Displaying the Structure of Molecules by Multidimensional Plots of Their Torsion Angles

Brian T. Luke

IBM Kingston, MS 284, Neighborhood Road, Kingston, New York 12401

Received July 23, 1992

Since any molecular structure can be uniquely determined by a list of the atoms, their connectivity, and the value of all necessary torsion angles, these torsion angles can be used to describe the changes that occur during Monte Carlo, molecular dynamics, geometry optimization, or virtually any other type of simulation or to compare different conformations of the same or related molecules. This paper shows how three different multidimensional plots can be used to represent various conformations of the tetrapeptide Met-enkephalin. In particular, the construction and results of a parallel coordinate plot, a concentric coordinate plot, and a compound Ramachandran plot are described.

INTRODUCTION

In many applications of molecular modeling, it is important to compare the results of different simulations or computational approaches. Though descriptors such as root mean squared deviation and percent similarity can yield meaningful information and plots can show differences in the three-dimensional structure of two molecules,[1] they do not show exactly where, or how, two or more structures differ. For example, two structures may appear to be structurally very different, though they may actually differ in the value of a single torsion angle. To assist in this determination, three different procedures for displaying some or all of the torsion angles of a molecule are presented.

In general, the complete description of the geometry of a molecule is given by either a list of N atom labels (names or numbers) and $3N$ Cartesian coordinates or N atom names and $3N-6$ internal coordinates. In many cases the internal coordinates consist of $N-1$ bond lengths, $N-2$ bond angles, and $N-3$ torsion angles, though several other sets of internal coordinates will yield the same unique conformation. Instead, if the atom labels and their connectivity are given, the required bond lengths, bond angles, and some of the torsion angles can be accurately estimated. Therefore, the atom labels, their connectivity, and a reduced set of torsion angles can be used to uniquely describe the conformation of a molecule since a partial geometry optimization that freezes the values of this set of torsion angles will always give the desired structure. The problem of comparing two or more structures reduces to graphically displaying this reduced set of torsion angles in each molecule or conformation.

The following discussion will examine how multidimensional plots are used to describe conformations of the pentapeptide Met-enkephalin (HIN-TYR-GLY-GLY-PHE-MET-OOH). While the application of these plotting methods is confined to displaying the structures of a polypeptide, some of the procedures can be used for virtually any polymer composed of well-established monomer units, and in many cases can be applied to displaying the structure of any molecule.

The next section describes how the appropriate torsion angles are chosen for a polypeptide and contains a discussion of procedures for reducing the total number of torsion angles needed to uniquely define a conformation and how degeneracy in a given angle is removed. As described above, the examples of multidimensional plots will focus on structures of the pentapeptide Met-enkephalin. These structures are generated by running a torsion searching program called PEPTOR. Though full details of this program will be given in a later publication,[2] a truncated description of how the different conformations are generated is given in the third section.

The fourth section details how a parallel coordinate plot[3] can be used to display one or more conformations of this pentapeptide. The fifth and sixth sections contain similar discussions of a concentric coordinate plot and a compound Ramachandran plot, respectively. This is followed by an overall comparison of the different multidimensional plots.

TORSION ANGLES IN POLYPEPTIDES

Since a polypeptide is composed of a prescribed joining of well-established amino acid residues, a simple list of the amino acid sequence is sufficient to uniquely determine the atom labels and all the connectivity in the molecule. Therefore, beyond the residue list (e.g., HIN-TYR-GLY-GLY-PHE-MET for Met-enkephalin), all that is needed is a set of required torsion angles.

The IUPAC-IUB Commission on Biochemical Nomenclature[4] stated that the backbone conformation in polypeptides will be determined by three torsion angles, denoted by the Greek letters ϕ, ψ, and ω. In addition, the side chain angles will be denoted by the Greek letter χ followed by the suitable subscripts. This group of torsion angles can be further reduced if it is realized that many amino acids have side chains that contain rings. The value of a torsion angle T(A-B-C-D) will not change very much if atoms B and C are part of a ring. The same would be true of any nonstandard amino acid residue that contained a double bond between atoms B and C.

PEPTOR evaluates the energy of a polypeptide using the ECEPP/2 empirical potential,[5] and the above example of a rigid torsion angle corresponds to Professor Scheraga's definition of what is excluded from the list of "rotatable bonds" in the parameterization of the ECEPP potential. A complete list of the amino acid residues handled by PEPTOR and their rotatable side chain angles is given in Table I.

The last column in this table lists the degeneracy of each side chain angle. For example, $\chi(2)$ in ABU is 3-fold degenerate since C_g is the carbon of a methyl group and has local C3v symmetry. This means that if the torsion angle is increased or decreased by 120°, the structure will be identical. To represent this torsion angle as a single point on the subsequent plots, all values of a generate angle will be examined, and the one that is closest to 180° will be used. This choice is completely arbitrary, but leads to a data point that is in the center of a parallel coordinate plot.

0095-2338/93/1633-0135$04.00/0

Table I. Rotatable Side-Chain Angles for ECEPP/2 Amino Acid Residues Stored in PEPTOR

