Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Alpha-Helical Topology and Tertiary Structure Prediction in Globular Proteins

Scott R. McAllister and Christodoulos A. Floudas

Abstract— **Within the field of protein structure prediction, the packing of α-helical proteins has been one of the more difficult problems. Distance constraints and topology predictions can be highly useful for reducing the conformational space that must be searched to find a protein structure of minimum conformational energy. We present a novel first principles framework to predict the structure of α-helical proteins. Given the location of the α-helical regions, a mixed-integer linear optimization model maximizes the interhelical residue contact probabilities to generate distance restraints between α-helices. A hybrid global optimization approach combines torsion angle dynamics with a deterministic global optimization technique (αBB) and a stochastic optimization technique (conformational space annealing) to minimize a detailed atomistic-level energy function subject to these constraints. Several improvements to this hybrid algorithm will also be described, including more robust initial point selection, the incorporation of side chain optimization techniques, and a streamlined parallel implementation. The proposed framework does not assume the form of the helices, so it is applicable to all α-helical proteins, including helices with kinks and irregular helices. The predictions of the proposed overall framework on a number of proteins, including the blind prediction of a four helical bundle designed from a combinatorial library, are presented.**

I. INTRODUCTION

The problem of protein structure prediction requires the identification of the three-dimensional structure of a protein given only its primary amino acid sequence. More than 42,000 structures have been experimentally-determined and deposited in the Protein Data Bank (PDB) [1], but there are numerous protein sequences that have yet to be studied or cannot be studied due to limitations of the current experimental techniques. The computational task of protein structure prediction requires searching through a vast conformational space for the native structure. First principles protein structure prediction approaches are limited by both the complexity of interatomic interactions and the overwhelming number of possible conformations that a polypeptide backbone can adopt [2]. The challenge of such a conformational search can be met through the application of powerful algorithms, such as the αBB branch-and-bound method for optimization used in this work [3]–[5], along with experimentally accurate information and models.

This work was supported by the National Science Foundation (R01 GM52032), the National Institutes of Health (R24 GM069736), and the US EPA (GAD R 832721-010). This work has not been reviewed by and does not represent the opinions of the funding agencies.

C. A. Floudas is the Stephen C. Macaleer '63 Professor in Engineering and Applied Science, Professor of Chemical Engineering, Princeton University, Princeton, NJ 08544, USA floudas@titan.princeton.edu

Proteins composed entirely of α-helical secondary structure, α-helical bundles, are one of the more difficult classes of proteins for structure prediction. Methods to predict the packing and structure of these proteins have received increasing attention. One approach derives distance constraints from multiple sequence alignment information to guide Monte Carlo searches for the native structures of small proteins [6]. They showed that native-like protein structures could be predicted by incorporating distance constraints into a tertiary structure prediction algorithm that uses a lattice-based reduced protein model.

A structure prediction strategy of identifying helical residues that form contacts has been proposed [7]. They divided the protein folding problem into two subproblems: creating a fold library of possible protein structures and choosing the best fold in the given library. The input of generic distance restraints of 5.0 to 11.0 Å between contacting helical residues was combined with a distance geometry procedure to generate a library of possible structures for several small helical proteins. The use of generic distance constraints in this range was sufficient to produce many native-like folds for each protein.

Another approach uses the prediction of triangular contact patches at the helix packing interface to restrict the conformational search [2]. They predict three-dimensional structures using torsion angle dynamics with inputs of the predicted secondary structure states and the predicted tertiary contacts. This approach resulted in the production of native-like folds for the majority of their 24 targets, with 14 of the 15 medium-sized proteins (80-100 residues long) having predicted structures within 6.5 Å root-mean-square deviation (RMSD). It has been suggested that an RMSD of about 6 Å is a reasonable target for small proteins [8]. The RMSD values from their approach were only based upon the helical residues of the protein.

An approach has been proposed that samples the entire conformational space of helical proteins [9]. This method imposes rules from a database of existing structures for the enumeration of all the geometrically possible three dimensional arrangements of the helices. A small, but positive, effect on the number of near native conformers resulted from the use of a refinement procedure.

α-helical protein structures can also be predicted through more general protein structure prediction algorithms. These methods are extensively outlined in two recent reviews [10]–[12].

