An Ecology-based Evolutionary Algorithm Applied to the 2D-AB off-lattice Protein Structure Prediction Problem

Rafael Stubs Parpinelli
Department of Computer Science
Santa Catarina State University, UDESC
Joinville, Brazil
Email: parpinelli@joinville.udesc.br

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

***Abstract*—This paper applies an ecology-inspired algorithm (ECO) 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, both homogeneously and heterogeneously. The aim is to search low energy conformations for the Protein Structure Prediction problem, concerning the 2D-AB off-lattice model. From the results, the heterogeneous configuration obtained the best conformations for almost all cases, possibly due to the use of different intensification and diversification strategies provided by different search algorithms.**

***Keywords*- ecology-inspired approach; cooperative search; protein structure prediction; 2D-AB model;**

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 limitations of exact methods for solving such a class of problems, the need for more robust techniques 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 provide population-based methods where each individual of a population represents a tentative solution to the problem to be solved. With such diversity of search strategies [1], it is possible to establish an analogy with the dynamics of biological ecosystems.

In [2], [3] the authors illustrate the potentiality of such ecological concepts presenting an ecology-inspired algorithm (ECO) for optimization. Aiming at applying the ECO approach to a more complex problem, the purpose of this paper is to search low energy conformations for the Protein Structure Prediction problem, concerning the AB off-lattice model.

The ECO algorithm is applied using five configurations. Four configurations employ the ECO algorithm homogeneously: ECO$_{ABC}$, employs the Artificial Bee Colony (ABC) algorithm [4]; ECO$_{PSO}$, employs the Particle Swarm Optimization (PSO) algorithm [5]; ECO$_{DE}$, employs the Differential Evolution algorithm [6]; and, ECO$_{jDE/BBO}$, employs the Biogeography-based Optimization algorithm hybrid with DE [7]. One configuration, ECO$_{All}$, employs a heterogeneous approach combining all four algorithms in which $1/4$ of the populations behaves according to one of these strategies.

Also, the overall best solutions are compared with other results found in literature.

II. The Ecological-Inspired Approach

The ecological-inspired algorithm, named ECO, represents a new perspective to develop cooperative evolutionary algorithms [8]. 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 modeled in two ways: homogeneous or heterogeneous. An 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 characterizes an heterogeneous model.

The ecological inspiration stems from the use of some ecological concepts, such as: habitats, ecological relationships and ecological successions [9][10]. 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. 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

978-0-7695-5092-3/13 $26.00 © 2013 IEEE
DOI 10.1109/BRACIS.2013.19

that occur between populations inside each habitat, and inter-habitats relationships that occur between habitats [9][10].

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.

Algorithm 1 shows the pseudo-code of the proposed approach. In line 1, NQ and NH refer to the number of populations and the number of habitats, respectively. In this algorithm, the ecological succession loop (lines 3 to 12) refers to iterations of the computational ecosystem. In line 4, evolutive period, each population evolves (generations/iterations) according to its own criteria. The metric chosen to define the region of reference is the centroid and represents the point in the space where there is a longest concentration of individuals of population i. For a detailed description refer to [11].

A key concept of the proposed ECO system is the definition of habitats (line 6 in Algorithm 1). 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 [2].

Once the habitats have been defined, the next step in Algorithm 1 (line 7) is the definition of the communication topology for each habitat that is probabilistically defined.

Algorithm 1 Pseudo-code for ECO

1: Consider $i = 1, \ldots, NQ$, $j = 1, \ldots, NH$ and $t = 0$;
2: Initialize each population Q_i^t with n_i random candidate solutions;
3: **while** stop criteria not satisfied **do** {Ecological succession cycles}
4: Perform evolutive period for each population Q_i^t;
5: Apply metric C_i to identify the region of reference for each population Q_i^t;
6: Using the C_i values, define the NH habitats;
7: For each habitat H_j^t define the communication topology CT_j^t between populations Q_{ij}^t;
8: For each topology CT_j^t, perform interactions between populations Q_{ij}^t;
9: Define communication topology TH^t between H_j^t habitats;
10: For TH^t topology, perform interactions between H_j^t habitats;
11: Increase t;
12: **end while**

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.

