

## Global Results in Space of Inter-Monomer Interactions for HP Lattice Model\*

**M. R. EJTEHADI**

*Institute for studies in Theoretical Physics and Mathematics,  
Tehran P.O. Box 19395-5531-IRAN*

**V. SHAHREZAEI and N. HAMEDANI**

*Institute for studies in Theoretical Physics and Mathematics,  
Tehran P.O. Box 19395-5531, -IRAN*

*Department of Physics, Sharif University of Technology,  
Tehran P.O. Box: 11365-9161-IRAN*

Received 23.04.2000

### Abstract

Using the ground state candidates for Hydrophobic-Polar lattice model on a two dimensional square lattice, we exactly enumerate the native states of proteins with length up to 20 for a wide range of energy parameters. We find the reduced set of contact maps, ground state candidates, perfectly stable sequences and designability of configurations.

### 1. Introduction

The proteins are bio-macromolecules, which are made from thousands of atoms. These atoms are in interaction with each other and water molecules, which surround them. A feasible approach to the problem of protein folding is based on a coarse-grained view. From this viewpoint the proteins are made from 20 types of monomers (amino acids). The most important point in this approach is the determination of the effective interactions between the amino acids [1].

The structural information for protein structures can be coded in a contact map [2]. A contact map is a binary  $L \times L$  matrix  $C$ . The element  $c_{ij}$  of this matrix is nonzero if  $i$ th and  $j$ th monomers are in contact. Because of the short-range nature of inter-monomer

---

\*Talk presented in Regional Conference on Mathematical Physics IX held at Feza Gürsey Institute, Istanbul, August 1999.

interactions, one can determine the configuration energy in terms of contacts. There are many papers, which study the thermodynamical and structural properties of proteins, by using contact maps [3].

It is well known that the biological functionality of proteins depends on the shape of their native states. The native structure is the unique minimum free energy structure for the protein sequence [4]. Thus searching the configuration space to find native states has been the subject of many papers. In most of previous works, the problem was studied for given values of inter-monomer energy parameters. As our knowledge about the effective interactions is not certain, and the native structures of proteins may be sensitive to these parameters [5], looking at the native states for different energy parameters is relevant [6, 7]. Recently, we have shown that the number of ground state candidates for any sequence is unexpectedly small [8]. This suggests that the problem can be studied for a wide range of interaction parameters by exact enumeration. We study this problem on a two dimensional square lattice. In this approach a protein structure is modeled by a self avoiding walk on the lattice, and, any pair of monomers which are nearest neighbors and are not adjacent according to sequence (non-sequential neighbor) are in contact.

The number of possible configurations for an  $L$ -mer is equal to the number of self-avoiding walks ( $N_{\text{SAW}}$ ) with  $L - 1$  steps. Since many of these walks give the same contact matrix, the number of possible contact matrices (physical maps)  $N_c$ , is much smaller, although it is still very large [9]. If one is interested only in the native structure of proteins, the set of the contact maps can be reduced further, by removing all maps, which have no chance to be a native state. We call the remaining maps, the *reduced set of contact maps* [10]. This reduced set of contact maps can be used in enumeration studies to find the possible ground states and the native states of proteins. In this paper we want to focus on some important and interested properties of reduced set of contact maps and the ground state candidates.

## 2. The reduced contact maps

The effective potential energies between the 20 types of amino acids can be described by a  $20 \times 20$  interaction matrix [1]. The energy of a given sequence  $\sigma$  in any structure can be determined from

$$E = \sum_{i,j} c_{ij} m_{\sigma_i \sigma_j}. \quad (1)$$

The  $c_{ij}$  and  $m_{ij}$  are respectively the elements of the contact matrix ( $C$ ) and the interaction matrix ( $M$ ). This shows that, all configurations, which have the same contact map, have equal energies. If we look at the energy spectrum of one sequence, the states corresponding to such maps are degenerate. We call such degeneracies, type-one degeneracies to distinguish them from other kinds of degeneracies [10]. If the energy of a sequence is minimum in such states, this sequence does not have a unique native state. Such sequences are not protein-like. The states corresponding to such *degenerate contact maps* can never be a native state, however, we cannot exclude them from our search, because

they compete with other maps. On the other hand, there are some maps, which cannot be the ground state, and do not have a role in the competition for the ground state. To see that, consider two contact matrices  $C_1$  and  $C_2$  and their subtraction ( $C' = C_1 - C_2$ ). We call  $C_2$ , a *component* of  $C_1$  if all elements of  $C'$  are non-negative ( $c'_{ij} = 0$  or  $1$ ). Note that  $C'$  has at least one non-zero element. Using equation 1, the energy of an arbitrary sequence  $\sigma$  in the configuration(s) corresponding to the map  $C_1$  can be written as:

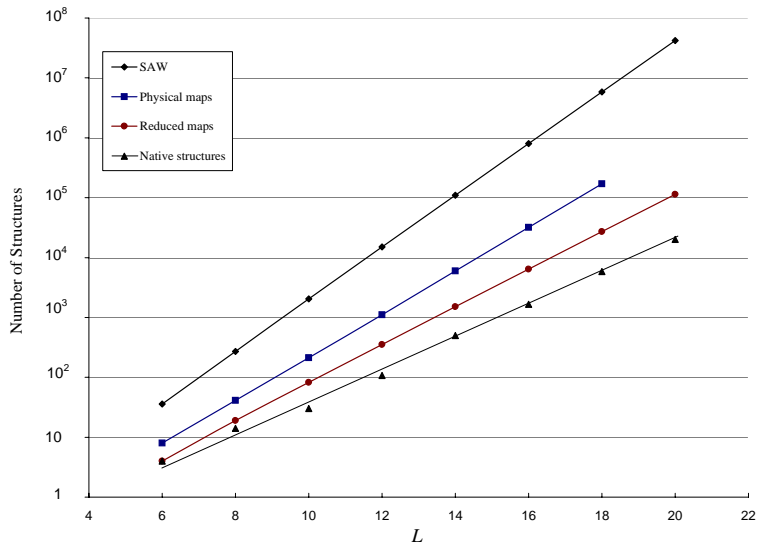
$$\begin{aligned} E_1 &= \sum_{i,j} c_{1,ij} m_{\sigma_i \sigma_j} \\ &= \sum_{i,j} c_{2,ij} m_{\sigma_i \sigma_j} + \sum_{i,j} c'_{ij} m_{\sigma_i \sigma_j} \\ &= E_2 + \sum_{i,j} c'_{ij} m_{\sigma_i \sigma_j}. \end{aligned} \quad (2)$$

