

研究要旨

化学物質のAmes変異原性を予測する手法として既存のQSARに加えて化学反応に基づく透明性の高い手法の導入が有効である。本課題では、令和6年度から芳香族アミンについてニトレニウムイオンの安定化がAmes変異原性に寄与する考えから構築された量子化学計算で得られる相対エネルギー値（ここではddEと呼ぶ）を閾値とするモデルについて部分構造依存性を検討し、有効性について議論を行ってきた。令和7年度はこれまでの知見を基にモデルの精度を高める検討を行い、既存の統計ベースならびにルールベースQSARモデルとの比較・検討を含めた外部検証を実施し、モデル活用の方針を提案した。

具体的には、芳香族アミン構造SMILESとAmes試験情報をQSAR Toolboxを用いてAmes変異原性データに分子量やアミノ基の個数や複素環を含まない等の制限を付けて収集し、322物質（Ames試験陽性：214、陰性：108）を学習データセットとし、ddEの閾値のみのモデルを導入した。今回用いた学習セットには芳香環ニトロを有する物質が58含まれ、うち85%にあたる45物質が陽性であったことから、ニトロ基を含む芳香族アミンは、ddEの値に関わらず陽性とするモデルのスキームもあわせて検討した。テストセットは既存の論文（Genes Environ. **46**, 24, 2024）で、構造情報が開示されており、かつ学習セットに含まれていない305物質（Ames試験陽性：192、陰性：113）を対象とした。外部検証の結果、ddEのスキームでは、感度（陽性を陽性と予測する能力）は既存の統計ベース・ルールベースQSARモデルよりも低くなるが、特異度（陰性を陰性と予測する能力）は既存のQSARモデルよりも高い値となった。検討の結果、（芳香環ニトロの有無の寄与の考慮に依存せず）ddEのスキーム、統計ベースモデル、ルールベースモデルの多数決（2モデル以上で陽性ならば陽性、2モデル以上で陰性ならば陰性）をとる方針のコンセンサスモデルで評価した場合に、感度と特異度のバランス（高いバランス精度）が得られることを示した。

A. 研究目的

化学反応をモデル化して量子化学計算で得られる活性化エネルギーや安定化エネルギーを用いた *in silico* による Ames 変異原性予測は、定量的構造活性相関 (QSAR) を補完する演繹の手法として力を秘めている。生体分子であるたんぱく質や DNA 等を考慮に入れたシミュレーション・量子化学計算は詳細なリスク評価では参考資料として有効となりうるが、化学物質をスクリーニング評価する手法としては、時間や計算コストそして再現性の問題から現実的ではない。また、行政判断において受入れられるためには説明可能で再現性があり透明性を担保した手法であることが重要である。

これまで、アニリン (PhNH₂) を基準として芳香族アミン (ArNH₂) のニトレニウムイオン (ArNH⁺) の安定性を評価する相対エネルギー値 ddE

$$ddE = \Delta E_{ArNH^+} + \Delta E_{PhNH_2} - \Delta E_{ArNH_2} - \Delta E_{PhNH^+}$$

を量子化学計算（半経験的分子軌道法）で求め、閾値よりも安定（エネルギー値が小さいの）であれば代謝活性化条件で DNA 付加体を形成する能力が高く変異原性ありと予測し、閾値や変異原性ありの条件を検討する報告がなされてきた（Genes Environ. **44**, 10, 2022 and reference therein, Genes Environ. **46**, 24, 2024）。特に、アニリンより

もニトレニウムイオンとして閾値よりもエネルギー的に安定で諸条件を満たすならば（例えば、Genes Environ. **46**, 24, 2024 では $ddE < -5$ kcal/mol 等）Ames 変異原性陽性とするモデルのスキームや、量子化学計算に基づきニトレニウムイオンの安定性を評価するために（Genes Environ. **44**, 10, 2022）統合計算システム MOE (Molecular Operating Environment, Chemical Computing Group Inc. (CCG 社)) の Scientific Vector Language (SVL) スクリプト (mut_nitre.svl) を用いて、ルーチン的に ddE を求める提案もなされてきた。

本研究課題においては、令和6年度（昨年度）は QSAR Toolbox の METI Japan データを用いて、芳香族アミンの Ames 変異原性予測スキーム案（バイナリモデル）を提案した。芳香族アミンの相対エネルギー値 ddE と構造的な特徴（いわゆる構造活性相関）を組み合わせた直観的なスキームにより、テストセットにおける感度をあげることができた。

令和7年度（今年度）は、これまでの知見を基に学習セットやテストセットを見直し、モデルの精度を高めるための検討を行った。既存の統計ベースならびに知識ベース QSAR モデルとの比較・検討を含めた外部検証を実施し、モデル活用方針を示

した。

B. 研究方法

モデルを構築するための学習セットの芳香族アミンについては、QSAR Toolbox (<https://qsartoolbox.org/>) v4.7.1を用いて収集をおこなった。収集する構造(SMILE情報)の基準はフェニル基にひとつだけアミノ基が結合する分子量500物質とした。更に、複素環にアミノ基が結合する物質、カチオン、2つ以上のアミノ基を持つ構造も除くこととした。学習セットのAmes試験結果については、OECD TG47 1準拠の菌株での結果のみ採用することとした。

