

経済産業省 委託業務報告書

平成 31 年度省エネ型電子デバイス材料の
評価技術の開発事業
(機能性材料の社会実装を支える
高速・高効率な安全性評価技術の開発)

調 査 報 告 書

令和 2 年 3 月

MIZUHO

みずほ情報総研株式会社

< 目次 >

1.	概要	1
2.	事業の内容・計画	1
3.	実施体制	1
4.	事業成果の活用等に関する調査・検討	2
5.	まとめ.....	18

1. 概要

「省エネ型電子デバイス材料の評価技術の開発事業（機能性材料の社会実装を支える高速・高効率な安全性評価技術の開発－毒性関連ビッグデータを用いた人工知能による次世代型安全性予測手法の開発－）」の研究開発について、基本計画に則り、研究開発の目標を達成するために、事業成果の活用等に関する調査・検討および本プロジェクトの効率的・効果的な推進に関する検討を行った。

2. 事業の内容・計画

(1) 産業界からの情報提供及び成果物の普及・活用に向けた調査・検討

プロジェクトの成果物が産業界等で活用されることを前提に、化学系企業等の事業者が本プロジェクトの成果物に対して求めるニーズについて調査した。また、事業者から動物実験データ等の情報提供を受けてシステム自体を発展させていくため、本プロジェクトで開発したプロトタイプシステムについて事業者からの意見を聴取し、情報の機密性の確保等について検討し、情報提供を推進するためのスキームの検討を行った。

なお、事業者の協力が得られた場合には、提供された実験データ等の情報の予測モデルへの実装を行うことを想定していたが、協力を得ることができなかった。

(2) システム仕様検討のための関連調査

本プロジェクトの研究開発戦略に反映させることを念頭に、関連する予測モデル等の *in silico* 手法の研究・開発・評価に関する日本及び米国・欧州の代表的な国での研究・開発状況の調査や情報収集を行い、本プロジェクトのシステム研究開発に資する。

3. 実施体制

本項目は、図 3-1 に示す体制で実施した。

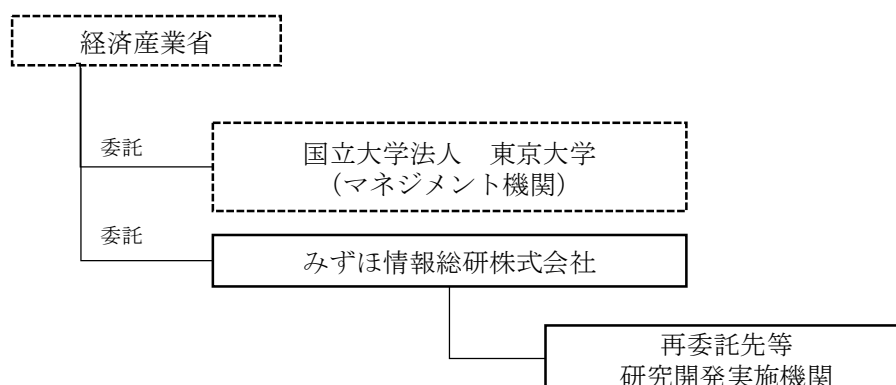


図 3-1 実施体制

4. 事業成果の活用等に関する調査・検討

(1) 産業界からの情報提供及び成果物の普及・活用に向けた調査・検討

プロジェクトの成果物が産業界等で活用されることを前提に、化学系企業等の事業者が本プロジェクトの成果物に対して求めるニーズについて調査を行った。

また、事業者から動物実験データ等の情報提供を受けてシステム自体を発展させていくため、情報の機密性の確保等について検討し、情報提供を推進するためのスキームの検討を行った。

① ニーズ調査の結果

化学系企業等の事業者に対するヒアリングを実施し、AI-SHIPS に対する意見を聴取した。ヒアリング日時は表 1、ヒアリング項目は表 2 の通り。

表 1 ヒアリングの日時

事業者	日時
A	2019年12月16日 15時30分～17時
B	2019年12月16日 15時～16時30分
C	2020年3月17日 10時～11時
D	2020年3月17日 15時～15時30分

表 2 ヒアリング項目

① 貴社における人健康 QSAR 等の予測モデルの利用状況・目的について
② 活用している QSAR 等ツールの詳細、予測対象エンドポイント、メリット、課題について
③ AI-SHIPS プロジェクトの予測対象について
④ AI-SHIPS プロジェクト予測モデルの活用方法について
⑤ プロトタイプ試用にあたっての条件（情報セキュリティ等）、懸念点
⑥ 使用料（システムの更新・メンテナンスのため）の相場観について
⑦ 自社データを取り込む機能の要否について
⑧ オーダーメイドの毒性予測支援サービスに対するニーズ、相場観について
⑨ 本プロジェクトで <i>in vitro</i> 実験値を取得する際の協力可能性について
⑩ その他ご意見・ご要望、AI-SHIPS に期待すること

