

20094/018A

厚生労働科学研究費補助金

化学物質リスク研究事業

化学物質リスク評価における(定量的)構造活性相関((Q)SAR)  
およびカテゴリーアプローチの実用化に関する研究

平成21年度 総括・分担研究報告書

研究代表者 本間正充

平成22(2010)年5月

## 目 次

### I. 総括研究報告

化学物質リスク評価における(定量的)構造活性相関((Q)SAR)およびカテゴリーアプローチ  
の実用化に関する研究 \_\_\_\_\_ **1**

本間 正充、林 真、江馬 眞、広瀬 明彦、小野 敦、吉田 緑、宮島 敦子

II. 研究成果の刊行に関する一覧表(別添4) \_\_\_\_\_ **69**

III. 研究成果の刊行物・別冊 \_\_\_\_\_ **71**

平成21年度厚生労働科学研究費補助金  
(化学物質リスク研究事業)

化学物質リスク評価における(定量的)構造活性相関((Q)SAR)

およびカテゴリーアプローチの実用化に関する研究

I.総括・分担研究報告書

平成 21 年度厚生労働科学研究費補助金（化学物質リスク研究事業）  
総括・分担研究報告書

化学物質リスク評価における(定量的)構造活性相関((Q)SAR)  
およびカテゴリーアプローチの実用化に関する研究

研究代表者	本間 正充	国立医薬品食品衛生研究所 変異遺伝部 室長
研究分担者	林 真	(財)食品農医薬品安全性評価センター センター長
研究分担者	江馬 真	(独)産業技術総合研究所 安全科学研究部門 招聘研究員
研究分担者	広瀬 明彦	国立医薬品食品衛生研究所 総合評価研究室 室長
研究分担者	小野 敦	国立医薬品食品衛生研究所 総合評価研究室 主任研究官
研究分担者	吉田 緑	国立医薬品食品衛生研究所 病理部 室長
研究分担者	宮島 敦子	国立医薬品食品衛生研究所 薬理部 主任研究官
研究協力者	鎌田 栄一	国立医薬品食品衛生研究所 総合評価研究室 室長
研究協力者	平田 睦子	国立医薬品食品衛生研究所 総合評価研究室 主任研究官
研究協力者	川村 智子	国立医薬品食品衛生研究所 総合評価研究室
研究協力者	高橋 美加	国立医薬品食品衛生研究所 総合評価研究室
研究協力者	菅谷 芳雄	国立環境研究所 環境リスクセンター 主任研究官
研究協力者	宮地 繁樹	(財)化学物質評価研究機構 安全性評価技術研究所 主任
研究協力者	茂木 邦夫	(株)シーティシー・ラボラトリーシステムズ
研究協力者	スイロイ ホセ	(株)富士通九州システムエンジニアリング

研究要旨

我が国の化学物質審査規制法では、年間の製造輸入量が 10 トン以下の新規物質（低生産量物質）については、ヒト健康影響に対するスクリーニング毒性試験の実施が求められていない。また、数十万種にもおよぶ既存化学物質の多くについてもヒト健康影響について十分な評価がなされておらず、早急な対応が求められているが、全ての化学物質について動物試験を行うには、莫大な費用や時間がかかる。

本研究では、化学物質の安全性を既存の毒性情報を活用することにより効率的かつ動物を使用せずに把握できる手法として期待されるカテゴリーアプローチや(定量的)構造活性相関((Q)SAR)の安全性評価における実用化に向けて、これまでの研究において構築済みの数種の (Q) SAR 評価モデルの改良や予測適用範囲の拡充、カテゴリーアプローチの安全性評価への適用による利用法・有用性の検証を行い、得られた成果をもとにこれらの手法を安全性評価に適用するためのガイダンスを提案することを目的としている。

本年度は、構造活性相関モデルの構築および具体的施策に適応にむけた改良として、*in vitro* 染色体異常誘発性に関しては、これまでに構築した知識ベースエキスパートシステムについて、7 化合物クラスの解析により、新たに 2 つのアラートと 5 つの Rapid Prototype アラートの構築に成功した。また、組織代謝シミュレータとの組み合わせによる予測精度の向上について検討を行った。肝障害予測については、AMES 試験予測において有用性が示された MultiCase を用いて新たに予測モデル構築を行うとともに、ADMEWORKS を用いた肝毒性予測モデルについて、学習データおよびモデル化パラメータ見直しによる精度向上の検討を実施した。ADMEWORKS については、既存の Ames 試験と染色体異常試験予測モデルについても精度向上に向けた再構築を実施した。また、病理組織学的所見をターゲットとした予測モデル構築やカテゴリーアプローチにおける毒性類似性の評価のため、既存化学物質のラットを用いた反復投与毒性試験報告書における病理組織学的所見を抜粋し、臓器ごとのシソーラスを構築した。一方、カテゴリーアプローチの有用性について検証するため、アミノフェノール類を対象として構造類似化合物の物理化学的性状及び環境有害性について既存情報を集収してカテゴリー評価にむけた精査を行った。

A. 研究目的

我が国では、昭和 48 年の化学物質の審査及び製造等の規制に関する法律（化審法）制定以降

に新たに流通した化学物質については事前審査が実施されてきており、ヒト健康影響に関しては、製造輸入業者は 3 種のスクリーニング毒性

試験（AMES 試験、染色体異常試験および動物を用いた 28 日間反復投与試験）結果の提出が義務付けられ、それに基づく安全性評価結果をもとにした規制が行われてきている。

しかし、化審法制定以前から市場に存在する化学物質（既存化学物質）については、国自ら必要な試験を実施して安全性評価を行い必要に応じて同法による規制措置を講じることとされているが、我が国では数万種の既存化学物質が流通しており、スクリーニング毒性試験の実施には費用も期間もかかることから、これまでに安全性評価が終了した化学物質はごく一部であり、ほとんどの既存化学物質については十分な安全性評価が実施されないまま使用されている。

既存化学物質の多くは、世界中で流通していることから、OECD ではいずれかの加盟国で生産量が多く、安全性情報が少ない既存化学物質の安全性点検を加盟各国で分担して行う高生産量化学物質安全性点検計画（HPV プログラム）を 1993 年より開始したが、現在までに評価が終了した化学物質は約 400 物質程度に留まる。

化学物質の安全性評価を従来の毒性試験によって行うためには、多額の費用と期間がかかるため、たとえ国際協力の下で実施したとしても数万種の既存化学物質について十分な評価を行うことは不可能であり、より効率的な方法が求められている。

更に我が国では平成 15 年の化審法改正により、それまでスクリーニング毒性試験の実施が求められていた製造輸入量年間 1t 以上 10t 未満の物質（低生産量物質）については、スクリーニング毒性試験の実施は不要とされ、既存情報に基づき審査が行われることとなった。しかし、安全性評価に利用可能な既存情報が入手可能である物質は限られており、結果としてヒト健康影響に関しては不明のまま流通している。しかも低生産量物質の申請数は、スクリーニング毒性試験を行った物質数を大きく上回っている。低生産量であっても、ヒト健康に対する影響を把握出来ることが望ましいが、その一方で経済的な観点や動物愛護の観点から全ての低生産量物質について動物試験を実施することは非効率的であり、効率的かつ動物を使用せずに毒性の強さを把握できる最良な方法が求められている。

受容体作用などの特異的作用を除き、化学物質の毒性は、化学物質と生体分子の非特異的相互作用によって引き起こされると想定され、分子構造と物理化学的性質が類似する化学物質同士は、類似する生体反応性すなわち毒性を有すると考えられる。このことから、構造類似化合物の既知の安全性情報から評価対象化合物の安全性を類推するカテゴリーアプローチによる安全性評価への利用が世界各国で検討されている。

さらには、部分構造や計算可能な物理化学的性質などと毒性発現との相関から評価対象化合物の安全性を予測する（定量的）構造活性相関（SAR）手法についても安全性評価における実用化が期待されている。

近年、安全・安心についての関心が高まる中、世界中で化学物質に対する懸念が広がっており、欧州ではすべての化学物質を対象とした規制が平成 19 年に施行されるなど、化学物質管理を巡る状況は大きく変化しつつある。

