MINIMIZATION OF POLYPEPTIDE ENERGY, III. APPLICATION OF A RAPID ENERGY MINIMIZATION TECHNIQUE TO THE CALCULATION OF PRELIMINARY STRUCTURES OF GRAMICIDIN-S*

BY ROY A. SCOTT,† GARRET VANDERKOOI,‡ ROBERT W. TUTTLE, PETER M. SHAMES, AND HAROLD A. SCHERAGA§

DEPARTMENT OF CHEMISTRY, CORNELL UNIVERSITY, ITHACA, NEW YORK

In the development of procedures for the calculation of polypeptide conformations, early work was based on hard-sphere potentials.[1-3] Subsequent calculations were made with more complete energy expressions.[4,5] At the same time, methods were explored for obtaining the conformation of minimum energy. In papers I[6] and II[7] of this series, Davidon's variable metric method[8] was used to minimize the energies of straight-chain[6] and cyclic[7] polypeptide structures. In this method, gradients are computed, and all of the variables are permitted to change simultaneously. Since the multidimensional energy surface contains many local minima, it is advantageous to have a rapid procedure for minimizing the energy so that a large number of randomly selected starting points may be employed. Such a procedure is described in this paper, and applied to the calculation of the structure of gramicidin-S. Since an exhaustive search of all possible local minima has not been made, the structures reported here must be regarded as preliminary ones.

Calculation of the Energy.—The amide geometry (including the planar *trans* conformation about the peptide bond) and energy functions of Ooi *et al.*[5] were used in the computations reported here. For nonbonded interactions, the parameters in Table II (columns 2 and 4) of reference 5 were taken. For the electrostatic interactions, the effective dielectric constant was set equal to 4.0. The hydrogen bond function[5] differs from that which was used in our earlier work[9] on gramicidin-S; hence, the energies given in the present paper are not directly comparable to those of the previous paper.[9] In contrast to the previous papers of this series,[6,7] specific solvation effects were omitted; in this work, the energy, rather than the free energy, is calculated.

Closure of the gramicidin-S ring was effected by including a fictitious, empirical loop-closing potential in the total energy. The C^α-C' bond of an arbitrarily selected residue was the site of the break in the ring, and the following potential function used to close the gap:

$$U_{gap} = A|r - r_0| + B(2 - \cos \alpha_{-1} - \cos \alpha_{-2}), \qquad (1)$$

where r_0 is the equilibrium length of the C^α-C' bond (1.53 Å), and r is the actual distance in angstrom units between the C^α and C' atoms; α_1 and α_2 are the angles $\tau[D_1 C^\alpha C']$ and $\tau[D_2 C' C^\alpha]$, D_1 and D_2 being dummy atoms attached to the C^α and C' atoms, respectively, in the directions of the bonding orbitals. The angles α_1 and α_2 become zero when the correct bond angles obtain. The parameters A and B are adjustable ones, which must be large enough to close the loop, but not so large as to dominate completely the total energy; out of the several sets of A and B values tried, the values $A = 12$ and $B = 100$ (for U_{gap} in kcal/mole) were selected.

The value of U_{gap} will differ from zero if there is any deviation from the proper bond distance or bond angles at the C^α-C' bond.

The total energy used for minimization included contributions from U_{gap} as well as torsional, nonbonded, electrostatic, and hydrogen bond energies.[5] However, the final (usually small) value of U_{gap} has been substracted from the final energies reported in this paper. All side-chain and backbone hydrogen atoms were included in the calculations. Rotation was permitted around all backbone and side-chain single bonds, except those leading to terminal methyl groups which were fixed in rotational minima; the energy was minimized by allowing for a continuous variation of all of the dihedral angles. The ε-amino groups of the two ornithine residues were uncharged. The procedures for transformation of coordinates were equivalent, but not identical, to those of Ooi *et al.*[5]

