

毒性試験の未来を考える
— (定量的) 構造活性相関による化学物質の変異原性評価 —

本間正充

The future of toxicity tests
Mutagenicity assessment of chemicals substances by (Quantitative) Structure Activity Relationship

Masamitsu Honma

Presently, more than 100,000 chemicals substances are industrially produced and present in our living environments. Some of these may have adverse effects on human health. Given the rapid expansion in the number of industrial chemicals, international organizations and regulatory authorities involved in the regulation of chemical substances have expressed the need for effective screening tools to promptly and accurately identify chemical substances with potential adverse effects without conducting actual toxicological studies. (Quantitative) Structure Activity Relationship ((Q) SAR) is one promising approach to predict the potential adverse effects of a chemical based on its chemical structure. Much effort has been invested in the development of (Q) SAR models for predicting Ames mutagenicity, among other toxicological endpoints, owing to the significant amount of the necessary Ames test data that have already been accumulated. The International Conference on Harmonization (ICH) M7 guideline for the assessment and control of mutagenic impurities in pharmaceuticals was established in 2014. It is the first international guideline that addresses the use of (Q) SAR models instead of actual toxicological study for human health assessment. Therefore, (Q) SAR models for Ames mutagenicity now require higher predictive power for identifying mutagenic chemicals. This review introduces the advantages and features of (Q) SAR, several (Q) SAR models for predicting Ames mutagenicity, and approaches to improve (Q) SAR models. Finally, we discuss the future of (Q) SAR and other advanced *in silico* tools in genetic toxicology.

Keywords: (Quantitative) Structure Activity Relationship ((Q) SAR) , Ames test, ICH-M7, rule-based, statistical based, prediction power, benchmark data set, deep learning, AI

1. はじめに

2015年7月に米国化学会 (American Chemical Society) の情報部門の化学情報の世界的権威である Chemical Abstracts Service (CAS) は化学物質データベースに、1億件目となる化学物質が登録されたこと

を発表した。それからわずか4年後の現在、その数は1億5千万を超えようとしている (<https://www.cas.org/>)。このペースは2分30秒で1個の新規の化学物質が登録されている計算となり、今後このペースはさらに加速することから、10年後にはその数は10億を超えると予測されている。これらの化学物質のうち、現在、約11万種類程度が工業的に生産され、我々の生活環境に存在しており、その数も増加している。環境中に放出された化学物質の中には地球環境や、生態系に影響を与え、地球温暖化や動植物資源の破壊を引き起こしたり、また、人間の健康に悪影響を及ぼしたりするものがある。従って、これら化学物質を適切に評価・管理する必要がある。

To whom correspondence should be addressed:

Masamitsu Honma; Division of Genetics and Mutagenesis, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki City, Kanagawa 210-9501, Japan; Tel +81-44-270-6672, Fax +81-44-270-6680; E-mail: honma@nihs.go.jp

る。2002年に南アフリカのヨハネスブルクで開催された地球サミット（WSSD）では、「予防的取組方法に留意しつつ、透明性のある科学的根拠に基づくリスク評価・管理手順を用いて、化学物質が人の健康と環境にもたらす著しい悪影響を最小化する方法で使用、生産されることを2020年までに達成することを目指す」（WSSD2020年目標）ことが合意されている。先進国の規制当局にはこのWSSD目標達成に向けた取り組みが求められている。日本では1968年に制定された「化学物質の審査及び製造等の規制に関する法律（化審法）」により、新規の化学物質に関しては、事前にその性状（分解性、蓄積性）や、環境及び人への健康影響を審査し、製造、輸入、使用等について必要な規制を行っている。化審法に基づき、年間、数万種類の新規化学物質が登録されているが、経済産業省、環境省、厚生労働省による合同審議会で審査されている化学物質は年間の製造・輸入量が10トンを超える約400種程度の高生産物質に限られ、その他の多くの化学物質は環境や人への健康影響は未評価にまま登録されているのが現状である。

化学物質の安全性は、一般的に動物、細胞、微生物等を用いた生物学的試験によって評価されるが、このような膨大な数の化学物質を試験することは、労力、時間、費用および動物福祉の問題を考慮すると非現実的である。WSSD目標達成には生物学的試験に頼らずに有害な化学物質を迅速、かつ正確に識別するための効果的なスクリーニングツールが必須である。（定量的）構造活性相関（(Quantitative) Structure Activity Relationship; (Q) SAR）は、有害作用を引き起こす可能性が高い化学物質を、その化学構造からインシリコで予測する手法である。現在、化審法においても各省が（Q）SARツールを用いて有害性予測を行っているが、その予測結果は参考として審議会に供されているだけであって、実用化には至っていない（表1）。

表1 化審法において利用されているQSARツール

	評価項目	QSARツール
経済産業省	分解性	BIOWIN5 BIOWIN6 CATALOGIC
	蓄積性	BCFWIN Amot-Gobas Model Baseline Model
環境省	生態影響	TIMES ECOSAR KATE
厚生労働省	人健康影響 (エームス変異原性)	DEREK Nexus CASE Ultra TIMES_Ames

2. 変異原性と（Q）SAR

多くの化学物質の毒性は、これ以下であれば毒性影響が見られないレベル、すなわち閾値が存在する。従って、曝露量が十分に少なければその毒性による健康リスクはゼロと見なすことができる。一方、毒性の中でも変異原性（Mutagenicity）は化学物質による発がん過程の重要なメカニズムである突然変異誘発性であり、DNAと化学物質の化学反応性に基づく。突然変異はゲノムの不可逆的、かつ永続的変化であり、たった1つの突然変異からでもがん原細胞になり得る可能性があり、閾値が設定できない。別の言葉として遺伝毒性（Genotoxicity）がある。こちらは広義であり、DNAや染色体に損傷を与え、構造的、もしくは量変化を引き起こすが、必ずしも突然変異を誘発するわけではない。遺伝毒性があるが、変異原性が認められないものはDNAに対して損傷能力はあるが、がんや引き起こす遺伝子変化の直接的証拠にはならない。また、変異原性以外の遺伝毒性には閾値が設定できるとされている¹⁾。従って、曝露量が低くても変異原性の評価、特にその有無は化学物質の発がんリスクを評価する上で重要な決定因子となる。そのため、変異原性試験の代表的試験であるAmes試験は、食品中に微量に含まれる残留農薬、食品添加物、プラスチック容器からの溶出物、また医薬品中の不純物のような極めて低曝露の化学物質の安全性を評価する際には常に求められる。

