Using Expression Data for Quantification of Active Processes in Physiologically Based Pharmacokinetic Modeling

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

Systems Biology and Computational Solutions, Bayer Technology Services, Leverkusen, Germany

Received October 5, 2011; accepted January 31, 2012

ABSTRACT:
Active processes involved in drug metabolization and distribution mediated by enzymes, transporters, or binding partners mostly occur simultaneously in various organs. However, a quantitative description of active processes is difficult because of 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 that 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 that already represent detailed anatomical and physiological information. The new approach was evaluated using three publicly available databases: whole-genome expression microarrays from ArrayExpress, reverse transcription-polymerase chain reaction-derived gene expression estimates collected from the literature, and expressed sequence tags 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 organic anion-transporting polypeptide 1B1 and organic anion transporter 3, active efflux by multidrug resistance protein 2, and metabolism by sulfotransferases in liver, kidney, and/or intestine. Benchmarking of PBPK models based on gene expression data against alternative models with either a less complex model structure or randomly assigned gene expression values clearly demonstrated the superior model performance of the former. Besides 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 because of the physiological representation of protein-mediated processes.

Introduction
Computational models are playing an increasing role in pharmaceutical research and development, because 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 et al., 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 on the basis of a large amount of prior physiological and biological information (Poulin and Theil, 2002a,b; Willmann et al., 2003; Rodgers et al., 2005; Rodgers and Rowland, 2006; Nestorov, 2007; Eissing et al., 2011). The various prediction methods included in PBPK modeling are based on compound-deduced parameters and quantify absorption, distribution, metabolism, and excretion of a drug (Poulin and Theil, 2002a,b; Willmann et al., 2003, 2004, 2005; Rodgers et al., 2005; Rodgers and Rowland, 2006). Hence, all model parameters either are obtained from collections of literature data or are derived from a few physicochemical properties of a compound such as lipophilicity or molecular weight.

Whereas 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 ($K_D$), catalytic velocity ($k_{cat}$) and tissue-specific protein abundance ($E_{tissue}$). Furthermore, the same active process may simultaneously occur in various types of tissue all over the body, which hampers a rigorous representation in the underlying model structure all the more because binding affinity and catalytic velocity can only be measured in vitro. As of now, representation of active processes in PBPK models is restricted to isolated organs and processes.

In this study, we present a novel approach that 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, whereas catalytic parameters are described by a global

ABBREVIATIONS: PBPK, physiologically based pharmacokinetic; RT-PCR, reverse transcription-polymerase chain reaction; EST, expressed sequence tags; OATP1B1, organic anion-transporting polypeptide 1B1; MRP2, multidrug resistance protein 2; OAT3, organic anion transporter 3; SULT, sulfotransferase; PK, pharmacokinetic; AUC, area under the curve; IVIVE, in vitro-in vivo extrapolation; IVIVC, in vitro-in vivo correlation.
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 on the basis of 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 for 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 be used directly in PBPK model building: whole-genome expression arrays from ArrayExpress (European Informatics Institute, 2010, http://www.ebi.ac.uk/microarray-as/ae/), reverse transcription polymerase chain reaction (RT-PCR)-derived gene expression estimates from the literature (Nishimura et al., 2003; Nishimura and Naito, 2005, 2006), and expressed sequence tags (EST) from UniGene (National Center for Biotechnology Information, 2010, http://www.ncbi.nlm.nih.gov/UniGene). The consolidated data were stored in a database with three sections termed EST (UniGene), Array (ArrayExpress), and RT-PCR (literature), respectively (Fig. 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 because it is known to be a substrate for various transporters [organic anion-transporting polypeptide 1B1 (OATP1B1, SLC01B1), multidrug resistance protein 2 (MRP2, ABC2C), organic anion transporter 3 (OAT3, SLC22A8) and metabolic enzymes (sulfotransferases, SULTs)] that are expressed simultaneously in several organs (Hatanaka, 2000; Kivistö and Niemi, 2007) (Fig. 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 gastrointestinal tract were developed first (Fig. 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 with that of 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 (Fig. 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, 2004, 2005; Eissing et al., 2011). PK-Sim-generated PBPK models were exported and modified in MoBi (version 2.2; Bayer Technology Services GmbH) as described further below. Academic licenses of PK-Sim and MoBi are available free of charge. All optimizations and batch mode simulations were performed using MATLAB (version 7; The MathWorks, Inc., Natick, MA) and the MoBi Toolbox for MATLAB (version 2.0; Bayer Technology Services GmbH). All numerical calculations were performed using R (version 2.8.1, 2010; R Development Core Team, http://www.R-project.org). Text processing was done using the scripting language perl (version 5.10). All data were 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. To identify solutions close to global optima (Moles et al., 2003), 1000 repetitive optimi-

