

Multi-objectivization, Fitness Landscape Transformation and Search Performance: A Case of Study on the HP model for 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
+52 (834) 107 0220

Abstract

Multi-objectivization represents a current and promising research direction which has led to the development of more competitive search mechanisms. This concept involves the restatement of a single-objective problem in an alternative multi-objective form, which can facilitate the process of finding a solution to the original problem. Recently, this transformation was applied with success to the HP model, a simplified yet challenging representation of the protein structure prediction problem. The use of alternative multi-objective formulations, based on the decomposition of the original objective function of the problem, has significantly increased the performance of search algorithms. The present study goes further on this topic. With the primary aim of understanding and quantifying the potential effects of multi-objectivization, a detailed analysis is first conducted to evaluate the extent to which this problem transformation impacts on an important characteristic of the fitness landscape, neutrality. To the authors' knowledge, the effects of multi-objectivization have not been previously investigated by explicitly sampling and evaluating the neutrality of the fitness landscape. Although focused on the HP model, most of the findings of such an analysis can be extrapolated to other problem domains, contributing thus to the general understanding of multi-objectivization. Finally, this study presents a comparative analysis where the advantages of multi-objectivization are evaluated in terms of the performance of a basic evolutionary algorithm. Both the two- and three-dimensional variants of the HP model (based on the square and cubic lattices, respectively) are considered.

Keywords: Multi-objectivization, fitness landscape analysis, protein structure prediction, hydrophobic-polar model, multi-objective evolutionary algorithms, metaheuristics

1. Introduction

The term *multi-objectivization* was originally coined by Knowles *et al.* to refer to the process of reformulating a single-objective optimization problem in terms of two or more objective functions, *i.e.*, as a multi-objective problem [1]. This transformation can be either based on the addition of new *supplementary objectives* [2, 3], or it can be based on the *decomposition* of the original objective function of the problem [1, 4]. In either case, multi-objectivization may result in fundamental changes to the fitness landscape of the problem. Since the performance of search algorithms is dictated by their interaction with the underlying fitness landscape [5], multi-objectivization can thus significantly impact on the ability of these algorithms to solve a given optimization task.

It is commonly assumed that the higher the number of objective functions, the more difficult a problem is; and this is usually the case [6, 7]. A single-objective to multi-objective

transformation, however, has led to the development of more competitive search mechanisms. A considerable number of successful applications of multi-objectivization have been reported in the literature. For a recent review on applications of multi-objectivization, the reader can be referred to [8]. Multi-objectivization has been largely studied in the context of well-known combinatorial problems such as the traveling salesman problem [1, 2, 9, 10], the job-shop scheduling problem [2, 11], the bin packing problem [12, 13], the vehicle routing problem [14], and the shortest path and minimum spanning tree problems [15]. Also, multi-objectivization has found interesting applications in the fields of mobile communications [16, 17], computational mechanics [18], power system planning [19], structural topology optimization [20], computer aided manufacturing [21], robotics [22] and computer vision [23]. Finally, multi-objectivization has also been proposed to deal with bioinformatic problems, such as those related to gene regulatory networks [24] and protein structure prediction [25–33].

Recently, the concept of multi-objectivization was applied with success to the hydrophobic-polar (HP) model, a reduced representation of the protein structure prediction problem [34]. This model abstracts the fact that hydrophobicity is a major determinant of the folded state of proteins. Despite its limited bi-

*Corresponding author

Email addresses: mgarza@tamps.cinvestav.mx (Mario Garza-Fabre), gtoscano@tamps.cinvestav.mx (Gregorio Toscano-Pulido), ertello@tamps.cinvestav.mx (Eduardo Rodriguez-Tello)

ological significance, from the computational point of view this model still represents an interesting and challenging problem in combinatorial optimization [35, 36]. Three different multi-objectivization schemes for the HP model have been proposed, all of them based on the decomposition of the original energy (objective) function of the problem [37–39]. Decomposition introduces plateaus of incomparable solutions, an effect that can be exploited in order to overcome search difficulties such as that of becoming trapped in local optima [1, 4]. In this way, the use of alternative multi-objective formulations of the HP model has led to an important increase in the performance of search algorithms [37–39], motivating further research in this direction.

The present study significantly extends preliminary research regarding the multi-objectivization of the HP model. While previous analyses were concerned only with the benefits of multi-objectivization in terms of search performance [37–39], the primary goal of this study is to thoroughly investigate the potential effects that this transformation has on the characteristics of the problem. As pointed out before, multi-objectivization influences the comparability relation among solutions. As a means of illustrating and, to some extent, quantifying this effect, it is first evaluated how the comparability between the different defined fitness classes can be affected. The alteration in the comparability of solutions directly impacts on an essential property of the fitness landscape, *neutrality*. Hence, a detailed analysis is conducted with the aim of understanding multi-objectivization from a fitness landscape perspective, by focusing on neutrality. Finally, a comparative study is presented where the three multi-objectivization proposals are evaluated with respect to each other, and with respect to the conventional single-objective formulation of the HP model, in terms of the performance of a basic single-solution-based evolutionary algorithm.

The remainder of this document is organized as follows. Section 2 provides background concepts and sets the notation used in this study. The three studied multi-objectivization approaches for the HP model are described in detail in Section 3. Section 4 is devoted to the analysis of the effects of multi-objectivization. The comparative study which focuses on search performance is covered in Section 5. Finally, Section 6 discusses the main findings and presents the conclusions of this study. Appendices at the end of this document contain supplementary information with regard to implementation details of the considered search algorithms, performance measures, test instances, the methodology followed for the statistical significance analyses, and the utilized experimental platform.

2. Background concepts and notation

2.1. Single-objective and multi-objective optimization

Without loss of generality, a *single-objective optimization problem* can be formally stated as follows:

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}), \\ &\text{subject to} && \mathbf{x} \in \mathcal{X}_{\mathcal{F}}, \end{aligned} \quad (1)$$

where \mathbf{x} is a *solution vector*; $\mathcal{X}_{\mathcal{F}}$ is the *feasible set*, i.e., the set of all *feasible solution vectors* in the search space \mathcal{X} , $\mathcal{X}_{\mathcal{F}} \subseteq \mathcal{X}$; and $f : \mathcal{X} \rightarrow \mathbb{R}$ is the objective function to be optimized.

Similarly, a *multi-objective optimization problem* can be formally defined as follows:

$$\begin{aligned} &\text{Minimize} && \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x})]^T, \\ &\text{subject to} && \mathbf{x} \in \mathcal{X}_{\mathcal{F}}, \end{aligned} \quad (2)$$

where $\mathbf{f}(\mathbf{x})$ is the *objective vector* and $f_i : \mathcal{X} \rightarrow \mathbb{R}$ is the i -th objective function, $i \in \{1, 2, \dots, k\}$. Rather than searching for a single optimal solution, the task in multi-objective optimization is to identify a set of trade-offs among the conflicting objectives. More formally, the goal is to find a set of *Pareto-optimal solutions* \mathcal{P}^* , such that $\mathcal{P}^* = \{\mathbf{x}^* \in \mathcal{X}_{\mathcal{F}} \mid \nexists \mathbf{x} \in \mathcal{X}_{\mathcal{F}} : \mathbf{x} < \mathbf{x}^*\}$. The symbol “ $<$ ” denotes the *Pareto-dominance* relation [40]:

$$\begin{aligned} \mathbf{x} < \mathbf{x}' &\Leftrightarrow \forall i \in \{1, \dots, k\} : f_i(\mathbf{x}) \leq f_i(\mathbf{x}') \wedge \\ &\exists j \in \{1, \dots, k\} : f_j(\mathbf{x}) < f_j(\mathbf{x}'). \end{aligned} \quad (3)$$

If $\mathbf{x} < \mathbf{x}'$, then \mathbf{x} is said to *dominate* \mathbf{x}' . Otherwise, \mathbf{x}' is said to be *nondominated* with respect to \mathbf{x} , denoted by $\mathbf{x} \not< \mathbf{x}'$. The image of \mathcal{P}^* in the objective space is the so-called *Pareto-optimal front* or *trade-off surface*.

2.2. Multi-objectivization

Multi-objectivization refers to the process of reformulating an originally single-objective problem in terms of two or more objective functions [1].¹ Two main directions exist to perform such a transformation: (i) by incorporating additional information in the form of *supplementary objectives*, also referred to as artificial or helper objectives [2, 3]; or (ii) by means of the *decomposition* of the original objective function of the problem [1, 4]. In either case, the goal remains to solve the original problem, so that the original optima are to be also Pareto-optimal with regard to the alternative multi-objectivized formulation.

When multi-objectivization is based on the addition of supplementary objectives, the single-objective problem is restated as a multi-objective problem of the form $\mathbf{f}(\mathbf{x}) = [f(\mathbf{x}), g_1(\mathbf{x}), \dots, g_h(\mathbf{x})]^T$; where f is the original objective function of the problem and g_i denotes the i -th supplementary objective, $1 \leq i \leq h$. In the literature, this has been the most extensively studied approach to multi-objectivization. In a recent review [8], a distinction is made between multi-objectivization proposals where supplementary objectives are problem-dependent and are computed based solely on information from the solution under consideration [2, 9, 11, 18, 42], and those where they act as diversity measures [20, 22, 43–46]. Separate treatment is given also to those proposals where additional objectives are implemented to handle constraints [21, 47, 48].

In multi-objectivization by decomposition, the focus of this research, the original objective is fragmented into several different components, each to be treated as an objective function under the new alternative formulation of the problem. More formally, the problem is restated in terms of $d \geq 2$ objectives, $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_d(\mathbf{x})]^T$, such that the sum of all

¹The term *multi-objectivization* was originally coined by Knowles *et al.* [1], but the first studies on this kind of problem transformation date back to the work of Louis and Rawlins [41].

the new objectives equals the original objective function; *i.e.*, $f(\mathbf{x}) = \sum_{i=1}^d f_i(\mathbf{x})$, for all $\mathbf{x} \in \mathcal{X}$.² It has been demonstrated that the only possible effect of decomposition is the introduction of plateaus in the search landscape [4]. That is, originally comparable solutions may become incomparable (mutually nondominated) with regard to the new decomposed formulation. Some of the works reported in this direction include the following [1, 4, 23]. Also, multi-objective optimization approaches for the protein structure prediction problem fall into this category [25–31]. Note, however, that such approaches are based on detailed energy models. It was not until the proposal of the methods studied herein that multi-objectivization was applied to the particular HP model of this problem [37–39].

2.3. Fitness landscapes and neutrality

The notion of a fitness landscape, first introduced by Wright [49], has been found to be useful in understanding the most essential characteristics of certain optimization problems, or problem classes. By analyzing the fitness landscape, it is possible to gain further insight into problem difficulty as a means of explaining, or even predicting, the performance of search algorithms. Fitness landscape analysis is expected to provide important clues for guiding the development of more competitive search mechanisms, which are able to deal with (or to take advantage of) the particular characteristics of the given optimization task. Some fundamental definitions on this topic, which are relevant according to the scope of this study, are presented below. For a more comprehensive literature review on fitness landscapes analysis the reader can be referred to [50–54].

A *fitness landscape* can be generally defined in terms of a triplet $(\mathcal{X}, \mathcal{N}, \xi)$. The first element, \mathcal{X} , represents the set of all potential solutions to the problem, *i.e.*, the *search space*. The notion of connectedness among solutions in \mathcal{X} is introduced by the so-called *neighborhood structure*, $\mathcal{N} : \mathcal{X} \rightarrow 2^{\mathcal{X}}$, which maps each possible solution $\mathbf{x} \in \mathcal{X}$ to a set of solutions $\mathcal{N}(\mathbf{x}) \subseteq \mathcal{X}$. Hence, $\mathcal{N}(\mathbf{x})$ is referred to as the neighborhood of \mathbf{x} and each $\mathbf{x}' \in \mathcal{N}(\mathbf{x})$ is called a *neighbor* of \mathbf{x} . Finally, ξ denotes the evaluation scheme, consisting of 1) a measure (or set of measures) to serve as an indicator of the quality of the different solution candidates; and 2) a mechanism to impose an ordering relation given the adopted quality measure(s). As the evaluation scheme, in single-objective optimization a *fitness function* (usually directly related to the objective function of the problem) is considered and a simple ordering sets the preference relation among solutions.³ In the multi-objective context, however, a number of (conflicting) criteria determine the quality of solutions, so that defining an ordering relation is not as straightforward. The partial order induced by the Pareto-dominance relation is assumed in this study.

The fitness landscape of a problem can be studied in terms of different properties, being the *neutrality* property of particular

importance given the purposes of the present study. The standard definition of neutrality, in the single-objective case, refers to the degree to which a landscape contains connected areas of equal fitness [53]. Considering a broader notion to cover also the multi-objective case, neutrality can be understood as the result of the *incomparability* that the adopted evaluation scheme ξ induces. The term incomparability is used in this study to indicate the situation where no preferences can be imposed between a pair of solutions, so that these solutions are considered equivalent when evaluated under ξ . Two different solutions $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ are said to be *neutral* (incomparable), denoted by $neutral(\mathbf{x}_1, \mathbf{x}_2)$, if either they share the same fitness value (single-objective case), or they are nondominated, in the Pareto sense, with respect to each other (multi-objective case).

Having defined neutrality, a series of related basic concepts can be introduced as follows. The *neutral neighborhood* of a solution $\mathbf{x} \in \mathcal{X}$ is given by the subset of all its *neutral neighbors*: $\mathcal{N}_n(\mathbf{x}) = \{\mathbf{x}' \in \mathcal{N}(\mathbf{x}) \mid neutral(\mathbf{x}, \mathbf{x}')\}$. The total number of neutral neighbors of \mathbf{x} , *i.e.*, the cardinality of $\mathcal{N}_n(\mathbf{x})$, is known as the *neutrality degree* of \mathbf{x} , and the ratio of the neutrality degree to the size of the neighborhood is referred to as the *neutrality ratio*. A *neutral fitness landscape* is characterized by a large number of solutions presenting a high degree of neutrality. This leads to (potentially large) connected areas of incomparable solutions called *plateaus*, more formally referred to as *neutral networks*. Consider the *neutrality graph* $G = (\mathcal{X}, \mathcal{E}_n)$ where $\mathcal{E}_n = \{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{X}^2 \mid \mathbf{x}_2 \in \mathcal{N}_n(\mathbf{x}_1)\}$. Each connected component of the graph G corresponds to a different neutral network. In other words, a neutral network is a connected subgraph $G' = (\mathcal{X}', \mathcal{E}'_n)$ of G , $\mathcal{X}' \subseteq \mathcal{X}$ and $\mathcal{E}'_n \subseteq \mathcal{E}_n$, where 1) there exists a path connecting any pair of solutions $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}'$, and 2) there exists no edge $(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{E}_n \setminus \mathcal{E}'_n$ such that $\mathbf{x}_1 \in \mathcal{X}'$ and $\mathbf{x}_2 \in \mathcal{X} \setminus \mathcal{X}'$. The neutral network of a solution \mathbf{x} will be denoted as $NN(\mathbf{x})$. Finally, another important concept is that of a *neutral walk*. A neutral walk from \mathbf{x}_1 to \mathbf{x}_k refers to a sequence of solutions $\langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \rangle$ such that $\mathbf{x}_{i+1} \in \mathcal{N}_n(\mathbf{x}_i)$, $1 \leq i < k$. That is, a neutral walk represents a path on a neutral network.

2.4. The HP model for protein structure prediction

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 of a protein determines how it folds into a unique compact conformation responsible for its biological functioning. Among all the possible conformations that a protein can adopt, it is believed that the optimal conformation, often referred to as the native state, corresponds to the one minimizing the overall free-energy [55]. Thus, the *protein structure prediction* (PSP) problem, can be stated as the problem of finding the functional, energy-minimizing conformation for a protein given only its amino acid sequence.

