Pacific Symposium on Biocomputing 7:601-612 (2002)

AB INITIO FOLDING OF MULTIPLE-CHAIN PROTEINS

J.A. SAUNDERS, K.D. GIBSON, AND H.A. SCHERAGA
Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, U.S.A.

Our previous methodology for *ab initio* prediction of protein structure is extended here to treat multiple-chain proteins. This involved modification of our united-residue (UNRES) force field and our Conformational Space Annealing (CSA) Global Optimization procedure. Good results have been obtained for both a four-and a three-helix protein from the CASP3 exercise.

1 Introduction

Ever since Anfinsen's formulation of the thermodynamic hypothesis for protein folding[1], attempts have been made to compute the structure of a native globular protein as the global minimun of its free energy. These include knowledge-based approaches [such as homology modeling[2-7], threading[3,6,8], and combinations of these together with secondary-structure prediction[9]], and *ab initio* methods[10,11]. The latter are based only on energy, without the aid of knowledge-based information; the purpose of the *ab initio* approach is to provide an understanding of how pairwise and multibody (cooperative) interactions lead to the folded structure.

In the *ab initio* approach, it is computationally impossible at present to predict the 3D structure of an *all-atom* protein by energy minimization, Monte Carlo, or molecular dynamics procedures. However, by breaking the problem into its component parts, it is possible to achieve the realization of this computational goal. Therefore, a hierarchical approach to the computation of protein structure has been developed in our laboratory[10-14]. This hierarchy involves the following six steps, the key stage of which is the global optimization of an off-lattice simplified chain:

1. Using a virtual-bond representation of the polypeptide chain, described by a united-residue potential (UNRES)[13,15-17], and an efficient procedure (Conformational Space Annealing, CSA)[18,19] to search the conformational space of this virtual-bond chain rapidly, we obtain a family of low-energy UNRES structures. The combination of UNRES and CSA narrows the region of conformational space in which the global minimum is likely to lie, which can be achieved at this stage with the simplified virtual-bond model but not with an all-atom model.
2. Next, this very restricted part of conformational space is searched by first converting the virtual-bond chains of the low-energy structures obtained in the united-residue calculations to an all-atom representation, using our dipole-path method[20].

3. The backbone conformation is then optimized, subject to C^{α}- C^{α} distance constraints from the parent united-residue structure, by using the Electrostatically Driven Monte Carlo (EDMC) method[21], which incorporates elements of our Monte Carlo-plus-Minimization (MCM)[22,23] and Self-Consistent Electric Field (SCEF)[24] procedures.
4. All-atom side chains are then attached subject to the condition of non-overlap.
5. Loops and disulfide bonds are treated with the empirical loop-closing potential of ECEPP[25], or with an exact procedure[26,27].
6. Final energy refinement is carried out with the ECEPP/3 all-atom force field[25], plus the SRFOPT surface-hydration model[28], with gradual reduction of the C^{α}- C^{α} distance constraints of the parent united-residue structure (until they vanish at the end of the procedure).

For the success of the hierarchical algorithm outlined above, the united-residue force field must be able to capture the essential features of the interactions in proteins at a coarse-grain level. However, it is equally important to have a reliable all-atom force field that can reproduce the fine details of the all-atom chain in stage 6. Recently, we began to use our algorithms for the global optimization of crystal structures as tools to refine the parameters of all-atom potentials (so that the global minimum of the potential energy function is close to the experimental structure, and its energy is close to the experimental enthalpy of sublimation). These global-optimization algorithms were originally based on the deformation of the potential energy surface, and include the Diffusion Equation Method (DEM)[29], the Distance Scaling Method (DSM)[30] and its improved descendant, the Self-Consistent-Basin-to-Deformed-Basin Mapping (SCBDBM) method[31] which involves back-and-forth deformations and reversals, until self-consistency, with perturbations at each stage of reversal. More recently, we use the Conformation-Family Monte Carlo (CFMC) method[32] for crystal structure calculations.

Our methodology, and other types of approaches such as those of Skolnick and Baker[9,33] (and references cited therein), have had partial success in the CASP3 and CASP4 exercises.