本稿では主にAmes試験の（Q）SARについて述べる。化学物質のヒト健康影響評価に対する（Q）SAR研究ではAmes変異原性予測（Q）SARの研究が最も進んでおり、実用化段階にあるが、その理由は上記に示したように、Ames試験自体の重要性だけでなく、以下の点が挙げられる。

- 1) 変異原性化学物質は基本的に電子求引性化学構造を持ち、物理化学的性質から変異原性を説明可能な科学的メカニズムが存在する²⁾。
- 2) Ames試験結果は再現性がよく、他の試験に比べて結果が安定している。また、毒性試験の中でも試験データの蓄積数が多い。これら大量で安定した試験データを基に（Q）SARによる予測モデルが作りやすい。
- 3) 2014年に医薬品規制調和国際会議（ICH）において、医薬品に含まれる不純物のAmes変異原性評価に（Q）SARの利用がガイドライン（ICH-M7）に明文化された³⁾。このガイドラインがきっかけで、（Q）SARによる毒性予測、特に変異原性予測の研究がこの数年の間に大きく発展した。

（定量的）構造活性相関（(Q) SAR）は化学構造と毒性の相関に関する研究であり、構造的に類似した物質の

既知の活性に基づき、生物学的毒性試験を実施することなく評価対象物質の毒性を予測することである。化学物質の毒性には一般的に量的相関性があるため、(Q) SARによる毒性の予測は、エンドポイントとなる毒性の発現用量を予測することが本来の目的である。一方、Ames試験を含む変異原性試験（遺伝毒性試験）では、定量的評価ではなく、その結果は変異原性の有無（陽性もしくは、陰性）の2元性である。このような定性的な評価結果が(Q) SARモデルを作りやすいことも理由にある。すなわち、予測モデルの検証（正解、不正解）が容易である。従って、Ames変異原性予測は構造活性相関(SAR)とするのが正しく、(Q)をつけているのはこの理由による。本稿では、便宜のため以降QSARに統一して用いる。

3. QSARによるAmes変異原性の予測

化学物質の構造から変異原性、発がん性を予測する研究は古くから行われている。1960年代、James & Elizabeth Miller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求電子理論を唱えた⁴⁾。それ以降、化学発がんに関する研究は急速に進展した。Bruce 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Ames試験を確立した⁵⁻⁷⁾。Ames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。変異原性機序による発がん物質のほとんどは

Ames試験で陽性を示し、Millers仮説の範疇で妥当と考えられた。Millerの求電子理論に続いて、John AshbyとRay 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を開発した。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変異原性のSAとも考えられた^{8,9)}。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（図1）。現在、Ames変異原性QSARモデルはルールベース

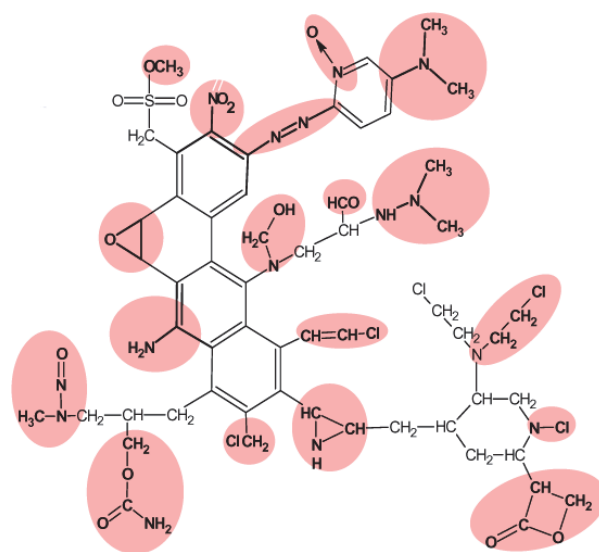
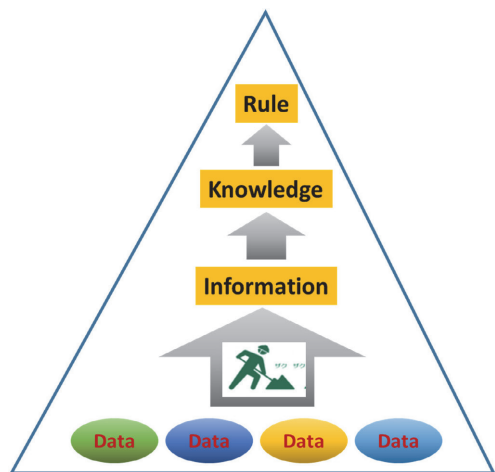


図1 Ashbyらによって提唱された発がん性を有する構造アラート

ルール（知識）ベースQSAR

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Ames試験結果の予測を行う。



統計ベースQSAR

化学物質の構造をフラグメントに分解後、その構造を幾何学的、電子的、物理化学的等の記述子（数値データ）に変換し、エームス試験陽性と相関性の高い記述子を用いて、多変量解析、パターン認識などの機械学習により試験結果を予測する人工知能型モデル。

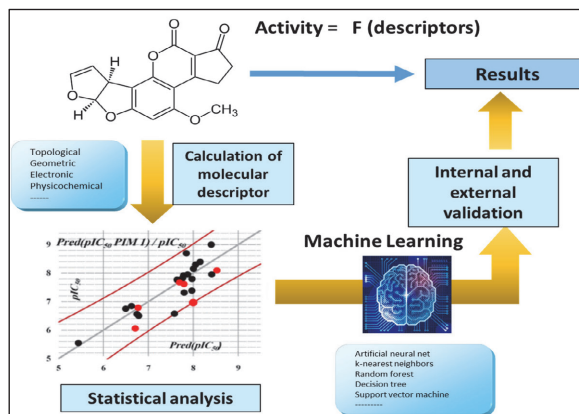


図2 ルール（知識）ベースQSARと統計ベースQSAR

と、統計ベースの2つのモデルが主であるが（図2）、ルールベースQSARはAshbyらが行ったように、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Ames試験結果の予測を行うものである。一方、統計ベースQSARは、化学物質の構造をフラグメントに分解後、その構造を幾何学的、電子的、物理化学的等の記述子（数値データ）に変換し、Ames試験陽性と相関性の高い記述子を用いて、多変量解析、パターン認識などの機械学習により試験結果を予測する人工知能型モデルである。

