進化論的アルゴリズムを利用したQSAR逆設計

長谷川 清^a、木村 敏郎^b、船津 公人^{c*}

^a中外製薬株式会社・鎌倉研究所（〒247-8530 神奈川県鎌倉市梶原200番地）

^b株式会社ワールドフュージョン（〒103-0013 東京都中央区日本橋人形町2-15-15）

^c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）

(Received September 17, 2008; Accepted February 3, 2009)

定量的構造—活性相関 (Quantitative Structure-Activity Relationship: QSAR) は、古くから行われており、成功例も多い。しかしながら、予測精度の高いモデルを構築すると、逆にモデルの解釈が難しくなり、設計が困難な場合が多い。そこで、任意の構造をコンピュータで発生させて、モデルの予測値が高い構造だけを自動発生するQSAR逆設計のシステムを構築した。

システムのコアとなる構造発生部分については、EA-Inventor (Evolutionary Algorithm-Inventor) を利用した。SMILEで表記した初期構造を入力すると、交差、変異などの操作で新たな構造を生成する。発生した構造をQSARモデルで予測して、予測値をスコアとして返す。スコアが高い構造は残して、さらに、新たな構造を生成する。このような操作を複数回繰り返して、高いスコアを持つ構造が十分そろったら終了する。

EA-Inventorを文献のトリプシン阻害剤データに適用した。スコアに利用するQSARモデルとして、CoMFA (comparative molecular field analysis) を採用した。学習セットと大きく異なる構造が発生した場合には、高いペナルティー値を加え、発生構造が学習セットから大きく逸脱しないように工夫をした。出力された構造は、トリプシンの構造と高い相補性を持っており、高い阻害活性が期待できる。今回は、阻害活性のQSARモデルの応用であるが、これに、ADME (absorption, distribution, metabolism, excretion) モデルや合成の難易度を加えることもできるので、より実用的な薬物設計が可能である。

キーワード: 定量的構造—活性相関、逆QSAR解析、EA-Inventor、トリプシン阻害剤、CoMFA, ADME

* funatsu@chemsys.t.u-tokyo.ac.jp

解析手法

1. はじめに

背景

定量的構造－活性相関 (Quantitative Structure-Activity Relationship: QSAR) は、古くから行われており、成功例も多い。[1,2] しかしながら、予測精度の高いモデルを構築すると、逆にモデルの解釈が難しくなり、設計が困難な場合が多い。特に、ニューラルネットワークのような非線形モデルの場合、構造記述子と阻害活性値の関係があらわには記述できないので、より深刻となる。そこで、任意の構造をコンピュータで発生させて、モデルの予測値が高い構造だけを自動発生するQSAR逆設計のシステムを構築した。

システムのコアとなる構造発生部分については、EA-Inventor (Evolutionary Algorithm-Inventor) を利用した。発生した構造をあらかじめ作成しておいたQSARモデルで予測して、予測値をスコアとして返す。スコアが高い構造は残して、さらに、新たな構造を生成する。このような操作を複数回繰り返して、高いスコアを持つ構造が得られるようにした。

2. 解析データと解析手法

解析データ

解析データとして、3-amidinophenylalanine誘導体のトリプシン阻害剤データを利用した。[3] トリプシン阻害剤の化学構造式を、図1に示した。置換位置は、R1, R2と二箇所ある。文献データは、トリプシン、トロンビン、ファクターXaと3種類のセリンプロテアーゼ阻害活性値がある。今回、出力された構造の妥当性を検証する目的で、amidinophenylalanineとのX線構造が知られているトリプシンにのみ着目した。

Figure 1: Chemical structure of trypsin inhibitor

進化論的アルゴリズム

遺伝的アルゴリズム (Genetic Algorithm: GA) は、スコアを極大 (あるいは極小) とする入力値を速やかに求めるための手法である。[4] 入力値は、0/1のビット列で表現し、より高いスコアになるようにビット列を進化させる。ビット列を進化させるための方法として「ランダムに選択したビットの状態を反転させる: Mutation」「2つのビット列をランダムに選択した位置で切断して入れ替える: Cross-over」が用いられる。極大値探索を「ビット列操作」で実現しているため、複雑な系の探索を単純な手続きで実現できるのが、GAの大きな特徴である。

進化論的アルゴリズム (Evolutionary Algorithm: EA) は、GAの考え方を踏襲して設計された分子構造探索のための計算手法である。[5] GAとの違いは、進化させる対象が分子構造となっている点にある。アルゴリズムの概略は以下のとおりである。

