,116,117] in which $F(Q)$ and $U(Q)$, respectively, were minimized (with all dihedral angles as independent variables and with a loop-closing potential) has been reviewed.[1,2,6] We describe here some further recent work on gramicidin S and on some synthetic cyclic polypeptides.

Further work has been done[80] on the minimization of $U(Q)$ of gramicidin S, starting with the conformation proposed by Hodgkin and Oughton,[118] Schwyzer,[119,120] and Schwyzer and Ludescher.[121] The structure of minimized energy[80] (designated GS_V) has an energy of −112 kcal/mol, compared to −98 and −100 kcal/mol, respectively, for two previously discussed[45,116] structures. Structure GS_V is compatible with the nmr results of Schwyzer and Ludescher[121] and of Stern, *et al.*,[122] although there are some differences. A structure, similar to GS_V, and designated GS_{V1} (with an energy of −114 kcal/mol), was also obtained by a procedure involving a statistical search of the $U(Q)$ surface rather than the selection of the Schwyzer conformation as a starting point. While the ring closure and twofold symmetry of structure GS_{V1} are excellent, and it has a lower energy than any found heretofore, it cannot yet be regarded as the most stable conformation of gramicidin S, since the whole $U(Q)$ surface (which contains many local minima) has not yet been explored.

A calculation[88] of the complete $F(Q)$ surface of cyclo(glycylglycylglycylprolylprolyl) has been carried out primarily to apply the procedures[12] (discussed in section II.A) for determining the most stable conformation. In this calculation $F(Q)$ was taken as $U(Q)$, approximation C (see section II.A), *i.e.*, rigid geometry, was used, and a condition of exact ring closure[87] was imposed, rather than applying a loop-closing potential (approximation B is presently being used[41] to extend the treatment of this molecule). Since this polypeptide has eight backbone dihedral angles and approximation C was used, the condition of exact ring closure makes two dihedral angles[87] (taken as the angles ψ_4 and ψ_5 between Pro and Pro and between Pro and Gly, respectively) independent. The conformational energy can then be plotted as contours on a two-dimensional diagram of ψ_4 *vs.* ψ_5. However, several such diagrams are required to represent the complete $U(Q)$ surface, since there are several sets of solutions of the six dependent dihedral angles for given values of the two independent angles. Thirteen local minima appeared within the range of 100 kcal of the global minimum in the energy surface. This is the first step of the three-step recipe given in section II.A. Three out of the 13 local minima (C′, D′, and E′ of Figure 8) have energies within *ca.* 1 kcal/molecule of the global minimum, the energies for all of the other ten conformations being much higher. In approximation C, the second step in the recipe is skipped, and the conformational entropy was calculated (in the third step of the recipe) for the three lowest minima; *i.e.*, det **G** was computed for all 13 minima, and det **F** only for minima C′, D′, and E′. It was found[19] that the values of $(1/2)R$ ln det **G** for the various minimum-energy conformations varied sufficiently to change the relative stabilities of some of the minimum-energy conformations when this entropy term is included, as indicated in section II.D. The energies and free energies of the three lowest minimum-energy conformations are shown in Table II.[19] Whereas conformation D′ has lower energy than the other two, the stabilities of conformations D′ and E′ are found to be essentially the same when the conformational entropy, $-(R/2)$ ln det **GF**, is included in the free energy. This means that, at 300°K, isolated molecules of

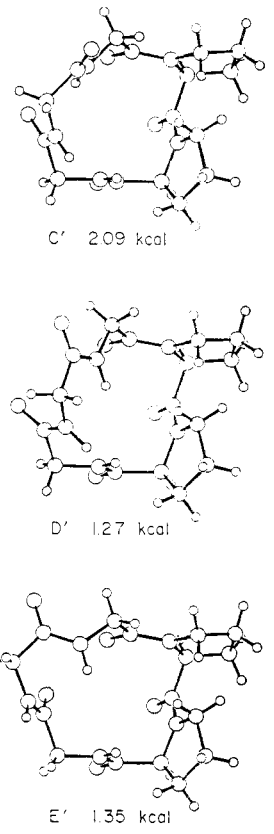

Figure 8. Conformations corresponding to the three lowest energy minima[19] of cyclo(Gly_3Pro_2). The values of the minimum energies are given in units of kcal/mol of molecule.

Table II

Energies and Free Energies of the Three Lowest Minimum-Energy Conformations of Cyclo(Gly_3Pro_2)[19]

	Local minimum		
	C′	*D′*	*E′*
E_{total}[a]	2.09	1.27	1.35
$-(R/2)$ ln det **F**[b]	3.12	4.77	4.99
$-(R/2)$ ln det **G**[b]	13.48	12.10	12.08
$-(R/2)$ ln det **GF**[b]	16.60	16.87	17.07
$E_{total} + (RT/2)$ ln det **GF**[c]	−2.89	−3.79	−3.77

[a] Total energy in units of kcal/molecule. [b] In entropy units/molecule. [c] Calculated for $T = 300$°K in units of kcal/molecule.

(116) A. M. Liquori, P. deSantis, A. L. Kovacs, and L. Mazzarella, *Nature*, **211**, 1039 (1966).
(117) G. Vanderkooi, S. J. Leach, G. Nemethy, R. A. Scott, and H. A. Scheraga, *Biochemistry*, **5**, 2991 (1966).
(118) D. C. Hodgkin and B. M. Oughton, *Biochem. J.*, **65**, 752 (1957).
(119) R. Schwyzer, *Chimia*, **12**, 53 (1958).
(120) R. Schwyzer, *Rec. Chem. Progr.*, **20**, 147 (1959).
(121) R. Schwyzer and U. Ludescher, *Biochemistry*, **7**, 2519 (1968).
(122) A. Stern, W. A. Gibbons, and L. C. Craig, *Proc. Nat. Acad. Sci. U. S.*, **61**, 734 (1968).

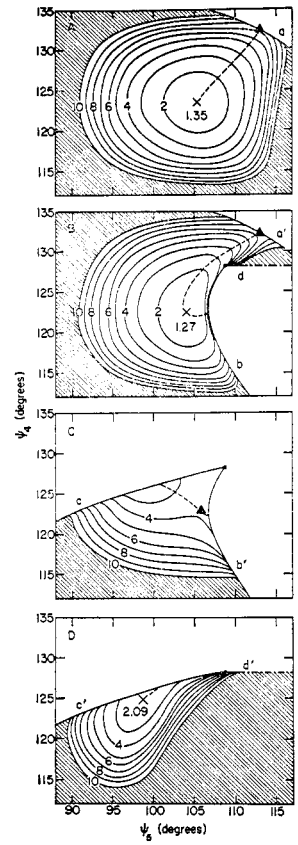

Figure 9. The contour diagram of the energy surface of cyclo-(Gly_3Pro_2) in the global energy minimum region.[19] Energies are given in units of kcal/mol of molecule. The three energy minima in this region are marked by an ×, and their energy values are given. Dotted lines show the easiest path ("reaction path") to go from one minimum to another. The two saddle points on the path are marked by the symbol ▲. The shaded region has energies higher than 10 kcal. See ref 19 for further description.

cyclo(Gly_3Pro_2) are a thermal mixture of an almost equal number of conformations D′ and E′ with a much smaller number of conformation C′. The energy contour diagram in the region of minima C′, D′, and E′ is shown in Figure 9; the detailed interpretation of these diagrams is given in ref 19.

Besides being of intrinsic interest, as far as the conformation of this cyclic pentapeptide is concerned, and also demonstrating the application of the principles discussed in section II.A, the complete energy surface will serve as a useful model for future tests of procedures for passing from one minimum to another (lower) one. The procedure[87] used for achieving exact ring closure is also applicable to the problem of introducing *local* deformations in the conformations of chain molecules; such a deformation technique may be useful in bypassing local minima.

During the course of the computations on closed rings,[87] it was proven that no cyclic tri- and tetrapeptides with planar trans peptide units having Pauling–Corey bond lengths and bond angles are geometrically capable of existing. Similar conclusions were reached by Ramachandran[123] and by Ramakrishnan and Sarathy,[124] who also treated cyclopentapeptides and cyclohexapeptides.

