１１章　構造ー活性相関支援システム

現在、構造ー活性相関支援システムとして様々なシステムが展開されている。前章までに解説した総てのアプローチがそれぞれ何らかの形でコンピュータと関係している事を考えれば当然である。ここではこれらのシステムについて細かく述べるつもりはない。これは既に序文でも述べたように、本著を単なるシステムの解説書や紹介書にしたくないからである。

システムの解説で苦労することは時代を越えた真実をいかに取り出して説明出来るかである。表面上の解説は単なるパンフレットと変わり無く、読者には無益であり、むしろ害にすらなってしまう。コンピュータの進歩はすさまじいものがあるがソフトウエア自体もひけを取らぬ位に移り変わりが激しい。既に解説したように、構造ー活性相関自体も時代を経るにつれて様々な形や新たな手法が展開されている。システムの表面だけの解説では時間のフィルターに取り残され、時間を越えた真実を伝えることは困難である。

本著では二つの構造ー活性相関支援システムを紹介する。一つは ADAPT（Automated Data Analysis using Pattern recognition Techniques）システムで、残るシステムは EMIL（Example-Mediated-Innovation for Lead Evolution）システムである。この選択には幾つか理由がある。第一に ADAPT は著者が約二十年にわたり利用しており、今なお現役であるという正に時代を越えたシステムであること。パターン認識の技術は解析の基本技術であり、構造ー活性相関の多くの手法にて利用されていること。そして、本システムであるならば著者にとりシステムの本質部分を取り出して説明することが出来る事がある。また、EMIL システムは非常にユニークなシステムであり、このシステムが持つ機能や思想を持つシステムは世界的にも例を見ない。著者は一時期このシステムの立ち上げに関与し、内容的に良く知っている事も選択の理由である。さらに、本著に掲載した殆どの解析事例はこの二つのシステムを用いて展開されており、本著の理解という点からも ADAPT と EMIL システムの解説は最適であると考える。

1.　パターン認識による構造ー活性／毒性／物性相関、およびケモメトリクス研
究支援システム：ADAPT

1.1　ADAPT システムの全体概要

ADAPT システムは化合物構造式と薬理活性、毒性、物性との相関を求めるための研究支援システムである。これらの研究の他にも、種々のスペクトルデータを用いて解析すれば一般的なケモメトリクス研究も可能である。

本システムは化学の分野に世界で最初にパターン認識を導入した三人の一人である P.C.Jurs 教授によって開発された。従って、システムの構成自体が Jurs

教授の長年にわたる化学パターン認識研究ノウハウそのものであり、パターン認識による化学解析を実施するに最適な構造をしている。化合物構造式（二次元）のシステムへの入力から始まり、数値データへの変換、特徴抽出後の各パターン認識の適用及び予測業務までの全作業工程をコンピュータ上で流れるように実行する事が可能である。ADAPT は独立した機能を持つ、約百本にも及ぶプログラム群から構成されており、実際の解析はこれらのプログラム群をデータの流れに従って順に実行することで実現される。ADAPT の基本的なシステム構成を図　に示す。

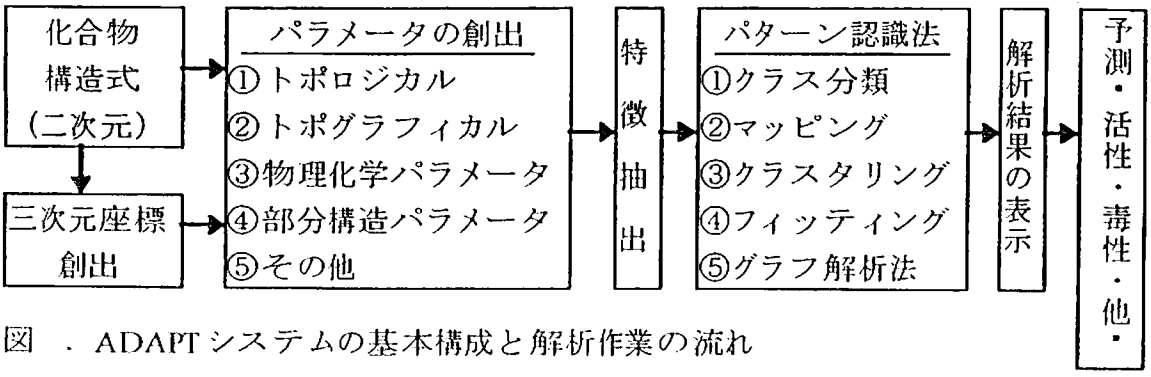


図　．ADAPT システムの基本構成と解析作業の流れ

また、表　に ADAPT の適用分やおよび関連機能をまとめたものを示す。

表　．ADAPT の適用分野およびシステム機能

適用分野／作業内容
構造ー活性相関
構造ー物性相関
構造ー毒性相関
構造ースペクトル相関
簡易配座解析

主機能
パターン認識
多変量解析
パラメータ創出
データセット作成
特徴抽出
データマネージメント

関連機能
化合物構造入力
化合物構造表示
簡易分子力学
簡易分子軌道法
エネルギーマップ
種々統計グラフ

1.2　ADAPT システムの特徴

ADAPT　システムの特徴は、化合物構造式と薬理活性データのシステムへの入力から解析結果の取り出しと活性予測までの一貫性にある。一連の解析作業中以下の二項目が特に強力であることが、ADAPT の活性・毒性・物性解析に利用出来るという汎用性を生み出している。

①二／三次元構造式から多数のパラメータ群を創出する機能

この機能はパターン認識解析に特に重要である。実際に解析を行うと、解析の律速段階はパラメータの獲得にあることがわかる。

ADAPT　は百種類以上におよぶパラメータを構造式から簡単に創出する。ま

た構造一活性相関での要因解析に重要な役割を果たす部分構造パラメータも独自に設定可能である。この結果、実用的にはADAPT内で発生されるパラメータだけで殆どの解析は実行可能である。