テストセットは既存の論文(Genes Environ. **46**, 24, 2024)で、今回用いた学習セットに含まれていない物質で評価した。ここで分子量500以上の物質は除いたが、フェニル基に2つ以上のアミノ基が結合している物質やカチオン、複素環も検証対象に含めることとした。

相対エネルギー値ddEはWindows環境でCCG社の統合計算化学システムMOEのmoe2024.0601において前述のスクリプトmut_nitre.svlを用いて、配座解析と量子化学計算を連続的に実施して算出した。更に、構造的特徴(芳香族ニトロ基が存在する構造: cNO_2 と記載、SMART表記 $\text{c}[\text{N+}] (=0) ([\text{O-}])$)について、MOEを用いて抽出・分離後、学習セットに対しddEを用いてMOEバイナリモデルジェネレーターを使用してモデルを構築した。

外部検証において、既存のQSARモデルと比較を行う為、テストセットを用いて統計ベースモデルと知識ベースモデルのAmes/QSAR予測を実施した。統計ベースモデルについては、MultiCASE社ソフトウェアCASE Ultra 1.9.2.1のAmes変異原性予測モデルGT1_BMUT 1.9.2.0を用いた。知識ベースモデルはLhasa社のソフトウェアNexus 2.8.0のDerek Nexus 6.4.2を用いた。GT1_BMUTではKnown Positive, Positive(陽性確率:50~100%), Inconclusive(陽性確率:40~50%)を陽性、Known Negative, Negative(陽性確率:0~40%)を陰性とし、Out of Domainは評価対象外とした。Derek Nexusでは「PROBABLE, PLAUSIBLE, EQUIVOCALを陽性とし、すべてのINACTIVEを陰性」とした場合と、「PROBABLE, PLAUSIBLE, EQUIVOCAL, INACTIVEでContains misclassified featuresまたはContains unclassified featuresを陽性とし、INACTIVEでNo misclassified or unclassified featuresを陰性」とした場合の両方で評価を行った。

考察で用いた昨年度構築したモデルとの比較・検討に必要な「芳香族炭素に SO_2 が結合した部分構造」と「芳香族アミンのアミノ基のオルト位置が両方とも水素の部分構造」についてはSMART表記 $\text{cS} (=0) (=0)$ と $[\text{ch1}]\text{c}([\text{Nh2v3}])[\text{ch1}]$ で定義し、Alvascience社のalvaDesc v3.0.12で構造の有無を抽出した。そのあと昨年度構築したスキーム案(バイナリモデル)に当てはめ、陽性・陰性の予測を行った。

(倫理面の配慮)

本研究では動物を用いた実験は実施しない。

C. 研究結果

収集した学習セット 322 物質のうち陽性は 214 物質(66%)、陰性は 108 物質(34%)であり、うち芳香族ニトロ cNO_2 を含むもの 53 物質では、陽性が 45 物質(85%)、陰性は 8 物質(15%)存在した。テストセット 305 物質のうち陽性は 192 物質(63%)、陰性は 113 物質(37%)であった。

図1に学習セットの ddE と Ames 試験陽性・陰性の関係を示した。横軸を ddE の値として、図の上側の線が Ames 試験陽性(Positive)の物質、図の下側が Ames 試験陰性(Negative)の物質を示している。全学習セットの図1aでは、ddEの値が低いほどAmes試験陽性の部分の密度が高まり、ddEの値が高いほどAmes試験陰性の密度が高まる傾向が見えるが、20~40 kcal/mol で Ames 試験陽性の数が増える傾向にある。 cNO_2 を除いた図1bでは、20~40 kcal/mol で Ames 試験陽性が消え、ddE と Ames 試験陽性・陰性の関係が明確になった。

図2にddEで陽性の閾値を $-\infty$ から $+\infty$ に変化させた真陽性率と偽陽性率のプロットであるReceiver Operating Characteristic(ROC)曲線を示した。図2aが cNO_2 を含む全322の学習セットの場合、図2bが学習セットから53物質の cNO_2 を含む物質を除いた場合である。 cNO_2 を含まない場合図2bでは、False Positive Rate (FPR:偽陽性率)が1に近い部分でTrue Positive Rate (TPR:真陽性率)が図2aの場合よりも大きくなり、赤線で示した $y=x$ の下側になる傾向が改善され、ROC曲線が描くカーブの下面積が大きくなった。

学習セットを用いてddEのみ(図3a)と cNO_2 の有無とddEを加味(図3b)して構築したモデルのスキームを真陽性(True Positive: TP)、真陰性(True Negative: TN)、偽陽性(False Positive: FP)、偽陰性(False Negative: FN)の数と共に示した。いずれのモデルも閾値はおおよそ $\text{ddE}=-6.6$ kcal/molで

