

## 科学研究費助成事業 研究成果報告書

平成 27 年 5 月 31 日現在

機関番号：14301

研究種目：挑戦的萌芽研究

研究期間：2013～2014

課題番号：25670073

研究課題名(和文) 進化的アルゴリズムによるPopPKモデリングの完全自動化

研究課題名(英文) Full-automatic PopPK modeling utilizing an evolutionary algorithm

## 研究代表者

山下 富義 (Yamashita, Fumiyoshi)

京都大学・薬学研究科(研究院)・准教授

研究者番号：30243041

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：ポピュレーション・ファーマコキネティック・モデルは、患者情報に基づく平均的な薬物動態パラメーター、個体間変動および個体内変動からなるモデルである。患者個人個人の血中薬物動態を予測するためには、より精度の高い患者情報と薬物動態パラメーターとの間の関係すなわち共変量モデルを見つけることが重要である。しかしながら、共変量候補の数が多いために、モデル開発は困難を極める。そこで本研究では、コンピューター上で数式を進化させることにより、臨床データから完全自動で共変量モデルを開発するシステムを完成させた。

研究成果の概要(英文)：A population pharmacokinetic model is composed of average pharmacokinetic parameters, their inter- and intra-individual variabilities. In order to predict blood concentration profile of a drug in each individual, we have to develop an accurate covariate model that represents quantitative relationship between pharmacokinetic parameters and patient information. However, it is a labor-intensive process because there are too many possibilities comprising covariate candidates. In the present study a full-automatic system for covariate modeling was developed, that utilizes an algorithm of evolving mathematical formulae on computers.

研究分野：薬物動態学

キーワード：母集団薬物動態解析 共変量モデル 遺伝的アルゴリズム 遺伝的数式プログラミング 個体間変動  
拡張最小二乗法

## 1. 研究開始当初の背景

ポピュレーション・ファーマコキネティクス・モデル (PopPK モデル) は、患者集団における平均薬物動態パラメーターと生理的変量との関係、および個体間・個体内変動を考慮した非線形混合効果モデルである。すなわち、クリアランスや分布容積といった薬物動態パラメーターに対して、その支配要因となる共変量を見出し、個体間変動に対する固定効果の影響を定量的に表現する。ただし、経験論的モデルである以上、共変量候補の数とともに可能なモデルの数が爆発的に増え、いわゆる NP 困難問題となってしまう。とりわけ、近年では患者背景に基づく個別化医療の重要性が叫ばれ、バイオマーカーによる層別化試験も実施される中、モデル構築のための共変量候補の数が爆発的に増えている。こうした現状において、手作業による共変量モデルの構築はかなり限られた範囲内でしか探索できず、合理的な方法の開発が強く望まれている。

研究代表者は、これまでにケモインフォマティクス技術を駆使した薬物動態の *in silico* 予測に関する研究を実践してきた。特にニューラルネットワークや遺伝的アルゴリズムといった機械学習法をいち早く当該分野に取り入れ、多様な化学構造を有する薬物の水への溶解性、Caco-2 細胞膜透過性、皮膚透過性などを精度よく予測できることを報告した。とりわけ、生物進化の自然淘汰をモデル化した遺伝的アルゴリズムは、構造活性相関における説明変数の探索にも有効であり、組み合わせ最適化問題への応用に強力なツールとなることを示した。

## 2. 研究の目的

本研究では、ポピュレーション・ファーマコキネティクス解析の核である共変量モデルの開発を、遺伝的アルゴリズムを利用してコンピュータに完全自動で行わせるという画期的なシステムを開発する。これにあたり、以下の4つの目標を掲げた。

- (1) 遺伝的アルゴリズムの利用にあたり、数式構造を遺伝子配列で表現する情報記述法を確立する。
- (2) 逐次提示される候補式のパラメーター最適化を行うルーチンを開発する。
- (3) パラメーター最適化にあたり、実験値との誤差を計算する拡張最小二乗法アルゴリズムを実装する。
- (4) 数千~数万回のパラメーター最適化を効率よく行うために、マルチスレッドアルゴリズムを開発する。

