Journal of Global Optimization, 4, 209–227 (1994)

Conformational Searches for the Global Minimum of Protein Models

DAVID M. FERGUSON, AMANDA MARSH, THOMAS METZGER
Department of Medicinal Chemistry and Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455

DAVID G. GARRETT AND KEITH KASTELLA
PARAMAX Systems Corporation, P.O. Box 64525, St. Paul, MN 55164

Abstract. The conformational space of two protein structures has been examined using a stochastic search method in an effort to locate the global minimum conformation. In order to reduce this optimization problem to a tractable level, we have implemented a simplified force field representation of the protein structure that drastically reduces the degrees of freedom. The model replaces each amino acid (containing many atoms) with a single sphere centered on the C_α position. These spheres are connected by virtual bonds, producing a "string of beads" model of the peptide chain. This model has been coupled with our stochastic search method to globally optimize the conformation of two common structural motifs found in proteins, a 22-residue α-helical hairpin and a 46-residue β-barrel. The search method described further reduces the optimization problem by taking advantage of the rotational isomerisms associated with molecular conformations and stochastically explores the energy surface using internal, torsional degrees of freedom. The approach proved to be highly efficient for globally optimizing the conformation of the α-helical hairpin and β-barrel structure on a moderately powered workstation. The results were further verified by applying variations in the search strategy that probed the low energy regions of conformational space near the suspected global minimum. Since this method also provides information regarding the low energy conformers, we have presented an analysis of the structures populated, and brief comparisons with other work. Finally, we applied the method to globally optimize the conformation of a 9-residue peptide fragment using a popular all-atom representation and successfully located the global minimum consistent with results from previous work.

Keywords: Protein conformation, molecular mechanics, conformational searching

1. Introduction

The protein folding problem is generally accepted to be one of the most difficult and challenging problems of molecular biophysics and biochemistry [30, 8, 13, 31, 15]. Over the years, a tremendous amount of effort has been devoted to solving this problem which is not surprising due to the ramifications of its solution. Although much insight has been gained regarding the folding pathway and folded states (or native state) of proteins from both computational and experimental studies, the problem remains unsolved [30, 8, 13, 31, 15, 40, 2, 7, 33, 4]. The approach that we have undertaken assumes that all information needed to fold the protein to the native state is contained in the primary sequence of the protein, which appears to be valid for the majority of small globular proteins [28]. Using energy expressions that capture this information and global optimization techniques we would like

to develop computational methodologies to predict the 3-dimensional structure of proteins from the primary sequence.

The computational problems involved, however, are tremendously complicated by the number of possible minima that exist for molecular structures with many degrees of freedom. One should also keep in mind that, unlike ground states that are typical of homogeneous materials (that can be modeled as Lennard-Jones clusters on a lattice) [18], proteins can adopt an unknown multitude of conformations and vary in sequence as well. The problem is simply intractable for even the smallest of proteins (approximately 60 amino acids) without substantial computational simplifications. While there are many ways to accomplish this task (see ref. 13 for a review), our interests lie in the development and application of simplified molecular representations that take advantage of the polymeric nature of the biomolecule and model the protein structure using transferable, continuous-space potential energy functions. This is not a new concept and has been used in recent studies by several investigators, as well as many others in earlier work, to model protein structures on a limited basis [29, 19, 14, 39, 5, 21, 17]. For the most part, these models stem from a much larger bulk of work from which modern molecular mechanics force fields have evolved [3, 26, 24].

The approach we shall follow may be one of the simplest, but is satisfactory as a starting point in our studies in global optimization. The model reduces the amino acid residue to a single sphere centered on the C_α carbon position. These spheres are connected by virtual bonds that replace the peptide linkage (see Figure 1) producing a "string of beads" model of the protein structure. The energetics of the structure are evaluated using a classical potential energy function that has been parameterized appropriately [3]. These are the foundations of the so called "C_α force field" that has regained popularity in recent work [29, 19, 14, 39, 5]. While the model is quite primitive it has two practical advantages: First, the number of degrees of freedom of the system is substantially reduced (with N equal to the number of residues in the polymer), shrinking the volume of conformation space drastically. Second, the simplified model should have a smoother potential energy surface with larger catchment basins for the minima [22]. While this is difficult to prove outright, it would seem valid since individual atomic interactions are replaced by an average interaction energy, smoothing the surface. This should not only aid in the location of the global minimum, but speed convergence of our conjugate gradient local optimization used in the global search methodology.

These computational considerations are important to the success of most global optimization techniques. The method that we shall describe not only takes advantage of these simplifications, but also exploits the isomeric nature of molecular conformations [6]. The structure of butane demonstrates the concept nicely. This 4 carbon structure has one unique rotatable bond, that adopts three conformations; gauche$^+$ ($\phi = 60^0$), gauche$^-$ ($\phi = -60^0$), and trans ($\phi = 180^0$), where ϕ is the torsion angle (see Figure 1).[22] These conformations differ by rotations about one degree of freedom only, the carbon-carbon single bond. This is the general way molecular conformations are defined: as rotations about torsional degrees of free-

dom. The fact that conformers are rotational isomers provides yet another route to simplify the global optimization problem and will be used advantageously in our scheme below.

The method that will be followed in this work is an off-shoot of the approach we have previously described for conformational searching [11, 12, 10]. In our early work, we found it possible to search the conformational space of small to medium size ring systems using external coordinates as random variables in a stochastic search. We were also successful in globally optimizing relatively large structures with this approach, however, it became clear that the method was inefficient for large systems, especially linear molecules such as peptides [27]. The approach to be described here uses a similar search flow-chart, but performs the stochastic search using torsional degrees of freedom as variables. This basic idea of using the torsion angles in a stochastic search originated in the seminal work of Li and Scheraga for the global optimization of a small pentapeptide [23]. Still and Guida also pioneered torsional searching techniques that proved to be highly successful for rather large ring systems [16, 32]. We plan to apply a stochastic torsional search method, coupled with a simplified molecular force field, to globally optimize the conformation of two protein structures, a 22-residue α-helical hairpin and a 46-residue β-barrel. These systems are not actual proteins, but represent common structural motifs that form subdomains in many proteins and are often considered to be supersecondary structures. They are not only of interest structurally, but have been studied by other computational methods as well, offering an opportunity for comparisons [29, 19, 14]. Our results should be enlightening, since the characterization of the global minimum conformation of structures of this size has proven problematic in the past. Our method also provides a route to locate low energy conformers that populate the potential surface near the global, providing further information regarding the nature of the protein model employed. Finally, we shall present results using this methodology coupled with a popular all-atom molecular force field to globally optimize a short peptide fragment 9-amino acid residues in length.

