

# 化学物質の毒性情報と構造活性相関予測

## Toxicity Information and Its QSAR Prediction of Chemical Substances

田辺 和俊<sup>\*,†</sup> 大森 紀人<sup>\*</sup> 小野 修一郎<sup>\*</sup> 松本 高利<sup>‡</sup>  
長嶋 雲兵<sup>§</sup> 上坂 博亨<sup>¶</sup> 鈴木 孝弘<sup>||</sup>

Kazutoshi TANABE Norihito OHMORI Shuichiro ONO  
Takatoshi MATSUMOTO Umpei NAGASHIMA  
Hiroyuki UESAKA and Takahiro SUZUKI

近年、化学物質の毒性に関する情報の取得が地球的な規模で喫緊の課題となっている。しかし、動物を用いる安全性試験は莫大な時間と費用がかかるため、毒性が未知の全ての化学物質について動物試験により毒性を評価することは不可能である。また、それらの情報を集録した毒性データベースにも様々な問題点がある。そこで、構造活性相関、特にコンピュータを利用した定量的構造活性相関 (QSAR) による毒性予測が化学物質管理の観点から重要になっており、多くの毒性予測システムが開発されている。しかし、QSAR による既存の毒性予測システムの成績は実用的には不十分であり、世界中の多くの研究者がこの問題に取り組んでいる。我々は既存のシステムより高精度かつ高汎用性の予測システムの開発を目指して、ニューラルネットワークを用いた毒性予測手法を研究している。しかし、この問題の解決には信頼度の高い毒性データの収集を始めとして多くの課題が横たわっている。

Recently, it is globally necessary to collect data on toxicities of numerous chemical substances available in market. But safety tests using animals take huge cost and period, and it is impossible to get toxicity data on all unascertained chemical substances by animal test. Also there are many problems in existing databases which collect toxicity data on chemical substances. As a screening method for animal tests, a toxicity prediction on a basis of quantitative structure-activity relationship (QSAR) is meaningful. Several toxicity prediction systems with QSAR have been developed, but the performances of most existing systems are highly questionable. In order to develop a toxicity prediction system for numerous chemical substances showing a higher performance than those of existing systems, we are applying a neural network technique. Several problems on toxicity information and its QSAR prediction of chemical substances are discussed.

キーワード：化学物質の毒性、毒性データベース、構造活性相関、QSAR、ニューラルネットワーク  
Toxicity of Chemical Substances, Toxicity Database, Structure-Activity Relationship, QSAR, Neural Network

### 1 はじめに

人類は新規の化学物質を続々合成し、我々の生活及び健康の向上に役立ててきた。し

かし、一方でそれら化学物質の毒性が現在、地球上で深刻な問題になっている。すでに1962年、Rachel Carsonは有名な書物“Silent Spring”<sup>[1]</sup>を出版し、その中で農薬を中心とする化学物質の毒性について人類に警告した。それにもかかわらず、有毒な化学物質に起因する様々な事故、事件が地球上のあらゆる地域で発生した。この原因は、大多数の化学物質がその毒性が未知のまま我々の周囲に存在しているからである。そのため、化学物質の

\* 千葉工業大学社会システム科学部経営情報科学科

† kazutoshi.tanabe@it-chiba.ac.jp

‡ 東北大学多元物質科学研究所

§ 産業技術総合研究所計算科学研究部門

¶ 富山国際大学地域学部

|| 東洋大学経済学部経済学科

毒性に関する情報の取得が地球的な規模で喫緊の課題となっている。

本論文では、化学物質の毒性情報とその予測技術の現状と課題、及びこの課題に関する我々の研究成果を総括する。

## 2 化学物質の毒性

有毒な化学物質による事故は古来より多発しているが、最近起きた多数の被害者が出た大事故としては表1に示すようなものがある。化学物質の危険度を定量的に示すリスクは、毒性の強さだけでは表されず、その毒物にどの程度接触するかという接触頻度あるいは接触確率と毒性との積で表される。表1に示す化学物質は、毒性はきわめて高いが、法律で管理が規制されているので、一般市民が接触する確率は低い。したがって、一般市民にとってこれらの化学物質のリスクはきわめて低い(ただし、ダイオキシンは以下のようにリスクが高いのでこれらには含まれない)。

一方、近年、日本では表2に示すような化学物質が環境中から検出され、大きな社会問題になった。これらは、環境中の存在量がごく微量であるにもかかわらず、一般市民が接触する頻度が高いため、リスクが高く、したがって監視が必要な化学物質である。

しかし、大多数の化学物質はその毒性が不明なまま我々の周囲に存在しているのが現状である。もちろん、医薬品をはじめ、食品添加物、化粧品、農薬、殺虫剤等は人体との関わりが深く、これらの毒性についてはそれぞれ厳しい基準が設けられている。また、その他の工業化学物質についても、中間体、製品、副生物、不純物を含め、環境汚染、食物連鎖等を通じての健康障害防止の面から、わが国では化審法(化学物質の審査及び製造等の規制に関する法律)により毒性に関し規制がなされている。

さらに、これらの化学物質の製造現場における労働災害防止のために、労安法(労働安全衛生法)に基づいた毒性調査、届出が定められている。米国では Toxic Substances Control Act(TSCA) に基づいて同様の審査が行われるが、新規化学物質の登録を管轄する Environmental Protection Agency(EPA) の発表では、申請される化学物質の数は毎年約 2,000 種にのぼるといふ<sup>[2]</sup>。

一方、化学物質の登録機関 Chemical Abstracts Service(CAS) が登録した化学物質の総数はすでに 2,700 万種類を越え、毎日 4,000 種類の化学物質が登録されている<sup>[3]</sup>。この 2,700 万種類という数は学術論文に登場した、すなわち化学の実験室で合成された化学物質の総

表1 有毒化学物質による最近起きた代表的な大規模事故・事件

化学物質名	事故・事件内容
イペリット, 青酸ガス, アルシン等	1974年以降, 中国や日本の各地で旧日本軍遺棄の毒ガスによる被害続出
ダイオキシン	1976年, イタリア・セブソの化学工場が爆発し大量放出, 被災者22万人, 住宅地域1,800haが汚染, 10年間居住禁止
シアン化合物	1984年, インド・ボパールの米国ユニオンカーバイドの子会社の農薬工場が爆発し噴出, 死者4千人, 被災者20万人
水銀等	1986年, スイス・バーゼルの薬品会社火災でライン川に流出, 流域のドイツ, フランス, オランダ等が大規模汚染
サリン	1995年, 日本・東京の地下鉄で散布, 死者12人, 被害者3,796人
ニトロベンゼン等	2005年, 中国・東北部の化学工場の爆発事故で河川に流出, ロシア極東のハバロフスクに到達

表2 環境中に存在するリスクの大きい代表的な化学物質

化学物質名	存在環境	発生原因
ダイオキシン	大気, 河川, 土壌, 農作物	廃棄物の焼却, 農薬等
ベンゾピレン	室内空気, 大気	タバコの煙, 自動車排ガス
フタル酸エステル	プラスチック製品	可塑剤
トリハロメタン	水道水	塩素殺菌
トリクロロエチレン	土壌, 地下水	半導体工場
ホルムアルデヒド	室内空気	建材等
アスベスト	大気, 室内空気	建材, 耐熱材等

数であり, それらすべての化学物質が一般市民にとって危険を及ぼす訳ではない. 一般市民にとってその毒性が問題になるのは市場で流通している化学物質であり, その総数は約10万種類程度である. しかし, この市場で流通している化学物質の中で毒性が判明している化学物質の総数は実はきわめて僅少(毒性の信頼性の高い物質数は数千にすぎない)である.

化学物質の毒性を評価するためには通常, 動物を用いた試験が行われる. しかし, 表3に示すように, 安全性の動物試験には長期の年月と巨額の費用が必要である<sup>[4]</sup>. また, 動物愛護の観点からも多数の動物を用いる安全性試験は好ましくない. したがって, 毒性が未知の約10万種類もの莫大な化学物質の全てについて動物試験により毒性を評価することは現実的には不可能である.

### 3 毒性データベース

動物試験による毒性評価の結果を収録した各種の毒性データベースがインターネット, CD-ROM, 冊子体等の媒体で公開されており, インターネットで利用可能な毒性データベースとしては表4に示すものがある<sup>[5-7]</sup>. これらのデータベースに収録されている化学物質については, 手軽に毒性情報を入手することが可能である. しかし, その内容や利用に関しては色々な問題点がある.

まず第1は, データの内容や形式が不統一であり, データベースが使いにくい点である.

表3 毒性評価のための動物試験の費用と期間(化学物質1種当たり)<sup>[4]</sup>

試験項目	費用(千円)	期間
発ガン性	187,500	3年
慢性毒性	114,000	1年半
生殖試験	45,500~ 60,000	3年
催奇形性	34,000	1年
生体内運命	30,000	6ヶ月~1年
亜急性毒性	7,500~9,500	5~6ヶ月
濃縮性	6,000~8,500	4~6ヶ月
変異原性	4,500	3ヶ月
急性毒性	900~4,500	3ヶ月
ミジンコ急性 毒性	750~2,700	3~4ヶ月
染色体異常	1,800~2,300	2~3ヶ月
生分解性	1,400~2,000	2~4ヶ月
魚類急性毒性	500~1,000	3ヶ月
エームス試験	500~800	45~60日

例えば, 発ガン性のデータベースについては, 表5に示すように, 発ガン性のランクが2段階から6段階までと異なり, また各段階の発ガン性の表現がまちまちなため, これらの違いを理解するのは困難であり, きわめて使いにくい.

