A Hybrid Genetic Algorithm for 2D FCC Hydrophobic-Hydrophilic Lattice Model to Predict Protein Folding

Md Tamjidul Hoque, Madhu Chetty, and Laurence S. Dooley

Gippsland School of Information Technology
Monash University, Churchill VIC 3842, Australia
{Tamjidul.Hoque, Madhu.Chetty,
Laurence.Dooley}@infotech.monash.edu.au

Abstract. This paper presents a Hybrid Genetic Algorithm (HGA) for the protein folding prediction (PFP) applications using the 2D face-centred-cube (FCC) Hydrophobic-Hydrophilic (HP) lattice model. This approach enhances the optimal core formation concept and develops effective and efficient strategies to implement generalized short pull moves to embed highly probable short motifs or building blocks and hence forms the hybridized GA for FCC model. Building blocks containing Hydrophobic (H) – Hydrophilic (P or Polar) covalent bonds are utilized such a way as to help form a core that maximizes the |fitness|. The HGA helps overcome the ineffective crossover and mutation operations that traditionally lead to the stuck condition, especially when the core becomes compact. PFP has been strategically translated into a multi-objective optimization problem and implemented using a swing function, with the HGA providing improved performance in the 2D FCC model compared with the Simple GA.

1 Introduction

Protein is a three dimensionally folded molecule composed of amino acids [1] linked together (called primary structure) in a particular order specified by DNA sequence of a gene. They are essential for functioning of the living cells as well as for providing structure. *Protein folding prediction* (PFP) is a problem of determining the native state of a protein from its primary structure and is of great importance [2] because three dimensionally folded structures determine the biological function and hence prove extremely useful in down streaming applications like drug design [3].

To investigate the underlying principles of protein folding, lattice protein models introduced by Dill [4] are highly regarded tools [5]. Protein conformation as a *self-avoiding walk* in the lattice model has been proven to be *NP-complete* [6] [7] so therefore a deterministic technique to folding prediction is not possible. Hence, a nondeterministic approach including robust strategies that can extract minimal energy conformations efficiently out of these models is of great importance. This is a very challenging task as there exist an inordinate number of possible conformations even for short amino acid sequences [8] [9].

So far, the most successful approach to the hard optimization problem like *PFP* is based on hybrid evolutionary approach [10]. A lattice model avoids the continuous conformational space that simplifies many of the required calculations. Among HP lattice models, 2D square and 3D cube models have been used mostly by the research

A. Sattar and B.H. Kang (Eds.): AI 2006, LNAI 4304, pp. 867 – 876, 2006.

community [10-13] for the sake of simplicity but the parity problem within this model complicated the design approach without much benefit [12-13]. It was shown [17] that triangular model is parity problem free, that is an odd indexed amino acid or residue in the sequence position can be the neighbor of both odd and even indexed residues in the sequence and vice versa. Later, a full proof of the famous *Kepler Conjecture* [15] [16] problem was completed which implies that the *FCC* is the densest sphere-packing model, where a residue can have 12 neighbors in a 3D space and 6 neighbors forming a hexagon in 2D. Clearly, the most compact hydrophobic core (*H-Core*) [17] can be represented by *FCC* model [5].

Although triangular and *FCC* models have similar properties, it is shown Section 3 that the optimal *H-Core* is hexagonal rather than triangular. Therefore, while search is carried out using GA, the on going sub-optimal conformation (i.e. chromosome of the population) is guided towards the formation of a hexagonal *H-Core*. The highly likely motifs are remapped within the conformation based on a dynamic nuclei defined as *H-Core Centre* (HCC) [12-13, 18].

The pull move has been shown to be a very effective operator in [12-13, 19] for square and cube HP lattice. Here, in 2D *FCC*, we show that the pull move is easily implemented and that makes the deevloping conformation less destructive while highly likely motifs are mapped. The only difficulty is the handling of floating point operation in the *FCC* model, which can be overcome by an effective programming technique. As our main focus is to show the effectiveness of our strategies, the focus is confined within the 2D *FCC* model rather than the 3D model which in turn permits easy explanation. Further, the properties of *FCC* model relate well [20-21] to protein conformation amongst the known lattice models.

