ELSEVIER

Drug Discovery Today: Technologies

Vol. xxx, No. xx 2013

*Editors-in-Chief*
**Kelvin Lam** – Blue Sky Biotech, Inc., Worcester, MA
**Henk Timmerman** – Vrije Universiteit, The Netherlands

DRUG DISCOVERY TODAY TECHNOLOGIES

Informatics, docking and data investigation

# Multi-objective optimization methods in drug design

Christos A. Nicolaou[1,*], Nathan Brown[2]

[1]Lilly Research Laboratories, Drop Code 1930, Eli Lilly & Company, Indianapolis, IN 46285, USA
[2]Cancer Research UK Cancer Therapeutics Unit, Division of Cancer Therapeutics, The Institute of Cancer Research, 15 Cotswold Road, Sutton, London SM2 5NG, UK

**Drug discovery is a challenging multi-objective problem where numerous pharmaceutically important objectives need to be adequately satisfied for a solution to be found. The problem is characterized by vast, complex solution spaces further perplexed by the presence of conflicting objectives. Multi-objective optimization methods, designed specifically to address such problems, have been introduced to the drug discovery field over a decade ago and have steadily gained in acceptance ever since. This paper reviews the latest multi-objective methods and applications reported in the literature, specifically in quantitative structure–activity modeling, docking, *de novo* design and library design. Further, the paper reports on related developments in drug discovery research and advances in the multi-objective optimization field.**

**Section editor:**
C. Ottman – Max Planck Society, Dortmund, Germany.

## Introduction

Research in multi-objective optimization (MOOP) technologies has experienced a major resurgence during the last two decades mostly because of the pressing needs of numerous scientific communities dealing with problems searching for solutions when multiple objectives exist [1]. MOOP technology was introduced to drug discovery informatics more than a decade ago and has since steadily gained acceptance. Its introduction challenged widely accepted practices aiming to optimize pharmaceutically relevant objectives one at a time, typically starting with potency. Reviews on the topic first appeared in [2] describing the use of MOOP in quantitative structure–activity relationship (QSAR) models, molecular docking, molecular library design and *de novo* design among others. Following those early attempts, research in the domain has produced numerous examples of methods and applications developed specifically to address the needs of the drug discovery researcher. This paper builds on previous work by the authors to review recent literature and document the main achievements in the field in recent years. *Multi-objective optimization and chemoinformatics* section introduces fundamental MOOP concepts, recent methodological advances and briefly discusses earlier applications and prior reviews in drug discovery. The following sections review a representative selection of MOOP methodology applications in chemoinformatics. The final section of the paper presents our conclusions and comments on MOOP and its future in the field.

## Multi-objective optimization and chemoinformatics

Multi-objective problems (MOP) are characterized by the presence of two or more primary objectives that solutions must satisfy. The score of each solution in a MOP consists of a vector whose dimensionality is equal to the number of objectives considered. In contrast to single-objective

*Corresponding author.: : C.A. Nicolaou (c.nicolaou@lilly.com), N. Brown (nathan.brown@icr.ac.uk)

1740-6749/$ © 2013 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.ddtec.2013.02.001 

Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

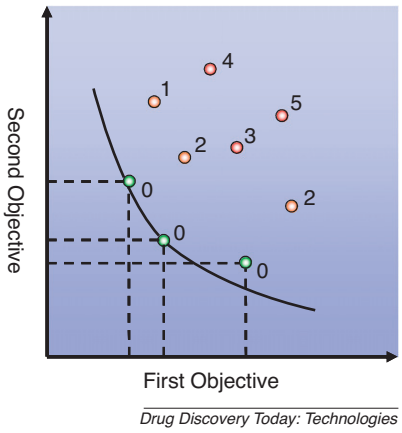


**Figure 1.** Identifying optimal solutions in the presence of multiple objectives requires the identification of non-dominated solutions that compromise best the objectives considered. The label to each solution refers to the number of other solutions that dominate them, that is, that have better scores in all objectives considered. Non-dominated solutions are labeled with zero.

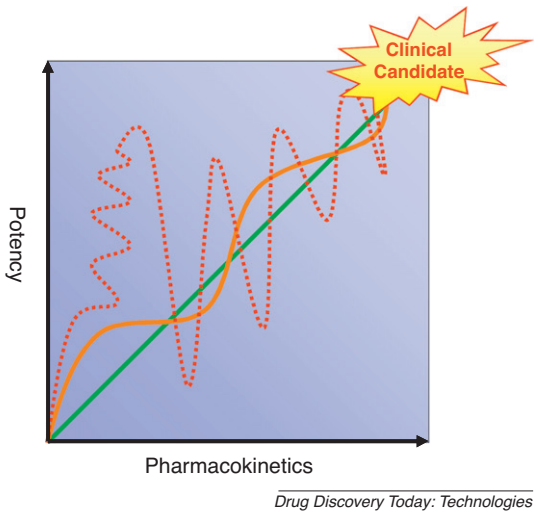


**Figure 2.** Sequential single-objective optimization of conflicting objectives often results in rounds of trial and error and a waste of resources and time (dashed line). Multi-objective optimization methods simultaneously search the solution space for satisfactory compromises to all objectives thus enabling faster convergence (bold continuous line) (adapted from [3]).

problems, where ranking candidate solutions is trivial (i.e. simple sort on the scalar objective value), ranking the candidates in a MOP is more challenging because of conflicting objective values. MOPs have multiple equivalent solutions, the so-called Pareto or non-dominated solutions, which represent different compromises of the various objectives (see Fig. 1). Pareto-ranking is the process of determining the rank of each solution through identifying the number of other solutions that dominate it, that is, the number of solutions that have better scores in all objectives considered. Pareto solutions have rank 0; all other solutions are given as rank the number of other solutions that dominate them in all objectives.

