

化合物の物性を多段階のルール適用で薬様化合物に成長させる web サービスの構築

新井 直樹[†] 吉川 舜亮[‡] 安尾 信明[‡] 中島 悠介[‡] 吉野 龍之介[§] 関嶋 政和[§]
 東京工業大学 工学部情報工学科[†] 東京工業大学院 情報理工学研究科 計算工学専攻[‡]
 東京工業大学 学術国際情報センター[§]

1. はじめに

創薬研究においては、まず標的蛋白質に対してヒット化合物の探索が行われるが、獲得した化合物は必ずしも薬物として適切な活性や物性を有しているとは限らない。そこで、探索によって得られたヒット化合物に対して構造活性相関 (Structure-Activity Relationship : SAR) の解析などを行うことで、薬物様に進化したリード化合物を獲得する。しかしその際には、ヒット化合物の周辺化合物を十分に含んだ化合物ライブラリが必要となる。

本研究では、仮想的な化学反応を入力化合物に対して多段階に適用することで、入力化合物の仮想周辺化合物ライブラリの構築を目的とする。さらに、このライブラリ構築システムを web サービスとして実装することで、ユーザがインタラクティブに化合物の物性値や合成ルールを確認しながら、ヒット化合物から合成可能な周辺化合物の仮想ライブラリを獲得可能とするのも目的としている。

2. 手法

2.1 化合物生成

本研究では、図 1 のように、仮想的な化学反応を入力化合物に対して多段階に適用していくことで新たな周辺化合物を獲得し、ライブラリを成長させる [1]。さらに、その際に特定の物性を増強する傾向を持った反応ルールを重点的に適用することで化合物を目的通り進化させる。

適用する反応として、実際に有機合成可能な化学反応を元として構築した仮想的なグラフ構造変換ルールを SMARTS 記法 [2] で記述したものをを用いた。この反応ルールを RDKit [3] によって入力化合物に適用する。

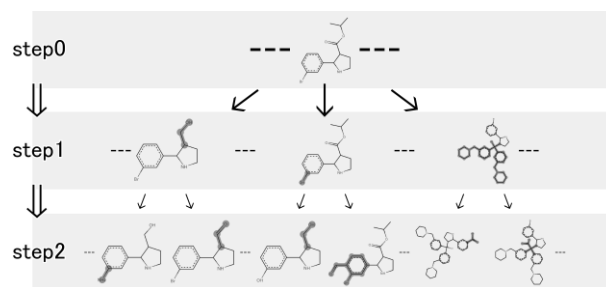


図 1 化合物生成の模式図

2.2 重複削除

ライブラリの構築過程では既に生成済みの化合物が再び生成されうる。そこで、生成済みか判定を行うために、RDKit を用いて化合物データを SMILES 記法の文字列に変換する。この文字列を Python の辞書型オブジェクトのキーとして利用することで、判定を行う。

2.3 構造保護

創薬研究では、標的蛋白質に結合能を示す構造を保持したまま新規の結合を獲得することが重要となる場合がある。そこで、指定された構造が保持されていない化合物を排除することで、構造が保持された化合物のみを獲得する。この機能は化合物の生成時もしくはライブラリ構築の終了時に行う。

2.4 生成過程の追跡

本システムは化合物の生成時にどの反応ルールでどの化合物から生成したか、データベースに記録する。このデータベースを参照することで化合物の生成過程の表示が可能となる。この機能はライブラリに含まれる化合物の合成計画を行う際に利用できる。

3. 実験

3.1 実験条件

システムによるライブラリ構築の検証を行うため、入力化合物としてナミキ商事の building block 統合データベース (2014 年 3 月版) から無作為に 1000 化合物抽出したものを利用した。

また、178 種類の全反応ルールのセットと水素結合供与体 (HBD) 増強傾向の 151 種類、水素結

Building web services to generate druglike analogue library

[†]Dept. of Comp. Sci., Sch. of Eng., Tokyo Tech

[‡]Dept. of Comp. Sci., Grad. Sch. of Information Sci. and Eng., Tokyo Tech

[§]GSIC, Tokyo Tech

合受容体 (HBA) 増強傾向の 30 種類、環構造増強傾向の 106 種類、の 3 つの増強セットをそれぞれ入力化合物に対して適用した。

3.2 動作環境

CPU : Intel Corei7-3770 (3.40GHz * 4)
メモリ : 16GB
OS : Ubuntu 15.04 (64 bit)
言語 : Python ver. 2.7.9, RDKit ver. 2015.3,
Django ver. 1.7.6