そうした状況下にあつて、我が国の化審法においても平成 21 年改正にあたり既存化学物質も含めた全ての化学物質の包括的管理制度が導入され、さらにはこれまで安全性評価の対象ではなかった「環境中で分解しやすい化学物質」についても安全性評価の対象とする一方で、その附帯決議においては、化学物質のスクリーニング評価に要する費用・期間の効率化や国際的な動物試験削減の要請にかんがみ、定量的構造活性相関の活用等を含む動物試験の代替法の開発・活用を促進することが求められている。

カテゴリーアプローチや SAR 手法を組み合わせた、効率的かつ信頼性の高い安全性評価スキームが実用化されれば、全ての化合物について多額の費用と時間がかかる安全性試験を行わなくても、安全性について有る程度の予測的評価を行うことが可能となり、さらに、安全性試験を実施すべき物質を選定して試験実施することで効率的な安全性評価が可能となる。

本研究では、先立つ研究における成果をもとにカテゴリーアプローチや（定量的）構造活性相関（(Q)SAR）の安全性評価における利用の実用化に向けて、構築済みの数種の（Q）SAR 評価モデルの改良や予測適用範囲の拡充、カテゴリーアプローチの安全性評価への適用による利用法・有用性の検証を行い、最終的に得られた成果をもとにこれらの手法を化学物質評価に適用するためのガイダンス案を提案することを目的とした研究を進めている。

本年度は、構造活性相関モデルの構築および具体的施策に適応にむけた改良として、*in vitro* 染色体異常誘発性の予測に関しては、これまでに構築した知識ベースエキスパートシステムの被覆率向上のため、予備的調査から選択された 7 種の化合物クラスの解析により、新たに 2 つのアラートと 5 つの Rapid Prototype アラートの構築に成功した。また、組織代謝シミュレータとの組み合わせによる予測精度の向上について検討を行った。肝障害予測については、AMES 予測モデルの一つとして有用性が示された MultiCase を用いて新たに予測モデル構築を行うとともに、ADMEWORKS を用いた肝毒性予測モデルについて、学習データおよびモデル化

パラメータ見直しによる精度向上に向けた検討を実施した。ADMEWORKS については、既存の Ames 試験と染色体異常試験予測モデルについても予測精度向上のに向けた再構築をあわせて実施した。

また、現在の肝毒性予測モデルは、肝重量や逸脱酵素を予測フェノタイプとして用いているが、病理組織学的所見をターゲットとした予測モデル構築やカテゴリーアプローチにおける毒性類似性の詳細評価における利用のため、これまでに公表されている既存化学物質のラットを用いた短期反復投与毒性試験報告書において報告されている病理組織学的所見を抜粋し、臓器ごとのシソーラスを構築した。

一方、カテゴリーアプローチの有用性について検証するため、アミノフェノール類を対象として選択し物理化学的性状及び環境有害性について構造類似化合物の既存情報を収集してカテゴリー評価適用の可能性について検討した。

## B. 研究方法及び結果

### 1. Derek for Windowsの染色体異常エンドポイントの被覆率向上のための新規アラートおよびRapid Prototypeアラートの開発に関する研究

#### 1. 1. 序論

Derek for Windows (DfW) は毒性予測のための知識ベースエキスパートシステムであり、知識ベースは、DfW により行われる毒性予測に寄与するアラート、例証化合物、および Reasoning rule から構成されている。これまでの研究結果から、AMES 試験結果予測に関しては、国立医薬品食品衛生研究所 (NIHS) で収集した試験データをもとに信頼性の高い予測アラートの構築に成功しており、さらに別に構築したメカニズムの異なる他の2つの予測手法と組み合わせることにより実用可能なレベルの予測評価が可能であることから、本手法の毒性予測における有用性が示された。さらに引き続き研究により、NIHS で収集した *in vitro* 染色体異常試験および 28 日間反復投与毒性試験データをもとにした解析より肝毒性および染色体異常の新規 DfW アラートと Rapid Prototype アラートの導出することで被覆率を向上することによりより信頼性の高い予測が可能であることが示された。本年度の研究では、NIHS より提供された *in vitro* 染色体異常試験データに基づき新たなアラートと Rapid Prototype アラートを構築することにより染色体異常アラートの被覆率を更に向上させることを目的とした。NIHS より 2008 年 7 月に提供された 76 化合物の *in vitro* 染色体異常試験結果を用いた実行可能性調査により、11 個の新たな化合物クラスの特

と 2 つの既存染色体異常アラートの改良の見込みがあるという結果が得られた。DfW における染色体異常アラートは、これまでの NIHS と他の組織との共同研究によって開発されてきたものであり、現在、染色体異常に関して 82 本のアラートを含んでいる。DfW の染色体異常アラートを開発するための新たな *in vitro* 染色体異常試験データの使用により、文献からの裏付け情報が不足している化合物クラスの被覆率を向上が可能である。

#### 1. 2. 方法

NIHS より提供された 270 化合物の構造情報とそれに付随する *in vitro* 染色体異常試験データについて化合物構造の整理を行った。幾つかの化合物については価数違反により再描画された。また、化合物 ID199299415 と 199899411 は重複しており、解析に当たって片方の化合物構造のみを使用した。SD ファイル (NIHS data\_Sept09.sdf) を作成し、これには *in vitro* 染色体異常試験結果に基づき、115 個の陽性化合物、146 個の陰性化合物、および、試験結果がはっきりしない 8 個の化合物、の合計で全 269 個の化合物が含まれている。NIHS data\_Sept09 に対して、Derek for Windows version 12 で毒性予測を行い、予測精度を算出した。

当プロジェクトの初期の段階で実施された実行可能性調査 2 から、まず 7 個の新たな化合物クラスが選択された(付録 D)。化合物クラスは以下の通りである。

- ・ Alkyl anthraquinones
- ・ Alkyl naphthalenes
- ・ Alpha,beta-unsaturated pyridines - Michael acceptors
- ・ Aromatic nitriles
- ・ Benzoic acids
- ・ Benzothiazole or derivatives
- ・ Styrene or derivatives

これらの化合物クラスは、予備的調査が、これらを裏付けるために入手可能な文献情報が殆んどもしくは全く無いということを示唆していること、従って、アラートが主に NIHS データのみに基づき構築できるか否かを証明するだろうということを基盤として選択された。二つの化合物クラスがデータセット全体の解析により更に特定され、それらは anilines と thiophosphates である。

化合物構造と試験結果に加えて、その多くの化合物に対する詳細な *in vitro* 染色体異常試験データが NIHS ウェブページ から入手可能で

ある。試験結果の評価は、最大耐容量 10mM に基づいて行われ、それは構造的もしくは数的異常を示す化合物が陽性であると見做される数値である。10mM よりも高い投与量は陰性として見なされる。加えて、個々の化合物クラスに対して、試験に供された化合物クラス内の化合物を更に特定するため、および、観測された染色体異常の解析の起点を成すかも知れない毒性発現の理論的根拠を探るために、公共データおよび文献情報を検索した。文献情報での追加的な裏付けデータが不足しているが、明白な構造活性相関が導出できるような化合物クラスが Rapid Prototype アラートとして構築された。

### 1. 3. 結果

全部で 2 個のアラートと 5 個の Rapid Prototype アラートが構築され、Derek for Windows version12 の知識ベース(この報告書では、DfW12 NIHS 151209\_1 と引用する)に組み入れられた。これらのアラートは *in vitro* 染色体異常アラートとして実装され、一つの新しいエンドポイントが DfW ガイドラインに従って、それらの Rapid Prototype アラートに対して作成された(rapid prototypes: chromosome damage *in vitro*)。これらのアラートと Rapid Prototype アラートに対する DfW rtf 形式レポートが付録 III に添付されている。

実行可能性調査にて特定された二つの化合物クラス(benzoic acids と aromatic nitriles)は、公共文献中の構造類似化合物に対する証拠情報が不足していること、そして、構造類似化合物に対する試験結果に矛盾が見られることにより、DfW 知識ベースには実装されなかった。