Traditionally, MOPs have been simplified by considering only one of the present objectives, typically, the one considered most important such as potency for drug discovery problems. This led to longer optimization cycles that attempted to address each objective sequentially often realizing that decisions taken earlier produced solutions that failed to meet the criteria imposed later and therefore required revisiting [3]. This realization led to the design of optimization techniques specifically developed to address MOPs by addressing all of the objectives simultaneously, the so-called multi-objective optimization methods (MOOP). In drug discovery, candidate molecules are modeled in multiple objectives and novel chemical entities subsequently prioritized for synthesis (Fig. 2).

Initially, the MOOP method of choice was to combine all objectives into one, for example, by a weighted sum approach, and transform the problem to a single objective one [2]. This aggregation approach enables the use of existing optimization techniques designed to address single objective problems at the expense of producing a single solution of the Pareto set which may or may not be appropriate for all objectives. More recent implementations of this approach use desirability functions to define a candidate solution desirability index based on the individual objective scores and, thus, reduce the complexity of the problem [4]. Pareto-based methods where the aim is to obtain the set of solutions on the trade-off surface between all objectives in a single run have also been introduced. The benefit of these methods is found in improved overall efficiency and in avoiding local minima related to each objective [2]. The challenge they face is the identification of a set of solutions effectively representing the true Pareto set. Interactive methods represent an alternative approach that requires user intervention during the optimization process. In effect, these methods use expert users in the place of the ranking and selection mechanisms used by fully automated techniques.

The presence of numerous objectives typically results in large complex solution spaces for MOPs. Consequently, most optimization methods used in MOP research resort to (meta-)heuristics that can provide a sufficiently good solution set in a reasonable amount of time. Most commonly, iterative, population-based approaches are used, for example, Evolutionary Algorithms (EA) [1].

Early implementations of Pareto-based methods were simple extensions of single objective optimization techniques complemented with Pareto-ranking. For example, in EAs, where solutions are identified through an iterative process combining individual fitness assessment, selection and reproduction, multiple objectives were handled by the addition of



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

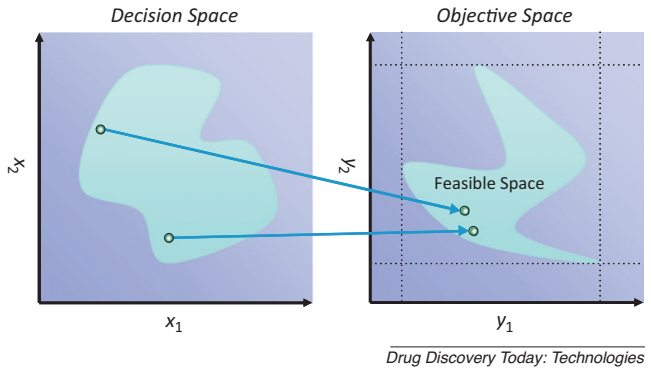


**Figure 3.** Solution diversity in decision and objective spaces: Solutions (represented by circles) of a multi-objective problem are mapped from decision space to objective space through objective functions. Proximity of solutions in objective space is not related to solution similarity in the decision space and vice versa.

a Pareto-ranking step. Intense research efforts have since led to several improvements. Secondary populations known as Pareto-archives, have been introduced in an effort to ensure that no non-dominated solution is lost during the optimization process [5]. Niching techniques aiming at the preservation of solution diversity in objective space have also been widely used [6]. More recently, the issue of population diversity in parameter space has been gaining attention because several scientific domains, including drug discovery, are interested in producing solutions that differ [7] (Fig. 3).

In the field of chemoinformatics MOOP techniques were adopted early on because of the requirements of drug discovery. Drugs are essentially molecules with a pharmacological profile that compromises numerous relevant objectives including potency, selectivity, pharmacokinetics and toxicity [8]. In line with the evolution of the field the first applications used composite MOOP methods and, soon after, Pareto-based methods using algorithms developed for other fields. Since then several applications spanning nearly all domains of computational chemistry and chemoinformatics have been reported that feature more advanced methods, often custom designed to drug discovery process requirements. A detailed overview of recent MOOP technology applications in this field is presented below with some emphasis placed on Pareto-based methods. A review on earlier MOOP applications in the field can be found by the same authors in [2] and also in [9] among others. Table 1 summarizes selected, representative applications.

