JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 9, Number 4, 2002

Pp. 613–620

# Minimum Energy Configurations of the 2-Dimensional HP-Model of Proteins by Self-Organizing Networks

BERRIN YANIKOGLU[1] and BURAK ERMAN[1,2]

## ABSTRACT

**We use self-organizing maps (SOM) as an efficient tool to find the minimum energy configurations of the 2-dimensional HP-models of proteins. The usage of the SOM for the protein folding problem is similar to that for the Traveling Salesman Problem. The lattice nodes represent the cities whereas the neurons in the network represent the amino acids moving towards the closest cities, subject to the HH interactions. The valid path that maximizes the HH contacts corresponds to the minimum energy configuration of the protein. We report promising results for the cases when the protein completely fills a lattice and discuss the current problems and possible extensions. In all the test sequences up to 36 amino acids, the algorithm was able to find the global minimum and its degeneracies.**

**Key words:** protein folding, structure prediction, self organizing maps.

## 1. INTRODUCTION

A PROTEIN IS A CHAIN OF AMINO ACID RESIDUES that folds into a specific three-dimensional structure (*native state* or *tertiary structure*), under favorable physiological conditions. The three-dimensional structure of a protein greatly determines the protein's functionality, but determining the three-dimensional structure experimentally is difficult and time consuming. Several research groups are working on computational methods for predicting the native state of a protein from its *primary structure*, the sequence of amino acid residues along the chain, commonly believed to uniquely determine the native structure. However, protein structure prediction is a difficult problem even with simplified models. If one assumes three possible rotational isomeric states for each amino acid, there are $3^N$ possible configurations, only one of which is the minimum free energy conformation determining the native state. Furthermore, the energy and the stability of the protein greatly depends on nonlocal interactions between amino acids far from each other along the chain. In fact, it is shown that finding the minimum free energy conformation of even a simplified model of a protein, specifically the 2D HP model studied in this paper, is an NP-complete problem (Crescenzi *et al.*, 1998).

[1]Laboratory of Computational Biology, Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli, 81474 Istanbul, Turkey.
[2]Present address: Departments of Chemistry & Mechanical Engineering, Koc University, Rumelifeneri Yolu, Sariyer 80910 Istanbul, Turkey.

Molecular dynamics simulation is the most detailed, but also the most time consuming computational technique applied to determine the three-dimensional minimum energy structure of real proteins. Recently, a small protein of 36 residues was brought from a random initial configuration to within a root mean square deviation of 1.5 Ångstroms from the known minimum energy configuration; however, the calculations took more than a year on a supercomputer (Duan *et al.*, 1998). Faster but coarser-scale simulation techniques such as the Monte Carlo technique and simulated annealing are also used in protein structure prediction (Li and Scheraga, 1987). Due to the complexity of the problem, scientists have also studied various simplifications. In particular, Dill *et al.* (1995) introduced the two- and three-dimensional hydrophobic-polar (HP) model which abstracts the problem by grouping the 20 possible amino acids used in proteins as hydrophobic (H) or hydrophilic (polar or P). As a further simplification, the lattice HP-models restrict the possible conformations to self-avoiding paths on a lattice where each amino acid residue along the chain occupies a vertex of the lattice. Two- and three-dimensional square lattice models using protein chains with hydrophobic and polar centers are the most widely studied simple models. Recently, the protein folding problem has been formulated in the form of the Traveling Salesman Problem (TSP) and approached using the elastic net algorithm (Ball *et al.*, 2001). This approach was successful in predicting the minimum energy structures of several real proteins and the simulation time scaled linearly with chain length.

In the present paper, we apply the Self-Organizing Map (SOM) algorithm to predicting the minimum energy configurations of simple 2D HP models of proteins. To our knowledge, this is the first use of SOM in the literature for determining compact minimum energy structures of proteins. SOM has been widely used for tasks such as clustering of protein sequences (Hanke *et al.*, 1996), classification of protein families (Andrade *et al.*, 1997), determination of structural motifs of protein backbones (Schuchhardt *et al.*, 1996) and gene expression (Tamayo *et al.*, 1999). It has also been successfully applied to optimization problems such as the well known Traveling Salesman Problem (Angeniol *et al.*, 1988; Jeffries *et al.*, 1994; Budinich 1996; Altinel *et al.*, 2000). Also, recently, Wriggers *et al.* (1998) applied the SOM algorithm for the first time to characterize the shape and density distribution of the occupied volume of large-scale protein assemblies in order to reconstruct blurred data from electron microscope images. On the other hand, Bohr and Brunak approach the protein conformation problem from a TSP perspective, but using relaxation techniques (Bohr *et al.*, 1989).

