

厚生労働科学研究費補助金 (化学物質リスク研究事業)
分担研究報告書

気道障害性を指標とする室内環境化学物質のリスク評価手法の開発に関する研究

気道障害性にかかる情報収集及び優先順位判定のための情報収集

研究分担者 小野 敦 岡山大学・医歯薬学総合研究科 教授

研究協力者 平岡結実 岡山大学・薬学部

研究要旨

本研究では生活環境を経由して暴露される可能性のある化学物質のうち室内環境汚染による健康影響が危惧される化学物質の詳細評価に向けた優先順位付けのためのハザード情報及び関連情報の網羅的スクリーニング調査を目的としている。これまでの研究により、網羅的情報収集のため JP-GHS 情報をデータベース化して気道障害性に関する情報について収集解析を行った結果、気道障害性についてはガイドライン化された試験法がないことから主に産業暴露などによるヒトにおける障害報告により区分されており、ヒトでの報告がない物質については、気道障害性に関しての区分がされていないことが確認された。JP-GHS 分類のうち関連する他の障害性との比較解析の結果、気道障害性のうち特に気道感作性については、知見の無い物質であっても、皮膚感作性に関する情報と化学構造から気道感作性評価のための優先順位付けスクリーニングの可能性が示された。本年度は、気道障害性として気道感作性よりも多くの物質について報告のある気道刺激性に着目して、GHS 分類結果をもとに関連する障害性である皮膚や眼刺激性について報告のある物質を対象として、気道刺激性との関連や、物理化学的性質や構造記述子を組合せた気道刺激性評価の優先順位付けスクリーニングのための予測モデル構築を行った。

A. 研究目的

本研究では生活環境を経由して暴露される可能性のある化学物質のうち室内環境汚染による健康影響が危惧される化学物質の詳細評価に向けた優先順位付けのためのハザード情報の網羅的な収集、及び気道障害性に関するハザード情報が得られていない物質について関連情報による補完を目的とし

た経気道曝露によるハザードのスクリーニング調査による化学物質の優先順位付けを進めている。気道障害性についてはガイドライン化された試験法がないことから主に産業暴露などによるヒトにおける障害報告により区分されており、ヒトでの報告がない物質についての気道障害性の有無は不明なままである。これまでの研究より、気道

障害性のうち特に気道感作性については、知見の無い物質であっても、関連するハザードである皮膚感作性に関する情報と化学構造から気道感作性評価のための優先順位付けスクリーニングの可能性が示された。本年度は、これまでの研究で構築した JP-GHS データベースを用いて、気道障害性のうち気道感作性よりも多くの物質について報告のある気道刺激性に着目して、GHS 分類結果をもとに関連する障害性である皮膚や眼刺激性について報告のある物質を対象として、気道刺激性との関連や、物理化学的性質や構造記述子を組合せた気道刺激性評価の優先順位付けスクリーニングのための予測モデル構築を行った。

B. 研究方法

1, 解析に用いたデータセット

これまでの研究により構築した、JP-GHS データベース分類結果から気道刺激性及び関連すると思われる障害性として眼刺激性、皮膚刺激性に関する情報を整理した。各々の刺激性の陽/陰性の判断は国連 GHS 文書に記載されている分類基準にのっとり、“区分 1～3”を陽性(+)、“区分外”を陰性(-)、“分類できない”を no data(nd)とした。気道刺激性については「特定標的臓器毒性」の区分 3(気道刺激性)または区分 1(呼吸器系)に分類されているもので、GHS 文書詳細に気道刺激と思われる記述のあるものを陽性(+)とした。データセットは Instant JChem を用いて作成した。GHS 分類結果の全 3815 物質のうち、重複しているもの、CAS 番号のないもの(901 物質)を除き、気道刺激性について報告のある物質と気道刺激性について報告がなく、眼刺激性、皮膚刺激性ともに陰

性の物質を組み合わせた化学物質リストを作成した(543 物質)。また、構造情報は QSARToolbox を使用して CAS 番号を基に検索して、作成した化学物質リストの smiles を取得し、KNIME を用いて sdf ファイルに変換して解析に用いた。本研究では皮膚刺激性、眼刺激性が陽性でかつ気道刺激性について報告のある物質を「陽性」とし、皮膚刺激性、眼刺激性が陰性でかつ気道刺激性について情報のない物質を「陰性」とした。その中で金属元素を含む物質(115 物質)、ハロゲン元素を含む物質(181 物質)を除外した合計 295 物質 (金属元素・ハロゲン元素を除外した過程については後述)のうち、KNIME で化学構造を読み込めなかった 4 物質を除外した合計 291 物質を解析に用いた。

2, 解析ソフト

本研究では、以下の解析ソフトを用いた。

・ QSAR Toolbox3.4

化学物質の安全性を評価するために構造情報や各国から提供された種々の毒性試験データが集積されているシステムである。本研究では QSAR Toolbox を使用して SMILES のデータを収集し、データセットを作成した。

・ Instant JChem(version17.1.16.0)

化学構造式データベースの構築・編集・管理を行うアプリケーションである。本研究では GHS 分類結果のエクセルファイルから作成したデータセットをインポートし、新たにクエリを作成してエクスポートした sdf ファイルを KNIME に読み込ませた。

・ KNIME Analytics Platform(version3.2.1)

ワークフロー型の機械学習および予測モデ

ル作成ツールである。オープンソースソフトウェアで誰でも無償で使うことができる。ノードが豊富で様々な形式のファイルを読み取り、データ処理することが可能である。本研究では KNIME を用いて予測モデルを作り、化合物の刺激性の有無を予測した。

