Applied Soft Computing 27 (2015) 543–552

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Evolutionary algorithms for *de novo* drug design – A survey

R. Vasundhara Devi [a], S. Siva Sathya [a], Mohane Selvaraj Coumar [b,*]

[a] Department of Computer Science, School of Engineering & Technology, Pondicherry University, Kalapet, Puducherry 605014, India
[b] Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Kalapet, Puducherry 605014, India

ARTICLE INFO

Article history:
Received 31 December 2013
Received in revised form 20 May 2014
Accepted 29 September 2014
Available online 7 October 2014

Keywords:
De novo drug design
Evolutionary algorithm
Genetic algorithm
Multi-objective optimization
Computer-aided drug design
Soft computing

ABSTRACT

The process of drug design and discovery demands several man years and huge investment. Computer-aided drug design (CADD) technique is an aid to speed up the drug discovery process. *De novo* drug design, a CADD technique to identify drug-like novel chemical structures from a huge chemical search space, helps to find new drugs by the optimization of multiple pharmaceutically relevant parameters required for a successful drug. As the search space is very large in the case of *de novo* drug design, evolutionary algorithm (EA), a soft computing technique can be used to find an optimal solution, which in this case is a novel drug. In this paper, various EA techniques used in *de novo* drug design tools are surveyed and analyzed in detail, with particular emphasis on the computational aspects.

1. Introduction

Discovering a drug is a lengthy and expensive process, taking over a decade and costing about 1 billion US dollars [1]. Particularly, the need to optimize multiple objectives simultaneously makes drug discovery a very challenging process. Traditionally, drugs have been discovered by searching through the natural products and synthetic chemical libraries of biologically active molecules. But nowadays, soft computing techniques are employed in the search for novel drugs resulting in smaller cost and reduced time [1,2].

Use of computers for drug design (computer-aided drug design, CADD) has resulted in several successfully marketed drugs, such as Captopril, Zanamivir, and Oseltamivi [3,4]. Particularly, the use of structure-based drug design (SBDD) technique is a popular and useful method, where drugs are designed *in silico* based on their interaction with the target protein in three dimensions. The *in silico* designed drugs, with optimized interaction with the target protein, are then synthesized in the lab and tested *in vitro*/*in vivo* for biological activity. The CADD process not only helps in minimizing the number of drugs actually synthesized and tested for identifying biologically active molecules, but also helps in identifying drugs which can be orally administered and has negligible toxicity. In summary, CADD helps to speed-up the overall drug discovery process and helps to find new drugs for diseases in a faster and cheaper way.

The term "*De novo*" is the latin word meaning "from the beginning", "afresh" or "anew". *De novo* drug design, a computer-aided drug design (CADD) technique, designs drug molecules from scratch, and hence the name "*De novo*". It provides better exploration of chemical space and helps in finding novel chemical structures for drug development, which cannot be found in any known chemical databases (natural products or synthetic chemical libraries). Hence the identified molecules/drugs will be unique in nature [5,6].

The first *de novo* drug design tool LEGEND was published in 1991, and currently over two dozens of such tools are known. The tools namely LigMerge [7], LEGEND [8], LUDI [9,10], SPROUT [11], BREED [12], and PRO_LIGAND [13] use structure-based drug design technique to derive new molecules. Among these tools, LEGEND uses atom by atom construction and the rest use fragment by fragment construction of the molecule/drug. The tools NEWLEAD [14] and PhDD [15] use ligand-based drug design technique to derive the new molecules. In addition to these tools, simulated annealing technique has been used in tools such as CONCERTS [16], SkelGen [17] and MOLig [18] using AMOSA [19]. Tools such as LigBuilder [20], LEA [21], ADAPT [22], PEP [23], SYNOPSIS [24], LEA3D [25], GANDI [26], TOPAS [27], Flux(1) [28], FLUX [29], MEGA [30] and EvoMD [31] use evolutionary algorithm for the drug design process, which forms the subject of this review.

De novo drug design requires optimization of several objectives simultaneously and independently to obtain an optimal solution. In this respect, evolutionary algorithm, a soft computing technique, is a problem-independent search technique, which could be applied to a variety of optimization problems from robot behavior to drug

* Corresponding author. Tel.: +91 413 2654950; fax: +91 413 2655211.
E-mail address: mohane@bicpu.edu.in (M.S. Coumar).

http://dx.doi.org/10.1016/j.asoc.2014.09.042
1568-4946/

design and discovery [2]. Hence, evolutionary algorithm has been used extensively for the past 15 years for finding the optimal solution in *de novo* drug design. This paper surveys the developments in the field of *de novo* drug design, with particular emphasis on the evolutionary algorithms involved in the design. Though there are few excellent reviews related to CADD [32–34], and the application of soft computing techniques in drug discovery process [2], they lack the coverage of all the *de novo* design tools available to date which utilizes evolutionary algorithms [5,6,35].

The rest of the paper is organized as follows: Section 2 describes the concept of *de novo* drug design, Sections 3–5 deal with evolutionary algorithm techniques and their types. Section 6 describes in detail the computational approaches used in various *de novo* drug design tools and Section 7 summarizes them. Finally, Section 8 concludes the paper.

