

## 2.2. DEREK における Chromosome damage alert の改良

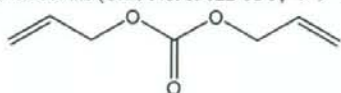
### 2.2.1. 目的及び方法

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 化合物について化審法ガテドラインに基づいて染色体異常試験を行った。試験を行った化合物を下記に示す。

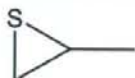
- a. Allyl ethyl ether (CAS No. 557-31-3, Alfa Aesar 製)



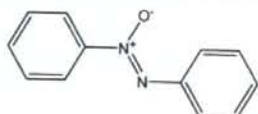
- b. diallyl carbonate (CAS No. 15022-08-9, シグマ製)



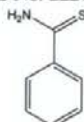
- c. Propylene sulfide (CAS No. 1072-43-1, 東京化成製)



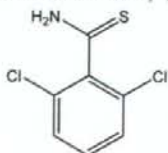
- d. Azoxybenzene (CAS No. 495-48-7, 和光純薬製)



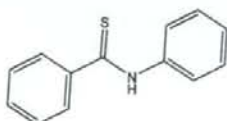
- e. Thiobenzamide (CAS No. 2227-79-4, 和光純薬製)



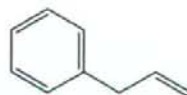
- f. chlorthiamid (CAS No. 1918-13-4, 和光純薬製)



- g. N-phenylthiobenzamide (CAS No. 636-04-4, 東京化成製)



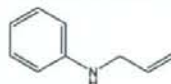
- h. Allylbenzene (CAS No. 300-57-2, 和光純薬製)



- i. N-allylaniline (CAS No. 627-37-2, シグマ製)



- j. N-allylmethylaniline (CAS No. 589-09-3, シグマ製)



### 2.2.2. 結果及び考察

試験結果を表 13 に示す。

表13. 染色体試験結果一覧

| 物質名                    | CAS No.    | 短時間処理法 |     | 数的異常      |     | D20<br>mg/mL |
|------------------------|------------|--------|-----|-----------|-----|--------------|
|                        |            | -S9    | +S9 | -S9       | +S9 |              |
| Allyl ethyl ether      | 557-31-3   | +      | +   | Equivocal |     | 0.11         |
| Diallyl carbonate      | 15022-08-9 | -      | +   | -         | -   | 0.021        |
| Propylene sulfide      | 1072-43-1  | -      | +   | -         | -   | 0.30         |
| Azoxybenzene           | 495-48-7   | -      | +   | -         | -   | 0.013        |
| Thiobenzamide          | 2227-79-4  | +      | +   | -         | -   | 0.44         |
| Chlorthiamid           | 1918-13-4  | -      | -   | -         | -   |              |
| N-Phenyl thiobenzamide | 636-04-4   | -      | -   | +         | +   | 1.5          |
| Allylbenzene           | 300-57-2   | -      | -   | -         | -   |              |
| N-allylmethyl amine    | 627-37-2   | +      | +   | -         | -   | 0.06         |
| N-allyl aniline        | 589-09-3   | -      | -   | -         | -   |              |

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。

Episulphides に関して、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、染色体数的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。

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。

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、その硫黄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。

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。

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。

以上の結果、O-Allyl compounds, Episuiphides 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。

### 3. 反復投与毒性試験における(Q)SAR モデルの開発に関する研究

#### 3.1. Derek for Windows の肝及び腎毒性予測精度改善のための Rapid Prototype 手法開発に関する研究

##### 3.1.1 目的及び方法

Derek for Windows は毒性予測のためのエキスパートシステムである。その知識ベースは、そのシステムにより行われる毒性予測にそれぞれ寄与するアラート、例証化合物および Reasoning rule から成り立っている。現行プロジェクトの目的は Derek for Windows の肝毒性および腎毒性の予測精度を改善するための Rapid Prototype を開発することである。

808 個の化合物に対する毒性試験データを要約した ISIS/Base データベース (NIHS Ames database1.db) を作製し、更にデータを追加して使

用した。(NIHS\_database\_UPDATED.db)。更にこのデータベース内に数多くの重複した化合物構造が認められたことより、データベースから取り除かれ、データベース (NIHS\_UPDATED\_Without\_duplicates\_FINAL\_280408.db) は、全部で 776 個の化合物を有し、Rapid Prototype 開発のデータソースとしてまず使用された。このデータベースは 29 個の Rapid Prototype 肝毒性アラートを導出するために使用された。

肝毒性および腎毒性が陽性であるための定義は、臨床的指標や病理組織学的な効果が認められる投与量に関係なく、データベース内の肝臓と腎臓の病理組織学的値の有無に基づく。このことから、病理組織学的欄に陽性値が存在する場合には化合物は陽性に分類し、値が無い場合には陰性に分類した。

Rapid Prototype 手法に対する最初の候補クラスが、クラスタリングソフトウェア"ChemTK"を使用して特定された。そのソフトウェアは、定義済みの設定(環、枝分かれした原子、立体構造形成能を持つフラグメント)に基づき与えられたデータセット中の化合物クラスの迅速な検索に対して有用なツールであり、また、各々の化合物クラスに対して平均的な活性に関する情報を用意することも可能である。解析において使用された設定は表 14 に示した。

表 14. ChemTK によるデータセット解析において使用された設定

| 手法               | 設定   |
|------------------|--|
| 環                | Min/Max = 1/3、 2/3* and 3/3  |
| 立体構造形成能を持つフラグメント | Min/Max fragment size = 3/10   |
| 枝分かれしたフラグメント     | Min/Max atoms in fragment = 4/8<br>Min/Max number of branched points = 0/2*<br>Min/Max number of bonds between terminal atoms = 1/7* |

SD ファイルが、化学構造・ID 番号および付加された。活性値欄を含むデータベース

(NIHS\_UPDATED\_Withoutduplicates\_FINAL\_280408.db) から作成された。活性値欄は肝臓の病理組織学値に対応しており、毒性が陽性である化合物に対しては値"1"が与えられ、陰性である化合物に対しては"0"が与えられた。その後、その SD ファイルは ChemTK で処理された。

遮断値が、潜在的に Rapid Prototype アラートに変換することができる化合物クラスを選択するために選ばれ、この値は 55% もしくはより多くの陽性化合物が化学物質クラス内に含まれるように設定された。

もしも、特定されたクラスが 55% 以下の陽性化合物しか有していなかった場合、そのクラスはこの解析からは除外した。残っている化合物クラスに対して、トキシコファーの範囲を導出するための更なる解析(例えば、ISIS/Base データベース内での部分構造検索など)が行われ、その後 Derek for Windows に実装された。

また、データセットに対して 55% もしくはより大きな基準値に適合しなかった化合物クラスを更に特定するために視覚的解析が行われた。しかし、それはまだ潜在的に有効な Rapid Prototype アラートとなるかも知れない。候補クラスは、官能基や類似な化学的環境を考慮すること、および、Derek for

Windows 知識ベース内の他の毒性エンドポイントに対する既存トキシコファーとのアナロジーによって特定された。特定された化合物クラスは先に記載した方法で処理され、Rapid Prototype アラートが Derek for Windows の知識ベース内に実装された。

上記とは異なる手法が、腎毒性 Rapid Prototype アラートを開発するために採用された。まず、既存の肝毒性アラート (Derek for Windows の先行バージョンである version10 と Rapid Prototype アラート) が腎毒性予測の判断材料として調査された。NIHS によって提供されたデータセットは、肝毒性 Rapid Prototype アラートを含む知識ベースに対して予測が実施され、腎毒性予測に対する肝毒性アラートの予測精度が調査された。最良の予測判断材料 (化合物クラス内の 55%を超える化合物を正しく予測したもの) が選ばれ、腎毒性 Rapid Prototype アラートとして開発された。さらに、視覚的解析が、肝毒性アラートに対して使用されたのと同じ基準を使用して、腎毒性に対する Rapid Prototype アラートをさらに開発するために行われた。