Heretofore, these methods have been applied primarily to single-chain proteins. Recently, we have started to develop methods to treat multiple-chain proteins. Since these methods extend the above single-chain approaches, we begin by summarizing our treatment of single-chain proteins, and then will describe our methodology to treat multiple-chain proteins.

2. Summary of the UNRES Force Field

2.1 UNRES Geometry

UNRES employs a simplified representation of polypeptide chains in which backbone peptide groups and entire side chains are represented by united atoms. The polypeptide backbone is modeled as a sequence of α-carbon atoms linked by virtual bonds 3.8 Å in length (corresponding to *trans* peptide groups), with a united peptide group located at the midpoint of each virtual bond. Attached to each backbone α-carbon is a united side-chain ellipsoid whose size and distance from the backbone are determined by the amino acid identity of the side chain. Each united residue has four degrees of freedom; the angle between successive backbone virtual bonds (θ) and the dihedral angle about each virtual-bond (γ) determine the backbone conformation, and two angles α and β determine the orientation of the side chain relative to the backbone. The united peptide groups and side chains serve as interaction sites for the force field, while the α-carbons are present only to define the geometry of the chain.

2.2 UNRES Force Field

The UNRES force field is derived as a restricted free energy (RFE) function, which corresponds to averaging the energy over the degrees of freedom that are neglected in the united-residue model. It is expressed by eq. (1):

$$U = \sum_{i<j} U_{SC_iSC_j} + w_{SCp}\sum_{i\neq j} U_{SC_ip_j} + w_{el}\sum_{i<j-1} U_{p_ip_j} + w_{tor}\sum_i U_{tor}(\gamma_i) + w_b\sum_i U_b(\theta_i) + w_{rot}\sum_i U_{rot}(\alpha_{SC_i},\beta_{SC_i}) + \sum_{m=2}^{N} w_{corr}^{(m)} U_{corr}^{(m)} \quad (1)$$

where $U_{SC_iSC_j}$ is the interaction energy between side chains, $U_{SC_ip_j}$ is an excluded-volume potential between side chains and peptide groups, $U_{p_ip_j}$ is the energy of average electrostatic interaction between peptide groups, U_{tor} is the intrinsic energy of rotation about the virtual C^{α}- C^{α} bonds, U_b and U_{rot} are the bending energy of the virtual-bond angles and the energy of different rotameric states of the side chains, respectively, and U_{corr} is the multibody or correlation energy arising from the loss of degrees of freedom when computing the restricted free energy.

The terms $U_{SC_iSC_j}$, U_b, and U_{rot} were parameterized based on distributions within a set of 195 non-homologous structures from the Protein Data Bank. The

excluded-volume term $U_{SC_i p_j}$ was parameterized to reproduce the correct backbone geometry in short model helices and sheets. The form of $U_{p_i p_j}$ was obtained by averaging the electrostatic interactions of the peptide groups, and it was parameterized by fitting the average restricted free energy surface of two interacting peptide groups, calculated with the all-atom ECEPP/3[25] force field. The expression for the torsional potential U_{tor} results from the cluster expansion of the RFE, and it has been parameterized by fitting to ECEPP/3[25] RFE surfaces.

Because UNRES is a coarse-grain force field, in which the interactions between the united side chains and peptide groups are mean-field potentials, multibody terms (U_{corr}) that capture the underlying physics of the hidden degrees of freedom are required for successful *ab initio* structure prediction. The RFE is

$$F(\mathbf{X}) = -RT \ln\left(\frac{1}{V_{\mathbf{Y}}}\int_{\Omega_{\mathbf{Y}}} \exp\left[-E(\mathbf{X};\mathbf{Y})/RT\right]dV_{\mathbf{Y}}\right) \tag{2}$$

where $\mathrm{E}(\mathbf{X};\mathbf{Y})$ is the all-atom ECEPP/3 energy function, $\mathbf{X}$ is the set of UNRES degrees of freedom, $\mathbf{Y}$ is the set of degrees of freedom over which the average is computed (e.g. the positions and orientations of solvent molecules, the side-chain dihedral angles, etc.), R is the gas constant, T is the absolute temperature, $\mathbf{\Omega_Y}$ is the region of the $\mathbf{Y}$ subspace over which the integration is carried out, and $V_{\mathbf{Y}}$ is the volume of this region. It can be expressed as a sum of cluster cumulant functions[34], which correspond to increasing order of correlations between component energy terms. The most significant correlation terms are those for electrostatic-local interaction correlations. By expanding these functions into a generalized cumulant series, approximate analytical expressions can be obtained; the lowest order cumulants are sufficient to capture the essential features of the cumulant functions. The multibody term U_{corr} is thus a set of electrostatic-local correlations of up to sixth order, the forms of which were derived from the cumulant expansion, and the parameters of which were found by fitting to appropriate ECEPP/3 RFE surfaces.

