PROTEINS: Structure, Function, and Bioinformatics 66:889–902 (2007)

Steiner Minimal Trees, Twist Angles, and the Protein Folding Problem

J. MacGregor Smith, Yunho Jang, and Moon K. Kim*
Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst, Amherst, Massachusetts

ABSTRACT **The Steiner Minimal Tree (SMT) problem determines the minimal length network for connecting a given set of vertices in three-dimensional space. SMTs have been shown to be useful in the geometric modeling and characterization of proteins. Even though the SMT problem is an NP-Hard Optimization problem, one can define planes within the amino acids that have a surprising regularity property for the twist angles of the planes. This angular property is quantified for all amino acids through the Steiner tree topology structure. The twist angle properties and other associated geometric properties unique for the remaining amino acids are documented in this paper. We also examine the relationship between the Steiner ratio ρ and the torsion energy in amino acids with respect to the side chain torsion angle χ_1. The ρ value is shown to be inversely proportional to the torsion energy. Hence, it should be a useful approximation to the potential energy function. Finally, the Steiner ratio is used to evaluate folded and misfolded protein structures. We examine all the native proteins and their decoys at http://dd.stanford.edu. and compare their Steiner ratio values. Because these decoy structures have been delicately misfolded, they look even more favorable than the native proteins from the potential energy viewpoint. However, the ρ value of a decoy folded protein is shown to be much closer to the average value of an empirical Steiner ratio for each residue involved than that of the corresponding native one, so that we recognize the native folded structure more easily. The inverse relationship between the Steiner ratio and the energy level in the protein is shown to be a significant measure to distinguish native and decoy structures. These properties should be ultimately useful in the *ab initio* protein folding prediction. Proteins 2007;66:889–902.** **© 2006 Wiley-Liss, Inc.**

Key words: Steiner trees; twist angles; protein folding; side chain conformations; minimum energy Conformations

INTRODUCTION

The protein folding problem is to take a linear string of amino acids (primary sequence) and examine how it folds in three-dimensional (3D) space. This 3D structure of the protein is encoded in its primary sequence.[1] It is important to know the 3D structure since the function of the protein is better understood and drugs for various diseases are developed either by searching an existing protein in a protein database or by designing a new protein that satisfies geometric characteristics of a ligand binding site on a receptor.

In principle, each protein has a unique native conformation. After the primary structure is determined, the secondary structure is composed of alpha helices, beta strands, and turns, which provide the next level of structure in proteins. While the tertiary structure is concerned with how these secondary structures interact, the quaternary structure is concerned with the final shape of its elements.[2] In this paper, the focus is placed on how Steiner minimal trees (SMTs) can aid in understanding some of the rules in the folding process.

As one can imagine, there are many computational approaches to the protein folding problem. Generally speaking, the approach to protein structure and prediction is multi-faceted. As opposed to empirical methods that are cost- and time-consumptive, computational approaches offer an alternative and complementary approach to understanding this predictive process.

As indicated in the morphological tree in Figure 1, there are essentially two main divisions: knowledge-based methods and ab initio methods. The latter are based on either directly or indirectly using the potential energy function of proteins to guide the search. The knowledge-based methods use well-known protein structures as templates or guides to predict structure. Knowledge-based methods are further subdivided into homology based, threading methods or learning methods.[3–7] Within these knowledge-based methods, various optimization tools are often used since the underlying feature of these approach is to "match" the candidate structure with an already-known structure.

The *ab initio* methods are somewhat more ambitious since they do not assume a pre-ordained structure; how-

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

*Correspondence to: Moon K. Kim, 160 Governors Drive, Amherst, MA 01003. E-mail: mkkim@ecs.umass.edu

Received 6 October 2005; Revised 14 August 2006; Accepted 15 September 2006

Published online 15 December 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21257

ever, they assume a known potential energy function guiding the search. Unfortunately, the known potential energy function is non-convex, so the search process among the many local optima presents a formidable challenge. In global optimization methods, recent successes have been made. These methods can generally be subdivided into deterministic (branch-and-bound) type techniques or stochastic methods such as simulated annealing and genetic algorithms.[8,9]

A topological network design approach to the modeling of proteins and the protein folding problem is presented in this paper. It is largely based on the Steiner tree properties of the amino acids and their peptide structures. Steiner trees are shown to be an important aspect of the underlying minimum energy conformation (MEC) of the proteins. For example, Figure 2 illustrates the Steiner structure of a protein segment comprised of cysteine, glycine, and leucine in the native folded state of a protein. The last letter in the atomic structure of the figure designates the amino acid within which the atom resides.

The reason for the interest in this problem is to better understand the rules and properties that underly the protein folding process so that one can forecast the secondary and tertiary structure of the protein sequence as it folds in space. Currently, empirical approaches to the determination of the geometry of protein sequences are time-consuming or expensive, so any new understanding and method that can assist in the identification of the 3D geometry of the proteins would be a significant addition to the tools of protein science.

Although what is presented in this paper is not a new complete algorithm for protein structure prediction, there are insights and properties of Steiner trees that should be helpful in either the knowledge-based or *ab initio* methods.

Fig. 1. Morphological tree of protein folding prediction methods.

METHODS

Steiner Trees

Steiner trees represent the shortest possible network in E^d to connect a given set of terminals $V = \{v_1, v_2, \ldots, v_N\}$,

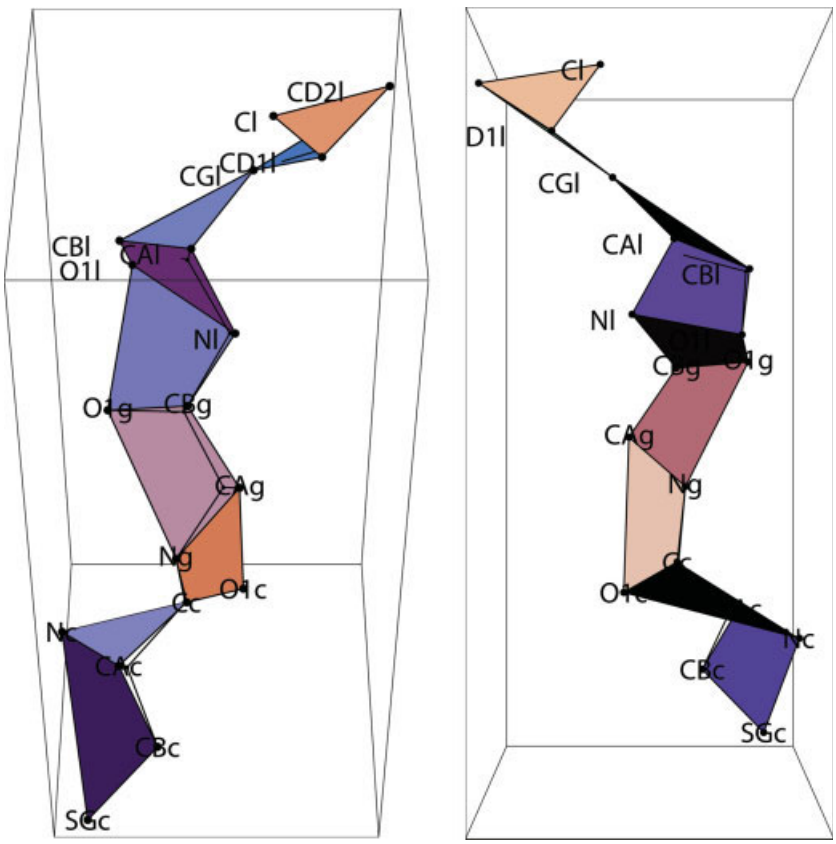

Fig. 2. Two views of the SMT and Steiner planes of a Cys-Gly-Leu tripeptide. Last letter in the atomic structure represents the amino acid type. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

where possible additional points called Steiner points from a set $S = \{s_1, s_2, \ldots, s_M\}$ may be used to further reduce the Euclidean distance connecting the given points in V. The problem is difficult (in fact it is NP-Hard)[10] because it is not known how many Steiner points to employ nor their location in space.