sym	name	χ	A-B-C-D	deg	sym	name	χ	A-B-C-D	deg
ABU	aminobutyric acid	1	N-Ca-Cb-Cg	1	HPU	H-Pro-up	1	Cg-Cg-Od1-Hd1	1
		2	Ca-Cb-Cg-Hg3	3	ILE	isoleucine	1	N-Ca-Cb-Cg1	1
ALA	alanine	1	N-Ca-Cb-Hb1	3			2	Ca-Cb-Cg1-Cd1	1
ARG	arginine$^+$	1	N-Ca-Cb-Cg	1			3	Ca-Cb-Cg2-Hg3	3
		2	Ca-Cb-Cg-Cd	1			4	Ca-Cg1-Cd1-Hd3	3
		3	Cb-Cg-Cd-Ne	1	LEU	leucine	1	N-Ca-Cb-Cg	1
		4	Cg-Cd-Ne-Cz	1			2	Ca-Cb-Cg-Cd1	1
		5	Cd-Ne-Cz-Nh1	2			3	Cb-Cg-Cd1-Hd3	3
		6	Ne-Cz-Nh1-Hh1	2			4	Cb-Cg-Cd2-Hd6	3
		7	Ne-Cz-Nh2-Hh3	2	LYS	lysine$^+$	1	N-Ca-Cb-Cg	1
ARX	arginine	1	N-Ca-Cb-Cg	1			2	Ca-Cb-Cg-Cd	1
		2	Ca-Cb-Cg-Cd	1			3	Cb-Cg-Cd-Ce	1
		3	Cb-Cg-Cd-Ne	1			4	Cg-Cd-Ce-Nz	1
		4	Cg-Cd-Ne-Cz	1			5	Cd-Ce-Nz-Hz3	3
		5	Cd-Ne-Cz-Nh1	1	LYX	lysine	1	N-Ca-Cb-Cg	1
		6	Ne-Cz-Nh1-Hh1	1			2	Ca-Cb-Cg-Cd	1
		7	Ne-Cz-Nh2-Hh2	2			3	Cb-Cg-Cd-Ce	1
ASN	asparagine	1	N-Ca-Cb-Cg	1			4	Cg-Cd-Ce-Nz	1
		2	Ca-Cb-Cg-Od1	1			5	Cd-Ce-Nz-Hz2	2
		3	Cb-Cg-Nd2-Hd2	2	MET	methionine	1	N-Ca-Cb-Cg	1
ASP	aspartic acid$^-$	1	N-Ca-Cb-Cg	1			2	Ca-Cb-Cg-Sd	1
		2	Ca-Cb-Cg-Od1	2			3	Cb-Cg-Sd-Ce	1
ASX	aspartic acid	1	N-Ca-Cb-Cg	1			4	Cg-Sd-Ce-He3	3
		2	Ca-Cb-Cg-Od1	1	NOR	norleucine	1	N-Ca-Cb-Cg	1
		3	Cb-Cg-Od2-Hd2	1			2	Ca-Cb-Cg-Cd	1
BAS	benzylasparatate	1	N-Ca-Cb-Cg	1			3	Cb-Cg-Cd-Ce	1
		2	Ca-Cb-Cg-Od2	1			4	Cg-Cd-Ce-He3	3
		3	Cb-Cg-Od2-Ce	1	ORN	ornithine	1	N-Ca-Cb-Cg	1
		4	Cg-Od2-Ce-Cz	1			2	Ca-Cb-Cg-Cd	1
		5	Od2-Ce-Cz-Cn2	2			3	Cb-Cg-Cd-Ne	1
CYS	cysteine	1	N-Ca-Cb-Sg	1			4	Cg-Cd-Ne-He2	2
		2	Ca-Cb-Sg-Hg	1	ORP	ornithine$^+$	1	N-Ca-Cb-Cg	1
CYX	cystine	1	N-Ca-Cb-Sg	1			2	Ca-Cb-Cg-Cd	1
GLN	glutamine	1	N-Ca-Cb-Cg	1			3	Cb-Cg-Cd-Ne	1
		2	Ca-Cb-Cg-Cd	1			4	Cg-Cd-Ne-He3	3
		3	Cb-Cg-Cd-Oe1	1	PHE	phenylalanine	1	N-Ca-Cb-Cg	1
		4	Cg-Cd-Ne2-He2	2			2	Ca-Cb-Cg-Cd1	1
GLU	glutamic acid$^-$	1	N-Ca-Cb-Cg	1	PRF	proline-flat			
		2	Ca-Cb-Cg-Cb	1	PRO	proline-down			
		3	Cg-Cg-Cd-Oe1	2	PRU	proline-up			
GLX	glutamic acid	1	N-Ca-Cb-Cg	1	SER	serine	1	N-Ca-Cb-Org	1
		2	Ca-Cb-Cg-Cd	1			2	Ca-Cb-Og-Hg	1
		2	Cb-Cg-Cd-Oe1	1	THR	threonine	1	N-Ca-Cb-Og1	1
		3	Cg-Cd-Oe2-Hde	1			2	Ca-Cb-Og1-Hg	1
GLY	glycine						3	Ca-Cb-Cg2-Hg3	3
HID	histidine	1	N-Ca-Cb-Cg	1	TRP	tryptophan	1	N-Ca-Cb-Cg	1
		2	Ca-Cb-Cg-Nd1	1			2	Ca-Cb-Cg-Cd1	1
HIE	histidine-2	1	N-Ca-Cb-Cg	1	TYR	tyrosine	1	N-Ca-Cb-Cg	1
		2	Ca-Cb-Cg-Nd1	1			2	Ca-Cb-Cg-Cd1	2*
HIP	histidine$^+$	1	N-Ca-Cb-Cg	1			3	Ce1-Cz-Oh-Hh	2*
		2	Ca-Cb-Cg-Nd1	1	VAL	valine	1	N-Ca-Cb-Cg1	1
HPD	H-Pro-down	1	Cb-Cg-Od1-Hd1	1			2	Ca-Cb-Cg1-Hg3	3
HPF	H-Pro-flat	1	Cg-Cg-Od1-Hd1	1			3	Ca-Cb-Cg2-Hg6	3

Glycine has no side-chain angles, and since the side chain of proline is part of a ring, ECEPP/2 assumes this residue has no rotatable side-chain angles. Since there is some motion in this five-membered ring, ECEPP/2 lets the user choose one of three different conformations for this residue and hydroxyproline.

The last point to mention about the side-chain angles is the χ(2) and χ(3) angles of tyrosine. Though these angles are nondegenerate when examined individually, if they both are changed by 180°, the structure will be unchanged. The convention adopted here is to make χ(2) as close to 180° as possible, adjusting χ(3) to leave the structure unchanged.

The backbone angles of the first and last residue may also have a degeneracy, depending upon what the amine and carboxy terminus residue groups are. For example, if the amine terminus is a protonated H_3N^+ group, the φ angle of the first residue will have a 3-fold degeneracy. The ECEPP/2 residue file has a nonplanar nitrogen if the amine terminus is simply an H_2N group, making the φ angle nondegenerate. If a different force field made this group planar, the φ angle would be 2-fold degenerate. If the carboxy terminus is a COO^- group, the ψ angle of the last residue will have a 2-fold degeneracy. Similarly, if the carboxy terminus group is a $-CH_3$, $-NH_2$, or $-N(CH_3)_2$, the ω angle of the last residue is 3-, 2-, or 2-fold degenerate, respectively. The convention adopted here is again to set all degenerate angles as close to 180° as possible.