III. Protein Structure Prediction Problem

Proteins are the basic structures of all living beings [12]. 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 processes that governs the process of forming protein structures [13].

Protein folding is the process by which a polypeptide chain is transformed into a compact structure that performs some biological function. It is 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 [14]. 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.

A. The AB off-lattice model

The AB off-lattice model was one of the first non-lattice model to represent protein structures. In this model the protein sequences are composed of only two species of monomers (ξ): 'A' for hydrophobic amino acids and 'B' for hydrophilic (or polar) amino acids. Although it is a very simplified representation of a real protein structure, this model is useful to verify some of the properties of proteins in the real world.

Monomers have an unit length of distance between them, in such a way that a monomer is connected to the next one in the chain through a bond that forms an angle relative to its predecessor.

In the AB model, a protein composed of N-monomers needs $N - 2$ angles to be represented. These angles are defined in the range $[-\pi,\ \pi]$. Figure 1 shows an example of a hypothetic protein with seven amino acids.

The model defines the energy values for the monomers: 'A' has energy 1 and 'B' has energy -1. Considering two

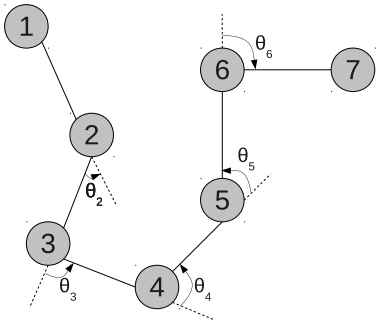


Figure 1. Generic representation of a hypothetic protein structure.

generic monomers i and j, and the types ξ_i and ξ_j, respectively, the interaction between the monomers leads to different values of potential energy (C). Positive values represent attraction and negative, repulsion: AA bonds have energy 1 (the monomers AA tend to attract each other strongly), BB bonds have energy 1/2 (they tend to attract each other weakly) and AB or BA bonds have energy -1/2 (they have a weak repulsion). The energy E of the structure of a protein with n monomers (n-mers) is given by Equation 1:

$$E(\vec{\theta},\vec{\xi}) = \sum_{i=1}^{n-1} V_1(\theta_i) + \sum_{i=1}^{n-2}\sum_{j=i+2}^{n} V_2(d_{ij},\xi_i,\xi_j) \quad (1)$$

Equation 1 postulates two types of intermolecular potential energies, terms V_1 and V_2. The former represents the backbone potentials. It is defined by Equation 2 and depends only on the angle between monomers. The latter, defined by Equation 3, represents the potential energy present in the non-bonded interactions and it is known as the Lennard-Jones potential.

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

$$V_2(d_{ij},\xi_i,\xi_j) = 4 \times (d_{ij}^{-12} - C(\xi_i,\xi_j) \times d_{ij}^{-6}) \quad (3)$$

where

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

Equation 4 is the potential energy due to the interaction between monomers i and j, and d_{ij} is the distance between these monomers in the chain, such that $i < j$.

A total of 5 synthetic protein sequences are used. These sequences have been previously used by other researchers [15], [16]. In Table I, N is the number of monomers of each sequence (13, 21, 34 and 55 amino acids-long sequences) and it is followed by the sequence itself.

Table I
BENCHMARK SEQUENCES FOR THE 2D-AB OFF-LATTICE MODEL

N	Sequence
13	ABBABBABABBAB
21	BABABBABABBABBABABBAB
34	ABBABBABABBABBABABBABABBABBABABBAB
55	BABABBABABBABBABABBABABBABBABABBAB BABABBABABBABBABABBAB

IV. COMPUTATIONAL EXPERIMENTS

The experiments are conducted using the benchmark sequences shown in Section III, Table I. All experiments reported are run in a computer with an Intel processor (Core2-Quad running at 2.8GHz) running Linux. The applications are developed using the C programming language.