これまでAmes変異原性QSARに関しては、学術的目的、商用目的で多くのQSARツールが開発されてきた。経済協力開発機構（OECD）ではQSARが今後の化学物質の有害性評価に積極的に利用されることを鑑み、QSAR 評価の国際調和の原則を示している（表2）。尚、この原則には拘束力がなく、その運用は各国に委ねられている。

4. 代表的なAmes変異原性QSAR予測ツール

4.1. Derek Nexus (Lhasa Limited, UK)：ルールベースQSAR

Derek (Deductive estimation of risk from existing knowledge) NexusはLhasa Knowledge Suiteの一部として市販されているルールベースのエキスパートQSARシステムである。ここに含まれる知識ルールは、構造アラート、化合物例および代謝活性化とメカニズムに関する知見を考慮して作られている。これら知識ルールは、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。Derek Nexus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（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 Nexus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。Derek Nexus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。また本システムがルールベースのシステムであるため、規定されたトレーニングセットや適用範囲はない。しかし、最近実装された構造分類機能により、ユーザーが陰性予測を実証することが可能となった¹⁰⁾。推論レベルが曖昧またはそれ以上のアラートを持つすべての化合物は、本システムにおいて陽性予測として処理されている。

4.2. Sarah Nexus (Lhasa Limited, UK)：統計ベースQSAR

Lhasa Knowledge Suiteには、Sarah Nexusと呼ばれる統計ベースのAmes変異原性モデルも含まれている。このモデルでは、自己組織化仮説ネットワーク（SOHN）を用いて、トレーニングセット内において活性または不活性への関連が認められた部分構造の存在に基づきクエリー化合物の活性仮説を提示する¹¹⁾。仮説を統合し、クエリー化合物の活性についての総合的に予測し、各予測の信頼性も示される。これは各仮説におけるクエリー化合物とトレーニングセット内の最近傍データとの類似性によって決定される。次に、個々の仮説の信頼性を統合し、クエリー化合物についての総合的予測および予測の信頼性が得られる（この総合的予測は、陽性、陰性、曖昧または適用範囲外である可能性がある）。ユーザーには、総合的予測に加え、仮説、関連するトレーニング

表2 QSAR 5原則（OECD QSAR Validation Principlesに基づく）

1. 予測するエンドポイントを明確にすること
予測モデルのデータセット／試験系を明確にすること（遺伝毒性、変異原性を予測するのではなく、エームス試験結果、染色体異常試験結果を予測する）。
2. 使用する予測モデルやアルゴリズムを明確にすること
モデルの種類や（ルールベース、統計ベース）、モデル構築に用いた手法（アルゴリズム、記述子等）を明らかにし、その透明性を担保すること。ただし、商用目的のモデルの場合、この情報は必ずしも公開されないことが多い。
3. 予測可能な適用範囲を明確にすること
QSARの予測性は、モデル構築に使用されたトレーニングセットに依存するため、精度の高い予測を行える化学物質の種類には限りがある。そのため、QSARモデルを適用できる化学構造の限界を明らかにすること（Out of Domainの明示）。
4. 予測モデルの性能を適正に評価すること
予測モデルの適合度と頑健性は、内部トレーニングセットによって評価すること。また、その予測性は、適切な外部データセットを用いて判定されるべきである。
5. 予測結果のメカニズムを示すこと
可能であれば、モデルの記述子と予測エンドポイントとのメカニズム的な関連性を示すこと。メカニズムによる解釈が可能であれば、原則-3の適用範囲の一部にもなり得る。

セットの例が、関連するメタデータと共に提示される。

4.3. CASE Ultra (MultiCASE Inc., USA) : 統計ベースQSAR

米国クリーブランドにあるMultiCASE社はQSAR研究の祖といわれるGilles Klopmanによって1966年に設立された。本QSARツールは、統計的手法をベースとし、機械学習技術によってトレーニングデータから自動的にアラートが抽出される。トレーニングに必要なデータは化学構造とその毒性ラベルのみである^{12,13)}。クエリー化学物質について予測される毒性の程度は、特定されたアラートだけでなくアラート周囲の構造的環境にも依存する。アラート周囲の構造的特徴は「モジュレーター」と呼ばれ、これらもトレーニングデータから自動的に学習される。このアルゴリズムでは、連続した毒性エンドポイントのQSARモデルを構築するために、様々な物理化学的パラメータおよび記述子が使用されている。Ames変異原性に関するCASE Ultraの主なモデルは、GT1_AT_ECOLI, GT1_A7B, PHARM_ECOLIおよびPHARM_SALMの4つのモジュールで構成されているが、2018年からの新バージョンでは*Salmonella/E. Coli* コンセンサスモデルであるGT1_BMUTが、2019年にはPharm_BMUTの新モジュールがリリースされた。

4.4. Leadscope Model Applier; LSMA (Leadscope Inc., USA) : 統計ベースQSAR

LSMAの統計的変異原性QSARモデルは、米国食品医薬品庁と共同で開発された統計ベースのQSARツールである。これらのモデルは、過去に公表されたAmes変異原性データのトレーニングセットから構築され、構造記述子は、(1) 事前に定義された構造的特徴、(2) 自動的に生成される化学骨格、(3) 外部知識、(4) 算出された特性に基づいている。選択した記述子(独立変数)およびAmes変異原性データ(応答変数)を用いて部分ロジスティック回帰モデルによりAmes試験結果を予測する。モデルからは、陽性の結果が得られる確率、活性化および不活性化に寄与する構造的特徴、ならびにトレーニングセットの化学物質に関する詳細な試験情報等を提示してくれるため、専門家の詳細なレビューが可能である。

4.5. TIMES_AMES (Bourgas University, Bulgaria) : ルールベースQSAR

TIMES_AMES QSARツールはブルガス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。Ames/QSARツールには、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。構造アラートの決定にはルールベースが利用され、変異原性の予測にはパターン認識法 (Mechanism-based Common REactivity Pattern; COREPA) が使用されている。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^{14,15)}。クエリー化合物をラットS9存在下でのAmes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。これがOASIS/TIMESの最大の利点である。

4.6. Toxtree (Istituto Superiore di Sanità, Italy) : ルールベースQSAR

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Ideacon社 (Bulgaria) によって開発された。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。Ames変異原性の予測にはBenigni/Bossaルールを実装している^{16,17)}。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。

