

# Elitist Animal Migration Optimization for Protein Structure Prediction based on 3D Off-Lattice Model

Ezgi Deniz Ülker

Department of Software Engineering, European University of Lefke, Mersin-10, Turkey

**Abstract**—Predicting the structure of protein has been the center of attraction for the researchers. The aim is to make a reliable prediction of the protein structure by obtaining the minimum energy values among amino acids interactions. According to the generated shape of amino acids, the functionality of the proteins can be determined. However, it is known as one of the most challenging tasks in the field of bioinformatics considering its high computation complexity. Metaheuristic algorithms are mainly preferred by researchers from various fields, since their performances are quite satisfactory in solving such complex problems. Animal Migration Optimization (AMO) algorithm is a metaheuristic approach which mimics the behavior of animals during the migration process. However, in this research to reach a high solution quality, an elitist version of Animal Migration Optimization (ELAMO) algorithm is considered and in particular it is applied to Protein Structure Prediction (PSP) problem. The performance of ELAMO is tested on some well-studied artificial and real protein sequences, and then compared with powerful optimization algorithms which are specially designed for solving PSP problem. The results show that ELAMO is quite capable in solving this problem. Hence, it can be used as an efficient optimizer for solving complex problems that require better solution quality in the field of bioinformatics.

**Keywords**—Animal migration optimization; bioinformatics; elitism; metaheuristics; protein structure prediction

## I. INTRODUCTION

In molecular biology, comprehending the structure of a protein sequence reveals the hidden functionalities of the life [1]. When the proteins are folded in different ways, the information necessary for understanding their functionalities will arise. Proteins are formed by the combination of amino acids which are connected by peptide bonds [2]. According to Christen Anfinsen's leading work, proteins can be found in the lowest energy levels which are called Gibbs energy level, when they are in three dimensional states [3]. Protein Structure Prediction (PSP) problem is located on finding this state by seeking the minimum Gibbs energy level. As the amino acid sequence becomes large, predicting the structure of a protein sequence becomes complex.

Researchers developed an approach called 'HP model' for protein folding prediction [4]. In the HP model, a protein description is made up of smaller pieces called monomers and which are either represented on 2D or 3D surface. 'H' and 'P' letters are used to define each of the monomers which are hydrophobic or polar, respectively. It is aimed to find the optimal structure of a given H-P chain that is defined as the maximum number of H-H bondings. Although the HP model is

specifically designed for solving protein folding with its simplicity, it does not provide satisfactory solutions for PSP. The problem is proved to be an NP-hard problem due to large number of amino acids sequences and requires quite efficient algorithms to solve them [5–7].

One of the biggest limitations of protein folding is having the multiple local optimum points in the free-energy space and the global optimum is located in between these points which is quite challenging to obtain [8]. In order to design a scheme by avoiding the large computational cost, such models with eliminated properties in protein folding have been preferred [8–10]. An accurate example of these kinds of models is the off-lattice model which is presented by Stlinger et al. [8]. The model is employed to simplify the protein folding.

Animal Migration Algorithm (AMO) is a bioinspired metaheuristic approach proposed by Li et al. [11]. It is founded on an animal's instinct to follow their close neighbors during the migration and has quite validated performances on many optimization problems. Despite of the noticeable properties of AMO, there may be some disadvantages such as low convergence rate by choosing the next possible solutions only among the current animal's neighborhood or having a less chance by finding the global optimum because of following the wrong neighbors. In order to avoid these limitations and make it guarantee that the algorithm converges to the global optimum in less number of iterations, Elitist Animal Migration Optimization (ELAMO) is proposed on the basis of an animal's instinct to follow their leaders, not only their closest neighbors [12]. ELAMO has a validated performance in solving combinatorial NP-hard problem. However, to the best of our knowledge neither AMO nor ELAMO have been proposed for solving bioinformatics problems, particularly, for the PSP problem. In this paper, ELAMO algorithm is adapted to bring another aspect in solving Protein Structure Prediction (PSP) problem by using 3D AB Off-Lattice Model.

The rest of the paper is designed as follows; in Section II, some important studies for solving the PSP over the years have been given. In Section III, three dimensional AB off lattice model and the adaptation of Elitist Migration Algorithm to PSP problem with the model equations are given. In Section IV, ELAMO algorithm's performance is compared on both synthetic and real protein sequences with some powerful optimizers. The obtained results with the visual representations of minimum energy configurations and discussions are given in detail in this section. Lastly, in Section V, the concluding remarks are given.

