An Improved Multiobjectivization Strategy for HP Model-Based Protein Structure Prediction*

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

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

Abstract. Through multiobjectivization, a single-objective problem is restated in multiobjective form with the aim of enabling a more efficient search process. Recently, this transformation was applied with success to the hydrophobic-polar (HP) lattice model, which is an abstract representation of the protein structure prediction problem. The use of alternative multiobjective formulations of the problem has led to significantly better results. In this paper, an improved multiobjectivization for the HP model is proposed. By decomposing the HP model's energy function, a two-objective formulation for the problem is defined. A comparative analysis reveals that the new proposed multiobjectivization evaluates favorably with respect to both the conventional single-objective and the previously reported multiobjective formulations. Statistical significance testing and the use of a large set of test cases support the findings of this study. Both two-dimensional and three-dimensional lattices are considered.

Keywords: Multiobjectivization, protein structure prediction, HP model.

1 Introduction

Protein structure prediction, PSP, is the problem of finding the native (energy-minimizing) conformation for a protein given only its amino acid sequence. The *hydrophobic-polar* (HP) *model* is an abstraction of this problem, where hydrophobicity is assumed to be the main stabilizing force in protein folding [6]. Even under this rather simplified model, PSP remains a challenging problem in combinatorial optimization [1, 3]. An extensive literature exists on the use of metaheuristics to address this problem, some of which is reviewed in [17, 22].

Multiobjectivization refers to the reformulation of single-objective problems in terms of two or more objective functions [15]. This transformation has been successfully used to deal with difficult optimization problems. Among them, there can be mentioned the traveling salesman problem [12, 13, 15], job-shop scheduling [13, 16], and problems in the fields of mobile communications [19], computational mechanics [9] and computer vision [21]. Multiobjectivization has also been proposed for the PSP [4, 5, 10, 20]. However, it was not until recently that this concept was applied to the particular HP model of this problem [7, 8].

* This research has been partially funded by CONACyT projects 105060 and 99276.

C.A. Coello Coello et al. (Eds.): PPSN 2012, Part II, LNCS 7492, pp. 82–92, 2012.

In [7], the originally single-objective HP model was restated in multiobjective form by decomposing the conventional energy (objective) function into two separate objectives. Such a decomposition relies on the fact that topological interactions on the lattice are only possible between amino acids whose sequence positions are of opposite parity.[1] This alternative formulation, called the *parity decomposition* (PD), showed very promising results, leading to an increased search performance in most of the conducted experiments. More recently, an improved multiobjectivization strategy for the HP model was proposed, the *locality decomposition* (LD) [8]. In LD, the decomposition of the HP model's objective 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.

Motivated from previous findings [7, 8], this paper introduces a novel multiobjectivization for the HP model, the *H-subsets decomposition* (HD). HD organizes the hydrophobic amino acids into different groups, the H-subsets. Then, the HP model's energy function is decomposed based on the correspondence of amino acids to the H-subsets. The suitability of this proposal is investigated. Through a comparative analysis, HD is evaluated with respect to the conventional single-objective formulation and the preceding PD and LD multiobjectivizations.

This paper is organized as follows. Background concepts are covered in Sect. 2. In Sect. 3, the new proposed multiobjectivization is described. Section 4 details the implemented algorithms and the performance assessment methodology. Results are given in Sect. 5. Finally, Sect. 6 provides some concluding remarks.

2 Background and Notation

2.1 The Hydrophobic-Polar (HP) Model

Proteins are chain-like molecules composed from 20 different building blocks called amino acids. The hydrophobicity of amino acids is a dominant force determining the functional, three-dimensional conformation of proteins. In the HP model [6], amino acids are classified either as hydrophobic (H) or polar (P). Protein sequences are thus of the form $S \in \{H, P\}^L$, where L is the length of the sequence. Valid protein conformations are modeled as *Self-Avoiding Walks* of the HP chain on a lattice; *i.e.*, 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.

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:

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

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, $e(s_i, s_j) = 0$. In hydrophobic topological contacts, two H amino acids $s_i, s_j \in S$ are nonconsecutive in S but adjacent in the lattice.

[1] This is true for the two-dimensional square and the three-dimensional cubic lattices.

