

厚生労働行政推進調査事業費補助金（化学物質リスク研究事業）
In silico予測手法の高度化とNew Approach Methodologyの活用に基づく化学物質の統合的
ヒト健康リスク評価系の基盤構築に関する研究

令和5年度 分担研究報告書

遺伝毒性評価に資するAmes/QSARの深化

研究分担者 古濱 彩子 国立医薬品食品衛生研究所・変異遺伝部 主任研究官
研究分担者 杉山 圭一 国立医薬品食品衛生研究所・変異遺伝部 部長
研究分担者 本間 正充 国立医薬品食品衛生研究所 所長
研究協力者 水野 忠快 東京大学大学院薬学系研究科 助教
研究協力者 菊池 陽介 東京大学大学院薬学系研究科 学部生

研究要旨

定量的構造活性相関（QSAR）を用いた Ames 変異原性予測（Ames/QSAR）の行政利用の拡大に向けて、発がん性の懸念が高く既存データにおいて陽性と陰性が混在する N-ニトロソアミン化合物の Ames 試験を継続的に実施し、既存の結果や QSAR 予測結果と比較しながら結果の整理および考察を行った。また、2020 年から実施している第 2 回 Ames/QSAR 国際チャレンジプロジェクト成果を総括し、論文化を果たした。あわせて、Transformer モデルによる Ames 変異原性予測への応用可能性を検討した。学習に用いた公開データでは既報と同程度の性能を示した一方、外部検証データでは公開データのラベルに引っ張られる傾向にあった。

A. 研究目的

Ames 変異原性を定量的構造活性相関（QSAR）によって *in silico* 予測し遺伝毒性評価を実施する流れは、ICH M7 ガイドラインに基づいた医薬品不純物の遺伝毒性評価や国内外の農薬のリスク評価等の分野において活用が進みつつある。化学物質の *in silico* による Ames 変異原性予測（Ames/QSAR）評価の活用を進めるためには高度化するなか、高品質のデータセットの使用、モデルの予測精度の更なる向上、予測結果の信頼性を評価する方法等、Ames/QSAR 深化が必要とされている。

本分担課題においては、これまで国際共

同研究（第 2 回 Ames/QSAR 国際チャレンジプロジェクト、以下「第 2 回プロジェクト」と呼ぶ）によるモデルの向上、並びに、Ames 試験データの解析を行い Ames/QSAR の高度化に資する情報を収集することとする。

更に、令和 5 年度においては、行政活用への検討が進められている統計ベースや知識ベースの Ames 変異原性予測 QSAR モデルによる検証に加え、深層学習ベースの QSAR モデルの利用可能性に関する検証も必要とされていることから、Transformer モデルの開発と応用可能性の検討も研究目的とする。

本研究を通じて、Ames/QSAR の結果を行政における判断基準として積極的に活用する統合的な遺伝毒性評価系開発の基盤を構築することを目指す。

B. 研究方法

B1. Ames 変異原性試験の実施

発がん性の懸念が高く既存データにおいて陽性と陰性が混在する N-ニトロソアミン 3 物質について令和 5 年度は Ames 試験を実施した。被験物質の名称・CAS 番号純度・用いた溶媒や構造は表 1 の通りである。Ames 試験は、OECD テストガイドライン 471 に準拠した 5 菌株 (TA98, TA100, TA1535, TA1537, WP2uvrA)、代謝活性化条件と非活性化条件による試験を実施した。最高用量を 5 mg/plate とした用量設定試験の結果をもとに本試験を実施し、2 回の試験試験実施で再現性の担保を図った。また、プレインキュベーションを 37 度 20 分実施し、代謝活性化条件では 10% Rat S9 を使用した。試験溶媒は水を第一選択肢としたが、水に不溶な場合にはジメチルスルホキシド (DMSO) を使用する方針とした。N-Nitrosodimethylamine (NDMA) については水と共に DMSO 溶媒で試験を実施・比較検討したが、DMSO 溶媒では本試験のみの実施とした。陽性の判定条件は、基本的には陰性対照から 2 倍以上のコロニー数の増大が観られた場合と定義した。

また、知識ベース QSAR モデルの DEREK Nexus (ver. 6.2.1, Nexus 2.5.2, Lhasa Limited 社 (英国))、及び統計ベース QSAR モデル CASE Ultra (GT1_BMUT モジュール, ver. 1.9.0.2, MultiCASE 社 (米国)) を用いた QSAR 評価をおこなった。更にブルガス大学 数理化学研究所 LMC (ブルガリア) の OASIS TIMES 2.31.2 における Ames mutagenicity v.18.18 モデルと Ames

Mutagenicity S9 activated kinetic v.03.03 モデルの Ames 予測結果との比較も行った。

B2. 第 2 回 Ames/QSAR 国際チャレンジプロジェクトの総括

第 2 回プロジェクトでは、最終的に表 2 で示した 11 か国の 21 チームが参加することとなった。うち 9 チームが第 1 回プロジェクトにも参加したチームである。

第 1 回プロジェクトの評価対象物質に相当する物質に対して安衛法試験結果に基づき本研究分担課題を通じて精査し、12,134 物質を第 2 回プロジェクトにおける学習データとして提供した。なお、非公開の安衛法試験結果の詳細等については、令和 3 年度および令和 4 年度の本課題の分担報告書ならびに厚生労働行政推進調査事業費 (化学物質リスク研究事業)「インシリコ予測技術の高度化・実用化に基づく化学物質のヒト健康リスクの評価ストラテジーの開発 (H30-化学-指定-005)」における平成 31 年度および令和 2 年度分担研究報告書「Ames/QSAR 予測性の向上と運用可能な Ames 変異原性予測のスキームの確立に関する研究」に記載されている。