According to experimental data all elements of interaction matrix  $M$  are negative [1]. Thus the second term in the rhs gives a negative contribution to energy, and,  $E_1 < E_2$ , for any sequence. Then map  $C_2$  can never be a ground state. One can find all component maps such as  $C_2$ , and remove them from the set of contact maps. Indeed such component maps are related to configurations, which can fold to more compact shapes without losing any of their old contacts. By this procedure the reduced set of contact maps is found [10]. In figure 1 the number of reduced maps ( $N_r$ ) are compared with the number of self-avoiding walks ( $N_{\text{saw}}$ ) and the number of physical maps ( $N_c$ ), on a two dimensional square lattice for sequences with length up to 20. Although, all of these quantities have similar behaviors, the growth rate of  $N_r$  is very slower than the others.

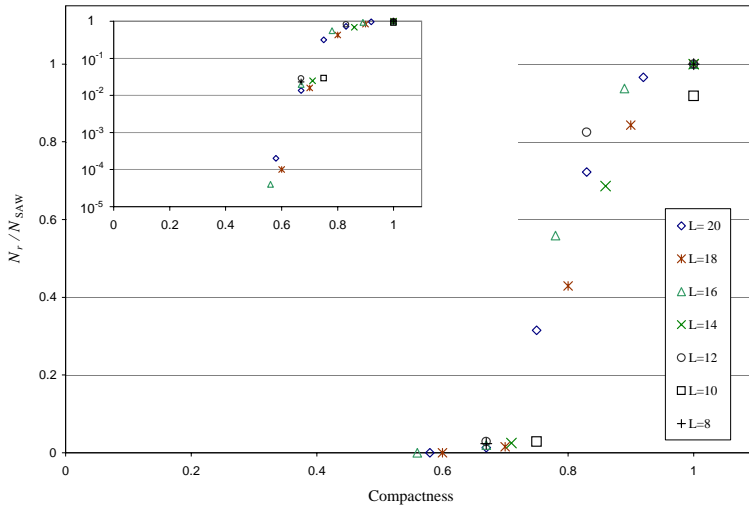
Let's consider the ratio of the number of contacts,  $b = \frac{1}{2} \sum_{i,j} c_{i,j}$ , to the maximum number of possible contacts for sequences of the same length,  $b_{\text{Max}}$ , as a measure for the compactness of configurations  $\Gamma$ ,

$$\Gamma = \frac{b}{b_{\text{Max}}}. \quad (3)$$

If one scales the number of reduced maps ( $N_r$ ) by the number of total structures ( $N_{\text{SAW}}$ ) at each compactness, a scale-independent behavior can be seen (figure 2). It also seems that there is a critical compactness, below which the compactness of the members of reduced set never drops. This shows why the results of studies on compact structure spaces are reasonable. We do not have an exact analytical proof, but it seems from these data that a transition occurs in the number of reduced maps, near the compactness of 0.8 and it vanishes for a compactness below 0.5.



**Figure 1.** The number of self avoiding walk structures, physical contact maps, reduced set of contact maps and native structures, *vs.* length of sequences.



**Figure 2.** The number of reduced maps that scaled by the number of all structures at each compactness, for sequences with length 8 to 20. There is a transition near to 0.8 and a cut off near to 0.5. The later can be seen better in logarithmic scale (inner graph).

### 3. The ground state candidates for HP model

The native states of proteins are to be found among the structures corresponding to the reduced set of contact maps. The sequence of the amino acids along the protein chain and their interactions have an essential role in the selection of a particular structure as the native state. In the coarse-grained viewpoint, the interaction between the amino acids is characterized by the effective energies. These effective interactions depend on the properties of the solutions. A relevant question is how sensitive the native structures are to changes in these interactions. We address this question by enumerating the possible ground states of protein sequences for a wide range of effective inter-monomer interaction energies.

Without any loss of generality, we use a hydrophobic-polar (HP) two-dimensional lattice model [11] in this paper. The general form of the interactions between  $H$  and  $P$  monomers in an HP model can be written as follows [8, 10]:

$$\begin{aligned} E_{HH} &= -2 - \gamma - E_c, \\ E_{HP} &= -1 - E_c, \\ E_{PP} &= -E_c, \end{aligned} \tag{4}$$

where  $E_{\sigma\sigma'}$  is the contact energy between monomers of types  $\sigma$  and  $\sigma'$ . These potential energies are only between non-sequential nearest neighbors. Here  $\gamma$  and  $E_c$  are the mixing and compactness potentials respectively, two parameters which are determined from experimental data. There are many publications based on this model, and in most of them the values of  $\gamma$  and  $E_c$  are fixed [11, 12]. Here, we consider them as two free parameters and discuss our results in terms of them.

