Calculation of the Conformation of the Pentapeptide Cyclo(glycylglycylglycylprolylprolyl). II. Statistical Weights

ARTICLE *in* MACROMOLECULES · DECEMBER 2003

Impact Factor: 5.8 · DOI: 10.1021/ma60017a603

3 AUTHORS, INCLUDING:

Nobuhiro Go
Japan Atomic Energy Agency
238 PUBLICATIONS **9,966** CITATIONS
SEE PROFILE

Peter N. Lewis
University of Toronto
61 PUBLICATIONS **2,107** CITATIONS
SEE PROFILE

Calculation of the Conformation of the Pentapeptide Cyclo(glycylglycylglycylprolylprolyl). II. Statistical Weights[1]

Nobuhiro Gō,[2] Peter N. Lewis, and Harold A. Scheraga[3]

Department of Chemistry, Cornell University, Ithaca, New York 14850.
Received June 17, 1970

ABSTRACT: The contributions to the statistical weight from librational degrees of freedom are calculated for conformations corresponding to local minima in the complete conformational energy surface of this cyclic molecule. The calculation, based on the small-fluctuation approximation, involves the computation of $-(R/2)$ ln det **G**, the elements of the matrix **G** being the coefficients of the kinetic energy term in the canonical expression for the Hamiltonian, and $-(R/2)$ ln det **F**, the elements of the matrix **F** being the second derivatives of the conformational energy function at the minimum point. The dependence of the values of $-(R/2)$ ln det **G** on various minimum-energy conformations is found to be large enough (the largest difference being 4.53 eu) to change the relative stabilities of some of the minimum-energy conformations when this entropy term is included. The two lowest minimum-energy conformations are found to have almost the same stability for one energy parameter set used, when the conformational entropy, $-(R/2)$ ln det **GF**, is included in the free energy. The validity of the small-fluctuation approximation is checked by computing the exact energy surface around the minimum points. The approximation is found to be a good one for the two most stable conformations. Effects of crystallization on the conformation of the individual molecule are discussed.

I. Introduction

In the first paper[4] of this series, the complete energy surface was calculated for the pentapeptide cyclo-(Gly-Gly-Gly-Pro-Pro), with fixed bond lengths, bond angles, and planar *trans*-amide groups. Thirteen local minima were found and located in that region of the energy surface which is within 100 kcal of the global minimum. In order to determine the relative stabilities of the molecular conformations corresponding to these minima, it is necessary to compare their free energies (or statistical weights), rather than their energies. A theoretical analysis of the contribution of internal vibrations to the statistical weights of equilibrium conformations of macromolecules has been made,[5] and one of the reasons for carrying out the calculation of the conformation of the cyclic pentapeptide treated in this series of papers is that this molecule is particularly suitable for applying the procedures given there[5] to determine the most stable conformation.

For conformations with small conformational fluctuations, the free energies, F, are given[5] by

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

where $F_1(Q_1)$ is the value of the conformational energy at a minimum point $Q = Q_1$, 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 derivatives of the conformational energy function $F_1(Q)$ at the minimum point. Since constant terms, which are independent of conformation, have been omitted from eq 1, this equation cannot be used to obtain absolute values of F but only to discuss differences in free energy for various conformations. From an examination of a few systems, the contributions of det **F** to the relative free energies have been found to be of the order of 1 kcal/residue at room temperature.[6,7] The calculation of det **G** has been carried out thus far only for helical conformations of a polyethylene-type molecule,[5] and it was found that the contribution from det **G** is about of the same order as that from det **F**. Therefore, in order to determine the most stable conformation of a molecule, it is sufficient to calculate the conformational entropy, $-(R/2)$ ln (det **GF**), only for local-minimum conformations within about 1 kcal/residue of the global minimum. For the cyclic pentapeptide under consideration here, three out of the thirteen local minima (C′, D′ and E′ in Table II of I) have such low conformational energies, the energies for all of the other ten conformations being much higher. Hence, det **F** was calculated here only for conformations C′, D′, and E′. However, det **G** was calculated for all thirteen minimum-energy conformations found in I, in order to obtain further information about the conformational dependence of det **G**, since our previous knowledge of this subject was based on only one example.[5] The calculations of det **GF** and a discussion of the relative stabilities of conformations C′, D′, and E′ (the first purpose of this paper) are described in section II.

Equation 1 was derived[5] by assuming that the molecule undergoes small conformational fluctuations about the minimum-energy conformation. One of the advantages of being able to calculate the *complete* energy surface[4] for the cyclic pentapeptide under dis-

(1) This work was supported by research grants from the National Science Foundation (GB-17388 and GB-7571X), from the National Institute of General Medical Sciences of the National Institutes of Health, U. S. Public Health Service (GM-14312), from the Eli Lilly, Hoffmann-La Roche, and Smith Kline and French Grants Committees, and from Walter and George Todd.