(123) G. N. Ramachandran, *Biopolymers*, **6**, 1494 (1968).
(124) C. Ramakrishnan and K. P. Sarathy, *Biochim. Biophys. Acta*, **168**, 402 (1968); *Int. J. Protein Res.*, **1**, 63, 103 (1969).

Work is now in progress in this laboratory[60] on the conformation of cyclohexaglycyl and on derivatives in which one or two glycines, respectively, are replaced by alanyl residues. Preliminary results indicate that cyclohexaglycyl may have several conformations of nearly equal energy, and thus exist as a mixture of these in solution.

As pointed out in section II.E, the calculations for gramicidin S, cyclo(Gly_3Pro_2), cyclohexaglycyl, etc., in packed crystalline arrays can be carried out by the procedure[103] which was applied to homopolymers; thus, comparison can be made between theoretical calculations and the results of X-ray diffraction studies.

I. PROTEINS

The same computer programs used in the calculations described in section II.H are applicable to single protein molecules of the size of lysozyme, with every backbone and side-chain dihedral angle taken as an independent variable. Also, the crystal packing program described in section II.E can be applied to protein *crystals*. In a much simpler calculation, short sections[86] (nonapeptides) of lysozyme are being treated independently to explore the role of short-range interactions in determining protein conformation (see section II.J).

On the other hand, the complete structure of lysozyme is being minimized[125] (to refine X-ray coordinates) by starting with a semirefined structure. The latter is obtained by a best fit of the X-ray data by a polypeptide chain of standard bond angles and bond lengths.[126] In this way, one may expect to obtain a refined structure without encountering the problem of multiple minima. Similar work is being carried on by Diamond[127] and by Levitt and Lifson.[128] While these procedures serve the very useful function of refining X-ray data, they, of course, do not fulfill the purpose of predicting protein conformation; for the latter, it will be necessary to obtain a satisfactory solution of the problem of multiple minima in the $F(Q)$ surface.

Calculations on hairpin-like structures of poly-L-alanine,[97] which may be regarded as incipient forms of the compact structures of globular proteins, have been mentioned in section II.E.

Also, in a preliminary stage, we are examining the preferred orientations of substrates in the clefts of enzymes[129] and the corresponding energies of enzyme-substrate complexes. For this purpose, the proteolytic enzymes chymotrypsin, trypsin, elastase, and thrombin are being considered. These enzymes appear to have similar amino acid sequences, but different specificities. A number of different substrates for each enzyme are placed in the cleft, and their minimum-energy conformations determined.

J. THE ROLE OF SHORT-RANGE INTERACTIONS

One approach to bypassing the problem of multiple minima in the $F(Q)$ surface is to try to find alternative methods of arriving at even a rough approximation to a protein structure, which can then be refined by current energy-minimization

(125) D. N. Silverman, K. D. Gibson, and H. A. Scheraga, unpublished work.
(126) N. Gō, S. Rosen, and H. A. Scheraga, unpublished work.
(127) R. Diamond, *Acta Crystallogr.*, **21**, 253 (1966).
(128) M. Levitt and S. Lifson, *J. Mol. Biol.*, **46**, 269 (1969).
(129) K. Platzer and H. A. Scheraga, unpublished work.

procedures. For this purpose, as mentioned in section II.A, the relative importance of short-range interactions in determining protein structure (already referred to in sections II.F and II.G) has been explored.[82,33,35,47] A short-range interaction is essentially one in which the energy depends on only one or two consecutive dihedral angles. As also discussed in a previous review,[6] it appears that the interactions of a side-chain R group with its neighboring backbone atoms play a major role (but not to the exclusion of longer range interactions) in determining the conformation of a residue.[32] The interactions between neighboring R groups seem to be of considerably lesser importance. This is borne out, not only in these computations on dipeptides,[32] but also (at least for alanine side chains) in calculations on the statistical weights[18,47] of helical and coil sequences, respectively. On the basis of such observations, the various residues were classified as h or c, depending on whether the right-handed α-helical conformation was, or was not, the preferred one. Using this classification,[33] it was found empirically that a helical sequence is initiated, and grows toward the C terminus of the protein, if four h's occur in a row (or four h's separated by at most one Gly), and that the occurrence of two c's in a row (or two c's separated by one or more Gly) terminates a helical sequence. With these rules, and with modification of the h and c designation of three residue types (based on X-ray diffraction studies of proteins), it was possible to assign correctly the h or c character of 78% of the amino acid residues in the four proteins whose structure was known at the time. A prediction[33] of the conformations of the residues of a staphylococcal nuclease, using these rules, was subsequently found[130] to be 70% correct.

Later, with the availability of several more protein structures, an experimental test[34] was made of these rules for helix initiation and termination, *i.e.*, of the tendency of small polypeptide segments of an intact protein to adopt a conformation which depends primarily on short-range interactions. In particular, with the larger sample available (seven proteins) it was possible to test the hypothesis that specific dipeptides (those that occur at least once at the C terminus of a helical sequence, and therefore have a strong tendency to act as "helix breakers") occur preferentially in the nonhelical regions throughout the protein sample. This is confirmed by the data of Figure 10, from which it can be seen that there is a sharp change in the probability of occurrence of dipeptides in the helical and coil regions, respectively, at the C terminus of the helical region. Further, whereas the overall content of coil residues in the seven proteins is ~70%, the dipeptides near the C terminus (in the group designated C+2 in Figure 10) have a 90% probability of occurring in the coil region. No such structure of the histogram is observed when single residues, rather than dipeptides, are considered; *i.e.*, the combination of *two* c residues appears to be required to terminate a helix, as proposed in the original rules.[33] While there is a similar discontinuity at the N terminus of a helical sequence (see Figure 11), there is no corresponding peak in the coil region, which indicates that helical sequences grow from the N toward the C terminus and thus are terminated at the C rather than at the N terminus. Additional evidence of the difference between the N- and C-terminal coil regions was obtained from a study of the chemical composition of these two

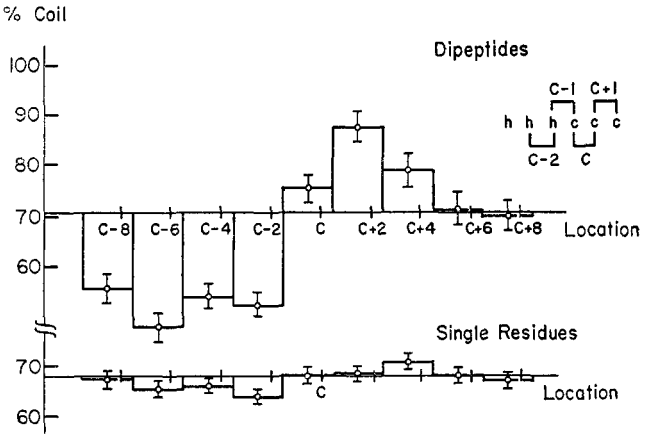

Figure 10. Distribution of dipeptides[34] near the C-terminal ends of helical segments. The base line for each histogram corresponds to the per cent coil for all dipeptides (or single residues, respectively) in the seven-protein sample.