Of special importance linking Steiner trees and proteins is the fact that the solutions of Steiner trees in E^3 are characterized by planes joining the terminal vertices together. In fact, one key to finding the optimal Steiner tree structure is to know what the twist angles between the various planes are. To demonstrate this angle importance, let us take 4 terminals in space, which are denoted as $\{V_i, V_j, V_k, V_l\}$, that are from an equilateral tetrahedron in Figure 3 along with their Cartesian coordinates:

$$V_i(1,1,1) \quad V_j(1,-1,-1) \quad V_k(-1,1,-1) \quad V_l(-1,-1,1). \tag{1}$$

The larger nodes represent the terminal vertices while the smaller nodes either represent equilateral points or Steiner points. One possible optimization approach to determine the Steiner tree is to find the maximum distance between the equilateral vertices of the edge pairs or, in essence, find the orientation of the circles through the equilateral reflection points that are furthest apart. The following optimization problem is presented, involving the equilateral reflection points $T_{ij} - T_{kl}$ of the Melzak Circles[11,12] with the constraints ensuring the equilateral edges are satisfied:

$$\text{Maximize } \Phi = \|T_{ij} - T_{kl}\| \tag{2}$$

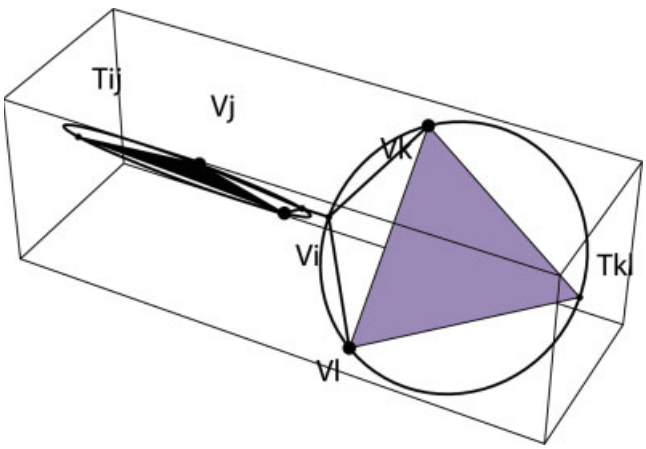

Fig. 3. SMT solution for $N = 4$. Two Melzak circles link pairs of terminals with Steiner points T_{ij} and T_{kl}. The twist angle between the planes defined by the Melzak circles is $\pi/2$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

subject to

$$\begin{aligned}
\left[(x_{ij} - x_i)^2 + (y_{ij} - y_i)^2 + (z_{ij} - z_i)^2\right]^{\frac{1}{2}} &= e_{ij} \\
\left[(x_{ij} - x_j)^2 + (y_{ij} - y_j)^2 + (z_{ij} - z_j)^2\right]^{\frac{1}{2}} &= e_{ij} \\
\left[(x_{kl} - x_k)^2 + (y_{kl} - y_k)^2 + (z_{kl} - z_k)^2\right]^{\frac{1}{2}} &= e_{kl} \\
\left[(x_{kl} - x_l)^2 + (y_{kl} - y_l)^2 + (z_{kl} - z_l)^2\right]^{\frac{1}{2}} &= e_{kl}.
\end{aligned} \tag{3}$$

This is a non-trivial optimization problem. What is important about this problem is captured in Figure 3. The circles represent Melzak circles that link pairs of terminals with equilateral vertices T_{ij} and T_{kl}. The twist angles between the planes defined by the Melzak circles are crucial to the SMT topology. For the $N = 4$ regular tetrahedron case, the twist angles between the planes are $\pi/2$. These twist angles define the Steiner tree topology for this point set. When one has point sets of $N = 4$ that deviate from the regular tetrahedron, the twist angles change and are in general difficult to compute. Therefore, as will be shown, the planes of the points in the Steiner tree structure are an important part of the solution process for the Steiner tree topology.

The following is a list of notation and definitions concerning Steiner trees in E^3 and higher dimensions.[13–15]

- M = Number of Steiner vertices from point set S.
- N = Number of given terminal vertices from the set V.
- FST = Full Steiner tree with the maximum number of Steiner points $M = N - 2$.
- MST = Minimum spanning tree with the number of Steiner points $M = 0$.
- $\rho_3(V) = SMT/MST$ Steiner ratio of a given terminal vertex set V in E^3.
- $M \leq N - 2$, $\forall E^3$.
- Angles subtended at each Steiner point are equal to 120° $\forall$, E^3.
- The ratio in the plane of dimension two $\rho_2 = \sqrt{3}/2$ is attained for equilateral triangles, ladders, and lattice configurations.

One important aspect of SMTs in E^3 is the Steiner ratio ρ_3 (V) of the SMT to an MST for a given terminal vertex set V. Since we work only in 3D, the subscript is dropped, and it is referred to as simply ρ. This value helps to determine whether the optimal SMT conformation has been achieved for a particular given terminal vertices.

SMTs and MECs

What is the crucial link between Steiner trees and MECs and how can this link be beneficial to predicting protein structure? It is argued that the geometric properties of Steiner trees can be used to approximate the potential energy function.

This relationship between SMTs and MECs is well described by Maxwell's theorem.[13] Let F_1, F_2, F_3, F_4 be unit forces acting at fixed terminals $\{v_1, v_2, v_3, v_4\}$ respectively. If one designs a network linking these terminals

with Steiner points $\{s_1, s_2\}$ that can be moved, then one seeks to find the location of the Steiner points and the network where these forces will be in equilibrium. The result is illustrated in Figure 4.

Theorem 1 *If one draws unit vectors from a Steiner tree in the direction of each of the lines incident to* $v_1, v_2, \ldots, v_N$, *and letting* F_i *denote the sum of the unit vectors at* v_i, *then in mechanical terms,* F_i *is the external force needed at* v_i *to hold the tree in equilibrium. The length of the tree* SMT *has the simple formula*

$$SMT = \sum_{i=1}^{N} v_i F_i. \quad (4)$$

If the forces at the vertices are not uniform, then the SMT acts only as a lower bound. This was discussed in detail in a previous paper.[16] Notice that in Maxwell's Theorem the function is separable in the force components. That is, the FST can be decomposed into its FST components. One way to construct a FST is to identify its FST components. Thus, if one can somehow subdivide the overall terminal set into its FST components, then the overall SMT can be computed by constructing the SMT for its FST components. In fact, as will be shown for protein structures, the FST components are largely composed of the backbone and amino acid components.

One of the surprising features of the link between Steiner trees and proteins is that the angles of the carbon and nitrogen atoms in the amide plane of the trans-peptide group satisfy the 120° angle property of the edges incident to the Steiner points (recall that angles subtended at each Steiner point are equal to 120°). In other words, the heavy atoms such as carbon and nitrogen function as Steiner points in protein backbone structures.[17] This is apparently due to the partial double-bond nature of the backbone plane. The fact that the peptide groups naturally form the trans-peptide conformation with few exceptions is also considered to be important.

The overall network of peptide bonds is normally configured as a linear polymer rather than a branched polymer chain. With this linear chain of peptide bond planes, the resulting geometry can be one of the secondary structures depending on the pattern of dihedral angle pairs (ϕ_i, ψ_i). Of course the complicating feature of the folded protein structure results from the geometry of side chains and how the entire structure packs together. Since the tertiary structure is uniquely determined by the interaction of the amino acids in 3D space, the relationship of Steiner trees to the amino acids themselves is examined in some detail.

We now see that the characteristics of proteins can be explained topologically by means of those Steiner properties, whereas the conventional protein folding problem has been considered as an optimization problem over a high-dimensional energy landscape.

RESULTS AND DISCUSSION

The results of the paper are composed of four parts. In the first part, a theoretical model of the amino acids is used to illustrate a regular angle property of the twist angles of planes within the amino acids. The planes are defined through the Steiner tree structure of each amino acid. What is also important about the theoretical model of the amino acids is that it contains the hydrogen atoms. In the second part, the Steiner ratio is shown to be inversely proportional to the torsion energy of the protein. This inverse relationship then affords a useful alternative measure of the energy function in proteins. The Steiner ratio becomes a surrogate for measuring energy in the protein structure.

In the third part, the amino acids from empirical protein structures are examined for their Steiner ratio. This is in contrast to the results generated in the first part

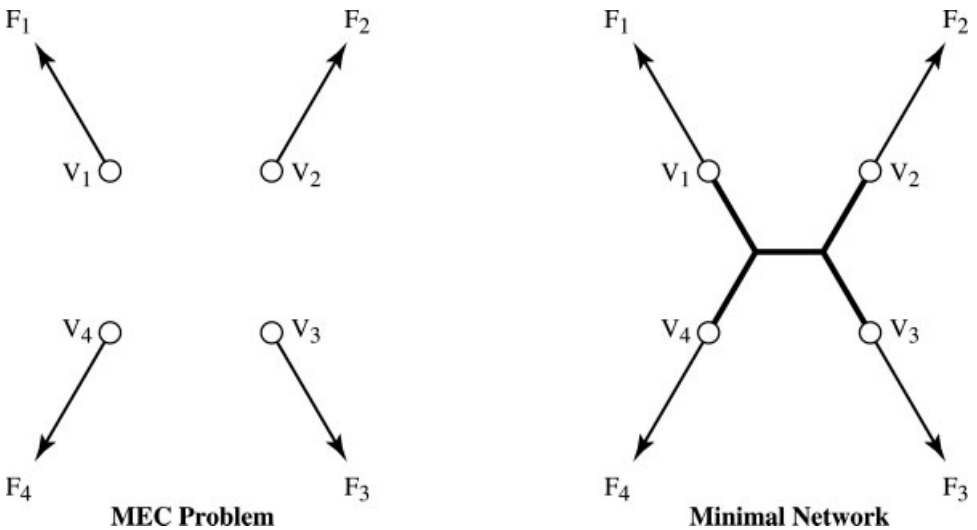

Fig. 4. Schematic of Maxwell's theorem. Given *V*, (Left) SMT ≡ MEC (Right). Two Steiner points are added into the given four vertexes to generate the Steiner Minimal Tree.

using a theoretical model of the amino acids. We compute the basic statistics including the mean and standard deviation of the Steiner ratio from samples of thousands of amino acids from PDB protein structures along with the entire probability distribution for each amino acid. The distinguishing feature of the empirical assessment is that the amino acids in the PDB data sets do not have the hydrogens, so this measure becomes a useful benchmark as it relates to the first part of the results.

