*IOSR Journal of Applied Physics (IOSR-JAP)*
*e-ISSN: 2278-4861.Volume 7, Issue 3 Ver. I (May. - Jun. 2015), PP 58-62*
*www.iosrjournals.org*

# Energy Landscape Theory for Protein Structure Prediction Using a Lattice Based Coarse-Grained Model

Samson O. Aisida and Oluwole E. Oyewande
[1]*Department of Physics and Astronomy, University of Nigeria, Nsukka, Nigeria*
[2]*Department of Physics, University of Ibadan, Ibadan, Nigeria*

***Abstract:*** *The spontaneous folding of protein has posed a fundamental problem popularly known as Levinthal paradox. Effort has been harnessed both in vivo and in vitro to obtain the native (functional) structure of proteins which is the most stable thermodynamically from the experimental and theoretical point of view. A coarse-grained simplified model has been a very useful tool for the study of protein structure by representing proteins as linear and self-avoiding chains which contain two types of monomers H (hydrophobic) and P (polar) on a lattice model. We present the results of numerical studies of the energy landscape on three sequences of amino acid and distinct funnels were generated with contact interactions to obtain the ground state conformation on a square lattice using a move-biased Monte Carlo simulation (MBMC). This method is very efficient and outperforms the conventional Monte Carlo method in the state-of-the-art results.*
***Keywords:*** *Coarse-grained, Energy landscape, Monte Carlo, Protein folding*

## I. Introduction

Proteins are macromolecules composed of sequence of amino acids displaying complex and heterogeneous interactions and serve as the workhorse of living cells. The folding of proteins to obtain the native (functional) structure is a fundamental problem in molecular biophysics. Effort has been harnessed both in vivo and in vitro to know how the protein machine work to obtain the native structure which is the most stable thermodynamically from the experimental and theoretical point of view. The aim of predicting the three-dimensional (native) structure of proteins from the sequences of their amino acid alongside their folding pathways has been one of the most important tasks in computational biophysics. The predicted structures are very crucial to pharmacology and medical sciences. This has been of keen interest to many researchers in Biophysics, Physics, Molecular biology and Biochemistry in the last decade in the area of experimental, theoretical and computational approaches since folding can be seen as the final step between genetic information and biological functions. The knowledge of 3-D conformation of protein is crucial to drug design. Aberration from this due to structural differences leads to "misfolding" the origin of most diseases such as Alzheimer, Huntington, Parkinson and Cancer related diseases e.t.c. Variations in proteins structures make people respond to drugs differently. Understanding these differences opens new possibilities in drug design, diagnosis and disease control. [1, 2, 3]

The structure of this paper is as follows: in section 2, we present the theoretical background via: a brief description of the physics of folding energy landscape, protein structure prediction (PSP) and coarse-grained model. The searching procedures and the results are presented in section 3 and 4 respectively. Finally, the conclusion is given in section 5.

## II. Theoretical Background

### 2.1 The Physics of folding energy landscape

The energy landscape of a random heteropolymer like the landscape of structural glasses ultimately resembles the most extreme case of energetic ruggedness, similar to random energy model introduced by Dorrinda to model spin glasses. An energy landscape is a surface defined over conformation space indicating the potential energy of each and every possible conformation of the molecule [2]. This approach is the theoretical manifestations of the interactions that contribute to the chemical processes. Greta et al [4] describe the landscape as rugged and broad rather than smooth surface that is reflected in the non-linear folding kinetics. The spontaneous folding of protein has posed a fundamental problem popularly known as Levinthal paradox. The energy landscape theory which entrenched the statistical mechanics nature of the folding process proffers a solution to this paradox. The protein structure prediction phenomenon was largely an experimental endeavor using x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy which were limited by some constraints until the formulation of an energy landscape theory of proteins by Joseph Bryngelson and Peter Wolynes in the late 1980s and early 1990s. Proteins which borrowed some concepts and techniques from physics of spin glass field are viewed as a system with minimal frustration in its native conformation. Proteins exhibit similar complexity with spin glass for which many different folding configurations may have almost the