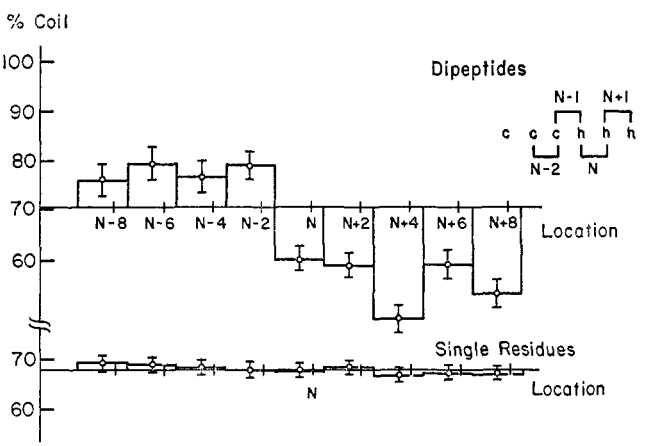

Figure 11. Distribution of dipeptides[34] near the N-terminal ends of helical segments. The base line for each histogram corresponds to the per cent coil for all dipeptides (or single residues, respectively) in the seven-protein sample.

regions; the N-terminal coil region was found to have essentially the same composition as coil regions throughout the protein sample, whereas the C-terminal region has a much greater percentage of polar and helix-breaking residues. Thus, the data on coil regions of known protein structures provide additional evidence of the importance of short-range interactions in determining protein conformation. A similar analysis of the distribution of amino acid residues among helical and nonhelical regions in globular proteins has recently been carried out by Ptitsyn.[131]

With this demonstration of the importance of short-range interactions, an attempt was made[35] to find a more quantitative basis (rather than the designation h and c, and rules associated with combinations of h and c) for determining the propensity of a given amino acid sequence to be helical or not. In particular, continuing the discussion of section II.G, it was shown that the parameters σ_A and s_A, for each residue of type A (obtained, in principle, from melting data on helices of homopolyamino and copolyamino acids), can be used to establish a correlation between the calculated values of $P_H(i)$

(130) D. C. Richardson, private communication.

(131) O. B. Ptitsyn, *J. Mol. Biol.*, 42, 501 (1969).

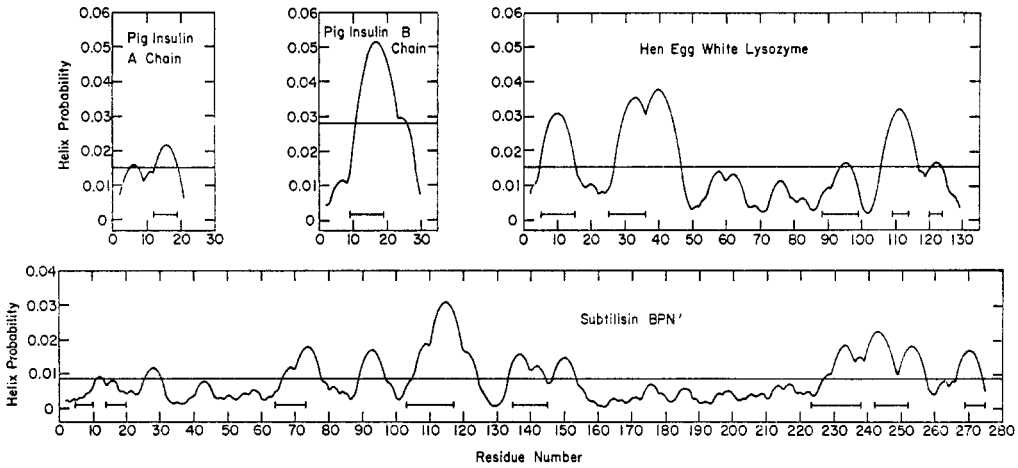

Figure 12. Helix probability profiles[35] for three proteins in the denatured form. The ordinates correspond to $P_H(i)$ computed from eq 11. The horizontal bars (|——|) denote those regions of the protein found to be in the right-handed α-helical conformation by X-ray diffraction analysis. The horizontal line is the computed value (eq 13) of θ_H.

of eq 11 for *denatured* proteins and the experimentally observed helical regions in the corresponding *native* structures. Pending the acquisition of experimental values of σ_A and s_A for all naturally occurring amino acids (as discussed in section III.D), a tentative assignment (based on the limited available data) was made, and the values of $P_H(i)$ and θ_H computed for 11 denatured proteins. Illustrative data are shown in Figure 12 for three denatured proteins. A strong positive correlation was found between those residues for which $P_H(i)$ exceeded the average helical content θ_H in the *denatured* condition and those parts of the sequence which were generally found to be helical in the *native* protein. In order to make this result quantitative, the following rule was used: a residue was predicted to be in the helical state in the native protein if its probability (*i.e.*, its propensity for being in a helical conformation in the denatured protein) exceeded that of the mean for the chain. On the basis of this criterion, the conformations of 68% of the residues in the 11 proteins were predicted correctly from a comparison of $P_H(i)$ and θ_H. Considering only those residues experimentally observed to be helical, 64% of these were predicted correctly for the 11-protein sample. These compare with values of 71 and 41%, respectively, when the earlier criterion,[88] discussed above, is applied to the same (larger) sample of 11 proteins. This suggests that the incipient helical regions in the denatured chain may serve to nucleate the folding to form the native protein (see section II.K). The procedure discussed here (based on eq 10–13, which incorporate the near-neighbor correlations characteristic of the one-dimensional Ising model) is being extended by including next-nearest-neighbor interactions.

A further study of the role of short-range interactions, which may shed light not only on their importance in determining the conformations of α-helical regions but of other regions as well, is one involving calculations of the conformational energies of 14 specific (consecutive, nonoverlapping) nonapeptide segments of lysozyme.[86] Each segment was first isolated from the rest of the lysozyme molecule. Then, using the dihedral angles (as determined by refinement[126] of the results of X-ray analysis of the intact protein) as the initial ones, $F(Q)$, including the hydration contribution, of each nonapeptide segment was minimized. This energy was then compared to those found by minimizing from conformations which were perturbed from the native conformation by varying the initial dihedral angles of the *central* residue. If short-range interactions play an important role in determining protein conformation, then the minimized energies of the perturbed systems (corresponding to conformations in other local energy minima) should be greater than those of the native systems. The preliminary, tentative results indicate that the energies of the native structure are the lowest in six of the first seven nonapeptides near the N terminus of lysozyme, which contains two of the three regular segments of α helix and the one β-pleated-sheet structure; on the other hand, the energies of most of the perturbed structures are the lowest in the remaining nonapeptides near the C terminus, which contains one regular segment of α helix and two short irregular segments (each one having one turn of helix, bad Pauling–Corey parameters, and no hydrogen bonds).

Thus, in the N-terminal region of the protein, where most of the α and β structure is present, the short-range interactions play a dominant role in determining conformation. Whether these interactions are of major importance only for regions with much α or β structure, or whether they are important near the N-terminus of most proteins (to aid in folding as the protein comes off the ribosome), will have to be determined on the basis of similar studies on other proteins. However, as discussed in section II.K, the nucleation centers for the folding of the chain need not be confined to the N-terminal portion.

It is possible that the procedure used to study the nonapeptides (in which the energy is minimized by varying the conformation of the central residue) may be applicable to the determination of the most stable conformations of such short sequences of amino acids. Then, by treating overlapping nonapeptides, it might be possible to cycle several times through the whole protein to arrive at a rough conformation, which would then serve as a starting point for conventional energy minimization. The possibility of using this approach to bypass the problem of multiple minima in the $F(Q)$ surface is being actively explored.[86]

K. NUCLEATION AND FOLDING OF PROTEIN CHAINS

The results presented in section II.J are of importance for an understanding of how the three-dimensional structure of a pro-

tein is nucleated in specific regions so that the whole chain may then fold up into the native structure.

