

進化的プログラミングを用いた創薬候補化学構造の創出

Structure Generation for Finding New Drug Candidates using Evolutionary Programming

吉丸 和孝
Kazutaka Yoshimaru

河合 健太郎
Kentaro Kawai

高橋 由雅
Yoshimasa Takahashi

豊橋技術科学大学 工学部 知識情報工学系
Department of Knowledge-based Information Engineering, Toyohashi University of Technology

We propose an approach to finding new drug candidates which are expected to have desirable biological activities. The method is based on evolutionary programming. In our approach a seed structure is submitted to generate an initial population. Then, individuals of the initial population are generated along with evolutionary operations, including mutation and recombination. It has two key features. A chemical structure is coded by an atomic connection table. SVM classifiers trained with a set of known drugs in advance are employed for the survivor selection of individual structures. Thus, the method allows us to generate possible candidate chemical structures of new drugs. Basic idea of the method and its implementation are illustrated with examples.

1. はじめに

創薬において、医薬品の候補となりうる化学構造だけでも 10^{60} 種類は存在するといわれている[Bohacek 96]. 従来のコンビナトリアルライブラリでは、あらかじめ生成された構造に対してのみ評価が可能であり、目的の活性を発現する候補構造が必ずしも含まれているとは限らない。そこで、目的の活性発現が期待される創薬候補構造を能動的かつ効率的に生成し、提示するための手法が望まれる。

本研究では、進化的プログラミングの技法を化学構造の創出に適用し、核となる基本構造を出発点に、任意の活性(一つあるいは複数の活性)の発現が期待される候補化学構造を創出する手法について検討を行った。

2. 分子進化システムの構築

2.1 グラフベースの進化的プログラミング

進化的プログラミング(以下 EP)は、自然界における淘汰・交叉・突然変異などを基にした進化計算の一種である。一般によく知られているビット列型の解をもつ GA とは異なり、化学グラフの結合表をそのまま染色体として用いている。図1に進化プログラミングの概要を示す。

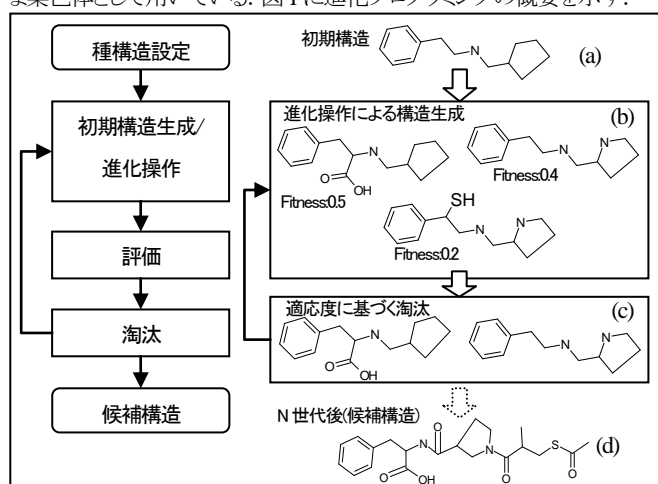


図1 分子進化的プログラミングの概要。初期構造(a)より突然変異にて個体集団を生成する(b)。集団に対し、適応度関数に従って評価・淘汰することで、新世代への生き残り個体を選抜する(c)。この操作を繰り返すことで、より適応度の高い個体集団への進化を図る(d)。

連絡先: 高橋由雅, 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1 豊橋技術科学大学 知識情報工学系, Tel: 0532-44-6878, taka@mis.tutkie.tut.ac.jp

2.2 初期構造

当研究室ではこれまでに、化学構造から側鎖を除外して得られる環を中心とした基本骨格グラフ Non-Terminal vertex Graph (NTG) と薬理活性との相関について明らかにしてきた[大野 02]. 基本骨格が変わると発現する活性も大きく変わることは経験的にも知られており、分子進化において基本骨格を保持することは機能発現との関わりに重要な意味をもつ。そこで、本研究では目的の活性と関連のある構造を種構造とし、その NTG を初期構造として与えた。また、生成されるすべての構造は、同一の NTG を保持するように設計をおこなった。

