Evaluation of Two Procedures for Selecting Starting Conformations for Energy Minimization of Peptides

Richard B. Davidson and S. Scott Zimmerman*

Department of Chemistry and Graduate Section of Biochemistry, Brigham Young University, Provo, Utah 84602

Received November 12, 1993®

In an effort to develop a computationally efficient method of finding low-energy minima of blocked dipeptides, we have evaluated two methods for generating starting conformations. The "representative" method involves (1) plotting points representing the locations of single-residue energy minima on two-dimensional conformation maps with the dihedral pairs ϕ–ψ, χ^1–χ^2, χ^3–χ^4, etc., as coordinates, (2) choosing a representative dihedral angle pair from each of the clusters of points on these maps, and (3) combining these representative single-residue points to form dipeptide starting conformations. The "concatenation method" uses all combinations of single-residue minima. Applying these two methods to blocked His-Arg and to blocked Asp-His, we found that (1) both methods successfully located what we assume to be the global minimum, (2) the two methods yielded essentially the same statistical averages of conformational properties, and (3) many low-energy minima located by each method were not located by the other method, although the representative method located many more low-energy minima than did the concatenation method.

INTRODUCTION

The general procedure for finding minimum-energy conformations in molecular modeling of peptides is to select a set of starting conformations and then, from those starting points, to minimize the total conformational energy as a function of the dihedral angles or of the atomic coordinates. Selecting starting conformations, however, is not straightforward, except for the simplest of molecules, because of the huge volume of conformational space. Leach[1] has classified the methods of selecting starting conformations into five groups: systematic methods, model-building methods, random search methods, distance geometry methods, and molecular dynamics. In this paper, we describe and compare two model-building methods for selecting starting conformations. Our goal is to test and improve current methods of finding as many peptide minimum-energy conformations as possible.

Typically, the conformational space of peptides has broad enough energy wells that, in concept, as exhaustive search for the global minimum and the other minima can be made by systematically selecting starting points at intervals of 10–25° in all the variable backbone dihedral angles and 40–90° in all the side chain angles. This approach, however, results in so many starting points that such an exhaustive search of conformational space is not feasible with present computer speeds, except for the blocked single residues (that is, the *N*-acetyl-*N'*-methylamino acid amides) of Gly, Ala, and Pro, that is, those amino acids with the fewest number of variable dihedral angles.

It is not necessary, however, to include all of conformational space in a systematic search, because large regions of ϕ–ψ space are high enough in energy to be considered forbidden (that is, they have extremely low probabilities of occurrence). For example, the ϕ–ψ conformational energy contour map of *N*-formyl-*N'*-methylalaninamide (f-Ala-NMe), shown in Figure 1, reveals large forbidden regions in the ranges of $\phi = 0° \pm 40°$, $\phi = 130° \pm 50°$, and $\psi = -120° \pm 30°$. In selecting starting points, one could ignore these forbidden regions and limit the starting points to 10–25° intervals within the allowed regions (for example, regions within 15 kcal/mol of the global minimum). Such an approach would yield, upon minimization of these starting conformations, all the important minima for even the largest blocked single residues, like Arg and Lys, without excessive computational times.