二つの benzoic acid 誘導体が NIHS データセット中で特定され、その双方とも *in vitro* 染色体異常試験で陽性と報告されている。(図 1)

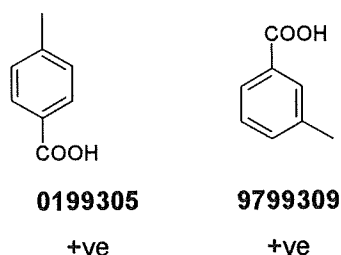


図 1: NIHS データセット内で特定された benzoic acid 誘導体

4-methyl benzoic acid (0199305) と 3-methyl benzoic acid (9799309) の詳細データについて、より詳しい調査により、これらの化合物がそれぞれ 13.2mM および 14.7mM の投与量にて構

造的な染色体異常を引き起こすことが明らかとなった。

幾つかのデータベース(NTP [1], Sofuni data book [2], Vitic [3])と組織内のデータセットが benzoic acid 官能基を含む化合物を探すために検索され、同じ化合物クラスに属する様々な置換基に対して雑多な試験結果が示されている。例えば、benzoic acid (1)は *in vitro* 染色体異常試験の結果がはっきりしなかった一方で、sodium benzoate (2)は陽性であると報告されている[2] (図 2)。

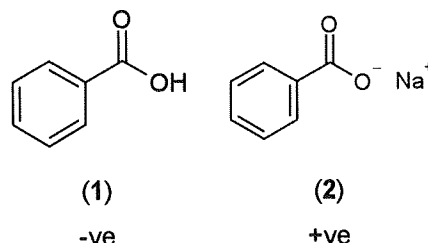


図 2: Benzoic acid (1)と sodium benzoate (2)

三つの化合物サブクラスが benzoic acid 誘導体群にて特定された:

Phthalate esters, Salicylates and Parabens.

Phthalate esters は、mono-2-ethylhexyl-phthalate [4]の例外を除き、一般に染色体異常試験で陰性を示す。染色体異常試験結果の多様性、および、多くの場合に陰性の試験結果が示されることを考えると、これらの化合物をカバーするアラートを導出することができないと結論付けられる。Salicylates は、NIHS 遺伝毒性プロジェクトにおいて構築された Rapid Prototype アラートによって当初カバーされたが、この化合物クラスは、その厳しく制限されたスコープ(適用範囲)のために、完全なアラート(Rapid Prototype でない、標準的なアラート)に変換されていない。Parabens、もしくは、p-hydroxybenzoic acid のエステルは、ミトコンドリアエネルギー生産攪乱物質として知られている[5]。これらの化合物は、酸化的リン酸化を脱共役させ、その結果として細胞内における ATP 産生を激減させる。細胞内における ATP の減少は、微小管の重合を妨げ、従って、紡錘体を分裂させることにより染色体異常を引き起こす。このミトコンドリア効果は、エステル基内のアルキル鎖が長くなると共に増加し、一方で、染色体異常はその逆の傾向を示す(より短いアルキル鎖は染色体異常試験で陽性を示す)。アラートがこの化合物クラスに対して構築可能かも知れないが、構造活性相関を調査するための更なる研究が必要とされるだろう。

四つの aromatic nitriles が NIHS データセットにおいて特定され、そのうちの三つが in vitro

染色体異常試験で陰性であり、残りの一つが陽性であると報告されている。(図 3 参照)

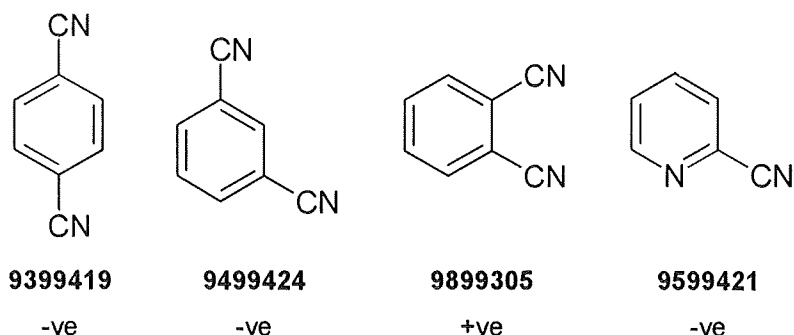


図 3: NIHS データセット内で特定された Aromatic nitriles

1,2-Dicyanobenzene (9899305)は、NIHS データセットにおいて in vitro 染色体異常試験で陽性であると報告されているが、そのデータの詳細な解析結果は、この化合物が倍数性を誘導し、構造的な染色体異常を引き起こさなかったことを示している。構造類似化合物の 1,3-dicyanobenzene (9499424) と 1,4-dicyanobenzene (9399419)は、染色体異常アッセイにおいて構造的・数的な異常のいずれも引き起こさず、3-cyanopyridine (9599421)についても同様に引き起こされなかった。

公的に入手可能な文献を検索することにより、in vitro 小核試験 [6] (図 4) における benzonitrile (3)の活性について記述している論文が発見された。その化合物は、サブマイクロモルの投与量レベルにおいて陽性を示し、主にキネトコア陽性微小核を生成し、また、ミリモルの投与量にも関わらず、チューブリン重合を阻害することが示されている。また、1,2-dicyanobenzene (9899305)はこのメカニズムを経由して倍数性を引き起こす可能性が考えられるが、この仮定に対する直接的な証拠は発見されていない。それに反して、chlorothalonil (4) (図 4) は構造的な染色体異常を誘導し [1]、それはこの化合物とグルタチオンのような細胞のサルファ求核剤との既知の反応性によるものと考えられる [7]。

この化合物クラス内の陽性化合物に対して得られた雑多な試験結果、および、続いて起こる多数のメカニズムに対する証拠情報から鑑みて、この化合物クラスに対するアラートを実装しないことにした。

二つの化合物クラスが完全なアラートへと変換された: alpha,beta-unsaturated pyridine, pyrimidine or 1,3,5-triazine と 2-thio-benzimidazole or -benzothiazole である。残りの化合物クラスは、文献からの裏付け情報が無いものの明白な構造活性相関を詳細なデータと NIHS より提供された化学構造から導出することができた。これの一つの例として、alkyl anthraquinones がある。DfW は、少なくとも一つの水酸基が分子内に存在しなければならないという制限が設けられている hydroxylated anthraquinones に関連したアラートを一件包含している。実験的な証拠情報は、これらの化合物により引き起こされる構造的な染色体異常が、代謝活性化を必要としていることを示唆している。従って、可能性のある毒性発現メカニズムは水酸化誘導体(活性酸素種の形成を通じての染色体異常活性ありとして既に知られている化合物)を形成するための芳香環における直接的酸化反応を含むかも知れない。この推測されたメカニズムに対する裏付け情報が不足しているため、NIHS より提供された化合物に基づき Rapid Prototype アラートが組み込まれた。

新しい知識ベースの予測精度を DfW version12 に対し、NIHS より提供された Kison データセットを使用して調査した。表 1 と表 2 は in vitro 染色体異常予測に対する、標準の DfW version12 と DfW12 NIHS 151209\_1 知識ベースそれぞれについての予測精度を示している。

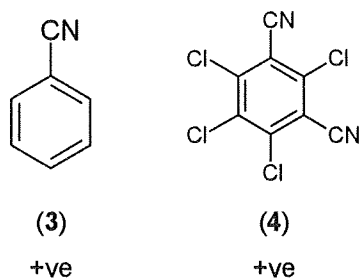


図 4: Benzonitrile (3)と chlorothalonil (4)



