

# 構造活性相関 ((Q) SAR) による遺伝毒性の予測

## 1. はじめに

化学物質の規制に関わる国際機関や、各国規制当局の最近の関心の焦点は、規制の対象となるすべての化学物質を実験的に試験することなく、有害作用を引き起こす化学物質を同定するための単純なスクリーニングツールを確立することにある。(定量的) 構造活性相関; (Quantitative) Structure Activity Relationship; (Q) SAR は、コンピュータトキシコロジーの重要な研究分野であり、有害作用を引き起こす可能性が高い化学物質を、その化学構造から *in silico* で予測する手法である。SAR は統合型毒性評価システムの重要な構成要素の 1 つであり、安全性評価が必要とされる化学物質の優先順位付けや絞り込みに有用である。また、動物実験の代替、もしくは最小化にも貢献できる。SAR は、創薬における探索試験段階での医薬品候補化合物の選択にも利用されている。現在、工業化学物質、農薬、食品添加物、化粧品材料、医薬品候補化合物の毒性予測のため多くのエキスパートシステムや SAR ツールが開発されている。

## 2. 発がん性、遺伝毒性予測と(Q)SAR

発がん性は、ヒトの健康にとって最も関心の高い毒性の一つであり、日常生活において暴露する可能性のある発がん性物質に関する(Q)SAR モデル化の試みは、その他のヒトにおける健康関連のエンドポイントのいずれに対する試みよりも多大な努力が払われている。化学物質の構造から発がん性を予測する研究の歴史も古い。すでに 1930 年頃には強力な発がん物質として知られているベンツピレンの物性と発がん性との関連を明らかにするための研究が行われている。当初は、吸収、蛍光、赤外、NMR スペクトル、イオン化ポテンシャル、電位、磁気異方性、化学反応性などと、発がん性との関係が詳細に調べられたが、発がん性を規定する性質を見つけることはできなかった。1940 年代に Schmidt および Pullman らはベンツピレン分子内の電子分布と発がん性に関係に注目し、一部の芳香族炭化水素の発がん性を合理的に説明することに成功した(1,2)。これら研究が発端となり、いわゆるベンツピレン分子のK領域や、Bay 領域と言われる部分構造と、発がん性との相関が明らかとなった。

1970 年代、James & Elizabeth Miller らは発がん性アルキル化剤の求電子性に注目し、多くの発がん性化学物質は、求電子性誘導体か、もしくは生体内でそれらに代謝されて、発がん標的組織において DNA やタンパク質などの求核性基と結合し、がんを引き起こすという理論を唱えた(3)。それ以降、化学発がんに関する研究は急速に進展した。重要な進歩の一つは、化学発がん研究の重要なツールであるげっ歯類を用いた発がん性試験に対して、より安価で短期間の代替バイオアッセイが確立されたことである。Ames は発がん性化学物質(アルキル化剤、インターカレーターなど)に感受性をもつ一連のサルモネラ変異株を開発し、いわゆるエームス試験を確立した(4)。エームス試験は、発がん性化学物質を検出する *in vitro* モデルといえる。当時の既知の発がん物質の多くは遺伝毒性機序によるものでありエームス試験で陽性

を示す化学物質の作用は、ほとんど **Millers** 仮説の範疇で妥当と考えられた。現在では作用機序の観点から、発がん性物質は以下の二つに分類される。1) 遺伝毒性発がん物質であり、DNA に直接損傷を与え、突然変異を誘発し、これが発がんの第 1 ステップになりうる。2) エピジェネティックな発がん物質であり、これは DNA とは共有結合はせず、直接的に DNA 損傷を起こすことは無い。またエームス試験のような標準的な遺伝毒性試験では通常陰性を示す。1) のカテゴリーに属す遺伝毒性発がん物質が **Miller** 仮説に従った特徴を持ち、それ自体が求電子性であるか、その代謝産物や代謝中間体が求電子性を持つ。

