Improved Bees Algorithm for Protein Structure Prediction Using AB Off-Lattice Model

Nanda Dulal Jana, Jaya Sil and Swagatam Das

Abstract Protein Structure Prediction (PSP) using sequence of amino acids is a multimodal optimization problem and belongs to *NP* hard class. Researchers and scientists put their efforts to design efficient computational intelligent algorithm for solving this kind of problem. Bees Algorithm (BA) is a swarm intelligence based algorithm inspired by the foraging behaviour of honey bees colony, already exhibits its potential ability for solving optimization problems. However, it may produce premature convergence when solving PSP like problems. To prevent this situation, Adaptive Polynomial Mutation based Bees Algorithm (APM-BA) has been proposed in this paper for predicting protein structure in 2D AB off-lattice model. In this strategy, each of best scout bees are mutated with adaptive polynomial mutation technique when their performances are no more improve during execution phase. The experiments are conducted on artificial and real protein sequences and numerical results show that the proposed algorithm has strong ability for solving PSP problem having minimum energy.

N.D. Jana (✉)
Department of IT, National Institute of Technology, Durgapur 713209, India
e-mail: nanda.jana@gmail.com

J. Sil
Department of CST, Indian Institute of Engineering Science & Technology,
Shibpur 711103, India
e-mail: js@cs.iiests.ac.in

S. Das
ECS Unit, Indian Statistical Institute, Kolkata 700108, India
e-mail: swagatam.das@isical.ac.in

1 Introduction

Proteins are the primary building blocks in all living organisms and represented by a sequence of 20 different amino acids. Biological functions of proteins are closely related with their structures which play an important role in drug design, disease prediction and many more [1, 2]. Therefore, protein structure prediction is an important research area in computational biology. Anfinsen's thermodynamic hypothesis [3] states that the native structure of a protein corresponds to the global minimum of the free energy surface of the protein. So, the protein structure prediction problem can be treated as a global optimization problem. This problem is of NP-hard [4] and its multimodality characteristics increase with the protein sequence length [5].

Experimental methods like X-ray crystallography and Nuclear Magnetic Resonance (NMR) are time consuming and expensive to predict the structure of proteins. Moreover, due to strict laboratory requirements and heavy operational burdens, it is not always feasible to determine the protein structure experimentally. Therefore, researchers focused on predicting protein structure from the given amino acid sequences by computational methods [6]. A successful study of computational methods in PSP reveals two facts. The first part is the consideration of physical model which corresponds to a potential energy function. The second part involves searching of global minimum of the potential energy function. In the literature, most of the physical models are grouped into two classes of residues: hydrophobic (non-polar or H) and hydrophilic (polar or P) instead of considering 20 different amino acids individually. The most widely used simplified model is HP lattice models [7]. An off-lattice model is a generalization of HP lattice model known as AB off-lattice model proposed by Stillinger et al. [8] where the hydrophobic and the hydrophilic residues are labelled by 'A' and 'B' respectively. Many computational intelligent algorithms have been applied on AB off-lattice model for finding global minimum energy to predict protein structure [9–18].

The Bees Algorithm (BA) [19, 20] is a swarm intelligence based algorithm inspired by the food foraging behaviour of honey bee colonies developed in 2005 by D. T. Pham. It performs a kind of exploitative neighbourhood search combined with random explorative search, successfully applied to many engineering optimization problems. However, it has the limitation of premature convergence due to lack of diversity in the search space when solving multimodal optimization problems. In this paper, to prevent premature convergence and improve the performance of BA, we have proposed, adaptive polynomial mutation based bees algorithm (APM-BA) for solving PSP problem in 2D AB off-lattice model. Adaptive polynomial mutation is applied on best scout bees which do not improve their visited site in a predefined limit, known 'trial' counter of inefficient search. As a result, there is a high chance to jump out from visited site to unvisited site and made exploration on the search space. Experiments are conducted on artificial and real protein sequences using 2D AB off-lattice model. The numerical results show that the proposed algorithm is suitable for solving PSP problem having minimum energy. Moreover, results are compared with other algorithms demonstrating efficiency of the proposed method.

The paper is organized as follows: In Sects. 2 and 3, basic principle of 2D AB off-lattice model and bees algorithm are describe, respectively. The details of APM-BA are presented in Sect. 4, followed by experimental setups and results in Sect. 5. Finally, the conclusion is drawn and future works are highlighted in the Sect. 6.