4.7. ADMEWORKS (FUJITSU KYUSHU SYSTEMS LIMITED, Japan) : 統計ベースQSAR

ADMEWORKSは株式会社富士通九州システムズが開発した2つの統計モデル (1つは高感度、もう1つは高特異性を有する) からなるコンセンサスモデルである。主として米国National Toxicology Program (NTP) から得られた1,977種類の化合物をトレーニングセットとして開発された。2つの統計モデルの結果が一致する (すなわち、両方が陽性または両方が陰性である) 場合に最終予測結果が提示される。それ以外は予測を行わない。コンセンサスモデルに加え、元のデータから陽性構造アラートと陰性構造アラートを抽出し、コンセンサスモデルで予測する前にフィルターとして使用している。化合物にいずれかの陽性アラートが含まれている場合は、直ちに「陽性」に分類され、それ以外でいずれかの陰性アラートが含まれている場合は、直ちに「陰性」に分類される。化合物にアラートが含まれていない場合、

コンセンサスモデルを用いて最終予測を行う。

4.8. ChemTunes•ToxGPS (Molecular Networks GmbH and Altamira LLC, USA) : 統計ベース QSAR

ChemTunesモデルは、規制関連、文献および主要な情報源から収集した*in vivo*および*in vitro*毒性データのToxGPS知識ベースに基づいている。予測は、Dempster-Shafer理論に基づく定量的な証拠の重みづけの手法を用いて統合した化学種アラートおよび作用機序に基づくQSARモデルに基づいている。このモデルでは、構造ベース (ToxPrint化学種など)、特性ベース (双極子モーメント、溶解度、logPなど)、および量子力学的記述子を使用する。ChemTunes•ToxGPSモデルは、ヒトの健康に関連する多様な毒性エンドポイントに利用可能である。決定理論アプローチは、不確かさを推定し、複数のモデルからの予測を組み合わせて記述されており、各証拠の情報源の信頼性が考慮されている。ここには示されていないが、ChemTunes•ToxGPS予測システムは、各予測に関連する不確かさの推定値も作成する。

4.9. MUT_Risk (Simulations Plus Inc., USA) : 統計ベースQSAR

MUT_Riskは、10の個別の人工神経ネットワークアンサンブル (ANNE) 分類モデル (5つの菌株で、ラットS9の存在下および非存在下での個別の10のAmes試験結果) による変異原性予測を要約したADMET Risk™スコアである。5つの±S9モデルペアのそれぞれが陽性分類をするごとにスコアにポイントを追加する。判定の閾値はユーザーが設定する。MUT_Risk-0はスコアが0より大きい場合に化合物を変異原性とみなすが、MUT_Risk-1はスコアが1より大きい場合に化合物を変異原性とみなす。これにより、各アプリケーションに応じて感度と特異度のトレードオフを調整することができる。寄与する10の各モデルは、それぞれ固有の範囲外フラグと不確かさの推定値を有する。

4.10. StarDrop Auto-Modeler (Optibrium Ltd., UK) : 統計ベースQSAR

StarDropは統合型QSARツールであり、知識ベース型の毒性予測機能としてDerek Nexus、統計ベース型予測機能としてAuto-Modelerがある。Auto-Modelerでは、構造情報と予測対象物性値を含むデータセットから、ほぼ自動的に独自予測式を作成することが可能である。つまり、記述子情報を用意する必要はなく、9種類の分子全体に関与する記述子 (分子量やlogP予測値など)、321種類のSMARTS形式の記述子 (原子種や特性

基の数) を自動生成する機能が含まれている。独自の記述子を追加することも容易であり、応用範囲は広い。解析手法には、10種類のContinuousモデルおよび3種類のCategoryモデルが含まれており、最良モデルを自動判定する機能も含まれている。これらの優れた構造情報処理機能や統計処理の自動化により、非専門家による物性予測モデル作成を支援するだけでなく、専門家の試行錯誤プロセスの効率化も支援する。

4.11. VEGA (Istituto di Ricerche Farmacologiche Mario Negri, Italy) : コンセンサスモデル

VEGAはイタリアのIstituto di Ricerche Farmacologiche Mario Negriで開発されたJavaベースのシステムで、無償でダウンロードすることができる。VEGAには、異なるデータセットによって開発された複数のQSARモデルが搭載されており、その計算結果は同一の表示方法で出力される。最新のVEGAバージョン 1. 1. 4の「Mutagenicity (Ames test) model」には以下の4つのモデルが搭載されており、最終的にはコンセンサスモデルがこの4つのモデルの結果に基づき結論を導く。

- 1) CAESAR : サポートベクターマシーン (SVM) による統計ベースQSAR
- 2) SarPy : 部分構造に基づくルールベースQSAR
- 3) ISS : Benigni/Bossa ルールに基づくルールベースQSAR
- 4) KNN : istKNN を用いた類推 (リードアクロス)

5. QSARツールの予測性の評価

Ames試験結果は陽性、陰性の2元性であるためQSARによる計算結果との一致性からその予測性を客観的に数値化し評価することができる。真の陽性 (TP)、偽の陽性 (FP)、偽の陰性 (FN)、および真の陰性 (TN) の2×2予測結果マトリックスを表3に示す。表4に感度 (陽性物質を検出する能力)、特異度 (陰性物質を検出する能力)、陽性および陰性予測率 (予測されたものの正解率)、精度、バランス精度、適用範囲、Mathews'

表3 QSARモデルの予測結果マトリックス

		実際の試験結果	
		陽性	陰性
QSAR 予測結果	陽性	真の陽性 (TP)	偽陽性 (FP)
	陰性	偽陰性 (FN)	真の陰性 (TN)
	予測不能 (OOD*)	—	—
総数	Total		

*Out of Domain

表4 QSARモデルの予測性評価指標

評価指標	計算方法と説明
感度 (Sensitivity)	$TP/(TP+FN)$ 陽性物質を陽性と判定できる能力
特異度 (Specificity)	$TN/(FP+FN)$ 陰性物質を陰性と判定できる能力
精度 (Accuracy)	$(TP+TN)/(TP+TN+FP+FN)$ 総合判定能力
バランス精度 (Balanced Accuracy)	$(\text{感度} + \text{特異度})/2$ 陽性、陰性のバランスを考慮した総合判定能力
陽性予測率 (Positive Prediction Value)	$(TP)/(TP+FP)$ 陽性予測結果の正解率
陰性予測率 (Negative Prediction Value)	$TN/(TN+FN)$ 陰性予測結果の正解率
Mathews'相関係数 (MCC)	$\frac{((TP*TN)-(FP*FN))}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$ モデル全体のパフォーマンス. $-1 \sim 1$ までの数字で表され, 0 以上が予測能力あり
適用率 (Coverage)	$(TP+TN+FP+FN)/\text{Total}$ 予測できた化合物の割合