same energy. Also, the concept of two-state transitions (conformational transition) from unfolded (denatured) state to the folded (Native) state is synonymous to phase transition in spin glass. The folding landscape topology is like a funnel using the principle of minimal frustration which underlies the thermodynamic illustration of protein folding. According to this illustration, protein negotiates a rough, funnel-shaped energy landscape during the folding process and ultimately settles in a state that, as much as possible satisfies the energetical constraints as a result of the covalent interactions. The concept of minimal frustration has been made quantitatively precise by using the statistical mechanics of spin glasses. The energy landscape theory declares that without much loss of kinetic information protein folding can be captured by one or a small number of reaction coordinates. This reaction co-ordinate is rarely obtainable experimentally [3, 5, 6, 7, 8, 9].

In an energy landscape, valleys in the folding funnel indicate stable low energy conformations while mountains indicate unstable high energy conformations (as shown in figure 2) and many proteins with fast folding kinetics are known to be within or near a downhill regime [2, 10, 11]. The energy of the conformations in the folding sequence and the reaction coordinate is expected to be proportional with some roughness that is introduced by non-native contacts. This correlation of energy and structure introduces a bias in favor of the native conformation as well as a bias against the non-native structure. Such a correlation is responsible for the funnel shape of the landscape [11].

**2.2 Protein structure prediction (PSP)**

The main goal of PSP is to model the free energy of the given amino acid chain and then to find minimum energy conformations. Experiment and theory show folding proceeds fairly directly to the native structure which is energetically very stable. The challenges in finding out the native (3D) structure from a given sequence is a problem that emanated from Anfinsen's discovery [12] "which says that the sequence of amino acids of a naturally occurring protein uniquely specifies its thermodynamically stable native structure". How the native three-dimensional structure emanated from the sequence of amino acid is a pursue in which experimental methods like X-ray crystallography, Nuclear magnetic resonance (NMR), Protein engineering e.t.c have been used to play vital roles in the analysis, but these methods are very slow, and capitally intensive when compared with the computational methods which was actually formulated to find the global minimal of a potential energy function. Although, PSP has been proven to be NP complete (i.e problem considered cannot be solved optimally within a reasonable time 'polynomial time') even from the application of the simplest hydrophobic-hydrophilic (HP) lattice model like 2D-square and 3D-cubic lattices [13].

**2.3 Coarse- grained model**

The processes involved in protein folding are very complex as a result of the large ($N\longrightarrow\infty$) number of the degrees of freedom. Sequel to this, a coarse-grained simplified model has been a very useful tool for the study of protein processes. In this case instead of representing each atom in the protein, definite groups of atoms can be treated as a single coarse-grained site where the interactions are approximated to capture the important physical features and the amino acid residues are coarse-grained by single monomers. The coarse-grained model is of two types, one is lattice-based and the other is off-lattice based model, both of which have two types of amino acids, hydrophobic (H) and polar (P). The lattice-based type is designed to understand the basic Physics governing the protein folding process. It can be use to extract vital dogma, make predictions and harmonize our understanding of many different properties of proteins. The major lattice-based coarse-grained model that provided deep insights into the physical principles of folding is hydrophobic-Polar (HP) model by Dill [14].

**2.4 HP lattice model**

The HP lattice model is a standard model with simple energy function; view it from the perspective of statistical mechanics, this model is very rich in thermodynamic behaviors. Since the HP chain is constructed specifically to represent an individual protein. Hence, thermodynamic properties of the chain depend on its chain length and the sequence of H and P monomers uniquely. In this model, focusing on the hydrophobic effect, protein is described by its sequence $\{\delta_i\}$ of N amino acids which takes the values H and P and $i$ is a monomer index, and their positions $\{x_i\}$. The bond vectors $\{b_i\}=\{r_{i+1}-r_i\}$, with fixed (unit) length. The energy of a structure is given by sequence – independent local interactions $\left(\mathcal{H}_{Loc}\right)$ and sequence-dependent nearest-neighbor contact (non-local) $\left(\mathcal{H}_{NonL}\right)$ interactions:

$$\mathcal{H}=\eta\mathcal{H}_{Loc}+\mathcal{H}_{NonL} \tag{1}$$

$$\mathcal{H}_{Loc}=2\sum_{i=2}^{N-1}(1-b_i\cdot b_{i-1}) \tag{2}$$

$$\mathcal{H}_{NonL} = \sum_{1\le i<j\le N} \in_{\delta_i\delta_j} (x_i, x_j)\pi_{ij} \qquad (3)$$

