

分子設計のための統合システム ToMoCo の開発

荒川 正幹*, 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 丁目 3 番 1 号

*e-mail: arakawa@chemsys.t.u-tokyo.ac.jp

**e-mail: funatsu@chemsys.t.u-tokyo.ac.jp

(Received: September 5, 2005; Accepted for publication: January 27, 2006; Published on Web: June 5, 2006)

分子設計や薬物設計においては、コンピュータを用いた効率的な解析が求められており、そのためのコンピュータシステムが数多く研究、開発されている。我々の研究室においても、ケモインフォマティクス、ケモメトリクスに関連する多くの研究を行っており、それらの解析を行うプログラムについても開発を進めている。ToMoCo (The total system for molecular designs by the computer chemistry laboratory) はこれらのコンピュータプログラムをひとつにまとめた分子設計のための統合システムであり、現在までに CoMFA 法を中心とした定量的構造活性相関解析 (QSAR) に関する手法が実装されている。本論文では ToMoCo に実装されている手法のうち、分子構造重ね合わせ機能、新規薬物候補構造自動創出機能、CoMFA 領域選択機能について説明し、これらの手法を用いて COX-2 阻害剤の解析を行った結果を示す。COX-2 阻害剤 54 サンプルに対して、構造重ね合わせを行うことにより活性配座を推定し CoMFA 法による 3D-QSAR 解析を行った結果、 $R^2=0.922$ 、 $Q^2=0.653$ の CoMFA モデルが得られた。そして、そのモデルを評価基準として用いて新規薬物候補構造の自動創出を行い、予測活性値 7.11 から 7.87 までの 200 個の構造を得た。また CoMFA 領域選択を行った結果、変数の数が 6 分の 1 以下に削減され、 $R^2=0.938$ 、 $Q^2=0.751$ というシンプルで予測的な QSAR モデルが構築された。

キーワード: 分子設計, CoMFA, 化学構造重ね合わせ, 領域選択, 構造生成, COX-2 阻害剤

1 概要

分子設計や薬物設計を行うにあたり、化学者が取り扱わなければならないデータや情報の量は増加の一途を辿っている。コンピナトリアルケミストリの出現や各種実験、測定装置の技術的な進歩によって、日々大量の化合物が合成、測定され、その情報がデータベースへと蓄積されている。開発期間の短縮、コストの削減を目指し効率的な分子設計を行うためには、これらの情報のなかから必要な知識を見つけ出し、有効に利用することが求められる。しかし大量のデータ中から必要な知識のみを抽出することは容易ではなく、コンピュータを用いたさまざまなデータ処理技術の適用が必要不可欠となっている。また見つけ出された知識を用いて設計の効率化を実現するためにコンピュータを

利用することの重要性も認識されはじめている。このように情報工学的な手法を化学の問題へと適用する分野は、ケモインフォマティクスあるいは情報化学と呼ばれており、その必要性が認識され注目を集めている。

当然のことながら、ケモインフォマティクス手法を用いた研究、解析においては、コンピュータソフトウェアが重要な役割を果たす。そのため、多くの企業がこの分野へ進出しており、様々な種類の市販ソフトウェアが利用可能となっている。また大学などで開発されたソフトウェアの一部は無償で公開されており、自由に利用することができる。しかしながら、必要とする機能を備えたソフトウェアが常に入手可能であるとは限らず、非常に高価であったり、希望する解析を行うことが出来なかつたりといった理由により、使用

できるソフトウェアが制限されているのが現状である。このような状況においては、自社内あるいは研究室内において要求を満たすソフトウェアを開発することが求められる。特に新規な手法を利用したり、そのアルゴリズムを改良したりする場合には必ずプログラミングが必要となる。

そこで我々の研究室においては、目的のために必要となるソフトウェアについて、研究室内で独自に開発を行っており、またそのソフトウェアについては、ある程度完成したところで外部へと公開している。これまでに、有機化合物の合成経路設計システムや反応予測システム [1-3]、自動構造解析システム [4]、ケモメトリックスシステム [5, 6]、などに取り組み一定の成果をあげている。合成経路設計システムは、有機化合物の合成経路をコンピュータ支援により設計するためのシステムである。目的とする構造を入力することにより、考えられる出発物質と合成経路が自動的に複数提案される。自動構造解析システムは、構造が未知の化合物の構造を解析するためのシステムであり、各種スペクトルなどの情報を入力として、その情報に矛盾しない化学構造を出力する。ケモメトリックスシステムは、各種統計解析などを行うためのシステムであり、PLS 法やニューラルネットワークによる解析が可能のため、主に構造活性相関解析 (QSAR) の分野などで利用されている。