・ PaDEL Descriptor

PaDEL Descriptor は 1875 個の分子記述子 (1444 個の 1D、2D 記述子および 431 個の 3D 記述子) と 12 個の Fingerprint (合計 16092 個の記述子) を計算するソフトウェアである。記述子およびフィンガープリントは、原子タイプの電気トポロジカル状態、Crippen による logP および MR、拡張トポ化学原子 (ETA)、McGowan 体積、分子線自由エネルギー関係、環の数、Laggner による化学的基礎構造の数、および Klekota および Roth によって同定された Fingerprint および化学構造の数などが計算される。本研究では Fingerprint を使用して物質の構造の情報を取得した。

・ EZR version1.36(Eazy R)

R コマンドを用いた統計解析ソフトである。生存解析、ROC 曲線、t 検定等医療統計で使われる解析を行うことができる [6]。本研究では ROC 曲線を生成し AUC からパラメーターの有用性を調べるために用いた。

3, 物理化学的性質の情報

化合物の物理化学的性質の情報は Chemid plus 及び PubChem で得られたものを使用した。Chemid plus 及び PubChem で得られた情報はモル数、Complexity、色、状態、溶解度、沸点、融点、LogP、TPSA、密

度、蒸気圧などで、この中でほとんどの化合物で情報がある 4 つの性質 (SlogP・Molweight・Complexity・TPSA) に注目した。

Chemid plus URL :
<https://chem.nlm.nih.gov/chemidplus/>
PubChem URL :
<https://pubchem.ncbi.nlm.nih.gov/>

4, KNIME を用いた予測モデルの構築

KNIME で図 1 に示すワークフローを作成して予測モデルの構築と検証を行った。本研究では“SDF Reader”に読み込ませるデータセットに以下の 3 つを用意した。

〈データセット〉

- ・ Estate FP : PaDEL Descriptor で EStateFingerprint を出力したデータセット
- ・ Substructure FP : PaDEL Descriptor で SubstructureFingerprint を出力したデータセット
- ・ Descriptor : KNIME の RDKit Descriptor Calculation で計算したデータセット

KNIME では RDKit を用いた計算を行うため、正確に計算できない金属元素を含む化合物とハロゲン元素を含む化合物をあらかじめ除外した全 295 物質のうち 60% をトレーニングセットとして用い、40% をテストセットとして用いた。トレーニングセットとテストセットは Bootstrap sampling によってランダムにサンプリングされているので、サンプリングに偏りがないように $\alpha \cdot \beta \cdot \gamma$ の 3 つを用意した。最適な記述子を決定する前に Random Forest のモデル数と深度を適切な数に設定した。

5, 予測モデルに用いるパラメーターの決

定

5.1 Fingerprint を用いる予測モデルのパラメーター

KNIME で設定したワークフローにおけるすべてのパラメーターの決定は、記述子の決定と同様に scatter plot でのグラフと Randomforest の tree での優先度、頻度から決定した。設定したワークフローに3つのデータセットをそれぞれ読み込ませ、始めにすべてのパラメーターを用いて予測モデルにかけ、RandomForestLearner における決定木から判別に関わっていないパラメーターを除外する。次に残ったパラメーターの中で判別に関わっている可能性が高いと考えられるパラメーターを scatter plot から選別し、予測率を保てる最小のパラメーター数に限定した。パラメーターの数をできる限り少なくしたのは、毒性試験の前段階でのスクリーニングを目的としていることから、より単純に、簡単に危険度を判別できるモデルを開発することが重要であると考えたためである。

5.2 RDKit DescriptorCalculation を用いる予測モデルのパラメーター

記述子に関しては、KNIME の RDKit DescriptorCalculation を使用してそれぞれの化合物について SlogP、SMR、LabuteASA、TPSA、AMW、HallKierAlpha など116個の記述子を計算し、KNIME の予測評価を繰り返して scatter plot でのグラフと Randomforest の tree での優先度、頻度から、最終的にこれらの中で気道刺激性との関連性が高いと判断した5つの記述子を使用した。

〈最終的に使用した記述子〉

SlogP : Crippen による分配係数。疎水性の指標。

TPSA(極性表面積) : 全ての極性原子 (主に酸素、窒素、それと結合する水素) の表面総和として定義される。細胞膜透過性の指標。

Num Lipinski HBD : Lipinski 則によるHBD(水素結合ドナー)数

Num Hetero Atoms : ヘテロ原子数

Num Heavy Atoms : 重原子数

5.3 モデルの精度評価

トレーニングセットと同じ物質をテストセットとしてセットして設定したモデルの精度を評価した。本研究では、以下の指標を用いてモデルおよびパラメーターの評価を行った。

True Positive(TP) : 真陽性

True Negative(TN) : 真陰性

False Positive(FP) : 偽陽性

False Negative(FN) : 偽陰性

Precision(精度・適合率) : 陽性と予測したデータのうち実際に陽性であるものの割合

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Recall(再現率) : 実際に陽性であるデータのうち陽性と予測されたものの割合

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Sensitivity 感度(真陽性率) : 実際は陽性のデータのうち、陽性と予測されるものの割合

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

Specificity 特異度(真陰性率) : 実際は陰性のデータのうち、陰性と予測されるものの割合

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

F-measure : 正確性と網羅性の総合的な評価

の尺度

$$F\text{-measure}=(2\text{Recall}*\text{Precision})/(\text{Recall}+\text{Precision})$$

Accuracy(正確度)：陽性/陰性の予測率

$$\text{Accuracy}=(\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FN}+\text{FP})$$