②特徴抽出（パラメータ選択）機能

パラメータが多いことは単に解析のスタートラインに立つことが出来たにしか過ぎない。解析の成否は、これらのパラメータ群から構造一活性相関に有効なパラメータ群のみを厳密に取り出す事が出来るか否かにかかっている。この特徴抽出機能に関しADAPTは第　章のパターン認識法でも述べたような種々の特徴抽出手法を備えている。

この全作業過程を実施するのに必要な時間は化合物数にもよるが、一般的な解析であれば、慣れると約二日程度で一次解析結果を得る事が可能である。このうち、化合物構造式と薬理活性データのシステムへの入力と三次元座標の創出作業に約一日程度必要である。残るパラメータの創出から最後の予測業務までの作業に約一日かかる。

1.3　ADAPTシステムによる解析の流れ

ADAPTシステムは、入力データとして化合物構造式と薬理活性データを用いる。標準的な解析の流れを図　に示す。この流れはシステムを使った時の操作に関するもので、得られた情報をどのように解釈し活用するかは研究者の知識とノウハウに大きく依存する。コンピュータによる総ての構造一活性相関解析システムの果たす役割は、単に良質な（使い方を誤れば単なるノイズ）情報を研究者に提供するにしかすぎない。

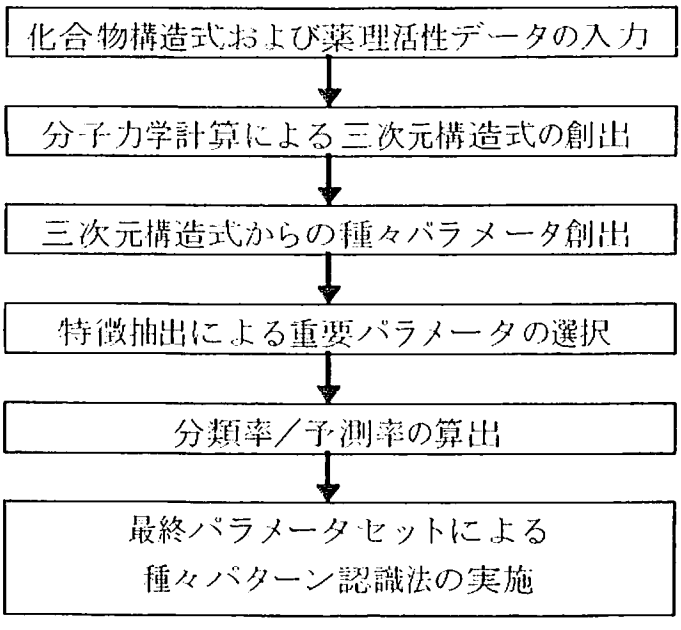


図　．ADAPTによる構造一活性相関解析の流れ

実際の解析は左図のような手順を経て行われる。最初に化合物の二次元構造式とその薬理活性データをグラフィック上から入力する。入力された二次元構造式は、簡易分子力学計算により三次元構造式へと変換される。この三次元構造式を用いて、様々なアルゴリズムに従って多数のパラメータが創出される。

幾つかの特徴抽出手法を経て最終的に統計的な信頼性基準を満たすパラメータ

セットが決定される。このパラメータセットを用いて分類率と予測率とを求める。同時に構造一活性相関の最終パラメータセットを用いて様々なパターン認識法を実行し、その出力結果等を参考に様々な情報解析を行う。

1.3.1 分子力学による三次元座標の生成

分子力学はグラフィックから直接入力された二次元構造式を三次元に立ち上げるのに利用される。ADAPT が備えている分子力学プログラムはパターン認識による解析の観点から以下に示す三つの特徴を持つ。

① ADAPTシステムが扱う化合物は必ず分子力学計算が出来る。

②定性的精度である。

③三次元立ち上げ時のテトラヘドラル炭素への導入を確実にする。

最初の①と②の項目は互いに密接に絡み合った問題である。パターン認識による解析で最も避けるべき事項は欠損データの存在である。パラメータ中に一つでも欠損データがあれば解析自体が出来なくなるか、そのパラメータを捨てるしかない。このように、欠損データの存在は解析そのものを実行不可能とするほどの重みを持つ。従って ADAPT に入力された化合物は少なくとも計算が出来なければその後の解析作業は出来なくなる。この点で①の項目は極めて重要である。また、パターン認識や多変量解析は同時に多数のパラメータを扱うアプローチである。勿論、個々のパターンの精度は高い方が望ましいが、解析結果の精度は用いたパラメータの最も精度の低いものに強く影響される。従って他の多くのアプローチと異なり、個々のパラメータの精度は解析に大きな影響を与える程のものではない。

一方で、分子力学手法は精度を高めようとするほどエネルギー計算の関数が複雑となり、計算に用いる力場パラメータ数が増大する。この結果、化合物のエネルギー計算時に用いるパラメータ不足が生じ易くなり、化合物構造の変化に対する汎用性が減少する。分子力学は、計算精度と化合物の変化に対する抵抗性は互いに反比例関係にあるといえる。

以上の事実より、ADAPT で採用される分子力学プログラムは化合物に対する汎用性（欠損データの回避）を第一優先とし、計算精度は定性的レベルで妥協している。パターン認識的な解析の観点から眺めるならば、精度は定性的レベルであっても、同一系列の化合物群に対しては常に同じ精度で安定的に計算できる事実の方が重要である。

これらの問題に対する答えとしてADAPTの分子力学計算は各歪み関数式としてパラメータ依存度の少ない式を採用している。この結果、化合物構造式の変化に対する抵抗力は高まっているが、計算精度は定性的レベルとなる。但し、原報によれば自由度の低い、リジッドな化合物（モルヒネ等）の３次元構造式

への立ち上げ精度はX線回折の結果と大差ないとしている。著者の実験では自由度の高い化合物（プロスタグランジン等）は初期座標の違いにより結果の座標データが変化するので注意が必要である。自由度の高い化合物群を用いた解析では、少なくとも初期座標は可能な限り似たものを利用するべきである。

