Bio-inspired optimization for the molecular docking problem: State of the art, recent results and perspectives

María Jesús García-Godoy [a], Esteban López-Camacho [a], José García-Nieto [a],

Javier Del Ser [b,c,d,*], Antonio J. Nebro [a], José F. Aldana-Montes [a]

[a] *University of Málaga, 29071 Málaga, Spain*
[b] *TECNALIA, 48160 Derio, Spain*
[c] *University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain*
[d] *Basque Center for Applied Mathematics (BCAM), 48009 Bilbao, Spain*

ABSTRACT

Molecular docking is a Bioinformatics method based on predicting the position and orientation of a small molecule or ligand when it is bound to a target macromolecule. This method can be modeled as an optimization problem where one or more objectives can be defined, typically around an energy scoring function. This paper reviews developments in the field of single- and multi-objective meta-heuristics for efficiently addressing molecular docking optimization problems. We comprehensively analyze both problem formulations and applied techniques from Evolutionary Computation and Swarm Intelligence, jointly referred to as Bio-inspired Optimization. Our prospective analysis is supported by an experimental study dealing with a molecular docking problem driven by three conflicting objectives, which is tackled by using different multi-objective heuristics. We conclude that genetic algorithms are the most widely used techniques by far, with a noted increasing prevalence of particle swarm optimization in the last years, being these last techniques particularly adequate when dealing with multi-objective formulations of molecular docking problems. We end this experimental survey by outlining future research paths that should be under target in this vibrant area.

1. Introduction

In the last few decades, the field of Bioinformatics has experienced an exponential growth of the amount of biological data available [1], leading to the rise of new challenges related to data storage, management, extraction and analysis. In particular, many Bioinformatics problems (e.g., molecular docking, protein structure prediction, phylogenetic inference, etc.) can be defined in terms of one or more objective functions that have to be minimized or maximized [2]. Most of such problems have been addressed in the literature by resorting to meta-heuristic techniques [3].

In the area of Genomics, the increase of the number of human DNA sequences available has allowed the detection of genes [4], which express proteins that can be considered as therapeutic targets for drug discovery. These candidate proteins are extracted by applying techniques, such as X-crystallography and Nuclear Magnetic Resonance spectroscopy (NMR). This has enabled to know many structural details of ligand–protein complexes, as well as the creation of databases such as the Protein Data Bank (PDB) [5]. This database stores information about biological macromolecule crystal structures, whose atom coordinates are utilized in *in silico* studies. These advances on laboratory and computational techniques have shed light on the approaches applied in drug discovery, and particularly molecular docking. This procedure consists in finding the best orientation and position of a given chemical compound (*ligand*) to a therapeutic target (*macromolecule*).

Molecular docking has significantly contributed to structure-based drug discovery as has been emphasized [6], postulated [7] and shown with successful protocols [8] in manifold contributions. Finding the ligand's best orientation and binding site can be formulated as an optimization problem, whose huge complexity has stimulated the adoption of non-exact search methods belonging to the family of meta-heuristics, such as Evolutionary Algorithms. Indeed, meta-heuristics have been applied to single- and multi-objective formulations of this problem, which is the main focus of this survey.

In this regard, addressing the optimization of a molecular docking problem requires to deal with two issues: the definition of the energy scoring function and the choice of the search method. The energy scoring function evaluates the quality of the ligand–macromolecule interactions through approximations to provide a compromise between accuracy and speed. The definition of energy scoring functions has been conducted over the last two decades to yield different main categories [9]: (1) force field-based methods, (2) empirical scoring functions, (3) knowledge-based potentials, and (4) descriptor-based scoring functions. These approaches can model the molecular docking simulations as rigid or flexible macromolecule–ligand conformations only if flexibility can be applied as degrees of freedom in the ligand or macromolecules bonds. Simulations assuming flexibility are more realistic than those in which flexibility is not endowed. The quality of the molecular docking solutions are measured in terms of the energy associated to the ligand–protein interactions (ΔG), which indicates the stability of the complex computed. Likewise, the RMSD (root-mean-square deviation) is also employed as a mean to gauge the distance between the co-crystallized ligands and the computed ones, measured in Angstroms (Å). The search method refers to the technique that is used to find the best solution, which represents the best ligand's conformation to the macromolecule in terms of energy. As anticipated above, the most widely adopted family of solvers for molecular docking are meta-heuristics, which have been used since the 90's [10]. Nevertheless, there is a lack of studies in which this kind of optimization techniques are reviewed towards providing biologists and researchers in applied metaheuristics with an organized corpus of references related to this research area, including a prospect of different perspectives and challenges.

Therefore, our motivation in this paper is to provide the audience with a comprehensive survey of contributed works related to bio-inspired meta-heuristic algorithms applied the molecular docking optimization problem. Our bibliographic study provides an overall overview, which we deem of utmost interest for both biologists and professionals in bioinformatics with strong background in the problem, as well as for computer scientists working on the wide field of bio-inspired meta-heuristics. The first target group of readers will find in this manuscript informed hints on which techniques can be more promising to be applied in practice, whereas the second group will discover a vast spectrum of features related with this real-world optimization problem, that can be useful to investigate and derive new heuristic techniques.

The contributions of this paper can be summarized as follows:

- The proposed survey focuses on different yet complementary perspectives: the different scoring functions that have been adopted in the literature, the diverse single/multi-objective formulations derived to model this problem, and the variety of bio-inspired algorithms used to deal with them. More than 115 references have been selected for revision, and catalogued accordingly.
- We provide practical guidelines to the audience by presenting and discussing results of an experimental case study

based on a three-objective formulation of the molecular docking problem on 75 instances with multi-objective meta-heuristics. A thorough algorithmic comparison with recommendations is included to support experts in the selection of algorithms.
- A prospective discussion is made to elaborate on open issues, trends and challenges with the aim of encouraging research community in this area to follow-up with molecular docking optimization.

The remainder of this paper is structured as follows: first, a background section devoted to bio-inspired optimization algorithms is included. Next, different categories of energy scoring functions are introduced, which is used as the baseline criterion to classify the literature in this field until today. Then we introduce and discuss different single- and multi-objective heuristic approaches applied to solve the molecular docking problem by sorting them by the number of objectives optimized in each reported contribution. Once we have examined the state of the art, we present and discuss the results obtained for the aforementioned molecular docking problem with three objectives. Finally, we discuss open issues, trends and challenged, and draw final remarks on the future of this particularly active research area.

2. Bio-inspired optimization algorithms

Bio-inspired metaheuristics [11] for solving complex optimization problems have become a very popular research topic in last decades. Among the myriad of methods proposed to date, classical techniques from the wide field of Evolutionary Computation – such as Genetic Algorithms (GA) or Differential Evolution – are arguably the most established solvers because of their versatility and relatively longer presence in the community. However, modern bio-inspired approaches such as Particle Swarm Optimization (PSO) [12], Artificial Bee Colony (ABC) [13], Cuckoo Search (CS) [14] and Firefly Algorithm [15] have started to attract more attention for continuous- and discrete-variable optimization problems, due to their prominent performance shown in manifold real-world problems. Most of these new metaheuristics hinge on search mechanisms that are rather different than those characterizing evolutionary search techniques, as we will expose briefly in what follows.

There are three main phases in most bio-inspired evolutionary algorithms: (1) initialization of each individual in a set of candidate solutions to the problem at hand (*population*), which is randomly generated according to some numerical representation (*solution encoding*); (2) each solution in the population is evaluated in terms of its fitness value, which can be used for ranking the individual solution within the population for the purpose of selecting or discarding it along the search process, (3) generation of a new population by recombination and/or mutation of solutions in the existing population, followed by the evaluation of the newly produced solutions (*offspring*) as per the fitness metric of the problem; (4) a replacement criterion is adopted and applied to select which solutions remain in the population and which ones are discarded due to their worse fitness. Selection, recombination, mutation, evaluation and replacement are repeated until a predefined stop condition is met. A general pseudocode of the whole optimization process undertaken by bio-inspired evolutionary algorithms is shown in Algorithm 1. At the end of this process, the best solution found so far along the search process is returned, which is declared to be the solution provided by the algorithm.

In regards to newer search heuristics like the ones exemplified above, their search process is rather guided by the interactions

Algorithm 1 Pseudocode of a generic evolutionary algorithm.

```
1: initializePopulation()
2: evaluatePopulation()
3: while not StopCondition() is met do
4:    selectParents()
5:    recombination()
6:    evaluateOffspring()
7:    replaceIndividuals()
8: end while
9: returnGlobalBestSolution()
```

between multiple agents endowed with simple behavioral rules for interaction and communication with their counterparts. It is the collective knowledge acquired by a set (*swarm*) of such agents where the search power of the overall algorithm resides, as their performed exploration benefits from the information about the problem transmitted and conveyed throughout the members of the swarm. A biological simile often lies underneath this wide class of solvers, such as the behavioral patterns observed in bird flocks or fish schools. This gives rise to the concept of *Swarm Intelligence* [16], which refers to all search algorithms that conform to the multi-agent procedure sketched previously.

3. The molecular docking problem

This section is devoted to the molecular docking problem. It is divided into three different subsections, concentrated on: (1) solution encoding strategy; (2) the energy scoring function, which can be classified according to four main categories; (3) the objectives to be minimized; and (4) the search method used to find the best ligand–macromolecule interactions (solutions).

3.1. Solution encoding

Molecular docking processes are aimed at finding an optimized conformation between the ligand (L) and the receptor (R) that results in a minimum binding energy. The interaction between L and R can be described by an energy function calculated from three components representing degrees of freedom: (1) the translation of the ligand molecule, involving the three axis values (x, y, z) in cartesian coordinate space; (2) the ligand orientation, modeled as a four variables quaternion including the angle slope (θ); and (3) the flexibilities, represented by the free rotation of torsion (dihedral angles) of the ligand and side chains of the receptor.

Fig. 1 illustrates the most commonly used strategy for solution encoding, which consists of a real-value vector of $7 + n$ variables. In this vector, the first three values correspond to the ligand translation, the next four values correspond to the ligand and/or receptor orientation, and the remaining n values are the ligand torsion dihedral angles. In order to reduce the computational cost, a grid-based methodology can be carried out by which the protein active site is embedded in a 3D rectangular grid, and by taking into account all the protein atoms. In this way, the protein contribution at any given point is obtained by trilinear interpolation in the grid cell. As a consequence of this process, the range of translation variables (x, y, z) is $[0 \cdots 120]$, which bounds effectively the limits of the coordinates of the grid space previously set for each problem (and in essence, reduces the size of its search space). Orientation (quaternion) and torsion variables are measured in radians and encoded in the range of $[-\pi, \pi]$.

3.2. Energy scoring functions for the molecular docking problem

The energy scoring function allows evaluating the solutions to the problem (ligand–macromolecule interactions), making approximations that establish a trade-off between accuracy and speed. In this context, a complete review by Liu et al. [9] categorizes different energy scoring functions reflecting the current progression in this field. This categorization is composed by the following categories: (1) force-field-based energy functions, (2) empirical or regression-based energy functions, (3) knowledge-based scoring functions and (4) descriptor-based scoring functions.

Functions that were first developed are included in the category of force-field-based energy functions proposed by Martin Karplus et al. [17,18]. This type of energy scoring function was proposed to calculate the ligand–macromolecule interactions through the inclusion of non-covalent terms, like the so-called Van der Waals interactions, the electrostatic energy and H-bond terms. Some molecular docking software programs incorporating this evaluation function are AutoDock [19], DOCK [20], COMBINE [21], GoldScore [22], MedusaScore [23] and CHARMM [24]. The following general functional form became more widely adopted for this type of energy function:

$$\Delta G = \Delta E_{vdw} + \Delta E_{electrostatic} + \Delta E_{H-bond} + \Delta E_{desolvatation} \tag{1}$$

The second type of evaluation energy functions is referred to as *empirical scoring function*. There are a few of exemplifying software frameworks equipped with this class of scoring functions, such as ChemScore [25], X-Score [26] and GlideScore [27]. The total score is composed of a sum of scores based on H-bond interactions, bonds in which metals and lipophilic atom contacts are involved and a set of penalty scores based on rigid rotatable bonds, ligand–macromolecule steric clashes, and other terms based on ligand–macromolecule interactions' constraints that could be useful, for example, to limit the search space of the considered search method (see Section 3.4). An example of equation that represents this method is as follows:

$$S_{total} = S_{H-bond} + S_{metal} + S_{lipophilic} + P_{rotatable} + P_{clashes} \tag{2}$$

A third category of energy functions corresponds to *knowledge-based scoring function*, as they are known in the reviewed literature. Software packages using these functions include DrugScore [28] and IT-SCORE [29]. The value of this energy function builds upon the sum of pairwise statistical potentials between protein and ligand, e.g.:

$$A = \sum_{n}^{lig} \sum_{m}^{macro} \omega_{ij}(r), \tag{3}$$

where *lig* and *macro* refer to the ligand and the macromolecule, respectively.

A fourth category is based on machine learning based-scoring functions that use modern Quantitative Structure–Activity Relationship (QSAR) analysis into the ligand–protein interactions. This type of energy function extracts and exploits descriptors from the ligand and macromolecule to improve the solutions of molecular docking simulations. This is an emerging strand in the area that can be very promising if combined with other energy scoring functions and/or search methods, such as meta-heuristics.