Owing to the stochastic nature of the proposed approach and other meta-heuristic algorithms, their performance cannot be evaluated by the result of a single run. Many trials with independent population initializations should be made to obtain an useful conclusion. Therefore, in this study the results are obtained in 30 trials.

The ecological-inspired algorithm (ECO) was tested using five configurations. All configurations implement the Algorithm 1 as described in Section II. The first configuration, ECO$_{ABC}$, employs the ABC algorithm homogeneously. The second configuration, ECO$_{PSO}$, employs the PSO algorithm homogeneously. The third configuration, ECO$_{DE}$, employs the DE algorithm homogeneously. The fourth configuration, ECO$_{jDE/BBO}$, employs the BBO algorithm hybrid with DE, also homogeneously. The fifth configuration, ECO$_{All}$, employs a heterogeneous approach combining all four algorithms in with $1/4$ of the number of populations behaves according to one of these strategies.

Also, the overall best solutions are compared with other results found in literature.

The parameters for the ECO algorithm are: number of populations (*N-POP*) that will be co-evolved, the initial population size (*POP-SIZE*), number of cycles for ecological successions (*ECO-STEP*), the size of the evolutive period (*EVO-STEP*) that represents number of function evaluations in each *ECO-STEP*, and the tournament size (*T-SIZE*).

The parameters used were *POP-SIZE* = 40, *N-POP* = 200, *ECO-STEP* = 6250, *EVO-STEP* = 800, and *T-SIZE* = 5. The number of dimensions D is 11, 19, 32, and 53 for the sequences of size 13, 21, 34, and 55, respectively. With this configuration, the total number of function evaluations was 5,000,000 for each population. Studies about the adjustment of parameters have not been carried out yet. Hence, all the parameters of the algorithm were defined empirically [11]. Default parameters recommended in the literature were used in the employed algorithms.

V. Results and Analysis

Table II shows the results obtained for the benchmark sequences using the ecological approaches. In this table, the first column identifies the sizes N of amino acids sequences and the remaining columns show the average and standard deviation obtained by each configuration followed by its best result. In bold the best result obtained for each sequence are shown.

With a close look in Table II, considering the 13 and 21 amino acids sequence, it is possible to notice that the ECO_{All} approach achieved the best results. For 34 and 55 amino acids sequences, both $\text{ECO}_{jDE/BBO}$ and ECO_{All} achieved equivalent results.

When using the heterogeneous model (ECO_{All}) for sequences 13, 21, and 34, the optimization process took advantage of using different intensification and diversification strategies provided by different search algorithms. However, although ECO_{All} has had the best average result for the largest sequence of 55 amino acids, the $\text{ECO}_{jDE/BBO}$ achieved the best result. Overall, the best results are obtained by $\text{ECO}_{jDE/BBO}$ and ECO_{All} approaches.

Although the results of using different search strategies have already shown to be promising, with this observation we realized that the use of different algorithms can be better explored if using some source of feedback from the optimization process during its course. The main concern is to use this heuristic information to better distribute the habitats formation and to better define the intra and inter habitats communication topologies. We have a strong insight that this can be done using the information contained in the hierarchical clustering procedure. We believe that biasing the dendrogram weights it should be possible to achieve different probabilities for habitats formation and communication topologies definition during the optimization process. This analysis is quoted as future research.

Table III shows the mean of the elapsed time in hours by each configuration for each sequence, followed by the respective standard deviations. The processing time employed by the ECO_{All} approach is around the average elapsed time employed by the other strategies. Also, the times shown in this table encourages the use of massive parallel strategies inside the ECO framework (e.g., using cluster of computers or Graphical Processing Units).