GENERATING POLYPEPTIDE CONFORMATIONS

By using the definition of a "rotatable bond" given in the previous section, a conformation of Met-enkephalin is uniquely given by its amino acid sequence, HIN-TYR-GLY-GLY-PHE-MET-OOH, and the values of its 24 rotatable bonds. Therefore, the potential energy hypersurface of this pentapeptide can be scanned by varying these 24 torsion angles. A separate program, PEPTOR, has been written to do this using the ECEPP/2 potential.[5] Though this program can be used several different ways, and a full description of the

program will be given in a latter publication,[2] this section contains a brief description of PEPTOR and one particular application to Met-enkephalin.

In simple minimization procedures, the *n*-dimensional problem (24-dimensional in the case of Met-enkephalin) is broken into *n* one-dimensional problems, and the minimum is found for each dimension sequentially. This procedure is repeated until convergence is achieved. PEPTOR[2] uses a similar procedure, but instead tries to find the global, and not the local, minimum in each dimension.

Throughout the PEPTOR run, each one-dimensional minimization varies only a single torsion angle. This is different from the minimization procedure proposed by Powell[6] where a linear combination of several variables can be used as a single dimension. Instead of finding the local minimum of each torsion angle, the full torsional surface is scanned using a constant step size. For example, if a step size of 6° is used, the energy of 60 points on the torsional potential surface of a single torsion angle is calculated. The lowest energy point and the two adjacent points are used in a parabolic fit to locate the "global minimum" with respect to this single dimension, and this torsion angle is set to this value while the others are examined.

Once this procedure has converged on a structure, the user has the option of minimizing it with the nonderivative method described by Powell.[6]

If the total energy of a polypeptide could be written as a sum of functions, where each function depends on only one torsion angle, this procedure would locate the global minimum in a single pass through the torsion angles and will not depend upon the order the angles are examined nor the starting conformation, assuming the scanning step size is small enough to properly locate the lowest energy point in each dimension. Unfortunately, this is not the case, and the results do depend upon which order the torsion angles are examined and the initial conformation.

Though this procedure may appear to be too simple to be useful, tests[2] show that, in certain cases, PEPTOR is able to produce structures that are virtually identical to the global minimum structure of Met-enkephalin.[7] Though these may be fortuitous results, further tests show that sets of PEPTOR structures consistently yield the global minimum when used as a "starting population" in a genetic optimization.[8]

In this examination, a random number generator is used to place the 24 rotatable bonds in a random order. The torsion angles are then scanned in this order. For example, the initial geometry is used to determine values for all 24 rotatable torsion angles. One of the angles is then varied from 0 to 360° using an input step size, and the other 23 angles are fixed at the initial value. This "torsional potential surface" is then scanned, and the torsion angle is set to the value yielding the lowest energy, as determined by the parabolic fit. A second angle is then varied with 22 of the angles held at their initial values and one, the first one examined, set to its "global minimum" value.

This procedure is continued until all 24 angles have been varied; after which one "cycle" has been completed. A new cycle is then started, and in this study, the angles are examined in the same order in each cycle. This process continues until the final energy after subsequent cycles is less than 0.1 kcal/mol.

To graphically show different conformations of Met-enkephalin, PEPTOR is run five times. In all cases the starting conformation has all backbone angles set to 180°, and the same seed is used for the random number generator. This means that all runs will scan each of the 24 torsion angles in the same order. The only difference in the jobs is the step size used in varying each angle from 0 to 360°. In the first job, the step size is 2° and is increased to 4, 6, 8 and 10° in each following job.

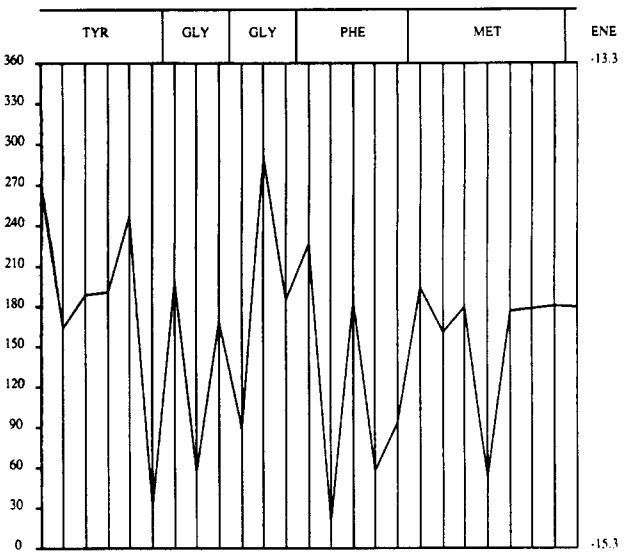

Figure 1. Parallel coordinate plot of the final conformation of Met-enkephalin obtained with a torsional step size of 2° and its energy (kcal/mol).

PEPTOR is written such that the initial torsion angles and the angles at the end of each scanning cycle are first corrected to give a unique value for a degenerate angle, as described above, and then written to a disk file. The final set of angles is also corrected and written to this file. In addition, the optimized angles, after correction, are written to this file if the user chooses to perform a final geometry optimization. Therefore, a single angle file can be used to examine the changes in each torsion angle after each cycle, or multiple files can be used to compare the final result obtained with different step sizes, with or without an optimization. The final structures presented here are those obtained from the scanning procedure only. In other words, no subsequent geometry optimization is performed.

PARALLEL COORDINATE PLOT

Knowing that a conformation of Met-enkephalin is uniquely given by the values of the 24 rotatable bonds, the value of a given torsion angle can be represented by a point on a one-dimensional number line that varies from 0 to 360°. If a separate line is used for each torsion angle, a single conformation of Met-enkephalin can be represented by points on these 24 lines. If the lines are drawn parallel to each other and the 24 points are connected by line segments, a parallel coordinate plot is generated.[3]

A parallel coordinate plot of one conformation of Met-enkephalin is shown in Figure 1. For each amino acid residue, the torsion angles are given in the order ϕ, ψ, ω, $\chi(1)$, $\chi(2)$, ... Figure 1 shows that for this conformation, the ϕ angle of TYR(1) is approximately 270°, while the ψ and ω angles are approximately 165 and 190°, respectively. The three side-chain angles of TYR(1) are given in the next three parallel lines, and they are followed by the three backbone angles of GLY(2). This continues until all 24 torsion angles are listed. The rightmost line is for the energy of this conformation, with the energy values listed to the right, since there is no restriction that the same type of data must be represented on each parallel