相関係数 (MCC) の計算法を示す. 優れたQSARモデルとは高い感度 (変異原性物質を確実に検出する), 高い陰性予測率 (陰性と予測されたものは実際に陰性である), 高い適用範囲 (できるだけ多くの化学物質を判定できる) を持つことが重要とされる. ただし, これらの数値は評価するデータセットのバランスに大きく依存するため注意する必要がある. 評価対象となるデータセットは一般に陰性データが多い. データセットの陽性, 陰性の割合が10%, 90%の場合, 感度が40%でも優れたQSARモデルと言える. バランス精度, MCCはデータセットの偏りを考慮した予測指標であり, 異なる評価対象のデータセットによるQSARツールの予測性の評価に有用である.

6. ICH-M7でのQSARの適用

2014年に医薬品規制調和国際会議 (ICH) において, 医薬品に含まれる不純物の変異原性評価にQSARの利用がガイドライン化された (ICH-M7). これはQSARが実際の生物学的試験の代替として人健康影響評価に用いることが許可された最初のガイドラインである³⁾. 医薬品中の含まれる不純物はきわめて微量であり, 種類も多く, また不安定である場合もあり, それらを単離, 精製し, 実際の生物学的試験で毒性を評価することは, 不可能であることが多い. 不純物の化学構造的な見えかればQSARで変異原性を評価することを許容する本ガイドラインはきわめて現実的であると言える. ICH-M7では

Ames変異原性の評価に知識ベースと統計ベースの両者のQSARツールの使用を求めている. これらQSARツールは先に述べたOECDバリデーション原則に従っていれば, どのQSARツールでも使うことができる. 二つのQSARツールの両方でAmes変異原性を示す警告構造がないことが示されれば, その不純物はAmes陰性と結論することができ, 更なる試験の必要ない. また, 両者で予測結果が異なる場合や, 予測ができなかった場合は, 専門家による適切なレビューにより最終結論を導くことができるとされている. しかしながら, 具体的な専門家判断の手法については示されていない. 我が国では, 日本環境変異原学会が中心となり, この専門家判断の手法の開発のためのワークショップを毎年開催している^{18,19)}. また, ICH-M7では, 現在Q&Aの策定を進めているが, ここにはQSAR結果の専門家判断の基準や要件等を示す予定である.

7. QSARツールの予測性の向上の目的としたAmes/QSAR国際チャレンジプロジェクト

2014年にICH-M7が制定されたことから, QSARは, これまでのAmes変異原性を予測するためのプレスクリーニングの目的から, 実際の化学物質の変異原性を評価するための試験として扱うことになった. 従って, QSARによるAmes変異原性予測にはこれまで以上の予測精度が要求される. しかしながら, 現実的には多くのQSARツールの予測精度は十分ではなく, 特に新規の化

学物質に対する感度は50%以下であると報告されている²⁰⁾。QSARの予測精度の向上させるためには、トレーニングデータとしての実際のAmes試験データの蓄積が重要である。多数、かつユニークな構造を持つ化学物質からなるトレーニングデータはケミカルスペースを拡大し、偽陽性や偽陰性を低減化し、予測精度を向上させる。また、その試験データの信頼性も重要である。現在、ウェブサイト等で公開されているAmes試験結果を有する化学物質数は約10,000程度であり、多くのQSAR開発者は基本的にこれらデータベースを基にQSARモデルを開発している²¹⁾。

我が国では新規化学物質の安全性は化審法だけでなく安衛法（労働安全衛生法）においても評価されている。安衛法は「職場における労働者の安全と健康を確保」する目的で、年間生産量もしくは輸入量が100kg以上の新規化学物質については、Ames試験の実施が要求され、強い変異原性が認められた化学物質に対しては、事業者には暴露を低減するための措置を求めている。1979年以降、安衛法に基づき実施されたAmes試験は25,000化学物質を超え、世界最大規模のAmes試験ビッグデー

タベースになっている。また、安衛法試験データは基本的には非公開で有り、これまでQSARの予測モデルの開発に利用されていない。変異遺伝部では厚生労働省厚生労働省・労働基準局安全衛生部・化学物質対策課化学物質評価室の許可を得て、このジャパנקオリティの試験データを世界中のQSARベンダーに提供し、Ames/QSARツールの予測精度の比較と向上のための国際チャレンジプロジェクトを実施した。世界7カ国から12のQSARベンダーが開発した17のQSARツールが本プロジェクトに参画した（表5）。プロジェクトは3回のフェーズに分けて2014年から2017年まで実施された。各フェーズ約4,000の化合物をQSARベンダーに提供し、QSARによる評価と検証を繰り返した（表6）。注目すべきは、その対象となる評価物質である。各フェーズの評価物質は無作為に振り分けたにもかかわらず、陽性物質の割合は約15%程度に一定している。このことは、我が国で流通する新規化学物質の約15%は変異原性を有し、おそらくその多くには発がん性があることが予測される。変異原性発がん性物質には閾値がないことを考えると、たとえ曝露量が低くとも発がんリスクは存在

表5 Ames/QSAR国際チャレンジプロジェクト参加機関

QSAR ベンダー	QSAR ツール名	QSARモデル
1. Lhasa Limited (英国)	a. Derek Nexus	ルール
	b. Sarah Nexus	統計
2. MultiCASE Inc (米国)	c. CASE Ultra statistical-based	統計
	d. CASE Ultra rule-based	ルール
3. Leadscope Inc (米国)	e. Leadscope statistical-based	統計
	f. Leadscope rule-based	ルール
4. Istituto di Ricerche Farmacologiche Mario Negri (イタリア)	g. CAESAR	統計
	h. SARPY	ルール
	i. KNN	統計
5. LMC - Bourgas University (ブルガリア)	j. TIMES_AMES	ルール
6. Istituto Superiore di Sanita (イタリア)	k. Toxtree	ルール
7. Prous Institute (スペイン)	l. Symmetry	統計
8. Swedish Toxicology Science Research Center (スウェーデン)	m. AZAMES	統計
9. FUJITSU KYUSHU SYSTEMS LIMITED (日本)	n. ADMEWORKS	統計
10. IdeaConsult Ltd. (ブルガリア)	o. AMBIT	統計
11. Molecular Networks GmbH and Altamira LLC (米国)	p. ChemTune•ToxGPS	統計
12. Simulations Plus, Inc (米国)	q. MUT_Risk	統計