---

**Fig. 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. GI, gastrointestinal.
defining an affinity constant implemented in PK-Sim by using the Michaelis-Menten kinetics equation and matic reactions and transport processes in various tissues of the body are further refined by active, protein-mediated, tissue-specific processes. Enzy-

- scaling factor $SF$ (micromoles per liter) corrects for the absolute in vivo protein concentration such that eq. 2 can be rewritten to:

$$V_{\text{Organ},i}^\text{max} = k_{\text{cat},i} \cdot SF \cdot e_{\text{Ref},i} = k_{\text{cat},i} \cdot e_{\text{Ref},i}$$

(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 $e_{\text{Organ},i}$ is scaled by an apparent catalytic rate constant $k_{\text{cat},i}$ (micromoles per liter per minute). Note that $k_{\text{cat},i}$ is a global model parameter, which is used for calculation of the tissue-specific maximum velocity (see eq. 4) and that SF implicitly considers translational efficacy and post-transcriptional modifications for a particular protein:

In conclusion, by using eq. 4 for prediction of in vivo $V_{\text{max}}$ values instead of the original Michaelis-Menten equation, only one single free parameter, i.e., $k_{\text{cat},i}$, instead of an i-fold consideration of $V_{\text{max}}$ for each organ ($V_{\text{Organ},i}^\text{max}$) 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 by applying this modeling approach using expression data from three different sources: whole-genome expression arrays from ArrayExpress (European Informatics Institute, 2010, http://www.ebi.ac.uk/microarray-as/aaf), RT-PCR-derived gene expression estimates from literature (Nishimura et al., 2003; Nishimura and Naito, 2005, 2006), and EST from UniGene (National Center for Biotechnology Information, 2010, http://www.ncbi.nlm.nih.gov/uni gene), Details for downloading, automatic processing, and curation of data from these resources are described further below.

### Randomized 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 were generated by random drawing from a log normal distribution. The log normal distribution was parameterized with mean and variance of experimentally observed expression data. Expression profile PBPK models as described earlier were used, and gene expression was replaced by random values.

### EST from UniGene

Relevant files from the human section of UniGene (Wheeler et al., 2003) were downloaded from the UniGene site. In a semia-

### TABLE 1

**Physicochemical properties of pravastatin integrated in all PBPK model variants**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular weight, g/mol</td>
<td>424.53</td>
</tr>
<tr>
<td>Fraction unbound</td>
<td>0.50 (0.46–0.57)$^a$</td>
</tr>
<tr>
<td>Solubility at pH 5, mg/L</td>
<td>180$^b$</td>
</tr>
<tr>
<td>$pK_a$</td>
<td>4.56$^c$</td>
</tr>
</tbody>
</table>

$a$ Data from Hatanaka (2000).
$^b$ Data from Serajuddin et al. (1991).
established database also contains metadata such as a dictionary of
proteins and gene synonyms and mappings for organ and tissue names,
which was retrieved from EntrezGene and UniProt/Swiss-Prot as the
primary sources.

Upon completion, the established customized expression database
usable for PBPK modeling contains 3.4 million measurements from
23,000 different genes in 62 different tissues and organs, respectively.

For each gene name, four 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 or 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 the literature.

Comparison of Expression Database Sections. The established
database has three sections representing different sources and exper-
imental techniques of measuring gene expression. To assess the dif-
ferences 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, whereas
correlations to EST values were insignificant (see Table 2 for mean correla-
tions values). Although 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 prop-
erties 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 (Wollmann et al., 2003, 2004,
2005, 2007, 2010; Edginton et al., 2006; Vossen et al., 2007; Eissing
et al., 2011). For the sake of reproducibility of the models shown in
this work, all independent parameters needed for PBPK model build-
ing and those subject to optimization are exhaustively listed in the
tables. Most notably, physicochemical parameters can be used as
direct input for PK-Sim in the graphical user interface for parameter-
ization of the underlying distribution model such that absorption,
distribution, metabolism, and excretion properties of a compound
are directly quantified.