ヒアリングにより得られた、事業者が AI-SHIPS に対して求めるニーズは表 3 の通り。

表 3 事業者が AI-SHIPS に対して求めるニーズ

項目	ニーズ
エンドポイントについて	<ul style="list-style-type: none"> ✓ 全身毒性は、肝、腎、血液だけではないので、神経毒性、心毒性などについては他の方法で補完しないとイケない。欠落しているエンドポイントがあった場合に、そのエンドポイントで影響がないという判断をするのは難しいからである。 ✓ 化審法で受け入れられるためには、他の臓器も必要ではないか。 ✓ 内分泌かく乱、副腎、神経毒性等も予測できるとよい。 ✓ 化審法では分解物の評価も必要である。親化合物との比較ができると、大きなメリットである。つまり、代謝物が親化合物より活性が低いということが分かれば、試験回避の説明ができる。分解物については、体内だけでなく、環境放出時の光やオゾンでの分解も評価も同時にできると化審法においては利用価値が上がるだろう。 ✓ 主要でないエンドポイントは、信頼性が低くてもよいので、信頼性と共に結果を示してくれれば利用価値はある。 ✓ 数値が得られるとよいが、精度次第である。できれば NOEL だけでなく NOAEL も知りたい。安全性の判断をするなら NOEL、次に行う毒性試験等の気づきに使うなら、NOAEL が欲しい。 ✓ Study NOEL が得られないと、使用場がかぎられるのでは。Study NOEL も出してほしい。 ✓ NOEL/LOEL はどちらであっても利用できる。GHS 分類に活用できるのは LO(A)EL なので、LOEL も一定のニーズはあるのではないか。
閾値について	<ul style="list-style-type: none"> ✓ 化審法対応であれば、閾値は 300 mg/kg、30 mg/kg でよいと思う。 ✓ 食品添加物や化粧品を考慮するのであれば、閾値 1000 mg/kg もあるとよい。OECD テストガイドラインの投与上限量 1000 mg/kg で毒性がでないか、ということと判定してくれると使い道があるのではないか。 ✓ リスク評価に用いる場合には、NO(A)EL の数値が得られる方がよい。AI-SHIPS では閾値 300 の 2 値分類とのことであるが、NOEL ≥ 300 mg/kg ということの信頼性が高ければ、NOEL=300 mg/kg として評価に用いることも可能だろう。ただし、他のエンドポイント（神経毒性等）が POD (Point of Departure) になる場合、300 mg/kg という数値は使えない。
AI-SHIPS プロジェクト予測モデルの活用方法	<ul style="list-style-type: none"> ✓ 物質を評価する場合、in silico や、read across、in vitro の結果等も合わせて、Weight of Evidence で総合的に判断することになる。その際に AI-SHIPS も in silico の 1 つとして取り入れるということになるだろう。 ✓ 製品の評価に AI-SHIPS を使っていくとなると、この結果のみでの判断はできず、反復投与試験のデータをとることになるだろう。むしろ反復投与試験の用量設定や、試験系の設計に役立つかもしれない。 ✓ スクリーニングとしての利用については、現状で使えるツールがあるので、AI-SHIPS に寄せる期待は大きくない。AI-SHIPS に最も期待するのは、化審法や海外規制の申請で受け入れてもらえるものにしていただきたいという点である。 ✓ 既存の in silico ツールの多くは、定量評価ができない。肝毒性アラートが出るツールはあるが、NOEL 等の値が出てこないで、その後の評価が難しい。そのため、AI-SHIPS では定量評価をしていただけるとよい。 ✓ 改正食品衛生法に基づく、容器包装材の添加剤等のリスク評価において、毒性予測に 28 日反復投与試験結果が必要である。その場面に AI-SHIPS が活用できるかもしれない。 ✓ 社内開発原料のスクリーニングだけでなく、規制当局への説明資料にも使える情報としたい。
精度について	<ul style="list-style-type: none"> ✓ スクリーニングであれば 8 割程度の精度があれば十分である。ただし、過小評価しないという前提は必要である。また誤回答が多くなってしまうのはどのような化合物なのかの情報を提供してほしい。(適用範囲、ケミカルスペースの明確化。当局の受け入れ性を指標に判断してもらいたい) ✓ 毒性発現機序の解釈に有用な情報が提示されるようなものであれば、スクリーニングの次の段階で利用することになるかもしれない。

4 事業成果の活用等に関する調査・検討

項目	ニーズ
QSAR Toolbox との連携について	<ul style="list-style-type: none"> ✓ QSAR Toolbox には、複数の計算結果を一度に得られるような仕組みがあり、AI-SHIPS の結果も QSAR Toolbox から得られると助かる。 ✓ QSAR Toolbox に AI-SHIPS が収載されると、広く認知されることで信頼性が高まるかもしれない。
ソフトウェアの使用条件（セキュリティ等）	<ul style="list-style-type: none"> ✓ 利用者がカスタマイズできるようなものであればインストール版になるだろう。 ✓ 単に利用するだけなら、社内規定上インストールすることのハードルが高いため、Web の方がよい。 ✓ 最終的には、インストール版がよい。プロトタイプのトライアルについては、秘匿情報を入力することもないので Web 版でもよいだろう。 ✓ インストール版の場合、データベースの種類（PostgreSQL 等）が重要である。HESS は脆弱性診断にひっかかるため社内許可を得るのが大変であった。
使用料	<ul style="list-style-type: none"> ✓ 例えば年間 10 万円程度で会員であれば、多くのメーカーが会員になってくれるのではないかと。 ✓ システムの更新・メンテナンスをしていかないといけないものなので、その費用を賄うための使用料ということであれば、高額でなければ、受け入れられるであろう。 ✓ 有償のツールは年間数百万の費用がかかる。AI-SHIPS のメリットとの兼ね合いであるが、それよりは安くなってほしい。 ✓ これを使えば化審法の試験が不要ということになれば、相当の価値があるだろう。
オーダーメイドの毒性予測支援サービス	<ul style="list-style-type: none"> ✓ AOP 説明や、データの解釈の支援サービスがあるとよい。 ✓ 必ずしも専門家や部門をかかえていない企業もあるので、あるとよいだろう。 ✓ 社内で実施するよりコストメリットがあれば考えるかもしれない。 ✓ 当局への申請などへの活用を視野に入れ、新システムの妥当性に関する一連の説明データ（当局側への説明や論文文化等の支援）まで提示してもらえると魅力が出てくる。
自社データを取り込む機能	<ul style="list-style-type: none"> ✓ 自社のデータを入れてデータベースとして管理できる機能があれば、ありがたい。 ✓ 自社データを取り込んで、さらにそのデータに基づく学習による独自のモデル開発を可能とするオプションを有料化していただくことも考えられる。
データの協力について	<ul style="list-style-type: none"> ✓ データを提供することによって精度が上がり、化審法届出に AI-SHIPS が活用できるということがあれば社内の納得は得られるだろう。 ✓ データ提供にあたっては、一般のユーザーにはその物質の骨格等の情報が見られないようにするとか、規制当局だけが見ることができるなど、CBI に配慮した仕組みがないと難しいだろう。
その他	<ul style="list-style-type: none"> ✓ in vitro 試験結果も予測モデルを作成しているようだが、予測に予測を重ねることになり、精度、信頼性が落ちることが危惧されるのではないかと。簡単な in vitro 試験を必要とするスキームでもよいのではないかと。 ✓ 5 年のプロジェクトでのターゲットは肝、腎、血液であるが、その後、全身毒性を評価できるように継続していくのか。全体のロードマップが知りたい。 ✓ どのようなアウトプットが見られるのか不透明である。早めにプロトタイプを出してほしい。 ✓ 化審法への適用をするということであれば、プロジェクト後の運用については、民間ではなく、公的な機関が管理すべきだろう。民間だと、メンテナンスやサービスの継続性に不安がある。 ✓ DEREK などの既存ツールとの差別化が必要だろう。 ✓ 国内に存在するあらゆるデータを収集してほしい。厚労省や農水省もデータを持っている。また、医薬品であれば代謝のデータ、農薬であれば環境中での代謝という情報も入っているので、それらも収集してほしい。 ✓ まずはラットの毒性（28 日反復投与毒性）の予測を達成してほしい。そのうえで、将来的に技術基盤を展開することによりヒトに対する毒性の予測もできればなおよい。（限られたエンドポイントでもよいの。例：代謝）