Derek for Windows の知識ベース内の Rapid

Prototype アラートの実装は、pattern、alert description、picture およびアラートの適用範囲を定義するための簡素なコメントから構成される。

### 3.1.2. 結果及び考察

全部で69個の Rapid Prototype アラートが開発され、Derek for Windows version11 知識ベースに実装され、その知識ベースはこの報告書において "DfW11 NIHS 201108\_1" と参照される34個の肝毒性 Rapid Prototype アラートと、35個の腎毒性アラートが実装された。

#### ・肝毒性 rapid prototype アラート

新しい知識ベースの予測精度が ISIS/Base データベース "NIHS5\_20080916.db" からのデータを使用して Derek for Windows version11 で調査された。表15と16は、それぞれ、標準の Derek for Windows version11 および DfW11 NIHS 201108\_1 知識ベースに対する肝毒性予測精度を示したものである。

表15. 肝臓の病理組織学的な所見に準じた分類に基づく、NIHS データベースに収録された化合物に対する Derek for Windows version11 の予測精度

| Derek for Windows<br>試験結果 | 陽性        | 陰性         | 合計  |
|---------------------------|-----------|------------|-----|
| 陽性                        | 69(23.0%) | 231        | 300 |
| 陰性                        | 51        | 380(88.2%) | 431 |

表16. 肝臓の病理組織学的な所見に準じた分類に基づく、NIHS データベースに収録された化合物に対する DfW11 NIHS 201108\_1 の予測精度

| Derek for Windows<br>試験結果 | 陽性         | 陰性         | 合計  |
|---------------------------|------------|------------|-----|
| 陽性                        | 184(61.3%) | 116        | 300 |
| 陰性                        | 88         | 343(79.6%) | 431 |

表16より、肝毒性予測精度が61%(184/300)に向上したが、その一方で、特異性が80%へと減少したことが分かる。特異性の減少は、Rapid Prototype アラートの実装に由来する追加的な37個の偽陽性予測によるものである。

#### ・腎毒性 rapid prototype アラート

新しい知識ベースの予測精度を ISIS/Base データベース "NIHS5\_20080916.db" からのデータを使用して Derek for Windows version11 と比較した。

表17と18は、それぞれ、標準の Derek for Windows version11 と DfW11 NIHS 201108\_1 知識ベースの腎毒性予測精度を示したものである。

表 1 7. 腎臓の病理組織学的な所見に準じた分類に基づく、NIHS データベースに収載された化合物に対する Derek for Windows version 11 の予測精度

| 試験結果 \ Derek for Windows | 陽性      | 陰性         | 合計  |
|--------------------------|---------|------------|-----|
| 陽性                       | 4(1.7%) | 226        | 230 |
| 陰性                       | 3       | 498(99.4%) | 501 |

表 1 8. 腎臓の病理組織学的な所見に準じた分類に基づく、NIHS データベースに収載された化合物に対する DfW11 NIHS 201108\_1 の予測精度

| 試験結果 \ Derek for Windows | 陽性         | 陰性         | 合計  |
|--------------------------|------------|------------|-----|
| 陽性                       | 143(62.2%) | 87         | 230 |
| 陰性                       | 69         | 432(86.2%) | 501 |

表 18 から、腎毒性予測感度が 62% に向上したが、一方で特異性は 85% 以上の値を維持していることが分かる。

Prototype アラートの予測精度を調査するために、二つの外部データセット ("A" と "B") に対して予測を実施し、肝毒性に対する予測精度 (表 19 と 21) と腎毒性に対する予測精度 (表 20 と 22) を示す。

・二つの外部データセットに対する予備的な結果  
DfW11 NIHS 201108\_1 に実装された Rapid

表 1 9. 肝臓の病理組織学的な所見に準じた分類に基づく、"A" データベースに収載された化合物に対する DfW11 NIHS 201108\_1 の予測精度

| 試験結果 \ Derek for Windows | 陽性  | 陰性  | 合計  |
|--------------------------|-----|-----|-----|
| 陽性                       | 73  | 165 | 238 |
| 陰性                       | 140 | 340 | 480 |

感度:  $73/238 = 31\%$ ; 特異性:  $340/480 = 71\%$ ; 一致率:  $413/718 = 58\%$

表 2 0. 腎臓の病理組織学的な所見に準じた分類に基づく、"A" データベースに収載された化合物に対する DfW11 NIHS 201108\_1 の予測精度

| 試験結果 \ Derek for Windows | 陽性  | 陰性  | 合計  |
|--------------------------|-----|-----|-----|
| 陽性                       | 47  | 119 | 166 |
| 陰性                       | 143 | 409 | 552 |

感度:  $47/166 = 28\%$ ; 特異性:  $409/552 = 74\%$ ; 一致率:  $456/718 = 64\%$

表 2 1. 肝臓の病理組織学的な所見に準じた分類に基づく、"B"データベースに収録された化合物に対する DfW11 NIHS 201108\_1 の予測精度

| Derek for Windows \ 試験結果 | 陽性 | 陰性 | 合計 |
|--------------------------|----|----|----|
| 陽性                       | 2  | 13 | 15 |
| 陰性                       | 4  | 11 | 15 |

感度:  $2/15 = 13\%$ ; 特異性:  $11/15 = 73\%$ ; 一致率:  $13/30 = 43\%$

表 2 2. 腎臓の病理組織学的な所見に準じた分類に基づく、"B"データベースに収録された化合物に対する DfW11 NIHS 201108\_1 の予測精度

| Derek for Windows \ 試験結果 | 陽性 | 陰性 | 合計 |
|--------------------------|----|----|----|
| 陽性                       | 3  | 8  | 11 |
| 陰性                       | 7  | 12 | 19 |

感度:  $3/11 = 27\%$ ; 特異性:  $12/19 = 63\%$ ; 一致率:  $15/30 = 50\%$

"A"データセットに対する肝毒性(表 19)と腎毒性(表 20)の予測感度は 30%前後に落ち込んでおり、その一方で特異性は 70%以上を維持している。表 19 からは、73 個の真陽性化合物と 140 個の偽陽性化合物が肝毒性アラートにヒットし、そのうち 51 個の真陽性化合物と 84 個の偽陽性化合物がそれぞれこのプロジェクトの一部として開発された Rapid Prototype アラートと関連していることが分かった。腎毒性に関しては、表 21 から 47 個の真陽性化合物と 143 個の偽陽性化合物が認められ、そのうち 44 個の真陽性化合物および 140 個の偽陽性化合物がそれぞれ Rapid Prototype アラートに関連している。肝毒性に対する標準 Derek for Windows アラートと Rapid Prototype アラート双方の予測精度に対するより詳細な調査により、Rapid Prototype アラートは既存の Derek for Windows アラートよりもよ

い予測精度を有することが分かった。

54%の肝毒性 Rapid Prototype アラートが、"A"データセットに対して 50%以上の陽性予測率を有している一方で、たった 39%の肝毒性 Derek for Windows アラートしか 50%を超える予測精度を有していない。この結果は、肝毒性および腎毒性双方に対する"DfW11 NIHS 201108\_1"の予測感度の向上は、既存の Derek for Windows version11 のアラートと比較した場合、Rapid Prototype アラートの存在に依るところが大きく、しかしながら、それはまた偽陽性予測の増加も招くことを示している。

