DMD Fast Forward. Published on January 31, 2012 as doi:10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



# Using expression data for quantification of active processes in physiologically-based pharmacokinetic modeling

Michaela Meyer, Sebastian Schneckener, Bernd Ludewig, Lars Kuepfer, Joerg Lippert

Systems Biology and Computational Solutions, Bayer Technology Services, Building 9115, 51368 Leverkusen, Germany

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



# Running title page

## Running title

Integration of expression data into PBPK models

## Corresponding author

*Lars Kuepfer, Systems Biology and Computational Solutions, Bayer Technology Services GmbH, Building 9115, 51368 Leverkusen, Germany, Tel +49 214 30 22745, Fax +49 214 30 64801, lars.kuepfer@bayer.com

## Number of text pages

21

## Number of tables

3

## Number of figures

4

## Number of references

40

## Number of words in the abstract

244

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



## Number of words in the introduction

715

## Number of words in the discussion

1192

## List of nonstandard abbreviations

| | | |
|---|---|---|
| ABCs | | ATP binding cassette transporter |
| ADME | | Adsorption, distribution, metabolism, and excretion |
| Ae | | Amount excreted (to the urine) |
| AUC | | Area under the curve |
| CEL | | An affymetrix file format |
| CLdegrad | | Clearance (degradation) |
| CLhep | H | eaptic clearance |
| CLren | Re | nal clearance |
| CLsult | C | learance through sulfotransferases |
| Cmax | M | aximum concentration |
| CYPs | Cy | tochromes |
| EST | E | xpessed sequence tags |
| GAPDH | | Glyceraldehyde 3-phosphate dehydrogenase |
| GI | G | astro-intestinal |
| HMG-CoA-reductase | | 3-hydroxy-3-methylglutaryl-coenzyme A reductase |
| iv | in | travenous |

| | |
|---|---|
| IVIVC | in vitro in vivo correlation |
| IVIVE | in vitro in vivo extrapolation |
| Kcat | cat alytic constant |
| Km | Th e Michaelis constant |
| logP | loga rithmic partition coefficient |
| MoBi | Molecular Biology modeling software |
| MRP2 | Mul ti resistence protein 2 |
| OAT3 | Organic anion transporter 3 |
| OATP1B1 | Organic anion transporter 1B1 |
| OATs | Organic anion transporter protein family |
| PAPS | 3' -Phosphoadenosine-5'-phosphosulfate |
| PBPK | P hysiology based pharmacokinetic |
| P-gp | P ermeability glycoprotein |
| PK | Ph armacokinetic |
| pKa | A cid dissociation constant |
| po | P er oral |
| PPIA | P eptidylprolyl isomerase A |
| RMSD | Root mean square deviation |
| RT-PCR | Reverse transcription-polymerase chain reaction |
| SULT isoform | Sulfotransferase isoform |
| SULTs | Su lfotransferase enzymes |
| UGTs | UDP-glu coronosyltransferase |

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



## Abstract

Active processes involved in drug metabolization and distribution mediated by enzymes, transporters or binding partners mostly occur simultaneously in various organs. A quantitative description of active processes, however, is difficult due to a limited experimental accessibility of tissue-specific protein activity *in vivo*. In this work we present a novel approach to estimate *in vivo* activity of such enzymes or transporters which have an influence on drug pharmacokinetics. Tissue specific mRNA expression is used as a surrogate for protein abundance and activity and is integrated into physiologically-based pharmacokinetic (PBPK) models which already represent detailed anatomical and physiological information. The new approach was evaluated using three publicly available databases: Whole genome expression microarrays from ArrayExpress, RT-PCR derived gene expression estimates collected from literature, and expressed sequence tags (EST) from UniGene. Expression data were preprocessed and stored in a customized database that was then used to build PBPK models for pravastatin in humans. These models represented drug uptake by OATP1B1 and OAT3, active efflux by MRP2, and metabolization by sulfotransferases in either liver, kidney and/or intestine. Bench-marking of PBPK models based on gene expression data against alternative models with either less complex model structure or randomly assigned gene expression values clearly demonstrated the superior model performance of the former. Besides an accurate prediction of drug pharmacokinetics, integration of relative gene expression data in PBPK models offers the unique possibility to simultaneously investigate drug-drug interactions in all relevant organs due to the physiological representation of protein mediated processes.

## *Introduction*

Computational models play an increasing role in pharmaceutical research and development, since they offer an efficient way for storing, representing and analyzing experimental data at each stage of (pre-)clinical development (Gabrielsson et al., 2010; Vicini, 2010; Eissing, 2011). Ideally, such models integrate the current state of knowledge, which can be validated by testing hypotheses computationally in an efficient and rational way.

Physiologically-based pharmacokinetic (PBPK) models describe the pharmacokinetic behavior of a substance within the human body based on a large amount of prior physiological and biological information (Poulin and Theil, 2002b; Poulin and Theil, 2002a; Willmann et al., 2003; Rodgers et al., 2005; Rodgers and Rowland, 2006; Nestorov, 2007; Eissing, 2011). The various prediction methods included in PBPK modeling are based on compound-deduced parameters and quantify absorption, distribution, metabolisation and excretion (ADME) of a drug (Poulin and Theil, 2002b; Poulin and Theil, 2002a; Willmann et al., 2003; Willmann et al., 2004; Rodgers et al., 2005; Willmann et al., 2005; Rodgers and Rowland, 2006). Most notably, all model parameters are hence either obtained from collections of literature data or are derived from few physiochemical properties of a compound such as lipophilicity or molecular weight.

While passive diffusion processes are merely driven by concentration gradients, several transport and degradation processes in the body are active, i.e. they are protein-mediated and net energy consuming. In contrast to passive processes, which are only dependent on a compound's physicochemistry and concentration gradient, active processes are governed by the binding affinity between a drug and a specific protein ($K_m$), catalytic velocity ($k_{cat}$) and tissue-specific protein abundance ($E_o$). Most notably, the same active process may simultaneously occur in

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



various types of tissue all over the body, which hampers a rigorous representation in the underlying model structure all the more since binding affinity and catalytic velocity can only be measured *in vitro*. As of now, representation of active processes in PBPK models is hence restricted to isolated organs and processes.

We here present a novel approach which allows the simultaneous consideration of active processes in multiple organs by taking gene expression data as a proxy for tissue-specific protein abundance. The new concept implies that protein availability and catalytic rate constants, which ultimately underlie enzyme and transporter activity, can be decoupled: Relative protein abundance can be set according to available expression data while catalytic parameters are described by a global kinetic rate constant that is adjusted during model establishment. On the one hand, this method enables direct estimation of *in vivo* enzyme and transporter activity based on data-based inclusion of tissue-specific protein abundance which represents first-hand experimental-measurements. On the other hand, the number of free model parameters, which would have to be measured or adjusted in case of an exhaustive mechanistic representation in multiple organs, is reduced significantly.

To this end, large-scale gene-expression data from publicly available sources were downloaded, processed, stored and customized such that they can directly be utilized in PBPK model building: Whole genome expression arrays from ArrayExpress (ArrayExpress, 2010), RT-PCR derived gene expression estimates from literature (Nishimura et al., 2003; Nishimura and Naito, 2005; Nishimura and Naito, 2006), and expressed sequence tags (EST) from UniGene. The consolidated data was stored in a database with three sections termed EST (UniGene), Array (ArrayExpress), and RT-PCR (Literature), respectively (Figure 1).

To evaluate the benefit of using gene expression data, different PBPK models of pravastatin (Singhvi et al., 1990; Everett et al., 1991; Mwinyi et al., 2004; Niemi et al., 2006) with increasing complexity were subsequently considered. Pravastatin was chosen since it is known to be a substrate for various transporters (OATP1B1 (*SLCO1B1*), MRP2 (*ABCC2*), OAT3 (*SLC22A8*) and metabolic enzymes (sulfotransferases) that are expressed simultaneously in several organs (Hatanaka, 2000; Kivisto and Niemi, 2007) (Figure 2). A basic PBPK model with a simple kidney and liver clearance and an extended basic PBPK model with additional consideration of metabolism in the gastro-intestinal tract were developed first (Figure 1). This workflow represents a typical PBPK modeling approach that focuses on the main metabolizing organs such as liver, kidney and intestine. The performance of these model alternatives was compared to three independently parameterized PBPK model variants for pravastatin including gene expression profiles obtained from the newly generated database and to alternative model variants generated with randomized gene expression values to exclude artifacts based on differences in systemic degree of freedom. All models were benchmarked by comparison of the simulation error as (Figure 1).

