

化学空間の可視化を利用した化学構造ジェネレータの開発

○三島 和晃、金子 弘昌、船津 公人

東京大学大学院工学系研究科 (〒113-8656 東京都文京区本郷 7-3-1)

【緒言】

創薬研究の初期段階において未知の活性化合物を探索することを目的として、バーチャルに多様な化学構造を作成する様々な構造ジェネレータが提案されている[1]。指定した条件を満たす構造を網羅的に発生させる MOLGEN[2] や evolutionally algorithm を用いて活性が高い化合物を探索する CoG[3] をはじめとして多くの試みがなされている。多くの構造ジェネレータでは、化学構造に対する評価指標として対象の構造との類似性や quantitative structure-activity relationships (QSAR)モデルによる活性の予測値などを使用しており、これらを算出するために構造を数値的に処理することを目的として様々な構造記述子を用いられている。適切な構造記述子を用いることで、構造に対する評価を行なうだけでなく、self-organization map (SOM)[4] や generative topographic mapping (GTM)[5] などの次元削減手法により可視化した化学空間上で構造間あるいは構造と活性との関係を視覚的に確認することができる。また counter propagation (CP)[6] を用いることで、化学空間の可視化と QSAR モデルの構築を同時に行なうことも可能である。

創薬研究においては、高い活性を示す化合物であっても、ADMET 特性などの諸条件のために医薬品として利用することが難しい場合がある。このような場合に備え高い活性を持つ様々な化合物が必要とされる。化学空間において既存の高活性化合物が存在する周辺の領域には高い活性を示しうる構造が数多く存在すると考えられ、その中から活性以外の条件も満たす化合物が得られると期待できる。

そこで本研究では、化学空間上で高活性化合物の周辺領域に存在する構造を効率的に生成することを目的とする構造ジェネレータを開発した。既存の高活性化合物群を初期シード構造として本ジェネレータを使用することで、化学空間上の目的領域に存在し高い活性を示すと考えられる構造を得ることが可能となる。ジェネレータの有用性を確認するため、リガンド結合活性のデータを用いた化学空間の可視化と構造生成を行なう。

【手法】

開発した構造ジェネレータの概要を図 1 に示す。構造ジェネレータを用いるにあたり、はじめに既存の化合物から構造記述子を算出し、次元削減手法により二次元平面への写像を行なう。二次元平面上で活性の高い化合物が集まる領域を目的領域として、目的領域付近にある化合物を構造ジェネレータに入力する初期シード構造として選択する。

選択したシード構造に対して重原子の付加などの操作が適用されることで新規構造が生成される。得られた構造に対してフィルタリングが行なわれたのち、残った構造は構造プールに追加される。

プールされている構造の中からは評価値にもとづき次世代のシード構造が選択される。各構造の評価値の算出においてはマップ上で目標とする座標からの距離 r を用いて、以下の式により算出する。

$$SCORE(r_i) = \exp\left(-\frac{(r_i - r_{min})^2}{\alpha\sigma_r^2}\right) \quad (1)$$

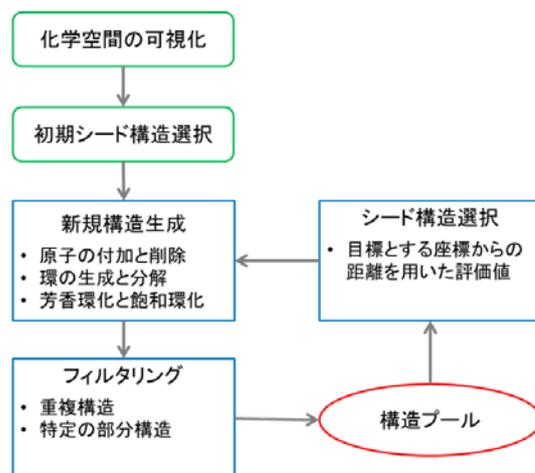


図 1 開発した構造ジェネレータの概要

ここで r_{min} は r の最小値、 σ_r^2 は r の分散である。また a は評価値の分布を調整するパラメータである。(1)式により、マップ上で目標とする座標からの距離が小さい構造ほど評価値が高くなる。

選択されたシード構造に対して前の世代と同様に新規構造生成、フィルタリングを行ないプールに追加する。このサイクルを繰り返すことにより、マップ上の目的領域に持続的に新規構造が生成される。