我々の研究室においてはこれらの他に、CoMFA 法を中心とした 3 次元構造活性相関に関する研究を行っており、関連するプログラムも開発されている。本論文では、これらのプログラムを統合する形で開発が行われている、分子設計統合システム ToMoCo (The Total System for Molecular Designs by the Computer Chemistry Laboratory)[7] についてその詳細を示す。ToMoCo は分子設計に関する一連の解析をひとつのシステム上で行うことを目的として開発しているソフトウェアであり、現在のところ CoMFA 法による 3 次元構造活性相関、HNN を用いた化学構造の重ね合わせ、LigConstructor による新規薬物候補構造の自動創出、GARGS 法による領域選択などの機能が実装されている。

本論文では、ToMoCo を用いた解析例として COX-2 阻害剤に関する解析の結果を示す。COX-2 阻害剤 54 化合物について、構造重ね合わせによる活性配座の推定を行った後、CoMFA 法によるモデルの作成を行った。そしてそのモデルを評価基準とし、高活性を示すと考えられる構造の自動生成を行った。また、遺伝的

アルゴリズム (GA) を用いて変数選択を行い、QSAR モデルにとって重要な領域の選択を行った結果についても示す。

2 手法

2.1 CoMFA 法 [8]

CoMFA (Comparative Molecular Field Analysis) 法は 3D-QSAR 解析における代表的な手法である。記述子の計算方法が単純であり、汎用性も高いことから、さまざまな種類の化合物群の QSAR 解析に適用され成果が報告されている。CoMFA 法における記述子の計算方法の概要を Figure 1 に示す。まず解析の対象とする化合物全体を含むように直方体領域を定義し、その内部に一定間隔の格子点を設定する。そして各格子点にプローブ原子 (probe atoms) と呼ばれる仮想的な原子を配置し、対象分子との間の立体的、静電的相互作用を求めて構造記述子とする。プローブ原子としては、通常電荷が +1 の sp^3 炭素が用いられる。

この構造記述子を説明変数とし、目的変数である活性値との間で PLS 法による回帰分析を行うことによって、QSAR モデルを構築する。こうして得られる相関モデルを用いることで、活性が未知である化合物の活性を予測することが可能となる。また回帰式における回帰係数の等高線図を描くことで、活性に大きな影響を与える領域を特定することが可能である。そしてこの結果に対して考察を加えることによって、より理想的な活性を示す化合物の設計を行うための指針となる情報を得ることができる。

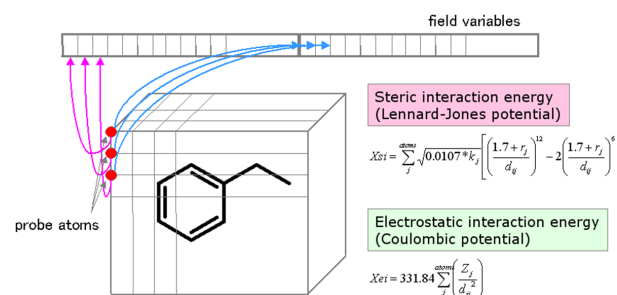


Figure 1. Summary of CoMFA method.

2.2 HNN を用いた分子構造重ね合わせ手法

CoMFA 法をはじめとする多くの 3D-QSAR 解析手法においては、解析の前処理として対象とする化合物について適切な重ね合わせ処理を行うことが必要となる。この分子構造の重ね合わせ処理は、QSAR 解析の結果に非常に大きな影響を与えるため、安定で予測的なモデルを構築するためには適切な重ね合わせを行うことが重要である。そのため、分子構造の重ね合わせを行うための手法については数多く提案されているが、重ね合わせの目的や対象とする分子の特徴などによって最適な重ね合わせ手法は異なるため、全ての場合に適用可能な手法は存在しないと言える。

そこで我々は、3D-QSAR 解析に適した重ね合わせを行うための手法として、HNN (Hopfield Neural Network)[9] による分子構造重ね合わせ手法を提案している [10]。この手法では、まず重ね合わせを行う 2 つの分子上に疎水性、水素結合供与性、水素結合受容性などの特徴を定義し HNN を用いてそれらの対応付けを行う。そして対応付けられた部位同士の距離の 2 乗和を最小化することによって、3 次元構造の重ね合わせを実現する。

HNN は人工ニューラルネットワークの一種であり、主に組み合わせ最適化問題において高い最適化能力が証明されている。HNN におけるニューロンは互いに全て結合されており、各ニューロンは 0 あるいは 1 の値を持っている。HNN の学習においては、まず最小化するエネルギー関数を次の形式で定義する。

$$E = -\frac{1}{2} \sum_i \sum_{j \neq i} W_{ij} S_i S_j + \sum_i \theta_i S_i \quad (1)$$

ここで W_{ij} はニューロン i と j との間の結合に対応する重み、 S_i はニューロン i の値、 θ は閾値、 E はエネルギー関数値である。すべてのニューロンの値をランダムに初期化した後、次式のニューロン更新ルールに従ってすべてのニューロンの値を繰り返し更新することによりエネルギー関数値の最小化を行うことが可能である。

$$\begin{aligned} \sum_i W_{ij} S_j - \theta_i > 0 : S_i = 1 \\ < 0 : S_i = 0 \end{aligned} \quad (2)$$

