Locality-based Multiobjectivization for the HP Model of Protein Structure Prediction

Mario Garza-Fabre Gregorio Toscano-Pulido Eduardo Rodriguez-Tello

Information Technology Laboratory, CINVESTAV-Tamaulipas. Parque Científico y Tecnológico TECNOTAM
Km. 5.5 carretera Cd. Victoria-Soto La Marina, Cd. Victoria, Tamaulipas 87130, MÉXICO
{mgarza, gtoscano, ertello}@tamps.cinvestav.mx

ABSTRACT

Even under the rather simplified HP lattice model, protein structure prediction remains a challenging problem in combinatorial optimization. Recently, the multiobjectivization of this problem was proposed. By decomposing the original objective function, a two-objective formulation for the HP model was defined. Such an alternative formulation showed very promising results, leading to an increased search performance in most of the conducted experiments. This paper introduces a novel multiobjectivization for the HP model which is based on the locality notion of amino acid interactions. Using different evolutionary algorithms, this proposal was compared with respect to both the conventional single-objective formulation and the previously reported multiobjectivization. The new proposed formulation scored the best results in most of the cases. Statistical significance testing and a large set of test cases support the findings of this study. Results are provided for both the two-dimensional square lattice and the three-dimensional cubic lattice.

Categories and Subject Descriptors

I.2.8 [**ARTIFICIAL INTELLIGENCE**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

Keywords

Multiobjectivization, protein structure prediction, HP model

1. INTRODUCTION

Proteins are fundamental elements of living organisms. These chain-like molecules are composed from a set of 20 different building blocks called amino acids. The specific sequence of amino acids determines how proteins fold into unique three-dimensional structures defining their biological functions [1]. The *protein structure prediction* problem, PSP, is the problem of finding the native (energy-minimizing) conformation for a protein given only its amino acid sequence.

GECCO'12, July 7–11, 2012, Philadelphia, Pennsylvania, USA.

The *hydrophobic-polar* (HP) *model* [12] is an abstraction of the PSP. This model captures the fact that the hydrophobicity of amino acids is one of the main driving forces determining the functional conformation of proteins. The prediction of protein structures using the HP model is a hard combinatorial optimization problem [3, 7]. Such a complexity has motivated the use of evolutionary algorithms and a variety of other metaheuristics to address this problem [25, 35].

Multiobjectivization refers to the process of reformulating a single-objective optimization problem as a multiobjective one [23]. This transformation has been successfully used to deal with difficult problems, such as the PSP [2, 8, 10, 16, 32]. However, it was not until recently that this concept was applied to the particular HP model of this problem for the first time [14]. In [14], the conventional energy (objective) function of the HP model was decomposed into two separate objectives based on the parity of amino acid positions in the protein sequence. This approach was named the *parity decomposition* (PD). Experimental results reported in [14] indicate that an important improvement in the search performance can be obtained by using such an alternative formulation, motivating further research in this direction.

In this paper, an improved multiobjectivization strategy for the HP model is proposed: the *locality decomposition* (LD). In LD, the decomposition of the HP model's energy function is carried out by segregating local from nonlocal amino acid interactions. This locality notion is based on the sequence distance between the interacting amino acids. The suitability of the proposed LD is investigated by comparing it with respect to both the conventional single-objective formulation and the preceding PD multiobjectivization [14].

The remainder of this document is organized as follows. Background concepts and notation are covered in Section 2. Section 3 summarizes related work. In Section 4, the new proposed formulation is described. Section 5 details the implemented algorithms, test cases and the performance assessment methodology. The results are presented in Section 6. Finally, Section 7 provides the conclusions of this study.

2. BACKGROUND AND NOTATION

2.1 The Hydrophobic-Polar (HP) Model

Amino acids can be classified either as *hydrophobic* (H) or *polar* (P) on the basis of their affinity for water. While the H amino acids tend to clump together on the inside of proteins, the P ones are usually found at the outer surface

interacting with the aqueous environment. Hydrophobicity is, therefore, a dominant force in the protein folding process.

In the HP model [12], proteins are abstracted as chains of H- and P-type beads. Protein sequences, which are originally defined over a 20-letters alphabet, are thus of the form $S \in \{H, P\}^L$, where L is the number of amino acids. Valid protein conformations are modeled as *Self-Avoiding Walks* of the HP chain on a lattice. That is, each lattice node can be assigned to at most one amino acid and consecutive amino acids in S are to be also adjacent in the lattice.

By emulating the hydrophobic effect, the HP model aims to maximize the interaction among H amino acids in the lattice. Formally, protein structure prediction under the HP model is defined as the problem of finding $c^* \in C$ such that $E(c^*) = \min\{E(c) \mid c \in C\}$, being C the set of all valid conformations. $E(c)$ denotes the energy of conformation c:

$$E(c) \quad = \quad \sum_{s_i, s_j \in S} e(s_i, s_j) \qquad (1)$$

where $e(s_i, s_j) = -1$ if s_i and s_j form a *hydrophobic topological contact*, denoted by $htc(s_i, s_j)$. Otherwise, $e(s_i, s_j) = 0$. A hydrophobic topological contact occurs when two H amino acids $s_i, s_j \in S$ are nonconsecutive in S but adjacent in the lattice. An example conformation for an HP chain of length $L = 20$ on the square lattice is shown in Figure 1.

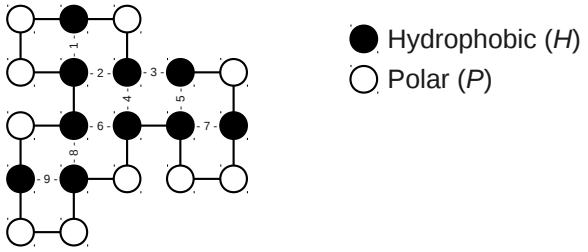

Figure 1: Hydrophobic topological contacts are numbered. The energy of this structure is $E(c) = -9$.

2.2 Single- and Multiobjective Optimization

A *single-objective optimization problem* can be stated as the problem of minimizing an *objective function* $f : \mathcal{F} \to \mathbb{R}$, where $\mathcal{F}$ is the set of all feasible solutions. The aim is to find the solution(s) $x^* \in \mathcal{F}$ such that $f(x^*) = \min\{f(x) \mid x \in \mathcal{F}\}$.