Where $\pi_{ij} = 1$, if monomers $i$ and $j$ are lattice neighbors and $\pi_{ij} = 0$ otherwise. $x_i$ defines the type of amino acid residue at position $i$. The energy depends on three parameters which determine the strength of non-local interaction $\in_{HH,} \in_{HP,}$ *and* $\in_{PP}$, according to Lau and Dill [15] $\in_{HH} = -1; \in_{HP} = \in_{PP} = 0$. While $\eta$ determines the strength of the local interactions. For $\eta = 0$, the model is identical to the HP model. The hydrophobic effect is modeled by having stronger attraction between HH pairs than between HP and PP pairs of amino acids; hence the hydrophobicity of the non-polar amino acids is considered as the main driving force of obtaining the tertiary structure of protein [16, 17].

We represent proteins as linear and self-avoiding chains which contain two types of monomers H (hydrophobic) and P (polar) on a two-dimensional lattice model, in which self-avoiding walk is used to represent the protein chain such that covalently linked residues occupy neighbor lattice sites. The energy of a conformation is the sum of the energies of pairwise contacts between monomers. Two monomers are defined to be in contact if they are neighbors on the lattice and not connected by a covalent bond. The energy of the contact depends only on the identity of the two amino acids residues involved. The interaction energies for residue pairs are determined from the statistical distribution of contacts in real proteins.

## III. The Optimization Searching Procedure

In this paper, we implement the move-biased Monte Carlo simulation (MBMC) on 2D-square HP lattice model. This method incorporates the coupled neighborhood search strategy (diagonal-pull move) for the protein structure prediction. The MBMC is a class of heuristic global optimization and a generation of Monte Carlo (MC) method. The implementation of MBMC as a tool to fold up given sequences on a square lattice model includes the set of selected benchmark protein sequences in table 1. The adopted method generates a parent conformation by simulating the evolution of the HP sequence in the conformational space. The move biased (coupled moves) is then applied to improve the parent conformation by building the hydrophobic-core and improve the premature conformation to obtain a new conformation by exploring its neighbourhood. The new conformation is accepted if and only if the H-H contact of the new conformation is higher than the parent conformation (in other words, the energy of the new conformation is lower than the parent conformation) otherwise, it is discarded and the process continues until the lowest energy is obtain.

**Table 1:** Selected 2D standard benchmark protein sequences [18]. $H_i$ and $P_i$ are Hydrophobic and polar amino acids respectively.

| Instances | N | PDB ID | Protein Sequence (H-hydrophobic, P-polar) | E[a] | E[b] | E[c] |
|---|---|---|---|---|---|---|
| 1 | 20 | SI-1 | $(HP)^2PH^2PHP^2HPH^2P^2HPH$ | -9 | -9 | -9 |
| 2 | 48 | SI-2 | $P^2HP^2H^2P^2H^2P^5H^{10}P^6H^2P^2H^2P^2HP^2H^5$ | -23 | -21 | -20 |
| 3 | 64 | SI-3 | $H^{11}(HP)^3P(H^2P^2)^2HP^2(H^2P^2)^2HP^2(H^2P^2)^2(HP)^2H^{12}$ | -42 | -42 | -35 |

[a] the ground state energy, [b]the MBMC energy, [c]the conventional Monte Carlo energy.

Using MBMC, we enumerate for a given sequence a set of optimal structure showing a compact hydrophobic core and shape. We implement the algorithm in Silverfrost FTN 95 compiler and run it on a laptop computer with an intel Pentium Dual-core CPU, 2.30 GHz processor and 4.00GB of RAM.