## II. RELATED WORK

It is known that many metaheuristic algorithms are verified to be quite efficient in solving complex and even NP-hard problems. In specific, it is found that Bee colony optimization algorithms and its variants are used to solve PSP problem by Li et al. [13]. Kalegari and Lopes proposed an improved Differential Evolution algorithm for solving PSP using 2D and 3D off-lattice models efficiently [14]. Another Differential Evolution algorithm variant is proposed by Rakhshani et al. for solving complex protein structure prediction problems [15]. Deep learning practices are also tested for the PSP by Senior et al. [16] and achieved promising results. Schauerperl and Denny performed an AI based protein structure prediction in drug discovery [17], Chowdhury et al. [18] solved protein prediction problem using a deep learning model and Weißenow et al. [19] have solved PSP problem using AI model accurately. Multi-meme algorithms are also adapted for solving PSP by Krasnogor et al. [20]. Lin and Zhang introduced a novel-hybrid global optimization method by forming Genetic Algorithm and Particle Swarm Optimization to solve PSP in which aiming to produce lower energy conformation levels [21]. Boiani and Parpinelli proposed a hybrid algorithm called cuHjDE-3D which is formed by self-adaptive Differential Evolution that uses jDE and Hooke-Jeeves Direct Search (HJDS) [22].

The literature review revealed that the standard metaheuristic algorithms have limited performance in solving PSP problem. Before attempting to solve the problem, researchers either modify or hybridize the standard algorithms. By introducing such boosted algorithms, the researchers aimed to use the strengths of the algorithms on PSP problem.

In this study, none of the machine learning methods are implemented. Instead, a modified version of AMO algorithm has been studied to observe how it evaluates the problem by using its parameters. One of the main contributions of this study is to adapt animal migration algorithm by enhancing its diversity using elitist approach for protein sequence prediction problem and achieving satisfactory results.

## III. MATERIALS AND METHODS

### A. Three Dimensional AB Off-Lattice Model

The AB off-lattice model is stimulated by HP model and considered as one of the useful models for solving Protein Structure Prediction Problem. When the AB off-lattice model was introduced, it was initially designed for 2-D protein structures. However, then the model was upgraded for solving 3-D models as well [8, 23].

In a protein sequence 20 types of amino acids exist which are categorized in two; hydrophobic and hydrophilic. This is simply performed according to their affinity to water. The amino acids then translated to two specialized monomers 'A' and 'B'. As K-D method proposes I, V, L, P, C, M, A, G are hydrophobic amino acids represented by letter A and D, E, F, H, K, N, Q, R, S, T, W, Y are hydrophilic amino acids represented by letter B [24]. The name AB of AB off-lattice model is because of these specialized monomers A and B.

The amino acids are bonded with each other by chemical bonds and can be placed anywhere in the 3D space. The reason

of that is called off-lattice is the positions of the amino acids which are not restricted by a lattice. In the AB off-lattice model, bondings are formed by set of angles; folding ( $\theta$ ) and rotation ( $\phi$ ). In a protein sequence which contains n monomers also contains n-2 folding angles and n-3 rotation angles. The optimal structure of AB off-lattice model produces the free energy which gives general information about the physical and chemical concept of protein sequences. Fig. 1 shows the representation of an artificial protein sequence ABAA with folding ( $\theta$ ) and rotation ( $\phi$ ) angles where 'A' and 'B' are hydrophobic and hydrophilic amino acids, respectively. Folding angles  $\{\theta_1, \theta_2\}$  and rotation angle  $\{\phi_1\}$  are needed to be optimized for having the minimum free energy level.

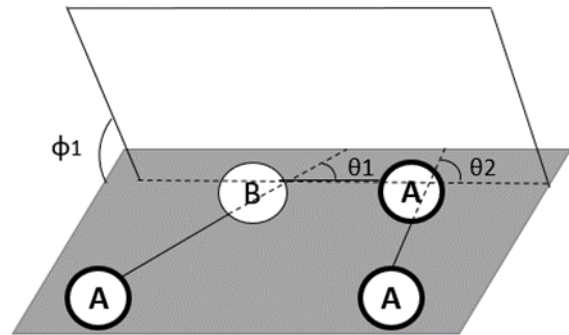


Fig. 1. 3D AB-off lattice model representation for ABAA artificial protein sequence.

Adaptation of protein structure prediction to a numerical optimization problem by 3D AB-off lattice model is done as follows,

$$Energy = \sum_{i=2}^{n-1} \frac{1-\cos \theta_i}{4} + 4 \sum_{i=1}^{n-2} \sum_{j=i+2}^n [r_{ij}^{-12} - C(\varepsilon_i, \varepsilon_j) r_{ij}^{-6}] \quad (1)$$

where  $\varepsilon_i$  is the characteristic of the  $i^{\text{th}}$  amino acid. If  $\varepsilon_i = 1$ , then  $i$  is a hydrophilic amino acid. If  $\varepsilon_i = -1$ , then  $i$  is a hydrophobic one. The folding angles ( $\theta$ ) are bounded  $[-180^\circ, 180^\circ]$ .