第 2 回プロジェクトで予測対象とした 1589 物質のデータは、陽性 (A 判定 ; Ames 試験における比活性値が 1,000 rev/mg 以上であり、強い陽性と判定される物質が 80 物質、B 判定 ; Ames 試験における比活性値が 1,000 rev/mg 未満である陽性と判定される物質が 156 物質) 236 物質であり陰性 (C 判定) 1353 物質となり、第 1 回プロジェクト以降に収集した安衛法試験結果 (非公開) に相当した。

参加 21 チームは 12,134 物質の Ames 試験陰性・陽性と化学物質構造情報を学習データ情報として機密保持契約の元で提供を受け、各チーム独自のモデルで予測対象

である 1,589 物質 (Ames 試験結果がわからない物質) の化学物質構造情報から Ames 陽性陰性予測を行い、国立医薬品食品衛生研究所変異遺伝部に予測結果を送付し、同変異遺伝部で評価 (感度・特異度等の精度の検証) を行った。なお、参加チームは複数のモデルによる予測結果を提出可としていたが、同変異遺伝部からは各チーム内での Best model を選択するように依頼していた。参加チームのそれぞれの Best model と複数モデルを含む約 50 モデルのパターンで統計量の考察を行った。

B3. Transformer モデルの開発と検証

- ①文献[K. Hansen ら, J. Chem. Inf. Model., 49 (2009) 2077] から 6512 物質の Ames 変異原性情報と SMILES 構造を抽出した。本データセットを以下 Hansen データと呼称する。
- ② ①の学習データ情報ならびに大規模化合物データベース ZINC、及び PubChem を活用し、SMILES を Transformer モデルに入力(※1)し、適切な構造記述子を抽出する。構造記述子の生成に用いた入力情報、モデルの詳細、出力結果も適宜とりまとめた。
- ③得られた構造記述子から標準的な機械学習モデル XGBoost(※2)により、Ames 変異原性予測モデルを構築した。実験データと比較し、感度・特異度・陽性率などの指標を計算しとりまとめた。
- ④令和 3～5 年度の本研究分担課題で実施した 18 物質の Ames 試験結果[以下 in-house データと呼称する]で外部検証を行い、③と併せて予測結果を総括した。

※1. それぞれ以下の条件を満たす化合物を用いている: ZINC[・塩除去 (元データの時点)・H, C, N, O, F, S, P, Si, Cl, Br, I, B, Sn 以外の原子を含む化合物の除外 (元データの時点)・SMILES 表記の長さ 250 まで]、PubChem[・塩除去・水素からウランまでの化合物・SMILES

表記の長さ 250 まで]

※2. Tree 系の非線形な機械学習モデル。デフォルトパラメーターでのモデルの性能が比較的高く、様々な機械学習タスクで用いられている。

(倫理面への配慮) 本研究は動物及びヒト試料等を用いた研究を行わないため対象外である。

C. 研究結果

C1. Ames 変異原性試験の実施

表 1 に示した通り、N-Nitrosufenfluramine と N-Nitrosodiethylamine (NDEA) は代謝活性化条件下において陽性であり、N-Nitrosodimethylamine (NDMA) は水溶媒、DMSO 溶媒の試験共に陰性であった。

NDEA と NDMA の Ames 試験結果のうち、代謝活性化条件下の結果について、陰性対照からの常用対数(log)比からのコロニー数の増大の比を図 1 に示した。NDEA は TA1535、WP2uvrA で陽性であり、TA100 は陽性ではないが用量依存性があった。NDMA は陰性だが、水溶媒では TA100 で用量依存的に増加する傾向がみられた。一方 NDMA の DMSO 溶媒では増加傾向はみられなかった (図 2)。

QSAR の予測結果表 1 に示す通り CASE Ultra、Derek Nexus、TIMES いずれもどの物質においても陽性予測であった。なお、NDMA と NDEA については最近実施された OECD テストガイドライン 471 に準拠した試験結果報告 (F. Bringezu, and S. Simon, *Salmonella typhimurium* TA100 and TA1535 and *E. coli* WP2 uvrA are highly sensitive to detect the mutagenicity of short Alkyl-N-Nitrosamines in the Bacterial Reverse Mutation Test, Toxicology Reports 9 (2022), pp. 250-255) との比較も行った。我々の試験との相違点

としては、プレインキュベーション時間が 37 度 60 分と 3 倍長いこと、NDMA は水、NDEA は水と DMSO で試験がされていること、プレート法試験がなされていること等が挙げられる。Bringezu らの結果では、NDMA、NDEA とともに TA100、TA1535、WP2*uvrA* (いずれも +S9) で陽性であり、プレート法の感度が低く、NDEA での溶媒依存性は顕著ではないという結果であり、我々の Ames 試験結果との相違がみられた。