(2) On leave of absence from the Department of Physics, Faculty of Science, University of Tokyo, Tokyo, Japan, 1967–1970.

(3) To whom requests for reprints should be addressed.

(4) N. Gō and H. A. Scheraga, *Macromolecules*, **3**, 188 (1970). Hereafter, this paper will be referred to as I.

(5) N. Gō and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969).

(6) K. D. Gibson and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **63**, 242 (1969).

(7) M. Gō, N. Gō, and H. A. Scheraga, *J. Chem. Phys.*, **52**, 2060 (1970).

cussion here is that we can check the validity of the assumption of small conformational fluctuations by calculating the energy surface in the vicinities of the local minima. This is carried out (the second purpose of this paper) and discussed in section III. Possible effects of crystallization on the conformation of the molecule are also discussed.

II. Relative Stabilities of Minimum-Energy Conformations

The independent variables for the calculation of det **GF** were selected on the basis of the following considerations. Since this cyclic pentapeptide contains 49 atoms, there are 141 *independent* internal variables when the molecule is treated as a flexible one. Using the definition of "soft" and "hard" variables of an earlier paper,[5] only two of the eight dihedral angles (soft variables) in the backbone are independent because of the condition of ring closure,[8] the other six being functions of the two independent dihedral angles; we choose ψ_4 and ψ_5 (see Figure 1 of I) as the two independent soft variables. The remaining 139 independent internal variables are hard ones; these are chosen as follows: all bond lengths (51 variables), all bond angles except three $\tau[HC^{\alpha}H]$'s, two $\tau[C^{\delta}NC^{\alpha}]$'s, two $\tau[HC^{\alpha}C^{\beta}]$'s, two $\tau[HC^{\beta}H]$'s, two $\tau[HC^{\gamma}H]$'s, two $\tau[HC^{\delta}H]$'s, five $\tau[C^{\alpha}C'N]$'s and three $\tau[C'NC^{\alpha}]$'s (75 variables), the dihedral angles $\omega[O_5C_5'N_1H_1]$, $\omega[O_1C_1'N_2H_2]$, $\omega[O_2C_2'N_3H_3]$, $\omega[O_3C_3'N_4C_4^{\alpha}]$, and $\omega[O_4C_4'N_5C_5^{\alpha}]$ around the partial double bonds (five variables), and five angles between a C'O bond and a plane defined by the two bonds $C^{\alpha}C'$ and C'N, and three angles between an NH bond and a plane defined by the two bonds C'N and NC^{α} (eight variables); see Figure 1 of I for the labeling of the atoms and other nomenclature.

The method outlined elsewhere[5] was used for the calculation of det **G**. In that method, 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 5; for the cyclic pentapeptide, the **G** matrix is 141 × 141. The elements of the **G** matrix were obtained, with the aid of a computer program, by calculating Wilson's **s** vectors for the given cartesian coordinates and connectivity (chemical bonds) of the atoms. Then the condition of rigidity was introduced (corresponding to the freezing of bond lengths and bond angles) by using eq 29 of ref 5 to obtain a 2 × 2 **G** matrix whose elements are functions of the hard variables; the computation of these elements required the inversion (performed numerically) of a 139 × 139 submatrix (corresponding to the 139 hard variables) of the original 141 × 141 matrix, and the multiplication by 2 × 139 and 139 × 2 submatrices.

The calculated values of $-(R/2)$ ln det **G** for the rigid molecule are given in Table I. It can be seen that there is no correlation between the values of E_{tot} and $-(R/2)$ ln det **G**. The largest difference in the values of $-(R/2)$ ln det **G** is 4.53 eu (between the values for conformations B and H), which is equivalent to 1.36 kcal at 300°K. This difference is not negligibly small compared with the differences in some of the conformational energies. Actually, the order of the relative stabilities of the minimum-energy conformations D and C changes when the term $-(RT/2)$ ln det **G** is included. In order to determine the real relative stabilities, another term, $-(RT/2)$ ln det **F**, must be added. However, the above point clearly indicates the importance of considering the term det **G**. The range of the change in the values of $-(R/2)$ ln det **G** observed in this cyclic pentapeptide, 4.53 eu/molecule or 0.9 eu/residue, is about the same as the one observed for helical structures of polyethylene-type molecules.[5]

TABLE I
VALUES OF $(R/2)$ Ln Det **G** AND RELATED QUANTITIES FOR MINIMUM-ENERGY CONFORMATIONS OF THE RIGID CYCLIC PENTAPEPTIDE

Conformation[a]	E_{tot}[b]	$-(R/2)$ ln det **G**[c]	$E_{tot} + (RT/2)$ ln det **G**[d]
A	42.52	12.86	38.87
A′	33.40	15.18	29.09
H	27.40	15.81	22.92
E	23.40	11.55	20.12
F	21.90	13.29	18.13
G	20.82	11.91	17.44
B	20.41	11.28	17.21
B′	19.80	13.51	15.97
D	19.08	14.11	15.08
C	18.93	11.44	15.68
C′	2.09	13.48	−1.73
E′	1.35	12.08	−2.08
D′	1.27	12.10	−2.16

[a] For a further characterization of the minimum-energy conformations, see Table II of I. [b] Total energies, E_{tot}, are taken from Table II of I and are given in kcal/molecule. [c] In entropy units/molecule. [d] A part of the quantity on the right-hand side of eq 1, given in kcal/molecule and calculated for T = 300°K. Only the differences of the values for different minima have meaning.