To obtain the distance between amino acids  $i$  and  $j$ , the following equation is used.

$$r_{ij} = \sqrt{[1 + \sum_{k=i+1}^{j-1} \cos(\sum_{l=i+1}^k \theta_l)]^2 + [\sum_{k=i+1}^{j-1} \sin(\sum_{l=i+1}^k \theta_l)]^2} \quad (2)$$

The following equation shows the basic rotations for two amino acids.

$$C(\varepsilon_i, \varepsilon_j) = \frac{1}{8}(1 + \varepsilon_i + \varepsilon_j + 5\varepsilon_i\varepsilon_j) \quad (3)$$

In AB-off lattice model, it is assumed that strong correlations between AA pairs result with the value of  $C(\varepsilon_i, \varepsilon_j) = 1$ , relatively weaker correlations between BB pairs result with the value of  $C(\varepsilon_i, \varepsilon_j) = 0.5$  and different pairs BA or AB pairs result with the value of  $C(\varepsilon_i, \varepsilon_j) = -0.5$ . Using the assumptions obtained through AB-off lattice model, the protein structure problem is converted into a numerical optimization problem that can be handled by evolutionary

optimization techniques. By having various ordering of distances and rotations, different energy levels of amino acids are obtained and as the algorithm iterates the optimum energy level is obtained.

### B. Adaptation of ELAMO to PSP Problem

Animal migration is a common behavior which belongs to animal herds to be used in discovering better places to live and reproduce. Animal Migration Optimization (AMO) algorithm based on this behavior and proved to be a validated optimizer in solving optimization problems [11]. Our approach is simply based on the main steps of AMO by including the elitism behavior in it. In the Elitist Animal Migration approach, the neighborhood structure of the standard AMO is reconstructed. Thus, the animals in the herd follow their leaders not only their close neighbors.

During the migration process, an animal's position depends on its neighbor. In standard AMO, migration is done by following five closest neighbors of each animal. However, in our elitist approach, an animal's instinct to follow the leader of the herd is essential. In a typical animal herd, there are three kinds of animal; Alpha ( $\alpha$ ), Beta ( $\beta$ ) and Omega ( $\omega$ ).

Alpha ( $\alpha$ ) is responsible from all animals in the herd such as finding the preys or discovering new life areas. If the alpha dies, a new leader is selected among the beta ( $\beta$ ) animals who are in charge after the alpha ( $\alpha$ ). The rest of the animals are considered as omegas ( $\omega$ ) who obey the rules of the herd. In the algorithm ELAMO, only  $\alpha$  and  $\beta$  animals are in charge of migration. The number of  $\alpha$  is 1 and the number of  $\beta$  is fixed to 4. Thus, all of the animals in the herd move towards to new life areas by following these leaders. As the algorithm iterates, new  $\alpha$  and  $\beta$  animals are selected with respect to their positions among the rest of the animals. The following figures; Fig. 2 and Fig. 3 demonstrate the neighborhood structure in AMO and ELAMO, respectively where each animal represented by a circle theoretically.

Elitist animal migration algorithm is built up on two fundamental steps; Animal migration and population updating.

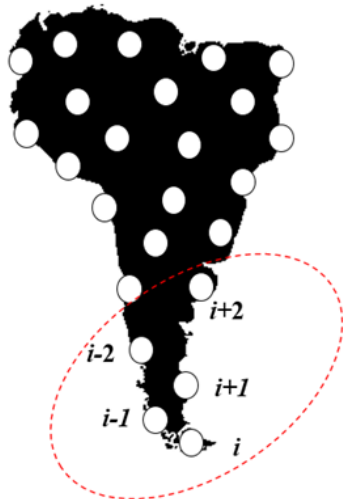


Fig. 2. The neighborhood structure of AMO algorithm where  $i$  is an animal and  $i \pm 1,2$  are the closest neighbors.

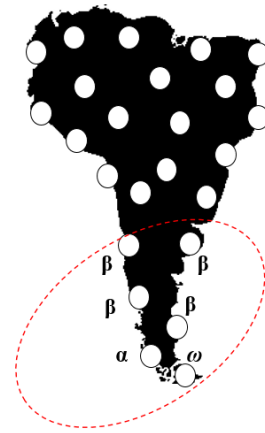


Fig. 3. The neighborhood structure of ELAMO algorithm.

In animal migration step, animals change their positions towards their  $\alpha$  and  $\beta$  animals as given in (4). In population updating step, displacement of animals is introduced. Some animals may be eliminated due to death or they may compete for their positions and the losers are discarded from the population. The new positions are updated according to their fitness values as it is shown in (5).

Animal migration step;

$$X_{i,G+1} = X_{i,G} + \delta(X_{neighborLeader,G} - X_{i,G}) \quad (4)$$

where  $\delta$  is a random number produced by Gaussian distribution and  $G$  is the dimension for each animal  $\in [1... D]$  and  $X_{neighborLeader}$  is the leader's position randomly selected from the neighborhood structure of an animal  $X_i$ .

Population updating step;

$$X_{i,G+1} = X_{betaRand_{a,G}} + rand(X_{alpha,G} - X_{i,G}) + rand(X_{betaRand_{b,G}} - X_{i,G}) \quad (5)$$

where  $X_{betaRand}$  is an animal selected randomly between beta animals,  $X_{alpha}$  is the position of alpha,  $rand$  is a random number in between 0, 1 and  $a \neq b$ .