## Quantitative structure–activity relationships

Since first proposed in the 1960s, Quantitative Structure–Activity Relationship (QSAR) models have gained widespread

**Table 1. Summary table.**

| Reference | MO type | Search method | Remarks/objectives |
|---|---|---|---|
| **Quantitative structure–activity relationships** | | | |
| Nicolotti *et al.* (2009) | Pareto-based | Evolutionary (genetic) algorithm | Selection of docking poses for 3D QSAR model construction |
| MO-QSPR: Manoharan *et al.* (2010) | Desirability, aggregative | Expert-driven search, desirability plot | QSAR and QSSR models |
| Soto *et al.* (2009) | Pareto-based and aggregative | Evolutionary (genetic) algorithm | Feature selection for QSAR model construction |
| Hii *et al.* (2011) | Pareto-based | Evolutionary (genetic) programming | *In silico* toxicity prediction; 'goodness-of-fit', model complexity |
| MTAR: Bajorath *et al.* (2012) | Visualization | Interactive, expert-driven visual inspection | Multiple target activity |
| DAD/TAD: Medina-Franco *et al.* (2011) | Visualization | Interactive, expert-driven visual inspection | Two or three target activities |
| **Docking** | | | |
| Mardikian *et al.* (2007) | Pareto-based | Evolutionary (genetic) algorithm | Protein-ligand bonding; van der Waals interactions |
| Boisson *et al.* (2008) | Pareto-based | Evolutionary (genetic) algorithm | Energy, surface |
| MOSFOM: Li *et al.* (2009) | Pareto-based | Evolutionary (genetic) algorithm | Energy, contact |
| **De novo design and inverse QSAR** | | | |
| MEGA: Nicolaou *et al.* (2009) | Pareto-based | Evolutionary (genetic) algorithm | Ligand and target based objectives |
| PLD: Ekins *et al.* (2010) | Pareto-based | Evolutionary (genetic) algorithm | Predicted activity, ADME related properties |
| Molecule Commander: van der Horst *et al.* (2012) | Desirability and Pareto-based | Evolutionary (genetic) algorithms | High predicted affinity, pharmacophore model, predictive models |
| MOLig: Sengupta *et al.* (2012) | Pareto-based | Simulated annealing | Internal energy, target interaction, similarity to scaffold, oral bioavailability |
| **Library design** | | | |
| Sharma *et al.* (2011) | Weighted-sum, aggregative | Deterministic annealing | Diversity, representativeness of compounds |
| MEGALib: Nicolaou *et al.* (2011) | Pareto-based | Evolutionary (genetic) algorithm | Selectivity between related targets |



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

**Table 1 (*Continued*)**

| Reference | MO type | Search method | Remarks/objectives |
|---|---|---|---|
| Venhorst *et al.* (2010) | Pareto-based | Evolutionary (genetic) algorithm | Structural diversity, chemical tractability, etc. |
| MSDS: Meinl *et al.* (2011) | Composite, aggregative | Iterative, greedy search | Activity, representativeness of compounds |
| CNS MPO: Wager *et al.* (2010) | Desirability, aggregative | Multi-parameter optimization | Fundamental physicochemical parameters (6) |
| MO-PSO: Namasivayam *et al.* (2012) | Pareto-based | Particle Swarm Optimization | SAR discontinuity index; ADME-related properties |

acceptance by the drug discovery community. QSAR models associate molecular descriptors to biological properties using statistical techniques and/or computational intelligence algorithms. Typically, QSAR models have been used for interpretation purposes, that is, to identify structure–activity relations in the available data. A second use has been as predictive models, that is, for the prediction of the biological property of new, untested chemical structures. MOOP techniques have been used in QSAR modeling in the past decade starting with MoQSAR which considered several conflicting objectives, including model accuracy and complexity, using a multi-objective genetic programming method and Pareto ranking [10].

In [11] a multi-objective genetic algorithm for the selection of appropriate docking poses for building 3D QSAR models is proposed. The method optimizes two objectives, namely the correlation of docking scores to biological activity and the averaged root mean squared deviation with respect to a defined co-crystallized inhibitor. The set of equivalent QSAR models produced are subsequently clustered and a representative set is selected. The method was successfully applied to the analysis of a well-known dataset that the authors had previously examined using other approaches. The results showed that the method was very effective overall and, in particular, the generation of trade-off 3D QSAR models with good sensitivity [11].

Manoharan *et al.* propose the multi-objective QSPR (MO-QSPR), which in combination with FB-QSAR, a fragment based QSAR methodology, is used for data interpretation and the design of new compounds [12]. The method is based on the preparation of QSAR models, the calculation of individual desirability functions and their combination via geometric mean. Next, a desirability plot is used to identify the optimal value ranges of descriptors that have major influence on predicted responses and, thus, support the interpretation of the available biological data. The authors also used the results of MO-QSPR to select specific fragments with optimal descriptor values and suggest promising analog designs that compromise activity and selectivity.

MO technology has also been applied to the problem of feature selection (FS) for QSAR models in [13]. The authors propose a two-step approach, consisting of the MO wrapper step, which performs feature searching and feature subset evaluation, and the validation step. In the MO wrapper step, a collection of multi-objective techniques, including aggregation and Pareto-based methods are used for feature subset searching. The authors report a detailed analysis of the proposed methods and an extensive comparison with other methods reported in the literature to conclude that the utilization on MO technology in either form is advisable for the problem of FS in the QSAR field.

