Arch. Bas. App. Med. 3 (2015) 37- 42

ARCHIVES OF BASIC
& APPLIED MEDICINE

Research Article

The Effect of Amino-acid Gyration on Protein Conformation

Aisida S.O[*] and Oyewande O.E[**]

[*] Department of Physics and Astronomy, University of Nigeria Nsukka
[**]Department of Physics, University of Ibadan, Ibadan

Abstract

Proteins are basically polymers of amino acids that are linked together through peptide bonds. The biological function of a given protein depends on its native structure, which has the lowest free energy level; hence finding out the structure is very important. The de novo protein method for predicting the structure of proteins has been proved to be nondeterministic polynomial time (NP)-hard even with short lattices. In order to facilitate a study of the factors influencing conformation of a protein to its native structure, we develop a Monte Carlo model of protein conformation on square lattice. We studied the effect of amino-acid gyration on the native conformation of a protein. We found that this model is very effective and gives compact structures for isotropic gyration.

Keywords: Protein folding, HP lattice model, Monte Carlo, directional probability.

INTRODUCTION

Proteins, like other biological macromolecules such as polysaccharides, lipids and nucleic acids are essential parts of organisms, exist and functions in aqueous environments, and participate in virtually every process within cells and act as the workhorse on which all life-form on earth relies (Tyers and Mann, 2003; Michael, 2003). Hence, the importance of proteins to life cannot be overemphasized. For instance, Hugh et al., (2013) observed the increased level of homocysteine, an amino acid with dementia for both Yoruba and African Americans, as a result of aberrant proteins. They observed that increase in homocysteine levels posed a significant increase in dementia risk among the elderly people in the observed communities. Ogunniyi et al., (2011) studied hypertension as a risk factor for dementia among the elderly in Nigeria communities which is also a result of aberrant protein in the body common to elderly people aged 55 and above.

The amino acid sequence of each protein determines how it folds into a unique three-dimensional (native) conformation, which is the minimal free energy state (Hans-Joachim and Dirk, 2007). A protein can then be unfolded or denatured by adding some denaturants like solvent, pH, temperature e.t.c. These denaturants change the protein into a flexible chain that has lost its natural shape. When the denaturant is removed, the protein refolds into its original conformation. More than half of the dry weight of a cell is made up of proteins of various shapes and sizes and a protein's specific folded three-dimensional (3D) shape enables it to perform specific tasks. The knowledge of 3-D conformation of protein is crucial to drug design (pharmacology), to an understanding of the causes of the misfolding due to structural differences, i.e the origin of most diseases (medical science), and has some technological implications.

All the information necessary to specify the three dimensional (native) structure, which determines outright the function of the protein, is contained in its amino acid sequence. Over the years, experimental techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been veritable tools to determine the 3-D structure of a protein. As potent as these techniques are, they are very expensive, time- consuming, laborious and restricted to some specific conditions. Since two decades ago, scientists have been devising computational approaches, which circumvent the experimental challenges to predict the 3-D structure of a protein. But the native structure of most proteins is still unknown.

This paper addresses the protein folding problem which consists of predicting the native structure of a protein from its sequence of amino acids. The problem involves conflicting constraints as well as rugged energy landscape (Anderea et al., 2001; Seno et al., 1996). The folded structure must be stable and fold within a reasonable time interval (Martin, 2011). We studied the influence of the gyrations of the protein sub-units (amino-acids) on the protein conformation with a straight forward self-avoiding walk (SAW) Monte Carlo procedure and test it within the frame work of a 2D square lattice backbone-only model for chains with up to N = 50 monomers. The fundamental point is to view the effect of some gyrations, which could enhance the 3D conformation.

MATERIALS AND METHODS

Theoretical background: Folding of protein is driven by non-bonded interactions which are represented as "contact energies" in the lattice model involving interactions between residues that are situated on adjacent (or nearest-neighbor) lattice site but are not covalently bonded together to each other.

