Multi-Objective Optimization Methods in *De Novo* Drug Design

C.A. Nicolaou*[,1,2], C. Kannas[2,3] and E. Loizidou[2,4]

[1]*Computation-based Science and Technology Research Center, The Cyprus Institute, 15 Kypranoros Street, Nicosia 1061, Cyprus;* [2]*Computer Science Department, University of Cyprus, 75 Kallipoleos Street, CY-1678 Nicosia, Cyprus;* [3]*Noesis Chemoinformatics, Metcohiou 66, Nicosia, CY-1461, Cyprus;* [4]*CySilicoTech Research Limited, Kronou 64, Nicosia, CY-2048, Cyprus*

Abstract: *De-novo* drug design (DND) is a complex procedure, requiring the satisfaction of many pharmaceutically important objectives. Several computational methodologies employing various optimization approaches have been developed to search for satisfactory solutions to this multi-objective problem varying from composite methods, which transform the problem to a single objective one to Pareto methods searching for numerous solutions compromising the objectives. In this review we initially focus on the DND problem and the challenges it poses to computational methods, followed by an examination of the reported methodologies and specific applications. Emphasis is placed on the multi-objective nature of the problem, related considerations and the solutions proposed by the drug discovery community.

Keywords: Multi-objective optimization, de novo design, drug discovery.

1. INTRODUCTION

De-novo drug (or ligand) design (DND) attempts to generate ligands from scratch based only on information about the pharmaceutical target site or known ligands [1]. The design products need to satisfy a number of objectives of crucial pharmaceutical importance. Among them is biological activity against the target of interest, selectivity to the specific target and a number of pharmacokinetic properties collectively known as ADME (Absorption, Distribution, Metabolism, Excretion) [2]. The multitude of design constraints turn DND into a multi-objective optimization problem of significant complexity, but also importance [3], since the method can potentially produce novel chemical designs representing a wide range of compromises of the supplied constraints and may therefore be used as an "idea generator" to support the lead discovery process.

The task faced by optimization methods used in the DND field is that of exploring a chemical search space consisting of all possible chemical compounds with drug-like characteristics and, identifying those that satisfy the specific problem constraints imposed. The size of this space makes a full enumeration impossible and so powerful search methods need to be applied to detect the best possible solutions in a limited amount of time. In addition to searching an immense, complex space for solutions satisfying multiple, often conflicting objectives, DND methods need to implement virtual synthesis engines that produce chemically synthesizable structures, a task that has proven challenging to this day [4].

*Address correspondence to this author at the Computation-based Science and Technology Research Center, The Cyprus Institute, 15 Kypranoros Street, Nicosia 1061, Cyprus; Tel: +357-22208644; Fax: +357-22208625; E-mail: cnicolaou@cs.ucy.ac.cy

In the following sections an overview of the de novo design field is given placing special emphasis on multi-objective methods. Section 2 describes the main challenges and considerations DND methods need to deal with. Section 3 reviews the algorithmic approaches used by the DND community to address the presence of multiple objectives in the process. The next section briefly presents applications using single-objective optimization approaches while section 5 describes in detail multi-objective methods of various types. The final section summarizes our conclusions on the progress performed in the DND field and suggests directions for future research.

2. DND METHOD CHALLENGES AND CONSIDERATIONS

The main challenges facing any method attempting to identify solutions to a given problem include representing the problem accurately, generating valid candidate solutions from the feasible solution space and assessing the quality of the proposed solutions. Satisfying these requirements in the DND setting requires, among others, the encoding of pharmaceutically relevant chemical structure assessment methods, the implementation of a virtual chemical structure generation engine, and the usage of an optimization method for exploring the chemical search space of interest. Fig. (**1**) presents the major steps of a traditional de novo design process.

Chemical structure assessment traditionally involved the calculation of the similarity of a structure to a known drug [5] or the prediction of its binding affinity to a pharmaceutical target receptor [6]. These types of scoring over-emphasized the biological activity potential of a structure and ignored the multitude of additional constraints necessary for a compound to become a drug. The advent of chemoinformatics in the last two decades has provided computational methods for the calculation of numerous compound properties including drug-likeness [7], ADME
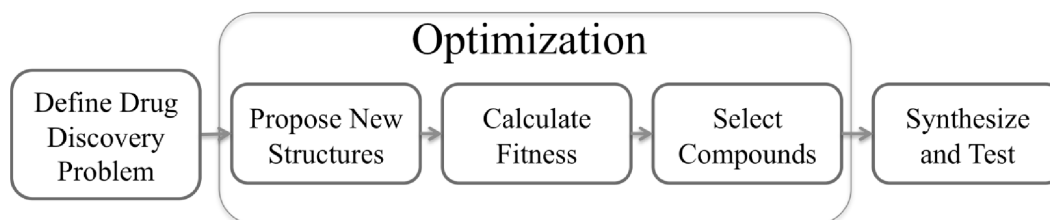
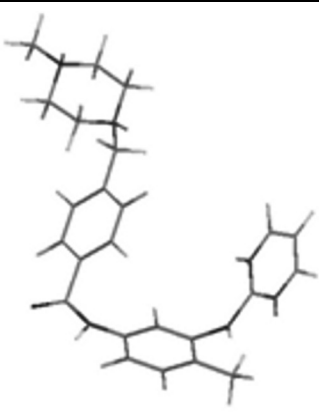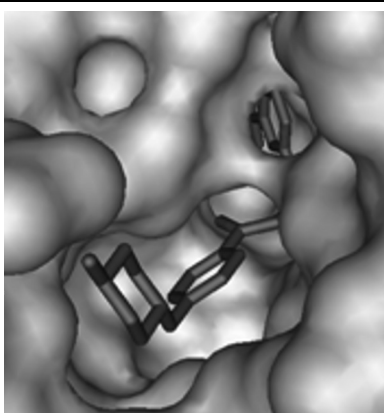**Fig. (1).** General diagram of the *de novo* drug design process.

and toxicity properties [3] which can readily be used in a search procedure initiated as part of a de novo design effort [8]. Table **1** summarizes the various chemical structure scoring methods.