It has been argued that the following relations should hold between inter-monomer energies:

$$\begin{aligned} E_{HH} &< E_{HP} < E_{PP}, \\ E_{HH} + E_{PP} &< 2E_{HP}. \end{aligned} \tag{5}$$

These arguments are based on the compactness of the native states [13] and some calculations on  $20 \times 20$  inter-monomer interaction matrix  $M$  [14]. These restrict  $\gamma$  and  $E_c$  to positive values ( $\gamma, E_c > 0$ ).

At first sight, it might seem possible to arrive at any native state for a given sequence by changing  $\gamma$  and  $E_c$ . But when we consider the geometrical properties of the ground state, we will find that these parameters are not powerful enough to select any configuration as the native state. In other words, the native states are stable against the change of interaction parameters.

If we consider  $H = -1$  for hydrophobic monomers and  $P = 0$  for polar monomers, a given sequence can then be represented by a binary vector  $(\sigma)$  [6]. The energy of this sequence in a configuration characterized by a contact matrix  $C$ , can be written as:

$$E = -m - a\gamma - bE_c, \tag{6}$$

where  $m$ ,  $a$  and  $b$  are three integers, related to  $\sigma$  and  $C$  as follows:

$$\begin{aligned} m &= -\sigma^t \cdot C \cdot \mathbf{1}, \\ a &= \frac{1}{2} \sigma^t \cdot C \cdot \sigma, \\ b &= \frac{1}{2} \mathbf{1}^t \cdot C \cdot \mathbf{1}. \end{aligned} \quad (7)$$

It can be seen that  $m$  is equal to the number of all non-sequential neighbors of  $H$  monomers in the configuration,  $a$  is the number of  $H-H$  contacts and  $b$  is the number of all contacts. It can be shown that the following inequalities hold between these parameters [16].

$$m - b \leq a \leq \frac{m}{2} \leq b. \quad (8)$$

Equation 6 suggests that the energy levels of a given sequence can be described by three integer numbers  $(m, a, b)$ . It is highly probable that these states are degenerate. There are three types of degeneracy:

- Type 1:  $C = C'$

In which case two or more configurations with different shapes have the same contact matrix. These configurations will remain degenerate for any sequence, and any choice of  $\gamma$  and  $E_c$ . These are the configurations corresponding to the degenerate maps already mentioned in section 2. This type of degeneracy, is more probable for configurations with low compactness (see figure 2). Note that we are not talking about the configurations which are related to each other by spatial symmetries, i.e. rotation, reflection, etc., for our purpose such configurations are identical.

- Type 2:  $(m, a, b) = (m', a', b')$  but  $C \neq C'$

In this case one particular sequence has the same  $m$ ,  $a$  and  $b$  values in two or more configurations. This degeneracy persists for any value of  $\gamma$  and  $E_c$ , but may disappear for another sequence. Although, this degeneracy depends on sequence coding, the  $b = b'$  condition is purely geometrical, and is a necessary condition for this degeneracy.

- Type 3:  $E = E'$ , but  $(m, a, b) \neq (m', a', b')$

One sequence has the same energy in two different states  $(m, a, b)$  and  $(m', a', b')$ , provided that  $\gamma$  and  $E_c$  obey the following relation:

$$(m - m') + (a - a')\gamma + (b - b')E_c = 0. \quad (9)$$

This degeneracy is related to both sequence coding  $\sigma$  and inter-monomer interactions.

The first type of these degeneracies is completely geometric. The second one depends on both geometry and the amino acids' coding sequence. These two types do not depend

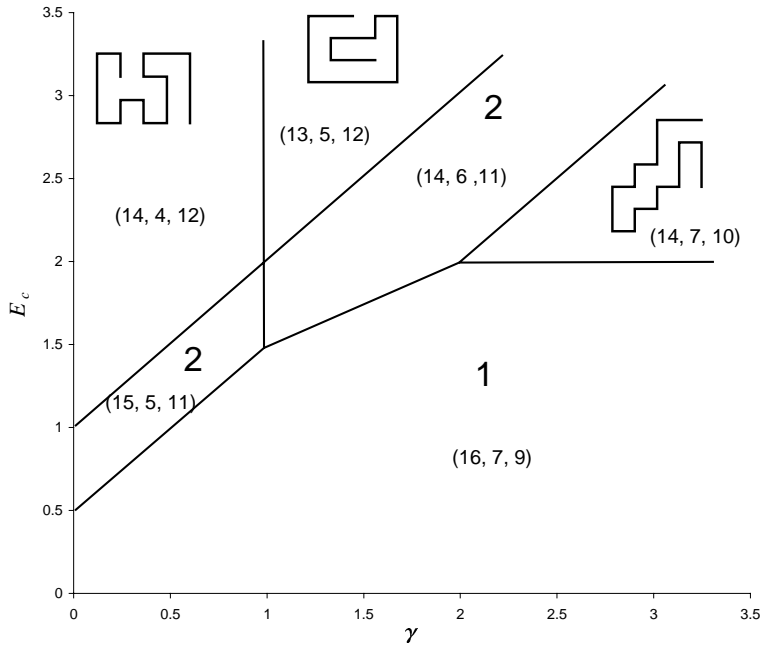
on the values of the interaction energies. Thus, in the energy spectrum of any sequence there are some states, which are degenerate independently from the potential. If the ground state of a particular sequence is one of these degenerate states, that sequence does not have a unique native structure.