第2の問題点はデータの量と質の点である. この種のデータベースの中で最大規模のものはRegistry of Toxic Effects of Chemical Substances(RTECS)のデータベースである. こ

表4 主なオンライン毒性データベース<sup>[5-7]</sup>

データベース名	物質数	構造データ	構造検索	URL
NTP	500種以上	あり	不可	<a href="http://ntp-server.niehs.nih.gov/">http://ntp-server.niehs.nih.gov/</a>
TOXNET	多数	あり	可	<a href="http://toxnet.nlm.nih.gov/">http://toxnet.nlm.nih.gov/</a>
Gene-Tox	3,000種以上	あり	可	<a href="http://toxnet.nlm.nih.gov/">http://toxnet.nlm.nih.gov/</a>
CCRIS	8,000種以上	あり	可	<a href="http://toxnet.nlm.nih.gov/">http://toxnet.nlm.nih.gov/</a>
IRIS	500種以上	あり	可	<a href="http://toxnet.nlm.nih.gov/">http://toxnet.nlm.nih.gov/</a>
RTECS	70,000種以上	あり	可	<a href="http://toxnet.nlm.nih.gov/">http://toxnet.nlm.nih.gov/</a>
CPDB	1,300種以上	なし	不可	<a href="http://potency.berkeley.edu/cpdb.html">http://potency.berkeley.edu/cpdb.html</a>
GAP	600種以上	あり	不可	<a href="http://www.epa.gov/gap-db/">http://www.epa.gov/gap-db/</a>
IARC	800種以上	なし	不可	<a href="http://monographs.iarc.fr/">http://monographs.iarc.fr/</a>
NCI	37,000種以上	あり	可	<a href="http://cactus.nci.nih.gov/">http://cactus.nci.nih.gov/</a>

表5 各種発ガン性データベースにおける発ガン性ランクと表現の違い

データベース名	発ガン性ランク	説明
IARC	1	Carcinogenic to humans
	2A	Probably carcinogenic to humans
	2B	Possibly carcinogenic to humans
	3	Not classifiable as a human carcinogen
	4	Probably not carcinogenic to humans
EU	1	Known as a human carcinogen
	2	Should be regarded as if a human carcinogen
	3	Possibly carcinogenic to humans
EPA	A	Carcinogenic to humans confirmedly
	B1	Probably carcinogenic to humans
	B2	Carcinogenic to animals, but unknown to humans
	C	Possibly carcinogenic to humans
	D	Not classifiable as a human carcinogen
	E	Not carcinogenic to humans confirmedly
NTP	A	Known as a human carcinogen
	B	Reasonably anticipated as a human carcinogen
ACGIH	A1	Carcinogenic to humans confirmedly
	A2	Carcinogenic to humans suspectedly
	A3	Carcinogenic to animals, but unknown to humans
	A4	Not classifiable as a human carcinogen
	A5	Not suspected as a human carcinogen
日本産業衛生学会	1	Carcinogenic to humans
	2A	Probably carcinogenic to humans
	2B	Possibly carcinogenic to humans

のデータベースは米国の National Institute of Occupational Safety and Health(NIOSH) が製作しており、現在、約 17 万件のデータが収録されている。しかし、この数はデータを抽出した論文数の累計であり、化学物質の種類は 17 万よりかなり少ない。さらに、このデータベースには様々な試験法で得られた毒性データが原論文からそのまま収録されており、統一的な試験法で評価されたデータが収集されていない。したがって、このデータベースのデータは信頼性の点で問題がある。

統一的な試験法で化学物質の毒性を評価して収録したデータベースとしては National Toxicology Program(NTP) のデータベースがある。NTP は米国で 1978 年に設立された毒性試験研究機関であり、統一的な動物実験の計画、実施、分析において世界的なリーダーの位置を占めている。したがって、このデータベースに収録されているデータはきわめて信頼性が高い。しかし、残念ながら、このデータベースに収録されている化学物質の数はそれほど多くない。例えば、発ガン性のデータが収録されている化学物質の総数は僅か 1,000 種類程度である。したがって、市場で流通している化学物質の中で発ガン性が判明しているものは 1% 程度であり、残りの 99% は発ガン性が不明なまま我々の周囲に存在している。

第 3 の問題点は化学物質の検索方法である。どのデータベースでも化学物質の名称、分子式、CAS 番号等の文字情報の検索が主体である。しかし、化学物質の命名法はきわめて複雑であり、データベースに登録されている名称でもれなく検索することは一般に困難である。CAS 番号は化学物質を簡単明確に定義できる番号であるが、任意の化学物質について CAS 番号を調べることは容易ではない。化学物質の構造を用いれば容易に化学物質を検索できるが、構造式検索機能があるデータベースでは化学構造を線形表記して入力する必要があり使いやすくない。

## 4 化学物質の毒性の予測

### 4.1 構造活性相関による毒性予測

化学物質の構造からその毒性を予測することができれば、動物試験にかける化学物質の数を絞り込むことによって毒性情報の効率的な取得が可能になり、化学物質の安全管理の観点から大きな意義がある。また、新規の化学物質を合成、製造する前にその毒性を事前に評価することも可能になり、新規化学物質の開発にとっても意義が高い。このような観点から、化学物質の毒性をその構造から予測する手法の研究開発が欧米では活発に行われている。

この基礎になっているのは、「類似の構造をもつ化学物質は類似の生理活性を生じる」という構造活性相関 (Structure-Activity Relationship, SAR) の概念である。化学物質の活性発現は化学物質が生物体に吸収され、作用を引き起こす部位に輸送され、標的部位と化学反応を引き起こすことにより生ずる。化学物質は細胞膜を透過して標的部位に達するが、細胞膜は脂質に富んでいるため、油に溶けやすい (親油性) 物質は比較的速く多量に細胞膜を通過する。一方、標的部位と化学物質との反応は薬物—受容体相互作用、いわゆる鍵と鍵穴の関係にたとえられるように化学物質の分子構造に依存する場合が多い。したがって、よく似た分子構造をもつ化学物質は同じ生体内での標的に対して同様な反応を引き起こすと考えられる。

しかし、化学物質が生体内で活性を発現する機構を解明して化学物質の構造から活性を予測することは現段階ではきわめて困難である。それは、生体内での化学物質の挙動には生体内の様々な受容体や酵素等が関与し、吸収、代謝、輸送、結合等、多くの現象が関係するが、生体側の構造を含め未だ解明されていない現象が多いからである。そこで、現状の活性予測手法としては、化学物質と受容体等との相互作用に基づいた分子レベルでの理論的アプローチでなく、既知データの活用により、化学物質の構造と活性との関係を統計

的に推測する SAR の手法である。SAR の考え方は 1868 年に Crum-Brown と Fraser が化学物質の生理活性を化学構造の関数と仮定したことに始まり、20 世紀の初めから医薬品や農薬等の薬理活性を同族の化学物質の構造に関連させて解明するために発展した。

活性を定量的に予測する定量的構造活性相関 (Quantitative Structure-Activity Relationship, QSAR) は 1964 年の Hansch と Fujita<sup>[8]</sup> 及び Free と Wilson<sup>[9]</sup> の論文に始まり、近年のコンピュータの発達により急速に進展した。QSAR による薬理活性分子の設計、いわゆるドラッグデザインについては多くの適用例がある<sup>[10-12]</sup>。最近では QSAR の手法が化学物質の薬理活性と正反対の毒性の解明・予測にも用いられるようになってきた。

化学物質の毒性は化学物質と受容体との化学反応の結果と考えられる。したがって、同じ毒性機構の化学物質について構造を反映す

る記述子を用いて毒性データとの回帰式を算出し、同じ反応過程をもつ毒性が未知の化学物質の毒性を予測することが可能であり、この方式が QSAR による毒性予測の基礎である。QSAR の方法を化学物質の毒性の予測に用いる場合は構造毒性相関 (Quantitative Structure-Toxicity Relationship, QSTR) と呼ぶこともある。化学物質の構造と毒性との相関を QSAR の手法で解析した研究は多数発表されている<sup>[13-15]</sup>。

その際に用いられる化学物質の構造を特徴づける記述子としては表 6 に示すものがあり、これらの記述子は化学構造からソフトウェアを用いて自動的に求めることができる。また、これらの記述子と毒性データとの相関を解析する統計手法としては表 7 に示すような様々な手法があり、これらの方法が単独あるいは組み合わせて用いられる。

表 6 QSAR でよく用いられる記述子

特徴・性質	記述子
構造的特徴 <sup>[16]</sup>	各種の原子・結合・部分構造の有無またはその個数、分子量等
物理化学的性質	沸点, 融点, 蒸気圧, 溶解度, 分配係数 (通常は $\log P$ がよく用いられる) 等
電子的性質 <sup>[17]</sup>	置換基定数 (Hammett の $\sigma$ 等), 酸解離定数, 酸化還元電位, 結合次数, 電子密度, 正味電荷, 電気陰性度, 有効分極率, 分子軌道 (特に HOMO や LUMO) のエネルギー, 局在化エネルギー, イオン化ポテンシャル等
立体的性質	置換基定数 (Taft の $E_s$ 等), van der Waals 半径, 分子容, 表面積, 分子屈折, カップ形状指数, STERIMOL 等
トポロジカル的性質 <sup>[18]</sup>	分子結合指数等

