# A Heterogeneous Parallel Ecologically-inspired Approach Applied to the 3D-AB off-lattice Protein Structure Prediction Problem

César M.V. Benítez
Bioinformatics Laboratory
Federal Technological University of Paraná
Curitiba, Brazil
Email: cesarbenitez@utfpr.edu.br

Rafael Stubs Parpinelli
Applied Computing Graduate Program
Santa Catarina State University
Joinville, Brazil
Email: parpinelli@joinville.udesc.br

Heitor Silvério Lopes
Bioinformatics Laboratory
Federal Technological University of Paraná
Curitiba, Brazil
Email: hslopes@utfpr.edu.br

***Abstract*—This paper applies a heterogeneous parallel ecology-inspired algorithm (pECO) to solve a complex problem from bioinformatics. The ecological-inspired algorithm represents a new perspective to develop cooperative evolutionary algorithms. Different algorithms are applied to compose the computational ecosystem in a heterogeneous model. The aim is to search low energy conformations for the Protein Structure Prediction problem, concerning the 3D-AB off-lattice model. Being a problem that demands a lot of computational effort, a parallel master-slave architecture is employed in order to allow the application of the computational ecosystem in a reasonable computing time. From the results, the pECO approach obtained the best conformation for the 13 amino-acid long sequence and competitive results for the other sequences.**

## I. Introduction

Many Bioinformatics problems are featured mainly to be non-linear and strongly constrained. This is the case of the protein structure prediction problem approached in this paper. Due to the lack of exact methods for solving such a class of problems, the need for robust heuristic methods arises. Along decades, Evolutionary Computation (EC) and Swarm Intelligence (SI) have provided a large range of flexible and robust optimization methods, capable of dealing successfully with complex optimization problems. Both EC and SI are population-based methods in which each individual of a population represents a tentative solution to the problem to be solved. In recent years several other SI algorithms have appeared [1]. With such diversity of search strategies, it is possible to establish an analogy with the dynamics of biological ecosystems.

In [2] the potentiality of some ecological concepts are illustrated presenting an ecology-inspired algorithm (ECO) for optimization. In order to explore the use of different search strategies cooperatively in an ecologically inspired context, in this work, a parallel and heterogeneous application of ECO (pECO) is developed and applied to a hard optimization problem from bioinformatics. The aim is to search low energy conformations for the Protein Structure Prediction (PSP) problem, concerning the 3D-AB off-lattice model. Being a problem that demands a lot of computational effort, a parallel master-slave architecture is employed in order to allow the application of the computational ecosystem in a reasonable computing time.

Four population-based approaches are employed cooperatively in a heterogeneous model: the Artificial Bee Colony algorithm (ABC) [3], Particle Swarm Optimization algorithm (PSO) [4], Differential Evolution (DE) [5], and a hybrid Differential Evolution / Biogeography-based Optimization algorithm (jDE-BBO) [6].

This paper is organized as follows: Section II describes the 3D-AB off-lattice PSP problem; Section III describes the ECO approach as well as its parallel version pECO; Section IV shows how the experiments were conducted; Section V shows the results obtained and the analysis made; finally, in Section VI some conclusions and future directions are pointed.

## II. Protein Structure Prediction

Proteins are the basic structures of all living beings [7]. They are composed of a chain of amino acids that are linked together by means of peptide bonds. Several amino acids exist in nature, but only 20 are proteinogenic. They can be classified into two classes, according to their affinity to water: Hydrophilic (or Polar) and Hydrophobic. According to this behavior, one can conclude that the hydropaticity of the side chain is one of the main process that governs the process of forming protein structures [8].

Protein folding is the process by which a polypeptide chain is transformed into a compact structure that performs some biological function. Its known that better understanding the protein folding process can result in important medical advancements and development of new drugs.