Amino acids can be classified on the basis of their affinity for water. *Hydrophilic* or *polar* amino acids (*P*) are usually found at the outer surface of proteins. By interacting with the aqueous environment, *P* amino acids contribute to the solubility of the molecule. In contrast, *hydrophobic* or *nonpolar* amino

²Though other different decompositions are possible, this definition ensures that the original optimum coincides with one of the Pareto-optima in the multi-objective version of the problem [1].

³In this study, a fitness function is assumed to be always maximized (the goal is to search for the fittest solution candidate).

acids (H) tend to pack on the inside of proteins, where they interact with one another to form a water-insoluble core. This phenomenon, usually referred to as *hydrophobic collapse*, is a major driving force in protein folding, representing the reasoning and motivation behind the hydrophobic-polar (HP) model of the PSP problem studied in this paper [34, 56].

In the HP model, proteins are abstracted as chains of H - and P -type beads. Protein sequences, which are originally defined over a 20-letters alphabet, are now of the form $S = \langle a_1, a_2, \dots, a_\ell \rangle$, where $a_i \in \{H, P\}$ denotes the i -th amino acid and ℓ is the length of the sequence. A valid protein conformation is modeled as a *self-avoiding walk* on a given lattice, that is, as an embedding of the protein chain on the lattice such that the following two properties are satisfied: (i) *self-avoidance*, two different amino acids can not be mapped to the same lattice position; and (ii) *connectivity*, consecutive amino acids in S are to be also adjacent in the lattice. This research work focuses on both, the two-dimensional square and three-dimensional cubic lattices. With the aim of emulating the so-called hydrophobic collapse, the goal in the HP model is to maximize the interaction among H amino acids in the lattice. Such interactions are to be referred to as *H - H topological contacts*. Two H amino acids a_i and a_j are said to form a topological contact if they are nonconsecutive in S (i.e., $|j - i| \geq 2$) but adjacent in the lattice. The objective is thus to find a valid structure where the number of H - H topological contacts, $HHtc$, is maximized. Adhering to the notation of the field, an energy function, to be minimized, is defined as the negative of $HHtc$.

Let \mathcal{X} be the set of all potential protein conformations, i.e., the search space, and let $\mathcal{X}_{\mathcal{F}}$ denote the subset of all the feasible states ($\mathcal{X}_{\mathcal{F}} \subseteq \mathcal{X}$). PSP under the HP model can be formally defined as the problem of finding $\mathbf{x}^* \in \mathcal{X}_{\mathcal{F}}$ such that $E(\mathbf{x}^*) = \min\{E(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}_{\mathcal{F}}\}$. The energy function $E : \mathcal{X} \rightarrow \mathbb{R}$ maps each possible conformation $\mathbf{x} \in \mathcal{X}$ to an energy value; formally,

$$E(\mathbf{x}) = \sum_{a_i, a_j} e(a_i, a_j), \quad (4)$$

where

$$e(a_i, a_j) = \begin{cases} -1, & \text{if } a_i \text{ and } a_j \text{ are both } H \text{ and} \\ & \text{they form a topological contact;} \\ 0, & \text{otherwise.} \end{cases}$$

As an example, the optimal structure for a protein sequence of length $\ell = 20$ on the two-dimensional square lattice is presented in Figure 1. This example corresponds to sequence 2d4, one of the test cases considered in this study, see Appendix A.1.

An *internal coordinates representation* based on *absolute moves* has been implemented in this study. A protein conformation is encoded as a sequence of moves specifying the lattice position for each amino acid with regard to the preceding one (the position of the first amino acid is fixed). On the three-dimensional cubic lattice, conformations are thus defined as sequences in $\{F, B, L, R, U, D\}^{\ell-1}$, to denote the forward, backward, left, right, up and down moves from one amino acid to the next. Only moves $\{F, B, L, R\}$ are considered when using the two-dimensional square lattice, see Figure 2.

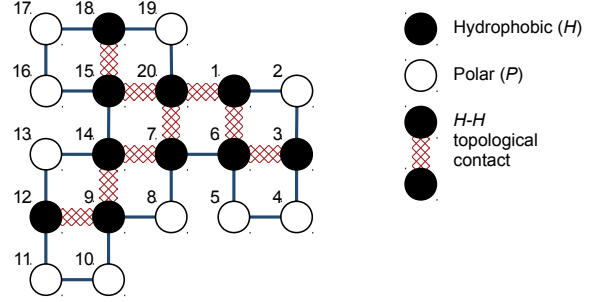


Figure 1: Optimal conformation for sequence 2d4 of length $\ell = 20$ on the two-dimensional square lattice. Black and white beads denote H and P amino acids, respectively. Amino acids have been numbered from 1 to ℓ according to their positions in the protein sequence S . The energy of this conformation is $E = -9$, since there are 9 H - H topological contacts, $HHtc = 9$.



Figure 2: Internal coordinates representation based on absolute moves. Encoding scheme (left). Example conformation encoded as FLLFRFRB (right).

The prediction of protein structures based on the HP model is a hard combinatorial optimization problem which has been proved to be NP-complete [35, 36]. An extensive literature exists on the use of metaheuristics to address this problem, including genetic algorithms [57, 58], memetic and hybrid algorithms [59–61], tabu search [62], ant colony optimization [63, 64], immune-based algorithms [65, 66], particle swarm optimization [67, 68], differential evolution [69, 70], estimation of distribution algorithms [71, 72], artificial plant optimization [73], and firefly-inspired algorithms [74].

3. Multi-objectivization of the HP model

Three different multi-objectivization schemes for the HP model are analyzed in this study: the parity decomposition [37], the locality decomposition [38] and the H-subsets decomposition [39]. As their name indicates, these approaches are based on the decomposition of the original energy (objective) function of the HP model. Decomposition, as discussed in Section 2.2, has the potential effect of introducing incomparability among solution candidates. In this way, these alternative multi-objective formulations of the problem can be useful as a means of accepting degrading moves (i.e., replacement of a solution with an inferior one), and thus can be implemented as a mechanism to prevent search algorithms from becoming trapped in local optima. The parity, locality and H-subsets decompositions are described in detail in Sections 3.1, 3.2 and 3.3, respectively.

3.1. Parity decomposition

In the two-dimensional square and three-dimensional cubic lattices, the two variants of the HP model covered by this re-

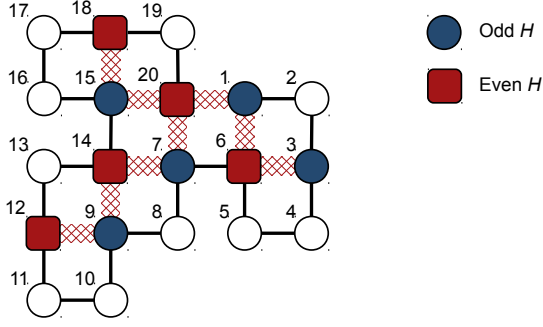


Figure 3: Parity decomposition. Interactions on the lattice are only possible between amino acids whose sequence positions are of opposite parity. In this example conformation, $f_1 = 0$ and $f_2 = -9$.

search project, adjacencies (topological contacts) are only possible between amino acids whose positions in the protein sequence are of opposite parity, see Figure 3. Based on this fact, a two-objective formulation $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x})]^T$ is defined over the set of all potential protein conformations $\mathbf{x} \in \mathcal{X}$:

$$f_1(\mathbf{x}) = \sum_{a_i, a_j} e(a_i, a_j), \quad i \equiv 0 \pmod{2}, i < j; \quad (5)$$

$$f_2(\mathbf{x}) = \sum_{a_i, a_j} e(a_i, a_j), \quad i \equiv 1 \pmod{2}, i < j; \quad (6)$$

where both $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are to be minimized and $e(a_i, a_j)$ represents the conventional energy contributions of the HP model as defined in Section 2.4. That is, the objective function f_1 accounts only for H - H topological contacts between pairs of amino acids a_i, a_j , where i , the sequence position of amino acid a_i , is even ($i < j$). On the contrary, f_2 is defined for those cases where such the i -th sequence position is odd. Note that the sum of the two new alternative objectives equals the conventional energy function of the HP model presented in Section 2.4 (i.e., $E(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$), which is in accordance with the decomposition approach for multi-objectivization, see Section 2.2. Figure 3 presents the optimal conformation for protein sequence 2d4 on the two-dimensional square lattice. In the particular case of this conformation, the values for the objective functions are $f_1 = 0$ and $f_2 = -9$.

3.2. Locality decomposition

In this multi-objectivization scheme, the conventional energy function of the HP model is decomposed based on the locality notion of amino acid interactions. An H - H topological contact between amino acids a_i and a_j can be considered to represent either a *local* or a *nonlocal* interaction. This classification depends upon whether or not the sequence distance between a_i and a_j (i.e., $|j - i|$) is within a given maximum δ , see Figure 4.

From this, a two-objective problem formulation, $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x})]^T$, is defined for every potential pro-

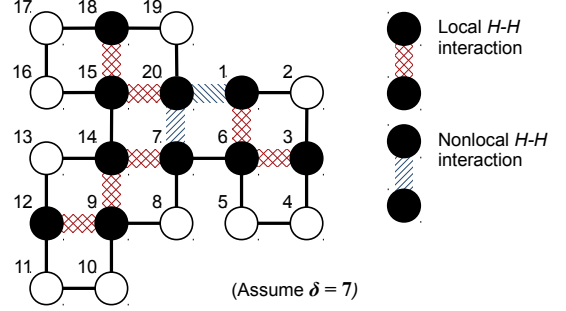


Figure 4: Locality decomposition ($\delta = 7$). This structure presents seven local and two nonlocal H - H interactions. Therefore, in the particular case of this example, $f_1 = -7$ and $f_2 = -2$.

tein conformation $\mathbf{x} \in \mathcal{X}$:

$$f_1(\mathbf{x}) = \sum_{a_i, a_j} e(a_i, a_j), \quad j - i \leq \delta, i < j; \quad (7)$$

$$f_2(\mathbf{x}) = \sum_{a_i, a_j} e(a_i, a_j), \quad j - i > \delta, i < j; \quad (8)$$

where functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are both to be minimized and $e(a_i, a_j)$ denotes the conventional energy contributions defined previously in Section 2.4. Thus, the objective function f_1 is defined for all the local interactions, whereas the objective function f_2 accounts for the nonlocal ones. In this way, the evaluation of the example conformation illustrated in Figure 4 under this alternative formulation leads to objective values $f_1 = -7$ and $f_2 = -2$. Note that $E(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, which is consistent with the decomposition approach for multi-objectivization, as defined in Section 2.2.

It is worthy to mention that parameter δ plays a decisive role for the behavior of this proposal. In [38], an analysis was conducted in order to investigate the influence of varying this parameter on the performance of two different evolutionary algorithms. All odd values in the range $[3, 21]$ were considered in the analysis.⁴ As a result, the best performance for the evaluated algorithms were observed when δ assumes low values around 7.

3.3. H-subsets decomposition

In the H-subsets decomposition, all H amino acids in the protein sequence are first assigned to one of two possible groups, \mathcal{H}_1 or \mathcal{H}_2 . In this study, \mathcal{H}_1 and \mathcal{H}_2 are to be referred to as the H-subsets and such an assignment of H amino acids to these groups is to be called the H-subsets formation process. Figure 5 illustrates one of different H-subsets formation strategies which are described later at the end of this section.

Once the H-subsets have been formed, an alternative two-objective formulation $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x})]^T$ of the problem can

⁴In the two-dimensional square and the three-dimensional cubic lattices, a topological contact can only occur if the sequence distance between the amino acids is odd and at least equal to 3.

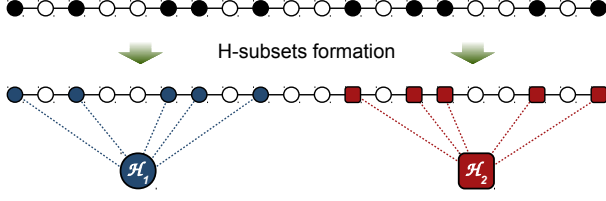


Figure 5: H-subsets decomposition. H-subsets formation (FIX strategy).

be defined as follows ($\mathbf{x} \in \mathcal{X}$):

$$f_1(\mathbf{x}) = \sum_{a_i, a_j \in \mathcal{H}_1} e(a_i, a_j) + \sum_{a_i, a_j \in \mathcal{H}_2} e(a_i, a_j), \quad (9)$$

$$f_2(\mathbf{x}) = \sum_{a_i \in \mathcal{H}_1, a_j \in \mathcal{H}_2} e(a_i, a_j), \quad (10)$$

where $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are both minimization functions and $e(a_i, a_j)$ denotes the conventional energy contributions of the HP model (Section 2.4). In this way, function f_1 accounts for H - H topological contacts where the two amino acids belong to the same H-subset, either to \mathcal{H}_1 or to \mathcal{H}_2 . On the contrary, f_2 is defined for H - H topological contacts between amino acid pairs where each amino acid belongs to a different H-subset. Notice that $E(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, which adheres to the definition of the decomposition approach for multi-objectivization, as provided in Section 2.2.

Given the assignment of H amino acids to the H-subsets illustrated in Figure 5, the structure in Figure 6 presents four H - H topological contacts defined between amino acids belonging to the same H-subset, while the remaining five H - H interactions occur between amino acids from different H-subsets. In this example the objective values are thus $f_1 = -4$ and $f_2 = -5$.

The H-subsets formation process plays a major role for this decomposition proposal. Different strategies can be adopted in order to accomplish this task; three of them have been explored in [39], namely (i) FIX: the first half of H amino acids in S are assigned to \mathcal{H}_1 , all others to \mathcal{H}_2 , as shown in Figure 5; (ii) RND: each H amino acid can be assigned to \mathcal{H}_1 or to \mathcal{H}_2 with equal probability; and (iii) DYN_k: it is based on the RND strategy, but the H-subsets are recomputed after k iterations of the search algorithm without achieving an improvement. From the results reported in [39], the best performance can be achieved by implementing the DYN_k strategy. This suggests that the effect of decomposition for allowing algorithms to escape from local optima can be further enhanced by changing the landscape dynamically throughout the search process.

4. Effects of multi-objectivization

This section is devoted to investigating the potential effects that can be achieved by multi-objectivization. Although three different multi-objectivization schemes for the HP model are covered by this research work, only the locality decomposition (defined in Section 3.2), using a value of $\delta = 7$, is considered in this section due to the high computational demands of the conducted analyses. The locality decomposition has been found, in

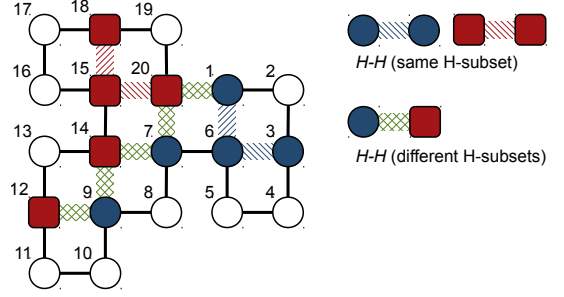


Figure 6: H-subsets decomposition. Four H - H topological contacts are defined between amino acids from the same H-subset, while the other five are given between amino acids from different H-subsets. Therefore, $f_1 = -4$ and $f_2 = -5$.

the authors' previous work [38, 39], to provide a quite promising behavior, as it will also be shown in subsequent sections of this document. For convenience, hereafter the locality decomposition will be simply referred to as the multi-objective (MO) formulation of the problem. Similarly, two (relatively) small test instances, sequences 2d4 and 3d1, are investigated in this section (refer to Appendix A.1 for details).⁵ It is expected, however, that other different multi-objectivization proposals and test cases can be explored with similar results.