2 AB Off-Lattice Model

AB off-lattice model, known as toy model proposed by Stillinger et al. in 1993 [8] and widely used for predicting the structure of a protein sequence due to its simplicity. In this model, 20 amino acids are classified into hydrophobic and hydrophilic residues, labelled as 'A' and 'B' respectively. Two residues are linked by rigid unit-length bonds and the angle between two bonds can change freely in two dimensional Euclidean space. An n length protein sequence is represented by (n-2) bend angles $\theta_2, \theta_3, ..., \theta_{n-1}$ at each of the non-terminal residues. Each bend angles θ_i are in the range $-180°$ to $180°$ and $\theta_i = 0$ represents linearity in the successive bonds. The bend angel, $\theta_i \in [-180°, 0)$ and $\theta_i \in (0, 180°]$ represent rotation of amino acids in clockwise and counter clockwise direction respectively. A 2D off-lattice model of a protein sequence with length 9 shown in Fig. 1. The AB off-lattice model represents the intra-molecular potential energy for each molecule with backbone bend potentials V_1 and nonbonded interactions V_2. Amino acids along the backbone can be conveniently encoded by a set of bipolar variables ξ_i. If $\xi_i = 1$, the ith amino acid is A while $\xi_i = -1$, it is B. Hence, the total potential energy function Φ for any n length protein sequence is expressed using Eq. 1.

Fig. 1 2D off-lattice model of a protein sequence with length 9
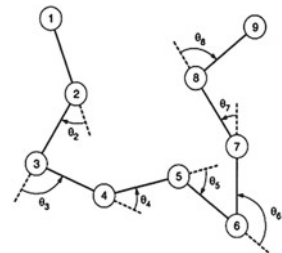
$$\Phi = \sum_{i=2}^{n-1} V_1(\theta_i) + \sum_{i=1}^{n-2} \sum_{j=i+2}^{n} V_2(r_{ij}, \xi_i, \xi_j) \tag{1}$$

Where V_1 is the bending potential, independent of protein sequence as defined by Eq. 2

$$V_1(\theta_i) = \frac{1}{4}(1 - cos\theta_i) \tag{2}$$

The nonbonded interactions V_2 have a species-dependent Lennard-Jones 12, 6 form, described in Eqs. 3 and 4 respectively.

$$V_2(r_{ij}, \xi_i, \xi_j) = 4[r_{ij}^{-12} - C(\xi_i, \xi_j)r_{ij}^{-6}] \tag{3}$$

$$C(\xi_i, \xi_j) = \frac{1}{8}(1 + \xi_i + \xi_j + 5\xi_i\xi_j) \tag{4}$$

Where r_{ij} denotes the distance between i^{th} and j^{th} residue of the chain. For an AA pair, $C(\xi_i, \xi_j) = 1$ regarded as strongly attracting for an AB or BA pair while $C(\xi_i, \xi_j) = -0.5$, regarded as weakly repelling and for a BB pair, $C(\xi_i, \xi_j) = 0.5$, regarded as weakly attracting. Our objective is to find the minimum value of Eq. 1, representing lowest free energy of the structure of a protein.

3 Bees Algorithm (BA)

Bees Algorithm [19] is a swarm intelligence based algorithm inspired by the foraging behaviour of honey bees used for finding global optimum solution for a given optimization problem. Scout bees i.e. candidate solutions are randomly generated in the search space and the quality of the visited locations depend on the fitness value. The generated solutions are ranked and other bees are recruited from the neighbourhood having the highest ranking locations on the search space. This algorithm locates the most promising solutions and selectively explores their neighbourhoods looking for the global minimum of the fitness function.

The population contains N number of scout bees which are randomly scattered with uniform probability across the search space. Therefore, the jth element of the ith solution X_i is expressed using Eq. 5.

$$X_i^j = X_{min}^j + \left(X_{max}^j - X_{min}^j\right) \times rand(0, 1), j = 1, 2, ..., D \tag{5}$$

Where X_{min}^j and X_{max}^j denote the lower and upper bound of the jth element and D denotes the dimension of any X_i. Each scout bees evaluates using the fitness function. After initialization of the scout bees, BA enters into a cycle which is composed of four phases [19]. Following phases are executed sequentially until stopping condition is met.