**Author for correspondence: +234 8063957679*
E-mail: Samson.aisida@unn.edu.ng

These interactions manifest in the gyrations of the amino-acids and translate to directional probabilities in our SAW model. It is generally accepted that protein folding is driven mainly by the hydrophobic effect, which is the tendency of protein monomers to be repelled by water molecules. But, what is really lacking is an understanding of specific physical mechanism or principles governing the folding process. The main idea of our model is the simulation of these gyrations in terms of directional probabilities that maps the simulated structure to the native structure, which, to our knowledge is a new approach. The mechanism may not uniquely obtain the folded structure, but it will drive it towards a sort of basin of attraction, which will give the basis for convergent evolution of the conformation.

In lattice models, a protein of length x and lattice L is usually considered as being defined by its sequence $\sigma = (\sigma_1, \ldots\ldots, \sigma_x) \in L^x$ where the nature $k(\sigma_i) = -1$, if subunit σ_i is hydrophobic and 0 otherwise. σ_1 is the N-terminal and σ_x the C-terminal residue assigned to a run of at least three helix residues or three strand residues and k is the number of contact. Its (coordinated) structure $S = (S_1, \ldots\ldots\ldots, S_x)$ where S_i represents the position of the ith monomer and the Euclidean distance between the coordinates S_i and S_j is represented as $\xi(S_i, S_j) = |S_i - S_j|$

A general theoretical model is the two-dimensional hydrophobic-polar (HP) model which was first proposed by Dill (Dill, 1985) as a free energy model where the low energy conformation is favored with a hydrophobic core by allowing the hydrophobic residues, which are less ionic and have low affinity for water, to cluster inside while the hydrophilic/polar residues which are ionic and bond well with water are at the surface. In this model protein folding and conformation are attributed to the repulsion or attraction between the hydrophobic (H) unit or the polar (P) unit respectively and water molecules.

In the HP model, the 20 kinds of amino acids are grouped into two letters, H for hydrophobic monomers and P for polar monomers. HP model is feasible if one can associate a protein sequence with a unique ground state energy minimum and the sequence must be able to attain the minimum energy structure in a short time interval. The "contact" energy E of a given conformation is typically calculated by summing the values of energies over all nonbonded contacts in the lattice (Seno et al., 1996; Irback et al., 1998; Mann et al., 2008, 2009)

$$\mathrm{E} = \sum_{neighbors} \sigma_{ij} \Delta(q_i, q_j) \tag{1.1}$$

where q_i and q_j denote the locations of residues at i and j and

$$\Delta(q_i - q_j) = 0 \tag{1.2}$$

unless residues i and j are on adjacent vertices of the lattice. While σ_{ij} indicates the nonbonded neighboring interaction between the residues of i and j. These contact interactions are typically on the order of $k_B T$. The conformation of the protein is then self-avoiding walk (SAW) on a 2D lattice such that its energy E which depends on the topological neighboring contacts between hydrophobic amino acids is minimal.

To obtain a conformation that is stable with unique ground state energy minimum in this model, a directional probability $P(x_0 / \sigma)$ plays a vital rule. This probability is given by

$$P(x_0 / \sigma) = \frac{\exp[-\mathrm{H}(x_0, \sigma)/T]}{Z(\sigma)} \tag{1.3}$$

Where x_0 denotes the target structure, T the temperature, $\mathrm{H}(x_0, \sigma)$ is the energy and the partition function $Z(\sigma)$ is given by

$$Z(\sigma) = \sum \exp[-\mathrm{H}(x, \sigma)/T] \tag{1.4}$$

The objective function of this model is to maximize the number of H-H contacts, that is, the number of adjacencies in the lattice between hydrophobic amino acids. For a given sequence, a structure is called native if and only if it's $E(H, P)$ is minimal among all structures of the sequence (Abkevich et al., 1995). Also, the energy of a given conformation (ζ) is defined as the number of topological neighboring (TN) contacts between those of Hs (i.e the adjacent hydrophobic amino acids that are not neighbors in the sequence) in the lattice, denoted by H-H with $E_{HH} = -1$, and $E_{HP} = E_{PP} = 0$. If a conformation is denoted as $\zeta = \zeta_1, \zeta_2, \ldots, \zeta_x$; $\zeta_i \in [H, P]$ and $i \in \{1, 2, \ldots\ldots, n\}$, where ζ_i is H if ith amino acid is hydrophobic and P if it is polar. Therefore, if we have λ such H-H TN contacts, its energy $E(\zeta) = \lambda(-1)$, the energy evaluation focuses on hydrophobicity only; a conformation with the highest number of H-H contacts is the one with the lowest free energy.