あり、閾値以上を陰性・閾値以下が陽性と判断された。図3bでは図3aと比較するとTPが143から181と38増え、TNが85から77と8減る。このため、表1で示すようにcNO₂の有無を加味すると感度(陽性を陽性と予測する能力)は改善(66.8%→84.1%)され、特異度(陰性を陰性と予測する能力)は低下(78.7%→71.3%)した。他の指標では陽性適中率が86.1%から85.4%に減少する以外は、cNO₂の有無を加味することによりddEを用いたモデルのスキームでのAmes変異原性予測のパフォーマンスが向上した。

表1には学習セットで構築したモデルをテストセットで検証した結果もあわせて記載した。ddEにcNO₂の有無を加味したスキームを用いることで、ddEのみのモデルと比較すると特異度が0.9%減少した以外、他の指標ではパフォーマンスが上がる事が示された。

テストセットで検証したモデルのTP, TN, FP, FNの割合については棒グラフ(図4)で示した。なお、CASE Ultra GT1_BMUTモデルでは1物質がOut of Domainのため、評価対象外となり、Derek Nexusでは1物質が予測できない入力構造(Structure could not be normalized)であったため評価から除外した。ここで注目するのは、既存QSARモデル(GT1_BMUTとDerek Nexus)では真陽性(TP)が多く感度が高い傾向(95%と82%)であるが特異度は低めの傾向(69%と66%)であり、ddEモデルのスキーム(cNO₂の有無を加味した場合を含む)とは対照的な傾向が見られたことである。なお、Derek NexusではINACTIVEでContains misclassified featuresまたはContains unclassified featuresを陽性とした場合、偽陽性の数は増減せず、真陽性の数のみが増加し、特異度は変化せず、感度のみが82%から89%に上昇した。よって、今回のテストセットにおける評価の範囲内では、INACTIVEで懸念があるContains misclassified featuresまたはContains unclassified features(以下、「DN Contains」と記載)を陽性とみなすことが有効であり、以後の議論ではDN Contains結果を取り上げることとする。

ddEとcNO₂の有無を加味したモデルのスキーム、GT1_BMUT、DN Containsのうち、2モデル以上で陽性を予測した場合は陽性、2モデル以上で陰性を予測したものを陰性としたコンセンサスモデルを構築した。ここで、ひとつ以下の陽性・陰性の場合(CASE Ultra GT1_BMUTでOut of Domainの物質で、Derek NexusとddEではcNO₂の有無を加味したスキームで結果が分かれる場合またはDerek Nexusの予測がで

きない構造でGT1_BMUTとddEではcNO₂の有無を加味したスキームで結果が分かれる場合が該当)は評価しないこととした。テストセット得られたコンセンサスモデルのTP, TN, FP, FNの割合は図4右端に記載し、そのパフォーマンスをddEとcNO₂の有無を考慮に入れたモデルと比較した結果を(考察で議論する昨年度のコンセンサスモデルを加味した結果と合わせて)図5に示した。コンセンサスモデルでは、ddEとcNO₂の有無を加味しただけのモデルよりも真陰性が減り、偽陽性が増えることから、特異度でコンセンサスモデルは劣るが、全体的にパフォーマンスが向上した。なお、『ddEの閾値のみのモデル、GT1_BMUT、DN Containsのいずれかうち、2モデル以上で陽性を予測した場合は陽性、2モデル以上で陰性を予測したものを陰性としたコンセンサスモデル』についても同様の評価を行ったが、テストセットのTP, TN, FP, FNの割合は『ddEとcNO₂の有無を加味したスキームのコンセンサスモデル』と同一であり、パフォーマンスは同等となった。

D. 考察

昨年度報告したスキーム案は、「芳香族炭素にニトロ基が結合した部分構造;芳香族炭素にSO₂が結合した部分構造;芳香族アミンのアミノ基のオルト位置が両方とも水素の部分構造」を有する物質をそれぞれ順に「陽性;陰性;陽性」とみなし、該当する構造が存在しない物質はddE=-6.62 kcal/molを閾値として陽性・陰性を判定するバイナリモデルであった(令和6年度当該課題・分担報告書 図2, 図3参照)。今年度構築したモデルの閾値は昨年度構築したバイナリモデルの閾値とほぼ同じだった。また、過去の論文で提唱された-5 kcal/mol(Genes Environ. **46**, 24, 2024)にも近い値であった。昨年度構築したバイナリモデルで今年度対象としたテストセットによるパフォーマンス評価をすると表1で示す通り感度78.1%、特異度67.3%となり、昨年度用いたテストセットでの数値(感度71%、特異度60%)よりも高い値となった。また、今年度構築した「ddEとcNO₂の有無」モデルのスキームと比較すると、昨年度構築した「芳香族炭素にSO₂が結合した部分構造」と「芳香族アミンのアミノ基のオルト位置が両方とも水素の部分構造」を考慮に入れることで感度は増加したが、特異度は減少した結果となった。