表 7 QSAR でよく用いられる統計解析手法

種類	線形解析	非線形解析
教師付学習法	重回帰分析 (MLR), 主成分回帰 (PCR), 判別分析, 部分最小二乗法 (PLS), 適応最小二乗法 (ALS), SIMCA, 分類・回帰木 (CART), 線形学習機械等	ロジスティック回帰, ニューラルネットワーク (ANN), 遺伝的アルゴリズム (GA) 等
教師なし学習法	主成分分析 (PCA), クラスタ分析, 因子分析, k-近隣 (KNN) 法等	コホーネンマップ, 自己組織化マップ, 非線形マッピング等

## 4.2 毒性予測システム

上記のような実用上の理由から QSAR の手法を用いて化学物質の構造から毒性を予測する種々のシステムが多数開発されている。これらのシステムで予測可能な毒性は急性毒性、慢性毒性、発ガン性、変異原性等、色々な種類があるが、その内でも特に発ガン性を中心に、代表的なシステムの予測成績を紹介する。その理由は、ガンは人類の死因の第1位であり、かつ発症まで時間がかかるため最も重大な毒性であること、及び表3に示すように、発ガン性の動物試験は最長の期間を要し、QSAR による予測が最も効果的であることからである。

### 4.2.1 TOPKAT

TOPKAT(Toxicity Prediction by Komputer Assisted Technology)<sup>[19-22]</sup>は米国 Health Designs Inc. の Enslein らによって開発された各種の毒性の予測システムである。重回帰分析(Hansch-Fujita 法)や判別分析法を用いて急性毒性、変異原性、発ガン性、催奇形性、生分解性等の予測を行うことができる。システムにはそれぞれの毒性データベースと構造記述子のセットが用意されている。

標準的な記述子としては部分構造記述子の他に、電子的記述子としては電子密度、正味電荷、グループ電気陰性度、有効分極率等、トポロジカルな記述子としては分子結合指数、形状記述子としてはカップ形状指数が用いられる。これらの記述子は TOPMOS II プログラムにより化合物の構造を SMILES(Simplified Molecular Input Linear Entry System) で入力することにより自動的に生成される。

TOPKAT による発ガン性の予測成績については、計 705 化合物を用いた再現性が 99.6%、cross-validation による予測率が 98.2% という高い数値が公表されている。しかし、米国の National Institute of Environmental Health Sciences(NIEHS) の発ガン性予測テストの結果では TOPKAT の予測的中率は 58% であった。予測成績がこのように低くなった原因は不明であるが、TOPKAT の予測モデルの学

習に用いられた化合物が NIEHS のテスト化合物とかなり異なっていたことが考えられる。

### 4.2.2 CASE

CASE(Computer Automated Structure Evaluation)<sup>[23-27]</sup>は米国 Case Western Reserve 大学の Klopman によって開発されたシステムである。CASE においては、化合物の構造の特徴を表す構造記述子の選定と抽出が完全にコンピュータにより自動的に行われ、毒性に関する知識が皆無の場合でも容易に予測モデルが生成される。

構造記述子は log P や分子結合指数等の若干の物性推算値の他は、多数の部分構造が機械的に抽出される。複雑な構造の化合物では部分構造の総数は莫大な数となるので、ベイズ統計理論を用いて毒性に寄与する構造記述子を選別する。最近では、CASE をさらに発展させて、代謝体の影響も考慮する Multi-CASE (Multiple Computer Automated Structure Evaluation)<sup>[28]</sup>も開発されている。

米国 NIEHS の発ガン性予測テストの結果では、CASE の予測的中率は 49% であった。その原因は余りにも機械的な記述子の生成と、莫大な数の記述子からの選別による chance correlation(偶然の相関)の混入と考えられる。

### 4.2.3 REPAD

REPAD<sup>[29]</sup>はイタリアの Instituto Speriore di Sanita の Benigni により開発された発ガン性の半定量的予測法である。この方法では化合物の親電子的反応性の簡易推算値 Ke と発ガン性警戒構造 SA の両者を用いて判定を行う。Ke の数値が 3.0 より大きい化合物は発ガン性が高く、小さい化合物は発ガン性が低いと判定する。また、SA は必ずしも全ての種類の化合物の発ガン性を説明できるものではないが、変異原性のデータとの相関から発ガン物質の指標として利用している。REPAD の予測成績については NIEHS の予測的中率は 57% であり、TOPKAT と同程度であった。

#### 4.2.4 COMPACT

COMPACT(Computer-Optimized Molecular Comparative Analysis of Chemical Toxicity)<sup>[30]</sup>は Lewis らにより開発された発ガン性予測法である。この方法では、化学物質による発ガン機構においてチトクロム P450-I が重要な役割を演じるとして、チトクロム P450-I の基質や Ah レセプター作用物質であるかどうかの識別により発ガン性を予測する。

この理論に基づき、立体的特性については平面性に関する形状記述子である Area/Depth<sup>2</sup>, 電子的特性については HOMO と LUMO のエネルギー差  $\Delta E$  を用い、それらの数値の比を COMPACT 比と名付け、その値が 0.20 より大きい場合は+, 0.15 より小さい場合は-, 0.15 と 0.20 の中間の時は不明と判定する。NIEHS による発ガン性予測テストでは COMPACT の予測的中率は 54% であり, CASE や REPAD と同程度であった。

#### 4.2.5 FALS

FALS(Fuzzy Adaptive Least Squares)<sup>[31-34]</sup>は森口らにより開発された毒性予測システムである。化学物質の毒性データは-, +, ++, +++ のような等級の形で扱うことが多いため、順序があるカテゴリーデータを解析するために適応最小二乗(ALS)法を考案した。さらに、大量の毒性データの予測成績の向上を目指してファジイ推論の技法を ALS に取り入れた Fuzzy ALS(FALS)を開発した。所属する化合物の分布を正規凸集合とみなし、釣り鐘型のメンバーシップ関数で表す。構造記述子としては、分子量、疎水性定数、各種の部分構造等の記述子を用いて線形識別関数(予測モデル)を算出する。

化学物質を構造的特徴により炭化水素化合物群、環状ヘテロ化合物群、ニトロ・ニトロソ・N-オキサイド化合物群、ハロゲン化合物群、アルコール・フェノール・エーテル化合物群、カルボニル化合物群、非芳香族アミン化合物群、含酸素イオウ化合物群の 8 つのクラスに分類して個別の相関式を作成することにより、予測の精度を高めている。

予測テストの結果については、NTP の発

ガン性データ(陽性 155, 陰性 91 物質)を用いた識別では, false positive が 6.1%, false negative が 4.9% であり, leave-one-out 予測では false positive が 13.4%, false negative が 8.9% という結果が公表されている。判別の誤りのうち, 陽性を陰性と誤る false negative が特に問題になると思われるが, false negative は予測でも 8.9% と良好な結果を示している。NIEHS の発ガン性データを用いた予測テストでは FALS の的中率は 68% であり, これまでのシステムよりは成績が高い。

### 4.3 予測システムの成績評価

#### 4.3.1 NIEHS の公開テスト

1990 年に NIEHS の Tennant らは NTP による発ガン性試験化合物 44 種について国際的に広く呼びかけて発ガン性予測の公開テストを提案した<sup>[35]</sup>。これが NIEHS の公開テストの第 1 ラウンドである。このテストには幾つかのグループが参加し, それぞれの方法で予測した結果が公表された。1993 年に NTP によるテストの結果が発表され, 国際ワークショップにおいて予測の成功や失敗の原因が議論された。さらに, 1994 年に NIEHS の Wachsmann により発ガン性予測の第 2 ラウンドが提案され, 1995 年末を期限として世界各国の毒性予測研究チームに参加が呼びかけられた。

最近発表された第 1 ラウンド及び第 2 ラウンドの集計結果<sup>[36]</sup>を表 8 に示す。このうち, Tennant らのシステムや RASH は変異原性等のデータから活性-活性相関(Activity-Activity Relationship)により発ガン性を予測するものであり, 構造データのみからの予測ではない。また, SA(発ガン性警戒構造)も用いているが, これにはあいまいな点があり, 最終的には判定者の知識, 経験に依存する。Weisburger, Lijinsky, DEREK も専門家の意見を取り込んで定性的な予測を行う。これに対して, Ke, Benigni, CASE, TOPKAT, COMPACT 及び FALS ではコンピュータにより発ガン性が自動的に予測される。また, Benigni の REPAD は半定量的方法であるが, これも