②情報提供を推進するためのスキームの検討

事業者に対するニーズ調査の結果から得られた、事業者からの情報提供を求めるにあたっての課題及び情報提供推進スキーム案は表 4 の通り。

表 4 事業者からの情報提供を求めるにあたっての課題及び情報提供推進スキーム案

課題	課題の説明	情報提供推進スキーム案
事業者が情報提供を行うメリットの確保	<ul style="list-style-type: none"> ✓ インハウスデータを外部に提供するために社内の承認を得る必要があり、そのためにはメリットを社内に十分に説明する必要がある。 ✓ 事業者側のメリットとして、一番大きな事項は、AI-SHIPS が化審法に活用されることであり、その場合には化審法への具体的な活用予定及び時期等について示していく必要がある。 ✓ その他、事業者が情報提供を行うメリット（例：協力事業者として公表等）を検討していく必要がある。 	<ul style="list-style-type: none"> ✓ 事業者からのデータをデータベース化するサービスを実施する。（例：試験報告書を HESS データ化し提供する） ✓ 協力事業者として公表を行う。 ✓ 提供データに対して、in vitro 試験を実施し、提供された化学物質付近のケミカルスペースの信頼性を向上させる。
事業者提供データの秘匿性の確保	<ul style="list-style-type: none"> ✓ 現時点でシステム構成が不明のため、提供データの秘匿性が十分に確保されないのではないか（提供した化合物が類似化学物質としてシステム内に表示される等）、という事業者の懸念がある。 ✓ 当該懸念を払拭するために、具体的なシステムの仕様の確定（提供データは統計処理されており、ユーザーには提示されない等）が必要である。 	<ul style="list-style-type: none"> ✓ 秘匿性が確保されたプロトタイプを事業者に提示する。
提供方法の明確化	<ul style="list-style-type: none"> ✓ 事業者にとって、どのレベルの情報（試験報告書、要約等）やどのような形式で提供すればよいか不明確である。 ✓ そのため、具体的な提供様式を策定する必要がある。 	<ul style="list-style-type: none"> ✓ HESS データ形式を提供し、さらに入力のためのガイダンス及び入力支援を実施する。

なお、本年度は事業者から提供データが得られていないため、予測モデルへの実装は行っていない。

(2) システム仕様検討のための関連調査

本プロジェクトの研究開発戦略に反映させることを念頭に、本プロジェクトのシステム研究開発に資するべく関連する予測モデル等の *in silico* 手法の研究・開発・評価に関する日本及び米国・欧州の代表的な国での研究・開発状況の調査や情報収集を行った。

①日本

最新の国内の事例としては、2015年より日本医療研究開発機構（AMED）による創薬支援推進事業「創薬支援インフォマティクスシステム構築」¹において実施されている、「多階層データの統合モデリングによる薬物動態予測システムの構築」（水口ら）および「肝毒性予測のためのインフォマティクスシステム構築に関する研究」（山田ら）の開発事例がある。

1) AMED「創薬支援インフォマティクスシステム構築」—多階層データの統合モデリングによる薬物動態予測システムの構築—

国立研究開発法人医薬基盤・健康・栄養研究所 AI健康・医薬研究センターセンター長 水口 賢司氏がリーダーをつとめ、理化学研究所、慶應義塾大学、京都大学が参画するプロジェクトである。当該プロジェクトの「薬物動態基礎データベースと薬物動態予測モデルの構築」について、概要を表 5 に示す。

¹ 創薬支援推進事業「創薬支援インフォマティクスシステム構築」研究課題評価 令和元年度実施 事後尾評価報告書（創薬支援インフォマティクスシステム構築事業研究課題評価委員会、2020年）

表 5 薬物動態予測システムの国内事例

プロジェクト名称	多階層データの統合モデリングによる薬物動態予測システムの構築					
開発関係者 (敬称略)	水口賢司 (医薬基盤・健康・栄養研究所) 泰地 真弘人 (理化学研究所) 大澤匡範 (慶應義塾大学) 中津井雅彦 (京都大学)					
使用データベース ^{3,2}	プラットフォーム : DruMAP (Drug Metabolism and pharmacokinetics Analysis Platform) ³ 合計 30,000 超の化合物に係るデータ、製薬企業のインハウスデータ 【基本版データベース】 ・化学構造式 ・各種公開データベース (ChEMBL, KEGG Drug, hERGcentral など)からのキュレーションデータ ・市販化合物を用いた新規取得データ 【企業提供データ (約 24,300 化合物)】 ・構造記述子 ・実創薬データ (溶解性・代謝安定性・血漿タンパク結合・hERG 阻害など) ・企業提供化合物を用いた新規取得データ					
モデル名称	fu,p class ⁴	fu,p ⁴	fu,brain ⁵	Fa ⁶	Papp ⁶	D-sol ⁶
入力	Mordred ⁷ , PaDEL ⁸		CDK ⁹ , Mordred ¹⁰ , PaDEL ¹¹			
予測手法	Random Forest ¹²		Gradient Boosting	Random Forest	R-SVM	L-SVM
出力	Low (0.001-0.05), Medium (0.05-0.2), High (0.2-1.0)	fu,p 値	fu,brain 値	Low (0-0.2), Moderate (0.2-0.7), High (0.7-1.0)	Low (< 10 ⁻⁵ cm/s) High (> 10 ⁻⁵ cm/s)	Low (< 10 µg/mL) High (> 10 µg/mL)
予測精度	Accuracy: 0.676, True positive rate in Low class: 0.826	R ² = 0.691	R ² = 0.630, RMSE = 0.477	Accuracy: 0.836 Kappa: 0.560	Accuracy: 0.810 Kappa: 0.601	Accuracy: 0.811 Kappa: 0.628

² 第 6 回保健医療分野 AI 開発加速コンソーシアム「医薬健栄研における AI 関連研究」(米田、2019)

³ <https://drumap.nibiohn.go.jp/>

⁴ Predicting fraction unbound in human plasma from chemical structure: improved accuracy in the low value ranges (Watanabe et al, 2018)

⁵ Computational model to predict the fraction of unbound drug in the brain (Esaki et al., 2019)

⁶ Constructing an in silico three-class predictor of human intestinal absorption with Caco-2 permeability and dried-DMSO solubility. J. Pharm. Sci. (Esaki, T. et al., 2019)

⁷ Mordred: a molecular descriptor calculator. J. Cheminform. (Moriwaki et al, 2018)

⁸ Yap, C. W., PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. (Yap, C. W, 2011)

⁹ The Chemistry Development Kit (CDK): an open-source Java library for chemo- and bioinformatics. J. Chem. Inf. Comput. Sci. (Steinbeck, C et al., 2003)

¹⁰ Mordred: a molecular descriptor calculator. J. Cheminform. (Moriwaki et al, 2018)