それぞれの Rapid Prototype の予測精度を解析し、その結果が表 23 に要約した。

表 2 3. "A"データセットに対する"DfW11 NIHS 201108\_1"の陽性予測率の点から見た Rapid Prototype アラートの予測精度の要約

| エンドポイント                             | Rapid prototype アラートの予測精度 |    |
|-------------------------------------|---------------------------|----|
| 肝毒性<br>(34 個の Rapid Prototype アラート) | ヒットせず                     | 8  |
|                                     | 0%正しい                     | 6  |
|                                     | 1-49%正しい                  | 6  |
|                                     | >50%正しい                   | 14 |
| 腎毒性<br>(35 個の Rapid Prototype アラート) | ヒットせず                     | 12 |
|                                     | 0%正しい                     | 5  |
|                                     | 1-49%正しい                  | 11 |
|                                     | >50%正しい                   | 7  |

予測精度毎に分解したアラートの要約 (表 23) から、8 個の肝毒性 Rapid Prototype アラートと 12 個の腎毒性 Rapid Prototype アラートがヒットしなかった。一方で、双方の毒性エンドポイント内のおよそ 15 個の Rapid Prototype アラートはより低い (<50%) 予測精度を示しているということが分かる。これら予測精度の低いアラートをより詳細に解析すると、幾つかの Rapid Prototype アラートはトレーニングデータセット内の比較的少量の化合物に基づいて作成されていることが分かる。例えば、Hep16、Hep07、Nephro17、Nephro27 がそれに当たる。一般に腎毒性 Rapid Prototype アラートはより少ない真陽性化合物のトレーニングデータセットに基づいている一方で、殆どの肝毒性 Rapid Prototype アラートはトレーニングデータセット中のよりばらついた個数の化合物に基づいていることが分かる。Rapid Prototype アラートの予測精度は、アラートを開発するために使用されたトレーニングデータセット中の化合物の個数が適用範囲に影響を及ぼすので、これが部分的には関連しているかも知れず、そして、これは将来の Rapid Prototype 手法に対して考慮されるかも知れない。

### 3.1.3. 結論

34 個の肝毒性および 35 個の腎毒性 Rapid Prototype アラートが Derek for Windows に実装された。全ての Rapid Prototype アラート実装後のトレーニングデータセットに対する予測精度は、肝毒性に対して感度 61%、特異性 80%であり、腎毒性に対して感度 62%、特異性 86%である。

このプロジェクトの開始前には Derek for Windows 知識ベースは 40 個の肝毒性アラートを含んでおり、Rapid Prototype アラート開発により 74 個まで増加した。このプロジェクトの開始前までは系統的に開発されていなかった腎毒性に対しては、35 個の Rapid Prototype アラートの実装により、Derek for Windows の予測精度が目覚ましく向上した。NIHS により提供された私有データセットからの Rapid Prototype 開発の程度は

NIHS 保有のデータセット中のケミカルスペースに依存しており、そして、更なる開発は、双方の予測感度がおよそ 60%の水準に達した後に、限定的であることが分かった。

二つの外部データソースを使用しての、当プロジェクトで開発された Rapid Prototype アラートの評価により、肝毒性および腎毒性双方の予測感度および特異性の減少が認められた。幾つかの標準 Derek for Windows アラートに対しても同様の減少が見られたので、これらの所見は Rapid Prototype 手法の使用に直接起因するものではあり得ない。

Rapid Prototype アラートの適用範囲を精密に定義することができる程度は、トレーニングデータセット中のそれを支援する化合物の個数と特徴によって制約される。テストデータセットでの結果は、この適用範囲の改良がいくつかの Rapid Prototype アラートに対して求められることを示唆している。その他の場合において、予測精度の改善は、個々のアラートの適用領域の考慮の結果として予想されるかもしれない。最後に、病理組織学的な結果に準じた分類を越えた、予測精度の更なる調査は実施に値するかも知れない。

当プロジェクトは、ある適切なトレーニングデータセットが与えられた場合、Rapid Prototype 手法を Derek for Windows 知識ベース中の既存および新たな毒性エンドポイントに対するアラートの予測精度を改善するために使用することができるということを示した。新しいテストデータセットを、これらの Rapid Prototype アラートをさらに保障もしくは改良するために使用できる可能性があり、それらに続く標準 Derek for Windows アラートの開発を可能にする。適用領域の考慮といった追加的な技法が、予測精度の信頼度を解釈したり、Rapid Prototype アラートの適合性や品質の調査を手助けしたりするために使用されるかもしれない。

## 3.2. 肝毒性予測モデルの構築

### 3.2.1. モデルの構築 (1)

反復投与毒性試験結果から、Leadscopeを用いて肝毒性に対する QSAR 予測モデル構築を行った。808 件の主に 28 日間反復投与試験のデータ (LOAEL 値) を計算式「 $-\text{Log}(\text{数値}+4)$ 」により変換し、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。連続数値データからモデルの構築を試みた他、連続データをバイナリ化したモデルの構築も試みた。

#### ・対象化合物

モデル構築用のトレーニングデータセット (内部データセット) の化合物数は 808 であった。

#### ・予測モデルのまとめ

各モデルの内部データ予測精度を表24に示す。

表 24. 各モデルの内部データ予測精度

| 条件    | トレーニングセット   | モデル名             | Feature数 | Features 選択方法   | Concordance | Sensitivity | Specificity |
|-------|-------------|------------------|----------|---|-------------|-------------|-------------|
| 条件 1  | Liver_conv  | Run#10           | 207      | Automatic   | 87.1        | 50.3        | 96.7        |
| 条件 2  | Liver_conv  | Run#11           | 156      | Manual(Mean が 0.3-0.6 となる Features を除外)                               | 87.3        | 43.6        | 98.6        |
| 条件 3  | Liver_conv2 | Run#12           | 207      | Automatic   | 77.5        | 53.4        | 91.4        |
| 条件 4  | Liver_conv2 | Run#4            | 624      | Manual(Mean が 0.3-0.6 となる Features を除外)                               | 80          | 62.3        | 90.2        |
| 条件 5  | Liver_conv2 | Run#13           | 283      | Manual(Mean が 0.1-0.6 となる Features を除外)                               | 80          | 56.3        | 93.4        |
| 条件 6  | LGW         | Run#14           | 193      | Automatic   | 74.9        | 88.3        | 56.9        |
| 条件 7  | LGW         | Run#15           | 134      | Manual(Mean が 0.3-0.6 となる Features を除外など)                             | 74.6        | 91.1        | 52.8        |
| 条件 8  | LG          | Run#16           | 207      | Automatic   | 73.5        | 82.4        | 65.7        |
| 条件 9  | LG          | Run#18           | 179      | Manual(Mean が 0.3-0.6 となる Features を除外など)                             | 71.3        | 82.8        | 61.3        |
| 条件 10 | LG          | Run#20           | 187      | Manual(Mean が 0.3-0.7 で Z-Score が $\pm 2$ 範囲内となる Features のみを除外)      | 73.8        | 84.4        | 64.5        |
| 条件 11 | LG          | Run#21           | 179      | Manual(Z-Score が小さい ( $\pm 2$ ) Features を全て削除)                       | 73.0        | 85.1        | 62.4        |
| 条件 12 | LW          | Run#17           | 189      | Automatic   | 73.8        | 83.8        | 62.8        |
| 条件 13 | LW          | Run#19           | 145      | Manual(Mean が 0.3-0.6 となる Features を除外など)                             | 70.5        | 86.9        | 52.8        |
| 条件 14 | LW          | Run#22           |          | Manual(Mean LW が 0.294-0.7 で Z-Score が $\pm 2$ 範囲内となる Features のみを除外) | 73.5        | 83.5        | 62.6        |
| 条件 15 | Liver_conv2 | Predictive Model | 207      | Automatic   | 77.3        | 68.5        | 84.9        |