昨年度構築したバイナリモデル、GT1_BMUT、DN Containsのうち、2モデル以上で陽性を予測した場合は陽性、2モデル以上で陰性を予測したものを陰

性としたコンセンサスモデルのパフォーマンス (感度：92% 特異度：71%, 図5) については、今年度検証したddEのみ (又はddEとcNO₂) モデルのスキームを含むコンセンサスモデル (感度：91% 特異度：75%) の方が感度と陰性的中率を除いて改善する傾向が高く、後者の方が感度と特異度のバランスが取れた (高いバランス精度の) 結果となった。偽陰性を減らすという意味では、昨年度構築したバイナリモデルを含むコンセンサスモデルの方が優れた性能を持っているが、今年度検証したモデルを含むコンセンサスモデルも感度が90%を超えており、十分なパフォーマンスだと考えられる。つまり、既存の統計ベースと知識ベースQSARモデルと組み合わせる場合には、ddEの閾値のみを考慮に入れることでバランスが取れたモデルになると期待される。これは、統計ベースと知識ベースQSARモデルで芳香族ニトロの様なアラート構造や活性を減少させる方向に働く部分構造の有無など経験的に既知の情報が考慮されているためだと説明することできる。ddEモデルでは、複数の官能基等が寄与した電子的な影響を加味してアミノ基がニトレニウムイオンを形成する際の安定性でAmes変異原性を予測する手法であり、分子量が大きかったり立体障害をもたらす構造の影響等を加味できなかったり等の制限はある (Genes Environ. **46**, 24, 2024, Genes Environ. **44**, 10, 2022) が、ルール化することが難しい複数の官能基の影響を取り込むことが可能である。既存QSARモデルのパフォーマンスも良好であるが、ddEモデルにおいてはテストセットで真陰性の割合が増えかつ既存QSARモデルと概念がことなるため、お互いに相補的な役目を担うことができるという期待がある。

最後に、知識ベースモデルに比べて統計ベースモデルのパフォーマンスが高い理由としてはテストセットのうち約7割が既知 (Known) のAmes変異原性を反映させた予測であり、陽性確率よりも既知情報を優先させて陽性・陰性を定めた予測結果を含むことも一因だと考えられる。

E. 結論

芳香族アミンのニトレニウムイオンの安定性を評価する相対エネルギーddEを指標としたAmes変異原性予測モデルのパフォーマンスを評価した。既存のQSARモデルと組み合わせることにより、テストセットで感度・特異度のバランスがとれた (高いバランス精度の) 結果が得られることを示した。ddEに

芳香族ニトロcNO₂などの部分構造の有無を含めたモデルについてはddEを単独で用いるときには重要な要素であるが、既存のQSARモデル (統計ベース・知識ベース) と組み合わせたコンセンサスモデルではddE単独のモデルで十分なパフォーマンスが得られることを示した。芳香族アミンの予測手法としては「統計ベースモデル+知識ベースモデル+ddEの指標 (部分構造の加味なし)」が有効なことを示した。

F. 健康危険情報

該当なし。

G. 研究発表

1. 論文発表

1. Koyama N, Furuham A, Kruhlak NL, Shinada NK, Izawa K. Potential for Computational Genotoxicity: A Report on Symposium 3 of the 53rd Annual Meeting of the Japanese Environmental Mutagen and Genome Society (JEMS), 2024. *Genes Environ.* 2026 Mar 12;48(1):8. doi: 10.1186/s41021-026-00358-y.
2. Sugiyama K, Furuham A, Horibata K, Tsuda M, Izawa K, Tsuji G, Demizu Y, Matsushita K, Toyoda T, Hirabayashi Y, Saito Y, Honma M. Assessment of mutagenic potential of puberulic acid contaminated in red yeast rice (beni-koji) health food supplements. *Mutagenesis.* 2026 Feb 5;41(3): 110-118. doi: 10.1093/mutage/geag005.

2. 学会発表

1. 古濱彩子: (Q) SARの基礎とQAFへのいざない. 日本環境変異原ゲノム学会・微生物 変異原性 試験研究会 (JEMS/BMS研究会) 第73回定例会 (2025. 11)
2. 古濱彩子, 亀山暁子, 三島雅之, 杉山圭一, 本間正充: 化学物質審査での活用を目指した芳香族アミンのAmes変異原性評価手法の提案 その2. 日本環境変異原ゲノム学会第54回大会 (2025. 11)