Monomer Gyration Move-Biased (MGMB) model

We model the gyration of the monomers, due to the interactions of nonbonded monomers, as the directional probabilities of a SAW. SAW is a random walk that is prohibited from revisiting an old site previously visited. SAW was first proposed about 5 decades ago as a standard model of a linear polymer molecule of long chain in a good solvent (Madras and Sokal, 1988; Madras and Slade, 1993; Alan, 1996). However, a lot of software used for protein structure prediction iterates on SAW subsets but our model differs in the inclusion of a mapping process to account for the fact that different protein-solvent systems correspond in our model to different sets of the directional probabilities and SAW length. Our model also include a post-SAW relaxation move in which case the SAW generated structure relaxes into vacant sites via pull moves (move-biased).

In our approach we simulate a protein conformation as a SAW on a square lattice to model the folding process but if the walk terminates before a specified length required for the map, we retrace the step and try it again until it continues and terminates on the specified length.

In this paper we analyze the influence of the variation of the probabilities, of the four possible directions which a SAW may take, on the sequence length (length of simulated protein).

We let $\alpha = \frac{P_u}{P_d}$, where

$P_u = 0.005 \cdots 1/2, P_d = 1/4$ so that

$$0.02 \le \alpha \le 2.0 \quad (1.5)$$

$\beta = \frac{P_r}{P_l}$, where

$P_r = 1/2 \cdots 0.005, P_l = 1/4$ so that

$$2.0 \ge \beta \ge 0.02; \quad (1.6)$$

$$d = \frac{\alpha}{\beta} \quad (1.7)$$

Where P_u, P_d, P_r, and P_l are the probabilities of up, down right and left step respectively.

Large d represents very low beta relative to alpha (since the maximum of alpha is 2.0) d increases as beta tends to 0.02 this means that two directions are formed for high d, probability of up and that of down (see figure 1). For low beta, the probability of right is lower than left; that is why the sequence length may be short because two directions are favored. The fluctuation in the sequence length as a function of alpha in figure 2(a) is due to the other factors (directions). If all other directional probabilities are fixed; then the fluctuations will give way to a more steady variation. A run for such parameter revealed that some variations still which can be explained is of a competition between the three directions. Figure (2b) shows a gradual decrease in the sequence length as alpha increases which confirms that the variation in the sequence length stabilizes with reduction in the competition between the directions (for this figure we used $P_d = 0.02$, $P_r = 0.02$, and $P_l = 0.25$ to simulate a situation in which only the left direction is dominant). Figure (3) shows the variation of the walker with respect to the sequence and the directional probabilities.

Numerical Simulation: The folding simulation has a specific role and the main goal is to mimic the structure formation process of a protein. It is more of the structural changes rather than the final result. In our lattice model we used Markov chain Monte Carlo to simulate the folding.

We compute the energy $E(\zeta, \sigma)$ from the following procedures

1. For (i=1 to maximum step) do
2. If (step not equal to 1) then
3. Call random number ($\Re$) [0,1]
4. If $\Re < W_1$ then
5. Row=row+1, the walker goes down
6. Else, if $\Re < \{W_1 + W_2\}$, then
7. Row=row-1, the walker goes up
8. Else, if $\Re < \{W_1 + W_2 + W_3\}$ then
9. Column=column+1, the walker goes right
10. Else, column =column-1, the walker goes to the left

A periodic boundary condition is applied: if row $r = L+1$ then set $r = 1$, and if $r = 0$ then set $r = L$. The same thing applies to the column.