Miller の求電子理論に続いて、Ashby と Tennant は発がん化学物質に対する構造アラート (Structural Alert; SA) と、発がん性予測のコンパイル (SAR モデル) を開発した(5)。発がん性に対する SA は、化学物質の発がん性活性に関連した分子官能基、または部分構造と定義され、同時に発がんの主要なステップである DNA への損傷や、突然変異の誘発をもたらす遺伝毒性の SA とも考えられた。Ashby は米国 National Toxicology Program の 222 の化学物質の中からげっ歯類発がん性試験陽性と強い相関性を示す 18 種類の SA を同定した。Fig. 1 に "Ashby's polycarcinogen" と呼ばれる 18 の全ての SA を持つ仮想の究極発がん物質を示す(5)。Toxnet 公開データベース、Gold/ Zeiger の Carcinogenic Potency Database、イタリアの Istituto Superiore di Santa の動物がん原性 ISSCA データベースに存在する 698 のエームス試験データと、878 のげっ歯類発がん性試験データを用いて Ashby の SAR モデルの検証を行ったところ、発がん性試験結果とは 65% の一致に留まったのに対して、エームス試験結果とは 78% が一致した(6)。このことは、エームス試験用サルモネラ菌株は DNA 応答性に対して正しくデザインされており、Miller の提唱する発がん性物質の求電子理論を適切に反映することを示している。他のバイオマーカーについては SAR モデルとの相関性が低い。このような事実から、SAR モデルの予測率の検証にはエームス試験が使われることが多い。

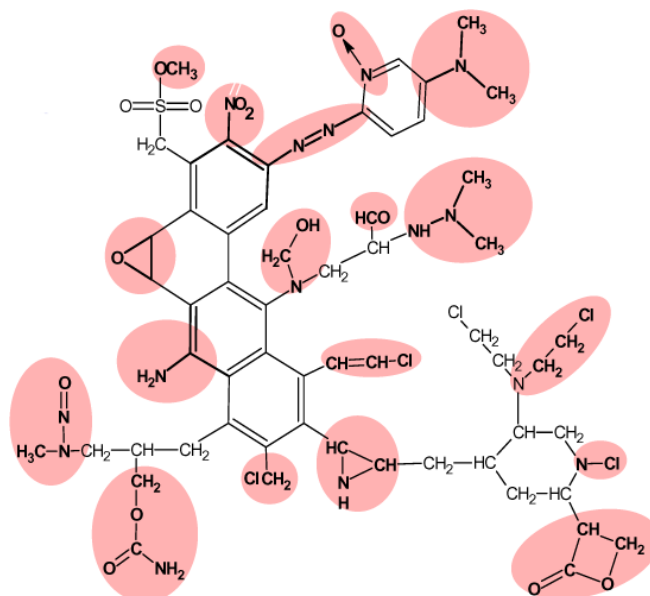


Fig. 1

### 3. 代表的な SAR モデルによるエームス試験の予測

変異原性と発がん性との関連を示す部分構造が、AshbyとTennantによって初めて同定されて以来、これらは標準的な「構造アラート」とされ、その多くが実際の(Q)SARツールに取り入れられている。最も一般的に利用される(Q)SARは、(1) 既知データから陽性をもたらす特徴的な部分構造を定義し、ルール化された経験則から、定性的にエームス試験結果の予測を行うルールベースシステム、(2) 化学物質の構造をフラグメントに分解後、パラメータ（数値データ）に変換し、エームス試験陽性と相関性の高いパラメータを用いて、多変量解析、パターン認識、ニューラルネットワークにより試験結果を予測する人工知能型システム、である。これまで、多くの(Q)SARシステムが、商用および無料ソフトウェアとして世界中で開発されてきた。ICH-M7では特定の(Q)SARシステムの使用を明言してはいないので、どのような(Q)SARシステムも利用可能である。以下に代表的な(Q)SARシステムとその特徴を紹介する。