本研究の目標は PopPK モデルの自動開発システムの構築であるが、(3) の臨床薬物動態データに対して拡張最小二乗法による最適化計算を行うという点を除けば、アルゴ

リズムの大半は一般性を有している。したがって、システム自体をオブジェクト指向に基づいて開発することによって、基本コンポーネントを様々な問題に対して利用可能となる。医薬品開発に限っても、化合物の構造活性相関、経口製剤の *in vitro-in vivo* 相関、処方設計における実験計画・応答曲面法などが想定される。本研究の実施にあたっては、臨床薬物動態データの解析アルゴリズムの実装はかなり複雑であるために、まずは数式自動最適化アルゴリズムの動作性を構造活性相関解析に適用することで行うこととした。その上で、本アルゴリズムを PopPK 解析に適用し、共変量モデル作成における数式自動最適化の有効性を検証した。

## 3. 研究の方法

(1) 遺伝的アルゴリズムを利用した数式最適化アルゴリズムの開発：まず、木構造に基づく数値計算アルゴリズムを考案、実装した。数式はその性質上木構造で表される点に着目し、木のノード部分には演算子・関数・変数のいずれかが割り当てられているものと仮定した。この際に、各ノードは下位のノードに値を要求し適切な演算を行って要求元である上位ノードに値を返すと定義する。この定義によって、根ノードに値を要求するだけで命令が下位に波及し、再帰的な演算を通じて数式演算が可能となる。本アルゴリズムをオブジェクト指向言語である JAVA 言語を用いて実装した。

前述の定義に従うと、ノードに割り当てられた素子が演算子である場合は2つ、関数の場合は1つの下位ノードを持ち、変数は下位ノードを持たずに値だけを保持することになる。このように下位ノードの数がノードタイプによって定まっているために、遺伝コードの要素の順に従って、数式木に幅優先でノードの割り付けを行うと、記号列から数式木を再構成したり、あるいは逆変換することも可能となる (図1)。本相互変換アルゴリズムをコンピュータに実装した。

遺伝的アルゴリズムは、常法に従ってコンピュータに実装した。すなわち、まず遺伝子

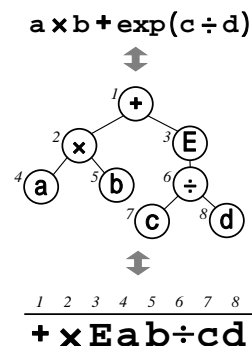


図1 数式の遺伝的コーディング

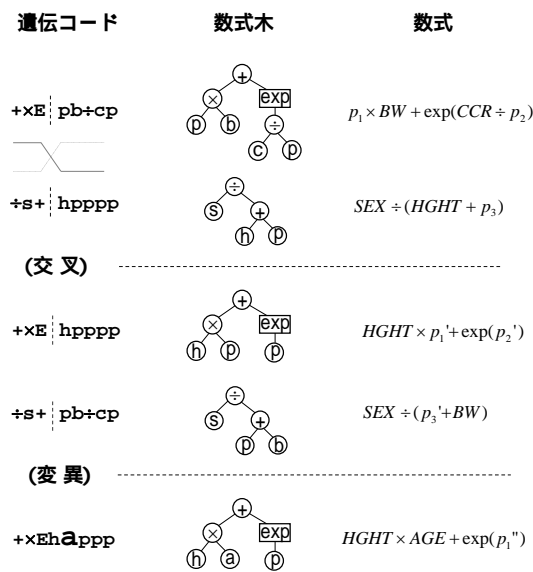


図2 数式の進化過程

の優劣を適合度によって評価した上で、ルールトワイール法により適合度の高いものを高確率で選択し、次に選ばれた2対の親を交叉、変異させて新しい遺伝コード(子)を作成する。なお、交叉および変異の確率は、計算問題ごとに最適化できるように、任意に設定できるインターフェースを開発した。図2は、遺伝コードが与えられたときの数式の進化の様子を示したものである。