1-4244-1498-9/07/$25.00 ©2007 IEEE.

II. METHODS

Given a prediction of the α-helix locations within a protein (which can be attained through a variety of methods [13]–[16]), we attempt to predict the three dimensional structure of this protein. The main thrusts of this approach are (A) the prediction of hydrophobic residue pairs that form interhelical contacts using a mixed-integer linear programming formulation and (B) the tertiary structure prediction of the protein using a deterministic hybrid global optimization algorithm to minimize an atomistic level energy function.

A. α-helical Contact Prediction

The problem of α-helical topology prediction can be separated into three main parts. First, a dataset of known α-helical bundle proteins must be created. From this dataset, hydrophobic-to-hydrophobic interhelical contact probabilities can be developed. These probabilities are then used in a two-level mixed-integer linear programming framework to maximize the residue contact probabilities subject to a number of constraints. A summary of the important details of the approach are presented here. A more detailed description of the method is available elsewhere [17].

1) Dataset Selection: A database PDB set of 318 helical protein structures was compiled to generate probabilities for specific hydrophobic-to-hydrophobic residue contacts between helices of the same helical protein. The set of proteins was gathered from a number of sources: 20 from Table 2 of Zhang et al. (2002) [2]; 7 from Table 1 of Huang et al. (1999) [7]; 62 from the CATH database [18]; and 229 from the PDB Select 25 Database [19].

2) Probability Development: The probabilities are calculated for both a contact at position i (denoted as a PRIMARY contact) and any associated contacts in the helical wheel position (denoted as WHEEL contacts). Figure 1 shows the PRIMARY contacts and a subset of the possible WHEEL contacts for an antiparallel helical interaction.

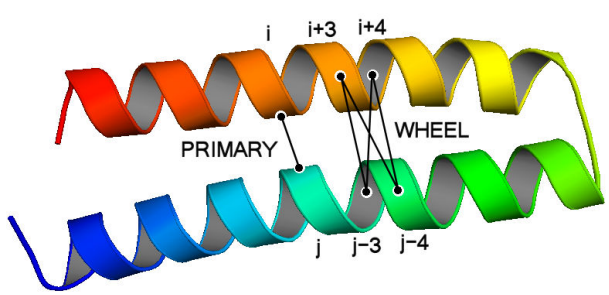

Fig. 1. Two interacting α-helices in the protein 1ROP (PDB). The helices here interact in an antiparallel fashion. The hydrophobic residues i and j form a PRIMARY contact and the residues (i+3), (i+4) can each interact with $(j-3)$, $(j-4)$ to form WHEEL contacts if both residues of a given pair are hydrophobic. All figures have been created with PyMol.[20]

By avoiding assumptions about the form of the helix, such as representing the helix as a simple cylinder, the proposed method is able to address the most difficult cases, including irregular helices that bend or kink. In this way, representing the problem as a set of PRIMARY and WHEEL contacts provides a significant advantage over other methods.

The number of occurrences of hydrophobic-to-hydrophobic minimum interhelical distances was counted to determine the occurrence frequencies of each of the 36 possible hydrophobic pairs. The frequency for each pair was then split into two groups, parallel and antiparallel, based on the relative direction of the interacting helices. The PRIMARY probabilities are calculated as the occurrence frequency of a specific hydrophobic pair with a specific directionality divided by the total number of hydrophobic-to-hydrophobic contacts.

The interacting helices in the database PDB set were also used to generate conditional WHEEL probabilities. Given that two α-helices have a valid minimum distance PRIMARY contact, the conditional probability that the residues on the same side of the helical wheel form a hydrophobic-to-hydrophobic contact was determined. These probabilities were calculated by considering the number of hydrophobic-to-hydrophobic WHEEL contacts and the total number of possible WHEEL contacts for every specific helix to helix interaction individually and calculating the probability for each interhelical residue contact by averaging over the total number of such contacts after the entire database PDB set has been considered.