The third type is not actually a degeneracy at all. Equation 9 corresponds to a line in the parameter space of  $E_c$  and  $\gamma$ . This line is a level crossing line. Degeneracy actually occurs only on the line, and a highly accurate fine-tuning is needed to reach a point on this line. For the two sets of interaction energy parameters on the two sides of this line, the energy ordering of the states is different. For any pair of states such an ordering line exists. By drawing all ordering lines in the space of  $E_c$  and  $\gamma$ , this space is divided into many ordering zones. We are only interested in the ground state, which means that many of these ordering lines are not relevant. Some of them only govern the ordering of the excited states. By removing the irrelevant lines, one gets a diagram which shows the ground state cells (Fig. 3). As mentioned before changing the inter-monomer interaction parameters inside any of these cells does not change the ground state. By looking at the whole energy space, one can find all possible ground states and their corresponding cells. Any such cell in the space of energy parameters is associated with one ground state candidate. The number of cells is equal to the number of ground state candidates ( $G_c(\sigma)$ ). By drawing such diagrams, one can easily find the ground state for any choice of  $E_c$  and  $\gamma$ . Fig. 3 shows this diagram for a 20-mer. In this example there are only six possible ground states. The cells marked with the numbers “1” and “2” correspond to type-1 and type-2 degenerate states respectively, therefore there is no unique native structure for these cells. The sequence in this example has 3 non-degenerate states. These structures are shown in the figure. It is possible that all the ground state candidates of a given sequence are degenerate. These sequences constitute universally bad sequences i.e. for any set of interaction parameter values they do not have a native structure. Any sequence which is not a bad sequence, we call a *good sequence*. Nearly 54% of the sequences of length 20 are good sequences, i.e. for some specific set of energy parameters they have a native state.

The interesting point in figure 3 is that the number of ground state candidates is very small. The largest value of  $G_c$ , for sequences with length 6, 8, 10, 12, 14, 16, 18, 20 are 1, 1, 1, 3, 4, 5, 6, 7 respectively. Fig. 4 shows the histogram of  $G_c(\sigma)$  for all sequences with  $L = 20$ . The light gray area in this figure shows the result for all  $2^{20}$  sequences, and the dark area shows the results for good ones. From this diagram it can be seen that the mean value of  $G_c(\sigma)$  is very small. The average of  $G_c(\sigma)$  for various lengths is shown in figure 5. However, the data in hand is not enough to draw a reliable conclusion about the number of ground state candidates for sequences of large length, but the average number does not seem to grow very rapidly, and the growth rate appears to be linear. Comparison of the average value of  $G_c(\sigma)$  for these sequences with the number of all configurations (i.e. for sequences with length 20 the number of sequences is on the order of  $10^8$ ), shows that the geometric constraints play an important role in selecting a state as the ground state. The reason that there are few ground state candidates for any sequence can be given by a geometrical argument. Consider a three dimensional space with axes  $X$ ,  $Y$

and  $Z$ . Any state is represented by a point with coordinates  $a$ ,  $b$  and  $m$  in this space (Fig. 6). All the states will be inside a pyramid according to equation 8. For any value of  $\gamma$  and  $E_c$ , let's consider the following plane perpendicular to the vector  $(\gamma, E_c, 1)$ :

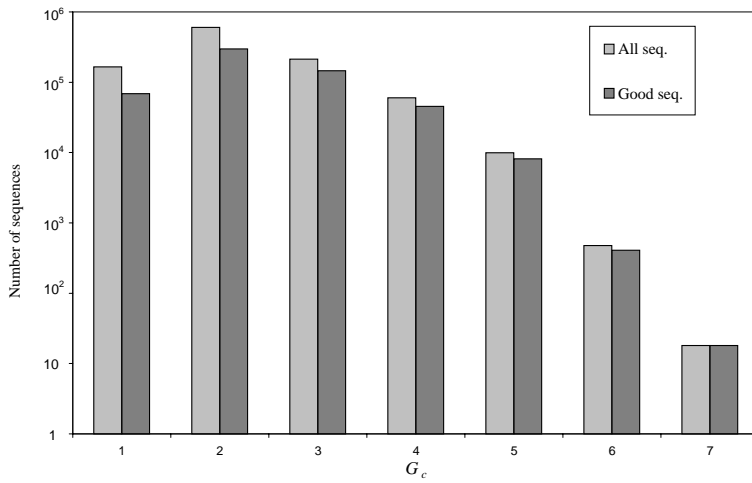
$$z = -\gamma x - E_c y + (z_0 + \gamma x_0 + E_c y_0). \tag{10}$$



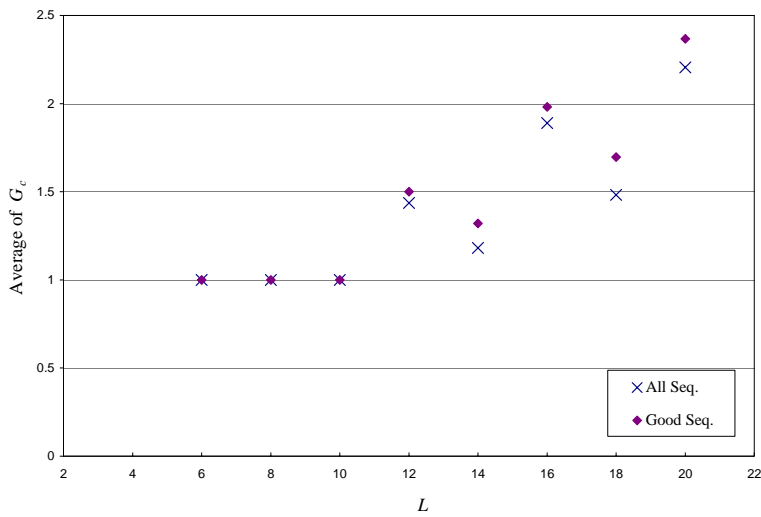
**Figure 3.** The space of energy parameters for sequence  $HPPPHPHPHPPPHPHPHPH$  is divided to six cells. The integer numbers  $(m, a, b)$ , inside any cell indicate the ground state corresponding to the cells. Three of these states are degenerate. The types of degeneracies for degenerate states and shape of structures for non-degenerates are indicated in the cells.