## *Materials and Methods*

**Software.** The PBPK model for pravastatin was built using the commercial software tool PK-Sim® Version 4.2 (Bayer Technology Services GmbH, Leverkusen, Germany) (Willmann et al., 2003; Willmann et al., 2004; Willmann et al., 2005; Eissing, 2011). PK-Sim® generated PBPK models were exported and modified in MoBi® (version 2.2; Bayer Technology Services) as described further below. Academic licenses of PK-Sim® and MoBi® are available free of charge. All optimizations and batch mode simulations were carried out using Matlab (version 7;

MathWorks, Natick, MA) and the MoBi® Toolbox for Matlab (version 2.0; Bayer Technology Services). All numerical calculations were performed using R version 2.8.1. Text processing was done using the scripting language perl, version 5.10. All data was stored in a Microsoft Access Database for consistent handling.

**Parameter identification**. Optimizations were based on a simulated annealing algorithm as provided by the MoBi® Toolbox for Matlab. In order to identify solutions close to global optima (Moles et al., 2003), 1000 repetitive optimizations were performed for each model version which were based on randomly distributed starting guesses. The root mean square deviation (RMSD) relative to the various experimental data, henceforth called simulation error, was considered as objective function for all optimizations.

**PBPK model development and parameterization**. All PBPK models described in the following consider the pharmacokinetic characteristics of absorption, metabolization and excretion of the simulated drug and assume an average 30 year old European male subject with a body weight of 73 kg and a height of 176 cm. Physiological parameters describing basic model structure such as organ volumes, blood flow rates or surface permeabilities are provided by the software tool PK-Sim® (Willmann et al., 2003). Furthermore, the biophysical properties of pravastatin as summarized in Table 1 are integrated in all PBPK model variants. Lipophilicity (logP) was reported to be in the range from -0.23 at pH 7 to 1.6 at pH 5 (Serajuddin et al., 1991) and was set as fit parameter in all PBPK models. Parameter identification for all PBPK models was performed considering experimental plasma concentration-time profiles and urinary secretion data after intravenous (iv) and oral (po) application of 9.9 and 40 mg pravastatin, respectively (Singhvi et al., 1990; Mwinyi et al., 2004; Niemi et al., 2006).

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



**Basic PBPK model.** A basic PBPK model for pravastatin was constructed which considers clearance processes in kidney and liver as first order rate equations. The model thus represents an exemplary PBPK model with little structural complexity based on mainly phenomenological observations. The adjusted parameters were lipophilicity, intestinal permeability and degradation in kidney and liver.

**Extended basic PBPK model.** The basic PBPK model was extended by two additional processes known to influence metabolization and distribution of pravastatin: First, acid catalyzed degradation in the gastrointestinal (GI) tract was added as a first order clearance process ($CL_{degrad}$) in the intestinal lumen to account for chemical and enzymatic degradation responsible for the low bioavailability of pravastatin of 19.1 % (Singhvi et al., 1990). Second, hepatic uptake of pravastatin by the organic anion transporting polypeptide 1B1 (OATP1B1) (Niemi et al., 2006) was implemented as Michaelis Menten kinetics with a $K_m$ value of 11.5 µM (Nakai et al., 2001).

**Expression profile PBPK model.** The extended basic PBPK model is further refined by active, protein mediated, tissue specific processes. Enzymatic reactions and transport processes in various tissues of the body are implemented in PK-Sim® by using the Michaelis-Menten kinetics equation and defining an affinity constant $K_m$ and a maximum velocity $V_{max}$ for each process. In contrast to $K_m$ values, experimentally determined in vitro $V_{max}$ values cannot be used in the PBPK model without further scaling, since they depend on the total enzyme or transporter concentration, henceforth referred to as $E_0$.

Our concept of using gene expression data as a proxy for protein abundance is based on the definition of the maximum velocity $V_{max}$ [$\mu$mol/l/min]. Based on the Michaelis-Menten equation, $V_{max}$ is dependent on both the total enzyme or transporter concentration $E_0$ [µmol/l] and the catalytic rate constant $k_{cat}$ [1/min]:

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



$$V_{\max} = k_{cat} \cdot E_0 \,. \tag{1}$$

Assuming that $k_{cat}$ is not influenced by *in vivo* factors, the tissue-specific maximum velocity $V_{\max}^{Organ,i}$ is defined by:

$$V_{max}^{Organ,i} = k_{cat} \cdot E_0^{Organ,i} \,. \tag{2}$$

Replacing $E_0^{Organ,i}$ by relative expression values $e_{rel}^{Organ,i}$ [-] multiplied by a scaling factor *SF* [µmol/l] corrects for the absolute *in vivo* protein concentration such that equation 2 can be rewritten to obtain:

$$V_{\max}^{Organ,i} = k_{cat} \cdot SF \cdot e_{\mathrm{Re}l}^{Organ,i} = k_{cat}^{*} \cdot e_{\mathrm{Re}l}^{Organ,i} \,. \tag{3}$$

Equation 3 describes the new approach for estimating absolute tissue-specific activities of enzymes or transporter according to which the relative tissue-specific distribution $e_{\mathrm{Re}l}^{Organ,i}$ is scaled by an apparent catalytic rate constant $k_{cat}^{*}$ [µmol/l/min]. Note, that $k_{cat}^{*}$ is a global model parameter, which is used for calculation of the tissue-specific maximum velocity (see equation 4) and that *SF* implicitly considers translational efficacy and post-transcriptional modifications for a particular protein.

$$\left.\begin{array}{l} V_{max}^{Organ,1} = e_{Rel}^{Organ,1} \\ V_{max}^{Organ,2} = e_{Rel}^{Organ,2} \\ \quad\vdots \\ V_{max}^{Organ,i} = e_{Rel}^{Organ,i} \end{array}\right\} \cdot k_{cat}^{*} \tag{4}$$

In conclusion, by using equation 4 for prediction of *in vivo* $V_{max}$ values instead of the original Michaelis-Menten equation only one single free parameter, i.e. $k_{cat}^{*}$, instead of an *i*-fold consideration of $V_{max}$ for each organ ($V_{max}^{Organ,i}$) has to be determined during parameter

identification. This novel modeling approach therefore significantly reduces the amount of free parameters to be estimated during model building.

Three alternative models were built applying this modeling approach using expression data from three different sources: Whole genome expression arrays from ArrayExpress (ArrayExpress, 2010), RT-PCR derived gene expression estimates from literature (Nishimura et al., 2003; Nishimura and Naito, 2005; Nishimura and Naito, 2006), and expressed sequence tags (EST) from UniGene (Unigene, 2010). Details to download, automatic processing and curation of data from these resources is described further below in the Materials and Methods part.

**Randomised expression PBPK models.** Alternative PBPK models were generated with randomized gene expression for further benchmarking and to exclude structural differences of the PBPK model variants leading to different model performance. Random gene expression data was generated by random drawing from a lognormal distribution. The lognormal distribution was parameterized with mean and variance of experimentally observed expression data. Expression profile PBPK models as described earlier were used and gene expression replaced by random values.

**Expressed sequence tags (EST) from UniGene**. Relevant files from the human section of UniGene (Wheeler et al., 2003) were downloaded from the UniGene site. In a semi-automated approach, all balanced, normalized or subtracted libraries were excluded from further analysis. For the remaining libraries, data was extracted into a database that contains, for a given UniGene cluster, identifier and a particular library, respectively, counts per cluster and total count per library. Meta data of libraries was mapped to a standardized vocabulary such that age of the sample donor, health state and physiological location of the sample can be included. This approach allows to select a subset of libraries according to criteria like sample site and donor

properties. The expression value of a particular gene of a subset of libraries was calculated as the ratio of the sum of sample counts to the sum of total counts, thereby correcting for differences in library sizes.

**Microarray data from ArrayExpress.** Raw data from 36 different healthy tissues profiled by microarry analysis (Ge et al., 2005) and produced using the Affymetrix HG-U133A platform was downloaded as CEL files from ArrayExpress (Parkinson et al., 2007) and processed using R/Bioconductor package simpleaffy (Wilson and Miller, 2005). For normalization, the mas5 algorithm (Hubbell et al., 2002) was applied. Probeset identifiers were mapped to UniGene identifiers (Wheeler et al., 2005) and entrez gene identifier (Maglott et al., 2007). If multiple probeset identifiers were mapped to the same gene, the mean expression over the probesets was used per gene.

**RT-PCR data from literature.** Gene expression values for cytochrome P450 family enzymes and other proteins of known pharmacokinetic relevance from different tissues were extracted from published data (Nishimura and Naito, 2006), likewise metabolizing enzymes (Nishimura et al., 2003), and ATP binding cassette and solute carrier transporter (Nishimura and Naito, 2005), respectively. Here, probes originate from human tissue samples and expression values were estimated using real time reverse transcription polymerase chain reaction (RT-PCR). The published values were given as relative expression levels with respect to a reference housekeeping gene, i.e. either GAPDH (Nishimura et al., 2003) or PPIA (Nishimura and Naito, 2005; Nishimura and Naito, 2006).

