

平成 29 年度厚生労働科学研究費補助金（化学物質リスク研究事業）  
分担報告書

研究課題名：化学物質のヒト健康リスク評価における（定量的）構造活性相関および、  
カテゴリーアプローチの実用化に関する研究

**分担研究課題名：エームス変異原性の予測精度の向上に関する研究**

研究分担者 本間 正充 国立医薬品食品衛生研究所変異遺伝部 部長  
研究協力者 北澤 愛莉 国立医薬品食品衛生研究所変異遺伝部

**研究要旨**

エームス試験予測QSARツールの向上を目指した国際共同研究を2014年から実施中である。日本で実施された安衛法に基づくエームス試験データから、QSARツールの開発に有用と考えられる12,140化学物質をデータベース化した。平成29年度はこのうち4,409化合物を用いて、エームス試験結果の予測精度を競うPhase IIIトライアルを行った。世界6カ国から、11のQSARベンダーがAmes/QSAR国際競争のPhaseIIIトライアルに参画し、19のQSARツールについて予測精度の評価を行った。多くのQSARツールの予測精度は比較的高く、実際のエームス試験との一致率は平均76%であった。しかしながら、Phase Iでの一致率は平均75%、Phase IIでの一致率は平均80%であり、PhaseIIからの予測率は低下した。その理由として特異性の低下が考えられた。PhaseIIIの化合物の5%分子量800以上のエームス陰性化合物で有り、その95%がエームス陰性であるが、これらを正しく陰性と予測できなかったことが原因と考えられる。事実、感度に関してはPhaseIIよりわずかな改善が認められた。3回のトライアルにより、本共同研究は成功裡に終了した。さらに、様々な構造を持つ化学物質での試験データを収集し、トレーニングデータとしてQSARツールに組み込み、ケミカルスペースを拡大させることが、QSARの予測精度の向上に繋がるものと考えられる。また、トレーニングデータのいくつかには間違っていて、陰性、もしくは陽性とラベルされているものが確認された。これらデータはQSARモデルの開発に大きな障害となる。正確な試験結果データのみからなるベンチマークデータセットを整備し、トレーニングデータとすることが精度の高いQSARモデルの開発に重要である。

**キーワード**：エームス試験、安衛法、化審法、QSAR、国際共同研究

**A . 研究目的**

現在、人によって新たに作り出された化学物質は登録されているものだけでも1億4千万種類を超え、これは1日約4000種類のペー ス で 増 え 続 け て い る

( <http://www.cas-japan.jp/> )。これまでの調査では新規化学物質の約5%は強い変異原性を有し、人間に対して発がん性を示すと推察されている。このような変異原性化学物質を効率的、且つ正確に検出し、規制の

対象とすることは公衆衛生上重要な取り組みである。近年、IT（情報技術）の進歩により、化学物質の変異原性をインシリコで評価することが可能となっている。EU 各国および米国EPA においては、既存化学物質のリスク管理の目的で、安全性評価未実施の物質を対象にカテゴリーアプローチおよび、定量的構造活性相関（QSAR）の利用が検討されている。特に、医薬品における変異原性不純物の評価にQSAR の利用を明文化したICH-M7ガイドラインが2015年1月から適用されたことから、一般化学物質のヒト健康リスク評価・管理への適用にも拍車がかかると思われる。

国立医薬品食品衛生研究所・変異遺伝部ではこれまで我が国で行われた エームス試験データを収集し、大規模データベースを再構築してきた。本エームス試験データベースは2万化学物質以上からなり、現在では世界最大規模のデータベースである。また、個々の試験はOECDガイドラインに従い、GLP管理下で行われた試験であるため、その信頼性も高い。この最大且つ、信頼性の高いエームス試験データはエームス試験ベンチマークデータセットとして内部利用している。このデータベースの中で、未公開データを、外部評価データとしてQSARベンダーに提供し、世界中で汎用されているQSARツールの予測性の検証と、改良目的として国際共同研究を実施した。

未公開のエームス試験データは主に安衛法に従って、化学・製薬メーカーが要求され、自社、もしくは外部の試験受託機関で、GLPで試験されたデータである。安衛法では、労働者が曝露する可能性がある年間100kg以上の中間体を含む化学物質について試験