Several other outstanding nondeterministic approaches such as a number of versions *of Monte Carlo* (MC), *Simulated Annealing* (SA), *Tabu Search with GA* (GTB) and *Ant Colony Optimization* [23] are available for square and cube models. However, for the *FCC* model the GA is preferred as it consistently outperforms other approaches [10-13, 22], though it is noted that any approach including GA faces difficulties in the hard optimization problem like *PFP*. These will be investigated and overcome by using hybridization technqiues.

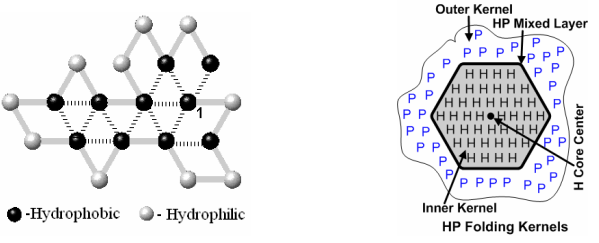
2 Two Dimensional FCC HP Lattice Model

Based on the observation that the hydrophobic forces dominate during protein folding, the HP model has been introduced by Dill [4] with amino acids being represented as a reduced set of *H* (Hydrophobic or Non-Polar) and *P* (Hydrophilic or Polar) only. The protein conformations of the sequence are placed as a *self-avoiding walk* (SAW) on a 2D hexagonal pattern in the *FCC* [20] model. The energy of a given conformation is defined as a number of topological neighboring (TN) contacts between those Hs, which are not sequential with respect to the sequence. The *PFP* can be formally defined as follows.

Given amino-acid sequence, $s = s_1, s_2, s_s, \cdots, s_m$, a conformation c needs to be formed where, $c^* \in C(s)$, energy $E^* = E(C) = \min\{E(c) \mid c \in C\}$ [23].

Here, m = total amino acids in the sequence and $C(s)$ is the set of all valid (i.e. SAW) conformations of s. If the number of TNs in a conformation c is q then the

value of $E(c)$ is defined as $E(c) = -q$. In a 2D *FCC* HP model (Fig. 1, (a)), a non-terminal and a terminal residue both having 6 neighbors can have a maximum of 4 TNs and 5 TNs respectively.



(a) (b)

Fig. 1. (a) Conformation in 2D FCC HP model shown by solid line. Dotted line indicates TN. Fitness = -(TN Count) = -15. (b) 2D metaphoric HP folding kernels for the FCC model.

It is well known [17] that the *H*s form the protein core so freeing up energy while the *P*s, have an affinity with the solvent and so tend to remain in the outer surface. This paper visualizes the folded protein through the 2D *FCC* HP model as a three-layered kernel (Fig. 1(b)). The inner kernel called the *H-Core* [17] [24], is compact and mainly formed of *H*s while the outer kernel consists mostly of *P*s. The *H-Core Centre* (*HCC*) is defined in Section 4.2 as the average of the coordinates of *H*s. The composite thin layer between the two kernels consists of those *H*s that are covalent bonded with *P*s, which for the purpose of this paper is referred to as the *HP mixed layer*.

3 Optimal Shape of 2D H-Core in FCC Model

In this section, a proof is developed induction basis, for the optimum shape of the *H-Core* for FCC model compared with triangular model, rejecting other possible shapes for obvious reasons [12-13]. The sequence for the sake of this proof is assumed to be a segment of Hs only and it is a variation of that presented in [17, 12-13]. Table-1 shows that the *H-Core* tends to form a hexagonal rather than triangular shape and has maximal |fitness| whereas both approaches can have the same number of neighbors so that an optimal *H-Core* shape is hexagonal. The positioning of the *H*s inside the core (assuming a hexagonal boundary) can be categorized as *H* at the corner, *H* on the edge and *H* inside the interior which will respectively have 3, 4 and 6 neighbouring sides each.