**Statistical analysis of expression data.** For easier visual comparison of gene expression data from different experimental platforms, scaling was performed such that the geometric mean of expression values equals a constant value for each data source. Ideally, the measured expression

of a g ene over different tissues and organs should be independent from the experimental platform. To estimate the platform effect, the Pearson correlation between gene expression profiles of two data sources is calculated. All calculations are performed using R (Development Core Team, 2010) and the *cor* function from the *stat* packages.

**Demographic information.** The Nishimura data was generated from samples of Caucasian origin, males and females of age 15 to 68 with no health related information available. Demographic annotations in UniGene are of variable extent, but specific for each library. Essentially, all ethnicities are represented. The data that was used from ArrayExpress in t his study did not include demographic or disease related information.

## *Results*

**Establishment of the expression database.** At first, gene expression data were retrieved from publicly available databases such as UniGene (Wheeler et al., 2005), ArrayExpress (Ge et al., 2005) or extracted from literature (Nishimura et al., 2003; Nishimura and Naito, 2005; Nishimura and Naito, 2006) (Figure 1). These data sources were chosen since they either provide large scale expression data or focus specifically on selected proteins which are of particular interest for pharmacokinetics such as metabolizing enzymes (CYPs, UGTs, etc.) or transport proteins (ABCs, OATs, etc.). Notably, all three data sources use a different type of assay technique (expressed sequence tags EST, microarrary data, and RT-PCR measurements (Nishimura et al., 2003; Nishimura and Naito, 2005; Nishimura and Naito, 2006) and therefore require specific workflows for data screening, processing and storage which was done in a second step (Figure 1). Download of data and subsequent reconciliation was done in a semi-automated way (see Materials and Methods). In addition to the quantitative expression data, the newly generated database also contains metadata such as a dictionary of protein and gene synonyms, and mappings for organ and tissue names which was retrieved from EntrezGene and UniProt/Swiss-Prot as primary sources.

Upon completion, the established customized expression database utilizable for PBPK modeling contains 3.4 million measurements from 23,000 different genes in 62 different tissues and organs, respectively. For each gene name, 4 synonyms are recorded on average. Based on the three original source databases, the newly generated database consists of three sections: EST (22,726 different genes resp. splice variants), Array (14,000 gene symbols), and RT-PCR (306 different genes). The core set of genes shared by all three resources contains 266 gene symbols. The

diversity of 57 tissues and organs is largest in the UniGene EST dataset, versus 26 tissues for the Microarray section and 24 tissues for RT-PCR data from literature.

**Comparison of expression database sections.** The established database has three sections representing different sources and experimental techniques of measuring gene expression. To assess the differences between the data sources, we calculated the Pearson product-moment correlation $r$ of the log of expression across several tissues. The correlation coefficient $r$ between two data sources was calculated for each of the 266 genes common to all databases. Reasonably high correlations were found between RT-PCR and Array data, respectively, while correlations to EST values were insignificant (see Table 2 for mean correlation values). While the correlation between genes of the subset used for PBPK modeling is significant between RT-PCR and Array data, this is only true for 65% of the larger set of common genes. However, only 14% of genes from the EST dataset and any other data source show a significant correlation.

**Establishment of benchmark PBPK models.** Pravastatin, a HMG-CoA-reductase inhibitor, was chosen as an exemplary case study compound because its pharmacokinetic behavior is largely dependent on various protein-mediated processes in different organs making it a demanding example for PBPK model building. The different PBPK models all consider the same physicochemical properties of pravastatin as summarized in Table 1. The underlying model structure of the distribution models in PK-Sim® and collection of physiological parameters such as blood flow rates or organ volumes, have been described in detail before (Willmann et al., 2003; Willmann et al., 2004; Willmann et al., 2005; Edginton et al., 2006; Vossen et al., 2007; Willmann et al., 2007; Willmann et al., 2010; Eissing, 2011). For the sake of reproducibility of the models shown in this work, all independent parameters needed for PBPK model building and those subject to optimization are exhaustively listed in the tables below. Most notably,

physicochemical parameters can directly be used as input for PK-Sim® in the graphical user interface for parameterization of the underlying distribution model such that ADME properties of a compound are directly quantified.

To evaluate the benefit of using gene expression data for quantification of active processes we established various PBPK models with increasing model complexity and level of detail. The performance of these PBPK model variants are described in the following. Thousand independent stochastic optimizations were performed for each model alternative. Since the focus was on identification of a global optimum in each case, only the overall best solution was given thus neglecting statistics of distribution of the 1000 parameters which rather reflect algorithmic performance of the optimizer then providing insights in robustness of the solution.

**Basic PBPK model.** First, we constructed a basic PBPK model for pravastatin as described in Materials and Methods with four adjustable parameters and little structural complexity based on mainly phenomenological observations. The PBPK model with the smallest simulation error out of 1000 independent optimization runs (Figure 3) showed an acceptable description of experimental PK curves after intravenous application of pravastatin but failed to describe the measured oral plasma concentration-time profiles with a sufficient level of accuracy despite extensive parameter optimization (Figure 4A). This is mirrored in the PK parameters determined for the basic *iv* and *po* PBPK models that are summarized in Table 3 and compared to the experimentally determined values from the clinical studies. The model PK parameters after *iv* dosing are within the error range of the experimental values, whereas area under the curve (AUC) values after *po* dosing of pravastatin are significantly overestimated. Excretion to urine was predicted rather well. The amount excreted to urine (*Ae*) predicted by this basic PBPK model was 36.0 % and 7.1 % after *iv* and *po* administration, respectively. Experimental values reported from

clinical studies for urinary pravastatin contents were 47 % and 7.3 %, respectively. Bioavailability, however, was clearly overestimated with a predicted value of 88.6 % (experimental: 19.1 %) which in summary leads to the conclusion that the overall model structure of this basic PBPK model is inappropriate and needs further modifications.

**Extended basic PBPK model.** We next constructed an extended version of the basic pravastatin PBPK model, which additionally takes metabolization in the gastro-intestinal tract and hepatic uptake by OATP1B1 into account (see Material and Methods for detail). 1000 repetitive independent optimizations yielded a PBPK model with a minimum simulation error, which is significantly lower than that of the previously fitted basic PBPK model structure (Figure 3). In accordance with this improvement in simulation error, predicted plasma concentration-time profiles obtained for this extended basic PBPK model were somewhat closer to the experimental data compared to the basic PBPK model (Figure 4A). The extended basic PBPK model, however, is still not matching the experimental values. It evidently overestimates exposure after *po* dosing, and overestimates the bioavailability (simulated: 92%, experimentally determined: 19.1 % (Singhvi et al., 1990)). Only the amount excreted to urine seems to match the experimental values (Table 3), which cannot be considered as meaningful taking into account the obviously poor performance of this PBPK model. In consequence, additional model modifications integrating further biochemical and physiological knowledge on pravastatin PK had to be considered in order to obtain a PBPK model able to fully describe the pharmacokinetics of this drug.

**PBPK model with measured expression profiles.** In order to capture the physiological processes underlying pravastatin PK at a high mechanistic level of detail we extended the basic PBPK models by the enzymatic metabolization and active transport processes depicted in Figure 2. Pravastatin is transported not only by OATP1B1 but also by the multidrug resistance-

associated protein 2 (MRP2) and the organic anion transporter 3 (OAT3) (Kivisto and Niemi, 2007). Additionally, enzymatic degradation catalyzed by sulfotransferases (SULTs) in intestinal epithelium, liver and kidney occurs which leads to the formation of inactive isomers in the presence of the cofactor adenosine-3′-phosphate 5′-phosphosulfate (PAPS) (Hatanaka, 2000). Following the new modelling concept outlined in the Material and Methods part, expression data from the newly established customized database were used to parameterize protein abundance in selected organs of known pharmacokinetic relevance for three independent models (Array PBPK model, RT-PCR model, EST PBPK model). The relative gene expression values retrieved from our customized database and integrated in the three PBPK model variants are summarized in Table 4. Since it is not yet known which SULT isoform is responsible for pravastatin metabolization, relative expression data from the three resources reported for SULT1A1, SULT1A2, SULT1A3, SULT1B1, SULT1E1, and SULT2A1 were summed up in the relevant organs (Table 4). For each gene, expression values for different organs are scaled to a maximum value of 1. For OAT3, maximum expression was observed in kidney, OATP1B1, MRP2 and SULTs were strongest expressed in the liver. Following the novel concept for integration of relative gene expression data into PBPK models intrinsic $V_{max}$ values are represented by the product of the global value $k^*_{cat}$ (Table 5) and the expression value specific for each organ (Table 4). These intrinsic $V_{max}$ values represent the actual determinants for transport or metabolizaton rates occurring at a given organ site.