Hii *et al.* presented a multi-objective method for the *in silico* prediction of toxicity [14]. The method uses a multi-objective Genetic Programming (MOGP) implementation to maximize the 'goodness-of-fit' to data and minimize the model complexity. In particular, the method combines the multi-gene symbolic regression variation of genetic programming with a widely used MOEA. Results on a publicly available aqueous toxicity dataset demonstrate that the proposed approach generates compact QSAR models with comparable performance to models reported in the literature [14].

Several visual SAR analysis applications have been developed recently, initially for the identification of 'activity cliffs' (i.e. structurally similar compounds with significantly different biological activity on a specific target) [15], and later expanded to a higher order data structure, the so-called 'activity ridge', combining possible pairwise activity cliffs formed across subsets of structural analogs of varying potency [16]. While the initial focus of both activity 'cliffs' and 'ridges' has been single target SAR they have been modified to accommodate multiple activities such as those resulting from polypharmacology/multi-target drug discovery projects. A successful application to the analysis of large, publically available multi-target datasets has been reported in [17]. Gupta-Ostermann and Bajorath [16] introduced the Multi-Target Activity Ridge (MTAR) concept and applied it to the analysis of a high-dimensional kinase inhibitor data set. The authors report that the method successfully identified several MTARs in the dataset each consisting of numerous compounds with target differentiation potential. In a parallel effort, Medina-Franco *et al.* report on the development and application of the Dual and Triple Activity-Difference (DAD/TAD) maps that approach the activity landscape generation problem by first representing pairwise activity differences for all compounds in a data set and then overlaying structure similarity information [18]. The authors have applied this



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

technique for the characterization of the multi-target SAR of 299 diverse compounds on three monoamine transporters.

As evident by the numerous methods described above, the need for consideration of multiple objectives in QSAR has clearly been realized. MO techniques are used to select features in an effort to use the best possible subset of descriptors, and, to develop models compromising multiple objectives. Custom MO techniques are also developed to enable visual SAR analysis of multi-target datasets. The advent of polypharmacology and multi-target drug discovery seems to be further contributing to the application of MO technology in QSAR, a trend that can only increase given recent developments in the field [19].

## Docking

Given a protein structure of interest it is possible to predict potential binding modes of small molecules for which the binding modes are unknown. The optimization process explores the translation and rotation of ligand conformations in addition to the potential conformers that the ligand may adopt in three-dimensional (3D) space. The search space of possible solutions is vast and the methods by which the poses are scored are important to effectively search that space and appropriately rank the results. MOOP approaches are applied to the scoring and ranking of each of the potential docking poses. The benefit of applying MOOP is to reach a family of global optima more rapidly and with results that more appropriately consider the parameters that are important for scoring binding modes.

Each docking pose may be scored using a multitude of parameters. Indeed, many of the commonly applied scoring functions are combinations of parameters oftentimes weighted according to empirical scores of significance. The original scoring function in GOLD (Genetic Optimization for Ligand Docking), GoldScore, includes several terms, including: van der Waals (vdW) interaction between the protein and ligand atoms; the internal energy of the ligand (vdW and torsion, with internal H-bonding if required) and the H-bond interaction energy between the protein and the ligand, giving:

$$\text{GoldScore} = [\text{protein-ligand H-bond}] + (1.375 \times [\text{protein-ligand vdW}]) + [\text{ligand internal}]$$

However, these consider each of the terms as a combined function, for which the limitations have been covered already.

The non-dominated sorting genetic algorithm (NSGA) was developed by Deb (2000) [6] and later applied to drug discovery problems including ligand docking. NSGA-II is an approach to overcome the computational complexity of many other multi-objective algorithms that tend to $O(MN^3)$. There is also a recognized lack of elitism strategies and the need to specify a sharing parameter to ensure an even distribution of equivalent solutions. Mardikian demonstrated that applying the vdW interactions and the combined electrostatic and hydrogen bonding between the protein and the ligand in a bi-objective system (S. Mardikian, PhD Thesis, The University of Sheffield, 2007). The bi-objective NSGA-II algorithm applied in this study re-capitulated experimental crystallography in several test cases of varying difficulty. The work found that consideration of the docking results could assist in understanding how the terms can affect docking success. The key finding was the interpretation of which parameters are most significant in docking for different systems thereby permitting the optimization of the docking methodology in the diagnostic mode that could then be used to apply term weightings to be used in the prognostic mode.

The work of Boisson *et al.* [20] again apply a bi-objective scoring function in a genetic algorithm for ligand docking, this time using terms for energy and surface. This work uses the NSGA-II algorithm as before, but also considers the Indicator-Based Evolutionary Algorithm (IBEA) [21]. IBEA was found to out-perform NSGA-II in this application, which confirms results from other applications of IBEA in the literature.

More recently, Li *et al.* [22] reported a multi-objective docking algorithm, Multi-Objective Scoring Function Optimization Methodology (MOSFOM), combining two scores: energy and contact from the DOCK program [23]. The method uses Pareto ranking and is demonstrated to out-perform single scoring functions, particularly in the enrichment of actives in the top 2% of recalled compounds.