The dihedral angles ψ_4 and ψ_5 were chosen as the independent variables in the calculation of the energies of the rigid cyclic pentapeptide.[4] For the calculation of det **F**, the conformational energies were calculated at grid points of a square lattice in the space of ψ_4 and ψ_5 in the vicinity of the three lowest minima C′, D′, and E′. An interpolation formula[9] was used to fit the calculated energy values by a quartic function of ψ_4 and ψ_5 at each minimum point. The second–fourth-order derivatives of the energy surface were computed from these quartic functions. The use of the fourth-order derivatives will be discussed in section III. The values of $-(R/2)$ det **F**, calculated from the second derivatives, and the free energies of the minimum-energy conformations are given in Table II. When the conformational entropy is included, the difference in energies of the lowest minimum D′ and the next lowest minimum E′ is reduced from 0.08 kcal (for the conformational energies) to 0.02 kcal (for the conformational free energies). This value is very small compared with RT = 0.6 kcal for T = 300°K, while the difference 0.9 kcal in the free energies of minima C′ and D′ is slightly larger than 0.6 kcal. This means that isolated molecules of cyclo(Gly-Gly-Gly-Pro-Pro) are a

(8) N. Gō and H. A. Scheraga, *Macromolecules*, **3**, 178 (1970).

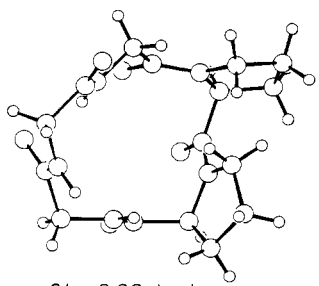
(9) J. B. Scarborough, "Numerical Mathematical Analysis," Johns Hopkins Press, Baltimore, Md., 1966, Chapter 7.

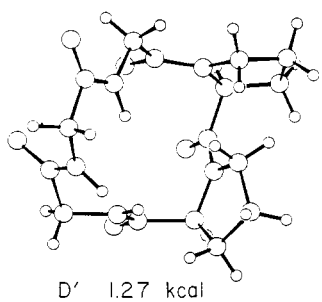
TABLE II
ENERGIES AND FREE ENERGIES OF THE THREE LOWEST MINIMUM-ENERGY CONFORMATIONS

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

[a] In entropy units/molecule. [b] Taken from Table I. [c] Total energy in units of kcal/molecule taken from Table II of I. [d] Calculated for T = 300°K in units of kcal/molecule.

thermal mixture of an almost equal number of conformations D′ and E′ with a much smaller number of conformation C′. However, before this conclusion can be accepted, we must answer two questions: (a) is the uncertainty in the values of the energy parameters used in the conformational energy calculations large enough to change the final result, and (b) is the small-fluctuation approximation upon which eq 1 is based good enough to warrant the use of the values of $-(R/2)$ ln det **GF** given in Table II as conformational entropies.





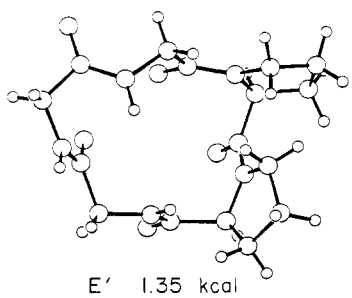


Figure 1. Conformations corresponding to the three lowest energy minima. The values of the minimum energies are given in units of kilocalories per mole of molecule.

TABLE III
THE MEAN-SQUARE AMPLITUDES OF FLUCTUATIONS IN THE DIHEDRAL ANGLES ψ_4 AND ψ_5 AT 300°K[a]

	Local minimum		
	C′	D′	E′
$\langle(\Delta\psi_4)^2\rangle$	3.0	12.1	6.9
$\langle\Delta\psi_4\Delta\psi_5\rangle$	1.8	0.5	0.1
$\langle(\Delta\psi_5)^2\rangle$	3.9	3.6	7.8

[a] Units are (deg)².