3) Interhelical Contact Model: Two mixed-integer linear programming (MILP) optimization problems were formulated to predict interhelical contacts. The first MILP problem, denoted as the Level 1 Model, identifies the set of the interhelical PRIMARY contacts that results in the maximum probability value subject to the constraints of the model. This set of contacts is enhanced by further selecting the set of most probable WHEEL contacts predicted in the Level 2 Model. This second model also provides a method to choose between results of the Level 1 model with equal probabilities.

The Level 1 Model is formulated with two sets of binary variables that represent both the existence of any contact between two helices as well as residue specific contacts. The binary variables y^a_{mn} and y^p_{mn} are activated if the helices m and n of the same protein interact (have at least one PRIMARY contact) in an antiparallel or a parallel fashion, respectively. When the hydrophobic residue pair (i,j) forms a PRIMARY contact, the binary variables w^{mn}_{ij} become active.

Equation 1 is the objective function of the Level 1 Model. The goal is to maximize the sum of the probable hydrophobic contacts by activating combinations of the variables corresponding to residue contacts, w^{mn}_{ij}, and variables corresponding to helical contacts,y^a_{mn} and y^p_{mn}. The probability of a parallel or antiparallel contact is indicated by the parameters $p^p_{ij;mn}$ or $p^a_{ij;mn}$, respectively.

$$\max \quad \sum_m \sum_n y^a_{mn} \cdot \sum_i \sum_j w^{mn}_{ij} \cdot p^a_{ij;mn} \qquad (1)$$

$$+ \quad \sum_m \sum_n y^p_{mn} \cdot \sum_i \sum_j w^{mn}_{ij} \cdot p^p_{ij;mn}$$

$$y^a_{mn}, y^p_{mn}, w^{mn}_{ij} = \{0,1\} \qquad (2)$$

The resulting product of binary variables in this equation is nonlinear. The objective function can be reformulated as a linear objective function by introducing a second pair of variables, as described elsewhere [17], [21].

This objective function is subject to a number of constraints to restrict the predicted set of contacts to a physically meaningful arrangement. First, the variables are constrained by a number of residue contact rules. Every hydrophobic amino acid of helix m, i, is restricted to have at most one PRIMARY contact with another hydrophobic amino acid of helix n, j. Any pairs of contacts (i, j) and (i', j') cannot both be selected as PRIMARY contacts if (i', j') is in a WHEEL position to (i, j). The difference between these two pairs of contacts is also restricted to limit the size of kinks in the protein backbone that result from a differing separation between (i and i') and (j and j').

There are a number of restrictions that must be imposed on helical interactions as well. A given helix, m, may only be involved in a limited number of helix-to-helix interactions, based on the number of hydrophobic residues it contains and their positions. A minimum number of loop residues must separate two α-helices to yield helical interactions of a given orientation. Two helices must be restricted to interact in a parallel or antiparallel fashion, if they interact at all, but not both. Several constraints are also imposed to restrict the topology of helical interactions based upon transitive rules.

Further constraints can be introduced to extend the utility of the model. An upper limit on the number of PRIMARY contacts that two interacting helices m and n can have is by specified a parameter that can be varied. A second parameter is introduced to eliminate a number of helical interactions from the Level 1 solutions to effectively loosen the helical packing. Obtaining a rank-ordered list of the most likely sets of helical contacts is more desirable than a single solution. Integer cut constraints are added into the model after each successive solve to exclude the previous solution from the feasible solution space.

The Level 2 objective function is formulated to select the most probable hydrophobic WHEEL contacts based upon a fixed set of hydrophobic PRIMARY contacts. This maximization is also formulated as a mixed-integer linear programming problem, which is constrained to allow a maximum of one WHEEL contact per PRIMARY contact.

B. Tertiary Structure Prediction

1) Previous Approach: Knowledge of the secondary structure of a protein can be used to set bounds on the dihedral angles and distances within the protein. The helical residues are constrained with an i, $(i+4)$ C^{α}-C^{α} distance of 5.5 to 6.5 Å where i is a given helical residue and $(i+4)$ is the residue four places away in the primary sequence. This restraint represents the backbone hydrogen bond that results from helix formation. For the residues that are predicted to be helical, the backbone dihedral angles are restricted to [-90,-40] for ϕ and [-60,-10] for ψ.