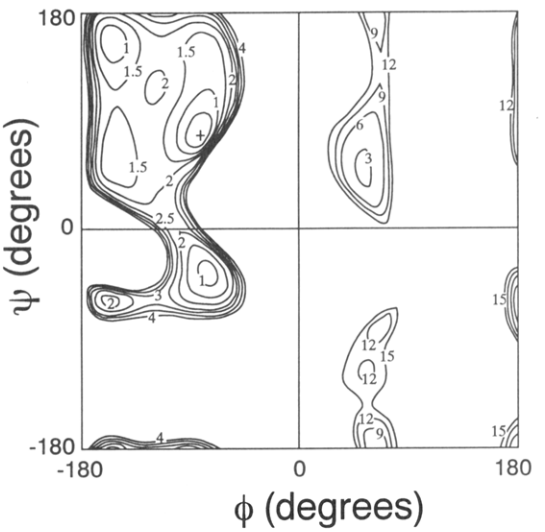

Figure 1. Conformational energy contour map of *N*-formyl-*N'*-methylalaninamide in the ϕ–ψ plane relative to the global minimum marked with "+". The calculated absolute energy E_0 at the global minimum is –1.39 kcal/mol. To show the details of conformational space in lowest-energy regions of the ϕ–ψ plane, contour lines on the left half of the map are drawn at 0.5 kcal/mol increments up to 4 kcal/mol above E_0; to show details in the high-energy regions, contour lines on the right half of the map are drawn at 3.0 kcal/mol increments up to 12 kcal/mol above E_0.

The location and relative energies of the minima of the blocked single residues of the 20 naturally occurring amino acids are well characterized. For example, Vásquez et al.[2] used ECEPP/2[3] to compute the minima of all 20 naturally occurring amino acids. For small blocked single residues of a-Gly-NMe, a-Ala-NMe, and a-Pro-NMe, Vásquez et al.[2] presented data on *all* minima, regardless of their relative energies; for the large blocked amino acids of Arg and Lys, they presented only those minima within 3 kcal/mol; and for

® Abstract published in *Advance ACS Abstracts,* May 15, 1994.

0095-2338/94/1634-1009$04.50/0

all the other blocked amino acids, they presented only those minima within 5 kcal/mol. These data have proven invaluable in our studies of di- and tripeptides. Our experience suggests, however, that excluding minima with energies above 5 kcal/mol in Arg- and Lys-containing peptides results in failure to find all the important low-energy minima in the peptides. When necessary, therefore, we have extended the data to include minima up to 15 kcal/mol above the global minimum.

In spite of the large body of published work on molecular modeling of blocked amino acids and peptides, no one has developed a totally satisfactory method for selecting starting conformations. The difficulty lies in trying to satisfy the following three criteria simultaneously:

1. Computation times for minimizing the total conformational energy of all starting conformations must be reasonable. In the absence of unlimited time on supercomputers, molecular modeling studies must somehow restrict the number of starting conformations.

2. At least one of the starting conformations must, upon energy minimization, lead to the global minimum. Finding the global minimum is, in almost any kind of conformational analysis study, a requirement fundamental to determining the conformational properties of the molecule.

3. The chosen starting points should also lead to an ensemble of minima that is statistically representative of the whole population relative to the conformational properties deemed important to the aims of the research. (Earlier work demonstrated the validity of using local minima as a representative sampling over all conformational space for statistical analyses.[5]) Such statistical properties might include probabilities of hydrogen bond formation, probabilities of β-bend formation, average end-to-end distances, or the proportion of the sample found in a given conformation.

A mathematically rigorous guarantee of satisfying these criteria is not possible. From a consideration of Boltzmann statistics, however, it is clear that, for accurate statistical mechanical results, the ensemble of minima must include a great majority of those minima with total conformational energies within 3 kcal/mol of the global minimum, and should presumably include the global minimum itself.

One common method for selecting starting conformations for energy minimization of small peptides is to combine the low-energy minima of the blocked single residues that make up the peptide. We call this procedure the "concatenation method". Vásquez and Scheraga[4] employed this method as part of their "buildup" approach to find minima of oligo- and polypeptides. While the concatenation method (which for dipeptides is essentially the same as the build-up method) has proven successful in locating all low-energy minima for several Ala-, Gly-, and Pro-containing dipeptides,[5] we will show in this paper that, for some molecules—specifically, blocked His-Arg and blocked Asp-His, the concatenation method fails to find many of the minima within 3–4 kcal/mol of the global minimum. We will also describe in this paper a new, more comprehensive method, which we call the "representative method", for selecting starting conformations. We apply both the concatenation and the representative methods to the large blocked dipeptides His-Arg and Asp-His and then compare and contrast the results obtained from the two methods.