The present approach is similar in spirit to the elastic net approach of Ball *et al.* (2001) Specifically, given a sequence of residues of type H or P, we try to find the minimum energy configuration of the protein by posing the problem as that of finding the Hamiltonian path of the lattice nodes, maximizing the attraction between the hydrophobic residues. Similar to the elastic net algorithm, the simulation time for the present approach scales linearly with sequence length, making it possible to apply the algorithm a large number of times and to choose the predicted structure with the least energy. The reason for choosing a 2D lattice HP model, rather than three-dimensional real proteins, as was the case in the elastic net problem, is to illustrate the basic features of SOM as applied to a minimization problem. Here, we show, in the form of a proof of principle, that SOM can enumerate the compact structures and degeneracies of the HP model, up to 36mers, in an efficient way. Its application to three-dimensions and to real proteins is straightforward, and work in this direction is in progress.

## 2. THE HP MODEL

A real protein molecule is a sequence of amino acids chosen from an alphabet of twenty different amino acids. Each amino acid (*bead*) consists of the repeating unit of the backbone and a unique side chain determining the properties of the amino acid and bonded to the backbone carbon atom (the alpha carbon). The simplest model that is used to represent the real protein chain is the HP model, which consists of beads along the chain such that each bead is either hydrophobic (H) or polar (P). Thus, the HP model partitions the twenty amino acids into H and P classes, according to statistically determined preferences of amino acids with respect to hydrophobicity—whether the amino acid likes to be near or away from the water in the three-dimensional protein structure. The HP model is a simple exact model, introduced by Dill *et al.* (1995), to study features important to the folding of proteins.

In a real protein, every neighboring bead along the chain is joined by a covalent bond. Once covalent bonds are formed, the primary structure of the chain is fixed. The tertiary structure of the chain, on the other hand, keeps changing during folding, until the biologically favorable three-dimensional configuration

is obtained. As the tertiary structure keeps changing, two beads separated by several beads along the chain may come close to each other and form an intramolecular bond, also referred to as a *nonbonded contact.* Unlike covalent bonds, which do not break once they are formed, the intramolecular bonds may form, break, and reform during folding. An $i$th and a $j$th bead can make an intramolecular bond only if $|i-j| \geq 3$.

In the HP model, there are three different types of intramolecular bonds: H–H, H–P and P–P. Hydrophobic (H) amino acids try to escape the water and bury themselves in the interior of the three dimensional protein molecule. This is equivalent to maximizing the number of H–H bonds in the HP-model. Polar (P) amino acids, on the other hand, tend to remain at the surface of the molecule where their tendency to contact water molecules is maximized. The hydrophobic properties of amino acids are commonly enforced in the HP-model by considering the H–H bond as favorable and the other two as neutral. Here, we follow the same practice and take the energy of the H–H bond as $-\epsilon_{HH}$, a negative quantity, and the energies of the H–P and P–P bonds as zero.

### *Lattice embedding*

Further simplification is possible by embedding the HP model on a lattice, where each chain configuration is a self-avoiding walk on the lattice. Two- or three-dimensional square lattices, as well as triangular and hexagonal lattices, have been used in different models. For the sake of simplicity, we consider HP chains which have their minimum energy states on a 2D plane, though the approach may readily be extended to the three-dimensional case, using a 3D topology for the Self-Organizing Network. Embedding of the protein in a lattice allows the definition of the intramolecular bond length as the length of the lattice edge. In the HP lattice models, the minimum energy configuration of the protein is uniquely defined as the lattice structure with maximum number of H–H bonds.

## 3. OUR APPROACH

We try to predict the structure of 2D proteins that are designed HP sequences such that the minimum energy configuration lies on a 2D square lattice. In this simplified model, the problem of finding the structure is equivalent to finding which bead occupies which lattice vertex. In the minimum free energy conformation, each city can be occupied by only one bead (satisfying the excluded volume effect) and the HH contacts are maximized.