Question b will considered in section III, and an affirmative answer will be given. However, the answer to question a is also an affirmative one. In the previous paper,[4] a change in the radius of the hydrogen atom, R_H, from 1.200 to 1.275 Å was examined, as one of the most likely variations in the energy parameters. When that reasonable change in the value of R_H was made, minimum E′ was found to have an energy lower by 1.07 kcal than that of minimum D′ in contrast to the case, when R_H = 1.200 Å is used, where minimum E′ had an energy higher than that of minimum D′ by 0.08 kcal. This means that we cannot draw any definite conclusions as to the order of the energies of minima D′ and E′ because of the uncertainty in R_H. However, considering that the best value of R_H is most probably in the range bracketed by 1.200 and 1.275 Å, and that the relative energies of various conformations are expected to be most sensitive to the change in R_H (among various possible changes in the energy parameters) as discussed in the previous paper,[4] we obtain some idea of the range of the values of the relative energies of minima C′, D′, and E′ for reasonable changes in the energy parameters. As a result, we reach the following conclusions: isolated molecules of cyclo(Gly-Gly-Gly-Pro-Pro) are a thermal mixture of minimum-energy conformations C′, D′, and E′ with D′ or E′ being most abundant (including the possibility of both being more or less equally abundant), and with C′ being least abundant.

Figure 1 shows the conformations corresponding to minima C′, D′, and E′. Other characterizations of these minimum-energy conformations were given in Table II of I.

The range of fluctuations of the dihedral angles can be calculated in terms of the second derivatives of the energy surface at the minimum point by the formula

$$\begin{pmatrix} \langle(\Delta\psi_4)^2\rangle & \langle\Delta\psi_4\Delta\psi_5\rangle \\ \langle\Delta\psi_5\Delta\psi_4\rangle & \langle(\Delta\psi_5)^2\rangle \end{pmatrix} = kT\begin{pmatrix} C_1 & C_2 \\ C_2 & C_3 \end{pmatrix}^{-1} \quad (2)$$

where $\langle\ldots\rangle$ designates an average value of an enclosed quantity over various conformations fluctuating around the mean conformation corresponding to the minimum point, $\Delta\psi_4$ and $\Delta\psi_5$ are the deviations of ψ_4 and ψ_5 from that of the minimum point, C_1 and C_3 are the second derivatives of the energy surface at the minimum point with respect to ψ_4 and ψ_5, respectively, and C_2 is the mixed second derivative with respect to ψ_4 and ψ_5. Table III shows the calculated values of the mean-square amplitudes of the fluctuations. It is seen that the dihedral angles ψ_4 and ψ_5 fluctuate from the mean values by a few degrees. These fluctuations are the source of the second term on the right-hand side of eq 1. The correlation between the value of $-(R/2)$ ln det **F**

for a minimum-energy conformation given in Table II and the mean amplitudes of fluctuations for the same conformation given in Table III [the bigger the fluctuation, the bigger the value of $-(R/2)$ ln det **F**] should be noted. The structures shown in Figure 1 represent the mean conformations over the fluctuations.

At this point, it is worth making a general comment about calculations of stable conformations of polypeptides and proteins. There are three different aspects to such computations. First, we must determine the functional form (and the values of the parameters) of the conformational energy as a function of the conformation of the molecule. Second, we must find and locate minimum points on the given conformational energy surface. Third, we must calculate the conformational entropy at each minimum point obtained in the second step in order to determine the most stable conformation. In the case of the cyclic pentapeptide under consideration here, the second step was carried out by calculating the complete energy surface[4] (thereby surmounting the problem of multiple minima in conventional energy minimization procedures[10]), and the third step is treated in this paper (following an earlier prescription[5]). However, while much work is currently being carried out in this laboratory and elsewhere to treat the first problem, there still remain uncertainties in the values of the parameters of the conformational energy function which prevent us from drawing conclusions about the conformation of cyclo(Gly-Gly-Gly-Pro-Pro), beyond those already reported above. It should also be noted that the present calculation is the most complete one 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**). We can make the calculation even more complete (a) by allowing the molecule to be flexible (variable bond lengths and angles), and (b) by introducing the effects of solvation for isolated molecules in solution, or (b′) by introducing the effects of crystallization for molecules in the crystalline form. The theoretical basis of procedure a was given elsewhere,[5] and it is now being applied in this laboratory to the pentapeptide treated in this series.