METHODS

Throughout this paper, we use the nomenclature and conventions adopted an IUPAC-IUB Commission.[6] The blocking groups for the dipeptides in this study were acetyl and formyl at the amino terminus and methyl amide at the carboxyl terminus. In this paper these blocking groups are designated a-, f-, and -NMe, respectively. Thus, f-His-Arg-NMe is *N*-formyl-*N'*-(methylhystidyl)argininamide.

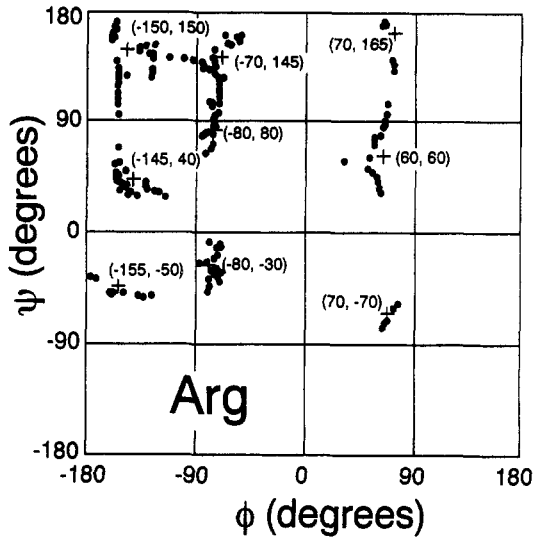

Figure 2. Locations of minima in the ϕ–ψ plane of f-Arg-NMe. Nine points (marked by "+") were selected to represent the clusters of minima in each region. It is instructive to compare this diagram with Figure 1.

Molecular Modeling Force Field. The force field employed to calculate all the conformational energies was based on the potential energy equations, parameters, geometries, and minimization method of ECEPP/2.[3] We used the default dialectric constant for ECEPP/2, which has an effective value of 4. No solvation was included in the calculations. The minimization algorithm is MINOP,[7] with a convergence criterion of 0.01 kcal/mol·rad). When minimization was complete (due to convergence), the resulting conformation was assumed to be a minimum in conformational space, without further testing.

Software. The software used to calculate the conformational energies is the nongraphical portion of PepCAD,[8] a computer program written in C for easy porting across various hardware platforms. PepCAD uses the potential energy equations, parameters, and geometries of ECEPP/2. The results of energy minimization using PepCAD yield a minimum-energy conformation with a total conformational energy E. The value of E for the assumed global minimum of each peptide is designated as E_0. The relative energy of each minimum is defined as $\Delta E = E - E_0$.

Hardware. Computations were performed on a Sun Microsystems SPARC Station 1 and on an IBM RISC 6000 machine.

Contour Maps and Single-Residue Minima. To create the conformational energy contour maps, we calculated the conformational energies using PepCAD and plotted the results using CONTOUR.[9] From the contour maps, the minimum-energy regions of ϕ–ψ were easily located. We followed the method of Vásquez et al.[2] to calculate the minima for the blocked single residues of Arg and Lys and used the data of Vásquez et al.[2] for the minima of the blocked single residue of Asp.

Representative Method. A plot of the locations of single-residue minima in the ϕ–ψ plane of conformational space reveals that the minima are clustered together in a few areas, as shown in Figure 2, which is a plot of the locations of the minima of f-Arg-NMe in the ϕ–ψ plane. Plotting the minima in planes generated by pairs of side-chain dihedral angles such as χ^1–χ^2 or χ^3–χ^4 show that the individual angles are