The main control parameters of ELAMO are  $\delta$  and  $rand$  which influence the population by having equilibrium between the diversity; exploring new possible areas in the search space and intensity; focusing the search area around the leaders  $\alpha$  and  $\beta$ .

The adaptation of ELAMO to PSP problem is given in the Fig. 4. One of the main motivations is to reach the higher optimization level by changing the neighborhood structure. In standard AMO, the closest neighbor's position might be used just because they are considered as the closest neighbors and even if their positions are relatively worse than the others. However, in ELAMO, the best positions are chosen at each iteration and are followed by all of the animals.

As it is explained clearly in no-free lunch theorem [25], when an algorithm's performance is sufficient in some aspects, the performance may not reach to that level as it is expected for the other aspects. In ELAMO, individuals are discarded from the population as the elitist feature requires even in the

beginning of the iterations. Therefore, it may be considered in ELAMO the balance of diversification and intensification may be negatively affected by the loss of individuals in the earlier stages of optimization.

<p>Step 1. Initialization of animal position (<math>X</math>) of population (<math>NP</math>) with dimension (<math>D</math>).</p> <p>Step 2. Generation of sequence order <math>S</math> from the positions of animals.</p> <p>Step 3. Applying AB-off lattice model to derive the energy levels as fitness values.</p> <p>Step 4. Updating the position of each animal according to neighbors by using the following;</p> $X_{i,G+1} = X_{i,G} + \delta(X_{neighbor,G} - X_{i,G})$ <p>where <math>\delta</math> is a random number produced by Gaussian distribution and <math>G</math> is the dimension for each animal <math>\in [1..D]</math>.</p> <p>Step 5. Replacing the positions, if the updated positions are better.</p> <p>Step 6. Introducing new animals to the population according to a predefined probability while others are leaving by using the following;</p> $X_{i,G+1} = X_{betaRand,G} + rand(X_{alpha,G} - X_{i,G}) + rand(X_{betaRand,G} - X_{i,G})$ <p>where <math>X_{betaRand}</math> is an individual selected randomly between beta animals, <math>rand</math> is a random number in between 0 and 1.</p> <p>Step 7. Generation of new fragment order <math>S</math> according to the updated positions of animals.</p> <p>Step 8. Applying AB-off lattice model to derive the energy levels as fitness values.</p> <p>Step 9. Replacing the positions, if the newly introduced animal's position is better.</p> <p>Step 10. Repeating steps 4-9 until a stopping criterion is met.</p> <p>Step 11. Finding the order of the proteins with the angles which corresponds to the lowest energy level.</p>
---

Fig. 4. The main steps of ELAMO algorithm for PSP problem.

A herd with  $\alpha$ ,  $\beta$  and  $\omega$  animals correspond an AB-off lattice model for a set of sequences. All of the animals in the herd are the potential solutions in solving PSP by corresponding distances, angles and interactions between particles. According to these values, energy levels are derived. In PSP, the objective function is the energy level function and the optimal solution refers to the lowest energy value.

#### IV. RESULTS AND DISCUSSION

In this section, two sets of analyses were used. First, a set of artificial Fibonacci protein sequences which have been used as benchmarks commonly in the literature is studied [8, 26] and then real protein sequences were analyzed. The data sets are experimentally examined structures for testing the efficiency of the methods for PSP problem.

A list of benchmark sequences with the 'A' and 'B' monomers are given in Table I where N is the sum of the monomers. A comparative study is performed in Table II to observe the performance of ELAMO with respect to some powerful optimization algorithms; Improved Particle Swarm Optimization (EPSO) [27], Internal Feedback strategy based on Artificial Bee Colony Algorithm (IF-ABC) [28], Combination of Genetic Algorithm and Particle Swarm Optimization (GAPSO) [21] and standard AMO which ELAMO is originated from. It is also important to note that the results of the compared algorithms included to the comparison table as they appeared in their original studies.

All simulations are implemented on an Intel Core i5 CPU with 4 GB RAM running at 3.10 GHz by C++ language. All benchmark sequences are evaluated for 30 independent runs with random initial points. All protein sequences are optimized by AMO and ELAMO for 50,000 number of iterations.

TABLE I. DETAILS OF THE ARTIFICIAL PROTEIN SEQUENCES

Coded Sequence	N
ABBABBABABBAB	13
BABABBABABBABBABABBAB	21
ABBABBABBBABBABABBABABBABBABABBAB	34

Fig. 5 is the representation of the changing lowest free energy values for three artificial protein sequences by the algorithms chosen for comparison. It can be clearly seen from both Table II and Fig. 5 that ELAMO has reached better energy values than the other algorithms for all artificial sequences.