Transport by OATP1B1, MRP2 and OAT3 is implemented as Michaelis-Menten kinetics with the $K_m$ values set to the literature value of 11.5 µM (Nakai et al., 2001), 223 µM (Yamazaki et al., 1997), and 27.2 µM (Nakagomi-Hagihara et al., 2007), respectively. Metabolization catalyzed by

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



SULTs was incorporated as first order clearance ($CL_{SULT}$) in intestine, liver and kidney in the PBPK model since no experimental $K_m$ value was available.

Incorporation of these in-depth structural modifications in the pravastatin PBPK model thus lead to a significant decrease in simulation error (Figure 3). The simulated plasma concentration-time curves obtained for the three PBPK models containing relative expression data and the PK parameters estimated from these plots showed very good accordance with experimental data (Figure 4, Table 3). The simulation curves obtained from all three PBPK models for intravenous application of pravastatin slightly underestimate total clearance of the drug. This might be due to the fact, that experimental data were only available up to 3 h after injection and drug clearance had to be estimated basically on plasma concentration values for oral application, which were measured until 12 h after oral dosing. The simulated curves obtained after oral application of pravastatin show a clearly better agreement with experimentally measured plasma concentration-time profiles. All three PBPK models lead to a slight underestimation of AUC both after intravenous and oral application of pravastatin, and consequently of $C_{max}$ after oral dosing, which is, however, still in the experimental range. Although the plasma concentration-time curves predicted by the three PBPK models integrating relative gene expression data are almost undistinguishable, values for the fraction excreted to urine after *iv* and *po* administration show slight differences. The EST PBPK model (Figure 4F) showed best prediction of urinary secretion (48 % and 6.2 % after inravenous and oral dose, respectively, compared to 41-47 % (Singhvi et al., 1990) and 7.3 % (Mwinyi et al., 2004)). The Array and the RT-PCR PBPK models, however, somewhat underestimated the fraction excreted to urine, especially after intravenous application. Bioavailability of pravastatin was 19.1 % as reported by Singhvi et al (Singhvi et al., 1990) and

was well predicted by all three PBPK models (Array PBPK model: 23.0 %, RT-PCR PBPK model: 22.3 %, EST PBPK model: 19.6 %).

The decreased simulation error in the PBPK model with measured expression profiles indicates a relevant role of gut wall metabolism and efflux transport in the pharmacokinetic behavior of pravastatin. These processes lack in the extended basic model, which might explain the overestimation of the bioavailability in th is model variant due to a weak description of pravastatin pharmacokinetics after oral administration in this case. While consideration of OAT3 and OATP1B1 appears to be sufficient in the case of intravenous administration alone, we found gut wall metabolism and efflux transport, respectively, to be necessary in order to decrease the overall simulation error for both forms of pravastatin administration.

**PBPK models with randomized expression profiles.** It cannot be excluded, that the decrease in simulation error in the PBPK models including gene expression profiles is due to an increase in the systemic degree of freedom by consideration of additional parameters rather than to an inclusion of physiological information, leading to the increased model performance.

To analyze this possibility, we generated 15 P BPK models integrating randomly generated expression profiles. These random expression profiles were drawn from a log normal distribution with the estimated mean and standard deviation of the set of genes and tissues under consideration. Free parameters of each of the 15 models were independently optimized, which again underwent 1000 optimisation runs. The performance of the models based on experimentally measured gene expression profiles was compared to the models including random profiles. For statistical purposes the minimal error among all 1000 optimisation runs was recorded for ea ch model. The mean minimal errors of a ll 15 PBPK models including random profiles was significantly higher ($p<0.05$) than the mean of the minimal error of the three PBPK models based

on experimentally measured gene expression profiles (Figure 3). Table 5 lists optimized PBPK model parameters for the best and worst random model in comparison to the physiological models. The parameters are very different indicating the necessity of a potentially non-physiological distortion of the PBPK model with questionable predictivity of such models.

Note that independent simulated annealing optimizations lead to different sets of parameters for a specific expression profile. Figure 3 indicates the range of simulation errors by dashed lines over optimisations. The quality of models depends on the model structure and expression profiles, and less on the parameter configuration. As can be seen in Figure 3, the best performing random PBPK model had a minimal error which was in the range of the PBPK models parameterized with experimental expression values. The PK curves and PK parameters obtained for this model also showed a good agreement with experimental values (Figure 4B, Table 3) even though a completely different relative expression profile was used in this random PBPK model. The PK curves and calculated PK parameters obtained for the worst performing random PBPK model, however, demonstrate a strong deviation from the experimental ones (Figure 4C, Table 3). Exposure, reflected by the AUC values in Table 3, excretion to urine and bioavailability are significantly underestimated. Despite a larger number of free systemic parameters compared to the basic and basic extended PBPK models, it was not possible to identify a parameter set for the worst random PBPK model that leads to a sufficient model performance. In the worst case, an unfavorable random expression profile could lead to a complete negligence of the active processes such that the simulation error of the basic model represents an upper bound in that case.

## *Discussion*

Passive mechanisms such as diffusion or membrane permeation are generally well described in PBPK models by simple consideration of a substance's physicochemical properties (Willmann et al., 2005). Active, protein-mediated processes, however, are much more difficult to capture in a rigorous quantitative sense. This is because experimental assessment of the governing molecular kinetics is almost impossible in an actual *in vivo* environment. We describe here a novel approach that considers the expression of genes in various tissues and organs in the human body. Taking this publicly available data as a proxy for protein abundance, active processes in PBPK models in the form of Michaelis-Menten processes at multiple organ sites all over the body can be quantified and allow a truly mechanistic description of the mechanisms driving drug pharmacokinetics.

For modeling of active processes, an organ specific $V_{max}$ is required for each PK-relevant protein in order to quantify catalytic efficiency. $V_{max}$ can be obtained by multiplying the global rate constant $k_{cat}^{*}$, which is organ-unspecific, and organ-specific protein abundance which can be approximated from relative gene expression. Most importantly, $k_{cat}^{*}$ implicitly takes into account translational efficacy and posttranscriptional modifications of the same protein thus ensuring relative comparability across multiple organs. Hence, $k_{cat}^{*}$ is a composite parameter of translational efficacy and the enzymatic turn over number. As such, it cannot be measured directly *in situ* or extrapolated from *in vitro* measurements but needs to be identified computationally. It is interesting to note that $k_{cat}^{*}$ of OATP1B1 varies most between the different models which however reflects the resulting simulation error. Also, the enzyme and transporter

activities (multiplication of expression profiles, table 4, and corresponding $k_{cat}^{*}$ values, table 5) show a very high correlation. Values of $k_{cat}^{*}$, which are given in table 5, represent the overall best solution out of thousand independent stochastic optimizations which were performed for each model alternative and which may reflect concurrent and partly synergistic, yet non-linear processes. Hence, even though $k_{cat}^{*}$ values quantifying hepatic uptake by OATP1 may be considerable lower for the PBPK model with expression profiles from ArrayExpress then for the two other model alternatives this observation may well be abrogated by additional mechanisms. Possible explanations could for example be a reduced metabolisation in the gut wall or an decreased OAT3-mediated uptake in the kidney. While a formal validation of the parameters identified is not possible, the significant decrease in simulation error for the three PBPKs models with measured expression profiles in comparison to the PBPK models with random expression profiles shows the relevance of using measured expression data for PK simulations in the case of pravastatin (see dashed line for range of simulation error in Figure 3).

The use of relative expression data to predict *in vivo* activity of enzymes and transporters is not common in PBPK modeling. One study used fractional expression values of P-gp in heart and brain compared to intestine for calculating intrinsic tissue clearances (Shirasaka et al., 2008). Another study mentioned the use of expression data gene expression levels of transporter proteins and metabolic enzymes in stomach, kidney, and small intestine for building a pravastatin PBPK model (Das et al., 2008). However, both studies remain rather unspecific in the applied workflow and moreover lack a systematic approach for the general inclusion of expression data in PBPK models.

The general approaches typically used to predict *in vivo* activities of enzymes or transporters are e.g. allometric scaling, *in vitro-in vivo* extrapolation (IVIVE)/*in vitro-in vivo* correlation (IVIVC) of clearance or transporter activity, and physiology-based direct scaling (Keldenich, 2004; Ekins et al., 2007; Pelkonen and Turpeinen, 2007), with the latter two being the most cost-effective and accurate approaches (Zuegge et al., 2001). In the case of the IVIVE/IVIVC scaling method experimental $K_m$ and $V_{max}$ values or intrinsic clearances obtained from *in vitro* assays (i.e. performed with tissue homogenates, microsomal fractions, cell lines, or recombinant proteins) are scaled using factors accounting for the average number of cells per gram tissue or mg protein per gram tissue, respectively. This method mostly offers good estimates of protein activities, if *in vitro* data to the protein of interest in a specific organ is available. A major drawback of this approach is, however, that only for a limited amount of proteins such experimental data has been measured so far and that *in vitro* clearances or transporter activities are available only for liver as the major metabolizing organ or intestine as the essential organ limiting oral drug bioavailability (Bruyere et al., 2010; Fan et al., 2010; Obach, 2011). The great advantage of the PBPK modeling approach integrating relative gene expression data from all organs of the human body offers the possibility to estimate *in vivo* activity of enzymes and transporters in each organ. This new concept combines the advantages of physiological scaling techniques with the high accuracy of mRNA profiling techniques such as RT-PCR, Microarray methodologies and EST methods and leads to PBPK models integrating an in-depth structural level of physiological detail and being easy to adjust thanks to the global fit parameters defined for each protein-mediated process $k_{cat}^{*}$. Therefore, this new approach presents an innovative possibility to predict *in vivo* activities in all organs in a systematic, comprehensible and reproducible way under consideration of whole-body physiology information even if no *in vitro* data are at hand at all.