In discussing the question of the nucleation and folding of a protein chain, consider first the formation of an α-helical section of j residues in a *homopolymer*. In the nearest-neighbor Ising model,[106,107,182] the statistical weight of such a section is $v_j = \sigma s^{j-2} = \sigma' s^j$. In addition, there is a combinatorial entropy which stabilizes such a helical section; this entropy arises from the fact that this helical sequence can appear in many locations in the chain. However, in a copolymer, helical sequences tend to be more localized in specific parts of the chain,[182] with a higher probability the greater the values of σ_A and s_A for the specific residues in the given local amino acid sequence. In the absence of long-range interactions, the formation of a helical region in a *given part* of a *protein* chain depends primarily on the factor $\sigma' s^j$ rather than on the combinatorial entropy. Since $\sigma' \backsim 10^{-4}$, v_j can approach unity only if s^j approaches 10^4. But s varies only from $\backsim 1.02$ to $\backsim 0.98$ throughout the transition range of a homopolymer. Thus, even with the value of 1.02 for s, j would have to be ~ 500 in order that s^j approach 10^4. This is simply a statement of the well-known fact that helical sequences have to be very large to be stable—*in a system behaving according to the one-dimensional Ising model.* In such a system, the most probable length of a helical sequence in the middle of the transition range (*i.e.*, at $s = 1$) is $\backsim \sigma^{-1/2}$ (or $\backsim 100$ residues), and shorter helical sequences have a lower probability of occurrence.[106,107,182] If one wanted to assign a value much greater than 1.02 to s, in order to achieve stability for *short* helical sections, one would encounter the dilemma that a chain with such helical sequences would not be predicted to denature in the accessible temperature range of 0–100°, whereas most proteins are observed to do so. However, short helical sequences ($\lesssim 10$ residues long) are known to occur in native proteins. How can we account for this and, at the same time, for the longer helical regions in myoglobin, etc.? Clearly, a short helical sequence cannot be stable in a polypeptide conforming to the one-dimensional Ising model. However, our view[85] is that, because a protein is not a homopolymer, small regions in the amino acid sequence of the protein can be densely populated with helix-making residues, which make it more likely that an α helix will form in that region compared to other regions, but that the stabilization of these incipient α-helical structures into actual short helical sequences in the native protein is accomplished by means of specific long-range interactions. It is just these long-range interactions which lead to globularity and thus the inapplicability of the one-dimensional Ising model in a native protein. On the other hand, this model is still applicable to a copolymer of, say, A and B units, in those cases in which the presence of a second component does not disrupt the one-dimensional character of the α-helical form of the first component.[114,115,182]

The correlations between the regions of higher helical propensity in the denatured state and the helical regions in the native structure (shown in Figure 12[85]) suggest that, during renaturation, the protein chain acquires *specific* long-range interactions which tend to stabilize these short helical regions. The difficulty of forming this initial long-range interaction (nucleation) introduces naturally the aspect of cooperativity with respect to the formation of the three-dimensional globular structure. In other words, the incipient formation of α-helical regions (among those residues with a propensity to be helical), stabilized by specific long-range interactions, may constitute the nucleation process for the refolding of the protein chain. The remainder of the protein molecule could then fold around these stabilized helical regions. Alternatively, in a protein with low helix content, other backbone conformations (*e.g.*, β structures) may possibly serve as nucleation centers.

Again it should be noted that, although the one-dimensional Ising model can be applied to the denatured state, it is physically unrealistic for systems in which long-range interactions (beyond nearest or next-nearest neighbors) are operative. In this light, the previously mentioned correlations lend further support to the substantial role of the nearest-neighbor interactions in determining the overall three-dimensional structure of the native protein. If there were no long-range interactions, *e.g.*, if a portion of a protein molecule were cleaved from the rest of the chain, the protein presumably could not assume its native conformation; this was the explanation provided earlier[88] for the failure of three isolated portions of the myoglobin chain to assume the conformation which they have in the intact native protein.[183] One encounters numerous examples for this requirement that most of the protein chain be present in order that the molecule fold into its native conformation. For example, the S peptide and S protein of ribonuclease do not separately have the conformations which they do in the native ribonuclease molecule; however, when added together, the two fragments interact to stabilize a three-dimensional structure which resembles that of native ribonuclease.[184] Such a bimolecular reaction has a reasonable chance of leading to a properly stabilized structure. However, the extra translational and rotational entropy loss required in a trimolecular reaction makes it less likely that three fragments can associate to form a structure resembling the native protein; this is presumably the reason that the apomyoglobin structure could not be regenerated[183] by mixing the three fragments together. Another example of successful refolding involves the bimolecular reactions among fragments of a staphylococcal nuclease.[185] In experiments, in which refolding to the native structure accompanies the re-formation of disulfide bonds, ribonuclease[186] and pro-insulin,[187] for which the refolding process is unimolecular (with the whole chain being intact), fold properly. However, when a portion of the pro-insulin molecule is removed, it is more difficult for the separated A and B chains to form the insulin molecule in the required bimolecular reaction.[188]

In summary, the process of nucleation in a given *intact* protein chain requires the formation of α-helical (or possibly other types of structural) regions in specific parts of the amino acid sequence. These α-helical regions become stabilized only in the folding process involving long-range interactions between various parts of the chain. The removal of a significant part of a protein chain would prevent such long-range interactions from stabilizing the folded structure, and the remaining part of the protein chain could not acquire the conformation which it has in the intact native protein.

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

(133) R. M. Epand and H. A. Scheraga, *Biochemistry*, 7, 2864 (1968).
(134) F. M. Richards, *Proc. Nat. Acad. Sci. U. S.*, 44, 162 (1958).
(135) C. B. Anfinsen, Jr., private communication.
(136) C. J. Epstein, R. F. Goldberger, and C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.*, 28, 439 (1963).
(137) D. F. Steiner, *New England J. Med.*, 280, 1106 (1969).
(138) P. G. Katsoyannis and A. Tometsko, *Proc. Nat. Acad. Sci. U. S.*, 55, 1554 (1966).

It is hoped that the ideas presented in sections II.J and II.K will provide a basis for obtaining a rough overall conformation of a protein, to which the mathematical procedures mentioned in section II.A can then be applied for refining the structure without encountering the problem of multiple minima in the $F(Q)$ surface. It appears from the discussion in section II that further understanding of the factors which influence the folding of a polypeptide chain in a solvent or in a crystal is being acquired.

III. Experimental

A: INTRODUCTION

The work described in section III constitutes a necessary experimental counterpart to the theoretical approach discussed in section II. It is being carried out with synthetic polyamino acids for two purposes, *viz.*, (1) to provide experimental information which is required in the computations, and (2) to provide an experimental test of predictions made from the computations. While many experimental data on small molecules (*e.g.*, barriers to internal rotation, crystal structures, etc). form the basis for selection of the parameters of the empirical energy functions, we will not discuss them here since they are already in the literature and were discussed in previous reviews.[1-6] Instead, we will confine our discussion here to the recent use of synthetic homopolymers and copolymers of amino acids to obtain experimental values of the parameters σ and s for the various amino acids (which, themselves, depend on the empirical energy functions, as outlined in section II.F, and are useful for computations such as those described in section II.G) and to verify some of the predictions made in the computations about helix sense and side-chain conformation. Since the solvent (*i.e.*, water) plays such an important role in determining protein conformation, the experiments were confined, as much as possible, to water-soluble polymers.

In order to determine the parameters σ and s for all naturally occurring amino acids, it would be desirable to have water-soluble homopolymers of all such amino acids for studies of thermally induced helix–coil transitions. Unfortunately, while essentially all such homopolymers can be synthesized,[189] many of them are not water soluble or, if so, are not α helical or, if so, do not melt in an accessible temperature range. The experiments described herein were designed to circumvent these difficulties. Water solubility was achieved by incorporating the amino acid under study into block, regular-sequence, or random copolymers with a water-soluble carrier amino acid. Block copolymers were used for those polyamino acids which form α helices naturally and which can be melted; for those which do not, resort was had to the use of water-soluble, α-helical carrier polymers in which the desired amino acid was introduced as a second component at random. This device assured that the desired amino acid would be in the α-helical conformation in water and, by proper selection of the host polymer, that the resulting copolymer would melt in an accessible temperature range. The block copolymers could be treated by the Lifson–Roig[107] or Zimm–Bragg[106] theories to obtain v and w, or σ and s, respectively; the use of random copolymers in the manner indicated above depended on the recent success[114,115] in formulating theories of the helix–coil transition of a random copolymer. At the present time, the random copolymer seems to offer the best experimental approach to circumventing the difficulties cited above, and to provide values of σ and s in water for any desired amino acid.