ここで W_{ij} および θ_i については、エネルギー関数の定義より自動的に決定される。

HNN は高い最適化能力を示すが、しばしば局所解へと収束してしまう現象が見られる。そこで、HNN

の改良としてボルツマンマシンと呼ばれる方法が提案されている。これは、HNN の学習に確率的な要素を取り入れた手法である。式 (2) によるニューロン値の更新において、下記の式による計算結果に基づいて値の更新をするかどうかを決定する。

$$p = 1 / (1 + \exp(\Delta E / T)) \quad (3)$$

ここで ΔE は更新を行った場合のエネルギー関数値 E の変化量、 T はネットワーク温度である。各ニューロンの更新において [0 1] の一様乱数を発生させ p が乱数値より大きい場合にのみ値の更新を行う。学習の始めの段階では T の値を比較的大きく設定し解空間全体の探索を試み、学習が進むにつれ T の値を減少させていき局所的な探索を行う。これにより精度の悪い局所解へ落ち込む可能性を減らすことが可能となる。本重ね合わせ手法においては、このボルツマンマシンを利用している。

2.3 新規薬物候補構造自動創出機能

CoMFA 法などにより QSAR モデルが得られた場合、そのモデルを用いてより理想的な活性を示すと考えられる化合物を探索することが可能となる。これは、化学者の提案する化合物について QSAR モデルによる予測活性値を計算することによっても可能ではあるが、コンピュータを用いることでさらなる効率化が実現する可能性がある。つまり、コンピュータによって自動的に多数の候補となる化学構造を生成し、それらの予測活性値を計算していくのである。

このような探索を行うための手法として、LigBuilder[11] が提案されている。LigBuilder は、進化的計算手法を用いて新規薬物候補構造の自動創出を行うシステムである。標的タンパク質の活性部位周辺に候補構造の基となる比較的小さな構造を配置し、それを成長させることでレセプターに適応した新規薬物候補構造を自動的に生成する。X 線結晶構造解析などにより標的タンパク質の 3 次元構造およびその活性部位が明らかになっている場合には、ポケットに適した有力な構造を得ることが可能である。しかし一方で、標的タンパク質のポケット情報が得られていなければ候補構造を作成することができないという問題がある。

そこで我々は、このような状況においても新規薬物候補構造の創出が可能なシステムとして LigConstructor[12] を提案している。LigConstructor では、CoMFA モデルによる予測活性値を評価値として用いて構造を

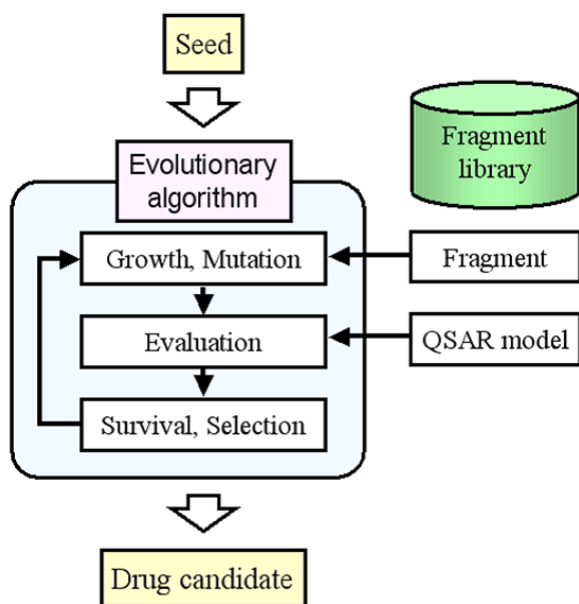


Figure 2. Summary of the LigConstructor.

進化させることによって、活性が高いと期待される薬物候補構造を生成する。そのためタンパク質の3次元構造が明らかになっていない場合においても適用することが可能である。

Figure 2 に LigConstructor の処理の流れを示す。

システムへの入力、提案される構造のもととなる適当な初期構造である。この構造に対して、構造結合と突然変異の処理が行われ、新規な候補構造が複数生成される。構造結合では、フラグメントライブラリに格納されているフラグメントがランダムに選択され、構造の一部に結合される。突然変異では、構造中の原子がランダムに選択され、その原子の種類が変更される。こうして生成された各候補構造について、QSARモデルにより予測活性値が計算される。そして、その値をもとにして淘汰と選択が実行される。淘汰では、予測活性値が他と比較して低い個体が削除され、選択では、予測活性値が高い個体の中から次の世代の親となる個体を選ばれる。以上の処理を指定された回数繰り返すことにより、最終的に高い予測活性値を示す構造が新規薬物候補構造として提案される。