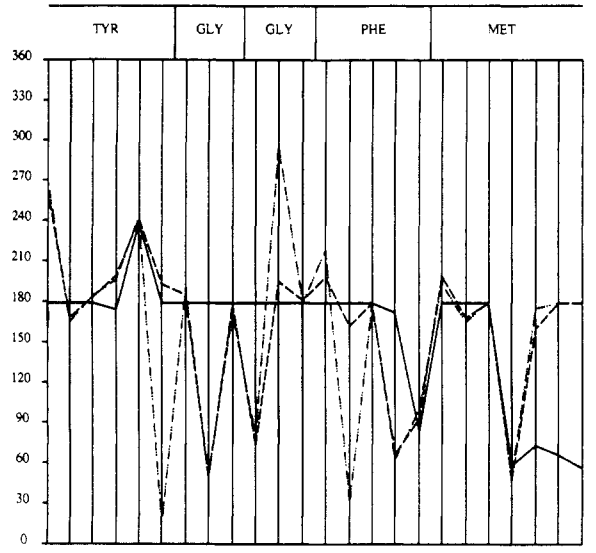

Figure 2. Parallel coordinate plot of the initial conformation of Met-enkephalin and the structure after the first and second scanning cycles with a step size of 2°.

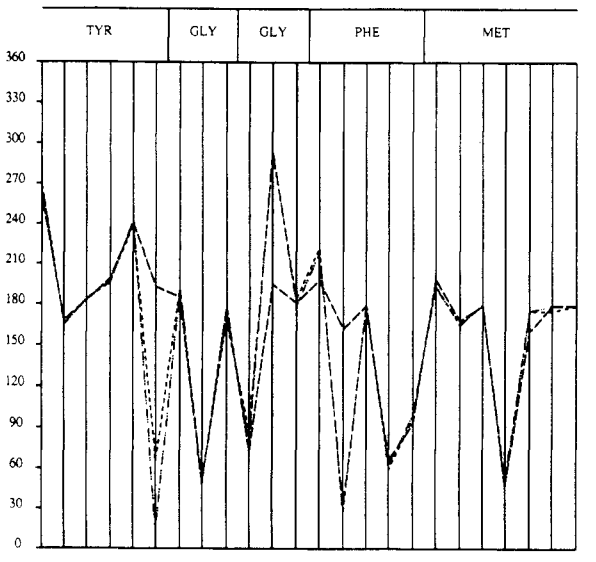

Figure 3. Parallel coordinate plot of the structure of Met-enkephalin after the first, second, and third scanning cycles with a step size of 2°.

line. From the position of the final line segment, this conformation has an energy of approximately −14.3 kcal/mol.

The line segments should not be taken to imply that the value of one torsion angle leads to the next since each coordinate, or vertical line, is assumed to be independent of the rest. Instead the line segments are there to show which set of 24 torsion angles belongs to the same structure and are very useful when multiple conformations are compared.

For example, Figure 2 shows the structure of the initial conformation (given points connected by solid line segments), and the structure after the first (dashed line) and second (dash-dot-dot line) scanning cycle when a 2° step size is used. It is clear from this picture that the structures obtained after the first and second cycle are quite similar to each other; the major differences are in the $\chi(3)$ angle of TYR(1), the ψ angles of GLY(3) and PHE(4), and the $\chi(1)$ angle of PHE-(4).

Figure 3 is simply one cycle further in the scanning process than Figure 2. In other words, the initial structure is removed, and the structure at the end of the third cycle (a dotted line) is included. This figure shows that the structures at the end

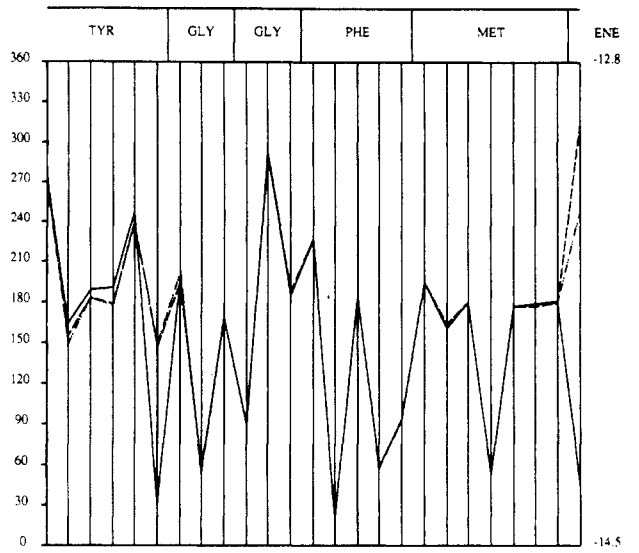

Figure 4. Parallel coordinate plot of the final conformations of Met-enkephalin after scanning with step sizes of 2°, 4°, and 6°.

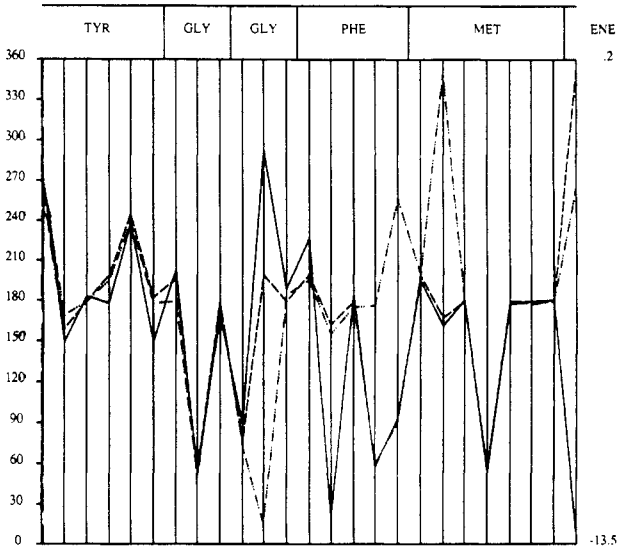

Figure 5. Parallel coordinate plot of the final conformations of Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

of the second and third scanning cycles are virtually identical, except for a difference of about 50° in $\chi(3)$ of TYR(1). Further plots would show that there is very little change in the torsion angles after the third scanning cycle even though 12 cycles were needed to make the energy change less than 0.1 kcal/mol.