2.2 Single-Objective 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 $\mathcal{F}$ denotes the set of all feasible solutions. The aim is to find those $x^* \in \mathcal{F}$ 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 $\mathbf{f}(x) = [f_1(x), \ldots, f_k(x)]^T$, where $f_i : \mathcal{F} \to \mathbb{R}$ is the i-th objective function, $i \in \{1, \ldots, k\}$. 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 defined as follows: $x \prec y \;\; \Leftrightarrow \;\; \forall i : f_i(x) \leq f_i(y) \;\wedge\; \exists j : f_j(x) < f_j(y), \;\; i, j \in \{1, \ldots, k\}$. If $x \prec y$, then x is said to *dominate* y. Otherwise, y is said to be *nondominated* with respect to x, denoted by $x \not\prec y$. The image of $\mathcal{P}^*$ in the objective space is the so-called *Pareto-optimal front*, also referred to as the *trade-off surface*.

2.3 Multiobjectivization

Multiobjectivization concerns the reformulation of single-objective problems as multiobjective ones [15]. This is done either by adding *supplementary objectives* [2, 13], or through the *decomposition* of the original objective function [11, 15]. 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 in the multiobjective version of the problem.

The present study 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 $f_i : \mathcal{F} \to \mathbb{R}, i \in \{1, \ldots, k\}$ such that $f(x) = \sum_{i=1}^{k} f_i(x), \forall x \in \mathcal{F}$. As the only possible effect [11], plateaus may be defined in the search landscape. That is, originally comparable solutions may become incomparable (mutually nondominated) with regard to the decomposed formulation. Decomposition has been proven to be effective as a means of escaping from local optima [11, 15].

3 The H-Subsets Decomposition

In this section, an improved multiobjectivization by decomposition proposal for the HP model is presented. First, all H amino acids in the protein sequence are assigned to one of two groups, namely H_1 or H_2. The H_1 and H_2 groups are to be referred to as the H-subsets. From this, a two-objective problem formulation, $\mathbf{f}(c) = [f_1(c), f_2(c)]^T$, is defined over the set of valid protein conformations $c \in C$:

$$f_1(c) = \sum_{s_i, s_j \in H_1} e(s_i, s_j) \;+\; \sum_{s_i, s_j \in H_2} e(s_i, s_j) \;\;, \tag{2}$$

$$f_2(c) = \sum_{s_i \in H_1, s_j \in H_2} e(s_i, s_j) \;\;, \tag{3}$$

where $f_1(c)$ and $f_2(c)$ are to be minimized and $e(s_i, s_j)$ was defined in Sect. 2.1.

That is, the objective function f_1 accounts for hydrophobic topological contacts $htc(s_i, s_j)$ where both the s_i and s_j amino acids belong to the same H-subset. On the contrary, f_2 is defined for those cases where s_i and s_j belong to different H-subsets. Note that $E(c) = f_1(c) + f_2(c)$ for all $c \in C$, which is consistent with the decomposition approach for multiobjectivization, see Sect. 2.3.

The organization of H amino acids into the H-subsets can be accomplished following different strategies, several of which are evaluated in Sect. 5.1.

4 Experimental Setup

4.1 Algorithms

A basic evolutionary algorithm (EA), the so-called (1+1) EA, is used to investigate the suitability of the proposed multiobjectivization (see pseudo-code below). First, an initial parent individual c is generated at random. Iteratively, an offspring c' is created by randomly mutating c at each encoding position with probability $p_m = \frac{1}{L-1}$. The new individual c' is rejected only if it is strictly worse than the parent individual c, otherwise c' is accepted as the starting point for the next generation. Such a discrimination between c and c' can be based either on the conventional, single-objective energy evaluation, or it can be based on the Pareto-dominance relation if using a multiobjective problem formulation. Only solutions representing valid protein conformations are accepted during the search.