3.1 Waggle Dance

Say, N number of visited sites are ranked based on fitness information and B number of sites with highest fitness (i.e. minimum measure) are selected for local search. The local search is performed by other bees (foragers) that are directed to the neigh-

bourhood of the selected sites. Each scout bees that are returned from one of the B best sites performs the 'waggle dance' to recruit nest mates for local search. The scout bees visit the first E elite (top-rated) sites among the B sites by recruiting E_r bees for neighbourhood search. The remaining $(B - E)$ sites that visited by the scouts recruit $B_r \leq E_r$ bees for neighbourhood search.

3.2 Local Search

For each of the B selected sites, the recruited bees are randomly placed in a neighbourhood of the high fitness location marked by the scout bees. This neighbourhood is defined as an D-dimensional hyper box of sides $a_1, a_2, ..., a_D$, centred on the scout bee. For each neighbourhood, the fitness is evaluated by the recruited bees. If one of the recruited bees lands in a position of higher fitness than the scout bee then the recruited bee is treated as the new scout bee. At the end of the local search, only the fittest bee is retained. The fittest solution visited so-far is therefore, considered as a representative of the whole neighbourhood.

3.3 Global Search

In the global search phase, $(N - B)$ number of bees are placed according to Eq. 5 across the search space for new solution. Random scouting represents the exploration capability of the BA.

3.4 Population Update

At the end of each cycle, the population is updated from two groups. The first group comprises the B scout bees which are associated with the best solution of each neighbourhood and represents the results of the local exploitative search. The second group is composed of the $(N - B)$ no. of scout bees associated with a randomly generated solution and represents the results of the global explorative search.

4 Adaptive Polynomial Mutation Based BA (APM-BA)

BA was originally designed as a numerical optimization technique based on foraging behaviour of honey bees and proved its robustness and efficiency to solving non-linear, real-valued function optimization problems. However, when dealing with multimodal problems, quality of solution is affected as the number of iterations

increases and it suffers from premature convergence. The situation occurs when all the best visited sites converge in a small region of the search space, forcing them to converge to the global best point found so far which is not a global optima. BA sometimes suffers from premature convergence due to many local minima in the search space. In general, convergence is a desirable property that recruited bees of best visited sites and allowed to search near the global minimum as time progresses. Unfortunately, in the context of many local minima, the scout bees of the best visited sites are trapped in one of the local minima and fail to explore more promising neighbouring minima. To enhance the exploration capability and to avoid being trapped into local optima, a mutation strategy is necessary to increase the diversity of the best scout bees in the search space. With this observation, an adaptive polynomial mutation based bees algorithm (APM-BA) has been proposed in this paper to prevent premature convergence. We define a neighbourhood structure of each of the say, B selected site in the local search processes of APM-BA. For example, the jth component of ith selected site creates neighbourhood by Eq. 6.

$$X_i^j = X_i^j + rand(-2, 2) \qquad (6)$$

We assume a parameter *trial* representing the number of iterations lead to inefficient search before better position is derived. If the ith best scout bee finds a better site, *trial(i)* is set to zero; otherwise, it is incremented by one for the next iteration. However, the searching competence of a best scout bee should not be evaluated by the quality of its current site i.e. the fitness value but by the efficiency of current search i.e. by the *trial* counter. Finally to avoid premature convergence, we employed adaptive polynomial mutation on i^{th} best scout bee when a specific number of times the ith best scout bee cannot improve its current position.

4.1 Adaptive Polynomial Mutation (APM)

In adaptive polynomial mutation strategy [21], the jth dimension of a ith candidate solution X_i is mutated with polynomial mutation as expressed in Eq. 7.

$$X_i^j(t+1) = X_i^j(t) + \left(X_{max}^j - X_{min}^j \right) \times \delta_j \qquad (7)$$

Where t represents current iteration number, X_{max}^j and X_{min}^j are the upper and lower bound of jth component of X_i while δ_j represents the polynomial function calculated using Eq. 8.

$$\delta_j = \begin{cases} (2r_j)^{\frac{1}{(\eta_m+1)}} - 1 & r_j < 0.5 \\ 1 - [2(1-r_j)]^{\frac{1}{(\eta_m+1)}} & r_j \geq 0.5 \end{cases} \qquad (8)$$

η_m is the polynomial distribution index and r_j represents uniformly distribute random number in (0,1). The probability of δ_j is calculated using Eq. 9.