Figure 4 shows a parallel coordinate plot of the final structures when scanning step sizes of 2, 4, and 6° are used. The 4° result is virtually identical to the 6° structure, and they differ from the 2° conformation mainly at $\chi(3)$ of TYR-(1). This means that an overlay of the structures would show very good overlap except for the position of the phenol hydrogen.

A parallel coordinate plot of the final structures obtained with step sizes of 6, 8, and 10° is shown in Figure 5. All three structures have similar values for the rotatable bonds in TYR-(1) and GLY(2) but have very different values for the ψ angle of GLY(3). In addition, the 8° and 10° structures have similar values for the ψ angle of PHE(4), but they are different than in the 6° structure. Conversely, the 6° and 8° structures agree with each other but disagree from the 10° structure in the values of $\chi(1)$ and $\chi(2)$ of PHE(4) and the ψ angle of MET(5).

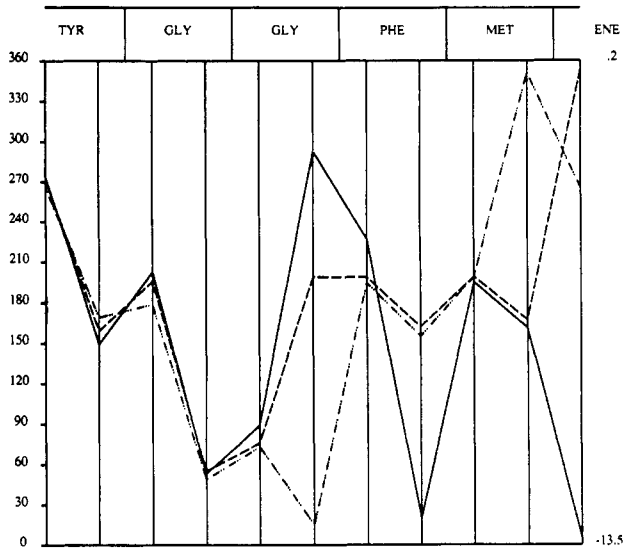

Figure 6. Parallel coordinate plot of the ϕ and ψ backbone angles of the final conformation of Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

As can be seen from Figure 5, a parallel coordinate plot gives a very good representation of exactly where two or more structures differ from each other; assuming care has been taken in setting consistent values for degenerate torsion angles. Though similar information can be obtained from a table of the difference in torsion angles, a separate comparison would have to be made for each pair of structures. A parallel coordinate plot allows for the easy comparison of three or more conformations.

To facilitate a comparison of a parallel coordinate, a concentric coordinate, and a compound Ramachandran plot, Figure 6 contains a parallel coordinate plot of the ϕ and ψ backbone angles of each residue in the final structures obtained with step sizes of 6, 8, and 10°. This plot is obtained from Figure 5 by simply taking the first two angles, or coordinates, from each residue.

CONCENTRIC COORDINATE PLOT

The only potential drawback of a parallel coordinate plot is the lack of a unique value for any torsion angle. Even though a convention has been adopted where degenerate torsion angles are given a single value (the one closest to 180°), there still is the problem of mapping a periodic function onto a one-dimensional line. This results in data points at the top of any line of a parallel coordinate plot being virtually identical to data points at the bottom of that line.

To remove this drawback, the plots in Figures 1–6 should actually be rolled into a cylinder so that the top and bottom of each plot meets. Though this would be more accurate, visualization on either a terminal or a piece of paper would be difficult. If parallel lines on a cone are used instead of a cylinder and the cone is oriented so that the vertex points towards you, a concentric coordinate plot is generated.

The conventions adopted in this study are as follows. A concentric coordinate plot is generated as a series of equidistant, concentric circles. The innermost circle is the first coordinate displayed [e.g., the ϕ angle of TYR(1)], and the last coordinate is the outermost circle. The 0° value is set at the top of each circle (i.e., the 12 o'clock position), and a right-handed axis is used as viewed down the cone from its vertex (i.e., the angles increase as you move clockwise around the circle). In addition, points on each circle corresponding

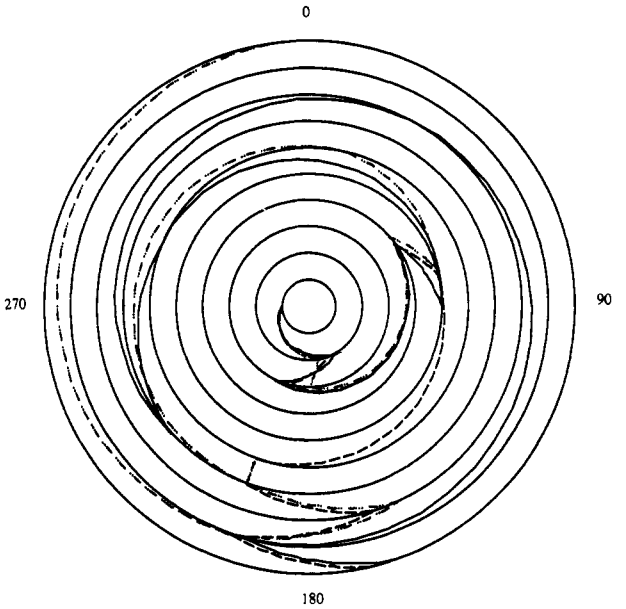

Figure 7. Concentric coordinate plot of the ϕ and ψ backbone angles of the final conformation of Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

to the same structure will be connected by an arc; the direction traversed by the arc is chosen so that the arc length is a minimum.

A concentric coordinate plot of the data shown in Figure 6 is displayed in Figure 7. The ϕ and ψ angles of each residue are displayed for the final structures obtained with step sized of 6, 8, and 10°. Both Figures 6 and 7 show very similar ϕ and ψ angles for TYR(1) and GLY(2) and a similar ϕ angle for GLY(3). Though all three structures have very different values for the ψ angle of GLY(3), and 8° structure (dashed line) and the 10° structure (dash-dot-dot line) have very similar values for the next three backbone angles. It is clear from both plots that the final backbone angle, ψ of MET(5), is very different in the 10° structure than the final 6 and 8° conformations.

Since a full circle is needed for each variable in a concentric coordinate plot, only about one-half as much data can be simultaneously displayed as in a parallel coordinate plot. On the other hand, the variables can be grouped into any sets that would yield the most information since each variable is assumed to be independent of the rest and the lines, or arcs, only intend to show which data points belong to the same structure.