The development of appropriate scoring functions for small molecule docking challenges remains an active area of research. Several approaches have investigated multi-objective scoring functions. Through the optimization of families of solutions that populate the trade-off surface in solution space, it is possible to more appropriately select optimal solutions. Multi-objective methods obviate the need for weightings in the functions that, because they are derived empirically, are not appropriate for all model cases. It can only be expected that multi-objective scoring functions will be the focus of continued interest in ligand docking to further improve docking performance.

## *De novo* design

*De novo* design is an effort to design chemical compounds using knowledge on the pharmaceutical target site or its available ligands [2]. To design products that are truly of interest to the drug discovery researcher several additional, pharmaceutically important objectives should be taken into account, including selectivity to the specific target, pharmacokinetic properties, synthetic accessibility, etc. Following the introduction of MOOP technology in [24] several new methods have appeared in the literature.



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

Nicolaou *et al.* described MEGA (Multi-objective Evolutionary Graph Algorithm), a method which combines evolutionary algorithms with local search techniques [7]. MEGA initially applies all available objectives on an initial set of compounds to obtain a list of scores for each individual. Individual Pareto-rank and population diversity are calculated with the latter given special consideration through the implementation of a clustering process operating in parameter space, directly on the chemical structures. Parent solutions are sampled from each cluster using individual ranking thereby ensuring the selection of diverse molecules. New solutions are generated using graph-specific mutation and crossover. MEGA also introduced the use of Pareto archive to preserve all promising solutions found throughout evolution [5]. Experimental results on the design of selective Estrogen Receptors have shown the ability of the method to generate a diverse set of compromising solutions covering the Pareto-front [7]. Kruisselbrink *et al.* also addressed the issue of diversity in parameter space through the use of a crowding operator based on compound similarity measurements to evolve structurally diverse niches of molecules [25].

A similar approach is followed by Ekins *et al.* who report on the development of the Pareto Ligand Designer (PLD) [26]. The method takes as input a set of reference molecules and calculates objective values that are used for defining individual Pareto-ranks. Non-dominated solutions are identified and stored in a Pareto-archive. In a next step, the working population is evolved using an extensive set of molecular transformations and the molecules that survive a number of property filters form a new working population. The process then iterates to calculate Pareto-rank and non-dominated solutions and continues until some completion criteria are met. The authors present results from the application of PLD to optimization experiments of two, three and four objectives, while maintaining biological activity [26].

Recently van der Horst *et al.* described a MOEA method they call Molecule Commander and applied it to generate potential A1 adenosine receptor antagonists [27]. The method used multiple objectives including a pharmacophore model and support vector machine models based on molecular fingerprints for subtypes of adenosine receptor. The objectives were grouped into logical groups by means of desirability indexes to reduce the dimensionality of the problem. Pareto ranking was used in combination with niching to select the compounds to be used as input for the next generation cycle. It is worth pointing out that the niching process used employs clustering in parameter space to ensure the selection of diverse chemical structures. Similar to the methods described above, new structures are developed using molecular perturbation techniques. Thresholds on physicochemical properties as well as chemical rules were used to as hard filters to remove compounds with an undesired profile.

In [28], MOLig, a method using multi-objective simulated annealing is presented. MOLig represents solutions using a tree-like structure and employs a collection of solution perturbation operators such as substitution, deletion, extension and shuffling to generate new chemical designs. In the cited example, MOLig was used to simultaneously optimize the internal energy of a candidate, its interaction energy with the given target, its chemical similarity to a reference scaffold and its oral biovailability. The authors report that the method was compared to several *de novo* design methods and was found to outperform them [28].

DND has long been criticized for generating unrealistic and unattractive chemical designs, that is, impossible to synthesize or with obvious pharmaco-chemical issues. As a response the DND community adopted MO technology early on to incorporate additional pharmaceutical objectives and, thus, produce compounds with an acceptable overall chemical profile. This trend has been steadily increasing in recent years and numerous innovative methods have been introduced with varying success. Currently, the use of MO technology and the inclusion of multiple objectives in the design process in one form or another is the norm; it is our firm belief that this will remain the case in the future with added emphasis placed on new objectives relating to synthesizability and chemical novelty.

## Library design

Typically, molecular modelers and medicinal chemists have far more design ideas than they could possibly realize in synthesis. Historically, exploration of the virtual libraries without enumeration has essentially proceeded as a local search, making small changes. It is important to however consider the space of compounds that could be synthesized to understand what modifications may result in beneficial endpoints, such as potency, anti-targets, toxicity, solubility and many more.

Computational molecular library design (MLD) aims at designing a collection of chemical products given a starting set of chemical structures and a set of objectives reflecting the desired characteristics of the resulting library. The method has two variations. In the first, the starting set of chemical structures consists of chemical reagents and rules to guide the virtual synthesis of chemical products (see Fig. 4). Alternatively, the initial set may consist of large libraries of real or virtual compounds. A typical goal in both cases is diversity, that is, obtaining a collection of products representative of the chemical space accessible by the input provided. Other common objectives include QSAR and QSPR models and drug likeness, often imposed to focus the resulting library to specific regions of the chemical space more relevant to the targeted drug discovery problem. A frequent problem is the possible generation of far too many virtual products that meet the imposed objectives. Sampling methods and