¹¹ Yap, C. W., PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. (Yap, C. W, 2011)

¹² Random Forest. Machine Learn. (Breiman, L., 2001)

2) AMED「創薬支援インフォマティクスシステム構築」—肝毒性予測のためのインフォマティクスシステム構築に関する研究—

国立研究開発法人医薬基盤・健康・栄養研究所 山田弘氏がリーダーとなり、国立研究開発法人 医薬基盤・健康・栄養研究所、国立研究開発法人 産業技術総合研究所、明治薬科大学、熊本大学が参画するプロジェクトである。主に肝毒性を対象に計 4 つの予測モデルが構築されており、ここでは肝毒性に係る 3 つのモデルについて、表 6 に概要を記す。

表 6 肝毒性予測システムの国内事例

プロジェクト名称	肝毒性予測のためのインフォマティクスシステム構築に関する研究		
開発関係者 (敬称略)	山田弘 (医薬基盤・健康・栄養研究所) 福井一彦 (産業技術総合研究所) 植沢芳広 (明治薬科大学)		
公開データベース	<ul style="list-style-type: none"> ● DILI-cSEARCH¹³ (Drug-Induced Liver Injury information Cross-SEARCH system): 医薬基盤・健康・栄養研究所および熊本大学が開発。Open TG-GATEs、DrugMatrix database、Library of Integrated Network-Based Cellular Signatures (LINCS)、Liver Toxicity Knowledge Base (LTKB) から収集、整理・統合した DB。 ● TOXPILOT¹⁴ (TOXic Process InterpretabLe OnTology): 医薬基盤・健康・栄養研究所が作成。オントロジー工学を応用して開発した肝毒性作用機序解釈支援のための知識システムであり、オントロジーライブラリ、RDF データベース、情報を可視化するための WEB アプリケーションシステムから構成される。 ● toxBridge¹⁵: 産業技術総合研究所が作成した、肝毒性 in vivo/vitro ブリッジングに着眼したデータベース。 		
モデル名称	DILI-PANEL ¹⁶	toxRank ¹⁷	LIVER/MIE-QSAR ¹⁸
概要	in vitro のヒト肝細胞における遺伝子発現データをもとに、肝毒性フェノタイプを予測およびオントロジーに基づいた毒性機序マップを出力。	in vivo のラットを用いた試験(単回、反復)および in vitro のヒト肝細胞またはラット肝細胞を用いた試験データをもとに、化合物投与時の変動遺伝子(up と down 分子刻印)を入力として、類似化合物と有意遺伝子・パスウェイ等を出力。	化学構造情報を入力とし、JAPIC-FAERS および TOX21 の DB をもとに肝毒性を予測。
入力	CEL ファイル(Affymetrix マイクロアレイ (HG-U133_Plus_2)のデータファイルかつヒト肝細胞の 24 時間暴露のデータ)	マイクロアレイのデータ (ヒトのチップ: Affymetrix 社 HG-U133_Plus_2、ラット: Rat230_2)	化合物の構造情報 (SMILES,SDF,MOL)
使用データベース	DILI-cSEARCH、TOXPILOT	toxBridge, Open TG-GATEs	PubChem、JAPIC AERS
予測手法	機械学習 (サポートベクターマシン)	超幾何分布解析	機械学習
出力	①肝毒性マーカーパネル: 肝毒性フェノタイプ毎の判定結果、②作用機序マップ: 想定される毒性機序と機序解釈支援知識システム	類似度をもとにしたランクマトリクス、インスタンスに対するクラスタリング、ネットワーク解析結果 (Single Data Point only)、エンリッ	訓練データの散布図、毒性予測の信頼度、毒性の予測値

¹³ <https://dili-csearch.nibiohn.go.jp/>

¹⁴ <https://toxipilot.nibiohn.go.jp/>

¹⁵ <https://toxbridge.molprof.jp/top>

¹⁶ <https://dili-panel.nibiohn.go.jp/manual?locale=ja>

¹⁷ https://toxrank.molprof.jp/toxrank/static/ranking/help/ToxRank_UserGuide_JPN_ver1.0.pdf

¹⁸ http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new

<https://u-lab.my-pharm.ac.jp/~mminfo/study01/>

4 事業成果の活用等に関する調査・検討

	TOXPILOT における各種関連情報へのリンク、③化合物ランキング：肝毒性化合物中でのランキング、④ヒートマップ：肝毒性化合物との遺伝子発現パターン比較	チメント解析 (Single Data Point only)	
--	---	---------------------------------	--

②米国

米国においては、2008年より Tox21 プロジェクト (The Toxicology in the 21st Century) が実施された。当該プロジェクトは、EPA、NIH (NCAT を含む)、NCGC、FDA による共同プロジェクトとして、迅速な毒性評価手法の開発を目的としたものである。Tox21 ではロボットを用いることによって、1536 ウェルマイクロタイタープレートを1日当たり 100 枚測定可能なアッセイ系を開発し、約 10,000 の化合物について 50 超の定量的ハイスループットスクリーニングアッセイを実施し、その膨大なアッセイデータを収載した公開データベースは AI-SHIPS プロジェクトにおいても活用されている。また、ToxCast では Tox21 のサブセットとなる約 1800 の化合物について 1200 に及ぶアッセイを追加実施し、そのデータを蓄積している。

この Tox21 および ToxCast のデータを使用した米国のプロジェクトとして、近年 HTTK (High Throughput Toxicokinetics) プロジェクトが開始された。米国の TSCA が The Frank R. Lautenberg Chemical Safety for the 21st Century Act による改正を受けた際に、EPA はリスク評価実施時に暴露状況を考慮することが義務付けられており、この暴露状況の試算において、暴露後のヒトの体内の化合物濃度予測が必要となっている。これに資するデータを予測するため、HTTK プロジェクトでは、体内の薬物動態予測モデル (PBTK モデル) の検討、in vitro の試験結果から in vivo への外挿方法 (IVIVE) の研究、およびオープンソースのデータとモデルの提供を目標として研究が進められている^{19,20}。

1) Tox21 Data Challenge

米国国立衛生研究所 (NIH) の先進トランスレーショナル科学センター (National Center for Advancing Translational Sciences : NCATS) は、Tox21 のデータを用いて毒性予測モデルの精度を競うコンペティション「Tox21 Data Challenge 2014」を 2014 年 7 月から 2014 年 11 月にかけて主催した。これは、モデル構築のための教師データを提供し、化学構造情報から毒性を引き起こす可能性のある化合物の細胞および生化学経路への干渉をどの程度予測できるか、研究者によるデータ分析を競争的に「クラウドソース」というものである。予測対象は、核内受容体とストレス応答パスウェイに関する 12 のアッセイ (表 4-7) で、18 か国の参加者がこれらのアッセイによる約 8,000 種類の化合物の活性と構造情報を用いて予測モデルの構築を行った。さらに参加者には、アッセイ結果が未公開の化合物約 300 種類の化学構造 (テストセット) が与えられた。このテストセットについて参加者が予測した結果は、予測対象となる