Minimization Procedure.—The minimization procedure was designed to provide very rapid minimization from a large number of starting conformations. It consists of three parts: (1) a grouping of variables into subsets, (2) a line-search technique, and (3) a procedure utilizing the line searches to explore a hypervolume. The rationale behind this approach is that the hypersurface, which represents the energy of a polypeptide, is very irregular, and probably has many local minima; as a result, it is probably impossible to proceed downhill from arbitrary points on the surface to the global minimum. Therefore, it is desirable to rapidly search the surface in order to locate the low energy local minima, rather than to carefully follow the surface downhill from an arbitrary starting point. In the procedure used here, many randomly selected starting points are taken, and the energy is rapidly, but crudely, minimized from each of them. The resulting conformations of high energy are discarded, and the low-energy ones are then used as starting points for a more careful energy minimization by the same procedure.

The dihedral angles of the backbone and side chains are the independent variables; those of an individual peptide unit[10] constitute a subset of the complete set of dihedral angles. At times, each of such subsets is divided further into two subsets, one for the backbone and the other for the side-chain dihedral angles. In this way, backbone and side-chain angles can be varied independently. An exception is made in the case of proline, which has only one bond (the C^α-C') about which rotation can occur; this variable is usually appended to the subset of the adjacent peptide unit. In the minimization procedure, the variables of only one subset are varied at a time; those of all other subsets are held constant. In this way, pairwise atomic interaction energies had to be computed only for those pairs of atoms whose interatomic distance changed (i.e., those on opposite sides of the peptide unit whose dihedral angles are being varied). No computation was required for those interatomic distances which did not change as a result of a change of the variables of the given subset (i.e., those on the same side of the peptide unit whose dihedral angles are being varied). By thus limiting the number of pair interactions which had to be computed during a change of the variables of any given subset, a considerable reduction of computer time was achieved, compared to that which would be required if all the dihedral angles were varied simultaneously. The computer program was arranged so that the subsets could be taken in any order. In a *full* iteration, all the variables of each peptide unit subset were permitted to vary, subset by subset; likewise, a *side-chain* iteration consisted of one

cycle through all side-chain subsets, and a *backbone* iteration consisted of one cycle through all backbone subsets.

The line search technique provides the energy minimum along a given line in hyperspace. It is based on quadratic interpolation between three points which bracket an energy minimum along the given line, i.e., the middle point has a lower energy than the two outer ones. A parabola is fitted to the energies at these three points, and the position along the line of the energy minimum of the parabola is found by equating the derivative of the parabolic function to zero. Since the actual surface is usually such that quadratic interpolation does not yield the exact minimum, the following additional feature is added to the interpolation: if the minimum of the parabola occurs within a specified tolerance of one of the three selected points, then the nearest original point is taken as the minimum; otherwise, the energy is computed at the calculated minimum, and the interpolation procedure is repeated using the calculated minimum as the middle of the three points required to construct a new parabola. This precaution is necessary in order to guard against problems arising from a significant departure of the energy surface from the assumed parabolic shape.

Having specified the manner of selecting the subsets, and the procedure for finding a minimum along a line, we now consider how to search for an energy minimum in a hypervolume in which all variables of a single subset are varied. A point in hyperspace represents a given set of dihedral angles; a line in hyperspace is completely defined by two points in hyperspace. Additional points along the same line may be found through the use of standard techniques of analytic geometry. The procedure used here for selecting lines in hyperspace is a simplified version of "general partan."[11] *Without* computing gradients, a line is chosen in an arbitrary direction, and the minimum along this line is found by the procedure described above. A second arbitrarily chosen line is placed parallel to the first one, and the minimum along this second line is also found. A line is then placed through the minima on the two parallel lines, and the minimum along this third line is found. This completes an iteration in two dimensions. For three or more dimensions, a fourth line is placed through an arbitrary point but parallel to the third line. A minimum is found on this line, and a fifth line is formed through the minima of the third and fourth lines. Determination of the minimum along the fifth line completes an iteration for three dimensions. This process is continued for higher dimensions by placing a sixth line parallel to the fifth, etc. In practice, the arbitrarily placed lines are conveniently chosen, i.e., the first line is taken so as to vary only the first dihedral angle, and the subsequent nth line is chosen either as required to pass through the minima of the two preceding lines, or as a parallel line (the lines of even n), which is found by incrementing the $[(n/2) + 1]$th variable. This simple means of placing the lines was the method of choice, since it was found to give equally as good results as those obtained with various more complicated line-placement procedures. By the time that a search is made along the last line for a given subset, all the variables of the subset are changing simultaneously. This procedure is then continued systematically through all the subsets in an arbitrary sequence.