2. The Protein Model

The basic theory behind the simplified model employed in our calculations stems from classical molecular mechanics force fields that model molecular interactions using "effective" pairwise-energy functions. One of the more widely applied force fields was developed by Kollman and coworkers for simulating the physical properties of proteins and nucleic acid structures [37]. The potential energy expression adopted includes harmonic bond stretching and angle bending terms, a Fourier series for torsional energies, and a Lennard-Jones 6-12 function for non-bonded interactions 1,4 and higher. (The original also included a Coulombic contribution.) The function takes the following form (see Table 1 and Figure 1 for parameters):

$$\sum_{i,j} K_l \Delta l_{i,j}^2 + \sum_{i,j,k} K_\theta \Delta\theta_{i,j,k}^2 + \sum_{i,j,k,l} \frac{V_n}{2}[1+\cos(n\phi_{i,j,k,l}+\gamma_{i,j,k,l})] + \sum_{i,l} \frac{A}{r_{i,l}^{12}} - \frac{B}{r_{i,l}^{6}}$$

where K_l and K_θ are the bond stretching and angle bending force constants, V_n is the n-fold torsional constant with a phase shift of γ, and A and B are the Lennard-Jones coefficients.

The simplified force field we have implemented in this work also includes these terms, but not on an all-atom basis. The atoms of the amino acid are instead replaced by a single sphere, centered at the C_α carbon. These spheres are connected by virtual bonds that replace the peptide linkage to model the protein as a "string of beads". Solvation effects, excluded volume effects (steric requirements), and other physical chemical properties of the amino acids must be accounted for by virtue of the parameterization. While this model is very primitive, it is a reasonable simplification for preliminary studies of protein folding since many structural motifs, such as those studied here, contain amino acids that have similar properties or patterns of amino acids that repeat. Although we could go to great lengths to increase the resolution of this model, this does not appear necessary to model the two structures of interest here. In fact, our approach, as well as others before us, reduce the number of C_α bead types to three, hydrophobic (PHB), hydrophilic (PHL), and bend or turn types (BND) which are neutral, to account for the main classes of amino acids found in nature. These bead types make up the composition of the two structural motifs studied here and are arranged in a sequence that allows the hydrophobic interactions to be packed in an optimal configuration, mimicking patterns reminiscent of those found in protein structures. The sequences studied are given in Table 1. We should make it clear, however, that these are not actual amino acid sequences taken from known proteins.

One of the premises for applying simplified force fields to model protein structures is that the global minimum conformation computed should correspond to the native folded state of the protein. The disposition of the polymer chain to reproduce a desired global minimum must be accomplished through careful empirical parameterization of the force field. Although this may seem straightforward, it is often difficult to ensure that the parameter set is correct, due in part to the lack of efficient and/or reliable global optimization techniques to test the model. We have chosen to start with the parameters reported by Honeycutt and Thirumalai for their studies of the same 46-residue β-barrel structure. The parameters employed for the α-helical hairpin were arrived at by modifying the energetics of the C_α angles and torsions to account for the local interactions of the α-helix. The optimal value for the C_α torsion angles were assumed to be 60^0, approximating the average distribution derived from naturally occurring proteins [36]. This bias in the torsional energy expression was captured using a simple phase shift of 240^0 for the single-fold cosine term (see Table 1) [35]. All parameters used in our study were evaluated using molecular dynamics calculations and fine-tuned if necessary. The final values are given in Table 1, along with other pertinent data regarding the force field.

The protein structures were built using the AMBER program modules PREP, LINK, EDIT, and PARM [1]. A special version of the PARM program was used to allow specific Lennard-Jones A and B coefficients to be defined between C_α bead

types [9]. The execution of this cascade of programs produces a cartesian coordinate file and parameter topology file that are required by the molecular mechanics and dynamics computational program package SPASMS [34]. The structures were left in the linear arrangement furnished by the EDIT program and subsequently optimized using the conjugate gradient method encoded in SPASMS. All non-bonded interactions were included in the calculations and a scale factor of 1/2 was applied to the 1,4 bead interactions (as was done by Kollman and coworkers [37]). Although we found this empirically motivated scaling to be non-essential for the success of the calculations, it seemed to produce better results and likely aids in the formation of turns and gauche states. No constraints were applied during the conjugate gradient minimization allowing full relaxation of the geometry.

3. Global Optimization Method

The global optimization method we have adopted employs a stochastic search method to explore the conformational space available to molecules, thereby characterizing a collection of conformers, including the global minimum. The basic scheme for our search is derived from our previous work in conformational searching and analysis with the major difference being that internal torsional coordinates, not cartesian coordinates, are the random variables [11]. Another important difference is the implementation. We have chosen to embed our algorithms in the SPASMS program package for these calculations due to the computational efficiency of the routines. The FORTRAN code of this program package has been highly optimized which not only decreases computation times but may prove beneficial in the future for parallelization.

The optimization problem can be formulated as follows:

$$GLOBALMIN\ E = f(\phi)$$

subject to

$$-180 < \phi_i < 180$$

where ϕ_i represents the i^{th} torsion angle in the molecular structure. Although the energy function (1) contains terms other than torsions (i.e. bond lengths, angles and non-bond distances) a molecular configuration is uniquely defined by the set of torsions. Also, evaluation of the energy function and gradient requires only the set of torsions as input (given an initial set of external coordinates).