表 8 NIEHS による発ガン性予測テストにおける各システムの的中率 (%)<sup>[36]</sup>

	システム名 (開発者)	陽性	陰性	合計
第 1 ラウンド	Tennant et al. <sup>[35]</sup>	61.9	78.3	70.5
	Ke (Bakale & McCreary) <sup>[37]</sup>	64.3	64.7	64.5
	Weisburger <sup>[38]</sup>	95.0	26.3	61.5
	RASH (Jones & Easterly) <sup>[39]</sup>	75.0	47.1	60.6
	DEREK (Sanderson & Earnshaw) <sup>[40]</sup>	80.0	38.1	58.5
	TOPKAT (Enslein et al.) <sup>[41]</sup>	75.0	43.8	57.1
	DEREK hybrid (Sanderson & Earnshaw) <sup>[40]</sup>	50.0	61.9	56.1
	Lijinsky <sup>[38]</sup>	61.9	47.8	54.5
	COMPACT (Lewis et al.) <sup>[30]</sup>	33.3	68.2	52.5
	CASE (Rosenkranz & Klopman) <sup>[25]</sup>	47.1	50.0	48.7
	Benigni <sup>[42]</sup>	47.1	31.8	38.5
第 2 ラウンド	SHE (Kerckaert et al.) <sup>[43]</sup>	57.1	70.6	66.7
	R2 (Lee et al.) <sup>[44]</sup>	40.0	73.3	65.0
	Benigni et al. <sup>[45]</sup>	71.4	61.1	64.0
	R1 (Lee et al.) <sup>[44]</sup>	40.0	71.4	63.2
	Bootman <sup>[46]</sup>	28.6	68.4	57.7
	Tennant et al. <sup>[47]</sup>	57.1	57.9	57.7
	OncoLogic (Woo et al.) <sup>[48]</sup>	42.9	57.9	53.8
	Ashby <sup>[49]</sup>	42.9	57.9	53.8
	Huff et al. <sup>[50]</sup>	28.6	52.6	46.2
	FALS (Moriguchi et al.) <sup>[51]</sup>	28.6	50.0	43.5
	RASH (Jones & Easterly) <sup>[52]</sup>	42.9	38.9	40.0
	KeE (Benigni et al.) <sup>[42]</sup>	100.0	12.5	39.1
	DEREK (Marchant) <sup>[53]</sup>	71.4	26.3	38.5
	COMPACT (Lewis et al.) <sup>[54]</sup>	71.4	26.3	38.5
	BeO (Benigni et al.) <sup>[42]</sup>	57.1	25.0	34.8
	Purdy <sup>[55]</sup>	57.1	25.0	34.8
	Progol et al. <sup>[56]</sup>	50.0	25.0	31.8
	CASE (Zhang et al.) <sup>[57]</sup>	14.3	20.0	18.2

短時間に予測が可能である。

この NIEHS の公開テストで用いられた物質数が第 1 ラウンドで 44, 第 2 ラウンドで 30 であり, 各種手法の成績評価には十分とはいえない。しかし, このテストからは, 他の実測データや専門家の意見を用いる半経験的システムの予測率が高く, 完全に自動化されたシステムはまだ予測率がきわめて低いと結論できる。

#### 4.3.2 Predictive Toxicology Challenge

上記の NIEHS の公開テストでは用いられた物質数が少数であったため, ドイツ・フライブルグ大学の C. Helma らは多数の物質を用いる公開テストを世界各国の研究者に呼びかけて Predictive Toxicology Challenge (PTC) 2000-2001<sup>[58]</sup>を開催した。このテストの特徴は, 主催者がモデル学習用及びテスト用の化学物質の構造, 記述子, 発ガン性データをイ

ンターネット上に公開し、参加者はその共通のデータセットを用いて自己のモデルの予測結果を提出し、主催者がその結果を集計するという公開コンテストであり、このコンテストは以下の3段階で行われた。

### ① Data Engineering

発ガン性試験データ (male, female の rat, mouse に対するもの) のある化学物質について構造のデータを主催者が公開し、これらに対する記述子を一般から公募した。その結果、7種類の記述子群、総数 7000 個以上の記述子データが公開された。物質数は、学習用が

NTP データベースから抽出した 471 種、テスト用が FDA (Food and Drug Administration) データベースから抽出した 185 種である。

### ② Model Construction

学習用化合物の記述子と発ガン性データを用いて参加者が各自のモデルを作成した。14のグループが 26~31 のモデルを提出した。

### ③ Model Evaluation

発ガン性データを伏せたテスト用化合物について参加者が各自のモデルで発ガン性を予測した結果を提出し、主催者がその結果を集約し ROC 分析 (図 1) によって評価した。し

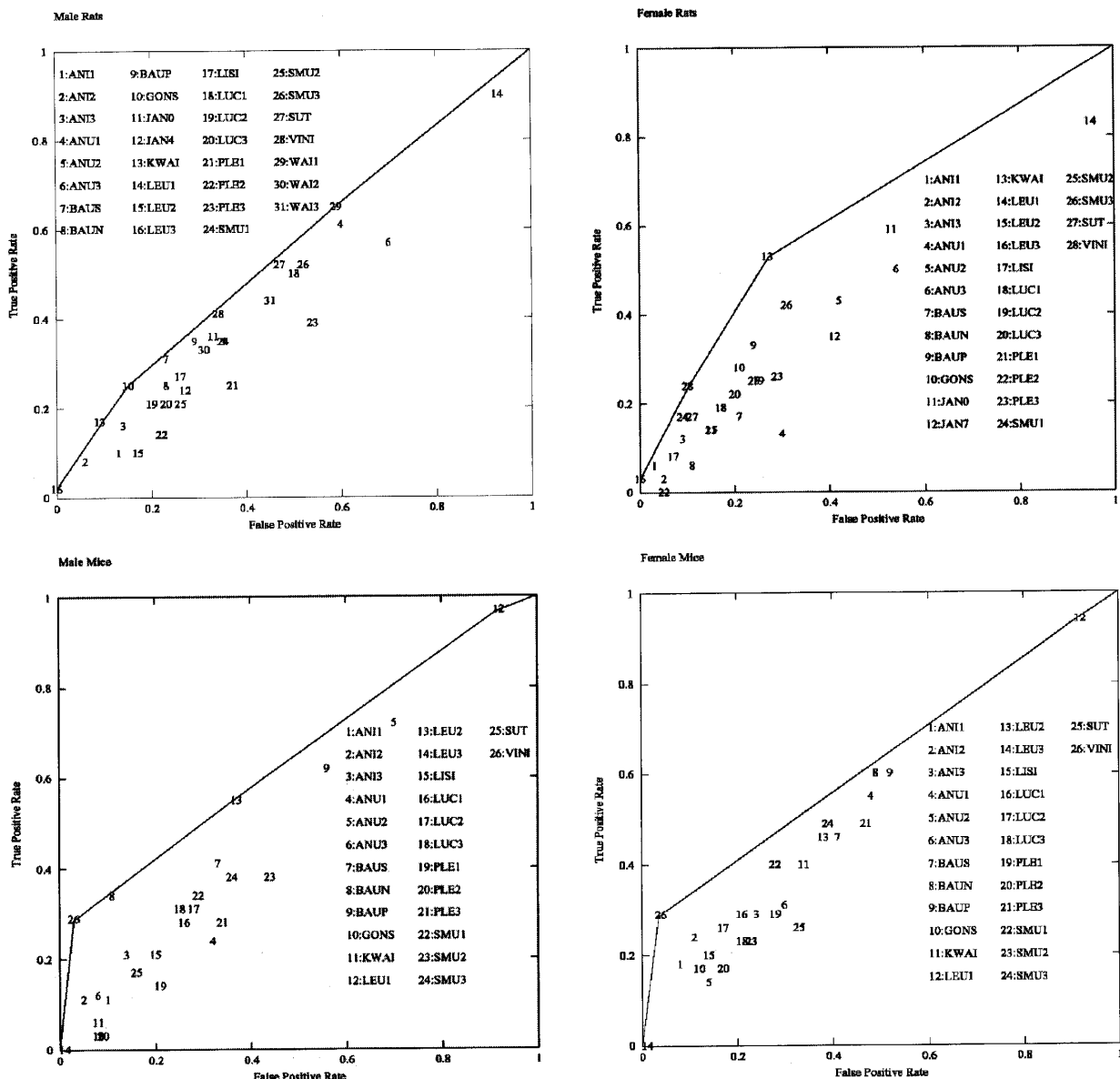


図 1 PTC コンテスト参加者の予測結果の ROC 分析図

かし、参加者の予測モデルの大半は ROC 分析図の左下に位置しており、発ガン陰性を陽性と誤判別する false positive は少ないが、陽性を陰性と誤判別する false negative が多いことが分かる。この原因として、FDA のテスト化合物の構造が NTP の学習用化合物とかなり異なっていたこと、及び発ガン機構の複雑さと構造情報のみを使うことが指摘されているが、これらの予測手法において false negative が多い点は実用的には問題である。

## 5 政府機関等における QSAR の利用<sup>[59]</sup>

QSAR による化学物質の毒性予測は製薬・化学産業ばかりでなく、各国政府機関における化学物質の管理・規制にも用いられつつあるので、その現状を概観する。

### 5.1 米国・カナダ

米国では Interagency Testing Committee(ITC) という機関が TSCA によって規制されるべき化学物質を決定しており、その際に QSAR を活用している。その他、EPA, NTP, Agency for Toxic Substances and Disease Registry (ATSDR), FDA, NIOSH 等の機関でも QSAR を利用している。

EPA は世界中の政府機関の中で QSAR を最も広く活用している機関であり、TSCA のもとで新規化学物質のリスクと暴露を実証することが義務付けられている。米国で製造、加工していると報告された化学物質は TSCA インベントリーに収録され、TSCA ではインベントリーに記載されていない物質を新規化学物質として商業目的でこれを製造、輸入する事業者に対して Premanufacture Notification(PMN) を提出することを求めている。

EPA の中で PMN を行っているのが Office of Pollution Prevention and Toxics (OPPT) であり、OPPT は化学物質の発ガン性の評価に OncoLogic を用いている。OncoLogic は EPA の研究者が開発したシステムで、発ガンの機構

に関する広汎な知識を活用した一連の QSAR を用いて、一般的な有機化合物以外の繊維、高分子、金属等についても発ガン性の評価が可能である。

EPA では環境毒性については EPISUITE というソフトウェアパッケージを開発し、公開している<sup>[60]</sup>。この中には、EPIWIN(SMILES 入力を変換するインタフェース)、AOPWIN(大気反応速度の評価)、BCFWIN(生物濃縮係数の評価)、BIOWIN(生分解性の評価)、ECOSAR(水棲毒性の評価) 等、多数のソフトウェアが含まれている。