C2. 第 2 回 Ames/QSAR 国際チャレンジプロジェクトの総括

第 2 回プロジェクト参加 21 チームが提出した約 50 モデル全結果の統計値を整理した。結果は、表 3 の通りである。

第 2 回プロジェクトの結果は特異度 (陰性物質を陰性と判定できる能力) が高いが、感度 (陽性物質を陽性と判定できる能力) が低い傾向であったうえ、第 2 回プロジェクトは前回第 1 回プロジェクトよりも参加チーム間でのモデルの感度、F1 Score (適合率 [モデル予測が陽性のうち実際に陽性だった物質の割合] と感度の調和平均) の差が顕著であった。陰性と陽性の試験データの比率が 85 : 15 と偏りがあることが、感度が低い要因の一つではあるが、全モデルの平均値は 50% をしたまわった。各チームが選択した Best model については、第 1 回と第 2 回のプロジェクト両方に参加したチームで比較的良好な結果であった。これらの結果をとりまとめ、2023 年 SAR QSAR Environ Res 誌に掲載された。

C3. Transformer モデルの開発と検証

本研究で用いたモデルの概念図と計算フローを記載した概要を作成し、別紙 1、及び 2 に記載した。

最初の検討として、これまでに水野グル

ープにて構築した ZINC データ学習済みの Transformer モデル (以下 ZINC モデルと記載) を用い、Hansen データの化合物の数値化に取り組んだ結果、翻訳精度 (SMILES を符号化して記述子を得た後に元の SMILES へと復号化する際の精度、モデルによる SMILES 表記の理解度の指標, [0, 100]) は 25.5% であり、著しく低くなった。同指標は PubChem データ学習済みモデル (以下 PubChem モデルと記載) でも同様に低く、30.6% であった。一方、ZINC と PubChem では、一般にケミカルスペースが広いとされる後者の翻訳精度が高かったことから、それぞれの Transformer モデルが受容可能な SMILES 表記の違いに起因していることが示唆された。

このように翻訳精度としては低いものの、これまでの水野グループの成果より、既存の多くの化合物関連のタスクは必ずしも化合物構造の全長の情報を必要とせず、部分的な構造情報でも精度が出ることが明らかとなっている (Nemoto S, Mizuno T, Kusuhara H. Investigation of chemical structure recognition by encoder-decoder models in learning progress. *J Cheminform.* 2023 Apr 12;15(1):45.). そこで翻訳精度が低いこれらのモデルを用いて Hansen データの記述子を生成し、当該記述子を入力とした XGBoost による Ames 変異原性情報の予測を行った結果、変異原性陽性・陰性の精度 (分類精度) は ZINC モデル、及び PubChem モデル由来の記述子を使った分類モデル (以下、ZINC-XGBoost モデル、及び PubChem-XGBoost モデルと称する) で 84.5%、及び 83.9% を示し、既存の Ames 試験に関する予測モデルと遜色ないことが明らかとなった (Huang K, Fu T, Gao W, Zhao Y, Roohani Y, Leskovec J, Coley CW, Xiao C, Sun J, Zitnik M. Artificial intelligence

foundation for therapeutic science. Nat Chem Biol. 2022 Oct;18(10):1033-1036).

ZINC-XGBoost, 及び PubChem-XGBoost モデルを用い、国立医薬品食品衛生研究所が提供する 18 物質の in-house データで Ames 変異原性情報の予測の外部検証を行った結果、分類精度はそれぞれ 61.1%, 及び 55.6%であった。なおパーセンテージとしては比較的大きな差に見えるが、実際判定がモデル間で異なった化合物は1つのみである。in-house データのうち、Hansen データと重複する化合物で、かつ陽性・陰性のラベルが食い違う化合物について特に誤っていた他、比較的陽性を返しやすいモデルとなった。

D. 考察

D1. Ames 変異原性試験の実施

QSAR の活用拡大の観点から、Ames 変異原性試験結果の精緻化が必要となる物質として、陽性の懸念は高いが陰性の実試験報告がなされている物質など判定の難しい物質が挙げられる。つまり、Ames/QSAR 予測が難しいアラート構造がある物質の精査が必要である。加えて、発がん性の高い物質に対する情報の拡充も重要となる。これらの物質に対し、Ames 試験を行い、情報を更新・拡充することは Ames/QSAR の改善に直接寄与する。この典型である N-ニトロソアミン化合物に注目して、これまで令和 3, 4 年度で合計 15 試験を実施したうえで QSAR の解析結果による考察を加え、既存の情報の更新に資する提案を行うべく学術論文投稿に向けて整理を行っている。さらに令和 5 年度は典型的な N-ニトロソアミン類である NDMA、NDEA を含む 3 物質の N-ニトロソアミンの Ames 試験を実施した。NDMA や NDEA は一般に S9 代謝活性化条件で陽性であることが知られて