B. BLOCK COPOLYMERS

Initial efforts to achieve water-soluble α-helical polymers resorted to block copolymers.[140-142] For example, a block of poly-L-alanine was incorporated between two blocks of poly-DL-glutamic acid[140] or poly-DL-lysine;[141] such copolymers are water soluble at neutral pH. By choosing D,L copolymers for the end blocks, and by adjusting the pH so that glutamic acid and lysine, respectively, are charged, the end blocks were nonhelical, and the thermally induced helix–coil transition of the central α-helical poly-L-alanine block could be followed by optical rotatory dispersion (ORD) measurements. From a study[92] of the melting of such copolymers in salt-free water, using samples with several chain lengths of the central poly-L-alanine block, it was found that short chains (*ca.* ten residues long) of poly-L-alanine are nonhelical, but longer chains form α helices whose melting is characterized by the parameters (assumed temperature-independent) for the transition of a residue from a coil to a helical state shown in Table III. The melting data for poly-L-alanine in salt-free water[92] were also used[47] to determine ΔH_B and ΔS_B, as described in section II.F. When salt was present, the polymers were more stable,[92] presumably because of interhelical hydrophobic bonding in the hairpin-like structures which can form when the electrostatic repulsion between the charged end-blocks is reduced by salt; it was felt that the hairpin-like structures do not exist in salt-free water.

Table III

Parameters for Transitition from the Coil to the Helical State for Several Nonpolar Homopolymers in Water[144]

Polyamino acid	v	ΔH, *cal/mol*	ΔS, *eu*
Polyglycine[a]	0.016	−488	−3.15
Poly-L-alanine[b]	0.012	−188	−0.55
Poly-L-leucine[c]	0.05 to 0.011	+100	+0.70 to 1.00
Poly-L-valine[a,d]	0.011	+212	+1.45

[a] Computed at 25°. [b] Experimental.[92] [c] Experimental.[144] [d] Forms a β structure in water but an α helix in 98% methanol;[143] hence these parameters are for a hypothetical α helix in water.

The existence of deca-L-alanine in the nonhelical form was already cited in section II.D as being compatible with the limited exploration of the conformational energy surface of this polymer in water, *viz.*, that extended, rather than α-helical, structures seem to have a higher statistical weight when the chain is so short.

Similar block copolymers were prepared with poly-L-valine,[143] poly-L-leucine,[142,144] and poly-L-phenylala-

(139) E. Katchalski, M. Sela, H. I. Silman, and A. Berger, in "The Proteins," Vol. II, 2nd ed, H. Neurath, Ed., Academic Press, New York, N. Y., 1964, p 405.

(140) W. B. Gratzer and P. Doty, *J. Amer. Chem. Soc.*, 85, 1193 (1963).
(141) N. Lotan, A. Berger, E. Katchalski, R. T. Ingwall, and H. A. Scheraga, *Biopolymers*, 4, 239 (1966).
(142) H. E. Auer and P. Doty, *Biochemistry*, 5, 1708, 1716 (1966).
(143) R. F. Epand and H. A. Scheraga, *Biopolymers*, 6, 1551 (1968).
(144) S. E. Ostroy, N. Lotan, R. T. Ingwall, and H. A. Scheraga, *ibid.*, 9, 749 (1970).

nine,[142,145] respectively, replacing poly-L-alanine; all of these polymers form α-helical structures under appropriate conditions. The poly-L-valine results are discussed in section III.E. The experimental parameters for the helix–coil transition in poly-L-leucine are shown in Table III, and the difference in the melting behavior of poly-L-leucine and poly-L-alanine is illustrated[144] in Figure 13. This difference arises primarily from nonbonded interactions (including hydrophobic bonding) between the nonpolar side chain and neighboring backbone (and side-chain) atoms in the helical and randomly coiled forms. No data, comparable to those in Table III, were reported for poly-L-phenylalanine.[142,145]

Using the experimental data for poly-L-alanine[92] and theoretical values[146] of the thermodynamic parameters for hydrophobic bond formation, the parameters for the formation of the hypothetical α helix of polyglycine (shown in Table III) were computed. Using this contribution for the backbone (*i.e.*, polyglycine) and theoretical values[146] for the valyl side chain, the data in Table III for poly-L-valine were computed. Also, the experimental data in Table III for poly-L-leucine are compatible[144] with theoretical values computed from these parameters for polyglycine and from theoretical values for the contribution from the leucyl side chain. Table III illustrates how the increased size of the nonpolar side chain (with a larger contribution from hydrophobic bonding) makes a larger contribution to the stability of the α helix; this stability is reflected, among other ways, by a higher helix content at a given temperature for chains of comparable length.

While the data of Table III were obtained with block copolymers having charged end blocks which may affect the computed thermodynamic parameters, similar results were obtained for glycine and alanine from melting of neutral random copolymers[147] (see section III.D) and for alanine from titration data for a neutral random copolymer.[148]

The conclusions of Table III are in agreement with those of Gō, *et al.*,[47] that polyglycine is in the random coil form in water in the temperature range of 0–100°, but that poly-L-alanine is helical and melts in this range. Quantitative comparison is difficult because v, ΔH, and ΔS were assumed to be temperature independent in the calculations of Table III but not in those of Gō, *et al.*;[47] also, there may be small differences between the parameters of Table III (computed for the Lifson–Roig[107] quantities v and w) and those of Gō, *et al.*[47] (computed for the Zimm–Bragg[106] quantities σ and s).

C. REGULAR-SEQUENCE COPOLYMERS

While a special treatment[114,115] is required for a copolymer of, say, A and B units if the latter are distributed at random over the chain, the Lifson–Roig[107] and Zimm–Bragg[106] theories for homopolymers can be applied to regular-sequence copolymers, $(A_mB_n)_x$, since the A_mB_n unit in the chain can be regarded as a "monomer" and, of course, the amino acid sequence in such a chain is known. As indicated in section III.A, the use of a regular-sequence copolymer is one possible approach toward circumventing the aforementioned difficulties in obtaining σ and s from experiments on homopolymers. As an initial

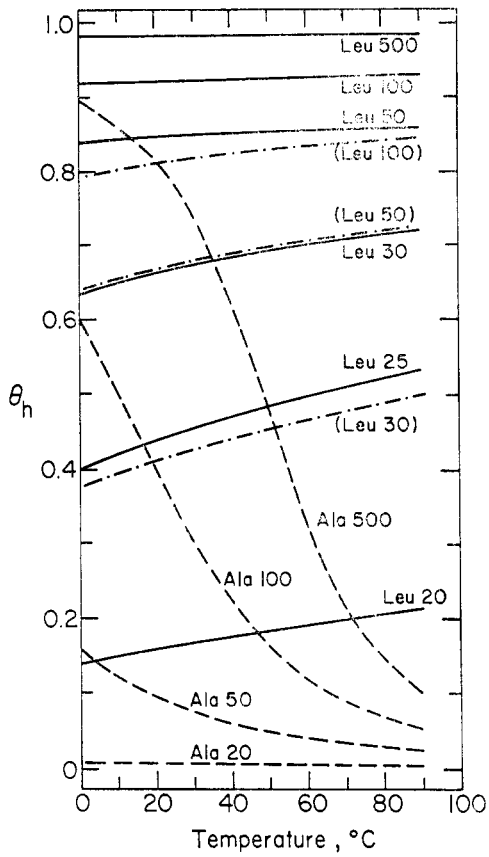

Figure 13. Theoretical curves[144] for helix content, θ_H, *vs.* T (which are based on matching of experimental data[92,144]) for poly-L-alanine and poly-L-leucine of various DP's. Poly-L-alanine: $\Delta H = -188$ cal/mol, $\Delta S = -0.55$ eu, $v = 0.012$. Poly-L-leucine: $\Delta H = +100$ cal/mol, $\Delta S = +1.00$ eu, $v = 0.013$ (———); $\Delta H = +100$ cal/mol, $\Delta S = +0.70$ eu, $v = 0.04$ (—·—·—).

study,[149] the regular-sequence copolymer poly(Leu-Leu-Lys)$_x$ was examined. For comparison, the analogous random-sequence copolymers poly(Leu_2Lys_1)$_x$, poly(Leu_1Lys_1)$_x$, and poly(Leu_1Lys_2)$_x$ were also studied. All of these polymers are water soluble.