をすることを求めており、その化学物質の種類は多岐にわたるため、極めて有用なデータベースである。安衛法のエームス試験データは2万化合物を超えるが、このうち、高分子ポリマー、混合物、金属化合物を除き、QSARモデルの開発に有用と考えられる12,140の低分子化学物質のデータベースを構築した。

国際共同研究では、この試験データを3回に分けて、エームス試験結果を伏せたままでQSARベンダーに提供する。各QSARベンダーは計算した予測結果を、本研究班に伝え、本研究班でその予測率を評価する。評価後、各QSARベンダーに実際のエームス試験結果を開示する。各QSAR開発者はその結果を基に、QSARモデルの改良や、トレーニングセットのアップデートを行い、次のトライアルに参加する。このようなトライアルを繰り返すことにより、全てのQSARツールの予測性の向上が期待できる。

H27年度は、3,902化合物について第1回目(Phase I)のトライアルを行った。H28年度はこのPhase Iトライアルの結果を各QSARモデルに組み込み、新たに3,829化合物についてチャレンジを行った(Phase IIトライアル)。Phase IIでの予測率は、Phase Iよりも向上していた。

今回、最後のPhaseIIIトライアルを実施した。4,409化合物リストを世界6カ国、11のQSARベンダーに提供し、19のQSARモデルが最後のチャレンジを行った。

## B. 研究方法

### B-1. エームス試験データ

労働安全衛生法に基づき実施された変異原性試験の結果評価書から入手した、Ames

試験判定結果約20,000のうち、12,140を電子データ化し、QSARで計算できるようにSMILES形式の構造式情報を追加したEXCELファイルを作成し、さらにQSARソフトが直接読み込めるSDファイル形式に変換した物を作成した。

今回のPhase IIIトライアルでは、上述の12,140物質のうち、Phase I、IIで利用した使用した7,731物質を除く、4,409物質のリストをQSARベンダー提供し、にエームス試験結果予測のコンペティションを実施した。Phase の4,409物質を含む、全12,140物質の内訳を表1に示す。これら物質のエームス試験結果は以下の3つに分類される。

- Class A: エームス試験における比活性値が1,000 rev/mg以上の強い陽性と判定される物質 (236物質)
- Class B: エームス試験における比活性値が1,000 rev/mg未満の陽性判定物質 (393物質)
- Class C: エームス試験における陰性判定物質 (3,780物質)

## B-2 共同研究参加機関

Phase III 国際共同研究は世界6カ国、11のQASRベンダーが共同研究に参画した。PhaseIIまで参加したProus社 (スペイン) は都合により参加を辞退した。Phase IIIでは合計19のQSARツールが4,409物質のエームス試験結果を予測にチャレンジした (表2)。

## B-3. 評価結果の判定基準

各QSARモデルの陽性、陰性の判定予測表記方法は統一されていない。判定予測表記の他に、In Domain、Out of Domain、

判定不能、予測不能についても、各QSARツールで表記方法が異なり、判定予測の条件付けも異なる。ブルガス大のTMESとアルタミラのChemTunesは他のQSARツールの様に判定予測にIn Domain、Out of Domainが含まれた表記ではなく、判定予測とDomain情報が別々である。特にChemTunesには判定予測の結果があってもOut of Domain = no ならば予測不能扱いの物質も存在する。上記の分類は最終的にはすべての判定結果を”陽性 (positive)”、”陰性 (negative)”、不適用 (unapplicable)”に分類した。不適用化合物は、予測対象化合物からは外し、陽性もしくは陰性と判定されたもので“適用率 (applicability)”を計算した。

また、基本的な能力評価として、感度 (Sensivity)、特異度 (Specificity)、全体の一致率 (Concordance) を用いた。感度はエームス陽性を陽性と判定できる能力であり、さらに変異原性の強い物質を検出できる感度 (A-Sensivity) を別に計算した。

## **C. 結果**

### C-1. Phase III 4,409化合物のQSARによるエームス試験結果の予測

表3に11のQSARベンダーが提供するQSARツールによるエームス試験予測結果を示す。ブルガス大学のTIMESは代謝物をメカニズムベースで予測するツールを含んでおり、情報となるドメイン構造が少ないため、ドメイン内とドメイン外で別々に計算した。ラーサ社、マルチケース社、リードスコープ社は知識ベース、統計ベースの2つのQSARツールを提供した。また、ラー