$$P(\delta_j) = 0.5(\eta_m + 1)(1 - |\delta_j|)^{\eta_m} \tag{9}$$

By varying η_m, the perturbation can be varied in the mutated solution. If the value of η_m is large, a small perturbation of a variable is achieved. To achieve gradually decreasing perturbation in the mutated solutions, the value of η_m is gradually increased. The following rule presented in Eq. 10 is applied to achieve the proposed adaption policy known as adaptive polynomial mutation.

$$\eta_m = (80 + t) \tag{10}$$

To improve the performance of BA, we used adaptive polynomial mutation on ith best scout bees if the $trial(i) > D$. The mutated solution mX_i^j is obtained by Eq. 11.

$$mX_i^j = X_i^j + (X_k^j - X_i^j) \times \delta_j \tag{11}$$

In Eq. 11, the ith best scout bee exchange information with the kth one in its jth component where $k \neq i$. If $f(mX_i) \leq f(X_i)$, then i^{th} best scout bee X_i is replaced by mX_i and the $trial$ counter is set to 0 i.e. $trial(i) = 0$.

   The pseudo-code of APM-BA for protein structure prediction is given in Algorithm 1. Since, protein structure prediction is a minimization problem, fitness values are ranked in ascending order.

5 Experiments and Results

In order to evaluate the performance of the proposed algorithm, the experiments are performed on both artificial and real protein sequences for protein structure prediction in 2D AB off-lattice model.

5.1 Artificial Protein Sequence

Fibonacci sequence known as artificial protein sequence considered usually as benchmark for the protein structure prediction problem in AB off-lattice model [22]. A Fibonacci sequence is defined recursively by
   $S_0 = A, S_1 = B, S_{i+1} = S_{i-1} * S_i$
Where $'*'$ is the concatenation operator. The first few sequences are $S_2 = AB$, $S_3 = BAB$, $S_4 = ABBAB$ and so on. Hydrophobic residue 'A' occurs in isolation along the chain, while hydrophilic residue 'B' occurs either isolated or in pairs and the

Algorithm 1 APM-BA Algorithm

1: Initialize the parameters of BA, Maximum iterations G, Number of candidate solutions N, Dimensions D and the inefficient *trial* counter $trial(i) = 0$.
2: Initialize Population X by Eq. 5 and evaluate fitness $f(x)$
3: Ranked fitness values $f(x)$ and X
4: **for** $iter = 1$ to G **do**
5:   **for** $i = 1$ to E **do**
6:     **for** $j = 1$ to E_r **do**
7:       Generate neighbourhoods according to Eq. 6
8:     **end for**
9:     Select best neighbourhood, X_{ngh}
10:    **if** $f(X_{ngh}) < f(X_i)$ **then**
11:      update X_i and $f(X_i)$
12:      $trial(i) = 0$
13:    **else**
14:      $trial(i) = trial(i) + 1$
15:    **end if**
16:    **if** $trial(i) > D$ **then**
17:      generate mX_i according to Eq. 11
18:      **if** $f(mX_i) < f(X_i)$ **then**
19:        update X_i and $f(X_i)$
20:        $trial(i) = 0$
21:      **end if**
22:    **end if**
23: **end for**
24: **for** $i = 1$ to $B - E$ **do**
25:    **for** $j = 1$ to B_r **do**
26:      repeat lines from 7 to 22
27:    **end for**
28: **end for**
29: **for** $i = 1$ to $N - B$ **do**
30:    generate according to Eq. 5 and evaluate fitness $f(X_i)$
31: **end for**
32: Ranked fitness values $f(x)$ and X
33: **end for**
34: Output the best solution

---

molecules have a hierarchical string structure. Artificial protein sequence lengths are obtained through the Fibonacci numbers $n_{i+1} = n_{i-1} + n_i$. In the experiment, we have considered artificial protein sequence given in Table 1 with different lengths, say 13, 21 and 34.

