JOURNAL OF COMPUTATIONAL PHYSICS 12, 491–497 (1973)

Minimization of Polypeptide Energy. XII. The Methods of Partial Energies and Cubic Subdivision*

G. M. CRIPPEN† AND H. A. SCHERAGA‡

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

Received October 24, 1972

Two techniques are described for finding relatively low minima of the energy of a polypeptide. The first, the method of "partial energies," in which successively more components of the total energy function are included in the minimization procedure, is applicable primarily to empirical conformational energy calculations. The second, the method of "cubic subdivision," in which high energy regions are eliminated from the space to be searched, is a more general global minimization algorithm. The results of both methods on "tetraglycine" are presented, and trials on other problems are discussed.

INTRODUCTION

A major problem in conformational energy calculations on polypeptides and proteins is the existence of many minima in the potential energy surface, thereby making it very difficult to locate the global minimum [1]. Therefore, much attention has recently been devoted [2–5] to the development of methods for searching for the global minimum of the energy of a polypeptide. These methods have proved successful for molecules whose energy depends on a small number of variables. However, these methods tend to waste time, by examining in great detail some regions which are unlikely to contain the global minimum. We present two new algorithms, those of "partial energies" and "cubic subdivision," which are more efficient in this respect. In principle, these two methods (and also previous ones [2–5]) could be applied to calculations on polypeptides of any size, with anticipated success in locating the global minimum, but practical restrictions on

* This work was supported by research grants from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312), from the National Science Foundation (GB-28469X1), and from Walter and George Todd.

† Predoctoral Fellow of the National Institute of General Medical Sciences, National Institutes of Health, 1967–1971. Present address: Room 831 HSW, Cardiovascular Research Institute, School of Medicine, University of California, San Francisco, California 94122.

‡ To whom requests for reprints should be addressed.

computer time and memory limit the thoroughness of the search, as governed by certain parameters (adjusted by the user) in each method. However, different algorithms give varying degrees of success for the same computational effort, depending on the particular problem at hand. The two methods to be described in this paper were outstandingly successful on two different types of problems, as will be shown.

METHODS

Partial Energies

Since the empirical energy function of a polypeptide is a very complicated one [1], we can simplify the search for the global minimum of the energy by initially including only the "most important" energy term, and then adding in the remaining terms in successive stages. By the most important energy term, we mean a portion of the total energy, the minimization of which can greatly affect the conformation of a large part of the molecule. Thus, by this criterion, nonbonded interactions are considered "unimportant," because only small alterations of most conformations are required to relieve atomic overlaps, even though the energy change in the process may be very large. The method of "partial energies" consists of the following steps: (a) Select many starting conformations, perhaps at random, if there is no better way; (b) starting with these, minimize only the "most important" energy term for each conformation; (c) starting with the conformations obtained in the previous step, minimize the previous energy term plus the next most important one; (d) iterate step (c) until the entire energy function is used in the minimization procedure. In conformational energy calculations on polypeptides, the energy terms (in decreasing order of importance) are considered to be: (a) disulfide bond stretching, bending, and torsion (to insure the proper closing of cystine bridges); (b) hydrogen bonding; (c) electrostatic interactions; (d) nonbonded interactions; and (e) hydration energy and intrinsic rotational potentials. This ordering is chosen because gross structural alterations may well be required to close disulfide bridges, but then, smaller changes are needed to form hydrogen bonds, after which even smaller changes are necessary to optimize the electrostatic energy, and so on.

The algorithm begins by first properly connecting the disulfide bridges, which is an easy task. Then, keeping the disulfide bridges formed, the best hydrogen bonds possible are formed. Subsequently, slight rearrangements are introduced to improve the electrostatic interactions. Finally, small alterations are made to relieve any atomic overlaps. The basic idea is that a structure must have well closed disulfide bridges and good hydrogen bonding in order to be of especially low energy; the relatively intricate demands of the nonbonded and electrostatic interactions can

be met without drastic alterations of the basic structure, so that their inclusion at the start would complicate the problem unnecessarily.

It may be noted that a different view is taken in the energetic refinement of the x-ray coordinates of proteins [6]. There, the nonbonded energy (with the inclusion of hydrogen-bonding and disulfide-bond potentials) is first minimized to reduce atomic overlaps.

Cubic Subdivision

The basic idea in this procedure is to eliminate high-energy regions, with as little computer time as possible, and then to search the remaining low-energy regions more thoroughly. In the earlier "spotting" method [4], this was done by accurately locating surfaces of constant energy, which required numerous energy evaluations at points far from the global minimum; in "cubic subdivision," the policy is to spend as little time as possible examining unpromising regions.