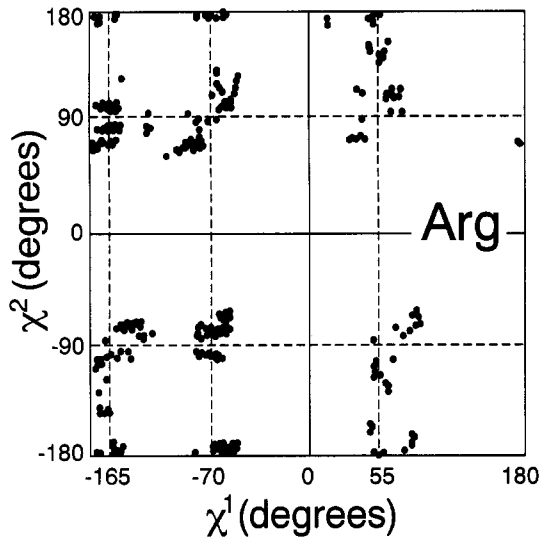

Figure 3. Locations of minima in the χ^1–χ^2 plane of f-Arg-NMe. The minima all lie near $\chi^1 = -165°$, $-70°$, and $55°$ and near $\chi^2 = -180°$, $-90°$, and $+90°$. (Note that $+180°$ and $-180°$ are identical angles.)

restricted to relatively small intervals of χ. The χ intervals for the peptide f-Arg-NMe are shown in Figure 3. The existence of these clusters and intervals suggests another method of selecting starting points, which we call the representative method. In the representative method, we select one point to represent all the points in each densely populated region of the blocked single-residue ϕ–ψ map. (Although these different minima are close together in one plane of conformational space, they are not close together in other planes; each point is a distinct minimum in multidimensional conformational space.) Thus, we have chosen nine representative ϕ–ψ values for Arg, one for each of the nine clusters in Figure 2, and 11 for Asp and His, one for each of the 11 clusters in the ϕ–ψ maps of Asp and His (not shown). In the case of the Asp χ^1 and χ^2 angles, the minima are not as well grouped as the side-chain angles of Arg and His. For this reason a series of starting angles at 60° intervals was chosen to represent the χ^1 and χ^2 minimum-energy values.

The representative method, therefore, differs from the concatenation (build-up) method in the following way. The representative method uses all combinations of *representative* ϕ, ψ, and χ values of single residues as starting conformations for the dipeptides, independent of which combinations of those dihedral angles actually lead to single-residue minima. Many of those combinations are, in fact, single-residue minima, but many of them are not. The concatenation method, on the other hand, uses only combinations of single-residue minima and therefore does not generate as many starting conformations.

Screening Procedure. Both the concatenation and the representative methods can yield starting conformations with physically impossible overlapping atoms or bonds. Attempting a complete minimization of these starting conformations not only wastes computer time but in some cases yields such a high energy that its numeric value exceeds the highest possible value on the computer system. Therefore, we have developed a screening procedure to discard the unpromising starting conformations.

The screening procedure consists of relaxing the minimization convergence criterion from the normal 0.01 kcal/(mol·rad) to a higher value of 1.0 kcal/(mol·rad) for the gradient and limiting the maximum number of iterations (energy-determining function calls) to 15. Typically, from 20 to 50% of the starting conformations will be processed to convergence in this first pass. A few of the conformations will typically yield an initial energy that exceeds the highest possible value represented by the floating-point hardware of the computer system. These conformations are discarded. The remaining unconverged minima are sorted by energy, and duplicate conformations are removed. This completes the screening procedure. The most valuable feature of the screening procedure is that about 20% of the starting points converge to conformations that are duplicated. By removing these redundancies and discarding conformations with excessively high energies, we can save considerable time in subsequent processing. We then process all the remaining conformations in a second pass, using a convergence criterion of 0.01 kcal/(mol·rad) and allowing up to 500 iterations. Both the screening procedure (on the SPARC station) and the final processing (on the RISC 6000) of 5000 conformations requires 12–18 h of computer time. CPU time for convergence from a given starting point to a relative minimum depends roughly linearly on the number of dihedral angles but also depends on the distance between the starting conformation and the minimum. We made no effort to determine whether one method surpassed the other in the ability to select starting points close to relative minima, but there is no reason to believe that one method is better than the other on the basis of convergence time.