6. ROC 分析

ROC (Receiver Operating Characteristic) 曲線とは、いくつかのスクリーニング方法の相対的な有効性を視覚的に判定する基準の 1 つである。1 つのスクリーニング方法について陽性・陰性の基準値を最小値から最大値まで段階的に変えると、偽陽性率 (= 1-特異度) も感度も 0 から 1 まで変わるので、偽陽性率を横軸に、感度を縦軸にとって線で結ぶと、基準値の変化に対応する曲線を引くことができる。この曲線ができるだけ左上を通る方がスクリーニングとしての有効性が高く、この曲線の最も左上の点が陽性・陰性を分けるカットオフポイントとして最も有効性が高いと判断される。また、AUC (Area under curve; 曲線下面積) を計算し、AUC が大きいほど精度が高い。本研究では、EZR を使用して ROC 曲線と AUC を算出し、スクリーニングに必要なカットオフ値を決定した。

C. 研究結果と考察

1. 皮膚刺激性/眼刺激性と気道刺激性の関係

気道刺激性、皮膚刺激性、眼刺激性に関して報告のある物質について解析を行った結果、気道刺激性の報告がある物質の多くで皮膚刺激性または眼刺激性が陽性であることが分かった。気道刺激性について報告がある物質のうち、82.63%が眼刺激性陽性

であり、60.06%が皮膚刺激性陽性であった(図 1)。

- ・気道刺激性あり：691 物質
- ・眼刺激性陽性：1258 物質
- ・皮膚刺激性陽性：853 物質
- ・気道刺激性かつ眼刺激性陽性：571 物質
- ・気道刺激性かつ皮膚刺激性陽性：415 物質
- ・気道刺激性かつ皮膚刺激性かつ眼刺激性陽性：396 物質

この結果から、眼刺激性を有する化合物は気道刺激性がある可能性が高いと考えられた。また、気道刺激性の報告はないが眼刺激性陽性の物質(687 物質)や、眼・皮膚ともに刺激性陽性の物質(370 物質)についても気道刺激性がある可能性が高いと考えられた。すなわち、眼刺激性/皮膚刺激性陽性物質と気道刺激性物質との間に共通の特徴を見つけることができればそれを基にスクリーニングすることが可能になると考察された。シックハウス症候群の病態・症候の定義では、「①皮膚や眼、咽頭、気道などの皮膚・粘膜刺激症および、②全身倦怠感、めまい、頭痛・頭重などの不定愁訴」となっており、眼刺激と気道刺激の関係性は、粘膜刺激を共通の原因とすると考察された。

2. 金属元素またはハロゲン元素を含む化合物の毒性

解析の結果から、金属元素またはハロゲン元素を含む化合物の多くが刺激性について陽性であることがわかった。

金属元素

刺激性陽性：115 物質中 109 物質

ハロゲン元素

刺激性陽性：181 物質中 129 物質

よって、金属元素あるいはハロゲン元素を有する化合物は気道刺激性がある可能性が高いと考えられる。TiO₂、Al₂O₃、CeO₂等、金属酸化物ナノ粒子の気道刺激性について評価されている物質もあり、金属元素を有する化合物は注目すべきだあると考えられる。また、ハロゲン化合物は農薬や殺虫剤として使用される物質に多く生活環境を經由した暴露を考えた場合、評価の優先順位が高いと考察された。

3、KNIME を用いた評価モデル

方法 2.4 で示したワークフローに Estate FP、Substructure FP、Descriptor の 3 つのデータセットを処理した結果を表 1 に示す。全体で False Negative と予測された物質は重複しているものを除き 39 物質あった。Estate FP、Substructure FP では陽性物質は高い精度で予測されているが、陰性物質の予測精度が低く、全体的な accuracy は 75%程度である。一方、Descriptor では陽性物質の予測精度は他の 2 つに比べて劣るが、陰性物質の予測精度が優れており、全体の accuracy は約 83%となった。本研究では刺激性陽性物質をスクリーニングすることが目的であり、高い真陽性率を保つことが重要である。よって陽性物質において高い予測精度を示したこの結果は良いものと判断した。

表 2 には予測モデルで偽陰性判定された化合物を示す。EStateFP、SubstructureFP は構造的特徴について学習・予測していて、descriptor は物理化学的特徴について学習・

予測しているため、それぞれの系で予測できるものとできないものがある。例えば、構造的な観点(Estate、Substructure FP)からエポキシを有する化合物は陽性であると全て判断されたが、物理化学的な観点(Descriptor)からでは偽陰性を示すものがあった。また、複数の系で偽陰性となる物質は 3 物質のみであった。このことから、物理化学的性質と構造的性質の 2 つの観点から特徴を把握することでより高い精度でスクリーニングすることが可能になると考えられた。

4、物理化学的性質

Chemid plus および PubChem を用いて化合物の物理化学的性質の情報(色、状態(液体、固体など)、沸点、融点、溶解度、密度、疎水性など)を収集し、物理化学的性質の観点から陽性・陰性・偽陰性を示す物質の特徴を考察した。Chemid plus に情報が記載されている物性のうち、十分な情報がある 4 つの性質(SlogP・Molweight・Complexity・TPSA)について解析を行った。

図 2 には、横軸に Molweight、縦軸に TPSA/SlogP/Complexity を各々設定したグラフを示す。図 2 から陽性/陰性物質の分布は、Molweight が小さい物質に陽性が多く、大きい物質に陰性が多く、TPSA のグラフ(図 2.1)では陽性/陰性/偽陰性のどれも全体に分布していて、特に特徴がみられなかった。SlogP のグラフ(図 2.2)では、陰性物質は全体に分布しているが陽性物質は SlogP 値が比較的 low、高値(約 4.0 以上)になるとほとんど存在しないことが明らかとなった。また、Complexity のグラフ(図 2.3)より、MolWeight が約 150 以下の範囲で

Complexity が低いと陽性物質が多く、高いと陰性物質が多く、MolWeight が約 150 より大きい物質には偽陰性を示す物質が多い傾向が示された。次に、パラメーターの有用性を判断し、カットオフ値を設定するために ROC 分析により AUC を算出した。表 3 より、Molweight、Complexity、SlogP は比較的高い AUC の値が得られたため、これらのパラメーターは有用であると判断した。TPSA については AUC の値が低く、有用ではないと判断した。