③の特徴は二次元構造式を三次元構造式に立ち上げる時に発生する分子力学上の問題に起因する。炭素原子の場合、二次元平面構造(Square planar)から三次元に立ち上がる時に二通りのケースが考えられる。図　で示される様に、一つはテトラヘドロン体(Tetragonal)で、もう一つは四本の結合が一方向にまとまった四角錐(Square pyramidal)の形状を取る場合である。

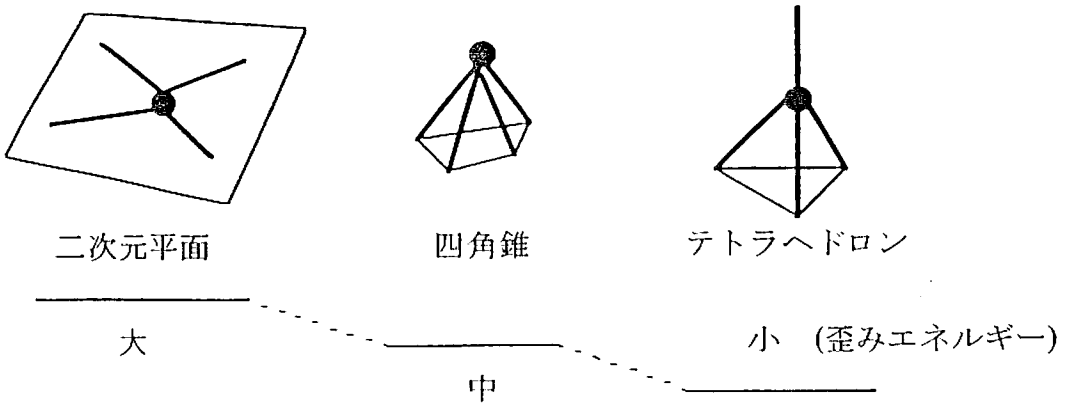


図　. 三種の形状と歪みエネルギーとの関係

エネルギー的には、テトラヘドロン体も四角錐も平面形状より安定である。従って、分子力学計算により一旦四角錐の形状に落ち込んだ炭素原子は、平面形状を経てテトラヘドロンに戻る可能性は少ない。分子力学を用いて二次元座標から三次元座標データを化学的に満足する形で創出するにはこの問題の解決が必要である。 ADAPT で採用されている分子力学計算はこの問題を回避すべく、四角錘の歪みエネルギーの方が平面構造よりも高くなる特別の歪み関数を導入している。この関数の導入により、二次元から三次元への構造式立ち上げの信頼度を高めている。

以下に ADAPT データ用いている歪みエネルギー項目を示す。

Estrain ＝ Eangle ＋ Ebond ＋ Enonbond ＋ Etortion ＋ Estereo ＋ Ehybrid

個々の関数の計算式については原著[*1]を参考されたい。これらの項目中 hybrid 項が平面構造をテトラヘドラルへと導くために導入されたものである。

＊1 ： W.Todd.Wipke, Peter H. Gund, Thomas M. Dyott and Joseph M. Verbalis, Doctors Thesis, Princeton University 08540.

1.3.2　構造式からパラメータへの変換

パターン認識や多変量解析を用いた解析で最も重要なのはパラメータであ

る。いかにパターン認識法や多変量解析手法が強力であっても、用いるパラメータの質が悪ければ精度の高い解析は望めない。また、パラメータ数が少なければ、解析自体が出来なくなる可能性が高まる。前節で述べた分子力学と同様に、パラメータも目的を限定して吟味すれば、解析目的に対する精度も必然的に高まる。一方で、解析目的の汎用性は減少する。ADAPT では多種多様な解析に耐えるように化合物構造式が持つ様々な情報を可能な限り数値データとして取り出すべく 25 本のプログラムを用いて構造情報をパラメータ化している。

パラメータを創出するプログラムの種類とそのプログラムから創出されるパラメータの数をまとめた表を表　に示す。

表　．ADAPT システムのパラメータ創出プログラムとパラメータ数

トポロジカル	トポグラフィカル	物理化学	部分構造	その他
DMALP 5	DMGEO 6	DMVOL 1	DMSSS 2	DMFRAG 19
DMCON 9	DMOMI 14	SAVOL 2	DMPATH 1	CALC 1
DKAPPA 6	DSYM 3	HMO 17	DMELEC 4	
DMCHI 19×3	LOVERB 2	MRFRAC 1	DMREF 1	
DMWP 5	SHADOW 6	DSC 3		
		STRAIN 6		
		CHARGE 4		
		MPOLR 1		
		CPSA 25		
4 4 (82)	3 1	6 0	8 (部分構造)	2 0

表中、プログラムの種類が略号で記載されているが、ここではこれら略号の説明は行わない。このように ADAPT では一つの構造式から百種類以上ものパラメータ群を簡単に創出する事が出来る。しかも、細かなオプションを設定すれば更にパラメータ数は増える。また、部分構造パラメータは一つの部分構造ごとに八種類のパラメータが創出される。一般的に一回の解析で利用される部分構造は約十種類程度であるので、この部分構造パラメータだけでも約百種類近くのパラメータが創出され、パラメータ数は更に増大する。このように、創出出来るパラメータの多さが ADAPT の最大の特徴である。この結果、化合物の構造式から得られる殆どの情報は取り出されていると考えられる。従って、安心して次のパターン認識による解析へと進む事が出来る。勿論、ADAPT 以外の外部からパラメータ群を導入することも容易に出来る。

1.3.3　ADAPT で創出されるパラメータ群の特徴

ADAPT が創出するパラメータ群の特徴は、構造ー活性相関分野で利用される Hansch-Fujita パラメータや CoMFA パラメータ等と異なり、薬理活性と直結

する、あるいは薬理活性との相関を前提としたパラメータではない事である。単に構造式の原子同志の結合情報、三次元的情報、その他の物理化学的情報等を様々な観点から網羅的に取り出しているにすぎない。これがADAPTが薬理活性のみならず、毒性、物性、スペクトル等の様々な種類の特性データとの構造相関研究で利用可能という高度な汎用性を生み出す原因である。このように多数のパラメータを扱うアプローチでは、個々のパラメータと薬理活性の相関情報は、特徴抽出を経て最終的に重要なパラメータとして選択された時点で初めて検討される。