Further, our concern is to compute the probability of an *H* to be appearing at a corner and on an edge. It is to be noted that for a hexagon, the number of residues at corner remain fixed but the number of residues on edge increase with the increasing size of the hexagon or the increasing number of *H*s. The total residues (T_t) within a hexagon in relation to 6 residues at corners and 6*t* residues on the edges can be expressed based on the recurrent equation (1), where $t = 0, 1, 2, 3, \ldots$ and $T_{-1} = 1$.

$$T_t = 6 + 6t + T_{t-1} \tag{1}$$

Therefore, the probability of an *H* being at corner is given by equation (2),

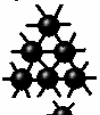
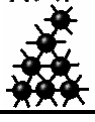
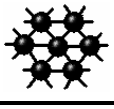
$$\Pr_{corner(t)} = \left(6/\left(T_t - T_{t-1}\right)\right) \tag{2}$$

and the probability of an *H* being on the edge is given by equation (3),

$$\Pr_{edge(t)} = 1 - \left(6/\left(T_t - T_{t-1}\right)\right) \tag{3}$$

where the actual number of total *H*s is n_H, such that $T_{t-1} < n_H \leq T_t$.

Table 1. *H-Core* conformations comparison between triangular versus hexagonal, while *m* indicates the number of residues and losses indicate non-bonded neighboring positions

m	Triangular Shape		Hexagonal Shape	
	Conformation	Losses	Conformation	Losses
4		16		14
5		20		16
6		18		18
7		22		18

4 Highly Likely Sub-conformation for HP Mixed Layer

To form the cavity of *H-Core*, it is intuitive to think of placing the *P* of a *-HP-* segment on the opposite side of *H* with respect to the current *HCC*, while searching for the desired conformation. However, with such a straightforward placement, the cavity would tend to form a circular shape, which is not the desired hexagonal form. To address these problems, motif or sub-conformation that is highly probable to a subsequence (defined in Fig. 2) is forced to remap. The main idea is to form immediate TN and place *P* as far away as possible from *HCC* while concomitantly placing *H* as near as possible to *HCC*. To implement the same strategies in *FCC* HP model, various *moves* [12-13] are further simplified and merged which is the benefit of the *FCC* model over the HP-square or cube model.

Furthermore, the enforced placement of sub-conformations is not easy because their location in the lattice model is discrete and it can destruct already achieved sub-optimal sub-conformation. To address these problems, two broad categories of subsequences are defined; gS_H and gS_P, where $g \in \mathrm{N}$ (N is natural number). These two categories completely cover the *HP mixed layer* including outer kernel. Let S_H

and S_P represent segments of *H* and *P* respectively. A segment refers to a contiguous string of length g, e.g. $2S_H$ means *-PHHP-*, i.e. $g = 2$ with the two boundary residues being of the opposite type. g is divided into even g_e and odd g_o numbers. For $1S_p$, $1S_H$, $2S_P$ and $2S_H$, there are few possible sub-conformations, so only highly potential sub-conformations (shown in Fig. 2) are chosen, based on embedded TN and core formation concepts. Collectively they are called *H-Core Boundary Builder Segments* (HBBS) and are mapped to potential sub-conformations which are referred to as the *H-Core Boundary Builder sub-Conformation* (HBBC). *HBBC* forms part of a corner (especially when $g = 1$ and through the composition with other group having $g = 2$) and an edge (especially when $g = 2$ and with the composition of the former group) of the *H-Core* boundary. The selection for mapping *HBBC* into *HBBS* can be probabilistically applied using (2) and (3). Due to the absence of the parity problem in the FCC model, we were able to simplify the desired sub-conformations and reduce their numbers compared to [12-13].

(a) –HPH- (b) –PHP- (c) –PHHP- (d) –HPPH-

Fig. 2. Potential motifs sub-conformation for (a) $1S_p$ (b) $1S_H$ (c) $2S_P$ (d) $2S_H$. ●, ○ and ■ indicate an *H*, a *P* and the approximate position of HCC, respectively. Dotted line indicates alternate connection.