表6 Ames/QSAR国際チャレンジプロジェクト評価対象物質

Ames 試験結果	Phase I (2014-2015)	Phase II (2015-2016)	Phase III (2016-2017)	合計 (2014-2017)
陽性	556 (14.5%)	562 (14.7%)	629 (14.3%)	1757 (14.4%)
陰性	3,336 (85.5%)	3,267 (85.3%)	3,780 (85.7%)	10,383 (85.6%)
合計	3,902	3,829	4,409	12,140

する。毎年数多くの新規化学物質が生産され、生活環境中に放出されるが、これらの中から変異原性物質を同定し、ヒトへの暴露を最小限にするように管理することが公衆衛生上重要であると考え。安衛法では強いAmes試験陽性を示す化学物質については法に基づいてWeb上で一般公開し、事業者に注意を呼びかけている。これら強い変異原性物質はAmes/QSAR国際チャレンジプロジェクトのホームページでも見ることができる (<http://www.nihs.go.jp/dgm/amesqsar.html>)。Ames/QSAR国際チャレンジプロジェクトの3回のトライアルの結果、全てのQSARツールの予測精度は著しく改良された²²⁾。多くは60%以上の感度、80%以上の特異度、70%以上のバランス精度を示した。このことは本プロジェクトが成功裡に終わったことを意味する(表7)。安衛法に基づ

表7 Ames/QSAR国際チャレンジプロジェクト結果

	Phase I	Phase II	Phase III
Sensitivity (%)	56.7 (38.6-70.0)	58.0 (41.6-72.1)	57.1 (31.7-67.6)
Specificity (%)	77.7 (62.5-91.5)	84.2 (64.9-92.8)	79.9 (60.7-93.0)
Accuracy (%)	74.7 (63.6-83.9)	80.3 (65.8-87.7)	76.7 (68.0-87.3)

くAmes試験は現在でも年間1000物質以上実施されている。今後、蓄積するデータによりチャレンジプロジェクトが継続的に行われることを望む。また、企業が持つ非公開のAmes試験データも貴重なデータベースである。化学物質の毒性試験データは公衆衛生上、公共性が高く、公開が望まれる。実現すればQSARの予測率は格段に向上するだけでなく、化学物質の安全性の透明化が図られる。

7. 化学物質安全性ビッグデータベースの構築とAI、ディープラーニングを用いた医薬品・食品・生活化学物質のヒト安全性予測基盤技術の開発

国立衛研ではこれまで長年にわたって整備してきた大規模で信頼性の高いヒト健康に係る毒性試験データを統合・拡充したビッグデータベースの構築を目指している。2018年からはインハウス研究の一つとして、このデータベースの構築と、医薬品・食品・化学物質3分野にまたがるレギュラトリーサイエンスに基づく安全性評価の専門的知見並びに高精度の安全性研究の経験と、Artificial Intelligence (AI)、ディープラーニング等の人工知能を統合した安全性予測プラットフォームの開発と実用化のための研究を開始した。このプラットフォームは医薬品、食品、生活化学物質を対象としており、規制が異なる分野の知見を統合することで、毒性評価の迅速

化・効率化・高度化を実現し、医薬品等の副作用の未然防止、生活・食品化学物質の信頼性の高い安全性評価基準の設定、さらには我が国の産業競争力強化に貢献できることが期待される。

多くの毒性試験データの中でもAmes試験データが最も多いことから、本プロジェクトでもディープラーニングを用いたAmes試験予測モデルの開発を行っている。プロトタイプモデルでは化学構造のSMILES標記を畳み込ニューラルネットワーク (Convolutional Neural Network; CNN) を用いて特徴量の自動抽出を行っている。通常CNNは画像分類に用いられるが、SMILES等の文字情報をベースとしたテキスト分類にも応用可能である。まずSMILES標記の1文字を70次元の one-hot ベクトルに変換する。その後、最大文字数による正規化を行い「70×最大文字数」の二次元配列として画像と同様に畳み込を行う。この手法により、文字の前後にどの文字が記されているかの情報を含めることができるため化学構造に関する特徴量が検出できる。SMILES-CNNモデルで16,651化学物質のAmes試験データを学習させ、新規の2,000物質の評価データを用いてその精度を検証したところ、感度55%、特異度80%、精度77%の結果を得ることができた。この予測率は他のQSARツールに匹敵する。現在、学習データセットの充実、説明変数の追加により予測精度の向上を目指している。また、SMILES-CNNだけでなく、3次元化学構造と隣接する元素や部分構造を多次元に解析するGraph-CNNモデルの導入も試みている²³⁾。

8. さらなる予測率向上のためには何が必要か？

これまで述べたようにAmes変異原性の予測精度はQSARで80%、SMILES-CNNで77%程度である。この精度をさらに改良し、100%に近づけるためには何が必要であろうか？問題は予測モデルにあるのではなく、Ames試験結果そのものにあると考える。Ames試験の試験機関間の再現性は一般に80-85%と報告されている^{24,25)}。Ames試験に限らず生物学的試験にはばらつきや、生物学的不安定は常に存在し、100%結果が再現できること少ない。むしろ、80-85%の再現性は他の毒性試験よりも高いと言える。予測精度が80%であることを考えると、この数字はすでに飽和状態にあり、このまま試験データの蓄積によりQSARモデルが改良されたとしてもさらなる予測性の向上は困難である。予測率の向上にはQSARモデルの改良だけでなく、毒性試験の改良や試験結果の再評価が重要である。Ames試験の問題点はその結果が陽性、陰性の2元性であるため、弱い陽性反応が、場合によっては陰性と判定されることがある。一般にAmes試験陽性は陰性対照の2倍以上を陽性の基準