Similarly, a *multiobjective optimization problem* can be defined as the problem of minimizing an *objective vector* $\mathbf{f}(x) = [f_1(x), f_2(x), \ldots, f_k(x)]^T$, where $f_i : \mathcal{F} \to \mathbb{R}$ is the i-th objective function, $i \in \{1, 2, \ldots, k\}$. The goal is to find a set of *Pareto-optimal solutions* $\mathcal{P}^* \subset \mathcal{F}$, such that $\mathcal{P}^* = \{x^* \in \mathcal{F} \mid \nexists x \in \mathcal{F} : x \prec x^*\}$. The symbol "$\prec$" denotes the *Pareto-dominance* relation, which is given by:

$$x \prec y \quad \Leftrightarrow \quad \forall i \in \{1, 2, \ldots, k\} : f_i(x) \leq f_i(y) \quad \wedge \quad \exists j \in \{1, 2, \ldots, k\} : f_j(x) < f_j(y) \qquad (2)$$

If $x \prec y$, then x is said to *dominate* y. Otherwise ($x \not\prec y$), y is said to be *nondominated* with respect to x. The image of $\mathcal{P}^*$ in the objective space is called the *Pareto-optimal front*.

2.3 Multiobjectivization

Multiobjectivization concerns the reformulation of single-objective optimization problems in terms of two or more objective functions [23]. This can be done either by adding *supplementary* (also called artificial or helper) objectives [4, 21], or through the *decomposition* of the original objective function [17, 23]. In either case, multiobjectivization introduces fundamental changes in the search landscape, usually leading algorithms to perform a more efficient exploration. However, the goal remains to solve the original problem, so that the original optima are to be also Pareto-optimal with regard to the multiobjectivized version of the problem.

This work is based on the decomposition approach. A single-objective problem, with a given objective function $f : \mathcal{F} \to \mathbb{R}$, is restated in terms of $k \geq 2$ objectives $f_i : \mathcal{F} \to \mathbb{R}, i \in \{1, 2, \ldots, k\}$ such that $f(x) = \sum_{i=1}^{k} f_i(x)$, for all $x \in \mathcal{F}$. As the only possible effect [17], plateaus may be defined in the search landscape. That is, originally comparable solutions may become incomparable (mutually nondominated) with regard to the decomposed formulation. Multiobjectivization by decomposition has been proven to be effective as a means of escaping from local optima [17, 23].

3. RELATED WORK

There has been a great deal of research on the use of metaheuristics to solve the HP model of the PSP. This includes genetic algorithms [18, 33], memetic and hybrid algorithms [6, 19], immune-based algorithms [9], ant colony optimization [31], particle swarm optimization [5], differential evolution [28] and estimation of distribution algorithms [27]. Some of the literature in this regard is reviewed in [25, 35].

Multiobjectivization has been successfully applied in order to solve difficult optimization problems. Among them, there can be mentioned well-known combinatorial problems such as the traveling salesman problem [20, 21, 23], job-shop scheduling [21, 24] and bin packing problems [30], as well as important problems in the fields of mobile communications [29], computational mechanics [15] and computer vision [34]. Multiobjectivization has also been proposed for the PSP [2, 8, 10, 16, 32]. However, it was not until recently that the first multiobjectivized formulation for the particular HP model of this problem was reported [14]. Such an HP model's formulation is briefly described in Section 3.1.

3.1 The Parity Decomposition

In the two-dimensional square and the three-dimensional cubic lattices, topological contacts are only possible between amino acids whose sequence positions are of opposite parity. Based on this fact and following the multiobjectivization by decomposition approach, Garza-Fabre *et al.* [14] proposed a two-objective formulation, $\mathbf{f}(c) = [f_1(c), f_2(c)]^T$, for $c \in C$:

$$f_1(c) = \sum_{s_i, s_j \in S} e(s_i, s_j) \qquad \text{for } \boldsymbol{i \equiv 0 \pmod 2}, i < j \qquad (3)$$

$$f_2(c) = \sum_{s_i, s_j \in S} e(s_i, s_j) \qquad \text{for } \boldsymbol{i \equiv 1 \pmod 2}, i < j \qquad (4)$$

where both $f_1(c)$ and $f_2(c)$ are to be minimized and $e(s_i, s_j)$ was defined in Section 2.1. Function f_1 accounts only for hydrophobic topological contacts $htc(s_i, s_j)$ where i, the sequence position of amino acid s_i, is even. On the contrary, f_2 is defined for those cases where such the i-th sequence position is odd. Notice that $E(c) = f_1(c) + f_2(c)$ for all $c \in C$.

4. THE LOCALITY DECOMPOSITION

In this section, a novel multiobjectivization strategy for the HP model is proposed. The conventional energy (objective) function of the HP model is decomposed based on the

locality notion of amino acid interactions. A hydrophobic topological contact $htc(s_i, s_j)$ can be considered to represent either a *local* or a *nonlocal* interaction. It depends on whether or not the sequence distance between the amino acids s_i and s_j (*i.e.*, $|j-i|$) is within a given maximum δ. From this, a two-objective formulation, $\mathbf{f}(c) = [f_1(c), f_2(c)]^T$, is defined over the set of valid protein conformations $c \in C$:

$$f_1(c) = \sum_{s_i, s_j \in S} e(s_i, s_j) \qquad \text{for } \boldsymbol{j - i \leq \delta}, i < j \tag{5}$$

$$f_2(c) = \sum_{s_i, s_j \in S} e(s_i, s_j) \qquad \text{for } \boldsymbol{j - i > \delta}, i < j \tag{6}$$

where $f_1(c)$ and $f_2(c)$ are both to be minimized and $e(s_i, s_j)$ has been previously defined in Section 2.1.

That is, f_1 is defined for local interactions, whereas f_2 accounts for the nonlocal ones. Note that the sum of the two proposed objectives equals the conventional energy function defined in Section 2.1 ($E(c) = f_1(c) + f_2(c),\ \forall c \in C$). This is in accordance with the decomposition approach for multiobjectivization. It should also be noted that δ plays a decisive role for the behavior of this proposal. Thus, the impact of varying this parameter needs to be investigated.

5. EXPERIMENTAL SETUP

5.1 Test Cases

A total of 30 HP instances were considered. Out of them, 15 are for the two-dimensional square lattice and the other 15 are for three-dimensional cubic one. Tables 1 and 2 present the full sequences, their length (L) and the optimal or best known energy value (E^*), to the authors' knowledge.

Table 1: HP instances for the 2D square lattice.