Computer science has an important role here, proposing models for studying the Protein Structure Prediction (PSP) problem [9]. Nowadays, the simulation of computational models that take into account all the atoms of a protein is frequently infeasible, even with the most powerful computational resources. Consequently, several simplified models that abstract the protein structure have been proposed. One of them is the AB off-lattice model. [10] employed neural networks, Monte Carlo search and biologically inspired methods using the 2D-AB off-lattice model. An extended 3D version was presented by [11]. Recently, [12] introduced an improved implementation of tabu search with the 3D-AB *off-lattice* model, using the energy function proposed by [13], obtaining

good performance. This paper is focused on the 3D-AB *off-lattice* model.

### A. The 3D-AB Off-lattice Model

The 2D AB off-lattice model was introduced by [13] to represent protein structures. In this work, we use the 3D generalization of the 2D AB off-lattice model proposed by [14]. In this model each residue is represented by a single interaction site located at the C$\alpha$ position. These sites are linked by rigid unit-length bonds ($\hat{b_i}$) to form the protein structure. The three-dimensional structure of an $N$-length protein is specified by the $N-1$ bond vectors $\hat{b_i}$, $N-2$ bond angles $\tau_i$ and $N-3$ torsional angles $\alpha_i$, as shown in Figure 1. These angles are defined in the range $[-180^o, 180^o]$.

The 20 proteinogenic amino acids are classified into two classes, according to their affinity to water (hydrophobicity): 'A' (hydrophobic) and 'B' (hydrophilic or polar). This model does not describe the solvent molecules. However, solvent effects such as the formation of the hydrophobic core are taken into account through interactions between residues, according to their hydrophobicity (species-dependent global interactions).

When a protein is folded into its native conformation, the hydrophobic amino acids tend to pack inside the protein, in such a way to get protected from the solvent by an aggregation of polar amino acids that are positioned outwards. Interactions between amino acids take place and the energy of the conformation tends to decrease. Conversely, the conformation tends to converge to its native state, in accordance with the Anfinsen's thermodynamic hypothesis [15].

The energy function of a folding using the 3D-AB Off-lattice Model is given by [14]:

$$\begin{aligned} E(\hat{b_i};\sigma) = E_{Angles} + E_{torsion} + E_{LJ} = \\ -k_1 \sum_{i=1}^{N-2} \hat{b_i} \cdot \hat{b_{i+1}} - k_2 \sum_{i=1}^{N-3} \hat{b_i} \cdot \hat{b_{i+2}} \\ + \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} 4\varepsilon(\sigma_i, \sigma_j)(r_{ij}^{-12} - r_{ij}^{-6}) \end{aligned} \tag{1}$$

where
$r_{ij}$ represents the distance between $i$th and $j$th residues; $\sigma$ = $\sigma_0$, ..., $\sigma_N$ form a binary string that represents the protein sequence.

$E_{Angles}$ and $E_{torsion}$ are the energies from bond angles and torsional forces, respectively; and are given, respectively, by Equations 2 and 3.

$$E_{Angles} = -k_1 \sum_{i=1}^{N-2} \hat{b_i} \cdot \hat{b_{i+1}} \tag{2}$$

$$E_{torsion} = -k_2 \sum_{i=1}^{N-3} \hat{b_i} \cdot \hat{b_{i+2}} \tag{3}$$

$\hat{b_i}$ represents the $i$th bond that joins the $(i-1)$th and the $i$th residues, and it is represented by the vector $\hat{b_i} = \vec{r_i} - \vec{r_{i-1}}$, and $k1 = -1$; $k2 = +1/2$.

The species-dependent global interactions are given by the Lennard-Jones potencial ($E_{LJ}$); for pairs of $i$th and $j$th residues separated by a distance of $r_{ij}$.

$$E_{LJ} = \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} 4\varepsilon(\sigma_i, \sigma_j)(r_{ij}^{-12} - r_{ij}^{-6}) \tag{4}$$

Where $\varepsilon(\sigma_i, \sigma_j)$ is chosen to favor the formation of the hydrophobic core ('A' residues). Thus, $\varepsilon(\sigma_i, \sigma_j)$ is 1 for AA interactions and 1/2 for BB/AB interactions.