Derek for Windows \ 試験結果	陽性	陰性	合計
陽性	45	70	115
陰性	14	132	146

感度: 45/115 = 39%; 特異性: 132/146 = 90%  
一致率: 177/261 = 68%

表 1: NIHS Kison データセット内の化合物に対する Derek for Windows version 12 の予測精度

Derek for Windows \ 試験結果	陽性	陰性	合計
陽性	59	56	115
陰性	17	129	146

感度: 59/115 = 51%; 特異性: 129/146 = 88%  
一致率: 188/261 = 72%

表 2: NIHS データベース内の化合物に対する DfW12 NIHS 151209\_1 の予測精度

表 2 から、染色体異常予測の感度が 51%へと増加した一方で、特異性が 88%へとわずかに減少したことが見て取れる。特異性に対するこの僅かな減少は、Rapid Prototype アラートの実装に起因する 3 件の付加的な偽陽性予測結果によるものである。

外部データセットに対する予備的な調査結果

DfW12 NIHS 151209\_1 に組み込まれたアラートと Rapid Prototype アラートの予測精度を評価するために、NIHS より提供された全部で 53 個の化合物から成る外部データセット (Sinki データ) が予測に供され、in vitro 染色体異常に対する予測精度 (表 3) が算出された。

Derek for Windows \ 試験結果	陽性	陰性	合計
陽性	4	22	26
陰性	3	24	27

感度: 4/26 = 15%; 特異性: 24/27 = 89%  
一致率: 28/53 = 53%

表 3: NIHS より提供された Sinki データセット内の化合物に対する DfW12 NIHS 151209\_1 の予測精度

Sinki データセットに対する染色体異常予測の感度 (表 3) が 15%である一方で、特異性は 88%を超えたままである。表 3 から、ごく限られた個数の真陽性化合物 (4 化合物) と偽陽性化

合物 (3 化合物) が染色体異常アラートもしくは Rapid Prototype アラートにヒットしたことが見て取れる。これらの 7 個の化合物はこの研究プロジェクトにて組み込まれたアラートもしくは Rapid Prototype アラートのいずれにもヒット

トしなかった。この結果は、トレーニングデータセットに包含されているケミカルスペースがテストデータセット化合物から大きく離れているだけでなく、既存の DfW アラートによりカバーされているケミカルスペースからも遠いことを示している。結果として、改良の見込みがある既存の DfW アラートと新しい化合物クラスの特定がこの新しいデータセットに対する染色体異常予測感度を向上させるかも知れない。

#### 1. 4. 結論

染色体異常予測に対する二つのアラートと五つの Rapid Prototype アラートが DfW に組み込まれた。これらのアラート実装後のトレーニングデータセットに対する染色体異常予測精度は、感度 51%、特異性 88%であった。

NIHS より提供された私有データセットからのアラート開発の限界と被覆率はトレーニングデータセット化合物に包含されているケミカルスペースの多様性と文献からの裏付け証拠情報に依存する。Rapid Prototype アラートの開発に対しては、文献からの証拠情報は考慮しておらず NIHS データセット中の詳細な *in vitro* 染色体異常試験データのみを考慮している。従って、これらのアラートによりカバーされているケミカルスペースはトレーニングデータセット内にて得られる化合物クラスに制限される。詳細な試験データは、10mM の範囲を超える投与量で陽性を示すと報告されている化合物 (benzoic acid) を選別する際、および、化合物クラス (aromatic nitriles) 内の構造活性相関を評価する際に有用であると示された。

あるテストデータセットを使用しているアラートと Rapid Prototype アラートの予測精度評価結果は、テストデータセット化合物内のケミカルスペースの被覆率の不足を示唆している。偽陰性化合物の更なる解析は、テストデータセット化合物内に包含されている化合物クラスの幾らかの洞察を提供し、DfW におけるそれらの被覆率を改善させるかも知れない。

このプロジェクトは、私有データセットが DfW 知識ベースにおける染色体異常アラートの被覆率を改善させるために使用できることを示した。新しいテストデータセットを入手できると、それを Rapid Prototype アラートを保証もしくは改良するために使用することができ、それに引き続く標準 DfW アラートの開発を可能にする。

#### 参考文献

- [1] NTP; National Toxicology Program Web Server; available at "<http://ntp-server.niehs.nih.gov/>".

- [2] Sofuni; Revised Edition 1998 Data Book of Chromosomal Aberration Test *in Vitro*, Sofuni T (editor), Life-Science Information Center, Tokyo
- [3] *Vitic 4.6*, Lhasa Limited, UK
- [4] Philips BJ *et al*, *Environmental Health Perspectives* 1986, 65, 263-266
- [5] Nakagawa Y and Moldeus P, *Biochemical Pharmacology* 1998, 55 (11), 1907-1914
- [6] Bonacker D *et al*, *Archives in Toxicology* 2004, 78, 49-57
- [7] Arvanites AC and Boerth DW, *Journal of Molecular Modeling* 2001, 7, 245-256

## 2. Optimized Approach Based on Structural Indices Set (OASIS) と Tissue Metabolic Simulator (TIMES) による染色体異常誘発性の予測に関する研究

### 2. 1. 染色体異常誘発性とそのモデル化

遺伝毒性の評価は、化学物質の承認及び登録における安全性確認に重要である。すべての遺伝毒性エンドポイントを検出しようとする単一の試験はないため、遺伝毒性の評価にはエンドポイントが異なる一連の *in vitro* 及び *in vivo* 試験が推奨されている。遺伝毒性の予測に使用されるコンピュータプログラムは、これまで主にバクテリアでの復帰突然変異試験 (エームス試験) の結果を変異原性の指標として予測することに焦点を合わせており、これに関しては比較的良好な予測率が達成されている。エームス試験に対するモデルの性能が優れていることは、この試験システムに関して利用可能なデータが豊富であること、遺伝学的エンドポイントの基礎をなしている分子メカニズムが比較的良好に理解されていること、化学構造と遺伝毒性を直接的に関連付けることが可能であることなどによる。一方、染色体異常を検出に関してはモデル開発のプロセスがより複雑である。染色体異常試験は標準化された試験であるが、利用可能な実験データベースはエームス試と比較し著しく少ない。さらに、化合物と DNA の直接的相互作用に加え、化合物と DNA 複製及び転写に関与する酵素や、染色体分配に関与するその他の核タンパク質 (例、ヒストンタンパク質) との相互作用などのその他のメカニズムも染色体異常を引き起こす場合がある。染色体異常は、分裂中期における染色体の構造的・特性的結果として観察可能となり、構造的又は数的異常として認識される。従って、染色体異常を引き起こす化学物質の予測のためのモデル化に多様な計算的アプロ

一が必要である。

## 2. 2. これまでの染色体異常誘発性を予測するためのQSARモデル

染色体異常を予測するための、化学物質のカテゴリ化に関しては、統計学に基づくものが大部分であり、いわゆる機械学習 (ML) モデルによる解析が適用される。これらのモデルは、化学構造をコードしている多数の分子記述子 (構造的指標) の同定、重要な構造的情報をコードしている指標の選択 (例、重複した情報の除去)、そして同定された重要な記述子を用いた分類モデルの構築に基づく。分子記述子は、研究対象である化合物の分子形態、幾何学的配置、及び電子構造を反映する。遺伝的アルゴリズムなどの異なる特徴を選択するアプローチは、多様な重要性を有する方法として知られており、最も理論的な指標を特定するために使用されている。ここでは、多数の分類モデルを使用する。ML の高い統計学的性能にも関わらず、これらのモデルでは作用機序に関する知識に基づくものではない。

統計学に基づいた遺伝毒性の予測に使用される別の方法として、コンピュータ自動構造評価 (Computer Automated Structure Evaluation : CASE) システムがある。最近では、CASE システムは、デンマーク環境保護庁において、染色体異常に関するモデルを開発に利用されている。このシステムでは、モデル化された影響のモジュレータとして作用するフラグメント (物理化学的パラメータ) を検出する。石館データコレクションから採取された約 500 種類の化合物を用いてトレーニングを行ったところ、感度 ( $\leq 60\%$ ) 及び特異度 (82%) と比較的低い予測率であった。この低い予測率は、根底にある 2 次元 (2D) フラグメントベースの記述子が、染色体異常試験において陽性の結果を引き起こす多数のメカニズムを表現していないためと考えられる。