Simply speaking, we formulated the problem to be similar to the solution of the Traveling Salesman Problem (TSP) by SOM, where neurons in a one-dimensional ring move towards the cities such that when the network converges, the mapping between the *ordered* neurons on the ring and the cities indicate the order of the cities to be visited (Angeniol *et al.*, 1988; Jeffries *et al.*, 1994; Budinich *et al.*, 1996; Altinel *et al.*, 2000).

### *TSP by SOM*

Given the locations of N cities, the TSP is to find a closed path that goes through all of these cities such that the travel distance is minimum. In the SOM approach to TSP, the cities are presented one by one, randomly, to the network and the closest neuron (bead) and its topological neighbors update their weights (or coordinates) to move closer to the input city, in proportion with their distance to the city and with their distance to the closest neuron. These topology-preserving updates to the SOM help find good (within a few percent of the optimal) approximate solutions to the TSP and are computationally efficient. More specifically, at each iteration of the algorithm, a randomly selected input city, k, is presented to the network. Then, the neuron, $i$, minimizing

$$\|r_i - r_k^o\| \tag{1}$$

is selected as the *winning* neuron. Then, each bead $j$ will move towards the input city, k, in proportion to its distance to the input city and the winning neuron, according to the following update rule, following Angeniol's notation (Angeniol *et al.*, 1988).

$$\Delta r_j = e^{-n_j^2/G^2}(r_j - r_k^o) \tag{2}$$

Here, $G$, which is the only adjustable parameter, is a gain variable and $n_j$ is the distance of node $j$ to node $i$ along the line of beads. $G$ works as a temperature variable: it is decreased from a high initial value, attracting the winning bead and its neighbors with almost equal strength, to 0, at which point only the winning bead moves towards the input city. The distance $n_j$ of the bead, $j$, to the winning bead, $i$, defined as the number of amino acids in between the two beads, also affects the strength with which the bead, $j$, is pulled: as $n_j$ gets larger, bead $j$ is pulled with less strength.

### *Protein structure prediction by SOM*

In our problem, the lattice vertices form the cities, and the beads form the neurons. We formulated the problem so that the minimum energy conformation corresponds to the *Hamiltonian path* of the vertices of the lattice, maximizing the HH contacts. When the algorithm converges, the location of the visited cities indicates the shape of the protein.

The Hamiltonian path of n nodes (*cities*) is a path passing through the given n cities, such that the path length is minimal. The minimum path length criterion implies that each city is visited only once. The Hamiltonian path problem differs from the TSP in that in the former there is a starting and an ending point in the path and these two may be far apart from each other, whereas the TSP considers cyclic paths. Inasmuch as the protein molecule has two unique termini, called the N-terminal and the C-terminal, and since these two termini are not necessarily close in space in the minimum energy conformation, the Hamiltonian path formulation is more suitable than the TSP formulation.

Specifically, we choose a square lattice with M vertices in one dimension and N vertices in the other. We try to embed a linear protein chain of MxN beads of known primary structure on this lattice, such that (i) each bead coincides with one and only one city, (ii) each covalent bond coincides with a lattice edge, and (iii) the number of nonbonded H–H contacts in the final folded structure is maximum. The lattice embedding provides the excluded volume effect and the Hamiltonian path criterion satisfies the near equality of covalent bond lengths (3.9 Å on the average in real life, represented by one edge of the lattice). The total of MxN vertices constitutes the cities in the Traveling Salesman Problem, and the beads represent the neurons in the Self Organizing Map.

We tried two approaches: (i) using the HH attraction as a penalty term, in addition to the update mentioned above, and (ii) using three separate iterations, where the standard SOM algorithm is followed by two iterations designed to exclusively introduce the protein interactions. The second method achieved better results that are described in this paper.

Let $r_i = (x_i, y_i)$ and $r_j^o = (x_j^0, y_j^0)$ represent the Cartesian coordinates of the $i$th bead and the $j$th city. With the lattice edge taken as unity, the initial configuration of the chain (neurons) is chosen randomly according to

$$x_i = M(0.5 - u_i), 1 \le i \le M$$
$$y_i = N(0.5 - u_i), 1 \le i \le N \tag{3}$$

where $u_i$ and $v_i$ are uniformly distributed random numbers in the interval (0,1). The cities $(x_j^0, y_j^0)$ are the coordinates of the uniformly distributed vertices of the lattice.