Results.—The minimization procedure was applied to two special starting conformations and to 27 randomly selected ones of gramicidin-S. The two special con-

formations were those suggested by Vanderkooi *et al.*[9] (designated as GS_I) and by Liquori *et al.*[12] (designated as GS_{II}). Structure GS_I had been obtained by generating approximately 280 sterically allowed conformations[9] having a twofold axis of symmetry,[13] and then evaluating the energy of each of them; perturbations of the dihedral angles of the lowest-energy structure out of the 280 led to the final structure. Structure GS_{II} was obtained by use of an assumed "stereochemical code,"[12] together with the same symmetry requirement.[13]

The coordinates published by Liquori *et al.*[12] were given to the nearest 0.1 Å. The low accuracy of the coordinate data prevented the reconstruction of the structure by calculation of the dihedral angles, on account of the variable bond lengths represented by the coordinates. Also, with the exception of proline, no coordinates were given for side-chain atoms beyond the β-carbon.[12] These problems were circumvented by building a non-space-filling model from the data available in the paper of Liquori *et al.*[12] In order to close the ring, it was necessary to change the dihedral angles of the α-helical portions of the backbone by about 10° from the values for the standard α-helix ($\phi = 127°$, $\psi = 128°$) (in fact, Liquori *et al.*[12] also mentioned that such a ring-closure adjustment was necessary). From this model, the dihedral angles shown in parentheses in the right-hand part of Table 1 were obtained; this is the structure which will be called GS_{II}.

Starting conformations for the *side chains* of GS_I and GS_{II} were obtained by selecting those which gave the lowest energy of the structure, after permutation through all the minimum positions of the torsional potentials of each side chain independently.

Extensive energy minimization, starting from GS_I, led to a similar conformation which will be designated GS_{Ia}. The structure obtained by energy minimization from GS_{II} will be called GS_{III}, since it differs from GS_{II} in certain important respects (see below). The dihedral angles (in degrees, defined according to the proposed standard conventions[10]) for both of these structures are given in Table 1, together with the angles of the starting conformations. No energy values are given for the starting conformations since they were artificially high, because of improper loop closure. We have not included illustrations of GS_{Ia} and GS_{III}, since their gross appearances are the same as the conformations GS_I and GS_{II}, respectively, from which they were obtained; both of the starting conformations have already been illustrated.[9,12] The loop closure obtained (at the leucyl residue) was excellent in both *final* conformations; the bond length across the gap deviated by less than 0.01 Å from the desired value (1.53 Å), and the bond angles were within 2° of the desired values.

Structure GS_{Ia} contains two good hydrogen bonds between the CO groups of the prolyl residues and the NH groups of the ornithyl residues, as previously reported for GS_I also.[9] The average of these two N...O distances is 2.82 Å and the average HNO angle is 11°. This N...O distance is close to the average value of 2.93 ± 0.11 Å given by Wallwork[14] for NH...O hydrogen bonds observed in crystals. On the other hand, structure GS_{III} does not contain any strong hydrogen bonds. The average N...O distance is 3.47 Å and the average HNO angle is 15° for the possible α-helix-type hydrogen bonds between the prolyl oxygens and the phenylalanyl NH groups. This N...O distance is considerably larger than the range of N...O distances found in crystals.[14] The conceivable 3_{10}-helix-type hydrogen bonds be-