染色体異常誘発性を有する化学物質の分類の識別を促進する構造アラートの規則を同定する試みが、多くの研究者により行われている。化合物の反応性に基づきアラートを直接的に同定する代わりに、ここでは化合物の分子構造にコードされている情報を用いている。この目的のために、分子の各部分による活性への寄与を計算する形態的サブストラクチャー分子デザイン (Topological Sub-Structural Molecular Design : TOPS-MODE) と呼ばれる分類モデルが開発された。TOPS-MODE はサブストラクチャー的な性質を有するため、この QSAR モデルは、エンドポイント値が結合寄与の合計値として計算される増分法とみなすことができる。このアプロー

チは、1 セットの化合物の染色体異常誘発活性に対して著しく寄与する領域を同定するためのものである。これにより、化学物質より誘発される染色体異常の原因である可能性のあるアラートを含む 22 の規則を得ることが可能となった。これらの規則により、潜在的な染色体異常誘発性を有する特定の種類の化合物の同定が可能となり、エキスパートシステムにおいて実施可能となった。

既存の知識によるリスクの演繹的推定 (Deductive Estimation of Risk from Existing Knowledge : DEREK) は、機構的規則に基づくアプローチを用いたエキスパートシステムである。このシステムでは、化学構造との定性的関連性に基づき専門家により抽出された「If-Then-Else」規則を適用している。最近の改良において、染色体異常を予測可能なモジュールが、変異原性及び発癌性を予測するモジュールに追加された。専門家の提案により得られた約 100 種類のアラートのプロトタイプ、そして複数の *in vitro* 染色体異常試験データコレクションの解析が提案されている。

染色体異常の予測に関してレビューされたモデルの解析により、これらのモデルはスクリーニングを目的とする場合には有用であるものの、以下の欠点があることが指摘されている。第 1 に、大多数のモデルは統計学的性質を有するため、純粋に機構的な基礎に基づく構造アラート構造アラートの明確な同定が不可能である。一部のモデルにおけるアラートの同定は、染色体異常を引き起こす可能性のある多数のメカニズムと関連していない。第 2 に、染色体異常誘発活性に特異的なメカニズムの同定がなされていない。このため、同定されたフラグメントのほぼすべては DNA 反応性に関連した既知の構造アラート構造アラートを表しており、タンパク質結合を生じるアラートは規定されていない。第 3 に、モデルの作成に使用されたトレーニング用セットにおいて、親構造として活性を誘発する化合物と、代謝活性化を必要とする化合物の区別がなされていない。すべての実験的研究において、代謝活性化を含むか否かにかかわらずデータは、同一の生物活性化合物の分類に含まれていた。このため、化合物の代謝活性化は、これまで作成されているモデルにおいてはシミュレートされていない。

## 2. 3. 組織代謝シミュレータ (Tissue Metabolic Simulator ; TIMES)

TIMES モデルは確率的アプローチに基づく。これは、階層的に並べられたトランスフォーメーションのリスト、そしてこれらを実行するサブストラクチャーとマッチしたエンジンか

らなる。ブルガス大学が開発した確率的アプローチでは、トランスフォーメーションの階層は、実証された代謝トランスフォーメーションのデータベース、又は消失速度に関するデータを再現する過程において判定されたトランスフォーメーション確率により規定されている。トランスフォーメーション確率は、代謝試験の時間枠における様々な反応の発現の可能性に付随する速度定数と関連している。トランスフォーメーションは独立しており、経時的に行われると仮定されている。各分子のトランスフォーメーションは、親のサブ分子フラグメント、トランスフォーメーション産物、及び抑制マスク (inhibiting masks) からなる。抑制マスクは、反応抑制において役割を果たす。マスクとして割り当てられたフラグメントが標的のサブフラグメントに付着された場合、親化合物のトランスフォーメーションが妨げられる。代謝反応を促進又は抑制するグループの存在は、主要なトランスフォーメーション数を著しく増加させる。中間代謝及び臓器特異的な代謝において知られている有機官能基の数は 60 種類以下であるが、多官能性化合物に関して起こる可能性のある反応は無数である。位置異性も、組み合わせが爆発的に多くなる原因の一つである。現在、343 種類の主要なトランスフォーメーションが、ラット肝臓における代謝のモデル化に使用されている。これらのトランスフォーメーションは、2 種類の主要な反応に分類されている。第 1 の分類には、試験の時間尺度において非常に速い速度で起こる 41 種類の非生物のおよび酵素制御反応が含まれる。高度に反応性のグループのトランスフォーメーション及び中間体はここに含まれる。互変異性などの様々な化学平衡プロセスもまた、この種類のトランスフォーメーションに含まれる。反応の第 2 の分類には、酸化的、酸化還元、還元的、加水分解的、及び合成反応などの第 1 相及び第 2 相解毒メカニズムの 302 種類の代謝トランスフォーメーションが含まれる。

シミュレータは、親分子を、発現する確率が最も高いトランスフォーメーションと関連した反応フラグメントとマッチさせることから開始される。マッチが同定された場合、分子は代謝され、トランスフォーメーション産物は次の変換工程の親分子として扱われる。この手順は、新規に形成された化合物に対し、継続的に実施されたトランスフォーメーションの確率の積が、ユーザーにより規定された閾値に達するまで繰り返される。初めに、親化合物はトランスフォーメーションリストにかけられ、付随するサブストラクチャーに適合するすべてのトランスフォーメーションが親化合物に対して実行され、

第 1 レベルの代謝物のリストが産生される。産生されたこれらの各代謝物は、第 2 レベルの代謝物を産生するために同一のトランスフォーメーションリストにかけられ、これが繰り返される。数学的形式論は、トランスフォーメーションが経時的に起こるという仮定に基づいている。つまり、最も可能性の高いトランスフォーメーションが最初に親化合物に適用され、その後残りの未代謝親分子がより確率の低い第 2 のトランスフォーメーションを受け、これが繰り返される。代謝シミュレータの反応確率は、哺乳類 (主にラット) の肝代謝に関する 332 種類の実証されたマップを含むデータベースを再現するように調整されている。トレーニング用セットが実証されたマップを再現する程度が、シミュレータの性能を規定し、これは新規の実験的エビデンスにより調節可能である。同様に、シミュレータの個々のトランスフォーメーションが実証された代謝物を再現する速度に基づき、産生された代謝物及び代謝マップの信頼性を評価するアセスメントが導入されている。

#### 2. 4. Optimized Approach Based on Structural Indices Set (OASIS) と Tissue Metabolic Simulator (TIMES) による染色体異常誘発性の予測

ブルガス大学のメケニアン博士らは、DNA との共有結合的相互作用に関して過去に作成されているモデル (エームス試験モデル) を、染色体異常誘発活性を引き起こすタンパク質との付加的な相互作用のメカニズムを考慮した新規のモデルを組み合わせた新しい染色体異常を予測する QSAR を開発した。タンパク質との付加的なメカニズムとしてはトポイソメラーゼ阻害、及び DNA に関与する核タンパク質 (例、ヒストンタンパク質) との相互作用を計算に入れた。これらメカニズムは DNA 合成を妨害し、染色体異常を引き起こす。この関連性を考慮し、染色体異常の QSAR による予測のために一連のモデルが開発された。これは、構造学的指標セットに基づく最適化アプローチ (Optimized Approach Based on Structural Indices Set : OASIS) の組織代謝シミュレータ (tissue metabolic simulator : TIMES) 構築基盤に組み入れられているエームス試験モデルと、染色体異常を引き起こす付加的な相互作用メカニズムを説明するためのモデルを組み合わせたものである。いずれのモデルも、古典的な反応性アラートの概念に基づくが、特定化されたアラートの一部は、DNA 又は核タンパク質と直接相互作用するが、その他は残りの分子によるアラートの活性化の程度を評価する 2 又は 3 次元定量的構造活性相関モデルとの組み合わせで構成され