Ideally, the generation of chemical structures in silico requires putting together compound designs that not only satisfy chemical rules (e.g. correct valences, molecular stability) but are also synthesizable. DND algorithms proposed in the literature typically use a predefined collection of molecular fragments as building blocks and a set of synthesis rules to increase the synthetic feasibility potential of the designed chemical structures [4]. Often, the molecular fragments are extracted through the fragmentation of a collection of molecules from a given database [9] although some methods also use simple atoms and bonds [10]. Optionally, information on reaction points of the fragments is also kept. The virtual synthesis of molecules needs to produce acceptable chemical designs by combining in an intelligent manner available building blocks selected in

a random fashion, or with a probabilistic bias based on the frequency of occurrence of each fragment.

The ability to design virtual compounds and score them against a number of pharmaceutically relevant objectives enables the implementation of de novo design techniques aiming to generate chemical structures occupying a specific region of the chemical space and, possessing a desirable biological profile. To this end, a number of optimization algorithms have been used including monte carlo [11, 12], graph search [13] and, overwhelmingly in recent years, evolutionary algorithms (EA) [8, 14]. It is worth stressing the size and complexity of the chemical search space DND methods have to explore. Bohacek *et al.* estimated the size of the space to be in the order of 10^{60} [15] while Reymond *et al.* [16] have introduced a global chemical universe database, the GDB, which enumerates actual lists of all molecules that are possible up to a certain size following simple constraints of chemical stability and synthetic feasibility. The group has published GDB-11, an enumeration of all molecules up to 11

Table 1.   Chemical Structure Assessment Categories

DND Objective Categories		
Ligand-based	**Receptor-based**	**Property-based**
		Lipinski properties (MW, logP, number of H-bond donors/acceptors**)** **Selectivity** **Polar surface area, pKa** **etc**
Score is based on the chemical similarity of the designed compounds to a known active molecule. Anticancer agent Imatinib is shown as an example of a known active molecule	Score is based on the predicted binding affinity of the designed compounds to a known pharmaceutical receptor. The docked structure of Imatinib (stick representation) at the active site of kinase AbI (surface representation is shown as an example	Score is based on calculated pharmaceutically relevant property(ies) such as Lipinski's properties, selectivity to the target of interest and ADME properties

atoms with C, N, O, F that has approximately 26.4 million compounds and GDB-13 with atoms C, N, O, Cl, S that has about 980 million compounds [16]. Moreover, the presence of multiple conflicting objectives results in a complex, non-uniform search space since multiple, equivalent solutions may be present at different regions of the space [1]. Equally significant is that many objective functions used in DND are not very precise and therefore the demand for optimality to these objectives may result in noisy solution spaces and increased risk of excluding valid solutions. It is therefore apparent that any optimization method used needs to combine efficiency with robustness to complex multimodal search spaces typical in real-life multi-objective problems.

3. HANDLING THE MULTI-OBJECTIVE NATURE OF DND

In a multi-objective problem, multiple equivalent solutions representing different compromises among the objectives are possible [8]. This is especially true when the objectives considered are in conflict, e.g. when improving performance on one objective tends to worsen performance in another. These multiple 'best' solutions, known as non-dominated, have no other solutions that are better than them in all of the objectives considered. Solutions are said to be dominated if there exist one or more solutions in the set that exhibit better performance in all objectives. The set of non-dominated solutions is also known as the tradeoff surface or the Pareto front named after the engineer/economist V. Pareto who introduced the domination concept. Fig. (**2**) illustrates the concept of nondominated solutions and the Pareto front.
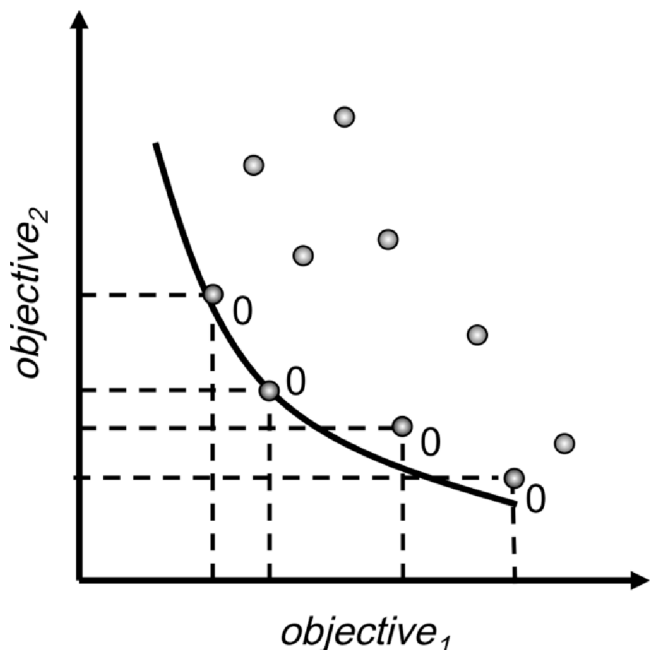

Fig. (2). A bi-objective minimization problem and a set of solutions (circles). Non-dominated solutions are labeled '0'. Multiple equivalent 'best' solutions are possible in multi-objective problems representing different compromises between the considered objectives. The solutions form the so-called Pareto front or trade-off .