It was found that, whereas poly-L-lysine is in the helical form only above a pH range in which most of the ϵ-amino groups are neutralized, this pH range is lowered as the leucine content of the copolymer increases. Since the average distance between the lysine charges is increased by the intervening leucine residues, the electrostatic repulsion, which favors the coil form, is reduced as the leucine content increases. Similarly, whereas charged poly-L-lysine[150,151] assumes the helical form when methanol is added, to the extent of ~90%, to an aqueous solution, the concentration of alcohol required to convert the copolymers to the helical form decreases as the leucine content increases. Finally, while all of these polymers were observed to melt, the data had not yet been analyzed (at the time this article was written) by the Lifson–Roig[107] or Zimm–Bragg[106] procedure for the regular-sequence copolymer, or by the Lehman–McTague[114] or Poland–Scheraga[115] procedure

(145) H. J. Sage and G. D. Fasman, *Biochemistry*, **5**, 286 (1966).
(146) G. Nemethy and H. A. Scheraga, *J. Phys. Chem.*, **66**, 1773 (1962).
(147) V. S. Ananthanarayanan, R. H. Andreatta, D. Poland, K. Platzer, and H. A. Scheraga, unpublished work.
(148) H. Sugiyama and H. Noda, *Biopolymers*, **9**, 459 (1970).
(149) J. L. Olpin, N. Lotan, R. H. Andreatta, and J. Alter, unpublished work.
(150) R. F. Epand and H. A. Scheraga, *Biopolymers*, **6**, 1383 (1968).
(151) F. J. Joubert, N. Lotan, and H. A. Scheraga, *Physiol. Chem, Phys.*, **1**, 348 (1969).

for the random-sequence copolymers. The melting behavior[149] is different from that shown in Figure 13 for poly-L-leucine, for which the helix content increases slightly with increasing temperature. In the case of the regular- and random-sequence copolymers, the strong stabilizing effect of leucine in poly-L-leucine is "diluted out" by the intervening lysine residues, so that the helix content decreases with increasing temperature.

It is of interest that the regular-sequence poly(Leu-Leu-Lys)$_x$ and the random-sequence poly(Leu$_2$Lys)$_x$ behave similarly in all of the properties cited in the previous paragraph. This is in accord with the expected behavior[114,115] for copolymers of two amino acids in which there is no tendency for a given type of amino acid to accumulate in *large* sequences. From this point of view, the random-sequence copolymer (see section III.D) can provide similar information (with less effort expended on the synthesis) to that obtained from the regular-sequence copolymer (of two amino acids).

Yaron[152] has studied a similar regular-sequence copolymer, in which alanine replaced leucine, and obtained similar results, the only differences being those attributable to the difference between alanine and leucine.

D. RANDOM COPOLYMERS

If one has two water-soluble, α-helical homopolymers which melt in an accessible temperature range, then one can compute the melting curve for all compositions of random copolymers of these two amino acids, using the parameters for the helix–coil transition of the homopolymers,[114,115] and compare the results with experimental data on the copolymers. Alternatively, if a desired homopolymer of, say, A units does not satisfy any of the requirements of water solubility, α helicity, or proper melting range, then the A units can be incorporated into a random copolymer with, say, B units if the latter do satisfy these requirements. Suitable copolymers formed from the "host" B units and "guest" A units can have increasing amounts of A, as long as the resulting random copolymers still satisfy the requirements mentioned above. The melting data for the copolymers and for the homopolymer of B units then give σ and s for the A units. In this section, we report two types of experimental results: (1) those in which both homopolymers (and the copolymers of all intermediate compositions) satisfy the three requirements, and (2) those in which only the "host" homopolymer satisfies the three requirements; in the latter case, the values of σ and s for the hypothetical "guest" homopolymer can be determined[114,115] from those of the "host" homopolymer and from the influence of the "guest" residues on the melting behavior of the random copolymers.

As suitable "host" homopolymers, use[153] has been made of the hydroxypropyl- and hydroxybutylamine derivatives of poly-L-glutamic acid, PHPG and PHBG, respectively.[154–156] These homopolymers satisfy the three requirements, PHBG having a higher melting temperature than PHPG in water.

Von Dreele, *et al.*,[153] studied random copolymers of PHPG and PHBG covering the whole composition range between that of the two homopolymers. The melting curves of the copolymers, computed from the values of σ and s for the homopolymers, were found to be in good agreement with the experimentally determined curves. These experiments provide confidence in the application of the theory[114,115] to determine σ and s for those amino acids which do not form water-soluble α-helical structures, or do not melt between 0 and 100°.

In using PHPG and PHBG as a "host," the selection is made on the basis of whether the "guest" residue is expected to be a helix-breaker or helix-maker. PHPG is used for helix-makers, since the homopolymer PHPG melts at low temperature and the "guest" residues bring the melting point into the intermediate temperature range between 0 and 100°. On the other hand, since PHBG melts higher, it is used with "guest" residues which are helix-breakers; the resulting melting curves appear at lower temperature than that for the homopolymer PHBG, but still above 0°.

The first amino acid, to which the "host-guest" technique was applied, was glycine,[147] which does not form an α-helix as a homopolymer. Some melting curves of copolymers of PHBG and glycine are shown in Figure 14. It can be seen that the melting point of the α-helical homopolymer PHBG is depressed by the addition of a very small amount of glycine. Even qualitatively, this implies that glycine is a helix breaker with a low value of s at room temperature, as also is implied by the data of Table III for polyglycine, and as was deduced from the calculations reported in section II.F.

For a study of alanine, PHPG was used as the "host" polymer.[147] Larger amounts of alanine (up to 50%) than glycine were incorporated. The presence of increasing amounts of alanine raised the melting point of PHPG, implying that alanine is a helix-former. However, from the fact that large amounts of alanine were required to raise the melting point, we see that alanine is not a *very strong* helix-former. A preliminary analysis of the melting data for these copolymers indicates that the parameters agree with those of Table III, based on the melting of a block of poly-L-alanine.

Whereas leucine and valine do not melt as homopolymers (*e.g.*, see Figure 13), the incorporation of these residues into PHPG yields α-helical polymers which do melt.[157] Preliminary data for valine (for copolymers containing about 3 to 10% valine) indicate that it is a stronger helix-former than alanine.

Poly-L-serine forms a β structure.[7] However, when incorporated (to the extent of up to 15%) in PHBG, the copolymers are α-helical and melt at lower temperatures than PHBG, indicating that serine is a helix-breaker.[158] Similar experiments are in progress with tyrosine,[157] lysine,[159] aspartic acid,[157] and glutamic acid[157] as "guest" residues, and experiments with the remaining naturally occurring amino acids are planned. In the case of copolymers containing ionizable "guest" residues, *e.g.*, lysine, the values of σ and s can be determined for both the charged and uncharged forms. Further, from the titration behavior of such polymers in water, we should be able to obtain more detailed information about the electrostatic potential (and the effect of salt thereon) in water than was obtained from a study of poly-L-lysine (see section III.H).

An important advantage in using PHPG and PHBG is that they are soluble not only in water but also in organic solvents. Hence, the melting behavior of the copolymers, both in water

(152) A. Yaron, private communication.
(153) P. H. Von Dreele, N. Lotan, D. Poland, V. S. Ananthanarayanan, R. H. Andreatta, and H. A. Scheraga, unpublished work.
(154) N. Lotan, A. Yaron, A. Berger, and M. Sela, *Biopolymers*, 3, 625 (1965).
(155) N. Lotan, Ph.D. Thesis, Weizmann Institute of Science, Rehovoth, Israel, 1966.
(156) N. Lotan, M. Bixon, and A. Berger, *Biopolymers*, 8, 247 (1969).
(157) J. Alter, R. H. Andreatta, and H. A. Scheraga, unpublished work.
(158) L. J. Hughes, R. H. Andreatta, and H. A. Scheraga, unpublished work.
(159) M. Dygert, R. H. Andreatta, and H. A. Scheraga, unpublished work.

and in organic solvents, should provide information about the effect of a change of solvent (from water to an organic one) on σ and s for the "guest" residues.