【結果と考察】

開発した構造ジェネレータの有用性を確認するため、GVK[7] のデータベースより得られた 635 個の化合物のアデノシン A_{2A} レセプターに対するリガンド結合活性のデータを用いたケーススタディを行なった。構造記述子として PubChem[8] のフィンガープリントのうち Section 1 から Section 5 までを使用した。前処理として分散が 0 の変数と相関係数が 1 となる変数の組のうち片方の変数を除去したのち、GTM を用いて化学空間を可視化した。

GTM により構築した二次元マップが図 2 である。マップ上で色が赤い部分は活性が高いことを表わす。活性の予測値の算出には、マップ上の近傍の 3 点の活性の実測値を平均して予測値とする 3-nearest neighbors 法を用いた。ここで図 2 に黒いプロットで示した特に高い活性の実測値を持つ 2 つの構造を選択し、構造ジェネレータに入力する初期シード構造とした。評価値算出の際には目標とする座標を 2 点の重心とした。

構造ジェネレータにより生成された約 10,000 個の構造のマップ上への写像を図 3 に示す。初期シード構造とした 2 点の付近から重心方向に多くの構造が生成されており、化学空間上の目的領域に効率的に構造を生成することができたと考えられる。なお、一部の構造は目的領域から大きく外れている。この原因として、マップが非線形であることや、フィンガープリントが化学構造の連続的な変化の情報を十分に反映していない可能性などが考えられる。

【結言】

本研究では、化学空間上で複数の既存の高活性化合物に囲まれた領域に存在する新規構造の生成を目的とした構造ジェネレータの開発を行なった。リガンド結合活性のデータを用いたケーススタディにおいて GTM による化学空間の可視化を行ない、初期シード構造と目的領域を定めた。マップ上の重心からの距離による構造の評価をもとに、目的領域内に多数の構造を生成することに成功した。本研究で開発した構造ジェネレータを用いることで、効率的な創薬研究が可能となることが期待される。

【参考文献】

- [1] B. Pirard, *Expert Opin. Drug Discov.*, 6, 225, 2011
- [2] T. Wieland, A. Kerber, R. Laue, *J. Chem. Inf. Comput. Sci.*, 36, 413, 1996.
- [3] N. Brown, B. McKay, J. Gasteiger, *J. Comput. Aided Mol. Des.*, 18, 761, 2004.
- [4] T. Kohonen, *Self-Organizing Maps 3rd Ed.*, Springer, 2000.
- [5] C. M. Bishop, M. Svensen, C. K. I. Williams, *Neural Computation*, 10, 215, 1998.
- [6] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, New York, 1999.
- [7] <http://www.gvkbio.com/informatics.html>
- [8] <http://pubchem.ncbi.nlm.nih.gov/>

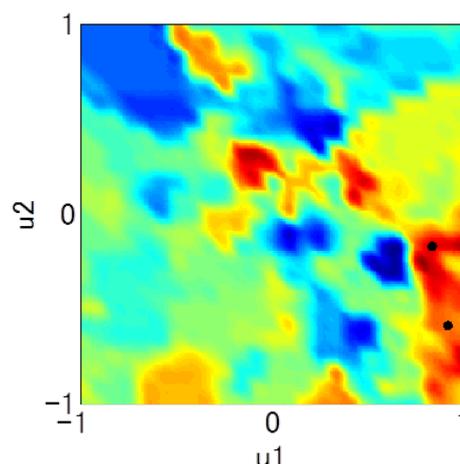


図 2 GTM と 3-NN 法による活性マップ

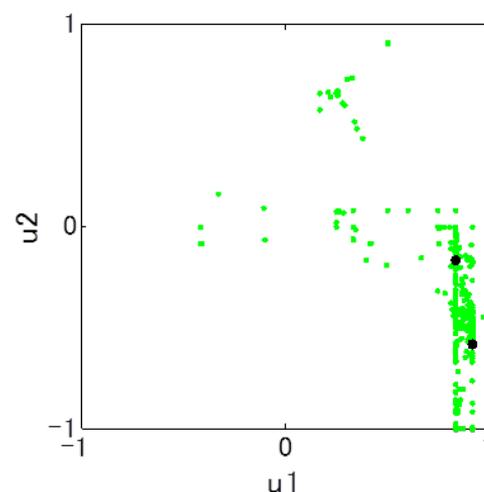


図 3 生成された構造のマップ上への写像