2.4 遺伝的アルゴリズムを用いた領域選択機能

CoMFA 法では、対象とする分子群をすべて含む直方体を定義し、その内部に設定された各点において構

造記述子を計算する。しかし多くの場合、活性値を決定するために重要な役割を果たす変数はごく一部であり、その他の領域に対応する変数はあまり多くの情報を持っていないと考えられる。PLS モデリングにおいて、説明変数のなかにこのような冗長な記述子を含めることは、モデルの精度や予測性を悪化させる要因となるため、あらかじめ不要であることがわかっている領域については、モデルから除外することが望ましい。

そこで我々は、活性の説明に重要な役割を果たしている領域を自動的に選択するための手法として、GARGs (genetic algorithm-based region selection) 法を提案している [13, 14]。GARGs 法では、CoMFA フィールドをいくつかの小領域に分割し、GA を用いて活性を説明するために有効な領域の探索を行う。各小領域を遺伝子の各ビットに割り当て、交差、突然変異、評価、淘汰、選択を繰り返すことによって選択領域の最適化を行う。各遺伝子の評価値としては、遺伝子において選択されている領域の記述子のみを用いた場合の PLS モデリングにおける Q^2 値を使用する。GARGs 法を用いて領域を選択することによって、シンプルで予測性の高い構造活性相関モデルを構築することができ、より精度の高い分子設計が可能となることが期待される。

3 COX-2 阻害剤の解析

3.1 COX-2 阻害剤

ここでは前章において述べた手法を用いて、COX-2 (cyclooxygenase-2) 阻害剤の解析を行った結果を示す。COX はアラキドン酸をプロスタグランジン H_2 に変換する生体内酵素であり、体内の多くの細胞に構造的に発現する COX-1 と、誘導酵素である COX-2 などに分類される。COX-2 を選択的に阻害することにより、副作用の少ない抗炎症剤が開発できると考えられており近年さかんに研究が行われている [15, 16]。我々も COX-2 阻害剤に関する QSAR 解析を行っており、その結果を報告している [17]。骨格の異なる 3 種の化合物シリーズからなる COX-2 阻害剤 54 化合物について、配座解析を行った後、HNN を用いた構造重ね合わせ手法を適用し、活性配座の推定を行った。そしてその配座を用いて CoMFA モデルを作成し、 $R^2=0.922$ 、 $Q^2=0.653$ という良好なモデルが得られている。本論文ではこの結果をもとにして行った新規薬物候補構造の自動創出と、GARGs 法による領域選択の結果について示す。

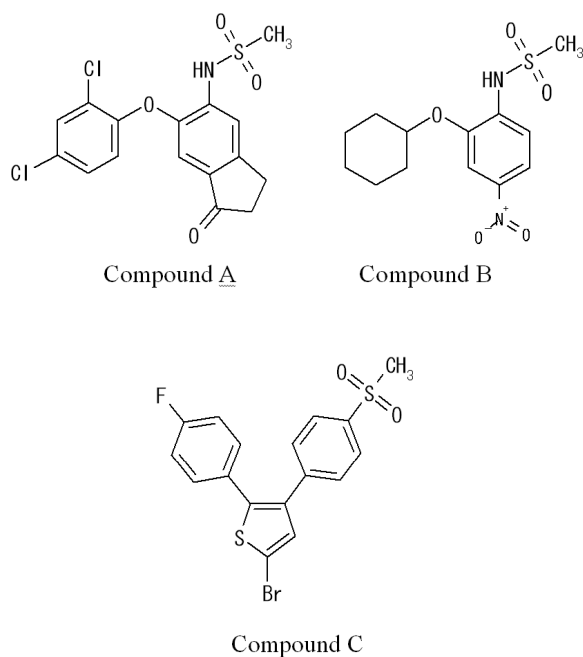


Figure 3. COX-2 inhibitors.

Figure 3 に代表的な 3 種の化合物を示す。解析に用いたその他の化合物は、これらの誘導体である。

3.2 CoMFA の結果

まず、この 3 つの構造について SPARTAN[18] を用いて配座探索を行った。その結果それぞれの構造に対して、57、21、47 種類の配座が得られた。次に、化合物 A と B、B と C の間でのすべての配座の組み合わせについて、HNN による構造重ね合わせを実行し、もっとも類似度の高い配座を活性配座として選択した。そして、決定された 3 つの構造の配座をもとに、他の化合物の配座を決定し、CoMFA 法による QSAR モデルの構築を行った。CoMFA フィールド値の計算においては、電荷+1 の sp^3 炭素をプローブ原子として用いている。フィールド値の計算範囲は各軸方向に 2Å の余裕をとり、 $20\text{\AA} \times 16\text{\AA} \times 19\text{\AA}$ 、計算間隔は 1Å とした。得られた CoMFA 記述子を用いて PLS 法によるモデルを作成した結果、 $R^2=0.922$ 、 $Q^2=0.653$ の 4 成分モデルが得られた。これは文献 [17] において得られているモデルと同じものである。