The w's are constant weights that balance the contributions of the different kinds of interactions. They are determined by simultaneous Z-score-and-gap optimization[35, 36] on multiple targets. This procedure produces a set of weights that maximize both the ratio of the difference in energy between the native structure and the mean energy of non-native structures to the standard deviation of the energy distribution of the non-native structures, and also the difference in energy between the native structure and the lowest energy non-native structure. The native structure is actually a set of UNRES structures within a certain rmsd cutoff of the experimental structure, and the non-native structures are generated through a global conformational search.

3. Summary of CSA

Conformational Space Annealing is a powerful global optimization method that has been used successfully with UNRES in the CASP blind structure prediction exercises. It is a variation on a genetic algorithm in that it maintains a population of structures and combines portions of the conformations of existing "parent" structures to generate new "offspring" structures, which then replace conformations in the population if they have lower energy.

CSA is distinguished from other optimization methods, however, in its use of a cutoff distance that decreases over the course of a search. Also, conformations are always locally minimized in energy using Gay's Secant-type Unconstrained Minimization Solver (SUMSL)[37]. A distance measure is defined which measures the similarity between two conformations; for UNRES, this distance is defined as the average angular difference between two conformations for all backbone dihedral angles in the chain. If a new structure is produced that is within the cutoff distance of an existing structure, the new one either replaces the old similar structure, if the new one has lower energy, or it is rejected. The effect of this cutoff distance is that, at the beginning of a search, when the cutoff distance is large, CSA will maintain a population that is sparse in the conformation space; i.e. the search will be global. As the cutoff distance decreases, a higher degree of similarity will be allowed between structures in the population, which effectively searches smaller regions for low-energy conformations. CSA thus begins as a global search for potential fold families, and ends as a more local search for the lowest-energy representatives of the best fold families.

4. Performance of Hierarchy in CASP3 and CASP4

The foregoing methodology has been tested in the CASP3[10,13,14] and CASP4[11] exercises. In their evaluation of our performance in CASP3, Orengo et al.[38] cited our predicted submission for HDEA (target 61) as "most impressive…., using more classical *ab initio* methods….(which use) no information from sequence alignments, secondary structure prediction, or threading".

Our predictions in CASP3 were all submitted for largely α-helical targets because we could not predict β structures at that time, i.e. in 1998. Since then, we have improved our UNRES force field with additional terms in the cumulant expansion of the free energy, and have achieved *partial* success in predicting β structures[11] in CASP4, i.e., we were able to predict *part* of the β structure in β and α/β proteins. Work is now in progress to extend our heirachical procedure (with improvements in our cumulant-based UNRES force field, in our global optimization procedures, and in our all-atom potential function) to try to extend our current

predictions for 100-residue α proteins with an rmsd of 4-6Å to 250-residue α, β, and α/β proteins with an rmsd of 2-3Å.

5. Extensions to multiple-chain Proteins

5.1 Extensions to UNRES

To apply the UNRES force field to multiple-chain proteins, new energy terms had to be added to represent the interactions between separate chains. The forms, parameters, and weights of these interchain terms are the same as the single chain terms because the physical bases of the intra- and interchain interactions are the same. However, the purely local interactions (i.e. U_b, U_{rot}, and U_{tor}) have no interchain counterparts. The single-chain multibody term U_{corr} includes correlations for residues that are both adjacent (e.g. electrostatic-local correlations around a β-turn) and separated in sequence (e.g. electrostatic-local correlations between β-sheet strands). The multiple-chain U_{corr} term includes nonadjacent electrostatic-local correlations between chains. The new UNRES energy function is:

$$\begin{aligned} U = & \left(\sum_k \sum_{i<j} U_{SC_{k,i}SC_{k,j}} + \sum_{k<l} \sum_{i,j} U_{SC_{k,i}SC_{l,j}} \right) \\ & + w_{SCp} \left(\sum_k \sum_{i \neq j} U_{SC_{k,i}p_{k,j}} + \sum_{k<l} \sum_{i,j} U_{SC_{k,i}p_{l,j}} \right) \\ & + w_{el} \left(\sum_k \sum_{i<j-1} U_{p_{k,i}p_{k,j}} + \sum_{k<l} \sum_{i,j} U_{p_{k,i}p_{l,j}} \right) + w_{tor} \sum_k \sum_i U_{tor}(\gamma_{k,i}) \\ & + w_b \sum_k \sum_i U_b(\theta_{k,i}) + w_{rot} \sum_k \sum_i U_{rot}(\alpha_{SC_{k,i}}, \beta_{SC_{k,i}}) \\ & + \sum_{m=2}^{N} w_{corr}^{(m)} \left(\sum_k U_{corr}^{(m)} + \sum_{k<l} U_{corr,nonadj}^{(m)} \right) \end{aligned} \tag{3}$$

where the indices k and l denote chains, the indices i and j denote residues, and the $U_{corr,nonadj}$ term represents the components of U_{corr} corresponding to correlations between nonadjacent residues. In addition to the angles that define the

internal conformation of a chain, a set of Euler angles and a translational vector are required for each chain to specify the packing arrangement of the multiple-chain protein.

5.2 Symmetry optimizations

Many multiple-chain proteins contain symmetries in the spatial arrangements of the individual chains. The presence of such symmetries allows optimizations that can greatly increase the speed of energy evaluation, gradient calculation, and local minimization of the energy[39]. For example, in a tetramer of identical chains with four-fold rotational symmetry, the intrachain energy of only a single chain need be computed because each chain has the same internal conformation. Although there are six possible interchain interaction energies between four chains, in this symmetric case only two are unique. Likewise, fewer gradient calculations have to be carried out because some contributions to the gradient can be computed simply by rotating the gradients from other symmetry-related interchain interactions. For identical chains with symmetric packing, the number of independent variables is greatly reduced, and the topology of the energy surface is simplified, which decreases the number of energy evaluations required for local minimization of the energy.

The types of symmetry considered here are rotational, screw, one-chain affine, and two-chain affine symmetries[39]. The packing variables in these symmetries consist of the orientation of the symmetry axis, the displacement of the axis from the origin, the radial distance of the first chain in the group from the axis, the rotation angle to the first chain in the axial frame, the axial shift to the first chain, the Euler angles of the first chain, the axial shift between chains, the axial rotation between chains, and, for two-chain affine symmetry, the Euler angles and position in the axial frame of the second chain. A protein may consist of several independent groups of chains, each with its own symmetry. Also, two or more of these groups can be constrained to share a symmetry axis.

5.3 Extensions to CSA

New ways of generating conformations have been added for the purpose of exploring the space of packing arrangements. The initial, random bank of non-clashing structures is generated with random packing as well as random internal chain conformations. In the stage of CSA when new “offspring” conformations are created by combining the variables of existing conformations, several methods are used to generate new chain packings. A new packing is initially copied from a seed conformation and then perturbed by taking packing variables from another

conformation. The perturbations can be implemented in several sub-sets of the packing variables.

In this study, the distance between conformations for multiple chains was the same as that defined for single chains, viz., the average angular difference between two conformations for all backbone dihedral angles. The primary goal was to determine the folded structure of a monomer (in its oligomeric state), and this measure reveals the backbone similarity of monomer conformations even in systems with different packings. Also, in the work reported here, the packing variables were highly coupled to the internal conformations being packed; thus, structures with differing internal conformations generally adopted different packing arrangements.

6. Results

Multiple-chain UNRES has been tested successfully on two homo-oligomeric systems that were targets in the CASP3 exercise. In both cases, the search for the native fold was carried out only at the united-residue level, without conversion to an all-atom representation.