	Sequence	L	E^*
2d1	$H_2P_5H_2P_3HP_3HP$	18	-4
2d2	$HPHPH_3P_3H_4P_2H_2$	18	-8
2d3	$PHP_2HPH_3PH_2PH_5$	18	-9
2d4	$HPHP_2H_2PHP_2HPH_2P_2HPH$	20	-9
2d5	$H_3P_2HPHPHP_2HPHPHP_2H$	20	-10
2d6	$H_2P_2HP_2HP_2HP_2HP_2HP_2HP_2H_2$	24	-9
2d7	$P_2HP_2H_2P_4H_2P_4H_2P_4H_2$	25	-8
2d8	$P_3H_2P_2H_2P_5H_7P_2H_2P_4H_2P_2HP_2$	36	-14
2d9	$P_2HP_2H_2P_2H_2P_5H_{10}P_6H_2P_2H_2P_2HP_2H_5$	48	-23
2d10	$H_2(PH)_4H_3P(HP_3)_3(P_3H)_3PH_4(PH)_4H$	50	-21
2d11	$P_2H_3PH_8P_3H_{10}PHP_3H_{12}P_4H_6PH_2PHP$	60	-36
2d12	$H_{12}PHPH(P_2H_2P_2H_2P_2H)_3PHPH_{12}$	64	-42
2d13	$H_4P_4H_{12}P_6(H_{12}P_3)_3HP_2H_2P_2H_2P_2HPH$	85	-53
2d14	$P_6HPH_2P_5H_3PH_5PH_2P_4H_2P_2H_2PH_5PH_{10}$ $PH_2PH_7P_{11}H_7P_2HPH_3P_6HPH_2$	100	-48
2d15	$P_3H_2P_2H_4P_2H_3PH_2PH_2PH_4P_8H_6P_2H_6P_9$ $HPH_2PH_{11}P_2H_3PH_2PHP_2HPH_3P_6H_3$	100	-50

Table 2: HP instances for the 3D cubic lattice.

	Sequence	L	E^*
3d1	$HPHP_2H_2PHP_2HPH_2P_2HPH$	20	-11
3d2	$H_2P_2HP_2HP_2HP_2HP_2HP_2HP_2H_2$	24	-13
3d3	$P_2HP_2H_2P_4H_2P_4H_2P_4H_2$	25	-9
3d4	$P_3H_2P_2H_2P_5H_7P_2H_2P_4H_2P_2HP_2$	36	-18
3d5	$P_2H_3PH_3P_3HPH_2PH_2P_2HPH_4PHP_2H_5PHPH_2P_2H_2P$	46	-32
3d6	$P_2HP_2H_2P_2H_2P_5H_{10}P_6H_2P_2H_2P_2HP_2H_5$	48	-31
3d7	$H_2(PH)_4H_3P(HP_3)_3(P_3H)_3PH_4(PH)_4H$	50	-32
3d8	$PH(PH_3)_2P(PH_2PH)_2H(HP)_3(H_2P_2H)_2PHP_4(H(P_2H)_2)_2$	58	-44
3d9	$P_2H_3PH_8P_3H_{10}PHP_3H_{12}P_4H_6PH_2PHP$	60	-52
3d10	$H_{12}PHPH(P_2H_2P_2H_2P_2H)_3PHPH_{12}$	64	-55
3d11	$P(HPH_2PH_2PHP_2H_3P_3)_3(HPH)_3P_2H_3P$	67	-56
3d12	$P(HPH)_3P_2H_2(P_2H)_6H(P_2H_3)_4P_2(HPH)_3$ $P_2HP(PHP_2H_2P_2HP)_2$	88	-72
3d13	$P_2H_2P_5H_2P_2H_2PHP_2HP_7HP_3H_2PH_2P_6HP_2HP$ $HP_2HP_5H_3P_4H_2PH_2P_5H_2P_4H_4PHP_8H_5P_2HP_2$	103	-56
3d14	$P_3H_3PHP_4HP_5H_2P_4H_2P_2H_2(P_4H)_2P_2HP_2H_2P_3H_2PHPH_3$ $P_4H_3P_6H_2P_2HP_2HPHP_2HP_7HP_2H_3P_4HP_3H_5P_4H_2(PH)_4$	124	-71
3d15	$HP_5HP_4HPH_2PH_2P_4HPH_3P_4HPHPH_4P_{11}HP_2HP_3HPH_2P_3$ $H_2P_2HP_2HPHPHP_8HP_3H_6P_3H_2P_2H_3P_3H_2PH_5P_9HP_4HPHP_4$	136	-80

5.2 Algorithms

The simplest version of an EA, the so-called (1+1) EA, is described in Algorithm 1. First, an initial individual c is generated at random. At each generation, a new individual c' is created by means of mutation. If c' is at least as good as c, then c' is accepted as the starting point for the next generation. Depending on the problem formulation, this acceptance criterion is to be based either on the conventional energy evaluation or on the Pareto-dominance relation.

Algorithm 1 Basic (1+1) Evolutionary Algorithm.

```
1: choose c ∈ C uniformly at random
2: repeat
3:    c' ← mutate(c)
4:    if c' not worse than c then
5:       c ← c'
6:    end if
7: until < stop condition >
```

A variant of the above described (1+1) EA is presented in Algorithm 2. An external archive stores the nondominated solutions found along the evolutionary process. The archiving strategy influences the search behavior of the algorithm in such a way that the mutant c' is only accepted if it is not dominated by any individual in the archive. If accepted, c' is included in the archive and all individuals dominated by c', and those mapping to the same objective vector $\mathbf{f}(c')$, are removed. Note that this archiving strategy makes only sense for the multiobjectivized problem formulations.

Algorithm 2 Archiving (1+1) Evolutionary Algorithm.

```
1: choose c ∈ C uniformly at random
2: A ← {c}
3: repeat
4:    c' ← mutate(c)
5:    if ∄ĉ ∈ A : ĉ ≺ c' then
6:       A ← {ĉ ∈ A : c' ⊀ ĉ ∧ f(ĉ) ≠ f(c')} ∪ {c'}
7:       c ← c'
8:    end if
9: until < stop condition >
```

It was also considered a genetic algorithm (GA), whose general structure is given in Algorithm 3. First, an initial parent population P of size N is randomly generated. At each generation, the fittest individuals in P are selected for mating (*selection-for-variation*). Then, a children population P' is created by applying the variation operators. Finally, parents and children compete for a place in the new population (*selection-for-survival*). When applied to the single-objective problem formulation, selection is driven by the conventional energy value of the candidate conformations. Regarding the multiobjective formulation, the discrimination among individuals is to be based on *nondominated sorting* and *crowding distance*, as in the NSGA-II [11].

Algorithm 3 Genetic Algorithm.

```
1: choose P ⊂ C : |P| = N uniformly at random
2: while < stop condition > do
3:    P̂ ← selection-for-variation(P)
4:    P' ← variation(P̂)
5:    P ← selection-for-survival(P ∪ P')
6: end while
```

A representation of absolute moves was adopted. Conformations are encoded as sequences in $\{U, D, L, R, F, B\}^{L-1}$,

denoting the up, down, left, right, forward and backward lattice positions for an amino acid with regard to the preceding one. Only directions $\{U, D, L, R\}$ are used in the two-dimensional case. The implemented genetic operators are as follows. One-point crossover (only for the GA) is applied with a given probability p_c. In mutation, each encoding position is randomly perturbed with probability p_m. In all cases, only valid solutions are accepted during the search.