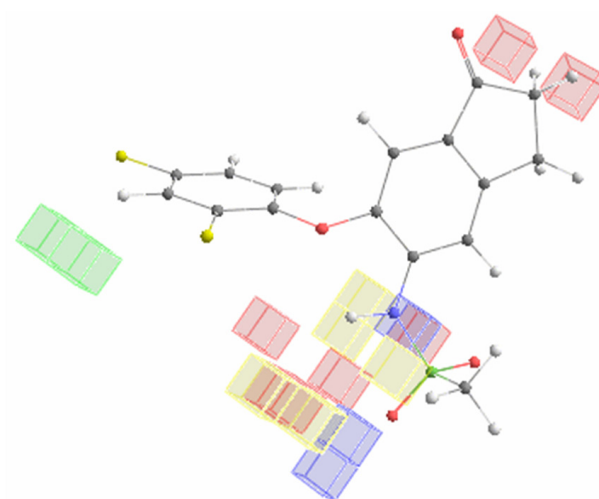


Figure 4. The contour plot of the PLS model.

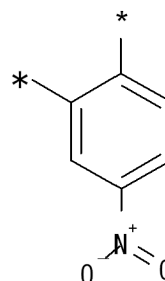


Figure 5. The initial structure for the LigConstructor.

Figure 4 にこのモデルの回帰係数値の等高線図を示す。黄色が立体的相互作用の正の係数、緑が負の係数、赤色と青色が静電的相互作用の正と負にそれぞれ対応している。また図中の分子は Figure 3 の分子 A である。COX-2 に関してはいくつかの結晶構造が知られているが、この等高線図については文献 [17] においてそのなかのひとつ (PDB コード: 6COX) と比較することによって妥当性が示されている。

3.3 新規薬物候補構造の自動創出

得られた CoMFA モデルを評価基準として用い、活性が高いと思われる構造を LigConstructor によって自動的に生成した結果を示す。

初期構造としては Figure 3 の分子 B の一部を用いた。その構造を Figure 5 に示す。図でアスタリスクが描かれている箇所が結合部位であり、フラグメントをこの位置に追加していくことで構造の進化が行われる。

LigConstructor を実行するためのパラメータは、予備実験を行いその結果をもとに決定した。各世代における個体数は 200、最大世代数は 10、構造出力数は 200 とした。構造出力数とは、最終的に提案される構造の数であり、探索中に生成されたすべての構造について類似度によるクラスタリングが行われ、その結果をもとに多様な構造が選択されて提案される。またエリート率を 10% とし、各世代で評価値が上位 10% の個体については、無条件で次世代へ引き継ぐものとした。構造結合率と突然変異率は共に 50% とし、フラグメントの付加と原子種の変更を同じ割合で行った。類似構造カット率は 90% に設定した。これは、集団内の構造の多様性を保つために行う類似構造カット処理の閾値を指定するものである。一定以上の類似度を示す構造が生成された場合には、その一方を削除することによって集団の多様性を保ち、幅広い範囲の探索を実現している。また分子量の範囲を指定して、一定の分子量を持つ構造のみを生成することも可能であるが、今回は制限を設けていない。

構造結合において用いるフラグメントとしては、Figure 6 に示す構造を含む 47 の構造を指定した。フラグメントライブラリは、構造生成の対象に応じてユーザが自由に設定することが可能であるが、今回は Lig-

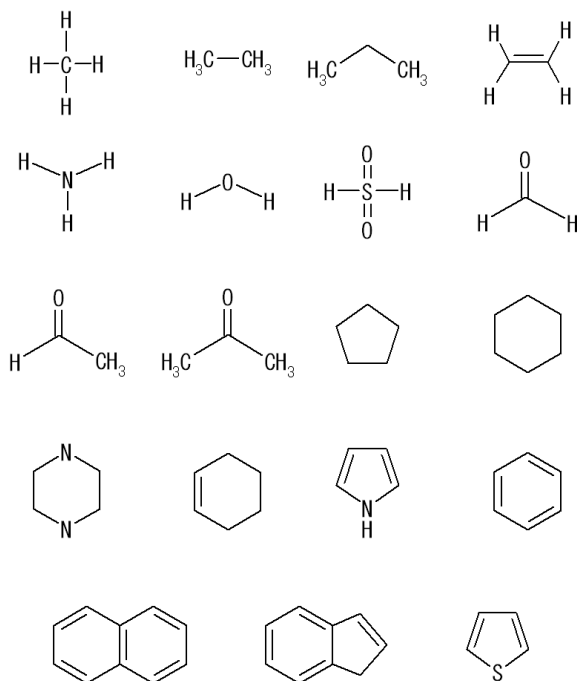


Figure 6. Fragment structures used in the LigConstructor.