As stated in Section 2.4, the quality of a candidate solution in the HP model is evaluated in terms of an energy function, E , defined as the negative of the total number of H - H topological contacts that the encoded protein structure presents, $HHtc$. Nevertheless, the use of positive rather than negative values, as well as the adoption of the term fitness (to be maximized) rather than that of energy (to be minimized), is considered more appropriate for the analysis here reported. Therefore, in the remainder of this section the fitness of a solution \mathbf{x} , $Fitness(\mathbf{x})$, will assume the value of $HHtc(\mathbf{x})$, i.e.,

$$Fitness(\mathbf{x}) = HHtc(\mathbf{x}) = -E(\mathbf{x}). \quad (11)$$

It is worthy to mention that the term fitness is used in this study to refer to the quality of solutions under the conventional single-objective (SO) evaluation scheme of the HP model.⁶ In addition, it is important to briefly introduce the concept of a *fitness class*; a solution $\mathbf{x} \in \mathcal{X}_{\mathcal{F}}$ will be said to belong to the fitness class c if it presents a fitness value of $Fitness(\mathbf{x}) = c$.

The analyses conducted in this section are based on an initial set of sampled solutions. Hence, the implemented sampling methodology is first introduced in Section 4.1. In Section 4.2, the effects of multi-objectivization are investigated in terms of how this transformation influences the comparability relation among solutions. Finally, Section 4.3 evaluates the extent to which such an alteration in the comparability of solutions can be translated into fundamental changes on the fitness landscape structure of the problem.

⁵Notwithstanding, given the absolute moves encoding described in Section 2.4, the size of the search space for these (relatively) small instances is huge, namely 4^{19} for 2d4, and 6^{19} for 3d1.

⁶Although alternative multi-objective formulations of the HP model are studied, the goal remains always to solve the original single-objective problem.

Algorithm 1 Sampling of the initial solution sets.

Input: M **Output:** S

```
1:  $S \leftarrow \emptyset$ 
2:  $\mathcal{FC} \leftarrow \{\text{Fitness}(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}_{\mathcal{F}}\}$ 
3: while  $|S| < M$  do
4:    $\mathcal{U} \leftarrow \text{search\_algorithm}()$ 
5:   for all  $c \in \mathcal{FC}$  do
6:      $\mathcal{U}_c \leftarrow \{\mathbf{x} \in \mathcal{U} \mid \text{Fitness}(\mathbf{x}) = c\}$ 
7:      $\hat{\mathbf{x}} \leftarrow \arg \max_{\mathbf{x} \in \mathcal{U}_c} \text{diversity}(\mathbf{x}, S)$ 
8:     if  $\text{diversity}(\hat{\mathbf{x}}, S) > 0$  then
9:        $S \leftarrow S \cup \{\hat{\mathbf{x}}\}$ 
10:    end if
11:  end for
12: end while
```

4.1. Sampling of initial solutions

The implemented sampling strategy was conceived by taking into account the following considerations: (i) a total of M different feasible solutions for the given problem instance are to be generated; (ii) the M generated solutions are to be, if possible, evenly distributed over the different available fitness classes (all fitness classes should be well represented in the collected sample); and finally, (iii) the diversity among solutions belonging to the same fitness class should be maximized.

Algorithm 1 outlines the adopted sampling strategy. The procedure starts by initializing the sample set S and by identifying the set of all possible fitness classes for the given problem instance, \mathcal{FC} (lines 1 and 2 in Algorithm 1). Iteratively, a search algorithm is executed and all solutions that this algorithm reaches during the search process are kept in \mathcal{U} (line 4). Then, the subset \mathcal{U}_c of solutions in \mathcal{U} belonging to each possible fitness class $c \in \mathcal{FC}$ is identified (line 6). Finally, the solution $\hat{\mathbf{x}} \in \mathcal{U}_c$ that best contributes to increasing the diversity in S , if any, is included in the sample (lines 7 to 9). This process continues until completing the required sample.

Any metaheuristic could be implemented as the embedded search method. An Iterated Local Search (ILS) algorithm [75], based on the SO problem formulation, was used in this study; refer to Appendix B.2 for details. Due to its distinctive exploration behavior, the ILS method can potentially reach a different local optimum at each iteration. Each time the ILS was invoked during the sampling procedure, this algorithm was allowed to run for a total of 5×10^5 solution evaluations.

The diversity contribution estimates have been partially based on the diversification mechanism proposed by Chira [76]. Instead of measuring diversity in genotype (encoding) space, in [76] diversity was computed from the *contact fingerprint* of candidate solutions. The contact fingerprint for a solution is given by the binary vector \mathbf{cf} , where each component $cf_i \in \{0, 1\}$ indicates whether a particular pair of amino acids in the encoded structure defines a topological contact or not. Vector \mathbf{cf} considers as many components as the total number of amino acid pairs which can potentially form a topological con-

Table 1: Sample sets generated for instances 2d4 and 3d1. Instance 2d4 involves 10 different fitness classes, and instance 3d1 involves 12.

Seq.	Fitness class												Total
	0	1	2	3	4	5	6	7	8	9	10	11	
2d4	119	118	119	119	119	119	119	119	47	2	-	-	1,000
3d1	84	83	83	83	83	83	83	84	84	84	83	83	1,000

tact.⁷ The use of the contact fingerprint rather than the encoding of solutions certainly fosters the development of more effective diversity promotion mechanisms. This can be explained by the fact that very different encodings may represent the same protein structure (after rotation or reflection). Note, however, that significantly different structures may also present the same contact fingerprint vector if they share the same set of topological contacts. This has motivated the use of a more fine-grained version, referred to in this study as the *distance fingerprint*.

The distance fingerprint for a given solution is defined by the vector \mathbf{df} , each of whose components df_i measures the distance between the lattice coordinates of a particular pair of amino acids. The Manhattan distance was employed for this sake. A total of $\binom{\ell}{2} - 2\ell + 3$ components describe the distance fingerprint vector \mathbf{df} ; *i.e.*, only amino acid pairs (a_i, a_j) such that $|j - i| \geq 3$ require to be considered. Finally, the diversity contribution for a new candidate \mathbf{x} with respect to the already collected sample S , $\text{diversity}(\mathbf{x}, S)$, has been computed as the minimum Hamming distance (H_d) between the distance fingerprint vector of \mathbf{x} and that of any $\mathbf{x}' \in S$ with the same fitness value as \mathbf{x} . Formally,

$$\begin{aligned} \text{diversity}(\mathbf{x}, S) = \min\{ & H_d(\mathbf{df}(\mathbf{x}), \mathbf{df}(\mathbf{x}')) \\ & \mid \mathbf{x}' \in S \wedge \\ & \text{Fitness}(\mathbf{x}) = \text{Fitness}(\mathbf{x}') \}. \end{aligned} \quad (12)$$

The size of the sample was set to $M = 1,000$ for both the 2d4 and 3d1 instances. Ideally, it is expected to generate a sample such that about $M/|\mathcal{FC}|$ different solutions represent each possible fitness class. This was the case of the sample set constructed for the three-dimensional instance 3d1, where a total of 83 or 84 different solutions were obtained for each of the $|\mathcal{FC}| = 12$ available fitness classes, see Table 1. Note, however, that due to the funnel-like energy landscape which characterizes the HP model [77], not all fitness classes for some of the instances can be equally sampled. As detailed in Table 1, only a reduced number of solutions with a high fitness value ($\text{Fitness} = 8$ and $\text{Fitness} = 9$) were obtained when sampling the search space of the 2d4 instance, so that a greater number of representatives for the remaining fitness classes were accepted in order to complete the M required solutions.

4.2. Incomparability of solutions

The three multi-objectivization proposals considered in this study are all of them based on the decomposition of the original energy (objective) function of the HP model, see Section

⁷In order for an amino acid pair (a_i, a_j) to form a topological contact, i and j need to be of opposite parity and $|j - i| \geq 3$.

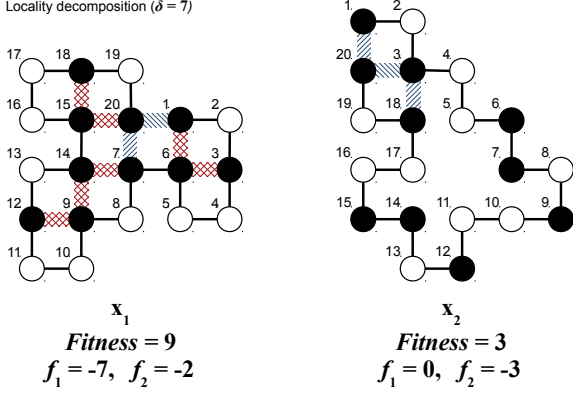


Figure 7: This figure illustrates how a pair of originally comparable solutions \mathbf{x}_1 (left) and \mathbf{x}_2 (right), with $\text{Fitness}(\mathbf{x}_1) = 9$ and $\text{Fitness}(\mathbf{x}_2) = 3$, can become incomparable by multi-objectivization (locality decomposition, $\delta = 7$).

3. As stated in Section 2.2, the only possible effect that can be achieved through decomposition is that originally comparable solutions may become incomparable (nondominated in terms of the Pareto-dominance relation) under the new multi-objective formulation of the problem. To illustrate this, consider the example provided in Figure 7. In this figure, conformation \mathbf{x}_1 (to the left) presents 9 *H-H* topological contacts, while conformation \mathbf{x}_2 (to the right) involves only 3. These originally comparable solutions (*i.e.*, \mathbf{x}_1 is clearly superior to \mathbf{x}_2) have become mutually nondominated when comparing them under the multi-objective formulation defined by the locality decomposition ($\delta = 7$). The goal of this section is not only to illustrate such an effect of decomposition, but also to explore the extent to which solutions belonging to different fitness classes can become incomparable as a consequence of this transformation.

An experiment was performed as follows. A sample set of $M = 1,000$ different solution candidates was generated by implementing the methodology detailed previously in Section 4.1. Then, all possible pairwise comparisons among the sampled solutions were performed. From this, it was computed the ratio between the number of incomparable solution pairs found and the total number of pairwise comparisons carried out; this measure is to be referred to as the *incomparability ratio* (IR). IR is thus defined in the range $[0, 1]$, and $\text{IR} = 1$ indicates that all the evaluated solution pairs were found to be incomparable. This experiment was replicated for both the conventional single-objective HP model formulation, SO, and the alternative multi-objective formulation, MO. The comparison of solutions under the MO formulation relies on the Pareto-dominance relation and the locality decomposition. Both the 2d4 and 3d1 instances were considered. The obtained results are summarized in Table 2. As it can be seen from this table, 70,616 out of the $\binom{1,000}{2} = 499,500$ total pairs of sampled solutions for instance 2d4 became incomparable when evaluated under the MO formulation. This represents an IR increase of 0.14 with regard to the conventional SO formulation. Similarly, multi-objectivization increased the IR measure by 0.13 when focusing on the 3d1 instance. Such an increase of 0.13 results from the 63,760 comparable solution pairs for which the original prefer-

Table 2: Incomparability ratio (IR) presented by the studied SO and MO formulations of the HP model. For both the 2d4 and 3d1 test instances, a total of $\binom{1,000}{2} = 499,500$ solution pairs have been evaluated.

	SO	MO	Increase (MO-SO)
2d4	0.11 (57, 132)	0.26 (127, 748)	0.14 (70, 616)
3d1	0.08 (41, 168)	0.21 (104, 928)	0.13 (63, 760)

ence relation has been suppressed.⁸

Finally, the results obtained using the MO formulation are broken down in Figures 8 and 9 in order to gain further insights into the likelihood of incomparability taking place among different fitness classes. Figures 8 and 9 show (for the 2d4 and 3d1 instances, respectively) the IR measure computed separately for each possible pair of fitness classes (*i.e.*, fitness classes with respect to the conventional SO formulation). Heat maps in these figures are symmetric along the diagonal. From Figures 8 and 9, it is possible to see that multi-objectivization makes incomparability possible even for pairs of solutions which are distant with respect to their fitness values. As an example, Figure 8 highlights that incomparability has been introduced between fitness classes 3 and 9 of instance 2d4 (as illustrated in Figure 7). A similar scenario can be found with regard to fitness classes 4 and 10 from the collected sample for sequence 3d1, see Figure 9. Note, however, that the closer the fitness classes for the selected pairs of solutions, the higher the IR values indicated in these figures (the highest IR values appear close to the diagonal). This can be understood by the fact that the increase in the fitness distance between a pair of solutions increases also the probability for a solution to be dominated (in the Pareto sense) by the other. Finally, it should be observed that in the case of both the 2d4 and 3d1 instances, no solution at fitness class 0 became incomparable with respect to any other solution from a higher fitness class. This is due to the fact that any solution \mathbf{x}_1 with $\text{Fitness}(\mathbf{x}_1) = 0$, $f_1(\mathbf{x}_1) = 0$ and $f_2(\mathbf{x}_1) = 0$ after applying decomposition, will always be dominated by any other solution \mathbf{x}_2 with $\text{Fitness}(\mathbf{x}_2) > 0$, no matter how $\text{Fitness}(\mathbf{x}_2)$ is decomposed into the new set of objectives.

4.3. Fitness landscape analysis

As discussed in Section 4.2, multi-objectivization by decomposition exerts an influence on the comparability relation over the search space, in such a way that solutions from different fitness classes may become incomparable when evaluated under the new multi-objective formulation of the problem. This notion of incomparability is equivalent to that of neutrality used in the context of fitness landscapes analysis. Two solutions are said to be neutral (*i.e.*, incomparable) if either they share the same fitness value (single-objective case), or they are Pareto-nondominated with respect to each other (multi-objective case), see Section 2.3. Hence, the potential effect of multi-objectivization, previously described in terms of introducing incomparability among solutions, will be referred in

⁸Note that, by definition, only solutions having the same fitness value are incomparable under the SO formulation of the problem.

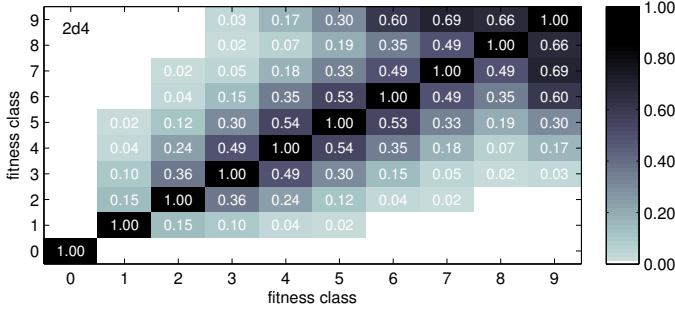


Figure 8: IR among the different fitness classes. 2d4 test instance.

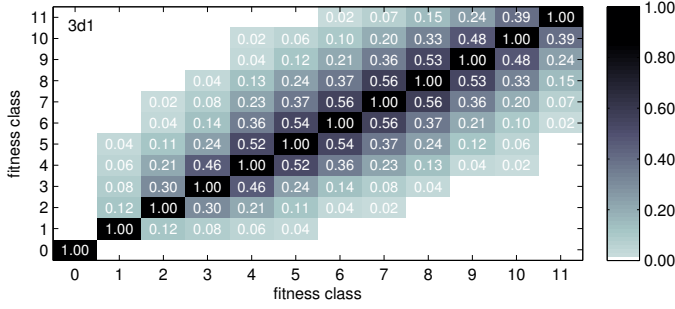


Figure 9: IR among the different fitness classes. 3d1 test instance.

this section to as that of increasing the neutrality in the fitness landscape. This section is intended to contribute in understanding and, to some extent, quantifying such an effect of multi-objectivization on the HP model’s fitness landscape.