¹⁹ HTTK: R Package for High-Throughput Toxicokinetics (US EPA, 2017)
https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=337720&Lab=NCCT

²⁰ The High Throughput Toxicokinetic (HTTK) R Package (US EPA, 2019)
<https://www.epa.gov/sites/production/files/2019-06/documents/comptoxcop-wambaugh-httk-r-package-062719-draft5.pdf>

アッセイごとに ROC 曲線下面積によって評価され、評価結果は登録者にフィードバックされた。この作業を繰り返すことにより、予測精度を段階的に向上させることが可能となる。

表 4-7 Tox21 Data Challenge の予測対象

<p>核内受容体</p> <ul style="list-style-type: none">(1) エストロゲン受容体αリガンド結合部位 (ER, LBD)(2) エストロゲン受容体α全配列 (ER, full)(3) アロマターゼ (aromatase)(4) 芳香族炭化水素受容体 (AhR)(5) アンドロゲン受容体全配列 (AR, full)(6) アンドロゲン受容体リガンド結合部位 (AR, LBD)(7) ペルオキシソーム増殖因子活性化受容体γ (PPAR-γ) <p>ストレス応答パスウェイ</p> <ul style="list-style-type: none">(8) 転写因子 Nrf2/アンチオキシダント応答配列 (Nrf2/ARE)(9) 熱ショック因子応答配列 (HSE)(10) ATPase ファミリー, AAA ドメイン含有 5 (ATAD5)(11) ミトコンドリア膜電位 (MMP)(12) p53
--

最終的なモデル評価では、40 チームから 378 のモデルが提出され、ほぼすべてのモデルで予測精度が 80%を超え、いくつかのモデルでは 90%を超える結果となった。優勝チームの用いた記述子やアルゴリズムを表 8 に示す²¹。

²¹ Ruili Huang, Menghang Xia et al., *Front. Environ. Sci.*, [https://doi.org/10.3389/fenvs.2015.00085\(2016\)](https://doi.org/10.3389/fenvs.2015.00085(2016))

表 8 Tox21 Challenge において優勝チームが用いた記述子とアルゴリズム²¹

項目	内容
構造記述子	<ul style="list-style-type: none"> • MOE (Chemical Computing Group Inc., Montreal, Canada) • ChemAxon (Chem Axon LLC., Cambridge, MA) • Dragon (Talete SRL, Milan, Italy) • PaDel (Yap,2011⁸) • RDKit²² • PubChem fingerprint²³ • GSFrag (Tetko et al.,2005²⁴) • ISIDA fragments (Ruggiuetal., 2010²⁵) • ESTATE indices (HallandKier, 1995²⁶) • Alog PS (Tetko and Tanchuk, 2002²⁷) • CDK (Steinbeck et al., 2003⁹) • Inductive descriptors (Cherkasov, 2005²⁸) • Adriana.Code (Molecular Networks GmbH, Erlangen, Germany) • QNPR (Thormann et al., 2007²⁹) • MERA,andMerSy (Bartashevich et al., 2002³⁰)
アルゴリズム	<ul style="list-style-type: none"> • Random Forest (Breiman,2001¹²) • Deep neuralnetworks (Schmidhuber, 2014³¹) • support vector machines (SVM) (Cortes and Vapnik, 1995³²) • Elastic Nets (Zou and Hastie, 2005³³) • Gradient Boosting Decision Trees (Friedman, 1999³⁴) • Extra Trees (Geurts et al. ,2006³⁵) • Associative neural networks (Tetko, 2008³⁶) • k-Nearest Neighbors (Altman, 1992³⁷)

²² <http://www.rdkit.org/>²³ <https://pubchem.ncbi.nlm.nih.gov/>²⁴ Tetko, I. V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., et al. (2005). Virtual computational chemistry laboratory—design and description. *J. Comput. Aided Mol. Des.* 19, 453–463. doi:10.1007/s10822-005-8694-y²⁵ Ruggiu, F., Marcou, G., Varnek, A., and Horvath, D. (2010). ISIDA property-labelled fragment descriptors. *Mol. Inform.* 29, 855–868. doi:10.1002/minf.201000099²⁶ Hall, L. H., and Kier, L. B. (1995). Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comput. Sci.* 35, 1039–1045. doi:10.1021/ci00028a014²⁷ Tetko, I. V., and Tanchuk, V. Y. (2002). Application of associative neural networks for prediction of lipophilicity in ALOGPS2.1 program. *J. Chem. Inf. Comput. Sci.* 42, 1136–1145. doi:10.1021/ci025515j²⁸ Cherkasov, A. (2005). Inductive QSAR descriptors. Distinguishing compounds with antibacterial activity by artificial neural networks. *Int. J. Mol. Sci.* 6, 63–86. doi:10.3390/i6010063²⁹ Thormann, M., Vidal, D., Almstetter, M., and Pons, M. (2007). Nomen est omen: quantitative prediction of molecular properties directly from IUPAC names. *Open Appl. Inform. J.* 1, 28–32. doi:10.2174/1874136300701010028³⁰ Bartashevich, E. V., Potemkin, V. A., Grishina, M. A., and Belik, A. V. (2002). A Method for multiconformational modeling of the three-dimensional shape of a molecule. *J. Struct. Chem.* 43, 1033–1039. doi:10.1023/A:1023611131068³¹ Schmidhuber, J. (2014). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi:10.1016/j.neunet.2014.09.003³² Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi:10.1007/BF00994018³³ Zou, H., and Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc. B* 67, 301–320. doi:10.1111/j.1467-9868.2005.00503.x³⁴ Friedman, J. H. (1999). Greedy function approximation: a gradient boosting machine. Available online at: <http://statweb.stanford.edu/~jhf/ftp/trebst.pdf>³⁵ Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi:10.1007/s10994-006-6226-1³⁶ Tetko, I. V. (2008). Associative neural network. *Methods Mol. Biol.* 458, 185–202. doi:10.1007/978-1-60327-101-1_10³⁷ Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* 46, 175–185.