ている。各アラートの使用は、相互作用の機構的解釈により検証されている。ラット肝臓 S9 の代謝シミュレータとの併用することにより、親化合物としては活性を誘発しないが、代謝的に活性化された化合物によって誘発された染色体異常誘発性を予測することが可能である。モデルを評価するために、化学物質を代謝活性化の有無別に識別したトレーニング用セットを用いた。代謝活性化を伴わないモデルの性能は、感度 77%、及び特異度 82%であった。一方、代謝シミュレータと組み合わせたモデルに関しては、感度は 75%であったが、特異度は 56%程度であった。このモデルの低い特異度を、特定のアラートと関連付けることは残念ながらできていない。モデルの性能は、さらなる S9 活性化化合物をトレーニング用セットに組み入れるごとに改善されると考えられる（特に代謝活性化後に染色体異常陰性となる化合物）。また、染色体異常試験結果陽性の多くには実際の遺伝毒性とは無関係の非特異的反応も多く含まれることから、信頼性の高い試験データセットの利用も予測率の改善に重要と考えられる。

## 2. 5. 最高濃度の低減化に伴う染色体異常試験陰性化合物のQSARでの予測

染色体異常試験結果陽性化合物の中で 1 mM 以上の高濃度のみで陽性を示すものは、非生理的条件下による非特異的な陽性反応であり、実際の遺伝毒性とは無関係との意見が多いことから、医薬品の遺伝毒性ガイドラインに関しては最高濃度の 1mM までの低減化が検討されている。これに伴い、これまでの染色体異常データベースの変更が必要となり、QSAR でのアラートの変更も必要となるかもしれない。また、非

特異的陽性結果が減少することから予測率の向上が期待できる。一方、化学物質によっては逆に高濃度でのみ陽性を示す潜在的遺伝毒性物質を見逃す可能性があり、その予測には QSAR が有効であるかもしれない。

1994年から2006年までに我が国で OECD ガイドラインに従い、GLP 管理の下に *in vitro* 染色体異常試験が行われた 249 化合物を調査した。この中で陽性結果を示すものは 113 化合物 (45%) であり、この陽性化合物のうち 1mM 以上で陽性を示すものは 54 化合物 (48%) であった。1mM に最高用量を低減化することによってこれら化合物は陰性と判定される。この 54 化合物が陰性と判断される科学的妥当性、および健康リスクへの影響を検討した。54 化合物のうちエームス試験陽性化合物が 14 化合物、OECD ガイドラインで要求する濃度 (10mM) 以上で試験され、陽性となったものを除くと 37 化合物であった。37 化合物中、20 化合物で観察された陽性反応は低 pH (5 化合物)、高い細胞毒性 (13 化合物)、倍数体の誘発 (2 化合物) と関連した非特異的陽性反応と判断された。残り 17 化合物に関しては陰性とすべき合理性は染色体異常試験結果、他の遺伝毒性試験データからは認められなかった。17 化合物については DEREK、TIMES によって染色体異常誘発性に関する QSAR アラートの検索を行った。DEREK では 3 化合物、TIMES では 10 化合物 (代謝中間体を含む) でアラートが検出された (表 1)。TIMES での特異性は 56% と低いことが報告されており、この全てに遺伝毒性があるとは考えにくい、いくつかの化合物 (#99, #91) に関しては、さらなる検証が必要と考えられる。

表 1 DEREK、TIMES による染色体異常誘発性に関する QSAR アラートの検索結果の比較

Chemical	CA Alerts by DEREK	CA Alerts by TIMES
<b>Positive only with S9</b>		
1,4-Dibromobenzene (#65)		CA (M)
Dibutyl adipate (#70)		
2-Mercaptobenzimidazole (#71)		CA (P, M)
Sorbitan monooleate (#76)		
N,N-Dimethylbenzylamine (#79)		CA (M)
2-(Diethylamino)ethyl methacrylate (#83)		CA (M)
(Methacryloyloxyethyl) trimethylammonium chloride (#99)	CA	CA (P)
Ethenyltrimethoxysilane (#111)		CA (M)
<b>Others</b>		
2-Hydroxypropanenitrile (#60)		
N-Methylaniline (#68)		CA (M)
Methacrylic acid, monoester with propane-1,2-diol (#75)		CA (P, M)
2,4-Dinitrophenol (#91)	CA	CA (P, M)
Trimethylamine (#92)		
2-Chlorophenol (#96)	CA	
p-Nitrophenol sodium salt (#98)		CA (P, M)
2-(Di-n-butylamino)ethanol (#110)		

P: Parent, M: Metabolite

### 3. MULTICASEによる肝毒性予測の検討

#### 3. 1. はじめに

MULTICASE (MultiCASE Inc.) は、化合物の薬理学的および毒性学的な性状を予測できる独自の CASE 理論に基づいて FDA と共同開発されたシステムである。

本研究では、MULTICASE を用いて化合物の反復投与毒性を予測することを目的として、実際に行われた反復投与毒性試験の結果を用いてデータベースを構築し、その予測精度の検証を行った。

#### 3. 2. 対象とした項目

反復投与毒性試験において、肝臓の病理組織学的変化が認められた化合物を活性あり(陽性)とし、その最小毒性量(LOEL)を活性値として使用した。

#### 3. 3. MULTICASEについて

##### 3.3.1 予測方法の詳細

予測方法の詳細は、以下の通りである。

- ① データベースに入れた化合物の部分構造を認識し、部分構造ごとに生物活性値を統計処理する。これを biophore とする。ここで、データベース中にある同一 biophore を持つ分子の活性値の平均が算出される。
- ② biophore ごとに生物活性値を修飾する因子(同一の biophore をもつ分子の平均活性値を強めたり弱めたりする)を見つける。これを modulator とする。この modulator は、biophore の修飾子となる部分構造や、logP、HOMO/LUMO エネルギー、Biophore 上の荷電密度、location of hydrogen donors、location of hydrogen acceptors、location of lipophilic centers などの化学的性質から見つけ出す。
- ③ 化学物質の活性値 =  $\Sigma$ (biophore の平均活性値 + 各 modulator 固有の修飾値) で計算される。

##### 3.3.2 データベース作成

以下のように 4 つのデータフィールドを設け、1 化合物につき 1 行の形式の ASCII テキストファイルを作成し、それを読み込ませることによってデータベースを作成した。

CAS number\*1[5 スペース]活性値\*2 [5 スペース] -S(Smiles code\*3)[5 スペース]化合物名\*4

\*1CAS number : -(ハイフン)を除いて入力。  
不明な場合は"0"を入力。

\*2 活性値 : MULTICASE は化合物を不活性、限界活性、活性として分類する。20 未満の値は不活性(inactive)、30 を超える値は活性(active)とみなされ、20~30 の値は限界活性(marginal)とされる。最高値は 80。

\*3Smiles code : 必ずプリフィックス-S を入れる。180 文字以上はプログラムが受け付けないので、その化合物はデータベースから削除しておく。

\*4 化合物名 : 30 文字以内にする。30 文字を超えると、プログラムによって切り捨てられてしまう。

##### MULTICASE データベースの例

0	44.92555294	-SN2=NC=CN2	9109005A
1072680	44.47378329	-SCN%10N=CC(C)=C%10	9209008A
0	41.34977357	-SCC1=CC(C)=NN1	0707503B
754052	41.34977357	-SC=C[S](C)(C)C	0110004A
27646806	33.20771791	-SOCC(C)(C)N([H])C	0303005A

##### 3.3.3 予測精度に関する用語とその定義

- sensitivity 感度 : 動物試験の陽性結果を陽性と正しく予測した割合 (%)
- specificity 特異性 : 動物試験の陰性結果を陰性と正しく予測した割合 (%)
- concordance 全体の一致率 : 動物試験の陽性結果を陽性、陰性結果を陰性と正しく予測した割合 (%)
- applicability 適用性 : 全体のデータ数に対し、予測したデータ数の割合 (%)

##### 3.3.4 使用したデータ

###### • 新規化学物質

反復投与毒性試験データがある 1259 個のうち、エラーになる物質(金属を含む物質、SMILES が 180 文字以上の物質、毒性判断のためのフラグメントを持たない物質、MULTICASE が立体異性を認識できないため重複になる物質)を削除した 1231 個を使用した。

