

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)に関する研究

平成19年度 総括・分担研究報告書

主任研究者 林 真

平成20(2008)年3月

目 次

I. 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)に関する研究	1
--	---

林 真、江馬 眞、鎌田 栄一、広瀬 明彦

II. Appendices

Appendix I Allyl ethyl ether のほ乳類培養細胞を用いる染色体異常試験	29
Appendix II Diallyl carbonate のほ乳類培養細胞を用いる染色体異常試験	47
Appendix III Propylene sulfide のほ乳類培養細胞を用いる染色体異常試験	61
Appendix IV Azoxybenzene のほ乳類培養細胞を用いる染色体異常試験	79
Appendix V チオベンズアミド、 <i>N</i> -フェニルチオベンズアミド、アリルベンゼン 染色体異常試験実施報告書	93
Appendix VI Chlorthiamid のほ乳類培養細胞を用いる染色体異常試験	101
Appendix VII <i>N</i> -allyl aniline : ほ乳類培養細胞を用いる染色体異常試験	117
Appendix VIII <i>N</i> -allylmethyl amine : ほ乳類培養細胞を用いる染色体異常試験	127
Appendix IX Leadscape Prediction Data Miner を用いた肝毒性予測モデルの構築	137
III. 研究成果の刊行に関する一覧表	193
IV. 研究成果の刊行物・別刷	197

平成 19 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)に関する研究

I. 総括・分担研究報告書

平成 19 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)に関する研究

主任研究者 林 真 国立医薬品食品衛生研究所 変異遺伝部 部長
分担研究者 江馬 眞 国立医薬品食品衛生研究所 総合評価研究室 室長
分担研究者 鎌田 栄一 国立医薬品食品衛生研究所 総合評価研究室 主任研究官
分担研究者 広瀬 明彦 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 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 秋山 浩子 (株)シーティシー・ラボラトリーシステムズ
研究協力者 石和 俊 (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：2003 年の化審法改正により 1t から 10t 未満の物質(低生産量物質)については、分解性・蓄積性・既知見に基づく審査が行われることとなった。この低生産量物質の申請数は、スクリーニング試験結果を行った物質を大きく上回っており、低生産量であってもその毒性を、ヒト健康の上から把握する必要がある。動物愛護の観点からも動物を使用せずに毒性の強さを把握する最良な方法として(定量的)構造活性相関((Q)SAR)が最有力候補であり、低生産量物質のみならず数十万種ともいわれている既存化学物質の評価にも役立てることが出来る。

化審法での申請時に要求されるスクリーニング試験[AMES 試験、染色体試験、28 日間反復投与毒性試験]の内、AMES 試験及び染色体試験については評価方式の異なる 3 種類の市販の(Q)SAR モデルを組み合わせることによる予測の決定樹を提案しており、本研究では、その精度を上げるためのモデルの改良を行ってきている。ADMEWORKS システムの改良の一環として、Ames よび染色体異常予測モデルにおいて SVM 法を用いた場合に比較的 concordance の高いモデルを作成することができた。また、染色体試験に関しては、6 つの化合物グループに分類される 10 化合物について染色体異常試験を行い、DEREK における Chromosome damage alert の適用範囲の拡大もしくは改良を行うための有用な知見を収集することができた。

反復投与毒性試験については、DEREK における肝および腎毒性における alert の検索を行い、肝毒性および腎毒性について各々 11 種の新しい Alert 候補を抽出することができ、約 30%の Sensitivity の増加が見込まれることが示された。ADMEWORKS システムでは肝重量を指標とした場合の SVM 法により比較的高い concordance のモデルを作成することができた。さらに QSAR 構築ツール Leadscape Prediction Data Miner を用いた肝毒性予測モデル構築では、モデルを相補的に組み合わせることやモデルが予測する計算値の閾値を設定することにより availability を制限し、予測精度を高くできる可能性が示された。

A. 目的

昭和 48 年 10 月 16 日に「化学物質の審査及び製造等の規制に関する法律」（化審法）が制定され、平成 15 年に化審法改正が行われた。それに関連して、衆参両議院の経済産業委員会は化審法改正に対する付帯決議で「既存化学物質の安全性点検については、・・・その有害性評価の計画的推進を図ること」と明文化された。しかし、1 化合物のスクリーニング試験を実施するにあたり約 1 億 8 千～2 億 4 千万円の費用が必要であり、時間的にも最低で半年を必要としている。未点検の既存化学物質を国内の委託ラボの総力を挙げて安全性試験を実施したとしても年間約 100 物質が限度である。従って、既存化学物質の全てについて試験を行うと、単純計算で 200 年かかることになる。

OECD の高生産量化学物質安全性点検計画 (HPV) においては、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の協力で行うもので、1993 年度から 3 年間で 154 物質について安全性評価に必要な試験を各国の分担で行うことで開始された。その後、ICCA (国際化学工業協会協議会) もこの取り組みに参画し、2002 年 11 月の時点で 350 以上の化学物質の評価を行うことを表明している。この様に国際協力の下に既存化学物質の安全性評価済み物質を加算しても全ての物質の評価作業が終了するまでにはほど遠い時間と費用がかかる。

この問題の解決策の一案として構造活性相関 (SAR) が考慮されている。信頼される SAR モデルが開発されれば、多額の費用と時間がかかる安全性試験を行わなくても、その評価を行うことが出来る。また更に、これらの SAR モデルを使用して大量の物質数から安全性試験を実施しなければいけない物質選定も可能となる。

更に、平成 15 年の化審法改正により、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、既存の毒性情報に基づき審査が行われることとなった。しかし、この低生産量物質の申請数は、ス

クリーニング試験結果を行った物質を大きく上回っている。低生産量であっても、その毒性をヒト健康の上からは把握することが望ましく、効率よく評価を行うことが期待される。また、動物愛護の観点からも動物を使用せずに毒性の強さを把握できる最良な方法として(定量的)構造活性相関((Q)SAR)が有力候補である。

