

1. 3D QSAR 概論

① 3D QSAR の歴史

定量的構造-活性相関といわれる新たな分野を開いたのは Hansch-Fujita 法であったのと同様に、3D QSAR も一つの解析手法が発表されたために確立された新分野である。この最初のアプローチが 1988 年 R.D.Cramer らにより発表された CoMFA (Comparative Molecular Field Analysis)法である。この CoMFA 法は、その後急速に全世界に普及していった。現時点では 3D QSAR という言葉はこの CoMFA 法とほぼ同意義で用いられている。

著者は 1992 年に 3D QSAR 手法として新たに導入された“領域効果”概念に基づく SARDE (Structure-Activity Relationships by Domain Effects)法を提唱した。同時に、この SARDE を実施するプログラムを開発し、実際のデータを用いて 3D QSAR 解析を行っている。この SARDE 法は定量的構造-活性相関分野では一般的な概念である“効果”の概念を見直すことで導かれた一連の手法の総称である。SARDE については本文中で改めて解説する。また 1995 年、中馬らは CoMFA 法の三次元格子の代わりに Voronoi 多面体を利用することで CoMFA 法と比較して大幅な次元減少を実現した。同時に相互作用エネルギーの簡易評価法を組み込み、PLS 法による次元減少と線形重回帰による解析を行う 3D QSAR を提唱している^{*1)}。

* 1 : 中馬 寛、唐沢 真美、熊沢 智、第 23 回構造活性相関シンポジウム講演要旨集、頁 266-269 (1995)

ここでは最初に CoMFA 法について述べる。続いて著者が開発した SARDE 法について、その基本となる“効果”概念の説明とその特徴と解析例について解説する。

② 3D QSAR の概要

3D QSAR はグラフィック中心のアプローチと、Hansch-Fujita 法に類似した線形および非線形重回帰法による解析を融合した最新の定量的構造-活性相関手法である。なお、この 3D QSAR の言葉に対応するものとして従来からの Hansch-Fujita 法は Traditional QSAR と呼ばれる。

従来から、ドラグレセプター理論を基本としたグラフィック中心のアプローチはフィッティングを前提とし、活性との直接的な繋がりが無いために定性的構造-活性相関に分類されてきた。概念やイメージ的に理解し易いという利点はあるが、強力な要因解析力がなく、定量的な評価や活性予測が出来ないといった限界があった。一方、Hansch-Fujita 法はこれらの主要要件を満たすが、イメージ的に取り付きにくく、パラメータを集めるのが大変であるといった問題点を抱えていた。このように構造-活性相関としては両局に位置する二つの代

表的な解析手法の長所のみを取り上げて、新たな定量的構造-活性相関手法として完成させたのが 3D QSAR である。

表 . Hansch-Fujita 法とドラグレセプター理論によるアプローチの特徴比較

解析の特徴	Hansch-Fujita 法	ドラグレセプター理論によるグラフィック的アプローチ
解析精度	定量的レベル	定性的レベル
定量的活性予測	可能	不可能
取り付き安さ	困難	簡単

③ 3D QSAR の特徴

前節で述べたように現時点での 3D QSAR はグラフィックを駆使したアプローチで三次元化合物群を解析対象とし、レセプターサイトとリガンド分子間に働く相互作用の評価を行う。ここまでの特徴はドラグレセプター理論によるアプローチとほとんど変わらないが、3D QSAR ではこれらのグラフィカルな要因に加えて線形/非線形重回帰法による要因解析、および定量的な活性予測が可能という Hansch-Fujita 法的な特徴をも合わせ持つ。

このように三次元化合物を解析対象とするので 3D、そして線形および非線形重回帰を用いて定量的な解析をすることから QSAR を、この二つを合わせて 3D QSAR という名称も納得できる。また、この組み合わせが 3D QSAR の特徴でもある。

2. CoMFA 法

CoMFA 法は 1989 年、Cramer により発表された^{*)}。Cramer は以前、主成分分析を利用して得られる BCD パラメータを用いた構造-活性相関を提唱している。CoMFA 法の中で用いられる PLS 法は流れ的にはこの主成分分析の変形であり、構造-活性相関分野で利用されてきた SIMCA 法を展開した Wold 親子により開発された手法である。CoMFA 法の特許は Cramer と Wold の連盟で登録されており、CoMFA の展開に Wold が果たした役割は大きいと考えられる。

2.1 CoMFA 法の基本

CoMFA 法は、第 1 章で説明したレセプターサイトに関する情報がない時に行なわれるレセプターマッピングと極めて類似した基本概念と手続きを持つ。しかし、レセプターマッピングの目的はあくまでもレセプターサイトデータのフィッティングであるのに対し、CoMFA 法の基本はレセプターサイトとそのレセプターサイトにフィッティングしたリガンド化合物間に働く非結合相互作用に基づく相関解析にある。従って CoMFA からは、レセプターサイトにお

けるリガンド化合物の三次元構造およびその周辺空間と薬理活性との相関情報を求めることが出来る。

CoMFA はレセプターサイトで最適フィッティングが行われたことを前提とし、その状態におけるリガンド化合物群とレセプターサイト間に働く非結合相互作用を薬理活性との相関関係の元に回帰式に反映させ、ドラグレセプター理論に基づくアプローチでは実行困難な回帰式による要因解析と定量的な活性予測を可能とする手法である。さらに、リガンド化合物周辺空間の要因解析はグラフィック表示の助けを借りながら研究者が視覚的に解析するもので、ドラグレセプター理論における視覚的要因解析と同じである。ただ、このグラフィックによる解析が行われる前に、電子的および立体的効果が解析に重要か否かの判定が回帰式を用いて行われている点が異なる。

2.2 CoMFA 法の手順

手続き的には①化合物の三次元座標創出、②三次元化合物群の重ね合わせ、③静電相互作用のための原子電荷の計算、④これら化合物群の周辺空間に三次元格子空間を設け、その格子の交差点上にプローブ(Probe)原子（一般的には炭素原子を用いる）を置く、⑤格子点上の各プローブ(Probe)原子とリガンド化合物上の各原子間に働くファンデルワールス (van der Waals) 反撥力を求めた値を各格子点単位で求め、立体相互作用評価用のフィールド (Steric Field)とする。同様に、③で求めておいた原子電荷と各プローブ原子 (+1に設定) との静電ポテンシャル値を求めて各格子点の値とし、静電相互作用評価用のフィールド (Electro Static Field)とする。続いて、⑥ P L S (Partial Least Squares) 法を用いて⑤で求められた各格子点の値を立体および電子の各フィールドごとに次元圧縮し、次元数 (数百～数千次元 : パラメータ) を数次元まで落とす。⑦この P L S より得られた最終パラメータ群を用いて線形/非線形重回帰式を求める。⑧要因解析は⑦の線形/非線形重回帰ニクラス分類より重要と判定されたフィールドについて三次元ディスプレイ上に等高線図 (Contour Map) を描き、このディスプレイを参考にしつつ議論する。ただし、この時点での解析は従来のドラグレセプター理論によるディスプレイ利用のアプローチと同じであり、内容的には活性に寄与する、寄与しないレベルの定性的解析となる。また、⑨活性未知の新規化合物がある場合は、⑦で求めた回帰式を用いて活性予測を行う。以上、CoMFA 法による一連の解析作業の流れ図を図 1 に示す。