5.3 Performance Assessment

For all the experiments, 100 independent executions were performed. The results are evaluated in terms of the best obtained energy value (β), the number of times that this solution was found (f) and the arithmetic mean (μ). Additionally, the *overall average performance* (OAP) measure [13] was adopted in order to assess the overall behavior of the studied approaches. OAP is defined as the average ratio of the obtained mean values to the optimum (E^*). Formally:

$$\text{OAP} = \frac{100\%}{|T|}\left(\sum_{t\in T}\frac{\mu(t)}{E^*(t)}\right) \quad (7)$$

where T is the set of all test cases. Thus, OAP = 100% suggests the ideal situation where the optimum solution for each instance was reached during all the performed executions.

Statistical significance analysis was conducted as follows. First, *D'Agostino-Pearson's omnibus* K^2 test was used to evaluate the normality of data distributions. For normally distributed data, either *ANOVA* or the *Welch's t* parametric tests were used depending on whether the variances across the samples were homogeneous (*homoskedasticity*) or not. This was investigated using the *Bartlett's* test. For non-normal data, the nonparametric *Kruskal-Wallis* test was adopted. A significance level of $\alpha = 0.05$ was considered.

6. RESULTS

6.1 Results for the (1+1) EA

In this section, the (1+1) EA is used for comparing among the three studied HP model's formulations: the conventional single-objective formulation (SO), the parity decomposition (PD) [14], and the locality decomposition (LD) being proposed. Results are also presented for the archiving version of the (1+1) EA, which applies only for PD and LD. A fixed mutation probability of $p_m = \frac{1}{L-1}$ and a stopping condition of 100,000 evaluations were considered in all cases.

Given the importance that parameter δ has on the behavior of the LD approach (see Section 4), the best adjustment for this parameter is first investigated. A total of 10 odd values for δ have been evaluated, starting from 3.[1] Figure 2 presents the overall average performance (OAP) obtained using LD for the different values of δ. Results are provided for both the basic and the archiving (1+1) EA. Also, the performance of the SO formulation is shown as a baseline.

It is evident from this figure that an important increase in performance has been obtained by using the new proposed multiobjectivization. For the different values of δ, LD reached the best results when using the basic, non-archiving variant of the algorithm. However, even using the archiving

[1]In the 2D square and the 3D cubic lattices, a topological contact can only occur if the sequence distance between the amino acids is odd and at least equal to 3.

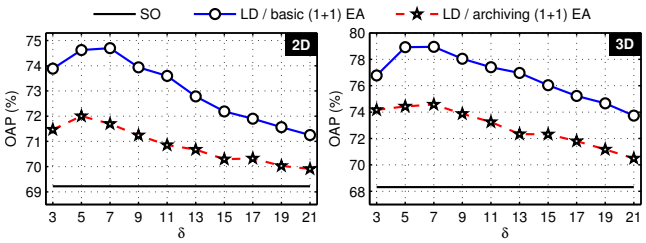

Figure 2: Varying the distance parameter δ.

(1+1) EA, LD performed better in all cases compared to the conventional SO formulation. It can be seen that the highest OAP values were obtained at $\delta = 7$, in most of the cases. In addition, notice that the performance of the algorithms gradually declined with the increasing value for δ.

The distance parameter δ was set to 7 for further analysis. Tables 3 and 4 detail the obtained results for all the two-dimensional and three-dimensional test cases, respectively. For each instance, these tables show the best energy (β), the frequency (f) and the mean (μ) achieved using the different formulations. The lowest μ obtained for each instance has been **shaded**. Finally, the OAP measure evaluates the overall performance of the formulations, see Section 5.3.

From these tables, it is possible to note that the proposed LD outperformed both the conventional SO formulation and the previously reported PD multiobjectivization in most of the cases. By using LD, the basic (1+1) EA reached the lowest average energy for 12 out of the 15 two-dimensional instances, leading to an OAP increase of $(74.70 - 69.22) = 5.48\%$ with regard to the SO formulation (see Table 3). As previously stated, the use of the archiving strategy within the (1+1) EA seemed not to be favorable for the proposed LD multiobjectivization. Nevertheless, even using this algorithm it was possible for LD to obtain better results than the SO and PD formulations for most of the instances.

As shown in Table 4, the proposed LD scored the best average performance for all the three-dimensional instances when using the basic (1+1) EA. This was also reflected as an OAP increase of 10.65% with respect to the conventional SO formulation. Just as it happened for the two-dimensional instances, the advantages of LD were not as remarkable when using the archiving (1+1) EA. However, the results of LD for this algorithm were still competitive; the OAP measure was improved by 6.28% over the SO, being this the second best performance achieved for the three-dimensional test cases.

Finally, Tables 5 and 6 outline how the SO, PD and LD formulations compare statistically with respect to each other in all the test cases. Each row in these tables compares two formulations, say A and B, which is denoted as "A/B". If a significant performance difference exists between A and B, the corresponding cells are marked either as + or − depending on whether such a difference was in favor of or against A. Empty cells indicate that there was not a statistically important difference between the approaches. The rightmost column shows the overall results of this analysis.

As can be seen from Table 5, both PD and LD significantly outperformed the conventional SO formulation in most of the cases when using the basic (1+1) EA. The proposed LD achieved statistically better results than SO in 28 out of the 30 adopted test cases. PD performed significantly better than SO for 20 of the instances. By comparing the multiobjectivized formulations, there was a significant performance difference in favor of LD for 23 of the test cases.

Table 3: Results for the basic and the archiving (1+1) EA on the two-dimensional benchmarks.