Constructor において用意されている標準的なフラグメントのみを用いた。

以上のような設定で候補構造の生成を行った結果、予測活性値が 7.11 から 7.87 までの 200 個の構造が得られた。

Figure 7 に、得られた 200 構造の活性値のヒストグラムを示す。活性値が 7.1 から 7.4 の間に多くの構造が集まっており、活性がそれ以上の構造については次第に少なくなっている。また得られた構造の活性値の平均は 7.36 であった。QSAR モデルを作成するために用いた化合物についての活性値の範囲は 3.5 から 8.7 であり、平均値は 6.75 であることから、高い活性を示す構造が選択的に生成されていることがわかる。

Figure 8 に LigConstructor によって生成された構造とその予測活性値の例を示す。図の右下部分が初期構造として与えた構造である。初期構造をもとにして、多様な構造が生成されている様子を確認することができる。これらの構造を参考に化学者が最適化を行うことで、医薬品開発の効率化が期待される。

しかしながら、ヘテロ原子の数が多すぎる構造や、複雑すぎる構造などが少なからず生成されており、これを防ぐためになんらかの制約を加える必要があると考えられる。この問題は今後の課題である。

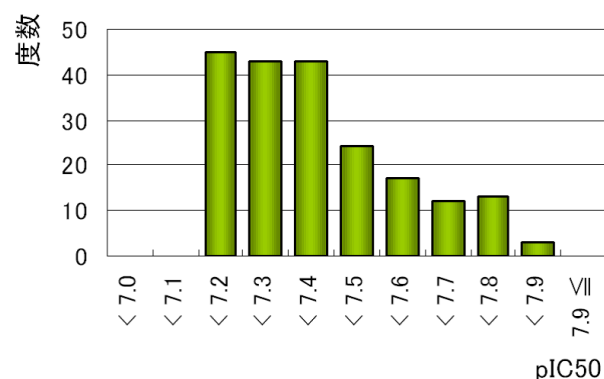


Figure 7. Histogram of activity values of the generated structures.

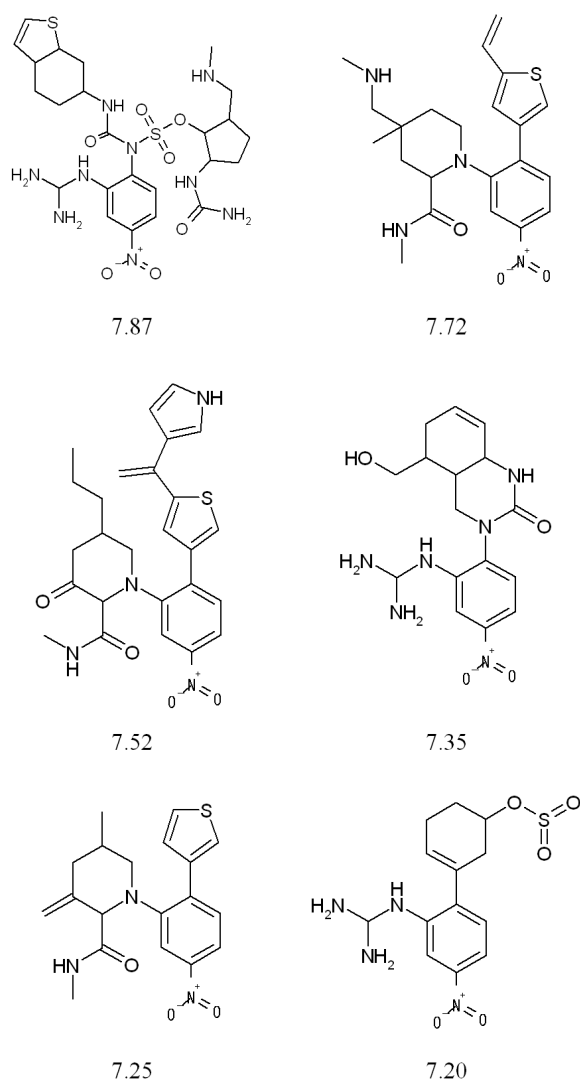


Figure 8. Generated structures and their activities.

3.4 GARGSによる領域選択

CoMFA 法においては活性値と関係のない領域に対応する記述子を用いることは、モデルの精度、予測性に対して悪影響を与えるため、適切な領域選択を行うことが重要となる。ここでは、COX-2 阻害剤のデータに対して GARGS 法による領域選択を行った結果を示す。前節と同様の設定で CoMFA フィールド値を計算した後 GARGS 法を適用した。

GARGS 法においては、はじめに CoMFA フィールドをいくつかの小領域に分割する。フィールド計算の範囲は前述のように $20\text{\AA} \times 16\text{\AA} \times 19\text{\AA}$ とほぼ立方体であるため、ここでは X、Y、Z の各方向について等しく 4 分割とした。つまり、各小領域は一辺が 4\AA から