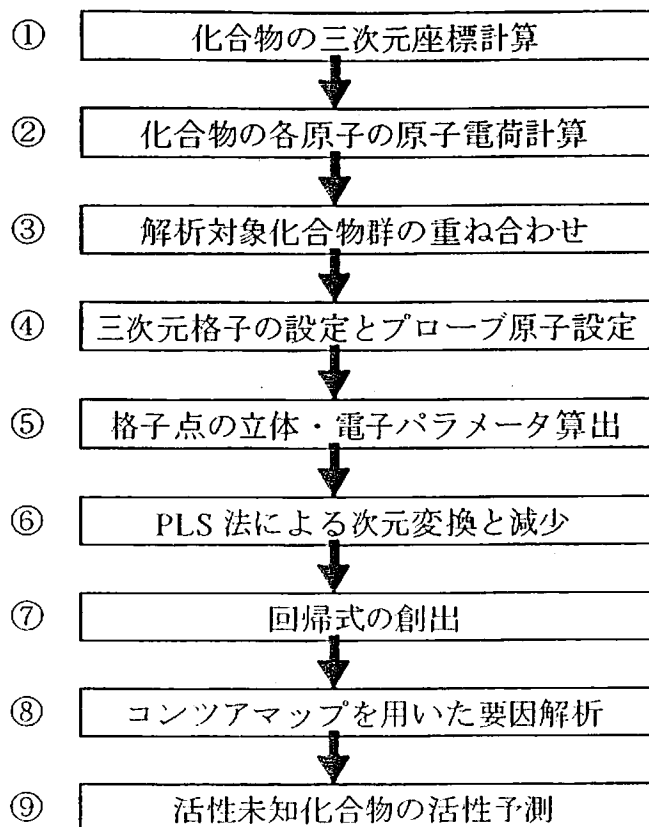


図 . CoMFA 法による解析の流れ図

この流れ図は 3D QSAR の典型的な作業形態を示している。CoMFA が発表された後に開発された 3D QSAR 手法の殆どは、多少の違いはあってもこの流れ図と大きな差異は無い。

また、作業の形態からいうならば①の三次元座標の計算から、⑤の格子点パラメータの算出までは若干の差異はあるが、ドラグレセプター理論によるアプローチの一つであるレセプターマッピングを行う過程で行われる手続きと変わらない。レセプターマッピングでは、この格子点データはフィッティングを行う為のモニターとして用いるが、CoMFA は、

この格子点データに次元圧縮をかけて次元を縮小することで統計的な処理を可能としている。この点が両アプローチの根本的な違いである。

2.3 CoMFA パラメータ

①概要

統計的な処理を解析手順に含む限りパラメータの存在は必須である。また、このパラメータの種類や起源などの差異により解析手法としての特徴が出て、得られる情報の内容も変化する。

CoMFA 法は、レセプターサイトとリガンド化合物との非結合相互作用の評価に基づく構造-活性相関手法である。従って、CoMFA 法で利用されるパラメータはこの非結合相互作用情報を取り出すものでなければならない。この観点から CoMFA では化合物の周辺空間上に“場 (Field)”を設け、この場に関する情報の数値データ化を実現させている。この手続きで創出されるパラメータを CoMFA パラメータと呼ぶ。

CoMFA パラメータはレセプターサイトとリガンド化合物上の各原子間に

おける非結合相互作用を、電子的およびファンデルワールス反撥力の二要素で評価するものである。これらの二要素に加え、最近では疎水相互作用のパラメータである LogP 値を三次元化合物の周辺空間上に分散させた疎水性ポテンシャルが実用化されている。この疎水性ポテンシャルの創出には“HINT”プログラム^{*1)}が良く利用される。

* 1) Kellogg, G.E., Semus, S.F. and Abraham, D.J., J.Comput. -Aided Mol. Design, 5, 545- 552 (1991).

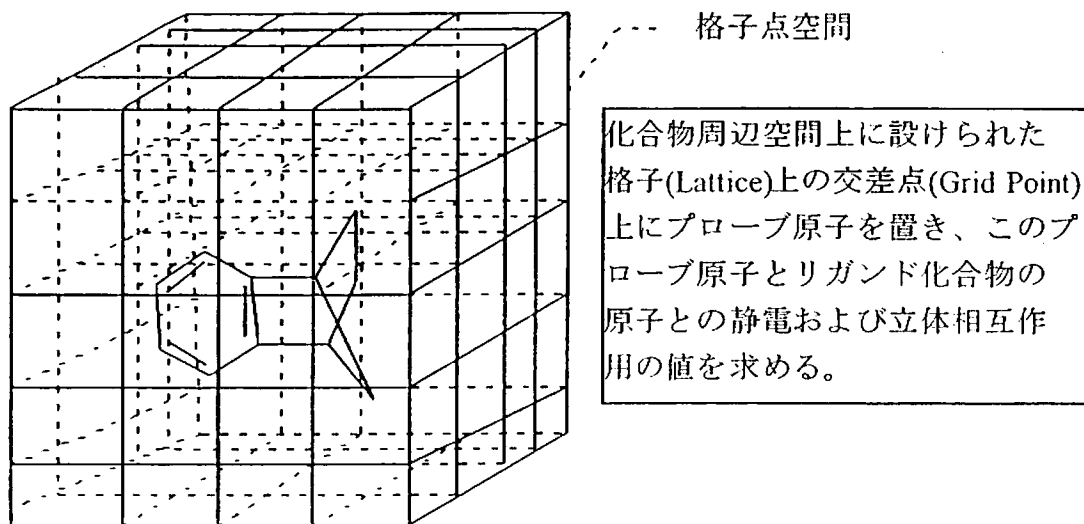


図 . 三次元化合物周辺空間の三次元格子

② CoMFA パラメータの具体的な創出手続き

化合物周辺空間に設定された三次元格子の各格子点上にプローブ原子（通常は炭素）を置き、このプローブ原子とリガンド化合物間に働く静電相互作用と、ファンデルワールス相互作用をその格子点上に置ける二つのパラメータとする。

これを総ての格子点について計算を行う。従って、三次元格子が X, Y, Z の軸それぞれが 10 点ずつで構成されていれば $10 \times 10 \times 10 = 1000$ の格子点について静電および立体のパラメータが求められ、総数で 2000 次元のデータとなる。

この 2000 次元のデータをすべての解析対象化合物について求めて、CoMFA パラメータの完成となる。

プローブ原子の電荷を +1 とすると、静電相互作用を求める式は以下のようになる。

$$E_{\text{electro}} = F \sum_{i=1}^N \frac{Q_i}{R_i}$$

ここで、 Q_i はリガンド化合物の i 番目の電荷、 R_i はプローブ原子と i 番目の原子との距離、 F はエネルギー単位に関する定数で、332.2 の値を持つ。