	Basic (1+1) EA						Archiving (1+1) EA			
	SO		PD		LD		PD		LD	
Seq.	β (f)	μ	β (f)	μ	β (f)	μ	β (f)	μ	β (f)	μ
2d1	-4 (4)	-2.70	-4 (6)	**-2.71**	-4 (3)	-2.69	-4 (5)	-2.69	-4 (2)	-2.67
2d2	-8 (18)	-6.81	-8 (24)	-7.04	-8 (31)	**-7.16**	-8 (21)	-7.00	-8 (21)	-6.99
2d3	-8 (11)	-7.00	-8 (48)	**-7.45**	-9 (2)	-7.39	-8 (24)	-7.12	-8 (22)	-7.05
2d4	-9 (8)	-6.84	-9 (4)	-6.95	-9 (11)	**-7.23**	-9 (6)	-6.88	-9 (14)	-7.13
2d5	-9 (3)	-6.92	-10 (2)	**-7.08**	-9 (1)	-7.06	-9 (1)	-6.99	-8 (14)	-6.89
2d6	-8 (14)	-6.81	-9 (1)	-6.87	-9 (2)	**-7.30**	-9 (1)	-6.89	-9 (1)	-6.95
2d7	-7 (26)	-5.79	-8 (6)	-5.90	-8 (7)	**-6.17**	-8 (5)	-5.80	-8 (10)	-6.08
2d8	-13 (1)	-9.97	-13 (1)	-10.23	-13 (4)	**-10.61**	-13 (1)	-10.12	-13 (1)	-10.13
2d9	-18 (5)	-14.23	-19 (2)	-15.20	-20 (2)	**-16.29**	-18 (5)	-15.02	-21 (1)	-15.64
2d10	-18 (2)	-13.79	-18 (1)	-14.06	-19 (1)	**-15.07**	-17 (4)	-13.76	-18 (1)	-14.40
2d11	-30 (2)	-24.39	-30 (7)	-25.43	-32 (1)	**-27.80**	-31 (1)	-25.32	-32 (1)	-25.80
2d12	-29 (1)	-23.82	-30 (1)	-25.12	-30 (4)	**-26.61**	-30 (1)	-24.63	-29 (2)	-24.91
2d13	-41 (1)	-33.81	-41 (1)	-34.54	-44 (1)	**-38.09**	-42 (1)	-34.18	-41 (1)	-35.34
2d14	-41 (1)	-30.80	-39 (3)	-32.18	-39 (2)	**-34.41**	-41 (1)	-31.72	-39 (1)	-32.58
2d15	-40 (1)	-31.71	-40 (3)	-32.70	-39 (7)	**-34.97**	-40 (1)	-32.57	-41 (1)	-33.60
OAP	69.22%		71.39%		**-74.70%**		70.47%		71.70%	

Table 4: Results for the basic and the archiving (1+1) EA on the three-dimensional benchmarks.

	Basic (1+1) EA						Archiving (1+1) EA			
	SO		PD		LD		PD		LD	
Seq.	β (f)	μ	β (f)	μ	β (f)	μ	β (f)	μ	β (f)	μ
3d1	-11 (57)	-10.48	-11 (69)	-10.64	-11 (94)	**-10.94**	-11 (64)	-10.51	-11 (60)	-10.52
3d2	-13 (23)	-11.30	-13 (34)	-11.70	-13 (66)	**-12.53**	-13 (27)	-11.59	-13 (42)	-11.87
3d3	-9 (57)	-8.48	-9 (70)	-8.65	-9 (95)	**-8.95**	-9 (62)	-8.51	-9 (73)	-8.66
3d4	-18 (10)	-15.19	-18 (13)	-15.74	-18 (46)	**-16.97**	-18 (8)	-15.30	-18 (15)	-15.96
3d5	-30 (2)	-23.87	-30 (1)	-25.38	-31 (1)	**-27.53**	-30 (1)	-24.56	-32 (1)	-25.68
3d6	-29 (1)	-22.79	-29 (2)	-24.42	-31 (1)	**-26.66**	-28 (3)	-23.64	-31 (1)	-24.64
3d7	-25 (6)	-20.64	-27 (1)	-22.07	-28 (1)	**-24.31**	-27 (1)	-21.22	-27 (4)	-22.77
3d8	-35 (1)	-27.34	-36 (1)	-29.02	-36 (2)	**-31.98**	-35 (1)	-27.96	-35 (5)	-29.98
3d9	-46 (1)	-37.20	-47 (1)	-40.03	-47 (3)	**-42.88**	-47 (1)	-38.81	-49 (1)	-41.59
3d10	-45 (1)	-35.59	-46 (1)	-37.69	-50 (1)	**-43.29**	-43 (2)	-36.51	-49 (1)	-40.28
3d11	-38 (2)	-30.17	-39 (2)	-32.65	-41 (1)	**-36.10**	-38 (2)	-31.17	-40 (3)	-34.82
3d12	-47 (1)	-36.22	-49 (1)	-39.85	-53 (1)	**-46.13**	-48 (1)	-38.09	-50 (1)	-42.55
3d13	-40 (1)	-29.97	-41 (1)	-31.31	-40 (1)	**-35.42**	-38 (1)	-29.94	-45 (1)	-33.52
3d14	-43 (4)	-34.51	-48 (1)	-36.97	-50 (2)	**-43.98**	-47 (1)	-35.04	-49 (2)	-40.83
3d15	-51 (1)	-37.26	-52 (1)	-42.11	-57 (1)	**-47.42**	-50 (1)	-40.43	-54 (1)	-44.90
OAP	68.31%		72.20%		**-78.96%**		70.00%		74.59%	

Table 5: Statistical analysis for comparing the three HP model's formulations. Basic (1+1) EA.

	2D benchmarks															3D benchmarks														Overall
	2d1	2d2	2d3	2d4	2d5	2d6	2d7	2d8	2d9	2d10	2d11	2d12	2d13	2d14	2d15	3d1	3d2	3d3	3d4	3d5	3d6	3d7	3d8	3d9	3d10	3d11	3d12	3d13	3d14	3d15	
PD/SO		+							+	++		++				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**20+ 0−**
LD/SO	+	+	+	+	+	+	+	+	+	+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**28+ 0−**
LD/PD				+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**23+ 0−**

Table 6: Statistical analysis for comparing the three HP model's formulations. Archiving (1+1) EA.

	2D benchmarks															3D benchmarks														Overall
	2d1	2d2	2d3	2d4	2d5	2d6	2d7	2d8	2d9	2d10	2d11	2d12	2d13	2d14	2d15	3d1	3d2	3d3	3d4	3d5	3d6	3d7	3d8	3d9	3d10	3d11	3d12	3d13	3d14	3d15	
PD/SO								+	+	++	+							+		+		+		+		+					**8+ 0−**
LD/SO						+		+	+	+	+	+	+	+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	**22+ 0−**
LD/PD									+	++		+						+	+	+	+	+	+	+	+	+	+	+	+	+	**15+ 0−**

Regarding the archiving (1+1) EA, Table 6 shows that the proposed LD significantly improved the search performance in 22 and 15 of the instances when compared to SO and PD, respectively. PD's results for 8 of the test cases were statistically superior to those obtained by the SO formulation.

6.2 Results for the Genetic Algorithm

In this section, the results for the implemented genetic algorithm (GA) are analyzed. Three different problem formulations are compared: the conventional single-objective (SO), the recently proposed parity decomposition (PD) [14], and the locality decomposition (LD) proposed in this paper.

LD is sensitive to the distance parameter δ (see Section 4). Preliminary testing was conducted in order to investigate the value of δ providing the best performance for the GA. Due to space limitations, details of such an analysis were not included in this document, but the obtained results suggest that $\delta = 7$ is a convenient adjustment for this parameter.

Furthermore, different settings for the GA are evaluated in order to identify the most appropriate conditions for the compared approaches. Three different recombination and mutation probabilities were considered: $p_c = \{0.8, 0.9, 1.0\}$ and $p_m = \{\frac{1}{L-1}, 0.01, 0.05\}$. Also, the effects of preventing duplicate individuals (clones) from the population are analyzed. Thus, 18 parameter configurations for the GA are investigated. The population size was fixed to $N = 100$ in all cases, and the algorithm was allowed to run until a maximum number of $100,000$ evaluations was reached. Figures 3

and 4 present (2D and 3D, respectively) the overall average performance (OAP) measure obtained by the studied formulations when varying the different parameters of the GA.