4.1 Probabilistic Constrained Fitness

While searching for an optimum conformation, if a sub-conformation corresponding to a particular sub-sequence exists in the *HP mixed layer* for a developing conformation, it is rewarded, otherwise penalized. This measure of fitness is referred to as the *Probabilistic Constrained Fitness* (PCF), so if any member of a *HBBC* corresponds to the related sub-sequence and the *H*s are nearer to HCC than the *P*s, then *PCF* will be decreased by 2 as reward, otherwise it will be penalized by an increase of 1 for a non-desired sub-conformation and 2 for a proper shape but having opposite of the desired directions (i.e. the position of *P*s are closer to *HCC* than *H*s).

4.2 Definitions for the Implementation of the Sub-conformations

To implement or remap *HBBC*, a number of terms need to be defined.

(i) *HCC* is the *H-Core centre*, calculated as the arithmetic mean of the coordinates of all *H*s as in (4). Before enforcing to map a sub-conformation, the *HCC* (i.e. x_{HCC}, y_{HCC}) is updated to place *H* near *HCC* and *P* as far as possible from *HCC*.

$$x_{HCC} = \frac{1}{n_H}\sum_{i=1}^{n_H} x_i \quad \text{and} \quad y_{HCC} = \frac{1}{n_H}\sum_{i=1}^{n_H} y_i \tag{4}$$

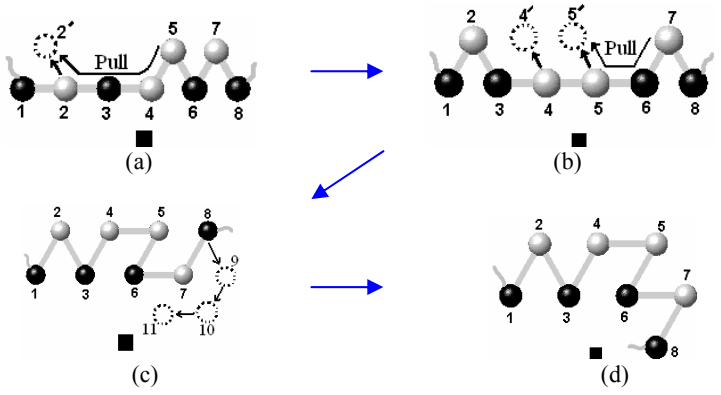


Fig. 3. The subsequence -123- in (a) need to remap to sub-conformation $2S_p$. If the position $2'$ is free then 2 can be placed at $2'$ and a pull (indicated in (a)) applied towards the higher indexed end. The pull moves 3 to 2, 4 to 3 and 5 to 4 and then finds a valid conformation without pulling further leaving (b). The |fitness| of (b) is increased by 1. In (b) assume, $4'$ and $5'$ are free positions and the segment 3 to 6 can be recognized as $2S_H$. To enforce a mapping to highly probable sub-conformation, 4 and 5 can be shifted to $4'$ and $5'$ respectively applying a pull move as indicated in (b) with the in (c). Thus 8 in (c) propagates through position 9, 10, 11 to give (d) which has increased the |fitness| by 2 with respect to (a).

(ii) *Pull move*: This has been shown to be very effective in [19] especially for enforcing any sub-conformation [12-13]. There are basically two-folded benefit: a) implementing the sub-conformations or motifs b) less distortion of other parts due to pulling required which may be in an optimal position as demonstrated in Fig. 3. For the *FCC* model we used the redefined pull move for the same purpose. As the parity problem is absent in *FCC* model, the pull move does not need to be moved diagonally to start as an ordinary pull to the next neighbor performs the same. This is because in *FCC* without the parity problem and with more neighbors, it is very likely to get a valid conformation, without need to propagating the pull often up to the terminal residue.

5 Implementation and Experiments

Although the additional constraint (*PCF*) formulates the multi-objectivities, the implementation is such that it ultimately maximizes the goal of original fitness $|F|$. The search process is mainly divided into two alternative phases namely, Phase 1 in which *F* dominates *PCF* and starts building the core. In the alternate Phase 2, *PCF* dominates which takes care of the proper formation of the *HP* mixed layer. Further, the enforcement to *HBBC* is performed in phase 2 since *PCF* helps the change sustain and stabilize. The *HBBC* implantation is done only if they are not found according to the highly likely sub-conformations for the corresponding sub-sequences. This action may reduce the already achieved fitness *F*, but it is expected that it will help reformulate a proper cavity that will maximize the *H* bonding inside core, which while

shifting to the favorable phase would maximize the fitness $|F|$. As the phases alternate throughout the search process, the impact becomes such that *F* and *PCF* come up with common goal that is highly likely to be optimal. The total or combined fitness is defined as,

$$TF = \alpha(t) * F + \beta(t) * PCF \tag{5}$$

where *t* is t^{th} generation while search is carried out by GA. To alter the weight of α and β to dominate *F* and *PCF* over each other, a swing function (equation (6)) is used.