For example, the final torsion angles selected by PEPTOR after multiple scans with step sizes of 6, 8, and 10° are displayed for each residue in Figure 8–12. Figure 8 shown that for TYR(1), all step sizes yield very similar conformations, with the largest differences in ψ, $\chi(1)$, and $\chi(3)$. Figure 9 shows that the conformations of GLY(2) are virtually identical in all three structures. The difference in all three structures for the value of ψ in GLY(3) is clearly shown in Figure 10.

Similarly, Figure 11 shows that one structure (the one found with a step size of 6°) has a very different value of ψ for PHE(4) than the other two, while another structure (the 10° step size conformation) has different values of $\chi(1)$ and $\chi(2)$. Figure 12 displays the final torsion angles of MET(5) and shows that the structure found with a 10° step size has a different ψ value, but all other angles are virtually the same in each structure.

COMPOUND RAMACHANDRAN PLOT

If only the ϕ and ψ angles are to be examined, as was done in Figures 6 and 7, a Ramachandran plot[9] has an advantage

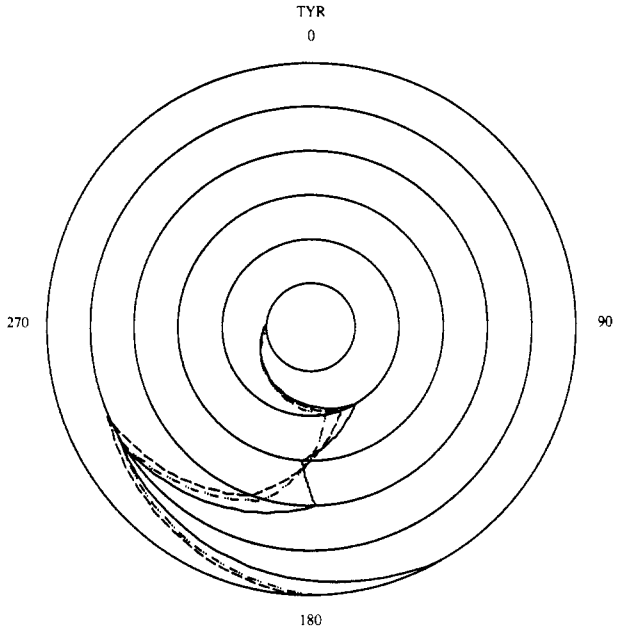

Figure 8. Concentric coordinate plot of the final conformations of TYR(1) in Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

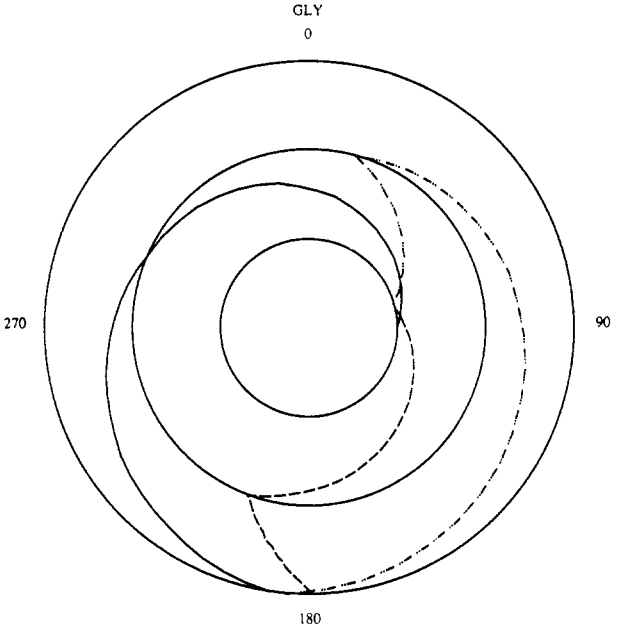

Figure 10. Concentric coordinate plot of the final conformations of GLY(3) in Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

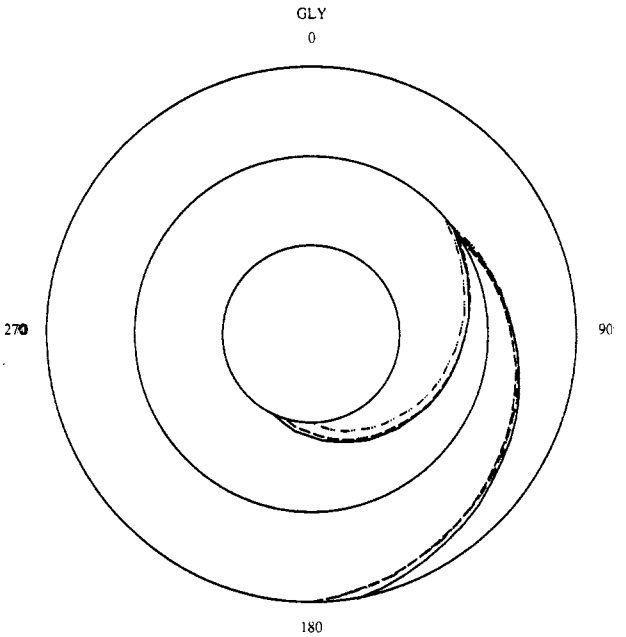

Figure 9. Concentric coordinate plot of the final conformations of GLY(2) in Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

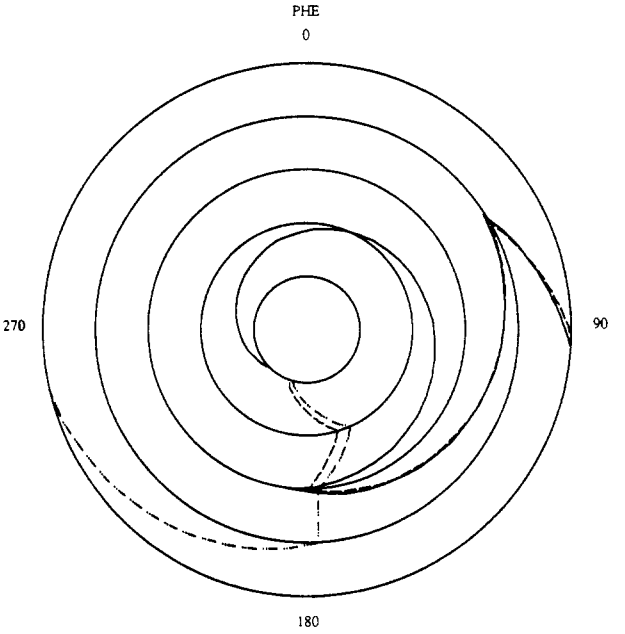

Figure 11. Concentric coordinate plot of the final conformations of PHE(4) in Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

over either a parallel coordinate or a concentric coordinate plot in that regions of this (ϕ, ψ) space have been associated with "standard" backbone conformations (e.g., a right-handed α-helix). A quick examination of a Ramachandran plot can aid in the rapid classification of a polypeptide segment.