Table 1. f-His-Arg-NMe Conformations with $\Delta E \leq x$

concatenation method starting 130 977		representative method starting 513 216	
x	n	x	n
1	4	1	4
2	38	2	44
3	253	3	341
4	958	4	1 275
5	2 725	5	3 797
6	6 191	6	8 789
7	11 755	7	16 487
8	19 094	8	26 326
9	27 222	9	36 959
10	35 963	10	47 683
11	44 681	11	57 892
12	53 411	12	67 595
13	61 822	13	76 930
14	69 585	14	85 326
15	76 682	15	92 668

Statistical Methods. Boltzmann statistics were used to determine the probabilities of distribution of conformations, bend probabilities, and hydrogen bond frequencies, following the methods described in Zimmerman et al.[5] The approximation employed here does not take into account differences in the breadth of the minima, but earlier results[5] showed that this is a reasonable approximation.

RESULTS AND DISCUSSION

Number of Conformations. Both the concatenation and the representative methods located the identical lowest-energy conformation, which we assume is the global minimal. Table 1 lists the numbers of conformations of f-His-Arg-MNe that were located within specified energy ranges relative to the energy E_0 of the global minimum. Both methods located 4 minima within 1 kcal/mol of the global minimum, but the representative method located a total of 341 minima within 3 kcal/mol of the global minimum while the concatenation method located only 253 within the same energy range. The number of minima found by the representative method exceeds the number found by the concatenation method at all energy levels (see Table 1). Within 15 kcal/mol of E_0 the concat-

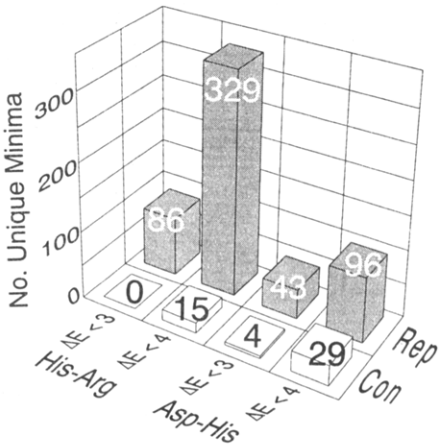

Figure 4. Number of unique minimum-energy conformations located by the representative ("Rep") and concatenation ("Con") methods for energies within 3 and 4 kcal/mol of the global minimum for the blocked dipeptides His-Arg and Asp-His.

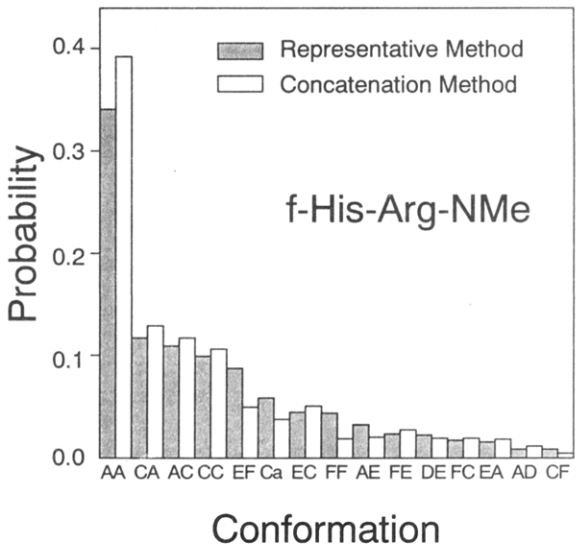

Figure 5. Calculated probabilities of backbone conformations of f-His-Arg-NMe determined by the concatenation and representative methods.

enation method located a total of 76 682 minima, while the representative method located a total of 92 668 minima. The concatenation method, however, is more efficient in view of the fact that 59% of the starting conformations lead to unique minima within 15 kcal/mol of E_0 compared with only 18% for the representative method.

The same general conclusions result when we compare the two methods in locating minima of f-Asp-His-NMe. The representative method located 319 minima within 3 kcal/mol of E_0 and a total of 5394 minima within 15 kcal/mol of E_0, while the concatenation method found only 280 minima within 3 kcal/mol of the E_0 and a total of 6242 minima within 15 kcal/mol of the E_0. Again, the concatenation method was more efficient in finding minima; 63% of its starting conformations lead to unique minima within 15 kcal/mol of E_0 compared with only 9% for the representative method.