NTP は NIEHS, NIOSH, FDA 等の活動を包含する政府間プログラムである。これらの機関から提出された化学物質について QSAR を用いて毒性を評価している。また、市販の毒性予測のソフトウェアについて成績評価を行っている。ATSDR においても QSAR を活用して化学物質の毒性の評価を行っている。FDA では Center for Drug Evaluation and Research(CDER) において薬品の発ガン性の評価の目的に TOPKAT や Multi-CASE をテストしている。

カナダでは 1999 年に施行された法律で化学物質の規制に QSAR 活用の検討を開始し、化学物質の蓄積性、生物濃縮性、毒性の評価のために、ECOSAR, TOPKAT 等の市販ソフトウェアの成績評価を行っている。

### 5.2 EU

EC においては、化学物質のリスクアセスメントを推進するために 1996 年に新規及び既存化学物質のリスク評価に関するテクニカルガイダンス文書(TGD)を公表した。その第 4 章では、QSAR の利用分野、利用方法等について規定している。そこでは、化学物質の環境毒性や環境寿命、例えば魚毒性等については QSAR の利用に関して詳しく記されているが、生体毒性については記されていない。

EU における現行の化学物質の規制システムでは、新規及び既存の化学物質は同一の試験を受けることが要求されていない。そこ

で、リスクアセスメントに関して新たなシステム REACH(Registration, Evaluation and Authorization of CHemicals) が計画されている。その際には QSAR を活用した毒性予測手法が重要な役割を果たすと期待されている。

EU の内、デンマークでは、毒性物質の判別に QSAR を活用しており、TOPKAT や Multi-CASE 等の商用ソフトウェアの予測精度の検証実験を行っている。ドイツにおいても、新規化学物質の毒性評価のために QSAR の利用が検討されており、急性毒性の QSAR モデルを開発している。

### 5.3 日本及びその他

わが国においては、製品評価技術基盤機構(NITE)、化学物質評価研究機構(CERI)、日本化学物質安全・情報センター(JETOC)等の機関で、QSAR による毒性予測及びそのための毒性データベースの開発を行っている。

OECD においても化学物質の毒性評価のために QSAR モデルやデータベースの開発を進めている。

## 6 発ガン性予測に関する我々の研究

### 6.1 ニューラルネットワークによる発ガン性予測

一般に QSAR 手法を用いる毒性予測には、予測精度、汎用性、処理時間という相反する 3 種類の問題があるが、これらをすべて解決するのは難問である。これらの問題に関係する選択肢は、構造-毒性の相関の解析にどのような手法を用いるかという点と、相関解析に有効な構造記述子をいかに選定するかという点である。上記のような既存手法の予測成績の低さの原因の 1 つは、構造-毒性の相関解析に重回帰分析等の線形解析手法を用いているためと考えられる。しかし、化学物質の構造と毒性との相関は単純な線形関係とは考えにくく、むしろ非線形関係と考えたほうが妥当である。したがって、非線形解析手法を

用いることにより、既存のものよりも精度が高い予測手法を開発できる可能性がある。

そのような非線形解析手法の一つとしてニューラルネットワークがあり、薬理活性の解析や予測等の分野では多くの適用例<sup>[61-63]</sup>があるが、毒性予測に応用した研究は少ない。そこで、我々は化合物の構造情報のみから毒性を高い精度で予測する手法を開発するために、ニューラルネットワークを用いて構造と毒性、特に発ガン性の相関を解析する研究を行ってきた。発ガン性を取り上げたのは、前記の 2 つの理由の他に、発ガン性を高精度で予測する手法が未開発なためである。

ニューラルネットワークを用いて化学物質の発ガン性を予測する手法の概念図を図 2 に示す。入力層、中間層、出力層各 1 層から成る 3 層構造のニューラルネットワークを使用し、入力層に化合物の構造を特徴づける記述子を入力し、出力層には各化合物の発ガン性の有無を教師データとして入力する。学習用の化合物群について、誤差逆伝播法のアルゴリズムを用いて、発ガン性データの出力値と教師データとの誤差を最小にするよう、ネットワークの重みと閾値の修正を行う。この学習が完了すると、テスト用の化合物群の記述子データを入力して発ガン性データの出力値を求め、実測値との比較からの中率を計算する。

ニューラルネットワークを用いて発ガン性予測を検討した我々の研究の結果を以下に紹介する。

### 6.2 結合記述子を用いる解析<sup>[64]</sup>

我々はまず、記述子として最も単純な結合の数だけを用い、ニューラルネットワークの有効性を検証した。NTP のデータベースから、原子 C, H, Cl のみから成る有機塩素化合物 41 種を抽出し解析に用いた。記述子としては各化合物における C-C, C=C, C=C(芳香族), C-H, C-Cl 結合の数, log P 及び分子量を用い、ニューラルネットワークに入力して学習した。Leave-one-out test を行った結果、83%の予測的中率を得た。

この方法では記述子が化学構造から簡単に求まるので、迅速な予測が可能である反面、このような単純な記述子ではこれ以上の予測的中率の向上は困難である。例えば、図3に示す構造異性体の対は発ガン性の有無が異なっているにもかかわらず、結合数が同じであるため、leave-one-out test を行うと実測値を説明できない。また、ベンゾピレンをはじめとする縮合多環芳香族炭化水素化合物 (PAH) ではベンゼン環の結合様式の違いによる多数の構造異性体があるが、このような構造の違いによる発ガン性の有無を結合記述子のみで説明することは不可能である。

### 6.3 量子化学的記述子を用いる解析<sup>[65]</sup>

そこで次に、このような構造異性体の構造

の違いを説明するために量子化学的記述子を用い、同じ41種類の有機塩素化合物の発ガン性データをニューラルネットワークで解析した。用いた記述子は分子軌道計算等から求まる Gibbs 標準生成自由エネルギー、イオン化ポテンシャル、LUMO エネルギー、HOMO-LUMO のエネルギー差、Connolly 体積、分子量、log P の7種類である。Leave-one-out test を行った結果、的中率は93%となり、量子化学的記述子を用いれば上記の構造異性体の発ガン性データの違いが予測できることが分かった。

しかし、ここで用いた量子化学的記述子を求めるためにはMOPAC等による量子化学計算が必要であり、結合記述子を用いる場合のように迅速な予測は困難である。特に内部回転異性体が存在する化合物では、分子軌道エネルギーや Connolly 体積等の記述子の値が回

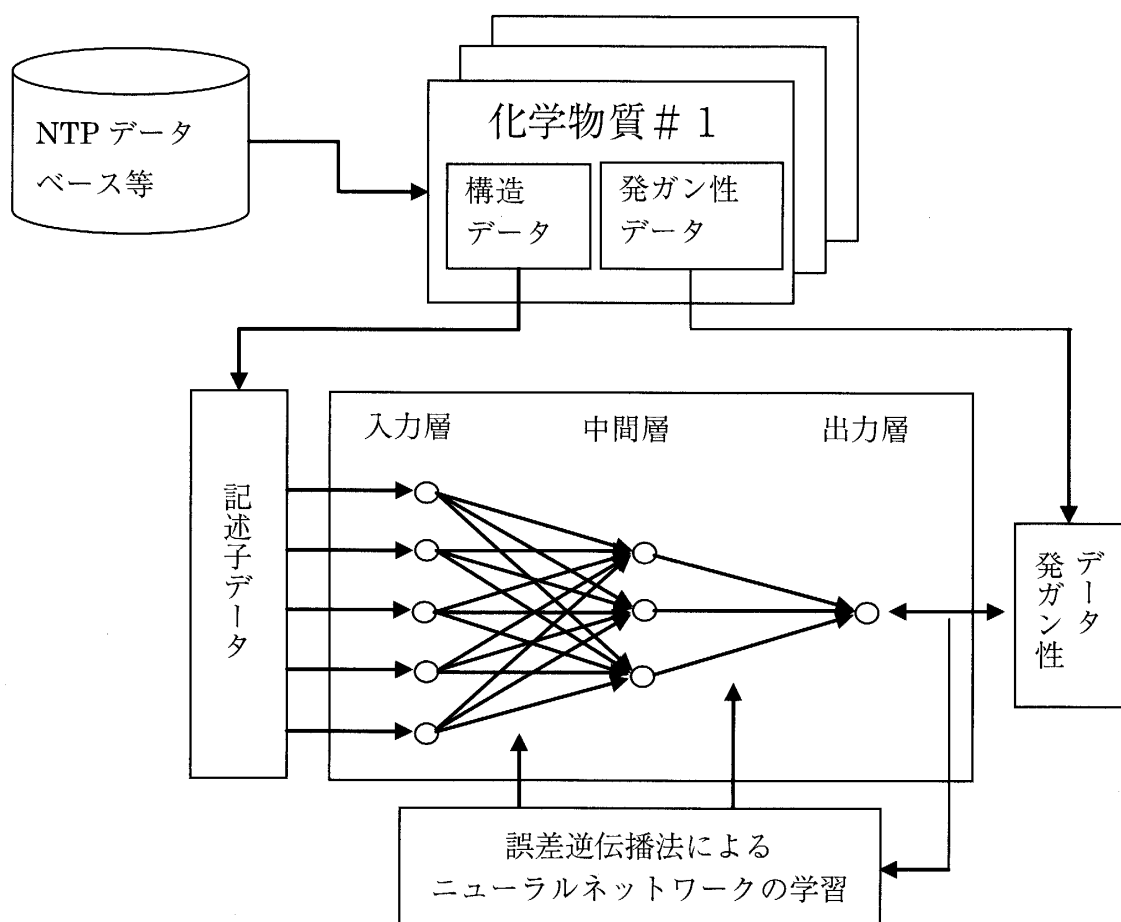


図2 ニューラルネットワークを用いる化学物質の発ガン性予測の概念図

転異性体間で異なるので、最安定の回転異性体を探索する操作が必要になる。しかし、内部回転異性体が多数存在する化合物について最安定の異性体を探索するにはかなり時間がかかる場合があり、この点が課題である。