## IV. Results

The three (3) sequences investigated are described by obtaining the folding funnel characterized by many local minima and few global minima. Figure 1 shows the folding funnel of the sequences investigated and the conformation of few global minima. Figure 2 gives the conformations with unique global energy minimum of the selected sequences.

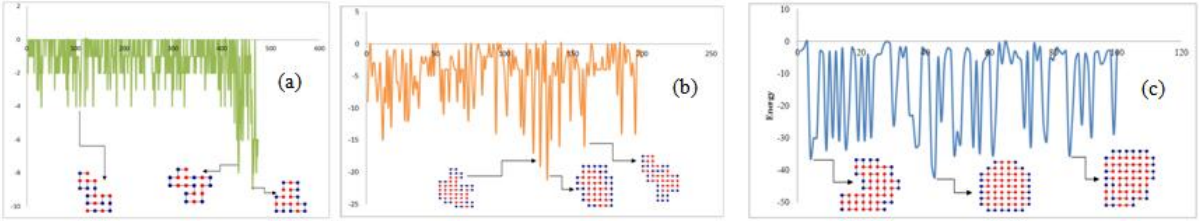

**Fig. 1** (color online) the energy landscape and conformations of selected sequences (a) N = 20 (b) N = 48 and (c) N = 64. Each funnel represents a conformation of energy against the number of iterations

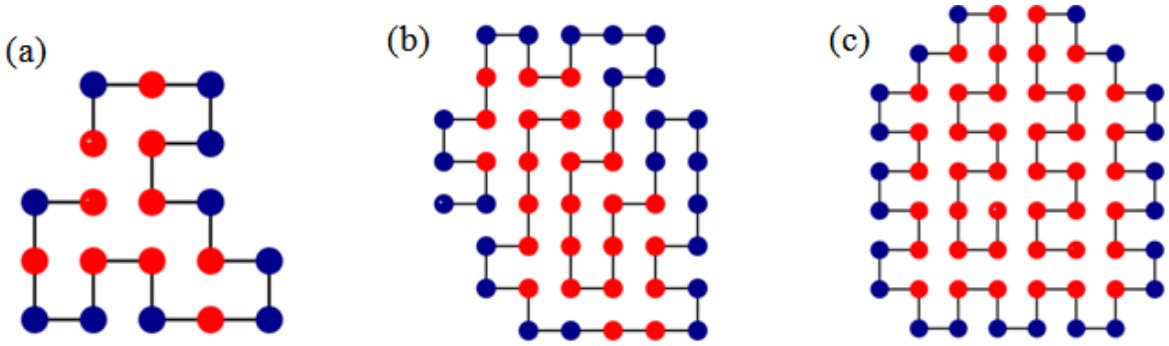


**Fig. 2** (color online) Shows the unique ground state energy conformations for the 3 sequences (the blues are polar while the reds are hydrophobic, the plus and the minus signs are the starting and the ending points respectively)

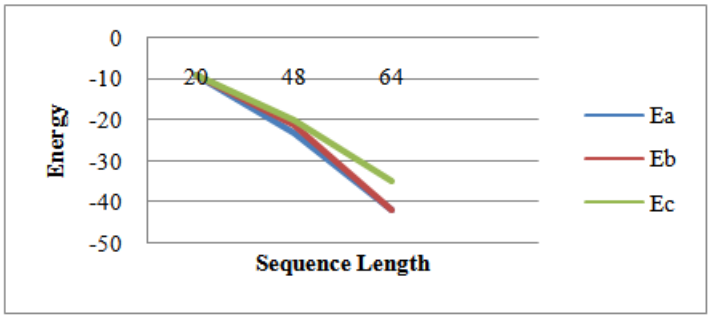


**Fig. 3** (color online) the performance of MBMC with CMC method; the results are calculated for at least 100 iterations.

From figure 3, it is obvious that the MBMC finds different lower free energies than the CMC for the three instances. We noticed that our method has achieved very good progress especially with higher sequences.