Predicted tertiary contacts, such as those established by the interhelical contact model, can also be imposed as distance constraints. These constraints are imposed to bound the C^{α}-C^{α} distance to between 5.0 Å and 12.0 Å for the PRIMARY and WHEEL contacts that have been predicted.

The tertiary structure prediction stage is a combination of the αBB global optimization algorithm, a stochastic global optimization method, and a molecular dynamics approach in torsion-angle space [22]. The basic formulation is the minimization of the force field energy over torsion angle space, subject to upper and lowering bounding constraints on these angles as well as distance constraints in Cartesian space. A torsion angle representation of the model significantly decreases the size of the independent variable set, while only modestly increasing the model complexity.

The use of the αBB global optimization algorithm [3], [4], [23]–[25] guarantees the identification of the global minimum solution by a convergence of upper and lower bounds on the potential energy minimum. The upper bound on the global minimum is obtained by constrained nonlinear minimization on any protein structure. The lower bound is determined by creating a valid convex underestimating function and identifying its minimum function value. The algorithm converges by successively partitioning regions of conformational space at every level of a branch and bound tree.

The rapid identification of feasible, low energy conformers through torsion-angle dynamics methods can significantly improve the performance of the αBB method. In addition, stochastic optimization methods may further improve the performance of the upper bounding approach of the formulation. One such hybrid global optimization method, described as an alternating hybrid, has been recently introduced [22], [26]. It combines the deterministic αBB approach with the stochastic approach of conformational space annealing [27]–[31]. Conformational space annealing balances genetic algorithm techniques of mutations and crossovers with simulated annealing to identify low energy conformers. A hybrid algorithm of this form retains the deterministic guarantees of convergence while yielding a much more efficient search for the native state [26].

A parallel implementation of this hybrid global optimization approach has been developed [22]. In this approach, two processors are allocated as control processors, to control both the αBB iterations and the CSA iterations. The αBB control processor maintains the lower and upper bounds as well as a list of the lower-bounding subregions for the αBB approach. A subset of the overall processors are assigned αBB subproblems by the control processor, receiving a bounded region, and performing torsion angle dynamics followed by constrained local minimizations of the convex lower bounding problem and of the upper bounding problem. The CSA control processor establishes a bank of conformers from the intermediate solutions of the αBB subproblems. The remaining processors receive trial conformations from the CSA control processor from this initial bank and perform constrained local minimizations. The algorithm can then be terminated when the control processors establish the convergence of the lower and upper bounding problems or

when an iteration or time limit is hit.

The ASTRO-FOLD methodology has been successfully applied to a varied set of proteins throughout the range of small to medium-sized proteins [5]. The recent success of the ASTRO-FOLD method in a double blind prediction of a four-helix bundle reinforces the value of the approach [32].

2) Algorithmic Improvements: Several improvements to the alternating hybrid algorithm have been implemented, including more robust initial point selection, the incorporation of side chain optimization techniques, and a streamlined parallel implementation. The identification of initial feasible points is critical to the success and efficiency of the constrained nonlinear minimization algorithm. The basic torsion angle dynamics routine in the initial implementation of the framework has been replaced by a more detailed torsion angle dynamics annealing procedure by interfacing with the CYANA software package [33]. The number of structures that satisfy both the distance and dihedral angle constraints after this routine is significantly increased, especially in the cases of distance constraints with tight bounds or large numbers of distance constraints.

A rotamer optimization stage has been introduced in the algorithmic framework prior to the constrained nonlinear minimization. The goal of this stage is to remove any steric clashes that may exist between protein side chains, especially after mutations or crossovers have been applied through the conformational space annealing algorithm. The algorithm is implemented as a greedy selection of rotamer modifications that lead to energetic improvements. With the backbone fixed in its initial conformation, the rotamers are varied one residue at a time and a replacement is made if any energetic improvement is realized. In this way, rotamer optimization acts as an efficient local minimizer, with no guarantee that the optimal set of rotamers is chosen.