In an attempt to simplify the problem, most de novo design methods ignore the multi-objective nature of drug discovery and focus on designing molecules satisfying a single objective, either predicted binding affinity to a known protein target [17] or, similarity to a ligand [9, 18], A second category of methods recognizes the existence of multiple objectives in drug discovery and attempts to take them into account in the design process. In addition to protein-ligand docking score and similarity to a target molecule, fitness scores based on Quantitative Structure Activity Relationship (QSAR) functions [19], drug-likeness [20] and experiments [21] can also be used. The majority of multi-objective methods combine the numerous objectives into a single one prior to the application of an optimization method. These methods effectively decide a´ priori on the relative importance of each existing objective, often by associating a weight to each one of them, to generate a new, composite objective. Alternatively, some methods follow a different approach and strive to identify chemical structure solutions covering the Pareto front of the specific problem investigated. These Pareto optimization based approaches produce sets of solutions that represent different compromises of the objectives and allow the user to choose those that better match their goals. While this may be perceived as a problem, the availability of several candidate solutions in reality enables users to choose a´ posteriori those that meet their criteria best. Specifically in DND, the generation of multiple equivalent diverse end-product solutions is in fact preferable since it provides experts with numerous, alternative starting points for the lead optimization phase. In addition, this approach does not require the (often) artificial assigning of weights on the various objectives. A more recent method that follows the composite approach to multi-objective optimization uses desirability functions where a global desirability index for each candidate solution is obtained from the individual compound objectives [22]. Finally, some methods provide a progressive, user-directed de novo design approach where chemical intuition [23] is used as the fitness function with users interactively evaluating generated molecules.

Among the most commonly used optimization methods in DND are Evolutionary Algorithms (EA). These techniques follow the concepts of Darwinian evolution to gradually design a fit population of individuals subject to environmental pressure. EA based DND methods start with an initial population of small molecules and iteratively build new, more 'fit' compounds based on the fitness score of the previous set of molecules through modifications mimicking natural breeding (i.e. crossover operations) and mutations [1]. Since EA's simultaneously optimize a population of individuals during search they are particularly suitable for multi-objective problems where a set of solutions that covers the Pareto front needs to be found. Consequently, the preservation of population diversity is a major issue in multi-objective EA and has recently been attracting attention in the context of DND as well [1, 24].

Also popular, especially in early DND methods, are graph-based combinatorial search approaches that explore the solution space using techniques such as breadth- or depth-first search to generate chemical designs meeting the

constraints imposed [4]. In DND, these approaches use a set of fragments with given connection points and combine them to form new chemical designs. The combination of the fragments takes place by successively enumerating the possible attachments to other fragments at each connection point and retaining the best performer(s). Depending on the search technique used new structures can be gradually 'grown' from initial fragments or 'linked' by first selecting promising fragments and then connecting them with 'linker' fragments.

There exist numerous reviews on DND methods and applications. The interested reader is referred to [25], [26] and [27]. Reviews for older methodologies can be found in [4] and [28]. In the following sections we describe a selection from the methods proposed in the literature with emphasis on recent applications and Pareto-based approaches. The presentation of the methods is organized with respect to the general methodology followed to address the multiple objectives involved in drug discovery.

4. SINGLE OBJECTIVE DND METHODS

The majority of DND approaches reported in the literature ignore the presence of multiple objectives in the pharmaceutical process and focus on optimizing a potency related objective. Broadly, methods following this approach fall into two categories depending whether the objective pursued is ligand or target-based.

Ligand-based approaches use known ligands to define their objective functions and measure the fitness of designed structures. A representative ligand-based de novo design method is TOPAS (TOPology Assigning System), which uses 2D fragments derived from known drug molecules and an EA method to design molecules similar to a target chemical structure [14]. The method, as well as its ancestor Flux (Fragment-based Ligand Building reaXions) [9], uses retrosynthetic analysis [29] to generate the fragment 'genes' and keeps information about the type of bonds at each attachment point. Candidate compounds are then evolved via operations that take into account chemical synthesis rules. TOPAS used exclusively mutation in the form of fragment substitution. Flux uses a richer collection of genetic operators that enable recombination of parent molecules via crossover, and changes in the number of fragments a molecule contains. In both algorithms the fitness function was based on the chemical similarity of the candidate compounds to a known active molecule. Certain drug-likeness rules are taken into account for compound selection. A recent application of Flux [30] indicated that the method can design compounds with substantial structural differences from the initial target thus achieving the so-called scaffold-hopping goal. Similar approaches have been reported in [10] and [21].

Target-based approaches rely on the availability of a detailed description of the target receptor of interest to design chemical structures predicted to bind well. The evaluation of the design products is performed via one or more docking/scoring methods that give an indication of the likely affinity of each virtual compound to the receptor [6]. Search-based optimization methods, such as EAs are

frequently used, especially in more recent approaches reported in the literature, in combination with a virtual compound synthesis engine, to design compounds satisfying an objective function based on docking [31-33]. Alternatively, some methods use the available knowledge on the receptor site to identify and characterize its regions that can be involved in chemical interactions [13]. Following, an incremental construction approach can be used to generate virtual compounds through coupling key receptor regions with molecular fragments that can theoretically interact and form chemical bonds. Special chemical substructures, known as linkers, are used to link the molecular fragments taking into account geometry considerations. The knowledge of the receptor site may also be used to derive a model of compounds with predicted binding affinity in which case the goal of the de novo design process is to generate compounds matching that model. Pro_Ligand used a fragment-based approach and a depth-first search method to incrementally design ligands fitting a model derived from a target receptor site or a collection of highly similar actives [34]. The method generates the ligands by matching fragments with the model components and constructs virtual molecules using standard chemical rules. In later publications the method was complemented by a post-processing EA module that further evolves the designed compounds using a limited set of evolutionary operations [35]. More recently, GARLig (Genetic Algorithm using Reagents to compose LIGands) proposed a self-adaptive EA for library design that also uses docking scores as objective functions [36]. The method decorates a given scaffold with fragments also provided as input, to design virtual compounds with high predicted binding affinity to a target receptor.