### B. Encoding of Candidate Solutions

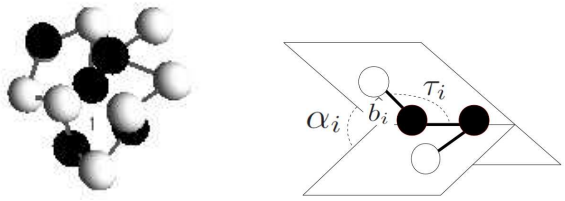
An important issue when using population-based approaches for a given problem is the encoding of the candidate solutions. The encoding has a strong influence not only in the size of the search space, but also in the complexity of the problem, due to the presence of unknown epistasis between variables that form the individuals (solution vectors). In this work, a given conformation of the protein is represented as a set of bond rotation and torsion angles over a three-dimensional space, as shown in Section II-A. Considering the folding of a protein with $n$ amino acids, an individual will represent the set of bond rotation and torsion angles. An individual has $(2n-5)$ variables, such that positions $P_1$ to $P_{n-2}$ represent the bond rotation angles, and $P_{n-1}$ to $P_{2n-5}$ represent the torsion angles.

To represent the position of the amino acids, three-dimensional Cartesian coordinates are defined by a vector ($x_i$, $y_i$, $z_i$). The first and second amino acids of the primary structure are set at the origin and point (0, 1, 0), respectively. Next amino acids are positioned at Cartesian coordinates relative to their predecessors and obtained by 3D geometrical transformations.

## III. The Ecological-Inspired Approach

The ecological-inspired algorithm, named ECO, represents a new perspective to develop cooperative evolutionary algorithms [2]. The ECO is composed by populations of individuals (candidate solutions for a problem being solved) and each population evolves according to an optimization strategy. Therefore, individuals of each population are modified according to the mechanisms of intensification and diversification, and the initial parameters, specific to each optimization strategy. The ECO system can be modelled in two ways: homogeneous or heterogeneous. A homogeneous model implies that all populations evolve in accordance to the same optimization strategy, configured with the same parameters. Any change in the strategies or parameters in at least one population characterises a heterogeneous model.

The ecological inspiration stems from the use of some ecological concepts, such as: habitats, ecological relationships and ecological successions [16]. Once dispersed in the search space, populations of individuals established in the same region constitute an ecological habitat. For instance, in a multimodal hyper-surface, each peak can become a promising habitat for some populations. A hyper-surface may have several habitats.




(a) (b)

Fig. 1. Example of a hypothetic protein structure (a) and Definition of $\hat{b_i}$, $\tau_i$ and $\alpha_i$ (b, Adapted from [14]). White balls represent the polar residues and black balls represent the hydrophobic residues. The backbone and the connections between elements are shown in black lines.

As well as in nature, populations can move around through all the environment. However, each population may belong only to one habitat at a given moment of time $t$. Therefore, by definition, the intersection between all habitats at moment $t$ is the empty set.

With the definition of habitats, two categories of ecological relationships can be defined. Intra-habitats relationships that occur between populations inside each habitat, and inter-habitats relationships that occur between habitats [16].

In ECO, the intra-habitat relationship is the mating between individuals. Populations belonging to the same habitat can establish a reproductive link between their individuals, favouring the co-evolution of the involved populations through competition for mating. Populations belonging to different habitats are called reproductively isolated.

The inter-habitats relationship are the great migrations. Individuals belonging to a given habitat can migrate to other habitats aiming at identifying promising areas for survival and mating.

In addition to the mechanisms of intensification and diversification specific to each optimization strategy, when considering the ecological context of the proposed algorithm, the intra-habitats relationships are responsible for intensifying the search and the inter-habitats relationships are responsible for diversifying the search.

Inside the ecological metaphor, the ecological successions represent the transformational process of the system. In this process, populational groups are formed (habitats), relations between populations are established and the system stabilizes by means of the self-organization of its components.

A key concept of the proposed ECO system is the definition of habitats. The ECO approach uses a hierarchical clustering algorithm to set-up the habitats where each cluster represents a habitat. Hence, the habitats are defined probabilistically taking into account the distance information returned by clustering algorithm [17].

