

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 事後評価報告書

令和2年1月

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

— 目次 —

1. 事業の概要
2. 評価の概要
3. 各研究課題の評価結果

## 1. 事業の概要

本事業では創薬支援のための基盤強化を図るため、医薬品や化合物に関する情報を格納した統合型データベースを構築し、新規化合物の代謝、毒性、薬効領域等を予測する多元的構造活性相関の手法を開発し、創薬支援に活用することを目的としている。平成 27 年度から 5 カ年の計画で実施しており、平成 29 年度から企業連携により、膨大かつ貴重な企業データを取り組んだ実用的なシステムの構築に取り組んでいる。事業終了後の令和 2 年 4 月には、全データの検索・閲覧が可能なオンラインデータベース及び予測システムの公開を予定している。

## 2. 評価の概要

平成 27 年度に開始された「創薬支援インフォマティクスシステム構築事業」について、平成 29 年度に中間評価を実施し、予定通りもしくはそれ以上の進捗・成果があるとの評価を得た。今回、令和元年度に事後評価を実施し、研究課題評価実施要綱に基づき委員会を設置し、各研究課題の進捗状況・成果等を明らかにし、今後の研究成果等の展開及び事業の運営の改善に資することを目的に、評価を実施した。

本評価報告書は、以下の 3 課題について評価結果を取りまとめたものである。

No.	研究課題名	分担研究者（研究代表者は◎）
1	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）
2	インフォマティクスとシミュレーションを融合した多面的心毒性予測システムの構築	◎本間光貴（理化学研究所） 関野祐子（東京大学）
3	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）

なお、評価委員会においては、各研究課題の評価にあたり、次ページの各評価項目に基づき、総合的に評価が実施された。

## 評価項目

### ①研究達成状況について

- ・研究計画に対する達成状況はどうか

### ②研究成果について

- ・予定していた成果が着実に得られたか
- ・創薬支援ネットワークに資するものであるか
- ・成果は新技術の創出に資するものであるか
- ・必要な知的財産の確保がなされたか

### ③実施体制

- ・研究代表者を中心とした研究体制が適切に組織されていたか
- ・十分な連携体制が構築されていたか

### ④今後の見通し

- ・今後、研究成果のさらなる展開が期待できるか

### ⑤事業で定める項目及び総合的に勘案すべき項目について

- ・生命倫理、安全対策に対する法令等を遵守していたか

### ⑥総合評価

10段階評価により、①～④及び下記の事項を勘案して総合評価する

## 3. 各研究課題の評価結果

各研究課題の評価結果は次ページのとおりである。なお、本事業の統括評価については、「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（研究代表者：水口賢司）」の事後評価報告書に記載した。

日本医療研究開発機構  
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」  
事後評価報告書

## I 基本情報

研究課題名：(日本語) 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  
(英語) Development of a pharmacokinetics prediction system using multiscale integrated modeling

研究実施期間：2015年9月28日～2020年3月31日(予定)

研究代表者 氏名：(日本語) 水口 賢司  
(英語) Kenji Mizuguchi

研究代表者 所属機関・部署・役職：

(日本語) 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

(英語) Kenji Mizuguchi, Director, Artificial Intelligence Center for Health and Biomedical Research

## II 研究の概要

### 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

本研究は対象化合物の化学構造から各種物性及び薬物動態関連パラメータを予測し、さらにこれらの結果を基に体内動態を予測できるシステムの構築を目的としている。

これまでに、化合物数 25,000 個規模のデータベースを構築し、そこからデータの検索、閲覧、各種薬物動態関連パラメータを予測するための Web 上の統合解析プラットフォーム DruMAP (Drug Metabolism and pharmacokinetics Analysis Platform) を一部公開した。データベースにはキュレーション済みの質の高い公共データのみならず、2015-2018 年度に本プロジェクトにおいて、統一した条件で収集した大規模新規実験データが収録されている。また、新規実験データ及び公共データベースのデータを用いて予測モデルの構築を行い、代謝安定性 (Esaki et al., Mol. Inform., 2019)、血漿タンパク結合率予測モデル (Watanabe et al., Mol. Pharm., 2018)、吸収率予測モデル (Esaki et al., J. Pharm. Sci., 2019)、脳ホモジネート予測モデル (Esaki et al., J. Chem. Inf. Model., 2019) に関する論文を発表した。並行して代謝酵素選択性予測モデル及び基質予測モデルを構築し、構築済みの 13 の予測モデルのうち 10 について、DruMAP への組み込みを完了した。中枢移行性予測については、P 糖タンパク質 (P-gp) 輸送能試験の結果を用いた補正の有用性を検討し、中枢移行性の予測方針及び方法を

決定した。この予測方法の有効性を検証するために、これまでに取得したラットのデータを用いて有効性を確認し、中枢移行性予測に必要な P-gp 輸送能予測モデルの構築に着手した。また、さらなる予測精度の向上を目的として、国内製薬企業複数社が保有する薬物動態関連データを収集した。この企業連携により取得したデータを用いて、インハウスのコンピュータ上でのモデル構築、予測、評価など一連の作業の自動化を目指し、これまでに記述子計算、予測モデル作成、予測実行の3つを行うプログラムを完成した。

The aim of this study is to predict various physicochemical properties and pharmacokinetic parameters from the chemical structure of a compound of interest, and to build a system that can simulate the pharmacokinetics of the compound reliably based on these results. We have developed a database with 25,000 compounds, released an initial version of the integrated analysis platform named DruMAP (Drug Metabolism and pharmacokinetics Analysis Platform) on the web. The database contains not only high-quality, curated public data, but also large-scale, novel experimental data that we collected under unified conditions. In addition, we have so far built and published four predictive models: metabolic stability (Esaki et al., Mol. Inform., 2019), plasma protein binding (Watanabe et al., Mol. Pharm., 2018), absorption rate (Esaki et al., J. Pharm. Sci., 2019), and brain homogenate binding (Esaki et al., J. Chem. Inf. Model., 2019). In parallel, predictive models for a metabolic enzyme (cytochrome P450; CYPs) and metabolic sites were built. Currently, 10 out of the 13 constructed predictive models have been incorporated into DruMAP. Regarding the prediction of brain penetration, the effect of the P-glycoprotein (P-gp) transport, obtained from *in vitro* assays, was examined, and an appropriate prediction method was defined. We have confirmed the effectiveness of this method using the observed rat *in vitro* data. Furthermore, to improve the prediction accuracy of pharmacokinetic parameters, large-scale pharmacokinetic data were obtained from a group of pharmaceutical companies in Japan. We have developed software to automate a series of operations such as model construction, prediction, and evaluation on in-house computers.