2.3 交叉

交叉の対象となる二つの構造(個体)をランダムに選択し、NTG上での交叉・組み換えにおいてもそのNTGフレームワークが結果として保持されるよう、両者の等価な結合部位での交叉・組み換えのみを許す方法を採用した。

2.4 突然変異

突然変異の対象となる任意構造を一つ選択し、次のいずれかの操作をランダムに選択する。

- 原子の付加(水素原子を除く)
- フラグメントの追加
- 原子/フラグメントの削除
- 原子種の変異
- 結合種の変異

原子の追加操作では、創薬において一般的な元素である C, N, O, S, P, X(ハロゲン)のうちから、ランダムに選択した1原子の付加(拡張)を行う。原子や結合の追加/変更の可否は、原子価チェックにより判断した。フラグメントの追加操作では、一般的な置換基や活性の発現に有意な部分構造を事前に定義したものをを用いる。このとき、各フラグメントは一つの疑似原子として取り扱われる。

2.5 適応度評価

分子進化プログラミングの実装に際しては、各世代集団を形成する個体としての化学構造の適応度をどのように評価するかは極めて重要な問題である。このとき、はじめに問題となるのは、目的とする化学構造の優劣を判断するための評価関数を定義する以前に、分子の構造特徴をどのように数値化するかということである。

化学構造特徴の数値記述表現(数値化)には、TFS 法[Takahashi 98]を用いた。TFS とは化合物の構造式から部分構造を列挙し、その数値的な特徴付けに基づいて化学構造のトポロジカルな構造特徴を

多次元数値ベクトルで表現したものである。ここでは、結合数 5 までの部分構造を列挙し、その質量数で特徴づけを行った。このように、本研究では候補化学構造の特徴表現に TFS 法を用いた。従って、ここでの適応度関数は TFS の関数として記述される。実際に用いた関数については後述する。

2.6 淘汰

適応度の評価を行った個体集合から任意の 4 つを選択し、のうち最も適応度高い個体のみを次世代集団のメンバーとして残し、その他は破棄するトーナメント方式を用いた。

3. 予備実験

3.1 概要

実装したアルゴリズムの進化操作が、定義される適応度に従って効果的に候補構造に到達できるか検証するため、TFS パターンの類似度を適応度関数とし、初期構造を抽出した親構造を伏せ、NTG 基本骨格から親構造の復元を試みた。実験は、各世代の個体数を 100、世代交代数 100 の進化計算を 100 セット繰り返し、各世代の適応度の平均と、目隠した親構造が再現された回数を求めた。

3.2 データセットと適応度関数

図 1(d)に示した NTG フレームの 3 つの頂点原子に側鎖を有する構造を親構造とし、その NTG 構造を初期構造とした。個体集団の進化操作のための定義フラグメントには、これら 3 つの側鎖および別途定義した 6 種の置換基をフラグメントセットとして用意した。

進化の方向を決定する適応度関数には、進化操作の過程で生成される候補構造の TFS ベクトルと、復元すべき元の構造の TFS ベクトルとの類似度を用いた。類似度には Tanimoto 係数を用いた。

$$S(x, y) = \frac{\sum x_i y_i}{\sum (x_i^2) + \sum (y_i^2) - \sum x_i y_i} \quad (0 \leq S \leq 1)$$

ここで、 x は個体(構造)TFS パターンベクトル、 y は目隠した親構造の TFS パターンベクトルを表す。

3.3 実験結果

100 回試行時の最良個体の適応度の平均は 0.955 で、15 世代程度で収束をみた。これより、元構造と高い類似度を持つ個体が早期に生成されているといえる。また、100 回の試行のうち、2 回は完全に元構造を復元した。初期構造の置換可能カ所は 20 であり、フラグメントセット(9 種類)のみの組み合わせでも 10^{20} 通り存在する。実際には他の変異操作も行われるため、候補解の数はさらに膨大になる。15 世代×100 試行で生成する個体数が 1.5×10^5 個であることを考慮すると、2 回という数字は決して少なくなく、シラミ潰し法に比べ効率的な探索を行っていることを示している。