Unique Conformations Found by Each Method. Figure 4 shows the number of f-His-Arg-NMe and f-Asp-His-NMe low-energy minima successfully located by one method but which the other method failed to locate. Within 3 kcal/mol of the energy E_0 of the assumed global minimum, the representative method found all the f-His-Arg-NMe minima that were found by the concatenation method plus 86 additional minima. Thus, the concatenation method found fewer than 75% of all minima within 3 kcal/mol of E_0 for this dipeptide. Within 4 kcal/mol of E_0, the representative method found 329 minima not found by the concatenation method, while the concatenation method found 15 minima not found by the representative method. In the case of f-Asp-His-NMe, the representative method failed to find 4 minima within 3 kcal/mol of the energy E_0 of the global minimum that were found by the concatenation method but found 43 that were overlooked by the concatenation method. This means that the concatenation method found fewer than 88% of the f-Asp-His-NMe minima within 3 kcal/mol of E_0.

Why does the representative method fail to find certain low-energy minima? In an attempt to answer this question we compared the locations in conformational space of the minima within 4 kcal/mol of E_0 that were found by the concatenation method but not found by the representative method. All these minima have dihedral angles that lie within the clusters of minima near the sites that were used in the representative method. Thus, this comparison failed to reveal any reason why the representative method did not find all the minima. There is, of course, no theoretical reason to expect that either method would yield all the relative minima. In fact, there is no way of knowing what percent of all the local minima has been found—or if the global minimum has been found, for that matter—short of using an exhaustive grid search of conformational space. Both methods failed to find all the global minima simply because conformational space is larger and contains more minima than are accounted for by either of the approximation methods. Given the number of dimensions of conformational space (equal to the number of variable dihedral angles) and the number of intervals required for each angle (e.g., at 20° intervals), the problem of finding all local minima seems intractable except for very small peptides. It is reassuring that both the concatenation and the representative methods independently yielded the same conformation of lowest energy, but we have no way of knowing with certainty that it is the global minimum.

Statistical Analysis. For the sake of discussion it is convenient to designate various regions of conformational space in the ϕ–ψ plane by letter codes.[10] With this convention, the backbone of a single residue consisting of an angle ϕ of –75° and an angle ψ of –65° would be in the approximate center of region A; region F has a center at about $\phi = -75°$, $\psi = 150°$, and region D is centered at $\phi = -145°, \psi = 55°$. Regions designated by lower-case letters are centrosymmetrically disposed relative to those with corresponding upper-case letters. Each dipeptide conformation has a pair of letter codes, for example FD, to describe the backbone conformations of the first and second residues.

With use of Boltzmann statistics, the predicted probabilities of finding the peptides we studied in their various conformations from one method or the other do not appear to be significantly different. Figure 5 compares the results of the analysis for the two methods on the basis of the probabilities of occurrence of backbone conformations of f-His-Arg-NMe. The distribution of conformations generated using the representative method is slightly broader than that generated by the concatenation method, at least partly because more conformations are found by the representative method. The same observation is true of the hydrogen bond frequencies (not shown). The most striking observation in Figure 5, however, is the similarity in the distribution of calculated minima generated by both the representative and concatenation methods. In terms of probabilities of occurrences of

Table 2. Calculated Average R Values and Bend Probabilities

	concatenation method		representative method	
peptide	av R	probability	av R	probability
f-Asp-His-NMe	6.8544	0.5658	6.9024	0.5377
f-His-Arg-NMe	6.6706	0.5647	6.7804	0.5313

conformations in each region of ϕ–ψ space, the two methods yield very similar results, even though each method failed to find all the low-energy minima.

Both dipeptides f-Asp-His-NMe and f-His-Arg-NMe are calculated to be most likely found in the AA conformation. In f-Asp-His-NMe, the AA conformation is usually stabilized by one hydrogen bond, occasionally by two hydrogen bonds, and in rare cases by three hydrogen bonds. In f-His-Arg-NMe, the calculated global minimum has the backbone conformation AA, but in f-Asp-His-NMe, the assumed global minimum has the conformation FD. Despite the low energy of this particular conformation, FD appears less frequently among all minima, so it ranks as the second most likely backbone conformation of f-Asp-His-NMe.