In the case of pravastatin we could show, that simple PBPK modeling approaches are not sufficient to account for the experimentally observed pharmacokinetics. Both the basic and the extended basic PBPK model could not describe the experimental observations with a sufficient level of accuracy (Figure 4). Only the consideration of all relevant organs expressing the proteins involved in metabolization and transport of pravastatin lead to PBPK models with high model performance and a correct description of pravastatin pharmacokinetics and mass balance.

The PBPK model of pravastatin described in this work is to our knowledge the only one reported so far, that accounts for all physiologically relevant protein mediated distribution and clearance processes reported for this compound. It considers transporters such as OATP1B1, MRP2 and OAT3 and sulfotransferases (Hatanaka, 2000; Kivisto and Niemi, 2007), integrates information to the cellular localization of these proteins (intracellular, basolateral, apical) and represents the relative abundance of these proteins in the following tissues: enterocytes, hepatocytes or kidney tubular cells. A previously established PBPK model aiming at predicting transporter-mediated clearance and distribution of pravastatin in humans (Watanabe et al., 2009) was rather simple compared to the one presented here. The PBPK model structure only considered six organ compartments (intestine, liver, lung, kidney, brain and muscle) with liver being divided into five units of extracellular and subcellular compartments in order to best describe experimentally determined hepatic availability. Hepatic uptake and canalicular efflux activity *in vivo* was estimated using the above describes IVIVE/IVIVC methodology. However, transporter activity in other tissues than liver was not considered.

The modeling results presented in this study clearly indicate that the use of expression profiles is beneficial for PBPK modeling because the resulting PBPK models show a superior model performance and the resulting models are more accurate and robust in the simulations. For

therapeutic substances that are actively transported or metabolized, such as pravastatin, basic PBPK models are challenged to reproduce and match experimental data. Notably, a comparison of simple PBPK models with PBPK models incorporating experimental expression data showed a significantly worse performance, underlining the general benefit of using as much experimental information during model building as possible. In addition, a comparison with random profile models shows that the low simulation error of expression profile PBPK models is not due to the number of free parameters. The minimal error obtained for the 15 random PBPK models ranges from that obtained for the PBPK models integrating experimental expression data to those determined for the basic PBPK model, a variability which is due to the diversity of random parameters and not explained by the number of free parameters. Only the best random profile model is on par with the experimental profile models, while those latter are very consistently showing very low simulation errors. The particular choice of experimental platform for gene expression profiling is of secondary importance since models based on different data base sources show similar low simulation errors. In addition, the range of simulation errors as achieved in independent parameter optimizations is small compared to the effect of using gene expression profiles versus basic PBPK modeling.

In conclusion, we believe, that usage of physiological information from various levels of biological organization ranging from gene expression to organ and whole-body scale has beneficial effects on the development of truly predictive computational models, which are needed to gain a comprehensive understanding of the processes governing drug distribution and drug action. Additionally, such models also offer a unique possibility for further mechanistic analyses such as investigations of drug-drug interactions occurring simultaneously in different tissues.

## *Acknowledgements*

We thank Deepti Chandraprakash and Christoph Niederalt for helpful discussions and comments.

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



## *Authorship contributions*

Participated in research design: Meyer, Schneckener, Ludewig, Kuepfer, Lippert

Conducted experiments: Meyer, Schneckener, Ludewig, Kuepfer

Contributed new reagents or analytic tools: -

Performed data analysis: Meyer, Schneckener, Ludewig, Kuepfer

Wrote or contributed to the writing of the manuscript: Meyer, Schneckener, Kuepfer, Lippert

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



## REFERENCES

ArrayExpress (2010) European Informatics Institute (EBI). http://www.ebi.ac.uk/microarray-as/ae/.

Bruyere A, Decleves X, Bouzom F, Ball K, Marques C, Treton X, Pocard M, Valleur P, Bouhnik Y, Panis Y, Scherrmann JM and Mouly S (2010) Effect of Variations in the Amounts of P-Glycoprotein (ABCB1), BCRP (ABCG2) and CY P3A4 along the Human Small Intestine on PBPK Models for Predicting Intestinal First Pass. *Mol Pharm.*

Das D, Dhurjati P and Wangikar P (2008) Prediction of pharmacokinetic behaviour by combining in vivo and in vitro data in physiologically based pharmacokinetic (PBPK) model: Parameter estimation and sensitivity analysis. *Jl of IISc* **88:**57-71.

Edginton AN, Schmitt W and W illmann S (2006) Development and evaluation of a g eneric physiologically based pharmacokinetic model for children. *Clin Pharmacokinet* **45:**1013-1034.

Eissing T, Kuepfer, L, Becker, C, Block, M, Coboeken, K, Gaub, T, Goerlitz, L, Jaeger, J, Loosen, R, Ludewig, B, Meyer, M, Niederalt, C, Sevestre, M, Siegmund, H-U, Solodenko, J, Thelen, K, Telle, U, Weiss, W, Wendl, T, Willmann, S, Lippert, J (2011) A computational systems biology software platform for multiscale modeling and simulation: integrating whole-body physiology, disease biology, and m olecular reaction networks. *Front Physio* **2**.

Ekins S, Ecker GF, Chiba P and Swaan PW (2007) Future directions for drug transporter modelling. *Xenobiotica* **37:**1152-1170.

Everett DW, Chando TJ, Didonato GC, Singhvi SM, Pan HY and Weinstein SH (1991) Biotransformation of pravastatin sodium in humans. *Drug Metab Dispos* **19:**740-748.

Fan J, Chen S, Chow EC and Pang KS (2010) PBPK modeling of intestinal and liver enzymes and transporters in drug absorption and sequential metabolism. *Curr Drug Metab* **11:**743-761.

Gabrielsson J, Green AR and Van der Graaf PH (2010) Optimising in vivo pharmacology studies--Practical PKPD considerations. *J Pharmacol Toxicol Methods* **61:**146-156.

Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM and Aburatani H (2005) Interpreting expression profiles of cancers by genome-wide survey of b readth of expression in normal tissues. *Genomics* **86:**127-141.

Hatanaka T (2000) Clinical pharmacokinetics of pravastatin: mechanisms of pharmacokinetic events. *Clin Pharmacokinet* **39:**397-412.

Hubbell E, Liu WM and Mei R (2002) Robust estimators for expression analysis. *Bioinformatics* **18:**1585-1592.

Keldenich J (2004) Prediction of human clearance (CL) and volume of distribution (VD). *Drug Discov Today Technol* **1:**389-395.

Kivisto KT and N iemi M (2007) Influence of drug transporter polymorphisms on pravastatin pharmacokinetics in humans. *Pharm Res* **24:**239-247.

Maglott D, Ostell J, Pruitt KD and Tatusova T (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **35:**D26-31.

Moles CG, Mendes P and Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res* **13:**2467-2474.

Mwinyi J, Johne A, Bauer S, Roots I and Gerloff T (2004) Evidence for inverse effects of OATP-C (SLC21A6) 5 and 1b haplotypes on pravastatin kinetics. *Clin Pharmacol Ther* **75:**415-421.

Nakagomi-Hagihara R, Nakai D and Tokui T (2007) Inhibition of human organic anion transporter 3 m ediated pravastatin transport by gemfibrozil and the metabolites in humans. *Xenobiotica* **37:**416-426.

Nakai D, Nakagomi R, Furuta Y, Tokui T, Abe T, Ikeda T and Nishimura K (2001) Human liver-specific organic anion transporter, LST-1, mediates uptake of pravastatin by human hepatocytes. *J Pharmacol Exp Ther* **297:**861-867.

Nestorov I (2007) Whole-body physiologically based pharmacokinetic models. *Expert Opin Drug Metab Toxicol* **3:**235-249.

Niemi M, Pasanen MK and Neuvonen PJ (2006) SLCO1B1 polymorphism and sex affect the pharmacokinetics of pravastatin but not fluvastatin. *Clin Pharmacol Ther* **80:**356-366.