#### 3.1. 経験に基づくルールベースシステム

##### DEREK

DEREK (Deductive estimation of risk from existing knowledge) は英国ラーサ社が開発・販売する商用のSARシステムである。DEREKは化合物の部分構造と毒性学的作用との間の既知の関係に基づく知識ルールを適用して、毒性作用を予測する。DEREKの知識ルールは、民間企業、大学、公的研究機関、および非営利団体によるデータや知識の提供により継続的開発されている。現行版のDEREK (v. 12) には、89の変異原性アラートが収められている。DEREKは、その知識ベースに構造パターンとしてコード化されたtoxicophore（すなわち、毒性作用に関与すると想定される構造アラート）を有する標的化合物の構造特性を比較することによって予測を導き出す。最終的な予測は、クエリー構造中のtoxicophoreの存在、ならびに分子特性を考慮に入れる推論スキームから導かれる。DEREKによる予測は通常、関連する文献を参照することにより妥当性が確認され、ユーザーはより信頼性の高い予測を得られる。しかし、クエリー化合物が活性である可能性が高いかどうかを、示されたアラートから推論するためには専門的知識が必要である。DEREKの主な長所は、予測の透明性、ルール開発にユーザーグループによる評価を受けている点、および新規ルールの追加が容易な点である。DEREKは陰性予測を行わない点に注意を要する。これは、予測される有害性がない場合、単に関連のあるアラートが同定されなかったことを示し、必ずしも有害性がないわけではない。DEREK陰性はクエリー化合物が単にDEREKでカバーされる化学空間の外にあることを意味している。これまでのDEREK for Windows (DfW)では、クエリー化合物がカバーされない際、ユーザーに知らせる警告はなかったが、最近のDEREK Nexusでは” nothing to report” と表示される。知識ルール開発のために共有される陰性データは、構造アラートに対する除外ルールの作成にも利用できる。ラーサ社は現在、DEREKによる陰性予測に取り組んでいる。DEREK Nexusの今後のバージョンには、境界空間及び対象を記述するためのモデル信頼性ドメインの定義が含まれる予定である。モデル信頼性ドメインは、膨大なデータセットに基づいて定められる。十分な非変異原参

考データセットからの化合物クラスと高い類似性を有する構造は、モデル信頼性ドメインを割り当てられる。したがって、将来、“nothing to report”は“negative”に変わる。

### ToxTree

ToxTreeは、化学物質をカテゴリーに分け、デシジョンツリーアプローチを適用して毒性を予測するJavaベースのオープンソースアプリケーションである。ToxTreeは、欧州委員会Joint Research Centerとの契約によりIdeaconsult社(ソフィア、ブルガリア)によって開発された(7)。ToxTreeは、化学物質の毒性の評価におけるコンピューターによる推定法のアプリケーションに関心がある科学研究者やその他の人々に対するサービスとして、自由に利用できる。変異原性の予測にはBenigni-Bossaルールを実装している(8)。DEREKと同様に、構造アラートをもたない化学物質を、陰性と評価することは理論的に困難である。

### OECD (Q)SAR Toolbox

経済協力開発機構(OECD)が提供する、各種試験データや毒性発現メカニズムに関する既知見を国際的に共有化するために開発された評価支援システムであり、2008年に一般公開された。蛋白結合やDNA結合などの反応様式に基づき、毒性発現の原因となる部分構造を認識する機能(プロファイラー)と、各種エンドポイントの実測試験データベースを備えている。変異原性に関するプロファイラーはBenigni-Bossaルールと、ブルガス大学が開発したOASIS DNA結合プロファイラーである(8,9)。また、OECD Toolboxは外部データを持つ複数のデータベースを備え、分類のサポートに加えて、データベース間での読み取りが可能である。エームス変異原に関する主なデータはa) イタリア高等健康研究所のISSSTYデータベースが持つ7,367種類の化学物質、b) 日本の国立医薬品食品衛生研究所のEXCHEMデータベースが持つ252種類の化学物質、c) ブルガリア、ブルガス大学のOASIS Genotox データベースが持つ7,500種類の化学物質、d) 米国カリフォルニア大学のCPDPデータベースが持つ1,530種類の化学物質、からなる。また、またOECD Toolboxは、166,000種類を超える化学物質の遺伝毒性および発がん性エンドポイントの予測データを含むデンマークEPAのデータベースも備えている。