The search algorithm has been coded as a subprogram in SPASMS and begins with a conjugate gradient optimization of the initial configuration furnished by the AMBER programs using the energy function (1). This local minima seeds the global search as the first conformation located, defining E and all ϕ_i at the start of the search. The general scheme implemented cycles through the following steps n times:

1. Randomize a variable set of torsional degrees of freedom.

j := *number_of_torsions*;
define current conformation, C;
$\phi := [\phi_1, \phi_2, ..., \phi_j]$;
$E := f(\phi)$

for i := 1 to j;
generate a random number *rand* $\in$ (-180,180);
$\phi_i^* := \phi_i$ + *rand*;
if ($\phi_i^* >$ 180) $\phi_i^* := \phi_i^*$ - 360
if ($\phi_i^* <$ -180) $\phi_i^* := \phi_i^*$ + 360
endfor

2. Locally optimize the new configuration by conjugate gradient over all cartesian coordinates, r

LOCALMIN $f(r(\phi^*))$

$\phi' := [\phi_1, \phi_2, ..., \phi_j]$;
$E' := f(\phi')$;

3. Examine conformation for novelty by energy or structural comparisons to all previously found conformers.

a) Energy based (option i or ii below):
generate a random number *rand* $\in$ (0,1);
compute $\Delta E := E' - E$;
if $\Delta E < 0$ or ($exp^{-\beta \Delta E} >$ *rand*) then
$\phi := \phi'$;
$E := E'$
endif
b) Structure based (option iii below):
C' := trial conformation (E', ϕ')
C_{set} := set of previously found conformations
if $E' < E_{limit}$ (where E_{limit} is a preset energy cutoff)
and $C' \notin C_{set}$ then
$\phi := \phi'$;
$E := E'$;
$C' \in C_{set}$
endif

4. Return to 1 with new conformation as starting point (resetting search counter, $n = 1$) or restore pre-randomized structure and return to 1 (updating $n = n + 1$) depending on result of 3.

5. Terminate search after a preset number of cycles, n, fail to produce a new conformation.

Three variants of this basic process have been developed. i) Global minimum search only. This technique accepts conformers in step 4 above if, and only if, the energy of the system decreases. No structural comparisons are performed and the search proceeds downhill only. ii) Global minimum search with Boltzmann probability. Similar to i, however, this search allows conformers with higher energies to be accepted in step 4 based on the Boltzmann probability factor ($P(\Delta E) = exp^{-\beta \Delta E}$) evaluated at 300K [20]. This provides a mechanism for the system to move uphill on the energy surface to search in a different direction for lower minima. iii) Conformational Search. This method accepts conformers in step 4 that first, are below a preset energy limit, and second, are novel by structural comparison with all previously found conformations. This search would typically be started from a low energy structure found by using method i or ii, and may be the most useful of all [32]. This procedure has the advantage that the search progresses as a walk across conformational space, not downhill in one direction only, allowing a global search of the low energy regions of conformational space. The search also recognizes when duplicate structures are found and provides data regarding the number of times certain minima, including the global, are located. In addition, the search can retrace previously found minima in step 4 above to backtrack out of "corners" in conformational space, increasing search efficiencies. (See ref. 28 for a detailed description.)

The searches require several parameters to be specified that can be configured for a variety of applications. Obviously, the most crucial deal with the torsional degrees of freedom that make up the random walk. These are specified as a list that may include all, or a subset of all the torsions in the molecule. Although the latter provides another method to reduce our search problem, significant assumptions must be made regarding the variable degrees of freedom; requiring some a priori knowledge that could be based on heuristics or chemical intuition. At the start of each cycle (described above) this set of torsions is rotated by a random increment selected from a uniform distribution between -180.0^0 to $+180.0^0$. The maximum rotation can also be limited to yield values that are more typical of those that produce conformational changes in molecules and we found that searches could be performed more efficiently by adjustment of this parameter. The randomized internal coordinates produce a multitude of possible configurations as the search progresses that, once locally optimized, become conformations. These conformations must be examined in step 4 and we have implemented two parameters to handle this assessment. The first is only applicable for type iii searches above and is specified as an upper energy cutoff. The second can be used with all searches and is defined as a torsional tolerance by which conformers are compared. If at least one torsion in the newly optimized structure differs from the corresponding torsion in all previously found conformers by this tolerance, the structure is considered novel and the appropriate action is taken in step 4. The final variable of our search defines the stopping criterion. We have adopted a straightforward method

that ends the search after a specified number of cycles, n, have failed to produce a new conformation as defined by the scheme above.

4. Results and Discussion

Although preliminary studies of linear hydrocarbon models using our global search methodologies had indicated that structures containing 20 beads would be easily searched, it was still surprising that the global minimum of the α-helical hairpin was located with such ease by this approach. Both global minimum searches with and without Boltzmann factor probabilities, determined the structure depicted in Figure 2 to be the lowest energy structure in nearly all independent searches performed. On rare occasions, the search did terminate early, but this is an unavoidable problem for probabilistic methods of this type [11, 32]. To further examine the energy surface, we also performed several conformational searches for low energy minima that, once again, confirmed this structure as the global minimum. The searches were found to be most efficient using a maximum random rotation of 120^0 (-60 to +60). This rotation was initially applied to all torsions in the system each search step, but we found that convergence to the global could be accelerated by choosing a random subset each step, instead. Global optimization generally required less than 2 CPU hours on a Silicon Graphics workstation. The minima found in a representative conformational search are reported in Table 2 along with the number of times the minima were revisited during the search. The conformations were judged to be novel by comparing all torsion angles to a tolerance of 10^0 (see methods above). As might be expected, the global minimum is revisited significantly more often than all others, suggesting that this structure is extremely prominent on the energy surface. In fact, all structures within 1 kcal/mol of the global share this same global fold, or α-helical hairpin structural motif, indicating that this general conformation predominates the density of states at low energy. The RMS differences and average torsions for the conformers are reported in Table 2 for comparison. We did notice, however, that some states slightly higher in energy contained helices that were compressed (the torsions were near eclipsed). This tendency may be due to the use of a 1,4 scaling factor of 1/2 that reduces steric repulsions involved in a torsion angle, but we have not examined this in detail.