いるが、過去に様々な条件で試験が実施され菌株や条件によって結果が異なる報告がある。例えば、TA1535 の Rat S9 では陰性だが、Hamster S9 では陽性であったり、30% Rat または Hamster S9 では TA100 と WP2uvrA/pKM101 陽性の報告がある。しかしながら、これまでの試験結果は試験条件が一定でなく (OECD テストガイドライン準拠の 5 菌株の試験ではない、試験の最高用量が原則 5mg/plate ではなく、333 µg/plate や 10000 µg/plate 等の試験等)、結果も陽性・陰性の両方が混在していた。そこで、今回 OECD テストガイドライン 471 に準拠した試験を実施したうえで、最近の報告である Bringezu らの結果と比較検討も行った。我々は 20 分だが、Bringezu らは 60 分のプレインキュベーション時間であった。いずれの時間も OECD テストガイドライン 471 に準拠しているが、プレインキュベーション時間 20 分では活性化には短い可能性があった。更に、我々も引用文献も 10% Rat S9 で実施したが、10% Rat S9 の代謝活性化では検証は不十分な可能性があり、近年海外の規制の場で採用されつつある Enhanced Ames Test (EAT) Conditions for N-nitrosamines を適用する検証も必要な道の一つと考えられる。EAT では試験菌株は TA98, TA100, TA1535, TA1537 と *E. coli* WP2uvrA (pKM101) を含むべきで、プレインキュベーションで時間は 30 分を推奨している。なお、EAT では S9 による代謝活性化では、ラット肝 30%S9 だけではなく、ハムスター肝 30%S9 の条件での実施が記載されている。S9 の種類や濃度も結果に影響していると考えられる。

D2. 第 2 回 Ames/QSAR 国際チャレンジプロジェクトの総括

第 2 回プロジェクトの Best model で感度が

60%以上の4モデルの共通点は、第1回プロジェクトに参加したチームであった。また、第1回プロジェクトに参加したチームのほうが Best models の予測能も高まる傾向がみられた。過去のチャレンジでの経験が精度向上に資する要因のひとつだと考えられる。試験結果の陽性の比率の低さもあるが新規物質の感度の低さが課題であることや、経験値が高いモデルで精度が高くなる傾向がみられることは、Ames/QSAR を行政の場で活用する際の基本情報として役立つと見込まれる。以上、モデル開発者のボトムアップにつながるプロジェクトを完結させた。

プロジェクトと Ames 試験結果の評価により得られた成果は、Ames/QSAR の結果を行政における判断基準として積極的に活用する統合的な遺伝毒性評価系開発の基盤として期待できる。

D3. Transformer モデルの開発と検証

Transformer モデルにより生成された記述子に基づく機械学習モデルは、既存の Ames 変異原性予測モデルと同程度の精度を示すことが明らかとなった。一方、今回用いた Transformer モデルは、低い翻訳精度が示すように Hansen データ内の化合物構造の全長を認識できていないことが確認された。その原因としては、化学言語モデルである今回用いた Transformer モデルが、Hansen データに固有の SMILES 表記のスタイルを把握できていない可能性が挙げられる。実際、これまでも、様々な化合物関連のデータセットが集約されている MoleculeNet (Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. 2017 Oct 31;9(2):513-530.)を用い、様々な化合物

関連のタスクを評価した際、データセットごとに翻訳精度が異なる事例が存在した。このような SMILES 表記の方言とも呼べるデータセットごとの揺らぎを研究し解決することは、適切な化合物構造情報の抽出につながり、Ames 変異原性予測モデルの発展に資すると期待される。

In-house データを用いた汎化性能評価では、Hansen データと重複する化合物が概ね Hansen データと同じラベルを示したことから、Hansen データを過学習していると推察される。一方、in-house データと Hansen データのラベルに食い違いが存在するように、そもそも Ames 変異原性試験の結果には揺らぎが存在する。用いる S9 の由来など実験条件がラベルに揺らぎを与えている可能性がある。機械学習による予測モデルは、学習データセットに依存するため。各研究機関での統一された条件でのデータについての外部検証をすることで、予測モデルの汎化性能評価が可能になると期待される。ラベルが揺らぐ場合、どちらを真と判断するかは極めて困難ではあるが、モデルがどの構造に基づいて判断しているかを把握することで、引き続き専門家による判断に資する知見を得ることが可能になると期待される。今回用いている Transformer モデルはアテンション機構(トークンと呼ばれる単位の集合で表現可能なデータの学習に際し、どのトークンにどれだけ重みをつけるかを同時に学習するというアルゴリズム)を有しており、これを可視化する方法論を実装することで、化合物構造中のどの部位に基づいて Ames 変異原性を判断しているか把握可能になると期待される。

E. 結論

QSAR の利用拡大に資する情報の拡充を目的として、発がん性の懸念が高いが、試験条件や文献によって Ames 試験結果が異なる N-ニトロソアミン化合物の Ames 試験を継続的に実施した。得られた試験結果と QSAR 予測の結果を比較し考察を進め、今後の評価に資する知見が蓄積された。特に今年度試験を実施した NMDA と DEDA の試験結果と既存の知見等から試験条件の相違についての考察を深めた。

第 2 回 Ames/QSAR 国際チャレンジプロジェクトを総括して成果を公表し、継続したプロジェクトの実施が Ames/QSAR の社会実装に有益となる情報の公開を果たした。

Transformer モデルによる記述子は、Ames 変異原性予測に資する知見が得られた。より精度を向上する方策として、SMILES 表記の方言への対処が今後の検討課題である。また Transformer モデルに特徴的なアテンション機構を用いることで結果の解釈性が大幅に向上し、特に変異原性のラベルがデータセット間で異なる際により総合的な解釈を可能にすると期待される。

F. 研究発表

F.1. 論文発表

1. Furuhama A, Kitazawa A, Yao J, Matos dos Santos CE, Rathman J, Yang C, Ribeiro JV, Cross K, Myatt G, Raitano G, Benfenati E, Jeliaskova N, Saiakhov R, Chakravarti SK, Foster RS, Bossa C, Battistelli CL, Benigni R, Sawada T, Wasada H, Hashimoto T, Wu M, Barzilay R, Daga PR, Clark RD, Mestres J, Montero A, Gregori-Puigjané E, Petkov P, Ivanova H, Mekenyan O, Matthews S, Guan D, Spicer J, Lui R, Uesawa Y, Kurosaki K, Matsuzaka Y, Sasaki S,

Cronin M, Belfield SJ, Firman JW, Spînu N, Qiu M, Keca JM, Gini G, Li T, Tong W, Hong H, Liu Z, Igarashi Y, Yamada H, Sugiyama K, Honma M. Evaluation of QSAR models for predicting mutagenicity: outcome of the Second Ames/QSAR international challenge project. SAR QSAR Environ. Res. 2023;34:983-1001.