H. 知的所有権の取得状況

1. 特許取得

該当なし

2. 実用新案登録

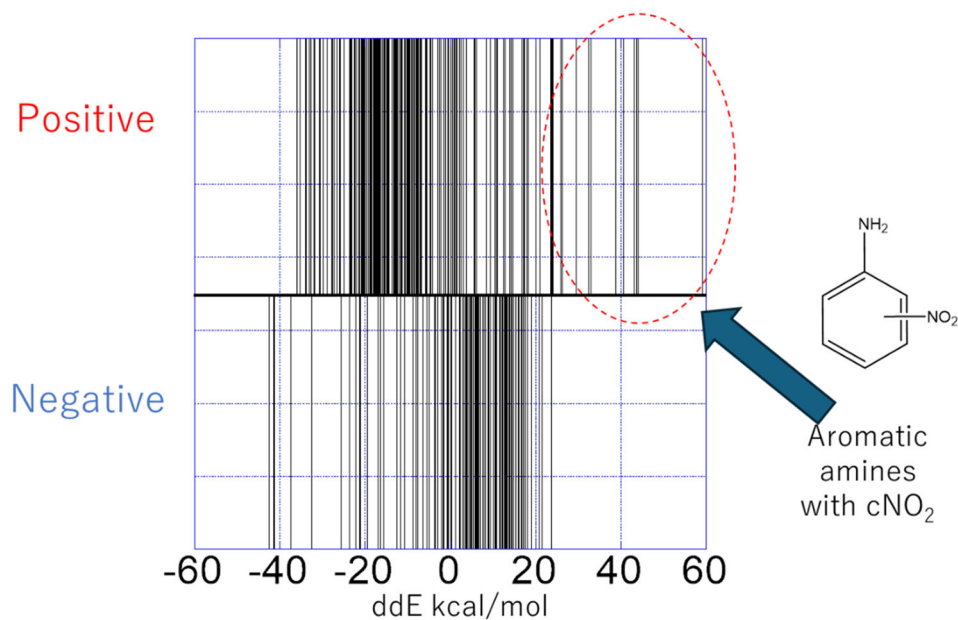
該当なし

3. その他

該当なし

(a)

学習セットデータ: 322



(b)

学習セットデータ
芳香族ニトロ(cNO₂)化合物を除く :269

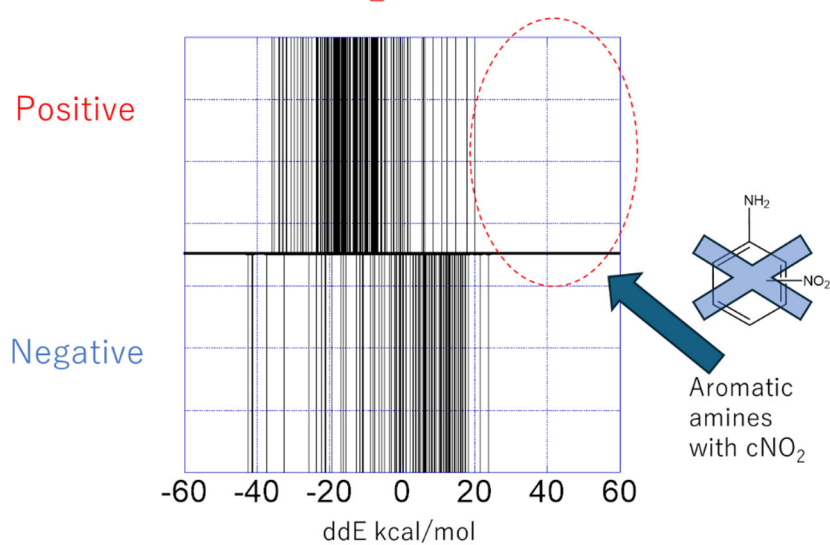
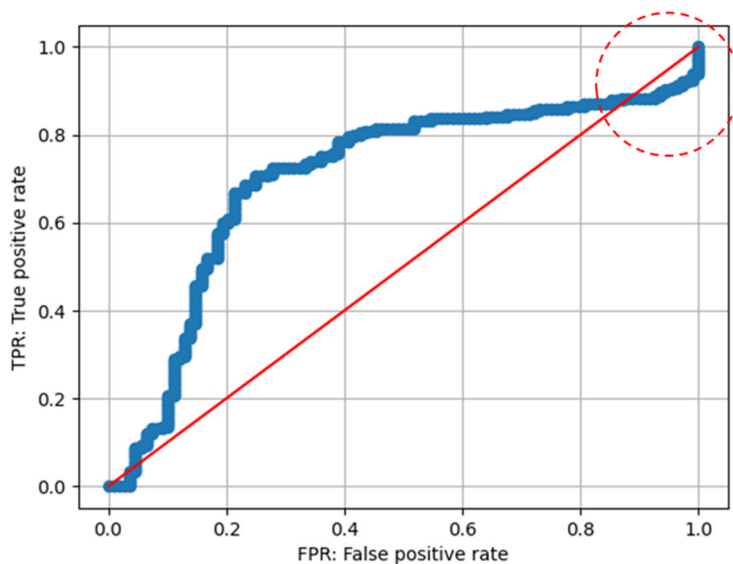


図1 学習セットの ddE(kcal/mol) と Ames 試験陽性(Positive)、陰性(Negative)の関係 (a)芳香族ニトロ(cNO₂)の有無を考慮に入れない学習セットの場合。(b)芳香族ニトロ(cNO₂)を含まない学習セットの場合

(a)

学習セットデータ: 322
ROC curve



(b)