5. 構造的性質

反応性の高い部分構造に着目して刺激性の有無について解析を行った結果を以下に示す。

エポキシ(刺激性 陽性：12 陰性：0)
シアン酸(刺激性 陽性：9 陰性：0)
カルボニル基(刺激性 陽性：43 陰性：28)
アミノ基(刺激性 陽性：17 陰性：8)
ヒドロキシ基(刺激性 陽性：50 陰性：15)
ピリジン(刺激性 陽性：6 陰性：2)
エーテル化合物(刺激性 陽性：13 陰性：20)
リン酸基(刺激性 陽性：6 陰性：18)

結果より、エポキシとシアン酸を含む化合物はすべて陽性であることからエポキシとシアン酸を含む化合物は陽性の可能性が高いと判断される。次に、これらの部分構造ごとに物理化学的性質 (Molweight、TPSA、SlogP、Complexity) と刺激性との関連について解析を行った結果、カルボニル基、エーテルを含む化合物で、物理化学的性質と刺激性物質の分布に関連が認められた。カル

ボニル基について SlogP が約 2 以下で陽性物質が多く、エーテルでは TPSA が約 30 以下で陽性が多い傾向が認められた (図 3.1、3.2)。また、どちらも Complexity が約 100 以下で陽性が多い傾向が認められた。しかし、それ以外の構造では特に特徴は得られなかった。一方、リン酸基を有する化合物は陰性物質が多いが、KNIME の予測結果では 6 つの陽性物質のうち 3 つが偽陰性を示しており、陽性物質が陰性物質と似た物理化学的性質を有することから注意が必要であると考察された。

E. 結論

本研究では、気道刺激性物質の詳細評価優先順位付けのための *in silico* スクリーニング目的として JP-GHS 分類をもとに気道刺激性の報告がある物質の構造的および物理化学的特徴について解析を行った。結果より、物理化学的性質と構造的性質の 2 つの観点から特徴を把握することでどちらか一方のみを用いるよりも高い精度でスクリーニングすることが可能であると結論付けられる。そして、化学物質の毒性試験を行う前段階で次の特徴を持つ物質に注意し、スクリーニングするべきであるという結論に至った。

- ・眼刺激性陽性物質
- ・金属元素またはハロゲン元素を有する物質
- ・Molweight \leq 132.21 かつ Complexity \leq 119.0 である物質
- ・Molweight $>$ 132.21 である物質

さらに特徴的な構造を含んだ物質において

は以下の点も注目すべきである。

- ・エポキシ基またはシアン酸基を有する物質
- ・カルボニル基を有し、Complexity \leq 162.000 または SlogP \leq 1.984 である物質
- ・エーテル結合を有し、TPSA \leq 29.460 または Complexity \leq 70.900 である物質

一方で、これらの特徴のいずれにも該当しない物質は刺激性を惹起する可能性が低く、リスク評価の優先度が低いと考えられる。

本研究においては、気道刺激性を有する可能性の高い物質を高精度で判別できるモデルを構築した。一方で、物理化学的性質と構造的性質など異なる性質に基づく、いずれの予測モデルにおいても偽陰性判定されたリン酸基を含む物質のように、現時点で GHS 分類された物質情報からは、陽性物質と陰性物質とで物理化学的性質および構造的な特徴に差がなく、判別が難しい物質群があることが明らかとなった。これらの物質については、陽性物質は少ないものの更なる情報の蓄積が必要である。また、本研究の解析においては、刺激性の有無のみを評価のエンドポイントとしており刺激性の強さやメカニズム等については考慮していない。今後は、個々の物質が刺激性を惹起するメカニズムに基づく評価スキームを構築することでスクリーニング評価のさらなる効率化と高精度化が期待される。

F. 研究発表

1. 論文発表

小野. 宏, 丸野内. 隼, 井口. 泰泉, 小野.

敦; ER STTA法 : hER α -HeLa-9903細胞を用いたエストロゲン受容体恒常発現系転写活性化試験法の評価報告書 *AATEX-JaCVAM* **6**, 1-27 (2017).

2. 学会発表

- A. Ono, S. Watanabe, T. Sugawara, K. Wakabayashi, Y. Tahara, N. Horie, K. Fujimoto, K. Kusakari, Y. Kurokawa, T. Sozu, T. Nakayama, T. Kusao, T. Kawakami, K. Kojima, H. Kojima, J. Richmond, N. Kleinstreuer, K. Bae-Hwa, Y. Yamamoto, M. Fujita and T. Kasahara :A multi-centre validation study of Amino acid Derivative Reactivity Assay (ADRA) - a novel in chemico alternative test method for skin sensitization. *Society of Toxicology 55th Annual meeting* (2018.March, San Antonio, USA)

小野 敦, 渡辺 真一, 菅原 経継, 若林 晃次, 田原 宥, 堀江 宣行, 藤本 恵一, 草薙 啓, 黒川 嘉彦, 寒水 孝司, 中山 拓人, 草生 武, 河上 強志, 小島 幸一, 小島 肇, J. Richmond, K. Nicole, K. Bae-Hwa, 山本 裕介, 藤田 正晴, 笠原 利彦 :新規in chemico皮膚感作性試験 ADRA法の多施設バリデーション試験 : 第2報. *日本実験動物代替法学会第30回年会* (2017.11, Tokyo)

- A. Ono, Y. Hiraoka, A. Hirose, and H. Jinno :Comparative analysis of respiratory, skin and eye irritation potential of chemicals using Japanese GHS classification.. *The 53nd Eurotox2017* (2017.9, Bratislava, Slovak)