パターン認識による解析という観点から見た時の、表　で分けられている五種類のパラメータ群の特徴を①分類能、②情報の取り出し易さ、③構造式の再現性の三つの観点から簡単にまとめてみる。但し、これらは総てのパラメータに当てはまるものではないことを予め述べておく。

①パラメータが持つ分類能力

パターン群の分類という観点から見るならば、最も分類能力の高いパラメータ群はトポロジカルパラメータである。続いてトポグラフィカルパラメータ群と物理化学的パラメータ群が中程度の分類能力を示す。分類能力が最も低いのが部分構造パラメータである。これは、部分構造の存在情報を基本としているために化合物中に同じ部分構造を持つ化合物同志の差別が付きにくい事や、部分構造以外の構造的な差異がパラメータに反映しにくいという特徴が分類率の低下に繋がっている。

②パラメータが持つ情報の取り出し易さ

構造ー活性相関解析に利用できる情報の、パラメータからの取り出しという観点で考える。トポロジカルパラメータからは具体的にイメージ可能な情報は見えてこない。一方、トポグラフィカルなパラメータ群はその生成過程から各パラメータが代表する情報が見えてくる。例えば分子主成分を用いたパラメータ群からは化合物の全体的な形状に関する情報が見えてくる。また、物理化学的パラメータ群はパラメータが示す値は必ず一つの物理化学的情報に結び付いている。同様に、部分構造パラメータは定義自体が部分構造との対応を前提としており、この点では極めて解読し易い特性を持つ。

③パラメータから構造式への再現性

パラメータの持つ情報を化合物の構造情報へと変換する機能は、ドラグデザインを行う上で極めて重要な機能である。この点で最も信用出来るのは部分構造パラメータである。予め指定された部分構造を基準にパラメータ化するので、部分構造パラメータ、即部分構造情報へと直結する。一方、トポグラフィカルパラメータは化合物の三次元形状とは結び付くが、細かな結合状態レベルの情報には結びつかない。また、物理化学的パラメータはケースバイケースではあ

るが構造式との結びつきは弱い部類に属する。トポロジカルパラメータは構造式は勿論のことで、形状情報への結び付きも殆ど不可能である。

以下に各パラメータ群の持つ前記特徴を簡単にまとめる。分類率や予測率を高めるならばトポロジカルパラメータを用い、情報の収集を優先にするならば部分構造パラメータを中心として、物理化学的およびトポグラフィカルパラメータを用いれば良い。一般的にはこれら総てのパラメータ群が共存する形で最終パラメータセットが決まる事が多い。

	トポロジカル	トポグラフィカル	物理化学的	部分構造
分類能力	大	中	中	小
情報の取り出し易さ	小	中	中	大
構造式の再現性	小	中	中	大

1.3.4　特徴抽出の実施

パターン認識による解析結果は解析目標に対して重要なパラメータ群が適切に選択されているか否かにより左右される。従って、実際に解析を行う上で最も大事な手続きがノイズパラメータを取り除いて、真に重要なパラメータのみを取り出す特徴抽出過程となる。

この特徴抽出の具体的な手続きはパラメータ群を対象にするものと、パターン（サンプル）群を対象にするものとがある。手続き的には、最初にパターンの特徴抽出を行い、ある程度パラメータ群を絞り込む。絞り込めた時点で、サンプルの選択（特徴抽出）を併用しながら最終パラメータセットを決定する。このパラメータ群の選択手法は特徴抽出の項にて三段階に分けられることは既に述べた。ADAPT ではこれらの機能を用いて各段階の特徴抽出手法を順に実行することで簡単に最終パラメータセットが得られる。ここでは、相関係数を用いてパラメータを選択する時の、パラメータの捨取選択技術に的を絞って述べる。

① 相関係数による特徴抽出の実施

相関係数による特徴抽出では採用するパラメータと、捨てるパラメータを決めることが必要である。相関が高い二つのパラメータは統計的な情報量の観点からは殆ど差がない。しかし、情報量的には同じであっても情報内容が異なるためにどちらのパラメータを捨てるかの判断が重要となる。この、パラメータの選択基準は解析目的により異なってくる。構造ー活性相関が目的である限り、最終パラメータセットが決定された後に行う要因解析での実施し易さを優先して考えることが大切である。即ち、要因解析で行うパラメータからの情報読み出しの負担を軽くするパラメータ群を優先的に選択すべきである。ドラグデザインで新薬のデザインが重要であれば構造式に直結するパラメータを選択するべきである。単純に分類だけが目的であればパラメータの選択に気を使う

必要はない。

パラメータの選択は前節でまとめた各パラメータ群の特徴を参考にして簡単に行う事が可能である。例えばドラグデザインが目的であれば、構造式への再現性が重要な問題となる。従って、構造再現性の高い部分構造パラメータが最重要となり、物理化学的およびトポグラフィカルパラメータが続く。最も重要性の低いパラメータはトポロジカルパラメータとなる。このような選択を行うことで、最終目標であるドラグデザインを行うに重要な情報が煮詰められてくる。この手続きが中途半端であるならば、真の目的であるドラグデザインの実施が困難となる。

② 解析手法に強く依存した特徴抽出手法

最終段階で実施される強力な特徴抽出手法は個々の解析手法の特性を利用した手法となる。例えば、二クラス分類を行う時は二クラス分類機である線形学習機械法の特性を利用した特徴抽出（バリアンスウエイト法、ウエイトサイン　法：パターン認識法の章参照）が最適である。

線形重回帰や非線形重回帰等のフィッティング手法に関してはこの解析手法に適した特徴抽出手法が必要である。F 値、T 値を指標とした手法や、総ての回帰式を計算してその良否を判定する手法等、自分のデータ内容に応じて特徴抽出手法を選択適用可能である。

1.3.5 種々のパターン認識法の実行

特徴抽出を経た最終パラメータセットを用いてクラス分類からマッピング、クラスタリング、そしてフィッティング等、種々のパターン認識法を実行する。一般的にパターン認識手法の数が注目されるが、実際の構造ー活性相関で利用される解析手法は多くなく、例えば、ALS 法、SIMCA 法、PLS 法（最初はスペクトル分析用として開発）等は構造ー活性相関解析用に特別に開発されている。また、名前は同じでも中身は構造ー活性相関用に改造された手法（例えば、Hansch らが用いている線形重回帰プログラム、市川が開発したニューラルネットワーク[*1, 2]等）も多い。

＊1）T. Aoyama, Y. Suzuki and H. Ichikawa, "Neural Networks Applied to Quantitative Structure-Activity Relationships", J. Med. Chem. , 33 , 905-908 (1990).