The global minimum located is consistent with the results of previous studies in which Monte Carlo, Brownian dynamics, and simulating annealing algorithms were used to explore the conformational space of the model [29, 19, 14]. A slightly different force field was used in those studies, so direct comparisons of structural energies is not possible. In any case, the differences are subtle and are apparently due to the packing of hydrophobic residues. Our structure has one of the helices slightly shifted vertically along the helix axis, allowing all contacts to be fully optimized whereas previous structures have the helices more or less aligned. We noted this tendency in a previous study of this structure, although the conformational change reported in that study was localized to a terminal end rotation. These discrepancies are minor and can be attributed, in part, to the differences in the

potentials used in this study with those used in the past. However, we should not rule out the possibility that our searches have been more exhaustive than those already reported.

It is also of value to examine the conformational states that were located during a typical global minimum search. The sequence of structures are given in Figure 3 along with the corresponding energies. At high energies, it is fairly obvious that substantial secondary structure has been formed in the absence of hydrophobic contacts. These contacts appear to play a larger role at lower energy in stabilizing the formation of the contact region as the structure converges to the global minimum. While one could make various arguments regarding the dynamics and mechanics of protein folding based on this sequence of events, such inferences would be very misleading. First, the force field has been derived to reproduce the folded state only and has not been parameterized to reproduce the folding pathway or structures that populate this pathway. Parameterization of this type is very difficult since very little is known about the intermediate states. Second, the structures located provide a single state that may, or may not, be a reasonable estimate of the average structure populated at that energy. No dynamic information is available. At high energies, these states are least likely to provide useful information since we would expect the density of states to be very large. However, the sequence does provide us with qualitative information that may be of value for future parameterization, especially as more information about the folding pathway becomes known.

The β-barrel structure poses a significantly larger problem to address with our search methods. Not only is the system substantially larger, but the folded state contains several turns, adding to the frustration in the system. The problem, however, was found to be easily tractable with our methodology. Global minimum searches consistently located the β-barrel structure depicted in Figure 4 to be the lowest energy conformation. Once again, the most efficient searches applied a maximum random rotation of 120^0 to a random subset of all torsion angles in the system as noted above. Search times did increase for this system, but not outrageously, showing a modest 5 to 10 fold gain over those noted for the α-helical hairpin.

Given the complexity of this system, we were careful to thoroughly explore the low energy regions of conformational space for new minima before concluding that the global minimum had been found. To do this, we arbitrarily chose several low energy structures, including the suspected global, to seed several conformational searches in which all or a subset of torsions were randomized. While we were successful in locating a multitude of structures having slightly higher energies (within 5 kcal/mol), we found no new, lower energy conformers, indicating that the structure given in Figure 4 is indeed the global minimum. This structure was further found to be the predominate minima on the potential energy surface by analysis of the conformational search statistics and once again (as was the case in the helix calculations), all minima within 1 kcal were in the same global fold. We did find that our low energy searches could be greatly accelerated by localizing the random torsions to the turn regions. Although this limits the search to exploring structures that contain the β-sheet secondary elements apparent in Figure 4, such an approach

has benefits for exploring larger, more complex systems by further reducing the computational problem.

These results are particularly interesting in light of recent work by Honeycutt and Thirumalai in which the folded states of the β-barrel structure were studied using the same model, with very slight differences in the bond stretching and angle bending terms [19]. The focus of that study, however, was not to locate the global minimum, but to examine the conformations that the system became trapped in as the protein chain was cooled from high temperature. They did indicate that the β-barrel structure is suspected to be the global minimum and report an energy of -49.57 kcal/mol for the conformation (using an energy scale of $\epsilon = 1$ kcal/mol from that study). This agrees with our results reported above, although our lowest energy reported is slightly higher (-48.3 kcal/mol). By performing several simple computations we determined that this energy difference stems from the use of a larger bending force constant and flexible bonds in our force field.

One of the main conclusions arrived at in that study was that many minima having similar conformations or global folds exist at significantly different energy levels [19]. We also found this to be true from our conformational searches of low energy structures. For the most part, the structures within 5 kcal/mol of the global minimum differed by slight twisting or angling of the β-sheets from the ideal arrangement. The structures near the end of this spectrum did contain obvious packing differences in the β-sheets, but we would still consider this minor. While it would be inappropriate for us to comment on the relationship of these minima to the folded states observed when cooling from high temperature, as was done by Honeycutt and Thirumalai, the results lend support to the conclusions of that study. However, we should once again point out that arguments based on the behavior of these models at higher energies could be misleading, for the reasons previously cited in this work, so our results should be applied with caution.

4.0.1. All-Atom Model

Recognizing that significantly more accurate force fields exist for simulating protein structures, we decided to apply our methods to a small peptide fragment made up of 9 alanines. This fragment is structurally simple and should have a strong propensity to form an α-helix using the Weiner et al. force field. Unlike our simplified models, this model contains a Coulombic interaction and we chose to use a distant dependent dielectric function to mimic solvent effects; albeit in a non-rigorous fashion. The global minimum search parameters were arrived at in a simple manner; all ϕ/ψ torsion angles (by which protein conformations are described) were defined as the random variables in the search. The maximum rotation was limited to 120^0, as suggested by our previous calculations. The global minimum located is given in Figure 5 and is in an α-helical conformation as indicated by the average ϕ/ψ angles of -48.1 ± 2.6^0 and -58.8 ± 3.8^0, respectively. These compare well with angles found in naturally occurring proteins which are typically between -50^0 and -60^0. The α-helix was also found to be the global minimum for this peptide fragment in

a previous study that used simulated annealing for optimization, supporting our result [38]. We also computed the C_α torsion angles to be approximately 48^0 which is slightly less than those reported earlier in this work and those derived from actual proteins ($\sim 57^0$) [36]. The compression evident from these angles may be due to the use of a non-physical dielectric function, but is not easily explained from our limited results.