#### ・組合せモデルの予測精度の検討

条件 4 は特異性が非常に高く感度が比較的低い。条件 7 は逆に感度が高く特異性が低い。これらのモデルを相補的に組み合わせて肝毒性を予測することで高い精度が得られると考えられた。その結

果、陰性/陽性が一致しない場合は予測不可能と見なし、予測可能なケースのみ集計した。モデルを組み合わせた場合、適用範囲が小さくなる代わりに予測精度が大幅に向上することが分かった (表 25)。



表 2 5. 条件 4 と条件 7 による組合せモデルの予測精度

|                  |   | 条件 4 と条件 7 モデルの組合せ |     |       | %           |       |
|------------------|---|--------------------|-----|-------|-------------|-------|
|                  |   | +                  | -   | total | concordance |       |
| Liver_conv2      | + | 172                | 15  | 187   | sensitivity | 91.98 |
|                  | - | 48                 | 196 | 244   | specificity | 80.33 |
| 予測可能な化合物の割合(%) : |   | 53.34              |     |       |             |       |
|                  |   | 条件 4 と条件 7 モデルの組合せ |     |       | %           |       |
|                  |   | +                  | -   | total | concordance |       |
| LGW              | + | 191                | 31  | 222   | sensitivity | 86.04 |
|                  | - | 29                 | 180 | 209   | specificity | 86.12 |
| 予測可能な化合物の割合(%) : |   | 53.34              |     |       |             |       |

・モデル適用範囲の検討

構築した予測モデルにより予測される Predicted Probabilities が 0.5 より大きい場合は「陽性」、0.5 より小さい場合は「陰性」と判断した上で予測精度が計算される。この判断基準を変更することで予測精度を上げることができると

予想された。従って、以下では条件 4(Run#4)および条件 7(Run#16)のモデルについて Predicted Probability のカットオフの値を変更することでより高い予測精度が得られるかを検討した。表 26 及び表 27 に Model 構築の結果を記載した。

表 2 6. 条件 4 モデルについて、判定のカットオフ値をマニュアル設定 (minimum probability needed for a positive call: 0.75 ; the maximum probability needed for a negative call: 0.25) にして予測した場合

|                  |   | 条件 4 モデル |     |       | %           |       |
|------------------|---|----------|-----|-------|-------------|-------|
|                  |   | +        | -   | total | concordance |       |
| Liver_conv2      | + | 118      | 27  | 145   | sensitivity | 81.38 |
|                  | - | 16       | 336 | 352   | specificity | 95.45 |
| 予測可能な化合物の割合(%) : |   | 61.51    |     |       |             |       |

表 2 7. 条件 7 のモデルについて、判定のカットオフ値をマニュアル設定 (同上) にして予測した場合

|                  |   | 条件 7 モデル |     |       | %           |       |
|------------------|---|----------|-----|-------|-------------|-------|
|                  |   | +        | -   | total | concordance |       |
| LGW              | + | 107      | 8   | 115   | sensitivity | 93.04 |
|                  | - | 9        | 140 | 149   | specificity | 93.96 |
| 予測可能な化合物の割合(%) : |   | 32.67    |     |       |             |       |

条件 7 による予測精度を計算したところ、concordance 93% となり条件 4 による予測精度と比べ若干小さい値となった。また、適用可能な化合物数も条件 4 のモデルの場合よりも少ないことが分かった。Positive と判断する基準を 0.75 以上から 0.6 以上に下げると、今度は false positive が増える問題が起きることも分かった。さらに、条

件 4 のモデルで、更に詳細に適用範囲を検討することで予測精度がどの程度改善できるのかを検討した。表 28 に各カットオフに対する予測精度を以下の表にまとめた。パターン 1~5 については、パターン 1 の予測精度が一番高いことがわかる。またパターン 6~10 では、パターン 6 が一番予測可能な範囲が広いことが分かった。

表 28. 適用範囲の検討

| パターン | 陰性      | 陽性      | 予測可能な割合(%) | concordance | sensitivity | specificity |
|------|---------|---------|------------|-------------|-------------|-------------|
| 1    | 0.10 以下 | 0.75 以上 | 42.45      | 94.17       | 96.72       | 92.76       |
| 2    | 0.20 以下 | 0.75 以上 | 54.83      | 93.23       | 89.39       | 94.86       |
| 3    | 0.25 以下 | 0.75 以上 | 61.51      | 91.35       | 81.38       | 95.45       |
| 4    | 0.30 以下 | 0.75 以上 | 65.97      | 89.49       | 74.68       | 95.73       |
| 5    | 0.40 以下 | 0.75 以上 | 79.95      | 84.52       | 58.42       | 96.4        |
| 6    | 0.25 以下 | 0.50 以上 | 73.64      | 87.06       | 87.08       | 87.05       |
| 7    | 0.25 以下 | 0.60 以上 | 67.70      | 89.58       | 85.08       | 91.8        |
| 8    | 0.25 以下 | 0.70 以上 | 63.74      | 91.07       | 83.13       | 94.65       |
| 3    | 0.25 以下 | 0.75 以上 | 61.51      | 91.35       | 81.38       | 95.45       |
| 9    | 0.25 以下 | 0.80 以上 | 60.15      | 91.56       | 80.15       | 96.00       |
| 10   | 0.25 以下 | 0.90 以上 | 55.32      | 92.84       | 74.53       | 98.53       |

### ・まとめ

本作業では、808 化合物に対する 28 日間反復投与毒性の肝毒性データをトレーニングデータセットとして 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。肝毒性に関連するデータ (LOAEL 値) を変換 ( $-\log(\text{数値}+4)$ ) し、連続的な数値データから予測モデルの構築を試みたが十分な精度のモデルを構築することはできなかった。連続データをバイナリ化 (1 か 0 に変換) し、このバイナリ値をトレーニングデータセットとした場合には、上記 808 化合物に対して感度の高いモデル (条件 7)、特異性の高いモデル (条件 4) などを構築することができた。肝毒性に関連するデータとしては、病理組織学的データ以外に GOT/GPT、肝重量変化をトレーニングデータセットとして加える場合と加えない場合を検討した。条件 4 と条件 7 の各モデルを相補的に組み合わせることで高い予測精度を得ることができた。また、モデルが予測の際に出力する Probabilities の閾値を設定することにより適用範囲を制限し、更に予測精度を高めることが可能なことが分かった。

### 3.2.2. モデルの構築 (2)

トレーニングデータ数を増やし、また、評価用に外部データセットを用意して、反復投与毒性予測モデルを構築し、検証を行った。

### ・予測対象とする毒性

#### 1) 肝毒性

組織病理学的検査における肝臓への影響 (項目 L)、血液生化学検査における GOT 値または GPT 値の高値 (項目 G)、臓器重量における肝臓 (絶対または相対) 重量の高値 (項目 W) をデータとして利用した。各項目は、最小毒性量 (LOAEL) である。

#### 2) 予測モデル構築のためのデータ項目

上記項目を単独で (L)、または、組み合わせて (LGW)、肝毒性を示す項目として用いた。

#### 3) 予測するデータ項目

構築したモデルで予測する値は、項目 L とする。Features の選択については、LGW を用いたモデル、および L を用いたモデルについて評価した。

### ・予測モデルとその検証

#### 1) 予測モデル構築に使用するソフト

LeadScope (LeadScope 社製) を用いて QSAR モデルの構築と QSAR モデルによる化合物の毒性予測を行った。

#### 2) 予測モデル構築のためのデータセット

予測モデル構築には既知の化合物の反復投与毒性試験データを用いた。表 29 に種類とデータ数を示す。

表 2.9. 既知の化合物の試験の種類とデータ数

|                        |           |
|------------------------|-----------|
| 28 日間試験                | 1,232 データ |
| 90 日間 (13 週間) 試験       | 28 データ    |
| 併合試験                   | 2 データ     |
| その他 (2 年間: 2、21 日間: 1) | 3 データ     |
| 計                      | 1,265 データ |