学習セットデータ
芳香族ニトロ(cNO₂)化合物を除く :269
ROC curve

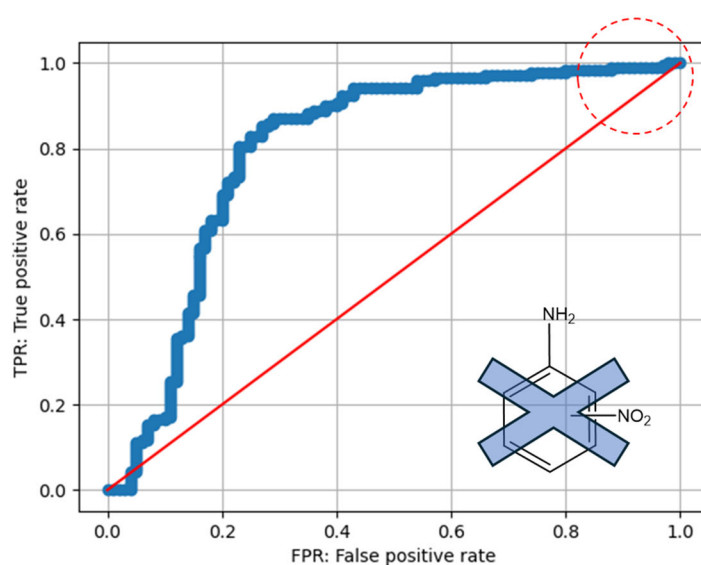
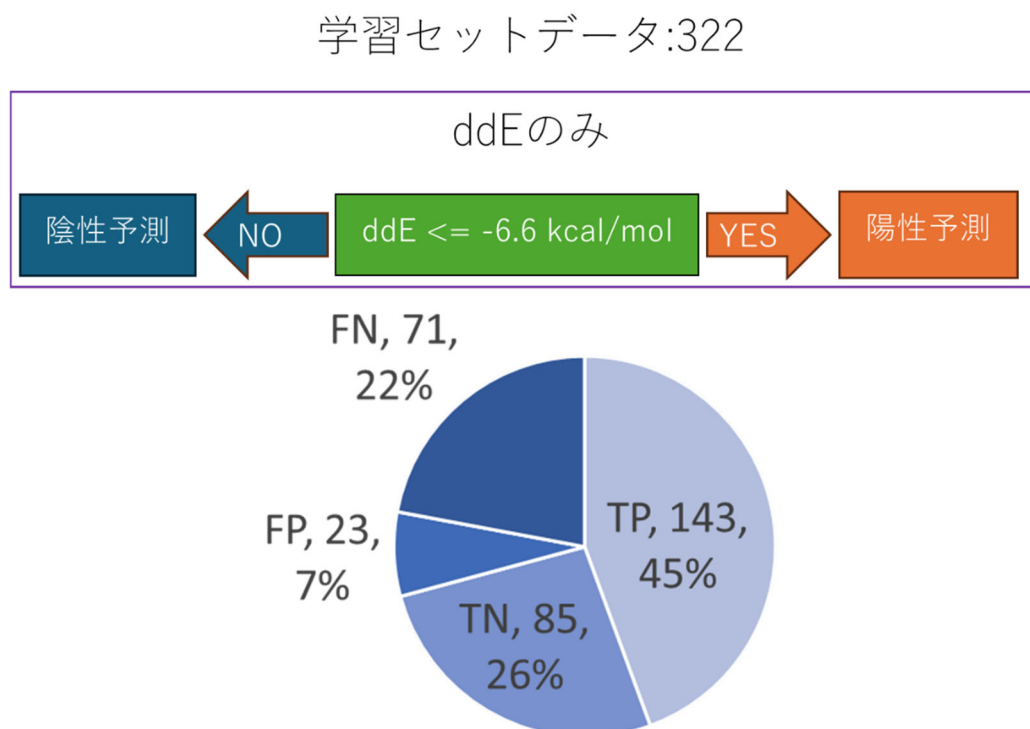


図2 Receiver Operating Characteristic(ROC)曲線 (a)芳香族ニトロ(cNO₂)の有無を考慮に入れない学習セットの場合。(b)芳香族ニトロ(cNO₂)を含まない学習セットの場合

(a)



(b)

