

1. その他のコンピュータによる構造-活性相関アプローチ

コンピュータによる構造-活性相関手法として、今まで上げた範疇に属さないアプローチも多数ある。また、本著で説明した構造-活性相関手法自体もより小さな構造-活性相関手法の集合体であることも多い。ここでは、本章では取り上げられる事になかった大きな構造-活性相関手法の一機能として組み入れられている単純なアプローチから、他のアプローチとは思想も手続きも全く異なるアプローチについてその代表的なものをまとめる。

2. コンピュータによるアプローチの分類

コンピュータによる構造-活性相関手法は一般的に①データベースを利用するアプローチ、②解析手法主体のアプローチとに大きく二分類される。ここではコンピュータの観点から簡単にまとめてみる。これらの中で、他の章では取り上げられなかったか、取り上げられてもこの章で改めて詳しい説明が必要なアプローチについて詳しく述べる。

2.1 データベース利用を主体としたアプローチ

データベースの種類は多数ある。データ内容だけみても文献、化合物、物性、薬理活性/毒性等様々なものがある。データベースはコンピュータ技術の観点から見ると奥が深く、これだけでも簡単に一冊の本になってしまう。特に化学分野のデータベースは他の分野とは異なり、化合物構造式を扱う点で非常に特殊な技術が要求される。興味のある人は是非専門書を一読する事を勧める。

①構造-活性相関やドラッグデザインでのデータベース利用形態

構造-活性相関でのデータベース利用事例は幾つか存在するが、一般的には他の解析手法に組み込まれ、一つの要素技術として利用される事が多い。当初は独立して利用されるが、他の解析手法が進歩するにつれて基本機能として組み込まれてしまう事が多い。最近展開された三次元化合物検索は当初独立したドラッグデザイン手法として開発されたが、現在では De novo デザインの基本技術の一つとなっている。ここではデータベース技術を用いた三種類のアプローチについて簡単にまとめる。

・部分構造出現頻度の利用

構造-活性相関で利用される最も古典的なアプローチは部分構造の出現頻度を調べるアプローチである。薬理活性データベースから目的活性を有する化合物群を取り出し、取り出された化合物群から様々な部分構造を取り出して、その出現頻度を調査する。出現頻度の高いものが目的薬理活性に重要な部分構造である。この場合解析対象構造式は二次元であること、および出現頻度だけで部分構造間の相互位置関係を考慮していないという点で構造-活性相関の観点上解析精度は高くない。しかし、全体的な傾向をつかむという点で貴重である。

最近再びこの部分構造検索機能が注目されている。既に第九章のコンビナトリアルケミストリ/IITS で述べたが、分子多様性の評価データベース利用されるパラメータとして、この部分構造検索から得られるパラメータが利用されている為である。

・類似化合物検索の利用

また、化合物データベース検索技術の進歩に伴い頻繁に利用されるようになった基本的なアプローチとして、類似化合物検索を行ってドラッグデザインを行うアプローチがある。これは、化合物の部分構造を基本とした様々な情報から検索化合物とデータベース中の化合物との類似度を求め、類似度の高い化合物群から順に取り出してリード候補化合物を選択するアプローチである。これは二次元化合物データベースを利用して行われる。

類似化合物検索技術の基本は第九章で述べた分子多様性の評価技術の応用で展開される。この類似化合物検索で常に議論となる問題は類似の尺度をどの観点で定義するかである。この定義は、類似検索自体の目的やこの検索結果をどう使うかなどの差異により大きく変化し、定義が変化すると用いるパラメータの種類や計算手法等が多様に変化する。また、単純な計算では定義出来ないような複雑な類似概念を扱う場合は、人工知能技術を用いるなどの手法自体の変更も必用となる。

・三次元化合物検索

現在、データベースで最も良く利用される形態は三次元データベースを用いたドラッグデザインである。このアプローチは当初、先に説明した二次元類似化合物検索によるドラッグデザインの拡張系として展開されたが、De novo デザインの一つの基本技術として利用される用になり急速に展開/普及してきた。三次元化合物データベース自体は高度のデータベース検索技術を必要とする。この検索にもパターン認識の技術を利用するアプローチがある。本著では三次元化合物検索のアルゴリズムには触れず、専門書にまかせる。ここでは実際に三次元化合物データベースを用いてドラッグデザインを行うという、利用者の観点からまとめる。

②ルールデータベース（人工知能：Artificial Intelligence）の利用

データベースとはかなり異なるが、ルールデータベース（知識ベース）というものがある。いわゆる人工知能に利用されるルールを集めたものを指している。人工知能は数式化出来ない、あるいは明確な形で手続き化出来ない複雑な事項をコンピュータで扱う時の技術である。この人工知能は奇妙に化学と縁が深い。人工知能の最初の本格システムとして開発されたものは DENDRAL というスペクトル解析システムである。このシステムは当時新たに開発された人工知能言語である LISP を用いてカリフォルニア州立大学スタンフォード分校の人工

知能研究グループにより構築され、人工知能システムのルーツとして位置づけられている。この DENDRAL システムは種々のスペクトルデータを入力データとし、可能性の高い化合物構造式を推測するものである。