さらに、実験データに適合する数式を探索する場合は、変数とパラメータを厳密に区別する必要がある。すなわち、パラメータは特定の処理を実行するために与えられる変数であり、実験データに適合するために調節されるものである。しかしながら、パラメータ推定において、例えば、 $p_1 + (p_2 - v_1)$  (図3、ケース2)では、 $p_1$ と $p_2$ を一意に決定することができず不定になる。本問題を

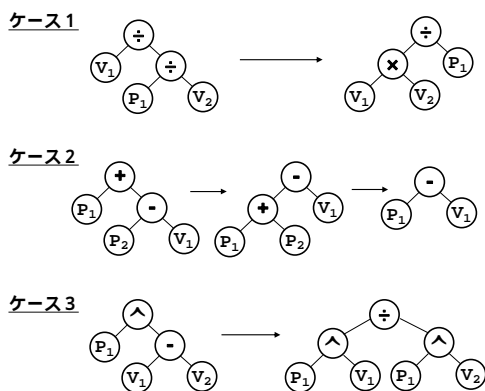


図3 数式木の等価変換の例

ケース2はパラメータを合一する例を示したものであり、これによってパラメータ最適化計算でパラメータ値が不定になるのを回避する。

解決するためにパラメータを合一するルールベースを開発した。最終的に実装されたルールの基本的概念を示すと、以下の点に集約される。

(1) 数式木中のパラメータは可能な限り上位の階層に集まるように計算順序を等価変換する。

(2) 上位階層に対してパラメータの存在を確認し、合一可能であれば式を等価交換して合一する。

(3) 可能な限り数式を分割し、演算の粒度を小さくする。

これらの点について四則演算や関数の性質を考慮してそれぞれ個別に処理アルゴリズムを実装した。具体的な例を図3に示す。

(2) 数式自動最適化アルゴリズムを利用したCYP3A4 阻害予測のための分類木の構築:

PubChem BioAssay データベース AID1851 には、proluciferin を基質として各種 CYP の発現系マイクロソームでの代謝阻害に関するハイスループットスクリーニングの結果 (12,464 化合物) が収載されている。CYP3A4 に対する阻害では、5,266 化合物が陽性、7,751 化合物が陰性と判断されている。本研究では、このデータを用いて再帰分割解析を行い、CYP3A4 阻害剤を判定する決定木の開発を試みた。

通常の再帰分割解析法では、各ノードでの分類基準は単一の変数とその閾値によって設定される。本研究では、より優れた基準を設定するために、分類基準は多変数非線形式も許すこととし、その分類基準式を先の数式最適化アルゴリズムにより探索した。分類効率は分割による多様性の減少で評価できる点に着目し、分類基準の優劣は情報エントロピー理論に基づく適合度指標で判定した。すなわち、分割による情報エントロピーの減少すなわち情報利得は、

$$\Delta IE = \left( \frac{N_{S_1}}{N_S} IE(S_1) + \frac{N_{S_2}}{N_S} IE(S_2) \right) - IE(S)$$

$$IE(S) = - \sum_{i=p,n} \frac{N_i}{N_S} \log \frac{N_i}{N_S}$$

で表される。ここで、 $N_S$  はグループ  $S$  中の化合物数、 $N_i$  は  $S$  中の阻害剤 ( $p$ ) ないし非阻害剤 ( $n$ ) の数である。情報利得は対数尤度であり、最終的な適合度指標としては AIC を用いることとした。

$$\begin{aligned} fitness &= -AIC \\ &= -2N_S \Delta IE - 2k \\ &\approx -2(N_{S_1} IE(S_1) + N_{S_2} IE(S_2)) - 2k \end{aligned}$$

$k$  は分類基準式に含まれるパラメータ総数であり、数式の複雑さに対するペナルティ項として働く。本基準の下でデータ分割を再帰的に繰り返し、最終的な決定木を作成した。得られた決定木の性能は、機械学習法として汎用されるサポートベクトルマシンの結果