また、ファンデルワールス反撥エネルギーは Lennard-Jones タイプの計算式を用いて求められる。

$$E_{steric} = \sum_{i=1}^N K_i \left[\left(\frac{V_i + 1.7}{R_i} \right)^{12} - \left(\frac{V_i + 1.7}{R_i} \right)^6 \right]$$

ここで、 K_i は i 番目の原子のファンデルワールス定数で、 V_i はファンデルワールス半径、 R_i はプローブ原子と i 番目の原子との距離である。

2.4 PLS法の概論と実際の解析手順

CoMFA アプローチでは PLS 法による次元圧縮が必須の手続きとなる。これは、統計手法に起因する問題を回避する一つの手段として必要となるのである。既に Hansch-Fujita 法の章にて説明したが、線形/非線形重回帰を実行する時にその解析の信頼性を保つ条件として、解析に用いるサンプル数を重回帰式で用いるパラメータ数で割った値は六以上あることが必要という制限があることは述べた。従って、CoMFA が数千~数万の格子点データを解析に用いるならばサンプル数もその六倍以上必要である。

構造-活性相関でサンプル数を多数揃える事は不可能であり、事実上統計的手法を解析に取り入れる事は不可能となる。従って、CoMFA アプローチではパターン数そのものを何らかの手法を用いて減少させる必要がある。この次元減少手法としては一般的に主成分分析があるが、この手法はパラメータ間の独立性を前提としたアプローチのため、CoMFA パラメータのような相関性の高いパラメータに直接利用することは出来ない。この、CoMFA パラメータの特徴に耐える解析手法として PLS 法が利用された。

① PLS (Partial Least Squares) 法^{*1)} 概論

PLS 法はスウェーデンの Wold 親子により開発されたプロジェクション (モデリングとも呼ばれる) のためのプログラムである。PLS 法は他の一般的な多変量解析と異なり、互いに相関性の高い多次元データの解析に適した解析手法である。ケモメトリクス分野では主としてスペクトル解析等に利用されてきたが、構造-活性相関分野での利用は CoMFA が最初である。本手法に関する解説はケモメトリクス^{*2, *3)} 関連の著書に詳しいのでそちらを参照されたい。本著では PLS 法を実際に使う時に必要となる事項を主眼にまとめる。

*1) S. Wold, H. Martens, H. Wold, The Multivariate Calibration Problem in Chemistry Solved by the PLS Method, in "Matrix Pencils (Lecture Notes in Mathematics, 973)", B. Kagstrom, A. Ruhe (Eds.), Springer-Verlag, Berlin, p.286

(1983).

* 2 : 宮下芳勝、佐々木慎一、 “ケモメトリックス (化学パターン認識と多変量解析)”、コンピュータ・ケミストリー シリーズ 3、共立出版株式会社、東京、(1995) .

* 3 : 相島

②クロスバリデーション (Cross Validation) による最適成分数の決定

PLS法の結果を線形重回帰で利用するためにはPLSから取り出す成分数を決定しなければならない。成分数決定の指標は、取り出した成分を用いた回帰式の評価により決定される。

回帰式の評価はその予測能力を指標としておこなわれる。この予測能力はパターン認識法の章で解説したリーブNアウト (Leave N Out)法を利用して得られた予測値と実測値との差である予測誤差 (PRESS : Prediction Residual Sum of Squares)を指標として測定される。この予測誤差は以下の式から求められる。式中Nは予測対象となるサンプル数であり、 Y_{obs} は実測値を、 Y_{pred} は予測値を示す。

$$\text{予測誤差 (PRESS)} = \sum_{i=1}^N (Y_{obs} - Y_{pred})^2$$

また、この PRESS の代わりに以下の式で求められる予測的相関係数の二乗 (Cross Validated Squared Correlation Coefficient r^2) を用いることが多い。最近ではこの Cross Validated r^2 を単に q^2 と表すことが多い。

$$q^2 \text{ (Cross Validated } r^2) = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - Y_{mean})^2}$$

Y_{mean} は実測値の平均である。式より、 q^2 の値が1の時完璧な予測が実現したことになり、0以下の値は実測値の平均を予測とした場合よりも結果が悪くないことを意味する。

PRESS および q^2 ともに総和は予測データセットについて取られる。この予測データセットの取り方はサンプル数が大きい時は独立して設定されるが、サンプル数が小さい時は先にも述べたようにリーブNアウト (Leave N Out)法等の手法により実施される。一般的にNは1の値を取ることが多い。

2.5 CoMFA 法実施上での問題点と限界

CoMFA アプローチはグラフィカルな手続きに Hansch-Fujita 法並の定量性を与えたアプローチとして、その果たす役割は大きい。しかし、CoMFA アプロ

一ち独特の特徴に起因した留意点も幾つか存在する。これらの特徴を認識した上での利用が必要である。

①化合物の三次元重ね合わせ

大きな問題点としては他の手法と比べ、解析結果が化合物の重ね合わせの違いにより大きく異なる場合があることである。これは重ね合わせ状態が異なると、たとえ同じ化合物を用いても三次元格子点の値が全体にわたって大きく変化するためである。これは CoMFA 法が化合物の周辺空間を細かく分断し、その各々についてパラメータを求めている限り回避出来ない問題点である。より大きな構造単位でパラメータを生成することが無い限り、化合物の形や位置の変化に左右されないパラメータの生成は不可能である。

②最終回帰式中の CoMFA パラメータの扱い

最大の限界は、最終回帰式に残った CoMFA パラメータから得られる情報量が少ないことである。つまり、CoMFA 法により形式上 Hansch-Fujita 法と同様な回帰式を得るが、CoMFA パラメータの回帰係数の強弱情報に基づく定量的評価は困難で構造-活性相関的には意味がない。単に、フィールド単位（電子、立体、疎水）の情報が使えるか、使えないかという定性的レベルの評価に止まる。

数学的に PLS で得られた各成分を元のパラメータ単位に分割することは出来るが、数千次元にも及ぶ線形重回帰型の式を得ても意味は無い。この限界は解析過程で PLS 法による次元減少を行う限り避けられない。

③パラメータの種類

もう一つの大きな限界は、CoMFA 法の基本に起因する。つまり、フィールドを前提として解析する限り、原理上使えるパラメータの種類が強く制限されることである。現在利用されているパラメータは電子的、立体的フィールドを基本とし、これらに加えて疎水性フィールドが利用され始めたところである。

2.6 CoMFA 法適用、および発表時に備えるべきデータ

①適用時

②発表時

CoMFA アプローチを適用するとき必要な情報は、重回帰式、サンプル数 (n)、標準偏差 (s)、相関係数 (r)、クロスバリデーション時の標準偏差 (s_{press}) と相関係数 (q^2)、および予測誤差 (PRESS)、成分数、個々のフィールドやパラメータの相対的な寄与率といった情報を明確にすることが必要である。以下に CoMFA による解析結果を発表する時の回帰式の典型的な書き方を示す。

Y (薬理活性) = 定数項 + 係数 パラメータ + [CoMFA フィールド名]

n = 、 s = 、 r = 、 S_{press} = 、 q^2 = 、 press = 、
number of component (成分数) = 、
contribution (寄与率) = パラメータ名 ; 値 : パラメータ名 ; 値 : . . .