すぐ後には合成デザインを行うシステム LHASA が E.J.Corey および W.T.Wipke を中心とするグループにより開発されている。この LHASA システムは科学分野で利用される一般的な言語の FORTRAN を用いていたため、人工知能分野で大きく取り上げられる事はないが、内容そのものは人工知能そのものである。LHASA は化学分野におけるコンピュータ利用の基本を確立した点で非常に有意義なシステムである。グラフィックを利用した化合物構造式の扱い、結合表を用いた化合物検索技術、その他数多くの先進的な技術（現在では基本技術となっている）がこの LHASA グループにより開発された。この LHASA システムの開発に関与した人たちの多くが企業や大学等でコンピュータケミストリの発展に貢献している。単なる技術のみならず、人材といった観点からも LHASA の果たした役割は大きい。

この人工知能技術は構造-活性相関やドラッグデザイン分野でも様々な形で利用されている。例えば、基本分野に関係するシステムとしては、二次元化合物から三次元化合物へと立ち上げる作業に人工知能を利用した AIMB、CONCORD、WIZARD、COBRA、CORINA 等のプログラム群^{*1~5)}がある。

個々の化合物の三次元立ち上げは一般的に分子力学法が利用される。しかし、分子力学法では計算時間がかかりすぎ、その安定構造式も初期座標の差異により大きく変化する。従って、化合物データベースのような数千~数十万といった大量の化合物群の三次元座標をまとめて一気に計算することは不可能であり、しかも化合物の構造的な統一性が要求される場面での利用には不適である。そこで、人工知能の力を借りることで高速、且つかなりの精度で初期座標に左右されずに二次元から三次元へと統一性を持った状態での構造式立ち上げが実現された。

現時点では三次元化合物データベースの殆どがこの人工知能システムを用いて座標を創出している。歴史的な背景もあり現在、殆どのデータベースは CONCORD を利用している、しかし一部のデータベースではその計算精度や計算可能な化合物の数や種類等の問題から CORINA 等を用いているものもある。一般の解析においてはデータの統一性が要求される。従って、複数の三次元化合物データベースを用いる時は、その三次元座標データがどの手法で求められたかに留意する必要がある。

* 1 : Wipke, W. Todd; Hahn, Mathew A. in "Artificial Intelligence Applications in Chemistry"; Hohne, B. Pierce, T., Ed.; ACS Symposium Series 306, American Chemical Society: Washington, D.C., 1986. "Analogy and Intelligence in Model

Building”, pp.136-146.

* 2 : Rusinko III, A., Sheridan, R. P., Nilakantan, R., Haraki, K. S., Bauman, N., and Venkataraghavan, R. (1989). Using CONCORD to construct a large database of three-dimensional coordinates from connection tables. *J. Chem. Inf. Comput. Sci.* 29, 251-255.

* 3 : Dolata, D.P., Leach, A.R. and Prout, K., *J. Comput.-Aided Mol. Design*, 1 (1987) 73.

* 4 : Leach, A.R. and Prout, K., *J. Comput. Chem.*, 11 (1990) 1193.

* 5 : Gasteiger, J., Rudolph, C., and Sadowski, J. (1990). Automatic generation of 3D atomic coordinates for organic molecules. *Tetrahedron Comp. Mrth.* 3 (6C), 537-547.

構造-活性相関解析そのものに人工知能を利用した事例は少ないが、最初の事例としては Teri E.Klein らによる レセプターマッピングと Hansch-Fujita 法を結びつけた人工知能システム KARMA (Kee Assisted Receptor Mapping Analysis)^{*}¹⁾がある。

* 1 : Klein, Teri E; Huang, Conrad; Ferrin, Thomas E; Langridge, Robert ; Hansch Corwin in “Artificial Intelligence Applications in Chemistry”; Hohne, B. Pierce, T., Ed.; ACS Symposium Series 306, American Chemical Society: Washington, D.C., 1986. “Computer-Assisted Drug Receptor Mapping Analysis”, pp. 147-158.

本システムは、過去に行われた Hansch-Fujita 解析の回帰式の情報から化合物 (リガンド) に関する様々な情報 (三次元的形状や置換基等に関する情報) をルールの形にし、レセプターマッピングを行う過程でこれらの情報を利用する。この KARMA で利用されているルールの例を以下に示す。

・ルール事例 1 : もしも分配パラメータの係数が約 1 の値を取るならば、リガンド化合物の置換基 X は殆ど脱溶媒和化 (Desolvation : 周辺に水分子が存在しない) している。

・ルール事例 2 : もしも化合物 Y の薬理活性が酵素 A において、関連酵素 B よりも小さいならば、置換基 X に関する立体障害の可能性が高い。

・ルール事例 3 : もしも疎水性パラメータが 0.5~1.0 であるならば、リガンドの置換基 X を取り巻く環境はくぼんだ形状をしている。

このようなルールを多数用いて、レセプターマッピングをする時の参考になっている。

また、ドラッグデザインを目指した人工知能システムとして藤田らは過去のドラッグデザイン過程でおこなわれた化合物の構造変換パターンをそのままルールとして用いる Emil システムを開発した^{*1)}。このシステムでは、入力された初期化合物が過去の構造変換パターン (ルール) に基づいて次々と構造変換さ

れて行く。

化合物構造式の変換パターンは事実上無限に存在する。過去に実行されて実績のある変換パターンを適用する事は、あたかもドラッグデザインに経験の深い研究者が化合物をみて自分のノウハウに従って構造式を変換し、より活性の高い化合物へと導くのと変わらない。Emil システムの詳細については第 11 章にて説明する。

* 1 : T.Fujita et al., In "QSAR and Drug Design' New Developments and Applications", Pharmaco Chemistry Library, Vol.23, T.Fujita., Ed., ERSEVIER, Amsterdam, pp 235-273, 1995.