また、ここで扱ったのはわずか41種類の有機塩素化合物であり、多種多様な化合物についてこのような高い中率で予測できるかどうかは疑問が残る。

#### 6.4 多種多様な化合物についての解析<sup>[66]</sup>

そこで、ニューラルネットワークを用いる方法が広範囲の化合物についてどの程度高い中率で予測できるかを検討するために、PTC 2000-2001のコンテスト<sup>[58]</sup>のデータをニューラルネットワークで解析し、コンテストの参加者の結果と比較した。このコンテストでは、多くの参加者が多種多様な化合物の発ガン性を学習・予測した結果が公表されているので、このコンテストの発ガン性データは、広範囲の化合物に対する我々の方法の適用性を検証し、他の方法の成績と比較する点できわめて有効なデータである。

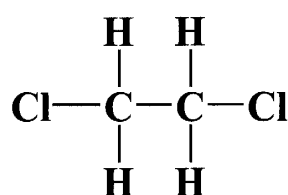
発ガン性及び記述子のデータは、このコン

テストにおいて公開されたデータの内の、参加者の予測的中率が最も低かった male rat のデータを用いた。物質数は、モデル学習用が陽性107、陰性180、テスト用が陽性45、陰性122、合計454である。コンテストでこれらの化合物について公開された記述子の中から量子化学的記述子である tReymers と Helma の記述子(表9)を用いた。

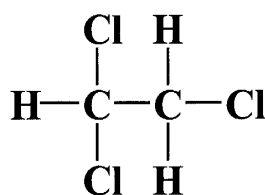
学習化合物についてこれらの記述子のデータを用いて学習したモデルに対して、テスト用化合物のデータを用いて求めた我々の手法の予測的中率を PTC の参加者の結果<sup>[58]</sup>と比較して表10に示す。参加者の解析法は、決定木、重回帰分析、カスケード法、相関ルール、ラティス探索、グラフ縮約、判別分析などであり、ほとんどが線形の解析手法である。

我々の方法は参加者のどれよりも高い予測的中率が得られたが、その理由としては、記述子の適切な選択とニューラルネットワークという非線形解析手法の適用の2点が考えられる。参加者の予測方法の大半では、非常に多数の記述子を単純な線形手法で解析しており、このような機械的な解析手法では多種多様な化合物の発ガン性を高い精度で予測するのは容易でないと考えられる。

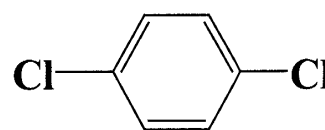
しかし、我々の手法も PTC 参加者の手法



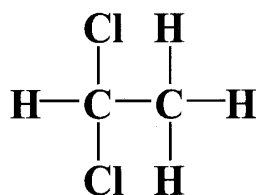
1,2-ジクロロエタン



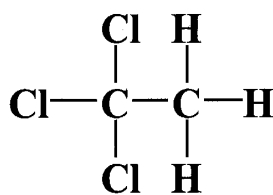
1,1,2-トリクロロエタン



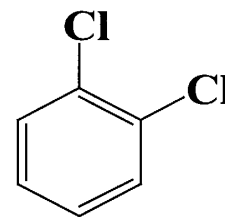
1,4-ジクロロベンゼン



1,1-ジクロロエタン



1,1,1-トリクロロエタン



1,2-ジクロロベンゼン

図3 結合記述子を用いた発ガン性予測において実測値(上段:陽性, 下段:陰性)と予測値(上段:陰性, 下段:陽性)が一致しない構造異性体の対

表9 PTCデータの解析に用いた量子化学的記述子

記述子群	記述子
tReymers	SLOGP, SMR, ROTBOND, FLEX, VOLUME, SURF_A, TPSA, HBD, HBA, LUMO, HOMO, DIPOLE, LOGD_2, LOGD_7, LOGD_10
Helma	log P, dipole, electronic energy, electronegativity, heat of formation, homo, homo-lumo, hybrid dipole, ionization potential, lumo, largest interatomic distance, molecular weight, point-chg. dipole, total energy, /POLARIZA, /DELTAHF, /STABIL, /STRAIN, total_access, non-polar_access, polar_access, perc_nonpolar

表10 PTCデータを用いた各発ガン性予測法の的中率(%)

手法名	陽性	陰性	合計
本研究	4.4	99.2	73.7
KWAI	16.9	90.9	71.1
ANI2	7.7	94.1	71.0
GONS	25.1	85.0	68.9
BAUS	31.2	77.3	64.9
LUC2	21.0	80.1	64.3
BAUN	25.1	77.1	63.1
BAUP	35.2	71.2	61.6
LISI	27.1	74.0	61.5
PLE2	14.1	78.1	60.9
JAN4	24.0	73.2	60.1
SMU2	21.0	74.0	59.8
WAI2	33.2	69.4	59.7
VINI	41.1	66.3	59.6
ANU2	35.0	65.3	57.2
SUT	52.3	53.1	52.9

も発ガン陰性化合物の的中率はいずれも好成績であるのに対し、陽性化合物の的中率はきわめて低い。この原因としては、学習用化合物の数が陽性 107 種類、陰性 180 種類と不均衡なため、モデルの学習が陰性化合物に偏って行われたためと考えられる。また、NTP の学習用化合物群が構造的に広範囲なバルクの化学物質を含み、かつ比較的小さな分子が多いのに対し、FDA のテスト化合物群が医薬品に偏り、化学構造の面でかなり異なっている点も指摘されている。

### 6.5 多段階の発ガン性データを用いる解析<sup>[67]</sup>

我々の手法が既存のものより優れているといっても、その的中率は 74% 程度であり、既存化学物質の安全管理、新規化学物質の毒性予測の実用目的からは決して十分な精度といえない。また、我々の方法でも発ガン陰性化合物の的中率は好成績であるのに対し、陽性化合物の的中率はきわめて低い。QSAR による毒性予測を動物試験のスクリーニングと位置付けるならば、発ガン陽性化合物を陰性と判定して見逃すことは好ましくないため、この結果は実用的にはきわめて問題である。

化学物質の毒性の中には QSAR 手法を用いた予測が良好な成績を示し、水棲毒性のように欧米諸国で化学物質管理に実用的に使用されている毒性もある。しかし、それらの毒性データは PTC コンテストで用いられた発ガン性データのように、毒性あり・なしの 2 値ではなく、毒性の強度に応じた連続変数データである。

したがって、発ガン性データについても発ガン性あり・なしの 2 値データではなく、発ガン性の強さに応じた連続値データを用いれば、既存手法より高い精度の発ガン性予測手法を開発できる可能性がある。表 5 に示したように、動物試験による発ガン性は IARC, EPA, NTP 等様々な機関で収集評価しているが、評価の信頼度は機関によって異なり、また同一の化学物質でも機関で異なる発ガン性のラン

クに評価されているものもある。

そこで、信頼度の高い機関による評価を中心に、各機関のランクを総合的に評価して、計9段階に格付けした発ガン性ランクデータベース (CRDB) を開発した。すなわち、表 11 に示すように、発ガン陽性は最も信頼度の高い A ランクから最も低い D ランクまで 4 段階、陰性も最も信頼度の高い I ランクから最も低い F ランクまで 4 段階、及び陽性・陰性の判定不明の E ランクの計 9 段階である。

量子化学的記述子を用いてこの発ガン性ランクデータをニューラルネットワークで解析した結果、平均誤差 0.007, 2 乗相関係数 0.999 という高い精度で広範囲の化学物質の発ガン性を予測できることが分かった。このことは、既存の発ガン性あり・なしの 2 値のデータを収録したデータベースではなく、ここで構築した多段階の発ガン性データベースが高精度の発ガン性予測手法の開発に有効であることを示している。

我々は現在、1,000 種類以上の広範囲の化学物質について発ガン性ランクデータベースを開発中であり、このデータを用いて多種多様な化学物質の発ガン性を高精度で予測する手法を開発する計画である。さらに、発ガン性を含む種々の毒性について、毒性の実測データの評価、有効な記述子データの選定、及び

相関解析手法の最適化を行うことにより、実用的な毒性予測システムを構築し、公開したいと考えている。

## 7 今後の課題

5 節に記したように QSAR による毒性予測システムは化学物質を取り扱っている産業界だけでなく、行政面でも利用されている。しかし、これまでの予測システムの成績は実用的には不満足である。その解決には以下の対策が必要である。

### 7.1 信頼性の高い毒性データの蓄積

QSAR による発ガン性予測に関してはこれまで数多くの研究が報告されている<sup>[68-78]</sup>が、それらの多くはハロゲン化炭化水素、フェノール類、ニトロソ化合物等、比較的少数の同族体、いわゆる congener を対象とするものである。このような場合は発ガンの機構がかなり類似している可能性があり、そのため発ガン性データと相関の高い記述子を見出すことが容易である。また、水棲毒性等のように、non-congener においても相関の高い log P 等の記述子を用いた単純な手法で予測できる場合もある。