2.7 CoMFA 法適用事例

3. SARDE 法による 3D QSAR

ここでは 3D QSAR の新たなアプローチとして、著者が開発した SARDE 法について述べる。

SARDE 法は構造-活性相関分野で基本となる“効果”に対する概念の見直しにより導かれたアプローチである。この手法は CoMFA 法が持つ限界と問題点の解決を意識しつつ設計されている。つまり、CoMFA 法では PLS 法を用いるために統計的な理由からフィールド効果の要因解析を定量的レベルで出来ないという欠点がある。一方、SARDE 法は手続きの工夫により定量的な要因解析が可能である。さらに、原理上多種多様のパラメータを用いることが可能で、解析の幅を広げることが可能である。

3.1 新たな構造-活性相関手法に関する基本概念の導入

①部分構造単位で議論する構造-活性相関事例

化合物の部分構造単位で活性との相関を議論する試みはかなり昔から当然のこととして行われてきた。例えばドラグレセプター理論で利用される受容体モデルは活性との議論を行うポイントは化合物全体でもなく、また置換基レベルの小さな単位でもない中間的な“部分領域レベル”となっている。また、古くから提唱されている仮説の多くは部分構造単位で議論されることが多い。例えば、芳香族多環縮合化合物の発癌性に関する電子密度情報の検討から得られた K 領域 (主発ガン要因)、L 領域 (副発ガン要因) とベイ領域の仮説は化合物の部分構造を基本とした議論に基づいている。

すなわち、レセプターサイトにおいてリガンド化合物はその構造全体で反応するのではなく、特定の領域 (部分構造、あるいは薬理作用団) を中心としてレセプターサイトと反応する。この事実は、薬理活性は特定の領域 (部分構造領域) が活性に寄与する割合の総和により説明可能であることを暗示している。

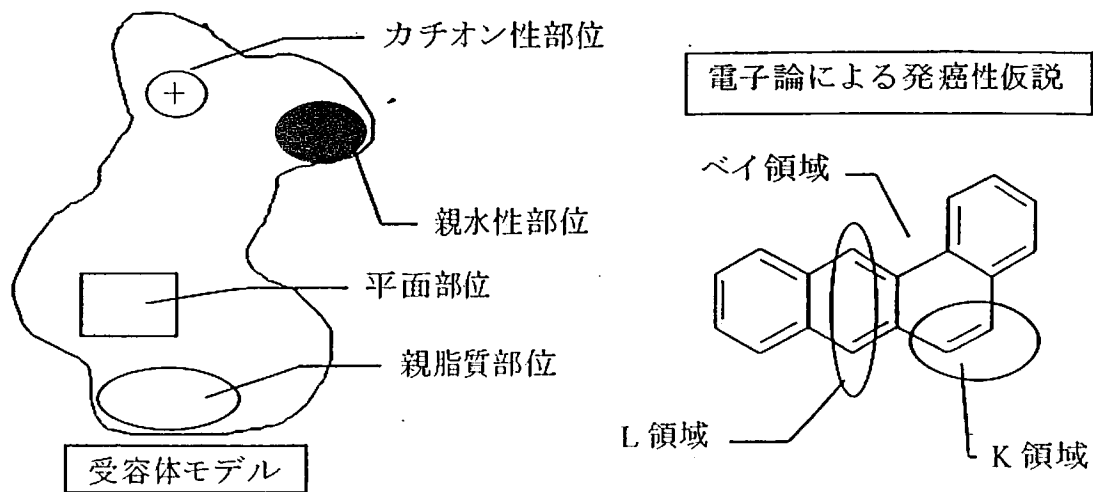


図 1 受容体モデルと電子論による発ガン性仮説に見る、領域単位で議論する構造-活性相関事例

つまり、この部分構造領域情報を用いて薬理活性との関係を導くならば以下の式で示される。

$$Y (\text{薬理活性}) = \sum_i a_i (\text{部分構造領域} \supseteq \text{薬理作用団}) + \text{Const.}$$

*部分構造領域は薬理作用団 (Pharmacophore) も含む、より大きな構造単位を意味する。

*領域は空間的な領域と、その領域の持つ特性との二要因から構成される。

ここで、Yは薬理活性を、 a_i は i 番目の領域が薬理活性 Y に寄与する活性の強さを表す。

② 新たな構造-活性相関手法の開発目標

このように、昔から薬理活性を論じる時には化合物の部分的な範囲 (領域) との関係で論じられてきた。これに対し現在実用化されている定量的構造-活性相関手法は、部分構造よりもさらに局所的な官能基レベルでの議論手法と、全構造レベルで議論する手法しか提供していない。この意味で、化合物の部分構造単位で薬理活性との相関を議論する手法を提供することは、構造-活性相関の観点上非常に望ましいことである。このような観点から著者は、以下の三点に留意して新たな解析手法の開発を試みた。

- 部分構造や領域単位で構造-活性相関を議論できる。
- 解析の基本型は、現在多くの研究者に利用されている 3D QSAR とする。
- 3D QSAR の持つ主要な限界をもクリアする。

以上の三項目を満たす手法として SARDE 法を開発した。以下に SARDE 法に

ついて、解析事例も交えて解説する。

3.2 定量的構造-活性相関の“効果”概念と、新たな効果概念の導入による SARDE-S および SARDE-F 法の開発

定量的構造-活性相関分野では“効果”という概念が存在する。例えば、Hansch-Fujita 法は“置換基効果”による定量的構造-活性相関と称される。つまり Hansch-Fujita 法においては、その活性の効果は基本骨格が固定された時の置換基の変化と結びつけて議論される。また、CoMFA 法は“場 (Field) 効果”による定量的構造-活性相関と呼ばれる。これは CoMFA 法が化合物周辺に設けられた“場” (例えば電子的、およびファンデルワールス反撥力) の変化と活性との相関を基本とするアプローチであるためである。

このように、著者が“領域効果”概念に基づく SARDE 法を提唱する以前は、Hansch-Fujita 法と CoMFA 法に関する二種類の効果のみが認識されていた。この二種類の効果を深く検討するならば、内容的に二種類の分類基準が存在することがわかる。一つは効果を及ぼす場所に関するもので、化合物構造式本体と化合物の周辺空間 (場 : Field) の二つが存在している。残る一つは、効果を発現する起因となる化合物の範囲 (領域) に関する基準である。この、最も小さな単位は置換基であり、続いて部分構造、そして全化合物構造の三段階にわけられる。この効果に関する二つの基準を用いてマトリクスを作成し、このマトリクスに現在実用化されている定量的構造-活性相関手法を入れたのが表である。

表中、×印の部分は物理的に不可能、あるいは無意味と考えられる。この中で、部分領域の列で化合物構造の行の部分構造に該当する SARDE-S、および場の行の部分空間に該当する SARDE-F の部分には既存の定量的構造-活性相関手法は存在しない。従って、この空白部分に該当する手続きを確立すれば新たな定量的構造-活性相関手法となる。なお、SARDE-S は Structure-Activity Relationships by Structure の略で、SARDE-F は Structure-Activity Relationships by Field の略である。