Basic (1+1) EA

choose $c \in C$ *uniformly at random*
repeat
  $c' \leftarrow mutate(c)$
  if c' *not worse than* c **then**
    $c \leftarrow c'$
  end if
until $<$ *stop condition* $>$

Archiving (1+1) EA

choose $c \in C$ *uniformly at random*
$A \leftarrow \{c\}$
repeat
  $c' \leftarrow mutate(c)$
  if $\nexists \hat{c} \in A : \hat{c} \prec c'$ **then**
    $A \leftarrow \{\hat{c} \in A : c' \not\prec \hat{c} \wedge \mathbf{f}(\hat{c}) \neq \mathbf{f}(c')\} \cup \{c'\}$
    $c \leftarrow c'$
  end if
until $<$ *stop condition* $>$

It was also considered an archiving variant of the above described (1+1) EA (see pseudo-code above). In this variant, an external archive stores the nondominated solutions found along the evolutionary process. At each generation, the offspring 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 multiobjective problem formulations.

A representation of absolute moves was adopted. That is, 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.

4.2 Test Cases and Performance Assessment

A total of 30 HP instances are used in this study (15 are for the two-dimensional square lattice and 15 are for the three-dimensional cubic one). Due to space

limitations, details of these instances are not provided here, but they are available online at `http://www.tamps.cinvestav.mx/~mgarza/HPmodel/`. For all the experiments, 100 independent executions were performed and the algorithms were run until a maximum number of 10^5 solution evaluations was reached. The results are evaluated in terms of the best (lowest) obtained energy (β), the number of times this solution was found (f) and the average energy (μ). Additionally, the *overall average performance* (OAP) measure was adopted. OAP is defined as the average ratio of the obtained μ values to the optimum (E^*). Formally:

$$\text{OAP} = \frac{100\%}{|T|}\left(\sum_{t\in T}\frac{\mu(t)}{E^*(t)}\right) \quad , \tag{4}$$

where T is the set of all the test cases. Larger OAP values are preferred. A value of OAP = 100% suggests the ideal situation where the optimum solution for each benchmark sequence 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 or not (*Bartlett's* test). For non-normal data, the non-parametric *Kruskal-Wallis* test was adopted. A significance level of $\alpha = 0.05$ was considered.

5 Results

In this section, the (1+1) EA is used in order to evaluate and compare the four different formulations of the HP model: the conventional single-objective formulation (SO), the recently reported parity (PD) [7] and locality (LD) [8] decompositions, and the H-subsets decomposition (HD) proposed in this paper.[2]

Given the importance that the H-subsets formation process has for the HD, different strategies are first investigated in Sect. 5.1. Then, Sect. 5.2 analyzes the impact of using the archiving strategy within the (1+1) EA for all the studied formulations. Finally, a detailed comparative analysis is presented in Sect. 5.3.

5.1 H-Subsets Formation

An important issue for the proposed HD is how H amino acids are organized into the H-subsets (H_1 and H_2). Therefore, the following strategies are investigated:

- FIX: the first half of H amino acids in S are assigned to H_1, all others to H_2. For an odd number of Hs, the one in the middle is assigned randomly.
- RND: each H amino acid is assigned to H_1 or to H_2 with equal probability.
- DYN$_k$: based on RND, but the H-subsets are dynamically and independently recomputed after k iterations of the algorithm without achieving an improvement. Different values for k are explored, $k = \{0, 10, 20, 25, 30, 50\}$, where $k = 0$ refers to the recomputation of the H-subsets at each iteration.

[2] LD depends on parameter δ. This parameter was set to $\delta = 7$ as suggested in [8].

Figure 1 presents the OAP measure obtained by the HD when using the above described strategies. 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.

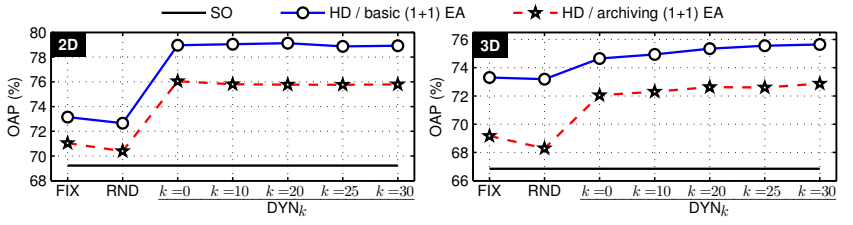

Fig. 1. Evaluating different strategies to form the H-subsets

It is evident from Fig. 1 that the proposed HD performed better in all cases compared to the conventional SO formulation. The highest OAP values were obtained when using the DYN_k strategy. That is, the ability of decomposition for allowing the algorithms to escape from local optima is further enhanced by changing the search landscape dynamically throughout the evolutionary process.