2. Furuhama A, Kasamatsu T, Sugiyama K, Honma. Curation of more than 10,000 Ames Test Data Used in the Ames/QSAR International Challenge Projects. In “QSAR in Safety Evaluation and Risk Assessment” “QSAR in Safety Evaluation and Risk Assessment” edited by Dr. Huixiao Hong. (Elsevier) Academic Press. 2023. p 365-372. DOI: 10.1016/B978-0-443-15339-6.00022-9

F.2 学会発表

1. Furuhama A, Maruyama T, Yamada T, Sugiyama K, Honma M: Evaluating Ames mutagenicity predictions for 1,3,5-tris(2,3-dibromopropyl)-1,3,5-triazinane-2,4,6-trione under the OECD QSAR Assessment Framework, The 20th International Workshop on (Q)SAR in Environmental and Health Sciences (June 7 2023, Copenhagen, Denmark)
2. 古濱彩子、杉山圭一、本間正充: 「ニトロソアミンに関する情報提供 -NDMA と NDEA の Ames 試験概要-」日本環境変異原ゲノム学会第 52 回大会 ICH M7/QSAR ワークショップ (2023.10.27・東京) 招待講演、(2023.11.11・福岡) ポスター

3. 古濱彩子、丸山（薦田）多恵子、山田 隆志、杉山圭一、本間正充:「:OECD QSAR 評価フレームワーク(QAF)の概要とケーススタディ」日本環境変異原ゲノム学会第 52 回大会（2023.11.12・福岡）ポスター

G. 知的財産権の出願・登録状況

（予定を含む。）

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

該当なし

表1 Ames試験被験物質と試験条件・結果並びにQSARモデルによる予測結果

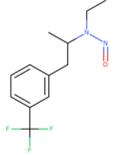
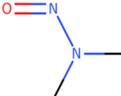
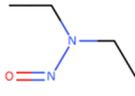
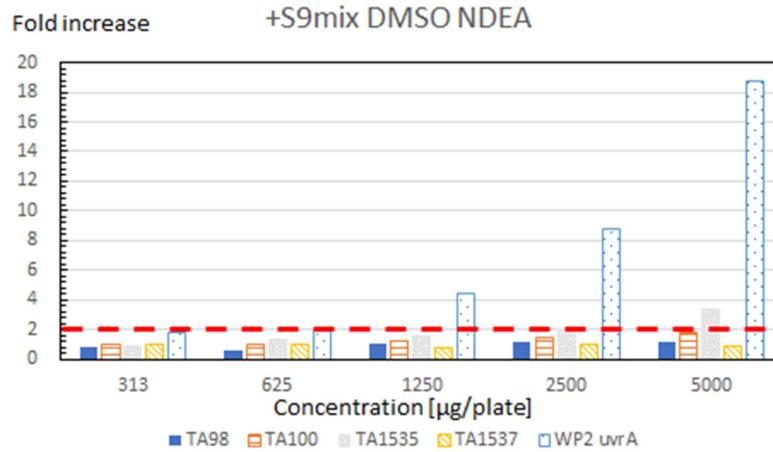
ID	Name	CAS RN	構造図	溶媒	Ames試験結果	陽性の菌株	最大比活性値	その用量	最大比活性値の菌株	CASE Ultra GT1_BMU T 1.9.0.2 (Prob%)	Derek Nexus 6.2.1 (Nexus 2.5.2)	TIMES Kinetic v03.03	TIMES non-Kinetic v18.18
1	N-Nitrosufenfluramine	19023-40-6		DMSO	陽性	WP2 uvrA + S9	256	78.1 μg/plate	WP2 uvrA + S9 (用量設定試験)	Positive (92.2)	PLAUSIBLE	in vitro Ames positive	in vitro Ames positive
2	N-Nitrosodimethylamine	62-75-9		水	陰性					Known Positive (99.2)	PROBABLE	in vitro Ames positive	in vitro Ames positive
				DMSO (本試験のみ)	陰性								
3	N-Nitrosodiethylamine	55-18-5		DMSO	陽性	WP2 uvrA TA1535 + S9	110	5000 μg/plate	WP2 uvrA + S9 (本試験)	Known Positive (99.2)	PROBABLE	in vitro Ames positive	in vitro Ames positive