サ社の統計ベースのSarahモデルはモジュールの違う3つのツールで計算した。アルタミラのChemoTunesも2つのモジュールで計算した。イタリアのマリオネグリ研究所は、3つの統計ベースのQSARツールで計算した。

17のモデルの内、Phase IIまでの結果をトレーニングセットしてモデルに組み込んだことを確認できたモデルは8モデルであった。特にDEREKやToxtreeのような知識ベースQSARモデルはルールの開発に時間がかかるためモデルの改善は困難である。一方、統計ベースモデルは簡単に組み入れることができる。特に高いA-sensitivityを示したのはブルガス大学にTIMES( 85.7% )であったが、これは予測のドメイン構造に限られ、低いApplicabilityを持つためである( 9.7% )。他に、高いA-sensitivityを示したのものとしてラーサ社のSarah(83.1%)が挙げられる。Sarahはsensitivityも最も高かった( 70.4% )。一方、Specificityに関してはMCASE社のCASE Ultra-statisricalが92.8%と最も高く、FUJITUのADMEWORKS ( 87.8% )、STSRのAZAMES ( 87.7% )が後に続いた。全体の一致率( concordance )はUltra-statisrical ( 87.3% )、AZAMES( 83.9% )が高かった。

表4にPhase I, II, IIIでのA-sensitivity、sensitivity、Specificity、Applicabilityをまとめた。Phase IでのConcordanceは平均75%、Phase IIでは80%、Phase IIIでは76%であり、Phase IIIでの全体的な予測性の改善は認められなかった。一方、Sensitivityに注目すると、Phase IIからPhase IIIでわずかに改善が認められた。

### C-3. QSARで検出できない強いエームス陽性化合物(偽陰性)

今回の4,409化学物質には236の強いエームス陽性化合物が含まれる(A判定)。しかしながら、このうち6化合物については、全てのQSARツールで、10化合物については1つを除くほとんどQSARツールで陰性、もしくは評価不能と判定された(図1)。これら化合物は”False Negative(偽陰性)”化合物と考えられる。

### C-4. QSARで検出できないエームス陰性化合物

今回の4,409化学物質のうち3,780化合物はエームス陰性である(C判定)。しかしながら、このうち13化合物については、全てのQSARツールで、別の10化合物については1つもしくは2つを除くほとんどQSARツールで陽性の判定がされた(図2)。これら化合物は”False Positive(偽陽性)”化合物と考えられる。

## **D. 考察**

これまで開発されたQSARツールの予測性の検証には、多くの場合、開発者が有する内部データによって検証されてきた。内部データはQSAR開発のためのトレーニングデータセットを含むため、その予測率が高いのは同然で有り、客観的評価となっていない。特に、化学メーカーや製薬企業が公開するデータの多くは、エームス陰性化合物で有り、それを基に開発されるQSARツールは、Specificityは高いものの、Sensitivityは低いか、もしくはその値は信頼性に乏しい。規制当局としては、スクリーニング試験であるQSARはSensitivityと

陰性予測率 ( Negative Prediction Value; NPV ) が十分に高いことが、偽陰性を避ける上で重要である。これにより、変異原性ハザードの高い化学物質が、安易に環境中に放出されること無く、適切に管理されることが可能となる。従って、外部データを用い、QSARモデル、QSARツールの精度を評価することは、QSAR開発ベンダーにとっても、QSAR利用者にとっても、そして規制側にとっても有益である。また、これら結果を各QSARベンダーで共有し、更に実際に試験データを開示することは、QSARベンダーの公正な競争により各QSARツールの予測性の向上に繋がる。本研究班では、このような研究を通じ、QSARベンダー、QSARユーザー、規制当局が“ウィン、ウィン、ウィン”の関係となることを目指す。