この Challenge によって構築された多数の予測モデルの予測結果を集約し、コンセンサスモデルが構築された。その性能は、受賞したモデルと同等程度との結果であったが、予測結果の集約方法によって性能が変化することが示された。例えば、個別モデルの結果を平均化した場合、コンセンサスモデルの性能は個別モデルの 86%を超える性能となったが、個別モデルの AUC-ROC が 0.8 超の優良なモデルのみを選択した場合には、96%を超える性能となった。また、予測対象アッセイ 12 種のうち 6 種について、コンセンサスモデルが個別の受賞モデルを超える性能を示し、「群衆の叡智 (wisdom of the crowd)」が実証されることとなった。

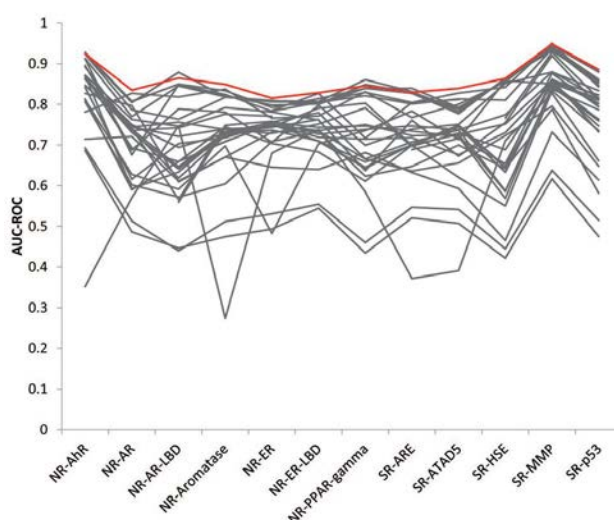


図 1 Tox21 Data Challenge の成果を活用したコンセンサスモデルの性能²¹
(灰色：個別モデル、赤：コンセンサスモデル)

2) HTKK

Tox21 のデータを使用した米国のプロジェクトとして、近年 HTKK (High Throughput Toxicokinetics) プロジェクトが開始された。このプロジェクトにおいては、PBPK モデルの検討を伴う *in vitro* の試験結果から *in vivo* への外挿方法の研究、およびオープンソースのデータとモデルの提供を目標として研究が進められている³⁸、オープンソースのフリーソフトウェアである R を利用したモデルを提供している。IVIVE の概念図を図 2 に、HTKK で提供される 4 つのモデルの概要を表 9 に記す。

³⁸ The High Throughput Toxicokinetic (HTTK) R Package (US EPA, 2019)
<https://www.epa.gov/sites/production/files/2019-06/documents/comptoxcop-wambaugh-httk-r-package-062719-draft5.pdf>

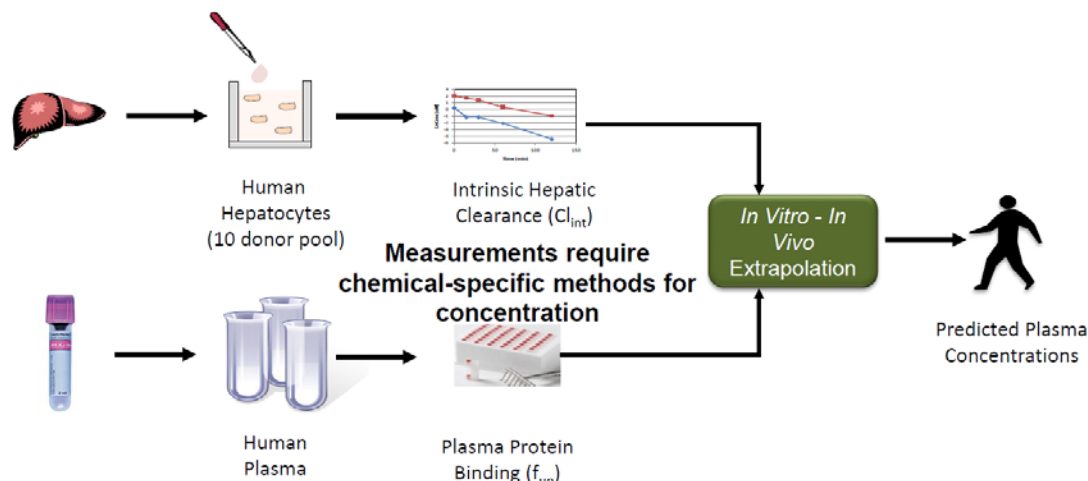


図 2 High Throughput Toxicokinetics による in vitro-in vivo 外挿の概略 (IVIVE)
 (出典：米国 EPA, Science Inventory, High-Throughput Toxicokinetics (HTTK)³⁹)

表 9 薬物動態予測システムに係る米国の事例^{20,40,41}

プロジェクト名称	The High Throughput Toxicokinetic (HTTK)			
概要	一般的な PBTK (Physiologically-based Toxicokinetic) モデルに in vitro のトキシコキネティクスデータを加えることで In vitro-in vivo の外挿手法 (IVIVE) の確立と、医薬品における有効用量の予測システム構築を目標とするプロジェクト。			
開発関係者	HTTK Team (US EPA, FDA, NTP, cyprotex が参画)			
モデル名称	PBTK モデル	1-compartment モデル	three-compartment モデル	three-compartment steady-state モデル
対象臓器	消化管、肝臓、肺、血管 (動脈、静脈)、腎臓	血漿	消化管、肝臓 (点滴投与のみ)	消化管、肝臓
入力	肝クリアランス、分配係数、非結合分画、ヘマトクリット値、分子量	分子量、排出速度、分布容積	肝クリアランス、分配係数、非結合分画、ヘマトクリット値、分子量	肝クリアランス、非結合分画、分子量
学習データ	933 物質 (医薬品および ToxCast でアッセイを実施した物質)			
予測手法	逆線量測定法、モンテカルロサンプリング			
出力	組織中濃度の動的予測、定常状態の予測			等価線量 (経口)

³⁹ https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=537346&Lab=NCCT

⁴⁰ htk: R Package for High-Throughput Toxicokinetics (Robert G. et al., 2017)
<https://www.istatsoft.org/article/view/v079i04>

⁴¹ https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NCCT&dirEntryId=345861

③欧州

欧州では、SEURAT-1 や EU-ToxRisk などのプロジェクトで動物実験に代わる試験方法、予測手法が検討されてきたほか、EFSA により反復投与毒性試験の予測システムとして OpenFoodTox の開発が進められている。

1) SEURAT-1

SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing-1) は欧州委員会および欧州化粧品工業会が出資し、70 の組織（企業、大学、研究機関、工業団体等）からなる研究コンソーシアムである。幅広い動物実験代替手法の検討を行っており、反復投与試験を研究対象とするプロジェクトは主に以下の3つが進行中である⁴²。

- COSMOS (The in silico toxicology project) Delivery of computational tools to predict effects of chemicals based on in silico calculations and estimation techniques
- NOTOX (The systems biology project) Development of systems biological tools for organotypic human cell cultures
- DETECTIVE (The biomarker project) Detection of endpoints and biomarkers for repeated dose toxicity using in vitro systems