と比較した。サポートベクトルマシンによるモデリングでは、統計解析ソフト R に実装された package “e1071” を用いた。

(3)完全自動 PopPK モデリング法の開発とトブラマイシン薬物動態解析への応用: PopPK モデル解析とは、固定効果と変量効果を考えた非線形混合効果モデルに基づく解析である。ここでこの目標は共変量による固定効果モデルを数式最適化アルゴリズムにより自動構築させることであり、拡張最小二乗法における尤度計算によりその適合度を判定した。

$$\begin{aligned} \text{fitness} &= -AIC \\ &= -OBJ - 2k \end{aligned}$$

$$OBJ = \sum_{i=1}^m \left\{ \frac{[y_i - f_i(\theta)]^2}{u_i(\theta, \sigma)} + \ln[u_i(\theta, \sigma)] \right\}$$

ここで、 $y_i$  は測定値、 $m$  は測定値の数、 $f_i(\theta)$  はモデルによる予測値、 $u_i(\theta, \sigma)$  は分散モデルを示す。

拡張最小二乗法を適用する際には、共変量モデルの中に含まれる定数項、各速度論パラメーターに対する個体間変動の分散、血中濃度の個体内変動の分散といった多数のパラメーターを同時に推定することが必要である。そこで、パラメーター最適化法には比較的安定な収束が期待できる準ニュートン法を採用したが、大規模パラメーター推定において多数の局所解の存在が問題となる。そこで、遺伝的アルゴリズムを用いたプレスクリーニングにより、適切な初期値を探索した。遺伝的アルゴリズムによるパラメーター最適化においては、個々の推定パラメーターに対して 16 ビット配列を仮定し、これを浮動小数点数に変換することによって最小二乗法計算を行った。図 4 には、遺伝コードが与えられたときにパラメーターおよび適合度

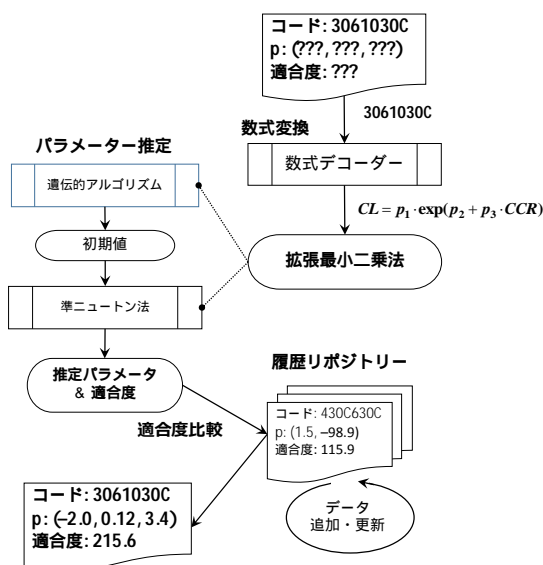


図 4 遺伝コードが与えられたときのパラメーター/適合度決定までの流れ

を決定するまでの流れをまとめた。なお、このプロセスでは大規模な臨床薬物動態データに対して解析を行うために、かなりの計算コストがかかる。そこで、JAVA のスレッドプログラミングを利用して、遺伝的アルゴリズムにおける各個体の適合度を並列計算で求められるようにした。

PopPK 解析には Aarons らが報告したトブラマイシンの繰返し静脈注射のデータを使用した (Br J Clin Pharmacol 28, 305, 1989)。なお、データはワシントン大学のホームページ (<http://depts.washington.edu/rfpk/service/datasets/index.html>) からダウンロードした。これは 97 人の被験者からなるデータで、体重(BW)、クレアチニンクリアランス(CCR)、年齢(Age)、性別(Sex)が背景情報として提示されている。