1231 個中、LOEL $\leq$ 1mg/kg が 3 個、1mg/kg<LOEL<1000 mg/kg が 344 個、LOEL $\geq$ 1000mg/kg が 98 個、所見なしが 786 個であった。

###### • 既存化学物質

254 個のうち、金属を含む物質と SMILES が 180 文字以上の物質を削除した 251 個を使用した。

251 個中、LOEL $\leq$ 1mg/kg が 1 個、1mg/kg<LOEL<1000mg/kg が 102 個、LOEL $\geq$ 1000 mg/kg が 15 個、所見なしが

133 個であった。

これらの物質を用いて、さまざまなパターンのラーニングセット、バリデーションセットを作成し、比較検討した。

### 3. 4. データベースの検証と考察

#### 3.4.1 ケース 1

新規化学物質 1231 個をラーニングセット、既存化学物質 251 個をバリデーションセットとした。  $0 \leq \text{LOAEL} < 1000 \text{mg/kg}$  を陽性、 $\text{LOAEL} \geq 1000 \text{mg/kg}$  と所見なしを陰性とした。

活性値は Table 1 のように設定した。

Table 1. LOAEL に基づく活性値

LOAEL	活性値
$1000 \text{mg/kg} \leq \text{LOAEL}$	10
$250 \text{mg/kg} \leq \text{LOAEL} < 1000 \text{mg/kg}$	30
$25 \text{mg/kg} \leq \text{LOAEL} < 250 \text{mg/kg}$	40
$5 \text{mg/kg} \leq \text{LOAEL} < 25 \text{mg/kg}$	50
$\text{LOAEL} < 5 \text{mg/kg}$	60

所見なしの活性値は、全て 10 とした。  
予測結果は Table 2 の通りであった。

Table 2. 既存化学物質の予測結果 (ラーニングセットは新規化学物質)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)		
					56.2 % (Concordance)	
試験結果	+	103	9	93	1	8.7 % (Sensitivity)
	-	148	14	132	2	89.2 % (Specificity)
		251	23	225	3	100 % (Applicability)
					251	

新規化学物質で作成したデータベースをラーニングセット、既存化学物質をバリデーションセットとして予測したところ、sensitivity 8.7%、specificity 89.2%と、ほとんどが陰性と予測される結果となった。

MULTICASE は、「生物活性は化学物質が持つ種々の化学的性質の複合的な反映であり、化学的性質は化学構造に基づく」という考え方で、生物活性が予測される。MULTICASE のデータベースを生物活性予測に十分機能させるためには、データベースに入れる物質数がある程度大きく(望ましくは 1000 個以上)し、活性を予測しようとする物質の部分構造をカバーしていることが必要とされる。

データベースに入れる物質数は 1231 個であったため、十分であると考えられる。しかし、このモデルは sensitivity が 8.7% という結果になった。原因として考えられるのは、ラーニングセットとした新規物質がバリデーションセットとした既存物質の部分構造をカバーしていないことであった。

そこで、ケース 2 からは、新規化学物質のみでデータベースを作成し、検討することとした。

#### 3.4.2 ケース 2

新規化学物質 1231 個を分子量を基準に 2 群に分け、それぞれをラーニングセット、バリデーションセットとした。  $0 \leq \text{LOAEL} < 1000 \text{mg/kg}$  を陽性、 $\text{LOAEL} \geq 1000 \text{mg/kg}$  と所見なしを陰性とした。

活性値は Table 3 のように設定した。

Table 3. LOAEL に基づく活性値

LOAEL	活性値
所見なし	10
$1000 \text{mg/kg} \leq \text{LOAEL}$	10
$1 \text{mg/kg} < \text{LOAEL} < 1000 \text{mg/kg}$	$y = -7 \ln(x) + 80$
$\text{LOAEL} \leq 1 \text{mg/kg}$	80

$1 \text{mg/kg} < \text{LOAEL} < 1000 \text{mg/kg}$  の物質の活性値は、 $y = -7 \ln(x) + 80$  で計算すると、約 30~80 となる。

#### 3.4.2.1 パターン 1

データベースを、それぞれ「活性あり：活性なし=2:3」になるように分子量を基準に 2 群に分け、A 群、B 群とした。

所見なしの数が多いので、所見なしのみ分子量の大きい方から順に 361 個削除した。

予測結果は Table 4 と Table 5 の通りであった。

Table 4. A 群の予測結果 (ラーニングセットは B 群)

		Mcaseモデル		判定不能 (Out of Domain or Borderline)	54.7 % (Concordance)	
		+	-			
試験結果	+	174	62	101	11	35.6 % (Sensitivity)
	-	261	71	176	14	67.4 % (Specificity)
		435	133	277	25	100 % (Applicability)
					435	

Table 5. B 群の予測結果 (ラーニングセットは A 群)

		Mcaseモデル		判定不能 (Out of Domain or Borderline)	57 % (Concordance)	
		+	-			
試験結果	+	173	77	81	15	44.5 % (Sensitivity)
	-	262	74	171	17	65.3 % (Specificity)
		435	151	252	32	100 % (Applicability)
					435	

B 群をラーニングセット、A 群をバリデーションセットとして予測したところ、sensitivity 35.6%、specificity 67.4%となった。

A 群をラーニングセット、B 群をバリデーションセットとして予測したところ、sensitivity 44.5%、specificity 65.3%となった。

新規化学物質で作成したデータベースをラーニングセット、既存化学物質をバリデーション

セットとして予測した場合と比較すると、sensitivity は大幅に上がった。

#### 3.4.2.2 パターン 2

A 群、B 群に、削除した所見なし 361 個を分子量を基準にして半分ずつ加え、C 群、D 群とした。

予測結果は Table 6 と Table 7 の通りであった。

Table 6. C 群の予測結果 (ラーニングセットは D 群)

		Mcaseモデル		判定不能 (Out of Domain or Borderline)	63 % (Concordance)	
		+	-			
試験結果	+	174	40	124	10	23 % (Sensitivity)
	-	442	71	348	23	78.7 % (Specificity)
		616	111	472	33	100 % (Applicability)
					616	

Table 7. D 群の予測結果 (ラーニングセットは C 群)

		Mcaseモデル		判定不能 (Out of Domain or Borderline)	62.8 % (Concordance)	
		+	-			
試験結果	+	173	42	121	10	24.3 % (Sensitivity)
	-	442	75	344	23	77.8 % (Specificity)
		615	117	465	33	100 % (Applicability)
					615	

D 群をラーニングセット、C 群をバリデーションセットとして予測したところ、sensitivity 23%、specificity 78.7%となった。

C 群をラーニングセット、D 群をバリデーション

セットとして予測したところ、sensitivity 24.3%、specificity 77.8%となった。

陰性のデータのみを増やすと、陰性とみなす部分構造が増えるためか、「活性あり：活性なし = 2 : 3」になるように 2 群に分けた場合より、



sensitivity は低くなった。

子量が大い方から 89 個ずつ削除し、E 群、F 群とした。

3.4.2.3 パターン 3

「活性あり：活性なし=1：1」になるように、A 群、B 群から所見なしのものを、それぞれ分

予測結果は Table 8 と Table 9 の通りであった。

Table 8. E 群の予測結果 (ラーニングセットは F 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)	61.6 % (Concordance)	
試験結果	+	174	100	62	12	57.5 % (Sensitivity)
	-	172	43	113	16	65.7 % (Specificity)
		346	143	175	28	100 % (Applicability)
					346	

Table 9. F 群の予測結果 (ラーニングセットは E 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)	59.8 % (Concordance)	
試験結果	+	173	102	53	18	59 % (Sensitivity)
	-	173	58	105	10	60.7 % (Specificity)
		346	160	158	28	100 % (Applicability)
					346	