#### ・予測モデル検証のためのデータセット

予測モデルの検証には、既存化合物の反復投与毒性試験データを用いた。また、予測モデル構築のためのデータセットとの重複はない。既存化合物の反復投与毒性試験の種類とデータ数は、表 30 の通りである。

表 3.0. 既存化合物の試験の種類とデータ数

|                  |         |
|------------------|---------|
| 28 日間試験          | 136 データ |
| 90 日間 (13 週間) 試験 | 2 データ   |
| 併合試験             | 113 データ |
| 簡易試験             | 6 データ   |
| 計                | 258 データ |

#### ・予測精度に関する用語

用語とその定義は以下の通りである。

- ・ Sensitivity 感度：動物試験の陽性結果を陽性と判定した割合 (%)
  - ・ Specificity 特異性：動物試験の陰性結果を陰性と判定した割合 (%)
  - ・ Concordance 全体の一致率：動物試験の陽性結果を陽性、陰性結果を陰性と判定した割合 (%)
- その他、
- ・ 予測可能な化合物の割合：適用させるデータ全数に対して実際に予測したデータ数の割合 (%)

#### ・予測モデルの構築

##### 1) データのバイナリ化

動物試験における肝毒性の強さは、全身的な症状を考慮した上で、その毒性の種類や症状の程度によって表現されるものである。したがって、化

合物間の肝毒性の強さは、LOEL の大小で単純には比較できない。このことから、各項目のデータを全身的な症状について考慮せず扱う場合には、連続的の数値として捉えるのではなく、そのデータをバイナリ化 (0: 毒性なし、1: 毒性あり、の二極化) して使用することは妥当であると考えられた。最高用量まで影響を示さなかった項目を 0 (毒性なし)、それ以外の項目を 1 (毒性あり) とした。

##### 2) Features の選択

記述子となる Feature 数が多いとモデルへの当てはまり (つまり、内部データセットによる精度) が良くなるが、外部データに対する予測は、モデルの複雑さが増すにつれ、あるところから精度が低くなるのが予想される。このことから、次に示す方法で、自動作成した Feature の数を減らすモデルの作成を試みた。

##### (1) 自動選択を用いる方法

(2) Feature の平均値が 0.5 付近 (0.400~0.600) の Feature を除外する方法：陽性と陰性の化合物が等量ずつ含まれる Feature は重要ではないと考え、平均値が 0.5 付近の Feature 除外しモデルを作成した。

(3) 類似 Feature を除外する方法：類似 Feature を LeadScope で自動的に除外し、モデルを作成した。

(4) Feature の Z スコアの絶対値が 2 より小さい Feature を除外する方法：Feature の Z スコアの絶対値が 2 より小さい Feature を有意ではないと考え除外し、モデルを作成した。

#### ・適用範囲の設定

算出される Predicted Probabilities (予測確率) は 0~1 の値である。Predicted Probabilities が 0.5 以上の場合を「陽性」、0.5 より小さい場合を「陰性」とした。この範囲を変化させることで、精度が良くなるかについて検討した。

##### 4) 種々のモデル検討

モデル LGW、モデル L の他に、それらのモデルを組み合わせたモデルについても検討した。

#### ・予測モデルの検証

予測モデルの構築と同様にデータをバイナリ化

して、モデルの検証に使用した。検証には項目 L を用いた。

### ・予測モデルの結果 (内部データセット)

#### 1) 結果および Features の選択による影響

構築データセットの陽性データ数、陰性データ数、および、その割合を表 31 に示す。選択された Features ごとの精度を表 32 と表 33 に示す。

表 3 1. 構築データセットの陽性(+)数と陰性(-)数

|     | +   | -   | 計    | +(%) | - (%) |
|-----|-----|-----|------|------|-------|
| L   | 453 | 812 | 1265 | 35.8 | 64.2  |
| LGW | 712 | 553 | 1265 | 56.3 | 43.7  |

表 3 2. モデル LGW の場合 (Features の選択方法ごと)

| Feature 選択方法 (Feature 数)    | 予測項目    | 予測値 (化合物数) |         |                    |             |             |
|-----------------------------|---------|------------|---------|--------------------|-------------|-------------|
| 自動 (387)                    | LGW     | 1 (863)    | 0 (402) | 75.7               | Concordance |             |
|                             | 実際の LGW | 1          | 634     | 78                 | 89.0        | Sensitivity |
|                             |         | 0          | 229     | 324                | 58.6        | Specificity |
| 0.5 付近の Feature 除外 (262)    | LGW     | 1 (842)    | 0 (423) | 76.0               | Concordance |             |
|                             | 実際の LGW | 1          | 636     | 76                 | 89.3        | Sensitivity |
|                             |         | 0          | 227     | 326                | 59.0        | Specificity |
| 類似 Feature 除外 (327)         | LGW     | 1 (842)    | 0 (423) | 77.9               | Concordance |             |
|                             | 実際の LGW | 1          | 637     | 75                 | 89.5        | Sensitivity |
|                             |         | 0          | 205     | 348                | 62.9        | Specificity |
| 予測可能な化合物の割合 (1265/1265*100) |         |            |         | 100.0 <sup>a</sup> |             |             |

<sup>a</sup>) Probability threshold を「陽性 $\geq$ 0.5、0.5 $>$ 陰性」とした場合

表 3 3. モデル L の場合 (Features の選択方法ごと)

| Feature 選択方法 (Feature 数)    | 予測項目  | 予測値 (化合物数) |          |                    |             |             |
|-----------------------------|-------|------------|----------|--------------------|-------------|-------------|
| 自動 (352)                    | L     | 1 (233)    | 0 (1032) | 72.8               | Concordance |             |
|                             | 実際の L | 1          | 171      | 282                | 37.7        | Sensitivity |
|                             |       | 0          | 62       | 750                | 92.4        | Specificity |
| 0.5 付近の Feature 除外 (345)    | L     | 1 (220)    | 0 (1045) | 71.5               | Concordance |             |
|                             | 実際の L | 1          | 156      | 297                | 34.4        | Sensitivity |
|                             |       | 0          | 64       | 748                | 92.1        | Specificity |
| 類似 Feature 除外 (293)         | L     | 1 (196)    | 0 (1069) | 71.3               | Concordance |             |
|                             | 実際の L | 1          | 143      | 310                | 31.6        | Sensitivity |
|                             |       | 0          | 53       | 759                | 93.5        | Specificity |
| 予測可能な化合物の割合 (1265/1265*100) |       |            |          | 100.0 <sup>a</sup> |             |             |

\*) Probability threshold を「陽性 $\geq$ 0.5、0.5 $>$ 陰性」とした場合

・適用範囲の検討結果

ける化合物の陽性数、陰性数を示す。

図 1～3 に、Predicted Probabilities の各区間にお

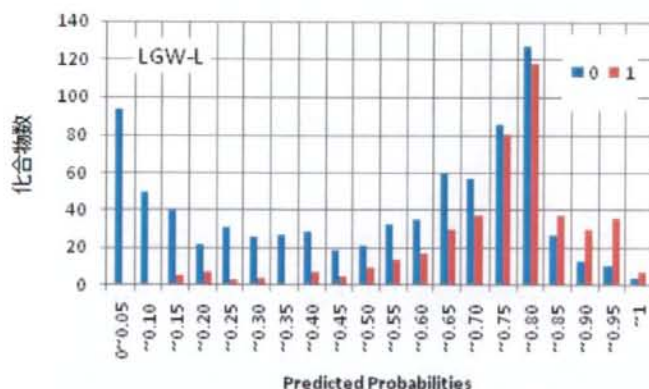


図 1. Predicted Probabilities の各区間における陽性、陰性の分布 (モデル LGW 自動-予測項目 L)