Several modifications of the parallel implementation of the alternating hybrid algorithm have also been made. The duties of the two control processors have been combined and assigned to a single overall control processor. This combination is effective because the work distributed to the work processors requires several orders of magnitude more time for processing than for the communication associated with delivering the work and receiving the results. Very little delay of the work distribution due to communication was observed. The initial implementation of the algorithm assigned a fixed set of processors to specific duties (αBB iterations, CSA iterations). Since no CSA work exists until a bank has been filled by low energy structures from the aBB algorithm, these processors would sit idle during the first rounds of the algorithm. This inefficiency has been removed by initially assigning all processors to perform iterations of the αBB method. Once the initial CSA bank of conformers is filled, the control processor sends a message to a subset of the work processors indicating that their duties have changed.

III. RESULTS

The proposed algorithmic framework was applied to 3 α-helical systems. Two of these proteins, 1HTA and 1NRE, are part of the set of test proteins for the initial evaluation of the interhelical contact prediction model [17]. The prediction of the third protein, S836, was done as blind test to evaluate the method. All three structure predictions were run for 60 hours on 50 3.0GHz processors of a Beowulf cluster.

A. α-helical proteins

The Level 1 and Level 2 interhelical contact prediction models were applied to 1HTA, a 67 amino acid protein. This protein contains 3 α-helices and is characterized by a long 28 residue second helix. The topology predictions are shown in Table I.

TABLE I

INTERHELICAL CONTACT PREDICTIONS FOR 1HTA

PRIMARY Contact	PRIMARY Distance	WHEEL Contact	WHEEL Distance	Helix-Helix Interaction
5I-28L	9.1	-	-	1-2 A
46L-62L	8.4	-	-	2-3 A

The contact model is able to predict two tertiary contacts that are nearby in the native structure. The first pair, residues 5 and 28, connect helices 1 and 2 with the hydrophobic pair of Isoleucine (I) to Leucine (L). A second hydrophobic contact of Leucine to Leucine is predicted for residue 46 in helix 2 with residue 62 in helix 3. These two tertiary contacts, in addition to 48 α-helical dihedral angle bounds, 14 residue-specific loop dihedral angle bounds and 37 local $i, (i+4)$ intrahelical distance restraints were imposed for the overall tertiary structure prediction algorithm. The lowest energy structure identified by the algorithm is presented in Figure 2, along with its alignment to the native 1HTA structure. The RMSD between these two structures is 6.70Å. Among the 23465 structures identified with an energy within 100 kcal/mol of the lowest energy structure was a structure with an RMSD to the native structure of 2.58Å.

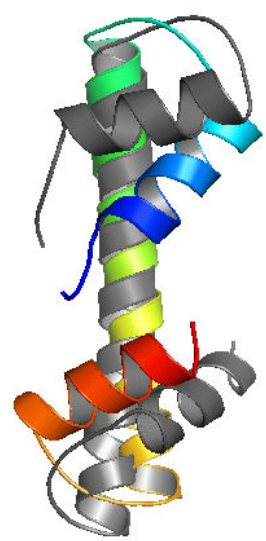

Fig. 2. Lowest energy predicted structure of 1hta (color) versus native 1hta (gray)

The interhelical contact prediction models were also applied to another 3 helix bundle protein, 1NRE, that contains 81 amino acids. The topology predictions are shown in Table II. The predictions of the contact model include four PRIMARY contacts and two WHEEL contacts. Five of the six contacts are below 10.0Å in the native 1NRE structure. It is noteworthy that the model is able to identify a parallel interaction between helices 1 and 3, validating the use of helix directionality as an important aspect of the model.

TABLE II

INTERHELICAL CONTACT PREDICTIONS FOR 1NRE

PRIMARY Contact	PRIMARY Distance	WHEEL Contact	WHEEL Distance	Helix-Helix Interaction
25L-49L	6.0	28L-45L	9.1	1-2 A
28L-83V	12.7	-	-	1-3 P
45L-85L	9.3	49L-81L	8.1	2-3 A
51I-77L	9.3	-	-	2-3 A

The six hydrophobic-hydrophobic tertiary residue contacts, along with 54 α-helical dihedral angle bounds, 12 residue-specific loop dihedral angle bounds and 42 local $i,(i+4)$ intrahelical distance restraints were imposed for the overall tertiary structure prediction of 1NRE. The structures presented in Figure 3 are the lowest energy structure and its alignment to the native 1NRE structure. These two structures have an RMSD of 6.63Å. The algorithm identified 9913 structures that have an energy within 100 kcal/mol of the lowest energy structure. The structure in this ensemble with the lowest RMSD to the native structure has an RMSD value of 3.52Å.