としているため(2倍法), 仮に最高誘発率が1.9倍で, 用量依存性, 再現性があってもルール上陰性と判断される²⁶⁾. また, 複数のデータが有り, 一つでも陽性を示した場合, 安全性を考慮して保守的に陽性と判断し, データベースに収載される. このような試験結果の評価は, 時として科学的妥当性を欠き, それらデータは学習データセットのノイズとなり, 正確な予測モデルの開発の障害となる. QSARであれ, AIであれ生物学的試験を学習データとする限り, そのデータの信頼性は予測率向上の最も重要な要素である. 信頼できるデータのみからなるベンチマークデータセットを学習データとし, モデルを開発することが重要である. QSARによる陽性の予測結果は, その根拠となるアラート構造とメカニズムを提示してくれる. そのメカニズムが正当である場合は実際の試験結果を疑い, 再評価する必要がある. 試験結果の判断はガイドラインに基づく定型的, 保守的な方法ではなく, メカニズムに基づく専門的な判断が重要である. 実際の試験結果とQSAR結果のクロストークが正しい評価に繋がる. そこから信頼性の高いベンチマークデータセットを構築し, そのデータセットを用いてQSARをリモデリングする. 新規の学習データセットを増やすことだけでなく, 試験結果とのクロストークとリモデリングを繰り返すことによりQSARやAIツールの予測性を限りなく向上するだけでなく, 実際の試験結果による判定を超えることができる(図3).

現実的にはインシリコでの予測が実試験を超えるにはまだ時間を要する. 現在, 筆者は実際に化学物質の変異原性を評価する場合, Ames試験とQSARの両方を

実施し, 両者の結果から最終的に結論を導くIntegrated Approachを推奨している(図3). 評価書や科学論文に報告される試験データの質は様々であり, その結果を単純に信用してはいけない. QSARを予測ツールとしてではなく, 実際の試験結果を最終的に評価するためのサポートツールとして利用するのである. これにより正しい結論を導くことができると考える. 現在, この手法を食品香料, 器具容器包装から溶出する化学物質の変異原評価取り入れ実践している. また, その最終評価の堅牢性を確認するために, 再試験を行い検証することもある.

9. おわりに

最初に現在地球上には一億五千万種類の化学物質が存在すると述べた. これら化学物質の毒性データをデータベース化すれば, まさにビッグデータベースである. それでは, 化学物質の種類は理論的にいくつ可能であろうか? これをケミカルスペースという. 分子量, 構成する元素数に依存するがその数は一般的に10の33乗から60乗と試算されている²⁷⁾. これは天文学的数字を超える. 当然, これら化学物質のすべての安全性を生物学的試験で評価することは全く不可能である. 従って, インシリコ手法が将来必ず主体となることは自明である. 10の40乗から60乗がケミカルスペースであるが, 囲碁スペースというものがある. それは, 10の360乗と計算されている. 現在, AIによるアルファ碁は人間の囲碁名人といい勝負をしている. これに比べると, 10の40乗から60乗のケミカルスペースの情報は圧倒的に少なく, その毒性の予

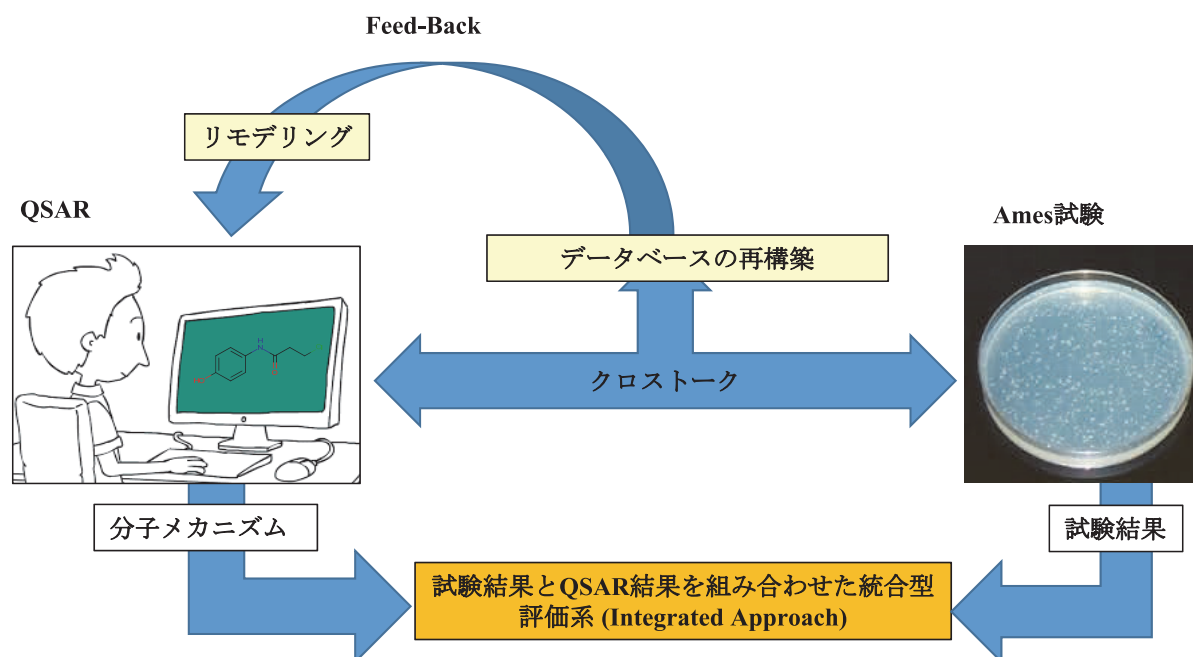


図3 Integrated Approachによる変異原性の評価

測など近い将来AIは簡単にやってくれるであろうと楽観視している。車が自動運転になる日が近づいている現在、動かない化学物質の性質をAIが予想できないわけがない。重要なことはAIに正しいドライビング技術を教えるのと同様に、正しい生物学的情報を与え、教育することである。これは容易ではないが、時間が解決してくれると思う。毒性学は21世紀中に生物学から化学へ間違いなく転換する（ことを望む）。遺伝毒性の分野はおそらくその先陣を切るであろう。

10. 謝辞

本総説はこれまで約10年間に渡るQSAR研究のうち、主にAmes試験QSARに関する研究成果をまとめたものである。研究には多くの公的資金の援助を受けた。ここに記載する、厚生労働科学補助金（H21-化学-一般-002, H24-化学-指定-010, H27-化学-指定-005, H28-食品-一般-001, H28-化学-一般-002, H30-食品-一般-003, H30-化学-指定-005）、AMED補助金（15mk0101052h0101, 18mk0101103j0401, 18am0101123j0102）、食品安全委員会補助金（平成30年度食品健康影響評価技術研究（課題番号：1803））。