表2 第2回プロジェクト参加チーム

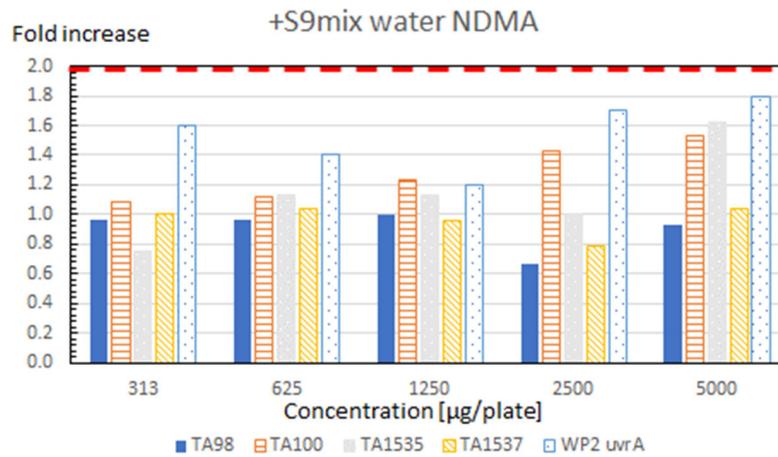
第1回プロジェクトにも参加した9チームを右側に示す

	チーム名	国	第1回プロジェクトに参加したか
1	Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences (SIOC, CAS)	中国	
2	Alttox Ltd.	ブラジル	
3	MN-AM	ドイツ/米国	Yes
4	Instem	米国	Yes
5	Istituto di Ricerche Farmacologiche Mario Negri IRCCS (IRFMN)	イタリア	Yes
6	IdeaConsult Ltd.	ブルガリア	Yes
7	MultiCASE Inc.	米国	Yes
8	Lhasa Ltd.	英国	Yes
9	Istituto Superiore di Sanità (ISS)	イタリア	Yes
10	Gifu University (岐阜大学)	日本	
11	Massachusetts Institute of Technology	米国	
12	Simulations Plus, Inc.	米国	Yes
13	Chemotargets	スペイン	
14	LMC - Bourgas University	ブルガリア	Yes
15	The University of Sydney	オーストラリア	
16	Meiji Pharmaceutical University (明治薬科大学)	日本	
17	Liverpool John Moores University	英国	
18	Evergreen AI, Inc.	カナダ	
19	Politecnico di Milano	イタリア	
20	National Center for Toxicological Research U.S. Food and Drug Administration (NCTR/FDA)	米国	
21	National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN 国立研究開発法人 医薬基盤・健康・栄養研究所)	日本	

(a)



(b)



(c)

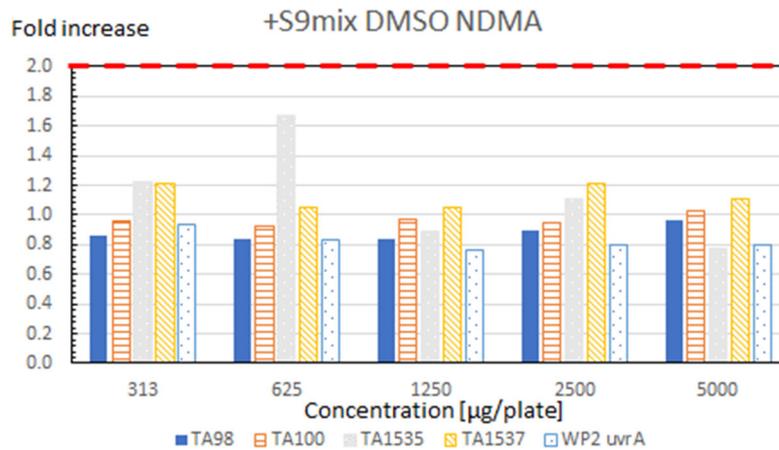
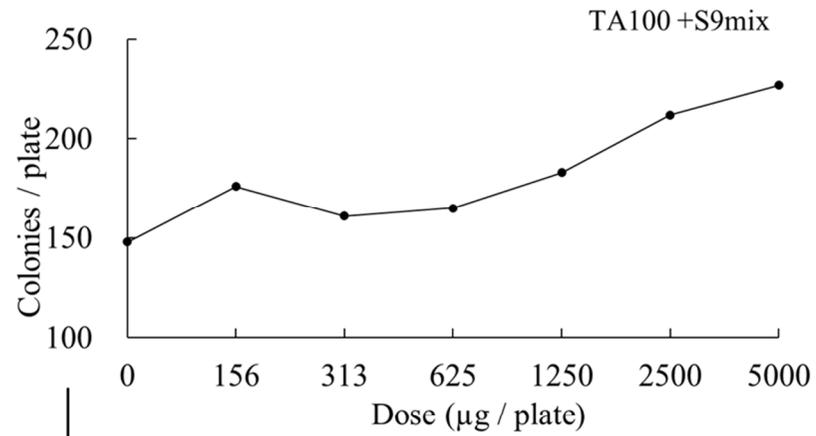
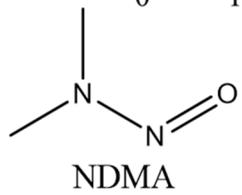
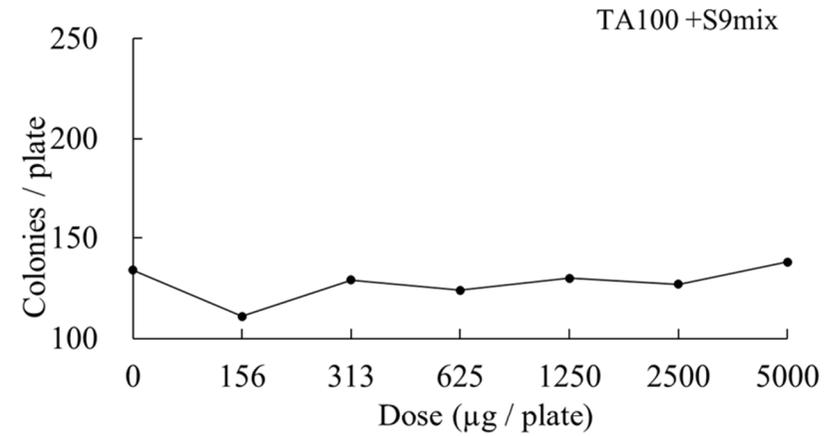