5. MULTI-OBJECTIVE DND METHODS

Among the DND methods that recognize the existence of multiple objectives several distinct subcategories exist. Broadly, these methods can be categorized according to the methodology used to address the various objectives. Composite methods, representing the overwhelming majority, transform the problem into a single-objective one by aggregating the multiple objectives into a single one. These methods in effect decide on the importance of each objective a' priori and use a prioritization scheme to calculate the contribution of each objective to the new one. Pareto-based methods attempt to identify compromise solutions among the various objectives and thus, avoid the problem of having to prioritize the multiple objectives. Instead, they produce multiple solutions which they present to the end user who can select those matching its requirements a' posteriori. Methods using desirability functions which provide a way to reduce objectives relative to user specified criteria and thus decrease the complexity of the problem have also been used in DND in combination with single- and multi-objective optimization methods. In the following sections we review these subcategories. A fourth section is focusing on interactive, progressive methods which give a pivotal role to the expert user who is actively engaged in the design process to evaluate and select solutions for further optimization. Table **2** presents the various categories of multi-objective optimization methods

Table 2.   Approaches Used by Multi-objective DND Methods to Address the Presence of Multiple Objectives

Multiple Objective Approaches			
Method Type	**User Interaction**	**Methods**	**Sample Applications**
A´ priori	Before optimization	Weighted-Average: Specification of expert knowledge by assignment of weights to objectives; aggregate objectives into a single one through linear combination of weighted averages of all objectives	Chemical Genesis [31], LEA [37], GANDI [33], FOG [20], PHDD [44]
		Desirability: Specification of expert knowledge by assignment of desirability to objectives; aggregate objectives into a single one through desirability functions	MOOP-DESIRE [47]
A´ posteriori	After optimization	Pareto-based: Optimization process takes place without usage of prior knowledge; produce set of optimal solutions; expert knowledge used to select set of desired solutions after optimization	COG [50], MEGA [1], PLD [55]
Progressive	During optimization	Interactive: Enable the user to interact with the optimization process to guide the search; user acts as the fitness function/scorer.	MoleculeEvoluator [23], Mobius [56]

used for DND organized according to the way they handle the multiple objectives.

5.1. Composite Methods

A straightforward approach in finding compromise solutions when numerous objectives are present is to transform the problem to a single-objective one by combining the multiple objectives. A common example of this approach is the weighted-sum-of-objective-functions method where a weight is associated a´ priori with each objective function and the weighted sum of the functions is taken as the new composite fitness function. An advantage of using such a scalarized objective function is that the same algorithms used for solving single-objective problems can be used for multi-objective problems. Drawbacks of the method include the need to select appropriate weighting for the different objectives even when the relation among them is not clear, and, the generation of 'best' solutions with no associated information of their placement on the Pareto front instead of the set of nondominated solutions.

One of the first methods using a composite fitness function was Chemical Genesis proposed by Glen and Payne in 1995 [31]. Their program uses molecular fragments to design molecules using a single-objective evolutionary algorithm. The fitness function used combines both receptor and ligand-based objectives. New chemical structure designs are produced through mutation, which allows structural modifications such as changing atom and bond types, inserting and removing fragments, and, crossover, which involves the exchange of fragments between two molecules [31].

Douguet *et al.* [37, 38] also used an EA-based search combined with an extensive set of mutation and crossover operators, and a set of repair mechanisms to ensure the validity of the produced chemical structures representations, especially with regard to branching and ring correctness. The original method, termed LEA (Ligand by Evolutionary Algorithm), represented compounds using the SMILES chemical language [39] and used a composite function consisting of ligand-based objectives and specifically QSAR model predictions. Molecular perturbation took place by modifying the SMILES compound representation. In subsequent work, LEA3D [38] was introduced which uses a pool of 3D fragments combined in a linear fashion. The method uses a composite fitness function taking into account both receptor and ligand-based constraints. ADAPT [32] used an EA to design compounds satisfying a composite criterion combining docking scores and simple physical properties like molecular weight and number of rotatable bonds. The method generated compounds by combining fragments from a user-supplied collection using both mutation and crossover operations to explore the chemical space. Similarly, LigBuilder [40], uses a multi-objective composite objective and an EA algorithm to build up ligands from a library of organic fragments by using growing and linking strategies. The new molecules are evaluated based on their binding affinities, estimated through an empirical scoring function, and the biological availability, evaluated based on a set of chemical rules.

An EA algorithm is also used by the method proposed by Feher *et al.* [41] for searching the chemical space. The commercially available structure generation program EA-Inventor [42] implements an evolutionary algorithm to generate new chemical structures from a seed set or previous generation and is agnostic to the single scoring function used. The specific application reviewed used a composite scoring function that comprised of the product of individual scores including, among others, ligand similarity to reference compounds, molecular weight and stereochemistry. Designed compounds containing fragments from a user-defined list of undesirable substructures were eliminated. The method was used for the design of selective norephinephrine re-uptake inhibitor ligands (SNRI) and binders to the gonadotropin releasing hormone (GnRH) receptor.

NovoFLAP [43] also used the EA-Inventor [42] as its chemical synthesis engine. In this implementation,

EA-Inventor uses a fragment library of 1300 fragments derived from known drugs and chemical transformation operators to generate chemical designs obeying valence rules. The new designs are evaluated using FLAP (Flexible Ligand Alignment Protocol) which is a composite scoring function aggregating measures based on the fit of overall molecular shape and pharmacophoric features to reference compounds.

GANDI (Genetic Algorithm-based de Novo Design of Inhibitors) joins a collection of predocked 3D fragments to a receptor site with a set of linkers to generate candidate molecules [33]. Individuals are represented as simple trees whose general shape and structure is restricted by the receptor site targeted. The method divides the working population into subpopulations and uses a parallel evolutionary algorithm with binary tournament selection of parents for selecting the predocked fragments and tabu-search to select the linkers [33]. The scoring function in GANDI is a linear combination of terms measuring both 2D and 3D properties of an individual.