#### 分子シミュレーションによる薬物代謝酵素シトクロム P450 (CYP) に対する薬物代謝の予測法開発

シトクロム P450 (CYP) は、薬物代謝に関与する酵素である。CYP による薬物の代謝部位を実験によって調べることは時間とコストがかかるという問題点がある。そのため、CYP の薬物代謝部位予測における計算科学の担う役割は大きい。そこで、我々は、分子シミュレーション (分子動力学シミュレーション、分子ドッキング等) を利用して、シトクロム P450 (CYP) による薬物の代謝部位を予測する手法の研究を行って

きた。CYP による薬物代謝は、ヘム分子を含む活性部位への薬物接近の過程（アクセシビリティ）と活性部位での薬物代謝反応の過程（リアクティビティ）が逐次的に起こることで達成される。この点に着目して、本研究では、①薬物のアクセシビリティと②薬物のリアクティビティを調査し、③ ①と②の知見を統合した薬物代謝部位の予測法を確立およびその適用調査を行った。これまで調査の結果、4つの CYP 種（CYP1A2、CYP3A4、CYP2D6、CYP2C9）に対する薬物の代謝部位を精度良く予測できる手法とその計算プログラムを開発した。また、この分子シミュレーションを高速に実行するための専用計算機 MDGRAPE-4 普及版の開発を行った。

We have studied a molecular simulation-based method to predict the sites of drug metabolism by cytochrome P450 (CYP). Our method adopts the concept that the drug metabolic mechanism by CYP consists of two sequential steps, access of a drug molecule into the active site (accessibility) and catalytic reaction in drug metabolism (reactivity). We thus have investigated the following three research topics by using molecular simulation: ① research of accessibility, ② research of reactivity, and ③ research of prediction of sites of drug metabolism based on both accessibility and reactivity evaluations. So far, we have developed the effective computational method to predict the site of drug metabolism by several CYP enzymes (CYP1A2, CYP3A4, CYP2D6, and CYP2C9). To accelerate MD simulations, we have also developed a special-purpose computer for MD simulations, MDGRAPE-4A.

### **Structure-free** による化合物ライブラリーデザイン法の開発

化合物構造をあらわに扱わず、セキュリティを確保した状態において、大規模なライブラリーの化合物データを収集・統合・解析するための計算化学の解析システムの構築を行った。公開されている 1,500 を超える創薬標的に対する化合物の生物活性予測法の開発を行った。これにより得られた生物活性情報からライブラリーの生物活性多様性を評価するシステムを開発した。リード化合物創成に繋がるコアライブラリーの作成を目的として、化学的類似性・物性・生物活性の 3つの観点から多様性を持つ化合物ライブラリーを作成する方法を完成させた。これを利用して、DISC ライブラリーを対象として、商用化合物ライブラリー（約 500 万化合物）から 2 万個程度の化合物によるコアライブラリーの作成を行い、その多様性の検証を行った。本研究課題の成果は DISC ライブラリーをはじめとする協働創薬への貢献が期待される。

We constructed a secure system of computational chemistry to collect, integrate, and analyze chemical compounds data in large-scale libraries without exact compound structures. We developed a method for predicting bioactivity of compounds against over 1,500 targets for drug discovery. By analyzing the resultant bioactivity information, we developed a system of assessing bioactive diversity of

libraries. We made a new method to construct a chemical library in consideration of its diversity in chemical similarity, molecular properties, and bioactivity, aimed for a core library for discovering lead compounds. Using this method, we constructed a core library to represent DISC library, by sampling about 20 thousand compounds from commercial compound libraries (about 5 million compounds) and confirmed its diversity. The result of this research project is expected to contribute to collaborative drug discovery such as DISC project.

### 製薬・医療関係文書からの情報抽出と基礎データベースの拡充

当研究は、CTD などの製薬・医療関係文書から、薬物動態・毒性パラメータを中心とした情報を抽出し、構造化を実施することにより、これまでに構築されてきた基礎データベースの拡充を図るため、以下の事項の調査検討およびプロトタイプシステム実装を実施した。

1. 抽出対象パラメータの検討
2. 製薬・医療関係文書に関する情報取得及び概要の調査
3. 製薬・医療関係文書からの情報抽出手法の検討とプロトタイプ作成
  - ① 構造化文書からの情報抽出手法の検討
  - ② 非構造化文書からの情報抽出手法の検討
  - ③ マニュアルキュレーションによる確認・修正を含むシステム設計

2018 年度までに実施した調査結果およびプロトタイプ実装を用い、情報抽出システムの本実装及び情報抽出を開始した。また、薬物動態・毒性パラメータ以外の情報を高精度に抽出するため、学習用データセットを作成するためのアノテーション、自然言語処理・機械学習手法を用いる情報抽出・構造化手法の構築を開始した。抽出した情報は汎用的に使用できる形式へ変換し、データベースへ格納する。

In order to expand basic database by extracting pharmacokinetic and toxicity parameters from pharmaceutical and medical documents such as CTD, we conducted the following research items.

1. Investigation of parameters to be extracted.
2. Acquisition and summary survey on pharmaceutical and medical documents
3. Investigation and prototype construction of natural language processing (NLP) method from pharmaceutical and medical document
  - ① NLP methods for extracting parameter values from structured documents such as XMLs.
  - ② NLP methods for extracting parameter values from unstructured documents such as PDFs.
  - ③ System and UI design including checking and correcting extracted parameter

values.