5. Conclusion

This study has demonstrated the applicability of a stochastic search method to the global optimization of two protein conformations. The method described takes advantage of the rotational isomerisms associated with conformational interconversions in molecules to reduce the variable degrees of freedom to torsion angles. As previously noted this approach is not new to the field of computational chemistry, however, the system sizes tackled here are much larger than those studied in the past by similar methods and require much less CPU time as well [29, 32]. This success can be attributed to the unique topology of the energy surface produced by the simplified force field. Unlike standard force field representations that produce relatively flat, jagged energy surfaces populated by many low energy conformations, the simplified model is parameterized to capture the average properties of the protein model, producing a smoother energy surface with larger catchment basins [22]. Conformational searches should therefore be more efficient and extendible to much larger systems than previously imagined; given this simplifying behavior. Our results strongly support these contentions and offer new evidence that simplified representations effectively reduce the problem of multiple minima.

Our results have also verified that the simplified force field developed yields the desired conformation of the α-helical hairpin and the β-barrel as the global minimum. This is an important finding since the basis for future force field development, and success in the application of optimization techniques, relies on our ability to reliably predict the folded state (or more generally, the experimentally significant state) as the global minimum. We have also verified the usefulness of a popular all-atom force field for the structure prediction of a small peptide fragment. This finding may not be new, but the observation that we find no structures lower in energy is important. However, the ability of all-atom force fields to predict the correct folded states where interchain contacts and solvent effects become important is still in question and we plan to examine these issues in further work.

While the results presented are encouraging, we still must address several formidable problems before this methodology can be applied to larger, more complex systems. Our limitations stem from two main factors. The first is a general concern to the field of computational chemistry, physics, and math and is due to the use of phenomenological energy functions to predict molecular conformations. Not only is the result of our optimization tied to the reliability of these functions, but success may also be linked to the topology of the energy surface. The second is due to the search methodology and may be the most crucial for us to address in the future.

The use of local, conjugate gradient optimization at each step of the search consumes a significant amount of computation time and is the main factor limiting the application of this methodology to larger structures. While we believe we can take advantage of further simplifications to protein structures it is clear that reductions in local optimization times will be necessary for us to realize our long term goals. We are currently looking into the suitability of the parallel algorithms developed by Rosen and coworkers for continuous minimization for inclusion into our scheme [41, 25]. An interdisciplinary effort is underway.

Acknowledgement

The authors would like to thank Ben Rosen and John Troyer for helpful discussions during the completion of this work. DMF also thanks 3M Pharmaceuticals for a faculty grant in partial support of this work.

References

[1] AMBER 3.0 Revision A: G. Seibel, U. C. Singh, P. K. Weiner, J. W. Caldwell, and P. A. Kollman, Regents of the University of California (1989).

[2] T. L. Blundell and L. N. Johnson, *Protein Crystallography*, Academic Press, New York (1976).

[3] U. Burkert and N. L. Allinger, *Molecular Mechanics*, ACS Monograph No. 177, American Chemical Society, Washington, D. C. (1982).

[4] T. E. Creighton, Stability of Folded Conformations, *Curr. Opin. Struct. Biol.*, **1**, (1991) 5-16.

[5] G. Crippen and M. E. Snow, A 1.8 Å Resolution Potential Function for Protein Folding, *Biopolymers*, **29** (1990) 1479-89.

[6] W. G. Dauben and K. S. Pitzer, *Steric Effects in Organic Chemistry*, ed. by M. S. Newman, Wiley, New York, 1956.

[7] K. A. Dill, Dominant Forces in Protein Folding, *Biochemistry*, **29** (1990) 7133-55.

[8] G. D. Fasman, *Prediction of Protein Structure and the Principles of Protein Conformation*, Plenum Press, New York (1989).

[9] David M. Ferguson and Peter A. Kollman, Can the Lennard-Jones 6-12 Function Replace the 10-12 Form in Molecular Mechanics Calculations?, *J. Comput. Chem.*, **5** (1991) 620-6.

[10] David M. Ferguson, William A. Glauser and Douglas J. Raber, Molecular Mechanics Conformational Analysis of Cyclononane Using the RIPS Method and Comparison with Quantum Mechanical Calculations, *J. Comput. Chem.*, **10** (1989) 903-910.

[11] D.M. Ferguson and D.J. Raber, A new approach to probing conformational space with molecular mechanics: random incremental pulse search, *J. Am. Chem. Soc.*, **111** (1989) 4371-8.

[12] D. M. Ferguson and D. J. Raber, Molecular Mechanics Calculations of Several Lanthanides: An Application of the Random Incremental Pulse Search, *J. Comput. Chem.*, **11** (1990) 1061-71.

[13] P. Flory, *Conformations of Biopolymers*, ed. by G. N. Ramachandran, Vol. 1, Academic Press, New York (1967).

[14] David Garrett, Keith Kastella, and David M. Ferguson, New Results on Protein Folding from Simulated Annealing, *J. Am. Chem. Soc.*, in press.

[15] C. Ghelis and J. Yan, *Protein Folding*, Academic Press, New York (1982).

[16] W. C. Guida, G. Chang, and W. C. Still, An Internal Coordinate Monte Carlo Method for Searching Conformational Space, *J. Am. Chem. Soc.*, **111** (1989) 4379.

[17] A. T. Hagler and B. Honig, On the Formation of Protein Tertiary Structure on a Computer, *Proc. Natl. Acad. Sci. USA*, **75** (1978) 554-8.

[18] M.R. Hoare, Structure and Dynamics of Simple Microclusters, *Advances in Chemical Physics*, **40** (1979) 49-135.

[19] J. D. Honeycutt and D. Thirumalai, The Nature of Folded States of Globular Proteins, *Biopolymers*, **32** (1992) 695-709.

[20] S. Kirkpatrick, C. D. Gelatt, Jr., and M. D. Vecchi, Optimization by Simulated Annealing, *Science*, **220** (1983) 671-80.