$$\delta(t) = A\,(1 + \cos\omega_m t)\cos\omega_0 t \tag{6}$$

where $\omega_m << \omega_0$, *t* = number of generation. The assignment of α and β is as,

$$\text{Phase 1: } \alpha(t) = \delta(t), \beta(t) = 1, \text{ when } \delta(t) > 0 \tag{7}$$

$$\text{Phase 2: } \alpha(t) = 1, \beta(t) = -\delta(t), \text{ when } \delta(t) < 0 \tag{8}$$

$$\text{Transient Phase: } \alpha(t) = 1, \beta(t) = 1, \text{ when } \delta(t) = 0 \tag{9}$$

For the typical value of $\delta(t)$ parameters are set as follows: amplitude *A*=30, ω_m = 0.004 and ω_0 =0.05. The value of *A* is selected as, $2A \geq \max\left(|F|, |PCF|\right)$ where the upper limit of *F* is set using (10), which has been extended from [14].

$$F = -\left(2n_H + n_T\right) \tag{10}$$

Here, n_H is the total number of hydrophobic residues in a sequence and n_T is the number of hydrophobic residues at the terminal positions and $0 \leq n_T \leq 2$. Note, the minimum value of both $|\alpha(t)|$ and $|\beta(t)|$ equal 1 are maintained and never set to zero in (7), (8) and (9), so preserving the sub-conformation or schema developed in the alternate phase, possessing good features.

The search procedure is given in Algorithm-I. A simple GA which is hybridized with population size [27] of 200 is chosen for all sequences. The elite rate = 0.10, p_c = 0.85, p_m = 0.5 and a single point mutation by pivot rotation [10-11] is applied.

Table 2. Predictability of SGA versus HGA maximum |fitness| from 10 runs is shown

Sequence	SGA	HGA	Conformation (HGA)
HPHPPHHPHPPHPHHPPHPH	-11	-15	Fig.4 (a)
HHPPHPPHPPHPPHPPHPPHPPHH	-10	-13	Fig.4 (b)
PPHPPHHPPPPHHPPPPHHPPPPHH	-10	-10	Fig.4 (c)
P3(H2P2)2P3H7P2H2P4H2P2HP2	-16	-19	Fig.4 (d)
P2(HP2H)2HP510HP5(H2P2)2HP2H5	-26	-32	Fig.4 (e)
H(HP)4H4PHP3HP3HP3HP4HP3HP3HPH4(PH)4H	-21	-23	Fig.4 (f)
P2H3PH8P3H10PHP3H12P4H6PH(HP)2	-40	-46	Fig.4 (g)
H12(PH)2((P2H2)2P2H)3PHPH12	-33	-46	Fig.4 (h)
HHHPPHPHPHPPHPHPHPPH	-11	-14	Fig.4 (i)

The implementation of crossover and mutation is same as in [10-13] but without any special treatment (e.g. cooling). Roulette wheel is used for selection procedure.

Simulations are carried out for benchmark 2D problems [25]. For each of the sequences, standard Simple Genetic Algorithm (SGA) [26] and HGA run parallelly together, and stop when all of them become non-progressive. Therefore they run for equal amount of time. As the parameters of the developed HGA were not tuned, we avoid comparison based on the number of iterations required. The goal is to compare the predictability of the developed HGA approach. Results are shown in Table 2 with the conformations corresponding to the maximum |fitness| achieved, for future comparison. HGA outperformed predictabilities of SGA significantly.