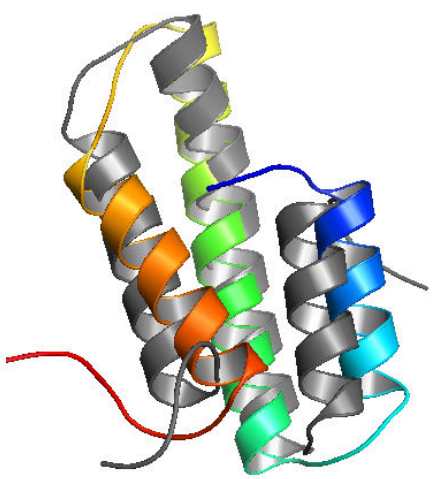

Fig. 3. Lowest energy predicted structure of 1nre (color) versus native 1nre (gray)

B. Blind studies

The hydrophobic effect is a major driving force for protein folding. As a result, de novo proteins have been designed using combinatorial libraries and a binary patterning of polar and nonpolar amino acids [34]. One such protein that was constructed in this fashion is S836, a 102 residue protein that has been shown to contain elements of α-helical structure by circular dichroism (CD) spectra [35]. Furthermore, this protein was shown to be quite stable by monitoring the CD spectra while increasing the concentration of guanidine hydrochloride, a denaturant.

The prediction of interhelical contacts for S836 was expected to be a challenge because the hydrophobic residues were chosen entirely from combinatorial libraries instead of being explicitly designed to have particular side chain interactions. The topology predictions are shown in Table III.

TABLE III

INTERHELICAL CONTACT PREDICTIONS FOR S836

PRIMARY Contact	PRIMARY Distance	WHEEL Contact	WHEEL Distance	Helix-Helix Interaction
5L-99L	7.2	9L-96L	7.7	1-4 A
9L-41I	7.3	12L-37L	8.8	1-2 A
12L-92I	7.0	-	-	1-4 A
16L-33V	9.7	-	-	1-2 A
37L-68L	9.3	40V-64F	6.6	2-3 A
47F-60M	10.7	44I-64F	10.4	2-3 A
57L-96L	7.4	61M-92I	9.4	3-4 A
71I-82L	7.6	68L-85V	8.2	3-4 A

The predictions of the contact model yielded a total of 14 hydrophobic-to-hydrophobic residue contacts, split into eight PRIMARY contacts and six WHEEL contacts. All but two of these predicted contacts have distances below 10.0Å in the native S836 structure. These 14 tertiary contacts were imposed for the blind structure prediction of S836, along with 77 α-helical dihedral angle bounds, 20 residue-specific loop dihedral angle bounds and 61 local $i,(i+4)$ intrahelical distance restraints. Figure 4 shows a superposition of the lowest energy structure and its alignment to the S836 native structure (once we had supplied our prediction, we received the native structure for evaluation purposes). The predicted structure has a RMSD of 2.84Å from the native structure. The algorithm also identified an ensemble of 18528 structures that have an energy within 100 kcal/mol of the lowest energy structure. The structure in this ensemble with the lowest RMSD to the native structure has an RMSD value of 2.39Å.

IV. CONCLUSIONS AND FUTURE WORKS

A novel optimization model for predicting interhelical contacts in α-helical globular proteins was used to generate tertiary distance constraints. These constraints, combined with dihedral angle and distance constraints based on secondary structure states, can then be used as part of a tertiary structure prediction algorithm. The success of this algorithmic approach has been illustrated on two different α-helical bundle proteins with 3 helices and a blind prediction of a 4-helix bundle designed using combinatorial libraries.

