iSyn: *De Novo* Drug Design with Click Chemistry Support

Hongjian Li
Dept of Computer Science and Engineering
Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
jackyleehongjian@gmail.com

Kwong-Sak Leung
Dept of Computer Science and Engineering
Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
ksleung@cse.cuhk.edu.hk

Chun Ho Chan
Dept of Computer Science and Engineering
Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
an_tony0522@yahoo.com.hk

Hei Lun Cheung
Deptt of Computer Science and Engineering
Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
alanchl603@gmail.com

Man-Hon Wong
Dept of Computer Science and Engineering
Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
mhwong@cse.cuhk.edu.hk

ABSTRACT

We present iSyn, an evolutionary algorithm that automatically designs *de novo* ligands with high predicted binding affinity and drug-like properties. It attempts to optimize candidate ligands in accordance with click chemistry and thus ensures chemical synthesizability. In addition to the existing genetic operators of mutation and crossover inherited from AutoGrow 3.0, our iSyn introduces four novel genetic operators to "cut" ligands in order to prevent them from becoming too large in molecular size, hence preserving drug-like properties. Moreover, iSyn interfaces with our fast docking engine idock, greatly reducing the execution time. We hope iSyn can supplement medicinal chemists' efforts.

iSyn was applied to optimizing candidate ligands against two important drug targets, *Tb*REL1 and HIV-1 RT, and managed to produce chemically valid ligands with high predicted binding affinities and drug-like properties. In the example of *Tb*REL1, the predicted free energy of the best generated ligand decreased from -9.878 kcal/mol to -13.985 kcal/mol after 3 generations. In the example of HIV-1 RT, the predicted free energy of the best generated ligand decreased from -5.427 kcal/mol to -12.488 kcal/mol after 2 generations, meanwhile the molecular mass dropped from 602.818 Da to 461.736 Da, so that the compound could be properly absorbed by human body.

iSyn is written in C++ and Python, and is free and open source, available at http://istar.cse.cuhk.edu.hk/iSyn.tgz. It has been tested successfully on Linux and Windows. In the near future we plan to implement a web-based user interface to facilitate its usage and to promote large-scale *de novo* drug design.

Categories and Subject Descriptors

J.3 [**LIFE AND MEDICAL SCIENCES**]: Biology and genetics

Keywords

Bioinformatics, Drug Discovery, Evolutionary Algorithms

GECCO '14, July 12-16, 2014, Vancouver, BC, Canada
ACM 978-1-4503-2881-4/14/07.
http://dx.doi.org/10.1145/2598394.2598398

1. INTRODUCTION

Drug discovery is a long-standing and costly industry. It generally requires US$1.8 billion over 13.5 years to develop a new drug.

We hereby present our new method iSyn for computationally designing *de novo* drugs with click chemistry support so as to aid the identification and optimization of predicted ligands in an automatic manner. iSyn guarantees synthetic feasibility by following approved reactions.

2. METHODOLOGY

iSyn features an evolutionary algorithm where multiple types of click chemistry rules and structural alterations are applied to the initial ligands selected from a fragment library to synthesize new ligands, which are then fed to our fast docking engine idock [1] , available as a service at http://istar.cse.cuhk.edu.hk/idock [2] , to predict their preferred conformations as bound to the target protein and to prioritize them in the ascending order of their predicted free energy. The lower the free energy, the higher the binding affinity. The ligands with the highest predicted binding affinity in the current generation directly survive into the next generation, while the remaining ligands are randomly chosen using the fitness proportionate algorithm. Our hybrid method not only implements elitism but also circumvents over fitting.