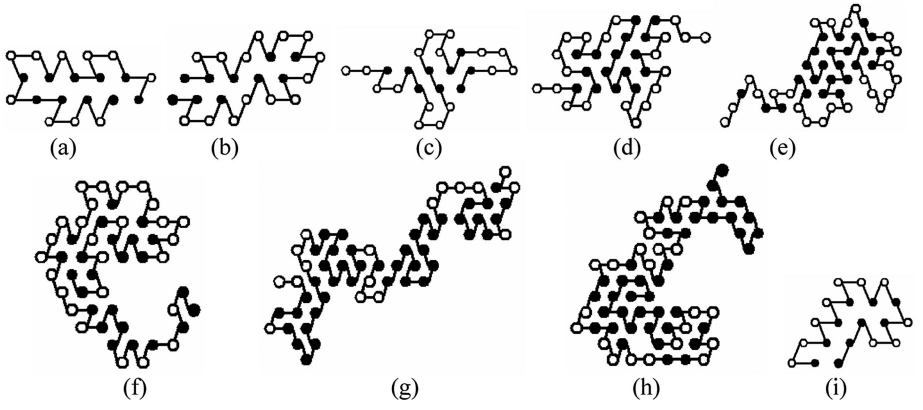


Fig. 4. (a) to (i) correspond to the conformation with maximum |fitness| achieved using HGA as indicated in Table-2

Algorithm-I. HGA for PFP Using 2D FCC Model

Input: Sequence *S*,
Output: Fitness of the optimum 2D FCC conformation.
COMPUTE ***PCF;*** COMPUTE ***A*** (amplitude)
t=0, F=0 /* Generation count and fitness initialization */
Fillup the population with random (valid) conformation possible for *S*.
While $F \neq Higher\_Target\_Value\_of\_F$ THEN
{ t = t + 1, COMPUTE $\delta(t)$, $\alpha(t)$, $\beta(t)$, *TF*
CROSSOVER and then MUTATION
IF $\delta(t) < 0$ THEN
{ FOR i =1 to *population_size* DO
Check $chromosome_i$ for any miss mapping of HBBC
IF miss-mapping true then
{ Re-map the sub-sequence to corresponding HBBC using move-sets. }}
COMPUTE *TF*
Sort, Keep Elite, Reduce Twins.
F ← Best fitness found from the population.
} END.

6 Discussion and Conclusions

There are two major drawbacks with SGA for *PFP*. GA computation is based on schema theorem, which states [10] that short, flexible schemata with above average performance will receive fast survival chance in the subsequent generations and schemata with below-average performance will decay nonlinearly fast too. So, one obstacle using SGA is that similarity within population grows which leads to stall or stuck condition, since crossover will most likely occur between twins. Those which are mutated are likely to be heavily dissimilar and would therefore be rejected by the selection process. To address this problem, twin removal was applied and elitism was used only to keep the found best. Secondly, as the optimum conformation is relatively compact, crossover and mutation confront more increasing collision and produce invalid conformation. Our specific implantation procedure of *HBBCs* moves the compact conformation without collision and the introduced move operator causes less destruction to the already gained fitness. The move creates probable reformation of the *H-Core* cavity to maximize the H-sides inside the *H-Core*. Hence, this approach makes change of the non-progressive situation in such a way, that it enhances the chance of gaining highly fitted conformations.

The novel strategies using HGA has also been extended for 2D *FCC* HP. It is shown that beyond removing the parity problem, *FCC* further simplifies the pull moves and reduces the need of higher number sub-conformations for remapping. As there is lack of previous works on 2D *FCC*, we compared the proposed HGA with the standard SGA, which shows significant improvement over predictability. Regarding future scope, parameter of HGA and the swing function are to be investigated further for optimization. It can also be extended for 3D *FCC* and real *PFP* as well. The overall framework we developed is robust enough and removes the causes of failure.