III. Energy Surface in the Global Energy Minimum Region

The three lowest energy local minima C′, D′, and E′ exist in one relatively small region of the whole conformational space, which will be referred to hereafter as the global energy minimum region. The energy map in this region is shown in Figure 2. In this region there are up to four different conformations for a given set of values of the two independent variables ψ_4 and ψ_5, which is the reason why four sheets of maps are needed in Figure 2 to show the energy surface of one continuous region. Common boundaries for the existence of solutions in two different sheets of maps are marked by a and a′, b and b′, c and c′. Thus, for example, the 8-kcal contour in Figure 2A, which vanishes at the boundary a, continues to exist as the 8-kcal contour in Figure 2B, which starts at the boundary a′. Lines d and d′ in Figures 2B and 2D, respectively, represent the same line. Thus, if one goes clockwise around the terminal point of the boundary (ψ_4 = 128.3°, ψ_5 = 108.7°) from Figure 2B, one travels into the region shown in Figure 2D. An easy way to visualize the construction of Figure 2 would be to cut the figures along the lines a, a′, b, . . . , d′ and paste them together keeping the correspondence between the

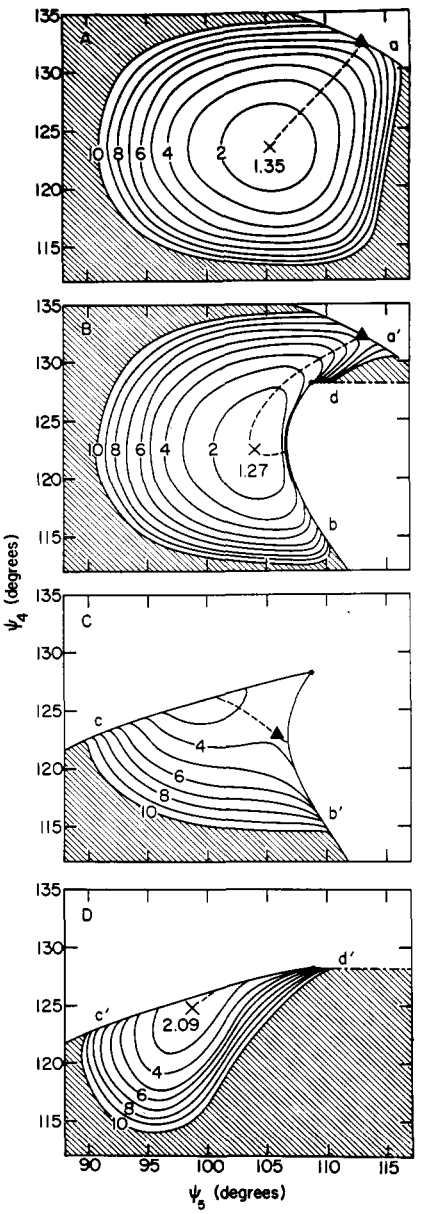


Figure 2. The contour diagram of the energy surface in the global energy minimum region. Energies are given in units of kilocalories per mole of molecule. The three energy minima in this region are marked by an ×, and their energy values are given. Dotted lines show the easiest path ("reaction path") to go from one minimum to another; these dotted lines are drawn in here more accurately than in Figure 7 of I, because of the finer grid used in the energy calculations in the present paper. The two saddle points on the path are marked by the symbol ▲. The shaded region has energies higher than 10 kcal. See text for further description.

(10) H. A. Scheraga in 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 43.

lines mentioned above. The locations of the three minima C′, D′ and E′, and their energies are shown. The shaded region has energies higher than 10 kcal/mol of molecule.

Equation 1 was derived on the assumption that the energy surface can be approximated by a multidimensional parabola in the vicinity of a minimum point where the statistical weight (the Boltzmann factor) is appreciable. In order to check the validity of this assumption for a real energy surface, the integral of the Boltzmann factor over an energy minimum region was calculated by three different methods for minimum E′.

The first method is the one used to derive eq 1. There, the energy function E (measured from the minimum value) is approximated by a two-dimensional parabola

$$E = {}^1/_2(C_1x^2 + 2C_2xy + C_3y^2) \tag{3}$$

where x and y are the two independent variables (*i.e.*, ψ_4 and ψ_5) and C_1, C_2, and C_3 are the same second derivatives of the energy function at the minimum point which appeared in eq 2; the integration is performed analytically in the whole range of the independent variables to give[11]

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-E(x,y)/kT\}\mathrm{d}x\mathrm{d}y = 2\pi kT(\det \mathbf{F})^{-1/2} \tag{4}$$

with

$$\det \mathbf{F} = \begin{vmatrix} C_1 & C_2 \\ C_2 & C_3 \end{vmatrix} \tag{5}$$

In the second method, the energy function E is expanded in a Taylor series around the minimum and approximated by a quartic function

$$E = {}^1/_2(C_1x^2 + 2C_2xy + C_3y^2) + {}^1/_6(D_1x^3 + 3D_2x^2y + 3D_3xy^2 + D_4y^3) + {}^1/_{24}(E_1x^4 + 4E_2x^3y + 6E_3x^2y^2 + 4E_4xy^3 + E_5y^4) \tag{6}$$

where the D's and E's are the third and fourth derivatives of the energy function at the minimum point with respect to x and/or y, and the Boltzmann factor is approximated by

$$\exp\{-E(xy)/kT\} \cong \left\{1 - \frac{1}{6kT}(D_1x^3 + 3D_2x^2y + 3D_3xy^2 + D_4y^3) - \frac{1}{24kT}(E_1x^4 + 4E_2x^3y + 6E_3x^2y^2 + 4E_4xy^3 + E_5y^4)\right\} \times \exp\{-(C_1x^2 + 2C_2xy + C_3y^2)/2kT\} \tag{7}$$