図1 代謝活性化条件下(Rat +S9)でのNDEA(a)、NDMA水溶媒 (b) ならびにNDMA DMSO溶媒(c) Ames試験 陰性対照からのコロニー数の増大の比 (本試験の結果) (2倍以上：赤点線部分以上の増大がみられるものが陽性)

溶媒：水



溶媒：DMSO



N-nitrosodimethylamine

図2 代謝活性化条件下(Rat +S9)でのNDMAの溶媒依存性

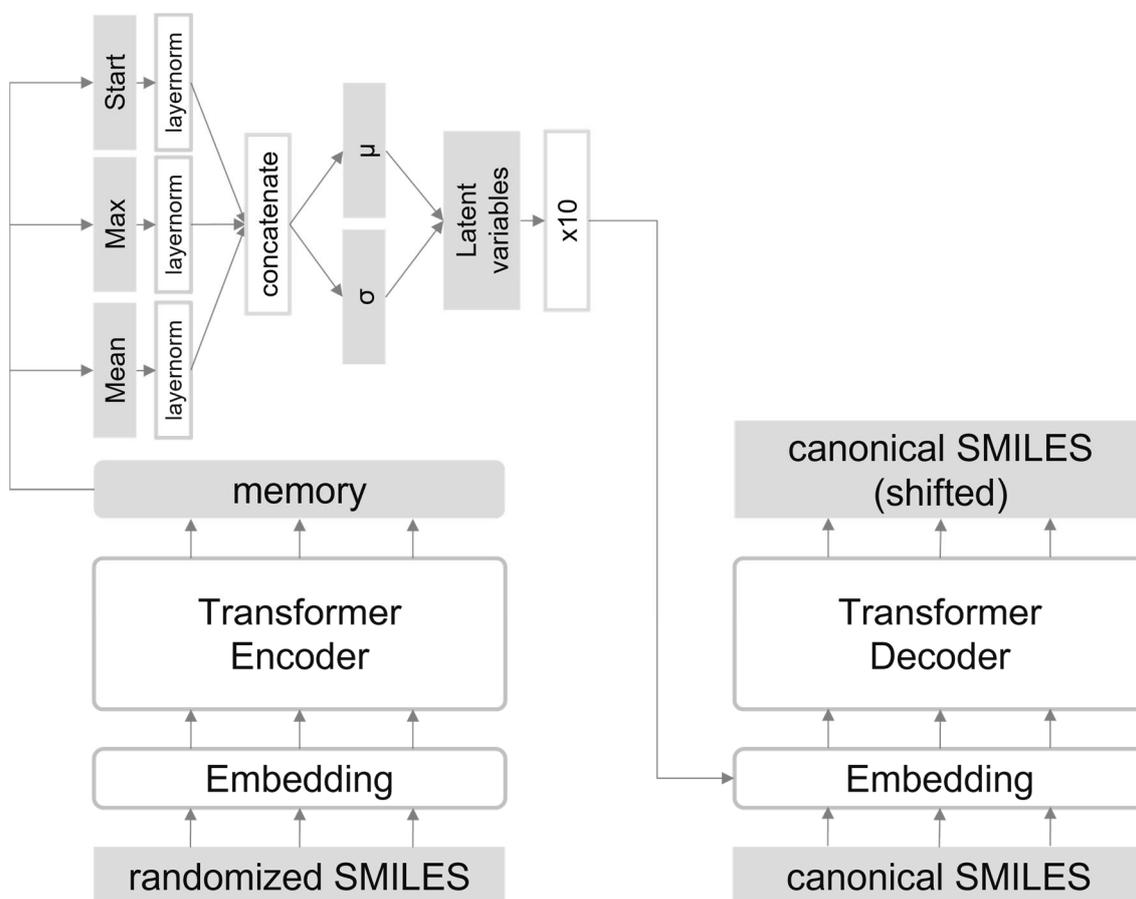
表3 第2回および第1回プロジェクトの結果概要（全モデルの平均比と値の幅）

	第2回プロジェクト			第1回プロジェクト(本間ら2019)		
	Ames試験結果を知る前に選択したBest models	Ames試験結果を知る前に選択したBest models (第1回プロジェクトに参加したチームのみ)	Ames試験結果を知った後に選択したBest models	フェーズI	フェーズII	フェーズIII
感度 (%)	46.3 (20.0-76.9)	56.2 (40.7-76.9)	52.0 (23.7-76.9)	55.8 (38.6-70.0)	56.6 (41.6-72.1)	56.6 (31.7-70.4)
特異度 (%)	87.5 (75.9-96.9)	83.9 (75.9-91.5)	85.2 (72.7-96.9)	78.7 (62.5-91.5)	85.4 (64.9-93.5)	80.1 (60.7-93.0)
精度 (%)	81.4 (70.5-86.8)	79.9 (73.6-85.9)	80.4 (71.5-86.8)	75.3 (63.6-83.9)	81.2 (65.8-87.7)	76.8 (61.1-87.3)
バランス精度 (%)	66.9 (49.6-78.5)	70.1 (62.8-78.5)	68.6 (59.2-78.5)	67.2 (62.1-72.5)	71.0 (64.0-78.9)	68.4 (62.0-74.4)
カバー率 (%)	93.0 (35.9-100)	84.2 (35.9-100)	92.3 (35.9-100)	86.7 (14.5-100)	85.5 (18.0-100)	86.0 (9.7-100)
F1スコア (%)	42.2 (16.8-53.8)	45.1 (34.8-53.8)	43.7 (30.4-53.8)	40.0 (31.8-48.9)	47.8 (36.9-57.9)	41.5 (31.7-51.5)
評価対象モデル数	21	9	22	18	21	19
予測対象物質数	1,589	1,589	1,589	3,902	3,829	4,409

詳細はFuruhama A, et al 2023. SAR QSAR Environ. Res. 34(12):983-1001参照。各チームが選択したモデルでは、結果を知る前に複数のモデルの結果のうち自分たちのモデルの中で一番よいものを選択してもらった結果では、1チームが提出せず、1チームがチーム内の2グループの結果を提示したため21モデルとなった。結果を知った後に複数のモデルの結果のうち自分たちのモデルの中で一番よいものを選択してもらった結果では、1チームがチーム内の2グループの結果を提示したため22モデルとなった。