DIHEDRAL ANGLES OF CONFORMATIONS OBTAINED

Residue	$\phi(N - C^\alpha)$	$\psi(C^\alpha - C')$	χ_1	χ_2	χ_3
			——GSI$_a$^{a,b}——		
L-val	53.1 (60)	303.8 (310)	303.9	...	...
L-orn	69.4 (70)	231.8 (240)	300.6	181.8	53.5
L-leu	243.7 (230)	240.0 (230)	311.9	302.7	...
D-phe	309.1 (320)	118.8 (100)	37.6	62.3	...
L-pro	123.0 (120)^e	303.0 (320)	...	...	...
L-val	74.0 (60)	307.2 (310)	308.8	...	...
L-orn	71.3 (70)	226.3 (240)	303.1	179.8	50.1
L-leu	243.2 (230)	230.0 (230)	310.7	302.6	...
D-phe	317.1 (320)	108.6 (100)	49.9	61.6	...
L-pro	123.0 (120)^e	325.1 (320)	...	...	...

^{a} The backbone dihedral angles of GS$_I$ are given in parentheses.
^{b} The energy of GS$_{Ia}$ is −96 kcal/mole.
^{c} The backbone dihedral angles of GS$_{II}$ are given in parentheses.
^{d} The energy of GS$_{III}$ is −98 kcal/mole.
^{e} Fixed angle of proline, i.e., set at 123°.

tween the valyl oxygens and the phenylalanyl NH groups in GS$_{III}$ are also weak or nonexistent because of a poor HNO angle: the average N...O distance is 2.92 Å, but the HNO angle is 58°. GS$_{III}$ is therefore a new structure for gramicidin-S; its general appearance is similar to that of GS$_{II}$, but it is distinctively different in that it lacks the α-helical hydrogen bonds which are an essential characteristic of the structure proposed by Liquori et al.[12]

An attempt was made to find a low-energy conformation containing two α-helical segments by putting the val, orn, and leu residues in the α-helical conformation ($\phi = 127°$, $\psi = 128°$, instead of the values listed in Table 1), and initially permitting the backbone dihedral angles of only the phe and pro residues to vary (in addition to all of the side-chain angles). Under these conditions, it was impossible to properly close the ring. When all of the backbone angles were then permitted to change, energy minimization led to the same GS$_{III}$ structure already described; no intermediate structure containing α-helical turns, which had even moderately low energy, was found. We therefore conclude that, with the present set of potential functions and molecular geometry, gramicidin-S is not at an energy minimum in the α-helical-type structure; however, a geometrically related structure, which lacks hydrogen bonds, is stable. Further, it is noteworthy that the same final structure (GS$_{III}$) was obtained from two different starting structures.

Besides starting with conformations GS$_I$ and GS$_{II}$, 27 randomly selected conformations were also considered. These were obtained by random number generation of all of the backbone dihedral angles for 27 conformations; no condition of symmetry was imposed on the starting conformations. These randomly generated structures had high energies because of atomic overlaps. The initial energy minimizations were carried out by truncating the side chains after the β-carbon atoms. A single backbone iteration on the truncated structures (Stage I) eliminated severe overlaps. Complete side chains were then added to the 12 conformations of lowest energy, the remaining 15 high-energy conformations being discarded. Minimization from the 12 full structures was then continued in two successive stages, II and III. Letting B and S designate backbone and side-chain iterations, respectively, the three stages correspond to the following cycles: I, B; II, SBS; III, BBSBB. In each of the 12 cases, the energy did not change by more than 1 kcal/mole in the final iteration. It is likely that the energies of some or all of these structures could

1

BY ENERGY MINIMIZATION FROM GS$_I$ AND GS$_{II}$

| Residue | ϕ(N − C$^\alpha$) | ψ(C$^\alpha$ − C′) | GS$_{III}$[c,d] | | |
			χ_1	χ_2	χ_3
L-val	106.7 (120)	148.4 (140)	298.5	...	...
L-orn	125.0 (120)	146.5 (140)	63.0	185.2	59.5
L-leu	104.5 (120)	120.0 (140)	178.9	185.2	...
D-phe	321.0 (300)	117.2 (110)	166.7	295.9	...
L-pro	123.0 (123)[e]	276.6 (300)	...	...	...
L-val	107.8 (120)	144.1 (140)	297.4	...	...
L-orn	128.9 (120)	145.0 (140)	62.0	185.8	59.6
L-leu	106.3 (120)	128.0 (140)	178.2	186.5	...
D-phe	327.3 (300)	116.3 (110)	166.0	295.9	...
L-pro	123.0 (123)[e]	277.9 (300)	...	...	...