Ames/QSAR国際チャレンジプロジェクトでは参加した12のQSARベンダー（表5）、とAmes試験データを提供して頂いた厚生労働省厚生労働省・労働基準局安全衛生部・化学物質対策課化学物質評価室に感謝する。また、実際にQSARを使っの多くの化学物質のAmes変異原性の計算や、Ames/QSAR国際チャレンジプロジェクト全般にわたってデータを管理してくれた派遣職員の北澤愛莉さんに感謝する。

11. 参考文献

- 1) 本間正充：食品中に含まれる化学物質のリスクアセスメントと遺伝毒性評価，FFI ジャーナル 2018;23:8-16.
- 2) Benigni R, Bossa C: Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. *Chemical reviews* 2011, 111 (4), 2507-36.
- 3) ICH-M7 (R1) (2017) ICH Harmonized Guideline. Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk. Current Step 4 version dated 31 March 2017. <https://www.ich.org/home.html>.
- 4) Miller A, Miller E C: Ultimate chemical carcinogen as reactive mutagenic electrophiles. In *Origin of Human Cancer*, Hiatt, H. H.; Watson, J. D.; Winsten, J. A., Eds. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1977; pp 605-627.
- 5) Mortelmans K, Zeiger E: The Ames Salmonella/microsome mutagenicity assay. *Mutation research* 2000, 455 (1-2), 29-60.
- 6) Ames BN, McCann J, Yamasaki E: Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test. *Mutation research* 1975, 31 (6), 347-64.
- 7) McCann J, Choi E, Yamasaki E, Ames BN: Detection of carcinogens as mutagens in the Salmonella/microsome test: assay of 300 chemicals. *Proceedings of the National Academy of Sciences of the United States of America* 1975, 72 (12), 5135-9.
- 8) Ashby J: The value and limitations of short-term genotoxicity assays and the inadequacy of current cancer bioassay chemical selection criteria. *Annals of the New York Academy of Sciences* 1988, 534, 133-8; (b) Ashby, J., Determination of the genotoxic status of a chemical. *Mutation research* 1991, 248 (2), 221-31.
- 9) Tennant RW, Ashby J: Classification according to chemical structure, mutagenicity to Salmonella and level of carcinogenicity of a further 39 chemicals tested for carcinogenicity by the U.S. National Toxicology Program. *Mutation research* 1991, 257 (3), 209-27.
- 10) Williams RV, Amberg A, Brigo A, Coquin L, Giddings A, Glowienke S, Greene N, Jolly R, Kemper R, O'Leary-Steele C, Parenty A, Spirkel HP, Stalford SA, Weiner SK, Wichard J: It's difficult, but important, to make negative predictions. *Regulatory toxicology and pharmacology* 2016, 76, 79-86.
- 11) Hanser T, Barber C, Rosser E, Vessey J, Webb SJ, Werner S: Self organising hypothesis networks: a new approach for representing and structuring SAR knowledge. *Journal of cheminformatics* 2014, 6, 21.
- 12) Klopman G, Rosenkranz HS: Structural requirements for the mutagenicity of environmental nitroarenes. *Mutation research* 1984, 126 (3), 227-38.
- 13) Klopman G, Tonucci DA, Holloway M, Rosenkranz HS: Relationship between polarographic reduction potential and mutagenicity of nitroarenes. *Mutation research* 1984, 126 (2), 139-44.
- 14) Mekenyan O, Dimitrov S, Serafimova R, Thompson E, Kotov S, Dimitrova N, Walker JD: Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 model. *Chemical*

- research in toxicology* 2004, 17 (6), 753-66.
- 15) Serafimova R, Todorov M, Pavlov T, Kotov S, Jacob E, Aptula A, Mekenyan O: 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* 2007, 20 (4), 662-76.
- 16) Benigni R, Bossa C: Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. *Mutation research* 2008, 659 (3), 248-61.
- 17) Serafimova R, Gantik M, Worth A: Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity. In *JRC Scientific and Technical Reports*, 2010; Vol. EUR 24427 EN.
- 18) Fukuchi J, Kitazawa A, Hirabayashi K, Honma M: A practice of expert review by read-across using QSAR Toolbox. *Mutagenesis* 2019, 34 (1), 49-54.
- 19) Mishima M, Hashizume T, Haranosono Y, Nagato Y, Takeshita K, Fukuchi J, Homma M: Meeting report, ICH M7 relevant workshop: use of (Q) SAR systems and expert judgment. *Genes and Environment* 2018, 40:19
- 20) Hillebrecht A, Muster W, Brigo A, Kansy M, Weiser T, Singer T: Comparative evaluation of in silico systems for ames test mutagenicity prediction: scope and limitations. *Chemical research in toxicology* 2011, 24 (6), 843-54.
- 21) Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T, Heinrich N, Müller KR: Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. *Journal of chemical information and modeling* 2009, 49 (9), 2077-2081.
- 22) Honma M, Kitazawa A, Cayley A, Williams RV, Barber C, Hanser T, Saiakhov R, Chakravarti S, Myatt GJ, Cross KP, Benfenati E, Raitano G, Mekenyan O, Petkov P, Bossa C, Benigni R, Battistelli CL, Giuliani A, Tcheremenskaia O, DeMeo C, Norinder U, Koga H, Jose C, Jeliaskova N, Kochev N, Paskaleva V, Yang C, Daga PR, Clark RD, Rathman J: Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. *Mutagenesis* 2019, 34 (1), 3-16.
- 23) 本間正充：化学物質毒性ビッグデータベースとインシリコによる毒性予測 IT・ビッグデータと薬学—創薬・医薬品適正使用への提言— 学術会議叢書 25 公益財団法人 日本学術協力財団 2019.
- 24) Kamber M, Fluckiger-Isler S, Engelhardt G, Jaech R, Zeiger E: Comparison of the Ames II and traditional Ames test responses with respect to mutagenicity, strain specificities, need for metabolism and correlation with rodent carcinogenicity. *Mutagenesis* 2009, 24 (4), 359-66.
- 25) Piegorsch WW, Zeiger E: Measuring intra-assay agreement for the Ames Salmonella assay. In *Lecture Notes in Medical Informatics, Vol. 43*, Hothorn, L., Ed. Springer: Heidelberg, Germany, 1991; pp 35-41.
- 26) Red Book II, Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food Redbook II Draft Guidance. U.S. Food & Drug. 1993.
- 27) Polishchuk PG, Madzhidov TI, Varnek A: Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of computer-aided molecular design* 2013, 27 (8), 675-9.