## V. Conclusion

The current theoretical framework to have a better insight into protein folding is based on the energy landscapes which govern both the folding kinetics and its thermodynamics. Protein folding is generally a complex process that requires a concerted effort between experiment and theory. In this paper we have demonstrated the strength of the MBMC on 2D lattice model in searching for the native state conformation of some selected sequences through their folding funnel. The method incorporates a coupled neighborhood search strategy (diagonal-pull moves) on premature conformations to obtain the ground energy conformations. Our results outperform the conventional Monte Carlo (CMC) method in all the instances.

## References

[1]. K. Takeshi, Spin correlations in a non-frustrated one-dimensional spin system and formation of the ground state as a model of protein folding, Physical A , 388, 2009, 129-136.
[2]. E. A. Scott, W. Johannes, & G. Frauke, Dynamic prestress in a globular protein, Plos Computational Biology , 2012
[3]. T.-A. Carlo, I. Ylva, J. Per, & G. Stefano, Folding and stability of globular proteins and implications for function, Current opinion in structural Biology, 19, 2009, 3-7.
[4]. H. Greta, P. W. Soren, C. N. Celestine, S. Kristian, & G. Stefano, An expanded view of the protein folding Landscape of PDZ domains, Biochemical and Biophysical Research Communication , 42, 2012, 550-553.
[5]. C. M. Allan & D. A. Ashok, Protein Folding at single-molecule resolution, Biochimica et Biophysica Acta , 1814, 2011, 1021-1029.
[6]. S. Plotkin, & J. Onuchic, Understanding protein folding with energy landscape theory-part 1: basic concepts, Q. Rev. Biophy , 35, 2002, 111-167.
[7]. R. Best, & G. Hummer, Coordinate-dependent diffusion in protein folding, Proc. Natl. Acad. Sci. USA , 107, 2010, 1088-1093.
[8]. B. S. Ginka, M. D. Ronan, N. V. Buchete & J. Kubelka, Dynamics of protein folding: Probing the kinetic network of folding-unfolding transitions with experiment and thoery, Biochimica et Biophysica Acta , 2011, 1814.
[9]. B. Schuler, & W. Eaton, Protein folding Studied by single- molecule FRET, Curr. Opin. Struct. Biol. , 18, 2008, 16-26.
[10]. B. M. Oren, M. D . Alexander, B. RouX, & W. Masaktsu, Computational Biochemistry and Biophysics (New York, United State of America: Eastern Hemisphere, 2001)

[11]. C. Sooyoung & L. Faming, Folding small Proteins via annealing stochastic approximation Monte Carlo, Biosystems , 105, 2011, 243-249.
[12]. C. Anfinsen, Principles that govern the folding of Protein chains, Science , 181, 1973, 223-239.
[13]. I. L . Jingfu, S. Beibei, L. Zhaoxia, H. Weibo, S. Yuanyuan, & L. Wenjie, Energy-landscape Paving for prediction of face-centered-cubic hydrophobic-hydrophilic lattice model proteins, Physical Review E , 052704, 2013
[14]. K. A. Dill, S. Ozkan, M. Shell, & T. Weikl, The Protein folding prblem, Biophys , 37, 2008, 289-326.
[15]. K. Lau, & K. Dill, A lattice Statistical Model for the Conformational and Sequence Spaces of Proteins, Macromolecules, 22, 1989, 3986-3997.
[16]. W. L. Ying, W. Thomas & L. P. David, Monte Carlo simulation of the HP model (The "Ising model" of protein folding), Computer Physics Communication, 182, 2011, 1896-1899.
[17]. T. Wust, D. Landau, C. Gervais, & Y. Xu, Monte Carlo simulations of systems with complex energy landscapes. Computer Physics Communication, 180, 2009, 475-479.
[18]. A. Shmygelska, H. Hoos, An ant colony optimization algorithm for the 2D and 3D hydrophobic polar protein folding problem, BMC Bioinformatics, 6:30, 2005.