Table IV shows the lowest energies obtained by the best ECO approaches along with the lowest energies obtained by other works using different methods. In this table, E_{PERM} is a Pruned-Enriched Rosenbluth Method [15]; E_{min} is the minimum energy obtained by the same method with subsequent conjugate gradient minimisation [15]; E_{ground} is the putative ground state energy obtained by Stillinger and Head-Gordon using a Monte Carlo method hybridised with Newtonian conjugate gradient minimisation [17]; PBHS is the best result obtained by a Population-Based Harmony Search algorithm [18]; DE-RI is the best result obtained by a Differential Evolutionary algorithm with a ring-island configuration [16]; DEadp is the lowest energy obtained using a Differential Evolution algorithm with self-adaptation of the F parameter and with other improvements [19]; ACMC is the lowest energy obtained using an Annealing Contour Monte Carlo Method [20]; CSA shows the energy using a Conformational Space Annealing approach [21]; ELP_+ is the lowest energy obtained using the improved energy landscaping paving method [22]. In bold the best results found in literature are shown.

One of the most challenging tasks when comparing different algorithms if to perform a fair comparison between them. To the best of our knowledge, a good base for comparison is to compare the results obtained by the algorithms using the same computational effort (e.g., processing time or number of function evaluations). Here the function evaluations criteria was used. Thus, it is worth remembering that for all ECO approaches the number of function evaluations was set to five millions (Section IV). That is the same number of function evaluations used for PBHS and DE-RI algorithms. For DEadp, the number of function evaluations was set to 35 millions that is a lot more than what was applied in the experiments. For the other methods no other information about the computational effort employed to obtain the results was found. Also, for all ECO approaches, all algorithms employed are in its canonical versions.

From Table IV it can be seen that the results obtained by the proposed approach are better than those of the E_{PERM}, E_{ground}, PBHS, and DE-RI for all four sequences, with the energy difference increasing gradually for longer chains. For sequences with length 13 and 21, the results obtained by the eco-approach were slightly better than that of E_{min} and competitive with the ACMC, CSA, and ELP_+ results. For other cases, 34 and 55 length sequences, however, the proposed approach cannot reach the energy yielded by E_{min}, ACMC, CSA and ELP_+.

The closest approach to the proposed method is the DE-RI, which employs a Differential Evolutionary algorithm with a ring-island configuration (sixth column of Table IV). From the results, it can be noticed that the eco-approach works very well when compared with the DE-RI approach.

Table V shows a comparison of the best solutions found in literature (E^*) with the best ECO solutions (ECO_{Best}). The fourth column ($\text{Diff}_{ECO_{Best}xE^*}$) shows the percent difference between E^* and ECO_{Best}. For the first sequence the difference observed is minimal, almost null. For the sequence of 21 amino acids the difference is null. For the two large sequences the differences are more accentuated, with 4.3494% and 12.4986%, respectively.

To evaluate visually the quality of the foldings produced by the ECO approaches, the best results shown in Table IV were used to draw the planar form of the sequence (conformation). A program in MATLAB was developed to

Table II
QUALITY OF SOLUTIONS USING ECO APPROACHES.

	ECO_{ABC}		ECO_{PSO}		ECO_{DE}	
N	Avg	Best	Avg	Best	Avg	Best
13	-3.1987±0.0010	-3.1990	-3.1990±0.0	-3.1990	-3.1990±0.0	-3.1990
21	-5.3743 ±0.5065	-6.1747	-5.1850 ±0.3569	-5.5056	-5.4402±0.0932	-5.5205
34	-8.2718 ±0.5404	-9.6805	-8.7419±0.4610	-9.8114	-7.8561±0.3671	-8.5590
55	-12.7603±0.5019	-13.3262	-13.5588±0.3304	-13.9440	-11.9394±0.8278	-13.4844

	$ECO_{jDE/BBO}$		ECO_{All}	
N	Avg	Best	Avg	Best
13	-3.1990±0.0	-3.1990	-3.2352±0.0356	**-3.2940**
21	-5.1049±0.4205	-5.5056	-6.1980±0.0	**-6.1980**
34	-9.8464±0.4861	**-10.3360**	-9.7185±0.5121	**-10.3360**
55	-14.9310±0.9884	**-16.5641**	-15.1982±0.5089	-15.8887

Table III
MEAN OF THE PROCESSING TIME OVER ALL RUNS (HOURS).