表 . 範囲および部位による“効果”の分類と定量的構造-活性相関手法

場所	範囲	置換基領域	部分領域	全構造領域
化合物構造		置換基	部分構造	全構造
		Hansch-Fujita 法	SARDE-S	Hansch-Fujita 法 パターン認識法
場 (Field)		×	部分空間	全空間
		×	SARDE-F	CoMFA 法

化合物を構造上のサイズの観点で考えた場合、置換基のような小さな単位と化合物の全体との間に部分構造という中間的な単位がある。しかし、当時の時点においてはこの部分構造を基本とした効果に基づく構造-活性相関手法は提唱されていなかった。そこで、この部分構造単位で活性との相関を議論出来る手続きを提供するならば、Hansch-Fujita 法、CoMFA 法に続く第三番目の定量的構造-活性相関手法となり得ることがわかる。この三番目の手法として提唱したのが SARDE 法である。

3.3 効果概念からみた定量的構造-活性相関手法の差異

効果という概念から Hansch-Fujita 法、CoMFA 法、SARDE 法特徴の関係を図的に表示したものが図 である。

Hansch-Fujita 法は化合物の全体構造で決まる種々の物理化学的パラメータによる議論と、基本構造式上に位置決めされた置換基の変化で議論するアプローチである。また、CoMFA 法はレセプターサイトでのリガンド化合物との非結合相互作用を化合物の周辺空間に設定された種々のフィールド情報を用いて解析する。手続き的には PLS を用いてフィールド単位に解析するので、CoMFA による解析は化合物全体を対象としたものとなる。もちろん、CoMFA では部分空間単位での議論を行うが、これはグラフィック表示を研究者が視覚的に解析する、従来手法の定性的な解析にしかすぎない。

SARDE-S 法は三次元化合物を任意の平面で分割し、要因解析はこの分割された三次元部分構造単位でおこなう。また、SARDE-F は化合物の周辺空間を部分空間単位に分割し、その単位で要因解析を行う手段を提供する。SARDE-S、SARDE-F とともに、部分領域 (構造、空間) 単位で活性特徴の相関を定量的に議論することが可能である。

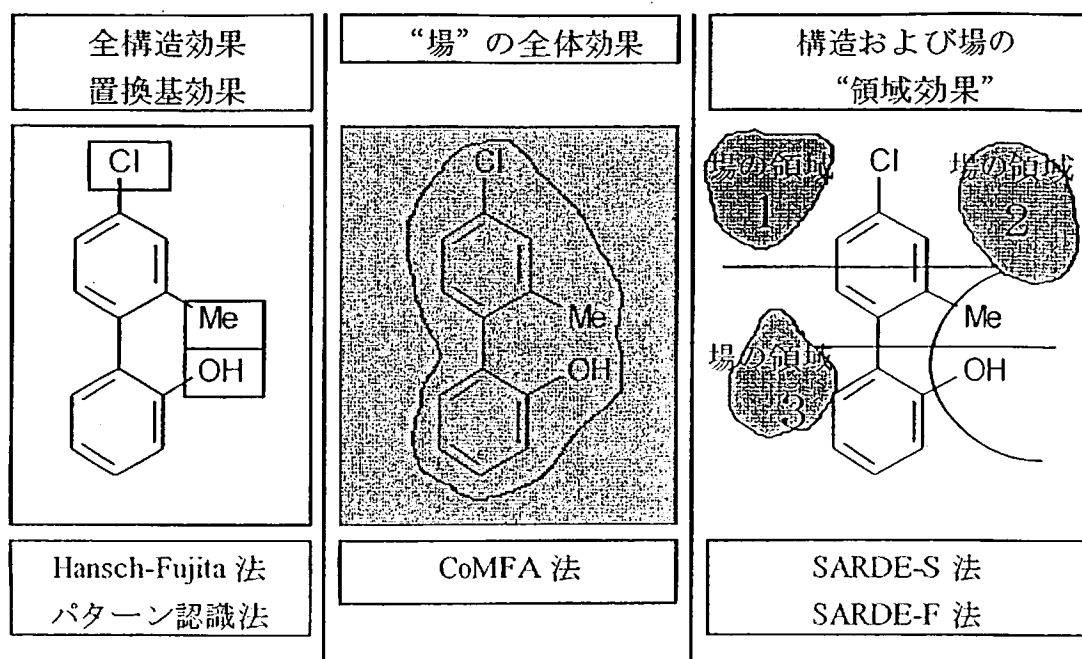


図 . 種々の“効果”と定量的構造-活性相関手法との関係

3.4 SARDE 法の具体的手続き

定量的構造-活性相関を行う時に利用される回帰式の形式を以下に示す。

Hansch-Fujita 法 $Y = D_1(\text{パラメータ } 1) + \dots + D_i(\text{パラメータ } i) + \text{Const.}$

CoMFA 法 $Y = [\text{場の効果 } 1] + [\text{場の効果 } 2] + \text{Const.}$

SARDE-S 法 $Y = D_1(\text{構造領域 } 1) + \dots + D_i(\text{構造領域 } i) + \text{Const.}$

SARDE-F 法 $Y = D_1(\text{空間領域 } 1) + \dots + D_i(\text{空間領域 } i) + \text{Const.}$

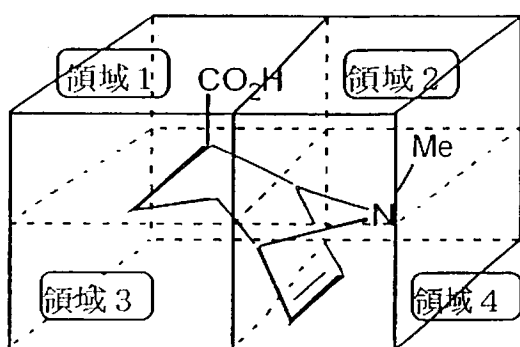
構造-活性相関の立場から考えるならば同じ回帰式にしても要因解析を精度高く（定量的レベル）行えるものであることが望ましい。構造-活性相関の観点で Hansch-Fujita 法と CoMFA 法で用いられる回帰式を比較した場合、Hansch-Fujita 法と同じ形式の回帰式であることが望ましい。この差異は、CoMFA 法が PLS による次元圧縮を行っていることに起因する。従って、あらたに導入される解析手法はこの作業のないものであることが必要である。SARDE 法で用いるパラメータは部分構造/空間単位で発生されるため、種類は多いが、独立性が高いために PLS による次元圧縮は必要ない。通常の特徴抽出手法の適用により回帰式を求めることが出来る。従って、Hansch-Fujita 法と同様に回帰係数の大小情報を用いることが可能であり、きめ細かな議論ができる。

SARDE-S は化合物の部分構造領域単位での解析を、また SARDE-F は化合物周辺空間の部分空間単位での定量的な議論を可能とし、より直感的で人間の感覚に近い形での要因解析が可能である。