Nishimura M and Naito S (2005) Tissue-specific mRNA expression profiles of human ATP-binding cassette and solute carrier transporter superfamilies. *Drug Metab Pharmacokinet* **20:**452-477.

Nishimura M and Naito S (2006) Tissue-specific mRNA expression profiles of human phase I metabolizing enzymes except for cytochrome P450 and phase II metabolizing enzymes. *Drug Metab Pharmacokinet* **21:**357-374.

Nishimura M, Yaguti H, Yoshitsugu H, Naito S and Satoh T (2003) Tissue distribution of mRNA expression of human cytochrome P450 isoforms assessed by high-sensitivity real-time reverse transcription PCR. *Yakugaku Zasshi* **123:**369-375.

Obach RS (2011) Predicting clearance in humans from in v itro data. *Curr Top Med Chem* **11:**334-339.

Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U and Brazma A (2007) ArrayExpress--a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* **35:**D747-750.

Pelkonen O a nd Turpeinen M (2007) In vitro-in vivo extrapolation of hepatic clearance: biological tools, scaling factors, model assumptions and correct concentrations. *Xenobiotica* **37:**1066-1089.

Poulin P a nd Theil FP (2002a) Prediction of ph armacokinetics prior to in vivo studies. 1. Mechanism-based prediction of volume of distribution. *J Pharm Sci* **91:**129-156.

Poulin P and Theil FP (2002b) Prediction of pharmacokinetics prior to in vivo studies. II. Generic physiologically based pharmacokinetic models of drug disposition. *J Pharm Sci* **91:**1358-1370.

R Development Core Team (2010): A Language and E nvironment for Statistical Computing. http://www.R-project.org.

Rodgers T, Leahy D and Rowland M (2005) Physiologically based pharmacokinetic modeling 1: predicting the tissue distribution of moderate-to-strong bases. *J Pharm Sci* **94:**1259-1276.

Rodgers T and Ro wland M ( 2006) Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *J Pharm Sci* **95:**1238-1257.

Serajuddin AT, Ranadive SA and Mahoney EM (1991) Relative lipophilicities, solubilities, and structure-pharmacological considerations of 3-hydroxy-3-methylglutaryl-coenzyme A

(HMG-CoA) reductase inhibitors pravastatin, lovastatin, mevastatin, and simvastatin. *J Pharm Sci* **80:**830-834.

Shirasaka Y, Sakane T and Yamashita S (2008) Effect of P-glycoprotein expression levels on the concentration-dependent permeability of drugs to the cell membrane. *J Pharm Sci* **97:**553-565.

Singhvi SM, Pan HY, Morrison RA and Willard DA (1990) Disposition of pravastatin sodium, a tissue-selective HMG-CoA reductase inhibitor, in healthy subjects. *Br J Clin Pharmacol* **29:**239-243.

Unigene (2010) National Center for Biotechnology Information (NCBI). www.ncbi.nlm.nih.gov/unigene.

Vicini P (2010) Multiscale modeling in drug discovery and development: future opportunities and present challenges. *Clin Pharmacol Ther* **88:**126-129.

Vossen M, Sevestre M, Niederalt C, Jang IJ, Willmann S and Edginton AN (2007) Dynamically simulating the interaction of midazolam and the CYP3A4 inhibitor itraconazole using individual coupled whole-body physiologically-based pharmacokinetic (WB-PBPK) models. *Theor Biol Med Model* **4:**13.

Watanabe T, Kusuhara H, Maeda K, Shitara Y and Sugiyama Y (2009) Physiologically based pharmacokinetic modeling to predict transporter-mediated clearance and distribution of pravastatin in humans. *J Pharmacol Exp Ther* **328:**652-662.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, DiCuccio M, Edgar R, Federhen S, Helmberg W, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pontius JU, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L and Yaschenko E (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **33:**D39-45.

Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA and Wagner L (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res* **31:**28-33.

Willmann S, Hohn K, Edginton A, Sevestre M, Solodenko J, Weiss W, Lippert J and Schmitt W (2007) Development of a physiology-based whole-body population model for assessing the influence of individual variability on the pharmacokinetics of drugs. *J. Pharmacokinet. Pharmacodyn.* **34:**401-431.

Willmann S, Lippert J and Schmitt W (2005) From physicochemistry to absorption and distribution: predictive mechanistic modelling and computational tools. *Expert Opin Drug Metab Toxicol* **1:**159-168.

Willmann S, Lippert J, Sevestre M, Solodenko J, Fois F and Schmitt W (2003) PK-Sim©: a physiologically based pharmacokinetic 'whole-body' model. *Biosilico* **1:**121-124.

Willmann S, Schmitt W, Keldenich J, Lippert J and Dressman JB (2004) A physiological model for the estimation of the fraction dose absorbed in humans. *J Med Chem* **47:**4022-4031.

Willmann S, Thelen K, Becker C, Dressman JB and Lippert J (2010) Mechanism-based prediction of particle size-dependent dissolution and absorption: cilostazol pharmacokinetics in dogs. *Eur J Pharm Biopharm* **76:**83-94.

Wilson CL and Miller CJ (2005) Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics* **21:**3683-3685.

Yamazaki M, Akiyama S, Ni'inuma K, Nishigaki R and Sugiyama Y (1997) Biliary excretion of pravastatin in r ats: contribution of the excretion pathway mediated by canalicular multispecific organic anion transporter. *Drug Metab Dispos* **25:**1123-1129.

Zuegge J, Schneider G, Coassolo P and Lave T (2001) Prediction of hepatic metabolic clearance: comparison and assessment of prediction models. *Clin Pharmacokinet* **40:**553-563.

## *FOOTNOTES*

Michaela Meyer and Sebastian Schneckener contributed equally to this work.

The authors acknowledge financial support by the German Federal Ministry of Education and Research for the grants [0313853B] (QuantPro Initiative), [0315280F] (FORSYS-Partner project 'A Systems Biology Approach towards Predictive Cancer Therapy'), and [0315747] (VirtualLiver).

## *FIGURES*

**Figure 1:** Overview of the steps involved in building a customized expression database and validation of the novel approach for integration of relative expression data into PBPK models.

**Figure 2:** Enzymes and transporters involved in metabolization and distribution of pravastatin which were integrated in the various PBPK models.

**Figure 3:** Comparison of minimal simulation errors of the different PBPK model variants: basic PBPK model (orange) that only considers first order clearance processes in kidney and liver, extended basic PBPK model (blue) that additionally considers degradation in intestinal lumen and hepatic uptake by OATP1B1, PBPK models integrating experimental expression data from three different resources (green), and fifteen PBPK models integrating random expression values (grey). Vertical bars indicate the minimal simulation error, while a horizontal dashed line shows the range of simulation errors for independent optimisations.

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



**Figure 4:** Comparison of simulated plasma concentration-time profiles from the different PBPK model variants of pravastatin after intravenous (coloured solid lines) and oral (coloured dotted lines) administration with experimental IV data (●) (Singhvi et al., 1990) and PO data (○) (Mwinyi et al., 2004): A) Basic PBPK model (orange) and extended PBPK model (blue), B) PBPK model integrating random expression values that showed smallest simulation error (Best random model from Figure 3, C) PBPK model integrating random expression values that showed largest simulation error ( Worst random model Figure 3), D-F) PBPK models integrating experimental expression data from three different resources: Microarray data from ArrayExpress (D) (ArrayExpress, 2010), RT-PCR data from literature (E) (Nishimura et al., 2003; Nishimura and Naito, 2005; Nishimura and Naito, 2006), and EST data from UniGene (F) .