6.1 Retro-GCN4 Leucine Zipper

The *retro*-GCN4 leucine zipper[40] (target 84 in CASP3) is an α-helical tetramer with 37 residues per chain. Two independent runs were carried out on both the tetramer and the monomer by itself. For the tetramer, four-fold rotational symmetry was assumed for both runs. The experimental structure shows that the native tetramer is very nearly, but not quite, rotationally symmetric, and consists of two slightly different monomer conformations.

In two runs of the monomer, totaling 76,000 minimizations, the lowest energy structure found (–174.2 kcal/mol) has a C^{α} coordinate rmsd for residues 2–37 (the first residue is not resolved in the experimental structure) of 13.4 Å from the native monomer structure. Of the 100 structures resulting from the two runs, none are closer to the native monomer conformation than 11.6 Å. The native monomer is a single, long helix, and in all cases the single-chain force field breaks the helix into multiple segments.

Two runs of the tetramer, totaling 177,000 minimizations, resulted in a lowest energy structure of –777.3 kcal/mol. This lowest energy structure has an rmsd of only 2.34 Å from the native tetramer (Fig. 1). A tetramer slightly closer to native (1.98 Å) is found as the ninth lowest energy, ~8 kcal/mol higher than the lowest. Comparing the lowest-energy tetramer results with the experimental results at the monomer level yields a monomer rmsd of 2.23 Å or 2.29 Å, depending on the experimental monomer used in the comparison. The monomer of the structure

closest to the native tetramer has an rmsd of 1.80 Å or 1.92 Å with the experimental monomers.

While the lowest energy structure from the monomer runs has an energy of –174.2 kcal/mol, the intrachain energy per chain for the lowest-energy tetramer is –149.7 kcal/mol. However, the packing energy due to interchain contacts decreases the total energy per chain in the tetramer to –194.3 kcal/mol, illustrating that the presence of additional chains results in a structure with monomers that are individually less stable than the isolated monomer conformation, but are packed such that favorable interchain contacts more than offset that loss in energy.

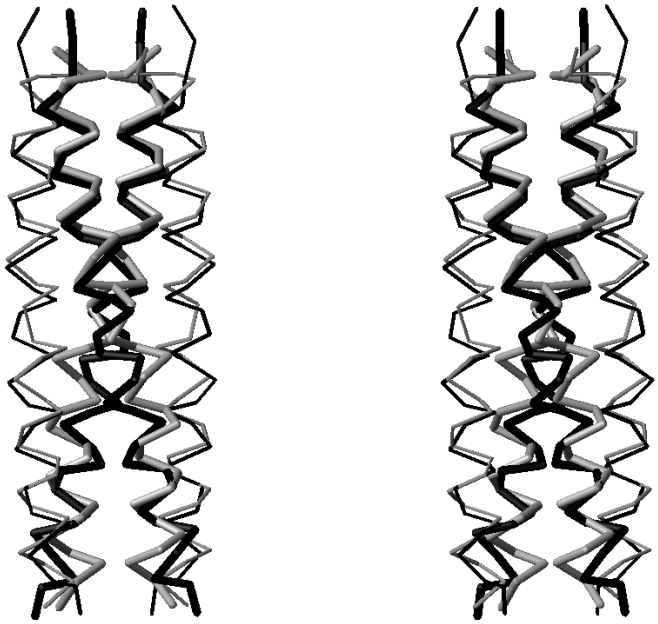

Fig. 1. Stereo view of computed structure of *retro*-GCN4 leucine zipper superposed on the x-ray structure[40].

6.2 Synthetic Domain-Swapped Dimer

Target 73 from CASP3 was a synthetic domain-swapped dimer, a designed α-helical dimer of 48 residues per chain that forms a three-helix bundle[41]. Two independent runs were carried out on both the dimer and the monomer by itself. Two-fold rotational symmetry was assumed for both dimer runs.

In two runs of the monomer, totaling 134,000 minimizations, the lowest energy structure found (–245.1 kcal/mol) has an rmsd from the native of 11.1 Å. A minimal-tree clustering of the 150 resulting structures at 3 Å yields 41 conformational families. The native family was found as the eighth lowest of the 41 family clusters. The lowest energy representative of this family, at –237.1 kcal/mol, is the closest to the native, with an rmsd of 2.79 Å. As was the case with the *retro*-GCN4 leucine zipper, the single-chain force field tends to break long helices to achieve better packing of the chain against itself in the absence of interchain contacts from the second monomer.