As detailed in Section 2.3, three important components define a fitness landscape (X, \mathcal{N}, ξ) . While the search space X and the neighborhood structure \mathcal{N} were kept constant in this study,⁹ ξ has been varied from the conventional single-objective, SO evaluation scheme of the HP model to the alternative multi-objective, MO one (based on the locality decomposition and the Pareto-dominance relation). By analyzing and comparing the landscapes induced by the SO and MO evaluation schemes, it will then be possible to evaluate the extent to which multi-objectivization has impacted on essential problem characteristics, those related to neutrality. Neutrality is here investigated by evaluating different properties of neutral networks (NNs); since neutral fitness landscapes, as those in the HP model, are known to be mainly described by their NNs [78]. Nevertheless, NNs in a neutral fitness landscape can be of a considerable size, so that their exhaustive exploration becomes computationally prohibitive even for relatively small problem instances. In the literature, NNs are usually sampled through neutral walks, *i.e.*, series of (neutral) neighboring solutions. In this study, however, an alternative approach was taken, as described below.

Given a sample set \mathcal{S} of M different solution candidates, collected following the methodology previously detailed in Section 4.1, the neutral network $NN(\mathbf{x})$ for each solution $\mathbf{x} \in \mathcal{S}$

⁹ X is given by the absolute moves encoding (Section 2.4). Likewise, $\mathcal{N}(\mathbf{x})$ is defined by all solutions which can be reached through a single change in the encoding of \mathbf{x} . Thus, $|\mathcal{N}(\mathbf{x})| = 3(\ell - 1)$ and $|\mathcal{N}(\mathbf{x})| = 5(\ell - 1)$ in the two- and three-dimensional cases, ℓ denoting the length of the protein sequence.

Algorithm 2 $pNN()$ - Partial NN computation

Input: \mathbf{x} , $depthLevel$, $maxDepth$

Output: $NN(\mathbf{x})$

- 1: $NN(\mathbf{x}) \leftarrow (\mathcal{V}, \mathcal{E}) : \mathcal{V} = \{\mathbf{x}\}, \mathcal{E} = \emptyset$
- 2: **if** $depthLevel < maxDepth$ **then**
- 3: **for all** $\mathbf{x}' \in \mathcal{N}_n(\mathbf{x})$ **do**
- 4: $NN(\mathbf{x}') \leftarrow pNN(\mathbf{x}', depthLevel + 1, maxDepth)$
- 5: $NN(\mathbf{x}) \leftarrow NN(\mathbf{x}) \cup NN(\mathbf{x}')$
- 6: $\mathcal{E} \leftarrow \mathcal{E} \cup \{(\mathbf{x}, \mathbf{x}')\}$
- 7: **end for**
- 8: **end if**

has been partially computed based on the $pNN()$ procedure outlined in Algorithm 2. As shown in this algorithm, $NN(\mathbf{x})$ is constructed recursively in a depth-first manner by allowing this procedure to reach a maximum defined depth level ($maxDepth$). The initially given solution \mathbf{x} is assumed to be at depth level 0, so that $depthLevel = 0$ is used in the first call to $pNN()$. At each call to the $pNN()$ method, $NN(\mathbf{x})$ is first initialized to the graph containing no edges and including the provided solution \mathbf{x} as the only node (line 1 in Algorithm 2). If the maximum allowed depth level has not been reached (line 2), the sub-network $NN(\mathbf{x}')$ for every neutral neighbor \mathbf{x}' of \mathbf{x} is obtained from a subsequent execution of the $pNN()$ method (by giving \mathbf{x}' as the new starting point and by increasing the value of $depthLevel$, see line 4). The resulting sub-network $NN(\mathbf{x}')$ is then merged with the parent network $NN(\mathbf{x})$ by means of a graph union operation, here denoted as \cup (line 5).¹⁰ Finally, edge $(\mathbf{x}, \mathbf{x}')$ is included in $NN(\mathbf{x})$ in order to establish the linkage between $NN(\mathbf{x})$ and the $NN(\mathbf{x}')$ sub-network. Partially computing the NN for a given solution \mathbf{x} , is equivalent to traversing all possible neutral walks departing from \mathbf{x} , by restricting the length of the walks to the maximum defined depth level ($maxDepth$).

The 2d4 and 3d1 test instances have been considered for this analysis, and the size of the initial sample sets was fixed to $M = 1,000$ in both cases. Thus, a total of 1,000 (potentially different) NNs for each of the instances have been explored by using both, the SO and MO evaluation schemes, as the basis for neutrality verification. In this way, changing the problem formulation from SO to MO will be reflected as an alteration in the properties of the sampled NNs. In the remainder of this section, the neutral network for a given solution \mathbf{x} will be either referred to as $NN_{SO}(\mathbf{x})$ or $NN_{MO}(\mathbf{x})$, depending on whether the neutrality relation among solutions was determined based on the SO or MO evaluation schemes during the network computation. Finally, in order to overcome the high computational cost of the conducted analysis, the maximum allowed depth level was set to $maxDepth = 10$ and $maxDepth = 7$ for the 2d4 and 3d1 test sequences, respectively.¹¹

The remaining of this section proceeds as follows. In Section 4.3.1, the fitness landscape transformation is first inves-

¹⁰Given $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2)$, the graph union operation $G_1 \cup G_2$ produces $G_3 = (\mathcal{V}_3, \mathcal{E}_3)$ such that $\mathcal{V}_3 = \mathcal{V}_1 \cup \mathcal{V}_2$ and $\mathcal{E}_3 = \mathcal{E}_1 \cup \mathcal{E}_2$.

¹¹Despite the use of such low $maxDepth$ values, the resulting NNs were considerably large, as it will be analyzed in Section 4.3.2.

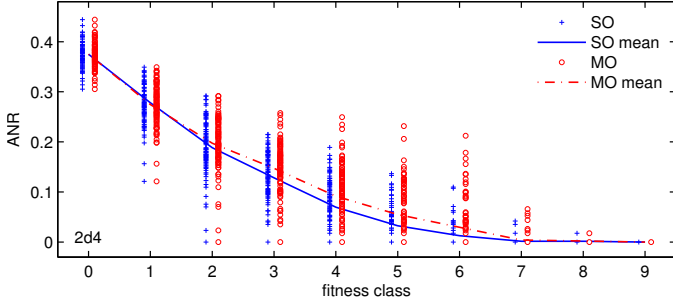


Figure 10: Average neutrality ratio (ANR). 2d4 test instance.

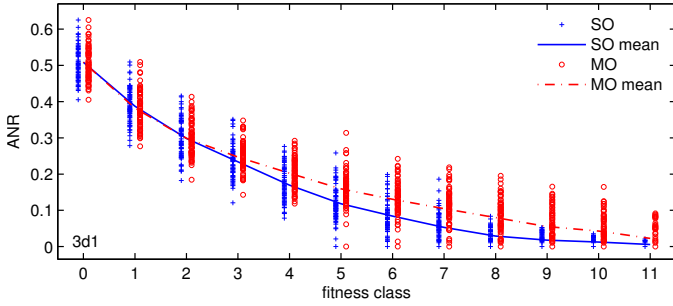


Figure 11: Average neutrality ratio (ANR). 3d1 test instance.

tigated in terms of how multi-objectivization impacted on the neutrality degree of solutions. This is captured by means of the average neutrality ratio. Then, Section 4.3.2 evaluates the extent to which such alteration on the neutrality degrees led to the increase in the size of the computed NNs. Finally, these increases in the size of the NNs are explained in Section 4.3.3 as a result of the connectivity that multi-objectivization introduces between NNs from different fitness classes.

4.3.1. Average neutrality ratio

As a means of evaluating the increase on neutrality caused by multi-objectivization, the *average neutrality ratio* (ANR) of the sampled NNs is investigated. The ANR is defined as the mean of the neutrality ratios (as defined previously in Section 2.3) considering all solutions in a NN [52, 79]. This measure assumes values in the range $[0, 1]$, where 1 corresponds to the highest neutrality. Figures 10 and 11 contrast, for the 2d4 and 3d1 instances, respectively, the ANR values obtained when using the SO and MO formulations (*i.e.*, the ANR values computed from $NN_{SO}(\mathbf{x})$ and $NN_{MO}(\mathbf{x})$, for all $\mathbf{x} \in \mathcal{S}$). In these figures, the obtained ANR values appear organized according to the fitness of the solution given as the starting point for the NN sampling. In addition, the mean of the ANR values in each fitness class is indicated for both the SO and MO formulations. From Figures 10 and 11, a general tendency can be perceived with regard to the neutrality of the HP model's fitness landscapes. In all the cases, ANR rapidly decreases with the increase in fitness. That is, while poor quality solutions (with low fitness values) are usually surrounded by a considerable number of neutral neighbors, leading to large NNs (see Section 4.3.2), solutions at the highest fitness classes tend to be more isolated

and enclosed by infeasible states.¹² In most fitness classes, it is evident from the figures that there was a slight increase in the ANR measure as a consequence of using the MO formulation (see fitness classes 2 to 7 of the 2d4 instance, and fitness classes 3 to 11 of the 3d1 instance). It is important to note that no increase in the ANR is possible for NNs at fitness class 0. This is due to the fact that, as discussed at the end of Section 4.2, a solution at this fitness class can not become neutral with respect to any other solution at a higher fitness class. After multi-objectivization, any solution \mathbf{x} with $Fitness(\mathbf{x}) = 0$ will still be considered inferior (dominated in the Pareto sense) with regard to any solution \mathbf{x}' with $Fitness(\mathbf{x}') > 0$. Thus, NNs for solutions at fitness class 0 will be exactly the same regardless of whether they are computed based on the SO or MO formulations (this applies also for subsequent analyses presented in Sections 4.3.2 and 4.3.3).

4.3.2. Size of the neutral networks

Despite the minor increases in the average neutrality ratio (ANR) obtained through the use of a multi-objective formulation, a small variation in the neutrality degree of solutions can still contribute significantly to increasing the size (number of solutions) of a NN. Therefore, in this section the size of the computed NNs is analyzed. For each sampled solution $\mathbf{x} \in \mathcal{S}$, the size of the $NN_{SO}(\mathbf{x})$ and $NN_{MO}(\mathbf{x})$ networks is shown in Figures 12 (for instance 2d4) and 13 (for instance 3d1). The results are presented separately for each fitness class, and the arithmetic mean in each of the cases is also indicated. Plots are given in a logarithmic (base 10) scale. These figures expose the high neutrality that characterizes the fitness landscapes in the HP model. Even when the sampling of NNs was restricted in this study by setting a maximum allowed depth level, as stated in Section 4.3, it is possible to see from the plots that NNs at fitness class 0 involve above 10^6 and around 10^8 solutions for the 2d4 and 3d1 instances, respectively. From Figures 12 and 13, it is also possible to confirm that, as suggested in Section 4.3.1, the size of the NNs is usually larger for low fitness classes, but neutrality tends to decrease as higher quality solutions are considered. An important increase in the size of the NNs can be observed in most of the cases due to the use of the MO formulation. As the plots indicate, NNs computed based on the MO formulation can be several orders of magnitude larger than those computed based on the SO formulation.

To go further in this analysis, the *neutral network size ratio* (NSR) is defined as the ratio of the size of $NN_{SO}(\mathbf{x})$ to that of $NN_{MO}(\mathbf{x})$. In fact, $NN_{MO}(\mathbf{x})$ will always be a supergraph containing all nodes and edges of $NN_{SO}(\mathbf{x})$, but including also those nodes and edges which result from the neutrality introduced by the multi-objectivization. Thus, $NN_{MO}(\mathbf{x})$ will have at least the same size as $NN_{SO}(\mathbf{x})$, so that NSR is defined in the range $[0, 1]$ and $NSR = 1$ indicates that no change in the NN

¹²The fitness, defined as the number of *H-H* topological contacts, is directly related to the compactness of the encoded structures; the higher the fitness value, the more compact the structure tends to be. Hence, it is easy to think that most perturbations to the encoding of a compact structure could lead either to an infeasible solution, or to a less folded state which worsens the fitness.

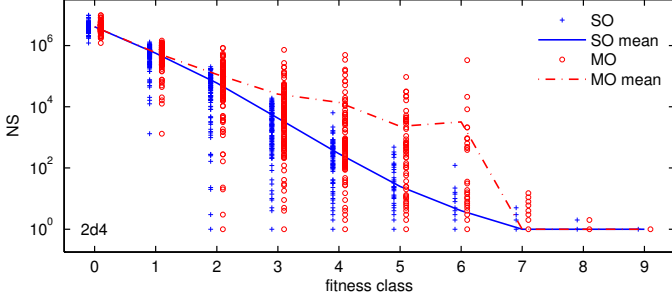


Figure 12: Size of the sampled neutral networks (NS). 2d4 instance.

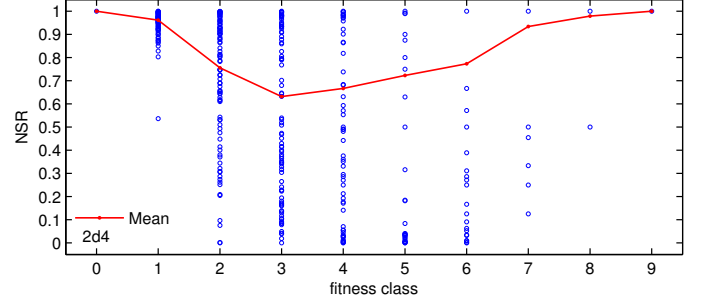


Figure 14: Neutral network size ratio (NSR). 2d4 test instance.

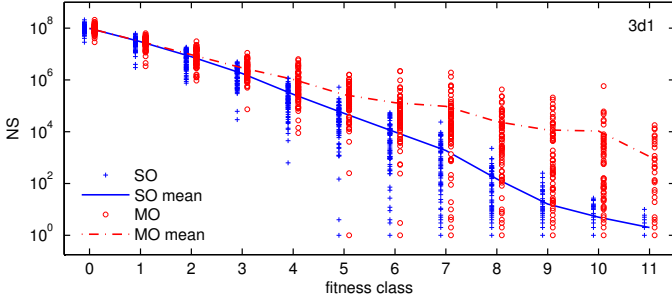


Figure 13: Size of the sampled neutral networks (NS). 3d1 instance.

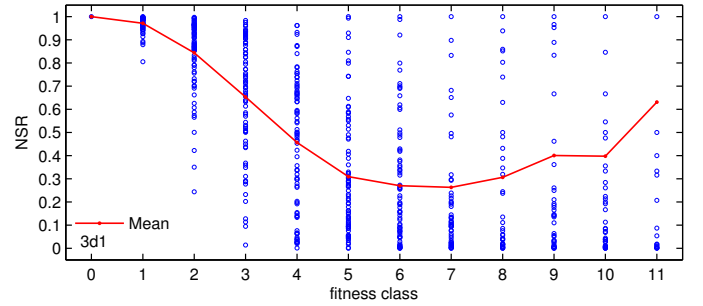


Figure 15: Neutral network size ratio (NSR). 3d1 test instance.

size was achieved when varying the problem formulation. Figures 14 and 15 show the NSR for all the sampled NNs along with the mean values calculated for each fitness class. Multi-objectivization led to an important rise in the size of the explored NNs for most fitness classes of the 2d4 instance (Figure 14). The sharpest increase in the NNs size can be observed at fitness class 3, for which the lowest average NSR value has been scored. In average, the size of $NN_{SO}(\mathbf{x})$ is only about 65% of the size of $NN_{MO}(\mathbf{x})$ when $Fitness(\mathbf{x}) = 3$. The impact of using the MO formulation becomes more evident when focusing on the 3d1 instance (Figure 15). The average NSR values for most fitness classes are below 0.5. This evinces that $NN_{MO}(\mathbf{x})$ at least doubled the size of $NN_{SO}(\mathbf{x})$ in the vast majority of the cases. Finally, it is possible to note from Figures 14 and 15 that a considerable number of very low NSR values (close to 0) have been accounted for, indicating a highly significant increase in the size of the corresponding NNs.