Once the habitats have been defined, the next step is the definition of the communication topology for each habitat that is probabilistically defined.

For a habitat with more than one population, intra-habitat communication occurs in such a way that each population inside the habitat chooses another population to perform communication. Here, the distance between populations influence directly the probabilistic decision. The closer two populations are from each other the higher is the chance of these two populations communicate. The opposite happens with farthest populations.

### A. Implementation of the Parallel Computational Ecosystem

In the parallel version of ECO, named pECO, the processing load is divided into several processors (slaves), under the coordination of a master processor. Each processor (master or slaves) is responsible for initializing the population, and performing the evolutive period of an population independently. The master processor is responsible for defining the communication topologies between populations and habitats.

In Algorithm 1 the italic instructions are processed by each processor (master or slaves) while bold instructions are processed by the master processor. In the first step, line 3, each slave initializes a population $Q_i$. In line 4, the master coordinates the ecological succession loop. The first step inside the ecological succession loop is the evolutive period that is carried by the slaves (line 5). In line 6, each slave calculates the centroids and send them to the master. Then, the master creates the habitats and defines the intra-habitats communication topology, lines 7 and 8 respectively. Once defined the intra-habitat communication topologies, the master coordinates the mating process, requesting the best individual of each specie and sending it to an adjacent specie. Next, each specie replaces a randomly selected individual by a new individual, which in turn, is generated through a genetic exchange between the received individual and an individual chosen using the tournament strategy (line 9).

Once the interactions between populations are finished, the master defines the inter-habitats communication topology (line 10). The master coordinates the migration process between habitats, requesting the best individual of a population randomly chosen from each habitat and sending it to an population randomly chosen of an habitat also randomly chosen (line 11). Finally, the ecological succession loop restarts.

## IV. Computational Experiments

All experiments done in this work were run in a cluster of 25 networked computers. Each computer has an Intel Core-2 Quad processor at 2.8 GHz, 2 GBytes RAM. All computers run a minimal installation of Arch Linux and used MPICH2 [1], version 1.0 for the implementation of the message passing interface. All algorithms were implemented in ANSI-C programming language.

[1] Available in: http://www.mcs.anl.gov/research/projects/mpich2/

**Algorithm 1** Pseudo-code for parallel ECO (pECO)

```
Start
Let i = 1, ..., NQ, j = 1, ..., NH and t = 0;
Initialize each population Q_i(t) with POP_i random candidate solutions;
while stop criteria not satisfied do {Ecological succession cycles}
    Perform evolutive period for each population Q_i(t);
    Idendify the region of reference C_i for each population Q_i(t);
    define_Habitats()
    For each habitat H_j(t) define the communication topology CT_j(t) between populations Q_i^j(t);
    interactions_C()
    Define communication topology TH(t) between H_j(t) habitats;
    interactions_H()
    t ← t + 1;
end while
End
```

Due to the stochastic nature of the algorithms compared in this work, 10 independent runs were done with different initial random seeds. For each run, an upper-bound for the number of fitness function evaluations was established to $5,000,000$.

### A. Benchmark sequences

In the experiments reported below, a total of 4 synthetic protein sequences were used. These sequences have been previously used by other researchers (for instance, [11], [18]) and they were based on the Fibonacci sequence. In Table I, $N$ is the number of monomers of the sequences (13, 21, 34 and 55 amino acids-long sequences).

TABLE I. BENCHMARK SEQUENCES FOR THE 3D-AB *off-lattice* MODEL

| $N$ | Sequence |
|---|---|
| 13 | ABBABBABABBAB |
| 21 | BABABBABABBABBABABBAB |
| 34 | ABBABBABABBABBABABBABABBABBABABBAB |
| 55 | BABABBABABBABBABABBABABBABBABABBAB BABABBABABBABBABABBAB |

### B. Control Parameters

The parameters for the pECO algorithm are: number of populations (*NQ*) that will be co-evolved, the initial population size (*POP*), number of cycles for ecological successions (*ECO-STEP*), the size of the evolutive period (*EVO-STEP*) that represents the number of function evaluations in each *ECO-STEP*, and the tournament size (*T-SIZE*) used to choose solutions to perform intra and inter-habitat communications. The values for these parameters were defined empirically with: *NQ = 200*, *POP = 50*, *ECO-STEP = 5,000*, *EVO-STEP = 1,000* , and *T-SIZE = 5*. The heterogeneous model of the pECO approach, combines all four algorithms (ABC-PSO-DE-jDE/BBO) in which $1/4$ of the populations behaves according to one of these strategies.