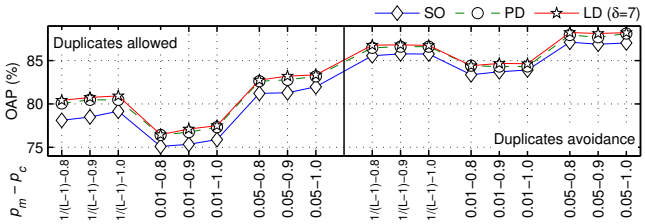

Figure 3: Settings for the GA, 2D benchmarks.

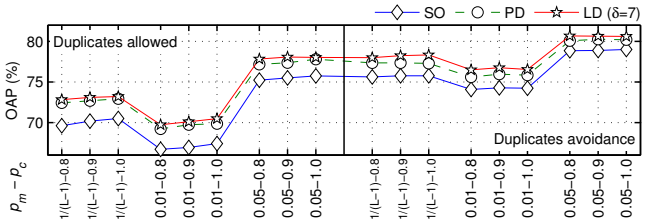

Figure 4: Settings for the GA, 3D benchmarks.

From these figures, both LD and PD performed better than the conventional SO formulation for all the different parameter configurations of the GA. By comparing between LD and PD, it can be noted that there was a performance difference in favor of the proposed LD, in all the cases. Some general observations can be made concerning the behavior of the GA. On the one hand, the algorithm seemed not to be seriously affected when varying the recombination probability. On the other hand, it responded positively to the increased mutation rate, being $p_m = 0.05$ the fixed value which provided the best performance in all the cases. Finally, the results were significantly improved in all the cases when duplicate individuals were removed from the population.

For a more detailed analysis, the parameters adjustment which allowed each of the formulations to reach the highest OAP value was selected. The obtained results are presented in Tables 7 and 8. These tables show the best energy (β), the frequency (f) and the mean (μ) obtained for each instance when using the different formulations. The lowest average energy achieved for each test case has been **shaded**. Also, the OAP measure is given at the bottom of the tables.

As shown in Table 7, the previously reported PD achieved the lowest average energy for 8 out of the 15 two-dimensional instances. The proposed LD presented the best performance for only 7 of the test cases. However, LD allowed the GA to reach the highest OAP value, which represents an increase of $(88.25 - 87.13) = 1.12\%$ over the SO formulation.

Regarding the three-dimensional instances, it can be seen from Table 8 that the best average performance of the algorithm was scored in most cases when using the proposed LD multiobjectivization. An OAP increase of 1.67% was obtained with regard to the conventional SO formulation.

Table 9 points out how the SO, PD and LD formulations are statistically compared to each other in all the instances. Each row in these tables compares two formulations, say A and B, which is denoted as "A/B". If a statistically significant difference exists between A and B, the corresponding cells are marked either as + or − depending on whether such a difference favors A or not. Empty cells indicate that there was not a significant difference between the approaches. The rightmost column presents the overall results of this analysis.

Table 7: Results for the GA, 2D benchmarks.

	SO		PD		LD	
Seq.	$\beta\ (f)$	μ	$\beta\ (f)$	μ	$\beta\ (f)$	μ
2d1	-4 (69)	-3.69	-4 (78)	**-3.78**	-4 (77)	-3.77
2d2	-8 (92)	**-7.92**	-8 (91)	-7.91	-8 (90)	-7.90
2d3	-9 (68)	-8.68	-9 (73)	-8.73	-9 (75)	**-8.75**
2d4	-9 (99)	-8.99	-9 (93)	-8.93	-9 (100)	**-9.00**
2d5	-10 (87)	-9.75	-10 (94)	**-9.89**	-10 (93)	-9.87
2d6	-9 (62)	-8.60	-9 (69)	-8.69	-9 (75)	**-8.75**
2d7	-8 (47)	-7.40	-8 (49)	**-7.47**	-8 (51)	**-7.47**
2d8	-13 (12)	-11.45	-14 (2)	-11.49	-13 (16)	**-11.62**
2d9	-21 (2)	-17.85	-23 (1)	-18.30	-22 (2)	**-18.45**
2d10	-21 (4)	-18.27	-21 (1)	**-18.54**	-21 (3)	-18.50
2d11	-34 (1)	-30.27	-34 (1)	**-30.54**	-34 (1)	-30.41
2d12	-36 (2)	-30.94	-35 (3)	-30.75	-36 (5)	**-31.56**
2d13	-49 (1)	-41.75	-48 (1)	**-42.57**	-47 (3)	-42.05
2d14	-44 (1)	-36.74	-43 (1)	**-37.74**	-41 (2)	-37.31
2d15	-43 (2)	-37.14	-43 (1)	**-38.28**	-43 (1)	-38.12
OAP	87.13%		88.13%		**88.25%**	

Table 8: Results for the GA, 3D benchmarks.

	SO		PD		LD	
Seq.	$\beta\ (f)$	μ	$\beta\ (f)$	μ	$\beta\ (f)$	μ
3d1	-11 (100)	**-11.00**	-11 (100)	**-11.00**	-11 (100)	**-11.00**
3d2	-13 (95)	**-12.94**	-13 (97)	**-12.94**	-13 (95)	-12.91
3d3	-9 (72)	-8.71	-9 (87)	-8.87	-9 (92)	**-8.92**
3d4	-18 (12)	-15.91	-18 (31)	**-16.54**	-18 (22)	-16.37
3d5	-32 (1)	-27.72	-32 (1)	-28.12	-31 (8)	**-28.37**
3d6	-31 (1)	-26.59	-30 (3)	-26.89	-31 (3)	**-27.24**
3d7	-30 (1)	-26.43	-29 (12)	-26.70	-31 (1)	**-26.85**
3d8	-37 (1)	-32.39	-37 (3)	-33.03	-40 (1)	**-33.53**
3d9	-50 (1)	-43.46	-50 (1)	**-44.56**	-49 (2)	-44.43
3d10	-52 (1)	-46.12	-53 (1)	-46.15	-52 (2)	**-46.95**
3d11	-41 (1)	-36.39	-43 (1)	-37.36	-43 (1)	**-37.62**
3d12	-50 (5)	-44.02	-54 (1)	-44.85	-52 (1)	**-45.26**
3d13	-41 (1)	-34.99	-43 (1)	**-35.78**	-42 (1)	-35.69
3d14	-51 (1)	-41.83	-50 (1)	-42.80	-49 (2)	**-43.09**
3d15	-52 (2)	-45.51	-56 (2)	-46.43	-54 (1)	**-47.17**
OAP	79.01%		80.26%		**80.68%**	

Table 9: Statistical analysis for comparing the three HP model's formulations. Genetic Algorithm.