＊1）T. Aoyama, Y. Suzuki and H. Ichikawa, "Neural Networks Applied to Quantitative Structure-Activity Relationship (QSAR) Analysis", J. Med. Chem. , 33 , 2583-2590 (1990).

ADAPT は構造ー活性相関上でのパターン認識の実行に充分な種類の解析手法と、構造ー活性相関用に改造された手法を備えている。これらの解析手法群は簡単な操作で即座に実行する事が出来る。また、パターン認識では計算結果が様々なチャートとして出力されることが多いが、これらのチャート上の各サ

ンプルポイントをヒットすればそのポイントに該当する構造式に変換され、そのチャート上に重ねて表示される。この機能は出力チャートを見て様々な要因解析を行うのには必須の機能である。この機能が無ければ各化合物は各チャート上での単なる点にしか過ぎず、全体的な傾向だけしか議論できなくなる。

1.3.6　活性予測

活性の予測の手続きは、①予測率を求める作業と、②実際に活性未知の化合物群の活性予測を行う場合の二通りがある。予測率は leave N out 法が実行出来るように設計されており、N の値を自由に設定する事が可能である。

新規化合物の活性予測は手続き的には通常の解析と同様な手順を取る。但し、発生するパラメータの種類は判別関数や回帰式で利用するものに限定される。この生成されたパラメータと、予めシステム内に保存しておいた判別関数や回帰式等を用いて活性の予測を実行する。判別関数や回帰式はそれぞれ最大 20 個保存可能である。

1.4　解析目標とシステム／プログラム開発目標との差異

ここでは ADAPT における解析プログラムのパラメータ扱いの差異とプログラムの差異について述べる。ADAPT では線形及び非線形重回帰に関係する一連の作業において、総てのパラメータはオートスケーリングされない生のデータを用いる。従って、オートスケーリングされたデータを用いて線形／非線形重回帰を行う時はオプションとして追加操作をする事が必要である。この処置は Hansch-Fujita 法を意識したものであり、ADAPT が単なるパターン認識解析システムでなく、構造ー活性相関解析を強く意識したシステムであることを意味する。一方、この線形／非線形重回帰以外の解析手法に関して ADAPT はオートスケーリングされたデータを用いるのが標準仕様であり、生のデータを用いる為にはオプションの設定が必要である。

表　. データ処理に関する ADAPT システムの仕様

解析手法	標準設定（デフォルト）	オプション設定
線形／非線形重回帰関連解析	生データ使用	オートスケーリング適用
左記以外のパターン認識解析	オートスケーリング適用	生データ使用

このように同じ基本機能を持つシステムであっても、解析目的が異なれば細かな点での仕様が異なってくる。一般的に、システムをそのシステムが開発目標とする内容と異なる作業に適用する場合には、例え適用が不可能ではないにしても細かな点での修正作業が必要となる。特に、目に見えない部分での細かな差異はシステムの操作過程で特に意識しない項目だけに忘れやすく、重大な間違いを起こす原因となりかねない。従って、自分が行う作業内容とシステム

が開発目標とする内容間とで差異の少ないシステムを用いることが効率の良い、失敗の無い解析を行うのに必要である。

以上のようなデータ設定の差異もあるが、解析プログラム自体もその使用目的の差異により使い勝手がかなり変化することも認識することが必要である。例えばADAPTには線形重回帰用プログラムとして二種類のプログラムが存在する。一つは Jurs 研においてケモメトリクス解析を意識して開発されたもの（MLRAプログラム）、残る一つはHansch研にて開発されたプログラム（IRAプログラム）である。両プログラムとも線形重回帰を実行するプログラムであるが、構造一活性相関を行う観点からはIRAの方がはるかに使いやすい。この差異は主としてプログラムの開発時に構造一活性相関解析の流れと、この作業に必要な手続きを意識して開発するか否かの差異により生じる。プログラムの開発目標と研究者の解析目標が一致した場合、ユーザ（研究者）は解析の流れを特に意識することなく自然に正しい解析を実行する事が出来る。これが、理想的なシステム開発であり、この実現には開発目標の作業内容や手続きに関する深い知識とノウハウが必要である。

システムのこのような差異は使ってみて初めてわかるものである。システムに組み込まれた高度なノウハウはちょうど隠し味のように、表面からは見えないが使うと味が出てくるものである。同じ解析を行うにしても、汎用を目的として開発されたプログラムを用いて特定の解析を行う事が困難であるのはこのような差異が大きく作用しているためである。

2. 過去の化合物展開ノウハウを利用したバイオアナログス化合物創出支援システム EMIL

2.1 EMIL システム

① EMIL システムの狙い（過去の構造変換事例ノウハウの利用）

現在に至るまでには膨大な数の医薬品が開発されてきた。構造式の一部を変化させる事で薬理活性が向上した例、低下した例、毒性が強まった例、等々の構造修飾情報と薬理活性との相関が多数の事例として存在する。この新薬開発過程では多くの研究者が試行錯誤を続け、その時点での最新、最強の技術を駆使している。即ち、医薬品開発過程の足跡自体が構造一活性相関に関するノウハウの結晶と考える事ができる。優秀な研究者はこれらの過去の情報を理解／整理することで自分のものとし、新たな医薬品開発の糧とする。