表 11 9 段階の発ガン性ランクデータを用いた予測値と設定値

ランク	発ガン性	化学物質数	発ガン設定値	発ガン予測値
A	++++	5	0.900	0.887 ~ 0.909
B	+++	70	0.800	0.719 ~ 0.818
C	++	39	0.700	0.693 ~ 0.749
D	+	55	0.600	0.584 ~ 0.625
E	+/-	29	0.500	0.472 ~ 0.517
F	-	21	0.400	0.386 ~ 0.423
G	--	34	0.300	0.293 ~ 0.320
H	---	44	0.200	0.190 ~ 0.214
I	----	124	0.100	0.031 ~ 0.200
合計		421		



それに対して、不特定の化合物群、いわゆる non-congener に対する発ガン性予測の問題では、化合物の構造は多様であり、発ガンの機構も複雑であり、それらの発ガン性を統一的に説明できる手法を見出すことはきわめて困難な問題である。一般の化学物質の毒性を高精度で予測するためには、多種多様な構造の non-congener についての毒性予測モデルの構築及びその信頼性の評価が必要であるが、上記のようにこれに成功している予測システムはまだない。

この原因として non-congeneric QSAR に共通して考えられる問題点としては

- (1) 学習用化合物の中にテスト用化合物に含まれる発ガン性(または反発ガン性)構造についての情報が不足している、
- (2) chance correlation(偶然の相関)の混入により学習用化合物セット中に含まれる情報が相関手法中に正しく抽出されなかった、などが考えられる。

これらの解決策としては(1)については学習用化合物の種類と数を増やすこと、(2)については発ガンの機構に直接または間接に関係があると思われる構造を中心に候補記述子を設定し、また相関手法に取り入れる記述子の選別の方法を再考すること等が挙げられる。ただし、現実にはデータベース、コンピュータのハードウェア、ソフトウェア及び発ガン機構に関する研究の進展等克服すべき課題は多い。

発ガン機構の完全な解明にはまだかなりの年月を要すると思われるが、これについては質量ともに十分な学習化合物データと、コンピュータの計算速度の飛躍的増大によってカバーできると思われる。多種多様な化合物群の発ガン性を統一的に予測できるような手法を開発する問題は多くの研究者が取り組んでいる最先端の研究課題である。

以上のことは発ガン性のみならず、様々な毒性の予測に共通する問題と思われる。最近、環境汚染物質をクラスⅠ(化学的に不活性な物質)、クラスⅡ(化学的にやや不活性な物質)、クラスⅢ(化学的に活性な物質)及びクラスⅣ(化

学的に特異的に活性な物質)に分類し、それぞれについて QSAR を検討することが提案されている。これは1つの有効な手段と思われるが、具体的な分類方法が問題であり、また毒性の種類によりメカニズムもかなり異なるので、さらに検討が必要であろう。もし可能であれば、毒性の種類ごとにメカニズムに立脚した構造記述子を用いたアプローチが望ましいことはいうまでもない。

## 7.2 新規解析技術の開発

化学物質の構造と毒性データとの相関を解析する技術としては、我々が用いているニューラルネットワークは強力な非線形解析手法であり、それまで解析困難であった多くの問題に応用されてきた。しかし、誤差逆伝播法による階層型ニューラルネットワークの学習においては、過剰学習の問題、収束の遅さ、局所最適解への収束、モデル選択の任意性等の問題が存在することが指摘されている。

一方、近年登場したサポートベクターマシン(SVM)はデータの集合を多次元の非線形空間に写像した後、線形モデルで識別を行う手法である<sup>[79]</sup>。この方法は、局所解の問題がないこと、機械学習方法として興味ある理論的なサポートがあること、実際問題への適用の際に比較的簡単に実施できること等の利点があり、階層型ニューラルネットワークの上記の問題を解決できる学習機械と期待されている。SVMは化学・生物学分野でも応用され始めているが、これまでは医療画像診断や遺伝情報解析等が主であり、毒性予測分野への応用は未だごく少数である<sup>[80]</sup>。

さらに、QSARによる毒性予測には、表7に示すように遺伝的アルゴリズム(GA)、分類木・回帰木(CART)等、多くの解析方法があり、新たなアルゴリズムが提案されている方法もある。したがって、これらの各種の解析手法を統合化した毒性予測システムの開発も今後の研究課題である。

### 7.3 知識工学的アプローチの展開

QSARによる毒性予測の方法論は、コンピュータを活用して膨大なデータを統計解析する機械的ルール導出アプローチと、専門家の知識を活用したエキスパートシステムのアプローチに二分することができる。前記のように、PTC コンテストの参加者の大半は、数千という膨大な記述子を用い、多種多様な化合物の発ガン性データとの相関分析を行っているが、このような機械的な解析法では発ガン性の予測に有効な記述子の効果が希薄になり、高精度のモデルを作成することが難しい。

そこで注目されるのは専門家の知識を活用したエキスパートシステムのアプローチであり、既存手法の中ではDEREKやOncoLogic等のシステムがこの範疇に入る。前記のように、DEREKもOncoLogicもまだ予測的中率はあまり高くないが、これからの発展性の点では機械的なルール導出アプローチよりも期待される<sup>[81][82]</sup>。

### 7.4 生体内における毒性発現機構の解明

生体内における化学物質の毒性の発現機構は単純ではなく、化学物質の種類によって様々な機構により毒性が発現すると考えられる。したがって、単一の予測モデルですべての化学物質の毒性を予測することには限界があり、予測成績の高いモデルを得るためには化学物質を毒性発現機構の種類によって分類し、それぞれについて予測モデルを求めることが必要である。そのためには生体内における化学物質の毒性の発現機構解明の基礎研究が不可欠である<sup>[83][84]</sup>。

## 8 おわりに

はじめに記したように、化学物質の毒性の予測技術の確立は、国内問題ではなく国際的な規模で重要な課題である。特にQSAR手法に基づく定量的予測は経済的、社会的、動物

愛護的等、様々な観点から急務の開発課題であり、その信頼性向上のためには国際協力体制による研究開発の飛躍的進展が切望される。

## 謝辞

本論文は、科学技術振興調整費知的基盤推進整備事業「化学物質安全特性基盤の確立に関する研究：構造活性相関による生体有害性予測手法開発」(平成9～13年度)、及び科学研究費補助金基盤研究(A)(1)14209022「化学物質の健康影響評価における非線形構造活性相関手法の研究」(平成14～17年度)の支援の下に行われた我々の研究成果を中心に、化学物質の毒性予測の現状をまとめたものである。終始ご指導いただいた北里大学名誉教授の森口郁生先生はじめ、これらの研究にご協力いただいた関係各位に感謝する。

## 参考文献

- [1] R. Carson: "Silent Spring", Mariner Books, 2002; 青樹築一(訳):「沈黙の春」, 新潮社, 1974.
- [2] C. M. Auer; M. Zeeman; J. V. Nabholz; R. G. Clements: SAR and QSAR in Environmental Research, Vol.2, pp.29-38, 1994.
- [3] <http://www.cas.org/EO/regsys.html>
- [4] <http://www.meti.go.jp/report/downloadfiles/g20704c03j.pdf> 及び <http://www.env.go.jp/council/05hoken/y053-05/mat04-4.pdf> より編集.
- [5] A. M. Richard; C. R. Williams: "Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens", edited by R. Benigni, pp.145-173, CRC Press, 2003.
- [6] K. L. E. Kaiser: "Predicting Chemical Toxicity and Fates", edited by M. T. D. Cronin and D. J. Livingstone, pp.17-29, CRC Press, 2004.

- [7] M. T. D. Cronin: "Predictive Toxicology", edited by C. Helma, pp.93-133, CRC Press, 2005.
- [8] C. Hansch; T. Fujita: *Journal of American Chemical Society*, Vol.86, pp.1616-1626, 1964.
- [9] S. M. Free Jr.; J. W. Wilson: *Journal of Medicinal Chemistry*, Vol.7, pp.395-399, 1964.
- [10] J. L. Fauchere: "Qsar: Quantitative Structure-Activity Relationships in Drug Design", John Wiley & Sons Inc, 1989.
- [11] T. Fujita: "Qsar and Drug Design: New Developments and Applications", Elsevier Science, 1995.
- [12] H. Kubinyi: "3D Qsar in Drug Design: Recent Advances", Kluwer Academic Pub., 1998.
- [13] 松尾昌季: 「QSAR(定量的構造活性相関)手法を用いた化学物質の手計算による毒性予測」, Life-Science Information Center, 1999.
- [14] L. Passerini: "Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens", edited by R. Benigni, pp.81-123, CRC Press, 2003.
- [15] R. Benigni: "Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens", edited by R. Benigni, pp.173-192, CRC Press, 2003.
- [16] T. I. Netzeva: "Predicting Chemical Toxicity and Fates", edited by M. T. D. Cronin and D. J. Livingstone, pp.61-83, CRC Press, 2004.
- [17] G. Schüürmann: "Predicting Chemical Toxicity and Fates", edited by M. T. D. Cronin and D. J. Livingstone, pp.85-149, CRC Press, 2004.
- [18] J. Devillers; A. T. Balaban: "Topological Indices and Related Descriptors in QSAR and QSPR", Gordon & Breach Science Publishers, 1999.
- [19] K. Enslein; P. N. Craig: *Journal of Environmental Pathology and Toxicology*, Vol.2, pp.115-121, 1978.
- [20] K. Enslein; P. N. Craig: *Journal of Toxicology and Environmental Health*, Vol.10, pp.521-530, 1982.
- [21] K. Enslein; T. R. Lander; M. E. Tomb; W. G. Landis: *Teratogenesis, Carcinogenesis and Mutagenesis*, Vol.3, pp.503-513, 1983.
- [22] K. Enslein; V. K. Gombar; B. W. Blake: *Mutation Research*, Vol.305, pp.47-61, 1994.
- [23] G. Klopman: *Journal of American Chemical Society*, Vol.106, pp.7315-7321, 1984.
- [24] G. Klopman; A. N. Kalos; H. S. Rosenkrantz: *Journal of Biochemical and Molecular Toxicology*, Vol.1, pp.61-81, 1987.
- [25] H. S. Rosenkrantz; G. Klopman: *Mutagenesis*, Vol.5, pp.333-362, pp.425-432, 1990.
- [26] G. Klopman: *Quantitative Structure-Activity Relationship*, Vol.11, pp.176-184, 1992.
- [27] G. Klopman; H. S. Rosenkrantz: *Mutation Research*, Vol.305, pp.33-46, 1994.
- [28] H. Rosenkrantz: "Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens", edited by R. Benigni, pp.175-206, CRC Press, 2003.
- [29] R. Benigni: *Mutagenesis*, Vol.6, pp.423-425, 1991.
- [30] D. F. V. Lewis; C. Ioannides; D. V. Parke: *Mutagenesis*, Vol.5, pp.433-435, 1990.
- [31] I. Moriguchi; S. Hirono; Q. Liu; Y. Matsushita; T. Nakagawa: *Chemical*