1. 初期分子構造入力
2. 初期分子構造のスコア計算
3. 分子構造の更新・新規構造生成
4. スコアの計算
5. 分子構造の順位付け
6. 手順3に戻る

EA については、EA-Inventor (Evolutionary Algorithm-Inventor) を利用した。EA-Inventorは、TRIPOS社より販売されており、ドッキング等と組み合わせた応用事例が報告されている。[6] 本研究では、EA-InventorとQSARモデルを組み合わせることによって、QSAR逆設計を検討した。初期構造は、SMILESで用意する。更新手法としては、33種類の操作 (mutation, cross-over)があり、何れかをランダムに選択して適用することで、手順3の「分子構造の更新」を行っている。図2に、代表的な操作の実例を示す。(i), (ii)は、原子タイプ、結合タイプの変更を示す。(iii), (iv)は、環構造の開裂、環のサイズの変更を示す。(v)は、フラグメントの追加を示す。フラグメントとしては、薬物を分解して発生させた1400個のライブラリが用意されている。したがって、これらの組み合わせで構築された構造は、ドラッグライクに近い。アルゴリズムの手順3-5を繰り返し実行することで、より高いスコアを与える分子構造が探索できるよう設計されている。図3に、EA-Inventorのアルゴリズム概要を示した。

Figure 2: Representative operations (i) modification of atom type, (ii) modification of bond type, (iii) breaking of ring, (iv) modification of ring size, (v) addition of fragment

Figure 3: Process of EA-Inventor

CoMFA

CoMFA (comparative molecular field analysis) は、分子の3次元立体構造を元にQSAR式を作成する、3次元構造活性相関 (3D-QSAR) 手法である。[7] まず、計算対象とする分子群の3次元構造を重ね合わせる。(図4、左図) 次に、各分子について分子周辺の状態を表すフィールドを作成する。(図4、右図) フィールドは、分子周辺に規則的に配置したグリッド点座標での立体相互作用・静電的相互作用の値として定義される。これらは、おのこの分子について、それぞれ「レナード・ジョーンズ・ポテンシャル」「クーロン・ポテンシャル」で計算され、これらの値はそのまま分子の特徴を表す構造記述子として利用される。構造記述子と阻害活性値との相関関係を、PLS (partial least squares) 解析することで、モデル式が得られる。

一度、モデル式が得られると、各グリッド点に対する重み係数を検討することで、活性発現に必要な構造的特徴を知ることができる。(等高線図) また、活性未知の分子の立体構造を重ね合わせしてフィールドを計算すれば、モデル式から活性予測値が得られる。

Figure 4: CoMFA (left hand side: superimposition of molecules, right hand side: calculation of molecular fields)

EAI-CoMFA

本研究で使用した各種手法は、SYBYLソフトウェアのマクロ言語SPL、Shell ScriptおよびPerl言語（もしくはawk言語）で記述して、EA-Inventorと連携させた。

EA-InventorとCoMFAモデルとを連携することで、新規構造を発生させた。予測を行う場合、分子重ね合わせが必須となるが、本研究では、amidinophenylalanine部分を重ね合わせポイントとした。また、共通部位を重ねた後、MMFF94力場にて分子構造のひずみを取り除くこととした。各分子のスコアとして、活性予測値から、ペナルティー絶対値を引いた値を利用した。

$$\text{スコア} = (\text{活性予測値}) - |\text{ペナルティー}| \quad (1)$$

ペナルティーは、発生した構造のフィールド値が、学習セットのフィールド値の範囲に対してどれだけ逸脱しているかを表す尺度である。

$$\text{ペナルティー} = \sum_q b_q \times X_{kq} \quad (2)$$

ここで、 q は、発生した構造 k のフィールドの中で、学習セットの範囲にないグリッドポイントである。 b_q は、対応するフィールドの係数値、 X_{kq} は、発生した構造 k の対応するフィールド値である。したがって、逸脱したフィールドが多いほど、ペナルティーが大きくなり、結果的にスコアが小さくなる。[8] 図5に、ペナルティー計算の概要を模式的に示した。

Figure 5: Calculation of penalty value from grid points

例えば、既知の分子構造から外れた紫の構造がついた分子が生成された場合を考える。赤で囲んだグリッド点からは「QSAR式作成時の値域を超える相互作用」が計算されることになる。そのような値を予測式に入力した場合、式の信頼性が保てなくなると考えられる。そこで、今回の計算ではこれらのグリッド点から計算される値を活性値から減じて分子の評価を行うこととした。

EAIInventorは、ユーザが簡単に条件を設定できる専用のグラフィックユーザインターフェイス (GUI) を持っている。基本的には、CoMFAのスコア値をEAIInventor内蔵のscore.logに保存するように設定するだけである。後は、初期構造、ライブラリー、成長部分、GAの世代数、集団数などをGUIで指定することで、種々の構造が、SMILE, MOL2, SDFなどで出力されるようになっている。

3. 結果と考察

文献データのトリプシン阻害剤72化合物について、amidinophenylalanine部分を重ね合わせて、CoMFAモデルを作成した。CoMFAモデルは、立体相互作用、静電相互作用の2つのフィールドを用いて、3成分のPLSで得た。相関係数値およびleave-one-outによるクロスバリデーション相関係数値は、それぞれ、 $R^2 = 0.829$ 、 $Q^2 = 0.592$ となった。図6に、実測値と予測値のプロット図を示した。阻害活性値の大きい領域で十分な予測性能が得られていることから、今回のQSARモデルに利用できると思う。