このような医薬品開発過程に内在する研究者のノウハウ（化合物変換情報）を利用出来るならば、理論や原理に基づいたアプローチに次ぐ新たなドラグデザイン手法となる可能性が高い。このように、理論や原理／定理、公式のような具体的な形とならないノウハウ情報を積極的に利用するコンピュータ技術

として人工知能がある。藤田らは理論や原理としてまとめる事が困難である新薬開発過程における化合物変換パターンを、化合物変換ルールとして置き換えた人工知能システム EMIL を開発した。

② EMIL システム概要

・医薬品の開発過程事例

医薬品でも農薬でもリード化合物が発見されただけでは薬物にならない。このリード化合物を、より活性が高く、毒性／副作用の無い薬物へと導く事が必要である。リード化合物の発見はこのような一連の作業のきっかけにすぎない。このように個々の薬物は開発の歴史を背負っており、この歴史をまとめて見ることで多くの情報がえられる。以下にこのような開発の流れを簡単にまとめたものをあげる。

化合物変換事例図

・化合物変換ルール事例（EMIL の持つ情報）

薬物の変換過程をコンピュータに情報として取り込む事が必要である。最も重要な情報は化合物の変換パターンであり、この情報を補足するデータとして、薬理活性や毒性／副作用等の変化、物性等の変化情報がある。化合物構造式の変化単位に、これらの付随情報をまとめて一つのルールとする。

ルールをシステムに入力する時は、これらの化合物変換事例をまとめて一構造変化単位に一枚のシートに書き込む。この化合物変換事例シートの一例を図に示す。

図　．構造式変換事例シート

現在、EMIL システムは医薬および農薬の分野において総数　　個のルールを備えている。これらのルールは更に細かく医薬は　個のグループに、農薬は

個のグループに細分化合物されている。EMIL が備えるルールの分類を図に示す。これらのルールは各カテゴリー順に順次適用され、状況に応じて医薬品すべて、農薬関連のルール総てを適用したり、医薬／農薬のルールを混在さ

せて適用する事も可能である。

図　．EMILの構造変換ルールの分類図

・EMILシステム構成

図　にEMIL システムの構成を示す。

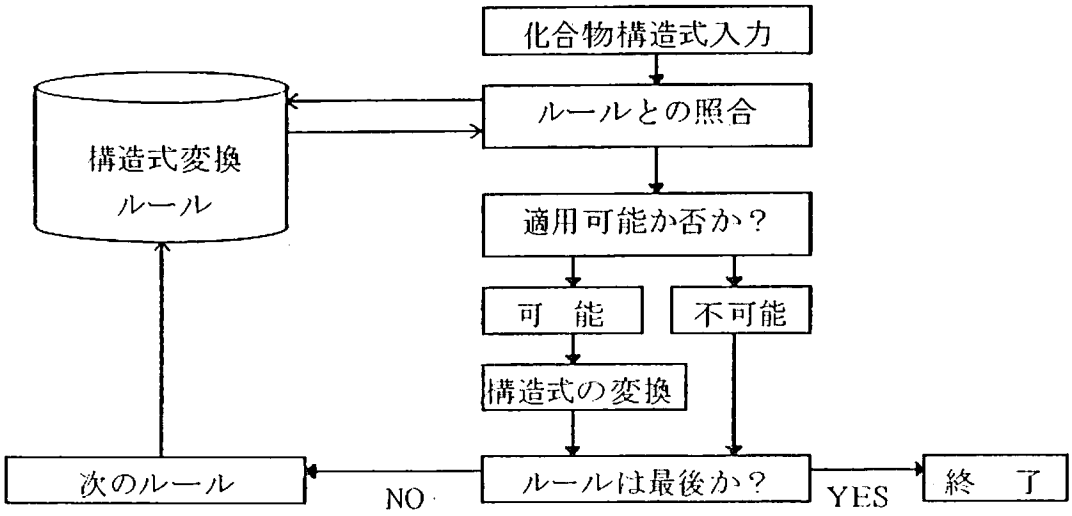


図　．Emilによる構造式変換ルールの適用（一段階のみ）流れ図

システムには構造変換の対象となる初期二次元化合物構造式が入力される。最初の構造式変換ルールがルールベースより取り出され、初期構造式が取り出された構造式変換ルールの適用条件を満たしていれば実際に構造変換がおこなわれる。用いたルールが最後でなければ新しいルールを読み出し、再び入力構造式への適用可能性をチェックする。構造ー活性相関の手続きを全ルールを適用するまで続ける。全ルールを適用した時点で構造変換の第一ステップが終了する。以下、第一段階で変換された個々の構造式を出発構造式として再び変換ルールの適用を試みる。構造ー活性相関の手続きを繰り返す。

本アプローチでは変換されて生成する化合物数はルールの適用回数が増えるに従って急速に増大する。従って、通常は各ステップ毎に出力された構造式を研究者が見て、意味のある、あるいは有意義な変換がなされている化合物構造式を取り出し、この取り出された化合物についてのみ次のステップに進ませる。

・バイオアナログス化合物創出事例

実際にEMIL システムを用いて構造変換させた事例を以下に示す。

図　．構造式の変換事例

③ EMIL システムの効果

・EMIL システムにおける構造式創出（段階的アプローチ）の特徴

EMIL システムは新規の化合物を創出するドラグデザインシステムであるが、De novo デザインで行われるような部品の組み合わせによる画一的な化合物創出は行わない。あくまでも現時点の化合物構造式をチエックし、その構造式に適用出来る構造変換の可能性を評価しながらステップバイステップで構造式を創出して行く。従って、創出化合物群は適用されたルール数に比例したツリー構造を有している。