To evaluate the benefit of using gene expression data for quantifi-
cation of active processes, we established various PBPK models with
increasing model complexity and level of detail. The performance of

<table>
<thead>
<tr>
<th>Gene set used for PBPK</th>
<th>EST/RT-PCR</th>
<th>EST/Array</th>
<th>RT-PCR/Array</th>
</tr>
</thead>
<tbody>
<tr>
<td>model (10 genes)</td>
<td>0.40 (14)</td>
<td>0.42 (17)</td>
<td>0.87 (100)</td>
</tr>
<tr>
<td>Common gene set (214 genes represented in more than 2 tissues)</td>
<td>0.21 (10)</td>
<td>0.27 (14)</td>
<td>0.66 (64)</td>
</tr>
</tbody>
</table>
these PBPK model variants are described in the following. For each model alternative, 1000 independent stochastic optimizations were performed. Because the focus was on identification of a global optimum in each case, only the overall best solution was given, thus neglecting the statistics of distribution of the 1000 parameters, which reflect algorithmic performance of the optimizer rather than providing insights into robustness of the solution.

**Basic PBPK Model.** First, we constructed a basic PBPK model for pravastatin as described under *Materials and Methods* with four adjustable parameters and little structural complexity based on mainly phenomenological observations. The PBPK model with the smallest simulation error of 1000 independent optimization runs (Fig. 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 (Fig. 4A). This is mirrored in the PK parameters determined for the basic intravenous and oral PBPK models that are summarized in Table 3 and compared with the experimentally determined values from the clinical studies. The model PK parameters after intravenous dosing are within the error range of the experimental values, whereas area under the curve (AUC) values after oral dosing of pravastatin are significantly overestimated. Excretion to urine was predicted rather well. The amount excreted to urine (A), 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 also takes into account metabolism in the gastrointestinal tract and hepatic uptake by OATP1B1 (see *Materials and Methods* for details). The 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 (Fig. 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 with the basic PBPK model (Fig. 4A). The extended basic PBPK model, however, still does not match the experimental values. It evidently overestimates exposure after oral dosing and overestimates the bioavailability (simulated 92% and 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 to integrate further biochemical and physiological knowledge on pravastatin PK had to be considered to obtain a PBPK model able to fully describe the pharmacokinetics of this drug.

**PBPK Model with Measured Expression Profiles.** To capture the physiological processes underlying pravastatin PK at a high mechanistic level of detail we extended the basic PBPK models by the enzymatic metabolism and active transport processes depicted in Fig. 2. Pravastatin is transported not only by OATP1B1 but also by MRP2 and OAT3 (Kivistö and Niemi, 2007). In addition, enzymatic degradation catalyzed by SULTs in intestinal epithelium, liver, and kidney occurs, which leads to the formation of inactive isomers in the presence of the cofactor 3'-phosphoadenosine-5'-phosphosulfate (Hatanaka, 2000). Following the new modeling concept outlined under *Materials and Methods*, 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, and 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. Because it is not yet known which SULT isoform is responsible for pravastatin metabolism, 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 most strongly 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 $k_{int}$ (Table 5) and the expression value specific for each organ (Table 4). These intrinsic $V_{max}$ values represent the actual determinants for transport or metabolization 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 SULTs was incorporated as first-order clearance ($Cl_{SULT}$) in intestine, liver, and kidney in the PBPK model because 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 (Fig. 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 (Fig. 4, D–F; Table 3).
ulation curves obtained from all three PBPK models for intravenous application of pravastatin slightly underestimate total clearance of the drug. This result 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 clearly show 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 maximum concentration ($C_{\text{max}}$) 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 indistinguishable, values for the fraction excreted to urine after intravenous and oral administration show slight differences. The EST PBPK model (Fig. 4F) showed the best prediction of urinary secretion [48 and 6.2% after intravenous and oral dosing, respectively, compared with 41 to 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.