4.3.3. Connectivity between neutral networks

The observed increments with regard to the size of the NNs, as analyzed in Section 4.3.2, can be understood as the result of allowing *neutral connections* to be established between NNs. While in the single-objective case all solutions in a NN share, by definition, the same fitness value, in a multi-objectivization scenario such a strict definition of a NN can no longer be supported. That is, given the adopted notion of neutrality for the multi-objective case, which is based on the Pareto-dominance relation (see Section 2.3), a NN constructed using as the basis the MO formulation may involve solutions from varying fitness classes, thus connecting the corresponding NNs. Consider two NNs, NN_1 and NN_2 , such that the fitness of NN_1 is $Fitness(NN_1) = A$ and the fitness of NN_2 is $Fitness(NN_2) = B$,

$A \neq B$. If, as a result of using the MO formulation, at least a solution \mathbf{x}_1 from NN_1 comes to be neutral with respect to a neighboring solution \mathbf{x}_2 which belongs to NN_2 , then this neutral connection between NN_1 and NN_2 will merge the two NNs together into a single NN where both the fitness classes A and B are represented. To further illustrate these ideas, refer to the example provided in Figure 16. This figure presents a series of neutral moves between neighboring solutions, *i.e.*, a neutral walk, based on the MO formulation. The neutrality that the MO formulation introduced between these solutions led to the formation of neutral connections between three different NNs: $NN(\mathbf{x}_1)$ at fitness class 5, $NN(\mathbf{x}_2, \mathbf{x}_3)$ at fitness class 4 and $NN(\mathbf{x}_4)$ at fitness class 6. The four solutions \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 and \mathbf{x}_4 , and all solutions belonging to their respective NNs, became part of a single larger NN as a consequence of multi-objectivization.

From the above introduced notion of neutral connections, it becomes relevant for this study to investigate the extent to which neutral connections took place, between the different fitness classes, during the performed sampling of NNs. Figures 17 and 18 summarize the results obtained for the 2d4 and 3d1 instances. The NNs constructed for each fitness class c were analyzed (*i.e.*, c is the fitness of the initially given solution for the NN computation), and these figures indicate whether neutral connections were identified between these NNs and NNs at each other possible fitness class c' . The total number of neutral connections found of each type, if any, is shown in parentheses. Diagonals in Figures 17 and 18 are used only as a reference (*i.e.*, all NNs connect to themselves at their corresponding fitness classes) to illustrate the single-objective case, so that all other connections not appearing along the diagonal are due to the landscape transformation. As an example, Figure 17 in-

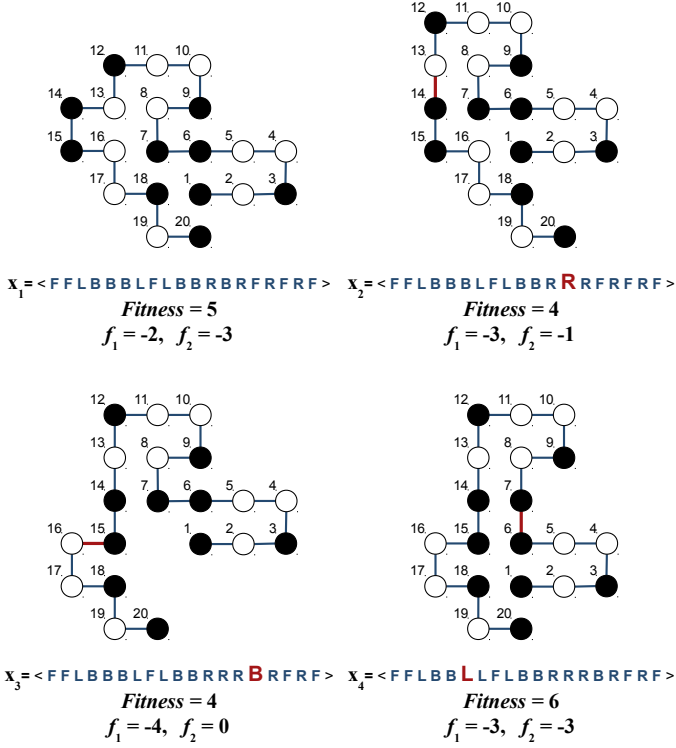


Figure 16: Neutral walk, based on the MO formulation, from a solution x_1 with $Fitness = 5$ to a solution x_4 with $Fitness = 6$. The connection of x_1 and x_4 was possible by traversing inferior solutions x_2 and x_3 with $Fitness = 4$. Neutral connections were formed between NNs at fitness classes 5, 4 and 6, leading to a single NN. The objective values, obtained by means of the LD multi-objectivization ($\delta = 7$), are presented for each solution.

icates (regarding sequence 2d4) that, out of the 118 sampled NNs for fitness class 1, only 16 formed neutral connections to NNs at fitness class 6. Through these figures it is then possible to gain an insight into the diversity of fitness classes that a NN, computed based on the MO formulation, may involve. Figures 17 and 18 highlight that a significant number of neutral connections were originated by multi-objectivization. On the one hand, NNs from fitness classes 2 to 6 of instance 2d4 presented neutral connections to all fitness classes between 1 and 7 (at least one connection in each of the cases can be observed from the plot). Note also that only a few neutral connections, all of them to inferior fitness classes, were produced from NNs at fitness classes 7 and 8. No NNs at fitness class 9 connected to others. On the other hand, a higher number of neutral connections were generated with regard to the 3d1 test instance. As it can be seen from Figure 18, connections were established, in one direction or the other, between almost all pairs of fitness classes. That is, even though the NNs computed for fitness classes 2 to 5 did not form connections to fitness class 11, multiple connections from class 11 to all such lower fitness classes were created. This points out the fact that inferior fitness classes are easier to reach than the superior ones (because of the funnel-like search landscape that characterizes the studied problem [77]).

A neutral connection from a fitness class c to a fitness class c' indicates that, given an arbitrary solution x with $Fitness(x) = c$, a neutral walk departing from x could potentially lead to a so-

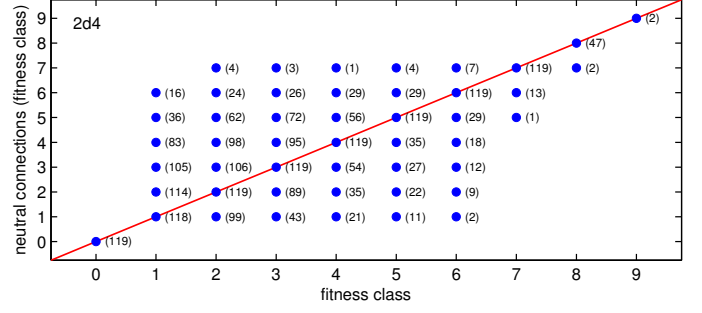


Figure 17: Neutral connections formed between fitness classes. 2d4 instance.

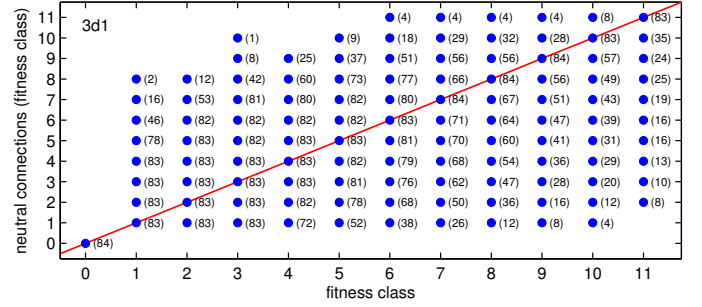


Figure 18: Neutral connections formed between fitness classes. 3d1 instance.

lution x' with $Fitness(x') = c'$. Nevertheless, as Figures 17 and 18 suggest, the more distant the fitness classes c and c' , the lower the likelihood that these classes can connect to each other through a neutral walk (in the plots, higher number of neutral connections are shown closer to the diagonal). Such a behavior is certainly accentuated if the length of the walks is bounded (as done in this study with the use of parameter *maxDepth*). In addition, although (relatively) distant fitness classes can directly connect to each other, *i.e.*, in a single step of the neutral walk, the increase in the fitness distance between a pair of solutions decreases the probability for these solutions to become incomparable after multi-objectivization, as analyzed at the end of Section 4.2. Thus, the connection between distant fitness classes is more likely to occur through a series of intermediate states. This point can be better explained by considering Figure 19. Taking as examples the fitness class 5 for instance 2d4 and fitness class 10 for instance 3d1, this figure illustrates how the neutral connections to the different fitness classes arose as each allowed depth level was reached during the NNs computation.¹³ The mean depth level at which connections to the different fitness classes were produced is also provided. It is possible to see from the plots that neutral connections to different fitness classes were given directly at depth level 1. Classes {2, 3, 4, 6} were directly connected from NNs at fitness class 5 of instance 2d4. Similarly, classes {4, 6, 7, 8, 9, 11} were connected in a single step from NNs at fitness class 10 of instance 3d1. Note, however, that the mean depth level values in these plots confirm the above suggested tendency, that the distance

¹³Similar results to those presented in Figure 19 were obtained for the different fitness classes of the considered test instances.

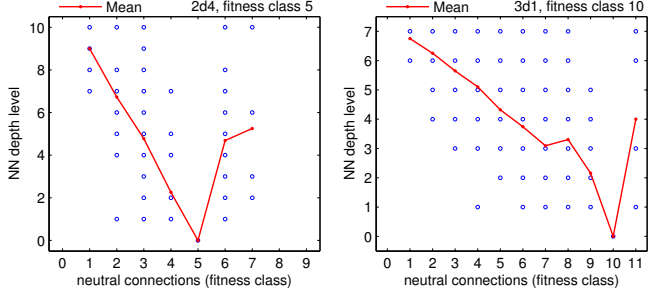


Figure 19: Neutral connections generated in relation to the depth levels reached during the NNs computation. Fitness class 5 of instance 2d4 (left) and fitness class 10 of instance 3d1 (right). Connections to the respective fitness classes (5 and 10) occurred at depth level 0 (starting point of the NN exploration).

between fitness classes is closely related to the number of neutral steps that will be usually required to connect different NNs.

Finally, it is important to address the question of how the formation of neutral connections may impact on the behavior of a search algorithm. As it has been seen, multi-objectivization can affect the neutrality relation for a given solution \mathbf{x} in two possible directions: (i) \mathbf{x} becoming neutral with respect to an inferior solution, *i.e.*, connection to a lower fitness class; or (ii) \mathbf{x} becoming neutral with respect to a superior solution, *i.e.*, connection to a higher fitness class. In the former scenario, multi-objectivization can be thought of as enhancing the mobility of the search algorithm. By connecting to lower fitness classes, the algorithm is allowed to traverse landscape areas which were originally inaccessible under the conventional SO evaluation scheme. In the neutral walk illustrated in Figure 16, for example, to move from fitness class 5 (\mathbf{x}_1) to fitness class 6 (\mathbf{x}_4) it was required to accept a degrading move to fitness class 4 (\mathbf{x}_2 and \mathbf{x}_3). In this way, the movement through inferior solutions constitutes the basis for a potential strategy to escape from local optima. It should be noted, therefore, that the design of the search algorithm will play a critical role for achieving success through the problem transformation. The new defined neutral paths can only be exploited if the algorithm is designed so as to accept moves between neutral solutions (or, in the words of Barnett, if the algorithm is able to crawl the NNs [80]). The formation of neutral connections to superior fitness classes, the later scenario, can be analyzed from two different perspectives. On the one hand, these connections (indeed all neutral connections in general) reduce what is called *selective pressure* in the context of evolutionary optimization [81], which can boost the exploration behavior of an algorithm. On the other hand, neutral connections to superior fitness classes can be interpreted as a loss in gradient information. Rather than benefiting from multi-objectivization, for instance, a search algorithm based on a strictly-better acceptance criterion could easily stagnate due to its inability to perform a proper discrimination. By relaxing the comparability relation among solutions, thus, multi-objectivization may also hinder the ability of an algorithm to identify a promising search direction.

Table 3: Results scored by the (1+1) EA when using the studied SO, PD, LD and HD formulations. Two-dimensional test cases.

Seq.	SO		PD		LD		HD	
	E_b (ν)	\bar{E}	E_b (ν)	\bar{E}	E_b (ν)	\bar{E}	E_b (ν)	\bar{E}
2d1	-4 (4)	-2.7	-4 (6)	-2.7	-4 (3)	-2.7	-4 (4)	-2.8
2d2	-8 (19)	-6.9	-8 (32)	-7.2	-8 (35)	-7.3	-8 (94)	-7.9
2d3	-8 (16)	-7.1	-8 (62)	-7.6	-9 (3)	-7.5	-9 (62)	-8.6
2d4	-9 (9)	-7.0	-9 (9)	-7.3	-9 (16)	-7.6	-9 (66)	-8.6
2d5	-9 (3)	-7.0	-10 (2)	-7.2	-9 (1)	-7.1	-10 (4)	-7.9
2d6	-8 (16)	-6.9	-9 (1)	-6.9	-9 (3)	-7.4	-9 (8)	-7.5
2d7	-7 (32)	-5.9	-8 (6)	-6.0	-8 (8)	-6.3	-8 (26)	-6.8
2d8	-13 (1)	-10.0	-13 (1)	-10.4	-13 (5)	-10.8	-13 (1)	-11.3
2d9	-18 (6)	-14.4	-19 (3)	-15.7	-21 (1)	-16.7	-21 (3)	-18.6
2d10	-18 (2)	-13.9	-18 (1)	-14.2	-19 (1)	-15.5	-20 (1)	-17.1
2d11	-30 (2)	-24.6	-32 (1)	-26.0	-33 (1)	-28.3	-33 (3)	-30.3
2d12	-29 (1)	-24.2	-30 (1)	-25.5	-32 (1)	-27.1	-32 (7)	-29.2
2d13	-41 (1)	-34.1	-42 (1)	-35.1	-44 (1)	-38.6	-47 (1)	-41.6
2d14	-41 (1)	-31.3	-39 (4)	-33.0	-39 (4)	-35.4	-43 (2)	-37.6
2d15	-40 (1)	-31.9	-40 (3)	-33.4	-42 (1)	-35.8	-40 (19)	-38.3
O-RMSE	31.28%		28.60%		25.12%		19.12%	

5. Search performance

The aim of this section is to investigate the influence that multi-objectivization can exert on the search behavior of meta-heuristic algorithms. To this end, the three studied multi-objectivization schemes, based on the parity (PD), locality (LD) and H-subsets (HD) decompositions, are evaluated and compared with respect to each other and with respect to the conventional single-objective (SO) formulation of the HP model. A basic single-solution-based evolutionary algorithm (EA), the so-called (1+1) EA, has been adopted for this sake.

All the implementation details of the (1+1) EA, as well as all the considered settings for the experiments reported in this section, are presented in Appendix B.1. In the (1+1) EA, the discrimination among candidate individuals can be either based on the conventional energy evaluation of the HP model, when using the SO formulation, or it can be based on the Pareto-dominance relation if using the alternative PD, LD and HD multi-objective formulations. Hence, the change in the problem formulation will be determinant for the performance of this algorithm. Details on the used test instances, performance measures, the statistical significance testing methodology and the utilized experimental platform are provided in Appendix A.