#### 4. 研究成果

(1) 数式自動最適化アルゴリズムを利用した CYP3A4 阻害予測のための分類木の構築: 通常、薬物治療では複数の薬剤が同時に投与されることが多く、臨床現場において薬物間相互作用が大きな問題となっている。CYP3A4 は薬物の酸化的代謝に関わるチトクロム P450 スーパーファミリーに属する分子種の一つであり、市販される薬物の肝代謝の約 50%に関わっている。したがって、CYP3A4 の阻害を介して薬物間相互作用が生じる可能性は非常に高い。そのため新薬開発でも薬物の併用を前提に CYP3A4 に対する阻害活性を予測することが求められる。そこで本研究では、CYP3A4 阻害活性に関する優れた構造活性相関予測モデルの開発を目的として、本数式自動最適化アルゴリズムを適用し網羅的な構造活性相関解析を行った。

まず、PubChem BioAssay AID1851 データセットに収載される化合物の構造情報に基づき、ADMET Predictor を用いて分子記述子を計算した。構造活性相関モデリングでは、一般に用いられる疎水性パラメーター (Moriguchi's partition coefficient; MlogP)、分子サイズ (McGowan's characteristic volume; VMcGowan)、水素結合能 (Number of hydrogen bond acceptors and donors; HBA and HBD)、イオン化状態パラメーター (Fraction of unionized, cationized, anionized molecules; FUnion, FCation, and FAnion) に関する計 7 種類の分子記述子と、CYP3A4 阻害活性との単相関で相関係数の高かった分極率 (Miller's Polarizability; PolarizM) と  $\pi$  Fukui index (Sum of absolute values of pi Fukui(+) indices on C; Pi\_AFPIc) を合わせた 9 種の分子記述子からなるデータセットを構築した。

AID1851 に収載される 12,464 化合物を訓練用 (正例 3,601、負例 5,110) とテスト用 (正例 1,548、負例 2,205) に分割し、CYP3A4

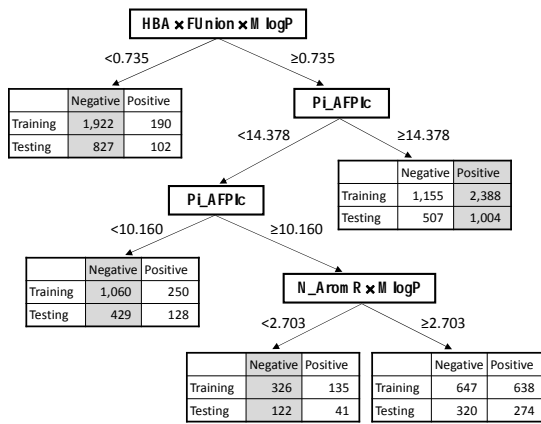


図5 CYP3A4 阻害活性の有無を判定するための決定木

表は決定木によって各ノードに分類された訓練用データおよびテスト用データの数をまとめている。

阻害活性に対する決定木モデルを構築した。訓練用データセットに対して、赤池情報量基準 (AIC) の減少が見られなくなるまで再帰的分割を繰り返したところ、図5に示す決定木が得られた。この決定木の分類正解率は、訓練用データに対して 72.8%、テスト用データに対して 72.0%であった。

第一分岐ルールには HBA x FUnion x MlogP が選ばれ、これが 0.753 を下回る群では 91%が CYP3A4 の阻害活性を有していない化合物であった。この結果より、水溶性が高く、水溶液中で電離してイオンになりやすいものや水素結合受容能の低いものは CYP3A4 阻害剤になりにくいことが示された。さらに、HBA x FUnion x MlogP が大きい場合でも、求電子性の高い炭素原子をもたなければ阻害剤となりにくい傾向にあった。

さらに、代表的な機械学習法であるサポートベクトルマシン (SVM) を用いて、同じデータを解析した。SVM は、元の特徴ベクトルを非線形カーネル関数によって変換し、写像空間においてマージン最大化を原理として線形識別する方法である。高い認識性能を有することから、構造活性相関解析においてしばしば利用されている。実際の解析で得られた SVM モデルの分類正解率は、訓練用データに対して 75%、テスト用データに対して 74%であった。本検討で構築された決定木の精度は先に示した通りであり、本モデルは SVM モデルと比べて遜色ない上、SVM と違って構造活性相関が数式として明示されるため、極めて有用なモデルと考えられた。