今回の4,409化合物の予測精度に関しては、ICH-M7を契機にエームス試験の予測性の向上に積極的に取り組んでいるラーサ社、マルチケース社、リードスコープ社のQSARツールの予測性に期待が持たれた。3社は、知識ベース、統計ベースの両者のモデルをコマーシャルベースで提供している。概して、3社のQSARツールはSensitivity、Specificityともバランスがよく、74~87%の一致率 ( Correspondence ) を示した。しかしながら、PhaseIからIIIまでの全体の一致率を見てみると、Phase Iでの一致率は平均75%、Phase IIでの一致率は平均80%であったが、PhaseIIIでは76%であり、PhaseIIからの予測率は低下した。その理由として特異性の低下が考えられた。PhaseIIIの化合物の5%分子量800以上のエームス陰性化合物で有り、その95%がエ

ームス陰性であるが、これらを正しく陰性と予測できなかったことが原因と考えられる ( 表2 )。事実、感度に関してはPhaseIIよりわずかな改善が認められた。3回のトライアルにより、本共同研究は成功裡に終了した。さらに、様々な構造を持つ化学物質での試験データを収集し、トレーニングデータとしてQSARツールに組み込み、ケミカルスペースを拡大させることが、QSARの予測精度の向上に繋がるものと考えられる。また、トレーニングデータのいくつかには間違っ、陰性、もしくは陽性とラベルされているものが確認された。これらデータはQSARモデルに開発に大きな障害となる。正確な試験結果データのみからなるベンチマークデータセットを整備し、トレーニングデータとすることが精度の高いQSARモデルの開発に重要である。

ほとんど全てのQSARツールの全てが陰性と間違っ判定したClass A化学物質が6化合物存在した。これは偽陰性物質で有り、この割合が多いことに規制当局がQSARの利用に対して最も懸念する部分である。Phase IIIでは、これはClassA物質の2.5%に相当する。この割合はPhase Iで10%、PhaseIIでは3.6%にも存在したことから、この場合でも共同研究は順調に推移し、偽陰性は減少していることが示された。すなわち、多くの試験データを組み入れエームス陽性のケミカルスペースを拡大することが、偽陰性を減らすことに繋がる。

## E. 結 論

エームス試験予測QSARモデルの向上を目指した国際共同研究を実施している。日本で実施された安衛法に基づくエームス試

験データから、QSARモデルの開発に有用と考えられる、12,140化学物質をデータベース化した。H29年度はこのうち4,409化合物について、エームス試験結果の予測精度を競うPhase IIIトライアルを行った。世界6カ国から、11のQSARベンダーがこの国際競争に参画し、19のQSARツールについて予測精度の評価を行った。多くのQSARツールの予測精度はプロジェクト開始前と比較して大幅向上されたが、PhaseIIからPhaseIIIの向上は顕著ではなかった。予測率の向上には、新規データセットをトレーニングデータとして単純に組み入るだけでなく、物質の特徴（分子量）や、既存のトレーニングデータを適切にレビューする必要がある。QSARモデルのさらなる改良には、科学的且つ、信頼性の高いベンチマークデータセットの開発が鍵である。

## F. 健康危機情報

なし

## G. 研究発表

### 誌上発表

1. Petkov, PI, Schultz TW, Honma M, Kirilov K, Kotov S, Mekenyan OG. Predicting in vitro genotoxicity by mouse lymphoma L5178Y thymidine kinase mutation assay (MLA): Accounting for simulated metabolic activation of chemicals. *Computational Toxicology*. 4, 45-53, 2017.  
Doi:10.1016/j-comtox.2017.10.002
2. Gadaleta D, Porta N, Vrontaki E, Manganelli S, Manganaro A, Sello G,

Honma M, Benfenati E. Integrating computational methods to predict mutagenicity of aromatic azo compounds. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev*. 35, 239-257, 2017. doi: 10.1080/10590501.2017.1391521.