References

1. Allen, et al.: Blue Gene: A vision for protein science using a petaflop supercomputer. IBM System Journal (2001), Vol 40, No 2
2. Pietzsch,, J.: The importance of protein folding, http://www.nature.com/horizon/proteinfolding/background/importance.html.
3. Petit-Zeman, S.: Treating protein folding diseases, http://www.nature.com/horizon/proteinfolding/background/treating.html.
4. Dill, K.A.: Theory for the Folding and Stability of Globular Proteins. Biochem.,1985, 24: 1501
5. Backofen, R. and Will, S.: A constraint-based approach to fast and exact structure prediction in three-dimensional protein models. Journal of Constraints, 11 (1) pp. 5-30, Jan. 2006.
6. Crescenzi, P. and et al.: On the complexity of protein folding (extended abstract). ACM, International conference on Computational molecular biology, 1998, pp.597-603.
7. Berger, B. and Leighton, T.: Protein Folding in the Hydrophobic-Hydrophilic (HP) model is NP-Complete. ACM, International conference on Computational molecular biology (1998)
8. Schiemann, R., Bachmann, M. and Janke, W.: Exact Enumeration of Three – Dimensional Lattice Proteins. Elsevier Science, 2005.

9. Guttmann, A.J.: Self-avoiding walks in constrained and random geometries. In "Statistics of Linear Polymers in Disordered Media" ed. B. K. Chakrabarti, Elsevier, 2005, pp:59-101.
10. Unger, R. and Moult, J.: On the Applicability of Genetic Algorithms to Protein Folding. IEEE, 1993, pp. 715-725.
11. Unger, R. and Moult J.: Genetic Algorithm for Protein Folding Simulations, JMB, 1993.
12. Hoque, M. T., Chetty, M. and Dooley, L. S.:A New Guided Genetic Algorithm for 2D Hydrophobic-Hydrophilic Model to Predict Protein Folding. 2005 IEEE CEC, pp. 259-266.
13. Hoque, M. T., Chetty, M. and Dooley, L. S.:A Guided Genetic Algorithm for Protein Folding Prediction Using 3D Hydrophobic-Hydrophilic Model, 2006 IEEE WCCI.
14. Agarwala, R and et. al.: Local Rules for Protein Folding on a Triangular Lattice and Generalized Hydrophobicity in the HP Model. J. of Computational Biology, 4(2) pp. 275 – 296, 1997.
15. Sloane, N. J. A.: Kepler's Conjecture Confirmed. Nature 395, 435-436, 1998.
16. Hales, T. C.: A proof of the Kepler conjecture. Annals of Mathematics, 162, no. 3 (2005), 1065–1185.
17. Dill, K.A., Fiebig, K.M., Chan H.S.: Cooperativity in Protein-Folding Kinetics. Proceedings of the National Academy of Sciences USA, Biophysics (1993) Vol 90, pp.1942-1946
18. Shmygelska, A.:Search for Folding Nuclei in Native Protein Structures. Bioinformatics, 2005, 21: i394 - i402.
19. Lesh, N., Mitzenmacher, M. and Whitesides, S.: A Complete and Effective Move Set for Simplified Protein Folding. RECOMB, 2003, Berlin.
20. Backofen, R. and Will, S.: Optimally Compact Finite Sphere Packings -- Hydrophobic Cores in the FCC. CPM2001, LNCS, pp. 257-272.
21. Koehl, P. and Levitt, M.:A brighter future for protein structure prediction, Nature Structural Biology 6, 108 – 111, 1999.
22. Unger R. and Moult, J.: Genetic Algorithm for 3D Protein Folding Simulations. 5th International Conference on Genetic Algorithms, 1993, pp. 581-588.
23. Shmygelska, A. and Hoos, H.H.:An ant colony optimization algorithm for the 2D and 3D hydrophobic polar protein folding problem. BMC Bioinformatics 2005, 6(30).
24. Flebig and Dill.: Protein Core Assembly Processes. J. Chem. Phys., 1993, 98(4), 3475-87.
25. Hart, W. and Istrail, S.: HP Benchmarks, http://www.cs.sandia.gov/tech_reports/compbio/tortilla-hp-benchmarks.html
26. Vose, M. D.:The Simple Genetic Algorithm. 1999, The MIT Press.
27. Digalakis, J. G. and Margaritis, K. G.: An experimental Study of Benchmarking Functions for Genetic Algorithms, Intern. J. Computer Math., 2002, 79(4), 403-416.