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

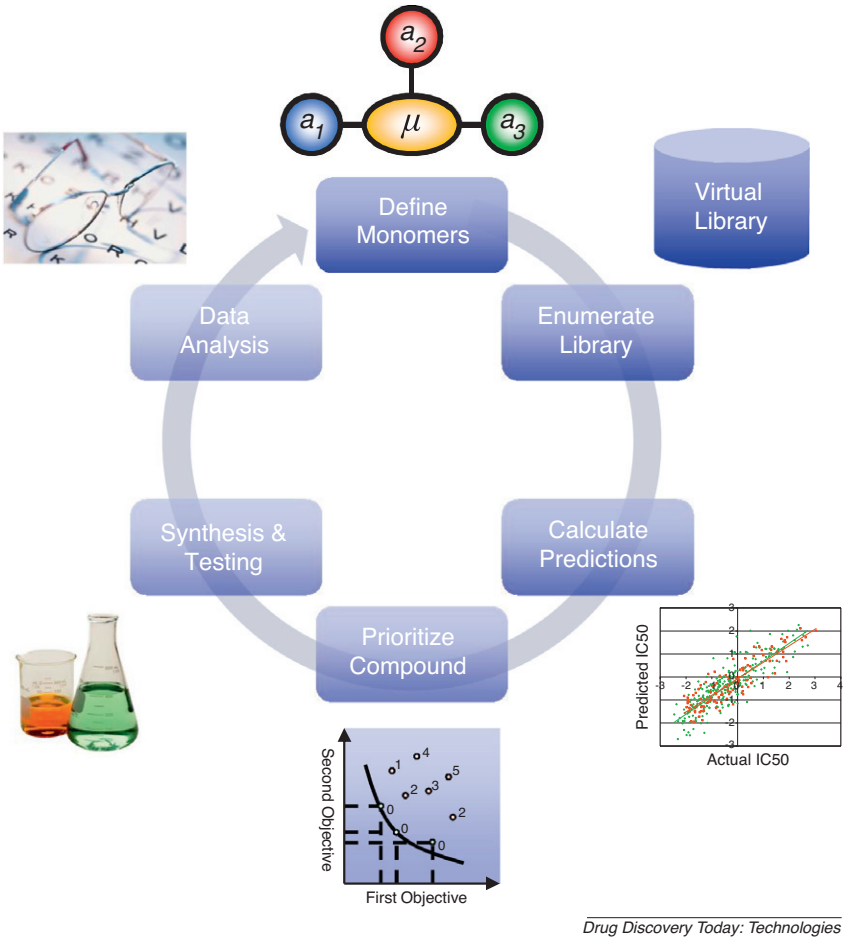


**Figure 4.** Schematic workflow of virtual library design and enumeration of compounds to be prioritized for synthesis. Predictions are made from relevant models and these values prioritized using a multi-objective prioritization strategy such as weighted-sum or Pareto ranking. The prioritized compounds may then be synthesized and tested and the results analyzed before cycle beginning again.

physicochemical filters can be used to reduce the number appropriately.

Multi-objective methods are often used for library design to balance the numerous objectives with the majority of them following the weighted-sum approach [2]. Pareto-based methods were first introduced with MoSELECT [29] which aimed to simultaneously address multiple objectives such as diversity, physicochemical properties and ease of synthesis. The method was proven effective at designing combinatorial libraries consisting of families of equivalent solutions for both diverse and focused library designs [29].

More recently, Sharma *et al.* proposed an algorithm that addresses simultaneously the two objectives of diversity and representativeness of compounds in the resulting library [30]. The method applies a deterministic annealing algorithm to identify clusters and truncate computations over the entire dataset to computations over individual clusters. Experimental results on lead-generation design problems show significant improvement and demonstrate the efficiency of the method. The authors have also performed an extensive analysis to quantify the trade-off between the error because of truncation and the computational effort [30].

Nicolaou *et al.* presented an extension to the MEGA *de novo* design algorithm, termed MEGALib, custom-designed to the library design problem [31]. The method uses MEGA but only takes fragments as building blocks and, by default, applies a set of well defined chemical rules for molecular synthesis. The fragments used by MEGALib contain information on reaction points to facilitate virtual synthesis and may be profiled using results from previous biological screening experiments to assign weights taken into account during virtual synthesis. The authors describe an application for the design of a focused library consisting of a diverse set of compounds meeting objectives associated with selectivity.

Venhorst *et al.* employed multi-objective optimization in the design of a diverse, high-efficiency fragment library from existing sources [32]. The authors generated pharmacophore models on a variety of targets and used them to select fragments from a virtual library of enumerated organic



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

molecules. Several properties of the candidate scaffolds were computed including structural diversity, chemical tractability and scaffold redundancy and Pareto-based optimization was used to select the final fragment library. The authors report usage of the library in virtual screening experiments and the identification of novel entities with high similarity to known inhibitors of the same target [32].

Meinl *et al.* present an interesting study of computational multi-objective methods that aim to select compounds from existing sources that are both diverse and, additionally, take molecular properties (e.g. activities) into account [33]. They propose the Maximum-Score Diversity Selection (MSDS) method which uses the 'erosion' concept to iteratively calculate a composite score for each individual molecule which, in turn, is used to determine the next selection of compounds. In the first iteration, the score simply consists of the activity of the molecules and in each of the following iterations scores of the remaining compounds are re-calculated and 'eroded' in proportion to the distance of each molecule to the ones already selected. The authors compare MSDS with implementations of various methods proposed in the literature, including a Pareto-based MOEA, on bi-objective problems. Their conclusions indicate that the proposed method is computationally far more efficient but at the same time produces comparable results to the MOEA.