In the fourth and final part of the analysis, the theoretical and the empirical models of the Steiner ratio for the amino acids are used to statistically compare native folded and decoy mis-folded proteins. All decoy data sets are compared with their theoretical and empirical Steiner ratios, and it is shown that the Steiner ratio can discriminate the native and decoy protein structures. Ultimately, it is through the Steiner ratio and its inverse relationship with the energy in the protein that a useful methodological tool is created in order to distinguish native and decoy structures.

A computer program based on a branch-and-bound optimization algorithm is used to evaluate the protein structures.[17] Starting with the fundamental geometry of the individual amino acids, an optimal Steiner structure is found by adding a maximum of $N - 2$ Steiner points so that the structure is a FST. One of the important properties of Steiner trees is that the optimal topology is dependent on the location of planes in space that determine the minimal length topology by connecting locally neighboring terminals and Steiner points.

Part 1: Twist Angle Properties of Steiner Planes

We found the Steiner tree solution for each amino acid and identified the planes of atoms corresponding to the Steiner tree structure. This allows us to characterize the internal dihedral angle structure of each amino acid. Thus, the Steiner structure and the twist angles are commensurate.

Glycine

The first acid examined in some detail is glycine (Gly). The reason this one is examined first is that it has the smallest number of atoms, $N = 10$ and its SMT structure is easy to compute.

The optimal 3D SMT structure is depicted in Figure 5, where the clusters of the FST components of the given terminals via the Steiner tree topology that interconnects them have been identified. Two different viewpoints are provided. The SMT topology provides more information about the geometric structure of the amino acid than just the chemical diagram. The Steiner points are thought to essentially define how the forces are transmitted through the molecular structure of the amino acid. Whether the Steiner points relate to some physical entity remains an open question. Defining the planes of interconnected atoms seems to be their primary role in order to define the conduit for the forces in E^3. The tetrahedral-planes isolate the FST components of the SMT acid structure. This helps explain why some of the carbon and nitrogen atoms are Steiner points in the SMT structure.

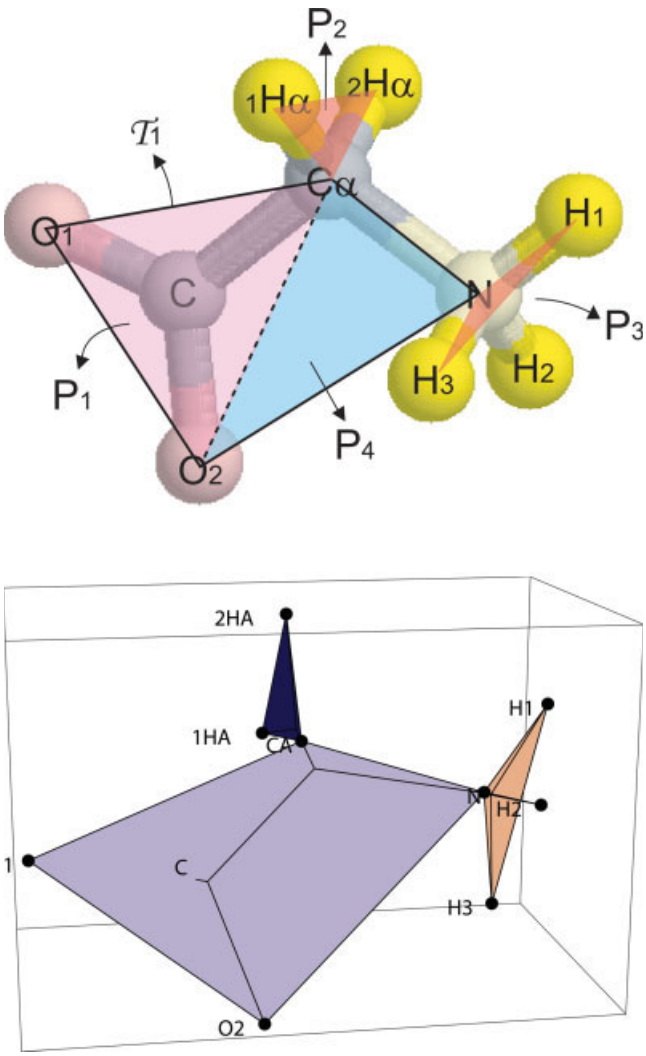

Fig. 5. Gly optimal structure. Steiner planes overlay the 3D ball and stick model of Gly (top) and the relative plane angles are measured from the 3D image generated by Mathematica (bottom). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

In Figure 5, the Melzak Circles have not been shown, but they are implicit in the diagrams. For example, one Melzak Circle in either of the diagrams of Figure 5 would pass through H_1, H_3 and the Steiner point linking them.

If the two tetrahedral-planes within glycine that are incident to the C_α are examined, the planes are essentially squashed tetrahedra with almost zero volume. This is illustrated in Figure 6. The constructed tetrahedra are ones that are composed of $\mathcal{T}_1 = \{C_\alpha, O_1, O_2, N\}$ and $\mathcal{T}_2 = \{{}_1H_\alpha, {}_2H_\alpha, C_\alpha, s_3\}$. In $\mathcal{T}_1$, H_1 was excluded since it did not lie in the plane of $\mathcal{T}_1$. Notice that C lies inside $\mathcal{T}_1$ as a Steiner point. It appears that $\mathcal{T}_1$ lies in a plane because of the double bond nature of the carbon-oxygen pair.

The volume of $\mathcal{T}_1$ is 0.0993 while that of $\mathcal{T}_2$ is 0.55616 $\times 10^{-7}$. If planes from these structures are defined, $P_1 = \{C_\alpha, O_1, O_2\}$ and $P_2 = \{C_{\alpha,1}\ H_{\alpha,2}\ H_\alpha\}$, the angle between their normals is 1.57055 radians, which is essentially $\pi/2$. Another plane from tetrahedron $\mathcal{T}_3$ as $P_3 = \{N, H_1, H_3\}$ can be defined and it is incident to the plane $P_4 = \{C_\alpha, N, O_2\}$, which has an angle of 1.0478 (approximately $\pi/3$). The angle between P_2 and P_4 is essentially $\pi/2$. Further, the angle between P_3 and P_4 is 1.04784 and P_2 P_3 is 1.048164 radians, which are both nearly $\pi/3$. Thus to summarize, the tables below organize the pairs of twist angles of the above planes. The first matrix is output from the software program MAPLE[18] in degrees and the second is the suggested rounded values in radians:

$$\begin{array}{c} P \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0.0 & 90.0 & 60.0 & 0.0 \\ & 0.0 & 60.0 & 90.0 \\ & & 0.0 & 60.0 \\ & & & 0.0 \end{pmatrix} \end{array} \quad \begin{array}{c} P \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} - & \pi/2 & \pi/3 & 0 \\ & - & \pi/3 & \pi/2 \\ & & - & \pi/3 \\ & & & - \end{pmatrix} \end{array}. \tag{5}$$

We suspect that one would not necessarily know the results about the twist angles within the amino acid unless one first computed the SMT structure, so it appears that these results are novel. The force planes as defined by the SMT components appear to be based on the covalent bonds of the atoms.

Finally, what appears to be happening is that the Steiner structure defines planes involving the atoms and these planes have certain regular dihedral angular relationships much as the Ramachandran angles in the peptide backbone.

Table I gives the detailed information about Gly and its Steiner coordinates. What is important is that the computer program locates the Steiner points coincident with the carbon and nitrogen atoms without any prior information or understanding that these would be Steiner points. The significance of the Steiner points is that they appear to be necessary to complete the overall structure of the amino acid according to Maxwell's Theorem. They help define the force planes in the amino acid structure much as the peptide bond planes in the network backbone structure. The Steiner points together with the given atomic terminals underly the MEC.

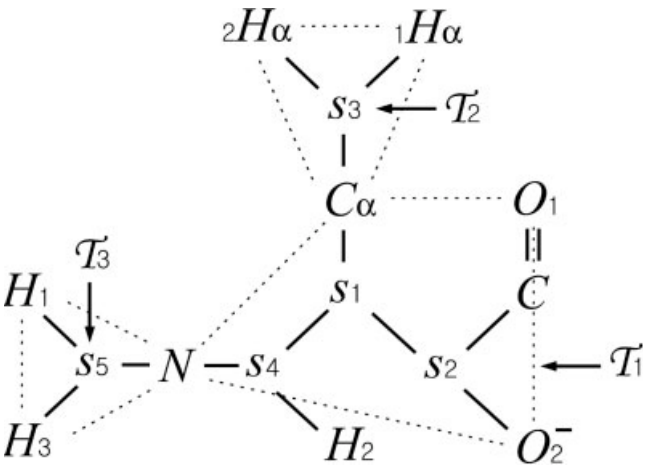

Fig. 6. Planar diagram of the Gly SMT. Four Steiner planes are defined. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Alanine

The second amino acid examined is alanine (Ala), which has the next smallest number of atoms $N = 13$. Figure 7 indicates the optimal Steiner topology and planes for Ala overlayed on its chemical diagram.