5\AA の直方体となり、領域の数は 64 となる。GA の学習では、各世代の個体数を 200 個体とし、100 世代までの最適化を行った。自然淘汰率は 20% であり、各世代で評価値が下位 20% の個体は淘汰される。その他、突然変異率は各ビットに対して 5%、各世代での交差の数は 50 とした。評価値は、選択された領域に対応する変数のみを用いた PLS 回帰において、leave-one-out 法によるクロスバリデーションで求められる Q^2 値を用いた。PLS 法における最大成分数は 5 に設定している。

また保護個体数は 10 に設定した。これは GARGS 法に特有の設定であり、保護処理において何個の個体を保護するのかを指定するためのものである。領域選択においては、評価値である PLS モデルの Q^2 値を最大化させるとともに、選択される領域の数をある程度少なくすることが要求される。しかし、CoMFA においては、変数の数が多いほど Q^2 値が増加する傾向があるため、 Q^2 値のみを評価値として GA を行った場合、選択される領域の数が非常に多くなってしまふことがある。この問題に対処するために行うのが保護処理である。選択領域数に注目して各個体を再評価し、選択されている領域数が少ない個体を保護し、優先的に次世代へと生き残らせる処理である。

GA による領域選択の結果を Table 1 に示す。選択領域数 15 から 7 までの 8 つの候補が得られ、最大の評価値は 0.753、最小は 0.743 であった。なお PLS の最適成分数はいずれのモデルも 5 となっている。これらのモデル中から、選択された領域数と R^2 値、 Q^2 値との関係をもとに検討した結果、選択領域数 10 のモデルを最終モデルとして採用することに決定した。このモデルの R^2 は 0.938、 Q^2 は 0.751 であり、すべての領域を用いた場合の PLS モデルでの値、 $R^2=0.922$ 、 $Q^2=0.653$ と比較して明らかな向上が見られた。よって、領域選択を行うことでモデルの精度、予測性を改善することが出来たと言える。また、選択された領域の数は 10 であり、全領域数 64 と比較すると 6 分の 1 以下に削減されている。Figure 9 に最適モデルにおいて選択された 10 の領域を示す。図中の分子は CoMFA モデリングに用いた全ての分子を重ね合わせたものである。直方体の CoMFA フィールドのうち、これらの分子周辺の領域が主に選択されていることを確認できる。図中央については領域が選択されていないが、この部分はほぼ全ての分子に共通の骨格であり、この領域が活性値に与える影響は少ないと考えられるため、妥当な結果であると言える。

Table 1. The result of the region selection by the GARGs

選択領域数	R^2 値	Q^2 値
15	0.937	0.753
14	0.937	0.753
13	0.937	0.753
12	0.937	0.753
11	0.939	0.752
10	0.938	0.751
8	0.934	0.743
7	0.922	0.743

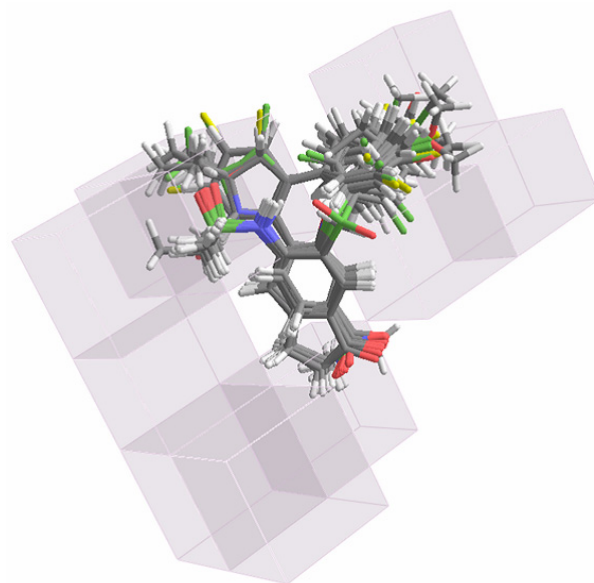


Figure 9. The selected regions by the GARGs.

これらのことから、シンプルで精度、予測性の高い CoMFA モデルを作成するための手法として、GARGs 法による領域選択は有効であると結論できる。

4 まとめ

本論文では、CoMFA 法などの 3D-QSAR 解析の前処理として重要な分子構造重ね合わせ手法と、それに続く CoMFA 法によるモデル構築、およびそのモデルを評価基準とした新規薬物候補構造自動創出を行う LigConstructor、領域選択を行うための GARGs 法について示した。そして実行例として、COX-2 阻害剤の QSAR 解析を行った結果を示した。

COX-2 阻害剤の解析では、まず複数の配座間での重ね合わせを行い、活性配座を推定した。そしてその配座を用いて CoMFA 法による解析を行い、 $R^2=0.922$ 、 $Q^2=0.653$ の CoMFA モデルを得た。さらに、このモデルを評価基準として用い、LigConstructor によって高活性を示すと思われる構造を自動的に多数生成した。その結果、予測活性値が 7.11 から 7.87 までの 200 個の構造を得ることができた。