In the first approach, (i), the simple SOM solution to the TSP is adopted with a penalty term derived from the HH attractions. Given a randomly selected input city, k, the winning neuron, $i$, and the displacements are chosen according to the modified SOM:

$$\alpha(r_i - r_k^o) + \zeta \sum_j \Gamma_{ij}(r_i - r_j) \tag{4}$$

Here, $\Gamma_{ij}$ is equal to $\epsilon_{HH}$ if $i$ and $j$ are both hydrophobic and zero if one of them is polar or if $|i - j| < 3$. The second term is intended to incorporate the force received by the $i$th bead in response to all the HH-attraction it receives. In other words, the winning neuron, $i$, is the one closest to the input city, subject also to the total HH-interaction it receives. However, convergence with this approach was more difficult than the approach described below.

In the second approach, (ii), which is found to be more successful, the SOM algorithm (without the penalty term) is followed by two other phases, inside a main loop of N iterations. The idea is to independently execute three steps to accomplish the various goals (excluded volume and equal covalent bond lengths, maximized HH-contacts). The first internal loop moves the beads and their covalently bonded neighbors towards the selected city, using the simple SOM algorithm, as follows:

1. Randomly choose a city, $k$.
2. Pick the bead, $i$, which is closest to the chosen city.
3. Move the bead, $i$, towards the city, $k$, according to

$$\Delta r_i = \alpha(r_i - r_k^o) \tag{5}$$

and move its neighbors $j$, such that $|i - j| = 1$, according to

$$\Delta r_j = \alpha e^{-\beta}(r_i - r_k^o) \tag{6}$$

where $\alpha$ modulates the size of the increment and $\beta$ is a constant that decreases linearly with the number of steps N of the algorithm.

In the second internal loop, the lengths of the covalent bonds are pushed towards unity according to:

$$\Delta r_i = \gamma e^{-\beta}[(l_{i,i+1} - 1)(r_{i+1} - r_i) - (l_{i,i-1} - 1)(r_i - r_{i-1})] \tag{7}$$

where, $l_{i,i+1}$ is the length of the bond between residues $i$ and $i + 1$ and $\gamma$ modulates the increments of $r$ under the bond potential. If bead $i$ is a terminal bead, then Equation 5 is applied only partly, as is obvious from the definition. In the third internal loop, the H–H bonds are turned on. If a pair of beads $i$ and $j$ is an HH pair, then the $i$th bead is moved according to:

$$\Delta r_i = \zeta e^{-\beta} \sum_j (r_j - r_i) \tag{8}$$

where, the summation is over $j$ and $\zeta$ modulates the increments of $r_i$ under the attractive H–H potential coming from all other H's. At the end of the three internal loops, the main loop is iterated until convergence to the given lattice. Our use of the SOM algorithm is similar to the KNIES algorithm of Altinel *et al.* (1999) in that it also preserves the global statistics of the problem at each iteration: we found that pulling the center of the locations of the beads to the center of the locations of the cities helps with the convergence.

## 4. RESULTS

Here we give results for three sequences, a 20mer embedded to a $4 \times 5$ lattice and two 36mers, 36-a and 36-b, each embedded to a 6x6 lattice, using the second approach. The corresponding native structures on the 2D square lattice are shown in Figures 1 and 2, where the black beads denote the H's and the white beads denote the P's. For the three cases, the same parameters are used: $\alpha = 0.02$, $\gamma = 0.01$, $\zeta = 0.002$,

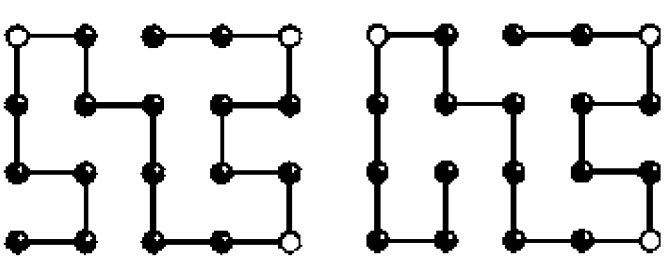

**FIG. 1.** Two minimum energy configurations for the 20mer, HHHHHPHHHHHHPHHHHPHH, on the 2D square lattice. The black beads denote the H's and the white beads denote the P's.