If this plane contains the point  $(x_0, y_0, z_0) = (a, b, m)$ , the  $z$  value on the  $Z$  axis will be equal to  $-E$ . Thus, to find the ground state it is enough to move this plane from above until it touches a state. This state is the ground state. When the plane is moving down, the first contact is occurred in one of the corners of the convex hull of the set of points (i.e. the polyhedral envelope of the states, see Fig. 6). It is possible that for special values of energy parameters in the first contact, the surface matches to one of the edges or planes of polyhedron. In these cases, the energy parameters are exactly in level crossing lines, which are introduced by eq. (9). Thus the possible ground states are in the corners of the convex hull. If  $\gamma$  and  $E_c$  can become negative, all the corner points which can be seen from the top view of this polyhedron, are ground state candidates, but clearly that for positive values of  $\gamma$  and  $E_c$  the number of possible ground states is even





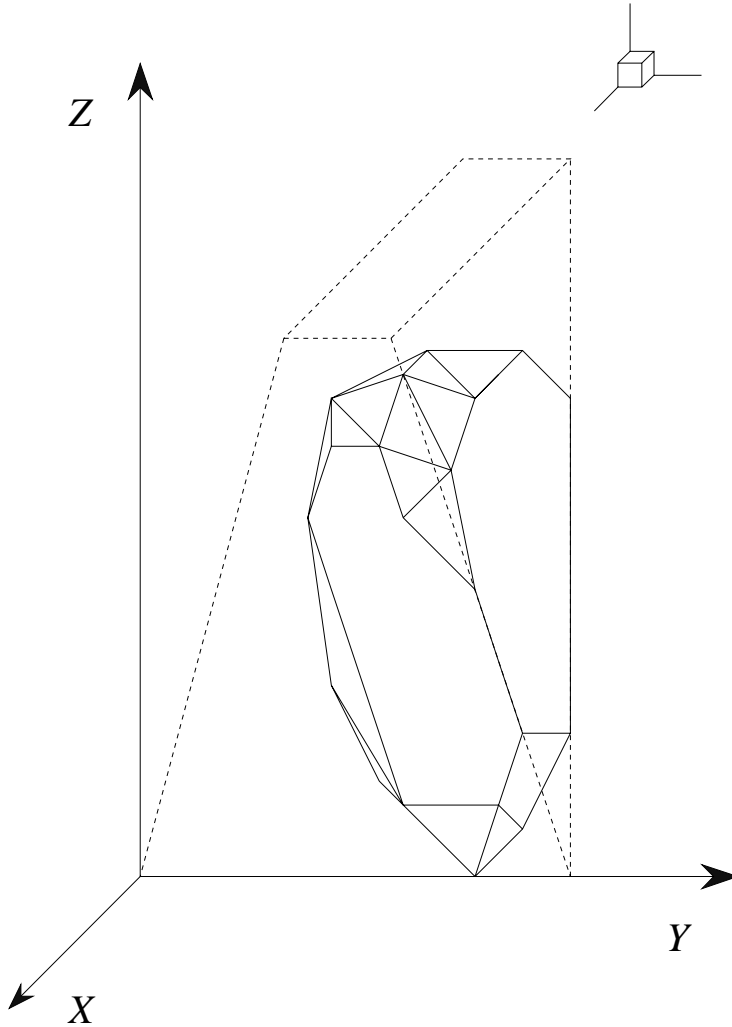
**Figure 4.** The histogram of the number of ground state candidates for 20-mers. The light and dark gray area show the results for all and good sequences respectively. There are some “good sequences” with only one ground state candidate.



**Figure 5.** The average of the number of ground state candidates for all and good sequences *vs.* length of sequences.

smaller. The cross section of the pyramid with a horizontal plane is a rectangle. For positive values of  $E_c$  and  $\gamma$  there is an upper limit for possible ground states. It is equal to the number of possible states in the biggest horizontal rectangular cross section of the

pyramid. The maximum number of contacts is of the order of the length of the sequence, i.e.  $b_{\text{Max}} \sim L$ . Thus this upper limit grows as  $L^2$ . This shows that the number of ground state candidates grows much more slowly than the number of configurations. For example, for 18-mers  $b_{\text{Max}} = 10$ , the biggest cross section is a  $6 \times 6$  rectangle. It thus gives 36 as the maximum number of ground state candidates.



**Figure 6.** State space of the particular sequence which is shown in figure 1. All states are inside a diamond like polygon inside a pyramid. Top viewed corner points of this polygon are the ground state candidates.

As figure 4 shows, there are some good sequences with  $G_c = 1$ . This means that for any set of energy parameter values, they have the same unique ground state. Indeed the native states of these sequences have perfect stability with respect to a change of the energy parameters. Our enumeration shows that these absolute native structures are to be found among the most compact structures. Although the ratio of the number of perfectly stable sequences to the number of all possible proteins decreases with increasing  $L$ , their actual number increases [10]. This suggests that for the proteins with typical lengths near that of natural proteins, perfectly stable sequences constitute a small but non-zero fraction of all possible sequences.

The existence of these sequences may answer some questions about protein folding. Their number is small compared with the huge number of the possible amino-acids sequences, their native states are highly compact and are stable against the changes in the inter-monomer interactions (i.e the properties of the solution).