Default parameters recommended in the literature were used in the algorithms employed. *POP* is a common parameter between all algorithms and is adjusted as previously mentioned. For ABC algorithm, there is only one control parameter, *limit = 100* [3]. For PSO algorithm, besides *POP*, the parameters were set to standard values[2]: inertia weight $W = 0.721$; cognitive and social components $\varphi_p = \varphi_g = 1.193$, respectively. For DE algorithm, the parameters are $F = 0.9$ ($F$ controls the amplification of the differential variation) and $CR = 1.0$ (crossover constant) with DE/rand/1/bin. And for jDE/BBO the parameters used are $I = E = 1.0$ (maximum possible immigration and emigration rates), $CR = 0.9$, $F = 0.5$, and $S_{max}$ =*POP* [6].

[2]Standard PSO (SPSO-07): http://www.particleswarm.info/Programs.html

## V. RESULTS AND ANALYSIS

In this section the results of our experiments are presented, as well as a comparison with the best results found in literature. The performance takes into account the average best solution found over all runs and the average processing time. Results are shown in Table II. In this table, first column represents the amino acids sequences ($N$), the second column shows the best results found in literature ($E^*$), the third column identifies the best-ever solution found by the pECO approach ($E_{best}$), the fourth column is the average value of the best solutions obtained in 10 runs ($E_{avg}$), the fifth column shows the average processing time in seconds ($t_p$), and the sixth column shows the percentage difference between $E^*$ and $E_{best}$.

TABLE II. RESULTS FOR THE 3D AB OFF-LATTICE MODEL. STRATEGY: ABC-PSO-DE-JDE/BBO

| $N$ | $E^*$ | $E_{best}$ | $E_{avg}$ | $t_p(s)$ | $diff(\%)$ |
|---|---|---|---|---|---|
| 13 | -26.507 [19] | -26.507 | -26.454 ± 0.17 | 3068.3 | 0.000 |
| 21 | -52.917 [20] | -52.029 | -51.334 ± 0.48 | 3904.5 | 1.678 |
| 34 | -97.7321 [21] | -94.35 | -91.845 ± 1.13 | 5882.9 | 3.460 |
| 55 | -173.9803 [21] | -159.726 | -157.476 ± 1.03 | 10093.45 | 8.191 |

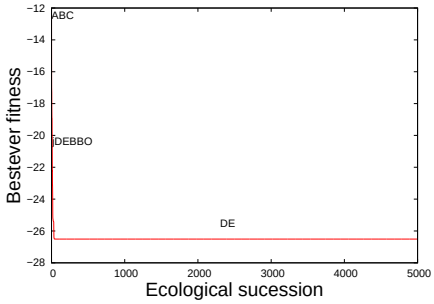
From sixth column of Table II, it is possible to notice that, for the two first sequences (13 and 21 amino acids) the differences observed are null and very small, respectively. For the two large sequences (34 and 55 amino acids) the differences are more accentuated, with 3.460% and 8.191%, respectively. Overall, the pECO approach found competitive results.