N	ECO_{ABC}	ECO_{PSO}	ECO_{DE}	$ECO_{jDE/BBO}$	ECO_{All}	ECO_{All-LM}
13	1.45±0.00	3.42±0.01	2.00±0.01	5.36±0.02	2.40±0.00	2.49±0.00
21	3.82±0.00	5.34±0.02	4.22±0.03	6.83±0.05	4.40±0.00	4.62±0.01
34	10.15±0.01	11.59±0.02	10.73±0.05	15.01±0.09	10.74±0.08	11.09±0.00
55	26.95±0.02	29.23±0.02	27.82±0.02	32.09±0.18	27.80±0.23	28.81±0.01

Table IV
COMPARATIVE OF BEST SOLUTIONS FROM DIFFERENT STRATEGIES WITH BEST ECO SOLUTIONS (ECO_{Best}).

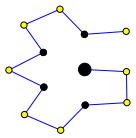
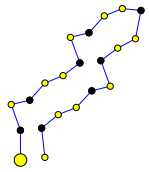
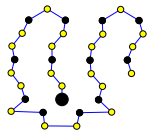
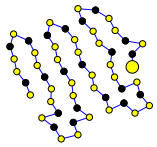
N	E_{PERM}	E_{min}	E_{ground}	PBHS	DE-RI	DEadp	ACMC	CSA	ELP_{+}	ECO_{Best}
13	-3.2167	-3.2939	-3.2235	-3.28	-3.2924	-3.1999	**-3.2941**	**-3.2941**	**-3.2941**	-3.2940
21	-5.7501	-6.1976	-5.2881	-5.96	-6.1979	**-6.1980**	-6,1979	**-6.1980**	**-6.1980**	**-6.1980**
34	-9.2195	-10.7001	-8.9749	-8.33	-9.6838	-10.5565	**-10.8060**	**-10.8060**	-10.7453	-10.3360
55	-14.9050	-18.5154	-14.4089	-11.51	-14.6847	-17.3133	-18.7407	-18.9110	**-18.9301**	-16.5641

Table V
COMPARATIVE OF OVERALL BEST SOLUTIONS FOUND IN LITERATURE (E*) WITH BEST ECO SOLUTIONS (ECO_{Best}).

N	E*	ECO_{Best}	$Diff_{ECO_{Best}xE^*}$
13	-3.2941	-3.2940	0.0018%
21	-6.1980	-6.1980	0.0000%
34	-10.8060	-10.3360	4.3494%
55	-18.9301	-16.5641	12.4986%

convert the string of angles into (x, y) coordinates and plot the structure. The larger dot represents the start of the sequence, black dots represent 'A' monomers and the yellow dots represent 'B' monomers. Recall that the energy of the folding is a function of the proximity of monomers, especially the 'A' monomers. Therefore, compact structures tend to have lower energy levels than those structures more dispersed. Table VI shows the best foldings obtained with the ECO implementations for sequences with 13, 21, 34, and 55 monomers. From Table VI it is possible to see that the hydrophobic A monomers tend to form a hydrophobic core in the 13 amino acids sequence or clusters of typically 3-5 monomers in other sequences. This can be explained by the fact that hydrophobic monomers are always flanked by the hydrophilic monomers along the sequence. This shows that the AB off-lattice model reflects the native characters of the real proteins in two-dimensions but it still is not perfect.

Table VI
BEST STATES FOUND BY ECO APPROACHES LISTED IN TABLE IV.