3.3. Objectives of the molecular docking problem

The main objective of the molecular docking problem is to find the best conformation between a ligand (L) and a receptor (R) in terms of energy and RMSD. This goal can be formulated as a single- or a multi-objective optimization problem. Therefore, the
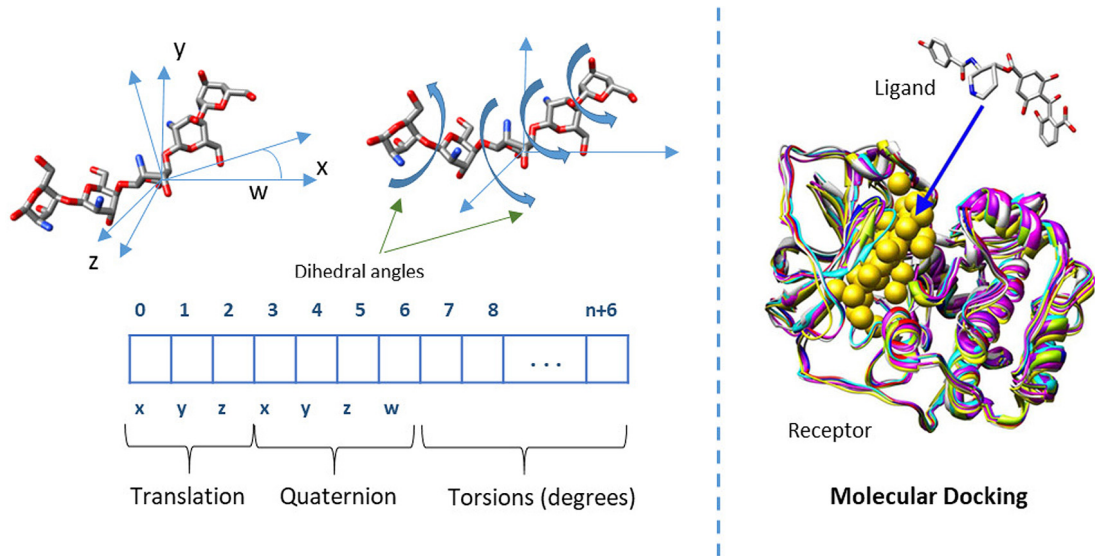
Fig. 1. Solution encoding. The first three values (translation) are the coordinates of the center of rotation of the ligand. The next four values correspond to quaternion and (θ). The rest of the values hold the torsion angles in degrees.

molecular docking problem can be formally expressed by letting A and B denote the ligand and protein molecule, respectively. If we define $f : \mathbb{R} \rightarrow \mathbb{R}$ as a scoring function that evaluates solutions with regards to a given score, and let $\mathcal{C}$ be the set of all feasible molecular docking solutions between A and B, then the molecular docking problem aims at finding a conformation $\overrightarrow{x} \in \mathcal{C}$ that satisfies $f(\overrightarrow{x}) \leq f(\overrightarrow{y}) \ \forall \overrightarrow{y} \in \mathcal{C} : \ \overrightarrow{y} \neq \overrightarrow{x}$.

According to the reviewed literature, if molecular docking is formulated as a single-objective optimization problem, the fitness function to optimize is usually set to the final binding energy ΔG. However, in the middle of 2000's, Grosdidier et al. [30] and Janson et al. [31] proposed new approaches in which multi-objective optimization was applied to the problem at hand by simultaneously minimizing several objectives that will be described below. These studies were subsequently followed by the work performed by Liu et al. [32] and Boisson et al. [33]. There are other remarkable studies published in the last decade that embraced this design choice, such as [12,34–37]. More recently, a new approach to the problem was proposed by López-Camacho et al. [38], where three objectives were first considered within the formulated molecular docking problem, thereby unleashing new perspectives over the field of multi-objective optimization strategies that could potentially improve the accuracy of drug discovery.

In a general multi-objective problem, it is necessary to define two spaces: (1) the decision space, that includes the set of feasible solutions; and (2) the objective space, which involves the fitness (scoring) values of the objectives to be optimized. Objectives formulated for molecular docking in the literature can be enumerated as follows:

- In Grosidier et al. [30], objectives that were minimized correspond to two scores resulting from a *SimpleFitness* and a *FullFitness*. The first objective computes the sum of the intramolecular energy (E_{intra}) of R and L, and the intermolecular energy (E_{inter}) of the complex R-L in Expression (4). The second objective is a term that corresponds to the solvation energy ΔG given in Expression (5):

$$SimpleFitness = E_{intra}^{R} + E_{intra}^{L} + E_{inter}, \tag{4}$$

$$FullFitness = SimpleFitness + \Delta G_{elec.solv}. \tag{5}$$

- In Janson et al. [31], Sandoval-Perez et al. [34] and García-Godoy et al. [36], the objectives used were:

$$E_{intra} = E_{intra}^{L} +_{intra}^{R}, \tag{6}$$

$$E_{inter} = E_{intra}^{R-L} +_{intra}^{R-L}, \tag{7}$$

where E_{intra} refers to the unbound and bound states of the R and L, and E_{inter} refers to the unbound and bound states of the complex R-L.

- In Boisson et al. [33], the authors proposed to optimize E_{inter} as per (7), and a surface criterion based on the hypothesis that the surface of L and R decreases because of the penetration of L into R.

- Liu et al. [32] elaborated on an approach that optimized the energy associated to the Van der Waals and torsion bonds, namely:

$$E_{L-R} = \Delta S_{conf} + W_{vdW}, \tag{8}$$

and a term based on the geometry dissimilarity (GD):

$$GS = \sqrt{\frac{\sum_{i=1}^{N} \|x_{indiv,i} - x_{input,i}\|^2}{N}}, \tag{9}$$

where $x_{indiv,i}$ and $x_{input,i}$ are the positions of the ith heavy atom in each individual and the input L-R conformer.

- Gu et al. [35]: The authors proposed an approach that minimizes the resulting score from three different scoring energy functions (see Section 3.2): the force-field-based, empirical-based and knowledge-based functions (see Eqs. (1), (2), (3)), which are treated as objectives during the docking optimization.

- In López-Camacho et al. [37] and García-Nieto et al. [12] the molecular docking problem is defined on the basis of two different metrics to be optimized, E_{inter} and the RMSD score. This score measures the average distance between the known ligand position in the receptor and the computed position of the docking ligand, which takes into account symmetry, partial symmetry (e.g. symmetry within a rotatable branch) and near-symmetry in a simple heuristic fashion. This score is given by:

$$RMSD_{ab} = \max(RMSD'_{ab}, RMSD'_{ba}), \tag{10}$$

where $RMSD'_{ab} = \sqrt{\frac{1}{N} \sum_i \min_j r_2^{ij}}$, and the sum is over all N heavy atoms in structure a, the minimum is over all atoms in structure a with the same element type as atom i in structure b.

- In García-Godoy et al. [38], a three-objective molecular docking problem is first posed and solved, based on the simultaneous optimization of E_{inter}, E_{intra} and the RMSD score defined above. Interestingly, this preliminary work unveiled the inherent benefits springing from considering a third objective in the molecular docking problem, thus settling the intuition in which this manuscript finds its motivation.

3.4. Meta-heuristics for the molecular docking problem

We have reviewed the most relevant works attending the applied meta-heuristics by considering two main categories: single- and multi-objective algorithms. This categorization can be very useful not only for practitioners that are interested in using one approach or another, but also for developers willing to implement innovative search methods. Furthermore, these approaches have been also organized taking into account the standard categorization of scoring functions provided by Liu et al. [9].

3.4.1. Single-objective algorithms

The use of molecular docking tools to predict the ligand-interactions were pioneered in the early 80's, but it was not until the 90's when researchers started to apply meta-heuristics to optimize single-objective formulations of this problem, mainly considering the free binding energy ΔG. During this decade, most published works were based on assessing new search methods that used different scoring functions, being one of the most widely used the one defined by Martin Karplus et al. [17,18], who proposed the use of force fields to compute the interactions between ligands and macromolecules. Some of these functions accommodate some degrees of freedom in the ligands' and macromolecules' bonds, which increases the complexity of the problem and the computation time required to solve it to a given accuracy, but yields more realistic molecular docking simulations.

In this context, studies such as [19,22,39–44] gravitated on the application of GA by minimizing the final binding energy. Previously, we have also mentioned that some of these early force-field scoring functions allow to apply flexibility to the ligands' conformations as the case of the studies performed in [19,22,39,42,43]. For example, the authors in [39] were the first to apply flexibility to three instances of the molecular docking problem, and the studies provided by Morris et al. [19] questioned which search method is more efficient in escaping from a local minimum, concluding that the Lamarckian Genetic Algorithm (LGA), a GA with Local Search (LS) obtained the best solutions, followed by GA and Simulated Annealing (SA), which failed to return better solutions in terms of energy and RMSDs.

This decade is also characterized by studies based on the use of the Monte Carlo algorithm. For example, [45] used a force-field based fitness function and Monte Carlo as search method in flexible ligands' instances. In [46], the authors also proposed a Monte Carlo algorithm to predict 20 molecules' interactions where flexibility was assumed.

In [47], a comparative analysis was carried out to assess the performance of several search methods: GA, Tabu Search (TS), SA and an Evolutionary Programming (EP) algorithm. The authors showed that the GA outperforms the rest of algorithms in terms of energy. Furthermore, the provided discussion suggested that GA is a very effective local search algorithm, but prone to falling in low-energy local minimum. TS, on its own, seems to perform a more global search, which permits to locate a global energy minimum, but less accurately than GA. The authors noted the importance of publishing studies about the search methods to improve future techniques to be applied to molecular docking simulations.

In the early 2000's, the use of EAs continued playing a prominent role, and a tendency of performing comparative analysis emerged at that time. For example, in [48], the authors proposed an EA that used global and local search strategies by integrating decreasing mutations and self-adaptive mutations. This approach was applied to a use case based on a flexible ligand-rigid protein complex. In [49], an EA was designed to minimize E_{inter}, evaluating the solutions by means of the CHARMM tool. The method is applied to three instances of the problem. Budin et al. [50] applied a method based on an EA by optimizing the energy of H-bonds of a set of complexes where flexibility was applied to the ligand (10 rotatable bonds). In [51], a new optimizer was proposed based on GA, Global Underestimator Method and Pole and applied to a set of 27 rigid instances. Solutions were evaluated by a force-field-based energy scoring function. In [28], the authors revolved on a new scoring function, which is part of the knowledge-based scoring function group, as well as on the LGA solver provided by AutoDock. Other remarkable studies include [52,53]. The former uses a multi-population GA and a force-field energy scoring function to evaluate the solutions from a set of 11 complexes with flexible ligands. The latter applies an GA-like method with an empirical energy function for seven complexes, where flexibility is applied to the ligand. At this point, we can conclude that these published works have several features in common: (1) the number of complexes is not enough to do a comparative analysis, (2) flexibility is only applied to the ligand's conformation; and (3) the number of algorithms applied to perform the comparative analysis is limited.

More recently, there is an uprising trend in the field to use Swarm Intelligence techniques rather than Evolutionary Computation approaches to tackle molecular docking based on optimizing a single objective. For example, Nesamalar and Chandran [13] introduced a Genetic Clustering combined with Bee Colony Optimization (BCO). The approach, coined as (GCBCO), was tested by using 10 docking instances from the PDB bind core set. The results showed that GCBCO outperforms other Swarm Intelligence algorithms (specifically, PSO and Ant Colony Optimization - ACO) in terms of energy. The authors in [54] utilized the AutoDock 4.2 energy function to include flexibility in the macromolecule and ligand. Another novel ingredient of this recent reference was a new search algorithm hybridizing features of a quantum-behaved particle swarm optimization (QPSO) and the local search method of Solis and Wets. A thorough analysis of the algorithm's performance was made by comparing the results obtained by QPSO with those rendered by the LGA optimizer embedded in AutoDock.

3.4.2. Multi-objective algorithms

The first studies to consider a multi-objective optimization approach appeared in 2007 [30,31]. In [30], Grosididier et al. presented EADock, which is a multi-objective formulation based on an EA and the minimization of two objectives from two different fitness functions: the total binding energy and the solvation energy as per Eqs. (4) and (5), respectively. In [31], a multi-objective approach based on PSO (namely, ClutMPSO) was shown to minimize efficiently E_{inter} and E_{intra}. The ligand–macromolecule interactions were evaluated with the early energy function of AutoDock 3.0. Other notable contributions to the multi-objective formulation of the molecular docking problem are [34,36], where the same objectives as in [31] are optimized, and [35]. In the first work, a force-field-based energy function is used to evaluate solutions, and the NSGA-II multi-objective meta-heuristic is used as the search method. In the second work, 11 instances were tested

by multi-objective solvers assuming flexibility in both the ligand and macromolecule. Search methods that were used ranged from NSGA-II to ssNSGA-II, SMPSO, MOEA/D and SMS-EMOA. Interestingly SMPSO, a Swarm Intelligence multi-objective optimization method that resembles the search mechanisms underlying the classical PSO solver, was shown to attain outperforming results in terms of energy and RMSD, when compared with the rest of algorithms. In the third work, the authors proposed a multi-objective approach (MoDOCK) that minimizes three fitness scores from different scoring functions. This approach is compared with single-objective approaches that used popular scoring functions, such as: DOCK, Glide, SurFlex, DOCK6 and DOCK6-F. The analysis indicated that MoDOCK dominated the rest of approaches in terms of RMSD.

In [37,55], the authors propose a multi-objective approach to optimize E_{inter} and the RMSD score using the energy function of AutoDock 4.2 to enable flexibility in the macromolecule and the ligand. The selected algorithm were NSGA-II, SMPSO, GDE3, and MOEA/D. In [37], it was again observed that SMPSO showed the best overall results in terms of energy and RMSD (value lower than 2 for successful docking results). Finally, in the most recently study reported in [38] three-objectives were considered for minimization: E_{inter}, E_{intra} and RMSD. The solvers utilized in this late study were SMPSO, MOEA/D and MPSO/D, whereas the AutoDock 4.2 function was considered to model flexibility in the macromolecule and the ligands. The obtained results in this preliminary study laid the foundations for more elaborated experiments as the one later presented and discussed in this manuscript.

4. A survey of molecular docking using bio-inspired optimization

This section classifies the most relevant articles that deal with the molecular docking problem using bio-inspired meta-heuristics, by conforming to the topics developed in the previous sections. In this way, the literature can be classified depending on three different taxonomic aspects: the first one deals with the scoring function used to evaluate the solutions produced by the different bio-inspired meta-heuristics; the second one reduces to the particular problem objectives that are under consideration in each work. The third classification criterion takes into account which family of bio-inspired meta-heuristic techniques have been applied to solve molecular docking problem.