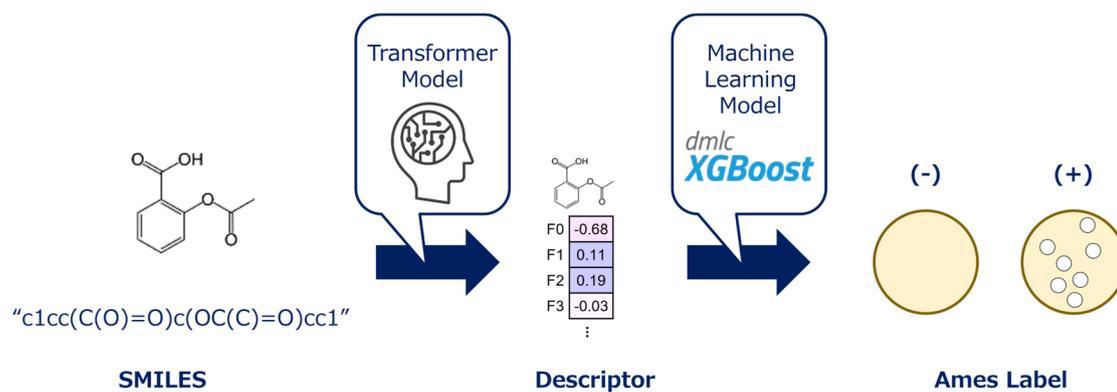
(別紙1) モデルの概念図

- Latent variables: 記述子
- μ , σ : 潜在表現の分布(正規分布)の平均, 標準偏差
- randomized SMILES: SMILES表記の開始順をランダムとしたもの (化合物構造としては対応するCanonical SMILESと同等)
- Canonical SMILES (shifted): Transformerモデルの予測対象。TransformerモデルはEncoder-Decoder構造をとっており, 入力したSMILESを潜在表現として符号化した後, 得られる固定長ベクトルから元のSMILESへと復号化する過程を学習している。復号化過程では, 一つ前までのトークン列から次にくるトークン列を予測するため, 元の入力データのトークンを一つシフトしたものがラベルとして用いられる
- Encoder-Decoder構造: データの符号化・復号化をタスクとする自己教師ありの学習構造
- トークン列: SMILES文字列情報の最小単位
- Embedding: SMILES表記を数値ベクトルへと変換する層
- LayerNorm: 学習安定化を目的にサンプル方向の正規化を行う層



(別紙2) 計算フロー

1. データセット内の化合物を表現するSMILES表記を取得する
2. 上記をTransformerモデルに入力し、記述子を得る
3. XGBoostモデルに対し、準備した記述子と対応するAmesラベル (変異原性有無)を入力し、変異原性予測を行う



(別紙3) Hansenデータを用いた性能評価

(%)	翻訳精度	分類精度	感度	特異度	陽性率
ZINC モデル	25.5	85.1	76.1	69.7	75.8
PubChem モデル	23.4	82.5	77.0	71.8	77.1

※精度： $\frac{TP+TN}{TP+FP+TN+FN}$

※感度： $\frac{TP}{TP+FN}$

※特異度： $\frac{TN}{FP+TN}$

※陽性率： $\frac{TP}{TP+FP}$

※TP:(Ames 予測の)真の陽性

※TN:真の陰性

※FP:偽陽性

※FN:偽陰性

(別紙 4) 18 物質の In-house データを用いた性能評価

(%)	翻訳精度	分類精度	感度	特異度	陽性率
ZINC モデル	88.9 (16/18)	61.1 (11/18)	76.9 (10/13)	20.0 (1/5)	71.4 (10/14)
PubChem モデル	94.4 (17/18)	55.6 (10/18)	69.2 (9/13)	20.0 (1/5)	69.2 (9/13)