Conformation AA is considered a bend because the distance R between the hydrogen atom of the formyl blocking group and the carbon of the N-methyl blocking group is $\leq$0.7 nm. Table 2 summarizes the statistical mechanical analysis of the dipeptides with regard to distance R and the bend probabilities. The statistical probability that a randomly chosen conformation will exhibit a bend is about 50% for both dipeptides.

CONCLUSIONS

Because the concatenation method fails to locate many more low-energy minima than does the representative method, the latter method appears, in general, to be a more effective method of modeling peptides. If the purpose for generating low-energy minima of dipeptides is for use as starting conformations for energy minimization of tripeptides and larger molecules, then the representative method is significantly better than the concatenation method.

On the other hand, the errors resulting from inaccuracies in the force field and in the basic assumptions of our molecular modeling approach (such as the assumption that the peptides are in the vapor phase) are likely to be greater than the differences found between the two methods with regard to statistical averages (see Table 2 and Figure 5). This leads to the conclusion that the concatenation method, with its fewer starting conformations and greater efficiency in finding minima, is probably adequate for determining statistical averages.

The results of this study emphasize the difficulty in mapping conformational space. Except for an exhaustive grid search, which is impossible for all but the simplest molecules, no known method provides absolute certainty for finding the global minima or all local minima.

ACKNOWLEDGMENT

This work was partially supported by the Brigham Young University Development Fund and the Brigham Young University Department of Chemistry. We would also like to acknowledge University support for computer time on the Sun Microsystems SPARC Station 1 and on an IBM RISC 6000 machine.

REFERENCES AND NOTES

(1) Leach, A. R. A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules. In *Reviews in Computational Chemistry*; VCH Publishers: New York, 1991; Vol. 2, pp 1–55.
(2) Vásquez, M.; Némethy, G.; Scheraga, H. A. Computed Conformational States of the 20 Naturally Occurring Amino Acid Residues and of the Prototype Residue α-Aminobutyric Acid. *Macromolecules* **1983**, *16*, 1043–1049.
(3) Némethy, G.; Pottle, M. S.; Scheraga, H. A. Energy Parameters in Polypeptides. 9. Updating of Geometrical Parameters, Nonbonded Interactions, and Hydrogen Bond Interactions for the Naturally Occurring Amino Acids. *J. Phys. Chem.* **1983**, *87*, 1883–1887.
(4) Vásquez, M.; Scheraga, H. A. Use of Buildup and Energy-Minimization Procedures to Compute Low-Energy Structures of the Backbone of Enkephalin. *Biopolymers* **1985**, *24*, 1437–1447.
(5) Zimmermann, S. S.; Shipman, L. L.; Scheraga, H. A. Bends in Globular Proteins. A Statistical Mechanical Analysis of the Conformational Space of Dipeptides and Proteins. *J. Phys. Chem.* **1977**, *81*, 614–622.
(6) IUPAC-IUB Commision on Biochemical Nomenclature. Abbreviations and Symbols for the Description of Conformation of Polypeptide Chains. *J. Mol. Biol.* **1970**, *52*, 1–17.
(7) Dennis, J. E., Jr.; Mei, H. H. W. Two New Unconstrained Optimization Algorithms Which Use Function and Gradient Values. *J. Optimiz. Theory Appl.* **1979**, *28*, 453–482.
(8) Feller, D. C.; Delmoe, E. F.; Zimmermann, S. S. Conformational Energy Analysis of Peptides Using Microcomputers: Description of PepCAD and Analysis of N-Formyl-N′-Methylalanineamide. *Comput. Chem.* **1989**, *4*, 337–341.
(9) Davidson, R. B.; Zimmerman, S. S. CONTOUR, a General Contour-Plotting Program for IBM-Compatible Microcomputers, and Its Application to Peptides. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 174–177.
(10) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. Conformational Analysis of the 20 Naturally Occurring Amino Acid Residues Using ECEPP. *J. Phys. Chem.* **1975**, *79*, 2361–2381.