	2D benchmarks															3D benchmarks															
	2d1	2d2	2d3	2d4	2d5	2d6	2d7	2d8	2d9	2d10	2d11	2d12	2d13	2d14	2d15	3d1	3d2	3d3	3d4	3d5	3d6	3d7	3d8	3d9	3d10	3d11	3d12	3d13	3d14	3d15	Overall
PD/SO			−					+					+++				++						++		++++						12+ 1−
LD/SO				+				+		+		+		+			++++						++	++++++++							16+ 0−
LD/PD			+										+												+							3+ 0−

Table 9 shows that PD obtained significantly better results than the SO problem formulation for 12 of the instances. Note, however, that PD was statistically inferior to SO in solving the 2d4 instance. For 16 out of the 30 adopted test cases, the proposed LD significantly increased the performance of the GA with respect to SO. Finally, the statistical analysis indicates that for only 3 of the adopted test cases there was a significant performance difference between PD and LD, all cases in favor of the proposed LD.

7. CONCLUSIONS AND FUTURE WORK

A novel multiobjectivization proposal for the HP model of protein structure prediction (PSP) was presented. In the proposed approach, called the locality decomposition, topo-

logical interactions on the lattice are classified either as local or nonlocal depending on the distance (in the sequence) between the amino acids involved. By grouping and isolating local interactions from the nonlocal ones, an alternative two-objective formulation of the problem was defined.

Different evolutionary algorithms (EAs) were implemented in order to investigate the suitability of the proposed locality decomposition. This approach was evaluated and compared with respect to both the conventional single-objective problem formulation and the recently proposed parity decomposition [14]. Experiments were conducted on both the two-dimensional square and the three-dimensional cubic lattices, and a large set of 30 HP model's instances was considered.

The proposed locality decomposition provided the best average performance of the implemented algorithms in most of the cases. Thus, the suitability of this approach was demonstrated. This supports previous evidence on the effectiveness of multiobjectivization to overcome search difficulties such as that of becoming trapped in local optima [17, 23].

Although competitive, both the parity and the locality decompositions were negatively affected by the use of the archiving strategy within the (1+1) EA. This is contrary to what is expected in multiobjective optimization, where archiving is essential for converging towards a set of trade-offs among the conflicting problem objectives [22, 26]. Nevertheless, in spite of being alternatively modeled and treated as a multiobjective problem, the HP model is actually a single-objective problem. Therefore, maintaining an approximation set of nondominated solutions becomes not as important. In addition, the archiving strategy influences the acceptance criterion of the algorithm in such a way that the introduction of plateaus, the only achievable effect of decomposition, may be partially reversed [17]. That is, some of the mutually incomparable solutions can be comparable to those in the archive. This could lead some parts of the plateaus to become inaccessible, thus restricting the exploration.

Even when the best results for the genetic algorithm were reached in most of the cases by using the proposed locality decomposition, the performance differences among the three compared formulations were not as impressive as those observed for the (1+1) EA. This can be explained by the fact that population-based approaches are inherently less susceptible to get stuck in local optima. On the other hand, the use of a multiobjective problem formulation enabled diversity promotion in the objective space (through the *crowding distance* operator [11]). This enhanced exploration and, to some extent, gives an explanation to the improvements that the parity and the locality decompositions achieved with respect to the conventional single-objective formulation.

To the best of authors' knowledge, the parity and the locality decompositions represent the first attempts on the use of multiobjective optimization techniques to solve PSP under the HP model. It is important to remark that the aim of this study was not to improve the state-of-the-art results, but rather to investigate the impact of using the proposed multiobjectivization on the resolution of this problem. From the obtained results, it is expected that the locality decomposition can be successfully incorporated in order to improve the performance of established state-of-the-art algorithms (such as those mentioned in Section 3). This issue needs to be investigated in order to derive more general conclusions. Also, the conflicting relationship between the objectives of the proposed formulation has to be analyzed. Finally, the multiobjectivization of the HP model by means of the addition of supplementary objectives remains unexplored. This can be seen as an interesting issue for future research.

8. ACKNOWLEDGMENTS

We would like to acknowledge support from CONACyT projects 105060 and 99276. Also, this research was partially funded by project 51623 from "Fondo Mixto Conacyt-Gobierno del Estado de Tamaulipas". Finally, we would like to thank to "Fondo Mixto de Fomento a la Investigación científica y Tecnológica CONACyT - Gobierno del Estado de Tamaulipas" for their support to publish this paper.

9. REFERENCES

[1] C. Anfinsen. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230, 1973.

[2] D. Becerra, A. Sandoval, D. Restrepo-Montoya, and L. Nino. A Parallel Multi-Objective Ab Initio Approach for Protein Structure Prediction. In *IEEE International Conference on Bioinformatics and Biomedicine*, pages 137–141, 2010.

[3] B. Berger and T. Leighton. Protein Folding in the Hydrophobic-Hydrophilic (HP) Model is NP-complete. In *International Conference on Research in Computational Molecular Biology*, pages 30–39, New York, NY, USA, 1998. ACM.

[4] D. Brockhoff, T. Friedrich, N. Hebbinghaus, C. Klein, F. Neumann, and E. Zitzler. Do Additional Objectives Make a Problem Harder? In *Genetic and Evolutionary Computation Conference*, pages 765–772, London, England, 2007. ACM.

[5] A. Băutu and H. Luchian. Protein Structure Prediction in Lattice Models with Particle Swarm Optimization. In *Swarm Intelligence*, volume 6234 of *Lecture Notes in Computer Science*, pages 512–519. Springer Berlin / Heidelberg, 2010.

[6] C. Chira. A Hybrid Evolutionary Approach to Protein Structure Prediction with Lattice Models. In *IEEE Congress on Evolutionary Computation*, pages 2300–2306, New Orleans, LA, USA, 2011.

[7] P. Crescenzi, D. Goldman, C. Papadimitriou, A. Piccolboni, and M. Yannakakis. On the Complexity of Protein Folding. In *ACM Symposium on Theory of Computing*, pages 597–603, Dallas, TX, USA, 1998. ACM.

[8] V. Cutello, G. Narzisi, and G. Nicosia. A Multi-Objective Evolutionary Approach to the Protein Structure Prediction Problem. *Journal of The Royal Society Interface*, 3(6):139–151, 2006.

[9] V. Cutello, G. Nicosia, M. Pavone, and J. Timmis. An Immune Algorithm for Protein Structure Prediction on Lattice Models. *IEEE Transactions on Evolutionary Computation*, 11(1):101–117, 2007.