If the (ϕ, ψ) coordinates of the first amino acid is labeled with the number "1", the (ϕ, ψ) coordinates of the second with "2", and so on, and the numbered points corresponding to the same polypeptide are connected with line segments, a compound Ramachandran plot is generated.

A compound Ramachandran plot of the initial conformation of Met-enkephalin and the resulting structures after the first two scanning cycles with a 2° step size is shown in Figure 13. Though it may not be easily noticed, the initial structure is represented by the set of overwritten numbers in the upper right corner since this structure has all backbone angles set to 180°. The structures generated after the first two scanning cycles show very similar (ϕ, ψ) values for the first two amino acid residues and similar ϕ but different ψ values for GLY(3).

This figure also shows that after the first scanning cycle the (ϕ, ψ) backbone angles for the last two residues are virtually the same. Though this similarity in backbone angles for these residues could be determined from the parallel coordinate or concentric coordinate plots, it is not nearly as noticeable as in the compound Ramachandran plot. The next scanning cycle changes the ψ angle of PHE(4), as it did with GLY(3), leaving the ϕ angle reasonably constant.

The structures obtained after the first three scanning cycles, with a 2° step size, are shown in a compound Ramachandran plot in Figure 14. Again it is clear that the backbone angles found after the second and third scanning cycles are virtually identical.

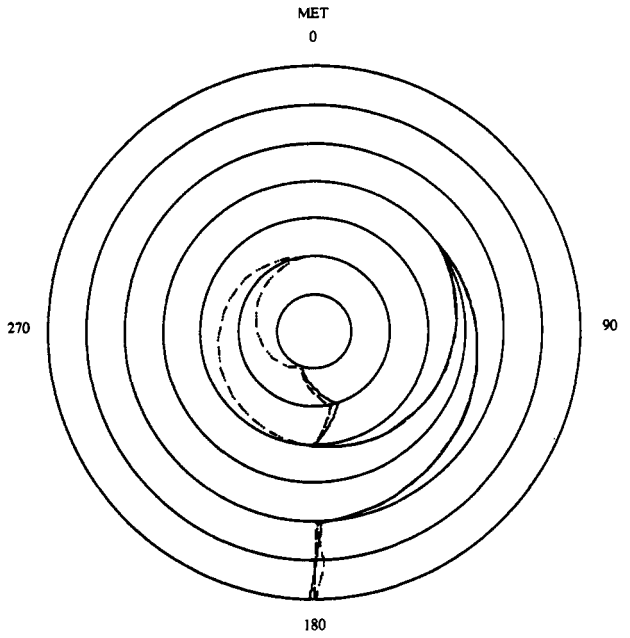

Figure 12. Concentric coordinate plot of the final conformations of MET(5) in Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

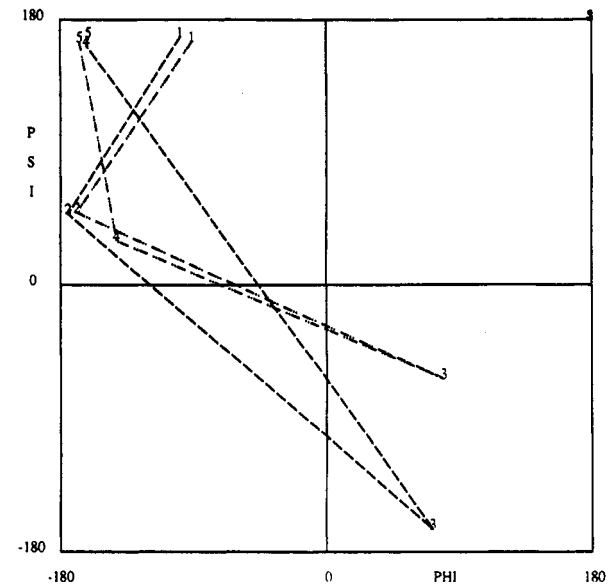

Figure 13. Compound Ramachandran plot of the initial conformation of Met-enkephalin and the structure after the first and second scanning cycles with a step size of 2°.

Figure 15 shows a compound Ramachandran plot for the final structures obtained with scanning step sizes of 6, 8, and 10°. All three structures have very similar (ϕ, ψ) values for TYR(1). Though the structure found with a 10° step size appears at first to have a very different value of ϕ than the other two for GLY(2), this is only because a Ramachandran plot should actually occur on the surface of a torus. Actually, the three structures have similar backbone angles for GLY-(2). All three structures have approximately the same ϕ value for GLY(3) but different ψ values. The structure found with an 8° step size has very similar (ϕ, ψ) values for both PHE(4) and MET(5). The 6° step size structure has similar (ϕ, ψ) values for MET(5) but a different ψ value for PHE(4). Conversely, the structure found with a 10° step size has similar (ϕ, ψ) values for PHE(4) but a different ψ value for MET(5).

CONCLUSIONS

This paper has proposed three different methods for graphically displaying some or all of the necessary torsion angles of a polypeptide. Before any of these methods can be widely used, a consensus has to be reached on what is the minimal set of torsion angles needed to determine the unique structure of a polypeptide, given its amino acid sequence. In addition, agreement is needed on the value to assign a torsion angle if degeneracy exists. In this study the value closest to 180° is chosen, but other options are available.

A parallel coordinate plot is very useful when you want to examine exactly what similarities or differences exist between two or more conformations of the same polypeptide or to examine all relevant torsion angles in a conserved amino acid sequence from different polypeptides or proteins. One slight drawback is that in this application, it is a two-dimensional representation of data points on the surface of a cylinder and it should be realized that data at the top of a given coordinate line is very close to data at the bottom.