(2) 完全自動 PopPK モデリング法によるトブラマイシン母集団薬物動態モデルの開発

Aarons �らは次に示す条件の下でトブラマイシンの PopPK モデルを導いた。  
(1) 臨床薬物動態データに対して 2-コンパートメントモデルを仮定する。  
(2) 全身クリアランス ( $CL_{tot}$ ) と中心コン

パートメントの分布容積 ( $V_1$ ) に対して共変量モデルを仮定する。

(3) 個体間変動、個体内変動に対しては比例誤差モデルを仮定する。

Aarons �らが提唱した PopPK モデルは、  
 $CL_{tot} = 0.0559 \times CCR$  (L/h/kg)

$V_1 = 0.327 \times WT$  (L/kg)

$k_{12} = 0.0119$  ( $h^{-1}$ ),  $k_{21} = 0.026$  ( $h^{-1}$ )

$\eta_{CL} : 31.4$  (%),  $\eta_{V1} : 0.266$  (%)

$\sigma : 21.2$  (%)

OBJ : -25.796

である。ここで、各パラメーターおよび個体間変動 ( $\eta$ )、個体内変動 ( $\sigma$ )、OBJ 値は NONMEM7.2 (ICON plc) によって推定されたものである。そこで、本研究では同じ制約条件の下で共変量モデルの探索を行うこととし、遺伝的アルゴリズムを利用した数式自動最適化アルゴリズムの有用性を検証した。

遺伝的アルゴリズムのパラメーターをそれぞれ世代数 1,000、個体数 50 (サブ集団数 5 x 各 10 個体)、染色体長 15、交叉点数 2、変異点数 2、移住頻度 5 世代毎、移住者数各 1 個体としてプログラムを実行したところ、図6に示す軌跡での数式の変遷が認められた。なお、この図は各世代において最大適合度を示した個体の共変量モデルを示したものである。世代を経る毎に適合度が上昇していく様子が示されている。なお、最小二乗法計算が収束しない個体もあるため、各世代の集団における平均適合度は計算しなかった。CL および  $V_1$  の共変量モデルに着目すると、それぞれに対して CCR, WT が常に数式内に組み込まれており、トブラマイシンの全身クリアランスにはクレアチニンクリアランスが、中心コンパートメントの分布容積には体重が重要な因子であることが示唆された。200 世代の計算ののち最終的に得られた共変量モデルに対して、計算結果の確認も含めて

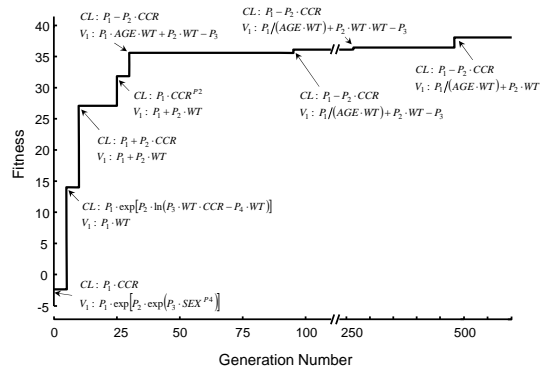


図6 完全自動 PopPK モデリングシステムを利用したトブラマイシン臨床薬物動態に関する共変量モデルの構築

図は進化過程の各世代での最良個体の適合度を示す。

NONMEM を使って再解析したところ、完全一致したパラメーター値が得られた。NONMEM の出力結果は以下の通りである。

$$CL_{tot} = 0.388 + 0.0534 \times CCR \text{ (L/h/kg)}$$

$$V_1 = 0.258 \times WT + 7040 / (WT \times AGE) \text{ (L/kg)}$$

$$k_{12} = 0.0167 \text{ (h}^{-1}\text{)}, \quad k_{21} = 0.0500 \text{ (h}^{-1}\text{)}$$