- and Pharmaceutical Bulletin, Vol.38, pp.3373-3379, 1990.
- [32] I. Moriguchi; S. Hirono; Y. Matsushita; Q. Liu; I. Nakagome: Chemical and Pharmaceutical Bulletin, Vol.40, pp.930-934, 1992.
- [33] I. Moriguchi; S. Hirono; Q. Liu; I. Nakagome: Quantitative Structure-Activity Relationship, Vol.11, pp.325-331, 1992.
- [34] 森口郁生: 薬学雑誌, Vol.114, pp.135-146, 1994.
- [35] R. W. Tennant; J. Spalding; S. Stasiewicz; J. Ashby: Mutagenesis, Vol.5, pp.3-14, 1990.
- [36] R. Benigni: "Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens", edited by R. Benigni, pp.259-282, CRC Press, 2003.
- [37] G. Bakale; R. D. McCreary: Mutagenesis, Vol.7, pp.91-94, 1992.
- [38] Anon.: "Predicting Chemical Carcinogenesis in Rodents — An International Workshop", NIEHS, 1993.
- [39] J. D. Jones; C. E. Easterly: Mutagenesis, Vol.6, pp.507-514, 1991.
- [40] D. M. Sanderson; C. G. Earnshaw: Human and Experimental Toxicology, Vol.10, pp.261-271, 1991.
- [41] K. Enslein; B. W. Blake; H. H. Borgstedt: Mutagenesis, Vol.5, pp.305-306, 1990.
- [42] R. Benigni: Environmental and Molecular Mutagenesis, Vol.19, pp.83-89, 1992.
- [43] G. A. Kerckaert; R. Brauninger; R. A. LeBoeuf: Environmental Health Perspectives, Vol.104, pp.1075-1084, 1996.
- [44] Y. Lee; B. G. Buchanan; H. S. Rosenkranz: Environmental Health Perspectives, Vol.104, pp.1059-1064, 1996.
- [45] R. Benigni; C. Andreoli; R. Zito: Environmental Health Perspectives, Vol.104, pp.1041-1044, 1996.
- [46] J. Bootman: Environmental and Molecular Mutagenesis, Vol.27, pp.237-243, 1996.
- [47] R. W. Tennant; J. Spalding: Environmental Health Perspectives, Vol.104, pp.1095-1100, 1996.
- [48] Y. T. Woo; D. Y. Lai; J. C. Arcos; M. F. Argus; M. C. Cimino; S. DeVito; L. Keifer: Journal of Environmental Science and Health — Part C: Environmental Carcinogenesis and Ecotoxicology Reviews, Vol.C15, pp.139-160, 1997.
- [49] J. Ashby: Environmental Health Perspectives, Vol.104, pp.1101-1104, 1996.
- [50] J. Huff; E. Weisburger; V. A. Fung: Environmental Health Perspectives, Vol.104, pp.1105-1112, 1996.
- [51] I. Moriguchi; H. Hirano; S. Hirono: Environmental Health Perspectives, Vol.104, pp.1051-1058, 1996.
- [52] T. D. Jones; C. E. Easterly: Environmental Health Perspectives, Vol.104, pp.1017-1030, 1996.
- [53] C. A. Marchant: Environmental Health Perspectives, Vol.104, pp.1065-1074, 1996.
- [54] D. F. V. Lewis; C. Ioannides; D. V. Parke: Environmental Health Perspectives, Vol.104, pp.1011-1016, 1996.
- [55] R. Purdy: Environmental Health Perspectives, Vol.104, pp.1085-1094, 1996.
- [56] R. D. King; A. Srinivasan: Environmental Health Perspectives, Vol.104, pp.1031-1040, 1996.
- [57] Y. P. Zhang; N. Sussman; O. T. Macina: Environmental Health Perspectives, Vol.104, pp.1045-1050, 1996.

- [58] <http://www.informatik.uni-freiburg.de/~ml/ptc/>
- [59] M. T. D. Cronin: "Predicting Chemical Toxicity and Fates", edited by M. T. D. Cronin and D. J. Livingstone, pp.413–427, CRC Press, 2004.
- [60] <http://www.epa.gov/oppt/exposure/docs/episuitd.htm>
- [61] J. Zupan; J. Gasteiger: "Neural Networks for Chemists", VCH, 1993; 田辺和俊; 長塚義隆 (訳): 「化学者のためのニューラルネットワーク入門」, 丸善, 1996.
- [62] J. Devillers: "Neural Networks in QSAR and Drug Design", Academic Press, 1996.
- [63] K. L. Peterson: "Reviews in Computational Chemistry, Volume 16", edited by K. B. Lipkowitz and D. B. Boyd, pp.53–140, Wiley-VCH, 2000.
- [64] 松本高利; 田辺和俊: JCP E Journal, Vol.11, pp.29–34, 1999.
- [65] 田辺和俊; 松本高利: Journal of Computer Chemistry, Japan, Vol.1, pp.23–28, 2002.
- [66] 田辺和俊; 大森紀人; 小野修一郎; 鈴木孝弘; 松本高利; 長嶋雲兵; 上坂博亨: Journal of Computer Chemistry, Japan, Vol.4, pp.89–100, 2005.
- [67] K. Tanabe; N. Ohmori; S. Ono; T. Matsumoto; T. Suzuki: Journal of Ecotechnology Research, Vol.11, pp.111–116, 2005.
- [68] D. Villemin; D. Cherqaoui; A. Mesbah: Journal of Chemical Information and Computer Sciences, Vol.34, pp.1288–1293, 1994.
- [69] X.-H. Song; M. Xiao; R.-Q. Yu: Computers and Chemistry, Vol.18, pp.391–396, 1994.
- [70] S. Hatrik; P. Zahradnik: Journal of Chemical Information and Computer Sciences, Vol.36, pp.992–995, 1996.
- [71] M. Vracko: Journal of Chemical Information and Computer Sciences, Vol.37, pp.1037–1043, 1997.
- [72] G. Gini; M. Lorenzini: Journal of Chemical Information and Computer Sciences, Vol.39, pp.1076–1080, 1999.
- [73] R. Vendrame; R. S. Braga; Y. Takahata; D. S. Galvao: Journal of Chemical Information and Computer Sciences, Vol.39, pp.1094–1104, 1999.
- [74] M. J.-Heravi; F. Parastar: Journal of Chemical Information and Computer Sciences, Vol.40, pp.147–154, 2000.
- [75] S. C. Basak; G. D. Grunwald; B. D. Gute; K. Balasubramanian; D. Opitz: Journal of Chemical Information and Computer Sciences, Vol.40, pp.885–890, 2000.
- [76] D. Bahler; B. Stone; C. Wellington; D. W. Bristol: Journal of Chemical Information and Computer Sciences, Vol.40, pp.906–914, 2000.
- [77] F. R. Burden; M. G. Ford; D. C. Whitley; D. A. Winkler: Journal of Chemical Information and Computer Sciences, Vol.40, pp.1423–1430, 2000.
- [78] F. R. Burden; D. A. Winkler: Chemical Research in Toxicology, Vol.13, pp.436–440, 2000.
- [79] N. Cristianini; J. Shawe-Taylor: "An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods", Cambridge University Press, 2000; 大北剛 (訳): 「サポートベクターマシン入門」, 共立出版, 2005.
- [80] N. Chen; W. Lu; J. Yang; G. Li: "Support Vector Machine in Chemistry", World Scientific Publishing Co. Inc., 2004.
- [81] R. D. Combes; R. A. Rofford: "Predicting Chemical Toxicity and Fates", edited by M. T. D. Cronin and D. J. Livingstone, pp.193–204, CRC

Press, 2004.

[82] S. Parsons; P. McBurney: "Predictive Toxicology", edited by C. Helma, pp.135-175, CRC Press, 2005.

[83] Y. T. Woo; D. Y. Lai: "Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens", edited by R. Benigni,

pp.41-80, CRC Press, 2003.

[84] Y.-T. Woo; D. Y. Lai: "Predictive Toxicology", edited by C. Helma, pp.385-413, CRC Press, 2005.

(2006年2月14日受付)

(2006年5月12日採択)