2.1.1 三次元化合物データベースを用いたドラッグデザイン

①三次元化合物検索の基本

“似た化合物は似た活性を示す”という言葉を知覚しているでしょうか（第三章第五節⑤参照）。この基本原理は単純ではあるが多種多様の構造-活性相関手法やドラッグデザインの基本となっている。ここで説明する三次元化合物データベースを用いたドラッグデザインの基本もこの原理を利用している。

つまり、化合物の三次元的な形状や三次元的な特性（官能基同志の距離等）が似ているならば同等の薬理活性が期待出来るという定理である。コンピュータの技術が十分に発達しなかった頃は、この似たという尺度は二次元構造式を基本として議論されてきた。しかし、本来化合物は三次元であり、レセプターサイトとのドッキングも三次元で議論するのが本質である。コンピュータ自体と情報化学の進歩が三次元化合物を対象とした化合物検索を可能とした。

三次元化合物の類似判定作業は人間が最も不得意な作業の一つであり、コンピュータを用いて大量の化合物群を高速に検索する事が出来るならば、その果たす役割は大きい。なお、三次元類似化合物検索の基本技術は二次元化合物を対象とする類似検索と比較してかなり高度な技術が要求される。

②三次元化合物検索の必要性

薬理活性に関し、特定の官能基間の三次元的な距離が重要であるという情報は多くの薬理活性についてかなり明確となっている。これはレセプターサイトとのドッキング時に、リガンド化合物が特定の構造を取ることが必要なためである。このような官能基間の距離といった具体的な情報を用いて、その要求を満たす化合物群を三次元化合物データベースから取り出してくるのが三次元化合物検索によるドラッグデザインである^{*1, 2)}。

* 1 : Yvonne C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 1992, 35, 2145-2154.

* 2 : Stu Borman, "New 3-D Search and De Novo Design Techniques Aid Drug

Development”, C & EN, 1992, August 10, 18-26.

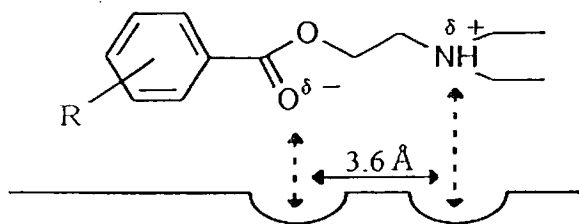


図 局所麻酔薬の活性発現のための三次元的位罫および電子的関係仮説
(山川浩司、他、“メヂシナルケミストリー”より抜粋)

例えば、図 に示されるように局所麻酔薬にはエステルケトンと四級アミン間の距離が 3.6 \AA 離れているという仮説^{*1)}が提唱されている。この仮説に従って三次元検索キーを作成することで要求仕様を満たす関連化合物群を取り出す事が可能となる。この三次元化合物検索で取り出された化合物群を研究者が総合的に評価し、新規リード候補化合物群の選択が完了する。

* 1 : J. Buchi, *Arzneim. Forsch.*, **10**, 6 (1960).

③三次元化合物検索の手続き

具体的な手続きは以下のようにして実行される。
・目的薬理活性を有する化合物を三次元化合物データベースより取り出す。
・取り出された化合物から代表化合物を選択し、この化合物を基本にその活性を発現するファーマコフォアを決定する。
・決定されたファーマコフォア同志の位置関係、ファーマコフォアの可動範囲の指定や、基準点の指定、類似度の設定等行い検索キーを作る。
・取り出された化合物のうち、最初に鋳型として用いた化合物と構造的に異なり、リード候補化合物となりうる化合物を決定する。以下、この順番に従って説明する。

・目的薬理活性を示す化合物を三次元化合物データベースから取り出す。この目的は、取り出された化合物を雛形として三次元検索キーを作成するためである。この時点での留意点は、薬理活性を特徴づける官能基群を網羅的に取り出せる化合物を選択する事である。また、目的が検索キーの構築に有るので、検索キーの構築に必要な条件を揃えていることも大事な基本要素となる。

・検索キーで利用するファーマコフォアやその他の検索に必要な小道具類を決定する。活性に本質的に必要なものを指定し、必要度の低いものは取り上げない方が化合物の自由度が増し、意外な化合物(真のリード候補化合物)を取り出す可能性も広がる。逆に、検索キーを細かく指定するほど意外な化合物の取り出しは困難になり、元の化合物に似た、変化のない化合物だけが選択される事になる。

・ファーマコフォア同志の位置関係、電子的な特性情報とその配向性、芳香

族環の中心位置や方向、置換基の可動範囲、その他の事項を許容範囲を含めて指定する。この許容範囲が一種の類似検索特徴抽出考えられる。許容範囲が大きい程変化の有る化合物が取り出され、小さい程変化のない化合物となる。

・作成された検索キーを用いて三次元検索を実施する。システムによってはここで類似検索オプションを持つこともある。検索により取り出された化合物は、最終的には研究者が構造式を確認することでリード候補化合物化合物否かを決定する。または、取り出された化合物群を見て新たな発想が研究者に湧けば目的は達成したことになる。