N = 13	N = 21
	
N = 34	N = 55
	

VI. CONCLUSIONS

This work applied an ecology-based approach to the PSP AB off-lattice model. The proposed algorithm uses cooperative search strategies where populations of individuals co-evolve and interact among themselves using some ecological concepts. Each population behaves according to

the mechanisms of intensification and diversification, and the control parameters, specific to a given search strategy. Six configurations of the ECO approach were employed and, when using the heterogeneous model (ECO_{All}) the search process gets more robust than the other approaches (e.g., for sequences 13, 21, and 34) possibly due the use of different intensification and diversification strategies provided by different search algorithms. Also, ECO_{All} has had the best average result for the large sequence of 55 amino acids.

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. This analysis is pointed as future research.

This work is still under development and as future work we intend to analyze the influence of the remaining control parameters (number of ecological successions, evolutive period, and number of populations) on the quality of solutions, as well as to add other search strategies in the proposed model. Currently, in order to bring more biological plausibility to the system, other ecological concepts are being modeled, and efforts are being done to eliminate control parameters. Also, the use of massive parallel strategies inside the ECO framework is a future trend.

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, pp. 1–16, 2011.

[2] ——, "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, J. P. et al., Ed., vol. 7637, no. 1. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 310–319.

[3] ——, "An ecology-based heterogeneous approach for cooperative search," in *Brazilian Symposium on Artificial Intelligence (SBIA)*, ser. Lecture Notes in Computer Science, L. Barros and M. F. andA.T.R. Pozo, Eds., vol. 7589, no. 1. Berlin, Heidelberg: Springer, 2012, pp. 212–221.

[4] 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.

[5] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization: an overview," *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, Jun. 2007.

[6] 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.

[7] 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.

[8] 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.

[9] M. Begon, C. R. Townsend, and J. L. Harper, *Ecology: from individuals to ecosystems*, 4th ed. Oxford, UK: Blackwell Publishing, 2006.

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

[11] R. S. Parpinelli and H. S. Lopes, "An eco-inspired evolutionary algorithm applied to numerical optimization," in *Proceedings of the Third World Congress on Nature and Biologically Inspired Computing*. Salamanca, Spain: IEEE Press, 2011, pp. 473–478.

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

[13] 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.

[14] 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. Berlin, Heidelberg: Springer-Verlag, 2008, vol. 1, no. 1, pp. 297–315.

[15] H. Hsu, V. Mehra, and P. Grassberger, "Structure optimization in an off-lattice protein model," *Physical Review E*, vol. 68, no. 2, pp. 1–4, 2003.

[16] D. H. Kalegari and H. S. 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.

[17] F. Stillinger and T. Head-Gordon, "Collective aspects of protein folding illustrated by a toy model," *Physical Review E*, vol. 52, no. 3, pp. 2872–2877, 1995.

[18] M. H. Scalabrin, R. S. Parpinelli, C. M. Benitez, and H. S. Lopes, "Population-based harmony search using gpu applied to protein structure prediction," *International Journal of Computational Science and Engineering*, 2012, to appear.

[19] D. H. Kalegari, "Algoritmo de evolução diferencial paralelo aplicado ao problema da predição da estrutura de proteínas utilizando o modelo ab em 2d e 3d," Universidade Tecnológica Federal do Paraná, 2010.

[20] F. Liang, "Annealing contour monte carlo algorithm for structure optimization in an off-lattice protein model," *Journal of Chemical Physics*, vol. 120, no. 14, pp. 6756–63, 2004.

[21] S.-Y. Kim, S. B. Lee, and J. Lee, "Structure optimization by conformational space annealing in an off-lattice protein model," *Physical Review E*, vol. 72, no. 1, p. 011916, 2005.

[22] J. Liu, S. Xue, D. Chen, H. Geng, and Z. Liu, "Structure optimization of the two-dimensional off-lattice hydrophobic-hydrophilic model," *Journal of Biological Physics*, vol. 35, no. 3, pp. 245–253, 2009.