Table 1 Artificial protein sequences

Artificial Protein	Sequence	Length
S_{13}	*ABBABBABABBAB*	13
S_{21}	*BABABBABABBABBABABBAB*	21
S_{34}	*ABBABBABABBABBABABBABABBABBABABBAB*	34

5.2 Real Protein Sequence

In order to measure the effectiveness of the APM-BA algorithm in predicting real protein structures, we select protein sequences from Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do). The PDB ID of these protein sequences are 1BXP, 1BXL, 1EDP and 1EDN respectively. The sequence information of these real proteins are given in Table 2. In the experiment, the same K-D method used in the literature [23] is adopted to distinguish the hydrophobic and hydrophilic residues of 20 amino acids in real protein sequences. The amino acids I, V, L, P, C, M, A, G are considered as hydrophobic *(A)* residues and D, E, F, H, K, N, Q, R, S, T, W, Y are hydrophilic *(B)* residues.

Table 2 Real protein sequences

Real Protein	Sequence	Length
1BXP	MRYYESSLKSYPD	13
1BXL	GQVGRQLAIIGDDINR	16
1EDP	CSCSSLMDKECVYFCHL	17
1EDN	CSCSSLMDKECVYFCHLDIIW	21

## 5.3 Parameter Settings and Initialization

The proposed APM-BA algorithm is compared with simple bees algorithm [20] and the algorithms which are already used for protein structure prediction such as CPSO [10], EPSO [11], IF-ABC [18] in 2D off-lattice model. All experiments are implemented in MATLAB R2010a and executed on an Intel Core (TM) 17-2670 QMCPU running at 2.20 GHz with 8 GB of RAM with Windows XP. The independent experiments of each algorithm is repeated 30 times with same initial population. The population size (N) for all approaches is fixed at 50 but the dimension D is different based on the respective length of protein sequences. The stopping criteria is same for all algorithms based on number of iterations (G). The number of iterations is defined [24] by $G = (D \times 10000)/N$. The parameters of the proposed algorithm are given in Table 3. The best of scout bees are going through adaptive polynomial mutation when *trial* counter of each of these scout bees are greater than D.

Table 3 APM-BA learning parameters

Parameters	Values
Number of scout bees (N)	50
Number of elite sites (E)	10
Number of best sites (B)	20
Number of recruited bees for elite sites (E_r)	10
Number of recruited bees for remaining best sites (B_r)	5

5.4 Results for Artificial Protein Sequences

Table 4 list the mean and standard deviation (Std.Dev) of 30 runs obtained by the
APM-BA along with the results obtained by CPSO, EPSO, IF-ABC and BA for
comparison. From Table 4, we can see that minimum energy obtained by the pro-
posed algorithm dominates all other algorithms for every artificial protein sequences.
Strong significant improvement has been observed in case of protein sequence length
21 and 34 by APM-BA. Three artificial protein sequences are placed in 4th, 3rd and
2nd position with respect to standard deviation by the APM-BA which measures
robustness of the minimum energy, as shown in Table 4. Therefore, the proposed
algorithm out performs than other algorithms as increase the length of artificial pro-
tein sequence. The convergence characteristics of each algorithm on artificial protein
sequences of lengths 13, 21 and 34 are shown in Fig. 2. The APM-BA exhibits better
convergence than other approaches.

(a) L=13   (b) L=21   (c) L=34
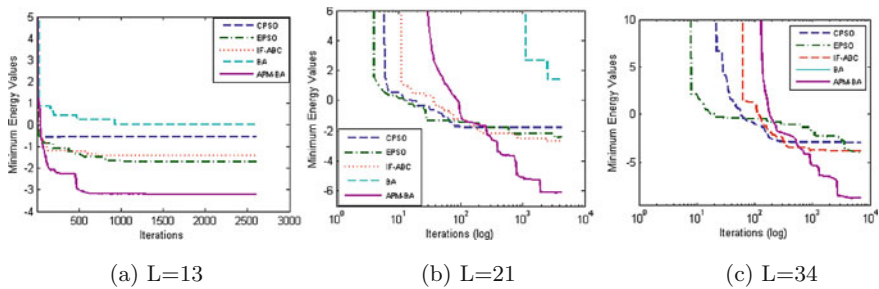
Fig. 2 Convergence graph of artificial protein sequence with different length

## 5.5 Results for Real Protein Sequence

The results obtained by the APM-BA for real protein sequences are summarized
in Table 5, along with the results of other algorithms used for comparison. It has
been observed that lowest energy obtained by the proposed algorithm is significantly
better than that of other algorithms like CPSO. EPSO, IF-ABC and BA. Therefore,
APM-BA is superior in solving real protein sequences. Based on standard deviation
for all real protein sequences compare to other algorithm, the proposed approach is
highly robust compare to other algorithms. The convergence characteristics of the
algorithms on real protein sequences are plotted in Fig. 3.