Based on the above survey conducted in FY2018, we started implementation of information extraction system and trial of information extraction. Furthermore, in order to extract information other than pharmacokinetics / toxicity parameters with high accuracy, we started development of information extraction method using NLP based on machine learning methods, including data annotation framework for creating training dataset for machine learning method. The extracted information will be converted to a universally format and stored in a database.

- 産学連携や研究成果の社会実装
  1. DruMAP (Drug Metabolism and pharmacokinetics Analysis Platform) 、有、URL: <https://drumap.nibiohn.go.jp/>
  2. 産学連携としては、国内製薬企業複数社が保有する薬物動態関連データを収集しデータベースを拡充した。この企業連携により取得したデータを用いてより高精度の薬物動態予測モデルの構築が可能となる。さらに、構築した予測モデル及びデータベースを、維持・発展が可能な状態で創薬支援に利用できるようにすることを目的として、IT 企業と共同でエコシステムの構築を行っている。

### III 評価結果

企業との連携が適切に行われ、さらにデータの収集と最新の統計解析技術の適用により、現時点において得られる最高レベルの予測モデルの構築が達成されようとしており、当初の計画に沿って順調に計画が進捗していることを評価する。一方、成果として得られた予測モデルが既存のモデルと比べて格段の進歩を示したとは判断できず、更なる進歩を期待したい。

生理学的薬物動態モデル (PBPK) 研究が進展している中、*in vitro* から *in vivo* 予測への移行、低分子から中分子への移行等研究のトレンドを先取りして研究を発展させて行っていただきたい。

### IV 統括評価結果

本創薬支援推進事業の薬物動態・心毒性・肝毒性に関するインフォマティクス構築の3プロジェクトを取りまとめ、各システムの構築を予定通りに進め、成果物を得たことを評価する。

今後は本研究の公開（無償あるいは商用版）により本邦の創薬研究に資することを期待する。

日本医療研究開発機構  
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」  
事後評価報告書

## I 基本情報

研究課題名：（日本語）インフォマティクスとシミュレーションを融合した多面的心毒性予測システムの構築

（英語）Construction of a multifaceted cardiotoxicity prediction system that combines informatics and simulation

研究実施期間：2015年10月1日～2020年3月31日(予定)

研究代表者 氏名：（日本語）本間 光貴

（英語）Teruki Honma

研究代表者 所属機関・部署・役職：

（日本語）国立研究開発法人理化学研究所 生命機能科学研究センター チームリーダー

（英語）Team Leader, Center for Biosystems Dynamics Research, RIKEN

## II 研究の概要

### 1. 研究概要

本研究課題では、心毒性の主要な原因タンパク質である hERG, Nav1.5, Kv1.5, Cav1.2 等のイオンチャンネルや iPS 由来心筋細胞のデータに基づいて最先端のインフォマティクスとシミュレーション的な手法を融合して多面的に予測することを目的としている。具体的な目標としては、上記の測定データについて慎重なキュレーションと統計処理を行った後に心毒性データベースとして統合し公開すること、心毒性指標（hERG, Nav1.5, Kv1.5, Cav1.2）の活性判別及び回帰予測モデルの構築と一部の公開、hERG モデルのシミュレーションの検討と計算プロトコールの公開の3点となる。

心毒性データベースに関しては、論文公知のデータに加えて、欠けているデータを補うために 3000 化合物以上の新規な hERG, Nav1.5, Kv1.5, Cav1.2 実測データを内部及び外注のアッセイにより取得し公開する。予測モデルの構築については、深層学習の一種であるマルチタスク学習・事前学習を含む最新の統計手法に基づいて機械学習予測モデルを構築する。また、予測値の提供だけでなく、創薬研究者が好む新規構造を発生することができるプログラムと連携させて hERG 回避可能な合成候補構造を提案できる、より実用的な手順も開発する。さらに、hERG の電子顕微鏡構造とそれに基づく分子動力学シミュレーションの結果からの hERG 構造アンサンブルを作成し、それらを用いた妥当なドッキング手順を開発し、上記の構造提案機能に融合する。これらの目標を達成しつつ、理化学研究所の創薬・医療技術基盤プログラムや AMED 創薬支援ネットワークのターゲットに適

用し、開発した心毒性予測プラットフォームを実際の創薬に応用する。

ヒト iPS 由来心筋細胞を用いた心毒性評価系については、マルチウエル型多点電極記録法による細胞外電位測定、無染色標本での収縮波形の測定、膜電位感受性色素による活動電位波形の測定を行って、心毒性を多角的に評価した。これらの方法を用いて、創薬の初期段階で高いスループットで候補化合物の心毒性を評価できる系を構築し、予備的な評価を行い、今後の普及を目指す。

## 2. 研究成果

### 2.1 心毒性データベースの構築と公開

文献公知データ及び本課題で新規に測定したデータの品質チェック等のキュレーション、複数データの統計処理による統合を行うことで、30 万件程度の世界最大規模の心毒性データベースを構築した。データベースには、hERG の他、Nav1.5, Kv1.5, Cav1.2 等のイオンチャンネルデータを含む。また、新規に公開された差分データを半自動的に更新するシステムも整備した。構築したデータベースについて Web ページ (<https://drugdesign.riken.jp/hERGdb/>) を介して公開し、後述の hERG 予測モデルも利用可能である。

### 2.2 多面的な心毒性予測システムの構築

心毒性データベースのデータに基づき、サポートベクターマシンと非優越ソート遺伝的アルゴリズムによる変数最適化を用いて、hERG 判別予測モデルを構築した。外部テスト用データセットを用いて、本予測モデルを現在市販されている他の心毒性予測プログラムと比較した結果、それらを大きく上回る予測性能（正答率：0.984, kappa 統計量：0.733, ROC(AUC): 0.962）を示すことに成功した。同様の構築手順によって、フリーの分子記述子を用いたモデルも構築し、上記の URL で公開した。hERG 回帰モデルについては、全データを用いた多様な化合物に対応できるモデル（グローバルモデル）が、 $R^2$ : 0.602、一部の骨格に特化したモデル（ローカルモデル）が  $R^2$ : 0.917 を示した。