Figure 6: Plot of observed and predicted inhibitory activities

CoMFAの2つのフィールドから計算される等高線を図7に示した。左図が、立体的等高線図であり、右図が、静電的等高線図である。

Figure 7: Steric contour map (+: green, -: yellow); Electrostatic contour map (+: blue, -: red)

これら等高線図から、従来法で化合物設計は可能であるが、図が複雑であるために容易ではない。また、実際に分子をモデリング・重ね合わせして予測する必要があり、かなりの労力を要する。

化学構造の変動が小さい置換基R1は、sulfone-naphtylに固定して、EA-Inventorで阻害活性が高い置換基R2を探索した。(図1) EA-Inventorの条件は、世代数10、集団数50で、5回実行した。EA-Inventorからの出力構造の一例を図8に示す。芳香環が圧倒的に多かった。各置換基の下にある数字は、CoMFAモデルによる阻害活性予測値である。また、括弧内の数字は、式(2)で計算されたペナルティー値である。いずれの置換基もペナルティー値が小さく、学習セットのR2から大きく逸脱していないことを表している。阻害活性値が一番大きい化合物11の阻害活性予測値は、6.91である。

R2 =

Figure 8: Output structures from EA-Inventor

図9に、X線構造が登録されている化合物11とトリプシンとの結合様式を示した。(PDB code: 1K1J) 出力された置換基R2は、いずれも、トリプシンのヒスチジン57とCH- π および π - π 相互作用があるので、高い阻害活性が期待できる。

Figure 9: Interaction between compound 11 and trypsin

図8に示した構造を元に、種々の構造を設計することは可能である。たとえば、benzothiazole, indoleは、benzofuranのような他の芳香環でも変換可能である。また、5員環のpyrrolidineから6員環のpiperidineへの変換も合理的である。このように、EA-Inventorから出力された構造は、次の分子設計のヒントになるので、幅広い応用が可能となる。

まとめ

EAとQSARモデルを連結して、逆設計のシステムを構築できた。QSARモデルは、CoMFA以外でも適用できるので、ニューラルネットワークのような非線形でモデルの解釈が困難な場合でも、実用的に利用できる。

今回は、阻害活性のQSARモデルの応用であるが、これに、ADME (absorption, distribution, metabolism, excretion) モデルや合成の難易度を加えることもできるので、より実用的な薬物設計が可能である。

参考文献

- [1] M.Arakawa; K.Hasegawa; K.Funatsu, *Curr.Comput.-Aided.Drug.Des.*, **3**, 254-262 (2007).
- [2] K.Hasegawa; K.Funatsu *Frontiers in Drug Design and Discovery Volume 4*, "In-vivo PK/PD and in-vitro ADME techniques", 2009, Bentham Science Publishers, Ltd.
- [3] B.Markus; J.Sturzebecher; G.Klebe, *J.Med.Chem.*, **42**, 458-477 (1999).
- [4] K.Hasegawa; M.Arakawa; K.Funatsu *Curr.Comput.-Aided.Drug.Des.*, **1**, 129-145 (2005).
- [5] G.Schneider; U.Fechner, *Nature Reviews*, **4**, 649-663 (2005).
- [6] Q.Liu; B.Masek; K.Smith; J.Smith, *J.Med.Chem.* **50**, 5392-5402 (2007).
- [7] R.D.Cramer III; D.E.Patterson; J.D.Bunce, *J.Amer.Chem.Soc.*, **110**, 5959-5967 (1988).
- [8] SYBYL 7.0, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.

Inverse QSAR Study Using Evolutionary Algorithm

Kiyoshi Hasegawa^a, Toshiro Kimura^b, Kimito Funatsu^{c*}

^aChugai Pharmaceutical Company, Kamakura Research Laboratories,
Kajiwara 200, Kamakura, Kanagawa 247-8530, Japan

^bWorld Fusion Co.,Ltd., 2-15-15Nihonbachi Ningyo-cho,
Chuo-ku, Tokyo 103-0013, Japan

^cThe University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan

Quantitative Structure-Activity Relationship (QSAR) has been widely used in molecular design and many successful applications have been known. However, when precise model would be made, its chemical interpretation becomes difficult. Therefore, we have developed inverse QSAR system to generate chemical structures with high inhibitory activities. EA-Inventor (Evolutionary Algorithm-Inventor) was used as structure generator in our system. In EA-Inventor, initial structures represented by SMILES strings are modified using cross-over and mutation operations. The inhibitory activities are predicted by QSAR model and their values are used as the scores. The structures with high scores survive and next new structures are generated. After these cycles, the higher scored structures can be obtained. We have applied EA-Inventor to Trypsin inhibitory data. CoMFA (comparative molecular field analysis) is employed as QSAR model. If new structures differ from training set, high penalty values are added to scores in order to guarantee them within chemical space. The generated structures are high complementarily to Trypsin structure and the result seems to be reasonable. This technology can be extended to ADME model and chemical feasibility filter.

Keywords: Quantitative Structure-Activity Relationship, Inverse QSAR system, EA-Inventor, Trypsin inhibitors, CoMFA, ADME

* funatsu@chemsys.t.u-tokyo.ac.jp