be lowered somewhat further by using a different minimization procedure, such as that described previously.[6, 7] The lowest energy obtained by this procedure was −58 kcal/mole; no two of the final conformations were the same or closely similar.

Discussion.—It was found in the course of the calculations that GS$_I$ and GS$_{Ia}$ are representative structures of a large number of closely related conformations of similar energy. Apparently, the energy surface contains many real or apparent local minima in the region corresponding to these structures, with their energies differing by only a few kcal/mole. The lowest energy found for any of the GS$_{Ia}$-type structures was −97 kcal/mole. Although all of the structures found in this region are roughly symmetrical, they do in general depart from precise symmetry (by up to ±20° for the one of energy −97 kcal/mole) for certain pairs of dihedral angles. The conformation given in Table 1 (GS$_{Ia}$) was selected for presentation on the basis of low energy (−96 kcal/mole) and fairly good symmetry (maximum departure from symmetry is ±11°). Several attempts were made to obtain a structure with better symmetry by taking the average values of the corresponding pairs of dihedral angles as a new starting point for minimization, but in each case minimization led once again to a nonsymmetrical structure. A calculation was also performed in which perfect symmetry was enforced during the course of the minimization; those conditions led to a conformation with an energy of −90 kcal/mole. This value is considerably higher than the minimal value of −97 kcal/mole found for a related nonsymmetrical structure.

One might be led to conclude from the above that the GS$_{Ia}$-type structure does not exist in crystalline gramicidin-S, since a twofold molecular axis of symmetry seems to be present there.[13] This is not necessarily the case, however, for the following reasons: (1) The calculations were performed on an isolated molecule rather than on a molecule in the crystalline environment; it is possible that complete symmetry would be restored on account of the *inter*molecular forces in the crystal. (2) Fixed bond lengths and angles were used in the calculations; small variations of these parameters could possibly restore complete symmetry to the lowest energy conformation. (3) The energy parameters used in these calculations are only first approximations; they will be revised in the future on the basis of results obtained from calculations on known crystal structures of small molecules. (4) Asymmetry is introduced into the molecule by the loop-closing potential function.

The GS_{Ia}-type structure for gramicidin-S may exist in solution. The many closely related forms, in which it can exist, would be expected to provide further stabilization of the structure through a significant conformational entropy contribution to the total free energy. This type of structure is quite similar to the one proposed by Schwyzer[15] to explain the results of the dimerization and cyclization reaction of gramicidin-S pentapeptide. Schwyzer's structure, in turn, is the same as the β-pleated-sheet type of structure preferred by Hodgkin and Oughton[16] to explain their X-ray diffraction data. Their structure differs from GS_{Ia} in that four intramolecular hydrogen bonds were proposed,[15,16] whereas GS_{Ia} contains only two.

The GS_{III} structure has a slightly lower energy (-98 kcal/mole) than has been obtained for any GS_{Ia}-type structure, and has good, but not perfect, symmetry (maximum deviation from symmetry is $\pm 4°$). It has already been pointed out that this conformation lacks hydrogen bonds, and that it is therefore not an α-helix type structure such as has been proposed by Liquori et al.[12] Its low energy arises from favorable nonbonded and electrostatic energy contributions, rather than from hydrogen bonding. This is a very tight structure, in contrast to the fairly loose GS_{Ia} structure: small variations in its geometry caused great increases in the energy. As a result, the conformational entropy of this structure would be expected to be much smaller than that of GS_{Ia}. It also means that the energy calculated for this structure may be very sensitive to changes in the nonbonded energy functions.