Since the experiments described above can be used with any amino acid as the "guest" units, the values of σ and s are being determined in this laboratory for all naturally occurring amino acids at, say, 25°. Thus, we should ultimately have available a quantitative measure of whether a given amino acid enhances helix formation, disrupts it, or behaves indifferently in this respect at the given temperature. Further, these parameters will then be used, in the manner discussed in sections II.G and II.J, to identify the nucleation centers in proteins, this approach being predicated on the assumption that σ and s for any amino acid are determined mainly by near-neighbor interactions.

E. VERIFICATION OF PREDICTED HELIX SENSES

In section II.D, the calculation of the helix senses of a large number of homopolyamino acids was discussed. In most cases, the experimental results were known and the results of the calculations agreed with experiment; however, the calculations provided an understanding of the factors which influenced the helix sense. On the other hand, in some cases, the experiments had not yet been carried out, and the calculations therefore served as a prediction. Hence, we cite here the experimental evidence which provides a verification of the predicted helix senses for these cases.

The predicted existence of the right-handed α-helical form of poly-L-valine[78] was verified by incorporating poly-L-valine into a block copolymer, between two blocks of poly-DL-lysine,[79] in the manner described in section III.B. Using optical rotatory dispersion (ORD) and circular dichroism (CD) data, about 50% of the short valine block of (DL-lysine hydrochloride)$_{18}$-(L-valine)$_{15}$-(DL-lysine hydrochloride)$_{16}$-glycine was found to be in the right-handed α-helical conformation in 98% aqueous methanol. Thus, as predicted, there is no steric hindrance preventing the formation of the α helix in this polyamino acid. Even more so than poly-L-leucine in water (see Figure 13), the α-helical form of poly-L-valine is thermally stable in 98% aqueous methanol.

In the case of poly-γ-*p*-chlorobenzyl-L-glutamate, the calculated[80] difference in energy between the left- and right-handed α-helical forms was quite small (and within the error of the calculations), with the energy for the right-handed helix being lower. Thus, while subsequent ORD measurements[90] on this polymer indicated that it does form a right-handed α helix in all the solvents studied, the verification of the predicted helix sense in this case is not, in itself, an establishment of the validity of the calculations; however, at least the calculated and experimental results are consistent with each other.

The predicted senses of the α-helical forms of the *o*- and *m*-chlorobenzyl esters of poly-L-aspartic acid,[49] discussed in section II.D, were checked by ORD and CD measurements.[91] From Figures 15 and 16 it can be seen that these polymers form left-handed α helices, as predicted, in contrast to the para isomer which forms a right-handed α helix; the same conclusion was reached on the basis of b_0 data.[91] The data of Figures 15 and 16 for the para isomer are in accord both with earlier experimental[83-86] and theoretical[80] results on this polymer.

The examples cited in this section provide confidence in the procedures (described in section II) used to calculate preferred conformations of homopolyamino acids.

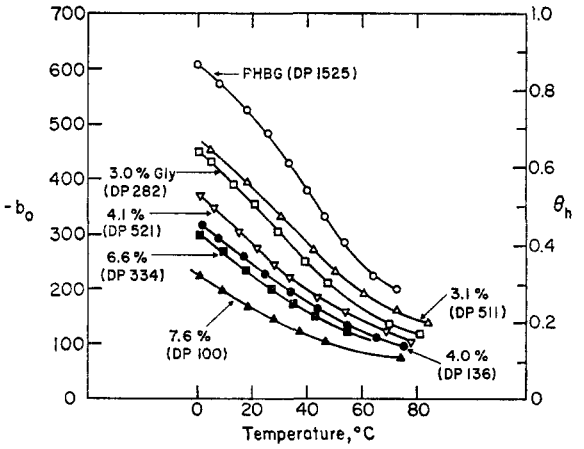

Figure 14. Temperature dependence of b_0 in aqueous solutions of PHBG and copolymers of HBG and glycine, for the compositions and DP's indicated.[147] The average fraction of helical units ($= -b_0/700$) for the copolymers is given on the right-hand ordinate.

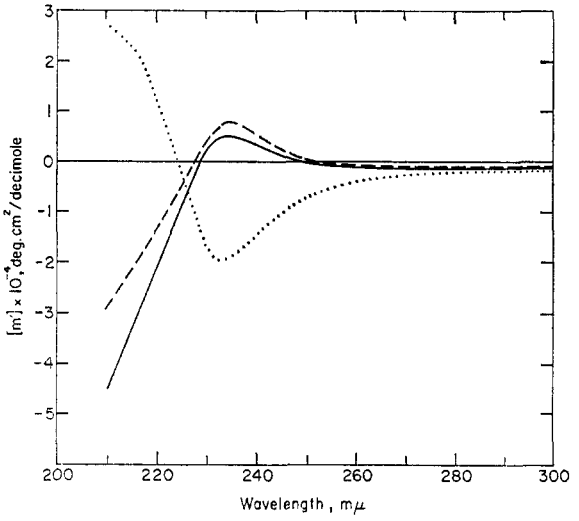

Figure 15. ORD of chloro-substituted poly-β-benzyl-L-aspartates in dioxane:[91] (———) ortho derivative at 40°, (– – –) meta derivative at 25°, (· · · · ·) para derivative at 25°.

F. ORIENTATION AND MOBILITY OF SIDE CHAINS

The procedures described in section II provide both the energy and conformation of the most stable structure of a polypeptide. In the case of α-helical homopolymers, the calculations yield information, for example, about the orientation of the side chains with respect to the helical backbone. It is, therefore, of interest to consider experiments designed to determine whether the side chains are relatively fixed or mobile and, if fixed, what the preferred orientation is. In making this comparison, it should be kept in mind that, so far, the influence of water has been included in the calculations only for polyglycine and poly-L-alanine (see section II.F); the calculations for the other homopolymers apply to a vacuum environment, or possibly to an organic solvent. The techniques that are useful for these experiments include measurements of dipole mo-

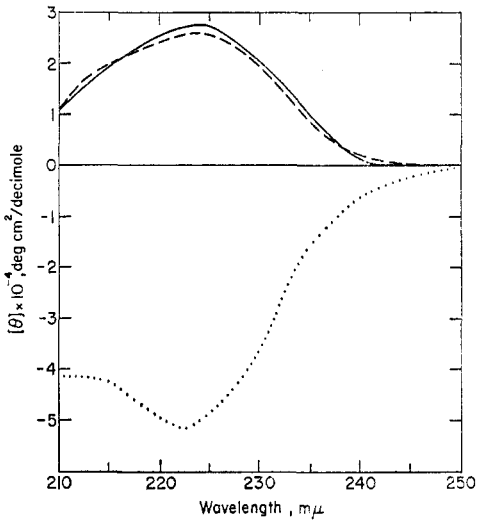

Figure 16. CD of chloro-substituted poly-β-benzyl-L-aspartates in dioxane[91] at 25°: (——) ortho derivative, (– – –) meta derivative, (• • • • •) para derivative.

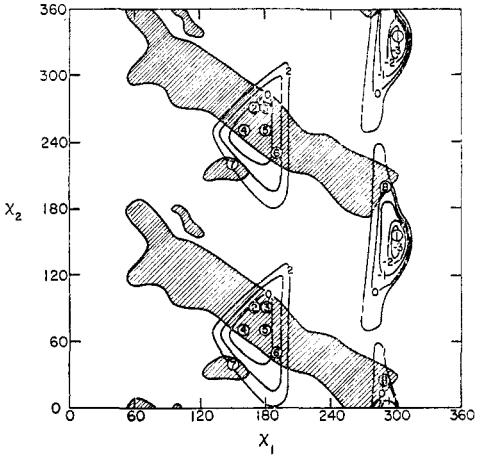

Figure 17. Energy contour map for the right-handed α helix (ϕ = 132°, ψ = 123°) of poly-L-phenylalanine.[164] Energies are in kcal/mol residue. The shaded areas contain those conformations for which the computed nmr shifts have the same sign as the observed ones for all of the side-chain protons.[164] See original paper for further discussion of this figure

ments, electric dichroism (ED), circular dichroism (CD), and nuclear magnetic resonance (nmr).

