Building of self-avoiding lattice models of proteins using a SCF-optimization

B.A. Reva[1,2], A.V. Finkelstein[3], D.S. Rykunov[4] and A.J. Olson[1]

[1]The Scripps Research Institute, 10666 North Torrey Pines Road, MB5, Ca 92037, USA; [2]Institute of Mathematical Problems of Biology; [3]Institute of Protein Research; [4]Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, 142292, Pushchino, Moscow Region, Russian Federation.

Introduction: Lattice models[1-3] are widely used for computational investigation of protein folding. In these models each chain link is imbedded in the points of a 3D lattice. However, it is not a trivial task to build a good lattice model of a protein molecule. The model has to satisfy the conditions of chain connectivity and self-avoidance while simultaneously minimizing the error function: $F_{err}(R_1, \ldots, R_N) = \sum_{i=1}^{N} (x_i - R_i)^2$ a difference between the coordinates of a protein chain $\{x_i\}$ and its lattice model $\{R_i\}$. The function F_{err} can be treated as a chain energy: the smaller this energy, the better the model. Thus, one needs to minimize F_{err} over $R_1,\ldots,R_N$ without violation of connectivity of the model chain. The condition of chain connectivity can be incorporated into the error function by the terms: $U_i(R_i, R_{i+1}) = \begin{cases} 0, \text{ if } & \left|\left|R_i - R_{i+1}\right| - d_i\right| \le \gamma \\ +\infty, & \text{otherwise} \end{cases}$. Here $i=1,\ldots,N-1$, d_i is the actual distance between links i and $i+1$ (for a protein α-carbon chain $d = 3.8$ **Å**), and γ limits an allowed deviation of inter-link distance in the model from its actual value. (In this study we use only $\gamma=0$).

The model chain energy, $F_{err} + \sum_{i=1}^{N-1} U_i(R_i, R_{i+1})$, can be minimize now by dynamic programing[4,7] because all the energy terms are "local": they contain either coordinate of only one link, or of two adjacent links and they do not contain "long-range" interaction terms. The global minimization algorithm[4] guarantees the minimal RMS deviation between on-lattice and off-lattice structure for continuous models, but it does not guarantee that the models found will not self-intersect.

In this work we suggest the SCF-optimization method for search for self-avoiding lattice models.

Methods: To avoid self-intersections we add the repulsive terms to the error function: $\Delta F_{err}(R_1, \ldots, R_N) = A \sum_{i>j+1}\sum \delta(R_i - R_j)$, where $A>>1$ is the value of the penalty for a link intersection; $\delta(x)=1$, *if* $x \le D_{min}$ and $\delta(x)=0$, if $x > D_{min}$.

Minimization of the $\Delta F + F + \sum U_i$ is done iteratively using Self-Consistent Field theory[6,7] where the pairwise link repulsions are represented as a **3D** field acting on the links. A 3D field acting on a link ***i*** in a lattice point ***R*** on the **s**-th iteration step is: $\Psi^s{}_i(R) = (x_i - R)^2 + A \sum_{j \neq i,\, i-1,i+1} \left[\sum_{R_j} W_j^{s-1}(R_j)\, \delta(R - R_j) \right]$, where $W_j^{s-1}(R_j)$ is the probability that a link ***j*** occupies a lattice point $\boldsymbol{R_j}$. The term $\boldsymbol{(x_i\text{-}R)^2}$ comes directly from the error function $\boldsymbol{F_{err}}$, the second term represents is a potential produced by the repulsion of links (in the first step of the iteration this term is not considered). Under the connectivity conditions the chain energy has the form: $\sum_i U_i(R_i, R_{i+1}) + \sum_i \Psi_i^s(R_i)$, which enables computation of the probabilities $\boldsymbol{W_i(R)}$ using the statistical mechanics of chain molecules[7,8]. At the next step, these $\boldsymbol{W_i(R)}$ functions are used to compute the modified probabilities $W_i^s(R) = \left(W_i(R) + W_i^{s-1}(R) \right)/2$, and then the modified field potentials, and so forth.

The field is refined by iteration to become self-consistent with the chain distribution[7], (i.e. the field value no longer changes with iterations), and, finally, dynamic programing is used to find the optimal lattice model as the "lowest-energy" chain pathway in this SCF-field[8].

Results: We have tested the method on the coarsest (the most difficult) lattices for proteins of a variety of structural types and show that the proposed method is adequate for building self-avoiding lattice models of proteins with a low RMS deviation from the actual structures.

References

1 Covell D., Jernigan R. *Conformations of folded proteins in restricted spaces*. Biochemistry **29**:3287-3294, 1990.

2 Kolinski A., Skolnick J. *Monte Carlo Simulation of Protein Folding. I. Lattice Model and Interaction Scheme.* Proteins **18**:338-352, 1994.

3 Hind D., Levitt M. 1994. *Exploring Conformational Space with Simple lattice Model for Protein Structure.* J.Mol.Biol. **243**:668-682.

4 Rykunov D.S., Reva, B.A., Finkelstein, A.V. *Accurate general method for lattice approximation of three-dimensional structure of a chain molecule.* Proteins **22:**100-109, 1995.

5 Reva B.A., Rykunov D.S., Olson A.J., Finkelstein A.V. *Constructing of lattice models of protein chains with side groups.*J.Com.Biol., in press, 1995.

6 Kubo R., *Statistical Mechnics*. North-Holland, Amsterdam, 1965

7 Finkelstein A.V., Reva B.A. *A search for the most stable folds of protein chains.* Nature **351**:497-499, 1991.

8 Finkelstein A.V., Reva B.A. *A search for the stable state of a short chain in a molecular field.* Prot. Engn. **5**:617-624,1992