Starting with a short chain (N=16) (*PHPPPPHPPHHHPPHP*) on a square lattice in which the non-degenerate ground state has been designed by (Seno et al., 1996; Irback et al., 1998) on the square lattice, this chain has both compact and non-compact 802075 different conformations for (Abkevich et al., 1995) which there is 456 good conformations and 1539 sequences out of all 2^N possible sequences. Other sequences considered are N = 24, 25 and 48 from (Unger and Moult, 1993; Toma and Toma, 1996). A conformation (ζ) is good if there is at least one sequence that has ζ as its nondegenerate ground state (Seno et al., 1996)

RESULTS AND DISCUSSION

We have introduced a new 2D square HP lattice model for the Monte Carlo simulation of protein conformations and demostrated the effect of amino-acid gyrations on the formation of the native state in proteins. The gyration is modelled as directional probability and was shown to play an important role in the formation of 3D structure of protein on which the function of a protein depends.

Figure 1 shows the variation of the probabilities on the sequence length. Figures 5 and 7 give the results of more compact structures where the native structure can be obtained, while figure 6, gives the results of non-compact structures, an aberration from the native state as a result of the intermediate state often known as "misfolding" caused by aggregation.

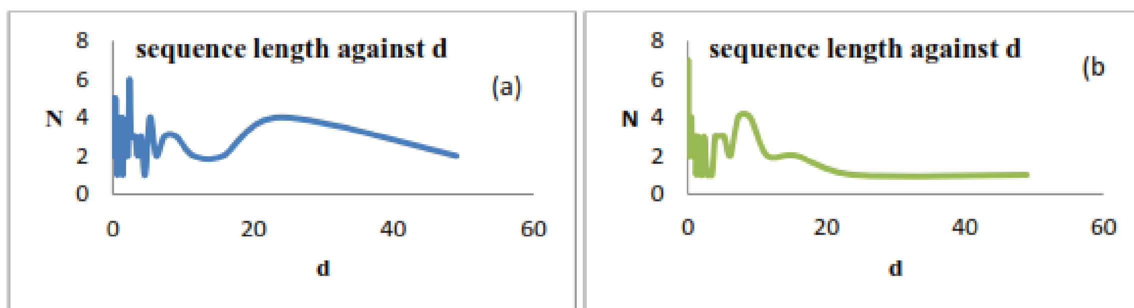

Figure 1
The plot of sequence length (N) against d when the realization is (a) 5 and (b) 10

Figure 2
The plot of sequence length (N) against (a) alpha (b) beta

Figure 4
Example of a conformation with unique ground energy state for N = 16 (adapted from Seno et al. 1996).

Figure 3
The plot of sequence length (N) against alpha

Figure 5
Examples of conformations for N=16 when the probabilities of the self-avoiding walker are equal; where the blue and the red atoms are polar and hydrophobic residues respectively. The plus and minus signs are the starting and ending point respectively. The walker tends to move equally

(a)

0	0	-1
0	0	1
0	0	2
0	0	3
0	0	4
0	0	5
0	7	6
0	8	0
10	9	0
11	0	0
12	0	0
13	14	0
0	15	0

(b)

0	0	15
0	0	14
0	0	13
0	0	12
0	0	11
0	0	10
0	0	9
0	0	8
0	0	7
0	5	6
0	4	0
2	3	0
1	0	0
-1	0	0

(c)

0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1	1	2	3	4	0	0	0	0	0	12	13	14	15
0	0	0	0	5	6	7	8	9	10	11	0	0	0

Figure 6
Examples of structures for N =16 where holes are the unoccupied site and -1 is the starting point (a) when the probability of down self-avoiding walker is greater than others, The walker tends to move downwards (b) when the probability of the up self-avoiding walker is greater than others, the walker tends to move upwards (c) when the probability of the right self-avoiding walker is greater than others, the walker moves to the right.