Table 4 Results for artificial protein sequence

Artificial protein	CPSO		EPSO		IF-ABC		BA		APM-BA	
	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
S_{13}	−1.653	0.518	−1.974	0.346	−1.846	0.190	0.020	0.228	−2.596	0.370
S_{21}	−3.352	0.741	−3.400	0.432	−3.062	0.264	1.491	0.850	−4.964	0.601
S_{34}	−5.284	1.302	−5.671	0.910	−4.586	0.363	2397.448	2845.384	−6.694	0.701

Table 5 Results for real protein sequence

Real protein	CPSO		EPSO		IF-ABC		BA		APM-BA	
	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
1BXP	-1.893	0.432	-2.110	0.250	-1.517	0.162	-0.118	0.349	-2.335	0.085
1BXL	-6.296	1.062	-6.255	1.161	-5.333	0.359	-0.696	1.087	-8.002	0.592
1EDP	-3.092	1.593	-3.773	1.438	-3.093	0.386	0.689	0.614	-6.230	0.518
1EDN	-4.232	1.325	-5.005	0.812	-4.232	0.482	2.135	1.058	-7.275	1.078

(a) L=13

(b) L=16

(c) L=17

(d) L=21
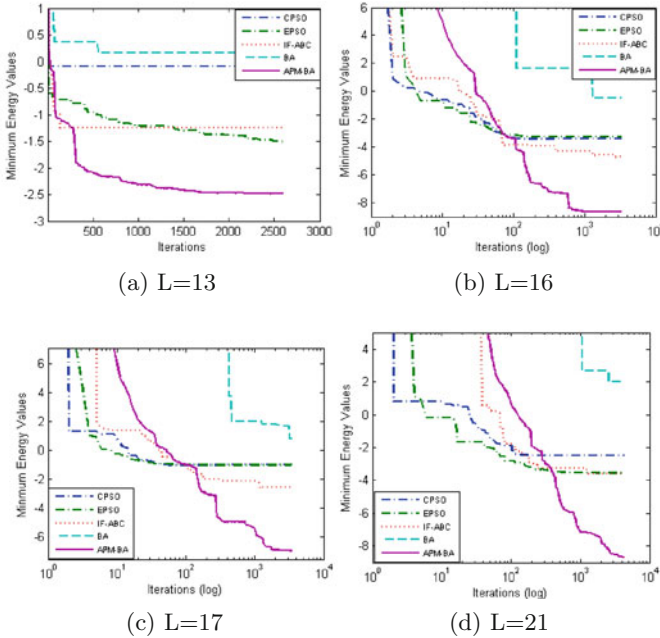
Fig. 3 Convergence characteristics of real protein Sequence with different length

6 Conclusions

In this paper, an improved BA revised by adaptive polynomial mutation strategy is introduced to optimize protein structure prediction in 2D AB off-lattice model. In APM-BA, adaptive polynomial mutation is applied to the best scout bees based on their inefficiency during the search processes. The proposed strategy is able to preventing stuck at local optima and made exploration on the search space. Experimental results confirm that APM-BA is significantly more effective for protein structure prediction problem with respect to artificial and real protein sequences. It should be noted that our study concerns few number of protein sequences with smaller lengths.

Predicting structure of more real protein sequences with larger lengths in 3D AB off-lattice model and investigations on the neighbourhood structure as well as the selection procedure of best visited sites of BA will be our future work.