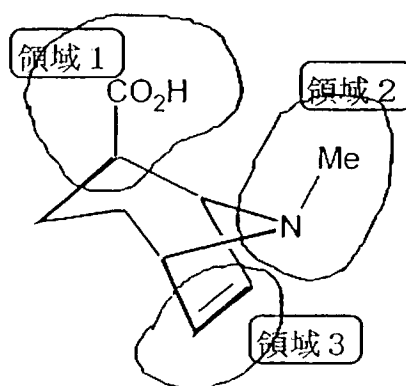
3.5 領域パラメータ

SARDE 法で利用されるパラメータを“領域 (Domain) パラメータ”と呼ぶ。このパラメータは解析手法に従って二種類に分類される。一つは、SARDE-S で利用されるもので、化合物の部分構造単位で発生される“部分構造領域パラメータ”である。また、SARDE-F で利用されるパラメータは化合物の周辺空間に設定されたフィールドを分割して発生される“部分空間領域パラメータ”である。ここでは SARDE-S で利用される“部分構造領域パラメータ”を例に取り説明する。

①領域パラメータの創出と種類



平面による部分構造分割



自由曲面による部分構造分割

図 ． 三次元化合物の部分構造領域への分割法

・化合物の分割

部分構造領域パラメータは三次元空間に設定された個々の空間内に占める化合物の部分構造ごとに算出される。従って、部分構造領域パラメータの創出における最初の操作は化合物群を三次元空間上で分割する事から始まる。

重ね合わせが済んだ化合物群を部分構造へと分割する作業の例を図 に示す。実際の化合物分割は平面による分割や、自由曲面による分割等が行われる。現在実用化されている SARDE システム上で実行可能な分割様式は、平面による分割だけである。この分割操作はグラフィック上で化合物を動かしながら研究者が切断して行く。

・分割された部分構造のパラメータ化

化合物が部分構造単位に分割されたならば、この部分構造単位にパラメータを発生させることが必要である。この部分構造単位での数値データへの変間は、パターン認識法で利用されていたパラメータ創出アルゴリズムをそのまま適用出来る。従って、部分構造領域パラメータの特性はパターン認識法で利用さ

れるパラメータと変わりはない。但し、分子軌道法から得られるパラメータ等は、一旦化合物全体について計算した後に各部分構造単位でパラメータ化することが必要である。

パターン認識法で用いられたパラメータとの本質的な違いは、パターン認識法でのパラメータは化合物自体と一対一で対応しているが、部分構造領域パラメータは部分構造と一対一で対応し、化合物自体とは一対複数で対応することである。

領域 1 トポロジカル パラメータ	CO_2H $\begin{array}{c} \\ \text{C} \end{array}$	領域 2 立体パラメータ	$\begin{array}{c} \text{Me} \\ / \\ \text{N} \\ \backslash \\ \text{C} \quad \text{C} \end{array}$
領域 3 電子的 パラメータ	$\begin{array}{c} \text{C} \\ / \\ \text{C} \text{---} \text{C} \end{array}$	領域 4 分子屈折率	$\text{C} = \text{C}$

図 . 領域 1 ~ 4 の部分構造と、発生されるパラメータ例

②領域パラメータの特徴

・領域パラメータの頑健性

SARDE 法で生成されるパラメータは本質的に頑健性を持っている。これは、領域パラメータが部分構造/空間単位で数値データ化されるためである。つまり、部分構造自体の変化には敏感に反応するが、部分構造の座標変化は極端な変化でない限りパラメータに影響を及ぼさない。これに対し、空間を細かく分割する CoMFA 法は化合物の些細な座標変化によりパラメータ値が大きく変化する。当然、解析結果も座標変化の影響を受ける。

この特性は SARDE 法が三次元化合物の重ね合わせによる初期座標の変化に影響を受けない手法であることを意味する。すでに 1.4.5 の①で述べたように CoMFA 法は化合物の重ね合わせに結果が強く影響されるアプローチであり、ほとんどの 3D QSAR アプローチはこの問題を有している。

・領域パラメータの種類

領域パラメータは化合物の構造に関するものと、フィールドに関するものがある。これらのパラメータのうち化合物構造式に関するパラメータは多種多様なものを利用することが可能である。パラメータの創出に利用される、構造式の数値データへの変換アルゴリズムはパターン認識法で利用されるものを一部修正する事で適用可能である。従って、部分構造領域パラメータを用いることで化合物の部分構造と薬理活性との関係に関する情報を多様な形で取り出

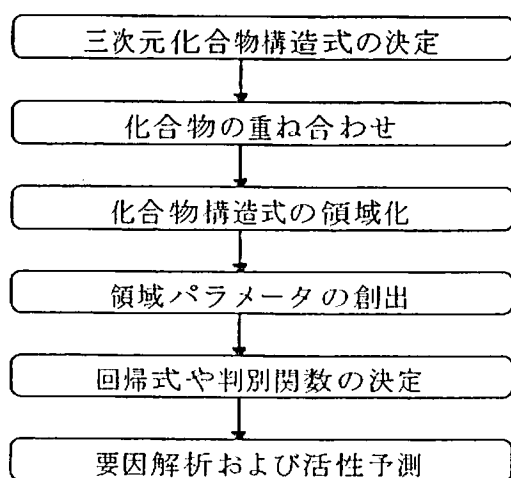
すことが出来る。

一方、部分空間領域パラメータはその部分構造を取り出すフィールドの種類に制限を受ける。現時点で実用化されているフィールド三種類程度であり、部分空間領域パラメータも三種類となる。

3.6 SARDE-S法の適用および事例

① SARDE法による解析概要

SARDE法による解析の流れを図に示す。



解析の流れは最初の三次元構造式の算出とその重ね合わせまでは一般の3D QSAR手法と全く同じである。これ以降がSARDE独自の操作に入る。先に重ね合わされた化合物群を三次元的な領域(ブロック)単位に分割する。この分制作業はグラフィックディスプレイと対話しながら行う。続いて、分割された個々のブロック単位に種々のアルゴリズムによりパラメータ化する。続いて、種々の特徴抽出手法によりパラメータを選択し、最終回帰式を求める。この回帰式を用いて要因解析および新規化合物の活性予測を行う。

図 SARDE法による解析流れ図

② SARDEによるCa²⁺アンタゴニストの定量的構造-活性相関解析

実験の目的: ジヒドロピリジン誘導体のCa²⁺アンタゴニストとしての作用について、その部分構造領域単位に活性との相関を論じる。

・薬理活性及び化合物構造式

SARDE法による解析事例を以下に述べる。解析ターゲット化合物はCa²⁺アンタゴニストの4-Phenyl-1,4-dihydropyridine誘導体を用いた*¹⁾。

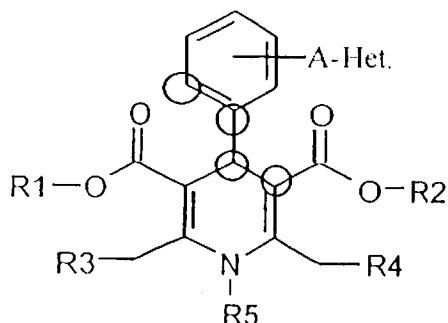


図 基本構造式

解析に用いた基本構造式を図に示す。化合物数は15化合物と少なく、かつ構造的な変化部位数が多い化合物群を意識的に用いた。これは、従来の定量的構造-活性相関手法ではこのように構造変化部位が多く、かつ化合物数が少ないデータセットを対象として解析することは極めて困難であるからである。

* 1 : Paolo Cozzi and et al., J. Med. Chem., 36, 2964-2972 (1993).