## 3.2. 統計ベースのQSARシステム

### MCASE (MultiCASE/MC4PC)

MCASE (MultiCASE/MC4PC ; Multiple computer aided structure evaluation) は、1996年にオハイオ州クリーブランドで創設されたソフトウェア会社MultiCASE 社が販売する(Q)SARシステムである(10)。MCASEにトレーニングセットの分子が入力され、プログラムによって処理されると、自動的に各分子構造をサブユニット(記述子)にフラグメント化される。化学構造はKLNもしくはSMILES線形表記法で入力できる。MCASEプログラムでは、活性(biophores ; 変異原性)及び不活性(biophobes ; 非変異原性)分子のフラグメントを比較し、主に活性分子の関与するbiophoresを特定する。さらに、モジュレーターと定義されるQSAR属性及び/又は分子フラグメント、すなわち、共通する構造特性を共有する化学物質の増強又は減退した活性

と関連する分子（部分）構造を特定する。最大5864までの構造で学習されたエームス試験結果予測のためのモデル又はモジュール（例えば、A2H）が市販されており、その多くは米国FDA/CDERのComputational Toxicologyグループと一緒に開発された。さらに、MCASEプログラムは、プロプライエタリデータを用いて、自分専用のモデルに訓練、構築することができる。クエリー化合物はフラグメントに分割された後、これまでに特定及び保存されたbiophores及びbiophobesと比較される。これらのフラグメントの存在の有無に基づいて、活性又はその欠如の確率を予測する。MCASEのアウトプットは、クエリー化合物に対して計算された物理化学的特性、活性フラグメント（biophores）及び/又は非活性化フラグメント（biophobes）及び/又は未知フラグメントの存在又は欠如を示す文字情報として提示される。活性、非活性化及び未知フラグメントの発生率は、対象又は適用ドメインに関する有用な指標となる。3つ以上の未知の構造フラグメントを有する場合は、「ドメイン外」と判断される。MCASEも変異原性の予測に関してこれら未知フラグメントの関連性について決定できないため、「ドメイン外」の予測は、ルールベースシステム同様に専門的知識が必要である。Biophoresとエームス試験結果との関連性は統計的であり、biophoreを導き出すために用いたトレーニングセットから、そのbiophoreが未知分子と関係するかどうかを評価するためには、やはり専門的知識が求められる。尚、現在MultiCASE社はMCASEの次世代モデルである“CASE Ultra”を販売している。これは、最新のWindows OSに対応し、マルチタスク機能を持つなど、使い安いで無く、データベース機能も改良されているため、予測率の向上も期待できる。

## LSMA

LSMA（Leadscope Model Applier）は、米国FDAと共同して作成された遺伝毒性の予測のため（Q）SARツールであり、トレーニングモデル（3575の構造を用いてトレーニングされたエームス変異原性QSARモデル）を含んでいる。Leadscopeは、プロプライエタリデータを用いるQSARモデルの構築に利用される予測因子及び統計的アルゴリズムが装備されているケモインフォマティクスプラットフォームであるLeadscope Predictive Data Miner（LPDM）を提供している。MCASEと同様、各分子構造をサブユニット（構成要素）にフラグメント化する。これらの構成要素は、比較的小さい～非常に大きい分子フラグメントのスペクトルとして表される。Leadscopeは、aLogP、極性表面積、水素結合ドナー及びアクセプターの数、Lipinskiルール違反、回転可能結合の数、親原子数及び親分子量という8つの分子特性を算出する。これらの特性は、偏最小2乗ロジスティック回帰モデルにおける記述子セットとして、フラグメントの集合と一緒に計算に用いられる。クラス所属確率に基づいて、「活性」及び「不活性」の予測が導かれる。用いられた特性の数が少ないため、LSMAはフラグメントに大きく重みを付けている。このプログラムは、トレーニング化合物について距離を測定することによって適用ドメインも評価する。所定の適用ドメイン外であることが分かっている化合物の予測はできない。