A. Ono, S. Watanabe, T. Sugawara, K. Wakabayashi, Y. Tahara, N. Horie, K. Fujimoto, K. Kusakari, Y. Kurokawa, T. Sozu, T. Nakayama, T. Kusao, T. Kawakami, K. Kojima, H. Kojima, J. Richmond, N. Kleinstreuer, K. Bae-Hwa, Y. Yamamoto, M. Fujita and T. Kasahara :Phase-1 of the validation study of Amino acid Derivative Reactivity Assay (ADRA): a novel *in chemico* alternative test method of skin sensitization.. *Tenth World Congress Alternatives and Animal Use in the Life Sciences* (2017.8, Washington State Convention Center, Seattle, Washington)

小野 敦, 渡辺 真一, 菅原 経継, 若林 晃次, 田原 宥, 堀江 宣行, 藤本 恵一, 草苺 啓, 黒川 嘉彦, 寒水 孝司, 中山

拓人, 草生 武, 河上 強志, 小島 幸一, 小島 肇, J. Richmond, K. Nicole, K. Bae-Hwa, 山本 裕介, 藤田 正晴, 笠原 利彦 :新規*in chemico*皮膚感作性試験 ADRA法の多施設バリデーション試験 : 第1報. *第44回日本毒性学会* (2017.7, Yokohama)

G. 知的所有権の取得状況

G-1. 特許取得

特になし

G-2. 実用新案登録

特になし

G-3. その他

特になし

表 1 KNIME を用いた予測モデルでの予測結果

Estate FP

	True Positive	False Negative	F-measure	Sensitivity	Specificity	Accuracy
α	103	4	0.831	0.963	0.255	0.734
β	102	0	0.819	1	0.211	0.717
γ	112	4	0.842	0.966	0.24	0.747

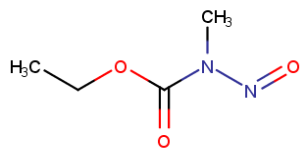
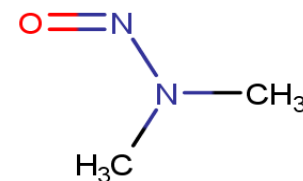
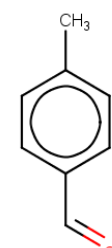
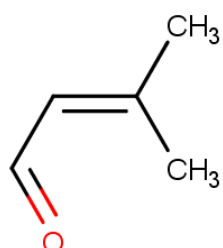
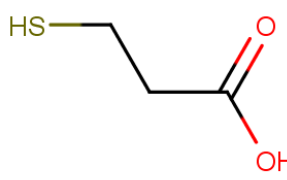
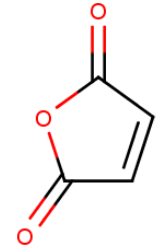
Substructure FP

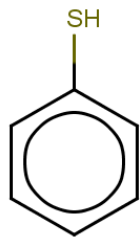
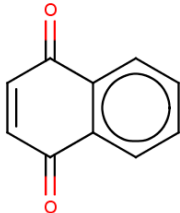
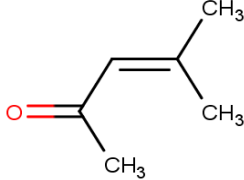
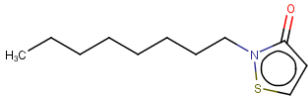
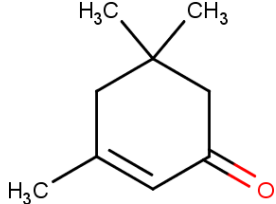
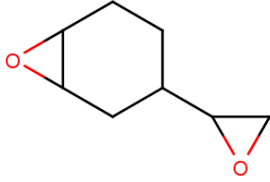
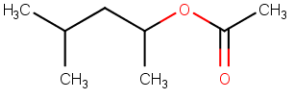
	True Positive	False Negative	F-measure	Sensitivity	Specificity	Accuracy
α	115	4	0.839	0.966	0.322	0.753
β	108	5	0.837	0.956	0.383	0.757
γ	113	10	0.825	0.919	0.345	0.735

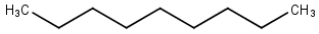
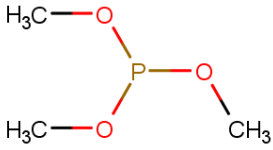
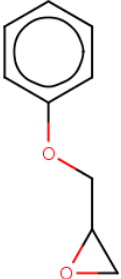
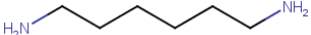

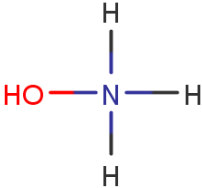
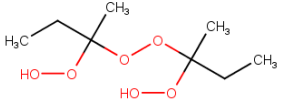
Descriptor

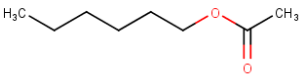
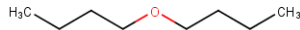
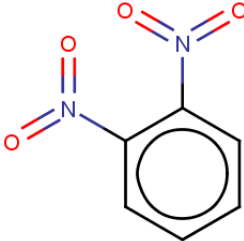
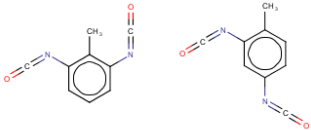
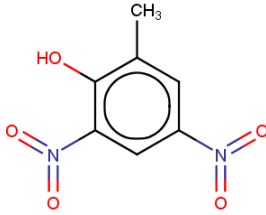
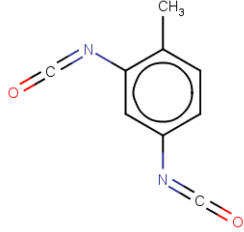
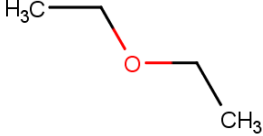
	True Positive	False Negative	F-measure	Sensitivity	Specificity	Accuracy
α	97	10	0.886	0.907	0.712	0.843
β	94	18	0.858	0.839	0.735	0.807
γ	103	9	0.88	0.92	0.604	0.825