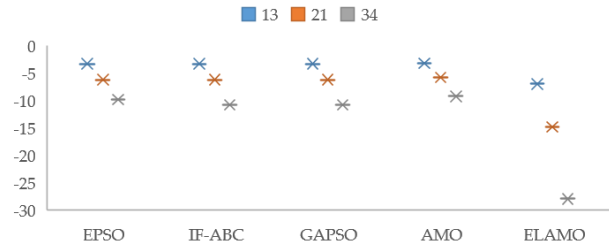


Fig. 5. A comparative free energy values obtained by different algorithms for Fibonacci sequences.

As shown in Fig. 5, almost all of the algorithms converge in similar rate in the first level of iterations; however as the iteration goes on convergence rate of ELAMO stands out for all benchmarks than the others. When the original algorithm AMO compared with the others, it is seen that all of the selected algorithms converge better than AMO.

All of the algorithms selected for comparison are either the hybrid version of one or more algorithms or the improved versions of the originals. AMO algorithm is the only one its performance was not boosted. It might be the main reason of the observation above which is about the lower convergence rate of AMO. On the other hand, this is a good indication that AMO has been improved efficiently to handle such kind of problems with high convergence rates.

In order to observe the effect of the added elitism feature on AMO, a detailed comparison between AMO and ELAMO is studied and shown in the Table III, where *Best* is for the lowest free-energy value obtained after 50,000 number of iterations, *Avg* is the averaged free-energy values of 30 independent runs and *Stdev* is for the standard deviation value.

TABLE II. THE LOWEST FREE-ENERGY VALUES OF BENCHMARK SEQUENCES OBTAINED BY THE ALGORITHMS

List of sequences	N	EPSO	IF-ABC	GAPSO	AMO	ELAMO
ABBABBABA BBAB	13	-3.294	-3.294	-3.294	-3.216	-6.982
BABABBABA BBABBABAB BAB	21	-6.198	-6.198	-6.210	-5.751	-14.811
ABBABBABB BABBABABB ABABBABBA BABBAB	34	-9.834	-10.806	-10.789	-9.219	-27.989

The results denote the improved solution quality of ELAMO over AMO as well as robustness. However, for all of the artificial sequences the *Best* and the *Avg* values of ELAMO are superior than the values AMO. However, only for the sequence with the length 34, the *Stdev* value is not satisfactory as it is expected. ELAMO algorithm produces encouraging results for the lowest-free energy values and the average values, but does not achieve the development of standard deviation for this sequence. When the *Stdev* values are compared for the sequence with the length 21, it is seen that AMO is quite close to ELAMO. The reason of this unexpected performance might be because of  $\alpha$  and  $\beta$  animals' convergence pace in the search space, while some of the  $\omega$  animals are still in the optimization process. As the iterations progress, the majority of  $\omega$  animals find their way towards the optimum.

TABLE III. COMPARATIVE RESULTS BEST, AVERAGE AND STANDARD DEVIATIONS ACHIEVED BY AMO AND ELAMO

AMO				ELAMO		
N	Best	Avg	Stdev	Best	Avg	Stdev
13	-3.21	-2.44	0.52	-6.98	-6.38	0.39
21	-5.75	-4.71	1.17	-14.81	-14.31	0.99
34	-9.21	-8.02	1.20	-27.98	-26.30	1.83

Researchers dealt with well-preferred real protein benchmark functions for a better analyzing of their algorithms and employed to their works [15, 28–31]. The same protein sequences were used in this study as benchmark functions and rewritten according to the K-D method where I, V, L, P, C, M, A, G are hydrophobic amino acids represented by letter 'A' and D, E, F, H, K, N, Q, R, S, T, W, Y are hydrophilic amino acids represented by letter 'B'. The list of the amino acid sequences were selected from the widely used Protein Data Bank (PDB) database with different lengths to make a more efficient comparison with the other algorithms. The IDs, lengths and contained amino acids are given in the Table IV.

In the study with real protein sequences, only one standard algorithm chosen for the thorough check. Instead, specially designed hybrid algorithms with significant performances are selected. Table V lists the lowest free energy values obtained by ELAMO as well as other competitive algorithms to make an extensive comparison. Convergence values can be seen even in earlier iterations in the Fig. 6.

TABLE IV. DETAILS OF THE REAL AMINO ACID SEQUENCES

PDB ID	N	Sequence
1AGT	38	AAAABABABABABAABAABBAABBABAABBB ABABAB
1HVV	75	BAABBBABBBBBBAABABBABBABBABAAAA ABBBABAABBABBABBABAABBABBBA BBBBABBB
1GK4	84	BABAABBBBBBABBABBBAABAABBBBB AABABBBABBBAABBABBBAABAABAAB ABAABBBBAABBBA
2EWH	98	AABABAAAAAABBBAAAAAABAABAABBAAB ABAAAABBBAAAAABAAAABBAAAABAAA ABAABBAABAAAAABAABA BBBABBAABAABA

TABLE V. COMPARISON OF THE LOWEST FREE-ENERGY VALUES FOR THE SEQUENCES OBTAINED BY THE ALGORITHMS

PDB ID	N	CMAES	L-SHADE	E-MASA-PAMS	ABC	ELAMO
1AGT	38	-34.45	-39.31	-41.28	-25.65	-50.82
1HVV	75	-27.35	-28.77	-	-27.35	-46.98
1GK4	84	-32.72	-40.26	-	-32.72	-49.89
2EWH	98	-	-	-125.53	-	-146.81



Fig. 6. A free energy values obtained by different algorithms for real protein sequences.