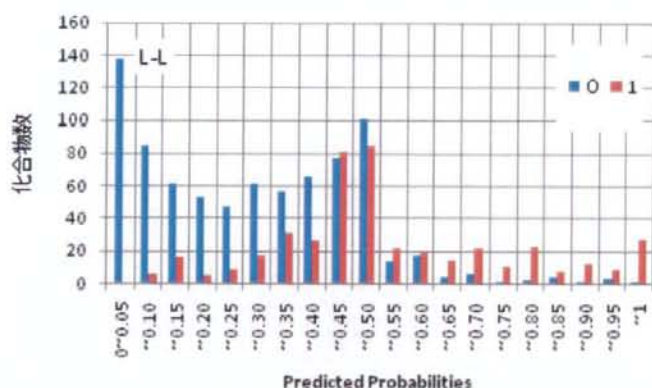


図 2. Predicted Probabilities の各区間における陽性、陰性の分布 (モデル L 自動-予測項目 L)

図 1 からモデル LGW\_自動では陽性/陰性基準を 0.5 から移動させてもバランスよく精度をあげる効果がないように考えられた。図 2 からモデル L\_

自動では 0.4 以上を陽性と判定した場合、効果的に陽性予測が増加すると考えられた。

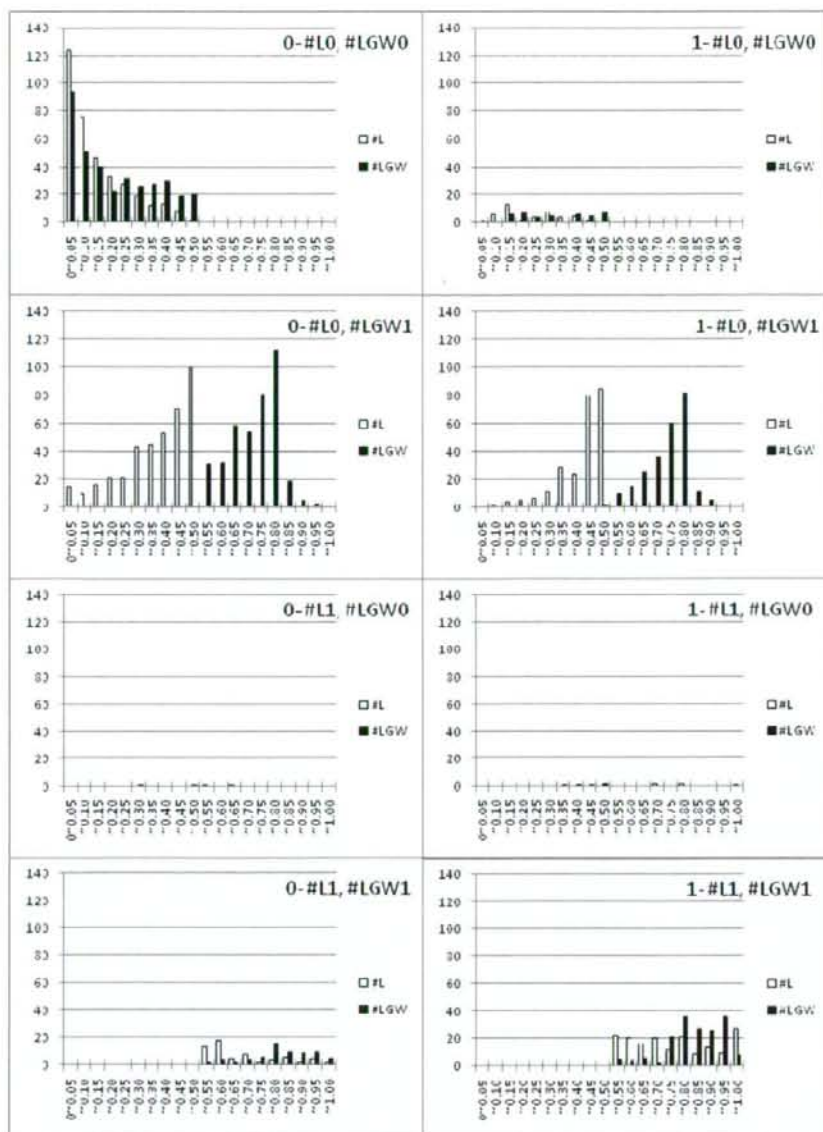


図 3. 組み合わせた場合の陽性、陰性の詳細な分布

図 3 より、組合せモデルにおいてモデル L\_自動では Predicted Probabilities が 0.4 以上を陽性と判定した場合、効果的に陽性予測が増加して感度が上がり、一方、陰性予測は減少して特異性が低下することが考えられた（モデル L'として検討）。モデル LGW\_自動については単独でモデルを使用する場合に Predicted Probabilities の基準を移動させても効果が弱いと考えられたことから、組合せモ

デルにおいてもモデル LGW\_自動については基準の移動は検討しないこととした。

モデル L\_自動とモデル LGW\_自動の結果を表 34 と表 35 に示す。同様に他のモデル L についても Predicted Probabilities の検討を行ったところ、モデル L\_0.5 およびモデル L\_類似は、モデル L\_自動から派生しており、モデル L\_自動と同様の分布を示した。結果として、モデル L\_自動、モデル

L\_0.5、モデル L\_類似とともに 0.4 以上を陽性、0.4 未満を陰性とした。修正した Predicted Probabilities を用いたモデルの名称は、モデル L' とした。

表 3.4. モデル L 自動とモデル L' 自動の場合 (Probability threshold 変更の影響)

|                             | 予測項目 | 予測値 (化合物数) |          |       |             |
|-----------------------------|------|------------|----------|-------|-------------|
| 陽性>=0.5、陰性<0.5 (初期設定)       | L    | 1 (233)    | 0 (1032) | 72.8  | Concordance |
| 実際の L                       | 1    | 171        | 282      | 37.7  | Sensitivity |
|                             | 0    | 62         | 750      | 92.4  | Specificity |
| 陽性>=0.4、陰性<0.4              | L    | 1 (578)    | 0 (687)  | 71.8  | Concordance |
| 実際の L                       | 1    | 337        | 116      | 74.4  | Sensitivity |
|                             | 0    | 241        | 571      | 70.3  | Specificity |
| 予測可能な化合物の割合 (1265/1265*100) |      |            |          | 100.0 |             |

表 3.5. モデル LGW 自動の場合

|                             | 予測項目 | 予測値 (化合物数) |         |       |             |
|-----------------------------|------|------------|---------|-------|-------------|
| 陽性>=0.5、陰性<0.5 (初期設定)       | L    | 1 (863)    | 0 (402) | 60.8  | Concordance |
| 実際の L                       | 1    | 410        | 43      | 90.5  | Sensitivity |
|                             | 0    | 453        | 359     | 44.2  | Specificity |
| 予測可能な化合物の割合 (1265/1265*100) |      |            |         | 100.0 |             |

#### ・複合モデル

前の結果から、組み合わせたモデルの陰性/陽性が一致しない場合を予測不可能と見なし、予測可能なケースのみ集計したモデルを複合モデルとし、その名称を表 36 に示す。新たに二段階モデルを構