Tables 3 and 4 detail the results that the four studied HP model formulations obtained for all two- and three-dimensional test instances. For each instance, these tables show the best obtained energy value (E_b), the number of performed executions where this solution was found (ν) and the arithmetic mean (\bar{E}). Also, the overall relative root mean square error (O-RMSE) is presented at the bottom of the tables in order to evaluate the general performance of the different formulations analyzed. Additionally, the lowest average energy for each of the instances and the lowest O-RMSE values have been **shaded** in these tables. As shown in Table 3, HD reached the best average performance for all the 15 two-dimensional instances. This is reflected as an O-RMSE decrease of $(31.28 - 19.12) = 12.16\%$ with respect to the conventional SO formulation of the HP

Table 4: Results scored by the (1+1) EA when using the studied SO, PD, LD and HD formulations. Three-dimensional test cases.

Seq.	SO		PD		LD		HD	
	E_b (ν)	\bar{E}	E_b (ν)	\bar{E}	E_b (ν)	\bar{E}	E_b (ν)	\bar{E}
3d1	-11 (65)	-10.6	-11 (87)	-10.8	-11 (99)	-11.0	-11 (100)	-11.0
3d2	-13 (28)	-11.6	-13 (50)	-12.1	-13 (89)	-12.9	-13 (100)	-13.0
3d3	-9 (71)	-8.7	-9 (77)	-8.8	-9 (96)	-9.0	-9 (100)	-9.0
3d4	-18 (12)	-15.4	-18 (23)	-16.2	-18 (68)	-17.4	-18 (88)	-17.8
3d5	-30 (2)	-24.4	-30 (2)	-25.9	-33 (1)	-28.6	-31 (2)	-28.5
3d6	-29 (2)	-23.2	-29 (3)	-24.8	-31 (1)	-27.6	-30 (1)	-27.3
3d7	-25 (6)	-21.1	-27 (1)	-22.9	-29 (1)	-25.3	-29 (1)	-24.8
3d8	-35 (1)	-27.7	-36 (1)	-30.1	-38 (1)	-33.6	-35 (6)	-32.6
3d9	-48 (1)	-38.1	-48 (1)	-41.1	-47 (8)	-44.7	-49 (1)	-45.0
3d10	-45 (1)	-36.2	-46 (2)	-38.8	-50 (2)	-45.5	-48 (3)	-43.3
3d11	-40 (1)	-31.0	-42 (1)	-33.7	-41 (3)	-37.9	-40 (2)	-36.9
3d12	-48 (2)	-37.3	-54 (1)	-41.4	-53 (3)	-48.5	-50 (1)	-45.3
3d13	-41 (1)	-30.7	-42 (1)	-32.5	-43 (1)	-37.1	-40 (2)	-36.3
3d14	-48 (1)	-35.5	-48 (4)	-38.3	-52 (1)	-46.4	-48 (2)	-43.4
3d15	-51 (2)	-38.6	-53 (1)	-44.0	-59 (1)	-50.3	-56 (1)	-48.7
O-RMSE	33.21%		28.66%		20.85%		21.66%	

Table 5: Statistical analysis for comparing the performance of the (1+1) EA when using the four studied HP model’s formulations.

	Two-dimensional instances															Three-dimensional instances															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		+	+					+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	23+0-
LD/SO	+	+	+			+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	28+0-
HD/SO	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	29+0-
LD/PD						+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	25+0-
HD/PD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	29+0-
HD/LD	+	+	+	+				+	+	+					+	+	+				-	-	-		-	-		-	-	-	16+9-

model. It can also be noted that LD and PD presented a lower average energy than SO in all the cases, improving the O-RMSE measure by 6.16% and by 2.68%, respectively. The LD formulation achieved the lowest average energy for 12 out of the 15 three-dimensional instances, see Table 4. The best results for the remaining test instances were obtained by using HD. In general, the results of the PD, LD and HD formulations are found to be quite competitive. These approaches improved the O-RMSE measure by 4.55%, 12.36% and 11.55% with respect to the SO formulation, respectively.

Table 5 outlines how the four studied 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 for a particular instance, the corresponding cell is marked either + 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 compared approaches. The rightmost column presents the overall results of this analysis. Out of a total of 30 instances, it can be seen from Table 5 that the multi-objective PD, LD and HD approaches significantly outperformed the conventional SO formulation of the HP model in 23, 28 and 29 of the cases, respectively. By comparing among the multi-objective formulations, the results of LD for 25 of the instances

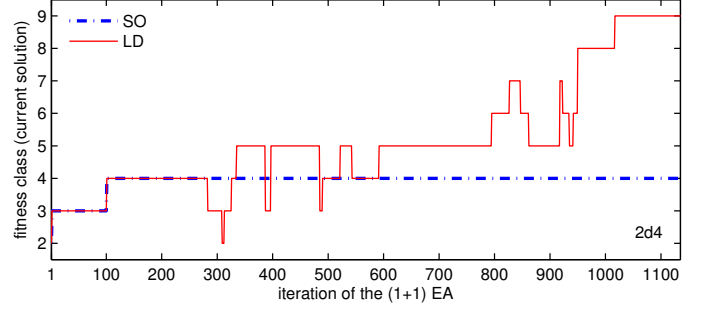


Figure 20: Online performance scored in a single execution of the (1+1) EA when using the SO and LD formulations. Two-dimensional 2d4 test instance.

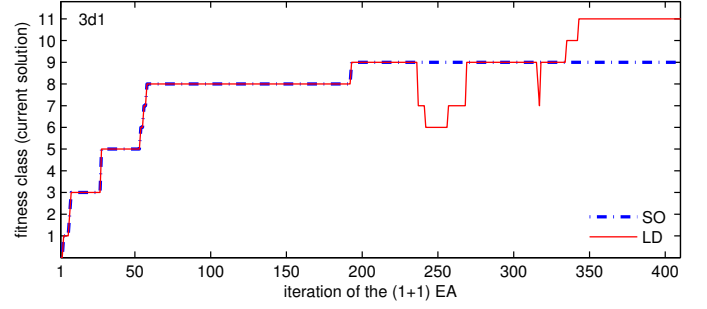


Figure 21: Online performance scored in a single execution of the (1+1) EA when using the SO and LD formulations. Three-dimensional 3d1 test instance.

were statistically superior to those obtained by PD. Compared with respect to PD, HD significantly increased the performance of the algorithm for all but one of the test sequences (2d1). Finally, HD was found to perform significantly better than LD in 16 of the instances, while there was a significant difference in favor of LD for 9 of the three-dimensional test cases.

Finally, the LD formulation is used as an example in order to illustrate how multi-objectivization affects the search behavior of the (1+1) EA. Plots in Figures 20 and 21 contrast the online (throughout the search) performance scored by the SO and LD formulations during a single execution of the (1+1) EA when solving, respectively, instances 2d4 and 3d1. Adhering to the notation used previously in Section 4, performance is measured as the fitness class of the current individual at each iteration of the algorithm.¹⁴ Notice that these figures cover only the first stages of the search process, where the algorithm reached the optimum solution of the considered test instances by using LD (fitness class 9 for instance 2d4 and fitness class 11 for instance 3d1). By focusing first on the convergence behavior achieved through the use of the conventional SO formulation, it is possible to observe from the plots that the algorithm exhibits the *punctuated equilibrium* that characterizes evolutionary dynamics on neutral fitness landscapes [82]. Such a behavior presents long periods of stasis in which the algorithm drifts along neutral networks, thus remaining in the same fitness class, punctuated by occasional transitions between networks [83]. As shown in the plots, the algorithm scored a similar behavior either using the SO or LD formulations at the beginning of the

¹⁴As it was defined in Section 4, $Fitness(\mathbf{x}) = HHtc(\mathbf{x}) = -E(\mathbf{x})$.

search process (both the SO- and LD-based executions started from the same initial randomly generated individual). It can be noted, however, that the neutrality (incomparability) introduced through the use of the LD formulation, as analyzed in Section 4, allowed the algorithm to accept degrading moves towards inferior fitness classes (in the figures, such degrading moves are shown as decreases in the corresponding LD curves). It is important to emphasize that such degrading moves were perceived as neutral moves from the perspective of the search algorithm. That is, each accepted degrading move represents the movement towards an individual that, despite having a lower fitness value, became part of the neutral network of the parent individual from which it was generated due to the problem transformation. The new defined neutral paths enhance exploration and eventually lead the algorithm to escape from local optima.

6. Discussion and final remarks

Multi-objectivization concerns the reformulation of a single-objective problem as a multi-objective one [1]. When applied to the particular case of study of this research project, the HP model for protein structure prediction, it has been reported that this transformation provides significant improvements in terms of the performance of search algorithms [37–39]. Motivated from these previous findings, the multi-objectivization of the HP model has been further investigated in this study.

This study has been divided into two main parts. The first part was devoted to address the primary aim of this study: to inquire into the potential effects of multi-objectivization as a means of understanding how this transformation can influence the behavior of search algorithms. When multi-objectivization is achieved through the decomposition of the original objective function of the problem, as considered in this study, incomparability among solution candidates can be introduced [4]. That is, originally comparable solutions may become incomparable when evaluated under the new multi-objective formulation of the problem. As a first step in understanding and quantifying such an effect, it was explored the extent to which incomparability may arise between the different fitness classes. To this end, a large set of sampled solution pairs were evaluated, out of which a considerable number became incomparable as a consequence of the multi-objectivization. It was also found that the more distant the fitness classes for a given pair of solutions, the lower the likelihood that the comparability relation between these solutions can be affected by multi-objectivization.

Introducing incomparability among solutions can be alternatively understood as the increase in the neutrality of the fitness landscape. Therefore, a detailed fitness landscape analysis was conducted in order to investigate how multi-objectivization has impacted on such an important problem characteristic. A large number of neutral networks (NNs) were sampled, and the main findings of their analysis can be summarized as follows. By rising the neutrality degree of solutions, multi-objectivization led to the formation of neutral connections between NNs from different fitness classes. That is, when originally comparable neighboring solutions become incomparable

by multi-objectivization, their corresponding NNs are merged together into larger connected neutral areas of the landscape.

The aforementioned effects of multi-objectivization may lead to different implications from the perspective of a search algorithm. On the one hand, introducing neutrality in the fitness landscape can be reflected as an enhancement in the exploration behavior of the algorithm. That is, by reducing the so-called *selective pressure* [81], an algorithm can be allowed to move through inferior fitness classes as a means of escaping from local optima. On the other hand, the increase in neutrality can also be understood as a loss in gradient information, so that multi-objectivization may also prevent algorithms from identifying promising search directions in some cases.

Understanding the potential effects and consequences of multi-objectivization could lead to the design of more effective search algorithms. To the best of the authors' knowledge, in most of the reported applications of multi-objectivization the original evaluation scheme of the problem is completely replaced by the alternative multi-objectivized one. This corresponds also to the approach assumed in the second part of this study, as it will be discussed later in this section. While the complete replacement of the evaluation scheme has reported very promising results in the literature, the above presented analysis suggests that a better strategy may involve applying multi-objectivization only under certain conditions during the search process; *e.g.*, when stagnation at a local optimum has been detected. Such a strategy could benefit from the potential effects of multi-objectivization as a means of escaping from local optima, at the same time that the gradient information is preserved in order to drive the search process in an effective manner. In the context of the HP model for protein structure prediction, a high degree of neutrality is inherently induced by the conventional evaluation scheme, as it was observed along the conducted analysis. Thus, another interesting approach may consist of alternating the use of multi-objectivization with the use of some other alternative formulation specifically designed to cope with the neutrality of the problem; such as those evaluated in [84]. Exploring these kind of strategies related to the partial (rather than total) use of multi-objectivization is considered as a possible direction for future research.

The second part of this study revisited the evaluation of multi-objectivization from the perspective of search performance, as reported in [37–39]. Three different multi-objectivization schemes for the HP model were compared and evaluated with respect to the conventional single-objective formulation of the problem. A basic (1+1) evolutionary algorithm was considered in this analysis. As a result, it has been found that the use of the alternative multi-objective formulations of the problem significantly increased the average performance of the implemented algorithm in most of the conducted experiments. The obtained results give further support to the suitability of multi-objectivization as a strategy to overcome search difficulties such as that of becoming trapped in local optima.

To the best of the authors' knowledge, the multi-objectivization approaches evaluated in this study represent the first efforts on the use of multi-objective optimization techniques to address the HP model of the protein structure predic-

tion problem. In addition, no previous work has been reported, as far as the authors are aware, where the potential effects of multi-objectivization are investigated through the explicit sampling and evaluation of the characteristics of the fitness landscape. The conducted analysis focused on neutrality. Extending this analysis to evaluate the problem transformation from the perspective of other different landscape properties, *e.g.*, ruggedness [85], can certainly be seen as a relevant research direction that will contribute to build a more comprehensive picture. Although the performed analysis was applied to a particular case of study, most of the findings regarding the fitness landscape transformation can be generalized to other problem domains. In this way, using the HP model as an example, this study is expected to contribute to the general understanding of multi-objectivization, a concept which represents a current and promising research direction.

Appendix A. Performance assessment

Appendix A.1. Test instances

A total of 30 well-known benchmark sequences for the HP model have been considered for the experimentation of this research project. Out of them, 15 are for the two-dimensional square lattice and the other 15 are for three-dimensional cubic lattice. Tables A.6 and A.7 present the full HP sequences, their length (ℓ) and the optimal or best known energy value (E^*) reported in the literature [59, 60, 66, 86–88].

Appendix A.2. Performance measures

Although alternative (multi-objectivized) formulations of the HP model are analyzed in this study, it is important to remark that the goal remains always to solve the original problem. Thus, the experimental results presented in Section 5 are evaluated in terms of the conventional energy function of the HP model defined in Section 2.4. Additionally, the *overall relative root mean square error*, O-RMSE measure has been adopted in order to assess the overall performance of the studied approaches. Before introducing the O-RMSE measure, the *relative root mean square error* for a given test instance t , $\text{RMSE}(t)$, can be formally defined as follows:

$$\text{RMSE}(t) = 100\% \sqrt{\frac{1}{R} \sum_{r=1}^R \left(\frac{E_r(t) - E^*(t)}{E^*(t)} \right)^2}, \quad (\text{A.1})$$

where $E_r(t)$ denotes the energy of the best solution found during a particular execution r , R is the total number of performed executions, and $E^*(t)$ is the optimal (or best known) energy value for instance t . Thus, RMSE expresses the performance of the evaluated approaches in a 0% to 100% scale, being $\text{RMSE}(t) = 0\%$ the preferred value for this measure.

Finally, O-RMSE is defined as the arithmetic mean of the RMSE computed over all the considered instances. Formally,

$$\text{O-RMSE} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \text{RMSE}(t), \quad (\text{A.2})$$

Table A.6: Test instances for the two-dimensional square lattice.