It can be seen from the figure that at the very first level of optimization all of the algorithms' convergence rates are very similar, but in the latter iterations elitism feature begins to appear in ELAMO and this affects the convergence rate in a very desirable way. The test parameters and results of the compared algorithms are accepted as they appear in the references [15, 28–31]. The best results for AMO and ELAMO are obtained over 30 runs. The stopping condition is set 200,000 iterations.

It can be acquired from both Table V and Fig. 6 that the performance of ELAMO can be distinguished from the others by having more precise solutions of the problems. Considering all of the results, it is possible to say that the algorithms' performances are very competitive but in the view of high solution quality, the performance of ELAMO is quite noteworthy.

In the comparison of ELAMO with one of the standard algorithms ABC, it is observable that ELAMO produces better results. For the other competitive algorithms, we can say that CMAES and L-SHADE performances are similar to each other, while E-MASA-PAMS superior than both and produces comparable results with ELAMO. However, in both analysing, we can see that ELAMO always contributes with the lowest free energy values and it can be a distinct evident by applying a set of modifications on the right steps of the AMO algorithm, influence the performance of ELAMO in a quite remarkable way. It is note to point that the compared algorithms can not be further analyzed since there is no *Stdev* or *Avg* reported in the literature.

In the light of findings, the visual representation of the folded protein structures by the best run of ELAMO are shown in the Fig. 7 (a) to (d). In the figures, green dots represent the hydrophobic monomer; 'A' and the purple represent the hydrophilic monomer; 'B'. As the figures reveal, the hydrophobic monomers are frequently enclosed by the

hydrophilic monomers in the folding structures. This is a natural phenomena for avoiding contact with water molecules and is verified by the following figures as well.

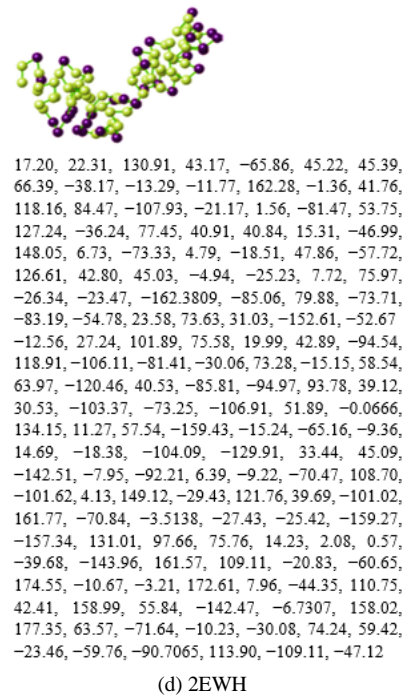
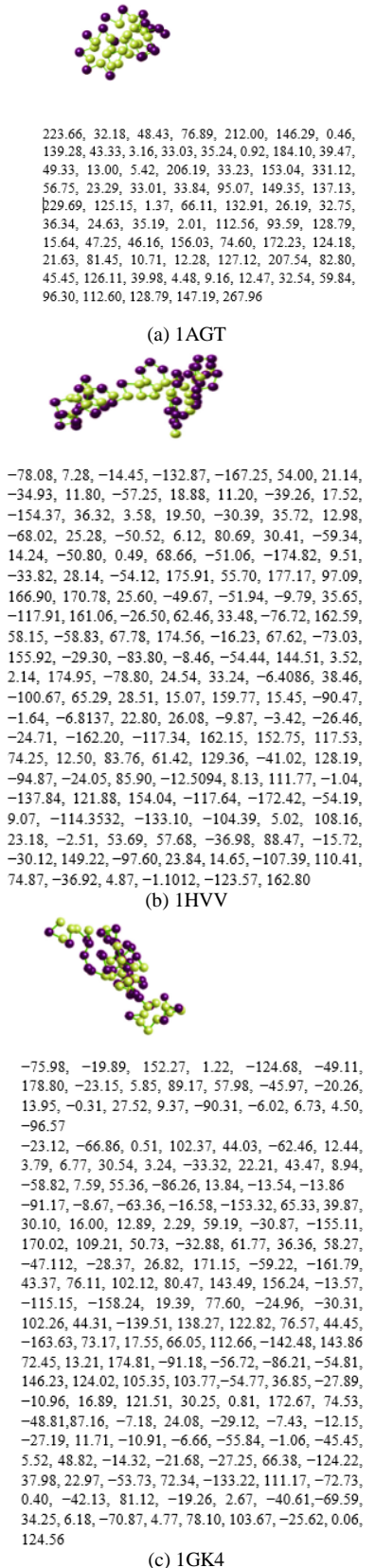


Fig. 7. Visual representations of minimum energy configurations for each protein sequence by ELAMO.