### Table 3

**Pharmacokinetic parameters**

<table>
<thead>
<tr>
<th>Pharmacokinetic Parameters</th>
<th>PBPK Model Variant</th>
<th>Reference Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intravenous dose of 9.9 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{AUC}_0-\infty$ μmol/l·h</td>
<td>0.35</td>
<td>0.24</td>
</tr>
<tr>
<td>$A_{\text{E0-48 h}}$ %</td>
<td>36.0</td>
<td>50</td>
</tr>
<tr>
<td>Oral dose of 40 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{AUC}_0-\infty$ μmol/l·h</td>
<td>1.02</td>
<td>0.73</td>
</tr>
<tr>
<td>$C_{\text{max}}$ μmol/l</td>
<td>0.64</td>
<td>0.83</td>
</tr>
<tr>
<td>$t_{\text{max}}$ h</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>$A_{\text{E0-12 h}}$ %</td>
<td>7.1</td>
<td>6.9</td>
</tr>
<tr>
<td>Bioavailability, %</td>
<td>88.6</td>
<td>92.0</td>
</tr>
</tbody>
</table>

$^a$ Data from Singhvi et al. (1990).

$^b$ Data from Mwinyi et al. (2004).
Experimental expression values are based on microarray measurements from ArrayExpress (European Informatics Institute, 2010, http://www.ebi.ac.uk/microarray-as/) and RT-PCR data from the literature (Nishimura et al., 2003; Mwinyi et al., 2004), and EST values from UniGene (2010, http://www.ncbi.nlm.nih.gov/unaligene). 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.

<table>
<thead>
<tr>
<th>Relative Expression in Different Organs</th>
<th>Experimental Gene Expression</th>
<th>Random Gene Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OATP1B1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>OAT3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>0.021</td>
<td>1.7 × 10⁻⁴</td>
</tr>
<tr>
<td>Kidney</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>MRP2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Kidney</td>
<td>0.462</td>
<td>0.361</td>
</tr>
<tr>
<td>SULT</td>
<td>0.391</td>
<td>0.131</td>
</tr>
<tr>
<td>Small intestine</td>
<td>1.980</td>
<td>0.590</td>
</tr>
</tbody>
</table>

(1990) and was well predicted by all three PBPK models (Array PBPK model: 23.0%, RT-PCR PBPK model: 22.3%, and 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 are lacking in the extended basic model, which might explain the overestimation of the bioavailability in this model variant due to a weak description of pravastatin pharmacokinetics after oral administration in this case. Whereas 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 to decrease the overall simulation error for both forms of pravastatin administration.

**PBPK Models with Randomized Expression Profiles.** The possibility 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, cannot be excluded. To analyze this possibility, we generated 15 PBPK models integrating randomly generated expression profiles. These random expression profiles were drawn from a log normal distribution with the estimated mean and variance of the experimental expression values. The PK curves and PK parameters obtained for this model also showed good agreement with experimental values (Fig. 4B; Table 3) even though a completely randomized PBPK model had a minimal error that 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 good agreement with experimental values (Fig. 4B; Table 3) even though a completely randomized PBPK model had a minimal error that 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 good agreement with experimental values (Fig. 4B; Table 3) even though a completely randomized PBPK model had a minimal error that was in the range of the PBPK models parameterized with experimental expression values.

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 optimizations. The quality of models depends on the model structure and expression profiles and less on the parameter configuration. As can be seen in Fig. 3, the best performing random PBPK model had a minimal error that 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 good agreement with experimental values (Fig. 4B; Table 3) even though a completely randomized 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 values (Fig. 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 with those in the basic and basic extended PBPK models, it was not possible to identify a