3.3 実験結果

表 1 はそれぞれの反応セットを適用した場合の出力化合物数と実行時間を示したものである。増強セットを適用した場合に化合物数が 0.01~0.39 倍、実行時間が 0.01~0.37 倍に減少しているが、これは適用する反応ルール数を絞ったことによるものである。

しかし、それぞれの増強セットが目的とする物性の平均値を比較すると、全ての増強セットにおいて目的とする物性をよりよく向上させていることが分かる。例として図 2 に HBA 数平均値をその増強セットと全反応ルールを適用した出力ライブラリについて比較したグラフを示す。

以上から、特定の物性を増強する反応セットに絞って適用することで効率的に目的とする物性値を増加させることが可能といえる。

3.4 web インターフェース

ライブラリ構築の起点とする化合物の入力や保護構造の指定には、SMILES や化合物ファイル及びそれらを含んだ zip ファイル等を利用する。これらの入力の後に、使用する反応セットと適用段階数を選択し、開始ボタンを押すことでライブラリ構築が開始される。ライブラリ構築の終了後は構築したライブラリを複数の形式でダウンロードすることが可能であり、また、ライブラリ内化合物の構造や物性値、化合物の生成過程を参照可能である。

web インターフェースの例として図 3、図 4 に化合物入力画面、生成過程の追跡画面を示す。

4. まとめ

本研究では入力化合物に対して仮想的な化学反応を適用することで、入力化合物から合成可能な仮想周辺化合物のライブラリを構築するシステムを開発した。

さらに、これを web サービスとして実装することで、煩雑な操作なしにライブラリ構築や指定構造の保護、化合物生成過程の追跡など創薬研究に有用な機能を利用可能にした。

参考文献

[1] 吉川舜亮, 安尾信明, 吉野龍ノ介, 関嶋政和: 方向性を持たせたグラフ構造変換による仮想化合物ライ

ブラリの構築の研究, 研究報告バイオ情報学 (BIO), 2015-BIO-41, (12), 1-6, 2015-03-13.

- [2] Daylight Theory: SMARTS, <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>
- [3] RDKit: Open-Source Cheminformatics Software, <http://www.rdkit.org/>
- [4] The Web framework for perfectionists with deadlines | Django, <https://www.djangoproject.com/>

表 1 各反応セットを適用した際の出力化合物数と実行時間

	出力化合物数[#]	実行時間[s]
全反応ルール	2732309	14649.44
HBD増強傾向	1001486	5689.76
HBA増強傾向	30442	183.35
環構造増強傾向	117537	972.99

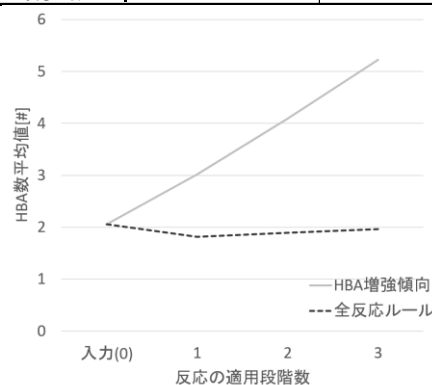


図 2 全反応ルールと増強セットにおける HBA 数平均値の比較

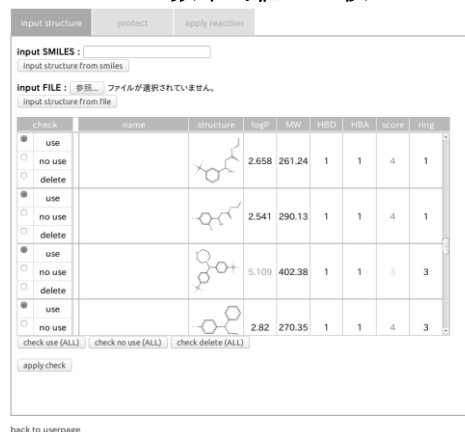


図 3 web インターフェース: 化合物入力

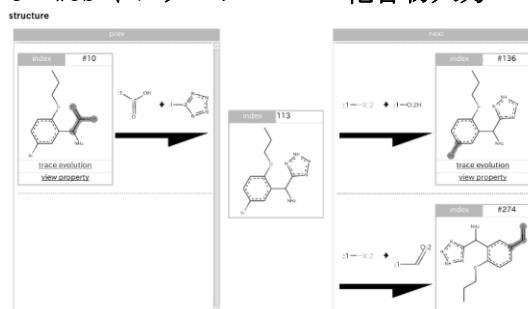


図 4 web インターフェース: 生成過程の追跡