For the two-dimensional instances, no important differences in performance can be observed when varying k. Regarding the three-dimensional test cases, the algorithms responded positively to the increased value for k. The DYN_k strategy with $k = 30$ was adopted for the experiments presented in Sects. 5.2 and 5.3.

5.2 The Impact of Archiving

This section aims at investigating the impact of using the archiving (1+1) EA rather than the basic version of this algorithm. The results are presented in Fig. 2, which contrasts the performance of these algorithms (in terms of the OAP measure) when using the four studied HP model's formulations.[3]

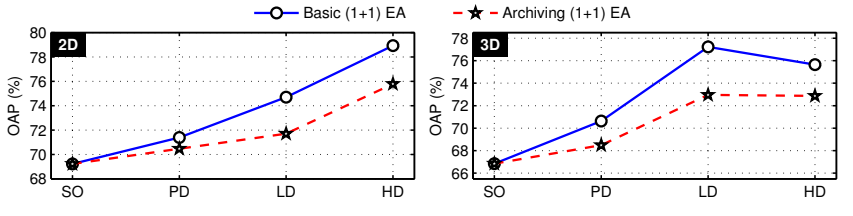

Fig. 2. Evaluating the impact of using the archiving strategy

From Fig. 2, it can be seen that an important increase in performance was obtained through multiobjectivization. The three multiobjective proposals (PD, LD and HD) improved the results for both the basic and the archiving (1+1) EA

[3] Although archiving is only useful in multiobjective scenarios, results of the archiving (1+1) EA applied to the SO formulation are shown only for illustrative purposes.

with respect to the conventional SO formulation. The proposed HD reached the highest OAP values at solving the two-dimensional instances. In contrast, the previously reported LD scored better results for the three-dimensional test cases.

Although competitive, the performance of PD, LD and HD was 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 [14, 18]. 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 [11]. 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 behavior of the algorithm.

5.3 Comparative Analysis

In all the cases, better results were obtained by using the basic (1+1) EA rather than the archiving (1+1) EA, see Sect. 5.2. Thus, the basic (1+1) EA is used for comparing in detail the four studied HP model's formulations. Tables 1 and 2 present, 2D and 3D respectively, the best (lowest) energy (β), its frequency (f) and the average energy (μ) obtained for each instance when using the different formulations. Also, the OAP measure is used to evaluate the overall performance of the approaches, see Sect. 4.2. The best (lowest) μ achieved for each instance, as well as the best (highest) OAP values, appear **shaded** in these tables.

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

			SO		PD		LD		HD	
Seq.	L	E^*	β (f)	μ	β (f)	μ	β (f)	μ	β (f)	μ
2d1	**18**	**-4**	-4 (4)	-2.70	-4 (6)	**-2.71**	-4 (3)	-2.69	-4 (4)	-2.70
2d2	**18**	**-8**	-8 (18)	-6.81	-8 (24)	-7.04	-8 (31)	-7.16	-8 (66)	**-7.65**
2d3	**18**	**-9**	-8 (11)	-7.00	-8 (48)	-7.45	-9 (2)	-7.39	-9 (28)	**-8.27**
2d4	**20**	**-9**	-9 (8)	-6.84	-9 (4)	-6.95	-9 (11)	-7.23	-9 (48)	**-8.19**
2d5	**20**	**-10**	-9 (3)	-6.92	-10 (2)	-7.08	-9 (1)	-7.06	-9 (7)	**-7.51**
2d6	**24**	**-9**	-8 (14)	-6.81	-9 (1)	-6.87	-9 (2)	**-7.30**	-9 (6)	-7.26
2d7	**25**	**-8**	-7 (26)	-5.79	-8 (6)	-5.90	-8 (7)	-6.17	-8 (21)	**-6.51**
2d8	**36**	**-14**	-13 (1)	-9.97	-13 (1)	-10.23	-13 (4)	-10.61	-12 (30)	**-11.00**
2d9	**48**	**-23**	-18 (5)	-14.23	-19 (2)	-15.20	-20 (2)	-16.29	-20 (3)	**-17.46**
2d10	**50**	**-21**	-18 (2)	-13.79	-18 (1)	-14.06	-19 (1)	-15.07	-18 (14)	**-16.25**
2d11	**60**	**-36**	-30 (2)	-24.39	-30 (7)	-25.43	-32 (1)	-27.80	-32 (2)	**-29.11**
2d12	**64**	**-42**	-29 (1)	-23.82	-30 (1)	-25.12	-30 (4)	-26.61	-32 (2)	**-27.99**
2d13	**85**	**-53**	-41 (1)	-33.81	-41 (1)	-34.54	-44 (1)	-38.09	-45 (1)	**-39.35**
2d14	**100**	**-48**	-41 (1)	-30.80	-39 (3)	-32.18	-39 (2)	-34.41	-40 (2)	**-35.40**
2d15	**100**	**-50**	-40 (1)	-31.71	-40 (3)	-32.70	-39 (7)	-34.97	-40 (7)	**-36.68**
OAP			69.22%		71.39%		74.70%		**78.93%**	