また、製薬企業 7 社との hERG データの共有と予測モデルの構築・提供の連携の枠組みが始まり、提供されるデータに基づいて汎用性が高く、特に hERG に結合しやすい骨格を有する化合物に対する予測性能を向上させた予測モデルの構築を試みた。その結果、1 社のみのデータに基づいて予測モデルを構築し、7 社のデータのテストセットで検証すると ROC が 0.597 であるのに対し、7 社のデータを用いて学習させることによって大きく性能が向上し、ROC が 0.800 を示すモデルを得ることに成功した。製薬企業には、上記のフリーの分子記述子版のグローバルモデル予測モデルと 7 社のデータによる予測モデルを提供した。

Nav1.5, Kv1.5, Cav1.2 の判別予測モデルについては、深層学習のマルチタスク学習も併用し、判別モデルにおいては、Kappa 統計量で 0.5 から 0.6 と、少ないデータ数（数百か

ら 1000 程度) にも関わらず、良好な予測精度のモデルを構築することができた。また、事前学習によって、Nav1.5 について、 $R^2: 0.613$  の回帰モデルも得ている。

hERG 回避のための構造提案機能については、創薬化学の論文掲載の構造展開事例をほぼ 100%再現することができる RLS 法を開発し、上記の hERG 判別モデルと組み合わせて、理化学研究所の創薬ターゲットに適用した。その結果、すでに終了したターゲットの検証で良好な結果を示した他、進行中のターゲットにおいて、実際に薬効を保った状態での hERG 回避に成功し、特許取得の原動力となった。

### 2.3 hERG 電子顕微鏡構造に基づくシミュレーション結果を利用した予測手法の開発

hERG の電子顕微鏡構造に基づいて分子動力学シミュレーションを実施し、シミュレーション結果とその後のドッキング結果から主要な hERG 阻害剤の変異実験を満たす hERG 構造を選択した。得られた hERG 構造アンサンブルに基づいて、ドッキングと相互作用記述子を用いた予測モデルを構築し、特定の hERG に結合しやすく予測が難しい骨格について、上記 2.2 で構築したドッキング結果を使わない hERG 判別予測モデル（この骨格に限定すると ROC: 0.869）を上回る予測性能（ROC: 0.949）を示すことを確認し、このような難しい予測対象の場合に、ドッキングを併用することの有用性を示した。また、ドッキングモデルを用いて、hERG 回避のための設計のアイデアを得ることも可能である。

### 2.4 化合物の保管・管理及び心毒性関連イオンチャンネルの試験の実施

化合物の保管管理システムとして、化合物の受け入れ、秤量、溶液化、純度検定を実施できる体制を整備した。心毒性測定用に 4000 化合物以上を購入して、化合物の溶液サンプルを調製した。外注用に 800 化合物以上の粉体試料を出庫し、in house 用のべ 1200 化合物以上の溶液試料を出庫した。心毒性実測データの取得については、外注による hERG binding、hERG autpatch の試験について、のべ 1300 化合物以上の測定を実施した。内部での hERG autpatch clamp 試験については、1500 化合物以上の測定を実施した。また、2017 年度からの計画変更に伴い、Nav1.5, Kv1.5, Cav1.2 の試験について、それぞれ 200 化合物、30 化合物、30 化合物に対して実施した。

### 2.5 iPS 由来心筋細胞を用いた創薬の早期に利用できるスクリーニング手法の開発

ヒト iPS 細胞由来心筋細胞を用いて、多点電極システムによる細胞外電位記録、無染色での心筋細胞の収縮波形記録と膜電位感受性色素を用いた活動電位のイメージングシステムを用いて、簡便かつハイスループットな心毒性リスク評価系を構築し、代表的な心毒性化合物、hERG 以外の要因で心毒性が出ている化合物について 30 化合物以上の評価を実施した。これらの方法を組み合わせることで、化合物の複数チャンネルへの結合や、キネティックスの遅い結合や、収縮力や興奮伝搬の抑制作用などを複合的に評価できる。化合物の hERG 阻害作用の IC<sub>50</sub> と各種の心毒性イベントの誘発濃度を比較することは、ヒト臨床試験にお

ける心毒性予測において重要な情報を提供するであろう。

## 2.1 Construction and release of cardiotoxicity database

The world's largest cardiotoxicity database consisting of more than 300,000 compounds was constructed by curating publicly known data and newly measured data, and integrating by statistical processing of multiple data. The database includes cardiotoxicity-related ion channel data such as hERG, Nav1.5, Kv1.5, and Cav1.2. In addition, a system for semi-automatically updating newly released data has been established. The constructed database was released via the web page (<https://drugdesign.riken.jp/hERGdb/>), and the hERG prediction model described later can also be used.

## 2.2 Construction of a multifaceted cardiotoxicity prediction system

Based on the data of the cardiotoxicity database, a hERG discrimination model was constructed using feature selection with support vector machine and NSGA-II. Using the external test set, this model showed significantly higher prediction performance (accuracy: 0.984, kappa statistic: 0.733, ROC (AUC): 0.962) compared with commercially available models. Another model using free descriptors was also constructed by the same procedure and published at the above URL. Regarding the hERG regression model, a model (global model) that can deal with various compounds using all data showed  $R^2$ : 0.602, and a model specialized for some skeletons (local model) recorded  $R^2$ : 0.917.

In addition, a new framework for sharing hERG data with seven pharmaceutical companies and the construction and provision of prediction models has begun. Based on the data provided, it is expected to enhance the prediction performance for compounds with common scaffolds to strongly bind to hERG. As a result, when a prediction model is built based on data from only one company and validated with the test set of data from seven companies, the ROC is 0.597, while learning by using all the data from seven companies greatly improves prediction performance (ROC: 0.800). The above-mentioned global prediction model of free molecular descriptor version and the prediction model based on data from seven companies were provided to the pharmaceutical companies.