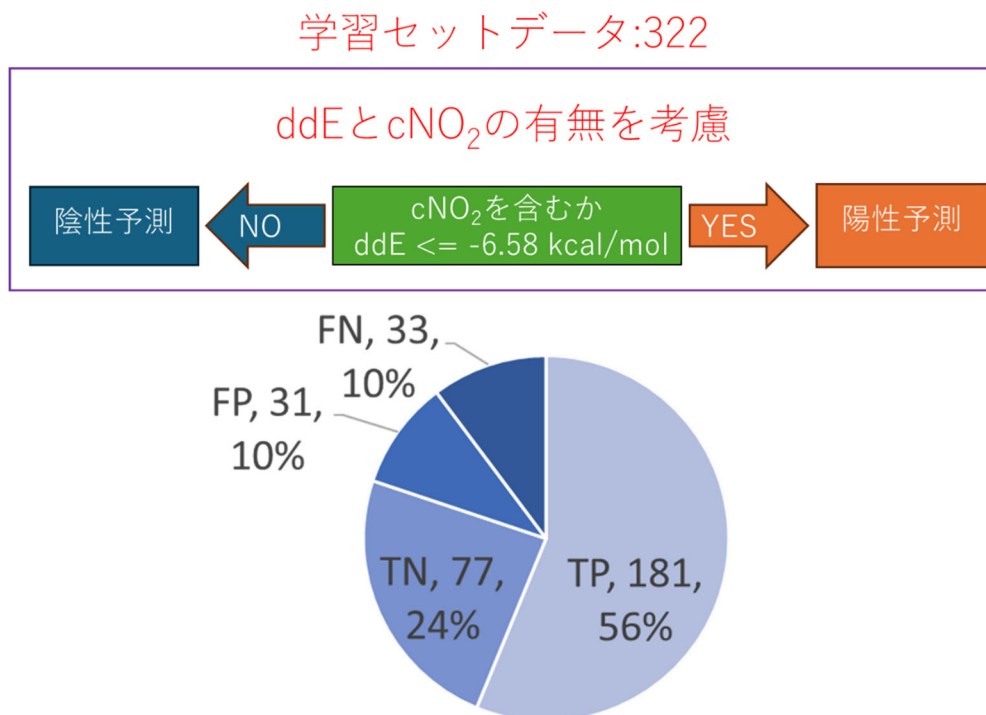


図3 学習セットを用いて構築した ddE モデルのスキームと真陽性(True Positive: TP)、真陰性(True Negative: TN)、偽陽性(False Positive: FP)、偽陰性(False Negative: FN)の数と割合(%) (a)芳香族ニトロ(cNO₂)の有無を考慮に入れない場合。(b)芳香族ニトロ(cNO₂)の有無を考慮に入れた場合

表1 ddE の閾値のみの評価、 ddE の閾値と cNO₂ の評価、昨年度構築したスキーム案（バイナリモデル）*での統計量

	ddE の閾値 -6.6 kcal/mol		ddE の閾値-6.58 kcal/molとcNO ₂		昨年度構築したスキーム案
データセット	学習	テスト	学習	テスト	テスト
物質数 (陽性:陰性)	322 (p:n=214:108)	305 (p:n=192:113)	322 (p:n=214:108)	305 (p:n=192:113)	305 (p:n=192:113)
感度(%)	66.8	68.8	84.6	71.9	78.1
特異度(%)	78.7	80.5	71.3	79.6	67.3
精度(%)	70.8	73.1	80.1	74.8	74.1
バランス精度(%)	72.8	74.6	77.9	75.8	72.7
陽性適中率(%)	86.1	85.7	85.4	85.7	80.2
陰性適中率(%)	54.5	60.3	70.0	62.5	64.4
MCC	0.43	0.48	0.56	0.50	0.45
F1スコア(%)	75.3	76.3	85.0	78.2	79.2

*昨年度構築したスキーム案（バイナリモデル）では芳香族ニトロ(cNO₂)の有無のほかに、「芳香族炭素にSO₂が結合した部分構造」と「芳香族アミンのアミノ基のオルト位置が両方とも水素の部分構造」を順に加味してddEの閾値を6.62 kcal/molとして評価した。（令和6年度当該課題・分担報告書 図2, 図3参照）