本研究では主に反復投与試験について、その((Q)SAR モデルの開発及び検討を行う。つまり、昨年度までの第一期の成果を踏まえて汎用性のある毒性指標についてモデルを作成し、反復投与毒性結果の為の決定樹作製を行う。AMES 試験および染色体異常試験については、更なるモデルの改良を行い、個々の(Q)SAR モデルの性能を上げると共に、信頼性と汎用性の高いモデルに基づく決定樹を再構築する。これらのモデルを組み合わせることにより、化審法で求められているスクリーニング毒性情報を効率よく提供することができ、安全性に対する危機管理体制の構築を目的とする。また、既存化学物質については毒性予測をデータベース(DB)化することで実試験実施における優先順位付けの情報提供を行う。

B. 方法

ADMEWORKS モデルの改良として、VM 法、KNN 法および ADA 法の 3 つの手法を用いての Ames 変異原性予測モデルの改良、および染色体異常予測システムの作成を試みた。

DEREK における染色体異常アラートの適用範囲の拡大を行うために、6 つの化合物グループに分類される 10 化合物について染色体異常試験を行った。

反復投与毒性試験については、反復投与毒性試験については、DEREK には反復投与毒性の予測モジュールがなく、MCase の場合にも、モジュールがヒト肝臓を標的としたデータベースであるためラットを用いた結果に対しては Sensitivity が非常に低い値しか得ることができないことが判明していた。そのため、“肝毒性物質”および“腎毒性物質”として、これら組織の病理組織学変化や

様々な関連指標（肝重量、血中生化学検査値）に対しての最小毒性量をまとめたデータベースを作成し、これらのデータを用いた Lhasa 社との共同研究により DEREK における肝および腎毒性における alert の検索を行った。さらに同データベースを用いて、Leadscope Prediction Data Miner と ADMEWORKS ModelBuilder による肝毒性予測モデル構築の構築を試みた。

倫理面への配慮

本年度の研究においては、*in vivo* およびヒト材料を扱う試験は実施していないことから、動物愛護およびヒトに対する倫理的な問題が生ずる可能性はない。

C. 結果と考察

化審法での申請時に要求されるスクリーニング試験[AMES 試験、染色体試験、28 日間反復投与毒性試験]の内、17 年度までの研究により、AMES 試験及び染色体試験については評価方式の異なる 3 種類の市販の(Q)SAR モデルを組み合わせることによる予測の決定樹を提案しており、AMES 試験については化審法調査会でその予測結果を参考資料として公表している。18 年度の研究では Ames 試験予測モデルに比較して精度の低い染色体異常予測モデルについての改良を検討し、DEREK でのアラート追加や ADMEWORKS 内の単独モデル改良を試みたが、さらなる検討の必要性が示された。そこで、19 年度は、DEREK における Chromosome damage alert の適用範囲の拡大もしくは改良を行うため、6 つの化合物グループに分類される 10 化合物について染色体異常試験を行い、一部を除き陽性結果を得ることができ、アラートの作成および改良に有用な知見を収集することができた。ADMEWORKS に値しての改良に関しては、まず、ModelBuilder を用いて 372 化学的パラメータと 1600 以上の部分構造パラメータを発生させ、モデル作成のためのクラスの特異的パラメータとして 43 件の部分構造及び化学的パラメータを選定した。これらを利用して SVM 法、KNN 法および ADA 法の 3 つの手法に対して、

高 sensitivity モデルと高 specificity になるように調整したモデルを各々作成し、さらに、分子量を 200 で区切った場合のモデルの精度の比較を検討した、その結果 SVM 法を用いた場合に比較的 concordance の高いモデルを作成することができることが示された。

反復投与毒性試験については、DEREK には反復投与毒性の予測モジュールがなく、MCASE の場合にも、モジュールがヒト肝臓を標的としたデータベースであるためラットを用いた結果に対しては Sensitivity が非常に低い値しか得ることができないことが判明していた。そのため、肝臓および腎臓の病理組織学変化を引き起こした物質をそれぞれ“肝毒性物質 (hepatotoxic compounds)”および“腎毒性物質 (nephrotoxic compounds)”として、この病理組織学変化に関して得られた最小毒性量 (lowest observed adverse effect level: LOAEL)を用いて予測を試みた。さらに、肝重量、血中 GOT/GPT およびコレステロール/トリグリセリドレベルを“肝毒性指標 (liver indicators)”、腎臓重量および BUN/CRN レベルを“腎毒性指標 (kidney indicators)”としてこれらの値の増加に関する LOAEL を用いた検討も行った。これらのデータを用いた Lhasa 社の共同研究により DEREK における肝および腎毒性における alert の検索を行うことを目的として、肝毒性および腎毒性について各々 11 種（4 種は肝臓と腎臓で重複）の新しい Alert 候補を抽出することができた。これらのアラートを用いた場合には、それぞれ約 30% の Sensitivity の増加が見込まれることが示された。ADMEWORKS システムを用いて、GOT、Cho および肝重量の増加や肝臓の病理変化の有無を指標として予測モデルの構築を試みた結果、肝重量を指標として場合に SVM 法により比較的高い concordance のモデルを作成することができた。さらに QSAR 構築ツール Leadscope Prediction Data Miner を用いて肝毒性予測モデルを構築した。Leadscope モデルについても sensitivity の高いモデルと specificity の高いモデルを別々に構築し、両モデルを相補的に組み合わせることやモデル

が予測する際の計算値の閾値を設定することにより availability を制限し、予測精度を高くできる可能性が示された。

1. 染色体試験

1.1. 方法

DEREK における Chromosome damage alert の適用範囲の拡大もしくは改良を行うため、6 つの化合物グループ (O-Allyl compounds, Episulphides, Azoxy compounds, Azoxy compounds, Allylbenzenes, N-Allyl compounds) に分類される 10 化合物について染色体異常試験を行った。