Table 2. Results for the basic (1+1) EA on three-dimensional benchmarks

			SO		PD		LD		HD	
Seq.	L	E^*	β (f)	μ	β (f)	μ	β (f)	μ	β (f)	μ
3d1	**20**	**-11**	-11 (57)	-10.48	-11 (69)	-10.64	-11 (94)	-10.94	-11 (100)	**-11.00**
3d2	**24**	**-13**	-13 (23)	-11.30	-13 (34)	-11.70	-13 (66)	-12.53	-13 (78)	**-12.75**
3d3	**25**	**-9**	-9 (57)	-8.48	-9 (70)	-8.65	-9 (95)	-8.95	-9 (99)	**-8.99**
3d4	**36**	**-18**	-18 (10)	-15.19	-18 (13)	-15.74	-18 (46)	**-16.97**	-18 (38)	-16.89
3d5	**46**	**-35**	-30 (2)	-23.87	-30 (1)	-25.38	-31 (1)	**-27.53**	-30 (1)	-27.26
3d6	**48**	**-31**	-29 (1)	-22.79	-29 (2)	-24.42	-31 (1)	**-26.66**	-29 (4)	-25.98
3d7	**50**	**-34**	-25 (6)	-20.64	-27 (1)	-22.07	-28 (1)	**-24.31**	-26 (4)	-23.58
3d8	**58**	**-44**	-35 (1)	-27.34	-36 (1)	-29.02	-36 (2)	**-31.98**	-35 (1)	-30.99
3d9	**60**	**-55**	-46 (1)	-37.20	-47 (1)	-40.03	-47 (3)	-42.88	-49 (1)	**-43.02**
3d10	**64**	**-59**	-45 (1)	-35.59	-46 (1)	-37.69	-50 (1)	**-43.29**	-46 (1)	-40.84
3d11	**67**	**-56**	-38 (2)	-30.17	-39 (2)	-32.65	-41 (1)	**-36.10**	-40 (1)	-35.16
3d12	**88**	**-72**	-47 (1)	-36.22	-49 (1)	-39.85	-53 (1)	**-46.13**	-48 (5)	-42.84
3d13	**103**	**-58**	-40 (1)	-29.97	-41 (1)	-31.31	-40 (1)	**-35.42**	-40 (1)	-34.25
3d14	**124**	**-71**	-43 (4)	-34.51	-48 (1)	-36.97	-50 (2)	**-43.98**	-46 (1)	-41.07
3d15	**136**	**-83**	-51 (1)	-37.26	-52 (1)	-42.11	-57 (1)	**-47.42**	-51 (4)	-45.47
OAP			66.84%		70.64%		**77.23%**		75.65%	

As shown in Table 1, the proposed HD reached the best average performance for 13 out of the 15 two-dimensional instances. This is reflected as an OAP increase of $(78.93 - 69.22) = 9.71\%$ with respect to the conventional SO formulation. HD allowed the OAP measure to be improved by 7.54% and by 4.23% with regard to the previously reported PD and LD multiobjectivizations, respectively.