## *TABLES*

**Table 1:** Physicochemical properties of pravastatin integrated in all PBPK model variants.

| Parameter | Value |
|---|---|
| Molecular weight [g/mol] | 424.53 |
| Fraction unbound (fu) | 0.50 (0.46-0.57 (Hatanaka, 2000)) |
| Solubility at pH 5 [mg/l] | 180 (Serajuddin et al., 1991) |
| $pK_a$ | 4.56 (Serajuddin et al., 1991) |

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



**Table 2**: Mean Pearson correlation between genes from two data sources across more than two tissues. In brackets, percentage of genes with correlation $p<0.01$.

| | EST / RT-PCR | EST /Array | RT-PCR/Array |
|---|---|---|---|
| Gene set used for PBPK model (10 genes) | 0.40 (14 %) | 0.42 (17 %) | 0.87 (100 %) |
| Common gene set (214 genes represented in more than 2 tissues) | 0.21 (10 %) | 0.27 (14 %) | 0.66 (64 %) |

Genes used for PBPK models: CYP3A4, SLCO1B1, SLC22A8, ABCC2, SULT1A1, SULT1A2, SULT1A3, SULT1B1, SULT2A1, and SULT1E1

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



**Table 3:** Pharmacokinetic parameters (AUC: area under the curve of plasma concentration, $A_e$: amount excreted to urine, $C_{max}$: maximum concentration, $t_{max}$: timepoint of peak concentration) after intravenous and oral application of pravastatin (9.9 and 40 mg, respectively) obtained for the different PBPK model variants in comparison to literature values (Singhvi et al., 1990; Mwinyi et al., 2004).

| Pharmacokinetic parameters | PBPK model variant | | | | | | | Reference value |
|---|---|---|---|---|---|---|---|---|
| | Basic | Extended basic | Array | RT-PCR | EST | Best Random | Worst Random | |
| | *Intravenous dose of 9.9 mg* | | | | | | | |
| $AUC_{0-\infty}$ [µmol/L·h] | 0.35 | 0.24 | 0.24 | 0.24 | 0.31 | 0.28 | 0.29 | 0.40 ± 0.07 [1] |
| $Ae_{0-48h}$ [%] | 36.0 | 50 | 27 | 35 | 48 | 22 | 1.4 | 47 [1] |
| | *Oral dose of 40 mg* | | | | | | | |
| $AUC_{0-6h}$[µmol/L·h] | 1.02 | 0.73 | 0.18 | 0.18 | 0.20 | 0.26 | 0.08 | 0.27 ± 0.16 [2] |
| $C_{max}$ [µmol/L] | 0.64 | 0.49 | 0.08 | 0.09 | 0.11 | 0.10 | 0.05 | 0.14 ± 0.09 [2] |
| $t_{max}$ [h] | 0.68 | 0.64 | 0.72 | 0.80 | 0.82 | 1.04 | 0.48 | 1.0 (0.8-2.0) [2] |
| $Ae_{0-12h}$ [%] | 7.1 | 6.9 | 4.1 | 4.2 | 6.2 | 4.1 | 0.1 | 7.3 ± 3.9 [2] |
| Bioavailiability [%] | 88.6 | 92.0 | 23.0 | 22.3 | 19.6 | 26.6 | 9.9 | 19.1 [1] |

[1] Singhvi et al., 1990
[2] Mwinyi et al., 2004

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



**Table 4:** Relative gene expression data for OATP1B1, OAT3, MRP2, and sulfotransferases (SULT) used for building three PBPK model variants of pravastatin. Experimental expression values are based on and microarray measurements from ArrayExpress, RT-PCR data from literature, and EST values from UniGene (Nishimura et al., 2003; Nishimura and Naito, 2005; Nishimura and Naito, 2006; ArrayExpress, 2010; Unigene, 2010). For comparison, the random expression values of PBPK model variants with the smallest and largest simulation error, respectively (Best and worst random models) are given.

| Relative expression in different organs | | Experimental gene expression | | | Random gene expression | |
|---|---|---|---|---|---|---|
| | | Array | RT-PCR | EST | Best random | Worst random |
| **OATP1B1** | Liver | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **OAT3** | Liver | 0.021 | $1.7 \cdot 10^{-4}$ | - | 0.243 | 0.046 |
| | Kidney | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **MRP2** | Liver | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Kidney | 0.462 | 0.361 | 0.534 | 54.5 | 0.000 |
| | Small Intestine | 0.391 | 0.131 | 0.876 | 4.366 | 0.071 |
| **SULT** | Liver | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Kidney | 0.252 | 0.086 | 0.290 | 0.005 | 0.414 |
| | Small Intestine | 1.980 | 0.590 | 1.132 | 0.007 | 7.647 |

DMD Fast Forward. Published on January 31, 2012 as DOI: 10.1124/dmd.111.043174
This article has not been copyedited and formatted. The final version may differ from this version.



**Table 5**: PBPK model parameters that were adjusted using literature data to plasma concentration and urinary excretion of pravastatin after intraveneous and oral doses of 9.9 mg and 40 mg, respectively (Singhvi et al., 1990; Mwinyi et al., 2004). Clearance processes ($CL_x$) are given as specific clearance rates [1/min], which can be calculated from intrinsic clearance rates [ml/min] by normalization to the volume of the compartment or organ in which metabolism or degradation takes place.

| Parameter | Basic | Basic extended | Array | RT-PCR | EST | Best random | Worst random |
|---|---|---|---|---|---|---|---|
| Lipophilicity | 1.00 | 1.00 | 0.84 | 0.79 | 0.99 | 0.60 | 1.1 |
| $P_{int}$ [cm/min] | $3.5 \cdot 10^{-3}$ | $3.50 \cdot 10^{-3}$ | $8.3 \cdot 10^{-3}$ | $6.4 \cdot 10^{-3}$ | $7.0 \cdot 10^{-3}$ | $5.9 \cdot 10^{-3}$ | $1.4 \cdot 10^{-2}$ |
| $CL_{hep}$ [1/min] | 22.6 | 22.6 | - | - | - | - | - |
| $CL_{ren}$ [1/min] | $7.4 \cdot 10^{-2}$ | $2.5 \cdot 10^{-1}$ | - | - | - | - | - |
| $CL_{degrad}$ [1/min] | - | 0.020 | 0.010 | 0.010 | 0.014 | 0.020 | 0.010 |
| $k^*_{cat,OATP1B1}$ [µmol/L/min] | - | $4.2 \cdot 10^{2}$ | $6.2 \cdot 10^{-1}$ | $1.1 \cdot 10^{3}$ | $7.0 \cdot 10^{2}$ | $2.1 \cdot 10^{2}$ | $1.6 \cdot 10^{-1}$ |
| $k^*_{cat,OAT3}$ [µmol/L/min] | - | - | $5.8 \cdot 10^{5}$ | $3.4 \cdot 10^{6}$ | $3.4 \cdot 10^{6}$ | $3.4 \cdot 10^{4}$ | $8.7 \cdot 10^{5}$ |
| $k^*_{cat,MRP2}$ [µmol/L/min] | - | - | 2.7 | 1.4 | 2.3 | 0.1 | 145 |
| $k^*_{cat,SULT}$ [µmol/L/min] | - | - | 0.6 | 4.6 | 2.8 | 4.9 | $2.3 \cdot 10^{-4}$ |