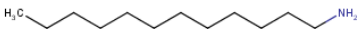
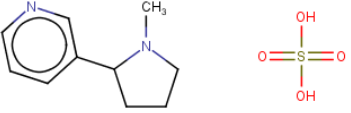
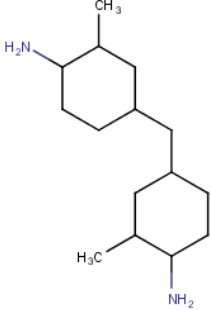
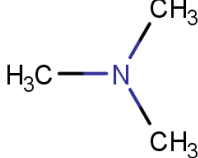
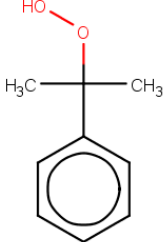
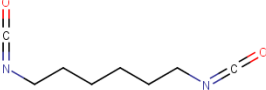
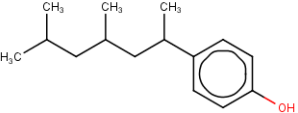
表2 3種の予測モデルで False Negative 判定された化合物

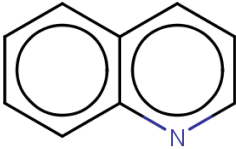
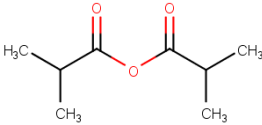
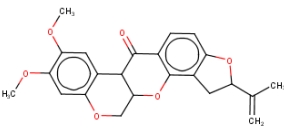
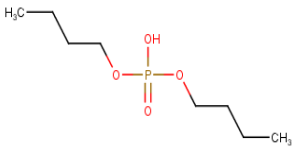
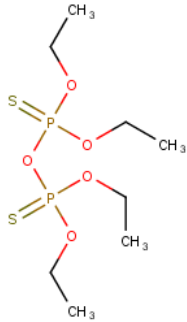
系	CAS	構造式
Estate FP	615-53-2	
	62-75-9	
Substructure FP	104-87-0	
	107-86-8	
	107-96-0	
	108-31-6	

	108-98-5	
	130-15-4	
	141-79-7	
	26530-20-1	
	78-59-1	
Descriptor	106-87-6	
	108-84-9	

111-84-2	
121-45-9	
122-60-1	
124-09-4	
1314-80-3	
1336-21-6	
1338-23-4	

142-92-7	
142-96-1	
25154-54-5	
26471-62-5	
534-52-1	
584-84-9	
60-29-7	

61788-46-3	
65-30-5	
6864-37-5	
75-50-3	
80-15-9	
822-06-0	
84852-15-3	

	91-22-5	
	97-72-3	
Substructure FP+Descriptor	83-79-4	
EstateFP+Substructure FP+Descriptor	107-66-4	
	3689-24-5	

Descriptors	AUC	最短距離の値
Mol Weight	0.799	132.210
Complexity	0.79	119.000
SlogP	0.704	1.773
TPSA	0.635	34.140

表 3 ROC 曲線の AUC と最短距離の値)

カルボニル基

パラメーター	AUC	最短距離の値	カットオフ値
Complexity	0.897	129.000	162.000(0.741, 0.902)
SlogP	0.846	1.766	1.984(0.643, 0.907)

※ (特異度, 感度)

エーテル

パラメーター	AUC	最短距離の値	カットオフ値
Complexity	0.785	70.900	70.900 (0.737, 0.923)
TPSA	0.829	29.460	29.460 (0.700, 0.923)

※ (特異度, 感度)

表 4 ROC 曲線の AUC と最短距離の値およびカットオフ値)