FOG (Fragment Optimized Growth), grows molecules by iteratively adding fragments in a statistically biased manner that implicitly takes into account the presence of multiple design objectives such as similarity to known ligands and synthesizability [20]. The algorithm relies on the calculation of the transition probabilities of a given growth fragment to other fragments in a database. Transition probabilities are based on connectivity statistics for fragments of interest from collections of small molecules. The selection of an appropriate training database of small molecules for fragmentation and calculation of connectivity statistics enables FOG to generate new compounds in a chemical space similar to the compounds in the training set. As currently implemented FOG can be incorporated as a synthesis engine in more elaborate fragment-based *de novo* design programs or, alternatively, it can be used as a standalone program to generate a virtual library of compounds of specific classes [20]. In a more recent publication, the same group suggested using FOG in combination with a composite fitness score mechanism which would take into account various objectives (synthetic accessibility, drug-likeness, solubility, etc.) to guide the search in relevant regions of the search space using an evolutionary algorithm [25].

More recently PhDD (Pharmacophore-based De novo Design method) has been developed to design new molecules based on the requirements of a 3D pharmacophore model [44]. The method generates molecules with high similarity to known pharmacophores by iteratively combining substructures derived from the fragmentation of known drugs. The new molecules are scored on their drug-likeness, bioactivity and synthetic accessibility. Drug-likeness is evaluated with Lipinski's rule of five [7] and the bioactivity from fitness values calculated according to specific formulas that describe the distance between the center of the fragment and that of the pharmacophore model. Synthetic accessibility is assessed by a complexity method that takes into consideration the contribution of rings, interatomic connections, atom types and chiral centers.

The weighted average approach to combine multiple objectives is also used in the Molecular Library Design area. A recent example is LoFT (Library optimizer using Feature Trees), a tool for focused combinatorial library design that uses a weighted multi-objective scoring function [45]. The method takes as input a fragment collection and generates compound libraries satisfying a composite criterion that incorporates consensus similarity scoring to multiple query molecules, and, maintains the products within desired property ranges. LoFT can use any of a number of optimization methods for searching the chemical space.

5.2. Desirability-Based Methods

Desirability-based methods is another approach to multi-objective optimization, where several variables are optimized simultaneously based on desirability functions. This method has primarily found applications in the building of QSAR models that can eventually serve as fitness functions to direct the de novo design of further molecules [46]. Recently, MOOP-DESIRE (Multi-Objective Optimization based on desirability estimation) was introduced based on which multiple objectives are simplified to a single one using independent desirability criteria [22, 47]. The MOOP-DESIRE methodology includes three main steps: 1) developing predictive models, 2) obtaining the global desirability (Di) from the individual desirability (di) for each compound, and for each biological property and 3) using the global desirability or the descriptor values as a template for a ranking algorithm that will rank new candidates.

The MOOP-DESIRE methodology can be applied to the screening of large and diverse data (High-Throughput Screening) in order to filter out the best ranked drug candidates that would combine potency, safety and bioavailabitiy. This was exemplified by Cruz-Monteagudo *et al.* who introduced the MOOP-DESIRE methodology. They applied this methodology in a global QSAR study that simultaneously considered the potency, bioavailability, and safety of 95 fluoroquinolines as well as, the analgesic, antiinflammatory, and ulcerogenic properties of fifteen 3-(3-methylphenyl)-2-substituted amino-3H-quinazolin-4-one. MOOP-DESIRE based optimization was also performed to a set of chlorophenyl derivatives in order to optimize their global antifungal profile and provide new chemicals with broad antifungal spectrum. Their results pointed to a hydroxyl group as being fundamental for the expression of antifungal activity [48]. Improvement of selectivity can also be achieved by MOOP-DESIRE and this was demonstrated by the work of Machado *et al.* Their study conducted the optimization of arylpiperazine derivates, inhibitors of both 5-HT1A and 5-HT2A serotonin receptors, in terms of selectivity and affinity towards the 5-HT2A serotonin receptor [49].

5.3. Pareto-Based Methods

Pareto-based multi-objective optimization has been introduced to de novo design with the system proposed by Brown *et al.* [50]. The system, named COG (Compound Generator), designs chemical structures using a multi-objective evolutionary algorithm and scoring functions based on similarity calculations to existing molecules of interest.

COG operates on a genetic graph molecular representation through mutations on both nodes and edges, and crossover. The method uses both molecular fragments and atoms/bonds as building blocks and imposes no constraint on the size or complexity of the chemical structures designed other than those required for the graph to represent a valid molecule [8]. COG has been applied successfully using quantitative structure-property relationship (QSPR) models to calculate individual compound fitness [51]. The same group reported an extension to this method that in addition to satisfying multiple molecular properties paid particular attention in designing molecules with structural differences to the query molecules [52].

The Multi-objective Evolutionary Graph Algorithm (MEGA) also operates on a genetic graph molecular representation [1]. MEGA combines evolutionary algorithms with local search techniques, to enable the use of problem-specific knowledge during the search and improve performance and scalability. The algorithm applies any number of objectives on the working population to obtain a list of scores for each individual. The list of scores is then used for the elimination of solutions with values outside the range allowed by the user-defined filters. Individual rank and population diversity are given special consideration through the implementation of a clustering process operating on the chemical structures. Solution generation takes place through graph-specific mutation and crossover. It is worth noting that MEGA maintains a secondary population, the so-called Pareto archive, which enables the preservation of promising solutions found throughout evolution and ensures that the results of the search process will contain the best solutions found. A variation of the method, MEGALib, which operates strictly with fragment building blocks and chemical rules for molecular synthesis has been used for multi-objective molecular library design [53]. The issue of maintaining chemical structure diversity among the solutions has also been addressed in the work of Kruisselbrink *et al.* who use a crowding operator based on compound similarity measurements to ensure the evolution of structurally diverse niches of molecules [24].