Probably, *speed-up* is the most widely used performance measure in parallel computing [22]. This measure aims at evaluating how much a parallel algorithm is faster than the equivalent sequential version. Speed-up ($s_m$) is defined as the time needed for running a given algorithm in one processor ($T_1$) divided by the running time of the same parallel algorithm, running in $m$ processors ($T_m$). In this work $m = 100$ processing cores. The speed-ups achieved were 1.8, 3.8, 4.73, and 11.93 for 13, 21, 34, and 55 amino acid-long sequences, respectively. A sublinear speed-up ($s_m < m$) behavior can be clearly identified here. Recall that a speed-up higher than one suggests that the parallelization of the algorithm decreases the overall computational cost. Ideally, the speed-up should be linear, but this is not possible in practice, since processors are not used only for processing, but also for other tasks of the underlying operating system [23] and, specially in this work, for *message-passing* communication between them. This is due to the fact that the number of species to be processed divided by the number of processors does not give a specie per processing core. It is important to recall that the speed-up can be improved including more processing cores. For instance, for $NQ$ populations, an equal number of processing cores can be used in order to achieve a quasi-linear speed-up. It is also possible to observe that the speed-up increases with the protein size. This is due to the relatively high time needed to transmit data between processes for small proteins, when compared with the processing load. Therefore, it is necessary to establish

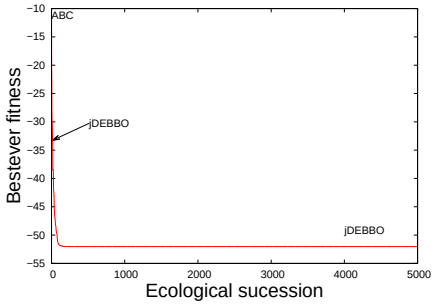
a load balance between the processing and communication loads between processes. Better speed-ups can be achieved for bigger proteins. This shows how computationally-intensive the problem is, thus justifying the parallel approach. The computed efficiency for pECO approach is 0.018, 0.038, 0.0473, and 0.1193 for 13, 21, 34, and 55 amino acid-long sequences, respectively. These values suggest that the processors are not being fully used all the time. In fact, speed-up and efficiency are a direct consequence of the balance between the processing load of the slaves and the communication load between master and slaves.

Figure 2 shows the convergence plot for all amino-acids tested. In this figure, the x-$axis$ shows the number of iterations and the y-$axis$ represents the best-ever value averaged over the same iteration of all runs. Analyzing these plots we can observe that for 13 and 21 amino-acids-long sequences the convergence is accentuated in the direction of a stagnation point in the earlier ecological successions. In the case of 13 amino acids the optimum value is reached and for 21 amino acids a solution very close to the global optimum is reached. For 34 and 55 amino acids-long sequences the convergence is not so abrupt as for the first two sequences but still with fast convergences. It is possible to verify that small improvements are achieved from half of the ecological successions forward. These convergence plot indicate that, in order to improve the results, strategies for maintaining diversity inside populations are required.
Also, Figure 2 shows for each sequence some labels indicating which algorithm achieved the best solution in each ecological succession. Once a different algorithm updates the best solution, a new label is added. For example, for the 13 amino acids-long sequence the ABC algorithm achieved the best solution until around succession 10; from successions 10 to around 2,500 the jDE/BBO algorithm achieved the best solution; and from successions 2,500 to 5,000 the DE algorithm achieved the best solution. Analysing these labels, it is possible to notice the coevolution between the different search strategies (ABC/PSO/DE/jDE-BBO) because they alternate in finding the best solutions. Possibly this is due to the peculiarity of each method in searching the space of solutions. It is also possible to conclude that, in this problem case, these strategies are quite complementary, even during few successions where one method dominates in obtaining better solutions. The convergence plots indicate that ABC and PSO algorithms are best suited for global search (initial ecological successions), whilst the DE and jDE/BBO algorithms are best suited for local search (final ecological successions).

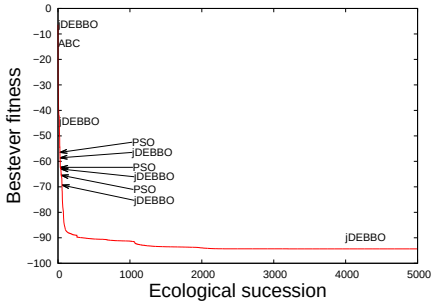
Figure 3 shows the evolution of the number of habitats for each ecological succession step. It is observed that, at the beginning of the optimization process, with the populations widely dispersed in the search space, there is a large number of habitats for all sequences. As the optimization process moves through the ecological successions, the populations tend to move through the search space converging to specific regions. As shown in Figure 3, the number of habitats decreases with the ecological succession cycles for 13 and 55 amino acids-long sequences, indicating that the populations tend to converge to points close to each other. This decay is inversely proportional to the complexity of the problem. For example, for the smallest sequence it is possible to notice a gradual convergence of populations. On the other hand, for the 55 amino acid-long sequence, due to its higher complexity, the populations are dispersed through the search space during all successions. This indicates that with more ecological successions the pECO approach could achieve even better results.