Sequence	ℓ	E^*
2d1 H ₂ P ₅ H ₂ P ₃ HP ₃ HP	18	-4
2d2 HPHPH ₃ P ₃ H ₄ P ₂ H ₂	18	-8
2d3 PHP ₂ HPH ₃ PH ₂ PH ₅	18	-9
2d4 HPHPH ₂ H ₂ PHP ₂ HPH ₂ P ₂ HPH	20	-9
2d5 H ₃ P ₂ HPHHPH ₂ HPHHPH ₂ H	20	-10
2d6 H ₂ P ₂ HP ₂ HP ₂ HP ₂ HP ₂ HP ₂ HP ₂ H ₂	24	-9
2d7 P ₂ HP ₂ H ₂ P ₄ H ₂ P ₄ H ₂ P ₄ H ₂	25	-8
2d8 P ₃ H ₂ P ₂ H ₂ P ₅ H ₇ P ₂ H ₂ P ₄ H ₂ P ₂ HP ₂	36	-14
2d9 P ₂ HP ₂ H ₂ P ₂ H ₂ P ₅ H ₁₀ P ₆ H ₂ P ₂ H ₂ P ₂ HP ₂ H ₅	48	-23
2d10 H ₂ (PH) ₄ H ₃ P(HP ₃) ₃ (P ₃ H) ₃ PH ₄ (PH) ₄ H	50	-21
2d11 P ₂ H ₃ PH ₈ P ₃ H ₁₀ PHP ₃ H ₁₂ P ₄ H ₆ PH ₂ PHP	60	-36
2d12 H ₁₂ PHPH(P ₂ H ₂ P ₂ H ₂ P ₂ H ₃) ₃ PHPH ₁₂	64	-42
2d13 H ₄ P ₄ H ₁₂ P ₆ (H ₁₂ P ₃) ₃ HP ₂ H ₂ P ₂ H ₂ P ₂ HPH	85	-53
2d14 P ₆ HPH ₂ P ₅ H ₃ PH ₅ PH ₂ P ₄ H ₂ P ₂ H ₂ PH ₅ PH ₁₀ PH ₂ PH ₇ P ₁₁ H ₇ P ₂ HPH ₃ P ₆ HPH ₂	100	-48
2d15 P ₃ H ₂ P ₂ H ₄ P ₂ H ₃ PH ₂ PH ₂ PH ₄ P ₈ H ₆ P ₂ H ₆ P ₉ HPH ₂ PH ₁₁ P ₂ H ₃ PH ₂ PHP ₂ HPH ₃ P ₆ H ₃	100	-50

Table A.7: Test instances for the three-dimensional cubic lattice.

Sequence	ℓ	E^*
3d1 HPHPH ₂ H ₂ PHP ₂ HPH ₂ P ₂ HPH	20	-11
3d2 H ₂ P ₂ HP ₂ HP ₂ HP ₂ HP ₂ HP ₂ HP ₂ H ₂	24	-13
3d3 P ₂ HP ₂ H ₂ P ₄ H ₂ P ₄ H ₂ P ₄ H ₂	25	-9
3d4 P ₃ H ₂ P ₂ H ₂ P ₅ H ₇ P ₂ H ₂ P ₄ H ₂ P ₂ HP ₂	36	-18
3d5 P ₂ H ₃ PH ₃ P ₃ HPH ₂ PH ₂ P ₂ HPH ₄ PHP ₂ H ₅ PHPH ₂ P ₂ H ₂ P	46	-35
3d6 P ₂ HP ₂ H ₂ P ₂ H ₂ P ₅ H ₁₀ P ₆ H ₂ P ₂ H ₂ HP ₂ H ₅	48	-31
3d7 H ₂ (PH) ₄ H ₃ P(HP ₃) ₃ (P ₃ H) ₃ PH ₄ (PH) ₄ H	50	-34
3d8 PH(PH ₃) ₂ P(PH ₂ PH) ₂ H(HP) ₃ (H ₂ P ₂ H) ₂ PHP ₄ (H(P ₂ H) ₂) ₂	58	-44
3d9 P ₂ H ₃ PH ₈ P ₃ H ₁₀ PHP ₃ H ₁₂ P ₄ H ₆ PH ₂ PHP	60	-55
3d10 H ₁₂ PHPH(P ₂ H ₂ P ₂ H ₂ P ₂ H ₃) ₃ PHPH ₁₂	64	-59
3d11 P(HPH ₂ PH ₂ PHP ₂ H ₃ P ₃) ₃ (HPH) ₃ P ₂ H ₃ P	67	-56
3d12 P(HPH) ₃ P ₂ H ₂ (P ₂ H) ₆ H(P ₂ H ₃) ₄ P ₂ (HPH) ₃ P ₂ HP(PHP ₂ H ₂ HP) ₂	88	-72
3d13 P ₂ H ₂ P ₅ H ₂ P ₂ H ₂ PHP ₂ HP ₇ HP ₃ H ₃ PH ₃ P ₆ HP ₂ HP HP ₂ HP ₅ H ₃ P ₄ H ₂ PH ₂ P ₅ H ₂ P ₄ HPH ₈ H ₅ P ₂ HP ₂	103	-58
3d14 P ₃ H ₃ PHP ₄ HP ₅ H ₂ P ₄ H ₂ P ₂ H ₂ (P ₄ H) ₂ P ₂ HP ₂ H ₂ P ₃ H ₂ PHPH ₃ P ₄ H ₃ P ₆ H ₂ P ₂ HP ₂ HPH ₂ HP ₇ HP ₂ H ₃ P ₄ HP ₃ H ₅ P ₄ H ₂ (PH) ₄	124	-75
3d15 HP ₅ HP ₄ HPH ₂ PH ₂ P ₄ HPH ₃ P ₄ HPH ₂ PH ₄ P ₁₁ HP ₂ HP ₃ HPH ₂ P ₃ H ₂ P ₂ HP ₂ HPH ₂ HPH ₈ HP ₃ H ₆ P ₃ H ₂ P ₂ H ₃ P ₃ H ₂ PH ₅ P ₉ HP ₄ HPH ₄	136	-83

where \mathcal{T} is the set of all test instances. Thus, O-RMSE = 0% suggests the ideal situation where the optimal solution for each instance was reached during all the performed executions.

Appendix A.3. Statistical significance analysis

The statistical significance analysis was conducted as follows. First, *D'Agostino-Pearson's omnibus K²* 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. Finally, a significance level of $\alpha = 0.05$ has been considered.

Appendix A.4. Experimental platform

The algorithms implemented in this study were coded in ANSI C and compiled with *gcc* using the optimization flag

-03. All experiments performed were run sequentially on the Neptuno cluster at the Information Technology Laboratory, CINVESTAV-Tamaulipas. This cluster is equipped with 10 InfiniBand interconnected nodes, each of which features 8 cores running at 2.66 GHz, has a total of 16 GB of RAM, and uses the CentOS distribution of the Linux operating system.

Appendix B. Algorithms

Appendix B.1. (1+1) evolutionary algorithm

The so-called (1+1) evolutionary algorithm is described in Algorithm 3. First, an initial parent individual \mathbf{x} is generated at random. At each generation, an offspring \mathbf{x}' is created by randomly and independently mutating \mathbf{x} at each encoding position with probability p_m . The new individual \mathbf{x}' is rejected only if it is strictly worse than the parent individual \mathbf{x} , otherwise \mathbf{x}' is accepted as the starting point for the next generation. Such an acceptance criterion can be either based on a single-objective discrimination between \mathbf{x} and \mathbf{x}' , or it can be based on the Pareto-dominance relation when implementing a multi-objective formulation of the problem.

Algorithm 3 Basic (1+1) evolutionary algorithm.

```

1: choose  $\mathbf{x} \in \mathcal{X}$  uniformly at random
2: repeat
3:    $\mathbf{x}' \leftarrow \text{mutate}(\mathbf{x})$ 
4:   if  $\mathbf{x}'$  not worse than  $\mathbf{x}$  then
5:      $\mathbf{x} \leftarrow \mathbf{x}'$ 
6:   end if
7: until  $< \text{stop condition} >$ 

```

In this algorithm, individuals encode protein conformations using an internal coordinates representation based on absolute moves, see Section 2.4. In mutation, each encoding position is randomly and independently perturbed with probability p_m . In all the cases, this probability is fixed to $p_m = \frac{1}{\ell-1}$, where $\ell - 1$ denotes the length of the individuals encoding. A maximum number of 5×10^5 solution evaluations was adopted as the stopping condition and a total of 100 independent executions were performed for all two- and three-dimensional instances. Finally, it is important to remark that only individuals encoding feasible conformations are considered during the search process; once mutation is to be applied to a particular encoding position, all possible perturbations to this position are explored in random order until a feasible conformation is obtained. If no change in this position leads to a feasible conformation, the original value is restored (infeasible offspring are rejected without consuming objective function evaluations). The initial feasible individuals are generated using a backtracking procedure reported in [89].

The LD and HD formulations are sensitive to the adjustment of some parameters. As pointed out in Section 3.2, the distance parameter δ of the LD formulation has been found in the authors' previous work to provide the best performance at low values around 7 [38]. Thus, the value of this parameter was set to $\delta = 7$ for the analysis here presented. Similarly, an important issue for the HD formulation is how the H-subsets formation

process is carried out. The DYN_k strategy, with $k = 30$, was considered in this study given the results reported in [39].

Appendix B.2. Iterated local search algorithm

Algorithm 4 outlines the general structure of a basic iterated local search (ILS) algorithm. The algorithm starts with a randomly generated conformation, denoted as \mathbf{x} . Then, a local search strategy (embedded heuristic) is applied to \mathbf{x} until a local optimum \mathbf{x}^* is found. At each iteration, a perturbation \mathbf{x}' of the current local optimum \mathbf{x}^* is obtained and used as the starting point of another round of local search. After each local search, the new local optimum solution found \mathbf{x}^{**} may be accepted as the new incumbent solution \mathbf{x}^* , based on a given acceptance criterion. This is repeated until a given stop condition is met.

Algorithm 4 Iterated local search (ILS).

```

1: choose  $\mathbf{x} \in \mathcal{X}$  uniformly at random
2:  $\mathbf{x}^* \leftarrow \text{LocalSearch}(\mathbf{x})$ 
3: repeat
4:    $\mathbf{x}' \leftarrow \text{Perturbation}(\mathbf{x}^*)$ 
5:    $\mathbf{x}^{**} \leftarrow \text{LocalSearch}(\mathbf{x}')$ 
6:    $\mathbf{x}^* \leftarrow \text{AcceptanceCriterion}(\mathbf{x}^*, \mathbf{x}^{**})$ 
7: until  $< \text{stop condition} >$ 

```

Three main components, which determine the behavior of an ILS algorithm, have to be defined: (i) the embedded local search heuristic; (ii) the perturbation strength; and (iii) the acceptance criterion. A best improvement local search algorithm was used as the embedded heuristic, and basic settings for the perturbation strength and acceptance criterion were adopted according to the authors' previous work reported in [84].

References

- [1] J. Knowles, R. Watson, D. Corne, Reducing Local Optima in Single-Objective Problems by Multi-objectivization, in: Evolutionary Multi-Criterion Optimization, Springer-Verlag, Zurich, Switzerland, 2001, pp. 269–283.
- [2] M. Jensen, Helper-Objectives: Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimisation, Journal of Mathematical Modelling and Algorithms 3 (4) (2004) 323–347.
- [3] D. Brockhoff, T. Friedrich, N. Hebbinghaus, C. Klein, F. Neumann, E. Zitzler, Do Additional Objectives Make a Problem Harder?, in: Genetic and Evolutionary Computation Conference, ACM, London, England, 2007, pp. 765–772.
- [4] J. Handl, S. Lovell, J. Knowles, Multiobjectivization by Decomposition of Scalar Cost Functions, in: Parallel Problem Solving from Nature, Vol. 5199 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Dortmund, Germany, 2008, pp. 31–40.
- [5] J. Watson, An Introduction to Fitness Landscape Analysis and Cost Models for Local Search, in: Handbook of Metaheuristics, Vol. 146 of International Series in Operations Research & Management Science, Springer US, 2010, pp. 599–623.
- [6] J. Knowles, D. Corne, Quantifying the Effects of Objective Space Dimension in Evolutionary Multiobjective Optimization, in: Evolutionary Multi-Criterion Optimization, Vol. 4403 of Lecture Notes in Computer Science, Matshushima, Japan, 2007, pp. 757–771.
- [7] H. Ishibuchi, N. Tsukamoto, Y. Nojima, Evolutionary Many-Objective Optimization: A Short Review, in: IEEE Congress on Evolutionary Computation, Hong Kong, 2008, pp. 2424–2431.