また、GA を応用した領域選択手法である GARGs 法によって、変数の数を 6 分の 1 以下に削減し、よりシンプルで予測的な QSAR モデルが構築可能であることを示した。領域選択によって得られたモデルは $R^2=0.938$ 、 $Q^2=0.751$ であり、通常の CoMFA モデルと

比較して良好なものであった。

これらの解析は、本研究室で開発を進めている分子設計統合システム ToMoCo で行ったものである。ToMoCo は Windows 上で動作するソフトウェアであり、コンピュータに比較的なじみのない化学者であっても容易に利用することができる。また複数の機能をひとつのシステム上に統合することによって、ソフトウェア間でのデータの変換などの作業が必要なくなり、一連の解析をスムーズに行うことのできる環境が構築されている。現在までのところ、過去に開発された手法を中心に実装を行っているが、今後は新たに開発される手法を実装するためのプラットフォームとしての役割も期待している。なお ToMoCo は製品化されており、ケムインフォナビ社 [19] によって販売されている。

参考文献

- [1] 船津公人, 佐々木慎一, *AIPHOS* - コンピュータによる有機合成経路探索, 共立出版株式会社 (1994).
- [2] 太田圭輔, 船津公人, 塚本幸治, 第 27 回情報化学討論会 講演要旨集, 57-60 (2004).
- [3] 勝見広幸, 太田圭輔, 山本史雄, 楠瀬直人, 船津公人, 第 27 回情報化学討論会 講演要旨集, 61-62 (2004).

- [4] 橘大樹, 船津公人, 増井秀行, 日本コンピュータ化学会 2003 春季年会 講演予稿集, 2P14 (2003).
- [5] T. Tanada, M. Arakawa, R. Nishimura, K. Funatsu, *J. Comput. Aided Chem.*, **1**, 35-46 (2000).
- [6] 山田吉朗, 荒川正幹, 船津公人, 第 27 回情報化学討論会 講演要旨集, 121-122 (2004).
- [7] 荒川正幹, 溝淵創一郎, 船津公人, 第 27 回情報化学討論会要旨集, 119-120 (2004).
- [8] R. D. Cramer III, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.*, **110**, 5959-5967 (1988).
- [9] J. J. Hopfield, D. W. Tank, *Biol. Cybern.*, **52**, 141-152 (1985).
- [10] M. Arakawa, K. Hasegawa, K. Funatsu, *J. Comput. Aided Chem.*, **2**, 29-36 (2001).
- [11] R. Wang, Y. Gao, L. Lai, *J. Mol. Model.*, **6**, 498-516 (2000).
- [12] 竹内英憲, 荒川正幹, 船津公人, 第 25 回情報化学討論会講演要旨集, 75-78 (2002).
- [13] T. Kimura, K. Hasegawa, K. Funatsu, *J. Chem. Inf. Comput. Sci.*, **38**, 276-282 (1998).
- [14] K. Hasegawa, T. Kimura, K. Funatsu, *J. Chem. Inf. Comput. Sci.*, **39**, 112-120 (1999).
- [15] D. L. Simmons, R. M. Botting, T. Hla, *Pharmacol. Rev.*, **56**, 387-437 (2004).
- [16] R. Garg, A. Kurup, S. B. Mekapati, C. Hansch, *Chem. Rev.*, **103**, 703-731 (2003).
- [17] M. Arakawa, K. Hasegawa, K. Funatsu, *J. Comput. Aided Chem.*, **3**, 63-72 (2002).
- [18] SPARTAN 5.0, Wavefunction, Inc., 18401 Von Karman Avenue, Suite 370, Irvine, CA, 92612.
- [19] (有)ケムインフォナビ,
<http://www.cheminfornavi.co.jp/>

Development of the Total System ToMoCo for Molecular Design

Masamoto ARAKAWA* and Kimito FUNATSU**

Department of Chemical System Engineering, The University of Tokyo

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan

**e-mail: arakawa@chemsys.t.u-tokyo.ac.jp*

***e-mail: funatsu@chemsys.t.u-tokyo.ac.jp*

In the field of molecular design or drug design, efficient methodologies with computer systems have been desired and various computer programs have been investigated. In our laboratory, we are researching chemoinformatics and chemometrics, and have been developing computer programs for these analyses. The ToMoCo (Total System for Molecular Designs by the Computer Chemistry Laboratory) is a total system for molecular design developed in our laboratory by integrating these in-house programs. The ToMoCo has functions that are related to quantitative structure-activity relationships such as CoMFA (comparative molecular field analysis), molecular structure alignment method using Hopfield neural network, region selection method by GARGS (genetic algorithm-based region selection), and automatic structure generation method by Lig-Constructor. In this paper, we describe these functions and the result of QSAR analysis of Cyclooxygenase-2 (COX-2) inhibitors using the ToMoCo.

Keywords: Molecular design, CoMFA, Structure alignment, Region selection, Structure generation, COX-2 inhibitors