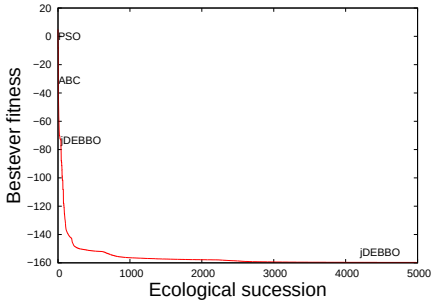

(a) Convergence for the 13 amino-acids-long sequence.




(b) Convergence for the 21 amino-acids-long sequence.

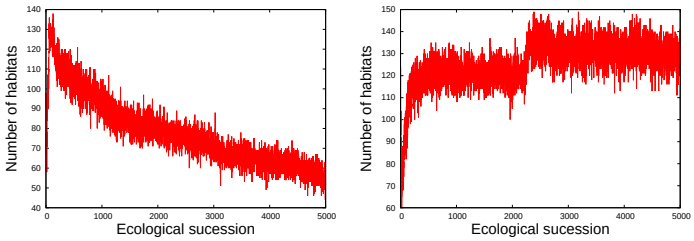



(c) Convergence for the 34 amino-acids-long sequence.




(d) Convergence for the 55 amino-acids-long sequence.

Fig. 2. Plots for the pECO convergence




(a) Convergence for the 13 amino-acids-long sequence.

(b) Convergence for the 55 amino-acids-long sequence.

Fig. 3. Plots for the number of habitats

## VI. Conclusion

The performance of a parallel ecologically-inspired optimization algorithm (pECO) was analyzed in this paper, under the task of minimizing the energy function of a Protein Structure Prediction problem, featuring the off-lattice 3D-AB model. Four population-based algorithms (ABC, PSO, DE, and jDE/BBO) were employed in an ecological heterogeneous model.

The results obtained were competitive with the ground solutions found in literature. Also, with analysis in the plot of habitats convergence, it is possible to verify that the pECO approach can achieve even better results if more ecological successions are used.

Parallel processing is essential to allow us to obtain good quality results in a reasonable computing time. As the size of the proteins increase, the computational demand will increase accordingly. Therefore, future works will consider the use of highly parallel approaches for dealing with the PSP problem, such as the use of GPGPU (General Purpose Graphics Processing Units) [24]. A hierarchical parallel approach will be explored, where the processing load related to the fitness function evaluations is also divided into several slaves which, in turn, are responsible to compute the fitness function of a number of individuals.

It is possible to highlight that the use of different algorithms can be better explored if using some source of feedback from the optimization process during its course to better distribute the habitats formation and to better define the intra and inter habitats communication topologies. This could be achieved using the heuristic information from the hierarchical clustering procedure or other KDD strategies to aid the self-organization process of the system. Also, the diversification of evolutive behaviors of the computational ecosystem, by inserting other algorithms, is a future research direction.

## References

[1] R. S. Parpinelli and H. S. Lopes, "New inspirations in swarm intelligence: a survey," *International Journal of Bio-Inspired Computation*, vol. 3, no. 1, 2011.

[2] R. S. Parpinelli and H. S. Lopes, "Biological plausibility in optimization: An ecosystemic view," *International Journal of Bio-Inspired Computation*, vol. 4, pp. 345–358, 2012.

[3] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108–132, 2009.

[4] J. Kennedy and R. Eberhart, *Swarm Intelligence*. San Francisco, CA, USA: Morgan Kaufmann, 2001.