The main source of information we have used to find the papers considered in this survey has been Scopus®.[1] The search terms utilized to query this database have been composed by combinations of "molecular docking", "protein docking", "genetic algorithm", "evolutionary algorithm", "swarm intelligence". From all the retrieved results, we have selected the most representative works for reaching informed insights and conclusions about the general trends of this literature strand.

As a starting point of our bibliographic analysis, Table 1 organizes related work according to scoring functions that have been used along the years. These scoring functions can be divided into four types as discussed in Section 3.2: force-field-based energy functions, empirical or regression-based energy functions, knowledge-based scoring functions and descriptor-based scoring functions. Works using force-field based energy functions are more common in the area, being the AutoDock scoring function the most prominent among them. Some studies resort to their own formulation of the force-field based function.

The bibliographic taxonomy follows in Table 2, although in this case the literature comprising multi-objective approaches

is categorized in terms of the selected optimization objectives, bearing in mind that when one single objective is considered, the intermolecular energy E_{inter} is always chosen.

Table 3 inspects the use of different bio-inspired meta-heuristics to solve the molecular docking problem from the 90's to the present day. It is remarkable that the use of GA has been prominent all over the history of the field, arguably by the particularly good results achieved by this solver when addressing molecular docking problems. Some approaches using the Monte Carlo algorithm (MC) were noticed at the dawn of the field, but this approach has fallen into disuse. PSO techniques have been used more frequently in recent years, as they were empirically found to improve existing methods at the time. It is worth noting that LS approaches have been also hybridized with other common meta-heuristics, hence tracing a promising research path around Memetic Computing [104], a paradigm within Bio-inspired Optimization that precisely embraces this hybridization as its core design principle.

5. Experimental study: optimizing three objectives

In the previous section, we have described the historical evolution of the research activity around the use of bio-inspired meta-heuristics to deal with the molecular docking problem. To provide readers interested in this subject with empirical evidences of the practical benefits derived from this research area, in this section we conduct a study where three state-of-the-art bio-inspired meta-heuristics are used to solve a molecular docking problem considering three objectives: the inter-molecular energy E_{inter}, the intra-molecular energy E_{intra} and the Root Median Square Deviation (RMSD) score. As commented previously, E_{inter} relates to the ligand–receptor affinity, E_{intra} characterizes the ligand deformity, and the RMSD score measures the difference of atomic distances between the co-crystallized ligand and the computed one.

The study involves both, a quantitative analysis of the solutions found by the algorithms (computer scientist perspective) and a qualitative analysis of some selected results (biologist perspective). Three multi-objective algorithms were selected for the experimental part of this study due to their renowned good performance in other multi-objective problems: SMPSO and MOEA/D, which have yielded good results when dealing with this problem in the recent past [36]; and MPSO/D, recently applied to the same problem optimizing only two objectives (E_{inter} and RMSD) and using rigid ligand–protein complexes [12]. We next describe them briefly for the sake of completeness:

- **SMPSO** is a multi-objective PSO algorithm that incorporates a velocity restriction mechanism. The approach allows producing new effective particle positions in those cases where the velocity becomes too high. This solver also includes additional features, such as the use of polynomial mutation as a turbulence factor and an external archive to store non-dominated solutions found during the search. In [116], a comparative analysis of SMPSO with five representative multi-objective solvers was performed in terms of the quality of the resulting approximation sets and the convergence speed to the Pareto front. The obtained results showed that SMPSO returned the best overall results in such terms.
- **MOEA/D** is a multi-objective EA based on decomposition [117]. This algorithm divides the multi-objective problem into a number of scalar optimization subproblems, which are solved simultaneously by evolving the population. At each generation, the best solution for each subproblem is obtained. The neighborhood relations among these subproblems are defined based on the distances between their aggregation coefficient vectors. The optimization of each subproblem is based on the information from the subproblems within its neighborhood.

Table 1
Classification of the literature in terms of scoring functions.

Year	Reference	Force field						Empirical	Knowledge (DrugScore)	Descriptor (QSAR)
		Other	DOCK	GOLD	AutoDock	AMBER	GROMOS			
1994	Xiao and Williams [56]	✓								
	Xiao and Williams [57]	✓								
	Knegtel1 et al. [58]	✓								
1995	Oshiro et al. [39]		✓							
	Clark and Ajay [59]				✓					
	Jones et al. [60]	✓								
	Gehlhaar et al. [42]	✓								
	Read et al. [46]	✓								
	Verkhivker et al. [41]	✓								
1997	Westhead et al. [47]	✓								
	Maddalena and Snowdon [61]									✓
	Jones et al. [22]			✓						
	Levine et al. [44]	✓								
1998	Vieth et al. [62]	✓								
	Morris et al. [19]				✓					
1999	Chen and Chi [63]	✓								
	Liu and Wang [45]	✓								
2000	Yang and Kao [48]				✓					
	Taylor and Burnett [49]	✓								
2001	Budin et al. [50]	✓								
	David et al. [51]	✓								
	Pegg et al. [64]	✓								
2002	Hou et al. [65]	✓								
	Sotriffer et al. [28]								✓	
2003	McGann et al. [66]							✓		
2004	Lee et al. [67]				✓					
	Li et al. [52]	✓								
	Yang and Chen [53]	✓								
	Magalhães et al. [68]					✓				
2005	Li et al. [69]	✓								
	Janson and Merkle [70]	✓								
	Li et al. [71]	✓								
	Li et al. [72]	✓								
	Chang et al. [73]	✓								
2006	Wiley et al. [74]	✓								
	Oduguwa et al. [75]	✓								
	Chen et al. [76]	✓								
2007	Grosdidier et al. [30]	✓								
	Zhao and Sanner [77]				✓					
	Kroemer [78]	✓								
2008	Kang et al. [79]							✓		
	Sung [80]				✓					
	Janson et al. [31]				✓					
2009	Liu et al. [32]	✓								
	Thiriot and Monard [81]							✓		
	Liu et al. [82]	✓								
	Tavares et al. [83]				✓					
2010	Fuhrmann et al. [84]	✓								
	Ling et al. [85]	✓								
	Atilgan and Hu [86]				✓					
	Archetti et al. [87]									✓
	Meier et al. [88]				✓					
	Boisson et al. [33]	✓								
	Mesmoudi et al. [89]				✓					
2011	Rondón et al. [90]					✓				
2012	Liu et al. [91]				✓					
	Lima et al. [92]	✓								
	Nesamalar and Chandran [13]				✓					
	Koohi-Moghadam and Rahmani [93]				✓					
	Liu et al. [94]				✓					
2014	Zhen-yu et al. [95]	✓								
2015	Fu et al. [54]				✓					
	Li et al. [96]	✓								
	Ravindranath et al. [97]				✓					
	Xu et al. [98]	✓								
	García-Godoy et al. [36]				✓					
	López-Camacho et al. [99]				✓					
2016	López-Camacho et al. [37]				✓					
	García-Nieto et al. [55]				✓					
	Guan et al. [100]				✓					
2018	Leonhart et al. [101]				✓					
	Guan et al. [102]				✓					
	García-Nieto et al. [12]				✓					
	Lopez-Camacho et al. [38]				✓					
	Fu et al. [103]				✓					

Table 2
Classification of the literature in terms of optimization objectives.

Year	Reference	E_{intra}	E_{inter}	$\Delta_{elec\ solv}$	S_{L-R}	ΔS_{conf}	E_{vdW}	GS	RMSD
2007	Grosidier et al. [30]	✓		✓					
2008	Janson et al. [31]	✓	✓						
2010	Boisson et al. [33]		✓		✓				
2013	Sandoval-Perez et al. [34]	✓	✓						
2015	García-Godoy et al. [36]	✓	✓						
	Liu et al. [9]					✓	✓	✓	
	Gu et al. [35]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2016	López-Camacho et al. [37]		✓						✓
2018	García-Nieto et al. [12]		✓						✓

Table 3
Classification of the literature in terms of the utilized meta-heuristics (ABC: Artificial Bee Colony)

Year	Reference	MC	EA			TS	SA	PSO	ACO	ABC	LS
			GA	DE	Other						
1994	Xiao and Williams [56]		✓								
	Xiao and Williams [57]		✓								
	Knegtel1 et al. [58]	✓									
1995	Oshiro et al. [39]		✓								
	Clark and Ajay [59]				✓						
	Willett [40]		✓								
	Jones et al. [60]		✓								
	Gehlhaar et al. [42]				✓						
	Read et al. [46]	✓									
1996	Clark and Westhead [43]				✓						
	Verkhivker et al. [41]		✓								
1997	Westhead et al. [47]		✓		✓	✓					
	Maddalena and Snowdon [61]		✓								
	Jones et al. [22]		✓								
	Levine et al. [44]		✓								
1998	Vieth et al. [62]	✓	✓								
	Morris et al. [19]		✓				✓				✓
	Szczerbicka et al. [105]		✓								
1999	Chen and Chi [63]		✓				✓				
	Liu and Wang [45]	✓									
2000	Yang and Kao [48]				✓						
	Taylor and Burnett [49]				✓						
2001	Budin et al. [50]		✓								
	David et al. [51]		✓			✓					
	Pegg et al. [64]		✓								
2002	Hou et al. [65]		✓			✓	✓				
	Sotriffer et al. [28]		✓								✓
2004	Lee et al. [67]	✓	✓								
	Schneidman-Duhovny et al. [106]		✓								
	Magalhães et al. [68]		✓								
2005	Li et al. [69]		✓								
	Janson and Merkle [70]		✓					✓			✓
	Li et al. [71]		✓								
	Li et al. [72]		✓								
	Chang et al. [73]			✓							✓
2006	Wiley et al. [74]		✓								✓
	Oduguwa et al. [75]			✓							
	Chen et al. [76]							✓			✓
2007	Grosdidier et al. [30]				✓						
	Zhao and Sanner [77]		✓								
	Kroemer [78]	✓	✓			✓					
2008	Kang et al. [79]		✓								
	Sung [80]		✓								
	Janson et al. [31]							✓			
2009	Liu et al. [32]							✓			
	Thiriot and Monard [81]		✓								
	Liu et al. [82]			✓							
	Tavares et al. [83]		✓	✓							✓
2010	Fuhrmann et al. [84]		✓								
	Ling et al. [85]		✓								
	Atilgan and Hu [86]		✓								
	Archetti et al. [87]				✓						
	Chang et al. [107]				✓						
	Meier et al. [88]							✓			
	Boisson et al. [33]		✓								
	Mesmoudi et al. [89]				✓						

Table 3 (continued).

Year	Reference	MC	EA			TS	SA	PSO	ACO	ABC	LS
			GA	DE	Other						
2011	Rondón et al. [90]		✓								
2012	Lima et al. [92]		✓								
	Nesamalar and Chandran [13]		✓					✓	✓		
	Liu et al. [94]		✓					✓			✓
2013	Sandoval-Perez et al. [34]				✓						
	López-Camacho et al. [108]		✓	✓				✓			
	Rakshit et al. [109]		✓					✓		✓	
	Liu et al. [110]									✓	
2014	Magalhaes et al. [111]		✓								
	Zhen-yu et al. [95]		✓								
2015	Fu et al. [54]							✓			✓
	Li et al. [96]		✓								
	Ravindranath et al. [97]		✓								
	Xu et al. [98]		✓								
	García-Godoy et al. [36]		✓	✓	✓			✓			
	López-Camacho et al. [99]		✓	✓				✓			
2016	López-Camacho et al. [37]		✓	✓				✓			
	García-Nieto et al. [55]							✓			
	Guan et al. [100]		✓								
	García-Godoy et al. [112]		✓	✓				✓			
	Peh and Hong [113]										✓
2017	Cecilia et al. [114]		✓								
	Guan et al. [115]		✓								
2018	Leonhart et al. [101]		✓								
	Guan et al. [102]			✓						✓	
	García-Nieto et al. [12]							✓			
	Lopez-Camacho et al. [38]			✓				✓			
	Fu et al. [103]							✓			

- **MPSO/D** is a relatively new multi-objective PSO algorithm based on decomposition [118]. This algorithm decomposes the objective problem into a set of sub-regions on the basis of a set of direction vectors. Each subregion has a solution that maintains a level of diversity. MPSO/D also exploits the crowding distance to calculate the fitness of the solutions for the selection operator, as well as the neighboring particles (solutions) of a given particle to infer the global best historical position found by the algorithm.

These three multi-objective algorithms have been implemented in C++ as part of the jMetalCpp[2] project [108], which is essentially a software framework for single- and multi-objective optimization meta-heuristics. The energy score evaluation has been performed by AutoDock 4.2 [119] which, as mentioned before, is a widely used tool for virtual drug discovery involving rigid and flexible docking simulations. The integration of AutoDock with jMetalCpp[3] provides a flexible capacity to use the majority of the meta-heuristics implemented for the molecular docking problem. Specifically, jMetalCpp is in charge of evolving solutions as per the search operators of the meta-heuristic algorithm under choice, only delegating in AutoDock the evaluation of a newly produced candidate.

This experimental use case was carried out over an extensive set of 75 molecular complexes that have been used in many other studies [119]. Table 4 summarizes this set of instances, including the PDB codes of these structures and the range of the crystalographic resolution in Å. Moreover, these complexes are classified in terms of the type of their ligand: *small*, *medium*, and *large* size inhibitors, as well as *cyclic urea* inhibitors. All these compounds are characterized by a flexibility of the ligand limited to 10 torsional degrees of freedom. These instances were previously preprocessed (both the macromolecule and the ligand), and their potentials were calculated for a three-dimensional square (x, y, z) grid of 120 Å per dimension and a grid spacing of 0.375 Å, using the AutoGrid4 software.