Two runs of the dimer, totaling 190,000 minimizations, found a lowest energy structure of –526.32 kcal/mol. This lowest-energy structure is a member of the native family, at 5.65 Å from the experimental dimer. Comparing only monomers of the lowest-energy structure and the native gives an rmsd of 3.33 Å. Within this same family, another structure at –517.99 kcal/mol is found that has an rmsd of only 3.16 Å from the native dimer. The closest monomer to the native belongs to another member of the native family, with an energy of –520.2 kcal/mol; the monomer rmsd for this structure is 2.19 Å.

The intrachain energy per chain in the lowest-energy structure is –224.08 kcal/mol, compared to the –245.1 kcal/mol lowest-energy isolated monomer. However, the total energy per chain is –263.16 kcal/mol, again demonstrating the importance of interchain interactions in stabilizing the native monomer conformation.

7. Summary and Conclusions

Without considering the interchain interactions, it is not possible to obtain a correct prediction of a multiple-chain protein. By extending our UNRES and CSA algorithms to take these interchain interactions into account, it is now possible to obtain reasonably accurate predictions of the structures of multiple-chain proteins, as illustrated here for the *retro*-GCN4 leucine zipper and a synthetic domain-swapped dimer.

Acknowledgments

We are indebted to J. Pillardy, C. Czaplewski, A. Liwo and J. Lee for helpful discussions. This work was supported by NIH grant GM-14312 and NSF grant MCB00-03722. Support was also received from the National Foundation for Cancer Research. J. A. Saunders was an NIH Biophysics trainee.

References

1. C.B. Anfinsen, *Science*, **181**, 223 (1973).
2. P.K. Warme, F.A. Momany, S.V. Rumball, R.W. Tuttle and H.A. Scheraga, *Biochemistry*, **13**, 768 (1974).
3. T.A. Jones and S. Thirup, *EMBO J*., **5**, 819 (1986).
4. D.A. Clark, J. Shirazi and C.J. Rawlings, *Prot. Eng*. **4**, 751 (1991).
5. M.J. Rooman and S.J. Wodak, *Biochemistry*, **31**, 10239 (1992).
6. M.S. Johnson, J.P. Overington and T.L. Blundell, *J. Mol. Biol*., **231**, 735 (1993).
7. D. Fischer, D. Rice, J.U. Bowie and D. Eisenberg, *FASEB J*., **10**, 126 (1996).
8. M.J. Sippl, *J. Comput.-Aid. Mol. Des*., **7**, 473 (1993).
9. J. Skolnick, A. Kolinski and A.R. Ortiz, *J. Mol. Biol*., **265**, 217 (1997).
10. J. Lee, A. Liwo, D.R. Ripoll, J. Pillardy, J.A. Saunders, K.D. Gibson and H.A. Scheraga, *Intl. J. Quantum Chem*., **77**, 90 (2000).
11. J. Pillardy, C. Czaplewski, A. Liwo, J. Lee, D.R. Ripoll, R. Kazmierkiewicz, S. Oldziej, W.J. Wedemeyer, K.D. Gibson, Y.A. Arnautova, J. Saunders, Y.-J. Ye and H.A. Scheraga, *Proc. Natl. Acad. Sci., U.S.A.*, **98**, 2329 (2001).
12. A. Liwo, M.R. Pincus, R.J. Wawak, S. Rackovsky, and H.A. Scheraga, *Protein Science*, **2**, 1715 (1993).
13. A. Liwo, J. Lee, D.R. Ripoll, J. Pillardy and H.A. Scheraga, *Proc. Natl. Acad. Sci., U.S.A.*, **96**, 5482 (1999).
14. J. Lee, A. Liwo, D.R. Ripoll, J. Pillardy and H.A. Scheraga, *Proteins: Structure, Function and Genetics*, **Suppl. 3**, 204 (1999).
15. A. Liwo, S. Oldziej, M.R. Pincus, R.J. Wawak, S. Rackovsky and H.A. Scheraga, *J. Comput. Chem*., **18**, 849 (1997).
16. A. Liwo, M.R. Pincus, R.J. Wawak, S. Rackovsky, S. Oldziej and H.A. Scheraga, *J. Comput. Chem*., **18**, 874 (1997).
17. A. Liwo, R. Kazmierkiewicz, C. Czapleswki, M. Groth, S. Oldziej, R.J. Wawak, S. Rackovsky, M.R. Pincus and H.A. Scheraga, *J. Comput. Chem*. **19**, 259 (1998).
18. J. Lee, H.A. Scheraga and S. Rackovsky, *J. Comput. Chem*., **18**, 1222 (1997).
19. J. Lee and H.A. Scheraga, *Intl. J. Quantum Chem.*, **75**, 255 (1999).
20. A. Liwo, M.R. Pincus, R.J. Wawak, S. Rackovsky and H.A. Scheraga, *Protein Science*, **2**, 1697 (1993).
21. D.R. Ripoll, A. Liwo and H.A. Scheraga, *Biopolymers*, **46**, 117 (1998).
22. Z. Li and H.A. Scheraga, *Proc. Natl. Acad. Sci., U.S.A.*, **84**, 6611(1987).
23. Z. Li and H.A. Scheraga, *J. Molec. Str. (Theochem)*, **179**, 333 (1988).
24. L. Piela and H.A. Scheraga, *Biopolymers* **26**, S33 (1987).
25. G. Némethey, K.D. Gibson, K.A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey and H.A. Scheraga, *J. Phys. Chem*., **96**, 6472 (1992).