Figure 8 depicts the optimal SMT structure of Ala from two different viewpoints. Ala again has an interesting collection of FST tetrahedra and corresponding planes, which form the SMT MEC structure as depicted in Figure 8. Figure 8 also illustrates the two central planar structures incident again to the C_α.

Some of the tetrahedral structures within alanine are examined through the decomposition of the Steiner tree topology in Figure 7. The first tetrahedron is $\mathcal{T}_1 = \{C_\alpha, C_\beta, N, s_1\}$ while the tetrahedron orthogonal to it is $\mathcal{T}_2 = \{C_\alpha, O_1, O_2, H_\alpha\}$. Within $\mathcal{T}_2$, C_3 lies inside as a Steiner point.

The volume of $\mathcal{T}_1$ is 0.0003704 and that of $\mathcal{T}_2$ is 0.01563. They are "squashed" tetrahedra with essentially zero volume that are almost orthogonal. Further, if the planes are defined where $P_1 = \{C_\alpha, C_\beta, N\}$ and $P_2 = \{C_\alpha, O_2, H_\alpha\}$, then the angle between the normals

TABLE I. Atom Coordinates and Optimal Steiner Points for Glycine

Atom coordinates	Steiner points
C_α: 0.066000 0.181000 0.855000	C_α: **0.066000 −0.181000 0.855000**
N: −1.335000 0.173000 0.469000	N: **−1.335000 0.173000 0.469000**
C: 1.036000 0.600000 0.002000	C: **1.036000 0.600000 0.002000**
O_1: 2.237000 0.457000 0.158000	
O_2: 0.601000 1.407000 −0.877000	***remaining Steiner points***
$_1H_\alpha$: 0.225000 1.247000 0.703000	s_1: 0.004612 0.012105 0.674124
$_2H_\alpha$: 0.225000 0.063000 1.903000	s_2: 0.915327 0.622915 −0.020424
H_1: −2.000000 −0.363000 1.053000	s_3: 0.095242 −0.256543 0.937432
H_2: −1.487000 1.186000 0.614000	s_4: −1.208799 0.308362 0.513966
H_3: −1.487000 −0.060000 −0.528000	s_5: −1.410222 0.102201 0.431010

Glycine SMT = 10.720631; MST = 10.775242; ρ = 0.994932.

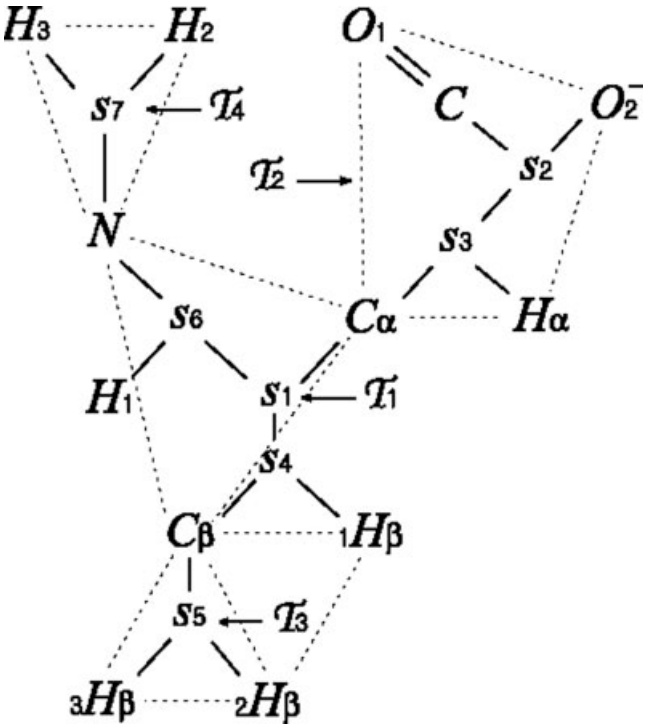

Fig. 7. Ala planar structures. Five Steiner planes associated with four tetrahedra are defined from the Ala SMT.

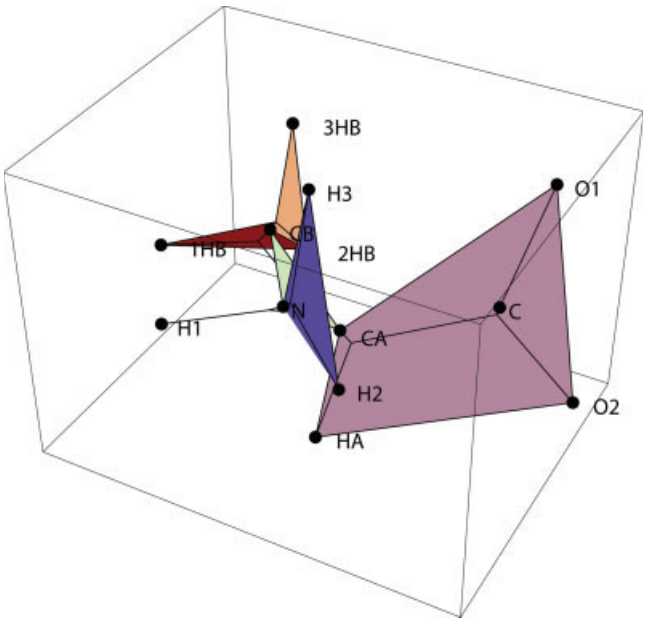

Fig. 8. 3D schematic of Steiner planes of the Ala SMT corresponding to the minimum energy conformation with $\chi_1 = -60^\circ$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

of the planes is 1.57073 radians or approximately $\pi/2$. Also, some additional planes are $P_3 = \{C_{\beta,2}\ H_{\beta,3}\ H_\beta\}$, $P_4 = \{N, H_2, H_3\}$, and $P_5 = \{C_{\beta,1}\ H_{\beta,2}\ H_\beta\}$ for the other tetrahedra. The planes and the atoms contained are slightly different from those of Gly but what follows appears to be a regular property of the twist angles of these two acids. Again, the first matrix is output from MAPLE in degrees, while the second is the radian approximations:

$$\begin{array}{c} P \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0.0 & 90.0 & 60.0 & 60.0 & 60.0 \\ & 0.0 & 60.0 & 60.0 & 60.0 \\ & & 0.0 & 60.0 & 0.0 \\ & & & 0.0 & 60.0 \\ & & & & 0.0 \end{pmatrix} \end{array}$$

$$\begin{array}{c} P \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} - & \pi/2 & \pi/3 & \pi/3 & \pi/3 \\ & - & \pi/3 & \pi/3 & \pi/3 \\ & & - & \pi/3 & 0.0 \\ & & & - & \pi/2 \\ & & & & - \end{pmatrix} \end{array} . \quad (6)$$

There appears to be a remarkably consistent result in that the twist angles are regular and only from a small subset of possible angular values. Table II gives the coordinates of the atoms and the optimal Steiner points. Again, the Steiner tree program did not a priori know the location of the Steiner points. The twist angles of the planes of atoms are then computed and displayed in a matrix.

As a result of computing the Steiner trees for all the amino acids, all have been classified into three categories based on the twist angles of the planes as defined by their SMTs (Fig. 9). The first group represents thirteen out of twenty amino acids and all have a remarkably regular set of twist angles. The second group has similar twist angles as the first group but an additional common pair, which appears to be $\{55^\circ, 73^\circ\}$ in Class I and $\{51^\circ, 68^\circ\}$ in Class II, respectively. It is observed that each amino acid in Class I has a side chain ring structure associated with the specific twist angles listed earlier, while additional twist angles of amino acids in Class II are strongly related to Steiner planes defined at the two distal branches of the side chains. We have lastly found that proline forms a class by itself with an irregular set of angles. Although, in this section, we investigated only Gly and Ala, the entire set of results about the classification of the amino acids is located in the supplementary material.

Part 2: Steiner Ratio vs. Torsion Energy

How the potential energy of the amino acid changes with variations in the torsion angles is of central interest. This, of course, relates directly to the earlier classification based upon the twist angles as defined by the Steiner structures. At the same time, the Steiner ratio is evaluated for each of the perturbed conformations. If it is true that variations in the side chain torsion angles correlate with an increase/decrease in the Steiner ratio, then the

TABLE II. Atom Coordinates and Optimal Steiner Points for Alanine

Atom coordinates	Steiner points
C_α: −0.134000 −0.047000 −0.101000	C_α: **−0.134000 −0.047000 −0.101000**
C_β: −1.379000 0.806000 0.122000	C_β: **−1.379000 0.806000 0.122000**
N: −0.093000 −1.143000 0.915000	N: **−0.093000 −1.143000 0.915000**
C: 1.098000 0.814000 0.038000	C: **1.098000 0.814000 0.038000**
O_1: 1.870000 0.931000 −0.900000	
O_2: 1.332000 1.418000 1.131000	***remaining Steiner points***
H_α: −0.164000 −0.479000 −1.099000	s_1: −0.345916 −0.115134 0.070850
${}_1H_\beta$: −2.267000 0.185000 0.022000	s_2: 1.087746 0.780693 −0.026563
${}_2H_\beta$: −1.409000 1.604000 −0.617000	s_3: −0.025105 −0.026176 −0.222490
${}_3H_\beta$: −1.349000 1.238000 1.120000	s_4: −1.394391 0.600278 0.100400
H_1: −0.937000 −1.734000 0.820000	s_5: −1.379000 0.919269 0.145862
H_2: 0.752000 −1.722000 0.764000	s_6: −0.256175 −1.116120 0.801090
H_3: −0.064000 −0.733000 1.865000	s_7: −0.012478 −1.158558 0.988639

Alanine SMT = 14.394162; MST = 14.478320; ρ = 0.994187.