Wager *et al.* present a new multi-parameter optimization (MPO) method which uses six fundamental physicochemical parameters to aid in the design of libraries for central nervous system (CNS) research [34]. The method, termed CNS MPO, constructs a desirability score for each of the six properties and a composite desirability score using summation. Desirability score ranges of optimal, less optimal, and undesirable values for CNS agents are defined for each property based on expert chemist input and validated on known CNS agents. CNS MPO attempts to balance the properties considered and account for the comprehensive profile of a compound. The authors report on tests that indicate that the method may be used as a design tool to identify compounds with a higher probability of success [34].

Recently, Namasivayam and Bajorath described the application of a multi-objective Particle Swarm Optimization (MOPSO) method to identify compound subsets from a larger set meeting several criteria [35]. The method combined PSO, a population-based EA that mimics the movement of birds in a flock [36], with Pareto ranking to optimize the selection of compound subsets. The objective functions used include a SAR discontinuity index derived from known activity values of the dataset under investigation and functions based on ADME-related molecular properties. Experiments on 10 datasets with varying SAR information content demonstrated that compound subsets with desired SAR and property distributions subsets can be extracted using the proposed method, provided that such compounds exist in the original dataset [35].

Library design has probably experienced the larger penetration of MO technology in recent years with an increasing number of publications describing MLD applications taking into account multiple objectives. As the body of pharmaceutical knowledge has increased to include multiple target activities for specific compound sets and numerous predictive ADME models, MLD methods have evolved accordingly. This evolution will surely continue as ongoing developments will only make available more pharmaceutical data and computational models that can improve the quality of molecular libraries if properly used.

## Conclusions

Drug discovery is inherently a multi-objective problem although, in previous years, it has been addressed in a sequential one-objective-at-a-time manner both experimentally and *in silico*. Recognition of this attribute, supported by technological advances in recent years, has caused a paradigm shift to commonly consider multiple pharmaceutical endpoints in parallel from the beginning of pharmaceutical projects. Nowadays, MOOP methodologies are frequently used [8,9,26] and whole projects aiming for a new type of drugs, those with a designed polypharmacology profile are initiated [19,37].

In our previous review over five years ago we stated in our conclusions that 'we anticipate that in the near future the pace of MOOP methodology adoption will increase considerably, and that standard chemoinformatics techniques will regularly be applied with the simultaneous optimization of multiple properties of interest to the discovery of new drugs' [2]. Clearly, the drug discovery community is still experimenting with the technology and more custom methods and applications need to be developed. However, major advancements have already been made and the methodology has proven its usefulness to a significant portion of researchers in the field. At an institutional level, a paradigm shift is already ongoing leading to a new, multi-objective optimization approach as indicated by the multi-target drug discovery initiative. The development of new tools to assist molecular modelers and medicinal chemists select their most appropriate candidates for synthesis has also begun and can only be of assistance to these scientists in understanding more fully the trade-off surfaces that are being explored in their design ideas.

We firmly believe that this path forward will continue in the coming years and that additional achievements, especially in newer more challenging drug discovery fields, will continue to be reported.

## Conflict of interest

The authors have no conflict of interest to declare.



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/

## Acknowledgements

CN would like to thank Christine Humblet (Eli Lilly & Co.) for reviewing the manuscript and providing useful feedback. NB is funded by Cancer Research UK Grant no. C309/A8274.