④三次元化合物検索キーの事例

三次元化合物検索で利用されるキーは二次元検索で利用されるものとはかなり異なっている。三次元検索で利用される検索条件は、原子や官能基間の距離関係、可動範囲の特定といった指定情報が中心となり、部分構造や官能基等の存在情報が中心の二次元検索とは基本が異なる。

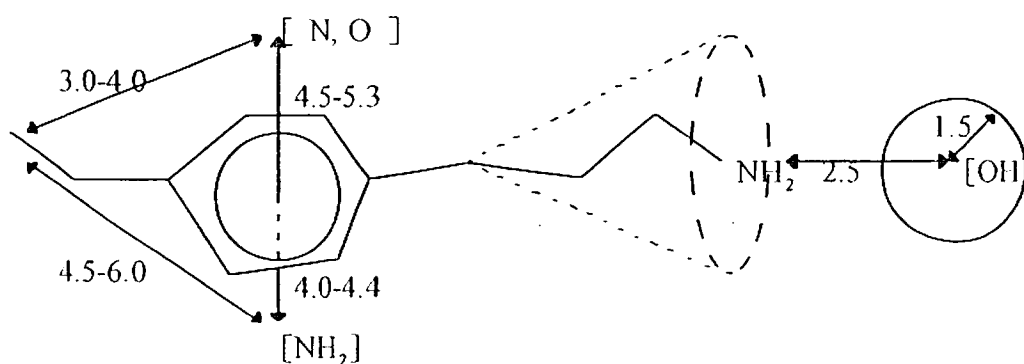


図 . 三次元検索キーの事例

三次元検索キーの簡単な例を図に示す。この検索キーではベンゼン環の中央を垂線が貫いている。この貫いている一端には窒素あるいは酸素原子が存在し、もう一端にはアミンが存在する。また、この垂線の両端からそれぞれ、3.0-4.0、4.5-6.0 オングストローム隔てた位置にベンゼン環の置換基から延びたエチレン基の末端メチルが存在する。さらに、ベンゼン環に結合したプロピルアミン基は円錐で囲まれた範囲内に有り、この末端アミンから 2.5 オングストローム離れた位置を中心とする半径 1.5 オングストロームの球の内部にヒドロキシ基が存在する。このような条件を満たした化合物群を三次元化合物データベースから取り出させるための検索キーである。

⑤三次元化合物検索事例

実際に三次元化合物検索を行った事例を紹介する。薬理活性は α -アミラーゼ阻害作用を有する新規薬物を三次元化合物検索にて取り出した例^{*1)}である。 α -アミラーゼ阻害作用を有するペプチドにおいて薬理活性を発現するため

には三つのアミノ酸の α および β 炭素間の距離と位置関係が重要であることが既に提唱^{*2)}されている。この事実を利用して三次元化合物検索を実施し、前記の三次元的要求事項を満足する新規化合物群を三次元化合物データベースから取り出してくる。

* 1 :

* 2 :

図 にデータベースから取り出されたトリペプチドの三次元構造を示す。この構造式において前記必要条件を満たす α 及び β 炭素の結合(図中の太線部分)を残して他の部分を消却する。残された三本の結合間の相互距離を幅を持たせて指定する。この三次元検索キーは図 の右に示されている。

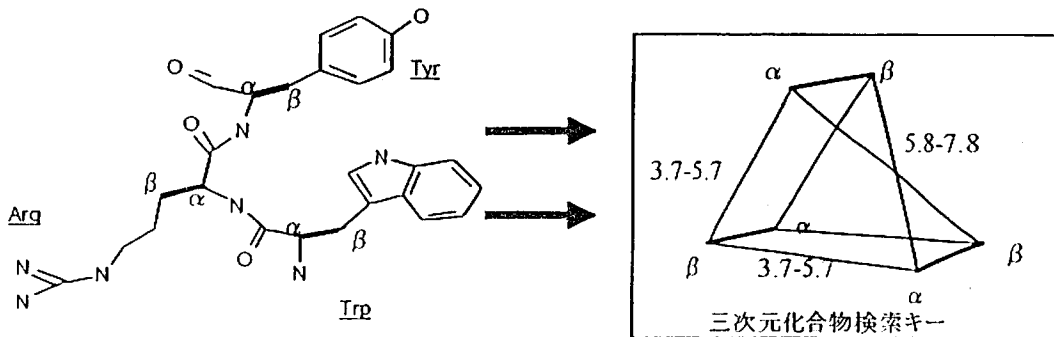


図 . α -アミラーゼ阻害作用ペプチドと、ペプチドから取り出された三次元検索キー

この三次元検索キーを用いて検索して取り出された化合物が図 に示され

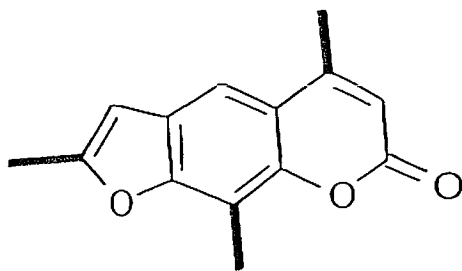


図 . 三次元化合物検索結果と検索キーを重ねた図

ている。図では化合物と検索キーが重ねて表示されている。図から基本骨格に直結する三個のメチル基が検索キーの結合に一致している事がわかる。このように、三次元化合物検索により、ペプチドの三炭素結合同士の位置関係を満たす、基本構造の全く異なる一般低分子化合物を簡単に取り出す事が可能である。