2) EU-ToxRisk

EU-ToxRisk は欧州を中心に多くの企業、研究機関や大学、国（米国や日本を含む）と連携した欧州の新しい作用機序に基づく毒性試験とリスクアセスメントの実施を進めるうえでフラッグシップとなるプロジェクトとして 2016 年から進行している⁴³。プロジェクトの目的を表 10 に示す。EU-ToxRisk の作業計画では、様々なケーススタディーの実施し、目的に応じた IATA (Integrated Approaches for Testing and Assessment) の確立及び Read-Across アプローチの改良を目指すとしている。

表 10 EU-ToxRisk プロジェクトの目的⁴⁴

- ヒト細胞の応答及び化学物質の有害性影響発現の包括的メカニズムの理解に基づく、非動物試験の化学物質評価手法を推進する。
- 化学物質の暴露から有害性影響の発現の関連を理解するために、細胞生物学、オミックス、システム生物学、計算科学を統合する。
- エンドポイントは、反復投与毒性（肝臓、腎臓、肺、神経系）、生殖発生毒性。
- プロジェクトでは、メカニズムに基づく化学物質安全性試験戦略の概念実証を提供する。
- 最終的な目標は、化学物質の有害性とリスク評価を可能にする信頼性の高い非動物試験戦略を提供する。

⁴² http://www.seurat-1.eu/media/What_is_SEURAT-1%20-%20Michael%20Schwarz.pdf

⁴³ <https://www.eu-toxrisk.eu/>

⁴⁴ <https://www.eu-toxrisk.eu/page/en/about-eu-toxrisk.php>

EU-ToxRisk プロジェクトでは、サブプロジェクトとして ToxTree⁴⁵、OpenTox⁴⁶、Caesar Project⁴⁷などが進行しており、これらに市販の DEREK や MultiCASE などのモデルが利用されている⁴⁸。

- ToxTree: 皮膚感作性、眼刺激性、変異原性等の予測システム
- OpenTox: 様々な毒性予測モデルをウェブ経由で利用可能な予測システム
- Caesar Project: 蓄積性、皮膚感作性、発がん性、変異原性、発達毒性の予測システム

3) OpenFoodTox のデータを活用したモデル構築

欧州食品安全機関（EFSA）は、有害性情報のための新しいワンクリックツールとして OpenFoodTox の開発を進めている。これは、緊急事態に際し、EFSA の科学的専門家とスタッフが科学的助言を提供する作業を容易にすること、また、毒性値の予測や、ヒトおよび環境リスク評価の閾値を分類するための QSAR モデルのような革新的な *in silico* ツールの開発を支援するためのデータソースとなることを意図したものである。概要を、図 3 に示す。

このデータベースには、2003 年から 2018 年 3 月までに発行された 4700 種類以上の化学物質についての 9000 件以上の評価結果を含む 1800 件以上の文書（意見・声明・結論）から抽出したデータが含まれている。このうち 90% について、代表的な化学構造情報（SMILES）が関連付けられており、化合物ごとに、毒性試験結果（NOAEL 等）、用量（TDI/ADI、ARfD 等）、遺伝毒性が閲覧可能である⁴⁹。

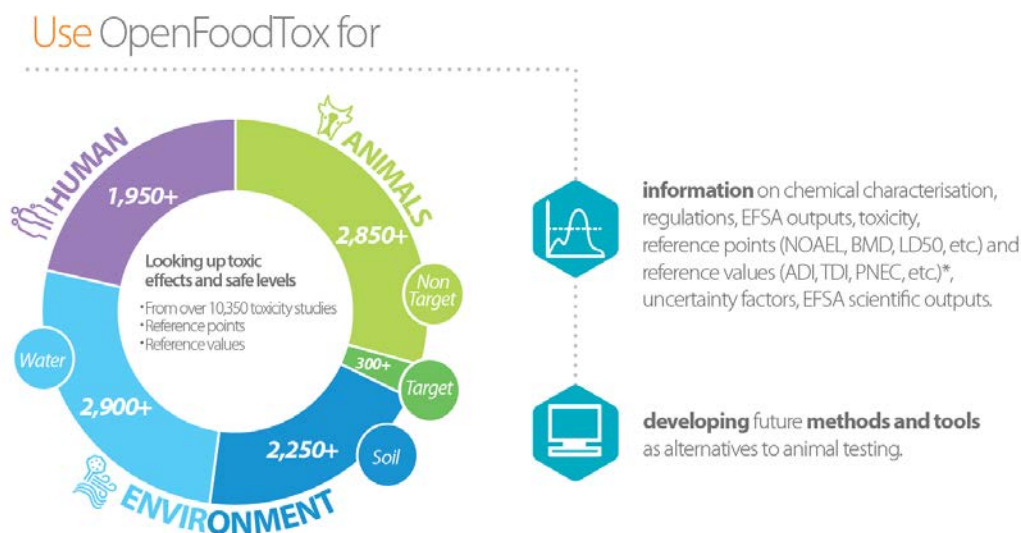


図 3 OpenFoodTox の目的⁵⁰

⁴⁵ <http://toxtree.sourceforge.net/>

⁴⁶ <https://opentox.net/>

⁴⁷ <http://www.caesar-project.eu/>

⁴⁸ 海外における計算科学的手法を用いた化学物質のリスク予測研究の現状と今後の展開（庄野、2019）

⁴⁹ 国立医薬品食品衛生研究所安全情報部 食品安全情報（化学物質）No. 3/ 2017（2017. 02. 01）

<https://www.nihs.go.jp/hse/food-info/foodinfonews/2017/foodinfo201703c.pdf>

⁵⁰ Infographic: OpenFoodTox – chemical hazards database

<http://www.efsa.europa.eu/en/discover/infographics/openfoodtox-chemical-hazards-database>