Once all the macromolecules and ligands have been prepared, we executed 30 independent runs for each pair of algorithm and instance. Table 5 indicates the parameter values of the three algorithms included in the benchmark. Each used a population/swarm size of 100 individuals (or particles), and were limited to a maximum of 1,000,000 fitness evaluations. The rest of parameters were set with their nominal values.

We used two quality indicators to evaluate the algorithms' performance: the Hypervolume (I_{HV}) [120] and the Unary Additive Epsilon Indicator ($I_{\epsilon+}$) [121]. I_{HV} gives a measure of both convergence and diversity, whereas the $I_{\epsilon+}$ takes into account the convergence degree of the obtained Pareto front approximations. As molecular docking is a real-world problem, the Pareto front is not known a priori. Consequently, in order to calculate these two measures for each instance, a reference front was generated using all the non-dominated solutions obtained for an instance using all executions of the three algorithms. For these indicators, we calculated the median and the Inter-Quartile Range (IQR) as statistical measures of central tendency and dispersion. Tables 6 and 7 list such statistics for I_{HV} and $I_{\epsilon+}$, respectively. The cells shaded in dark gray indicate the algorithm scoring the best median value for the corresponding indicator, whereas those in light gray refer to the algorithm achieving the second best median indicator value for each instance.

In the I_{HV} indicator, the higher the convergence and diversity degree of a front is, the higher (better) the resulting I_{HV} value will be. We can observe in Table 6 that MOEA/D obtained the best median value in 17 instances and the second best value in 18 instances. SMPSO obtained the best and second best median value in 38 and 17 instances, respectively. Finally, MPSO/D achieved the best and second best median value 20 and 40 times over the benchmark. Taking into account the $I_{\epsilon+}$ indicator, the higher the convergence of a front is, the lower (and better) the value of $I_{\epsilon+}$ will be. Having this in mind, it can be observed that MOEA/D obtained 14 best median values and 18 s best median values of this indicator. SMPSO obtained the better results with 43 and

Table 4

X-ray crystal structure coordinates taken from the PDB database and used in our docking experiments. They consist of 75 molecules with accession codes from the PDB database. The range of resolution (Å) of each subgroup is shown in the last column.

Ligand type	PDB code	Resolution (Å)
Small size	1a9m, 1aaq, 1b6l, 1b6m, 1bdl, 1bdq, 1bdr, 1gnm, 1gnn, 1gno, 1hbv, 1heg, 1hih, 1hpv, 1hsg, 1hte, 1kzk, 1sbg, 1tcx, 1zih, 1zir, 3aid	1.09-2.8
Medium size	1b6j, 1b6k, 1b6p, 1d4k, 1d4l, 1hef, 1hps, 1hxw, 1izh, 1izi, 1jld, 1k6c, 1k6p, 1k6t, 1k6v, 1mtr, 1mui, 2bpv, 2bpx, 4hvp, 4phv, 5hvp	1.75-2.8
Large size	1a94, 1hiv, 1hos, 1htg, 1hvi, 1hvj, 1hvk, 1hvl, 1hvs, 1hwr, 1ody, 1vij, 1vik, 3tlh, 7hvp, 8hvp, 9hvp	1.8-2.8
Cyclic urea	1bv7, 1bv9, 1bwa, 1bwb, 1dmp, 1g35, 1hpo, 1mes, 1meu, 1pro, 1qbr, 1qbt, 1qbu, 7upj	1.8-2.5

Table 5

Parameter setting used for SMPSO, MPSO/D and MOEA/D.

Algorithm	Parameter	Value
All	*Swarm size/Population size*	100 Particles/Individuals
	Maximum number of evaluations	1,000,000
SMPSO [116]	*Archive Size*	100
	C_1, C_2	1.5
	ω	0.9
	Mutation	polynomial mutation
	Mutation probability	1.66
	Mutation distribution index η_m	20
	Selection method	Rounds
MPSO/D [118]	C_1, C_2	$rand(1.5, 2.0)$
	ω	$rand(0.1, 0.5)$
MOEA/D [117]	μ	0.5
	C_r	1.0
	Mutation probability	$p_m = 1/$(number of decision variables)

13 best and second best median values respectively. MPSO/D obtained 18 best median values and 44 s best median values.

These observations on the obtained indicator values lead us to conclude that SMPSO achieves the best results (with a better convergence and diversity than the rest of solvers in the benchmark), followed by MPSO/D and MOEA/D. In order to assess whether the differences are statistically significants, we have also applied the Friedman's ranking and Holm's post-hoc multicompare tests. In this regard, we are able to assess which algorithms are statistically worse than the control one (i.e. the one ranking the best) and evaluate the results by applying a *p*-value of 0.05. Table 8 verifies that SMPSO is the best ranked algorithm according to Friedman's test for the two indicators, followed by MPSO/D. Therefore, SMPSO was used as control algorithm in the post-hoc Holm tests. In terms of I_{HV}, the adjusted *p*-values show that SMPSO is statistically better than MPSO/D and MOEA/D, given that the $Holm_{Ap}$ values are lower than 0.05 for both quality indicators.

Finally, we provide a visual hint of the solutions obtained in this benchmark in Fig. 2, which represents the intermolecular interactions between the ligand and the macromolecule. Image A in this figure shows a macrocyclic peptidomimetic inhibitor (represented with sticks) bound to the flexible HIV-protease (the flexible side-chains of the macromolecule are represented with sticks) (macromolecule). The values of E_{inter}, E_{intra} and RMSD are -19.81 kcal/mol, 0.47 kcal/mol and 0.51 Å, respectively. We have selected this solution because the ligand–macromolecule interaction is energetically stable and the ligand's deformity in terms of energy is less stable. The RMSD score corresponds to 0.51 Å, which is lower than 2 Å. This result elucidates that the computed ligand's position is a good prediction compared to the co-crystallized ligand. Furthermore, this solution can be selected by those practitioners interested in a lower ligand's deformity and a more stable ligand–receptor interaction. Image B shows a more detailed ligand–macromolecule interaction. The predicted H-bond is represented with green spheres. This corresponds to a non-covalent bond between the -NH of the amide bond of ASP29 of the macromolecule's active site and the atom O2 of the ligand.

In summary, this empirical study illustrates the potential of considering three objectives in the molecular docking problem, as well as the suitability of adopting multi-objective meta-heuristics towards producing a richer portfolio of possible ligand conformations.

6. Discussion and challenges

A first aspect for discussion in this survey is driven by the categorization of the energy scoring function in four main families according to [91]: (1) the force-field-based energy function; (2) empirical or regression-based energy functions; (3) knowledge-based energy scoring functions; and (4) machine-learning-based functions. As a result of our bibliographic analysis, we have observed that the force field-based energy function is the most used alternative to evaluate ligand–macromolecule binding solutions, with the one provided by AutoDock prevailing in the literature as almost a *de facto* choice. Furthermore, knowledge-based and descriptor-based functions are notably less utilized in the field. This can be useful as an overview for computer scientists to apply meta-heuristics using new evaluation methods such as the study provided by [87], where the authors applied a new QSAR-based-energy function and a search method based on Genetic Programming (GP).

When it comes to the algorithmic design of the search method, our survey has identified several meta-heuristic solvers being used in the literature, including the Monte Carlo algorithm, the family of Evolutionary Computation (GA, DE and others alike), Tabu Search, Simulated Annealing, Local Search methods, and Swarm Intelligence methods (PSO, ACO and ABC, among others). We have observed that Evolutionary Computation has been at the forefront of meta-heuristics applied to these problems, with a prominence of GA-based methods. By contrast, Swarm Intelligence (specially, PSO) is being explored with promising results in recent contributions.

Molecular docking can be also formulated as a single- or multi-objective optimization problem. According to the state of the art,

Table 6
Median and Inter-Quartile Range (IQR) of I_{HV}.

Category	Instance	MOEA/D	SMPSO	MPSO/D
Small-size	1A9M	$9.67e-01_{3.2e-02}$	$9.99e-01_{5.6e-04}$	$9.97e-01_{4.9e-03}$
	1AAQ	$9.95e-01_{2.1e-02}$	$9.92e-01_{1.8e-03}$	$9.95e-01_{2.0e-03}$
	1B6L	$9.76e-01_{7.4e-02}$	$9.88e-01_{3.6e-02}$	$9.85e-01_{4.4e-03}$
	1B6M	$6.51e-01_{3.0e-01}$	$9.56e-01_{1.9e-02}$	$9.06e-01_{7.2e-02}$
	1BDL	$8.01e-01_{2.2e-03}$	$8.59e-01_{3.8e-02}$	$7.79e-01_{2.8e-02}$
	1BDQ	$7.68e-01_{1.4e-01}$	$7.21e-01_{1.6e-02}$	$7.57e-01_{4.7e-02}$
	1BDR	$9.82e-01_{2.0e-01}$	$9.51e-01_{3.8e-03}$	$9.63e-01_{2.5e-02}$
	1GNM	$9.34e-01_{7.3e-02}$	$9.99e-01_{7.1e-04}$	$9.95e-01_{1.5e-03}$
	1GNN	$7.60e-01_{2.7e-01}$	$9.99e-01_{1.8e-03}$	$9.91e-01_{3.6e-03}$
	1GNO	$0.00e+00_{7.8e-01}$	$9.54e-01_{9.2e-03}$	$9.18e-01_{2.3e-02}$
	1HBV	$6.49e-01_{5.7e-01}$	$8.16e-01_{2.7e-02}$	$8.46e-01_{2.8e-02}$
	1HEG	$8.11e-01_{2.1e-02}$	$8.95e-01_{1.5e-02}$	$7.63e-01_{3.9e-02}$
	1HIH	$9.10e-01_{4.2e-02}$	$9.41e-01_{6.0e-03}$	$9.44e-01_{2.3e-02}$
	1HPV	$9.96e-01_{2.6e-02}$	$9.93e-01_{7.1e-04}$	$9.94e-01_{3.1e-03}$
	1HSG	$9.29e-01_{4.2e-02}$	$9.94e-01_{2.8e-03}$	$9.93e-01_{8.2e-03}$
	1HTE	$9.02e-01_{7.3e-02}$	$9.35e-01_{1.0e-02}$	$9.32e-01_{4.0e-03}$
	1KZK	$8.05e-01_{5.0e-02}$	$6.92e-01_{2.9e-02}$	$8.69e-01_{5.4e-02}$
	1SGB	$8.56e-01_{5.0e-02}$	$8.65e-01_{2.1e-02}$	$8.48e-01_{1.4e-02}$
	1TCX	$9.67e-01_{2.4e-03}$	$9.52e-01_{1.2e-02}$	$9.63e-01_{8.9e-03}$
	1ZIH	$9.37e-01_{8.1e-02}$	$9.88e-01_{6.5e-04}$	$9.87e-01_{1.1e-02}$
	1ZIR	$9.46e-01_{7.9e-02}$	$9.88e-01_{1.7e-03}$	$9.90e-01_{7.5e-03}$
	3AID	$9.90e-01_{8.9e-02}$	$9.97e-01_{1.0e-03}$	$9.87e-01_{5.5e-03}$
Medium-size	1B6J	$9.32e-01_{1.9e-01}$	$9.25e-01_{8.6e-03}$	$9.34e-01_{3.3e-02}$
	1B6K	$9.53e-01_{7.8e-02}$	$9.88e-01_{3.1e-03}$	$9.85e-01_{1.6e-02}$
	1B6P	$8.77e-01_{9.0e-02}$	$9.85e-01_{2.5e-03}$	$9.77e-01_{2.5e-02}$
	1D4K	$7.29e-01_{6.9e-02}$	$9.85e-01_{7.8e-03}$	$9.54e-01_{1.8e-02}$
	1D4L	$8.55e-01_{3.5e-02}$	$9.54e-01_{7.4e-03}$	$9.40e-01_{6.7e-02}$
	1HEF	$7.53e-01_{2.9e-02}$	$8.59e-01_{3.8e-02}$	$6.85e-01_{2.3e-02}$
	1HPS	$8.05e-01_{7.7e-01}$	$9.95e-01_{2.0e-04}$	$9.23e-01_{2.0e-01}$
	1HXW	$9.81e-01_{2.7e-02}$	$9.82e-01_{3.1e-03}$	$9.85e-01_{1.2e-02}$
	1IZH	$9.61e-01_{2.8e-02}$	$7.79e-01_{7.5e-04}$	$9.74e-01_{1.1e-02}$
	1IZI	$8.44e-01_{2.7e-01}$	$8.91e-01_{2.0e-02}$	$8.68e-01_{1.4e-02}$
	1JLD	$4.48e-01_{6.3e-02}$	$4.80e-01_{1.1e-01}$	$4.29e-01_{2.5e-02}$
	1K6C	$8.07e-01_{1.2e-01}$	$9.71e-01_{5.6e-03}$	$9.66e-01_{1.8e-02}$
	1K6P	$8.84e-01_{9.2e-02}$	$9.79e-01_{4.1e-03}$	$9.71e-01_{3.6e-02}$
	1K6T	$6.57e-01_{2.8e-01}$	$9.77e-01_{1.1e-02}$	$9.27e-01_{1.5e-01}$
	1K6V	$7.78e-01_{1.6e-01}$	$9.54e-01_{6.0e-03}$	$9.38e-01_{4.0e-02}$
	1MTR	$6.83e-01_{3.2e-01}$	$9.02e-01_{1.3e-02}$	$8.04e-01_{5.3e-01}$
	1MUI	$8.57e-01_{5.3e-01}$	$9.22e-01_{2.7e-03}$	$9.47e-01_{2.6e-02}$
	2BPV	$4.64e-01_{8.5e-02}$	$5.78e-01_{4.2e-02}$	$3.73e-01_{6.9e-02}$
	2BPX	$8.66e-01_{1.5e-01}$	$9.82e-01_{3.5e-03}$	$9.79e-01_{1.4e-02}$
	4HVP	$9.45e-01_{3.3e-01}$	$9.42e-01_{5.8e-03}$	$9.31e-01_{7.0e-03}$
	4PHV	$9.67e-01_{4.1e-02}$	$9.92e-01_{3.2e-04}$	$9.97e-01_{2.0e-03}$
	5HVP	$6.96e-01_{2.7e-01}$	$9.83e-01_{1.2e-02}$	$9.31e-01_{3.9e-02}$
Large-size	1A94	$6.79e-01_{2.7e-01}$	$7.54e-01_{1.9e-02}$	$7.15e-01_{1.1e-01}$
	1HIV	$9.39e-01_{1.2e-01}$	$9.75e-01_{7.1e-03}$	$9.57e-01_{3.2e-02}$
	1HOS	$8.40e-01_{1.3e-02}$	$8.99e-01_{1.4e-02}$	$8.02e-01_{1.1e-02}$
	1HTG	$1.00e-01_{8.4e-01}$	$9.42e-01_{1.2e-02}$	$9.52e-01_{1.1e-01}$
	1HVI	$9.70e-01_{4.1e-02}$	$9.78e-01_{2.3e-03}$	$9.84e-01_{3.7e-03}$
	1HVJ	$9.96e-01_{1.7e-02}$	$9.37e-01_{7.9e-03}$	$9.31e-01_{1.5e-02}$
	1HVK	$7.84e-01_{1.8e-02}$	$6.88e-01_{2.1e-02}$	$7.66e-01_{1.1e-02}$
	1HVL	$7.80e-01_{4.5e-02}$	$6.95e-01_{3.5e-02}$	$7.41e-01_{3.6e-02}$
	1HVS	$6.17e-01_{4.1e-02}$	$4.88e-01_{6.3e-02}$	$5.71e-01_{3.6e-02}$
	1HWR	$4.74e-01_{1.7e-01}$	$4.26e-01_{3.7e-02}$	$5.04e-01_{7.6e-03}$
	1ODY	$6.22e-01_{4.6e-01}$	$8.69e-01_{2.6e-02}$	$7.84e-01_{3.8e-02}$
	1VIJ	$5.22e-01_{5.5e-02}$	$7.03e-01_{5.4e-02}$	$4.18e-01_{5.9e-02}$
	1VIK	$4.86e-01_{5.0e-02}$	$5.25e-01_{6.7e-02}$	$4.67e-01_{5.9e-02}$
	3TLH	$8.59e-01_{3.5e-02}$	$8.99e-01_{1.2e-02}$	$8.27e-01_{2.2e-02}$
	7HVP	$9.82e-01_{1.7e-02}$	$9.68e-01_{3.8e-03}$	$9.82e-01_{7.4e-03}$
	8HVP	$3.17e-01_{5.2e-02}$	$4.61e-01_{1.9e-02}$	$2.47e-01_{3.3e-02}$
	9HPV	$7.91e-01_{1.6e-01}$	$7.80e-01_{9.3e-03}$	$7.63e-01_{4.9e-02}$
Cyclic-urea	1BV7	$8.07e-01_{8.3e-01}$	$7.15e-01_{2.4e-02}$	$7.48e-01_{4.7e-02}$
	1BV9	$9.55e-01_{8.6e-01}$	$9.29e-01_{5.2e-03}$	$9.51e-01_{1.9e-02}$
	1BWA	$4.35e-01_{8.9e-01}$	$9.25e-01_{5.4e-03}$	$9.49e-01_{3.8e-02}$
	1BWB	$8.80e-01_{1.9e-01}$	$9.72e-01_{2.1e-02}$	$9.77e-01_{6.1e-03}$
	1DMP	$3.07e-01_{7.2e-01}$	$5.27e-01_{1.3e-01}$	$6.88e-01_{3.2e-02}$
	1G35	$8.54e-01_{9.6e-02}$	$9.05e-01_{3.9e-03}$	$9.08e-01_{1.6e-02}$
	1HPO	$9.16e-01_{4.8e-02}$	$9.84e-01_{2.7e-03}$	$9.78e-01_{2.9e-02}$
	1MES	$5.19e-01_{1.7e-01}$	$3.67e-01_{7.4e-02}$	$4.94e-01_{4.9e-02}$
	1MEU	$9.93e-01_{4.0e-02}$	$9.93e-01_{4.2e-04}$	$9.96e-01_{1.9e-03}$
	1PRO	$9.08e-01_{9.0e-01}$	$7.38e-01_{6.3e-02}$	$8.85e-01_{1.5e-02}$
	1QBR	$9.91e-01_{1.2e-01}$	$9.84e-01_{1.2e-03}$	$9.88e-01_{6.6e-03}$
	1QBT	$0.00e+00_{8.9e-01}$	$7.92e-01_{1.5e-02}$	$8.34e-01_{9.3e-02}$
	1QBU	$6.04e-01_{4.0e-01}$	$5.06e-01_{2.8e-02}$	$5.58e-01_{7.2e-02}$
	7UPJ	$9.64e-01_{3.8e-02}$	$9.88e-01_{3.9e-03}$	$9.88e-01_{3.4e-03}$