4. 実データを用いた計算機実験

4.1 概要

当研究室では、TFS 表現を入力パターンベクトルとした、サポートベクタマシン(SVM)による多クラス活性分類・予測システムについて検討を進め、その有効性を明らかにしてきた [河合 07]。降圧剤の候補構造の創出を目的とし、ACE 阻害活性の有無、NEP 阻害活性の有無に関する SVM 分類機を作成し、両方の活性をもつ候補化学構造について創出を試みた。

4.2 適応度評価のための SVM 分類機

ここでは適応度評価に、別途、薬物構造データベース MDDR [MDL 01]を用いて作成した SVM 分類機を利用した。識別モデルは、クエリとして与えた構造の活性の有無をサポートベクトルからの距離と

して返し、距離が正であれば活性あり、負であれば活性なしと判断される。そこで、以下のように適応度関数を設定した。

$$fitness(q) = \sum s_i d_{q,i}$$

s_i : 活性*i*の発現を期待するなら 1.0、しないなら -1.0

$d_{q,i}$: クエリ*q*の活性*i*に関するサポートベクトルからの距離

4.3 実験結果

ACE 活性のみを発現する治験薬構造の基本骨格である NTG 構造(図 1(d))を初期構造とし、各世代の個体数を 200 として 5 回試行を行い、スコアの平均を算出した。各世代において最も高い適応度を示した個体の平均スコアについて、200 世代までの進化計算における変化を図 2 に示す。図 2 より、生成された候補構造については SVM による分類・予測モデルによって、ACE 阻害剤、NEP 阻害剤の両方の活性発現が期待できることが示された。これは、当初の狙い通りであり、ここでの分子進化プログラミングが有効に機能していることを示すものである。しかし、ここで生成された候補構造の実際の合成可能性については、別途、議論が必要かもしれない。

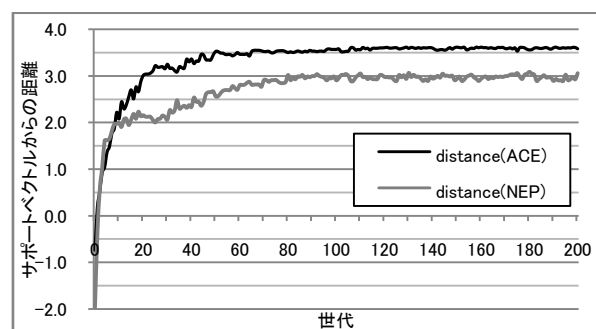


図2 計算機実験による各世代の最良個体における、サポートベクトルからの距離(距離が正であれば活性を有するとみなす)

5. まとめ

本論文では、進化的プログラミングを用いた創薬候補化学構造の創出に関する実験を行った。その結果、今回実装を行った進化的プログラミングによる構造創出手法が、適応度関数に応じて効率的に候補解を生成する能力を有し、さらに、SVM によるマルチクラス識別器の結果を適応度関数として用いることで、任意の活性を発現する創薬候補構造の効率的な創出可能性を示した。しかし、生成される候補構造の合成可能性に関する検証が不十分であり、今後の検討課題である。

参考文献

- [Bohacek 96] Regine S. Bohacek, Colin McMartin, and Wayne C. Guida: The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective, *Medicinal Research Reviews*, Vol. 16, No. 1, pp.3-50, 1996.
- [大野 02] 大野貴生, 高橋由雅, 第 30 回構造活性相関シンポジウム 公演要旨集, 41-42, 2002.
- [Takahashi 98] Y. Takahashi, H. Ohoka, and Y. Ishiyama : Structural Similarity Analysis Based on Topological Fragment Spectra, *In: R. Carbo and P. Mezey (Eds), Advances in Molecular Similarity 2*, pp.93-104, JAI Press, Stamford CT, (1998).
- [河合 07] 河合健太郎, 藤島悟志, 高橋由雅: “多クラス SVM を用いた薬物の活性予測システム”, 2007 年度人工知能学会全国大会 (第 21 回) 論文集, 3D7-5 (2007).
- [MDL 01] MDL, MDL Drug Data Report, 2001.1, (2001).