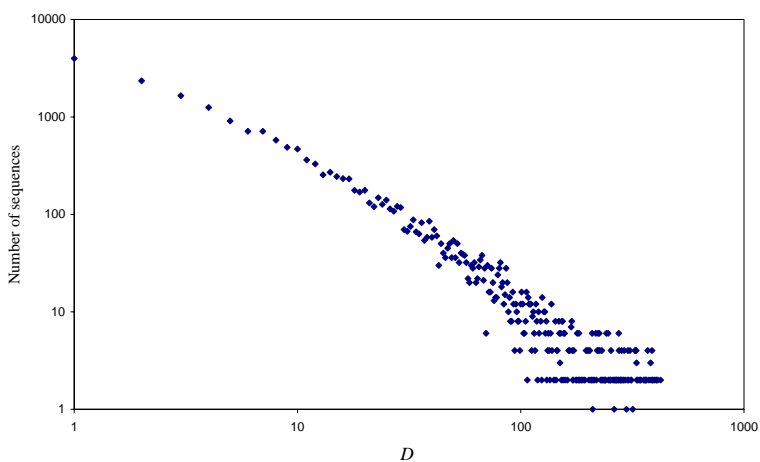
#### 4. Native structures

In section II we introduced the reduced set of contact maps. As it was shown the number of maps belonging to this set  $N_r$ , is very smaller than number of structures  $N_{SAW}$ . But the number of those structures which can be the native state, is still much less. The number of possible native structures,  $N_{native}$ , is shown in figure 1. In this figure all those structures which have been the native state of some sequence for at least one set of energy parameter values, have been counted.

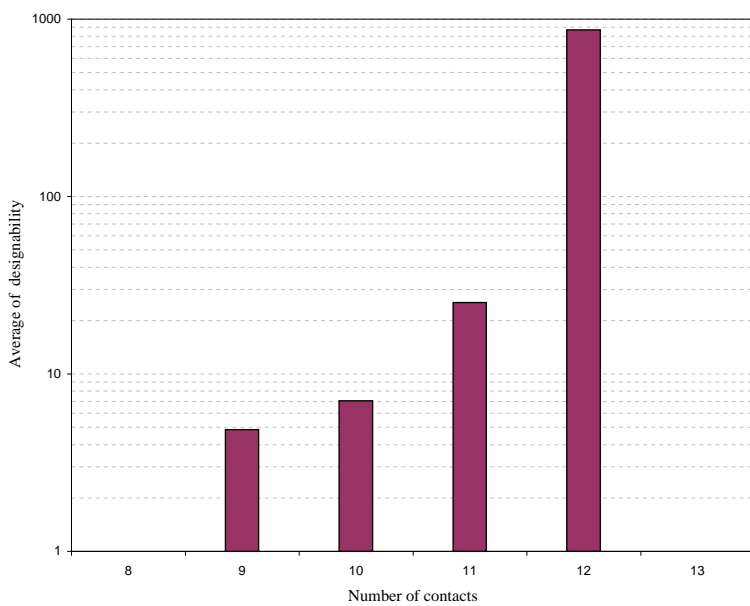
We can introduce a designability parameter  $D$  for these native states. However our definition is a bit different from the commonly used definition [15]. According to the common definition, designability shows how many times a structure is selected as the native state for a fixed set of interaction parameters. In our case we count how many times a structure becomes the candidate for a non-degenerate ground state.

Figure 7 shows the histogram of designability for structures with length 20. As one can see the results are very similar to those for a fixed set of energy parameters in the space of compact structures [6, 15]. The average designability as a function of compactness for  $L = 20$  is shown in figure 8. As the diagram shows the peak average designability occurs for the most compact structures and it falls sharply with decreasing compactness. Thus if one is only interested in highly designable structures, it is reasonable to search the space of compact structures.

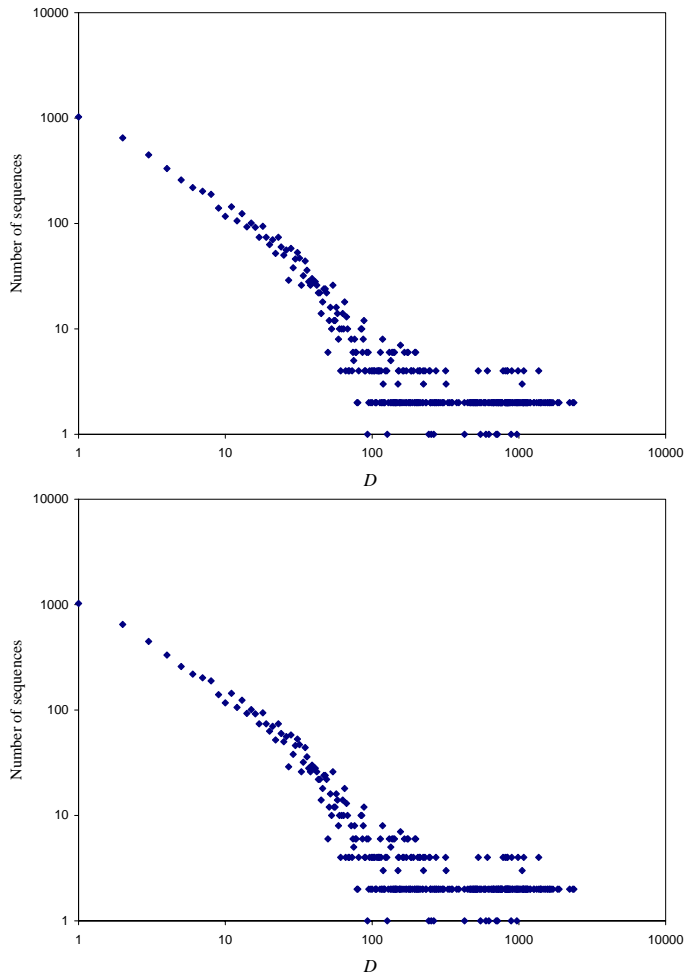
A very interesting result in designability of structures is that, Although by looking at the histogram of designability one can see that the designability of structures depends on the interaction parameters (Fig. 9), but the highly designable structures are fixed and don't change by changing the interaction parameters.