The conformational entropy contribution to the free energy of the isolated molecule may be an important consideration for gramicidin-S in solution, where all the energetically accessible states will be occupied to various extents, but it probably is of little or no importance for the crystalline state, where all of the molecules would be frozen into one, or at most two, conformations, as indicated by the crystal symmetry.[13] We may conclude from this that GS_{Ia} would be entropically favored in solution, but that both GS_{Ia} and GS_{III} are possible structures for the crystal, since they have nearly equivalent energies. It cannot be overemphasized, however, that these conclusions are strictly preliminary and tentative, on account of the reservations and possible sources of error already mentioned.

The data on the 27 randomly chosen starting conformations (deposited with the American Documentation Institute) serve only to illustrate a general procedure which can be applied to a polypeptide for which no structural information is already available. An insufficient number of starting points was used to make possible statistical predictions concerning the energy surface. Examination of the energies of these 27 conformations showed that none of them approached the energies of GS_{Ia} or GS_{III}. This indicates that there are many real or apparent local minima with energies much higher than the global minimum. (We do not mean to imply by this that GS_{Ia} or GS_{III} is necessarily at the global minimum, but only that the energy of the global minimum must be at least as low as the energies of these structures.) In the future, we hope to find out how many randomly chosen starting points must be taken for a given type of polypeptide, in order to yield a high probability of finding the global minimum.

Appendix.—The cartesian coordinates for GS_{Ia} and GS_{III} (Table 3), Table 2 (listing energies and bond gaps for 12 of the 27 randomly selected starting conformations), and Figure 1 (illustrating the course of energy minimization for one of these conformations) have been deposited with the

[1] Nemethy, G., and H. A. Scheraga, *Biopolymers*, **3**, 155 (1965).

[2] Scheraga, H. A., S. J. Leach, R. A. Scott, and G. Nemethy, *Disc. Faraday Soc.*, **40**, 268 (1965).

[3] Scheraga, H. A., R. A. Scott, G. Vanderkooi, S. J. Leach, K. D. Gibson, T. Ooi, and G. Nemethy, *Conformation of Biopolymers*, ed. G. N. Ramachandran (New York: Academic Press, 1967), p. 43.

[4] Scott, R. A., and H. A. Scheraga, *J. Chem. Phys.*, **45**, 2091 (1966).

[5] Ooi, T., R. A. Scott, G. Vanderkooi, and H. A. Scheraga, *J. Chem. Phys.*, **46**, 4410 (1967).

[6] Gibson, K. D., and H. A. Scheraga, these PROCEEDINGS, **58**, 420 (1967).

[7] *Ibid.*, p. 1317.

[8] Davidon, W. C., A.E.C. Research and Development Report, ANL-5990 (1959); R. Fletcher and M. J. D. Powell, *Computer J.*, **6**, 163 (1963).

[9] Vanderkooi, G., S. J. Leach, G. Nemethy, R. A. Scott, and H. A. Scheraga, *Biochemistry*, **5**, 2991 (1966).

[10] Edsall, J. T., P. J. Flory, J. C. Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, *Biopolymers*, **4**, 121 (1966); *J. Biol. Chem.*, **241**, 1004 (1966); *J. Mol. Biol.*, **15**, 399 (1966).

[11] Wilde, D. J., *Optimum Seeking Methods* (Englewood Cliffs, N.J.: Prentice-Hall, 1964), p. 134.

[12] Liquori, A. M., P. de Santis, A. L. Kovacs, and L. Mazzarella, *Nature*, **211**, 1039 (1966).

[13] Schmidt, G. M. J., D. C. Hodgkin, and B. M. Oughton, *Biochem. J.*, **65**, 744 (1957).

[14] Wallwork, S. C., *Acta Cryst.*, **15**, 758 (1962).

[15] Schwyzer, R., *Rec. Chem. Progr.*, **20** (3), 147 (1959).

[16] Hodgkin, D. C., and B. M. Oughton, *Biochem. J.*, **65**, 752 (1957).