References

1. Freitas, A.A., Wieser, D.C., Apweiler, R.: On the importance of comprehensible classification models for protein function prediction. IEEE/ACM Trans. Comput. Biol. Bioinform **7**(1), 172–182 (2010)
2. May, A., Pool, R., Dijk, E.V., Bijlard, J., Abeln, S., Heringa, J., Feenstra, K.A.: Coarse-grained versus atomistic simulations: realistic interaction free energies for real proteins. Bioinformatics **30**(3), 326–334 (2014)
3. Anfinsen, C.B.: Principles that govern the folding of protein chain. Science **181**(4096), 223–230 (1973)
4. Pierce, N.A., Winfree, E.: Protein design is np-hard. Protein. Eng. **15**(10), 779–782 (2002)
5. Rossi, G., Ferrando, R.: Searching for low-energy structures of nanoparticles: a comparison of different methods and algorithms. J. Phy. Condens. Matter **21**(8), 84208 (2009)
6. Dorn, M., e Silva, M.B., Buriol, L.S., Lamb, L.C.: Three-dimensional protein structure prediction: methods and computational strategies. Comput. Biol. Chem. **53**, 251–276 (2014)
7. Dill, A.K., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S.: Principle of protein folding: a perspective from simple exact models. Protein Sci. **4**(4), 561–602 (1995)
8. Stillinger, F.H., Head-Gordon, T., Hirshfel, C.L.: Toy model for protein folding. Phys. Rev. **48**(2), 1469–1477 (1993)
9. Kim, S.Y., Lee, S.B., Lee, J.: Structure optimization by conformational space annealing in an off-lattice protein model. Phys. Rev. **72**(1), 011916 (2005)
10. Liu, J., Wang, L., He, L., Shi, F.: Analysis of toy model for protein folding based on particle swarm optimization algorithm. In: Wang, L., Chen, Ke, S. Ong, Yew (eds.) ICNC 2005. LNCS, vol. 3612, pp. 636–645. Springer, Heidelberg (2005)
11. Zhu, H., Pu, C., Lin, X., Gu, J., Zhang, S., Su, M.: Protein structure prediction with epso in toy model. In: Second International Conference on Intelligent Networks and Intelligent Systems, 2009. ICINIS '09, pp. 673–676 (2009)
12. Liu, J.F., Xue, S.J., Chen, D.B., Geng, H.T., Liu, Z.X.: Structure optimization of the two-dimensional off-lattice hydrophobichydrophilic model. J. Biol. Phys. **35**(3), 245–253 (2009)
13. Cheng-yuan, L., Yan-rui, D., Wen-bo, X.: Multiple-layer quantum-behaved particle swarm optimization and toy model for protein structure prediction. In: 2010 Ninth International Symposium on Distributed Computing and Applications to Business Engineering and Science (DCABES), pp. 92–96 (2010)
14. Kalegari, D.H., Lopes, H.S.: A differential evolution approach for protein structure optimization using a 2d off-lattice model. J. Bio-Inspired Comput. **2**(3), 242–250 (2010)
15. Mansour, R.: Applying an evolutionary algorithm for protein structure pre-diction. Am. J. Bioinf. Res. **1**(1), 18–23 (2011)
16. Jana, N.D., S, Jaya: Hybrid particle swarm optimization technique for protein structure prediction using 2D off-lattice model. In: Panigrahi, Bijaya Ketan, Suganthan, Ponnuthurai Nagaratnam, Das, Swagatam, Dash, Shubhransu Sekhar (eds.) SEMCCO 2013, Part II. LNCS, vol. 8298, pp. 193–204. Springer, Heidelberg (2013)
17. Jingfa, L., Sun, Y., Li, G., Song, B., Huang, W.: Heuristic based tabu search algorithm for folding two-dimensional ab off-lattice model proteins. Comput. Biol. Chem. **47**, 142–148 (2013)
18. Li, B., Li, Y., Gong, L.: Protein secondary structure optimization using an improved artificial bee colony algorithm based on ab off-lattice model. Eng. Appl. Artif. Intell. **27**, 70–79 (2014)
19. Pham, D., Castellani, M.: The bees algorithmmodelling foraging behaviour to solve continuous optimisation problems. J. Bio-Inspired Comput. **223**(12), 2919–2938 (2009)

20. Pham, D., Castellani, M.: Benchmarking and comparison of nature-inspired population based continuous optimization algorithms. Soft comput. **18**, 871–903 (2014)
21. Saha, A., Datta, R., Deb, K.: Hybrid gradient projection based genetic algorithms for constrained optimization. In: 2010 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8 (2010)
22. Stillinger, F.H.: Collective aspects of protein folding illustrated by a toy model. Phys. Rev. E **52**(3), 2872–2877 (1995)
23. Mount, D.W.: Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001)
24. Liang, J.J., Qu, B.Y., Suganthan, P.N., Hernandez-Diaz, A.G.: Problem definitions and evaluation criteria for the cec 2013 special session on real-parameter optimization. Technical report DAMTP 2000/NA10, Nanyang Technological University, Singapore (2013)