The LD formulation achieved the lowest average energy for 11 out of the 15 three-dimensional benchmarks, see Table 2. The best results for the remaining four instances were obtained by the proposed HD. Although HD was inferior to LD in most of the three-dimensional test cases, with an OAP decrease of -1.58%, the results of this proposal are quite competitive; HD increased the OAP measure by 8.81% and by 5.01% over the SO and PD formulations, respectively.

Table 3 outlines how the formulations compare statistically with respect to each other in all the test cases. Each row in this table 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

Table 3. Statistical analysis for comparing the four studied HP model's formulations

	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		+							+		++			++			+	+	+	+	+	+	+	+	+	+	+	+	+	+	**20+ 0−**
LD/SO		+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**28+ 0−**
HD/SO		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**29+ 0−**
LD/PD					+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**23+ 0−**
HD/PD		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	**29+ 0−**
HD/LD		+	+	+	+		+	+	+	+	+	+	+	+	+	+	+				−	−	−		−	−	−	−	−	−	**15+ 9−**

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 3, the three multiobjective approaches significantly outperformed the conventional SO formulation in most of the cases. The proposed HD performed significantly better than SO in 29 out of the 30 test instances. Among the previously reported decompositions, the results of LD for 23 of the benchmarks were statistically superior to those obtained by PD. Compared with respect to PD, the proposed HD formulation significantly increased the performance of the algorithm for all but one of the instances (2d1). Finally, the proposed HD was statistically better than LD in 15 of the instances, while there was a significant difference in favor of LD for 9 of the three-dimensional test cases.

6 Conclusions

Multiobjectivization has proven to be a promising approach for solving difficult optimization problems. When applied to the hydrophobic-polar (HP) model, a simplified version of the protein structure prediction problem (PSP), this transformation has significantly improved the performance of search algorithms.

In this paper, a novel multiobjectivization for the HP model was proposed, called the H-subsets decomposition (HD). To the best of authors' knowledge, the HD formulation, together with the multiobjectivizations reported in [7, 8], represent the first efforts on the use of multiobjective optimization methods to address the HP model for protein structure prediction. The aim of this study was to investigate the impact of using the proposed HD multiobjectivization on the resolution of this problem. Through a comparative analysis, it has been shown that the HD formulation evaluates favorably in most of the cases with respect to the previously proposed multiobjectivizations for the HP model [7, 8].

Only basic evolutionary algorithms were adopted for the experiments presented in this paper. Nevertheless, from the obtained results it is expected that multiobjectivization can improve also the performance of established state-of-the-art algorithms for solving the HP model of the PSP. This issue needs to be thoroughly investigated in order to derive more general conclusions.

References

1. Berger, B., Leighton, T.: Protein Folding in the Hydrophobic-Hydrophilic (HP) Model is NP-complete. In: International Conference on Research in Computational Molecular Biology, pp. 30–39. ACM, New York (1998)
2. Brockhoff, D., Friedrich, T., Hebbinghaus, N., Klein, C., Neumann, F., Zitzler, E.: Do Additional Objectives Make a Problem Harder? In: Genetic and Evolutionary Computation Conference, pp. 765–772. ACM, London (2007)
3. Crescenzi, P., Goldman, D., Papadimitriou, C., Piccolboni, A., Yannakakis, M.: On the Complexity of Protein Folding. In: ACM Symposium on Theory of Computing, pp. 597–603. ACM, Dallas (1998)