[21] I. D. Kuntz, G. M. Crippen, and P. A. Kollman, Calculation of Protein Tertiary Structure, *J. Mol. Biol.*, **106** (1976) 983-94.

[22] M. Levitt and A. Warshel, Computer Simulation of Protein Folding, *Nature (London)*, **253** (1975) 694-8.

[23] Z. Li and H. A. Scheraga, Monte Carlo-minimization Approach to the Multiple Minima Problem in Protein Folding, *Proc. Natl. Acad. Sci. USA*, **84** (1987) 6611.

[24] S. Lifson and A. Warshel, Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules, *J. Chem. Phys.*, **49** (1968) 5116-29.

[25] R.S. Maier, J.B. Rosen, G.L. Xue, A Discrete-Continuous Algorithm for Molecular Energy Minimization, in Proceedings of *IEEE/ACM Supercomputing'91*, pp. 778-786, IEEE Computer Society Press 1992.

[26] S. R. Niketic and K. Rasmussen, *The Consistent Force Field*, ed. by G. Berthier and others, Springer-Verlag, New York (1977). E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations*, McGraw-Hill, New York (1955).

[27] In unpublished work we found the original algorithm of ref. 26 to be inefficient for large, linear hydrocarbons and di- and tri-peptide molecules. Our results indicated that the use of external coordinates was a limiting factor.

[28] P. L. Privalov, Stability of Proteins, *Adv. Protein Chem.*, **33** (1979) 167-241.

[29] A. Rey and J. Skolnick, Comparison of Lattice Monte Carlo Dynamics and Brownian Dynamics of Folding Pathways of α-helical Hairpins, *Chemical Physics*, **158** (1991) 199-219.

[30] F. M. Richards, The Protein Folding Problem, *Scientific American*, **264** (1991) 54-63.

[31] J. S. Richardson, The Anatomy and Taxonomy of Protein Structure, *Adv. Protein Chem.*, **34** (1981) 167-339.

[32] M. Saunders, K. N. Houk, Y. Wu, W. C. Still, M. Lipton, G. Chang, and W. C. Guida, Conformations of Cycloheptadecane. A Comparison of Methods for Conformational Searching, *J. Am. Chem. Soc.*, **112** (1990) 1419-27.

[33] J. Skolnick and A. Kolinski, Computer Simulations of Globular Protein Folding and Tertiary Structure, *Annu. Rev. Phys. Chem.*, **40** (1989) 207-35.

[34] SPASMS is a new molecular dynamics and mechanics program package authored by David Spellmeyer, William Swope, Erik-Robert Evensen, Dave Ferguson, and Peter Kollman. The FORTRAN source code and operations manual are available from the University of California, San Francisco.

[35] William C. Swope and David M. Ferguson, Alternative Expressions for Energies and Forces Due to Angle Bending and Torsional Energy, *J. Comput. Chem.*, **13** (1992) 585-94.

[36] J. M. Troyer and F. E. Cohen, Simplified Models for Understanding and Predicting Protein Structure, in *Reviews in Computational Chemistry*, Vol. 2, ed. by K. B. Lipkowitz and D. B. Boyd, VCH Publishers, New York, 1991.

[37] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, An All Atom Force Field for Simulations of Proteins and Nucleic Acids, *J. Comput. Chem.*, **7** (1986) 230-52.

[38] S. Wilson and W. L. Cui, Applications of Simulated Annealing to Peptides, *Biopolymers*, **29** (1990) 225-35.

[39] C. Wilson and S. Doniach, A Computer Model to Dynamically Simulate Protein Folding: Studies with Crambin, *Prot. Struct. Func. Gen.*, **6** (1989) 193-209.

[40] K. Wuthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York (1986).

[41] G.L. Xue, R.S. Maier, J.B. Rosen, Minimizing the Lennard-Jones Potential Function on a Massively Parallel Computer, in Proceedings of *1992 ACM International Conference on Supercomputing*, pp. 409-416, ACM Press, 1992.

Table 1. Force Field Parameters (PHB: hydrophobic residue; PHL hydropilic; BND: bend; parameters in α-helical hairpin which are the same with those in β-barrel are not shown).

β-barrel [(PHL-PHB-PHL-PHB-PHL-PHB-PHL-PHB-PHL-PHB-PHL)-$(BND)_3$-$(PHB)_9$ -$(BND)_3$ -(PHB-PHL-PHB-PHL-PHB-PHL-PHB-PHL)-$(BND)_3$-$(PHB)_9$]

angle	K_θ	θ_0	torsion	$V_n/2$	γ	n
PHB-PHB-PHB	20.0	105.0	X-PHB-PHB-X	1.2	0	3
PHB-PHL-PHB	20.0	105.0		1.2	0	1
PHL-PHB-PHL	20.0	105.0	X-PHL-PHL-X	1.2	0	3
PHB-PHB-BND	20.0	105.0		1.2	0	1
PHB-PHL-BND	20.0	105.0	X-PHB-PHL-X	1.2	0	3
PHL-PHB-BND	20.0	105.0		1.2	0	1
PHB-BND-BND	20.0	105.0	X-BND-PHB-X	0.2	0	3
PHL-BND-BND	20.0	105.0	X-BND-PHL-X	0.2	0	3
BND-BND-BND	20.0	105.0	X-BND-BND-X	0.2	0	3

bond	K_l	l_0	L-J coefficients	A	B
BND-PHB	100.0	1.0	PHB-PHB	4.0000	4.0000
BND-PHL	100.0	1.0	PHB-PHL	2.6666	-2.6666
BND-BND	100.0	1.0	PHL-PHL	2.6666	-2.6666
PHB-PHB	100.0	1.0	PHB-BND	4.0000	0.0000
PHB-PHB	100.0	1.0	PHL-BND	4.0000	0.0000
PHL-PHL	100.0	1.0	BND-BND	4.0000	0.0000