There are four types of reactions: crossover, addition, mutation, and cutting. The first three types are performed by two open source algorithms. Through the addition and mutation operations, the ligands could possibly "grow" too large in terms of molecular mass, and thus lose drug-like properties. For example, if a generated ligand has a molecular mass of over 500 Da, it is unlikely to be absorbed by human body and thus unlikely to be optimized into a drug. Unlike some other drug design algorithms which simply discard oversized ligands without decomposing them into moieties which could display inhibitory effects, iSyn implements four novel cutting reactions to split oversized ligands via a set of click chemistry reactions, which are ozonolysis of alkene, oxidation of alkene to carboxylic acid, acid anhydride to carboxylic acid, and hydrolysis of ester (Figure 1).

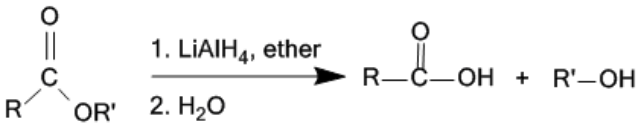

Figure 1: Hydrolysis of ester.

3. RESULTS AND CONCLUSIONS

To evaluate the utility of iSyn, we selected two critical drug targets, RNA editing ligase 1 from *T. brucei* (*Tb*REL1), the etiological agent of African sleeping sickness, and HIV-1 Reverse Transcriptase (HIV-1 RT), an enzyme used to generate complementary DNA from an RNA template. Figure 2 shows the evolutionary steps taken to synthesize the best ligand **Gen2_a4562** against *Tb*REL1 (Figure 3). Having undergone a series of mutation and addition reactions, the initial ligand was transformed to a new structure and its predicted free energy was optimized from -9.878 kcal/mol to -13.985 kcal/mol, proving that iSyn is capable of producing promising ligand structures.

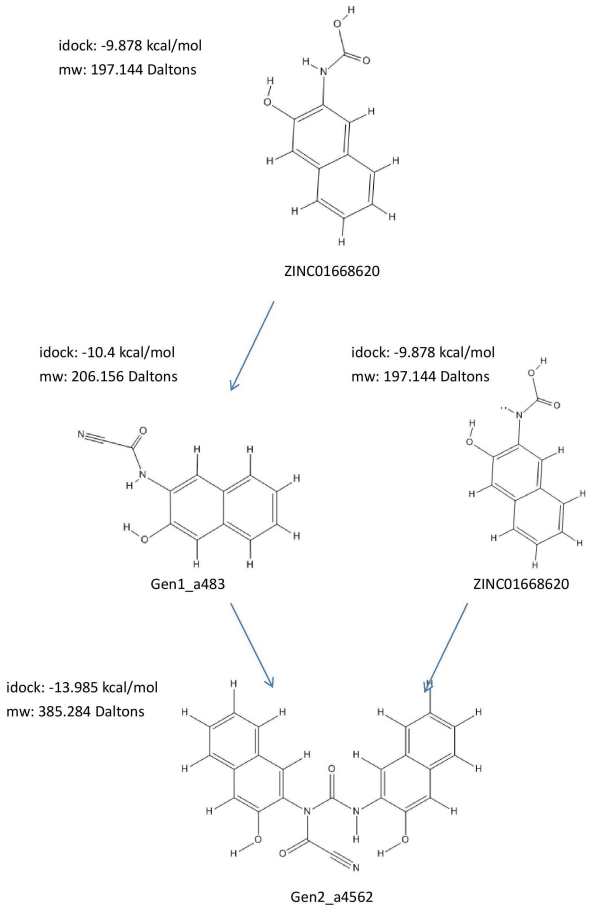

Figure 2: The evolutionary steps taken to generate high binding affinity ligands against *Tb*REL1 (PDB ID: 1XDN). The best ligand was Gen2_a4562 from generation 2 with a predicted binding affinity of -13.985 kcal/mol.