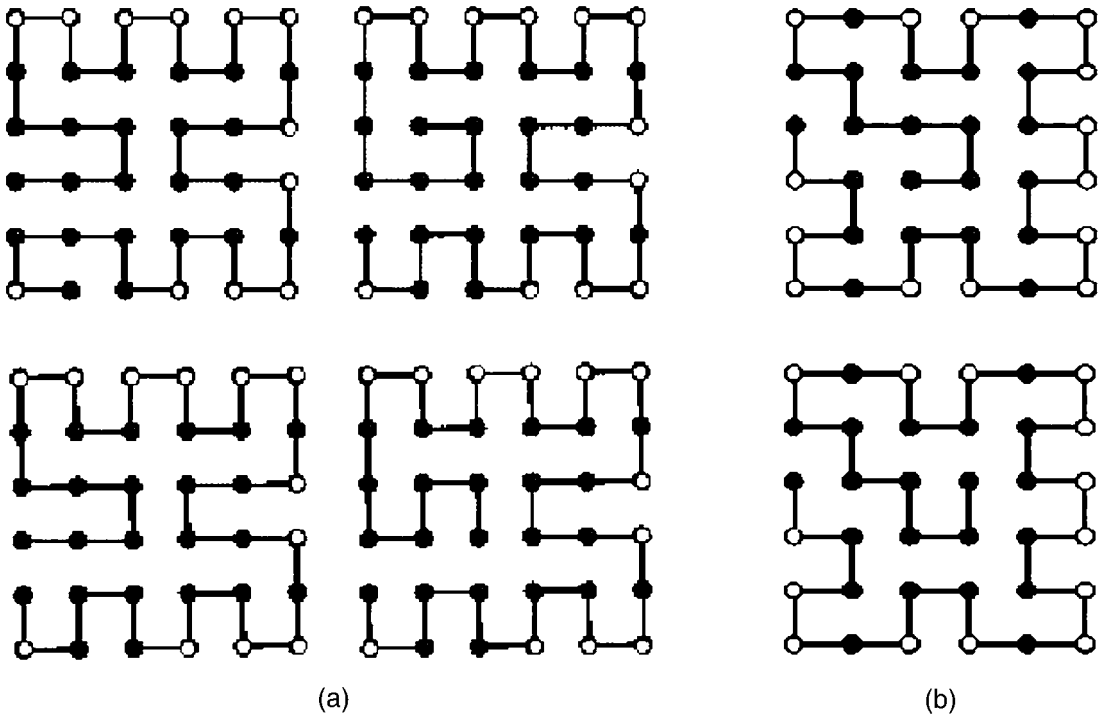


**FIG. 2.** a) Four minimum energy configurations with 21 HH-contacts, for the sequence 36-a, HPHHHHPHHP-PHPHHHHPHPPHHPPHHPPHHHHHHH, on the 2D square lattice. b) Four minimum energy configurations with 20 HH-contacts, for the sequence 36-b, HHHHHHHPHPHHPHPPHHPPHHPPHPHHPHPPHHPH, on the 2D square lattice. The black beads denote the H's and the white beads denote the P's.

the starting value of $\beta = 5$, $N = 100,000$. Here, $\beta$ is decremented linearly, such that it equates to zero at the end of N iterations. At the end of N steps, the number of H–H contacts is recorded and a new run is started. A total of 1,200 runs with different initial random configurations are performed for each of the three sequences. We report the distribution of the predicted configurations at different energy levels, after eliminating those runs that result in either bond crossings or diagonal crossings (diagonal edges of the lattice).

(i) The 20mer: The sequence of the 20mer is chosen as HHHHHPHHHHHHPHHHHPHH. Figure 1 shows the two different minimum energy configurations obtained for this 20mer on the $4 \times 5$ lattice, with the maximum number of contacts of 12. Out of the 1,200 runs that started with different random configurations, the algorithm found four configurations with 12 contacts, two for each of the configurations shown in Fig. 1. The distribution of the predicted configurations at different energy levels (number of contacts) were as follows: 12:4, 11:36, 10:295, 9:428, 8:187, and 7:60, where the number before the colon is the number of contacts and the one after is the frequency obtained in 1,200 runs.