α-helical hairpin [(PHB-PHB-PHL-PHL-PHL-PHB-PHB-PHL-PHL-PHB-BND-BND--BND-PHL-PHB-PHB-PHL-PHL-PHB-PHB-PHL-PHL-PHB]

angle	K_θ	θ_0	torsion	$V_n/2$	γ	n
BND-BND-PHB	10.0	104.0	X-PHB-PHB-X	1.2	0	3
BND-BND-PHL	10.0	104.0		1.2	240	1
BND-BND-BND	10.0	104.0	X-PHL-PHL-X	1.2	0	3
PHB-PHB-PHL	20.0	91.0		1.2	240	1
PHB-PHL-BND	20.0	91.0	X-PHB-PHL-X	1.2	0	3
PHL-PHL-PHB	20.0	91.0		1.2	240	1
PHL-PHB-BND	20.0	91.0	X-BND-PHB-X	0.2	0	3
			X-BND-PHL-X	0.2	0	3
			X-BND-BND-X	0.2	0	3

Table 2. Results of a representative low energy conformational search of the α-halical hairpin structure ([a] Number in parentheses represents the average deviation. [b] Bend 1 is the first of the four torsional angles (in degrees) which make up the bend region of the structure. [c] Number of times conformation was revisited during the entire search process. [d] RMS differences in cartesian space from the global minimum. Energy is in kcal/mol.)

Conf.	Helical Angle[a]	Bend 1[b]	Bend 2	Bend 3	Bend 4	# Hits[c]	RMS[d]	Energy
1	56.48 (4.69)	171.99	67.61	259.46	46.63	139		-8.5655
2	54.96 (6.02)	196.43	48.46	276.13	92.18	115	0.6529	-7.9933
3	51.37 (11.57)	180.74	51.05	257.00	59.55	29	0.4266	-7.7350
4	54.44 (9.13)	172.36	62.03	259.13	48.48	34	0.2954	-8.1411
5	53.83 (8.34)	174.33	50.89	255.30	64.87	36	0.5097	-7.7277
6	52.00 (9.26)	193.05	51.44	275.63	90.99	18	0.5504	-7.5950
7	53.93 (8.44)	176.07	49.14	255.25	66.47	31	0.5390	-7.7266
8	52.56 (11.71)	189.99	59.66	272.22	79.45	1	0.4432	-7.5250
9	51.13 (11.54)	175.71	93.51	279.73	42.66	13	0.4819	-7.5156

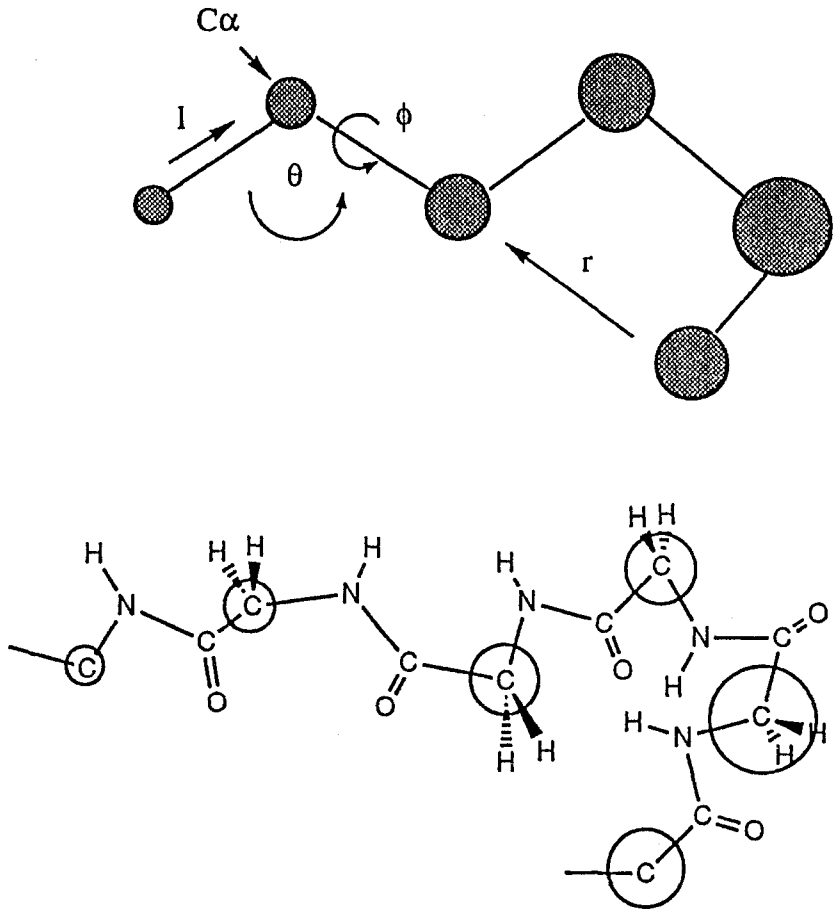

Figure 1. Typical C_α and all-atom representation of protein structures employed by popular models. The primary sequence is defined by teh order of connectivity of the amino acids (or C_α positions).

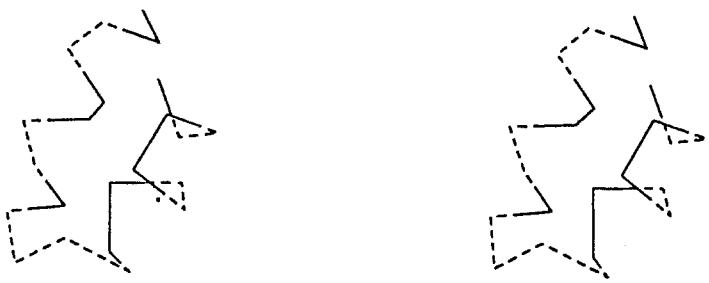

Figure 2. Stereo view of the global minimum of the 22-bead α-helical hairpin. The sequence is given in Table 1. Hydrophobic residues are represented by solid lines, hydrophilic and bend residues by dashed lines.