・分野の違う構造式変換ルールの適用効果

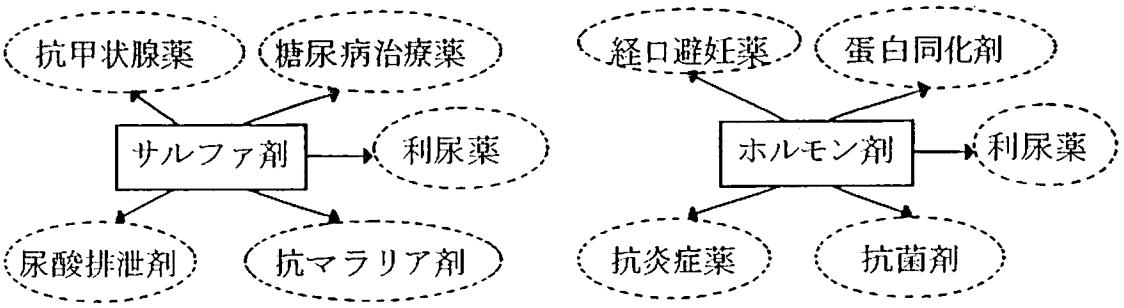
EMIL が持つ最大の特徴の一つとして、分野の異なる構造変換ルールを適用出来ることがある。EMIL 以外の総ての構造ー活性相関手法は、解析に用いた薬理活性から離れた解析／適用が出来ないことと対象的である。現在、EMIL の構造変換ルールは薬理活性群別を基本とし、その上に医／農薬別の構成からなっている。例えば、抗菌薬の構造変換に抗菌活性ルールを用い、続いて抗腫瘍活性のルールを適用し、さらには医薬から農薬へと構造変換ルールを拡大する事も可能である。図　に現時点で EMIL が持つ構造変換ルールの階層構造を示す。

図　xxx

構造変換に目的薬理活性分野以外の構造変換ルールを適用することは機能の過剰適用とは言えない。むしろ積極的に推進されるべきものである。このような変換により新規性の高いリード候補化合物が発見される可能性が高まる。また、一般の環境下では分野の異なる医薬品についての情報を効果的に利用する事は困難である。ましては、医薬品と農薬とでは研究者も研究手法も変わってくる。しかし、現実の薬理活性ではこのような垣根を越えた事例が多数存在する。このように、薬理活性や医薬／農薬の垣根を簡単に乗り越えられるのも EMIL の特徴である。

既に実用化された医／農薬間にも医薬として開発されたものが農薬として利用された事例は多数存在する。また、医薬品の改良過程で薬理活性の異なる

医薬品へと導かれた例は多数存在する。典型的な例としては抗菌性スルホンアミドがドラグデザイン過程で利尿薬、抗マラリア薬、抗甲状腺薬、糖尿病治療薬、尿酸排泄薬等の様々な薬品へと導かれている。また、ステロイド誘導体も本来のホルモン作用から、利尿薬、抗炎症薬、抗菌薬、経口避妊薬等の作用を持つ薬物が開発されている。



農薬／医薬品の相互変換も多数おこなわれている。

・知識（情報）の増大による成長

人工知能システムの特徴として、ルールを増やす事が出来る事がある。丁度人間においても知識が増えて、より高度／複雑な状況に対応出来る用になるように、人工知能システムもルールを増やすことで成長する。

但し、人工知能システム全般に共通であるが、ルールが増えると既存のルールとの競合が起こりやすくなる。また、ノイズの原因となるルールも増えてくる。ルールが増えると計算時間が急速に増大する等の副作用的な事象が生じてくる。このような点に留意しながら新たなルールを追加すればルール増大の効果を期待出来る。

・創出構造式の構造変化量

EMIL により出発化合物を中心とした構造変換がなされ、新規化合物群が創出される。一段階（ルールが一回適用される）のステップにおける構造修正量には限界がある。従って、化合物構造式の変化量（分子多様性）はルールが適用された回数に比例して大きくなる。

・創出化合物の数

EMIL では適用可能であれば一つのルール毎に化合物が創出される。従ってルールの数が増えればその分創出化合物数も増大する。例えばルールの適用率が 100%であれば、100 種類のルールを一回適用すれば 100 化合物が、創出された全化合物にさらにルールを一回適用すれば 100^2 化合物が創出される。N 回適用すると 100^N 化合物となり、コンピュータでも処理しきれない数となる。従って、システム上では一段階ごとに研究者が化合物をチェックし、選択され

た化合物群にのみ新たなルールを適用するインタラクティブな手法を取る。

・バーチャル（仮想）化合物ライブラリの構築

EMIL システムを用いる事でコンビナトリアルケミストリ／HTS で利用される化合物ライブラリを構築する事が出来る。EMIL で創出される化合物群はその殆どが合成されていない仮想上の化合物群であるので、バーチャルバイオアナロガス化合物ライブラリとなる。EMIL システムを用いた化合物ライブラリの構築については、EMIL システムより生成された化合物群の薬理活性予測を ADAPT システムにて行うアプローチについて第　章にて解説してある。

3. プログラムやシステムとのつきあい方

プログラムやシステムは自動車と同様、単なる便利な道具にしかすぎない。コンピュータが介在し、複雑な計算をして解析結果が綺麗なグラフィック等で出力されるとしばしば計算結果は絶対であるという錯覚に陥る。あなたは一度でも解析プログラムの結果を疑って、もう一度解析過程をチェックしたことが有りますか？と問いたい。一度も疑ったことがなければ、あなたは充分にコンピュータ信奉者であると言えよう。

ここでは ADAPT や EMIL に限らず、一般的にプログラムやシステムというものを利用する時に留意すべき事項について簡単にまとめる。

①プログラムの制限／限界事項の把握

化学分野の解析プログラムで万能なものは無いと言える。どのプログラムも何らかの形で制限項目や限界事項を持っている。これ自体はプログラムとして欠陥でも何でもない。プログラムを単に闇雲に使えば、個々のプログラムが持つ限界事項を簡単に越えて計算する。大きなシステムになるほど注意が必要である。プログラムが動かなければラッキーであるが、プログラムがそのまま動いてそれらしい答えを出す。これが最も恐ろしい事であるが意外と気がつかない。プログラムを使う時に留意すべき最も大事な点は、この制限や限界事項を正しく理解することである。