if molecular docking is approached as a single-objective optimization problem most cases embrace the final binding energy ΔG as the function to be minimized. In the multi-objective case, however, the reported experiments are more diverse in terms of the selected objectives, with assorted combinations of E_{inter}, E_{intra}, $\Delta_{elect_s olv}$, E_{L-R}, GS, fitness scores provided by three different scoring functions (force-field-based, empirical-based and knowledge-based) and RMSD. Either way, E_{inter} and E_{intra} comprise the combination of conflicting objectives that most authors have considered so far when modeling molecular docking as a multi-objective optimization problem.

Moreover, an insightful discussion has been held around the experiments and results obtained in a molecular docking problem consisting of three objectives to be minimized: E_{inter}, E_{intra} and the RMSD score. To the best of our knowledge, the present contribution corresponds to one of the only few published works that

Table 7
Median and Inter-Quartile Range (IQR) of $I_{\epsilon+}$.

Category	Instance	MOEA/D	SMPSO	MPSO/D
Small-size	1A9M	$3.32e-02_{3.2e-02}$	$6.40e-04_{5.6e-04}$	$2.83e-03_{4.9e-03}$
	1AAQ	$5.46e-03_{2.1e-02}$	$7.76e-03_{1.8e-03}$	$5.33e-03_{2.0e-03}$
	1B6L	$1.89e-02_{6.4e-02}$	$7.38e-03_{1.4e-03}$	$8.52e-03_{3.7e-03}$
	1B6M	$3.35e-01_{2.8e-01}$	$2.80e-02_{1.9e-02}$	$7.63e-02_{7.1e-02}$
	1BDL	$1.94e-01_{3.0e-03}$	$8.57e-02_{5.1e-02}$	$2.02e-01_{3.3e-02}$
	1BDQ	$2.15e-01_{1.9e-02}$	$2.26e-01_{3.0e-02}$	$2.17e-01_{2.1e-02}$
	1BDR	$1.77e-02_{2.0e-01}$	$4.91e-02_{3.8e-03}$	$3.74e-02_{2.5e-02}$
	1GNM	$6.60e-02_{7.3e-02}$	$6.48e-04_{7.1e-04}$	$4.72e-03_{1.5e-03}$
	1GNN	$2.40e-01_{2.7e-01}$	$1.40e-03_{1.8e-03}$	$8.93e-03_{3.6e-03}$
	1GNO	$1.15e+00_{1.6e+00}$	$2.41e-02_{2.7e-03}$	$4.47e-02_{1.8e-02}$
	1HBV	$2.91e-01_{5.4e-01}$	$9.00e-02_{7.3e-03}$	$1.08e-01_{3.0e-02}$
	1HEG	$1.70e-01_{2.9e-02}$	$7.08e-02_{1.9e-02}$	$2.07e-01_{5.7e-02}$
	1HIH	$5.61e-02_{4.2e-02}$	$4.56e-02_{1.2e-02}$	$4.28e-02_{1.0e-02}$
	1HPV	$4.03e-03_{2.6e-02}$	$7.37e-03_{7.1e-04}$	$6.00e-03_{3.1e-03}$
	1HSG	$7.14e-02_{4.2e-02}$	$5.60e-03_{2.8e-03}$	$7.14e-03_{8.2e-03}$
	1HTE	$6.29e-02_{2.6e-02}$	$4.59e-02_{1.0e-02}$	$5.17e-02_{3.7e-03}$
	1KZK	$1.94e-01_{5.0e-02}$	$3.06e-01_{3.7e-02}$	$2.80e-01_{1.1e-01}$
	1SGB	$1.38e-01_{3.2e-02}$	$1.01e-01_{2.3e-02}$	$1.43e-01_{5.9e-03}$
	1TCX	$3.12e-01_{1.8e-03}$	$3.94e-02_{4.5e-03}$	$3.18e-02_{4.8e-03}$
	1ZIH	$6.32e-02_{8.1e-02}$	$1.23e-02_{6.5e-04}$	$1.25e-02_{1.1e-02}$
	1ZIR	$4.68e-02_{7.8e-02}$	$7.33e-03_{1.1e-03}$	$5.54e-03_{3.4e-03}$
	3AID	$6.76e-03_{8.6e-02}$	$3.12e-03_{9.1e-04}$	$1.03e-02_{4.9e-03}$
Medium-size	1B6J	$5.75e-02_{1.1e-01}$	$6.48e-02_{8.7e-03}$	$5.56e-02_{1.5e-02}$
	1B6K	$4.11e-02_{7.6e-02}$	$7.64e-03_{2.5e-03}$	$9.32e-03_{1.5e-02}$
	1B6P	$1.30e-01_{9.5e-02}$	$8.45e-03_{2.4e-03}$	$1.34e-02_{2.0e-02}$
	1D4K	$2.51e-01_{6.8e-02}$	$1.09e-02_{2.9e-03}$	$3.28e-01_{1.6e-02}$
	1D4L	$1.07e-01_{3.3e-02}$	$3.38e-02_{7.0e-03}$	$4.55e-02_{3.1e-02}$
	1HEF	$2.31e-01_{3.4e-02}$	$1.02e-01_{5.1e-02}$	$2.97e-01_{3.4e-02}$
	1HPS	$1.88e-01_{6.3e-01}$	$4.48e-03_{1.9e-04}$	$7.13e-02_{1.7e-01}$
	1HXW	$1.86e-02_{2.7e-02}$	$1.82e-02_{3.1e-03}$	$1.52e-02_{1.2e-02}$
	1IZH	$3.66e-02_{1.7e-02}$	$2.17e-01_{2.1e-05}$	$2.03e-02_{1.2e-02}$
	1IZI	$1.26e-01_{1.5e-01}$	$9.31e-02_{5.3e-02}$	$1.25e-01_{1.2e-02}$
	1JLD	$5.45e-01_{5.1e-02}$	$4.63e-01_{1.7e-01}$	$5.58e-01_{2.5e-02}$
	1K6C	$1.84e-01_{1.4e-01}$	$2.41e-02_{5.2e-03}$	$2.69e-02_{1.8e-02}$
	1K6P	$1.09e-01_{9.1e-02}$	$1.12e-02_{2.4e-03}$	$1.57e-02_{3.0e-02}$
	1K6T	$3.38e-01_{2.9e-01}$	$1.86e-02_{1.1e-02}$	$6.81e-02_{1.4e-01}$
	1K6V	$2.04e-01_{1.6e-01}$	$2.47e-02_{4.4e-03}$	$3.29e-02_{2.8e-02}$
	1MTR	$2.56e-01_{2.9e-01}$	$5.50e-02_{3.2e-03}$	$1.25e-01_{4.8e-01}$
	1MUI	$1.30e-01_{5.0e-01}$	$6.47e-02_{2.2e-03}$	$4.13e-02_{2.5e-02}$
	2BPV	$4.07e-01_{1.5e-01}$	$1.89e-01_{5.5e-02}$	$4.29e-01_{5.7e-02}$
	2BPX	$1.29e-01_{1.5e-01}$	$1.44e-02_{3.4e-03}$	$1.77e-02_{1.4e-02}$
	4HVP	$4.24e-02_{3.1e-01}$	$4.22e-02_{6.2e-03}$	$4.24e-02_{3.2e-03}$
	4PHV	$3.26e-02_{4.1e-02}$	$8.30e-03_{3.2e-04}$	$2.53e-03_{2.0e-03}$
	5HVP	$2.84e-01_{2.9e-01}$	$1.63e-02_{9.1e-04}$	$5.31e-02_{3.7e-02}$
Large-size	1A94	$2.07e-01_{7.0e-01}$	$1.05e-01_{2.3e-02}$	$1.29e-01_{8.9e-02}$
	1HIV	$6.12e-02_{1.2e-01}$	$2.45e-02_{7.1e-03}$	$4.26e-02_{3.2e-02}$
	1HOS	$1.44e-01_{1.5e-02}$	$6.55e-02_{1.4e-02}$	$1.57e-01_{1.8e-02}$
	1HTG	$8.71e-01_{1.2e+00}$	$5.58e-02_{1.2e-02}$	$4.62e-01_{1.1e-01}$
	1HVI	$3.03e-02_{4.1e-02}$	$2.19e-02_{2.3e-03}$	$1.59e-02_{3.7e-03}$
	1HVJ	$4.20e-03_{1.7e-02}$	$6.27e-02_{7.8e-03}$	$6.89e-02_{1.5e-02}$
	1HVK	$1.89e-01_{3.2e-02}$	$2.92e-01_{2.3e-02}$	$2.11e-01_{2.0e-02}$
	1HVL	$1.95e-01_{4.2e-02}$	$2.83e-01_{3.6e-02}$	$2.39e-01_{3.4e-02}$
	1HVS	$3.62e-01_{3.8e-02}$	$4.95e-01_{6.9e-02}$	$4.11e-01_{3.9e-02}$
	1HWR	$5.11e-01_{9.3e-02}$	$4.02e-01_{2.8e-01}$	$4.70e-01_{9.0e-03}$
	1ODY	$2.62e-01_{3.0e-01}$	$8.63e-02_{1.6e-02}$	$2.08e-01_{2.7e-02}$
	1VIJ	$4.63e-01_{5.9e-02}$	$1.76e-01_{9.2e-02}$	$5.74e-01_{6.2e-02}$
	1VIK	$5.01e-01_{4.4e-01}$	$3.80e-01_{1.4e-01}$	$5.11e-01_{6.0e-02}$
	3TLH	$1.15e-01_{3.2e-02}$	$7.56e-02_{1.6e-02}$	$1.53e-01_{1.8e-02}$
	7HVP	$1.12e-02_{1.5e-02}$	$1.91e-02_{1.5e-03}$	$1.45e-02_{4.4e-03}$
	8HVP	$4.64e-01_{1.3e-01}$	$1.62e-01_{5.5e-02}$	$3.94e-01_{1.1e-01}$
	9HVP	$1.72e-01_{1.3e-01}$	$9.30e-02_{1.7e-02}$	$1.97e-01_{4.3e-02}$
Cyclic-urea	1BV7	$1.71e-01_{9.8e-01}$	$1.90e-01_{2.4e-02}$	$1.76e-01_{2.8e-02}$
	1BV9	$4.08e-02_{8.6e-01}$	$6.71e-02_{5.0e-03}$	$4.45e-02_{1.9e-02}$
	1BWA	$5.56e-01_{8.6e-01}$	$6.93e-02_{5.4e-03}$	$4.42e-02_{3.9e-02}$
	1BWB	$1.04e-01_{1.5e-01}$	$1.71e-02_{1.8e-03}$	$1.64e-02_{1.3e-03}$
	1DMP	$4.87e-01_{1.0e+00}$	$3.70e-01_{1.3e-01}$	$2.48e-01_{4.0e-02}$
	1G35	$8.64e-02_{3.2e-02}$	$8.29e-02_{6.3e-03}$	$7.57e-02_{5.2e-03}$
	1HPO	$7.24e-02_{4.7e-02}$	$9.79e-03_{1.4e-02}$	$1.37e-02_{2.4e-02}$
	1MES	$4.47e-01_{1.9e-01}$	$5.83e-01_{9.8e-02}$	$4.58e-01_{3.8e-02}$
	1MEU	$6.51e-03_{4.0e-02}$	$6.77e-03_{4.2e-04}$	$4.21e-03_{1.9e-03}$
	1PRO	$7.49e-02_{8.7e-01}$	$1.79e-01_{7.0e-02}$	$7.04e-02_{1.6e-02}$
	1QBR	$5.21e-03_{1.0e-01}$	$9.38e-03_{7.7e-04}$	$6.33e-03_{5.4e-03}$
	1QBT	$1.15e+00_{1.6e+00}$	$1.19e-01_{1.4e-02}$	$8.92e-02_{4.6e-02}$
	1QBU	$3.67e-01_{1.6e-01}$	$3.88e-01_{1.9e-01}$	$3.64e-01_{3.0e-02}$
	7UPJ	$2.98e-02_{3.3e-02}$	$7.07e-03_{1.1e-04}$	$7.52e-03_{3.3e-03}$