For the Nav1.5, Kv1.5, and Cav1.2 discrimination models, deep multi-task learning based on deep learning was used, the kappa statistic values showed

0.5 to 0.6 in all the models, in spite of the small number of data (200-1000 compounds). In addition, a regression model of  $R^2$ : 0.613 was obtained for Nav1.5 by the pre-training technology.

With regard to the structure proposal tool to release hERG inhibition, we developed RLS method that can reproduce almost 100% of the structural analogues included in several medicinal chemistry articles. We examined the combination of the RLS method and the hERG discrimination model described above using real drug targets. As a result, in addition to showing good results in the verification of the already completed target, it succeeded in releasing hERG affinity while maintaining the cell activity in the ongoing target, and became the driving force for patent application.

### 2.3 Development of design methods using simulations based on the hERG EM structure

Based on the electron microscope (EM) structure of hERG, molecular dynamics simulation was performed, and hERG structures satisfying the representative hERG inhibitors with mutation experiments were selected from the simulation results and subsequent docking results. Based on the obtained hERG structural ensemble, a hERG discrimination model using docking and interaction descriptors was constructed, and compared with already mentioned hERG discrimination model in 2.2 without docking results using the particular chemical series. It was confirmed that the docking-based model (ROC: 0.949) exceeds the model without docking results (ROC: 0.869 when limited to this chemical series), and the usefulness of docking information was demonstrated. It is also possible to obtain design ideas for avoiding hERG affinity based on the docking model.

### 2.4 Compound management and cardiotoxicity related ion channel testing

As a storage management system for compounds, we have established a system that can store, weigh, dissolve compounds and measure the purity. More than 4000 compounds were purchased for cardiotoxicity measurements and compound solution samples were prepared. Powder samples of more than 800 compounds were delivered for outsourcing, and solution samples of more than 1200 compounds were delivered for in-house use. Regarding new cardiotoxicity data, a total of more than 1300 compounds were assayed for hERG binding and hERG autopatch by outsourcing. For the in-house assay

for hERG autopatch clamp, more than 1500 compounds were tested. In addition, due to changes in the plan from FY2017, Nav1.5, Kv1.5, and Cav1.2 tests were conducted on 200 compounds, 30 compounds, and 30 compounds, respectively.

## 2.5 Development of screening methods for cardiac toxicities at the early stage of the drug discovery using human iPSC-derived cardiomyocytes

We evaluated more than 30 compounds including the CiPA-compound set and other representative cardiotoxicity compounds with multiple ion channel data, using a multiple-electrode array system for measuring field potentials, a newly-developed high-throughput image method for detecting contraction of non-stained cardiomyocytes, and a voltage sensitive dye imaging method to analyze ion channels components of action potential wave forms. We have shown the combination of these methods are strong tools to identify cardiac toxicity of drugs which have additional effects on hERG inhibition, such as multi-ion channel inhibitions, slow-binding kinetics, inhibition of contraction and conduction velocity. Comparison of hERG IC50 and the concentrations on which drugs induce toxic events would provide important information to predict adverse effects in clinical trials.

- 産学連携や研究成果の社会実装
  3. AMED Cardiotoxicity Database、有、<http://drugdesign.riken.jp/hERGdb/>
  4. 公開版 hERG 予測モデル、有、<http://drugdesign.riken.jp/hERGdb/>の hERG prediction タブ

## III 評価結果

研究代表者を中心とした研究体制が適切に組織されており、構造発生手法（RLS 法）の開発により、企業のメディシナルケミストが期待するレベルの予測と推奨を実証している点を高く評価する。

総じて、初期目標をよく達成しており、国際的な競争力を持つ正確なモデルを構築できた。創薬支援となる心毒性予測システムに関する新技術が開発されており、優れた成果が得られていると評価できる。

今後、開発された心毒性予測システムがアカデミア創薬のみならず製薬企業などでも実活用され、有効性がさらに確認されることを期待したい。

日本医療研究開発機構  
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」  
事後評価報告書

## I 基本情報

研究課題名：（日本語）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  
（英 語）Development of a Liver Toxicity Informatics System

研究実施期間：2015年9月28日～2020年3月31日

研究代表者 氏名：（日本語）山田 弘  
（英 語）Hiroshi Yamada 研究代表者 所属機関・部署・役職：  
（日本語）国立研究開発法人 医薬基盤・健康・栄養研究所  
医薬基盤研究所 トキシコゲノミクス・インフォマティクスプロジェクト  
プロジェクトリーダー  
（英 語）Project Leader  
Toxicogenomics Informatics Project  
National Institutes of Biomedical Innovation  
National Institutes of Biomedical Innovation, Health and Nutrition

## II 研究の概要

研究の成果およびその意義等

肝毒性は、市場撤退および臨床開発中止の主要因にも関わらず毒性発現予測および評価の確立が困難とされている。その理由により医薬品の効率的な開発および安全性確保のため、特に創薬早期において利用可能な *in silico* および *in vitro* 肝毒性評価技術の発展が強く望まれている。本研究では、アカデミア創薬支援を目的として、肝毒性に係る情報を提供するデータベースおよび肝毒性を予測するインフォマティクスシステムの開発を目標とした。

研究は、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、学校法人 明治薬科大学および国立大学法人 熊本大学が連携して推進した。

肝毒性の発症メカニズムは複雑で多岐に渡るため、医薬品候補化合物の肝毒性評価には多くの情報が必要であり、毒性予測は様々な視点から多面的に行う必要がある。そのため、本研究では様々な情報を提供するデータベースを構築すると共に、予測システムについては化学構造情報およびヒト肝細胞から取得した遺伝子発現情報から肝毒性を予測するシステムの開発を行った。後者の予測システムについては、肝毒性フェノタイプ別に予測するシステムと既知肝毒性化合物の遺伝子発現パターンとの類似性を予測するシステムの開発を目指した。加えて、単に予測結果を提示するだけではなく、結果の解釈を支援する毒性作用

機序解釈支援のための知識システムの開発も並行して行った。当知識システムの開発には、肝毒性研究分野において未だ世界的にも報告事例が無い毒性プロセスオントロジー研究を取り入れていることに大きな独創性がある。