A molecule (of fixed bond lengths and bond angles) with n bonds, about which rotation (from 0 to 2π radians) can take place, has a conformational space which may be regarded as an n-dimensional hypercube, each of whose edges in 2π radians in length. For a dipeptide, this is simply the familiar ϕ-ψ map [1, 2]. The energy is evaluated at two diagonally opposite corners of the hypercube (although, of course, these first two energies must be the same, since the two points correspond to the same conformation). For the dipeptide ($n = 2$), these are the lower left and upper right corners. In general, for any n, the first corner is the one having each coordinate algebraically as small or smaller than the corresponding coordinate of any other corner, and the second corner is the one with the largest coordinates. This choice of diagonal and the implicit choice of origin are arbitrary, and the use of other diagonals or origins might affect the results of the algorithm, particularly when applied to difficult problems; however, these effects have not been investigated. After evaluating the energies at these two points, the n-dimensional cube is then subdivided into 2^n smaller cubes, each having edges whose length is half of that in the original cube. The energies are then evaluated at two diagonally opposite corners of each subcube, using the same (above) criterion to pick the first and second corners of each subcube. Once the energy of any point (which may be a corner of more than one subcube) has been calculated, it is remembered as long as that information may be of any use, and the lowest energy so far found (the "interim global minimum") is remembered. According to various criteria, some cubes are eliminated from further consideration as being unlikely to contain points of energy lower than that at the interim global minimum, as discussed below. The remaining cubes are further subdivided, as before, and so on, until there remain only a few cubes of very small size, e.g., cubes whose edges are 10° long. The centers of these cubes are taken to be the starting points for ordinary energy minimization,

and the resulting structures should be of very low energy, and hopefully include the global minimum.

In order to devise a criterion for eliminating a cube from fruther considerations, we assume that the function being minimized is continuous, has continuous first and second derivatives, and, further, that there is an upper bound on the second derivative throughout the space being searched. This prevents the energy (which must have the computed values at the opposite corners of a cube) from dropping too low along the diagonal of the cube. In fact, there are singularities in the energy of a polypeptide, so that the function is not continuous everywhere, and the second derivatives are not bounded throughout. Nevertheless, in practice, the method works successfully on such systems. In addition, we assume that the energy at any point within a cube cannot be lower than the lowest value that a quadratic function could attain along the diagonal line connecting the two corners of the cube where the energy has been evaluated. This lowest possible energy is finite, since the second derivative of the energy is assumed to be bounded, and this assumption (based on a quadratic approximation) becomes increasingly accurate as cubes are further subdivided, and become-smaller and smaller. In fact, the lowest possible energy varies as the square of the cube edge. Each cube is tested to see if it is possible to find an energy along the diagonal line (without exceeding the upper bound on the second derivative) lower than the "interim global minimum;" if not, that cube is eliminated. If this process is allowed to go to completion, only one small cube of low energy will be given as the location of the global minimum. However, if the calculation is stopped somewhat before that, in general, there will be several cubes remaining (corresponding to the location of the few lowest minima), which is the more useful result for calculations on polypeptides.

In practice, "cubic subdivision" is a fast and thorough way of finding the global minimum, since no derivatives are required, and much of conformation space can be eliminated fairly quickly. The major difficulty with this method is the large number of quantities that have to be stored in the memory of the computer. If the original cube in the n-dimensional space represents the 0th subdivision, then after the mth subdivision (assuming that there have been *no eliminations*),

$$\text{number of points at which the energy has been evaluated} = N_c = (2^m + 1)^n - (2^n - 2), \tag{1}$$

$$\text{number of cubes} = N_x = 2^{mn}, \tag{2}$$

$$\text{maximum distance between points in the same cube} = r = 2\pi n^{1/2} 2^{-m} \text{ radians}, \tag{3}$$

and the lowest-energy point will have been located to an accuracy of

$$d = 2\pi(2^{-m}). \tag{4}$$

Hence, in a search, there will have been at most N_c energy evaluations and N_c conformations remembered (neglecting occasional m-fold duplication of some points). The global minimum will be located to an accuracy of d, and the energy function is assumed quadratic over a range r. The higher the stated value of the upper bound on the second derivative, the more thorough the search; the lower the bound, the more cubes will be eliminated in earlier iterations.

Results

Partial Energies

The method of "partial energies" was applied to "tetraglycine," the same six-variable molecule treated previously by the "spotting" algorithm [4]. Since this molecule contains no disulfide bonds, the "most important" energy term is the hydrogen-bonding one. Each CO group was allowed to interact with each NH group, except for O and H atom pairs, whose relative positions are affected by rotation about only one intervening single bond (so-called 1–4 interaction). The hydrogen-bonding energy function used [4] contains both an electrostatic and a nonbonding energy term for the interaction between the C′, O, N, and H atoms. No solvent interaction or intrinsic rotational potentials were included, so that comparisons could be made with the results obtained with the "spotting" technique [4]. Using exactly the same rigid geometry (fixed bond lengths, bond angles, and planar trans peptide groups) and energy parameters as before [4], and 20 different conformations A, chosen at random, the hydrogen bonding energies of these starting conformations were minimized to yield final conformations B. Of the 20 B's, seven were duplicates of the 13 others C, so that the seven were discarded. The total energies (hydrogen-bonding plus all other electrostatic and nonbonding energies) of these 13 conformations C were then minimized to yield 13 different conformations D.