## References

1 Yann, C. and Siarry, P., (eds) (2004) *Multiobjective Optimization: Principles and Case Studies*, Springer-Verlag
2 Nicolaou, C.A. *et al.* (2007) Molecular optimization using computational multi-objective methods. *Curr. Opin. Drug Discov. Dev.* 10, 316–324
3 Baringhaus, K-H. and Matter, H. (2004) Efficient strategies for lead optimization by simultaneously addressing affinity, selectivity and pharmacokinetic parameters. In *Chemoinformatics in Drug Discovery* (Oprea, T., ed.), pp. 333–379, Wiley-VCH
4 Cruz-Monteagudo, M. *et al.* (2008) Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries. *J. Comb. Chem.* 10, 897–913
5 Zitzler, E. and Thiele, L. (1999) Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Trans. Evol. Comput.* 3, 257–271
6 Deb, K. *et al.* (2000) A fast elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6, 182–197
7 Nicolaou, C.A. *et al.* (2009) De novo drug design using multi-objective evolutionary graphs. *J. Chem. Inf. Model.* 49, 295–307
8 Lusher, S.J. *et al.* (2011) A molecular informatics view on best practice in multi-parameter compound optimization. *Drug Discov. Today* 16, 555–568
9 Nicolotti, O. *et al.* (2011) Strategies of multi-objective optimization in drug discovery and development. *Expert Opin. Drug Discov.* 6, 871–884
10 Nicolotti, O. *et al.* (2002) Multiobjective optimization in quantitative structure–activity relationships: deriving accurate and interpretable QSARs. *J. Med. Chem.* 45, 5069–5080
11 Nicolotti, O. *et al.* (2009) Improving quantitative structure–activity relationships through multi-objective optimization. *J. Chem. Inf. Model.* 49, 2290–2302
12 Manoharan, P. *et al.* (2010) Rationalizing fragment based drug discovery for BACE1: insights from FB-QSAR, FB-QSSR, multi-objective (MO-QSPR) and MIF studies. *J. Comput. Aided Mol. Des.* 24, 843–864
13 Soto, A.J. *et al.* (2009) Multi-objective feature selection in QSAR using a machine learning approach. *QSAR Comb. Sci.* 28, 1509–1523
14 Hii, C. *et al.* (2011) Evolving toxicity models using multigene symbolic regression and multiple objectives. *Int. J. Mach. Learn. Comput.* 1, 30–35
15 Guha, R. and VanDrie, J.H. (2008) Structure–activity landscape index: identifying and quantifying activity cliffs. *J. Chem. Inf. Model.* 48, 646–658
16 Gupta-Ostermann, D. and Bajorath, J. (2012) Identification of multitarget activity ridges in high-dimensional bioactivity spaces. *J. Chem. Inf. Model.* 52, 2579–2586
17 Wassermann, A.M. *et al.* (2011) Comprehensive analysis of single- and multi-target activity cliffs formed by currently available bioactive compounds. *Chem. Biol. Drug Des.* 78, 224–228
18 Medina-Franco, J.L. *et al.* (2011) Multitarget structure–activity relationships characterized by activity-difference maps and consensus similarity measure. *J. Chem. Inf. Model.* 51, 2427–2439
19 Pujol, A. *et al.* (2010) Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol. Sci.* 31, 115–123
20 Boisson, J-C. *et al.* (2008) Parallel multi-objective algorithms for the molecular docking problem. arX@@iv:0811.0514 [q-bio.QM]
21 Zitzler, E. and Künzli, S. (2004) Indicator-based selection in multiobjective search. *Proceedings of the 8th International Conference on Parallel Problem Solving from Nature (PPSN VIII)* pp. 832–842, Springer
22 Li, H. *et al.* (2009) An effective docking strategy for virtual screening based on multi-objective optimization algorithm. *BMC Bioinform.* 10, 58
23 Ewing, T.J. *et al.* (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* 15, 411–428
24 Brown, N. *et al.* (2004) Graph-based. genetic algorithm and its application to the multi-objective evolution of median molecules. *J. Chem. Inf. Model.* 44, 1079–1087
25 Kruisselbrink, J. *et al.* (2009) Enhancing search space diversity in multi-objective evolutionary drug molecule design using niching. *GECCO 2009: Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation* pp. 217–224, ACM
26 Ekins, S. *et al.* (2010) Evolving molecules using multi-objective optimization: applying to ADME/Tox. *Drug Discov. Today* 15, 410–451
27 van der Horst, E. *et al.* (2012) Multi-objective evolutionary design of adenosine receptor ligands. *J. Chem. Inf. Model.* 52, 1713–1721
28 Sengupta, S. and Bandyopadhyay, S. (2012) De novo design of potential RecA inhibitors using multi-objective optimization. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 9, 1139–1154
29 Gillet, V.J. *et al.* (2002) Combinatorial library design using a multi-objective genetic algorithm. *J. Chem. Inf. Comput. Sci.* 42, 375–385
30 Sharma, P. *et al.* (2011) A scalable approach to combinatorial library design. In *Chemical Library Design, Methods in Molecular Biology*, (Vol. 685) (Zhou, J., ed.), pp. 71–89, Humana Press Inc.
31 Nicolaou, C.A. and Kannas, C. (2011) Molecular library design using multi-objective optimization methods. In *Chemical Library Design, Methods in Molecular Biology* (Zhou, J., ed.), pp. 53–69, Humana Press Inc.
32 Venhorst, J. *et al.* (2010) Design of a high fragment efficiency library by molecular graph theory. *ACS Med. Chem. Lett.* 1, 499–503
33 Meinl, T. *et al.* (2011) Maximum-score diversity selection for early drug discovery. *J. Chem. Inf. Comput. Sci.* 51, 237–247
34 Wager, T.T. *et al.* (2010) Moving beyond rules: the development of a central nervous system multi-parameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci.* 1, 435–449
35 Namasivayam, V. and Bajorath, J. (2012) Multi-objective Particle Swarm Optimization: automated identification of SAR-informative compounds with favorable physicochemical property distributions. *J. Chem. Inf. Model.* 52, 2848–2855
36 Kennedy, J. and Eberhart, R.C. (1995) Particle Swarm Optimization. *Proceedings IEEE International Conference Neural Networks IV (ICN95)* pp. 1942–1948, IEEE Conference Publications
37 Boran, A.D.W. *et al.* (2010) Systems approaches to polypharmacology and drug discovery. *Curr. Opin. Drug Discov. Dev.* 13, 297–309



Please cite this article in press as: Nicolaou, C.A. and Brown, N. Multi-objective optimization methods in drug design, Drug Discov Today: Technol (2013), http://dx.doi.org/