Table 1:
Experimental results for MGMB sequences where the probabilities are equal

Instances	N	H-P Sequence of amino acids	E^a	E^b
1	16	$PHP^4HP^2H^3P^2HP$	-4	-4
2	24	$H^2P^2HP^2HP^2HP^2HP^2HP^2HP^2H^2$	-9	-9
3	25	$P^2HP^2H^2P^4H^2P^4H^2P^4H^2$	-8	-8
4	36	$P^3H^2P^2H^2P^5H^7P^2H^2P^4H^2P^2HP^2$	-14	-14

H is hydrophobic and P is polar; E is the energy value; N is the sequence length
[a] the putative energy value; [b] the energy value obtained by MGMB

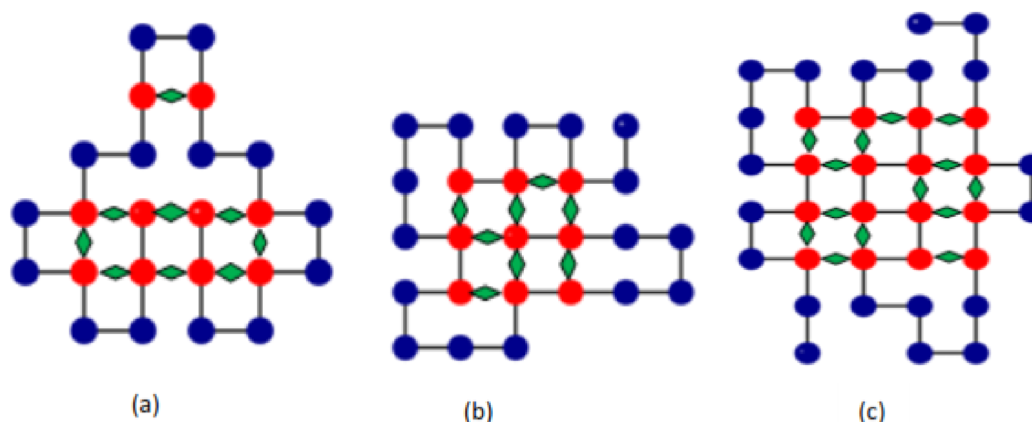

Figure 7
(color online) The optimal structures for (a) N= 24 (b) N = 25 and (c) N = 36 when the probabilities of the self avoiding walker are equal; where red atoms are hydrophobic, blues are polar, and the green diamonds are the non-bounded contacts between the hydrophobic residues. The plus and minus signs are the starting and the ending points. The walker tends to move equally to form compact conformations.

REFERENCES

Abdel-Naim, A. B, Abdel-Wahab, M. H, Attia, F. F. (1999). "Protective effects of vitamin E and probucol against Gentamicin nephrotoxicity in rats." Pharmacol Res., 40 (2): 183-187.

Allain C.C et al., (1974). Journal of Clinical Chemistry, 20/4, page 470-475.

Allan C, Deacon et Peter, Dawson J.G (1979). Journal of Clinical Chemistry, 25/6, page 976-984.

Baliga, R.; Ueda, N.; Walker, P.D, Shah, S.V. (1997). Oxidant mechanisms in toxic acute renal failure American Journal of Kidney Disease., 29:465-477.

Bishayee A. and Chanteerjee M. (1994). Protective effect of *Mikania cordata* root extract against physical and chemical factors induced gastric erosions in experimental animals. Planta Medica; 60:110-113.

Chang S.T, Buswell J.A. (1999). *Ganoderma lucidum* P. Karst (Aphylloromycetideae) – a mushrooms medicinal mushroom. Int.J.Med .Mushrooms, 1(2): 139-146.

Coles H.S.R, Burne J.F, Raff M.C (1993). Large scale normal cell death in the developing rat kidney and its reduction by epidermal growth factor Development. 118:777–784.

Conaldi P.G, Biancone L, Bottelli A, Wade-Evans A, Racusen L.C, Boccellino M, Toniolo A. (1998). HIV-1 kills renal tubular epithelial cells in vitro by triggering an apoptotic pathway involving caspase activation and Fas upregulation. Journal of Clinical Investigation. 102:2041–2049.