[10] R. Day, J. Zydallis, and G. Lamont. Solving the Protein Structure Prediction Problem Through a Multi-Objective Genetic Algorithm. In *IEEE/DARPA International Conference on Computational Nanoscience*, pages 32–35, 2002.

[11] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization:

NSGA-II. In *Parallel Problem Solving from Nature*, pages 849–858, Paris, France, 2000. Springer. Lecture Notes in Computer Science No. 1917.

[12] K. Dill. Theory for the Folding and Stability of Globular Proteins. *Biochemistry*, 24(6):1501–9, 1985.

[13] M. Garza-Fabre, E. Rodriguez-Tello, and G. Toscano-Pulido. Comparing Alternative Energy Functions for the HP Model of Protein Structure Prediction. In *IEEE Congress on Evolutionary Computation*, pages 2307–2314, New Orleans, LA, USA, 2011.

[14] M. Garza-Fabre, E. Rodriguez-Tello, and G. Toscano-Pulido. Multiobjectivizing the HP Model for Protein Structure Prediction. In *European Conference on Evolutionary Computation in Combinatorial Optimisation*, volume 7245 of *Lecture Notes in Computer Science*, pages 182–193. Springer Berlin / Heidelberg, Málaga, Spain, 2012.

[15] D. Greiner, J. Emperador, G. Winter, and B. Galván. Improving Computational Mechanics Optimum Design Using Helper Objectives: An Application in Frame Bar Structures. In *Evolutionary Multi-Criterion Optimization*, volume 4403 of *Lecture Notes in Computer Science*, pages 575–589. Springer Berlin / Heidelberg, Matshushima, Japan, 2007.

[16] J. Handl, S. Lovell, and J. Knowles. Investigations into the Effect of Multiobjectivization in Protein Structure Prediction. In *Parallel Problem Solving from Nature*, volume 5199 of *Lecture Notes in Computer Science*, pages 702–711. Springer Berlin / Heidelberg, Dortmund, Germany, 2008.

[17] J. Handl, S. Lovell, and J. Knowles. Multiobjectivization by Decomposition of Scalar Cost Functions. In *Parallel Problem Solving from Nature*, volume 5199 of *Lecture Notes in Computer Science*, pages 31–40. Springer Berlin / Heidelberg, Dortmund, Germany, 2008.

[18] M. Hoque, M. Chetty, A. Lewis, and A. Sattar. Twin Removal in Genetic Algorithms for Protein Structure Prediction Using Low-Resolution Model. *IEEE/ACM Transactions on Computational Biology and Bioinformatics,*, 8(1):234–245, 2011.

[19] M. Islam, M. Chetty, and M. Murshed. Novel Local Improvement Techniques in Clustered Memetic Algorithm for Protein Structure Prediction. In *IEEE Congress on Evolutionary Computation*, pages 1003–1011, New Orleans, LA, USA, 2011.

[20] M. Jähne, X. Li, and J. Branke. Evolutionary Algorithms and Multi-Objectivization for the Travelling Salesman Problem. In *Genetic and Evolutionary Computation Conference*, pages 595–602, Montreal, Canada, 2009. ACM.

[21] M. Jensen. Helper-Objectives: Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimisation. *Journal of Mathematical Modelling and Algorithms*, 3:323–347, 2004.

[22] J. Knowles and D. Corne. Properties of an Adaptive Archiving Algorithm for Storing Nondominated Vectors. *IEEE Transactions on Evolutionary Computation*, 7(2):100–116, 2003.

[23] J. Knowles, R. Watson, and D. Corne. Reducing Local Optima in Single-Objective Problems by Multi-objectivization. In *Evolutionary Multi-Criterion Optimization*, pages 269–283, London, UK, 2001. Springer-Verlag.

[24] D. Lochtefeld and F. Ciarallo. Helper-Objective Optimization Strategies for the Job-Shop Scheduling Problem. *Applied Soft Computing*, 11(6):4161–4174, 2011.

[25] H. Lopes. Evolutionary Algorithms for the Protein Folding Problem: A Review and Current Trends. In *Computational Intelligence in Biomedicine and Bioinformatics*, volume 151 of *Studies in Computational Intelligence*, pages 297–315. Springer Berlin / Heidelberg, 2008.

[26] M. López-Ibáñez, J. Knowles, and M. Laumanns. On Sequential Online Archiving of Objective Vectors. In *Evolutionary Multi-Criterion Optimization*, volume 6576 of *Lecture Notes in Computer Science*, pages 46–60. Springer Berlin / Heidelberg, 2011.

[27] R. Santana, P. Larranaga, and J. Lozano. Protein Folding in Simplified Models With Estimation of Distribution Algorithms. *IEEE Transactions on Evolutionary Computation*, 12(4):418–438, 2008.

[28] J. Santos and M. Diéguez. Differential Evolution for Protein Structure Prediction Using the HP Model. In *Foundations on Natural and Artificial Computation*, volume 6686 of *Lecture Notes in Computer Science*, pages 323–333. Springer Berlin / Heidelberg, 2011.

[29] E. Segredo, C. Segura, and C. Leon. A Multiobjectivised Memetic Algorithm for the Frequency Assignment Problem. In *IEEE Congress on Evolutionary Computation*, pages 1132–1139, New Orleans, LA, USA, 2011.

[30] C. Segura, E. Segredo, and C. León. Parallel Island-Based Multiobjectivised Memetic Algorithms for a 2D Packing Problem. In *Genetic and Evolutionary Computation Conference*, pages 1611–1618, Dublin, Ireland, 2011. ACM.

[31] A. Shmygelska and H. Hoos. An Ant Colony Optimisation Algorithm for the 2D and 3D Hydrophobic Polar Protein Folding Problem. *BMC Bioinformatics*, 6(1):30, 2005.

[32] C. Soares Brasil, A. Botazzo Delbem, and D. Ferraz Bonetti. Investigating Relevant Aspects of MOEAs for Protein Structures Prediction. In *Genetic and Evolutionary Computation Conference*, pages 705–712, Dublin, Ireland, 2011. ACM.

[33] R. Unger. The Genetic Algorithm Approach to Protein Structure Prediction. In *Applications of Evolutionary Computation in Chemistry*, volume 110 of *Structure & Bonding*, pages 2697–2699. Springer Berlin / Heidelberg, 2004.

[34] I. Vite-Silva, N. Cruz-Cortés, G. Toscano-Pulido, and L. de la Fraga. Optimal Triangulation in 3D Computer Vision Using a Multi-objective Evolutionary Algorithm. In *Applications of Evolutionary Computing*, volume 4448 of *Lecture Notes in Computer Science*, pages 330–339. Springer Berlin / Heidelberg, Valencia, Spain, 2007.

[35] X. Zhao. Advances on Protein Folding Simulations Based on the Lattice HP models with Natural Computing. *Applied Soft Computing*, 8(2):1029–1040, 2008.