1.2. 結果及び考察

O-Allyl compounds については、このアラート構造は rapid prototype alert である CD98 から発展したが、この alert は既存の in vitro 染色体異常試験データに基づいたものであり、その範囲は allyl ester 類および allyl alcohol に限られている。Allyl ethyl ether および diallyl carbonate について得られた陽性結果は、allyl ether 類及び carbonate 類の activity を裏付けている。推定されるメカニズムとしては、エポキシ化もしくはアリル酸化およびその後の求電子性中間体の形成が考えられる。そこで、Allyl ethyl ether と diallyl carbonate について染色体異常試験を行ったところ、両物質とも陽性物質であるという結果を得た (Appendix 1&2 参照)。

Episulphides に関しては、Propylene sulfide が in vitro 染色体異常試験で陽性結果を示せば、episulphides 構造が求電子性の官能基であることにより、一般的に active であることを裏付けることになると考えられる。この点に関して Propylene sulfide に対する染色体異常を行ったところ、染色体数の異常誘発能を有さないが、染色体構造異常誘発能を有する陽性の結果が得られた (Appendix 3 参照)。

Azoxy compounds に関しては、染色体異常試験データ集にて methylazoxymethanol acetate が陽性結果を示していることより、おそらく還元及び芳香族アミン類の生成メカニズムにより、azoxy 化合物が一般的に active であると考えられ、その代

表として Azoxybenzene が in vitro 染色体異常試験で陽性結果を示すことが明らかになれば、この構造アラートの妥当性を裏付けることが可能になると考えられた。この点に関して Azoxybenzene の染色体異常試験を行ったところ、短時間処理法 +S9 処理において、染色体構造異常の誘発が認められた (Appendix 4 参照)。

Aromatic thioamides について、染色体異常試験データ集にて ethionamide が陽性結果を示していることより、この芳香族チオアミド類が、その硫酸化および求電子性中間体の生成メカニズムにより一般的に active であると考えられ、Thiobenzamide, chlorthiamid および N-phenylthiobenzamide について、in vitro 染色体異常試験を行った。その結果、Thiobenzamide は、倍数性細胞を誘発しなかったが、濃度依存的に染色体の構造異常を誘発した (Appendix 5 参照)。一方、Thiobenzamide のアミノ基にベンゼン環が結合した N-phenylthiobenzamide は構造異常を誘発せず、分析した最高濃度でのみ倍数性細胞を誘発した (Appendix 5 参照)。しかしながら、N-phenylthiobenzamide は、培養液中で針状の沈殿を形成したことから、物理的な細胞分裂阻害により倍数性細胞が誘発された可能性があると考えられた。また、chlorthiamid は構造異常および倍数性細胞ともに誘発しなかった (Appendix 6 参照)。

Allylbenzenes に関しては、染色体異常試験データ集において、safrole と eugenol の陽性結果が得られていることより、ベンジル位の酸化および求電子性中間体の生成メカニズムにより、allylbenzene クラスが一般的に active であることを示していると考えられた。この点に関して Allylbenzene の染色体異常試験を行ったが、構造異常および倍数性細胞ともに誘発しない陰性の結果が得られた (Appendix 5 参照)。

N-Allyl compounds については、染色体異常試験データ集において、triallyl isocyanurate の陽性結果が得られていることより、エポキシ化やアリル基の酸化による求電子中間体の生成メカニズムによる N-allyl 化合物が一般的に active であることを

示していると考えられた。この点に関して、N-Allylaniline と N-allylmethylamine の染色体異常試験を行ったところ、N-allyl aniline は代謝活性化系の有無にかかわらず構造的染色体異常を誘発することが確認された(Appendix 7 参照)が、N-allylmethylamine では、代謝活性化系の有無にかかわらず構造的染色体異常を誘発しない陰性の結果を得た(Appendix 8 参照)。

以上の結果、O-Allyl compounds、Episulphides および Azoxy compounds についてはその構造アラートの補強が行われ、Aromatic thioamides と N-Allyl compounds のアラートについては、予測性を改良するための知見を得ることができたと考えられた。