この例で示された検索キーの情報は現在の三次元検索システムの機能から見れば非常に単純な方に属する。しかし、検索結果を見れば分かるように、三次元的な位置関係を満たす化合物を多数の化合物群から取り出す事は人間の力では不可能な事である。この観点から言うならば三次元化合物検索は、目的によっては非常に強力なものとなる事を意味する。

しかし、この三次元化合物検索は薬理活性を示すための必要条件は備えてい

るが十分条件を備えているものではないことに留意する事が必要である。しかも、三次元検索で利用される三次元的な必要条件は多くの場合単なる仮説にすぎない事が多い。実際の薬理活性に結びつけるためには、薬理活性を発現する上での様々な条件（電子的、立体的等）を可能な限り取り込んだ形での検索を行うことが必要である。この点で三次元化合物検索にも限界が有ることを理解して利用する事が必要である。

2.2 解析を主体としたアプローチ概要

解析主体のアプローチとして現在までに多種多様なアプローチが試みられている。ここではこれらのアプローチのうちで、より総合的な構造-活性相関支援システムの基本技術として採用されている技術二件（化合物の三次元重ね合わせ、ファーマコフォア同志の三次元的な位置関係の利用）と、アプローチ的に面白いと感じた“共通部分構造取り出し”にスポットを当てる。

コンピュータを用いた構造-活性相関に対する解析的なアプローチの基本技術として化合物の三次元重ね合わせがある。コンピュータのグラフィック機能が幼稚だった頃はコンピュータの重ね合わせ機能だけで十分に独立した構造-活性相関手法として通用した。この重ね合わせ機能をより高度に利用した初期の例としてAAA(Active Analogue Approach)がある。現在では、構造-活性相関の三次元重ね合わせ機能は3-D QSAR、ドラグレセプター理論によるアプローチ、De Novo デザイン等の要素技術として組み込まれている。

また、昔から積極的に試みられてきたアプローチとしてはファーマコフォア（あるいは特定の官能基群）同志の三次元的な位置関係を基本としたアプローチが行われてきた。このアプローチの最終点は、前節で述べた三次元化合物検索という形で収束したといえる。本説では三次元化合物検索にいたる前の段階でのアプローチについて簡単にまとめる。

三番目に取り上げるアプローチとして共通部分構造探索をあげる。このアプローチは、同一薬理活性を有する一連の化合物群からその共通部分構造を取り出す事が出来れば、その共通部分構造は薬理活性発現に重要なキー構造であるという前提の元に、コンピュータにこの共通部分構造を取り出させる。これはイメージ的に分かり易いアプローチであるが、これをコンピュータ上で実行するときの困難性について言及する。

2.2.1 三次元重ね合わせによるアプローチ

このアプローチも三次元化合物検索と同様に、“似た化合物は似た活性を有する”という仮説を基本としている。

コンピュータ上の手続き的には、最初に重ね合わせする化合物上の互いに対応する二点（通常は原子を用いる）を最低三ペア指定する。次に指定された原

子ペア間の距離が最小となるように二点間の距離をパラメータとし、最小二乗法やシンプレックス法等の最適化手法を用いて三角面同志が最小の距離で重なり合うように調整する。このアルゴリズムは三次元的な重ね合わせ作業を、原子間の距離をパラメータとした三次元データの最適化問題にすり替えたものである。この場合、重ね合わせる化合物として変形を許さない剛体を用いている。現在ではコンピュータの進歩と重ね合わせ技術の進歩により、化合物の変形を許したフレキシブル重ね合わせが行われている。このアプローチとしても種々の手法が試みられているが、主として第 3 章ドラグレセプター理論によるアプローチの 3.2 で述べたフレキシブルドッキング技術と同じ基本技術が用いられることが多い。ドッキングの場合はレセプターサイトにおけるリガンド化合物の安定化エネルギーを評価関数として用いたが、この場合は重ね合わされた時の重なり部分の体積等を用いる。この他、モンテカルロシミュレーション等の手法を併用しつつ、化合物の構造式をリラックスしながら柔体同志の重ね合わせを行う。

・ AAA (Active Analogue Approach)

このような重ね合わせ化合物群の共通体積を評価するプロセスを利用して構造-活性相関を行うアプローチとして、1980 年代初頭に展開された AAA (Active Analogue Approach) ^{*1)} がある。このアプローチは活性化合物群と非活性化合物群とを独立して重ね合わせ、その重なり領域を取る。次に、活性および非活性グループとの重なり領域同志を重ね合わせて両方の差異を取る。この手続きにより、活性領域と非活性領域の重なり共通部分を除いた活性に重要な領域と、非活性に重要な領域とが明確に区別する事が可能となり、構造-活性相関に関する議論や、ドラグデザインが可能となる。

* 1 :