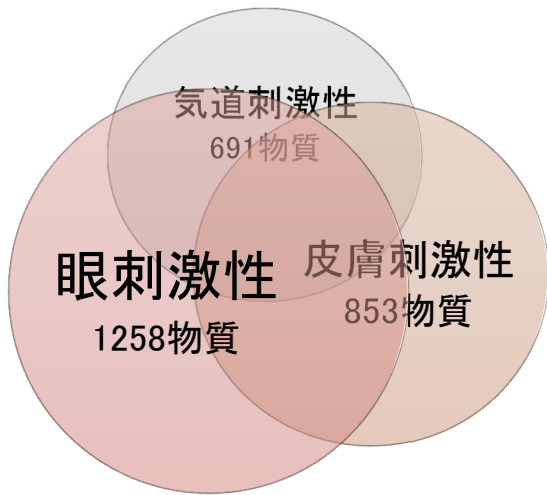


図1 気道、皮膚、眼刺激性の重なり

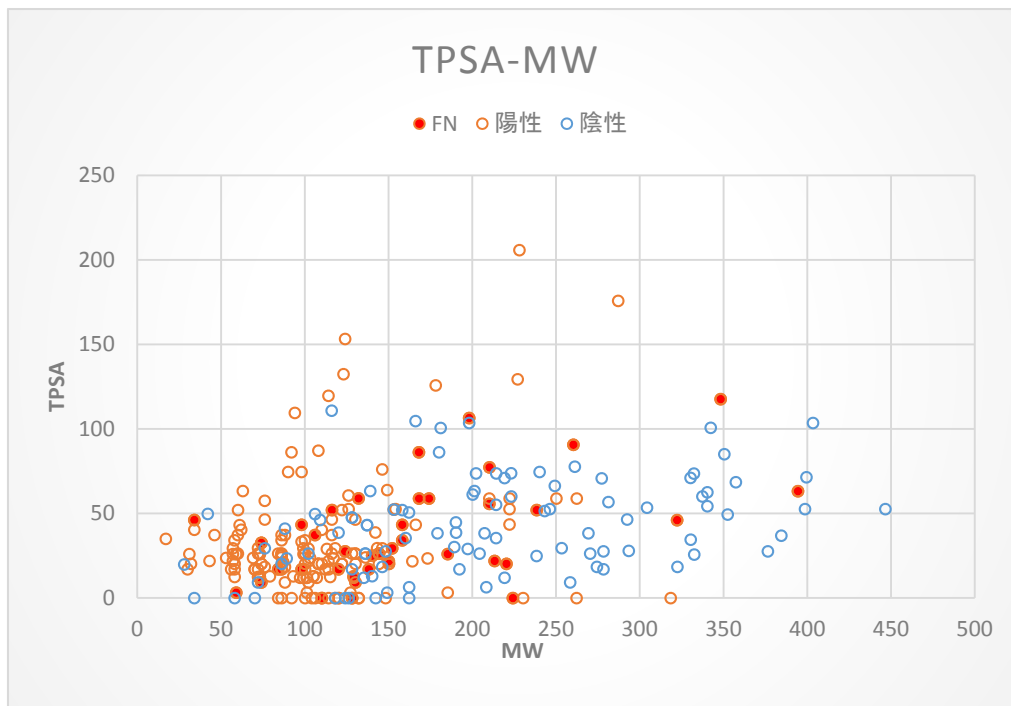


図 2.1 TPSA

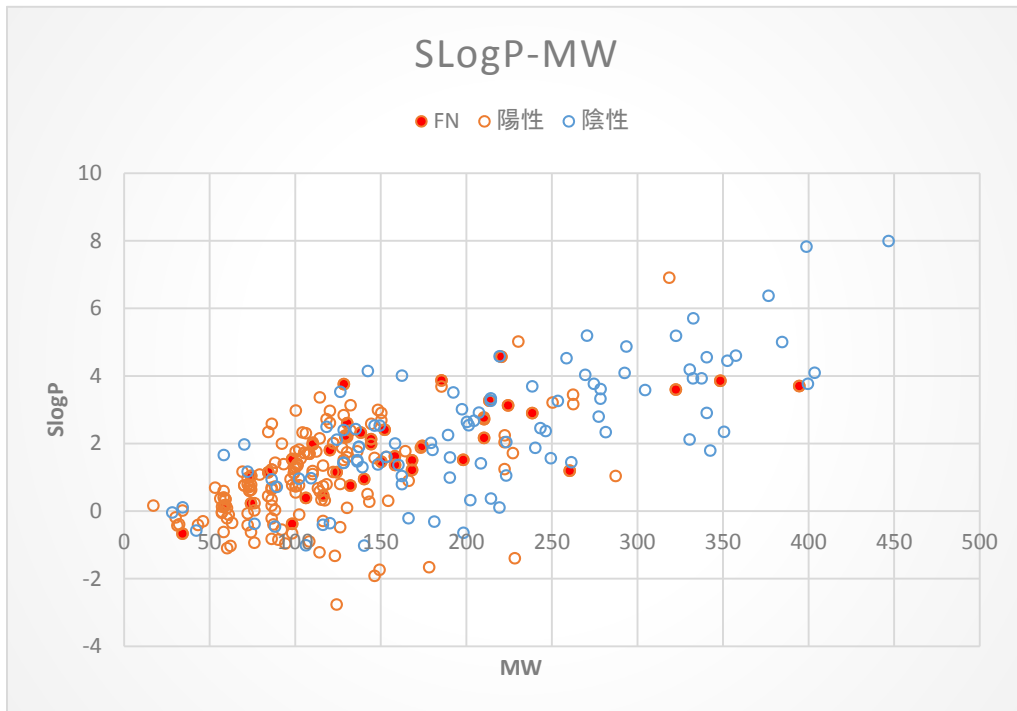


图 2.2 SlogP