F 群をラーニングセット、E 群をバリデーションセットとして予測したところ、sensitivity 57.5%、specificity 65.7%となった。

E 群をラーニングセット、F 群をバリデーションセットとして予測したところ、sensitivity 59%、specificity 60.7%となった。

3 パターンの中で、この「活性あり：活性なし=1：1」の sensitivity が一番高くなった。

3.4.2.4 パターン 4

sensitivity が一番高くなった E 群、F 群の活性値の幅を狭めてみることにした。LOAEL<=1 mg/kg の活性値を 80、1 mg/kg <LOAEL の活性値を  $y = -3\ln(x)+80$  にし、G 群、H 群とした。 $y = -3\ln(x)+80$  で計算すると、活性値は約 60~80 となる。

予測結果は Table10 と Table11 の通りであった。

Table 10. G 群の予測結果 (ラーニングセットは H 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)	59 % (Concordance)	
試験結果	+	174	94	69	11	54 % (Sensitivity)
	-	172	48	110	14	64 % (Specificity)
		346	142	179	25	100 % (Applicability)
					346	

Table 11. H の予測結果 (ラーニングセットは G)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)		
試験結果	+	173	113	51	9	65.3 % (Sensitivity)
	-	173	65	95	8	54.9 % (Specificity)
		346	178	146	17	101.5 % (Applicability)
					341	

### 3.4.2.5 パターン 5

E 群、F 群の活性値を以下のように設定することとした。

LOAEL<1000 mg/kg の活性値を全て 80 とし、

I 群、J 群とした。

予測結果は Table12 と Table13 の通りであった。

Table 12. I 群の予測結果 (ラーニングセットは J 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)		
試験結果	+	174	100	64	10	57.5 % (Sensitivity)
	-	172	48	112	12	65.1 % (Specificity)
		346	148	176	22	100 % (Applicability)
					346	

Table 13. J 群の予測結果 (ラーニングセットは I 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)		
試験結果	+	173	114	50	9	65.9 % (Sensitivity)
	-	173	63	99	11	57.2 % (Specificity)
		346	177	149	20	100 % (Applicability)
					346	

活性値を広く分布させるより、所見ありを 80、所見なしを 10 とした方が sensitivity が高くなる結果となった。

### 3.4.3 ケース 3

ケース 2 では、陽性データ 347 個、陰性データ 884 個としてデータベースを作成したが、ケース 3 では陽性データ数を増やすことを目的とし、1000mg/kg ≤ LOAEL を陽性としてデータベースを作成した。1000mg/kg ≤ LOAEL を陽性とした場合、陽性データ 445 個、陰性データ 786 個であった。

新規化学物質 1231 個を分子量を基準に 2 群に分け、それぞれをラーニングセット、バリデーションセットとした。

活性値は、Table 14 のように設定した。

Table 14. LOAEL に基づく活性値

LOAEL	活性値
所見なし	10
1mg/kg < LOAEL	$y = -7\ln(x) + 80$
LOAEL ≤ 1mg/kg	80

1mg/kg < LOAEL < 1000mg/kg の物質の活性値は、 $y = -7\ln(x) + 80$  で計算すると、約 30~80 となる。

#### 3.4.3.1 パターン 1

データベースを、それぞれ「活性あり：活性なし=2:3」になるように分子量を基準に 2 群に分け、A 群、B 群とした。

所見なしの数が多いので、所見なしのみ分子量の大きい方から順に 120 個削除した。

予測結果は Table15 と Table 16 の通りであった。

Table 15. A 群の予測結果 (ラーニングセットは B 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)	56.5 % (Concordance)	
試験結果	+	223	79	132	12	35.4 % (Sensitivity)
	-	333	84	235	14	70.6 % (Specificity)
		556	163	367	26	100 % (Applicability)
					556	

Table 16. B 群の予測結果 (ラーニングセットは A 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)	57.8 % (Concordance)	
試験結果	+	222	81	133	8	36.5 % (Sensitivity)
	-	333	71	240	22	72.1 % (Specificity)
		555	152	373	30	100 % (Applicability)
					555	

B 群をラーニングセット、A 群をバリデーションセットとして予測したところ、sensitivity 35.4%、specificity 70.6%となった。

A 群をラーニングセット、B 群をバリデーションセットとして予測したところ、sensitivity 36.5%、specificity 72.1%となった。

LOAEL $\geq$ 1000 mg/kg を陰性とし、「活性あり：活性なし=2:3」とした場合と比較すると、

sensitivity にあまり変化は見られなかった。

#### 3.4.3.2 パターン 2

A 群と B 群に、削除した所見なし 120 個を分子量を基準にして半分ずつ加え、C 群、D 群とした。

予測結果は Table 17 と Table 18 の通りであった。

Table 17. C 群の予測結果 (ラーニングセットは D 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)	59.7 % (Concordance)	
試験結果	+	223	77	133	13	34.5 % (Sensitivity)
	-	393	82	291	20	74 % (Specificity)
		616	159	424	33	100 % (Applicability)
					616	

Table 18. D 群の予測結果 (ラーニングセットは C 群)

		Mcaseモデル				
		+	-	判定不能 (Out of Domain or Borderline)	58.7 % (Concordance)	
試験結果	+	222	76	131	15	34.2 % (Sensitivity)
	-	393	78	285	30	72.5 % (Specificity)
		615	154	416	45	100 % (Applicability)
					615	

D 群をラーニングセット、C 群をバリデーションセットとして予測したところ、sensitivity 34.5%、specificity 74%となった。

A 群をラーニングセット、B 群をバリデーション

セットとして予測したところ、sensitivity 34.2%、specificity 72.5%となった。

パターン 2 は、1000mg/kg $\leq$ LOAEL を陰性とした場合よりも sensitivity が 10%程度高くなった。

### 3.4.3.3 パターン 3

「活性あり:活性なし=1:1」になるように、A 群、B 群から所見なしのものを、それぞれ分子量が大きい方から 111 個ずつ削除し、E 群、F

群とした。

予測結果は Table19 と Table 20 の通りであった。

Table 19. E 群の予測結果 (ラーニングセットは F 群)

		Mc caseモデル			判定不能 (Out of Domain or Borderline)	52.6 % (Concordance)
		+	-			
試験結果	+	223	110	86	27	49.3 % (Sensitivity)
	-	222	70	124	28	55.9 % (Specificity)
		445	180	210	55	100 % (Applicability)
					445	

Table 20. F 群の予測結果 (ラーニングセットは E 群)

		Mc caseモデル			判定不能 (Out of Domain or Borderline)	52.9 % (Concordance)
		+	-			
試験結果	+	222	101	102	19	45.5 % (Sensitivity)
	-	222	73	134	15	60.4 % (Specificity)
		444	174	236	34	100 % (Applicability)
					444	

F 群をラーニングセット、E 群をバリデーションセットとして予測したところ、sensitivity 49.3%、specificity 55.9%となった。

E 群をラーニングセット、F 群をバリデーションセットとして予測したところ、sensitivity 45.5%、specificity 60.4%となった。

A,B 群、C,D 群と比較すると、sensitivity は一番高くなったが、1000mg/kg ≤ LOAEL を陰性とした場合より低くなった。

### 3.4.3.4 パターン 4

sensitivity が一番高くなった E 群、F 群の活性値の幅を狭めてみることにした。

LOAEL ≤ 1 mg/kg の活性値を 80、1 mg/kg < LOAEL の活性値を  $y = -3\ln(x) + 80$  にし、G 群、H 群とした。 $y = -3\ln(x) + 80$  で計算すると、活性値は約 60~80 となる。

予測結果は Table21 と Table22 の通りであった。

Table 21. G 群の予測結果 (ラーニングセットは H 群)

		Mc caseモデル			判定不能 (Out of Domain or Borderline)	53.3 % (Concordance)
		+	-			
試験結果	+	223	115	92	16	51.6 % (Sensitivity)
	-	222	78	122	22	55 % (Specificity)
		445	193	214	38	100 % (Applicability)
					445	

Table 22. H の予測結果 (ラーニングセットは G)

		Mc caseモデル			判定不能 (Out of Domain or Borderline)	54.3 % (Concordance)
		+	-			
試験結果	+	222	107	107	8	48.2 % (Sensitivity)
	-	222	75	134	13	60.4 % (Specificity)
		444	182	241	21	100 % (Applicability)
					444	