4. Cutello, V., Narzisi, G., Nicosia, G.: A Multi-Objective Evolutionary Approach to the Protein Structure Prediction Problem. Journal of The Royal Society Interface 3(6), 139–151 (2006)
5. Day, R., Zydallis, J., Lamont, G.: Solving the Protein Structure Prediction Problem Through a Multi-Objective Genetic Algorithm. In: IEEE/DARPA International Conference on Computational Nanoscience, pp. 32–35 (2002)
6. Dill, K.: Theory for the Folding and Stability of Globular Proteins. Biochemistry 24(6), 1501–1509 (1985)
7. Garza-Fabre, M., Rodriguez-Tello, E., Toscano-Pulido, G.: Multiobjectivizing the HP Model for Protein Structure Prediction. In: Hao, J.-K., Middendorf, M. (eds.) EvoCOP 2012. LNCS, vol. 7245, pp. 182–193. Springer, Heidelberg (2012)
8. Garza-Fabre, M., Toscano-Pulido, G., Rodriguez-Tello, E.: Locality-based Multiobjectivization for the HP Model of Protein Structure Prediction. In: Genetic and Evolutionary Computation Conference, ACM, Philadelphia (2012)
9. Greiner, D., Emperador, J.M., Winter, G., Galván, B.: Improving Computational Mechanics Optimum Design Using Helper Objectives: An Application in Frame Bar Structures. In: Obayashi, S., Deb, K., Poloni, C., Hiroyasu, T., Murata, T. (eds.) EMO 2007. LNCS, vol. 4403, pp. 575–589. Springer, Heidelberg (2007)
10. Handl, J., Lovell, S.C., Knowles, J.D.: Investigations into the Effect of Multiobjectivization in Protein Structure Prediction. In: Rudolph, G., Jansen, T., Lucas, S., Poloni, C., Beume, N. (eds.) PPSN X. LNCS, vol. 5199, pp. 702–711. Springer, Heidelberg (2008)
11. Handl, J., Lovell, S.C., Knowles, J.D.: Multiobjectivization by Decomposition of Scalar Cost Functions. In: Rudolph, G., Jansen, T., Lucas, S., Poloni, C., Beume, N. (eds.) PPSN X. LNCS, vol. 5199, pp. 31–40. Springer, Heidelberg (2008)
12. Jähne, M., Li, X., Branke, J.: Evolutionary Algorithms and Multi-Objectivization for the Travelling Salesman Problem. In: Genetic and Evolutionary Computation Conference, pp. 595–602. ACM, Montreal (2009)
13. Jensen, M.: Helper-Objectives: Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimisation. Journal of Mathematical Modelling and Algorithms 3, 323–347 (2004)
14. Knowles, J., Corne, D.: Properties of an Adaptive Archiving Algorithm for Storing Nondominated Vectors. IEEE Transactions on Evolutionary Computation 7(2), 100–116 (2003)
15. Knowles, J.D., Watson, R.A., Corne, D.W.: Reducing Local Optima in Single-Objective Problems by Multi-objectivization. In: Zitzler, E., Deb, K., Thiele, L., Coello Coello, C.A., Corne, D.W. (eds.) EMO 2001. LNCS, vol. 1993, pp. 269–283. Springer, Heidelberg (2001)
16. Lochtefeld, D., Ciarallo, F.: Helper-Objective Optimization Strategies for the Job-Shop Scheduling Problem. Applied Soft Computing 11(6), 4161–4174 (2011)
17. Lopes, H.S.: Evolutionary Algorithms for the Protein Folding Problem: A Review and Current Trends. In: Smolinski, T.G., Milanova, M.G., Hassanien, A.-E. (eds.) Comp. Intel. in Biomed. & Bioinform. SCI, vol. 151, pp. 297–315. Springer, Heidelberg (2008)
18. López-Ibáñez, M., Knowles, J., Laumanns, M.: On Sequential Online Archiving of Objective Vectors. In: Takahashi, R.H.C., Deb, K., Wanner, E.F., Greco, S. (eds.) EMO 2011. LNCS, vol. 6576, pp. 46–60. Springer, Heidelberg (2011)

19. Segredo, E., Segura, C., Leon, C.: A Multiobjectivised Memetic Algorithm for the Frequency Assignment Problem. In: IEEE Congress on Evolutionary Computation, New Orleans, LA, USA, pp. 1132–1139 (2011)
20. Soares Brasil, C., Botazzo Delbem, A., Ferraz Bonetti, D.: Investigating Relevant Aspects of MOEAs for Protein Structures Prediction. In: Genetic and Evolutionary Computation Conference, pp. 705–712. ACM, Dublin (2011)
21. Vite-Silva, I., Cruz-Cortés, N., Toscano-Pulido, G., de la Fraga, L.G.: Optimal Triangulation in 3D Computer Vision Using a Multi-objective Evolutionary Algorithm. In: Giacobini, M. (ed.) EvoWorkshops 2007. LNCS, vol. 4448, pp. 330–339. Springer, Heidelberg (2007)
22. Zhao, X.: Advances on Protein Folding Simulations Based on the Lattice HP models with Natural Computing. Applied Soft Computing 8(2), 1029–1040 (2008)