研究の結果、以下に示す肝毒性関連のデータベース 3 種、ヒト肝細胞から取得した遺伝子発現情報から肝毒性を予測するシステム 2 種、化学構造情報から肝毒性と間質性肺炎を予測するシステム各 1 種の開発を達成した。このように多様な観点からの安全性への対処を示す肝毒性データベースおよび肝毒性予測システム群は世界に類を見ないものであり、創薬探索段階の早期から後期までの幅広い段階での安全性研究、さらには低分子医薬品のみならず、近年開発が盛んな高分子および中分子医薬品までもカバーした安全性研究に寄与すると考えられる。

#### 【肝毒性研究班ポータルサイト】

名 称： DILI-TOOLBOX

URL： <https://dili-toolbox.nibiohn.go.jp/>

概 要： 研究班で開発した全てのデータベースおよび予測システムの紹介を行うと共に、各システムへのアクセス環境を提供するポータルサイトである。

#### 【データベース 1】

名 称： DILI-cSEARCH (Drug-Induced Liver Injury information Cross-Search system)

開発機関： 医薬基盤・健康・栄養研究所、熊本大学

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

概 要： 肝毒性に関連する様々な情報を提供するデータベースである。情報は肝毒性関連データベース (Open TG-GATEs, DrugMatrix database, Library of Integrated Network-Based Cellular Signatures (LINCS)、Liver Toxicity Knowledge Base (LTKB)) から収集し、整理・統合して DILI-cSEARCH に登録した。本データベースは①各種大規模肝毒性関連データベースから収集した様々な情報が一元的に閲覧できる、②化合物、遺伝子発現データ、毒性データ (血液学検査、血液化学検査、病理検査等)、実験系 (in vivo、in vitro)、種 (ヒト、ラット)、暴露期間、暴露量など様々な項目について条件設定 (例えば、ある検査値が設定値を超えるもの) しながら複数選択し、絞り込み検索が行える機能を有している、③ ①②の情報および機能を活用することにより、in vitro (ラット、ヒト) —動物 (ラット) —ヒト、あるいは分子—細胞—臓器—個体など様々な視点から横断的および多面的に情報の閲覧が可能である、などの特徴を有している。

研究班では、データベース構築と並行して、体質性肝障害の患者由来 iPS 細胞から肝細胞に誘導する方法を確立し、当肝細胞から種々の情報を収集する研究を実施した。加えて、ラット肝細胞およびヒト肝細胞における肝毒性化合物暴露に対する miRNA の応答性に関する

研究も行った。これらの研究により得られた研究成果の概要も当データベースに掲載している。

#### 【データベース 2】

名 称： TOXPILOT (TOXic Process InterpretabLe OnTology)

開発機関： 医薬基盤・健康・栄養研究所

URL： <https://toxipilot.nibiohn.go.jp/>

概 要： 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。システムの提供する主な機能としてまず可視化機能が挙げられる。複雑な毒性作用機序の解釈を支援するためには、機序を可視化することは重要となる。そこで、肝毒性機序を体系化した毒性プロセスオントロジー (Toxic Process Ontology, TXPO) を構築し、その情報を核として RDF 化した概念間の各種関係を柔軟に可視化し、システム内でグラフを表示させる機能 (マップ表示機能) を実装した。ノードは作用機序テーマ毎に、毒性作用機序を構成するプロセスとその因果関係、プロセスと部分プロセス (全体一部分関係) の他、プロセスに関わる分子とその関係を表示できる他、関連する所見についても可視化できるよう実装している。オントロジーでは、機序を構成するプロセスの親概念は複数の機序で共通するプロセスとなることから、これらを計算機処理することで複数の機序を横断した汎用的なプロセス間の因果関係 (汎用機序) の可視化を実現した。汎用機序マップでは複数機序の共通プロセスを明示できるとともに、テーマ固有の原因も色毎に識別しながら情報を提供することから、機序の進行過程において複数の機序にまたがる重要な因子や従来把握できていなかったリスクの発見に貢献すると考えている。

#### 【データベース 3】

名 称： toxBridge

開発機関： 産業技術総合研究所

URL： <https://toxbridge.molprof.jp/top>

概 要： 肝毒性 in vivo/vitro ブリッジングに着眼したデータベースである。本研究では、公開データベース Open TG-GATEs に登録されている約 170 の化合物によるラット in vivo (肝臓) /vitro (肝細胞) 及びヒト in vitro (肝細胞) の遺伝子発現データの大規模数理解析を行った。肝毒性を予測するためのマーカー群としてのパネル構築を目的として、化合物種、暴露期間、暴露量ごとに遺伝子刻印およびそのパスウェイを推定し、また 4 種の肝毒性に関する生理的指標についても整理を行い、体系的にデータベース (toxBridge) に収納

した。toxBridgeの特徴としては、in vivo - in vitro や in vitro ヒト-ラット間のブリッジングに着眼し、発現プロファイルのブリッジ間における様々な相関指標をデータベースに格納しており、各種相関指標による化合物検索や遺伝子・パスウェイを検索する機能を有している点である。この機能により、相関指標を用いてどの化合物がブリッジングとして有意であるか、また相関指標が優位である化合物を選択することでブリッジングにおける分子群を選択し可視化可能としている。加えて既存情報である FDA が公開している DILI (Drug Induced Liver Injury) Rank を用い化合物を選択し、相関指標を用いることで肝毒性におけるマーカー群となりうる分子パネルの選定を可能としている。