26. K.D. Gibson and H.A. Scheraga, *J. Comput. Chem*., **18**, 403 (1997).
27. W.J. Wedemeyer and H.A. Scheraga, *J. Comput. Chem*., **20**, 819 (1999).
28. J. Vila, R. L. Williams, M. Vasquez and H.A. Scheraga, *Proteins: Structure, Function, and Genetics*, **10**, 199 (1991).
29. L. Piela, J. Kostrowicki and H. A. Scheraga, *J. Phys. Chem.*, **93**, 3339 (1989).
30. J. Pillardy and L. Piela, *J. Phys. Chem*. **99**, 11805 (1995).
31. J. Pillardy, A. Liwo, M. Groth and H.A. Scheraga *J. Phys. Chem*., **B103**, 7353 (1999).
32. J. Pillardy, C. Czaplewski, W. J. Wedemeyer and H.A. Scheraga, *Helv. Chim. Acta*, **83**, 2214 (2000).
33. P.M. Bowers, C.E.M. Strauss and D. Baker, *J. Biomolec. NMR*, **18**, 311 (2000).
34. A. Liwo, C. Czaplewski, J. Pillardy and H. A. Scheraga *J. Chem. Phys*., **115**, 2323 (2001).
35. J. Lee, D.R. Ripoll, C. Czaplewski, J. Pillardy, W.J. Wedemeyer and H.A. Scheraga *J. Phys. Chem. B*., **105**, 7291 (2001).
36. J. Pillardy, C. Czaplewski, A. Liwo, W.J. Wedemeyer, J. Lee, D.R. Ripoll, P. Arlukowicz, S. Oldziej, Y.A. Arnautova and H.A. Scheraga *J. Phys. Chem. B*, **105**, 7299 (2001).
37. D.M. Gay, *ACM Trans. Math. Software*, **9**, 503 (1983).
38. C.A. Orengo, J.E. Bray, T. Hubbard, L.LoConte, and I. Sillitoe, *Proteins: Struct., Funct., Genet.*, **Suppl. 3**, 149 (1999).
39. K.D. Gibson and H.A. Scheraga, *J. Comput. Chem*., **15**, 1414 (1994).
40. P.R.E. Mittl, C. Deillon, D. Sargent, N. Liu, S. Klauser, R.M. Thomas, B. Gutte and M.G. Grütter, *Proc. Natl. Acad. Sci. USA*, **97**, 2562 (2000).
41. N.L. Ogihara, G. Ghirlanda, J.W. Bryson, M. Gingery, W.F. DeGrado and D. Eisenberg, *Proc. Natl. Acad. Sci. USA*, **98**, 1404 (2001).