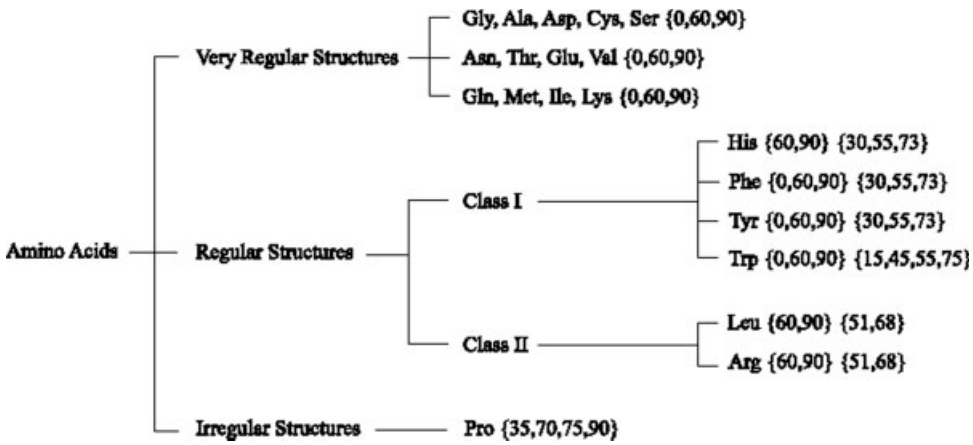

Fig. 9. Amino acid breakdown based on twist angles.

properties of Steiner trees should be a useful approximation to the potential energy function.

In order for this to occur, one can argue that the native state of an amino acid correlates strongly with the MEC of the amino acid. Changes in the torsion energy can then be measured by the Steiner ratio. A subset of the amino acids is examined next.

Alanine torsion angle variation

Figure 10 illustrates the variation in torsion energy of Ala with respect to its side chain torsion angle χ_1. The energy measures are captured by the software program Chembuilder.[19] Variations in the potential energy curve are plotted with variations in the torsion angle χ_1. Figure 10 also illustrates the corresponding changes in the Steiner ratio as a function of variations in the torsion angle. From these two plots, it is observed that the Steiner ratio is inversely proportional to the torsion angle.

Why should this be the case? As in Maxwell's Theorem, if a structure is a MEC, then perturbations in the Steiner structure will increase the Steiner ratio away from the MEC. The native Ala structure with a staggered conformation (set $\chi_1 = -60^\circ$) achieves the lowest energy value, or consequently the highest Steiner ratio, whereas the highest energy value occurs when the conformation is not staggered ($\chi_1 = 0^\circ$). Figure 11 shows the SMT of the highest energy conformation. As in the previous section, the planes of the atoms in the twist angles of Ala were utilized.

The most obvious difference of the non-perturbed conformation is the twisting of the planes involving the ${}_1H_\beta$, ${}_2H_\beta$ and ${}_3H_\beta$ atoms, which results in actually a shorter Steiner tree ($SMT^{-60^\circ} = 14.3942$ vs. $SMT^{0^\circ} = 14.3803$) and smaller Steiner ratio ($\rho^{-60^\circ} = 0.9942$ vs. $\rho^{0^\circ} = 0.9932$). Equation 7 displays the twist angle matrix for the perturbed Ala when $\chi_1 = 0^\circ$.

$$\begin{array}{c} P \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 00 & 90.0 & 60.0 & 60.0 & 60.0 \\ & 0.0 & \mathbf{33.6} & 60.0 & \mathbf{33.6} \\ & & 0.0 & 60.0 & \mathbf{2.4} \\ & & & 0.0 & 90.0 \\ & & & & 0.0 \end{pmatrix} \end{array} . \quad (7)$$

One can see that there are some new plane twist angels of about 33.6°. They were 60° before in the native

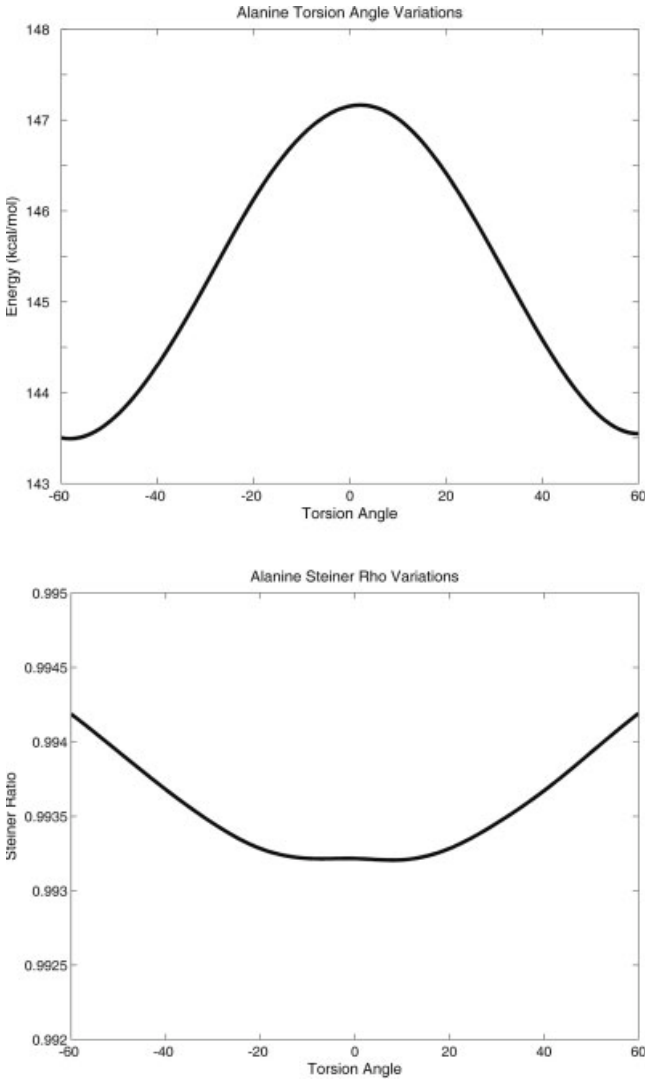

Fig. 10. Alanine torsion energy and Steiner ratio variations. The side chain torsion angle χ_1 is perturbed from a staggered conformation ($\chi_1 = -60^\circ$) to the next staggered one ($\chi_1 = 60^\circ$). Note that the Steiner ratio varies inversely proportional to the torsion energy.

Ala structure. The value of 2.4° computed by the software program in the right hand matrix is expected to be approximately zero.

Cysteine torsion angle variation

As with the Ala variations, variations were performed for the torsion angle with cysteine (Cys). Figure 12 illustrates the energy variations with respect to χ_1; notice that the Steiner ratio still varies inversely proportional to the torsion energy. The plane twist angles for the minimum and maximum energy conformations are also examined in Eq. (8), respectively. The irregular twist angles of {36.7°, 54.8°} appear in the maximum energy or the minimum Steiner ratio conformation. The tetrahedra and planes

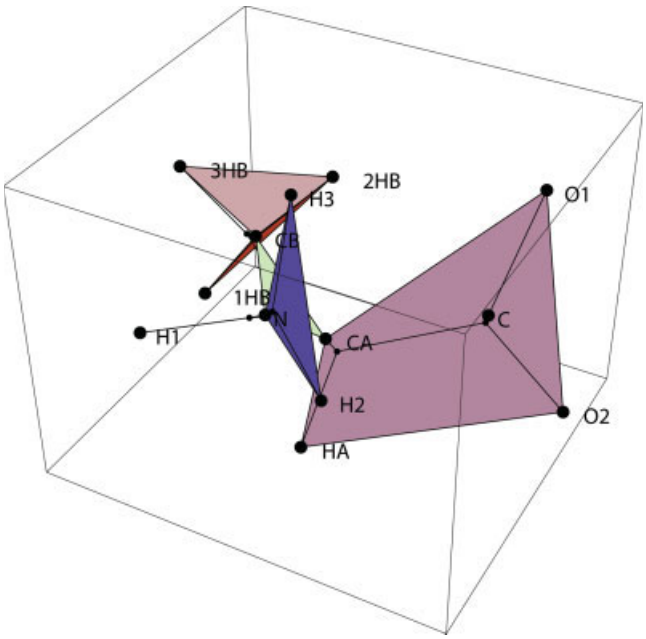

Fig. 11. The Ala SMT corresponding to the maximum energy conformation with $\chi_1 = 0^\circ$. An irregular twist angle of Steiner planes of 33.6° appears in this perturbed conformation when compared with the minimum energy conformation in Figure 8. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