## V. CONCLUSIONS

In this paper, an Elitist Animal Migration Optimization (ELAMO) is fitted to optimize the structure of protein sequences with 3D AB off-lattice model. According to the reformed structure of ELAMO, rather than following the neighbors, only the group leaders are followed. This movement results with high solution quality and the elimination of animals during the migration process makes the algorithm not to trap in local optimum points. To enable a more accurate comparison, standard AMO and effective algorithms are included in our experiment. Even though ELAMO has not been specially designed for solving PSP problem, its effectiveness in modelling of real protein sequences is successful.

ELAMO eliminates the animals whose positions are not adequate even in the early stages of optimization and the rest of the animals in the herd only can follow their leaders. This elimination brings faster convergence rate without trapping into local optimum points in the process. It is known that the optimization is built on consecutive iterative processes until the termination criterion is obtained. During this process, some individuals may be eliminated due to their undesired characteristics which may be improved in the latter iterations. In ELAMO, the elimination is performed even at the first level of iterations and the desired characteristics are also eliminated. It is important to keep in mind that this may bring low solution quality as the length of the protein sequence increases.

Also, when the problem complexity increases in terms of large number of amino acids, the algorithm's performance may not be as efficient as it is expected. It is because of the lack of proper design for PSP problem. In the future, ELAMO may need to be strengthened by adding some boosting steps to be efficient in solving more complex protein sequences.