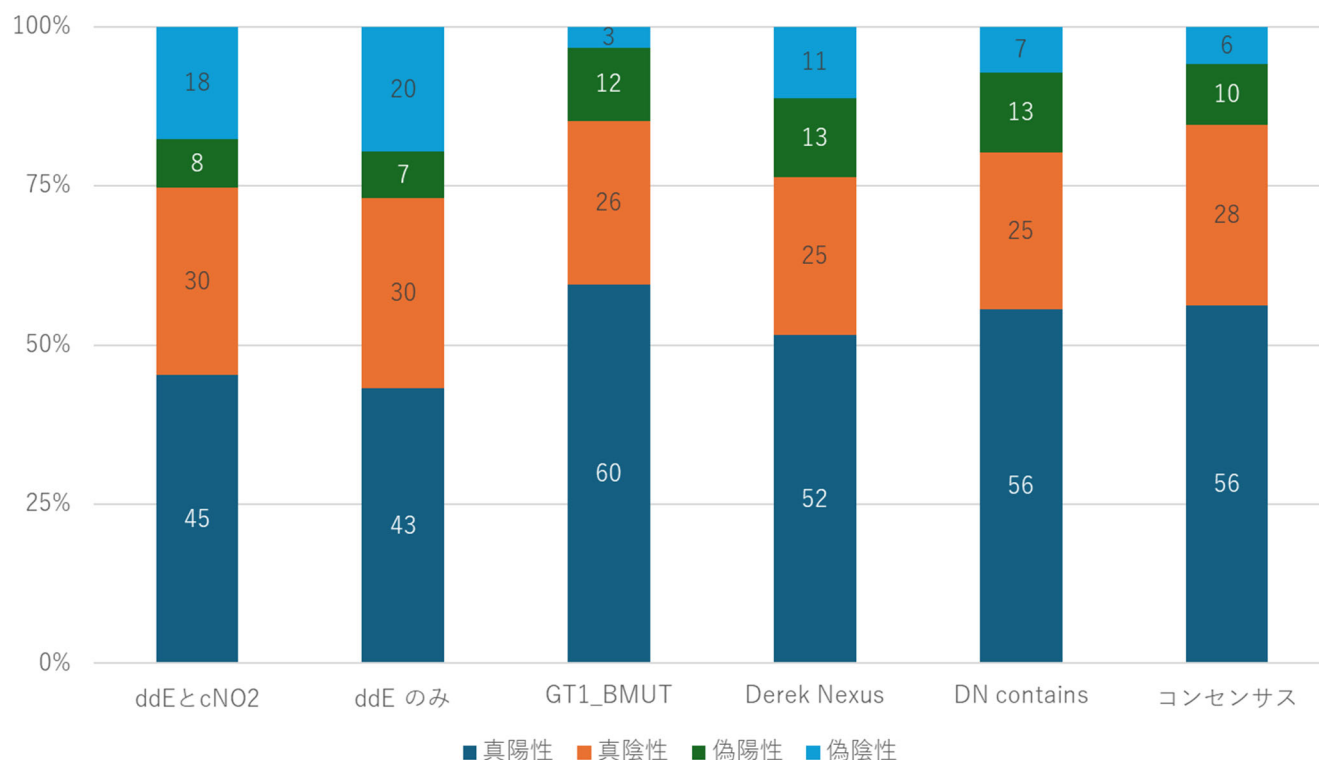


図4 テストセット(陽性:192 陰性:113)における各モデルの真陽性、真陰性、偽陽性、偽陰性の割合(%)
「ddE と cNO₂」および「ddE のみ」はそれぞれ表1の「ddE の閾値-6.58 kcal/mol と cNO₂」および「ddE の閾値 -6.6 kcal/mol」に該当。GT1_BMUTはCASE Ultraの統計ベースモデル、Derek Nexusは知識ベースモデルで、DN containsはDerek NexusでINACTIVEにおけるContains misclassified featuresとContains unclassified featuresを陽性とみなした場合。コンセンサスは『「ddE の閾値-6.58 kcal/mol と cNO₂」とGT1_BMUTとDN contains』の多数決で陽性・陰性を定義したコンセンサスモデル。コンセンサスモデルでは、「ddE の閾値-6.58 kcal/mol と cNO₂」を「ddE の閾値 -6.6 kcal/mol」モデルのスキームに置き換えてもコンセンサスモデルの真陽性、真陰性、偽陽性、偽陰性の割合(%)は変わらなかった。

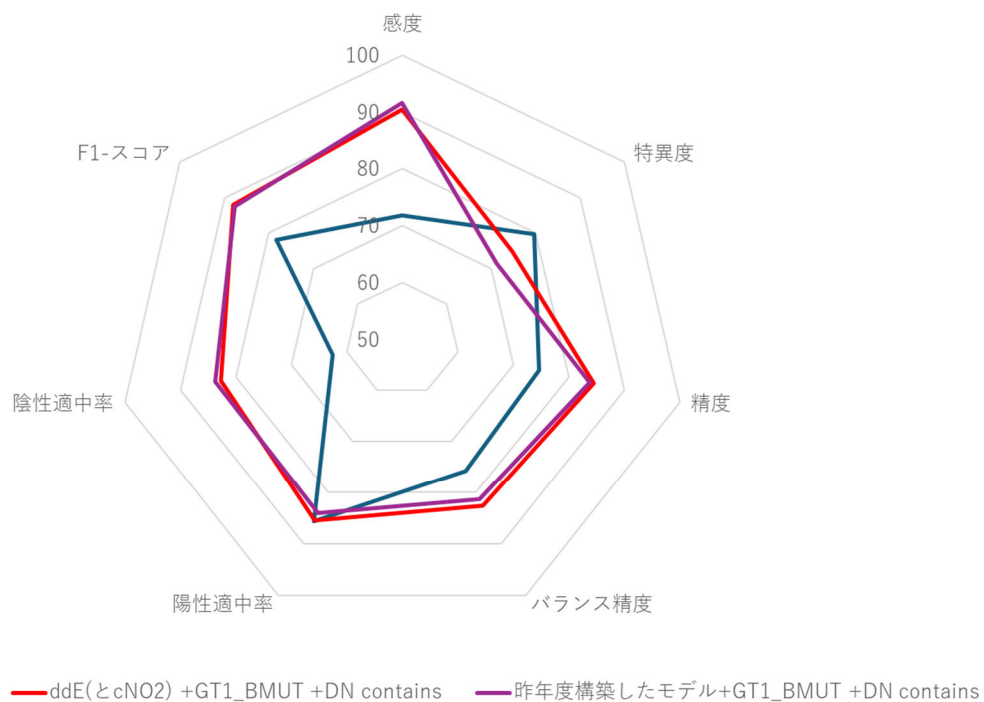


図5 表1のテストセットでのddEとcNO₂モデルのスキームのパフォーマンス(%)とコンセンサスモデルのパフォーマンス(%)の比較
 青線のddEとcNO₂は「ddEの閾値-6.58 kcal/molとcNO₂」を示し、赤線および紫線は、それぞれ今年度提案したモデルと統計ベースモデル・知識ベースモデルのコンセンサスモデルおよび昨年度構築したスキーム案(バイナリモデル)と統計ベースモデル・知識ベースモデルのコンセンサスモデルを示す。