If the side chain has a dipole moment, and if it has a preferred orientation with respect to the helical backbone whose amide groups also have a dipole moment, then the net dipole moment depends on the orientation of the side chains. While dipole moment measurements have been carried out for poly-L-tyrosine[160] and for several other polymers,[161] it is still too early to assess the conformational information which can be obtained from this technique; however, if some of the problems involved in the interpretation of dielectric constant measurements in solutions can be overcome, and if the side chain does have a preferred orientation, this method should be useful for giving the side-chain orientations.

Information about the permanent dipole moment and about the orientation, mobility, and electric properties of the side chains of helical polymers can also be obtained from ED measurements.[162] Preliminary measurements on helical poly-L-tyrosine in dioxane solution[162] yielded the dipole moment and an orientation factor, the interpretation of which in terms of mobility and conformational information requires further work, which is currently being carried on.

The nature and orientation of the side chain influence the rotational strength of, say, the n,π^* transition,[82] which is observed in CD and ORD experiments on polyamino acids. Chen and Woody[81] have carried out calculations of rotational strengths for poly-L-tyrosine and have deduced the screw sense of the helical backbone and the preferred orientation of the side chain.

Nmr measurements should prove useful in obtaining information about the mobility of the side chains of helical polymers. In the case of PHPG, PHBG, and the homologous ethyl derivative PHEG, variations in line width were used to distinguish among the mobilities of the different protons of the side chains.[163] For those helical polyamino acids with an aromatic side chain, such as poly-L-phenylalanine, it is possible to compute the ring current effect from neighboring aromatic residues for several conformations of the side chains and compare it with experimental values of chemical shifts, enabling deductions to be made about the orientations of the side chains.[164] Since a large range of possible side-chain conformations is consistent with the observed chemical shifts, conformational energy calculations offer the possibility of narrowing the possible range of conformations for the side chains. For example, computed energy contours are superimposed on the nmr data for poly-L-phenylalanine in Figure 17.[164] Presumably, the side chain adopts a conformation which lies both in the shaded region and in a low-energy region. Gibbons, *et al.*,[165] have proposed a similar combination of nmr data with conformational energy calculations to obtain information on polypeptides in solution.

G. CALORIMETRIC DATA ON HELIX–COIL TRANSITIONS

The enthalpy change $\Delta H°$ accompanying the conversion of a random-coil residue to a hydrogen-bonded helical one at the end of a long helical sequence determines the temperature dependence of the Zimm–Bragg parameter s, according to the van't Hoff relation. The calculation of this enthalpy change for polyglycine and poly-L-alanine in water was discussed in section II.F. Experimentally, $\Delta H°$ is determinable from melting curves of the homopolymer (requiring also a knowledge of σ)[106] or by a direct calorimetric measurement. It will be very important to have such calorimetric values, not only for a comparison with values determined from melting curves, but also to provide a direct experimental check of the kind of calculations described in section II.F. A variety of calorimetric data are accumulating,[166] the only data for polymers in water being $\Delta H°$ = −1100 for neutral poly-L-glutamic acid[166c] and

(160) J. Applequist and T. G. Mahr, *J. Amer. Chem. Soc.*, **88**, 5419 (1966).
(161) E. H. Erenrich and H. A. Scheraga, unpublished work.
(162) T. C. Troxell and H. A. Scheraga, *Biochem. Biophys. Res. Commun.*, 35, 913 (1969).
(163) F. J. Joubert, N. Lotan, and H. A. Scheraga, *Biochemistry*, **9**, 2197 (1970).
(164) D. N. Silverman and H. A. Scheraga, *ibid.*, in press.
(165) W. A. Gibbons, G. Nemethy, A. Stern, and L. C. Craig, *Proc. Nat. Acad. Sci. U. S.*, 67, 239 (1970).

$\Delta H^\circ = -1200$ for neutral poly-L-lysine.[166d] These experimental values will provide a useful comparison when the calculations described in section II.F are extended to these two polyamino acids.

H. ELECTROSTATIC EFFECTS

The importance of electrostatic effects in influencing the conformations of polypeptides has long been recognized. For example, poly-L-lysine is α-helical in water when the ϵ-amino groups are neutral, but randomly coiled when the amino groups acquire a charge.[7] Even when the interacting groups are neutral, but polar, electrostatic effects (describable in terms of partial charges on atoms of neutral groups) play an important role, *e.g.*, in the dipole–dipole interactions in poly-β-benzyl-L-aspartate and in its chloro derivatives, cited in section II.D. Numerous other similar examples could be mentioned. It is thus of importance to know how to compute the electrostatic interactions in conformational energy calculations. This is a difficult problem, discussed to some extent by Gō, *et al.*,[18] for example. In this section, however, we consider some experimental aspects of the problem.

Recently Liem, *et al.*,[167] attempted to assess the range of the electrostatic potential in poly-L-lysine, which is α helical over the whole pH range in 95% aqueous methanol[150] even when the ϵ-amino groups are charged.[151] Since the conformation does not change as the pH does, in this solvent, one has a known geometric array of charges at a given pH, and the difference between the titration curves of poly-L-lysine and the model compound *n*-butylamine is attributable to the electrostatic free energy of the charges on the polymer. A statistical mechanical treatment, using a matrix method, of the system showed how the theoretical titration curves approach the experimental one for poly-L-lysine as longer and longer range charge–charge interactions are taken into account. Thus, truncation of the electrostatic interactions at too short range can introduce an error in the electrostatic free energy. While the Debye–Hückel potential did not give satisfactory results, an empirically selected potential did.

While such an empirical potential would be applicable to conformational energy calculations in 95% methanol, it is desired to have one for use in water. For this purpose, the random copolymers of PHPG or PHBG with lysine, aspartic acid, and glutamic acid, respectively, mentioned in section III.D, should be useful, since they should be α helical over the whole pH range (at least for low concentrations of the "guest" residues) and their titration curves will yield the electrostatic free energy directly. As the lysine content of the copolymer increases, the titration curve should shift from that resembling *n*-butylamine toward that of a hypothetical α-helical poly-L-lysine in water. Such experiments are in progress, and it is hoped that they will provide a basis for selecting a suitable electrostatic potential for use in conformational energy calculations on polypeptides in water.

NOTE ADDED IN PROOF. Additional helix-probability profiles, of the type shown in Figure 12, have been computed for a number of proteins.[168] For example, it has been found that, despite variations in the amino acid sequence of cytochrome c proteins from a large number of aerobic organisms, the helix-probability profiles of these proteins are very similar; this suggests that they all have a similar three-dimensional structure, which presumably is retained throughout the evolution of these proteins. In addition, the helix-probability profiles[168] of lysozyme and α-lactalbumin are very similar to each other, despite many differences in amino acid sequence, supporting the view[169,170] that these two proteins have essentially the same three-dimensional structure. These considerations are being applied to a number of other proteins, for which amino acid sequence information is available, *e.g.*, γ-globulin and glutamate dehydrogenase.

Acknowledgment. The work described here was supported by the National Institutes of Health and by the National Science Foundation. This article was written in the spring of 1970 during the tenure of a National Institutes of Health Special Fellowship at the Weizmann Institute of Science, Rehovoth, Israel. The author gratefully acknowledges the hospitality of Professor Ephraim Katchalski during the period of this fellowship.

(166) See, for example, (a) F. E. Karasz, J. M. O'Reilly, and H. E. Bair, *Biopolymers*, 3, 241 (1965); (b) T. Ackermann and E. Neumann, *ibid.*, 5, 649 (1967); (c) G. Rialdi and J. Hermans, Jr., *J. Amer. Chem. Soc.*, 88, 5719 (1966); (d) P. Y. Chou and H. A. Scheraga, *Biopolymers*, in press.

(167) R. K. H. Liem, D. Poland, and H. A. Scheraga, *J. Amer. Chem. Soc.*, 92, 5717 (1970).

(168) P. N. Lewis and H. A. Scheraga, unpublished results.

(169) K. Brew and P. N. Campbell, *Biochem. J.*, 102, 258 (1967).

(170) W. J. Browne, A. C. T. North, D. C. Phillips, K. Brew, T. C. Vanaman, and R. L. Hill, *J. Mol. Biol.*, 42, 65 (1969).