表 . Ca²⁺アンタゴニズムの薬理活性データ

ID	Het.	A	Phenyl substit.	R1	R2	R3	R4	R5	IC ₅₀ (μ M) Ca ²⁺ ant
1	Im	CH ₂	meta	Et	OEt	H	H	H	0.06 ± 0.01
2	Im	CH ₂	meta	Me	OEt	H	H	H	0.26 ± 0.04
3	Im	CH ₂	meta	Me	OMe	H	H	H	0.85 ± 0.01
4	Im	CH ₂	meta	H	OEt	H	H	H	1.96 ± 0.30
5	Im	CH ₂	meta	Et	O(CH ₂) ₂ N(Me)Bzl	H	H	H	0.05 ± 0.01
6	Im	CH ₂	ortho	Et	OEt	H	H	H	0.46 ± 0.03
7	Im	CH ₂	meta	Et	OEt	H	H	H	1.67 ± 0.10
8	Im	CH ₂	meta	Et	OEt	H	H	Me	8.84 ± 0.20
9	Im	CH ₂	meta	Me	OEt	H	OMe	H	1.95 ± 0.20
10	Im	CH ₂	meta	Me	OEt	H	O(CH ₂) ₂ NHPh	H	3.03 ± 0.70
11	Im	CH ₂	meta	Me	OEt	H	O(CH ₂) ₂ NH ₂	H	2.37 ± 0.50
12	Im	CH ₂	meta	Et	OEt	Me	Me	H	0.97 ± 0.06
13	Py	CH ₂	meta	Et	OEt	H	H	H	0.10 ± 0.01
14	Py	CH ₂	meta	iBu	OiBu	H	H	H	1.50 ± 0.20
15	Py	CH ₂	meta	Et	O(CH ₂) ₂ N(Me)Bzl	H	H	H	0.06 ± 0.01

今回の解析に用いた全化合物の三次元構造式を図 1 に示す。この三次元構造式は ADAPT の簡易分子力学で二次元から三次元に立ち上げた後、MOPAC の AMI にて構造最適化を行っている。この正規化された三次元化合物同志を重ねあわせた。なお、重ね合わせの基準とした原子は、図 1 の基本構造式上の白丸で示された四原子である。

図 1 . 解析に用いた化合物

・化合物の領域化

これらの重ね合わされた化合物群を三次元ディスプレイ上に表示させながら化合物を適当な領域（ブロック）単位に分割する。図 2 にはこの三次元分割作業を行っている状態が示されている。この分割された化合物の領域を二次元

的に表示したのが図 である。この図では Dihydropyridine 誘導体が総計四つの領域に分割されている。

図 . 重ねあわせ化合物群の領域化：領域過程の図

図 . 重ね合わせ化合物群の領域化：各領域の三面図

図 . 各化合物の領域化と部分構造図：分割領域に帰属する部分構造図

化合物構造式の領域分割のポイントは構造一活性相関的に矛盾の無い分割を行う事である。この意味で、各領域は薬理活性的に意味のあるファーマコフオアレベル以上の構造単位を有している事が必要である。また、薬理活性的に興味のある構造単位があるならば取りあえず領域化しておき、実際の解析によりその領域が真に重要な働きをするか否かを確かめる事も可能となる。

今回はジヒドロピリジン基本骨格と 1、2、6 位置換機を一領域に、3 および 5 位置換基をそれぞれ一領域とし、4 位のベンゼン誘導体を一領域とした、総計四領域を設定した。

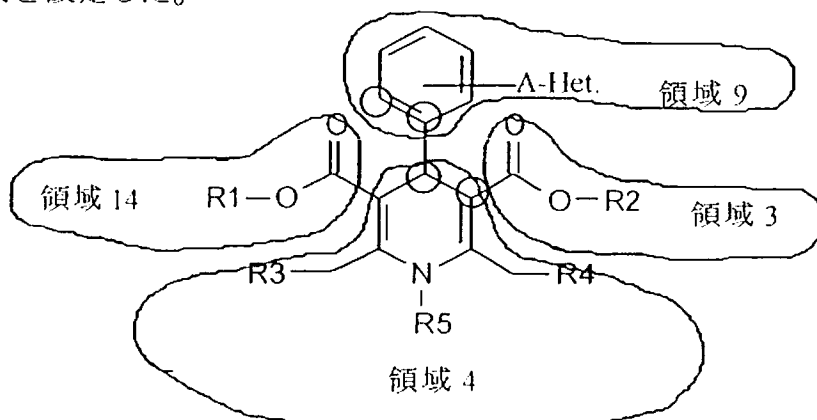


図 . 三次元的に分割された領域の二次元表示

・各領域のパラメータ化

各領域に属する部分構造を対象とし、種々のアルゴリズムによりパラメータ化する。SARDE はパターン認識法の章で説明した ADAPT システムを基本としており、従って SARDE で創出されるパラメータは一部を除いて ADAPT のパラメータと全く同じである。今回は個々の領域単位にトポロジカル、トポグラフィカル、および物理化学的パラメータの 71 パラメータを発生させた。なお、

パラメータ比較という観点で化合物の全構造を対象とした全構造パラメータも同じ71個発生させた。

・最終重回帰式

最終的に得られた重回帰式を以下に示す。

$$\text{LogCa}^{2+} = 7.32 (\text{SRMX : 領域4}) + 1.17 (\text{STR3 : 領域4}) - 11.69$$

$$N=13, R=0.96, S.D.=0.71, F_{(2,10)}=58.49$$

解析の過程で10番、および15番の化合物がアウトライヤーとして取り除かれている。解析の結果、領域4に関するパラメータのみが残り、それぞれSRMXがスーパーデローカライザビリティ（ラジカル：最大値）と、非結合歪みエネルギーである。また、パラメータが一個だけの場合の回帰式を以下に示す。

$$\text{LogCa}^{2+} = 7.88 (\text{SRMX : 領域4}) - 6.34$$

$$N=13, R=0.90, S.D.=1.07, F_{(1,11)}=45.62$$

パラメータは一個でも相関係数が0.90と良好な結果を示す。従って、この領域四におけるスーパーデローカライザビリティ（ラジカル：最大値）が活性発現に重要な役割を果たしている事がわかる。この事実はジヒドロピリジン誘導体がレセプターサイトにおいて酵素とジヒドロピリジン部位にてラジカル的な反応を生起している可能性を示唆している。

薬理活性が領域4だけですべて支配される事は考えにくい。前記二パラメータに続く三番目のパラメータとして領域9のパラメータ（酸素原子数）も寄与度は高くないが重要とされている。薬理活性の大部分が領域4データ決定されるにしても、領域9も何らかの形で薬理活性に寄与しているようである。