optimize more than two objectives [9,35]. The performed experimental study includes an extensive comparison of three state-of-the-art multi-objective meta-heuristic algorithms, from which we conclude that SMPSO performs best in terms of convergence and diversity, followed by MPSO/D. We have also provided a biological analysis based on the intermolecular H-bond interactions between ligand and macromolecule.

As a result of our experimental survey, we have identified several open challenges that are timidly considered in the specific literature or, in most cases, overlooked. An interesting aspect that has not been sufficiently elaborated to date is the generation of adapted initialization and recombination/mutation operators – in essence, how to compute the neighborhood of a given solution – so as to efficiently tackle molecular docking problems. Taking into account that common solution encoding strategies resort to
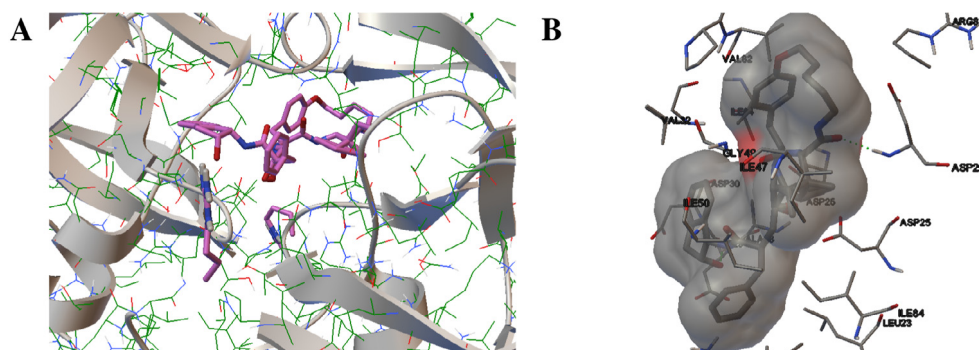
Fig. 2. Image A depicts the ligand's conformation in the HIV-protease active site. The ligand and the side-chains of the aminoacids are represented with sticks. Image B shows a more detailed representation of the ligand–macromolecule interactions. H-bonds are represented with green spheres.

Table 8
Average Friedman's rankings with Holm's Adjusted p-values (0.05) of compared algorithms for the test set of 75 docking instances. * indicates the control algorithm.

Hypervolume (I_{HV})			Epsilon ($I_{\epsilon+}$)		
Algorithm	Fri_{Rank}	$Holm_{Ap}$	Algorithm	Fri_{Rank}	$Holm_{Ap}$
*SMPSO	1.75	–	*SMPSO	1.68	–
MPSO/D	1.94	2.36e−01	MPSO/D	1.94	1.11e−01
MOEA/D	2.31	1.04e−03	MOEA/D	2.38	3.63e−05

well-delimited sets of variables, i.e. for translation, quaternion and torsions, with specific ranges of variables, it should be possible to generate specific search operators that take advantage of the inter/intra-dependency among such variables. In this sense, based on the geometrical restrictions observed in rotations they have not complete degree of freedom. Along with this, ranges of peptide conformations are limited according to Ramachandran's plot [122]. In light of these variable interactions and geometrical restrictions we envision a research niche focused on designing ad-hoc search operators that efficiently explore the landscape of the formulated molecular docking problem, thereby favoring unprecedented levels of search efficiency.

The review of the selected works has also revealed that, in general, algorithms are configured by using parameter settings utilized in similar studies. Furthermore, it is also widely noted that key parameter values are not explicitly indicated, thus making it very difficult to reproduce the reported results. These issues are often found in interdisciplinary works where experts in the application domain are not aware of the good practices that should be followed when comparing optimization algorithms. Consequently, the application of techniques for automatic parameter tuning [123] would help biologists when searching for the best algorithmic configuration for their practical docking problems. What is more, instead of configuring a particular technique according to some *application-agnostic* criterion (e.g. to take a popular optimization algorithm), a step further along this research path is the automatic design of metaheuristics [124] specifically tailored for molecular docking.

Finally, we have observed that most reported benchmarks among different bio-inspired metaheuristics do not comply with recommended practices in terms of methodology and statistical assessment of the gaps claimed to exist among the considered solvers. This is particularly concerning, as conclusions derived from such studies can be severely biased by an unadvertised lack of fairness in terms of parametric tuning and/or the statistical significance of detected performance differences. To prevent the community from misleading conclusions in this regard, we definitely advocate for a wide adoption of the methodological recommendations for analyzing benchmarks of metaheuristic algorithms [125], as well as the use of advanced Bayesian tests and

graphical tools for ranking and pairwise comparison among such techniques. Optimization frameworks contributed very recently to the community move indeed in this direction [126,127].

7. Concluding remarks

Molecular docking can be modeled as a complex optimization problem based on predicting the interaction between a ligand to a macromolecule, in terms of energy and RMSD. This modeling approach has been at the core of research in Bioinformatics, mainly aimed at discovering new compounds and therapeutic targets within the pharmaceutical industry. This paper has reviewed the state of the art related to the intersection between this problem and bio-inspired meta-heuristics. To this end the molecular docking problem relies essentially in two different design aspects: (1) the energy scoring function to be optimized and (2) the search method designed to efficiently solve the underlying optimization problem.

As evinced in our work, the extensive literature published in this topic has hitherto blossomed several problem flavors and algorithmic variants. This profitable background should suggest molecular biologists to embrace and apply these techniques to solve their own *in silico* molecular docking studies. However, there is still room for further improvement in what refers to several research avenues identified in this work. Among them, we have emphasized on hybrid meta-heuristics blending together modern bio-inspired solvers and local search methods specialized to deal with the complex geometry of docking problems. We foresee that this noted lack of hybrid approaches will mobilize future research efforts within the field of Bio-inspired Optimization, ultimately accelerating the emergence of radically new heuristic solvers adapted to the complex specificities of this family of problems. We have also stressed on the need for more principled comparison studies and statistically informed analysis when comparing different solvers for molecular docking problems. Unless robust comparison methodologies already used in the field of optimization heuristics are widely adopted by the community, a halo of doubt will remain over the claimed outperforming behavior of new, adapted or refurnished search algorithms for docking problems.

References

[1] C.S. Greene, J. Tan, M. Ung, J.H. Moore, C. Cheng, Big data bioinformatics, J. Cell. Physiol. 229 (12) (2014) 1896–1900, http://dx.doi.org/10.1002/jcp.24662.

[2] J. Handl, D.B. Kell, J. Knowles, Multiobjective optimization in bioinformatics and computational biology, IEEE/ACM Trans. Comput. Biol. Bioinform. 4 (2) (2007) 279–292, http://dx.doi.org/10.1109/TCBB.2007.070203.

[3] C. Blum, A. Roli, Metaheuristics in combinatorial optimization: overview and conceptual comparison, ACM Comput. Surv. 35 (3) (2003) 268–308, http://dx.doi.org/10.1145/937503.937505.

[4] T.H.S. Consortium, The sequence of the human genome, Science 291 (5507) (2001) 1304–1351.

[5] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.

[6] D.A. Gschwend, A.C. Good, I.D. Kuntz, Molecular docking towards drug discovery, J. Mol. Recognit. 9 (2) (1996) 175–186, http://dx.doi.org/10.1002/(SICI)1099-1352(199603)9:2<175::AID-JMR260>3.0.CO;2-D.

[7] B.K. Shoichet, S.L. McGovern, B. Wei, J.J. Irwin, Lead discovery using molecular docking, Curr. Opin. Chem. Biol. 6 (4) (2002) 439–446, http://dx.doi.org/10.1016/S1367-5931(02)00339-3.

[8] R.N. dos Santos, L.G. Ferreira, A.D. Andricopulo, Practices in molecular docking and structure-based virtual screening, in: M. Gore, U.B. Jagtap (Eds.), Computational Drug Discovery and Design, Springer New York, New York, NY, 2018, pp. 31–50, http://dx.doi.org/10.1007/978-1-4939-7756-7_3.

[9] J. Liu, R. Wang, Classification of current scoring functions, J. Chem. Inform. Model. 55 (3) (2015) 475–482, http://dx.doi.org/10.1021/ci500731a.

[10] J. Pérez-Serrano, B. Imbernón, J. Cecilia, M. Ujaldón, Energy-based tuning of metaheuristics for molecular docking on multi-GPUs, Concurr. Comput.: Pract. Exper. 30 (17) (2018) e4684, http://dx.doi.org/10.1002/cpe.4684.

[11] X.-S. Yang, Nature-Inspired Metaheuristic Algorithms, Luniver Press, 2008.

[12] J. García-Nieto, E. López-Camacho, M.J. García-Godoy, A.J. Nebro, J.F. Aldana-Montes, Multi-objective ligand-protein docking with particle swarm optimizers, Swarm Evol. Comput. (2018) http://dx.doi.org/10.1016/j.swevo.2018.05.007.

[13] E.K. Nesamalar, C.P. Chandran, Genetic clustering with bee colony optimization for flexible protein-ligand docking, in: International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), IEEE, 2012, pp. 82–87, http://dx.doi.org/10.1109/icprime.2012.6208291.

[14] X.-S. Yang, S. Deb, Multiobjective cuckoo search for design optimization, Comput. Oper. Res. 40 (6) (2013) 1616–1624, http://dx.doi.org/10.1016/j.cor.2011.09.026, http://www.sciencedirect.com/science/article/pii/S0305054811002905, Emergent Nature Inspired Algorithms for Multi-Objective Optimization.

[15] I. Fister, I. Fister, X.-S. Yang, J. Brest, A comprehensive review of firefly algorithms, Swarm Evol. Comput. 13 (2013) 34–46, http://dx.doi.org/10.1016/j.swevo.2013.06.001, http://www.sciencedirect.com/science/article/pii/S2210650213000461.

[16] J. Kennedy, Swarm intelligence, in: Handbook of Nature-Inspired and Innovative Computing, Springer, 2006, pp. 187–219.

[17] M. Karplus, D.L. Weaver, Protein-folding dynamics, Nature 260 (5550) (1976) 404–406, http://dx.doi.org/10.1038/260404a0.

[18] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (2) (1983) 187–217, http://dx.doi.org/10.1002/jcc.540040211.

[19] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (14) (1998) 1639–1662, http://dx.doi.org/10.1002/(sici)1096-987x(19981115)19:14<1639::aid-jcc10>3.0.co;2-b.

[20] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule-ligand interactions, J. Mol. Biol. 161 (2) (1982) 269–288, http://dx.doi.org/10.1016/0022-2836(82)90153-x.

[21] A.R. Ortiz, M.T. Pisabarro, F. Gago, R.C. Wade, Prediction of drug binding affinities by comparative binding energy analysis, J. Med. Chem. 38 (14) (1995) 2681–2691, http://dx.doi.org/10.1021/jm00014a020.

[22] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (3) (1997) 727–748, http://dx.doi.org/10.1006/jmbi.1996.0897.

[23] S. Yin, L. Biedermannova, J. Vondrasek, N.V. Dokholyan, Medusascore: an accurate force field-based scoring function for virtual drug screening, J. Chem. Inform. Model. 48 (8) (2008) 1656–1662, http://dx.doi.org/10.1021/ci8001167.