The approximation of eq 7 is good when the cubic and quartic terms in eq 6 are smaller than unity in the region where the last factor of the right-hand side of eq 7 is appreciable. The integration can be performed analytically for the approximate Boltzmann factor of eq 7 to give

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-E(x,y)/kT\}\mathrm{d}x\mathrm{d}y = \frac{2\pi kT}{(\det \mathbf{F})^{1/2}}\left[1 - \frac{kT}{8(\det \mathbf{F})^2}\{C_3^2E_1 - 4C_2C_3E_2 + 2(C_1C_2 + 2C_2^2)E_3 - 4C_1C_2E_4 + C_1^2E_5\}\right] \tag{8}$$

This formula will be examined to see if it gives us an improvement over eq 4. If it does, it is a useful formula, because it involves quantities calculated at a single point of the conformational space, and it can be generalized for cases of three and more dimensions.

The third method is the numerical integration of the Boltzmann factor whose values are known at grid points of a square lattice. The integration is replaced by the summation

$$\iint \exp\{-E(x,y)/kT\}\mathrm{d}x\mathrm{d}y = {}^1/_2(kT\Delta)^2\sum_i\sum_j \exp\{-E_{i,j}/kT\} \times \{D(0,1;\ 1,0;\ i,j) + D(1,0;\ 0,-1;\ i,j) + D(0,-1;\ -1,0;\ i,j) + D(-1,0;\ 0,1;\ i,j) + D(0,1;\ 1,1;\ i,j) + D(1,1;\ 1,0;\ i,j) + D(1,0;\ 1,-1;\ i,j) + D(1,-1;\ 0,-1;\ i,j) + D(0,-1;\ -1,-1;\ i,j) + D(-1,-1;\ -1,0;\ i,j) + D(-1,0;\ -1,1;\ i,j) + D(-1,1;\ 0,1;\ i,j)\} \tag{9}$$

with

$$D(k,l;m,n,i,j) = (E_{i+k,j+l} - E_{i,j})^{-1}(E_{i+m,j+n} - E_{i,j})^{-1} \tag{10}$$

where Δ is the lattice spacing and $E_{i,j}$ is the energy value at the grid point (i,j). The summation extends over a region around the minimum point, where the energy is not more than 5.5 kcal above the minimum value. The relative error arising from the neglect of the contribution from the region of higher energy is about 10^{-4}. The derivation of eq 9 is described in the Appendix. This formula is the most accurate of the three methods mentioned above However, the generalization of this method to cases of higher dimensionality is impractical, because the method requires the calculation of the values of a function in the whole range of an energy minimum region.

The results of the calculation for local minimum E′ are summarized in Table IV, with the result from eq 9 being the most accurate. The value obtained by the quadratic approximation (*i.e.*, by using eq 4) is seen to be about 10% off from the true value. This corresponds to an error of about 0.2 eu in the calculation of

TABLE IV
CONTRIBUTION TO THE STATISTICAL WEIGHT OF LOCAL MINIMUM E′ FROM CONFORMATIONAL SPACE, CALCULATED BY VARIOUS METHODS AT 300°K[a]

1.407×10^{-2}	By eq 4
1.216×10^{-2}	By eq 8
1.291×10^{-2}	By eq 9

[a] Units are (radian)^2.

(11) N. Gō, M. Gō, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.* **59**, 1030 (1968).

the conformational entropy of the minimum-energy conformation E′. An inspection of the value in Table IV shows that eq 8 is not much of an improvement over eq 4. As a conclusion, local minimum E′ can be treated as a conformation with small conformational fluctuations, *i.e.*, the energy surface can be approximated by a two-dimensional parabola in that region around the minimum point where the statistical weight is appreciable, if an error of 0.2 eu in the calculation of its entropy can be tolerated.

Before proceeding to a discussion of the validity of treating local minima C′ and D′ as conformations with small conformational fluctuations, we digress to consider further the properties of such arbitrary local minima in the two-dimensional problem (for which the energy function can be approximated by a two-dimensional parabola in the immediate region of the minimum point where the statistical weight is appreciable). Let us consider the quantity

$$I(E_0) = \iint_{E(x,y)<E_0} \exp\{-E(x,y)/kT\}\,dx\,dy \qquad (11)$$

which is the contribution to the statistical weight from that region of the conformational space where the energy is less than E_0 above the minimum value. When $E(x,y)$ is a quadratic function in x and y (*i.e.*, of the form of eq 3) it can be shown that