**TABLE 5**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Basic</th>
<th>Basic Extended</th>
<th>Array</th>
<th>RT-PCR</th>
<th>EST</th>
<th>Best Random</th>
<th>Worst Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipophilicity</td>
<td>1.00</td>
<td>1.00</td>
<td>0.84</td>
<td>0.79</td>
<td>0.99</td>
<td>0.60</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Cl&lt;sub&gt;s&lt;/sub&gt;, cm/min</strong></td>
<td>3.5 × 10⁻³</td>
<td>3.50 × 10⁻³</td>
<td>8.3 × 10⁻³</td>
<td>6.4 × 10⁻³</td>
<td>7.0 × 10⁻³</td>
<td>5.9 × 10⁻³</td>
<td>1.4 × 10⁻²</td>
</tr>
<tr>
<td><strong>Cl&lt;sub&gt;L&lt;/sub&gt;, l/min</strong></td>
<td>22.6</td>
<td>22.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cl&lt;sub&gt;urine&lt;/sub&gt;, l/min</strong></td>
<td>7.4 × 10⁻²</td>
<td>2.5 × 10⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>k&lt;sub&gt;cat&lt;/sub&gt;, OATP1B1, µmol⁻¹·min⁻¹</strong></td>
<td>0.020</td>
<td>0.010</td>
<td>0.010</td>
<td>0.014</td>
<td>0.020</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td><strong>k&lt;sub&gt;cat&lt;/sub&gt;, OAT3, µmol⁻¹·min⁻¹</strong></td>
<td>4.2 × 10⁸</td>
<td>6.2 × 10⁻¹</td>
<td>1.1 × 10⁸</td>
<td>7.0 × 10⁸</td>
<td>2.1 × 10⁸</td>
<td>1.6 × 10⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>k&lt;sub&gt;cat&lt;/sub&gt;, SULT, µmol⁻¹·min⁻¹</strong></td>
<td>5.8 × 10⁸</td>
<td>3.4 × 10⁸</td>
<td>3.4 × 10⁸</td>
<td>3.4 × 10⁸</td>
<td>8.7 × 10⁸</td>
<td>8.7 × 10⁸</td>
<td></td>
</tr>
</tbody>
</table>

PBPK model parameters that were adjusted using literature data to plasma concentration and urinary excretion of pravastatin after intravenous and oral doses of 9.9 and 40 mg, respectively (Singhvi et al., 1990; Mbwinyi et al., 2004). Clearance processes (Cl<sub>s</sub>) are given as specific clearance rates (liters per minute), which can be calculated from intrinsic clearance rates (milliliters per minute) by normalization to the volume of the compartment or organ in which metabolism or degradation takes place.

**TABLE 4**

Relative gene expression data for OATP1B1, OAT3, MRP2, and SULTs used for building three PBPK model variants of pravastatin.

PBPK model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Basic</th>
<th>Basic Extended</th>
<th>Array</th>
<th>RT-PCR</th>
<th>EST</th>
<th>Best Random</th>
<th>Worst Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipophilicity</td>
<td>1.00</td>
<td>1.00</td>
<td>0.84</td>
<td>0.79</td>
<td>0.99</td>
<td>0.60</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Cl&lt;sub&gt;s&lt;/sub&gt;, cm/min</strong></td>
<td>3.5 × 10⁻³</td>
<td>3.50 × 10⁻³</td>
<td>8.3 × 10⁻³</td>
<td>6.4 × 10⁻³</td>
<td>7.0 × 10⁻³</td>
<td>5.9 × 10⁻³</td>
<td>1.4 × 10⁻²</td>
</tr>
<tr>
<td><strong>Cl&lt;sub&gt;L&lt;/sub&gt;, l/min</strong></td>
<td>22.6</td>
<td>22.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cl&lt;sub&gt;urine&lt;/sub&gt;, l/min</strong></td>
<td>7.4 × 10⁻²</td>
<td>2.5 × 10⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>k&lt;sub&gt;cat&lt;/sub&gt;, OATP1B1, µmol⁻¹·min⁻¹</strong></td>
<td>0.020</td>
<td>0.010</td>
<td>0.010</td>
<td>0.014</td>
<td>0.020</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td><strong>k&lt;sub&gt;cat&lt;/sub&gt;, OAT3, µmol⁻¹·min⁻¹</strong></td>
<td>4.2 × 10⁸</td>
<td>6.2 × 10⁻¹</td>
<td>1.1 × 10⁸</td>
<td>7.0 × 10⁸</td>
<td>2.1 × 10⁸</td>
<td>1.6 × 10⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>k&lt;sub&gt;cat&lt;/sub&gt;, SULT, µmol⁻¹·min⁻¹</strong></td>
<td>5.8 × 10⁸</td>
<td>3.4 × 10⁸</td>
<td>3.4 × 10⁸</td>
<td>3.4 × 10⁸</td>
<td>8.7 × 10⁸</td>
<td>8.7 × 10⁸</td>
<td></td>
</tr>
</tbody>
</table>
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 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 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 these publically 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 \( V_{\text{max}} \) is required for each PK-relevant protein to quantify catalytic efficiency. \( V_{\text{max}} \) can be obtained by multiplying the global rate constant \( k_{\text{cat}}^{\text{global}} \), which is organ-unspecific, and organ-specific protein abundance, which can be approximated from relative gene expression. Most importantly, \( k_{\text{cat}}^{\text{global}} \) implicitly takes into account transcriptional efficacy and posttranscriptional modifications of the same protein, thus ensuring relative comparability across multiple organs. Hence, \( k_{\text{cat}}^{\text{global}} \) is a composite parameter of translational efficacy and the enzymatic turnover 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 the \( k_{\text{cat}}^{\text{global}} \) of OATP1B1 varies among the different models, which, however, reflects the resulting simulation error. In addition, the enzyme and transporter activities [multiplication of expression profiles (Table 4) and corresponding \( k_{\text{cat}}^{\text{global}} \) values (Table 5)] show a very high correlation. Values of \( k_{\text{cat}}^{\text{global}} \), which are given in Table 5, represent the overall best solution of 1000 independent stochastic optimizations, which were performed for each model alternative and which may reflect concurrent and partly synergistic, yet non-optimizations, which were performed for each model alternative representing the overall best solution of 1000 independent stochastic optimizations.