3. 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、Vol.223 No.01、8-16、2018

### 学会発表

1. 本間正充、Ames/QSAR International Collaborative Study、口頭、第7回国際遺伝毒性試験国際ワークショップ、東京、2107/11/8
2. 本間正充、In Silico Approaches in Genetic Toxicology -Progress and Future- 口頭、第12回国際環境変異原学会、韓国、2017/11/15
3. 本間正充、AOP-based Evaluation of Chemical Mutagenicity and Development of New Endpoints and Models 口頭、第12回国際環境変異原学会、韓国、2017/11/14
4. 本間正充、In Silico Approaches in Genetic Toxicology -Progress and Future- 口頭、第17回中国環境変異原学会年次大会、中国、2017/12/7
5. 本間正充、QSARの最近の進歩について、第2回ICH M7関連ワークショップ、東京、2017/5/23

## H. 知的財産権の出願・登録状況

なし

表 1 : Participants in Ames/QSAR International Challenge Project

QSAR Vender	QSAR Tool	Methodology
1. Lhasa Limited (UK)	a. DEREK Nexus	Rule
	b. SARAH Nexus	Statistical
2. MultiCASE Inc (USA)	c. CASE Ultra statistical-based	Statistical
	d. CASE Ultra rule-based	Rule
3. Leadscope Inc (USA)	e. Leadscope statistical-based	Statistical
	f. Leadscope rule-based	Rule
4. (Italy) IRCCS - Istituto di Ricerche Farmacologiche Mario Negri	g. CAESAR	Statistical
	h. SARPY	Statistical
	i. KNN	Statistical
5. LMC - Bourgas University (Bulgaria)	j. OASIS/TIMES	Hybrid*
6. Istituto Superiore di Sanita (Italy)	k. Toxtree	Rule
7. Prous Institute (Spain)	l. Symmetry	Statistical
8. Swedish Toxicology Science Research Center (Sweden)	m. AZAMES	Statistical
9. FUJITSU KYUSHU SYSTEMS LIMITED (Japan)	n. ADMWORKS	Statistical
10. IdeaConsult Ltd. (Bulgaria)	o. AMBIT	Statistical
11. Molecular Networks GmbH and Altamira LLC (USA)	p. ChemiTunes and ToxGPS	Statistical
12. Simulation Plus (USA)	q. Mut_Risk-0	Statistical