2. ADMEWORKS の改良

2.1. Ames 変異原性予測モデルの改良

2.1.1 データの整理

Ames 試験結果データ (化学構造と実測値) の中から、以下の手順に従って、解析に相応しくないデータを除外して、データの整理を行った。

①実測値がない、あるいは実測値が明らかでないデータの除外

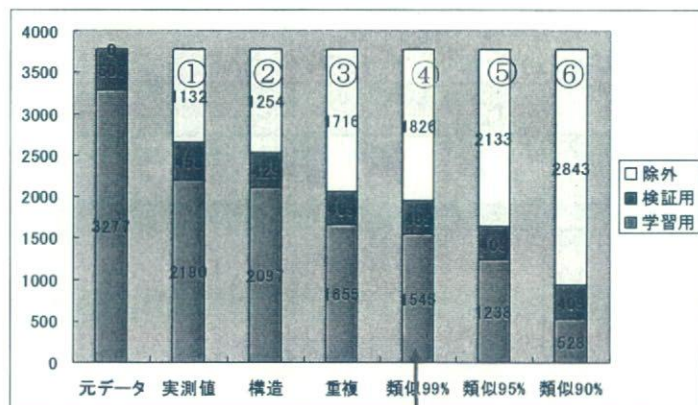
②金属系の原子、イオン、塩、壊れている構造の除外

③重複しているデータ、特に実測値が矛盾しているデータの除外

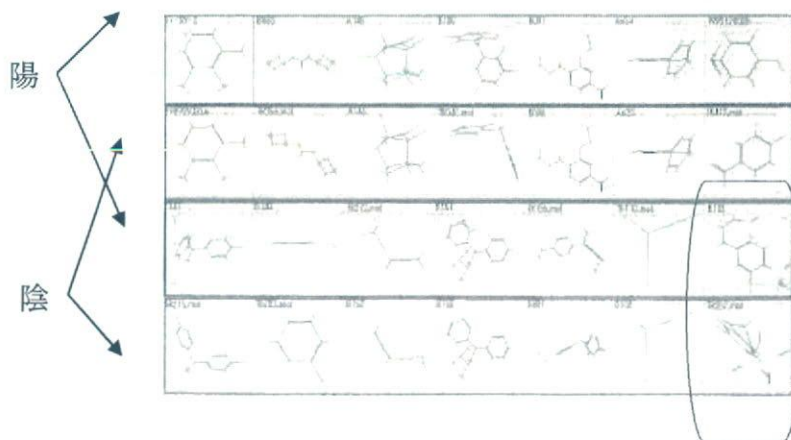
④検証用母集団と類似 (99%以上) しているデータの除外

⑤検証用母集団と類似 (95%以上) しているデータの除外

⑥検証用母集団と類似 (90%以上) しているデータの除外



実際の解析に使用した母集団



2.1.2 パラメータ作成

ADMEWORKSの開発PKGであるModelBuilderを用いて 372 化学的パラメータと 1600 以上の部分構造パラメータを発生した。

2.1.3 特徴・サンプル抽出

372 化学的パラメータの中から欠損値、ゼロ、高い相関のパラメータを除外してから、AdaBoostという手法によって、有効なパラメータ 6 件に絞った。また 1600 部分構造パラメータも、クラスの特異的パラメータ（片方のクラスにほとんどゼロ値しかない）を 28 件に絞った（詳細は別紙を参照）。

2.1.4 モデル作成

28 件の部分構造パラメータのみを利用して、高 sensitivity と高 specificity、2 本ずつのモデルを下記アルゴリズムに対して作成した。

また化学的パラメータと部分構造パラメータの両方を利用して、さらに SVM による 1 本のモデルを作成した AMES_SVM2。

2.1.5 モデルの検証

最終的には外部検証用母集団（409 件）の中、NIHS 検証用母集団（227 件）に対して最終的な外部検証を行った

モデル手法	高 Specificity モデル	高 Sensitivity モデル
SVM	AMES_CL_SVM_191	AMES_SVM_FN
KNN	AMES_CL_KNN_2	AMES_KNN_FN
ADA	AMES_ADA	AMES_ADA_FN

1) Ames予測モデルの外部検証結果

		AMES_C L_KNN_2	L_SVM_1 91	AMES_A DA	AMES_S VM_FN	AMES_K NN_FN	AMES_A DA_FN	AMES_S VM2
TOTAL	227	227	227	227	227	227	227	227
OK	227	225	225	225	225	225	225	225
NG	0	2	2	2	2	2	2	2
MW=ALL								
TP	43	11	15	10	39	38	40	22
FP	0	13	17	26	126	133	133	31
TN	184	169	165	156	56	49	49	151
FN	0	32	28	33	4	5	3	21
ALL	227	225	225	225	225	225	225	225
GRAY	0	0	0	0	0	0	0	0
GRAY RA*	0	0	0	0	0	0	0	0
Concordance = (TP+TN)/ALL		0.80	0.80	0.74	0.42	0.39	0.40	0.77
Specificity = TN/(TN+FP)		0.93	0.91	0.86	0.31	0.27	0.27	0.83
Sensitivity = TP/(TP+FN)		0.26	0.35	0.23	0.91	0.88	0.93	0.51
MW<=	200							
TP	33	9	11	7	31	29	30	17
FP	0	6	9	21	103	105	106	25
TN	128	120	117	105	23	21	20	101
FN	0	24	22	26	2	4	3	16
ALL	161	159	159	159	159	159	159	159
GRAY	0	0	0	0	0	0	0	0
GRAY RA*	0	0	0	0	0	0	0	0
Concordance		0.81	0.81	0.70	0.34	0.31	0.31	0.74
Specificity		0.95	0.93	0.83	0.18	0.17	0.16	0.80
Sensitivity		0.27	0.33	0.21	0.94	0.88	0.91	0.52
MW>	200							
TP	10	2	4	3	8	9	10	5
FP	0	7	8	5	23	28	27	6
TN	56	49	48	51	33	28	29	50
FN	0	8	6	7	2	1	0	5
ALL	66	66	66	66	66	66	66	66
GRAY	0	0	0	0	0	0	0	0
GRAY RA*	0	0	0	0	0	0	0	0
Concordance		0.77	0.79	0.82	0.62	0.56	0.59	0.83
Specificity		0.88	0.86	0.91	0.59	0.50	0.52	0.89
Sensitivity		0.20	0.40	0.30	0.80	0.90	1.00	0.50

2-2 染色体異常予測モデルの作成

2.2.1 データの整理

染色体異常結果データ（化学構造と実測値）の中から、以下に示す手順で、解析に相応しくないデータを除外して、データの整理を行った。

①実測値がない、あるいは実測値が明らかでないデータの除外

②金属系の原子、イオン、塩、壊れている構造の除外

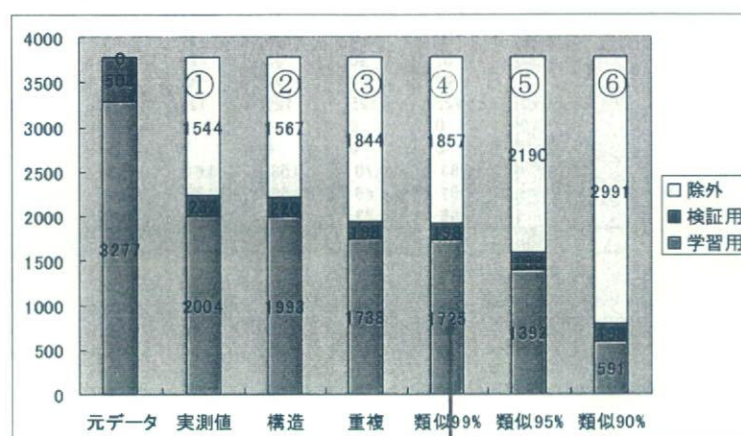
③重複しているデータ、特に実測値が矛盾して

いるデータの除外

④検証用母集団と類似（99%以上）しているデータの除外

⑤検証用母集団と類似（95%以上）しているデータの除外

⑥検証用母集団と類似（90%以上）しているデータの除外



実際の解析に使用した母集団

2.2.2 パラメータ作成

ModelBuilderを用いて 372 化学的パラメータと 1600 以上の部分構造パラメータを発生した。

2.2.3 特徴・サンプル抽出

372 化学的パラメータの中から欠損値、ゼロ、高い相関のパラメータを除外してから、AdaBoostという手法によって、有効なパラメータ 10 件に絞った。また 1600 部分構造パラメータも、クラスの特異的パラメータ（片方のクラスにほとんど