$$I(E_0)/I(\infty) = 1 - \exp(-E_0/kT) \qquad (12)$$

From this equation, we can conclude that, at 300°K, 90% of the statistical weight arising from the conformational space comes from the region where the energy is less than 1.4 kcal above the minimum value, when the problem is two-dimensional. Hence, if the energy function can be approximated by a quadratic function in the region where the value of the function is less than 1.4 kcal above the minimum value, and if an error of the order of 0.2 eu in the conformational entropy can be tolerated, eq 1 based on the assumption of small conformational fluctuations can be used for the evaluation of the conformational free energy. A crude, yet workable, method of estimating the range for which the quadratic approximation is valid is to determine the value of the function at a saddle point ΔE_{saddle} (whose energy is measured from the minimum point) with the lowest energy value around the minimum. The quadratic approximation of a function around a minimum point clearly fails to reproduce the energy function at saddle points. If $\Delta E_{\text{saddle}} > 1.4$ kcal, the small conformational fluctuation approximation works. If $\Delta E_{\text{saddle}} < 1.4$ kcal, the approximation fails. If $\Delta E_{\text{saddle}} \simeq 1.4$ kcal, the approximation must be used with caution. In the last case the maximum error involved in the small conformational fluctuation approximation would be about 1 eu.

In Figure 2 the easiest path to go from one local minimum to another (reaction path) is shown by dotted lines. The energy profile along this "reaction path" is shown in Figure 3A. Figure 3B is the same, but for the case where $R_H = 1.275$ Å is used instead of $R_H = 1.200$ Å (see section IIC of I). In both Figures 3A and 3B, ΔE_{saddle} for minima D′ and E′ is larger than 1.4 kcal. Therefore, the use of eq 1 for minima

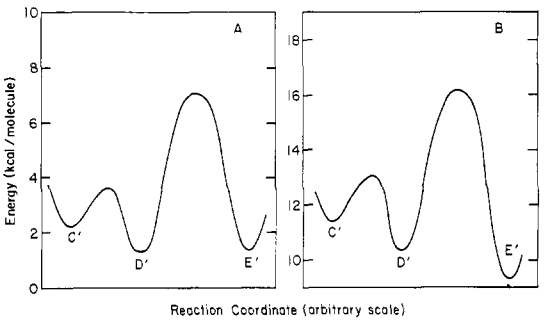


Figure 3. Energy profile along the "reaction path," (A) for $R_H = 1.200$ Å (the one for the dotted lines in Figure 2) and (B) for $R_H = 1.275$ Å.

D′ and E′ is good and would involve errors of the order of 0.2 eu. (For E′ this was already shown earlier in this section more directly.) For minimum C′, $\Delta E_{\text{saddle}} = 1.4$ kcal for $R_H = 1.200$ Å and $\Delta E_{\text{saddle}} = 1.7$ kcal for $R_H = 1.275$ Å. So, the use of eq 1 for minimum C′ may involve a maximum error of about 1 eu. However, this uncertainty in the free energy of the conformation corresponding to minimum C′ is sufficiently small so that the free energy of this conformation could not be lower than those of D′ and E′. Hence, we can safely exclude the possibility of minimum C′ being the most stable conformation of an isolated pentapeptide cyclo(Gly-Gly-Gly-Pro-Pro). Thus, the main conclusion made in section II about the most stable conformation of this molecule based on the small conformational fluctuation approximation is found to be correct. In more general cases, caution must be exercised about the use of eq 1, because the assumption of small conformational fluctuations is not necessarily always satisfied.

The main conclusion about the conformation of an isolated pentapeptide cyclo(Gly-Gly-Gly-Pro-Pro) arrived at above is not expected to hold without modification for the molecule in the crystalline state. Two factors, the intermolecular interactions in the crystal and the change of conformational fluctuations of the molecule upon crystallization, can cause the conformation of the molecule to change. When the molecules are packed in the crystal leaving little free space, a molecule can hardly have conformational fluctuations without affecting the conformations of nearby molecules. In such circumstances, the contribution to the free energy of a conformation from the librational degrees of freedom can differ greatly from the one for an isolated molecule. Various intermolecular interactions would not be strong enough to make conformations outside the global minimum region for an isolated molecule the most stable (because minima other than C′, D′, and E′ have energies which are higher by at least 18 kcal), but would be strong enough to change the appearance of the energy surface in the global energy minimum region. Among the various intermolecular interactions, intermolecular hydrogen bond formation would be the most important for this cyclic pentapeptide. This is because it is a strong interaction and also because there are many polar groups (CO and NH groups), which can form a hydrogen bond, sticking outside the molecule in conformations in the global minimum region (see Figure 1, and also Table II of I). The

conformation of the cyclic pentapeptide in question in a crystal would be the one in the global energy minimum region shown in Figure 1 which would optimize intermolecular hydrogen bond formation. The packing of molecules in a crystal must be considered explicitly in order to determine the conformation of a molecule in a crystal by a theoretical calculation.