表3: Summary of the results of prediction by QSAR models in Phase III\_4409

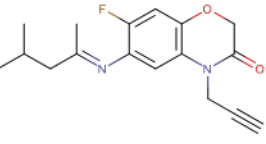
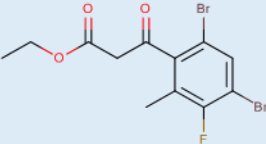
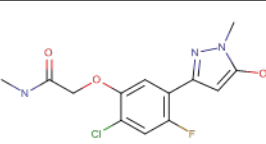
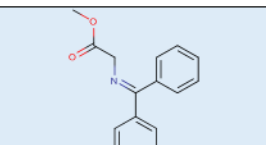
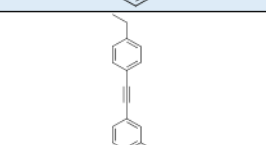
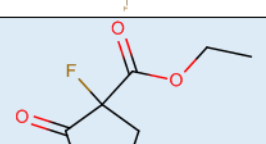
Builders	QSAR Models	QSAR_Results	Ames Strong positive A (++)	Ames Positive B (+)	Ames Negative C (-)	Ames data Total	Sensitivity [A] (%)	Sensitivity [A+B] (%)	Specificity (%)	Concordance (%)	Applicability (%)
Lhasa Limited (UK)	Derek Nexus v. 5.0.1	positive	107	177	833	977	70.8	54.7	83.3	79.2	100.0
		negative	89	216	3147	3432					
		total	236	393	3780	4409					
	Sarah Nexus v. 2.0.1 (model v.1.1.19)	positive	104	103	541	748	59.1	44.0	82.3	77.3	80.2
		negative	72	191	2524	2787					
		total	176	294	3065	3555					
	Sarah research prototype prediction	positive	147	169	649	965	83.1	70.4	74.4	73.8	67.7
		negative	30	103	1888	2021					
		total	177	272	2537	2966					
	Sarah Nexus v. 2.0.1 (model v.1.1.19)	positive	135	156	682	952	72.2	60.5	78.1	75.7	79.5
		negative	52	137	2383	2552					
		total	187	292	3025	3504					
MultiCASE Inc (USA)	Statistical approach: SALMIECOLI consensus  RULE BASED (GT_EXPERT)	Positive	132	116	238	486	69.8	50.6	92.8	87.3	85.9
		Negative	97	185	3081	3303					
		total	189	301	3299	3789					
Leadscope Inc (USA)	Leadscope gene tox statistical QSAR (Salmonella v3(internal candidate) and E.coli/TA102 v1)	positive	139	174	547	860	69.8	59.6	83.6	80.3	87.3
		negative	60	152	2779	2991					
		total	199	326	3326	3851					
Leadscope Inc (USA)	Leadscope gene tox expert alerts suite (Bacterial mutagenicity v2)	positive	143	153	434	730	69.4	55.5	87.1	82.8	88.4
		negative	63	174	2931	3168					
		total	206	327	3365	3898					
Istituto di Ricerche Farmacologiche Mario Negri (Italy)	CAESAR (training set based on Katus et al)	positive	184	241	1246	1671	78.0	67.6	67.0	67.1	100.0
		negative	52	152	2534	2738					
		total	236	393	3780	4409					
	SARPY (training set based on Katus et al)	positive	171	227	1484	1882	72.5	63.3	60.7	61.1	100.0
		negative	85	168	2298	2527					
		total	236	393	3780	4409					
	KNN (training set based on Hansen et al + strong positive in the first phase)	positive	145	179	1139	1463	61.7	51.8	69.6	67.1	99.2
		negative	90	212	2609	2911					
		total	235	391	3748	4374					
Bourgas University (Bulgaria)	TIMES AMES mutagenicity v.14.14. (In domain)	positive	18	13	46	77	85.7	47.0	87.3	81.0	9.7
		negative	3	32	315	350					
		total	21	45	361	427					
Bourgas University (Bulgaria)	TIMES AMES mutagenicity v.14.14. (All)	positive	151	163	822	1136	64.0	50.0	78.2	74.2	99.9
		negative	85	229	2953	3267					
		total	236	392	3775	4403					
Istituto Superiore di Sanità (Italy)	ToxTree 2.6.6	positive	173	210	1165	1548	73.3	60.9	69.2	68.0	100.0
		negative	63	183	2815	2861					
		total	236	393	3780	4409					
Swedish Toxicology Science Research Center (Sweden)	SveToxZAMES v2	positive	168	174	427	769	77.1	61.0	87.7	83.9	91.2
		negative	50	169	3033	3252					
		total	218	343	3460	4021					
FUJITSU KYUSHU SYSTEMS (Japan)	ADMWORKS AMES ver7.1.0	positive	100	105	348	553	60.6	46.3	87.8	82.2	74.5
		negative	65	173	2495	2733					
		total	165	278	2843	3286					
IdeaConsult Ltd. (Bulgaria)	Ambit consensus model	positive	166	212	748	1126	70.3	60.1	80.0	77.1	99.1
		negative	70	181	2994	3245					
		total	236	393	3742	4371					
Molecular Networks GmbH and Altamira LLC	Altamira/MN prediction	positive	180	206	624	1010	78.6	65.2	82.7	80.3	95.4
		negative	49	157	2989	3195					
		total	229	363	3613	4205					
Simulation Plus (USA)	MUT_Risk@5	positive	174	205	1018	1397	76.0	61.5	72.0	70.5	96.4
		negative	55	182	2617	2854					
		total	229	387	3635	4251					
Average							72.1	57.7	78.9	75.9	86.4
Minimum							59.1	44.0	60.7	61.1	9.7
Maximum							85.7	70.4	92.8	87.3	100.0



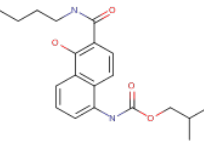
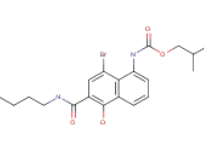
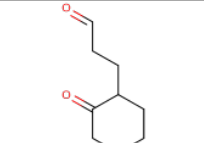
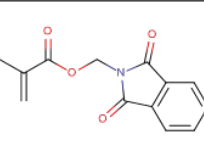
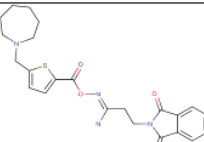
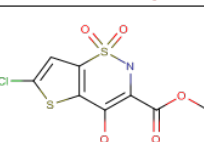
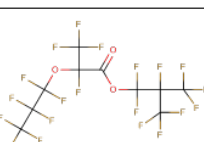
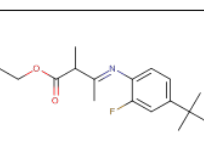
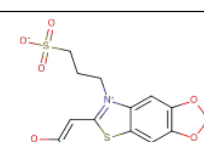
表4: Summary of the results of prediction by QSAR models in the Ames/QSAR challenge project\_12,140