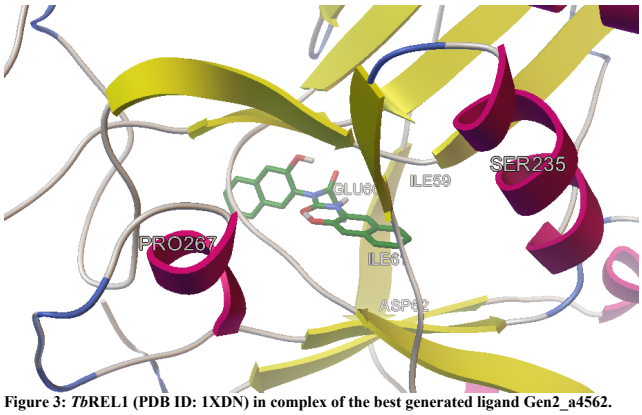

Figure 3: *Tb*REL1 (PDB ID: 1XDN) in complex of the best generated ligand Gen2_a4562.

In the example of HIV-1 RT, ligand **gen1_a13** was synthesized after the first generation (Figure 4). It had predicted free energy of -5.427 kcal/mol and a molecular mass of 602.818 Da. Because its molecular mass was higher than 500 Da, a threshold that discriminates between whether a compound can or cannot be properly absorbed by human body, it automatically underwent one of our four cutting reactions, i.e. the hydrolysis of ester, and got broken down into small child ligands, one of which had predicted free energy of -12.488 kcal/mol and a molecular mass of 461.736 Da, demonstrating that iSyn can simultaneously improve binding affinity and reduce molecular mass.

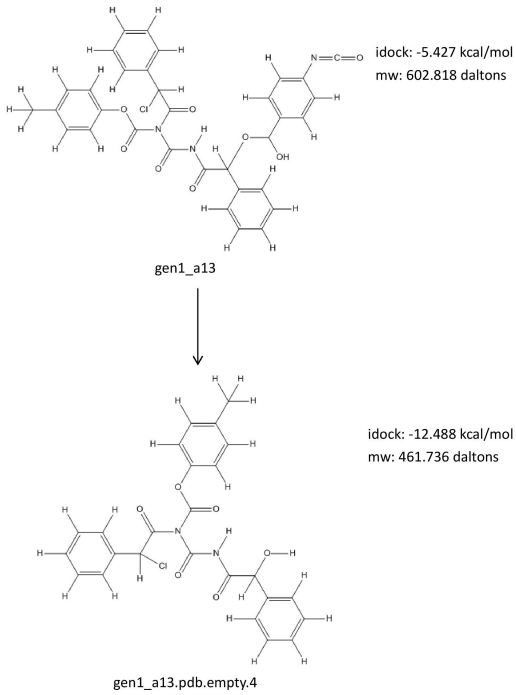

Figure 4: The evolutionary steps taken to cut a oversized ligand into smaller ligands which were subsequently docked against HIV-1 RT (PDB ID: 2ZD1). The resulting ligand had predicted free energy of -12.488 kcal/mol and a molecular mass of 461.736 Da.

In conclusion, we have developed iSyn to automatically generate novel ligands that are drug-like and chemically synthesizable using click chemistry reactions. We hope iSyn can supplement the efforts of medicinal chemists. In the near future, we plan to implement a web-based user interface in iSyn using our WebGL visualizer iview [3] , and port iSyn to our bioinformatics web platform istar at http://istar.cse.cuhk.edu.hk [2] .

4. REFERENCES

[1] Hongjian Li, Kwong-Sak Leung and Man-Hon Wong. idock: A Multithreaded Virtual Screening Tool for Flexible Ligand Docking. *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp.77-84, San Diego, United States, 9-12 May 2012.

[2] Hongjian Li, Kwong-Sak Leung, Pedro J. Ballester and Man-Hon Wong. istar: A Web Platform for Large-Scale Protein-Ligand Docking. *PLoS ONE*, 9(1):e85678, 2014.

[3] Hongjian Li, Kwong-Sak Leung, Takanori Nakane and Man-Hon Wong. iview: an interactive WebGL visualizer for protein-ligand complex. *BMC Bioinformatics*, 15(1):56, 2014.