Shin *et al.* [54] propose an evolutionary multi-objective optimization approach for the design of oligonucleotide probes. The method initially generates a random population, calculates the domination relations and forms an archive with the nondominated set. The algorithm then generates offspring through variation and selection, and, performs maintenance of the Pareto archive and the working population until the termination condition is met. In the variation step, parent solutions are selected from archive and population. Offspring are generated through uniform crossover and point mutation. The method, termed EvoOligo, has been applied and compared favorably to existing approaches to the oligo probe design problem [54].

Ekins *et al.* (2010) report on the development of the Pareto Ligand Designer (PLD) [55]. The method is based on the EA-Inventor [42] commercial structure generation tool discussed previously, combined with an evolutionary multi-objective optimization component. At initalization, a set of one or more reference molecules is provided as input and the optimization objectives are prepared. Next, the dominance relations are calculated and the nondominated solutions are stored in a Pareto archive. The working population is used for reproduction via an extensive set of molecular transformations. The resulting new structures are subjected to hard filtering to ensure that their structural features and property values fall within acceptable ranges. The new working population is formed by the molecules that survive the filters and the molecules stored in the Pareto archive. PLD has been used succesfully in optimization experiments to simultaneously improve the predicted values of two, three and four objectives, while maintaining biological activity [55].

5.4. Interactive Methods

Interactive, or progressive multi-objective optimization methods have also been used in the DND field. Two independent groups have published their research on the design and implementation of software that provides a user-directed de novo design approach. The MoleculeEvoluator [23] represents molecules using the SMILES chemical language [39] and emphasizes mutation operations while Mobius [56] uses a simple tree representation with molecular fragments as building blocks and a predefined blueprint to drive the generation of new chemical designs. In effect, the blueprint is a recipe for combining fragments to create a molecule; it is used specifically to limit the evolutionary operations performed and ensure that new structures will conform to the desired design. Both tools generate candidate compounds for which a variety of properties of pharmaceutical interest are calculated. Mobius also offers the ability to measure the fitness of the individuals on additional objective functions that can be added to the process. Selection is left to the user who is expected to visually assess the intermediary design products and assign a score for each of the candidate molecules based on their expert knowledge, taking into account the molecular structure and the associated property values calculated previously. In effect, the user is the multi-object optimizer and can therefore, focus the area of exploration to those regions deemed to be of most interest for the particular application [8].

Recently, Kruisselbrink *et al.* [57] reported an extension to the MoleculeEvoluator through a method that combines Pareto optimization with desirability indexes to reduce de novo design from a many-objective to a multi-objective optimization problem with a lower number of objectives. The method combines sets of objectives into logical groups by means of desirability indexes to obtain a more manageable optimization problem. It then applies a multi-objective EA Pareto optimization method, to generate trade-off results with reduced computational effort. The method has been applied to the design of estrogen receptor antagonists.

6. SUMMARY

De novo drug design attempts to generate new drugs from scratch using existing information on the biological targets or molecules of interest. The approach represents an additional effort to identify lead molecules complementary to more traditional methods such as biological and virtual screening. In this approach, a medicinal chemist is confronted with the difficult task of exploring a virtually

infinite chemical space, in the order of 10^{60} molecules, and come up with lead compounds that are synthesizable, have high affinity towards their target and have drug-like characteristics. To overcome this obstacle, de novo design methods employ chemoinformatics techniques to take into account as much biochemical knowledge as possible in the form of scoring functions. Most of these functions include information on the ligand-receptor interaction (primary target constraints) and attempt to approximate the free energy of binding between the ligand and the target protein. However, an effective drug molecule is subject to more objectives than the binding affinity, including favorable ADME and toxicity properties, synthetic accessibility and target selectivity. This clearly demonstrates the multi-dimensional optimization character of drug discovery and development.

The multitude of methods presented above are evidence to the interest the DND field has attracted in recent years. This growing interest, also represented by the wealth of applications reported in the literature, is a result of the improvements in the methods used and the quality of the results produced. We believe that the recognition of the multi-objective nature of the DND problem and the efforts that followed to incorporate numerous objectives to the design process have also contributed to the improvement of the DND track record. However, a number of challenges still remain. Prime among them are the development of efficient search methods capable of sampling the vast chemical space that needs to be explored, and, the synthesizability of the chemical structure designs proposed. The importance of the aforementioned challenges has already been recognized and intense research efforts are taking place to address them. The ability to reuse pre-existing relevant information or knowledge gained during the optimization process to guide the search process, decrease execution time and facilitate the discovery of solutions, has not received as much attention. A natural way to exploit prior information is through the appropriate encoding of available knowledge into computational objectives and their inclusion as part of a multi-objective optimization process. Monitoring the progress of the DND process and assessing the quality of solutions produced during optimization may also provide the means to identify problem-specific promising regions of the chemical space and better focus the search through improved, self-adaptive methods. Enhancements in these, as well as additional algorithmic domains, are sure to further improve DND application performance in the near future and contribute to an increased role in the drug discovery field. Such development will surely benefit the drug discovery process and, in combination with other technological advancements, contribute to a reduction of the resources required to discover leads and develop successful drugs.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

PATIENT CONSENT

Declared none.

REFERENCES

[1] Nicolaou, C. A.; Apostolakis, J.; Pattichis, C. S. De Novo Drug Design Using Multiobjective Evolutionary Graphs. *J. Chem. Inf. Model.,* **2009**, *49*, 295-307.