築し、その名称は複合モデルの「複合」を「二段階」に置き換え、タイプ名を付加したものとした。二段階モデルのタイプを図 4 に示す。モデル L' を使用した場合の名称は、L と同様に付けたのち「'」を付加した。

表 3.6. 複合モデルに使用した組合せ (モデル LGW とモデル L)

| モデル LGW            | モデル L              | 複合モデルの名称 |
|--------------------|--------------------|----------|
| 自動                 | 自動                 | 複合モデル A  |
| "                  | 0.5 付近の Feature 除外 | 複合モデル B  |
| "                  | 類似 Feature 除外      | 複合モデル C  |
| 0.5 付近の Feature 除外 | 自動                 | 複合モデル D  |
| "                  | 0.5 付近の Feature 除外 | 複合モデル E  |
| "                  | 類似 Feature 除外      | 複合モデル F  |
| 類似 Feature 除外      | 自動                 | 複合モデル G  |
| "                  | 0.5 付近の Feature 除外 | 複合モデル H  |
| "                  | 類似 Feature 除外      | 複合モデル I  |

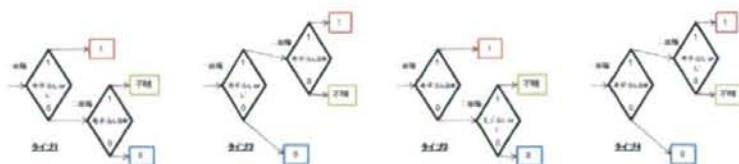


図 4. 二段階モデルのタイプ別フローチャート

モデル L (または L') とモデル LGW の組合せには、図 4 のように 4 通りある。実際の項目 L と項目 LGW の関係では、項目 L が陽性であるなら項目 LGW も陽性であり、項目 LGW が陰性であるなら項目 L も陰性である。したがって、二段階モデルのタイプとしてはタイプ 1 およびタイプ 4 が妥当と考えられた。

・ 予測モデルの検証 (外部データセット)

検証データセットにおける陽性データ数、陰性データ数、および、その割合を表 37 に示す。選択された Features ごとの検証結果を表 38~40 に示す。

表 3 7. 検証データセットの陽性(+)数と陰性(-)数

|   | +   | -   | 計   | +(%) | -(%) |
|---|-----|-----|-----|------|------|
| L | 121 | 137 | 258 | 46.9 | 53.1 |

表 3 8. モデル LGW の検証

| Feature 選択方法              | 予測項目  | 予測値 (化合物数) |        |      |             |             |
|---------------------------|-------|------------|--------|------|-------------|-------------|
| 自動                        | L     | 1 (218)    | 0 (40) | 50.8 | Concordance |             |
|                           | 実際の L | 1          | 106    | 15   | 87.6        | Sensitivity |
|                           |       | 0          | 112    | 25   | 18.2        | Specificity |
| 0.5 付近の Feature 除外        | L     | 1 (220)    | 0 (38) | 50.0 | Concordance |             |
|                           | 実際の L | 1          | 106    | 15   | 87.6        | Sensitivity |
|                           |       | 0          | 114    | 23   | 16.8        | Specificity |
| 類似 Feature 除外             | L     | 1 (217)    | 0 (41) | 51.2 | Concordance |             |
|                           | 実際の L | 1          | 106    | 15   | 87.6        | Sensitivity |
|                           |       | 0          | 111    | 26   | 19.0        | Specificity |
| 予測可能な化合物の割合 (258/258*100) |       |            |        |      | 100.0       |             |



表 3 9. モデル L の検証

| Feature 選択方法              | 予測項目 | 予測値 (化合物数) |         |       |             |
|---------------------------|------|------------|---------|-------|-------------|
| 自動                        | L    | 1 (37)     | 0 (221) | 52.7  | Concordance |
| 実際の L                     | 1    | 18         | 103     | 14.9  | Sensitivity |
|                           | 0    | 19         | 118     | 86.1  | Specificity |
| 0.5 付近の Feature 除外        | L    | 1 (47)     | 0 (211) | 55.0  | Concordance |
| 実際の L                     | 1    | 26         | 95      | 21.5  | Sensitivity |
|                           | 0    | 21         | 116     | 84.7  | Specificity |
| 類似 Feature 除外             | L    | 1 (22)     | 0 (236) | 52.3  | Concordance |
| 実際の L                     | 1    | 10         | 111     | 8.3   | Sensitivity |
|                           | 0    | 12         | 125     | 91.2  | Specificity |
| 予測可能な化合物の割合 (258/258*100) |      |            |         | 100.0 |             |

表 4 0. モデル L' の検証

| Feature 選択方法              | 予測項目 | 予測値 (化合物数) |        |       |             |
|---------------------------|------|------------|--------|-------|-------------|
| 自動                        | L    | 1 (177)    | 0 (81) | 55.0  | Concordance |
| 実際の L                     | 1    | 91         | 30     | 75.2  | Sensitivity |
|                           | 0    | 86         | 51     | 37.2  | Specificity |
| 0.5 付近の Feature 除外        | L    | 1 (179)    | 0 (79) | 56.6  | Concordance |
| 実際の L                     | 1    | 94         | 27     | 77.7  | Sensitivity |
|                           | 0    | 85         | 52     | 38.0  | Specificity |
| 類似 Feature 除外             | L    | 1 (165)    | 0 (93) | 56.6  | Concordance |
| 実際の L                     | 1    | 87         | 34     | 71.9  | Sensitivity |
|                           | 0    | 78         | 59     | 43.1  | Specificity |
| 予測可能な化合物の割合 (258/258*100) |      |            |        | 100.0 |             |

表 4 1. 複合モデルの検証

| モデル名              | 予測可能な化合物の割合 | Concordance | Sensitivity | Specificity |
|-------------------|-------------|-------------|-------------|-------------|
| モデル LGW_自動        | 100.0       | 50.8        | 87.6        | 18.2        |
| モデル L'_自動         | 100.0       | 55.0        | 75.2        | 37.2        |
| 複合モデル A'          | 80.2        | 53.6        | 87.3        | 21.0        |
| 二段階モデル A' (タイプ 4) | 82.2        | 53.8        | 85.6        | 23.1        |

複合モデルにおいて、感度は主にモデル L の Feature 選択の影響を受け、特異性はモデル LGW とモデル L の Feature 選択の影響を共に受けていた。二段階モデルではモデル I' (タイプ 4) において予測可能な化合物の割合、感度、特異性のバランスが比較的良かったが、その他のモデルにおいても二段階モデル I' (タイプ 4) と比べて精度に大差のないものが多く、二段階モデル A' (タイプ 4) でも遜色がないことがわかった。

種々のモデル構築方法における代表的なモデルの検証結果を表 41 に示す。モデル L\_自動において、予測可能な化合物の割合および全体の一致率が高く、感度と特異性のバランスも良いことがわかった。

| モデル手法 | Concordance | Sensitivity | Specificity |
|-------|-------------|-------------|-------------|
| SVM   | 99.2%       | 99.1%       | 99.3%       |
| KNN   | 88.4%       | 100%        | 80.6%       |
| ADA   | 100%        | 100%        | 100%        |

外部検証用母集団 (68 件) に対する検証結果は下記に示す様に、Sensitivity ではほぼ 80% を超える値を示していることから、肝に対して変化を及ぼす可

### 3.3.3. 結論

実験動物による反復投与毒性試験の結果を用い、肝毒性 (組織病理学的検査による肝臓影響) に関する反復投与毒性予測モデルを LeadScope によって構築し、その予測モデルの検証を行った。その結果、モデル作成用データセットとして項目 L (組織病理学的検査による肝臓影響) を用い、LeadScope による自動 Feature 作成を採用し、Predicted Probabilities が 0.4 以上の場合を陽性、0.4 未満の場合を陰性として構築したモデル (モデル L\_自動) の精度が比較的バランスよく高いことがわかった。