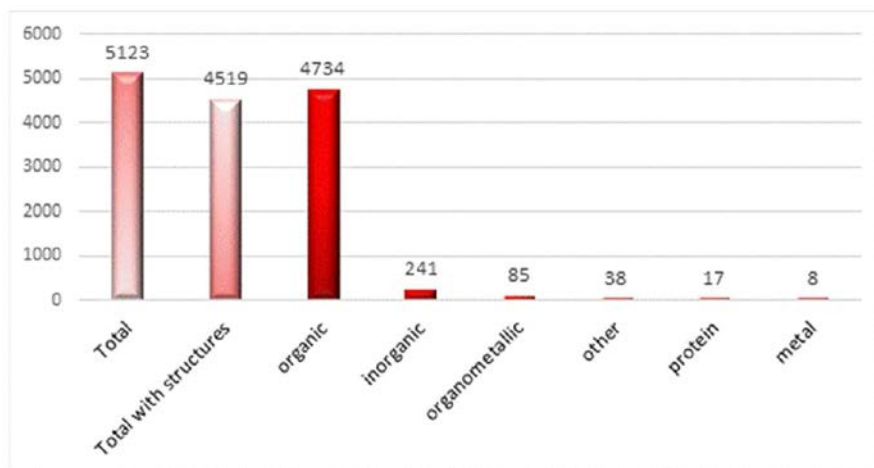


図 4 データベースを構成する化学物質の種類⁵¹

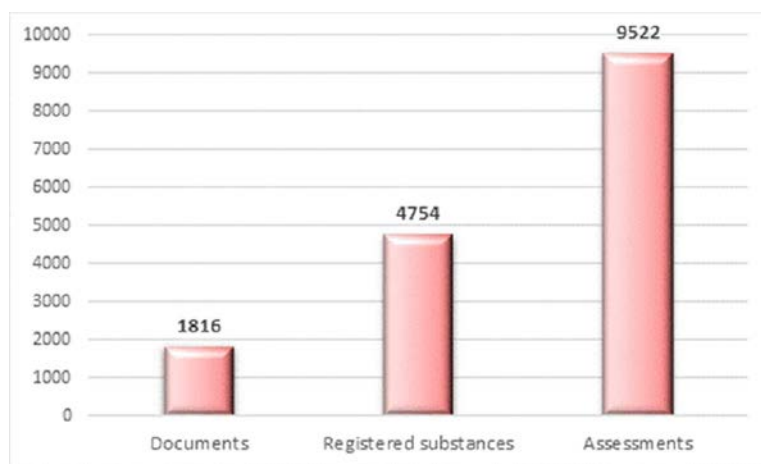


図 5 データベースを構成する文書数、物質数、評価件数（過去のものも含む）⁵¹

OpenFoodTox は、統一的にデータを収集し、構造化するために、OECD Harmonised テンプレートに従って構築されている。この OpenFoodTox に搭載されたラットの亜慢性毒性のデータを用いて QSAR モデルが開発された。一例として、The CORAL software について、表 11 に示す (Tropova et al., 2018 年⁵²)。また、VEGA プラットフォーム (Benfenati et al., 2017 年⁵³) を用いたグローバルモデルも開発された。

⁵¹ Update and maintenance of the EFSA's Chemical Hazards Database
<https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2018.EN-1438>

⁵² Toropova AP, Toropov AA, Marzo M, Escher SE, Dorne JL, Georgiadis N and Benfenati E, 2018. The application of new HARD-descriptor available from the CORAL software to building up NOAEL models. Food and Chemical Toxicology 112: 544-550.

⁵³ Benfenati E, Como F, Marzo M, Gadaleta D, Toropov A and Toropova A, 2017. Developing innovative in silico models with EFSA's OpenFoodTox database. EFSA supporting publication 2017: EN-1206. [19 pp.]. doi:10.2903/sp.efsa.2017.EN-1206.

表 11 哺乳類の慢性反復毒性予測システムに係る欧州の事例 (The CORAL software)

プロジェクト名称	OpenFoodTox ^{54, 55, 56}
概要	90 日反復投与ラットの亜慢性毒性試験の NOEL を予測
開発関係者 (敬称略)	EFSA、フラウンホーファー毒性・実験医学研究所 (ドイツ)、マリオネグリ薬理学研究所 (イタリア)
モデル名称	The CORAL software ⁵⁷
入力	CORAL 記述子→8 つの元素 (窒素、酸素、硫黄、リン、フッ素、塩素、臭素、ヨウ素) と化学結合の種類 (二重結合、三重結合、立体構造 (stereo chemical bond)) を用いたハイブリッドな記述子で、SMILES と hydrogen-suppressed graph (HSG) も考慮に入れている。
使用データベース	OpenFoodTox (128 物質、76 日暴露以上を抽出) に Fraunhofer @RepDose Database (362 物質、亜慢性試験を抽出) を加え、重複分削除したものを使用。最終的な物質数は 487 物質。
予測手法	モンテカルロ法
出力	NOAEL
予測精度	R ² 値→0.58-0.70 (トレーニングセット)、0.50-0.70 (バーチャルセット) coverage は 100%

⁵⁴ Openfoodtox: over 300 substances added to EFSA chemical hazards database
<http://www.efsa.europa.eu/en/press/news/180719>

⁵⁵ Editorial: OpenFoodTox – EFSA's open source toxicological database on chemical hazards in food and feed
<http://www.efsa.europa.eu/en/efsajournal/pub/e15011>

⁵⁶ Developing innovative in silico models with EFSA's OpenFoodTox database
<http://www.efsa.europa.eu/en/supporting/pub/en-1206>

⁵⁷ <http://www.insilico.eu/coral>

5. まとめ

本調査項目では、本プロジェクトの成果物の活用に向けて、事業成果の活用等に関する調査・検討および本プロジェクトの効率的・効果的な推進に関する検討を実施した。

特に、成果物のユーザーたる化学系企業等の事業者にとって有用なものとするためには、法制度における活用のあり方や、利用シナリオなど、具体的な活用方法の検討が重要となる。今回得られた事業者ニーズをもとに、次年度は具体的な予測システムの機能や利用スキームに落とし込んでいくこととなる。

また、関連する予測モデル等の *in silico* 手法の研究・開発・評価に関する日本及び米国・欧州の代表的な国での研究・開発状況など、各国、地域の先進的な研究開発の戦略に学ぶところは多い。米国では、ウルトラハイスループットスクリーニングアッセイを実施し、このデータを活用した研究開発が進められており、特に **Tox21 Data Challenge** のような競争的なスキームでの研究開発や、**HTTK** プロジェクトのような *in vitro* 試験結果から *in vivo* への外挿方法の研究が精力的に進められている。また欧州では、フラッグシップとなるプロジェクトとして **EU-ToxRisk** プロジェクトが進行しており、**AI-SHIPS** プロジェクトに類似するような *in vitro* 試験手法や予測モデル開発が進められている。こうした先進的な事例は、**AI-SHIPS** プロジェクトで得られたインビトロ試験結果の活用や予測モデル開発の技術的な課題解決等を検討する上で参考となるものであり、引き続き注視すべきと考えられる。

またさらに、技術的な側面だけでなく、こうしたプロジェクトの成果を諸外国が医薬、食品等他分野における有害性評価、リスク評価へと活用していくため、どのようなロードマップを展開していくのかという点にも注目する必要がある。こうした事例を参考にすることで、具体的な制度への適用だけでなく、データ形式の調和化や国際展開、他分野への応用といった展開を想定することで、プロジェクトの成果の幅広い活用へとつなげることが可能となるであろう。