defined for Cys after solving for the Steiner minimal tree are as follows:

$$T_1 := \{O_1, O_2, C_\alpha, H_\alpha\};\ P_1 := \{O_2, C_\alpha, H_\alpha\};$$
$$P_2 := \{C_\alpha, C_\beta, N\};\ P_3 := \{S_\gamma, C_{\beta,1} H_\beta\};$$
$$P_4 := \{N, H_1, H_2\};\ P_5 := \{N, H_2, H_3\};$$

$$\begin{array}{c} P \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0.0 & 90.0 & 60.0 & 60.0 & 60.0 \\ & 0.0 & 60.0 & 60.0 & 60.0 \\ & & 0.0 & 60.0 & 90.0 \\ & & & 0.0 & 60.0 \\ & & & & 0.0 \end{pmatrix} \end{array}$$

$$\begin{array}{c} P \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0.0 & 90.0 & \mathbf{36.7} & 60.0 & 60.0 \\ & 0.0 & \mathbf{54.8} & 60.0 & 60.0 \\ & & 0.0 & \mathbf{54.8} & 90.0 \\ & & & 0.0 & 60.0 \\ & & & & 0.0 \end{pmatrix} \end{array} . \quad (8)$$

From the experimental comparison of the two selected amino acids, we postulate that the torsion energy varies inversely with the Steiner ratio. All the remaining amino acids were also examined in a similar fashion and confirmed the proposed inverse relationship. Thus, if a native amino acid achieves a minimum energy conformation, variations in the torsion angle should be captured by the Steiner ratio denoted through its ρ value.

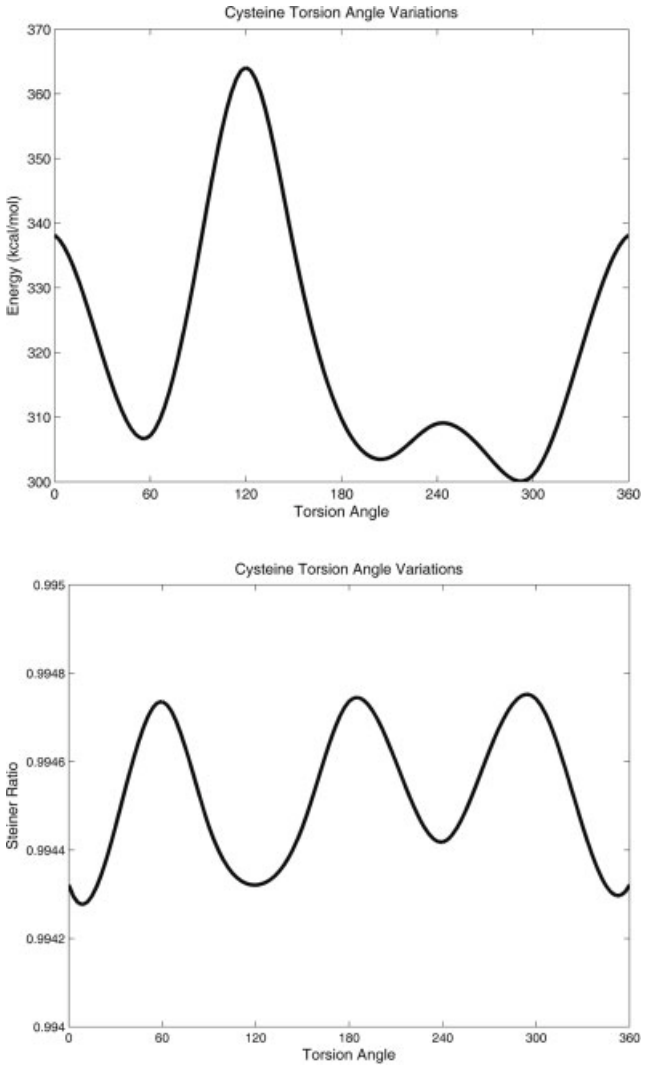

Fig. 12. Cysteine torsion energy and Steiner ratio variations. The χ_1 angle is perturbed from 0° to 360° because the side chain of Cys is not symmetric like Ala. Again, the inverse relationship between Steiner ratio and torsion energy is observed here.

Part 3: Theoretical vs. Empirical ρ Values

Table III shows the ideal ρ values for all 20 amino acids. This represents the best ρ values found so far with the computing capability presently available. Although not all the ρ_t values are optimal, they represent a best upper bound on the Steiner ratio, so they are still useful benchmarks.

Table III also includes the empirical ρ_e values statistically obtained from thousands of randomly selected amino acids from the PDB. The standard deviation σ of the empirical data sets is also included in the table and one recognizes that it is indeed very small. The mean and standard deviation will be useful in comparing whether the Steiner ρ_e value for certain PDB files is unusual or not. All 20 amino acids have been analyzed and a sample result with basic statistics and a histogram of the sample acid is presented in Figure 13. These sample results for all amino acids are also available from the authors.

In reading the last two columns of Table III, the $S' \subseteq S$ column represents the subset of Steiner points that correspond to the known carbon and nitrogen atoms in the amino acid, and the last column represents the combined number of carbon and nitrogen atoms within the amino acid.

With certain exceptions, the empirical ρ_e values are larger than the theoretical values since the theoretical ρ_t values include the hydrogen atoms. The exceptions are interesting because they largely have some of the carbon and nitrogen atoms as leaf vertices in the Steiner tree. The hydrogen atoms add structure and can greatly affect the Steiner ratio. The empirical ρ_e values along with the theoretical measure ρ_t are used as scoring markers in the next section to compare native and decoy protein structures.

Part 4: Protein Fold Recognition Using Steiner Ratio

To test the Steiner ratio as a surrogate for the potential energy function, folded and misfolded protein data sets are examined. They were obtained from the web site http://dd.stanford.edu, which contains 26 different data sets with folded and misfolded conformations. The given native folded structures are radically misfolded via a computer program, where the main chains are swapped between folds and the side chains are annealed using a Monte Carlo process.[20]

Native vs. Decoy Experiments

What is important to realize is that in the misfolded proteins all the detailed energy values in the potential energy equation, such as the bond lengths, bond angles, and torsion angles, are being perturbed so that the Steiner ratio varies with all these quantities. Although only a systematic analysis of the inverse relationship of the Steiner ratio with the side chain torsion angle was carried out, it appears that, in general, the Steiner ratio varies inversely with the potential energy of the entire system.

Can the Steiner ratio ρ be used as a predictive measure? As a first step, let's select three of the decoy data sets and compare the first three residues of the native with the decoy data sets. As one can see in Table IV, when residues 1-3 are examined for these three native proteins and their misfolded decoys, the ρ values for the decoys are mainly much higher than they should be and are higher than the native proteins. The energy computations are carried out for each structure using Swiss-Pdb Viewer, in which computations have done in vacuo with the GROMOS96 43B1 parameters set without reaction field.[21] In all cases, each misfolded protein segment (dipeptide or tripeptide in this context) has a higher Steiner ratio but lower energy value than the corresponding folded one. This inverse relationship is as expected with the Steiner ratio and the energy within the protein. A detailed view of

TABLE III. Ideal and Empirical Amino-Acid Results

Name	SMT	MST	ρ_t	ρ_e[a]	σ[b]	V	$S' \subseteq S$	$C + N$
(Ala)	14.394	14.478	**0.9942**[c]	0.99748	0.00099	13	4	4
(Arg)	31.162	31.357	**0.9938**	0.99751	0.00061	27	10	10
(Asn)	19.382	19.481	**0.9949**	0.99735	0.00070	17	6	6
(Asp)	17.292	17.391	**0.9943**[c]	0.99734	0.00064	15	5	5
(Cys)	16.395	16.541	**0.9912**[c]	0.99671	0.00108	14	4	4
(Gln)	23.071	23.186	**0.9950**	0.99754	0.00058	20	7	7
(Glu)	20.991	21.096	**0.9950**	0.99759	0.00056	18	6	6
(Gly)	10.720	10.775	**0.9949**[c]	0.99823	0.00157	10	3	3
(His)	23.291	23.518	**0.9904**[d]	0.98711	0.00046	20	8	9
(Ile)	25.417	25.586	**0.9934**	0.99603	0.00069	22	7	7
(Leu)	25.447	25.588	**0.9945**	0.99665	0.00072	22	7	7
(Lys)	28.912	29.102	**0.9935**	0.99665	0.00072	25	8	8
(Met)	23.801	23.990	**0.9918**	0.99245	0.00144	20	6	6
(Phe)	27.222	27.300	**0.9972**[d]	0.99841	0.00037	23	9	10
(Pro)	19.491	19.756	**0.9866**	0.97649	0.00209	17	6	6
(Ser)	15.696	15.792	**0.9940**[c]	0.99573	0.00100	14	4	4
(Thr)	19.359	19.494	**0.9931**	0.99556	0.00085	17	5	5
(Trp)	32.253	32.478	**0.9931**[d]	0.99190	0.00042	27	12	13
(Tyr)	28.473	28.562	**0.9969**	0.99853	0.00042	24	10	10
(Val)	21.752	21.885	**0.9939**	0.99592	0.00076	19	6	6

[a]Empirical ρ_e value is computed optimal for all acids except Trp.
[b]Estimated standard deviation.
[c]Those acids are optical structures.
[d]Those acids where some of the *C,N* atoms are leaf vertices.

the experimental results for the first comparison of "1bp2" and "1bp2on2paz" is provided in Table V.