ゼロ値しかない）を 38 件に絞った。合計 48 パラメータを合わせて、相関の高いパラメータを除外したら 43 件に絞った（詳細は別紙を参照）。

2.2.4 モデル作成

43 件の部分構造及び化学的パラメータを利用して、高 sensitivity と高 specificity、2 本ずつのモデルを下記アルゴリズムに対して作成した。

モデル手法	高 Specificity モデル	高 Sensitivity モデル
SVM	CA_SVM_FP	CA_SVM_FN2
KNN	CA_KNN_FP	CA_KNN_FN
ADA	CA_ADA_FP	CA_ADA_FN

2.2.5 モデルの検証

最終的な外部検証を行った。

最終的には外部検証用母集団（194 件）に対して

2) 染色体異常予測モデルの外部検証結果

		CA_SVM_ FP	CA_KNN_ FP	CA_ADA_ FP	CA_SVM_ FN2	CA_KNN_ FN	CA_ADA_ FN
TOTAL	194	194	194	194	194	194	194
OK	194	194	194	194	194	194	194
NG	0	0	0	0	0	0	0
MW=ALL							
TP	81	48	54	47	59	64	68
FP	0	29	37	41	52	63	65
TN	113	84	76	72	61	50	48
FN	0	33	27	34	22	17	13
ALL	194	194	194	194	194	194	194
GRAY	0	0	0	0	0	0	0
GRAY RA ⁺	0	0	0	0	0	0	0
Concordance = (TP+TN)/ALL		0.68	0.67	0.61	0.62	0.59	0.60
Specificity = TN/(TN+FP)		0.74	0.67	0.64	0.54	0.44	0.42
Sensitivity = TP/(TP+FN)		0.59	0.67	0.58	0.73	0.79	0.84
MW<=	200						
TP	64	42	47	41	52	55	57
FP	0	24	21	29	37	41	46
TN	61	37	40	32	24	20	15
FN	0	22	17	23	12	9	7
ALL	125	125	125	125	125	125	125
GRAY	0	0	0	0	0	0	0
GRAY RA ⁺	0	0	0	0	0	0	0
Concordance		0.63	0.70	0.58	0.61	0.60	0.58
Specificity		0.61	0.66	0.52	0.39	0.33	0.25
Sensitivity		0.66	0.73	0.64	0.81	0.86	0.89
MW>	200						
TP	17	6	7	6	7	9	11
FP	0	5	16	12	15	22	19
TN	52	47	36	40	37	30	33
FN	0	11	10	11	10	8	6
ALL	69	69	69	69	69	69	69
GRAY	0	0	0	0	0	0	0
GRAY RA ⁺	0	0	0	0	0	0	0
Concordance		0.77	0.62	0.67	0.64	0.57	0.64
Specificity		0.90	0.69	0.77	0.71	0.58	0.63
Sensitivity		0.35	0.41	0.35	0.41	0.53	0.65

2.3 反復投与毒性試験（肝毒性）予測モデル作成

2.3.1 データの整理

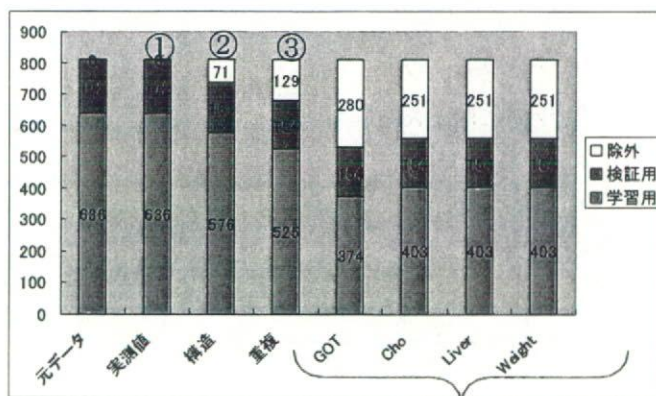
反復投与結果データ（化学構造と実測値）の中から、以下に示す手順で、解析に相応しくないデータを除外して、データの整理を行った。

①実測値がない、あるいは実測値が明らかでないデータの除外

②金属系の原子、イオン、塩、の除外

③重複しているデータの除外

また下記4種類の実験データ（GOT、Cho、Liver、Weight）に関して、学習母集団を別々に用意した。



実際の解析に使用した母集団

2.3.2 パラメータ作成

ModelBuilderを用いて372化学的パラメータと1600以上の部分構造パラメータを発生した。

2.3.3 特徴・サンプル抽出

372 化学的パラメータの中から欠損値、ゼロ、高い相関のパラメータを除外してから、AdaBoostという手法によって、有効なパラメータに絞った。また 1600 部分構造パラメータも、クラスの特異

的パラメータ（片方のクラスにほとんどゼロ値しかない）に絞った。（4 種類の実験データ（GOT、Cho、Liver、Weight）に関して、最終的に絞り込まれたパラメータは結果表を参照。）

2.3.4 モデル作成

2.3.3 で絞り込んだパラメータを利用して、各4 種類の実験データに対して、下記のようにモデルを作成した。

モデル手法	GOT モデル	Cho モデル	Liver モデル	Weight モデル
SVM	GOT_SVM_1	cho_SVM_1	liver_SVM_FN	WT_SVM
KNN	GOT_KNN_1	cho_KNN_1	liver_KNN_FN	WT_KNN
ADA	GOT_ADA_3	cho_ADA_1	liver_ADA_FN	WT_ADA

2.3.5 モデルの検証

て外部検証を行った。

最終的には外部検証用母集団（154 件）に対し

3) 反復毒性試験予測モデルの外部検証結果(1)