・制限／限界事項を理解すること

・制限／限界事項を越えた時のプログラムの処理方法

・制限／限界事項を越えた時、他のプログラムに与える影響

以上の内容を把握していれば殆どの場合に対処可能である。しかし、これらの項目中・はマニュアル等から分かるにしても、・や・を知ることは開発者かソースプログラムを解読する以外には基本的に難しい。しかし、これらの事項は個々のプログラムの基本原理を理解することで充分に予測可能である。また、数多くプログラムを動かす過程でノウハウとして蓄える事も重要である。

②プログラム稼働の為の準備や環境作り

道具であるプログラムは使いこなして初めてその本領を発揮できる。この使いこなすとは単にコンピュータを昼夜動かして元を取るという事ではない。個々のプログラムの内容を正しく理解して、その力が充分に発揮出来るように種々環境を整えることを意味している。つまり、プログラムが持つ機能を充分に発揮出来るように、データ設定や種々のオプションを解析状況に応じて変化させ、最高の計算環境下で解析出来るようにすることである。

・入力データ等の吟味をする

・データの内容に合わせてプログラムの種々設定を行う

入力データの吟味は非常に重要である。単にデータさえあれば入力してしまうようでは先の①の事項でトラブルに巻き込まれることが多いし、解析結果の精度も低下する。プログラムに限らず一般の解析でも、用いるデータの品質を一定水準に揃えることは基本中の基本である。

・のプログラムのオプション設定は功罪相半ばする事項である。オプションはプログラム開発者が様々な状況を設定して組まれている。中途半端な知識でオプションをいじれば開発者が意図した目的や環境と異なる状況で適用される可能性がある。総てがこのように重要なオプションではないが、充分なノウハウを身につけるまでは失敗覚悟で繰り返し学習する事が必要である。

③プログラムやシステムをじっくり使う

自動車も素晴らしい道具であるが、これを使うには免許がいる。大きなシステムを動かすとなると、ある意味では自動車以上のスキルが要求されるが実際には免許はいらない。最近のシステムは使いやすさを中心に設計されているので、初心者でもプログラムは動いてしまう。これはちょうど、運転免許も無いのに自動車を動かすようなものである。自動車も乗りこなすには充分の経験が必要であるように、プログラムやシステムを使いこなす為にもそれなりの経験が必要である。

最近はワープロでも新機能を盛り込んだものが次々と目移りするような位に出てくる。あなたは現在使っているワープロを十分に使いこなしていますか？と問いたい。化学分野でも新機能を盛り込んだシステムが続々と出てくるが、その陰にある基本原理は時代を越えて殆ど変化していない。少数であってもそのプログラムやシステムを十分に使いこなすことができれば、その陰にある基本原理や技術を理解できるようになる。基本原理や技術が身に付けば新たなプログラムやシステムに振り回される事無く、それらを正しく評価出来るようになる。次々と目移りしてプログラムやシステム放浪するのは、道具がリニューアルされて数も増えて仕事が出来るように感じるが、本人の実力向上には何の効果も及ぼさない。

④とにかく使ってみる

いろいろ書いたが道具は使わなければスキルは上がらない。車と違い失敗しても大事故になる事はない。せいぜいシステムがダウンするか、パソコンであればリセットすれば済むことである。自動車も本から学んだだけでは動かせない。プログラムも同じで、実際に動かしてディスプレイの出力を見て結果を出さなければスキルは向上しない。マニュアルにないデータを用いて実行し、そこで初めて真の学習が出来る。

欲張るならば、プログラムを動かしたならばその陰で動いている基本原理に興味をもってほしい。分子が二次元構造式から三次元構造式になったならば、何故そうなるか、分子力学とは、分子軌道法とは、化合物により計算出来るものと計算出来ない物が有るのは何故なのか等、プログラムを動かせば何かが出てくるはずである。プログラムは必ず何らかの基本原理に従って作成されている。これは論文や雑誌等を調べれば必ず出てくる。これらを学べば、プログラムの中身を知らずとも①で述べた制限や限界事項について分かるようになり、プログラムを暴走させたりダウンさせるようなことは少なくなるし、計算結果に過剰な期待をすることも、失望することも無くなる。

4. まとめ

ADAPT 及び EMIL システムについてその内容を簡単に説明した。これらのプログラムを通じて構造一活性相関支援システムに限らず、化学関連システム全般に興味を持っていただければ幸いである。

ADAPT も EMIL も使えば本当に便りになるシステムであることは事実である。

システムの長所や欠点を理解した上でこれらのシステムを使うことで、自分の研究の仕事の幅が広まることを是非実感していただきたい。また、システムの背景となっている基本技術を理解する事が出来れば、時代の変化に取り残されることなく常に最新かつ独自の研究が出来る。

例えば、化学分野におけるパターン認識の技術一つ取っても時代により大きな山や谷が存在している。Jurs らがパターン認識を化学分野に取り込んだ時、その適用対象は分析化学、特にスペクトル解析分野が中心であった。その後、化学パターン認識が急速に普及するとともに分析分野から一般化学や裁判化学（Forensic chemistry）までと適用分野が急速に拡大した。構造一活性相関への適用が始まったのもこの時代である。パターン認識法として構造一活性相関に一つの分野を築いたが、個々のパターン認識技術は本章中でも述べたように殆ど総ての構造一活性相関手法の中で何らかの形で利用されている。最近ではコンビナトリアルケミストリ／HTS における化合物ライブラリ構築の基本技術として注目されている。歴史は繰り返すというが、このコンビナトリアルケ

ミストリ／HTSで利用されているパターン認識技術はその昔、情報化学分野で展開されてきたパターン認識関連技術と殆ど変わるところは無い。勿論、ニューラルネットワークや遺伝的アルゴリズム等の新しいパターン認識法が利用されている点で目新しさは感じられるがそれだけである。前節で述べたように時代が変わってもその中身は大きく変化していないことの一つの例である。