[5] R. Storn and K. Price, "Differential evolution a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, Dec. 1997.

[6] W. Gong, Z. Cai, and C. X. Ling, "DE/BBO: a hybrid differential evolution with biogeography-based optimization for global numerical optimization," *Soft Computing*, vol. 15, no. 4, pp. 645–665, 2010.

[7] L. Hunter, *Artificial Intelligence and Molecular Biology*, 1st ed. Boston, USA: AAAI Press, 1993.

[8] H. Lodish, A. Berk, P. Matsudaira, C. Kaiser, M. Krieger, M. Scott, L. Zipursky, and J. Darnell, *Molecular Cell Biology*, 4th ed. New York, NY, USA: Freeman, 2000.

[9] H. Lopes, "Evolutionary algorithms for the protein folding problem: A review and current trends," in *Computational Intelligence in Biomedicine and Bioinformatics*, T. Smolinski, M. Milanova, and A.-E. Hassanien, Eds. Heidelberg, Germany: Springer-Verlag, 2008, vol. I, pp. 297–315.

[10] A. Torcini, R. Livi, and A. Politi, "A dynamical approach to protein folding," *Journal of Biological Physics*, vol. 27, no. 2–3, pp. 181–203, 2001.

[11] H. Hsu, V. Mehra, and P. Grassberger, "Structure optimization in an off-lattice protein model," *Physical Review E*, vol. 68, no. 3, p. id. 037703, 2003.

[12] X. Zhang and W. Cheng, "An improved tabu search algorithm for 3D protein folding problem," *Lecture Notes in Computer Science*, vol. 5351, pp. 1104–1109, 2008.

[13] F. Stillinger, T. Head-Gordon, and C. Hirshfeld, "Toy model for protein folding," *Physical Review E*, vol. 48, no. 2, pp. 1469–1477, 1993.

[14] A. Irback, C. Peterson, and F. Potthast, "Identification of amino acid sequences with good folding properties in an off-lattice model," *Physical Review E*, vol. 55, no. 1, pp. 860–867, 1997.

[15] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White, "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 47, pp. 1309–1314, September 1961.

[16] R. M. C. May and A. R. McLean, *Theoretical Ecology: Principles and Applications*. Oxford, UK: Oxford University Press, 2007.

[17] R. S. Parpinelli and H. S. Lopes, "A hierarchical clustering strategy to improve the biological plausibility of an ecology-based evolutionary algorithm," in *Ibero-American Conference on Artificial Intelligence (IBERAMIA)*, ser. Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence, PAVN, J. et al., Ed., vol. 7637, no. 1. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 310–319.

[18] D. Kalegari and H. Lopes, "A differential evolution approach for protein structure optimisation using a 2D off-lattice model," *International Journal of Bio-Inspired Computation*, vol. 2, no. 3/4, pp. 242–250, 2010.

[19] F. Liang, "Annealing contour monte carlo algorithm for structure optimization in an off-lattice protein model," *Chemical Physics*, vol. 120, pp. 6756–6763, 2004.

[20] M. Bachmann, H. Arkm, and W. Janke, "Multicanonical study of coarse-grained off-lattice models for folding heteropolymers," *Physical Review E*, vol. 71, pp. 1–11, 2005.

[21] S. Kim, S. Lee, and J. Lee, "Structure optimization by conformational space annealing in an off-lattice protein model," *Physical Review E*, vol. 72, pp. 1–6, 2005.

[22] H. Alba, *Parallel Metaheuristics: A New Class of Algorithms*. Wiley-Interscience, 2005.

[23] S. Roosta, *Parallel Processing and Parallel Algorithms: Theory and Computation*. New York, NJ: Springer-Verlag, 1999.

[24] C. Benítez, M.Scalabrin, H. Lopes, and C. Lima, "Reconfigurable hardware computing for accelerating protein folding simulations using the harmony search algorithm and the 3d-hp-side chain model," in *International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP)*, ser. Lecture Notes in Computer Science. Heidelberg: Springer, 2011, pp. 363–374.