[2] Baringhaus, K.-H.; Matter, H. In *Chemoinformatics in Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA, **2005**; pp. 333-379.

[3] Ekins, S.; Boulanger, B.; Swaan, P. W.; Hupcey, M. A. Z. Towards a New Age of Virtual ADME/TOX and Multidimensional Drug Discovery. *J. Comput.- Aided Mol. Des.,* **2002**, *16*, 381-401.

[4] Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov,* **2005**, *4*, 649-663.

[5] Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Model.,* **1998**, *38*, 983-996.

[6] Krovat, E. M.; Steindl, T.; Langer, T. Recent Advances in Docking and Scoring. *Current Computer - Aided Drug Design,* **2005**, *1*, 93-102.

[7] Lipinski C. A.; Lombardo F.; Dominy B. W.; Feeney P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.,* **1997**, *23*, 3-25.

[8] Nicolaou, C. A.; Brown, N.; Pattichis, C. S. Molecular optimization using computational multi-objective methods. *Curr. Opin Drug Discov. Devel.,* **2007**, *10*, 316-324.

[9] Fechner, U.; Schneider, G. Flux (1): a virtual synthesis scheme for fragment-based de novo design. *J. Chem. Inf. Model,* **2006**, *46*, 699-707.

[10] Globus, A.; Lawton, J.; Wipke, W. T. Automatic Molecular design using evolutionary algorithms. *Nanotechnology,* **1999**, *10*, 290-299.

[11] Todorov, N. P.; Dean, P. M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J. Comput. Aided Mol. Des.,* **1997**, *11*, 175-192.

[12] Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Boehm, H.-J.; Dean, P. M. A validation study on the practical use of automated de novo design. *J. Comput. Aided Mol. Des.,* **2002**, *16*, 459-478.

[13] Gillet, V. J.; Newell, W.; Mata, P.; Myatt, G.; Sike, S.; Zsoldos, Z.; Johnson, A. P. SPROUT: recent developments in the de novo design of molecules. *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 207-217.

[14] Schneider, G.; Lee, M. L.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput. Aided Mol. Des.,* **2000**, *14*, 487-494.

[15] Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.,* **1996**, *16*, 3-50.

[16] Reymond, J.-L.; van Deursen, R.; Blum, L. C.; Ruddigkeit, L. Chemical space as a source for new drugs. *Med. Chem. Commun.,* **2010**, *1*, 30.

[17] Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Van Aken, K.; Janssen, P. A. J. SYNOPSIS: SYNthesize and OPtimize System in Silico. *J. Med. Chem.,* **2003**, *46*, 2765-2773.

[18] Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. LEA3D: A Computer-Aided Ligand Design for Structure-Based Drug Design. *J. Med. Chem.,* **2005**, *48*, 2457-2468.

[19] Nachbar, Robert B. Molecular Evolution: Automated Manipulation of Hierarchical Chemical Topology and Its Application to Average Molecular Structures. *Genetic Programming and Evolvable Machines,* **2000**, *1*, 57-94.

[20] Kutchukian, P. S.; Lou, D.; Shakhnovich, E. I. FOG: Fragment Optimized Growth algorithm for the de novo generation of molecules occupying druglike chemical space. *J. Chem. Inf. Model,* **2009**, *49*, 1630-1642.

[21] Kamphausen, S.; Höltge, N.; Wirsching, F.; Morys-Wortmann, C.; Riester, D.; Goetz, R.; Thürk, M.; Schwienhorst, A. Genetic

algorithm for the design of molecules with desired properties. *J. Comput. Aided Mol. Des.,* **2002**, *16*, 551-567.

[22] Cruz-Monteagudo, M.; Borges, F.; Cordeiro, M. N. D. S.; Cagide Fajin, J. L.; Morell, C.; Ruiz, R. M.; Cañizares-Carmenate, Y.; Dominguez, E. R. Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries. *J. Comb. Chem.,* **2008**, *10*, 897-913.

[23] Lameijer, E.-W.; Kok, J. N.; Bäck, T.; Ijzerman, A. P. The molecule evoluator. An interactive evolutionary algorithm for the design of drug-like molecules. *J. Chem. Inf. Model,* **2006**, *46*, 545-552.

[24] Kruisselbrink, J.; Aleman, A.; Emmerich, M.; IJzerman, A.; Bender, A.; Baeck, Th.; Horst, E.v.d. Enhancing search space diversity in multiobjective evolutionary drug molecule design using niching. *Int. ACM Conf. GECCO,* **2009**, 217-224

[25] Kutchukian, P. S.; Shakhnovich, E. I. *De novo* design: balancing novelty and confined chemical space. *Expert Opin. Drug Discov.,* **2010**, *5*, 789-812.

[26] Loving, K.; Alberts, I.; Sherman, W. Computational approaches for fragment-based and de novo design. *Curr. Top Med. Chem.,* **2010**, *10*, 14-32.

[27] Hartenfeller, M.; Schneider, G. De novo drug design. *Methods Mol. Biol.,* **2011**, *672*, 299-323.

[28] Clark, D. E.; Westhead, D. R. Evolutionary algorithms in computer-aided molecular design. *J. Comput. Aided Mol. Des.,* **1996**, *10*, 337-358.

[29] Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP-Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Model,* **1998**, *38*, 511-522.

[30] Krueger, B. A.; Dietrich, A.; Baringhaus, K.-H.; Schneider, G. Scaffold-hopping potential of fragment-based de novo design: the chances and limits of variation. *Comb. Chem. High Throughput Screen,* **2009**, *12*, 383-396.

[31] Glen, R. C.; Payne, A. W. A genetic algorithm for the automated generation of molecules within constraints. *J. Comput. Aided Mol. Des.,* **1995**, *9*, 181-202.

[32] Pegg, S. C.-H.; Haresco, J. J.; Kuntz, I. D. A genetic algorithm for structure-based de novo design. *J. of Comput. Aided Mol. Des.,* **2001**, *15*, 911-933.