	GOT_SV M_1	GOT_KN N_1	GOT_AD A_3	cho_SVM _1	cho_KNN _1	cho_ADA _1
TOTAL	154	154	154	154	154	154
OK	154	154	154	154	154	154
NG	0	0	0	0	0	0
MW=ALL						
TP	20	13	20	34	34	14
FP	49	23	50	46	46	20
TN	71	97	70	60	60	86
FN	14	21	14	14	14	34
ALL	154	154	154	154	154	154
GRAY	0	0	0	0	0	0
GRAY RATE	0	0	0	0	0	0
Concordance = (TP+TN)/ALL	0.59	0.71	0.58	0.61	0.61	0.65
Specificity = TN/(TN+FP)	0.59	0.81	0.58	0.57	0.57	0.81
Sensitivity = TP/(TP+FN)	0.59	0.38	0.59	0.71	0.71	0.29
MW<=200						
TP	17	11	19	28	29	12
FP	38	19	39	39	38	15
TN	47	66	46	37	38	61
FN	14	20	12	12	11	28
ALL	116	116	116	116	116	116
GRAY	0	0	0	0	0	0
GRAY RATE	0	0	0	0	0	0
Concordance	0.55	0.66	0.56	0.56	0.58	0.63
Specificity	0.55	0.78	0.54	0.49	0.50	0.80
Sensitivity	0.55	0.35	0.61	0.70	0.73	0.30
MW>200						
TP	3	2	1	6	5	2
FP	11	4	11	7	8	5
TN	24	31	24	23	22	25
FN	0	1	2	2	3	6
ALL	38	38	38	38	38	38
GRAY	0	0	0	0	0	0
GRAY RATE	0	0	0	0	0	0
Concordance	0.71	0.87	0.66	0.76	0.71	0.71
Specificity	0.69	0.89	0.69	0.77	0.73	0.83
Sensitivity	1.00	0.67	0.33	0.75	0.63	0.25

3) 反復毒性試験予測モデルの外部検証結果(2)

	liver_SV M_FN	liver_KNN _FN	liver_ADA _FN	WT_SVM	WT_KNN	WT_ADA
TOTAL	154	154	154	154	154	154
OK	154	154	154	154	154	154
NG	0	0	0	0	0	0
MW=ALL						
TP	57	56	53	73	61	75
FP	40	38	44	1	22	21
TN	42	44	38	66	45	46
FN	15	16	19	14	26	12
ALL	154	154	154	154	154	154
GRAY	0	0	0	0	0	0
GRAY RATE	0	0	0	0	0	0
Concordance = (TP+TN)/ALL	0.64	0.65	0.59	0.90	0.69	0.79
Specificity = TN/(TN+FP)	0.51	0.54	0.46	0.99	0.67	0.69
Sensitivity = TP/(TP+FN)	0.79	0.78	0.74	0.84	0.70	0.86
MW<=200						
TP	45	48	45	56	50	60
FP	33	32	37	1	16	18
TN	28	29	24	46	31	29
FN	10	7	10	13	19	9
ALL	116	116	116	116	116	116
GRAY	0	0	0	0	0	0
GRAY RATE	0	0	0	0	0	0
Concordance	0.63	0.66	0.59	0.88	0.70	0.77
Specificity	0.46	0.48	0.39	0.98	0.66	0.62
Sensitivity	0.82	0.87	0.82	0.81	0.72	0.87
MW>200						
TP	12	8	8	17	11	15
FP	7	6	7	0	6	3
TN	14	15	14	20	14	17
FN	5	9	9	1	7	3
ALL	38	38	38	38	38	38
GRAY	0	0	0	0	0	0
GRAY RATE	0	0	0	0	0	0
Concordance	0.68	0.61	0.58	0.97	0.66	0.84
Specificity	0.67	0.71	0.67	1.00	0.70	0.85
Sensitivity	0.71	0.47	0.47	0.94	0.61	0.83

3. DEREK によるアラート抽出

3.1 方 法

反復投与毒性予測のための(Q)SAR モデルの開発

上記 3 つの (Q)SAR モデルのうち、DEREK について、778 化合物に関する反復投与毒性試験データを用い、肝毒性および腎毒性予測モデルの開発のための検討を行った。本研究では、特に肝臓および腎臓の病理組織学変化を引き起こした物質をそれぞれ“肝毒性物質 (hepatotoxic compounds)”および“腎毒性物質 (nephrotoxic compounds)”として、この病理組織学変化に関して得られた最小毒性量 (lowest observed

adverse effect level: LOAEL)を用いて予測を試みた。さらに、肝重量、血中 GOT/GPT およびコレステロール/トリグリセリドレベルを“肝毒性指標 (liver indicators)”、腎臓重量および BUN/CRN レベルを“腎毒性指標 (kidney indicators)”としてこれらの値の増加に関する LOAEL を用いた検討も行った。本研究で対象とした反復投与毒性試験データをそれぞれの指標について分類した結果を表 1 に示す。これらいずれの指標についても変化が見られなかった場合には陰性 (negative)とした。

表 1. 肝毒性および腎毒性指標に基づいた試験結果分類

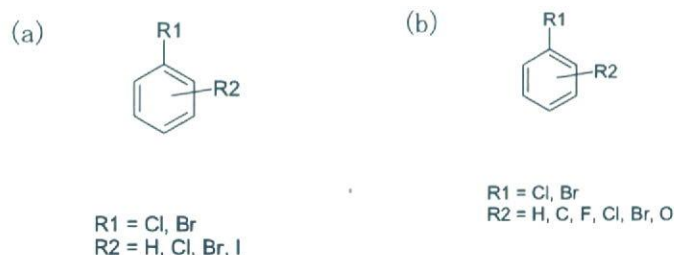
Liver/Kidney indicators	Experimental active	Experimental inactive
Hepatotoxicity		
Histopathology	293	515
Histopathology (LOAEL: <250 mg/kg)	154	654
Organ weight	378	430
GOT/GPT	193	615
Cholesterol/Triglyceride	213	595
Nephrotoxicity		
Histopathology	218	590
Histopathology (LOAEL: ≤250 mg/kg)	138	670
Organ Weight	252	556
BUN/CRN	125	683