The ASTRO-FOLD tertiary structure prediction algorithm has been improved by streamlining the parallel implementation, introducing an improved torsion angle dynamics routine for more robust initial point selection, and using a rotamer optimization stage to attain quick energetic improvements prior to applying nonlinear constrained local minimization

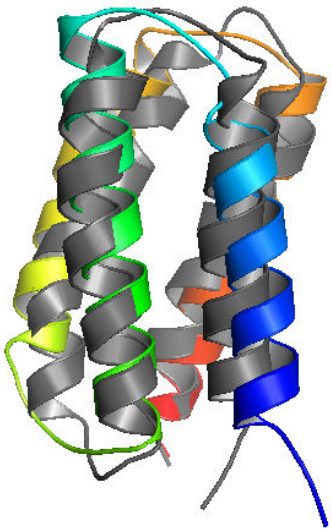

Fig. 4. Lowest energy predicted structure of S836 (color) versus native S836 (gray)

techniques. A larger set of protein structure predictions will be performed to further validate the method.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge Abigail T. Go and Professor Michael H. Hecht for providing the sequence of S836 and withholding the experimental structures until after we made our predictions.

REFERENCES

[1] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nuc. Acids Res*, vol. 28, pp. 235–242, 2000.

[2] C. Zhang, J. T. Hou, and S. H. Kim, "Fold prediction of helical proteins using torsion angle dynamics and predicted restraints," *Proc. Nat. Acad. Sci. USA*, vol. 99, pp. 3581–3585, 2002.

[3] C. S. Adjiman, S. Dallwig, C. A. Floudas, and A. Neumaier, "A global optimization method for general twice-differentiable NLPs. i. theoretical advances," *Comp. Chem. Eng.*, vol. 22, pp. 1137–1158, 1998b.

[4] C. S. Adjiman, I. P. Androulakis, and C. A. Floudas, "A global optimization method for general twice-differentiable NLPs. ii. implementation and computational results," *Comp. Chem. Eng.*, vol. 22, pp. 1159–1179, 1998a.

[5] J. L. Klepeis and C. A. Floudas, "Astro-fold: A combinatorial and global optimization framework for ab initio prediction of three-dimensional structures of proteins from the amino acid sequence," *Biophysical Journal*, vol. 85, pp. 2119–2146, 2003b.

[6] A. R. Ortiz, A. Kolinski, and J. Skolnick, "Fold assembly of small proteins using Monte Carlo simulations driven by restraints derived from multiple sequence alignments," *J. Mol. Bio.*, vol. 277, pp. 419–448, 1998.

[7] E. S. Huang, R. Samudrala, and J. W. Ponder, "Ab initio fold prediction of small helical proteins using distance geometry and knowledge-based scoring functions," *J. Mol. Bio.*, vol. 290, pp. 267–281, 1999.

[8] B. A. Reva, A. V. Finkelstein, and J. Skolnick, "What is the probability of a chance prediction of a protein structure with an rmsd of 6 angstrom?" *Fold Des*, vol. 3, pp. 141–147, 1998.

[9] B. Fain and M. Levitt, "A novel method for sampling alpha-helical protein backbones," *J. Mol. Bio.*, vol. 305, pp. 191–201, 2001.

[10] C. A. Floudas, "Research challenges, opportunities and synergism in systems engineering and computational biology," *AIChE J*, vol. 51, pp. 1872–1884, 2005.

[11] C. A. Floudas, H. K. Fung, S. R. McAllister, M. Mönnigmann, and R. Rajgaria, "Advances in protein structure prediction and de novo protein design: A review," *Chem Eng Sci*, vol. 61, pp. 966–988, 2006.

[12] C. A. Floudas, "Computational methods in protein structure prediction," *Biotechnol. Bioeng.*, 2007, in press.

[13] B. Rost and C. Sander, "Prediction of protein structure at better than 70% accuracy," *J. Mol. Bio.*, vol. 232, pp. 584–599, 1993.

[14] D. T. Jones, "Protein secondary structure prediction based on position-specific scoring matrices," *J. Mol. Bio.*, vol. 292, pp. 195–202, 1999.

[15] B. Rost, "Review: Protein secondary structure prediction continues to rise," *J Struct Biol*, vol. 134, pp. 204–218, 2001.

[16] J. L. Klepeis and C. A. Floudas, "Ab initio prediction of helical segments in polypeptides," *J. Comput. Chem.*, vol. 23, pp. 245–266, 2002.

[17] S. R. McAllister, B. E. Mickus, J. L. Klepeis, and C. A. Floudas, "A novel approach for alpha-helical topology prediction in globular proteins: Generation of interhelical restraints," *Prot Struct Funct Bioinf*, vol. 65, pp. 930–952, 2006.