#### 【予測システム1】

名 称： DILI-PANEL

開発機関： 医薬基盤・健康・栄養研究所

URL： <https://dili-panel.nibiohn.go.jp/>

概 要： in vitro (ヒト肝細胞) 遺伝子発現データから肝毒性マーカーパネル (ヒト肝細胞評価系での肝毒性マーカー群) を基に肝毒性フェノタイプ別 (リン脂質症など) に予測を行うデータ駆動型 (機械学習) 肝毒性予測機能に加え、予測結果に対応するオントロジーに基づいた毒性機序マップ提供機能を兼ね備えたシステムである。ユーザーは、WEB 上でマイクロアレイデータを DILI-PANEL にアップロードすることで予測結果を得ることができる。判定結果を文章で表示するとともに、解析結果を4種のタイル (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毒性機序マップ: 想定される毒性機序と機序解釈支援知識システム TOXPILOT に登録された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) で構成されるダッシュボード形式で表示する。各タイルをクリックすることにより、詳細な情報にアクセスすることができる。

#### 【予測システム2】

名 称： toxRank

開発機関： 産業技術総合研究所

URL： <https://toxrank.molprof.jp/>

概 要： toxRank は、遺伝子発現情報に基づき類似肝毒性化合物を検索する WEB サービスである。本プロジェクトにおいて計測された全遺伝子発現データからランクマトリクスを構築し、化合物投与時における変動遺伝子 (up と down 分子刻印) を入力として、遺伝子発現プロファイルを基に類似性検索を行い、類似化合物と有意遺伝子・パスウェイを連携させた。toxRank では、ランクマトリクスを参照することで、類似性を評価してその結果を表示させているが、この参照ランクマトリクスを動的に生成可能とする WEB 開発を行った。よって、利用者が肝毒性に重要と考える遺伝子や toxBridge で得られた分子パネルか

ら可変ランクマトリクスを生成することが可能であり、分子パネルによる類似性検索や DILI 等の情報を付加したクラスタリングおよびネットワーク解析による肝毒性予測に繋がっている。加えて、発現データの変動幅を考慮し、遺伝子変動を粗視化したパスウェイマトリクスによる類似性検索機能の実装も行っている。

#### 【予測システム 3】

名 称： LIVER/MIE-QSAR

開発機関： 明治薬科大学

URL： [http://mmi-03.my-pharm.ac.jp/tox1/prediction\\_groups/new](http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new)

概 要： 化学構造情報から肝毒性を予測するシステムである。肝毒性の *in silico* 予測を達成するため、ヒトにおける肝毒性の指標として大規模副作用データベース FAERS (JAPIC-AERS) を活用するとともに、Tox21 プロジェクトで公開されている有害性発現経路(Adverse Outcome Pathway; AOP)における核内受容体等の分子起始反応 (Molecular Initiating Event; MIE) 関連活性を利用した QSAR 予測モデルの構築を、以下に述べる戦略に基づいて実施した。本研究は①AOP の初期段階である種々核内受容体・ストレス応答パスウェイを含む MIE を化学構造情報から予測する QSAR モデルの構築、②副作用情報データベースに含まれる肝毒性関連副作用と MIE の関係の網羅的解析と統計モデルの構築、の二段階で構成される。先ず Tox21-MIE データベースと副作用データベースを統合した解析用データベースを構築した後、本データベースに基づき MIE と肝毒性の関係を解析するとともに、MIE 活性に基づく肝毒性予測モデルを構築した。上記の構造-MIE 活性相関予測モデルと MIE-肝毒性相関予測モデルを組み合わせることにより、化学構造から肝毒性を予測するシステムの構築を達成した。以上の肝毒性予測戦略において推算される各 MIE 予測値は、毒性学的に重要な記述子として機能するため、化学構造のみから直接肝毒性を予測するモデルに比して堅牢な予測結果が期待できる。

#### 【予測システム 4】

名 称： LUNG/MIE-QSAR

開発機関： 明治薬科大学

URL： [http://mmi-03.my-pharm.ac.jp/tox1/prediction\\_groups/new](http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new)

概 要： 化学構造情報から医薬品に起因した間質性肺炎を予測するシステムである。本研究は、LIVER/MIE-QSAR の構築で用いた手法の汎用性を検証するために実施したものであり、そのため LUNG/MIE-QSAR と、同様の戦略に基づいてシステム構築を行った。

(英文)

In order to contribute to the improvement of the ability related to safety evaluation in an academic drug discovery, the five-yearly program for “Development of a Drug Discovery Informatics System” supported by Japan Agency for Medical Research and Development (AMED) was started in 2015. In this program, a collaborative project of the National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), National Institute of Advanced Industrial Science and Technology (AIST), Meiji Pharmaceutical University and Kumamoto University for “Development of a Liver Toxicity Informatics System” was organized. The project established 3 databases related to liver toxicity (DILI-cSEARCH, TOXPILOT, toxBridge) and 4 toxicity prediction systems (DILI-PANEL, toxRank, LIVER/MIE-QSAR, LUNG/MIE-QSAR). All databases and prediction systems can be accessed via portal site of the project (DILI-TOOLBOX, <https://dili-toolbox.nibiohn.go.jp/>). Details of the databases and prediction systems developed by the project are as follows.

[Database 1]

Name	DILI-cSEARCH (Drug-Induced Liver Injury information Cross-SEARCH system)
Developer	NIBIOHN and Kumamoto University
URL	<a href="https://dili-csearch.nibiohn.go.jp/">https://dili-csearch.nibiohn.go.jp/</a>
Description	DILI-cSEARCH is a freely available database that provides comprehensive information about liver toxicity caused by drugs and chemicals. This database offers cross search access to information of Open TG-GATEs, DrugMatrix, LINCS and LTKB (Liver Toxicity Knowledge Base) including in vitro/in vivo experimental conditions, toxicological profiles, gene expression profiles, and much more. Available data may be searched in various ways.

[Database 2]

Name	TOXPILOT (TOXic Process InterpretabLe OnTology)
Developer	NIBIOHN
URL	<a href="https://toxipilot.nibiohn.go.jp/">https://toxipilot.nibiohn.go.jp/</a>
Description	TOXPILOT is a toxic process interpretable knowledge system. Focusing on hepatotoxicity, TOXPILOT provides visualization maps of the toxic course, general course map that visualizes general toxic courses common to multiple specific toxic courses, and systematized knowledge for hepatotoxic processes based on Toxic Process Ontology, TXPO.