2. *De novo* drug design

In *de novo* drug design, the drug molecules are constructed in two ways, using either atoms or molecular fragments [6,35]. When the drug is constructed by the atom based method, the chemical search space is huge and it is possible to arrive at a number of novel chemical structures. However, atom based method not only requires more time to construct these novel structures, but also results in a number of molecules which are not synthesizable.

Fragment based method, the preferred drug construction route in many *de novo* drug design tool is carried out by assembling fragments. The molecular search space is comparatively lesser and the solutions obtained have better ADME (Absorption, Distribution, Metabolism and Excretion) properties and notable diversity. *De novo* drug design is to generate novel molecular structure with desired biological activity and match the binding pattern of the biological target. Based on the binding pattern, there are two different *de novo* design techniques: Structure-based and ligand-based methods [6,35]. The overall classification of *de novo* drug design is shown in Fig. 1.

Structure-based *de novo* drug design develops new ligands using information from the three dimensional (3D) structure of a protein target without the prior knowledge of other ligands. A common structure-based design strategy is to construct the molecule directly in the binding site of the target protein and evaluate the quality of the designed molecule by calculating the interaction energy with the target protein. While, in the ligand-based *de novo* design, the protein target structure is not known and the new molecule is constructed based upon similarity to a known ligand molecule.

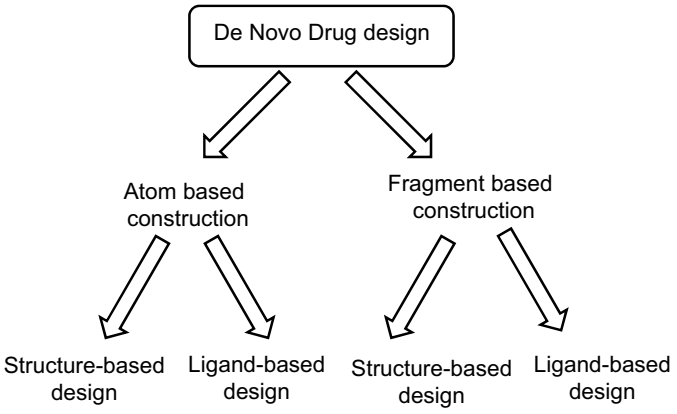

Fig. 1. Classification of *de novo* drug design techniques.

3. Evolutionary algorithm

An evolutionary algorithm (EA) is a population-based metaheuristic optimization algorithm inspired by the concepts of biological evolution such as reproduction, mutation, recombination and selection. They are randomized and stochastic techniques which simulate the evolutionary pressure of selection to select high fit individuals and use the evolutionary operators “crossover” and “mutation” to evolve better individuals. EAs generally start with a set (population) of random candidate solutions called individuals or chromosomes. In case of *de novo* drug design, chromosomes are represented by candidate chemical compounds. Evolutionary algorithms use the scoring function or the fitness function over the current population to determine the quality of the candidate solutions. Various selection methods such as “Roulette wheel”, “Tournament”, and “Rank” selection are used to populate the right candidates into the mating pool, where they are subjected to crossover and mutation. The crossover operator combines the genetic traits of two individuals (parents) and produces a new offspring. A mutation event introduces new piece of information into an existing population. Thus, the crossover and mutation operators provide variations among the current population and evolve better and better solutions. The evolution process continues till user-specified termination criterion.

4. Evolutionary algorithms for drug design

There are four major evolutionary algorithmic techniques. They are: genetic algorithm, genetic programming, evolutionary programming and evolutionary strategies. Apart from this classification, other swarm based algorithms [36] are also evolutionary in nature and get the idea from the collective behavior of Animals, Microbes, *etc.* Since most of the *de novo* drug design are based on genetic algorithm and evolutionary strategy, the following subsection deals with these concepts only.

4.1. Genetic algorithm

A genetic algorithm [37] is a search heuristic that mimics the process of natural selection and takes idea from Darwinian evolution. This is the most popular evolutionary algorithm technique. This heuristic is routinely used to generate useful solutions for optimization and search problems. The steps in GA are: initial population generation, fitness function evaluation, selection and breeding using genetic operations like crossover and mutation and check for termination conditions. Multi-dimensional optimization problems can be solved with the help of GAs where the different variables being optimized are encoded as bit vectors also termed as chromosome.

4.2. Evolutionary strategies

Evolutionary strategies (ES) is an EA technique using the adaptation and evolution concepts. In ES, the representation is problem-dependent. It uses mutation and selection in its iterations till a user-defined termination condition is reached. In contrast to GA, selection in ES is based on the fitness ranking instead of fitness values. The evolutionary strategy operates on the parent and the result of its mutants. The ES is (1,1)-ES when the parent is replaced by a comparatively better fit mutant. In (1,λ)-ES, λ mutants are generated and compete with the parent, wherein the best mutant becomes the parent of the next generation. As a result the current parent is disregarded.