- [8] C. Segura, C. Coello Coello, G. Miranda, C. León, Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimization, *4OR* 11 (3) (2013) 201–228.
- [9] M. Jähne, X. Li, J. Branke, Evolutionary Algorithms and Multi-Objectivization for the Travelling Salesman Problem, in: Genetic and Evolutionary Computation Conference, ACM, Montreal, Canada, 2009, pp. 595–602.
- [10] D. Lochtiefeld, F. Ciarallo, An Analysis of Decomposition Approaches in Multi-objectivization via Segmentation, *Applied Soft Computing* In Press.
- [11] D. Lochtiefeld, F. Ciarallo, Helper-Objective Optimization Strategies for the Job-Shop Scheduling Problem, *Applied Soft Computing* 11 (6) (2011) 4161–4174.
- [12] C. Segura, E. Segredo, C. León, Parallel Island-Based Multiobjectivised Memetic Algorithms for a 2D Packing Problem, in: Genetic and Evolutionary Computation Conference, ACM, Dublin, Ireland, 2011, pp. 1611–1618.
- [13] E. Segredo, C. Segura, C. León, Memetic Algorithms and Hyperheuristics Applied to a Multiobjectivised Two-dimensional Packing Problem, *Journal of Global Optimization* (2013) 1–26.
- [14] S. Watanabe, K. Sakakibara, A Multiobjectivization Approach for Vehicle Routing Problems, in: *Evolutionary Multi-Criterion Optimization*, Vol. 4403 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Matshushima, Japan, 2007, pp. 660–672.
- [15] F. Neumann, I. Wegener, Can Single-Objective Optimization Profit from Multiobjective Optimization?, in: *Multiobjective Problem Solving from Nature*, Natural Computing Series, Springer Berlin Heidelberg, 2008, pp. 115–130.
- [16] C. Segura, E. Segredo, Y. González, C. León, Multiobjectivisation of the Antenna Positioning Problem, in: *International Symposium on Distributed Computing and Artificial Intelligence*, Vol. 91 of Advances in Intelligent and Soft Computing, Springer Berlin / Heidelberg, Salamanca, Spain, 2011, pp. 319–327.
- [17] C. Segura, E. Segredo, C. León, Scalability and Robustness of Parallel Hyperheuristics Applied to a Multiobjectivised Frequency Assignment Problem, *Soft Computing* 17 (6) (2013) 1077–1093.
- [18] D. Greiner, J. Emperador, G. Winter, B. Galván, Improving Computational Mechanics Optimum Design Using Helper Objectives: An Application in Frame Bar Structures, in: *Evolutionary Multi-Criterion Optimization*, Vol. 4403 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Matshushima, Japan, 2007, pp. 575–589.
- [19] A. Trivedi, D. Sharma, D. Srinivasan, Multi-objectivization of Short-term Unit Commitment Under Uncertainty Using Evolutionary Algorithm, in: *IEEE Congress on Evolutionary Computation*, Brisbane, Australia, 2012, pp. 1–8.
- [20] D. Sharma, K. Deb, N. Kishore, Customized Evolutionary Optimization Procedure for Generating Minimum Weight Compliant Mechanisms, *Engineering Optimization* 46 (1) (2014) 39–60.
- [21] A. Churchill, P. Husbands, A. Philippides, Multi-objectivization of the Tool Selection Problem on a Budget of Evaluations, in: *Evolutionary Multi-Criterion Optimization*, Vol. 7811 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 600–614.
- [22] J. Mouret, Novelty-Based Multiobjectivization, in: *New Horizons in Evolutionary Robotics*, Vol. 341 of Studies in Computational Intelligence, Springer Berlin / Heidelberg, 2011, pp. 139–154.
- [23] I. Vite-Silva, N. Cruz-Cortés, G. Toscano-Pulido, L. de la Fraga, Optimal Triangulation in 3D Computer Vision Using a Multi-objective Evolutionary Algorithm, in: *Applications of Evolutionary Computing*, Vol. 4448 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Valencia, Spain, 2007, pp. 330–339.
- [24] S. Thomas, Y. Jin, Single and Multi-objective in Silico Evolution of Tunable Genetic Oscillators, in: *Evolutionary Multi-Criterion Optimization*, Vol. 7811 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 696–709.
- [25] R. Day, J. Zydallis, G. Lamont, Solving the Protein Structure Prediction Problem Through a Multi-Objective Genetic Algorithm, in: *IEEE/DARPA International Conference on Computational Nanoscience*, San Juan, PR, USA, 2002, pp. 32–35.
- [26] V. Cutello, G. Narzisi, G. Nicosia, A Class of Pareto Archived Evolution Strategy Algorithms Using Immune Inspired Operators for Ab-Initio Protein Structure Prediction, in: *Applications of Evolutionary Computing*, Vol. 3449 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Lausanne, Switzerland, 2005, pp. 54–63.
- [27] V. Cutello, G. Narzisi, G. Nicosia, A Multi-Objective Evolutionary Approach to the Protein Structure Prediction Problem, *Journal of The Royal Society Interface* 3 (6) (2006) 139–151.
- [28] V. Cutello, G. Narzisi, G. Nicosia, Computational Studies of Peptide and Protein Structure Prediction Problems via Multiobjective Evolutionary Algorithms, in: *Multiobjective Problem Solving from Nature*, Natural Computing Series, Springer Berlin Heidelberg, 2008, pp. 93–114.
- [29] J. Handl, S. Lovell, J. Knowles, Investigations into the Effect of Multiobjectivization in Protein Structure Prediction, in: *Parallel Problem Solving from Nature*, Vol. 5199 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Dortmund, Germany, 2008, pp. 702–711.
- [30] D. Becerra, A. Sandoval, D. Restrepo-Montoya, L. Nino, A Parallel Multi-Objective Ab Initio Approach for Protein Structure Prediction, in: *IEEE International Conference on Bioinformatics and Biomedicine*, Hong Kong, China, 2010, pp. 137–141.
- [31] C. Soares Brasil, A. Botazzo Delbem, D. Ferraz Bonetti, Investigating Relevant Aspects of MOEAs for Protein Structures Prediction, in: *Genetic and Evolutionary Computation Conference*, ACM, Dublin, Ireland, 2011, pp. 705–712.
- [32] B. Olson, A. Shehu, Multi-Objective Stochastic Search for Sampling Local Minima in the Protein Energy Surface, in: *International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, ACM, Washington DC, USA, 2013, pp. 430–439.
- [33] S. Sudha, S. Baskar, S. Krishnaswamy, Multi-Objective Approach for Protein Structure Prediction, in: *Swarm, Evolutionary, and Memetic Computing*, Vol. 8298 of Lecture Notes in Computer Science, Springer International Publishing, 2013, pp. 511–522.
- [34] K. Dill, Theory for the Folding and Stability of Globular Proteins, *Biochemistry* 24 (6) (1985) 1501–9.
- [35] B. Berger, T. Leighton, Protein Folding in the Hydrophobic-Hydrophilic (HP) Model is NP-complete, in: *International Conference on Research in Computational Molecular Biology*, ACM, New York, NY, USA, 1998, pp. 30–39.
- [36] P. Crescenzi, D. Goldman, C. Papadimitriou, A. Piccolboni, M. Yannakakis, On the Complexity of Protein Folding, in: *ACM Symposium on Theory of Computing*, ACM, Dallas, TX, USA, 1998, pp. 597–603.
- [37] M. Garza-Fabre, E. Rodriguez-Tello, G. Toscano-Pulido, Multiobjectivizing the HP Model for Protein Structure Prediction, in: *Evolutionary Computation in Combinatorial Optimization*, Vol. 7245 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Málaga, Spain, 2012, pp. 182–193.
- [38] M. Garza-Fabre, G. Toscano-Pulido, E. Rodriguez-Tello, Locality-based Multiobjectivization for the HP Model of Protein Structure Prediction, in: *Genetic and Evolutionary Computation Conference*, ACM, Philadelphia, PA, USA, 2012, pp. 473–480.
- [39] M. Garza-Fabre, E. Rodriguez-Tello, G. Toscano-Pulido, An Improved Multiobjectivization Strategy for HP Model-Based Protein Structure Prediction, in: *Parallel Problem Solving from Nature*, Vol. 7492 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Taormina, Italy, 2012, pp. 82–92.
- [40] V. Pareto, *Cours d'Economie Politique*, Droz, Genève, 1896.
- [41] S. Louis, G. Rawlins, Pareto Optimality, GA-easiness and Deception, in: *International Conference on Genetic Algorithms*, Morgan Kaufmann, 1993, pp. 118–123.
- [42] D. Lochtiefeld, F. Ciarallo, Multiobjectivization via Helper-Objectives With the Tunable Objectives Problem, *IEEE Transactions on Evolutionary Computation* 16 (3) (2012) 373–390.
- [43] L. Bui, M. Nguyen, J. Branke, H. Abbass, Tackling Dynamic Problems with Multiobjective Evolutionary Algorithms, in: *Multiobjective Problem Solving from Nature*, Natural Computing Series, Springer Berlin Heidelberg, 2008, pp. 77–91.
- [44] E. Segredo, C. Segura, C. Leon, A Multiobjectivised Memetic Algorithm for the Frequency Assignment Problem, in: *IEEE Congress on Evolutionary Computation*, New Orleans, LA, USA, 2011, pp. 1132–1139.
- [45] S. Wessing, M. Preuss, G. Rudolph, Niching by Multiobjectivization with Neighbor Information: Trade-offs and Benefits, in: *IEEE Congress on Evolutionary Computation*, Cancún, México, 2013, pp. 103–110.
- [46] T. Tran, D. Brockhoff, B. Derbel, Multiobjectivization with NSGA-II on the Noiseless BBOB Testbed, in: *Genetic and Evolutionary Computation Conference*, ACM, Amsterdam, The Netherlands, 2013, pp. 1217–1224.

- [47] H. Singh, T. Ray, R. Sarker, Optimum Oil Production Planning Using Infeasibility Driven Evolutionary Algorithm, *Evolutionary Computation* 21 (1) (2013) 65–82.
- [48] M. Garza-Fabre, G. T. E. Rodriguez-Tello, Handling Constraints in the HP Model for Protein Structure Prediction by Multiobjective Optimization, in: *IEEE Congress on Evolutionary Computation*, Cancún, México, 2013, pp. 2728–2735.
- [49] S. Wright, The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution, *Proceedings of the 6th International Congress of Genetics* 1 (1932) 356–366.
- [50] P. Stadler, Fitness Landscapes, in: *Biological Evolution and Statistical Physics*, Vol. 585 of *Lecture Notes in Physics*, Springer Berlin / Heidelberg, 2002, pp. 183–204.
- [51] S. Verel, P. Collard, M. Tomassini, L. Vanneschi, Fitness Landscape of the Cellular Automata Majority Problem: View from the Olympus, *Theoretical Computer Science* 378 (1) (2007) 54–77.
- [52] L. Vanneschi, Y. Pirola, G. Mauri, M. Tomassini, P. Collard, S. Verel, A Study of the Neutrality of Boolean Function Landscapes in Genetic Programming, *Theoretical Computer Science* 425 (0) (2012) 34–57.
- [53] E. Pitzer, M. Affenzeller, A Comprehensive Survey on Fitness Landscape Analysis, in: *Recent Advances in Intelligent Engineering Systems*, Vol. 378 of *Studies in Computational Intelligence*, Springer Berlin Heidelberg, 2012, pp. 161–191.
- [54] K. Malan, A. Engelbrecht, A Survey of Techniques for Characterising Fitness Landscapes and Some Possible Ways Forward, *Information Sciences* 241 (2013) 148–163.
- [55] C. Anfinsen, Principles that Govern the Folding of Protein Chains, *Science* 181 (4096) (1973) 223–230.
- [56] K. Lau, K. Dill, A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins, *Macromolecules* 22 (10) (1989) 3986–3997.
- [57] H. Lopes, Evolutionary Algorithms for the Protein Folding Problem: A Review and Current Trends, in: *Computational Intelligence in Biomedicine and Bioinformatics*, Vol. 151 of *Studies in Computational Intelligence*, Springer Berlin / Heidelberg, 2008, pp. 297–315.
- [58] F. Custódio, H. Barbosa, L. Dardenne, A Multiple Minima Genetic Algorithm for Protein Structure Prediction, *Applied Soft Computing* 15 (Feb.) (2014) 88–99.
- [59] N. Krasnogor, B. Blackburne, E. Burke, J. Hirst, Multimeme Algorithms for Protein Structure Prediction, in: *Parallel Problem Solving from Nature*, Vol. 2439 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, Granada, Spain, 2002, pp. 769–778.
- [60] M. Islam, M. Chetty, Clustered Memetic Algorithm With Local Heuristics for Ab Initio Protein Structure Prediction, *IEEE Transactions on Evolutionary Computation* 17 (4) (2013) 558–576.
- [61] M. Rashid, M. Newton, M. Hoque, S. Shatabda, D. Pham, A. Sattar, Spiral Search: A Hydrophobic-Core Directed Local Search for Simplified PSP on 3D FCC Lattice, *BMC Bioinformatics* 14 (Suppl 2) (2013) S16.
- [62] P. Pardalos, X. Liu, G. Xue, Protein Conformation of a Lattice Model Using Tabu Search, *Journal of Global Optimization* 11 (1) (1997) 55–68.
- [63] A. Shmygelska, H. Hoos, An Ant Colony Optimization Algorithm for the 2D and 3D Hydrophobic Polar Protein Folding Problem, *BMC Bioinformatics* 6 (1) (2005) 30.
- [64] M. Nardelli, L. Tedesco, A. Bechini, Cross-lattice Behavior of General ACO Folding for Proteins in the HP Model, in: *ACM Symposium on Applied Computing, SAC '13*, ACM, Coimbra, Portugal, 2013, pp. 1320–1327.
- [65] C. de Almeida, R. Gonçalves, M. Delgado, A Hybrid Immune-Based System for the Protein Folding Problem, in: *Evolutionary Computation in Combinatorial Optimization*, Vol. 4446 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, Valencia, Spain, 2007, pp. 13–24.
- [66] V. Cutello, G. Nicosia, M. Pavone, J. Timmis, An Immune Algorithm for Protein Structure Prediction on Lattice Models, *IEEE Transactions on Evolutionary Computation* 11 (1) (2007) 101–117.
- [67] N. Mansour, F. Kanj, H. Khachfe, Particle Swarm Optimization Approach for Protein Structure Prediction in the 3D HP Model, *Interdisciplinary Sciences: Computational Life Sciences* 4 (3) (2012) 190–200.
- [68] C. Zhou, C. Hou, Q. Zhang, X. Wei, Enhanced Hybrid Search Algorithm for Protein Structure Prediction Using the 3D-HP Lattice Model, *Journal of Molecular Modeling* 19 (9) (2013) 3883–3891.
- [69] H. Lopes, R. Bitello, A Differential Evolution Approach for Protein Folding Using a Lattice Model, *Journal of Computer Science and Technology* 22 (6) (2007) 904–908.
- [70] J. Santos, M. Diéguez, Differential Evolution for Protein Structure Prediction Using the HP Model, in: *Foundations on Natural and Artificial Computation*, Vol. 6686 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2011, pp. 323–333.
- [71] R. Santana, P. Larranaga, J. Lozano, Protein Folding in Simplified Models With Estimation of Distribution Algorithms, *IEEE Transactions on Evolutionary Computation* 12 (4) (2008) 418–438.
- [72] B. Chen, L. Li, J. Hu, A Novel EDAs Based Method for HP Model Protein Folding, in: *IEEE Congress on Evolutionary Computation*, Trondheim, Norway, 2009, pp. 309–315.
- [73] X. Cai, X. Wu, L. Wang, Q. Kang, Q. Wu, Hydrophobic-Polar Model Structure Prediction with Binary-Coded Artificial Plant Optimization Algorithm, *Journal of Computational and Theoretical Nanoscience* 10 (6) (2013) 1550–1554.
- [74] B. Maher, A. Albrecht, M. Loomes, X. Yang, K. Steinhöfel, A Firefly-Inspired Method for Protein Structure Prediction in Lattice Models, *Biomolecules* 4 (1) (2014) 56–75.
- [75] H. Lourenço, O. Martin, T. Stützle, Iterated Local Search: Framework and Applications, in: *Handbook of Metaheuristics*, Vol. 146 of *International Series in Operations Research & Management Science*, Springer US, 2010, pp. 363–397.
- [76] C. Chira, A Hybrid Evolutionary Approach to Protein Structure Prediction with Lattice Models, in: *IEEE Congress on Evolutionary Computation*, New Orleans, LA, USA, 2011, pp. 2300–2306.
- [77] K. Dill, C. H., From Levinthal to Pathways to Funnels, *Nature Structural Biology* 4 (1) (1997) 10–19.
- [78] M. Marmion, C. Dhaenens, L. Jourdan, A. Liefoghe, S. Vérel, On the Neutrality of Flowshop Scheduling Fitness Landscapes, in: *Learning and Intelligent Optimization*, Vol. 6683 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2011, pp. 238–252.
- [79] L. Vanneschi, M. Tomassini, P. Collard, S. Vérel, Y. Pirola, G. Mauri, A Comprehensive View of Fitness Landscapes with Neutrality and Fitness Clouds, in: *Genetic Programming*, Vol. 4445 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, Valencia, Spain, 2007, pp. 241–250.
- [80] L. Barnett, Netcrawling - Optimal Evolutionary Search with Neutral Networks, in: *IEEE Congress on Evolutionary Computation*, Vol. 1, 2001, pp. 30–37.
- [81] T. Back, Selective Pressure in Evolutionary Algorithms: A Characterization of Selection Mechanisms, in: *IEEE Congress on Evolutionary Computation*, 1994, pp. 57–62.
- [82] S. Gould, N. Eldredge, Punctuated Equilibrium Comes of Age, *Nature* 366 (6452) (1993) 223–227.
- [83] L. Barnett, Ruggedness and Neutrality – the NKp Family of Fitness Landscapes, in: *International Conference on Artificial Life*, MIT Press, Madison, WI, USA, 1998, pp. 18–27.
- [84] M. Garza-Fabre, E. Rodriguez-Tello, G. Toscano-Pulido, Comparative Analysis of Different Evaluation Functions for Protein Structure Prediction under the HP Model, *Journal of Computer Science and Technology* 28 (5) (2013) 868–889.
- [85] V. Vassilev, T. Fogarty, J. Miller, Smoothness, Ruggedness and Neutrality of Fitness Landscapes: from Theory to Application, in: *Advances in Evolutionary Computing*, Natural Computing Series, Springer Berlin Heidelberg, 2003, pp. 3–44.
- [86] J. Zhang, S. C. Kou, J. S. Liu, Biopolymer Structure Simulation and Optimization via Fragment Regrowth Monte Carlo, *The Journal of Chemical Physics* 126 (22) (2007) 225101.
- [87] C. Thachuk, A. Shmygelska, H. Hoos, A Replica Exchange Monte Carlo Algorithm for Protein Folding in the HP Model, *BMC Bioinformatics* 8 (1) (2007) 342.
- [88] T. Wüst, Y. Li, D. Landau, Unraveling the Beautiful Complexity of Simple Lattice Model Polymers and Proteins Using Wang-Landau Sampling, *Journal of Statistical Physics* 144 (2011) 638–651.
- [89] C. Cotta, Protein Structure Prediction Using Evolutionary Algorithms Hybridized with Backtracking, in: *Artificial Neural Nets Problem Solving Methods*, Vol. 2687 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2003, pp. 321–328.