Davis M A, Ryan D H. (1998). Apoptosis in the kidney. Toxicological Pathology; 26:810–825.

Garetz, S.L., Schacht, J. (1996). "Ototoxicty of mice and men" In Handbook of auditory research, ed. By R.R. Fay and A.N. Popper, Vol. VII: Clinical aspect of hearing, ed. By T.R. Van De Water, A. N. Popper and R.R. Fay, PP. 116-154, Springer New York

Gobe G.C, Axelsen R.A. (1987). Genesis of renal tubular atrophy in experimental hydronephrosis in the rat. Role of apoptosis Lab Investigation. 56:273–282.

Goljan, Edward F. (2011). *Rapid Review Pathology 3rd ed*. Philadelphia, PA: Elsevier. Page 241.

Heibashy, M. I. A. and Abdel Moneim, A. E. (1999). Kidney and liver function tests after late Dimethyl sulfoxide (DMSO) administration in rats with gentamicin induced acute renal failure. J. Egypt. Ger. Soc. Zool., 30(A): 35-48.

Heibashy, M.I.A.; El-Nahla, A.M.; Ibrahim, A.I. and Saleh, Sh.Y.A. (2009). Comparative study between dimethyl sulfoxide (DMSO), allopurinol and urate oxidase administration in nephrotoxic rats induced with gentamicin. 43rd Annual Veterinary Medical Symposium, College of Veterinary Medicine Nursing and Allied Health, Tuskegee University, Alabama, USA.

Hendriks J.G.E, van Horn J.R, van der Mei H.C, Busscher, H.J. (2004). "Backgrounds of antibiotic-loaded bone cement and prosthesis-related infection". *Biomaterials***25** (3): 545–556

Hoitsma A.J, Wetzels J.F and Koene R.A (1991). Drug induced nephrotoxicity. Aetiology, clinical features and management, Drug Safety, 1991; 6 (2): 131-147.

Koh J. H, Yu K. W, Suh H. J, Choi Y. M, Ahn T. S. (2002). Activation of macrophages and the intestinal immune

system by an orally administered decoction from cultured mycelia of Cordyceps sinensis. BiosciBiotechnol Biochem. 66:407–11.

Padmini M.P and Kumar J.U (2012). A histological study on gentamicin induced nephrotoxicity in experimental albino rets, Journal of Dental medicine; 1(1): 14-17

Paller MS. (1990). Drug induced nephropathies Med Clin North America, 74 (4):909-917.

Porter G.A, Bennett W.M. (1981). Nephrotoxic acute renal failure due to common drugs American journal of Physiology.

Searchy R.L., Reardon J.E., Foreman J. A, (1967). American Journal of Medical Technology, 33, 15-20.

Shirwaikar A, Malini S, Kumari S.C. (2003). Protective effect of Pongamia pinnata flowers against cisplatin and gentamicin induced nephrotoxicity in rats Indian Journal of Experimental Biology; 41(1):58-62.

Tietz N.W (2006). Clinical guide to laboratory test, 4th edition.,page 244-249.

Tietz N.W, Burtis C.A, Ashwood E.R, Saunders W.B (1999). Text book of clinical chemistry, 3rd edition. Page 1239-1241.

Veniamin M.P, Vakirtzi- Lemonias C, (1970). Chemical basis of the carbamidodiacetyl micromethod for estimation of urea, citrulline, and carbamyl derivatives. Clinical chemistry; 16 (1):3-6.

Vijay Kumar .K, Naidu M.U.R, Anwar Shifow A., (2000). Ratnaker Probucol protects against gentamicin induced nephrotoxicity in rats. Indian Journal of Pharmacology; 32 :108-113.

Yoshiyama, Teruaki Kobayashi and Fumiya Tomonaga. (1992). Chronotoxical Study of Gentamicin Induced Nephrotoxicity in Rats Journal of Antibiotics.

Young D.S., (1990). Effect of Drugs on Clinical laboratory tests, 4th edition. Page 3-599 to 3-609.