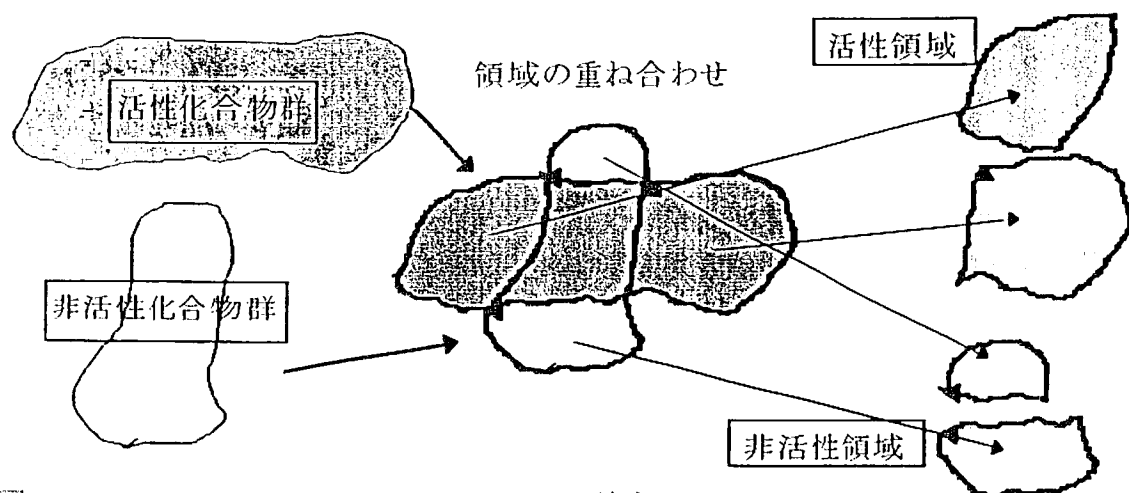


図 . AAA (Active Analogue Approach) の手続き

重ね合わせアプローチによる構造-活性相関はその性格上、同じ薬理活性で作用部位やメカニズムが同一であることが必要である。さらに、重ね合わせる化合物の三次元座標を厳密に計算する時には一般的に非経験的(*ab initio*)分子軌道法である Gaussian 等を利用する。しかし、Gaussian プログラムを用いた場合は最新のコンピュータを用いても小さな分子の計算が限界である(オプションにより異なる)。従って、化合物のサイズが比較的大きな一般低分子レベルになると半経験的分子軌道法である MOPAC 等の AM1 や PM3 を用いる事が多い。MOPAC にしても最安定構造を求めるには有る程度の時間がかかる。複数の化合物群をインタラクティブに重ね合わせする時や定性的に重ね合わせ状態を見る時には分子力学による計算で構造を決定する事も多いが、この場合でも最終的には分子軌道法を用いて構造最適化する事が学会や論文発表に必要となる。

2.2.2 化合物の二/三次元的類似性を基本としたアプローチ

化合物の類似という概念は構造-活性相関の基本であり、この概念に基づいて展開される構造-活性相関手法の数は多い。基本的には本章で述べた多くの解析手法(特にパターン認識や多変量解析を利用したアプローチ)は、物性を中心とした化合物構造式やその他の情報に関する多数のパラメータを用いた類似性の評価を基本としていると言える。ここでは類似性の基本を特に化合物の構造に限定してまとめる。

①官能基の存在情報の利用

類似の定義を化合物における官能基の存在情報に求めるならば、ある薬理活性系ではどのような官能基群が活性に重要な働きをするのかといった定性的な情報をつかむ事が可能である。このような目的で化合物データベース内の化合物群の官能基群を取り出してその統計を取るアプローチが取られている。このアプローチは化合物データベースが展開されてきた初期の段階においてしばしば利用された。

②化合物の三次元的な類似情報の利用

薬理活性の本質を化合物内の特定原子の三次元空間上の位置関係に集約することはドラグレセプター理論におけるレセプターサイトでのドッキングや受容体マップ等を考えれば当然の帰結である。このような考えに従った解析は比較的初期の段階から分子模型等を用いて展開されてきた。女性ホルモンのヘキサストロールとエストラジオールとの原子官能基距離を議論した事例^{*1)}。ベンゾチアゾシンの硫黄原子と窒素およびベンゼン炭素との距離データを基本とし、同様に比較分子上の距離データを取り出して、比較分子との誤差率を計算するアプローチ^{*2)}。この誤差率の小さい時、化合物は類似度が高い事に

なる。また、

具体的な原子や官能基ではなく、極性基、親油性フラグメント等のグループ間の距離関係を定義したアプローチ^{*3)}もある。

* 1 : C.O.Wilson, O.Gisvold and R.F.Doerge eds., "Textbook of Organic Medicinal and Pharmaceutical Chemistry", 7th ed., p746, J.B.Lippincott Co. (1977).

* 2 : 宮本秀一、吉本昌文、"新薬のリードジェネレーション (最新ドラッグデザイン)"、森口郁生、梅山秀明編、p.91、東京化学同人 (1987)

* 3 : Cosentino, U., Scolastico, C., Moro, G., Morosi, G., and Pitea, D., J. Mol. Struct.(THEOCHEM) 201, 199-212 (1989).