[18] C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton, "CATH - a hierarchic classification of protein domain structures," *Structure*, vol. 5, pp. 1093–1108, 1997.

[19] U. Hobohm and C. Sander, "Enlarged representative set of protein structures," *Prot. Sci.*, vol. 3, pp. 522–524, 1994.

[20] W. L. DeLano, *The PyMol Molecular Graphics System*. San Carlos, CA USA: DeLano Scientific, 2002, http://www.pymol.org.

[21] C. A. Floudas, *Nonlinear and Mixed-integer Optimization: Fundamentals and Applications*. Oxford University Press, 1995.

[22] J. L. Klepeis, M. J. Pieja, and C. A. Floudas, "Hybrid global optimization algorithms for protein structure prediction: Alternating hybrids," *Biophysical Journal*, vol. 84, pp. 869–882, 2003a.

[23] C. S. Adjiman, I. P. Androulakis, C. D. Maranas, and C. A. Floudas, "A global optimization method, αBB, for process design," *Comp. Chem. Eng.*, vol. 20, pp. S419–S424, 1996.

[24] C. S. Adjiman, I. P. Androulakis, and C. A. Floudas, "Global optimization of MINLP problems in process synthesis and design," *Comp. Chem. Eng.*, vol. 21, pp. S445–S450, 1997.

[25] I. P. Androulakis, C. D. Maranas, and C. A. Floudas, "αBB: A global optimization method for general constrained nonconvex problems," *J. Glob. Opt.*, vol. 7, pp. 337–363, 1995.

[26] J. L. Klepeis, M. J. Pieja, and C. A. Floudas, "A new class of hybrid global optimization algorithms for peptide structure prediction: Integrated hybrids," *Comp. Phys. Comm.*, vol. 151, pp. 121–140, 2003b.

[27] J. Lee, H. A. Scheraga, and S. Rackovsky, "Conformational analysis of the 20-residue membrane-bound portion of melittin by conformational space annealing," *Biopolymers*, vol. 46, pp. 103–115, 1998.

[28] ——, "New optimization method for conformational energy calculations on polypeptides: Conformational space annealing," *J. Comput. Chem.*, vol. 18, pp. 1222–1232, 1997.

[29] J. Lee and H. A. Scheraga, "Conformational space annealing by parallel computations: Extensive conformational search of met-enkephalin and the 20-residue membrane-bound portion of melittin," *Int J Quantum Chem*, vol. 75, pp. 255–265, 1999.

[30] D. Ripoll, A. Liwo, and H. A. Scheraga, "New developments of the electrostatically driven Monte Carlo method: Tests on the membrane-bound portion of melittin," *Biopolymers*, vol. 46, pp. 117–126, 1998.

[31] J. Lee, J. Pillardy, C. Czaplewski, Y. Arnautova, D. R. Ripoll, A. Liwo, K. D. Gibson, R. J. Wawak, and H. A. Scheraga, "Efficient parallel algorithms in global optimization of potential energy functions for peptides, proteins and crystals," *Comp. Phys. Comm.*, vol. 128, pp. 399–411, 2000.

[32] J. L. Klepeis, Y. N. Wei, M. H. Hecht, and C. A. Floudas, "Ab initio prediction of the three-dimensional structure of a de novo designed protein: A double-blind case study," *Prot Struct Funct Bioinf*, vol. 58, pp. 560–570, 2005.

[33] P. Güntert, C. Mumenthaler, and K. Wüthrich, "Torsion angle dynamics for NMR structure calculation with the new program DYANA," *J. Mol. Bio.*, vol. 273, pp. 283–298, 1997.

[34] M. H. Hecht, A. Das, A. Go, L. H. Bradley, and Y. Wei, "De novo proteins from designed combinatorial libraries," *Prot. Sci.*, vol. 13, pp. 1711–1723, 2004.

[35] Y. Wei, T. Liu, S. L. Sazinsky, D. A. Moffet, I. Pelczer, and M. H. Hecht, "Stably folded de novo proteins from a designed combinatorial library," *Prot. Sci.*, vol. 12, pp. 92–102, 2003.