Figure 1

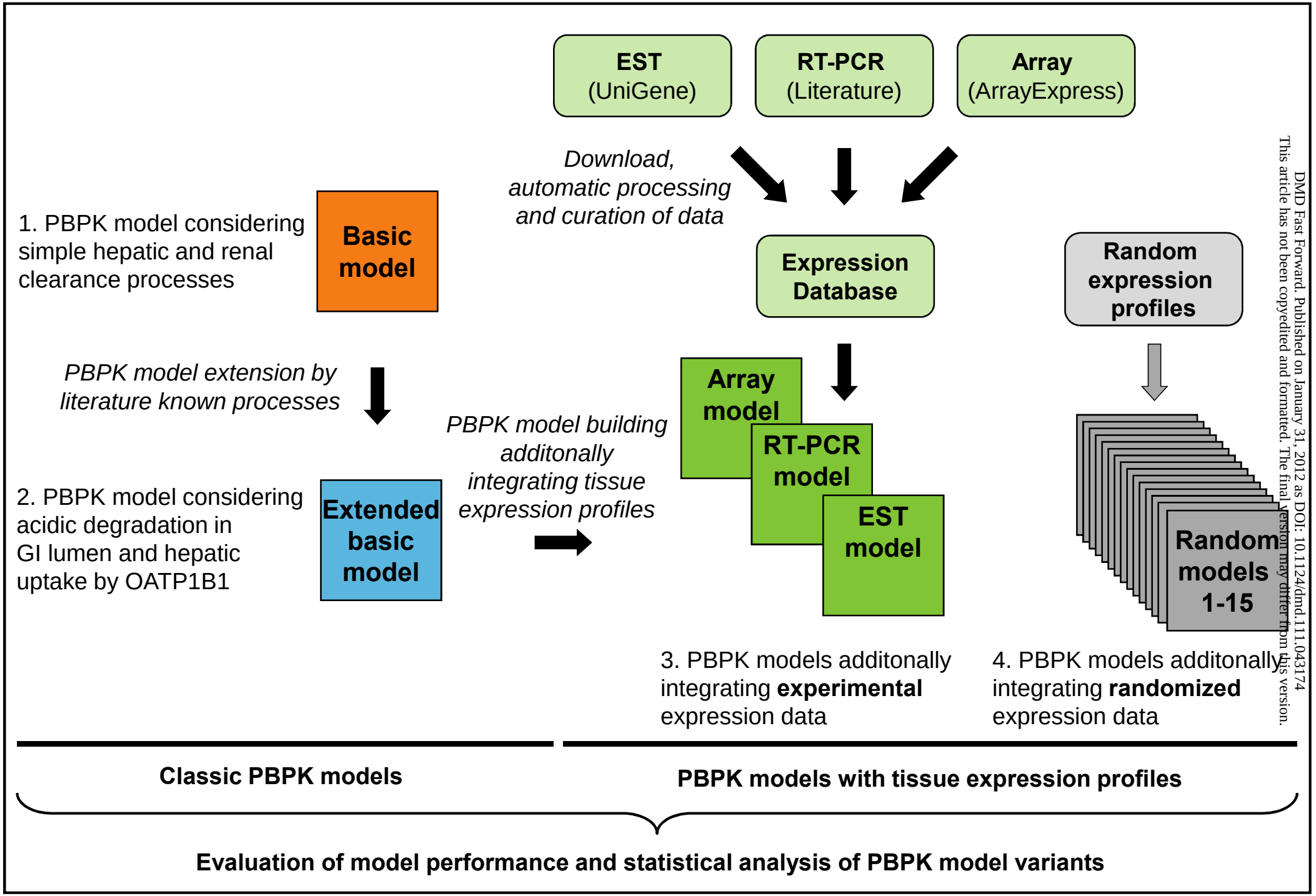

Figure 2

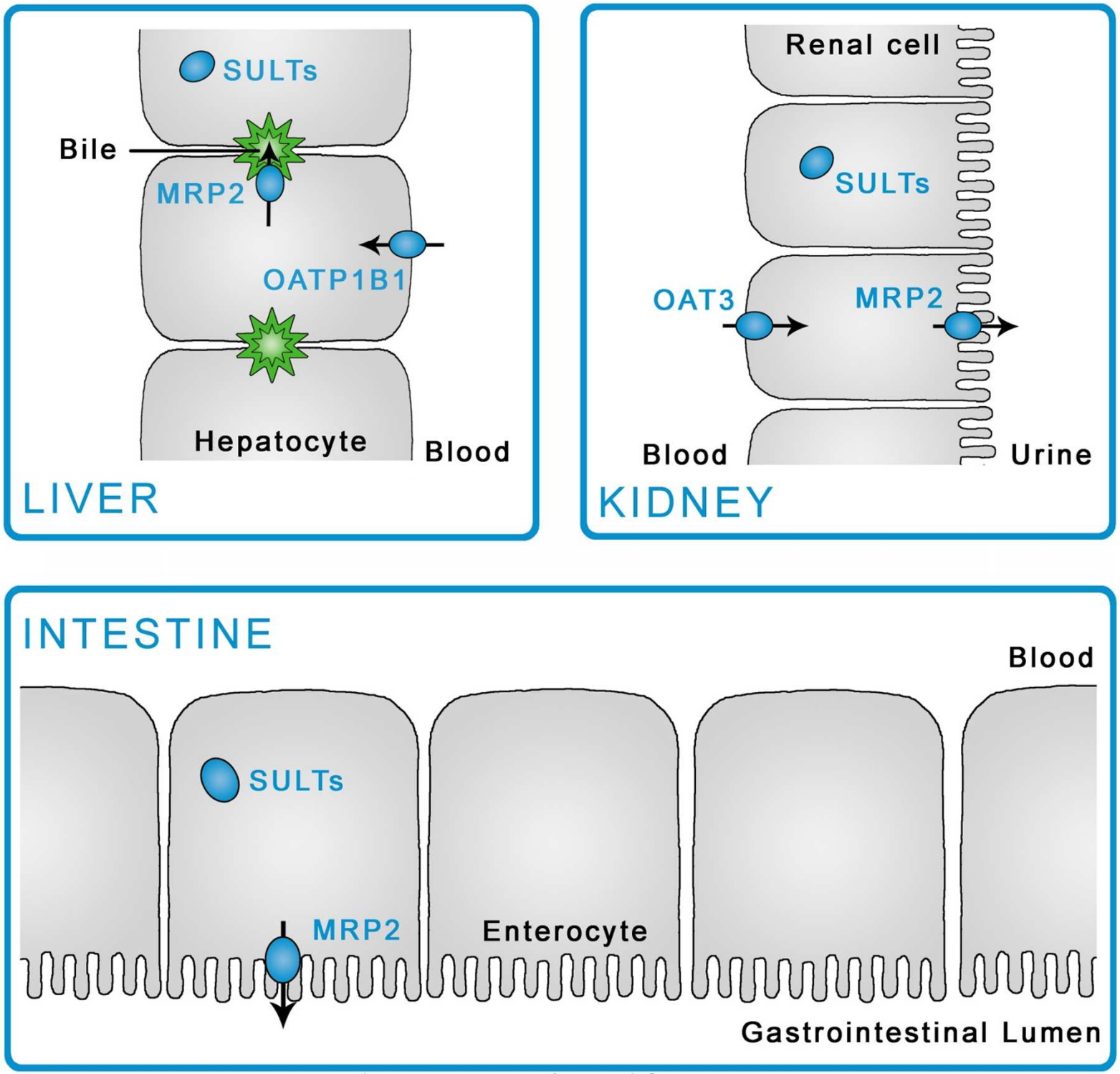

Figure 3

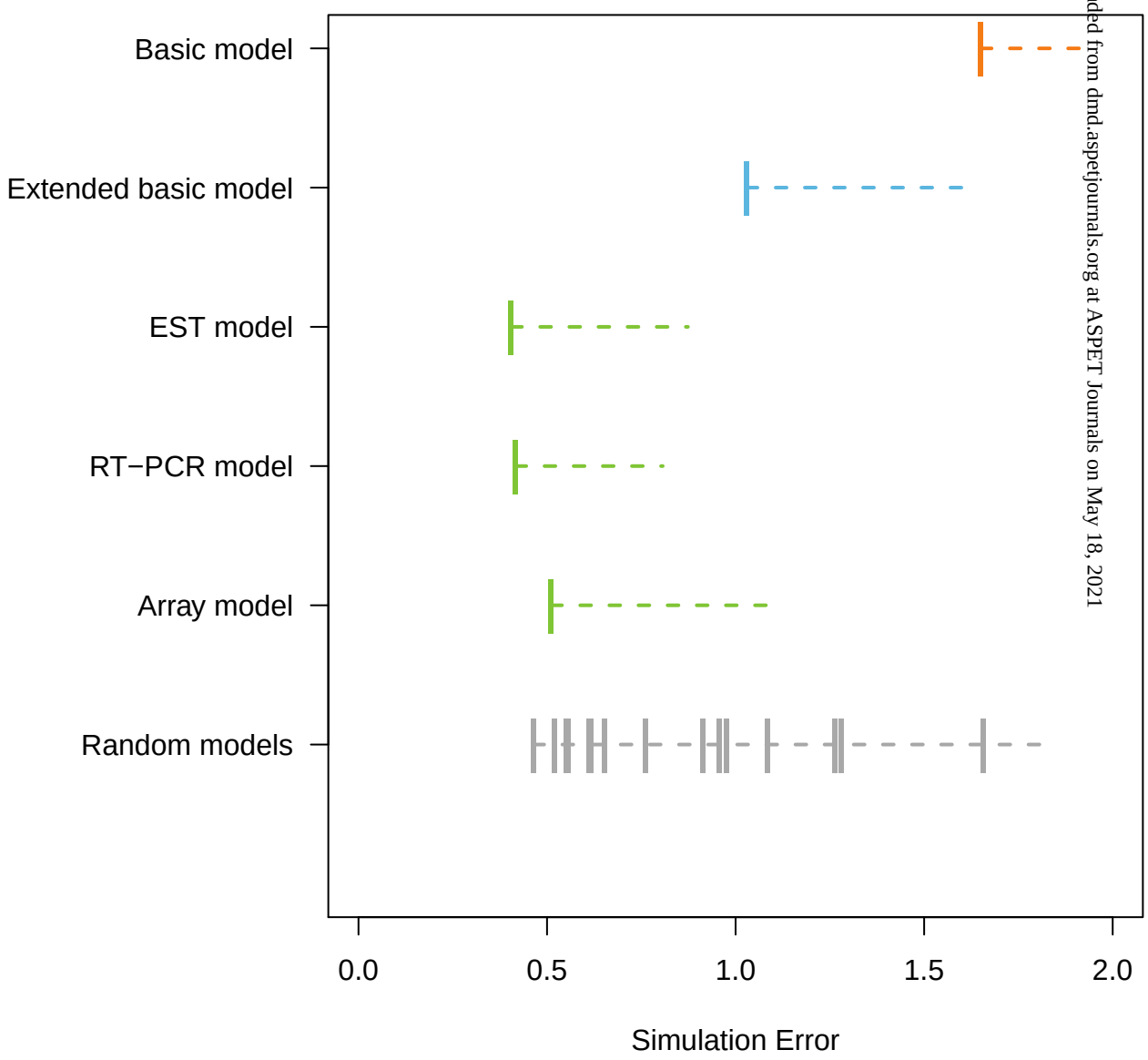

Figure 4