能性のある化合物の予測に使用できる可能性が示された。

### 1) SVM モデル

|      |   | SVM モデル |    |    |                      |
|------|---|---------|----|----|----------------------|
|      |   | +       | -  |    |                      |
| 試験結果 | + | 35      | 28 | 7  | 75.00% (Concordance) |
|      | - | 33      | 10 | 23 | 80.00% (Sensitivity) |
|      |   | 68      | 20 | 30 | 69.70% (Specificity) |

### 2) KNN モデル

|      |   | KNN モデル |    |    |                      |
|------|---|---------|----|----|----------------------|
|      |   | +       | -  |    |                      |
| 試験結果 | + | 35      | 29 | 6  | 72.06% (Concordance) |
|      | - | 33      | 13 | 20 | 82.86% (Sensitivity) |
|      |   | 68      | 26 | 26 | 60.61% (Specificity) |

### 3) ADA モデル

|      |   | ADA モデル |    |    |                      |
|------|---|---------|----|----|----------------------|
|      |   | +       | -  |    |                      |
| 試験結果 | + | 35      | 29 | 6  | 83.82% (Concordance) |
|      | - | 33      | 5  | 28 | 82.86% (Sensitivity) |
|      |   | 68      | 10 | 34 | 84.85% (Specificity) |

## 4. カテゴリーアプローチ手法に関する研究

### 4.1. 2-sec プチルフェノールの構造異性体に関するカテゴリーアプローチの検討

#### 4.1.1. 方法

##### 4.1.1.1. 物理化学性状

OECD の HPV プログラムにおける SIDS 項目 (Screening Information Data Set、初期評価要求項目) を参考に、下記の 9 項目について調査を行った。

- |          |         |
|----------|---------|
| ① 外観     | ⑥ 分配係数  |
| ② 融点     | ⑦ 対水溶解度 |
| ③ 沸点     | ⑧ 解離定数  |
| ④ 密度及び比重 | ⑨ 水中安定性 |
| ⑤ 蒸気圧    |         |

調査対象は、OECD の HPV プログラム実施マニュアルを参考に、以下の 12 の書籍及びデータベースとした。

- |   |  |
|---|--|
| ① CRC Handbook of Chemistry and Physics (84 <sup>th</sup> edition, CRC Press, 2003) | ③ Handbook of Environmental Data on Organic Chemicals (volume 1, John Wiley & Sons, Inc., 2001)  |
| ② Handbook of Aqueous Solubility Data (CRC Press, 2003)                             | ④ Handbook of Physical Properties of Organic Chemicals (CRC Press, 1997)   |
|   | ⑤ Hawley's Condensed Chemical Dictionary (15 <sup>th</sup> Edition, Wiley Inter-science, 2007)   |
|   | ⑥ Hazardous Substances Data Bank ( <a href="http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB">http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB</a> ) |
|   | ⑦ Lange's Handbook of Chemistry (6 <sup>th</sup> edition, McGraw-Hill, 2005)   |
|   | ⑧ The Merck Index (14 <sup>th</sup> Edition, MERCK & Co., Inc., 2006)  |
|   | ⑨ Perry's Chemical Engineers' Handbook (7 <sup>th</sup> edition, McGraw-Hill, 1997)  |
|   | ⑩ Sax's Dangerous Properties of Industrial Materials (11 <sup>th</sup> edition, Wiley Inter science,   |

2004)

- ⑪ SRC PhysProp Database  
(<http://www.syrres.com/esc/physdemo.htm>)
- ⑫ The Yaws Handbook of Physical Properties for Hydrocarbons and Chemicals (Gulf Publishing Company, 2005)

調査対象物質の一部については国（経済産業省）及び新エネルギー・産業技術総合開発機構により、一部の試験が実施されており、幾つかの物理化学的性状データが得られている。よって、以下に示すこれらの試験結果も参考にした。

- ・ 2-secブチルフェノール（被験物質番号 K-137）の物理化学性状の測定、1999年（試験委託書：通商産業省）
- ・ 2-secブチルフェノール（被験物質番号 K-1367）の1-オクタノールと水との間の分配係数試験、1999年（試験委託書：通商産業省）
- ・ p-secブチルフェノール（被験物質番号 K-479）の水中における解離定数の測定、2006年（試験委託書：新エネルギー・産業技術総合開発機構）
- ・ p-secブチルフェノール（被験物質番号 K-479）の1-オクタノールと水との間の分配係数試験、2006年（試験委託書：新エネルギー・産業技術総合開発機構）

さらに、構造活性相関手法が一般的に用いられている項目に関しては、OECDのHPVプログラムにおいて、比較的よく用いられている以下の計算ソフトを用いて、構造活性相関による計算値を求めた。

- ① 融点、沸点、蒸気圧：MPBPWIN (version 1.41)
- ② 分配係数：KOWWIN (version 1.67)
- ③ 対水溶解度：WATERNT (version 1.01)及び WSKOWWIN (version 1.41)
- ④ 解離定数：SPARC (version 3.1)

MPBPWIN、KOWWIN、WATERNT及び WSKOWWIN は米国 EPA のホームページ (<http://www.epa.gov/oppt/exposure/pubs/episuite.htm>)より入手したソフトウェアを用いた。また、SPARC は <http://fibmlc2.chem.uga.edu/sparc/index.cfm> におけるオンライン上の計算ソフトウェアである。

これらの結果より、調査対象物質が一つのカテゴリとして評価できるかどうかを、個々の項目毎に

検討した。

#### 4.1.1.2. 環境運命及び環境毒性

本検討で用いたデータ項目は、以下の SIDS 項目からさらに生態影響評価に用いるものを選択した。

##### 1) 環境運命及び経路

- 安定性 A. 光分解性  
B. 水中での安定性
- 移動と環境分布
- 好氣的生分解性
- BOD-5、COD 及び BOD-5/COD比
- 生物濃縮

##### 2) 環境毒性

- 魚類急性毒性
- 水生無脊椎動物急性毒性（ミジンコ）
- 水生植物（藻類）毒性
- 微小生物（バクテリア）への毒性
- 水生生物慢性毒性（\*）
  - A. 魚類慢性毒性
  - B. 水生無脊椎動物慢性毒性
- 陸生生物への毒性（\*）
  - A. 陸上植物毒性
  - B. 土壌生物毒性
  - C. その他の陸上生物（鳥類を含み、ほ乳類を除く）への毒性
  - D. 底生生物への毒性

（\*）：SIDSエレメントに準ずる項目

情報があれば記述すべき項目

本検討で用いたデータは、OECDのHPVプログラムによるSIDS初期リスク評価文書などのリスク評価書から入手した。さらに、環境運命及び経路については、化審法既存化学物質安全性（分解性・濃縮性）点検結果を、環境毒性については、環境省の生態影響試験結果一覧（平成20年3月版）及び米国EPAのEcotox Databaseを用いた。なお、既存のリスク評価情報の有無については、OECD e-chem-potalで検索することで点検した。

本検討はカテゴリーアプローチの生態影響評価への適用可能性について検討することを目的としているため、生態毒性情報について、試験報告書などの原典が入手できた場合には信頼性評価を行うこととした。ただし、カテゴリー構成物質の生態毒性値の類似性・法則性の検討を行う際には、連続数での毒性評価となることから水生生物への急性毒性値（L(E)C50）を重点的に用いることとした。試験データは数値の相互比較を行うため原則としてはOECDテストガイド