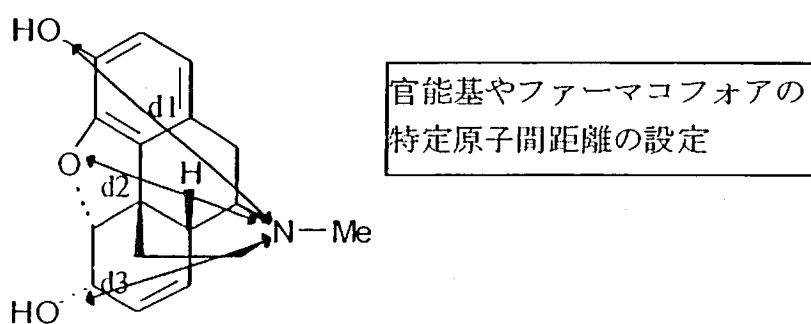


図 . 原子間距離指定によるモルヒネ類似化合物の特定

このような原子や官能基間の三次元的な特定による相関の研究は、現在は三次元化合物検索という形で実現している。先に述べたような三次元的な定義の検証も即座に結果が求められ、研究対象というよりもむしろ道具としての利用が進んでいる。

さらにこの類似性を単なる距離関係から化合物の全体的な形状情報まで拡張すると、化合物の直接的な重ね合わせの評価を行うことになる。例えば、重ね合わせた時の共通体積をモニターにして二分子間の類似性の評価を行う。前項の化合物重ね合わせ同様このアプローチだけで単独の解析手法となる事は殆ど無いが、レセプターマッピング、3-D QSAR 等の手法に組み入れられて活用されている。

2.2.3 共通部分構造の取り出しによるアプローチ

このアプローチは薬理活性が同一連の化合物群から、それらの化合物群に共通な要素（部分構造）を取り出してくるアプローチである。なお、この時の化合物は二次元構造式を基本としている。

このアプローチは、同一薬理活性を示す化合物群にはその薬理活性を示すための共通の部分構造が存在するはずという考えに基づいている。この薬理活性発現のための共通部分構造を一連の化合物群からコンピュータを用いて取り

出してくるのが本アプローチである。原理的には非常に明快で簡潔なアプローチであるが、実際にこれをコンピュータ上で完全に実現するためには多数の困難が付きまとう。以下に本アプローチが抱える幾つかの問題点をまとめる。

- ・共通部分構造のサイズの問題: 取り出された部分構造は誰もが認める有意義な部分構造であることが必要である。極端に言えば、原子や結合レベルに戻せば総ての原子は共通であるがこれで活性を議論することは出来ない。

- ・出現率: 出現率はある程度高い事が必要である。一般的に部分構造が複雑で大きな単位になるほど出現率は低下し、部分構造が小さくなるほど出現率は高くなる。真に活性に重要とされる範囲での部分構造選択が困難である。

- ・部分構造同志の相互関係に関する複合情報を取り出すことは極めて困難である。出来ない事はないが計算時間がかかる。

- ・原子や結合を一個一個トレースするコンピュータのアルゴリズムでは、複雑な構造を持つ化合物群の相互比較は極めて時間のかかる作業になる。人間が行う、化合物全体の構造を一瞬に判断するパターン認識による識別の方がコンピュータの逐次方式による判断より格段に優れている。

- ・の理由にもあるが、少ない数の化合物では人間の方が簡単に共通部分構造を取り出してしまうので、効果がアピールしにくい。

このように幾つかの重要な問題が完全に解決されないため、残念ながら現在ではあまり実行されていないのが現状である。このアプローチは基本的なスキルを必用とせず、活性に関する情報が構造情報という形で出力されるため、合成化学者には自分の仕事にフィードバックし易いという点で便利である。しかし、

- ・
- ・

にもあるように経験を積んだ合成化学者の洞察力に負けないシステムの構築は大変である。

2.3 その他のアプローチ

コンピュータを用いた構造-活性相関解析はいままで解説したアプローチの他にも多数のアプローチが試みられている。新しい基本技術が開発されればその技術を基本としてコンピュータの利用分野が開けてくる。コンピュータによる構造-活性相関はある意味では総合の技術であり、このような新しい技術がまっ先に適用されやすい分野の一つでもある。また、構造-活性相関を行う研究者は現状に満足することなく常により最新の、かつ完全な解析手法を求めてさまよい続けているという現実もある。これは構造-活性相関自体がまだまだ未熟であることの裏返しとも言えよう。

コンピュータがいろいろな目的に利用される道具であり、あらゆる分野で進歩する限りコンピュータを用いた構造-活性相関は増大し続ける。ここまで述べた以外に、現時点で行われている頻度の高い構造-活性相関手法を大まかにまとめると以下のようなになる。

①配座解析の技術を利用したアプローチ

分子力学（場合により分子動力学や分子軌道法との併用も行われる）により安定構造式を求めて、レセプターサイトとの結合情報を求めるアプローチである。ドラッグレセプター理論によるドッキング中心のアプローチと異なり、化合物の安定／活性配座解析との関係を中心として薬理活性との相関を追求するアプローチである。安定配座が決定されれば、その配座を取る化合物との構造的な類似性の評価により新たなリード候補化合物の開発が可能となる。薬理活性的に考えるならば活性配座解析の方が数段重要であるが、活性配座解析を確実に求めるアプローチは存在しない。

②分子動力学を基本としたシミュレーションによるアプローチ。

リガンド化合物のドッキングシミュレーションや水溶液中の安定および活性構造式等を分子動力学手法を用いて求める。現時点で分子動力学法は主として酵素自体の構造最適化や、レセプターサイトとリガンド分子の相互作用の探索に利用されている。全体が大きく移動する事が少なく、繰り返し単位の多い結晶等を扱う分野では数十ピコ単位のシミュレーションが行われている。しかし、柔体を扱い、かつ水素結合等の比較的弱い結合が複雑に関与し、多数の水分子を含み、構造が極めて複雑という悪条件を考えるならば、有る程度長い時間のシミュレーションを行わないと分子全体の動きのようなスケールの長い運動を捕らえることは難しく、実体を把握する事は困難かと思われる。現に Karplus^{*1)} は、生体高分子の動きには 10^{-15} ~ 10^{-3} 秒という広範囲の動きがあるとしている。

*1 : M.Karplus, J. A. McCammon, Ann. Rev. Biochem., 53, 263 (1983).