	Phase I	Phase II	Phase III
<b>A-sensitivity</b>	68.9 (51.4-82.8)	73.2 (58.5-89.5)	72.1 (59.1-85.7)
<b>Sensitivity</b>	55.8 (38.6-70.0)	57.6 (44.8-72.1)	57.7 (44.0-70.4)
<b>Specificity</b>	78.6 (62.5-91.5)	84.1 (64.9-93.5)	78.9 (60.7-92.8)
<b>Concordance</b>	75.3 (63.6-83.9)	80.2 (65.8-87.7)	75.9 (61.1-87.3)
<b>Applicability</b>	86.7 (14.5-100)	85.7 (18.0-100)	86.4 (9.7-100)

☒ 1: Class A chemicals, but all QSAR tools predicted “negative “ (false negatives)

Serial Id	ANEI No.	CAS#	Chemical_Name	Structure	Result	Mol Weight
6275	8-(7)-1104	NO_CAS	6-[(1,3-dimethylbutylidene)amino]-7-fluoro-4-(2-propenyl)-3,4-dihydro-2H-1,4-benzoxazine-3-one		A	302.349
7070	4-(7)-1776	NO_CAS	3-(4,6-dibromo-3-fluoro-2-methylphenyl)-3-oxopropionic acid ethyl ester		A	382.023
8218	8-(2)-1758	NO_CAS	2-[2-chloro-4-fluoro-5-(5-hydroxy-1-methyl-1H-pyrazol-3-yl)phenoxy]-N-methylacetamide		A	313.71
12639	7-(4)-1058	NO_CAS	2-[(diphenylmethylene)amino]acetic acid methyl ester		A	253.301
12798	7-(4)-1059	NO_CAS	4-[(4-ethylphenylethynyl]-1,2-difluorobenzene		A	242.269
20043	3-(3)-170	NO_CAS	ethyl 1-fluoro-2-oxo-cyclopentane-1-carboxylate		A	174.171

☒2: Class A chemicals, but almost QSAR tools predicted “negative “ (false negatives)

Serial	JANEI No.	CAS#	Chemical_Name	Result	Mol Weight	Structure
3561	5-1118	NO_CAS	N-[6-(N-butyl carbamoyl)-5-hydroxy-1-naphthyl] carbamic acid isobutyl ester	A	358.4	
3565	5-1119	NO_CAS	N-[8-bromo-6-(N-butyl carbamoyl)-5-hydroxy-1-naphthyl] carbamic acid isobutyl ester	A	437.3	
7351	3-(4)-433	NO_CAS	3-(2-oxocyclohexyl) propanal	A	154.2	
7762	8-(1)-2411	NO_CAS	methacrylic acid 1,3-dioxo-2H-isoindole-2-ylmethyl ester	A	245.2	
10063	8-(1)-2688	NO_CAS	N-[2-[N2-[5-(hexahydro-1H-azepine-1-yl)methyl]-2-thienyl carbonyloxy]amido]ethyl phthalimide	A	454.6	
11025	8-(7)-1337	NO_CAS	6-chloro-4-hydroxy-1,1-dioxo-1-thia (VI)-2H-thieno[2,3-e]-1,2-thiazine-3-carboxylic acid methyl ester	A	295.7	
12457	2-(6)-1576	NO_CAS	2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)propanoic acid 2-(trifluoromethyl)hexafluoropropyl ester	A	548.1	
13405	4-(7)-2114	NO_CAS	3-(4-tert-butyl-2-fluorophenylimino)-2-methylbutanoic acid ethyl ester	A	293.4	
15383	8-(7)-1520	NO_CAS	3-[[6-(2-ethoxy-1-butenyl)-1,3-dioxolo[4,5-f]benzothiazole-7-ium]-7-yl]propane-1-sulfonate	A	399.5	
15410	5-1372	NO_CAS	6-chloronaphthalene-2-sulfonic acid	A	242.7	