[24] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: THe biomolecular simulation program, J. Comput. Chem. 30 (10) (2009) 1545–1614, http://dx.doi.org/10.1002/jcc.21287.

[25] C.W. Murray, T.R. Auton, M.D. Eldridge, Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model, J. Comput. Aided Mol. Des. 12 (5) (1998) 503–519, http://dx.doi.org/10.1023/a:1008040323669.

[26] R. Wang, L. Lai, S. Wang, Further development and validation of empirical scoring functions for structure-based binding affinity prediction, J. Comput. Aided Mol. Des. 16 (1) (2002) 11–26, http://dx.doi.org/10.1023/a:1016357811882.

[27] R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, D.T. Mainz, Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes, J. Med. Chem. 49 (21) (2006) 6177–6196, http://dx.doi.org/10.1021/jm051256o.

[28] C.A. Sotriffer, H. Gohlke, G. Klebe, Docking into knowledge-based potential fields: a comparative evaluation of DrugScore, J. Med. Chem. 45 (10) (2002) 1967–1970, http://dx.doi.org/10.1021/jm025507u.

[29] S.-Y. Huang, X. Zou, An iterative knowledge-based scoring function for protein-protein recognition, Proteins: Struct. Function Bioinform. 72 (2) (2008) 557–579, http://dx.doi.org/10.1002/prot.21949.

[30] A. Grosdidier, V. Zoete, O. Michielin, Eadock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization, Proteins: Struct. Function Bioinform. 67 (4) (2007) 1010–1025, http://dx.doi.org/10.1002/prot.21367.

[31] S. Janson, D. Merkle, M. Middendorf, Molecular docking with multi-objective particle swarm optimization, Appl. Soft Comput. 8 (1) (2008) 666–675, http://dx.doi.org/10.1016/j.asoc.2007.05.005.

[32] Y. Liu, W. Li, Y. Wang, M. Lv, An efficient approach for flexible docking base on particle swarm optimization, in: 2009 2nd International Conference on Biomedical Engineering and Informatics, IEEE, 2009, pp. 1–7, http://dx.doi.org/10.1109/bmei.2009.5302670.

[33] J.-C. Boisson, L. Jourdan, E.-G. Talbi, D. Horvath, Single- and multi-objective cooperation for the flexible docking problem, J. Math. Modelling Alg. 9 (2) (2010) 195–208, http://dx.doi.org/10.1007/s10852-010-9128-y.

[34] A. Sandoval-Perez, D. Becerra, D. Vanegas, D. Restrepo-Montoya, F. Nino, A multi-objective optimization energy approach to predict the ligand conformation in a docking process, in: Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 181–192, http://dx.doi.org/10.1007/978-3-642-37207-0_16.

[35] J. Gu, X. Yang, L. Kang, J. Wu, X. Wang, MoDock: A Multi-objective strategy improves the accuracy for molecular docking, Alg. Mol. Biol. 10 (1) (2015) 8, http://dx.doi.org/10.1186/s13015-015-0034-8.

[36] M. García-Godoy, E. López-Camacho, J. García-Nieto, A. Nebro, J. Aldana-Montes, Solving molecular docking problems with multi-objective metaheuristics, Molecules 20 (6) (2015) 10154–10183, http://dx.doi.org/10.3390/molecules200610154.

[37] E. López-Camacho, M.J. García-Godoy, J. García-Nieto, A.J. Nebro, J.F. Aldana-Montes, A new multi-objective approach for molecular docking based on RMSD and binding energy, in: Algorithms for Computational Biology, Springer International Publishing, 2016, pp. 65–77, http://dx.doi.org/10.1007/978-3-319-38827-4_6.

[38] E.L. Camacho, M.J. García-Godoy, J.D. Ser, A.J. Nebro, J.F. Aldana-Montes, Multi-objective metaheuristics for a flexible ligand-macromolecule docking problem in computational biology, in: Intelligent Distributed Computing XII, Springer International Publishing, 2018, pp. 369–379, http://dx.doi.org/10.1007/978-3-319-99626-4_32.

[39] C.M. Oshiro, I.D. Kuntz, J.S. Dixon, Flexible ligand docking using a genetic algorithm, J. Comput. Aided Mol. Des. 9 (2) (1995) 113–130, http://dx.doi.org/10.1007/bf00124402.

[40] P. Willett, Genetic algorithms in molecular recognition and design, Trends Biotechnol. 13 (12) (1995) 516–521, http://dx.doi.org/10.1016/s0167-7799(00)89015-0.

[41] G.M. Verkhivker, P.A. Rejto, D.K. Gehlhaar, S.T. Freer, Exploring the energy landscapes of molecular recognition by a genetic algorithm: Analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes, Proteins: Struct. Function Bioinform. 25 (3) (1996) 342–353, http://dx.doi.org/{10.1002/(SICI)1097-0134(199607)25:3<342::AID-PROT6>3.0.CO;2-H}.

[42] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.R. Fogel, L.J. Fogel, S.T. Freer, Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming, Chem. Biol. 2 (5) (1995) 317–324, http://dx.doi.org/10.1016/1074-5521(95)90050-0.

[43] D.E. Clark, D.R. Westhead, Evolutionary algorithms in computer-aided molecular design, J. Comput. Aided Mol. Des. 10 (4) (1996) 337–358, http://dx.doi.org/10.1007/bf00124503.

[44] D. Levine, M. Facello, P. Hallstrom, G. Reeder, B. Walenz, F. Stevens, Stalk: an interactive system for virtual molecular docking, IEEE Comput. Sci. Eng. 4 (2) (1997) 55–65, http://dx.doi.org/10.1109/99.609834.

[45] M. Liu, S. Wang, MCDOCK: A Monte CArlo simulation approach to the molecular docking problem, J. Comput. Aided Mol. Des. 13 (5) (1999) 435–451, http://dx.doi.org/10.1023/a:1008005918983.

[46] R.J. Read, T.N. Hart, M.D. Cummings, S.R. Ness, Monte carlo algorithms for docking to proteins, Supramol. Chem. 6 (1–2) (1995) 135–140, http://dx.doi.org/10.1080/10610279508032529.

[47] D.R. Westhead, D.E. Clark, C.W. Murray, A comparison of heuristic search algorithms for molecular docking, J. Comput. Aided Mol. Des. 11 (3) (1997) 209–228, http://dx.doi.org/10.1023/a:1007934310264.

[48] J.-M. Yang, C.-Y. Kao, A family competition evolutionary algorithm for automated docking of flexible ligands to proteins, IEEE Trans. Inform. Technol. Biomed. 4 (3) (2000) 225–237, http://dx.doi.org/10.1109/4233.870033.

[49] J.S. Taylor, R.M. Burnett, Darwin: a program for docking flexible molecules, Proteins: Structure, Function, and Genetics 41 (2) (2000) 173–191, http://dx.doi.org/10.1002/1097-0134(20001101)41:2<173::aid-prot30>3.0.co;2-3.

[50] N. Budin, N. Majeux, A. Caflisch, Fragment-based flexible ligand docking by evolutionary optimization, Biol. Chem. 382 (9) (2001) http://dx.doi.org/10.1515/bc.2001.168.

[51] L. David, R. Luo, M.K. Gilson, Ligand-receptor docking with the mining minima optimizer, J. Comput. Aided Mol. Des. 15 (2) (2001) 157–171, http://dx.doi.org/10.1023/a:1008128723048.

[52] H. Li, C. Li, C. Gui, X. Luo, K. Chen, J. Shen, X. Wang, H. Jiang, GAsDock: a new approach for rapid flexible docking based on an improved multi-population genetic algorithm, Bioorg. Med. Chem. Lett. 14 (18) (2004) 4671–4676, http://dx.doi.org/10.1016/j.bmcl.2004.06.091.

[53] J.-M. Yang, C.-C. Chen, GEMDOCK: a generic evolutionary method for molecular docking, Proteins: Struct. Function Bioinform. 55 (2) (2004) 288–304, http://dx.doi.org/10.1002/prot.20035.

[54] Y. Fu, X. Wu, Z. Chen, J. Sun, J. Zhao, W. Xu, A new approach for flexible molecular docking based on swarm intelligence, Math. Probl. Eng. 2015 (2015) 1–10, http://dx.doi.org/10.1155/2015/540186.

[55] J. García-Nieto, E. López-Camacho, M.J.G. Godoy, A.J. Nebro, J.J. Durillo, J.F. Aldana-Montes, A study of archiving strategies in multi-objective PSO for molecular docking, in: Lecture Notes in Computer Science, Springer International Publishing, 2016, pp. 40–52, http://dx.doi.org/10.1007/978-3-319-44427-7_4.

[56] Y. Xiao, D. Williams, A comparison of GA and RSNR docking, in: Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence, IEEE, 1994, pp. 802–806, http://dx.doi.org/10.1109/icec.1994.349953.

[57] Y.L. Xiao, D.E. Williams, Molecular docking using genetic algorithms, in: Proceedings of the 1994 ACM Symposium on Applied Computing - SAC'94, ACM Press, 1994, pp. 196–200, http://dx.doi.org/10.1145/326619.326721.

[58] R. M.A.K.negtel, R. Boelens, R. Kaptein, Monte carlo docking of protein-DNA complexes: incorporation of DNA flexibility and experimental data, Protein Eng. Design Select. 7 (6) (1994) 761–768, http://dx.doi.org/10.1093/protein/7.6.761.

[59] K.P. Clark, Ajay, Flexible ligand docking without parameter adjustment across four ligand-receptor complexes, J. Comput. Chem. 16 (10) (1995) 1210–1226, http://dx.doi.org/10.1002/jcc.540161004.

[60] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, J. Mol. Biol. 245 (1) (1995) 43–53, http://dx.doi.org/10.1016/s0022-2836(95)80037-9.

[61] D. Maddalena, G. Snowdon, Applications of genetic algorithms to drug design, Expert Opin. Therapeutic Patents 7 (3) (1997) 247–254, http://dx.doi.org/10.1517/13543776.7.3.247.

[62] M. Vieth, J.D. Hirst, B.N. Dominy, H. Daigler, C.L. Brooks, Assessing search strategies for flexible docking, J. Comput. Chem. 19 (14) (1998) 1623–1631, http://dx.doi.org/10.1002/(sici)1096-987x(19981115)19:14<1623::aid-jcc8>3.0.co;2-l.

[63] J. Chen, H. Chi, Fast docking of drug molecules to their receptor, Chinese Sci. Bull. 44 (10) (1999) 904–908, http://dx.doi.org/10.1007/bf02885061.

[64] S.C.-H. Pegg, J.J. Haresco, I.D. Kuntz, A genetic algorithm for structure-based de novo design, J. Comput. Aided Mol. Des. 15 (10) (2001) 911–933, http://dx.doi.org/10.1023/a:1014389729000.

[65] T.-J. Hou, X.-J. Xu, L.-R. Chen, A comparison of three heuristic algorithms for molecular docking, in: Chinese Peptide Symposia, Kluwer Academic Publishers, 2002, pp. 65–68, http://dx.doi.org/10.1007/0-306-46880-8_15.

[66] M.R. McGann, H.R. Almond, A. Nicholls, J.A. Grant, F.K. Brown, Gaussian docking functions, Biopolymers 68 (1) (2003) 76–90, http://dx.doi.org/10.1002/bip.10207.

[67] K. Lee, C. Czaplewski, S.-Y. Kim, J. Lee, An efficient molecular docking using conformational space annealing, J. Comput. Chem. 26 (1) (2004) 78–87, http://dx.doi.org/10.1002/jcc.20147.

[68] C.S. de Magalhães, H.J.C. Barbosa, L.E. Dardenne, Selection-insertion schemes in genetic algorithms for the flexible ligand docking problem, in: Genetic and Evolutionary Computation – GECCO 2004, Springer Berlin Heidelberg, 2004, pp. 368–379, http://dx.doi.org/10.1007/978-3-540-24854-5_38.

[69] C. lian Li, Y. Sun, D. yun Long, X. cheng Wang, A genetic algorithm based method for molecular docking, in: Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2005, pp. 1159–1163, http://dx.doi.org/10.1007/11539117_156.

[70] S. Janson, D. Merkle, A new multi-objective particle swarm optimization algorithm using clustering applied to automated docking, in: Hybrid Metaheuristics, Springer Berlin Heidelberg, 2005, pp. 128–141, http://dx.doi.org/10.1007/11546245_12.

[71] C.-L. Li, Y. Sun, L. Zhang, X.-C. Wang, A parallel micro-genetic algorithm and its application, in: 2005 International Conference on Machine Learning and Cybernetics, vol. 5, IEEE, 2005, pp. 2880–2884, http://dx.doi.org/10.1109/icmlc.2005.1527434.

[72] C. lian Li, Y. sun, Y. shen Guo, F. ming Chu, Z. ru Guo, An entropy-based multi-population genetic algorithm and its application, in: Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2005, pp. 957–966, http://dx.doi.org/10.1007/11538059_99.

[73] D.T.-H. Chang, Y.-J. Oyang, J.-H. Lin, MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm, Nucleic Acids Res. 33 (Web Server) (2005) W233–W238, http://dx.doi.org/10.1093/nar/gki586.

[74] E.A. Wiley, M. MacDonald, A. Lambropoulos, D.J. Harriman, G. Deslongchamps, LGA-Dock/EM-Dock — Exploring Lamarckian Genetic algorithms and energy-based local search for ligand-receptor docking, Canad. J. Chem. 84 (3) (2006) 384–391, http://dx.doi.org/10.1139/v06-012.

[75] A. Oduguwa, A. Tiwari, S. Fiorentino, R. Roy, Multi-objective optimisation of the protein-ligand docking problem in drug discovery, in: Proceedings of the 8th annual conference on Genetic and evolutionary computation - GECCO'06, ACM Press, 2006, pp. 1793–1800, http://dx.doi.org/10.1145/1143997.1144287.