・領域パラメータと全構造パラメータとの比較

比較の為に化合物の全体構造から導かれる善構造パラメータを用いた結果を以下に示す。

$$\text{LogCa}^{2+} = 1.90 (\text{MOL8}) - 1.30 (\text{STR3}) + 4.12$$

$$N=13, R=0.61, S.D.=2.00, F_{(2,10)}=3.02$$

全構造式を対象とした解析では結果が非常に悪い。これは、この場合は化合物の全体で薬理活性を議論する事は効率が悪い事を示している。化合物全体の特性で構造-活性相関を論じる事自体に問題はないが、ケースバイケースに応じてパラメータを臨機応変に変える事が必要である。

3.7 SARDE 法適用上の利点と留意点

① SARDE 法適用上の利点

SARDE 適用上の利点は大まかな項目として以下の三点をあげることが出来る。

・回帰式の精度

SARDEは従来からの3D QSAR手法が抱えている本質的な欠陥をクリアしたアプローチである。これは、手続き過程でPLS法等の次元圧縮ツールを用いないため、回帰式の係数情報をHansch-Fujita法と同様な精度で利用できるためである。

- ・部分構造／部分空間単位での定量的構造-活性相関が可能

SARDE法では部分構造や部分空間単位での要因解析ができるので、きめ細かで理解し易い解析が可能となる。解析過程にPLS法等の次元圧縮手続を必要とする3D QSARアプローチは大きな(フィールド)単位での解析を余儀なくされ、部分領域レベルでの解析はグラフィック表示による視覚中心の定性的解析となる。

- ・座標データの変化に対する頑健性

SARDE法は化合物の初期座標の変化に強いアプローチである。CoMFAを含む多くの3D QSAR手法は化合物の重ね合わせによる初期座標の変化による影響を強く受ける。これは、空間上を細かく分割してパラメータ化しているため、微少な座標の変化であっても多くのパラメータの値が変化するためである。SARDEは大きなブロック単位でパラメータを創出するため、化合物の構造が多少変化してもブロック内で変化している限りパラメータが変化することは無い。つまり、SARDEでは初期座標の多少の変化は解析に全く影響を及ぼさない。

② SARDE法適用上での留意点

- ・領域設定の任意性

SARDE法は部分構造や部分空間単位での定量的な解析を実現する唯一の手法である。この部分構造や部分空間を決定する段階で任意性が入り込む。つまり、このような部分構造や部分空間を決めるのは人間であるという点である。この領域の切り方の違いにより解析結果が左右される。従って、実際の解析ではいろいろな切断パターンについて解析することが必要である。

この領域分けは化学的な常識を持っているならば途方もないパターンは生まれない。むしろ、常識的に活性に重要とされるもの、例えば Pharmacophore 等を中心としたグループなどを分割単位とすることが SARDE 法の効率的な利用法となる。

- ・領域間相互作用

SARDE法では各領域間は互いに独立している事が前提となり、領域間の相互作用を情報として取り出す機能はない。この点が現時点における SARDE の限界といえる。この問題に対する対応としては、相互作用が存在し、且つ薬理活性に重要と見なされる時はこれらの領域をまとめて一つの領域として解析

する事である。今後、この問題に関する根本的な対応を取るべき物と考えている。

3.8 3D QSAR 実施上での留意点

①化合物の重ね合わせでの留意点

3D QSAR 解析で注意が必要なのは最初のリガンド化合物の重ね合わせである。特にリガンド化合物群の構造的な差異が大きく、レセプターサイトの情報も無いときには重ね合わせの自由度が高くなる。SARDE 手法は重ね合わせの差異による影響に対して抵抗性を持つが、CoMFA 法は重ね合わせの差異が結果にも大きく影響する事が多いので特に注意が必要である。

②解析手法の基本の差異に基づく解析限界

3-D QSARに限らず解析手法が基本とする原理に注意する事が必要である。Hansch-Fujita 法は置換基効果を、CoMFA 法は場の効果を、SARDE 法は部分構造領域及び部分空間領域効果を基本としている。従って、解析の成否は解析目標である薬理活性を決定する要因がどの効果に支配されているかで決まる。この事実は、一時まじめに議論されたように、同じ薬理活性データを用いた場合の Hansch-Fujita 法と CoMFA 法との解析結果の比較には意味がないことを意味している。

つまり、置換基効果が支配的な事例を CoMFA や SARDE 法で解析しても成功するはずがない。逆に、場の効果が支配的な事例に Hansch-Fujita 法を適用しても良い結果は出てこない。このように、構造-活性相関解析では一つの解析手法によらず、多様な解析手法を試みる事が必要である。幸いなことに、以上の全アプローチをコンピュータ上で実施することはそんなに困難な事ではなくなってきた。

③情報と解析手法の差異の使い分け

同じ 3-D QSAR であっても CoMFA 法と SARDE 法とでは基本とする効果の種類、用いるパラメータの種類、解析手法と手続き、得られる情報の内容や精度が全く異なることを念頭に置くことが必要である。

Hansch-Fujita 法、CoMFA 法や SARDE 法を互いに比較し、その優劣を議論することの無意味さは既に理解していただけたかと思う。効果という基本概念を考えるならば、これらの定量的構造-活性相関手法は競合する物ではなく、互いに相補的な関係にあることを理解すべきである。自分の薬理活性に対して最も説明の付く解析手法での結果を採用すべきである。また、これら三種類の解析手法の基本をしっかりと理解することで誤った解析の可能性を少なくすることが必要である。

④様々な効果を具現するパラメータの混合使用の勧め

本章で説明した効果は総てパラメータの形で具現化される。薬理活性メカニズムは複雑で一つの効果だけが支配的な要因になる場合は少ないと考えるべきである。複数の効果が薬理活性に関与するならば、その解析も複数の効果を代表するパラメータ群を用いる必要がある。既に CoMFA 法では Hansch-Fujita 法の代表的なパラメータである logP を取り込んだ解析により CoMFA 単独の解析よりも良い結果を得ている。今後の定量的構造-活性相関解析は、Hansch-Fujita 法、CoMFA 法や SARDE 法で利用されるパラメータ群を混同して利用することが主流となるであろう。これにより、単独の効果を代表するパラメータ群を用いた狭い範囲の解析と比較して、より薬理活性メカニズムの実体に即した解析が可能となる。この実現には実験環境、特にコンピュータソフトウェアの整備が必要となるが、近い将来実現することを期待する。

4. まとめ

3-D QSAR は Hansch-Fujita 法的厳密性と要因解析力を備え、さらにグラフィックを多用したイメージ的な分かりやすさ/扱い易さという点で従来の構造-活性相関手法と比較して大きな改良点を持ったアプローチである。

3-D QSAR は確かに解析手法として魅力的で強力なものである。しかし、本論の“効果”の項中でも述べたように個々の構造-活性相関手法は独自の仮説に基づいて展開されており 3-D QSAR も例外ではない。この仮説の存在を忘れ、ただ闇雲にデータがあれば一つの解析手法を適用するというパターンは避けるべきであり、こうすれば不完全な解や偶然に振り回されて目的を見失うような事も無くなる。

3-D QSAR も手法として確立されつつあり、現在は特に大きな動きは無いように感じる。今後の傾向としては種々パラメータの混合使用の流れがより一層強まり、コンビナトリアルケミストリ/HTS の流れをにらみながら、予測機能を重視した展開が強まるものと考えている。