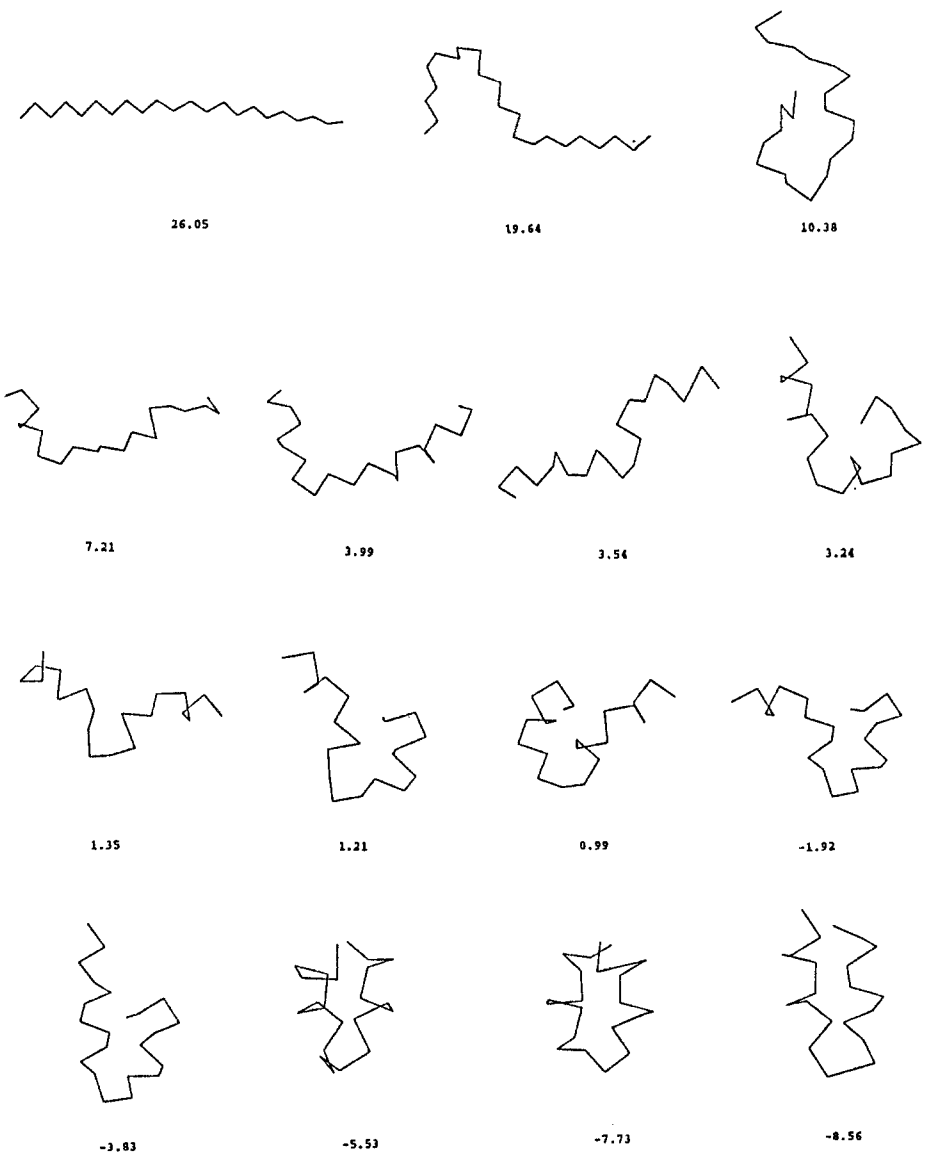

Figure 3. Structures and energies (kcal/mol) of selected conformers located during a typical global minimum search of the α-helical hairpin. The final structure is the global minimum..

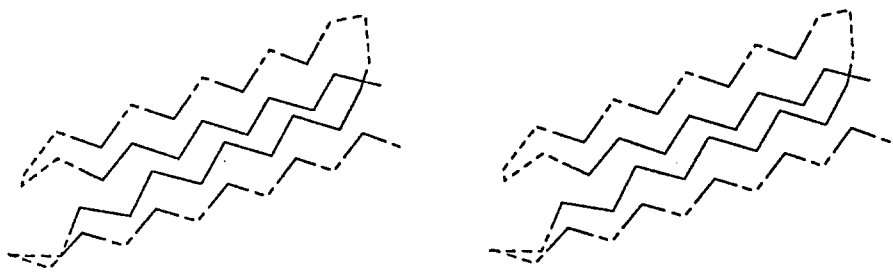

Figure 4. Stereo view of the global minimum of the 46-bead β-barrel. The sequence is given in Table 1. Hydrophobic residues are represented by solid lines, hydrophilic and bend residues by dashed lines.

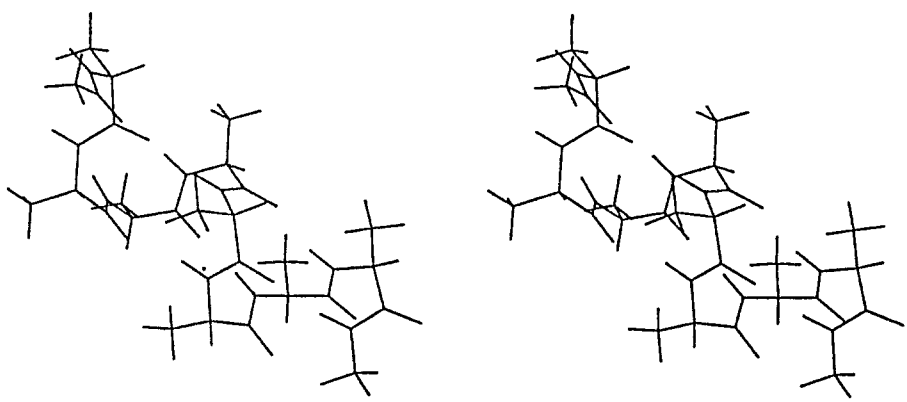

Figure 5. Stereo view of the global minimum of the 9-residue alanyl peptide fragment.