The changes in dihedral angles in going from C to D were generally 20° or less, but sometimes there were movements of over 60°. Hence, the addition of electrostatic and nonbonded energy terms occasionally induces more than minor alterations in the structure. The energies of the 13 final conformations D, ranged from +5.79 to −3.57 kcal/mole (with a mean of +0.22 and a median of +0.12 kcal/mole). The lowest-energy conformation (−3.57 kcal/mole) was $(\phi_1, \psi_1, \phi_2, \psi_2, \phi_3, \psi_3) = (-47°, 113°, 48°, 35°, 118°, -33°)$, which has about three partial hydrogen bonds. Approximately one hour of computer time on the IBM 360/65 computer was required for the whole procedure.

Cubic Subdivision

The method of "cubic subdivision" was tested first on an artificial "energy" function of two variables,

$$E(z_1, z_2) = \frac{1}{2.0001 + \cos z_1 + \cos z_2} + \frac{0.5}{3.01 + \cos z_1 + 2 \sin z_2} + \sin 3z_1 + \sin 3z_2, \quad (5)$$

which has 9 minima, 7 maxima, and 2 very high maxima resembling the infinite peaks caused by atomic overlaps. The algorithm successfully located the two lowest minima (within 10°, after only 136 function evaluations) by arbitrarily assigning the upper limit of the second derivatives as 10. Ordinary mapping would have required $36^2 = 1296$ energy evaluations for similar accuracy. Indeed, on this and similar problems, cubic subdivision far excels any other general global minimization algorithm tried in this laboratory [2–5]. When the method was applied to tetraglycine with the same geometry and energies as used before [4] (adjusting the upper bound of the second derivative to 20 kcal/radian^2, so that no more than 10,000 points had to be remembered), only one hypercube having edge of length 10° remained after one hour on the IBM 360/65 computer. Using this as the starting point for local minimization, a minimum which had an energy of −0.85 Kcal/mole was reached.

Discussion

Although the method of "cubic subdivision" works remarkably well on two-dimensional problems, it did not yield as low an energy minimum as the method of "partial energies" did in the same amount of computer time applied to tetraglycine. The known lowest-energy conformations, found by "spotting," and "partial energies," are located in subcubes which, unfortunately, were eliminated. On the other hand, the low-energy structure (−0.85 kcal/mole) obtained by "cubic subdivision" had a lower energy than most of the conformations found by the method of partial energies.

The best structure of tetraglycine found by "spotting" [4] had an energy of −3.33 kcal/mole, while the best one found by the method of partial energies (see Results section for conformation) had an energy of −3.57 kcal/mole. Of all the low-energy conformations found by both methods (spotting and partial energies), only one was common to both sets of computations, viz., the one with $(\phi_1, \psi_1, \phi_2, \psi_2, \phi_3, \psi_3) = (-71.5°, 64.7°, -71.5°, 65.0°, 71.7°, -64.5°)$ and an energy of −2.50 kcal/mole. With the exception of the lowest-energy structure

(−3.57 kcal/mole), the low-energy conformations (−3.33 to −2.50 kcal/mole) found by either method (spotting or partial energies) were essentially combinations of $(\phi, \psi) = (+60°, -60°)$ and $(-60°, +60°)$, which are axial and equatorial seven-membered hydrogen-bonded rings, respectively, for each of the three glycyl residues. The regular α-helix becomes the preferred structure only with longer polypeptide chains.

Conclusion

The method of "cubic subdivision" seems to be a good, very general technique for problems with six or fewer variables. The method of "partial energies" is apparently a better one for larger structures, for which almost all possible hydrogen-bonding schemes can be found, and may prove useful for small polypeptides, such as gramicidin S, oxytocin, vasopressin, etc. For larger molecules, it will be necessary to have a better initial estimate of the native conformation (perhaps by techniques based on near-neighbor interactions [1]) before the method of partial energies can be applied.

References

1. H. A. Scheraga, *Chem. Revs.* **71** (1971), 195.
2. G. M. Crippen and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **64** (1969), 42.
3. K. D. Gibson and H. A. Scheraga, *Comput. Biomed. Res.* **3** (1970), 375.
4. G. M. Crippen and H. A. Scheraga, *Arch. Biochem. and Biophys.* **144** (1971), 453.
5. G. M. Crippen and H. A. Scheraga, *Arch. Biochem. and Biophys.* **144** (1971), 462.
6. P. K. Warme and H. A. Scheraga, *J. Computational Phys.* **12** (1973), 49.