(ii) The sequence 36-a: This sequence is chosen as HPHHHHPHHPPHPHHHHPHPPHHPPHHPPHHH-HHHH. In Figure 2a, the four different minimum energy configurations for this 36mer on the $6 \times 6$ lattice are shown. The maximum number of contacts is 21. In 1,200 runs, the algorithm found four configurations with 21 contacts shown in Figure 2a. Using the above notation, the degeneracies are obtained with the following frequencies: 21:4, 19:1, 18:1, 17:2, 16:8, 15:13, 14:15, 13:33, 12:26, 11:26, 10:33, 9:8.

(iii) The sequence 36-b: This sequence is chosen as HHHHHHHPHPHHPHPPHHPPHPPHPHHPHP-PHHPH. In Figure 2b, the two different minimum energy configurations for the 36-b are shown. The maximum number of contacts is 20. In 1,200 runs, the algorithm found two configurations with 20 contacts shown in Figure 2b. The degeneracies are obtained as 20:2, 17:2, 15:5, 14:3, 13:18, 12:39, 11:64, 10:48, 9:47, 8:20, 7:6.

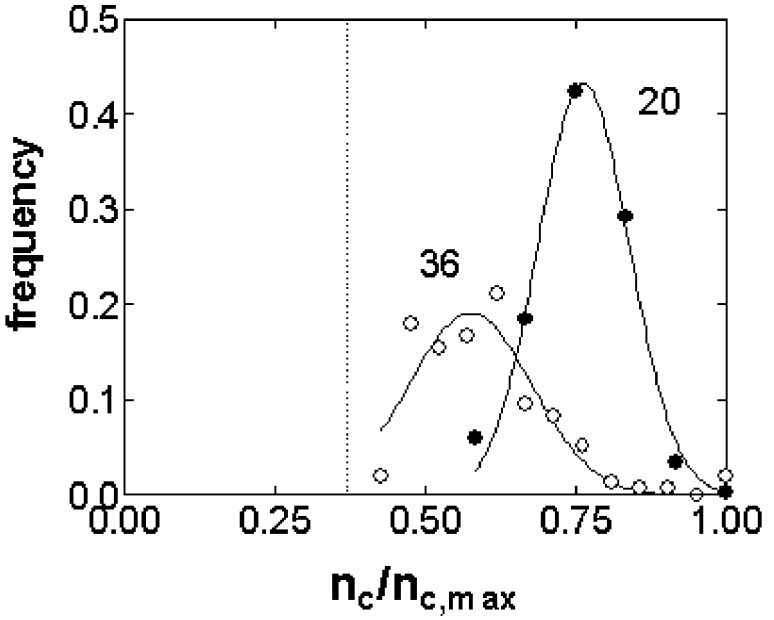


**FIG. 3.** Distributions of the configurations with different number of HH-contacts, found using the SOM approach. The x-axis is in relation to the (maximum) number of HH-contacts in the native state.

In Fig. 3, the frequency distribution of contacts is presented as a function of the ratio of number of contacts $n_c$ to the maximum number of contacts, $n_{c,max}$. The curves are Gaussian fits to the data points for the 20mer and for 36-a. The results for the 36-b were similar to those for the 36-a and are not shown. The vertical dotted line represents the lower bound obtained by Hart and Istrail (1996) for 2D HP chains, discussed in more detail below.

## 5. CONCLUSION AND DISCUSSION

In the present paper, we showed that SOM may be adopted as an optimization technique for determining the minimum energy configurations of model proteins. The method is efficient, as its running time scales linearly with the chain length. It also found the global minimum for all the test cases we have studied within the context of the present work, as well as the degeneracies of the minimum energy configurations. We are aware of a single previous study by Hart and Istrail (1996), where near-optimality has been quantified for HP-models. Though their algorithm *guarantees* a solution within 3/8 of the optimal, our approach using the SOM found the global minimum in all the test cases up to 36mers. Figure 3 shows that all of the results of calculations with the SOM algorithm fall above the 3/8 line.

The present approach is based on generating several configurations which are expected to yield low energy states. The temperature is high at the start of the simulations and is decreased gradually. At some value of the temperature, the chain finds a local energy minimum to which it settles. The distribution of energies obtained in this manner is shown in Fig. 3. Unlike the canonical distribution, this distribution is not a Boltzmann distribution, which is apparently due to the fact that the mechanisms of energy exchange between the chain and the lattice are not optimized. The only energy flow from the chain to the lattice is through a spring-like coupling of a given bead to its closest city. This energy exchange between the chain and the lattice is propagated to the chain only through the first covalently bonded neighbors of the chosen bead. The H–H pairs are also coupled to each other by linear springs. The changes in the HH energy are not coupled to the lattice, however. At each step, the H–H pairs are attracted towards each other by a small amount. Probably due to these assumptions, the relaxation of the chain energy is not as efficient as a system that obeys Boltzmann statistics.