REFERENCES

- [1] Sahu, S. S., Panda, G. A. Novel Feature Representation Method Based on Chou's Pseudo Amino Acid Composition for Protein Structural Class Prediction. *Comput. Biol. Chem.* 2010, 34 (5–6), 320–327. <https://doi.org/10.1016/j.compbiolchem.2010.09.002>.
- [2] Dehzangi, A., Paliwal, K., Lyons, J., Sharma, A., Sattar, A. A Segmentation-Based Method to Extract Structural and Evolutionary Features for Protein Fold Recognition. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2014, 11 (3), 510–519. <https://doi.org/10.1109/tcbb.2013.2296317>.
- [3] Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science* 1973, 181 (4096), 223–230. <https://doi.org/10.1126/science.181.4096.223>.
- [4] Dill, K. A. Theory for the Folding and Stability of Globular Proteins. *Biochemistry* 1985, 24 (6), 1501–1509. <https://doi.org/10.1021/bi00327a032>.
- [5] Chen, M., Huang, W.-Q. A Branch and Bound Algorithm for the Protein Folding Problem in the HP Lattice Model. *Genomics Proteomics Bioinformatics* 2005, 3 (4), 225–230. [https://doi.org/10.1016/s1672-0229\(05\)03031-7](https://doi.org/10.1016/s1672-0229(05)03031-7).
- [6] Istrail, S., Lam, F. Combinatorial Algorithms for Protein Folding in Lattice Models: A Survey of Mathematical Results. *Commun. Inf. Syst.* 2009, 9 (4), 303–346. <https://doi.org/10.4310/cis.2009.v9.n4.a2>.
- [7] Hart, W. E., Newman, A. Protein Structure Prediction with Lattice Models. *Handb. Mol. Biol.* 2006, 1–24. <https://doi.org/10.1201/9781420036275-45>.
- [8] Stillinger, F. H., Head-Gordon, T., Hirshfeld, C. L. Toy Model for Protein Folding. *Phys. Rev. E* 1993, 48 (2), 1469. <https://doi.org/10.1103/physreve.48.1469>.
- [9] Bryant, S. H., Lawrence, C. E. An Empirical Energy Function for Threading Protein Sequence through the Folding Motif. *Proteins Struct. Funct. Bioinforma.* 1993, 16 (1), 92–112. <https://doi.org/10.1002/prot.340160110>.
- [10] Levitt, M., Warshel, A. Computer Simulation of Protein Folding. *Nature* 1975, 253 (5494), 694–698. <https://doi.org/10.1038/253694a0>.
- [11] Li, X., Zhang, J., Yin, M. Animal Migration Optimization: An Optimization Algorithm Inspired by Animal Migration Behavior. *Neural Comput. Appl.* 2014, 24, 1867–1877. <https://doi.org/10.1007/s00521-013-1433-8>.
- [12] Ülker, E. An Elitist Approach for Solving the Traveling Salesman Problem Using an Animal migration Optimization Algorithm. *Turk. J. Electr. Eng. Comput. Sci.* 2018, 26 (1), 605–617. <https://doi.org/10.3906/elk-1705-61>.
- [13] Li, B., Lin, M., Liu, Q., Li, Y., Zhou, C. Protein Folding Optimization Based on 3D Off-Lattice Model via an Improved Artificial Bee Colony Algorithm. *J. Mol. Model.* 2015, 21, 1–15. <https://doi.org/10.1007/s00894-015-2806-y>.
- [14] Kalegari, D. H., Lopes, H. S. An Improved Parallel Differential Evolution Approach for Protein Structure Prediction Using Both 2D and 3D Off-Lattice Models. In 2013 IEEE Symposium on Differential Evolution (SDE), IEEE, 2013, 143–150. <https://doi.org/10.1109/sde.2013.6601454>.
- [15] Rakhshani, H., Idoumghar, L., Lepagnot, J., Brévilliers, M. Speed up Differential Evolution for Computationally Expensive Protein Structure Prediction Problems. *Swarm Evol. Comput.* 2019, 50, 100493. <https://doi.org/10.1016/j.swevo.2019.01.009>.
- [16] Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson, A. W., Bridgland, A. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* 2020, 577 (7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>.
- [17] Schauerl, M., Denny, R.A. AI-Based protein structure prediction in drug discovery: Impacts and challenges. *J. Chem. Inf. Modell.* 2022, 62(13), 3142–3156. <https://doi.org/10.1021/acs.jcim.2c00026>.
- [18] Chowdhury, R., et al. Single-sequence protein structure prediction using a language model and deep learning. *Nature Biotech.* 2022, 40(11), 1617–1623. <https://doi.org/10.1038/s41587-022-01556-z>.
- [19] Weißenow, K., Heinzinger, M., Rost, B. Protein language-model embeddings for fast, accurate, and alignment-free protein structure prediction. *Structure*, 2022, 30(8), 1169–1177. <https://doi.org/10.1016/j.str.2022.05.001>.
- [20] Pelta, D. A., Krasnogor, N. Multimeme Algorithms Using Fuzzy Logic Based Memes for Protein Structure Prediction. *Recent Adv. Memetic Algorithms* 2005, 49–64. [https://doi.org/10.1007/3-540-32363-5\\_3](https://doi.org/10.1007/3-540-32363-5_3).
- [21] Lin, X., Zhang, X. Protein Folding Structure Optimization Based on GAPSO Algorithm in the Off-Lattice Model. In 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2014, 43–49. <https://doi.org/10.1109/bibm.2014.6999246>.
- [22] Boiani, M., Parpinelli, R. S. A GPU-Based Hybrid JDE Algorithm Applied to the 3D-AB Protein Structure Prediction. *Swarm Evol. Comput.* 2020, 58, 100711. <https://doi.org/10.1016/j.swevo.2020.100711>.
- [23] Irbäck, A., Peterson, C., Pothast, F., Sommelius, O. Local Interactions and Protein Folding: A Three-Dimensional off-Lattice Approach. *J. Chem. Phys.* 1997, 107 (1), 273–282. <https://doi.org/10.1063/1.474357>.
- [24] Kyte, J., Doolittle, R. F. A Simple Method for Displaying the Hydrophobic Character of a Protein. *J. Mol. Biol.* 1982, 157 (1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0).
- [25] Wolpert, D. H., Macready, W. G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* 1997, 1 (1), 67–82. <https://doi.org/10.1109/4235.585893>.
- [26] Hsu, H.P., Mehra, V., Grassberger, P. Structure Optimization in an Off-Lattice Protein Model. *Phys. Rev. E* 2003, 68 (3), 037703. <https://doi.org/10.1103/physreve.68.037703>.
- [27] Zhu, H., Pu, C., Lin, X., Gu, J., Zhang, S., Su, M. Protein Structure Prediction with EPSO in Toy Model. In 2009 Second International Conference on Intelligent Networks and Intelligent Systems, IEEE, 2009, pp 673–676. <https://doi.org/10.1109/icinis.2009.172>.
- [28] 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.* 2014, 27, 70–79. <https://doi.org/10.1016/j.engappai.2013.06.010>.
- [29] Bošković, B., Brest, J. Differential Evolution for Protein Folding Optimization Based on a Three-Dimensional AB off-Lattice Model. *J. Mol. Model.* 2016, 22, 1–15. <https://doi.org/10.1007/s00894-016-3104-z>.
- [30] Lin, J., Zhong, Y., Li, E., Lin, X., Zhang, H. Multi-Agent Simulated Annealing Algorithm with Parallel Adaptive Multiple Sampling for Protein Structure Prediction in AB off-Lattice Model. *Appl. Soft Comput.* 2018, 62, 491–503. <https://doi.org/10.1016/j.asoc.2017.09.037>.
- [31] Jana, N. D., Sil, J., Das, S. Selection of Appropriate Metaheuristic Algorithms for Protein Structure Prediction in AB Off-Lattice Model: A Perspective from Fitness Landscape Analysis. *Inf. Sci.* 2017, 391, 28–64. <https://doi.org/10.1016/j.ins.2017.01.020>.