**Figure 7.** The histogram of number of structures with a given designability.



**Figure 8.** The average designability for structures with a given number of contacts, for  $L = 20$ .



**Figure 9.** The Histogram of designability for two different sets of energy parameters; a)  $\gamma = 0.3$  and  $E_c = 1.5$ , b)  $\gamma = 0.3$  and  $E_c = 20$ .

### 5. The space of Energy parameters, $E_c$ and $\gamma$

One of the important aspects of the work done in this paper, is that we can find the exact results for any range of energy parameters. The time it takes for this program to find the ground state candidates for all sequences by exact enumeration, is on the same order as that of the usual enumeration schemes for only one particular set of energy parameters. Because the average number of ground state candidates is very small, the determination of the native ground states for any range of interest only takes a little time. We found

the native states of all sequences of length 20, for all pairs of energy parameters within a  $12 \times 12$  square in arbitrary units, with a grid size of 0.1 (14400 points). The number of protein-like sequences (sequences which have unique ground states) is shown in figure 10. As one can see, there are jumps in the number of protein-like sequences. These jumps specify the borders of regions of relative stability within the space of energy parameters. The large changes in the number of protein-like sequences shows that when we cross these borders the ground states of many sequences change, and the degenerate ground states are replaced by non-degenerate ones (or vice versa). However, nothing can be said about the details of these changes. One can get some idea about what is happening on these border lines by comparing the contour plot for figure 10.a (figure 10.b) with the ordering lines diagram for one particular sequence (figure 3).

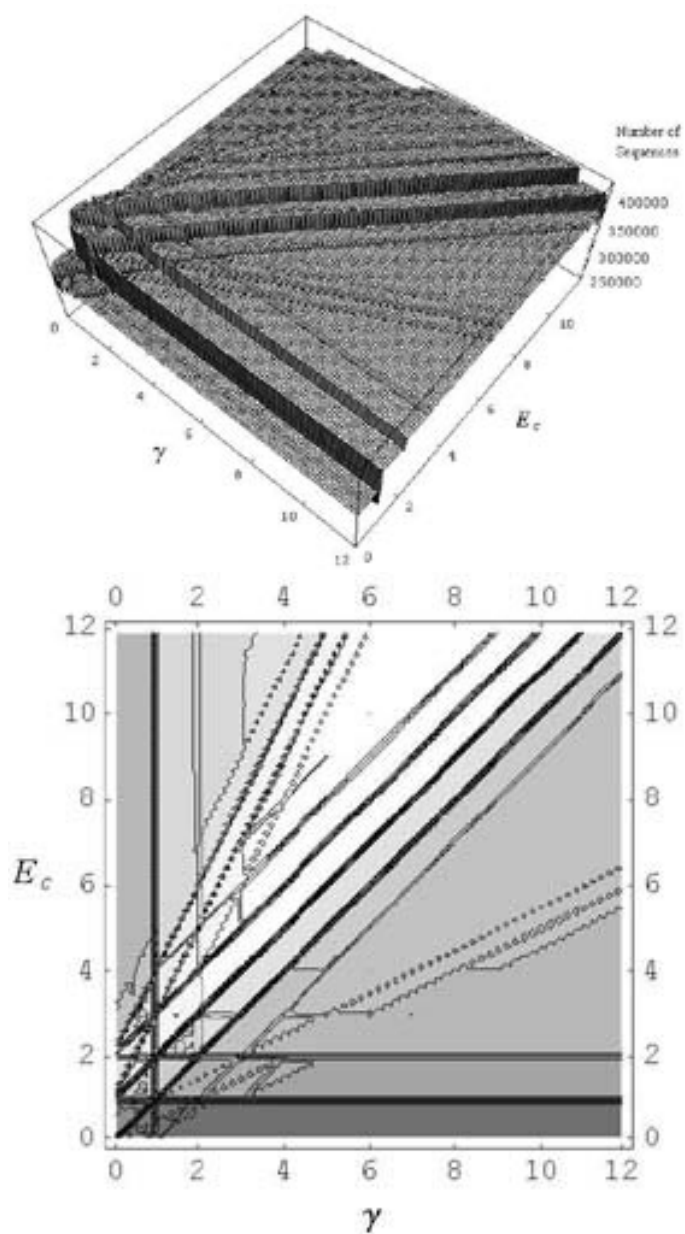
In addition to obtaining information about the sequences, with this procedure also finds the ground states. Since the energy parameters determine which states are the ground states, the number of structures which can be the native state of some particular sequence also depends on the energy parameters. Figure 11 shows the number of native states as a function of the energy parameters. The importance of compactness at for large values of  $E_c$  can also be seen in this diagram. Note that the smallest value for the number of native states is 503. This number corresponds to the number of most compact structures of length 20. Again, large jumps in the number of native states are observed. One can also find the average designability of the structures by dividing the data of figures 10 and 11 (the ratio of the number of sequences to corresponding number of native structures).

## 6. Conclusion

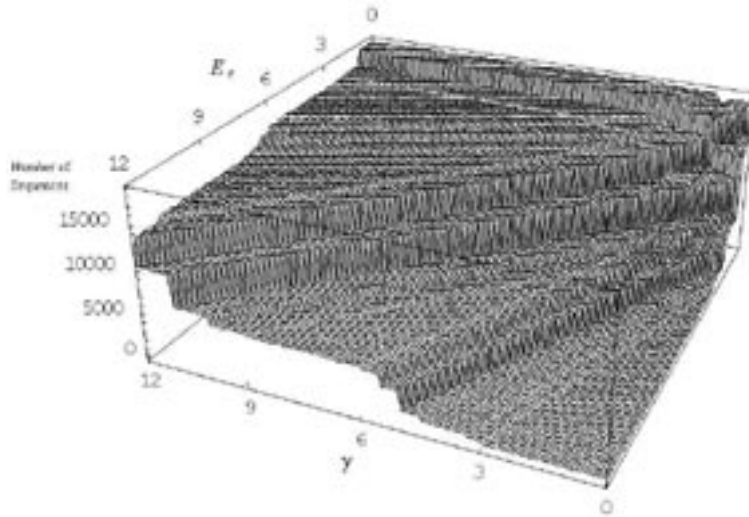
Due to the short-range nature of inter-monomer interactions, the configuration energy of protein sequences can be determined by using configuration contact matrices. In this paper, it has been shown that for this class of problems, where one is interested in native states of proteins, the space of physical contact maps can be reduced to a very smaller set by removing all irrelevant maps. We have found the reduced set of contact maps for sequences of lengths up to 20 in this paper by exact enumeration. This reduced set of contact maps shows a scale-independent behavior.