3.2 結果および考察

現在、Lhasa が開発中の alert には bromobenzene および chlorobenzene 類の肝毒性に関するものがある (表 2)。本研究で対照としたデータには bromobenzene 基を含む化合物が 17 種あり、このうち 10 種が肝毒性物質であったが、開発中の alert で正しく予測することが出来たのはこれらのうち 1 物質のみであった (偽陽性判定なし)。一方、この alert の領域を表 12(b) に示したように拡張することにより、すべての肝毒性物質を正しく予測することが出来た。本

研究で対象としたデータには chlorobenzene 類が 52 種あり、そのうち 28 種が肝毒性物質であった。単純構造を持つ chlorobenzene 類については、bromobenzene 類と同様に、表 12(b) のように alert 領域を拡張することで、新たに 20 種の肝毒性物質を正しく予測することが可能となると考えられた。その他の肝毒性化合物については他の官能基がその活性に寄与していると考えられる。このように bromobenzene および chlorobenzene 類の alert 領域を拡張することにより sensitivity を 10% 増加させることが出来た。

表 2. Bromobenzene および Chlorobenzene 類の肝毒性 alert

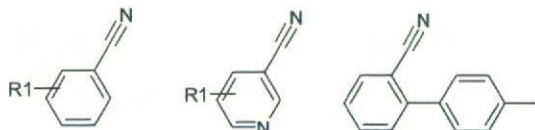


(a) 現在開発中の alert の領域, (b) alert 領域を拡張した場合

対象とした 778 化合物に関する試験結果を用いて、肝毒性に関連している構造部位の特定を試みた。ChemTK 3.1 を用いたところ、有効な構造フラグメントを抽出することが出来な

かったことから、目視による解析を実施した。その結果、以下に示す 11 種の新しい alert を特定することが出来た。

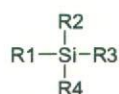
1) Aryl nitriles



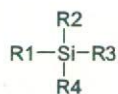
R1 = H, F, Cl, Br, CN

Hepatotoxic compounds: 10
Compounds with liver indicators: 3
Negatives: 3

2) Organosilicon compounds



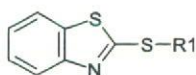
R1, R2 = C
R3, R4 = C, O



R1 = C, O
R2-R4 = O

Hepatotoxic compounds: 10
Compounds with liver indicators: 3
Negatives: 3

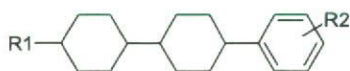
3) Benzothiazole sulphenamides and disulphides



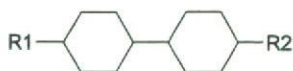
R1 = N, S

Hepatotoxic compounds: 3
Negatives: 2

4) Bicyclohexyl compounds



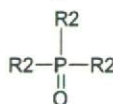
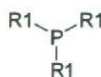
R1 = Et, Pr, =O
R2 = CH3, F, OCF3



R1 = C, =O
R2 = =O, CO2H, Et

Hepatotoxic compounds: 8
Compounds with liver indicators: 3

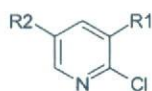
5) Organophosphorus compounds



R1, R2 = C, O

Hepatotoxic compounds: 11
Compounds with liver indicators: 9
Negatives: 7

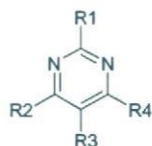
6) Pyridines



R1, R2 = N, CH₂N, CO₂H

Hepatotoxic compounds: 7
Compounds with liver indicators: 5
Negatives: 5

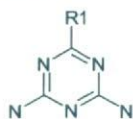
7) Pyrimidines



R1 = NH₂, SMe
R2, R4 = C, Cl, O
R3 = H, Cl

Hepatotoxic compounds: 5

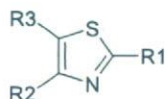
8) Triazines



R1 = C

Hepatotoxic compounds: 5
Compounds with liver indicators: 2
Negatives: 13

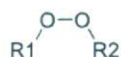
9) Thiazoles



R1 = C, H, Cl
R2, R3 = H, CO₂H, CH₂N, CH₂O

Hepatotoxic compounds: 2
Compounds with liver indicators: 5

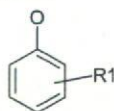
10) Peroxides



R1, R2 = C

Hepatotoxic compounds: 2
Negatives: 1

11) Phenols



R1 = H, C(alkyl)

Hepatotoxic compounds: 21
Compounds with liver indicators: 21
Negatives: 23

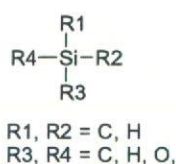
これらの新しい alert を用いた場合、90 種以上の肝毒性物質を正しく予測することが出来た。この結果は、sensitivity の 30% の増加を意味しており、一方で、specificity 96% を維持することが出来た。

DEREK ver. 9 に含まれる知識ベース (Windows 9 Knowledge 2006) を用いた結果、いずれの化学物質についてもその腎毒性も予測することは出来なかった。この知識ベースには雄ラットに特異的な $\alpha 2 \mu$ -globulin 腎症に関する 3 つの alert

(isopentyl group, polyhalogenated benzene, polychlorinated ethane) が含まれていることから、これらの alert を用いて解析を行った結果、17 種が陽性となったが、このうち 8 種は偽陽性 (毒性試験では陰性) であった。Sensitivity は 4.1 % と非常に低い値となった。

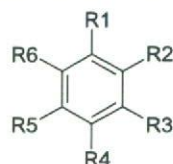
肝毒性同様に、目視による腎毒性関連構造部位の特定を試みた。その結果、以下に示す 11 種の新しい alert を特定することが出来た。

1) Organosilicon compounds



Nephrotoxic compounds: 7
Compounds with kidney indicators: 1
Negatives: 5

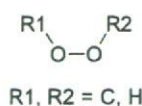
2) Halogenated benzenes



R1 = F, Cl, Br, I
R2-R5 = H, C, F, Cl, Br, I
R6 = H, C, F, Cl, Br, I, N, O, S
Aromatic ring fusions are not allowed

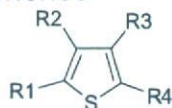
*There are over 70 compounds that contain a simple mono- or poly-halogenated benzene fragment. Approximately half of the compounds in this group show evidence for nephrotoxicity, while similar compounds are also negative.