Appendix

Equation 9 is derived in this appendix. Consider the integral on the left-hand side of eq 9 over a triangular region bounded by straight lines passing through the grid points (i,j), $(i + 1,j)$ and $(i,j + 1)$ of a square lattice; we shall designate this integral by $I(0,1;1,0;i,j)$ [as we did for $D(k,l;m,n;i,j)$ defined by eq 10]. If the lattice spacing Δ is chosen properly, the energy function $E(x,y)$ can be approximated by a linear function in x and y in this triangular region. The integration can then be carried out explicitly to give

$$I(0,1;1,0;i,j) = (kT\Delta)^2\{D(0,1;1,0;i,j)\exp(-E_{i,j}/kT) + D(-1,0;-1,1;i+1,j)\exp(-E_{i+1,j}/kT) + D(0,-1;1,-1;i,j+1)\exp(-E_{i,j+1}/kT)\} \quad (A.1)$$

Then the integration over the whole energy surface (*i.e.*, the left-hand side of eq 9) can be given by

$$\tfrac{1}{2}\sum_i\sum_j\{I(0,1;1,0;i,j) + I(1,0;0,-1;i,j) + I(0,-1;-1,0;i,j) + I(-1,0;0,1;i,j)\} \quad (A.2)$$

The right-hand side of eq 9 is obtained by rearranging terms in expression A.2.

The Configuration of Amylose and Its Derivatives in Aqueous Solution. Experimental Results

Kenneth D. Goebel and David A. Brant*

Department of Chemistry, University of California, Irvine, California 92664.
Received April 30, 1970

ABSTRACT: The unperturbed mean square end-to-end lengths $\langle r^2\rangle_0$ of sodium carboxymethylamylose (NaCMA) and diethylaminoethylamylose hydrochloride (DEAEA·HCl) have been determined in aqueous medium as a function of degree of polymerization (DP) and temperature using light scattering, osmotic pressure, and intrinsic viscosity measurements. A medium of high ionic strength (0.78 *M* NaCl) was employed in which reliable corrections for excluded volume expansion were possible using measured second virial coefficients. Degrees of substitution (DS) were sufficient to provide stable aqueous polymer solutions, but small enough to allow the derivatives to serve as suitable models for unmodified amylose. The characteristic ratio, $C_x = \langle r^2\rangle_0/xl^2$, where x is the DP and l is the distance between adjacent glycosidic bridge oxygens, was 5.3 and 6.4, respectively, for NaCMA (DS = 0.30) and DEAEA·HCl (DS = 0.20). No dependence of C_x on DP was observed for either derivative in the range 232–1710. In comparison, $C_x = 5.1$ and is constant in the DP range 513–10,800 for amylose. The respective temperature coefficients, d ln C_∞/dT, are -0.0066 and -0.0069 deg^{-1} for NaCMA and DEAEA·HCl. These results are interpreted in light of a theoretical treatment of amylose configuration presented in an accompanying paper.

Theoretical methods which provide an exact connection between the skeletal structure of linear polymer chains and their unperturbed configurational properties in dilute solution and in the bulk amorphous state have been extensively developed within the last decade.[1-3] Observable equilibrium chain properties may be calculated from the configuration partition function which is constructed from knowledge of backbone structural features and information about the conformational energies of appropriately chosen chain segments. Successful application of these methods to test structural hypotheses and hypotheses concerning contributions to the hindrance potentials for internal rotation about the skeletal bonds demands accurate determination of equilibrium properties such as, for example, the unperturbed mean-square radius of gyration or dipole moment of the chain. Studies reported here continue our efforts described earlier[4] to obtain appropriate experimental data for application of the statistical mechanical theory of polymer chain configuration to macromolecules of the polysaccharide class. In particular we seek information for amylose and derivatives on the chain length and temperature dependence of the characteristic ratio, $C_x = \langle r^2\rangle_0/xl^2$, of the unperturbed mean-square end-to-end distance $\langle r^2\rangle_0$ to the product xl^2 for the number of chain segments x of fixed length l.

Experimental attention is focused in the present work on two polyelectrolyte ether derivatives of amylose, sodium carboxymethylamylose (NaCMA),

* To whom correspondence should be addressed.

(1) M. Volkenstein, "Configurational Statistics of Polymeric Chains," S. N. Timasheff and M. J. Timasheff, Trans., Interscience, New York, N. Y., 1963.

(2) T. M. Birshtein and O. B. Ptitsyn, "Conformations of Macromolecules," S. N. Timasheff and M. J. Timasheff, Trans., Interscience, New York, N. Y., 1966.

(3) P. J. Flory, "Statistical Mechanics of Chain Molecules," Interscience, New York, N. Y., 1969.

(4) D. A. Brant and B. K. Min, *Macromolecules*, 2, 1 (1969).