$$\eta_{CL} : 32.5 \text{ (\%)}, \quad \eta_{V1} : 0.1 \text{ (\%)}$$

$$\sigma : 20.0 \text{ (\%)}$$

$$OBJ : -56.037$$

Aarons らのモデルと比較すると、 $CL_{tot}$  に定数項を、 $V_1$  に体重と年齢の積の逆数からなる項をそれぞれ追加したモデルとなっている。尤度比検定では、自由度 1 に対して OBJ 値が -3.84 以上減少すれば有意と判断される。今回のモデルでは、自由度が 2 増えているが、OBJ 値が約 -30 減少しており、明らかに有効な共変量が組み込まれたことがわかる。

本研究で開発したプログラムでユーザー自身が行うことは、単に臨床薬物動態データの入力とプログラムへの投入だけであり、最小二乗法における初期値探索も含め、共変量モデルの構築はコンピューターによって完全自動で行われる。ただ、科学的見地からは OBJ 値が最も小さいモデルが必ずしも適当とは言い難い。最終的には臨床薬理学者による判断が必要である。一つの工夫は最終的に出力されるモデルに不自然な共変量が組み込まれないような制約できるように設計した。もう一つの工夫は、計算過程で得られるすべての共変量モデルから、予め指定した数だけ上位のモデルを取り出すことができるように設計した。専門家がこれらのモデルを見ることでより良いモデルを見出すヒントになるものと期待している。今回開発した完全自動 PopPK モデリングシステムは、最終解を得るために用いるというよりはむしろ、網羅的な探索によるプレスクリーニングと位置づけ、良質で科学的な PopPK モデルを得るための手段と考えるのが適切であろう。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 2 件)

K. Yano, S. Mita, K. Morimoto, T. Haraguchi, H. Arakawa, M. Yoshida, F. Yamashita, T. Uchida, T. Ogihara, Multiple linear regression analysis indicates association of P-glycoprotein substrate or inhibitor character with bitterness intensity, measured with a sensor. J. Pharm. Sci., 査読有, 104, 2015, in press  
DOI:10.1002/jps.24232  
P. Buchwald, F. Yamashita, Bilinear

model for the size-dependency of the CYP3A4 inhibitory activity of structurally diverse compounds. Mol. Inform., 査読有, 33, 2014, 8-14  
DOI:10.1002/minf.201300132

〔学会発表〕(計 3 件)

F. Yamashita, S. Yoshida, M. Hashida, Computer-aided full-automatic construction of covariate models in population pharmacokinetic analysis, 19th North American ISSX Meeting and 29th JSSX Annual Meeting, 2014 年 10 月 22 日, San Francisco(USA)  
F. Yamashita, Utilization of machine learning techniques to pursue automatic PPK modeling, 2013 PAGJA-PAGK Joint Annual Meeting, 2013 年 11 月 28 日, Seoul(Korea)  
吉田秀哉、藤田淳人、山下富義、橋田 充、遺伝的数式プログラミング法を用いた CYP3A4 阻害活性に関する構造活性相関解析、第 28 回日本薬物動態学会年会、2013 年 10 月 9 日、タワーホール船堀(東京都)

〔図書〕(計 1 件)

浅井義之、山下富義、羊土社、Dr.北野のゼロから始めるシステムバイオロジー 第 10 講 PhysioDesigner による生理機能の多階層モデル構築とシミュレーション、2015、189

〔その他〕

ホームページ等

[http://dds.pharm.kyoto-u.ac.jp/Dds\\_Home/index.htm](http://dds.pharm.kyoto-u.ac.jp/Dds_Home/index.htm)

## 6. 研究組織

(1)研究代表者

山下富義 (YAMASHITA FUMIYOSHI)  
京都大学・薬学研究科・准教授  
研究者番号：30243041

(2)連携研究者

矢野育子 (YANO IKUKO)  
京都大学・薬学研究科・准教授  
研究者番号：50273446

天野 晃 (AMANO AKIRA)  
立命館大学・生命科学部・教授  
研究者番号：60252491