3) Peroxides



Nephrotoxic compounds: 1
Compounds with kidney indicators: 1
Negatives: 1

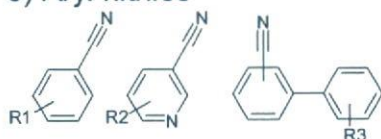
4) Thiophenes



R1-R4 = H, C, O, S, Br, N

Nephrotoxic compounds: 3
Compounds with kidney indicators: 2
Negatives: 3

5) Aryl nitriles



R1, R2 = H, C, F, Br, Cl, CN, O, N
R3 = H, C, O

Nephrotoxic compounds: 10
Negatives: 8

6) Aliphatic nitriles



R1 = C bearing C, H, S, O, N

Nephrotoxic compounds: 6
Compounds with kidney indicators: 2
Negatives: 9

7) Bicyclohexyl compounds



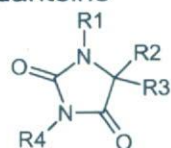
R1 = Et, Pr, =O, allyl
R2 = CH3, F, OCF3



R3 = =O
R4 = =O, CO2H, Et, O

Nephrotoxic compounds: 2, Compounds with kidney indicators: 7, Negatives: 2

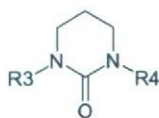
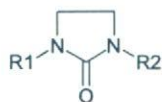
8) Hydantoins



R1, R2, R3, R4 = H, C

Nephrotoxic compounds: 2
Compounds with kidney indicators: 1
Negatives: 2

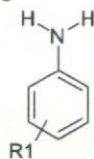
9) Cyclic ureas



R1-R4 = H, C

Nephrotoxic compounds: 3
Compounds with kidney indicators: 1
Negatives: 1

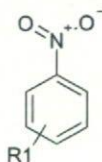
10) Anilines



R1 = C, H, F, Cl, Br, O, S, N, P

Nephrotoxic compounds: 17
Compounds with kidney indicators: 2
Negatives: 14

11) Aromatic nitro compounds



R1 = C, H, Cl, O, S, N

Nephrotoxic compounds: 7
Compounds with kidney indicators: 3
Negatives: 10

これらの新しい alert を用いた場合、およそ 70 種の腎毒性物質を新たに正しく予測することが出来た。この結果は、sensitivity 約 36%の増加を意味しており、一方で、specificity 81%を維持することが出来た。

現時点では、すべてのデータの解析は終了しておらず、今後の調査でさらに肝毒性および腎毒性に関わる alert の特定を進めることにより、より精度を向上させることが出来ると考えられる。

参考文献

Kirkland D, Aardema M, Henderson L, Muller L (2005) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I.

Sensitivity, specificity and relative predictivity, *Mutat.Res.* **584**, 1-256.

祖父尼俊雄 (1999) 「染色体異常試験データ集・改訂 1998 年版」、祖父尼俊雄 監修、林 真、松岡厚子 編集、(株)エル・アイ・シー (LIC)

4. Leadscope Prediction Data Miner を用いた肝毒

性予測モデルの構築

Leadscope 及び Leadscope Predictive Data Miner を用いて、28 日間反復投与毒性試験のデータセットから、肝毒性・腎毒性のように各エンドポイントに対する QSAR 予測モデル構築を行った。

- 前年度の作業で、808件の28日間反復投与毒性試験データの数値（LOAEL値）をExcelに出力し、計算式「 $-\text{Log}(\text{数値})+4$ 」によりデータ値を変換した。
- データがブランクの場合は10000とし、変換後は0になるようにした。
- 変換後に ISIS データベース「NIHS Ames database_New.db」を作成した。

4.1 対象化合物

◆トレーニングデータセット

化合物（808 化合物）

化合物数	808	
データ	構造情報+毒性試験データ	
毒性試験データ	フィールド名	フィールド Type
28 日間済み	check_28day	Integer (1 : 28 日間, 13 : 13 週間, 26 : 26 週間)
神経系	neuro	Real (高値=毒性強い、低値=毒性弱い)
RBC/Hb	RBC/Hb	Real (高値=毒性強い、低値=毒性弱い)
GOT/GPT	GOT/GPT	Real (高値=毒性強い、低値=毒性弱い)
BUN/CRN	BUN/CRN	Real (高値=毒性強い、低値=毒性弱い)
Cho/TG	Cho/TG	Real (高値=毒性強い、低値=毒性弱い)
肝重量	Weight_Liver	Real (高値=毒性強い、低値=毒性弱い)
腎重量	Weight_Kidney	Real (高値=毒性強い、低値=毒性弱い)
肝臓	Liver	Real (高値=毒性強い、低値=毒性弱い)
腎臓	Kidney	Real (高値=毒性強い、低値=毒性弱い)
甲状腺	Thyroid	Real (高値=毒性強い、低値=毒性弱い)
脾臓	Spleen	Real (高値=毒性強い、低値=毒性弱い)
胸腺	Thymus	Real (高値=毒性強い、低値=毒性弱い)
胃	Stomach	Real (高値=毒性強い、低値=毒性弱い)
腸	Gut	Real (高値=毒性強い、低値=毒性弱い)
精巣	Testis	Real (高値=毒性強い、低値=毒性弱い)
心臓	Heart	Real (高値=毒性強い、低値=毒性弱い)
副腎	Adrenal	Real (高値=毒性強い、低値=毒性弱い)
骨髄	bone	Real (高値=毒性強い、低値=毒性弱い)