If the relevant torsion angles are displayed for all residues in a polypeptide, as is done here, it is difficult to notice

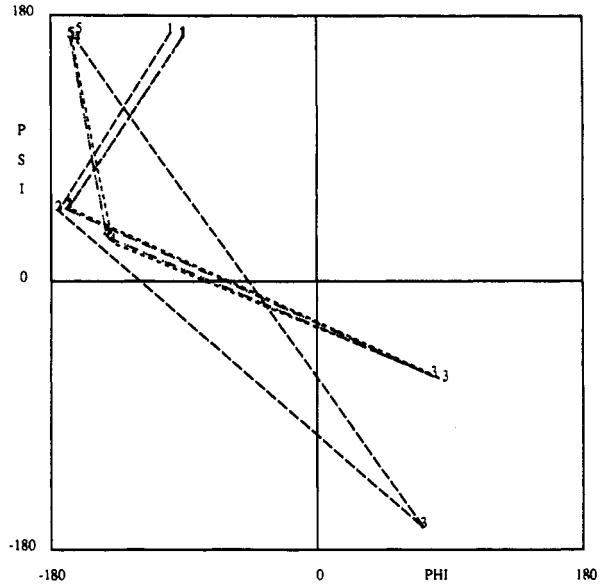

Figure 14. Compound Ramachandran plot of the structure of Met-enkephalin after the first, second, and third scanning cycles with a step size of 2°.

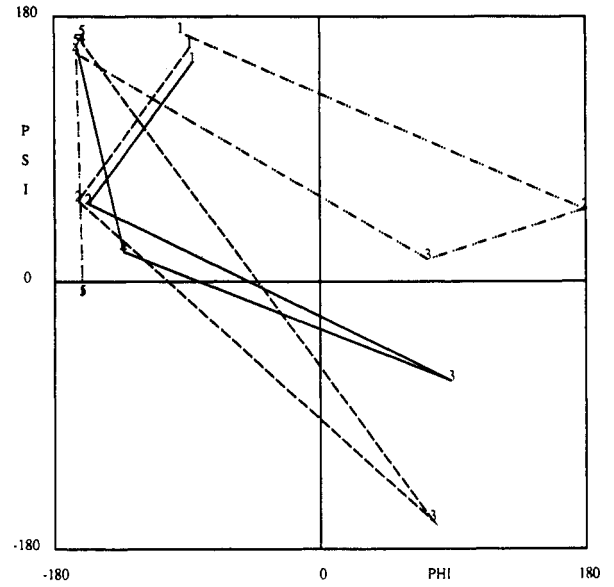

Figure 15. Compound Ramachandran plot of the final conformations of Met-enkephalin after scanning with step sizes of 6°, 8°, and 10°.

similarities in torsion angles between different amino acids. On the other hand, all torsion angles of a given amino acid can be displayed on the same set of coordinates for all occurrences of this residue in one or more polypeptides and would be useful in determining if side-chain angles are more or less conserved, or predictable, from protein to protein.

A concentric coordinate plot properly shows the periodicity of a given torsion angle but, because of its form, limits the number of coordinates, or torsion angles, that can be displayed/compared at the same time. As with a parallel coordinate plot, this type of display may be very useful in comparing the same amino acid residue from different positions in the same polypeptide or from different polypeptides.

A compound Ramachandran plot is very good for comparing the backbone angles of residues in the same or different polypeptides can be used to classify regions of a polypeptide. On the other hand, it suffers from the fact that it is a two-dimensional representation of the surface of a torus and does not allow for the display or comparison of any side-chain angles.

Two final points should be made about the use of any of the plots described in this paper. The first is that for large numbers of structures, or data points in any given coordinate, it is not necessary to connect the different coordinate points with line segments. This is only done to determine which values of a given coordinate belong to which structure. If you wish to simply examine the scatter in the torsion angles of a given amino acid in a large number of polypeptides, for example, data points on each coordinate would be sufficient.

In addition, these plots are presented for polypeptides but are not limited to describing this class of molecules. While a compound Ramachandran plot should be reserved for polymers where a pair of torsion angles gives all necessary information about the polymer's "backbone", a parallel coordinate or concentric coordinate plot can be used for virtually any structure.

The multidimensional plots displayed in Figures 1–15 are created by sending PostScript files to a laser printer. These files, as well as color plots on an appropriate monitor, are created by the program PLTTOR.[10] This program, which runs under DOS on a 386 or higher personal computer, and the scanning program PEPTOR[11] have been submitted for distribution through the Quantum Chemistry Program Exchange.

REFERENCES AND NOTES

(1) Luke, B. T. Displaying the Similarity or Difference in the Three-Dimensional Structure of Molecules. Manuscript in preparation.

(2) Luke, B. T. PEPTOR: A Program to Efficiently and Flexibly Scan the Torsional Hypersurface of a Polypeptide. Manuscript in preparation.

(3) (a) Inselberg, A. The plane with parallel coordinates. *Visual Comput.* **1985**, *1*, 69–91. (b) Inselberg, A.; Chomut, T.; Reif, M. Convexity Algorithms in Parallel Coordinates. *J. Assoc. Comput. Mach.* **1987**, *34*, 765–801. (c) Inselberg, A.; Dimsdale, B. Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry. *Proceedings of the First IEEE Conference on Visualization—Visualization '90*; 1990, pp 361–378.

(4) IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. *J. Biol. Chem.* **1970**, *245*, 6489–6497.

(5) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids. *J. Phys. Chem.* **1975**, *79*, 2361–2381.

(6) Powell, M. J. D. An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives. *Comput. J.* **1964**, *7*, 155–162.

(7) (a) Li, Z.; Scheraga, H. A. Monte Carlo-minimization Approach to the Multiple Minimum Problem in Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615. (b) Li, Z.; Scheraga, H. A. Structure and Free Energy of Complex Thermodynamic Systems. *J. Mol. Structure (THEOCHEM)* **1988**, *179*, 333–352.

(8) Luke, B. T. GENOPT: A Genetic Optimization Program for Finding the Global Minimun of a Polypeptide. Manuscript in preparation.

(9) Ramakrishnan, C.; Ramachandran, G. N. Stereochemical Criteria for Polypeptide and Protein Chain Conformation. *Biophys. J.* **1965**, *5*, 909–933.

(10) Luke, B. T. PLTTOR: A DOS Program that Generates Multidimensional Plots of PEPTOR Results. Submitted for publication in Quantum Chemistry Program Exchange.

(11) Luke, B. T. PEPTOR: A Program to Efficiently and Flexibly Scan the Torsional Hypersurface of a Polypeptide. Submitted for publication in Quantum Chemistry Program Exchange.