[33] Dey, F.; Caflisch, A. Fragment-based de novo ligand design by multiobjective evolutionary optimization. *J. Chem. Inf. Model,* **2008**, *48*, 679-690.

[34] Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. PRO-LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules. *J. Comput. Aided Mol. Des.,* **1995**, *9*, 13-32.

[35] Westhead, D. R.; Clark, D. E.; Frenkel, D.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B. PRO-LIGAND: an approach to de novo molecular design. 3. A genetic algorithm for structure refinement. *J. Comput. Aided Mol. Des.,* **1995**, *9*, 139-148.

[36] Pfeffer, P.; Fober, T.; Hüllermeier, E.; Klebe, G. GARLig: a fully automated tool for subset selection of large fragment spaces via a self-adaptive genetic algorithm. *J. Chem. Inf. Model,* **2010**, *50*, 1644-1659.

[37] Douguet, D.; Thoreau, E.; Grassy, G. A genetic algorithm for the automated generation of small organic molecules: drug design using an evolutionary algorithm. *J. Comput. Aided Mol. Des,* **2000**, *14*, 449-466.

[38] Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. LEA3D: A Computer-Aided Ligand Design for Structure-Based Drug Design. *J. Med. Chem.,* **2005**, *48*, 2457-2468.

[39] Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.,* **1988**, *28*, 31-36.

[40] Wang, R.; Gao, Y.; Lai, L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *J. Mol. Modeling,* **2000**, *6*, 498-516.

[41] Feher, M.; Gao, Y.; Baber, J. C.; Shirley, W. A.; Saunders, J. The use of ligand-based de novo design for scaffold hopping and sidechain optimization: two case studies. *Bioorg. Med. Chem.,* **2008**, *16*, 422-427.

[42] EA-Inventor software application, Tripos Inc, http://www.tripos.com. (Accessed March 14, 2011).

[43] Damewood, J. R., Jr; Lerman, C. L.; Masek, B. B. NovoFLAP: A ligand-based de novo design approach for the generation of medicinally relevant ideas. *J. Chem. Inf. Model,* **2010**, *50*, 1296-1303.

[44] Huang, Q.; Li, L.-L.; Yang, S.-Y. PhDD: a new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *J. Mol. Graph. Model,* **2010**, *28*, 775-787.

[45] Fischer, J. R.; Lessel, U.; Rarey, M. LoFT: similarity-driven multiobjective focused library design. *J. Chem. Inf. Model,* **2010**, *50*, 1-21.

[46] Nicolotti, O.; Gillet, V. J.; Fleming, P. J.; Green, D. V. S. Multiobjective optimization in quantitative structure-activity relationships: deriving accurate and interpretable QSARs. *J. Med. Chem.,* **2002**, *45*, 5069-5080.

[47] Cruz-Monteagudo, M.; Borges, F.; Cordeiro, M. N. D. S. Desirability-based multiobjective optimization for global QSAR studies: application to the design of novel NSAIDs with improved analgesic, antiinflammatory, and ulcerogenic profiles. *J. Comput. Chem.,* **2008**, *29*, 2445-2459.

[48] Saiz-Urra, L.; Bustillo Pérez, A. J.; Cruz-Monteagudo, M.; Pinedo-Rivilla, C.; Aleu, J.; Hernández-Galán, R.; Collado, I. G. Global antifungal profile optimization of chlorophenyl derivatives against Botrytis cinerea and Colletotrichum gloeosporioides. *J. Agric. Food Chem.,* **2009**, *57*, 4838-4843.

[49] Machado, A.; Tejera, E.; Cruz-Monteagudo, M.; Rebelo, I. Application of desirability-based multi(bi)-objective optimization in the design of selective arylpiperazine derivates for the 5-HT1A serotonin receptor. *Eur. J. Med. Chem.,* **2009**, *44*, 5045-5054.

[50] Brown, N.; McKay, B.; Gilardoni, F.; Gasteiger, J. A Graph-Based Genetic Algorithm and Its Application to the Multiobjective Evolution of Median Molecules. *J. Chem. Inf. Model.,* **2004**, *44*, 1079-1087.

[51] Brown, N.; McKay, B.; Gasteiger, J. The de novo design of median molecules within a property range of interest. *J. Comput. Aided Mol. Des.,* **2004**, *18*, 761-771.

[52] Brown, N.; McKay, B.; Gasteiger, J. A novel workflow for the inverse QSPR problem using multiobjective optimization. *J. Comput. Aided Mol. Des.,* **2006**, *20*, 333-341.

[53] Nicolaou, C. A.; Kannas, C. C. Molecular library design using multi-objective optimization methods. *Methods Mol. Biol.,* **2011**, *685*, 53-69.

[54] Shin, S.-Y.; Lee, I.-H.; Cho, Y.-M.; Yang, K.-A.; Zhang, B.-T. EvoOligo: oligonucleotide probe design with multiobjective evolutionary algorithms. *IEEE Trans. Syst. Man Cybern B Cybern,* **2009**, *39*, 1606-1616.

[55] Ekins, S.; Honeycutt, J. D.; Metz, J. T. Evolving molecules using multi-objective optimization: applying to ADME/Tox. *Drug Discov. Today,* **2010**, *15*, 451-460.

[56] Ecemis, M. I.; Wikel, J.; Bingham, C.; Bonabeau, E. A Drug Candidate Design Environment Using Evolutionary Computation. *IEEE Trans. Evol. Computat.,* **2008**, *12*, 591-603.

[57] Kruisselbrink, J. In *Evolutionary Multi-Criterion Optimization*; Lecture Notes in Computer Science; Springer Berlin/Heidelberg, **2009**; Vol. *5467*, pp. 453 - 467.