③最新のパターン認識法を適用したアプローチ

最近になり、ニューラルネットワークや遺伝的アルゴリズム等を用いて単発的に構造-活性相関解析を行うアプローチが方々で試みられている^{*1, 2)}。また、これらの技術は基本技術であるため他の構造-活性相関を行う過程で密かに利用される例が多い。例えば、遺伝的アルゴリズムは配座解析の最適化手法の一つとして利用されるし、レセプターサイトへのフィッティングでも利用されている（第 3 章 3.2.1 ③参照）。また、本著では述べていないがコンビナトリアルケミストリ/HTS でも化合物ライブラリ創出手続きの過程で利用されている^{*3, 4)}。さらに他の解析手法のアルゴリズム中に組み込まれたり、パターン認識で重要な特徴抽出手法の一つとして展開される事もある。

一方、ニューラルネットワークはその優れた分類能力から活性分類や予測等に利用される事例が増えている。遺伝的アルゴリズムやニューラルネットワークのいずれも従来手法による最適化や分類/予測に関して様々な改良点を持っており、この点で今後更に適用事例が増えるものと期待される。

* 1 : J.Devillers (Ed.), "Neural Networks in QSAR and Drug Design", Academic Press, London (1996).

* 2 : J.Devillers (Ed.), "Genetic Algorithms in Molecular Modeling", Academic Press, London (1996).

* 3 : Robert P. Sheridan and Simon K. Kearsley, "Using a Genetic Algorithm To Suggest Combinatorial Libraries", J. Chem. Inf. Comput. Sci., 35, 310-320 (1995).

* 4 : Lutz Weber, Sabine Wallbaum, Clemens Broger and Klaus Gubernator, "Optimization of the Biological Activity of Combinatorial Compound Libraries by a Genetic Algorithm", Angew. Chem. Int. Ed. Engl., 34, 2280-2282 (1995).

著者もニューラルネットワークによる構造-活性相関を試みており、良好な結果を得ている^{*1, 2)}。但し、ニューラルネットワークの適用に関してはチャンスコリレーションの入り込む可能性が高く、この点での留意が必要であるというのが著者の結論である。構造-活性相関分野では他の工学分野（一般的に数千~数万サンプルを扱う事が多く、チャンスコリレーションの入り込む余地は小さい）と異なり、極端に少ない数のサンプルを用いて解析する事が多いので特に注意深く適用すべきである。

遺伝的アルゴリズムは適用性が高く、解析目的によっては強力な武器になることは間違いない。但し、遺伝的アルゴリズムの運用（最適計算条件の探索）は大変な事が多く、この運用次第では期待した結果が得られない時もあるので注意が必要である^{*3)}。

* 1 : K. Yuta, "Structure-Activity Relationship Study of The Quinolone Antibacterials by The Neural Networks", CAMSE International Congress, Shibuya, Tokyo, Japan, Aug, 1990.

* 2 : 湯田 浩太郎、井関 京子、中山 唯夫、"ニューラルネットワークによる類似化合物の分類"、第12回情報化学討論会講演要旨集、P.164、大阪、1989年。

* 3 : 湯田 浩太郎、"拡張遺伝的アルゴリズムによるニューラルネットワーク（ローカルミニマからの脱出）"、第16回情報化学討論会講演要旨集、P.169、徳島、1993年。

3. まとめ

本章では前章までの一般的な構造-活性相関手法の分類に入らない、あるいは利用はされているが要素技術となり前面には出てこない手法について簡単にまとめた。これらのアプローチも実際に追求するとその背景にある技術の深

さは非常に深く、中途半端な解説では読者を混迷に陥れる可能性がある。興味ある方は分子力学／分子軌道法／分子動力学やパターン認識関連の専門書を一読されることを勧める。

繰り返すが、ここで取り上げたアプローチ以外にも多種多様の構造－活性相関手法が展開されている。今後もさらに新たなアプローチが展開される事は明白である。歴史的にコンピュータによる構造－活性相関においてはコンピュータ自体の進歩が新たな構造－活性相関手法を生み出すことが繰り返されている。この点で、コンピュータにおける最新技術をフォローする事は極めて重要となる。

コンピュータの進歩は日々前進し続け、その速度は一層高速になっている。現在はネットワーク関連の技術が急速に進歩している最中である。このネットワーク技術を用いて、全世界で同時に構造－活性相関をすることも実現に近づいている。また、コンピュータが最も苦手とする発想や連想を支援する技術も最近のコンピュータ技術を基本として更なる展開期を迎えようとしている。よりインテリジェントな機能を持つ新たな構造－活性相関手法が一日でも早く展開されることを望んでいる。