Using the reduced set of contact maps, the ground state candidates for all sequences were found in the HP model. The number of these ground state candidates is quite small. The ground state candidates divide the space of energy parameters into several cells. By finding this cell structure for all sequences, we have found the native states for all sequences of different lengths, for a wide range of energy parameters. Jumps are observed in the number of protein-like sequences. These jumps are related to boundaries of the aforementioned cells.

Another interesting result is that we find some sequences with absolute native states i.e. their native states are not sensitive to the values of energy parameters. Our results show that the number of such perfectly stable sequences grows with length, however, their percentage decreases.



**Figure 10.** The number of protein-like sequences of length 20, for given values of energy parameters in a  $12 \times 12$  square region (arbitrary units); a) Three dimensional plot, b) Contour plot.



**Figure 11.** The number of native states for sequences of length 20, for given values of energy parameters in a  $12 \times 12$  square region (arbitrary units).

Our results show that the existence of some highly designable structures is a geometrical property of such lattice models and it does not depend on interaction energy parameters.

Because the key tool used in this paper has been the structural information contained in the contact maps, the qualitative results can be generalized to all contact models, regardless of the details of the lattice and the contact rules. The dimension of state space is related to the model and the number of energy parameters. For example, if we look for the ground state in the space of compact configurations,  $E_c$  is an irrelevant parameter. In this special case the space of energy parameters is one dimensional (only  $\gamma$ ), and the space of states is two dimensional ( $a$  and  $m$ ) [6]. This argument can be generalized to models with more than two kinds of monomers, and also to off-lattice models. If the inter-monomer interaction has  $t$  free parameters, the energy levels can be described by  $t + 1$  integer. Similar to our case, these integers can be explained in terms of kind of contacts between the monomers. The simplest choice is the number of contacts between monomers type  $i$  and  $j$  ( $n_{ij}$ ), but there is no reason that it is the most convenient choice. Because of geometrical constraints on the number of contacts, there are some relations between these parameters similar to eq. 8, and our argument can be followed in the same way. The argument even can be generalized to models with  $n$ -body interaction ( $n > 2$ ). The introduction of  $n$ -body interaction only increases the difference the dimensionalities of state space and the space of energy parameters [16]. Therefore, quite generally, the ground state candidates of any given sequence are between the corner states of a hyper polyhedron in a hyper space which is very smaller than the number of all possible structures.



## Acknowledgement

We would like to thank S.E. Faez, R. Gerami, R. Golestanian, A.Yu. Grosberg, N. Heydari, M. Khorami, S. Rouhani and H. Seyed-Allaei for helpful comments.

## References

- [1] S. Miyazawa and A. Jernigan, *Macromolecules* **18**, 534 (1985).
- [2] S. Lifson and C. Sander, *Nature* **282**, 109 (1979).
- [3] For a recent review, see M. Vendruscolo and E. Domany, cond-mat/9901215.
- [4] C.B. Anfinsen, *Science* **181**, 223 (1973).
- [5] V.S. Pande, A.Yu. Grosberg and T. Tanaka, *J. Chem. Phys.* **103**, 9482 (1995).
- [6] M.R. Ejtehadi, N. Hamedani, H. Seyed-Allaei, V. Shahrezaei and M. Yahyanejad, *Phys. Rev. E* **57**, 3298 (1998); M.R. Ejtehadi, N. Hamedani, H. Seyed-Allaei, V. Shahrezaei and M. Yahyanejad, *J. Phys. A* **31**, 6141 (1998).
- [7] E.L. Kussell and E.I. Shakhnovich, Preprint Cond-mat/9904377.
- [8] M.R. Ejtehadi, N. Hamedani and V. Shahrezaei, *Phys. Rev. Lett.* **82**, 4723 (1999).
- [9] M. Vendruscolo, B. Subramanian, I. Kanter, E. Domany and J. Lebowitz, Preprint Cond-mat/9810285.
- [10] V. Shahrezaei, N. Hamedani and M.R. Ejtehadi, to appear in *Phys. Rev. E*, (Oct. 1999)
- [11] H.S. Chan, K.A. Dill, *J. Chem. Phys.* **90**, 492 (1989); H.S. Chan, K.A. Dill, D. Shottle, "Statistical Mechanics and Protein Folding", *Princeton Lectures on Biophysics*, W. Bialek ed., (World Scientific, 1992).
- [12] R. Melin, H. Li, N. S. Wingreen and C. Tang, Preprint Cond-mat/9806197.
- [13] K.A. Dill, *Biochemistry* **24**, 1510 (1985); K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas and H.S. Chan, *Protein Science* **4**, 561 (1995).
- [14] H. Li, C. Tang and N. Wingreen, *Phys. Rev. Lett.* **79**, 765 (1997).
- [15] H. Li, R. Helling, C. Tang and N. Wingreen, *Science* **273**, 666 (1996).
- [16] N. Hamedani, V. Shahrezaei and M.R. Ejtehadi, In Preparation.