The energy comparison in Table IV probably misleads one to consider the misfolded decoy as the best candidate for the native one, based on the energy minimization. This is counter-intuitive to the belief that the native folded protein abides by the minimum energy conformation implying that the potential energy functions, most of which have been defined in an empirical manner, are not perfect enough to be used as a predictive tool for the protein folding problem. Alternatively, the Steiner ratio in this experiment shows the consistent result that the ρ value for a native protein segment is much closer to the ideal value than that of the corresponding misfolded one. Here the ideal Steiner ratio is computed by averaging the ρ values of the amino acids involved.

At this point, the amino acids of all the 26 decoy data sets and their native counterparts are compared. The Steiner ratios of each residue in both native and decoy structures were calculated. These observed ρ values are compared with empirical and theoretical ρ values, respectively. Table VI presents the resulting root-mean-square-difference (RMSD) values of all 26 data sets. Decoy structures have significantly lower RMSD values with respect to empirical ρ values, while the RMSD values of native structures are slightly lower than those of decoy structures with respect to theoretical ρ values. This indicates that the native Steiner ratio is closer to the theoretical (ideal) value than is the decoy Steiner ratio. Figure 14 illustrates one sample experiment, namely "1bp2" and its decoy structure "1bp2on2paz" isolated from the entire set of the native and decoy Steiner values for all 123 residues of this sample data set. The empirical ρ value acts as an upper bound of the Steiner ratio of the native protein. Notice that the native Steiner ratio underestimates the decoy Steiner ratio and the decoy structure indicates a lower energy value (higher ρ) than that of the native structure. The experiment underscores the inverse relationship between the Steiner ratio and the energy level. Thus the Steiner ratio for each individual acid in the PDB structure is a useful indicator of the energy levels of the entire protein.

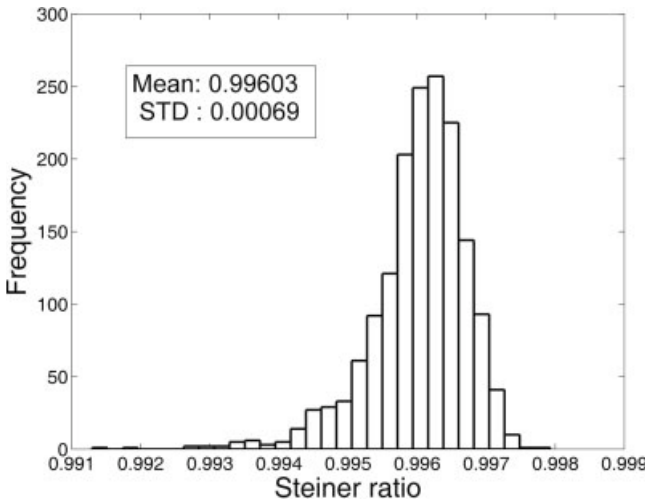

Fig. 13. Histogram of Ile empirical ρ values. Coordinates of 1628 Ile residues are randomly sampled from the PDB and those Steiner ρ values are calculated individually. The average and standard deviation of Ile empirical ρ values are also provided in the inset box.

An additional hypothesis is that it appears that the native structure more closely matches the theoretical ρ_t

TABLE IV. Comparison of Steiner ρ Values of Native and Misfolded Protein

Residue	Single	Dipeptide	Tripeptide	Single	Dipeptide	Tripeptide
		1bp2 (native)			1bp2on2paz (decoy)	
Ala	0.9968 (0.9942)[a]	–	–	0.9974		
Leu	0.9940 (0.9945)	0.9954 (0.9944)	–	0.9970	0.9972	
Trp	0.9925 (0.9931)	0.9930 (0.9938)	0.9945 (0.9939)	0.9925	0.9958	0.9960
$\sum E$			323.93 KJ/mol			234.22 KJ/mol
		1cbh (native)			1cbhon1ppt (decoy)	
Thr	0.9925 (0.9931)	–	–	0.9965		
Gln	0.9957 (0.9950)	0.9957 (0.9941)	–	0.9979	0.9972	
Ser	0.9890 (0.9940)	0.9936 (0.9945)	0.9955 (0.9940)	0.9958	0.9972	0.9969
$\sum E$			199.09 KJ/mol			104.37 KJ/mol
		2paz (native)			2pazon1bp2 (decoy)	
Glu	0.9934 (0.9950)	–	–	0.9981		
Asn	0.9980 (0.9949)	0.9958 (0.9950)	–	0.9974	0.9979	
Ile	0.9956 (0.9934)	0.9970 (0.9942)	0.9961 (0.9944)	0.9973	0.9975	0.9978
$\sum E$			234.66 KJ/mol			102.13 KJ/mol

[a]The value in parenthesis is the ideal Steiner ρ value.

TABLE V. Comparison of Detailed Energies in Folded and Misfolded Proteins

Residue	Bonds	Angles	Torsion	Improper	Nonbonded	Electro	Total
			Native Protein 1bp2				
HHT	0.000	6.183	7.540	0.000	0.00	6.52	20.239
ALA	0.107	0.944	1.936	0.359	−2.67	132.56	133.227
LEU	1.940	6.445	2.843	3.837	0.90	22.42	38.390
TRP	37.273	26.807	3.906	9.613	−12.21	61.31	126.706
OXT	0.000	0.000	0.000	0.000	−4.55	9.91	5.367
$\sum E$ (KJ/mol)	39.319	40.379	16.225	13.809	−18.53	232.72	323.929
			Misfolded Protein 1bp2on2paz				
HHT	0.000	6.183	7.540	0.000	0.00	7.50	115.980
ALA	0.204	2.545	0.218	0.158	−7.97	120.82	115.980
LEU	0.467	1.914	3.621	0.655	−13.29	15.94	9.303
TRP	1.894	26.546	3.260	6.073	−19.10	69.41	88.087
OXT	0.000	0.000	0.000	0.000	−1.97	1.59	−0.373
$\sum E$ (KJ/mol)	2.565	37.188	14.639	6.886	−42.32	215.26	234.221

value, since actually, the native structure has the hydrogens even though they are not recorded in the PDB structure. Therefore, although not a perfect measure, the theoretical Steiner ρ value by itself is a useful benchmark to distinguish correctly folded structures.

Experimental Error Detection

Besides the comparison of the native and decoy data structures, it appears that computing and graphing the Steiner ratio of all the amino acids in a native protein deposited on the PDB could indicate whether the PDB structure is an accurate indicator of the native structure. The graphs of data sets C ("1fdx" and its decoy "1fdxon5rxn") and U ("2paz" and its decoy "2pazon1bp2"), although the latter are not shown, are examined in the comparison with those of the native and decoy data sets, see Figure 15. Both proteins are interesting because they have statistical outliers that indicate the Steiner ratio for the amino acids in the PDB data are probabilistically unrealistic.

Using a standard normality test, Table VII compares the ρ values of the observed amino acids that are outliers and computes the probability that the observed amino acid ρ values would occur. In all instances, the probabilities are essentially zero. Even if one assumed that the theoretical values were distributed in a similar manner as the empirical ρ values, the probabilities of the observed conformation would still be essentially zero.

Thus, the computed ρ values should be a useful measure to indicate suspected conformational outliers or abnormalities in a protein structure obtained experimentally.

CONCLUSIONS

The SMT problem determines the shortest possible network to connect a given set of points in 3D space. This ge-

TABLE VI. RMSD of Decoy Sets With Respect to Empirical and Theoretical Steiner Ratios

Protein[a]	ρ_e vs. ρ_{native}	ρ_e vs. ρ_{decoy}	ρ_t vs. ρ_{native}	ρ_t vs. ρ_{decoy}
A	0.00226758084446	0.00080169662064	0.00307542421557	0.00380339175653
B	0.00252209632823	0.00086561543617	0.00294431229473	0.00389293920308
C	0.00899720953243	0.00084395144235	0.00806618892754	0.00446086803959
D	0.00202274768861	0.00071587065049	0.00309418402814	0.00387002790232
E	0.00234134600972	0.00069675754638	0.00242096050457	0.00327903313005
F	0.00273481490351	0.00076084043090	0.00379442693727	0.00389154181237
G	0.00266088230459	0.00139383439954	0.00298973701463	0.00429915675194
H	0.00307889201769	0.00095057552825	0.00296593827147	0.00369207755405
I	0.00218672115808	0.00085166858710	0.00292171495819	0.00364471880262
J	0.00206946838467	0.00082012901227	0.00306574382518	0.00346608680298
K	0.00247249335069	0.00077465697620	0.00330349898185	0.00401356341384
L	0.00247249335069	0.00078291542917	0.00330349898185	0.00424805526003
M	0.00095706956536	0.00076549658623	0.00290738480223	0.00329402594190
N	0.00210766152916	0.00075164613741	0.00348854407143	0.00381369332481
O	0.00101438455764	0.00087653677794	0.00339295324505	0.00360909024761
P	0.00101438455764	0.00083972931536	0.00339295324505	0.00371656865319
Q	0.00171440737911	0.00090986173929	0.00302500858994	0.00325749319617
R	0.00171440737911	0.00078263486918	0.00302500858994	0.00313377218503
S	0.00179952889265	0.00092412574496	0.00308037221334	0.00358489761612
T	0.00202460645252	0.00077607463340	0.00297905567680	0.00345544069529
U	0.00399540647016	0.00088136649262	0.00406085975718	0.00393308599333
V	0.00201920898040	0.00093025411562	0.00292958098021	0.00394733592996
W	0.00198190404311	0.00083021505716	0.00286781566249	0.00325789134694
X	0.00204184145641	0.00085082896671	0.00278786498424	0.00333892925091
Y	0.00209015991196	0.00091132128574	0.00308512686889	0.00363619416358
Z	0.00229482924807	0.00099494048770	0.00342678028616	0.00426266543070

[a]26 decoy sets are renamed in alphabetic order for convenience only.