## [Database 3]

Name	toxBridge
Developer	AIST
URL	<a href="https://toxbridge.molprof.jp/top">https://toxbridge.molprof.jp/top</a>
Description	toxBridge – a statistical interpretation of building relationship between in vivo and in vitro for the purpose of constructing a panel as a marker group for predicting hepatotoxicity. Based on the rat in vivo and human in vitro gene expression data, the significance of gene signatures and pathways were analyzed for about 170 medical compounds at four dose levels and time points. The characteristics of toxBridge are focused on in vivo-in vitro and in vitro human-rat bridging, and various correlation indices between the bridges in the expression profile are stored in the database. It has the function of searching for compounds and searching for genes and pathways. This function allows the selection of which compounds are significant as bridging using correlation indices and creates a molecular panel in bridging by selecting a compound having a superior correlation index.

## [Prediction system 1]

Name	DILI-PANEL
Developer	NIBIOHN
URL	<a href="https://dili-panel.nibiohn.go.jp/">https://dili-panel.nibiohn.go.jp/</a>
Description	DILI-PANEL is a prediction system of liver toxicity that was established to support the safety assessment and risk management for drug candidates. This system can automatically derive toxicological alerts from gene expression data of in vitro human hepatocytes using marker panel of liver toxicity. The marker panel of liver toxicity was established using the gene expression data of human primary hepatocytes in DILI-cSEARCH and machine learning methods. The central features of the DILI-PANEL are: <ol style="list-style-type: none"> <li>1. Prediction of potential risk of liver toxicity.</li> <li>2. Identification of potential mechanism of liver toxicity using liver toxicity mechanism maps.</li> </ol>

## [Prediction system 2]

Name	toxRank
Developer	AIST
URL	<a href="https://toxrank.molprof.jp/">https://toxrank.molprof.jp/</a>
Description	toxRank – rank analyses of drugs using rat in vivo/vitro and human in vitro gene expression data. Given a signature of query drug, toxRank computes

	the similarity with reference to expression profiles which are developed rank matrix data. It is possible to generate flexible rank matrix from genes that users consider important for hepatotoxicity and molecular panels obtained with toxBridge.
--	--

[Prediction system 3]

Name	LIVER/MIE-QSAR
Developer	Meiji Pharmaceutical University
URL	<a href="http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new">http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new</a>
Description	As part of QSAR, LIVER/MIE-QSAR system has served as predictor for 59 types of activities of molecular initiating events (MIE). Within the system, MIE activities such as nuclear receptors and stress response pathways are predictable through machine learning. The prediction effectively deals with the toxicity of drug-induced liver injury (DILI) based on the computation of their chemical structure. Consequently, the system can be used to target drugs of low molecular weight analogues and predicts their induction of cytotoxic liver injury, cholestatic liver injury, and liver cancer. On the other hand, the prediction of MIE can be generally applied in targeting a wider range of compounds of low molecular weight including pharmaceuticals and chemicals.

[Prediction system 4]

Name	LUNG/MIE-QSAR
Developer	Meiji Pharmaceutical University
URL	<a href="http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new">http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new</a>
Description	Based on the same method and considering hepatotoxicity, a LUNG/MIE-QSAR system for predicting interstitial lung disease was developed.

● 産学連携や研究成果の社会実装

1. ポータルサイト (DILI-TOOLBOX) : 研究班が開発した全システムの情報とアクセス環境を提供、有、URL : <https://dili-toolbox.nibiohn.go.jp/>
2. データベース (DILI-cSEARCH) : 医薬品および化学物質によって引き起こされる肝毒性に関する様々な情報を提供、有、URL : <https://dili-csearch.nibiohn.go.jp/>
3. データベース (TOXPILLOT) : 毒性作用機序解釈支援のための知識システム、有、URL : <https://toxipilot.nibiohn.go.jp/>
4. データベース (toxBridge) : 化合物を暴露した際の遺伝子発現情報をもとに、ラット間 (in vitro と in vivo) やヒト (in vitro) とラット間における橋渡しとなりうる刻印遺伝子やパスウェイを提示、有、URL: <https://toxbridge.molprof.jp/top>
5. 肝毒性予測システム (DILI-PANAL) : 肝毒性マーカーパネルを用いて、in vitro ヒト肝細胞の遺伝子発現データから肝毒性を予測、有、URL : <https://dili-panel.nibiohn.go.jp/>
6. 肝毒性予測システム (toxRank) : ユーザーが入力した遺伝子群の発現パターンに類似の発現プロファイルを持つ肝毒性化合物一覧を表示、有、URL : <https://toxrank.molprof.jp/>
7. 肝毒性予測システム (LIVER/MIE-QSAR) : 化学構造から肝毒性を予測する QSAR システム、有、URL : [http://mmi-03.my-pharm.ac.jp/tox1/prediction\\_groups/new](http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new)
8. 肺毒性予測システム (LUNG/MIE-QSAR) : 化学構造から間質性肺炎を予測する QSAR システム、有、URL : [http://mmi-03.my-pharm.ac.jp/tox1/prediction\\_groups/new](http://mmi-03.my-pharm.ac.jp/tox1/prediction_groups/new)

### III 評価結果

中間審査内容に柔軟に対応し、合理的で良い方向へ研究を導いたことを評価する。また、多様な角度から 3 種の DB、3 種の (肺毒性を入れれば 4 種の) 予測システムを構築し、一定の成果を挙げた。なお、他のプログラムとは異なり、本件 (特に、DILI-PANEL と ToxRank) は、純粋な in silico 予測モデルではないが、in vitro 遺伝子発現プロファイルからの in vivo 肝毒性予測技術は、創薬現場におけるニーズが高く、また比較的少数のデータでも実用レベルの精度が期待されることから、本研究成果の創薬研究への寄与は大きい。また、低分子創薬を超えた新しいモダリティによる創薬にも資する可能性がある点も評価する。

肝毒性関連データベースの開発及び肝毒性予測システムの開発、並びにそれらの公開について完了されており、これらの成果は、国際的にも先導的なものであることを高く評価する。