## SciQSAR（MDL-QSAR）

LSMAと同様に米国FDAと共同して作成された。SciQSARは、QSARの研究、開発のために必要なアルゴリズム及び統計ツールを含んでいる (11)。プログラムは、240を超える物理化学的、電子位相幾何学的 (electrotopological) E-state指標、連結性 (connectivity) 及びその他の記述子、構造類似性分析のための6つの異なる統計アルゴリズム、ならびに記述子変数選択のための遺伝的アルゴリズムツールを提供する。原子タイプE-state及び連結記述子によって提供される構造空間に基づいて、化学データベースを介した類似性検索が可能である。SciQSARは、試験分子とトレーニングデータセットの活性及び不活性分子間の距離によって決定される、適用ドメインの一基準である、クラス所属確率も計算してくれる。

### 3.3. その他のシステム

#### OASIS/TIMES

ブルガリアのブルガス大学のMekenyanらはルールベースと統計ベースのハイブリッド型の(Q)SARモデルを開発した (7)。このモデルは同大学が販売提供するOASIS/TIMESソフトウェアに組み込まれている。変異原性(Q)SARモデルには、構造アラート、それ以外の分子構造の影響を説明するための修飾因子とともに、DNAと構造アラートとの相互作用メカニズムが含まれている。このメカニズムは学術文献情報から抽出される。構造アラートの決定には専門的知識 (ルールベース) が利用され、変異原性の予測には統計ベースのパターン認識法

(Mechanism-based common reactivity pattern; COREPA) が使用されている。変異原性用のCOREPAモデルは、分子の3D立体構造と電子構造の包括的記述子と特定の警告グループに関連した局所反応性パラメータを併用している。OASIS/TIMESは、代謝経路に基づく肝代謝シミュレーターを備えている (Tissue Metabolite Simulator; TIMES)。このモデルで使用したトレーニングセットの化学物質は、代謝活性化なしで変異原性を持つもの、代謝活性化後に変異原性を持つもの、および代謝活性化の有無を問わず変異原性が無いものに分類される。クエリー化合物をラットS9存在下でのエームス変異原性を予測すると、化合物の代謝マップとともに、変異原性を持つと予測される代謝物を表示してくれる。他に、代謝の役割が考慮されるモデルがほとんどないことから、これはOASIS/TIMESの最大の利点である。代謝の予測には既存の文献データから得られた薬物代謝に関する変換率の情報を利用される。また、データがない場合は組み合わせアルゴリズムを用い、既知の代謝マップを適合性が最も高い変換確率に書き換える。

## 4. 参考文献

- 1) Schmit, O.: *Physik Chem B42*, 83 (1939)
- 2) Pullman, A. and Pullman, B.: *Rev Sci* 3, 117 (1946)
- 3) Miller, J.A. and Miller, E.C.: *Cancer* 47, 2327-2345 (1981)
- 4) Ames, B.N., *Cancer* 53, 2030-2040 (1984)
- 5) Ashby, J. and Tennant, R.W.: *Mutat Res* 204, 17-115 (1988)
- 6) Benigni, R. and Bossa, C: *Mutat Res*, 659, 248-261 (2008)

- 7) Lahl U & Gundert-Remy: The use of (Q)SAR methods in the context of 1297 REACH. *Toxicol. Mech. Methods* 18, 149–158 (2008)
- 8) Benigni R & Bossa C: Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. *Mutat. Res.* 659, 248–261 (2008)
- 9) Serafimova R et al: Identification of the Structural Requirements for Mutagenicity, by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals. II. General Ames Mutagenicity Model. *Chemical Research in Toxicology* 20, 662-676 (2007)
- 10) Klopman G: MULTICASE. 1. A hierarchical computer automated structure 1287 evaluation program. *Quant. Struct. Act. Relat.* 11, 176–184 (1992)
- 11) Contrera JF et al: In silico screening of chemicals for bacterial mutagenicity using electrotopological E-state indices and MDL QSAR software. *Regul. Toxicol. Pharmacol.* 43, 313–323 (2005)