[76] H.-M. Chen, B.-F. Liu, H.-L. Huang, S.-F. Hwang, S.-Y. Ho, SODOCK: swarm optimization for highly flexible protein–ligand docking, J. Comput. Chem. 28 (2) (2006) 612–623, http://dx.doi.org/10.1002/jcc.20542.

[77] Y. Zhao, M.F. Sanner, FLipdock: docking flexible ligands into flexible receptors, Proteins: Struct. Function Bioinform. 68 (3) (2007) 726–737, http://dx.doi.org/10.1002/prot.21423.

[78] R.T. Kroemer, Structure-based drug design: docking and scoring, Current Protein Peptide Sci. 8 (4) (2007) 312–328, http://dx.doi.org/10.2174/138920307781369382.

[79] L. Kang, H. Li, H. Jiang, X. Wang, An improved adaptive genetic algorithm for protein–ligand docking, J. Comput. Aided Mol. Des. 23 (1) (2008) 1–12, http://dx.doi.org/10.1007/s10822-008-9232-5.

[80] W.-T. Sung, Employing improved GA to promote molecular docking efficiency for drug design, in: 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, IEEE, 2008, pp. 37–40, http://dx.doi.org/10.1109/icbbe.2008.16.

[81] E. Thiriot, G. Monard, Combining a genetic algorithm with a linear scaling semiempirical method for protein–ligand docking, J. Mol. Struct. THEOCHEM 898 (1–3) (2009) 31–41, http://dx.doi.org/10.1016/j.theochem.2008.12.041.

[82] X. Liu, F. Bai, S. Ouyang, X. Wang, H. Li, H. Jiang, Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation, BMC Bioinform. 10 (1) (2009) 101, http://dx.doi.org/10.1186/1471-2105-10-101.

[83] J. Tavares, S. Mesmoudi, E.-G. Talbi, On the efficiency of local search methods for the molecular docking problem, in: Evolutionary Computation, Machine Learning and Data Mining in Bioinform, Springer Berlin Heidelberg, 2009, pp. 104–115, http://dx.doi.org/10.1007/978-3-642-01184-9_10.

[84] J. Fuhrmann, A. Rurainski, H.-P. Lenhof, D. Neumann, A new Lamarckian genetic algorithm for flexible ligand–receptor docking, J. Comput. Chem. 31 (9) (2010) 1911–1918, http://dx.doi.org/10.1002/jcc.21478.

[85] K. Ling, Z. Xiaoyu, C. Xi, W. Xicheng, An evolvement-based genetic algorithm for computer-aided molecular docking, AIP Conf. Proc. 1233 (1) (2010) 545–548, http://dx.doi.org/10.1063/1.3452231.

[86] E. Atilgan, J. Hu, Efficient protein-ligand docking using sustainable evolutionary algorithms, in: 2010 10th International Conference on Hybrid Intelligent Systems, IEEE, 2010, pp. 113–118, http://dx.doi.org/10.1109/his.2010.5600082.

[87] F. Archetti, I. Giordani, L. Vanneschi, Genetic programming for QSAR investigation of docking energy, Appl. Soft Comput. 10 (1) (2010) 170–182, http://dx.doi.org/10.1016/j.asoc.2009.06.013.

[88] R. Meier, M. Pippel, F. Brandt, W. Sippl, C. Baldauf, ParaDockS: a framework for molecular docking with population-based metaheuristics, J. Chem. Inform. Model. 50 (5) (2010) 879–889, http://dx.doi.org/10.1021/ci900467x.

[89] S. Mesmoudi, J. Tavares, L. Jourdan, E.-G. Talbi, Variable genetic operator search for the molecular docking problem, in: Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, Springer Berlin Heidelberg, 2010, pp. 1–12, http://dx.doi.org/10.1007/978-3-642-12211-8_1.

[90] P. Rondon, H. Arguello, R. Torres, Development of a zoned genetic algorithm for semi-flexible protein-ligand docking in drug design, in: 2011 6th Colombian Computing Congress, CCC, IEEE, 2011, pp. 1–4, http://dx.doi.org/10.1109/colomcc.2011.5936321.

[91] Y. Liu, L. Zhao, W. Li, D. Zhao, M. Song, Y. Yang, FIPSDock: A New molecular docking technique driven by fully informed swarm optimization algorithm, J. Comput. Chem. 34 (1) (2012) 67–75, http://dx.doi.org/10.1002/jcc.23108.

[92] A.N. Lima, E.A. Philot, D. Perahia, A.S.K. Braz, L.P. Scott, GANM: A Protein–ligand docking approach based on genetic algorithm and normal modes, Appl. Math. Comput. 219 (2) (2012) 511–520, http://dx.doi.org/10.1016/j.amc.2012.06.030.

[93] M. Koohi-Moghadam, A.T. Rahmani, Molecular docking with opposition-based differential evolution, in: Proceedings of the 27th Annual ACM Symposium on Applied Computing - SAC'12, ACM Press, 2012, pp. 1387–1392, http://dx.doi.org/10.1145/2245276.2231996.

[94] Y. Liu, W. Li, R. Ma, Particle swarm optimization on flexible docking, Int. J. Biomath. 05 (05) (2012) 1250044, http://dx.doi.org/10.1142/s1793524511001866.

[95] L. Zhen-yu, C. Ming-hua, K. Ling, Application of immune algorithm in molecular docking, in: Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics, SPAC, IEEE, 2014, pp. 122–125, http://dx.doi.org/10.1109/spac.2014.6982670.

[96] Z. Li, J. Gu, H. Zhuang, L. Kang, X. Zhao, Q. Guo, Adaptive molecular docking method based on information entropy genetic algorithm, Appl. Soft Comput. 26 (2015) 299–302, http://dx.doi.org/10.1016/j.asoc.2014.10.008.

[97] P.A. Ravindranath, S. Forli, D.S. Goodsell, A.J. Olson, M.F. Sanner, AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility, in: J.S. Fetrow (Ed.), PLOS Comput. Biol. 11 (12) (2015) e1004586, http://dx.doi.org/10.1371/journal.pcbi.1004586.

[98] W. Xu, A.J. Lucke, D.P. Fairlie, Comparing sixteen scoring functions for predicting biological activities of ligands for protein targets, J. Mol. Graph. Modelling 57 (2015) 76–88, http://dx.doi.org/10.1016/j.jmgm.2015.01.009.

[99] E. López-Camacho, M.J.G. Godoy, J. García-Nieto, A.J. Nebro, J.F. Aldana-Montes, Solving molecular flexible docking problems with metaheuristics: A comparative study, Appl. Soft Comput. 28 (2015) 379–393, http://dx.doi.org/10.1016/j.asoc.2014.10.049.

[100] B. Guan, C. Zhang, J. Ning, EDGA: a population evolution direction-guided genetic algorithm for protein–ligand docking, J. Comput. Biol. 23 (7) (2016) 585–596, http://dx.doi.org/10.1089/cmb.2015.0190.

[101] P.F. Leonhart, E. Spieler, R. Ligabue-Braun, M. Dorn, A biased random key genetic algorithm for the protein–ligand docking problem, Soft Comput. (2018) http://dx.doi.org/10.1007/s00500-018-3065-5.

[102] B. Guan, C. Zhang, Y. Zhao, An efficient ABC_DE_based hybrid algorithm for protein–ligand docking, Int. J. Mol. Sci. 19 (4) (2018) 1181, http://dx.doi.org/10.3390/ijms19041181.

[103] Y. Fu, Z. Chen, J. Sun, Random drift particle swarm optimisation algorithm for highly flexible protein-ligand docking, J. Theoret. Biol. 457 (2018) 180–189, http://dx.doi.org/10.1016/j.jtbi.2018.08.034.

[104] F. Neri, C. Cotta, Memetic algorithms and memetic computing optimization: A literature review, Swarm Evol. Comput. 2 (2012) 1–14, http://dx.doi.org/10.1016/j.swevo.2011.11.003.

[105] H. Szczerbicka, M. Becker, M. Syrjakow, Genetic algorithms: a tool for modelling, simulation, and optimization of complex systems, Cybern. Syst. 29 (7) (1998) 639–659, http://dx.doi.org/10.1080/019697298125461.

[106] B.S.P.D. Schneidman-Duhovny, B.S.P.R. Nussinov, B.S.P.H.J. Wolfson, Predicting molecular interactions in silico: II. Protein-protein and protein-drug docking, Current Med. Chem. 11 (1) (2004) 91–107, http://dx.doi.org/10.2174/0929867043456223.

[107] D.T.-H. Chang, J.-H. Lin, C.-H. Hsieh, Y.-J. Oyang, On the design of optimization algorithms for prediction of molecular interactions, Int. J. Artif. Intell. Tools 19 (03) (2010) 267–280, http://dx.doi.org/10.1142/s0218213010000182.

[108] E. López-Camacho, M.J. García-Godoy, A.J. Nebro, J.F. Aldana-Montes, jMetalCpp: optimizing molecular docking problems with a C++ Metaheuristic framework, Bioinformatics 30 (3) (2013) 437–438, http://dx.doi.org/10.1093/bioinformatics/btt679.

[109] P. Rakshit, A. Konar, A. Chowdhury, E. Kim, A.K. Nagar, Muti-objective evolutionary approach of ligand design for protein-ligand docking problem, in: 2013 IEEE Congress on Evolutionary Computation, IEEE, 2013, pp. 237–244, http://dx.doi.org/10.1109/cec.2013.6557576.

[110] Y. Liu, L. Zhao, M. Li, C. Zhao, Swarm intelligence for molecular docking, Int. J. Model. Ident. Control 18 (4) (2013) 357, http://dx.doi.org/10.1504/ijmic.2013.053541.

[111] C.S. de Magalhães, D.M. Almeida, H.J.C. Barbosa, L.E. Dardenne, A dynamic niching genetic algorithm strategy for docking highly flexible ligands, Inform. Sci. 289 (2014) 206–224, http://dx.doi.org/10.1016/j.ins.2014.08.002.

[112] M. García-Godoy, E. López-Camacho, J. García-Nieto, A. Nebro, J. Aldana-Montes, Molecular docking optimization in the context of multi-drug resistant and sensitive EGFR mutants, Molecules 21 (11) (2016) 1575, http://dx.doi.org/10.3390/molecules21111575.

[113] S.C.W. Peh, J.L. Hong, Protein ligand docking using simulated jumping, in: Computational Science and Its Applications – ICCSA 2016, Springer International Publishing, 2016, pp. 1–10, http://dx.doi.org/10.1007/978-3-319-42111-7_1.

[114] J.M. Cecilia, J.-M. Cutillas-Lozano, D. Giménez, B. Imbernón, Exploiting multilevel parallelism on a many-core system for the application of hyperheuristics to a molecular docking problem, J. Supercomput. 74 (5) (2017) 1803–1814, http://dx.doi.org/10.1007/s11227-017-1989-7.

[115] B. Guan, C. Zhang, J. Ning, Genetic algorithm with a crossover elitist preservation mechanism for protein–ligand docking, AMB Express 7 (1) (2017) http://dx.doi.org/10.1186/s13568-017-0476-0.

[116] A. Nebro, J. Durillo, J. Garcia-Nieto, C.C. Coello, F. Luna, E. Alba, SMPSO: A new PSO-Based metaheuristic for multi-objective optimization, in: 2009 IEEE Symposium on Computational Intelligence in Milti-Criteria Decision-Making, IEEE, 2009, pp. 66–73, http://dx.doi.org/10.1109/mcdm.2009.4938830.

[117] Y. Zhao, H.-L. Liu, Multi-objective particle swarm optimization algorithm based on population decomposition, in: Intelligent Data Engineering and Automated Learning – IDEAL 2013, Springer Berlin Heidelberg, 2013, pp. 463–470, http://dx.doi.org/10.1007/978-3-642-41278-3_56.

[118] C. Dai, Y. Wang, M. Ye, A new multi-objective particle swarm optimization algorithm based on decomposition, Inform. Sci. 325 (2015) 541–557, http://dx.doi.org/10.1016/j.ins.2015.07.018.

[119] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated Docking with selective receptor flexibility, J. Comput. Chem. 30 (16) (2009) 2785–2791, http://dx.doi.org/10.1002/jcc.21256.

[120] E. Zitzler, L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach, IEEE Trans. Evol. Comput. 3 (4) (1999) 257–271, http://dx.doi.org/10.1109/4235.797969.

[121] E. Zitzler, L. Thiele, M. Laumanns, C. Fonseca, V. da Fonseca, Performance assessment of multiobjective optimizers: an analysis and review, IEEE Trans. Evol. Comput. 7 (2) (2003) 117–132, http://dx.doi.org/10.1109/tevc.2003.810758.

[122] G. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1) (1963) 95–99, http://dx.doi.org/10.1016/S0022-2836(63)80023-6, http://www.sciencedirect.com/science/article/pii/S0022283663800236.

[123] A. Eiben, S. Smit, Parameter tuning for configuring and analyzing evolutionary algorithms, Swarm Evol. Comput. 1 (1) (2011) 19–31, http://dx.doi.org/10.1016/j.swevo.2011.02.001, http://www.sciencedirect.com/science/article/pii/S2210650211000022.

[124] L.C.T. Bezerra, M. López-Ibáñez, T. Stützle, Automatic component-wise design of multiobjective evolutionary algorithms, IEEE Trans. Evol. Comput. 20 (3) (2016) 403–417, http://dx.doi.org/10.1109/TEVC.2015.2474158.

[125] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm Evol. Comput. 1 (1) (2011) 3–18.

[126] G. Vrbančič, L. Brezočnik, U. Mlakar, D. Fister, I. Fister Jr, NiaPy: Python Microframework for building nature-inspired algorithms, J. Open Source Softw. 3 (2018).

[127] A. Benitez-Hidalgo, A.J. Nebro, J. Garcia-Nieto, I. Oregi, J. Del Ser, jMetalPy: A Python Framework for Multi-Objective Optimization with Metaheuristics, 2019, arXiv:1903.02915.