We have been able to apply the SOM algorithm to the case of compact proteins, where the chain completely fills the lattice. We were not successful in obtaining the global energy minimum structures

where the number of lattice points is larger than the number of beads and parts of the lattice nodes need to be unvisited. In this case, since the beads (nodes of the SOM) try to cover the input space (lattice), a bead is often located in between two lattice points. One solution to this problem may be to fill the $N$ lattice points by $N_1$ beads and $N - N_1$ solvent molecules. Another solution might be to use modified SOM algorithms where the number of nodes (beads) are adjusted as needed (Angeniol *et al.*, 1988), possibly using dummy beads to fill the parts of the lattice that need to be empty. Our work in devising such a modified SOM algorithm is in progress. The same approach would also address the simpler issue of unknown lattice dimensions for compact proteins.

## REFERENCES

Altinel, I.K., Aras, N., and Oommen, B.J. 2000. Fast, efficient and accurate solutions to the Hamiltonian path problem using neural approaches. *Comput. Oper. Res.* 27, 461.

Andrade, M.A., Casari, G., Sander, C., and Valencia, A. 1997. Classification of protein families and detection of the determinant residues with an improved self-organizing map. *Biol. Cybern.* 76, 441.

Angeniol, B., Vaubois, G.C., and Texier, J.-V. 1988. Self-organizing feature maps and the travelling salesman problem. *Neural Networks* 1, 289.

Aras, N., Altinel, I.K., and Oommen, B.J. 1999, The Kohonen network incorporating explicit statistics and its application to the travelling salesman problem. *Neural Networks* 12, 1273.

Ball, K.D., Erman, B., and Dill, K.A. 2001. The elastic net algorithm and protein structure prediction. *J. Comput. Chem.* 23, 77.

Bohr, H., and Brunak, S., 1989. A travelling salesman approach to protein conformation. *Complex Systems* 3, 9.

Budinich, M. 1996. A Self-organizing neural network for the traveling salesman problem that is competitive with simulated annealing. *Neural Comput.* 8, 416.

Crescenzi, P., Goldman, D., Papadimitriou, C.H., Piccolboni, A., and Yannakakis, M. 1998. On the complexity of protein folding. *J. Comp. Biol.* 5, 3, 423.

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., and Chan, H.S. 1995. Principles of protein folding—a perspective from simple exact models. *Protein Sci.* 4, 561.

Duan, Y., and Kollman, P.A. 1998. Pathways to a protein folding intermediate observed in a 1-millisecond simulation in aqueous solution. *Science* 282, 740.

Hanke, J., Beckmann, G., Bork, P., and Reich, J.G. 1996. Self-organizing hierarchic networks for pattern recognition in protein sequence. *Protein Sci.* 5, 72.

Hart, W.E., and Istrail, S. 1996. Fast protein folding in the hydrophobic-hydrophilic model within three-eights of optimal. *J. Comp. Biol.* 3, 53.

Jeffries, C., and Niznik, T. 1994. Easing the conscience of the Guilty. Net. *Comput. Oper. Res.* 21, 9, 961.

Li, Z., and Scheraga, H. 1987. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci*, 84, 6611–6615.

Schuchhardt, J., Schneider, G., Reichelt, J., Schomburg, D., and Wrede, P. 1996. Local structural motifs of protein backbones are classied by self-organizing neural networks. *Protein Eng.* 9, 833.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E., and Golub, T. 1999. *Proc. Natl. Acad. Sci. USA* 96, 2907.

Wriggers, W., Milligan, R.A., Schulten, K., and McCammon, J.A. 1998. Self-organizing neural networks bridge the biomolecular resolution gap. *J. Mol. Biol.* 284, 1247.

Address correspondence to:
*Berrin Yanikoglu*
*Laboratory of Computational Biology*
*Faculty of Engineering and Natural Sciences*
*Sabanci University*
*Orhanli, 81474 Istanbul, Turkey*

*E-mail:* berrin@sabanciuniv.edu