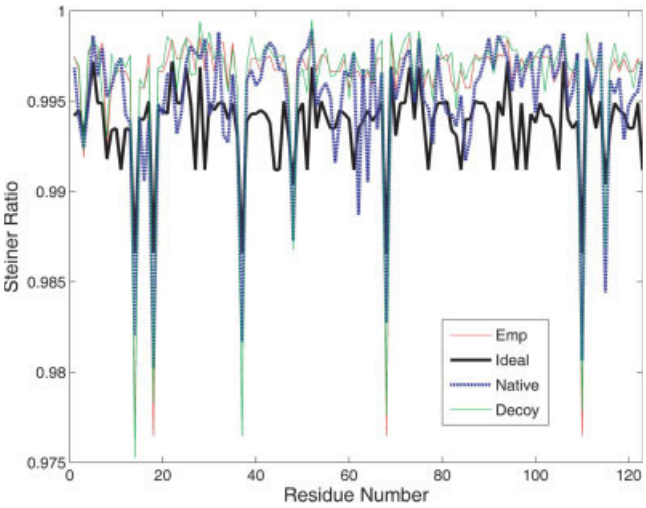

Fig. 14. Graphs of native and decoy Steiner ratios of Data Set A (1bp2 and 1bp2on2paz). The empirical ρ value acts as an upper bound while the theoretical (ideal) ρ value is a lower bound of the Steiner ratio of the native protein. Decoy Steiner values are much closer to the empirical plot. Native Steiner values largely follow the theoretical plot with some marginal error.

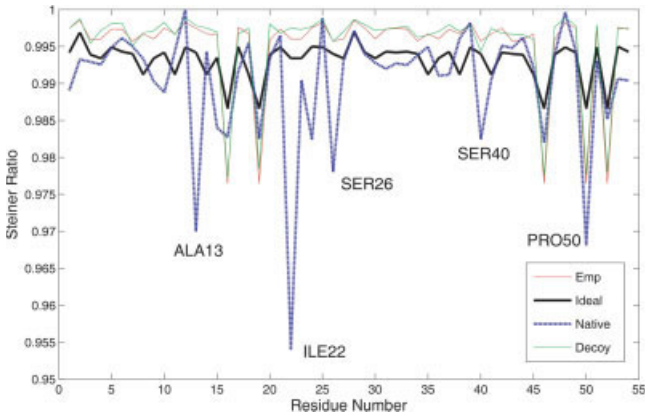

Fig. 15. Graphs of native and decoy Steiner ratios of Data Set C (1fdx and 1fdxon5rxn). Residue type and number of the five largest outliers are indicated. By comparison of Steiner ratio, we can detect abnormalities of the native structure corrupted by experimental noise.

ometry-based modeling technique is applied to the protein folding problem in this paper. The twist angles of planes determined by SMTs have been examined for 20 different amino acids. It has been shown that the subsets of these angles are from a limited set of values so that one can classify the amino acids into special clusters. This angular regularity can be extended to dipeptides, tripeptides, and larger protein structures.

Also, it has been shown that the Steiner ratio correlates with the torsion energy in the amino acid level. Systematic variation of the side chain torsion angle causes the energy value to vary inversely with the Steiner ratio, in fact, as the torsion energy goes down, the Steiner ratio goes up, and vice versa. Hence, the Steiner ratio should be a useful surrogate for the potential energy function.

TABLE VII. Statistical Comparison of Outliers in Proteins

Protein C	Observed ρ_e'	ρ_t	ρ_e	σ	Z-score[a]
ALA13	0.969951	0.9942	0.99748	0.00099	−27.807
ILE22	0.953953	0.9934	0.99603	0.00069	−60.981
SER26	0.978052	0.9940	0.99573	0.00100	−17.678
SER40	0.982450	0.9940	0.99573	0.00100	−13.280
PRO50	0.968086	0.9866	0.97649	0.00209	−4.021
Protein U	ρ_e'	ρ_t	ρ_e	σ	Z-score
MET84	0.982605	0.9918	0.99245	0.00144	−6.837
SER104	0.974016	0.9940	0.99573	0.00100	−21.714

[a]From the standardized normal distribution approximation of the probability distribution of the various acids, we can compute the Z-score $Z = (X - \mu)/\sigma$ which represents the probability of the particular acid have a ρ value below the empirical ρ_e value, and in particular $P(X \leq a) = \Phi((a - \mu)/\sigma)$ of the standardized normal distribution. By convention, a Z-value of $Z \leq -3.4$ or less corresponds to essentially a zero probability value.

Finally, the Steiner ratio acts as a predictive measure to distinguish the native folded protein structure. By comparison with the ideal Steiner ρ_t value and an empirically derived ρ_e value, they correctly distinguish the native proteins from their misfolded decoys that have even lower energy values. Although what is newly addressed in this paper is not a complete procedure for protein folding prediction, the Steiner properties are felt to be not only a useful scoring mechanism for protein structure and folding prediction, but also an initial step toward establishing a topological theory of protein folding.

REFERENCES

1. Anfinsen CB, Haber E, Sela M, White FH. The kinetics of formation of native ribonuclease during oxidation of the reduce polypeptide chain. Proc Natl Acad Sci USA 1961;47:1309–1314.
2. Stasiak A, Maddocks J. Best packing in proteins and DNA. Nature 2000;406:251–253.
3. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
4. Strickland DM, Barnes E, Sokol JS. Optimal protein structure alignment using maximum cliques. Oper Res 2005;53:389–402.
5. Blazewicz J, Hammer PL, Lukasiak P. Predicting secondary structures of proteins. IEEE Eng Med Biol Mag 2005;24:88–94.
6. Fischer D, Eisenberg D. Protein fold recognition using sequence-derived predictions. Protein Sci 1996;5:947–955.
7. Jones DT, Taylor WR, Thornton JM. A new approach to protein fold recognition. Nature 1992;358:86–89.
8. Klepsis JL, Pieja MJ, Floudas CA. Hybrid global optimization aLgorithms for protein structure prediciton: alternating hybrids. Biophys J 2003;84:869–882.
9. Pillardy A, Czaplewski C, Liwo A, Lee J, Ripoll DR, Kazmierkiewicz R, Oldziej S, Wedemeyer WJ, Gibson KD, Arnautova YA, Saunders J, Ye YJ, Scheraga HA. Recent improvements in prediction of protein structure by global optimization of a potential energy function. Proc Natl Acad Sci USA 2001;98:2329–2333.
10. Garey , MR, Johnson DS. Computers and intractability: a guide to the theory of NP-completeness. San Francisco: Freeman; 1979.
11. Melzak ZA. On the problem of Steiner. Can Math Bull 1961; 4:143–148.
12. Smith JM. Geometric optimization problem for Steiner minimal trees in E3. In: Pardalos P, editor. Approximation and complexity in numerical optimization. Holland: Kluwer Academic; 2000. pp 446–476.
13. Gilbert EN, Pollak HO. Steiner minimal trees. SIAM J Appl Math 1968;16:1–29.
14. Du DZ, Hwang FK, Weng JF. Steiner minimal trees on zig-zag lines. Trans Am Math Soc 1982;278:149–156.
15. Du DZ, Hwang FK. A proof of the Gilbert-Pollak conjecture on the Steiner ratio. Algorithmica 1992;7:121–135.
16. Stanton C, Smith JM. Steiner trees and 3-D macromolecular conformation. INFORMS J Comput 2004;16:470–485.
17. Smith WD, Smith JM. On the Steiner ratio in 3-space. J Comb Theory A 1995;69:301–332.
18. Maple. Waterloo: Maplesoft; 1988.
19. Chembuilder. San Diego: Interactive Simulations; 1997.
20. Holm L, Sander C. Evaluation of protein models by atomic salvation preference. J Mol Biol 1992;225:93–105.
21. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.