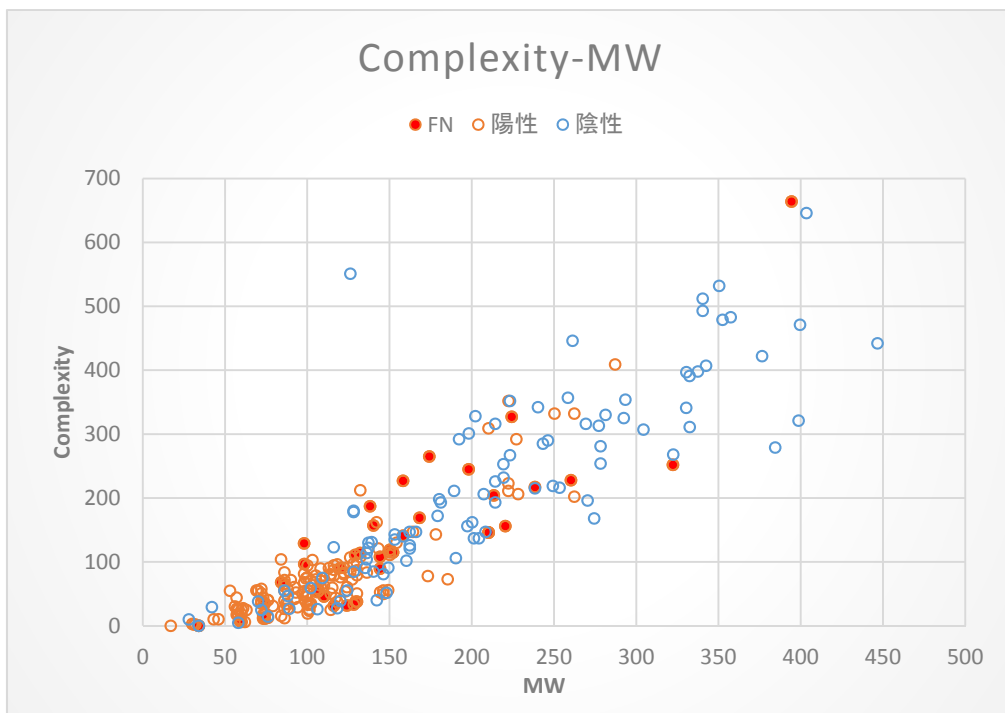


图 2.3 Complexity

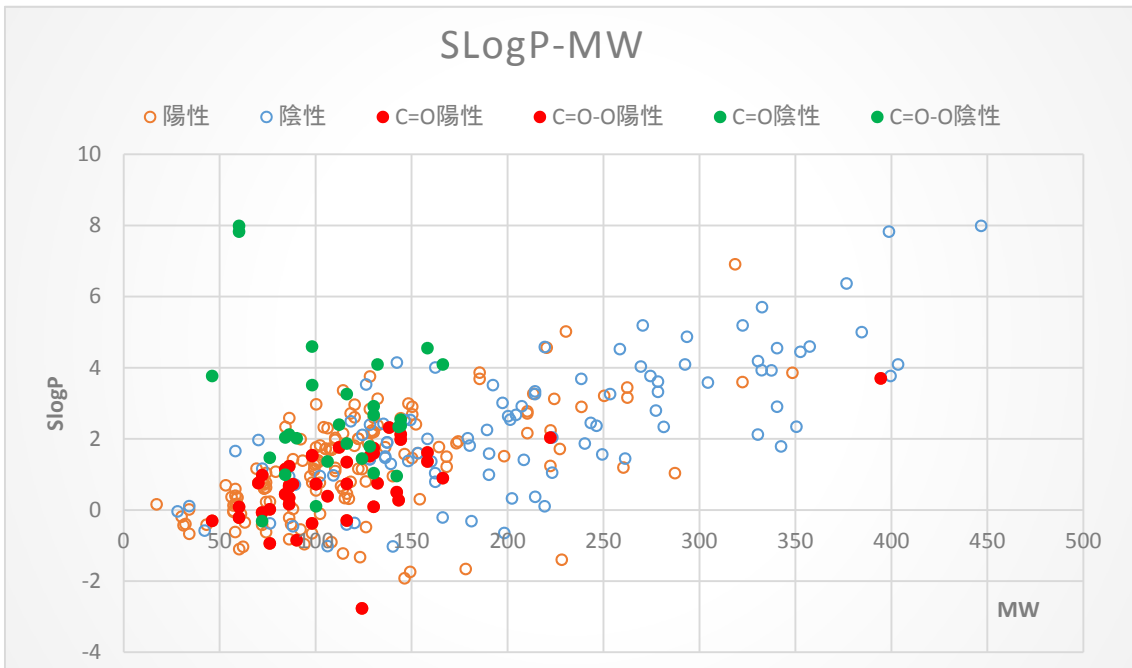


図 3.1 カルボニル基を含む物質の SlogP と刺激性陽性物質の分布

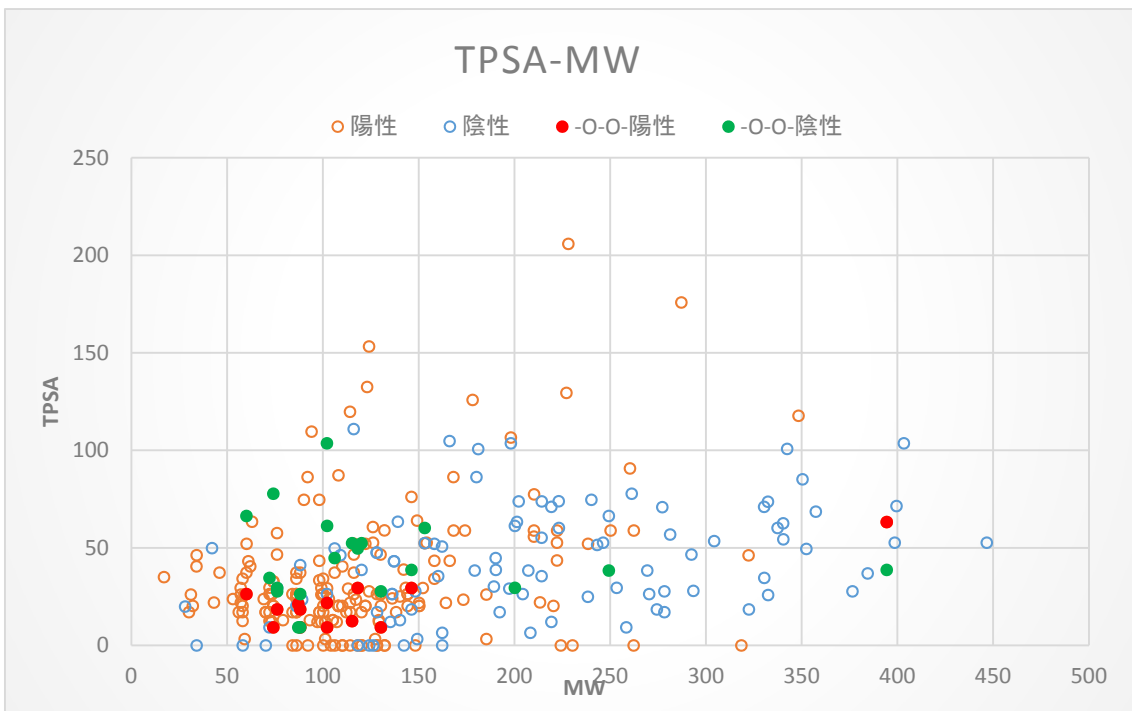


図 3.2 エーテル化合物の TPSA と刺激性陽性物質の分布