The PBPK model of pravastatin described in this work is to our knowledge the only one reported so far that accounts for all pharmacokinetically relevant protein-mediated distribution and clearance processes reported for this compound. It considers transporters such as OATP1B1, MRP2, and OAT3 and sulfotransferases (Hatanaka, 2000; Kivistö and Niemi, 2007), integrates information to the cellular localization of these proteins (intracellular, basolateral, and apical), and represents the relative abundance of these proteins in the following tissues: enterocytes, hepatocytes, or kidney tubular cells. A previously established PBPK model aimed to predict transporter-mediated clearance and distribution of pravastatin in humans (Watanabe et al., 2009) is rather simple compared with the one presented here. The PBPK model structure only considered six organ compartments (intestine, liver, lung, kidney, brain, and muscle) with 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 is 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 \( k_{\text{cat}} \). 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 available.

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 (Fig. 4). Only consideration of all relevant organs expressing the proteins involved in metabolism and transport of pravastatin leads 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 pharmacokinetically relevant protein-mediated distribution and clearance processes reported for this compound. It considers transporters such as OATP1B1, MRP2, and OAT3 and sulfotransferases (Hatanaka, 2000; Kivistö and Niemi, 2007), integrates information to the cellular localization of these proteins (intracellular, basolateral, and apical), and represents the relative abundance of these proteins in the following tissues: enterocytes, hepatocytes, or kidney tubular cells. A previously established PBPK model aimed to predict transporter-mediated clearance and distribution of pravastatin in humans (Watanabe et al., 2009) is rather simple compared with 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 to best describe experimentally determined hepatic availability. Hepatic uptake and can-
alicular efflux activity in vivo were estimated using the above-described IVIVE/IVIVC methodology. However, transporter activity in tissues other 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 superior model performance and the resulting models are more accurate and robust in their simulations. For therapeutic substances that are actively transported or metabolized, such as pravastatin, basic PBPK models are challenged to reproduce and match experimental data. Of note, a comparison of simple PBPK models with PBPK models incorporating experimental expression data showed significantly worse performance, underlining the general benefit of using as much experimental information as possible during model building. 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 basic PBPK model, a variability that 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, whereas the latter consistently show very low simulation errors. The particular choice of experimental platform for gene expression profiling is of secondary importance because models based on different database sources show similar low simulation errors. In addition, the range of simulation errors as achieved in independent parameter optimizations is small compared with the effect of using gene expression profiles versus basic PBPK modeling.

In conclusion, we believe that use of physiological information from various levels of biological organization ranging from gene expression to the 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. In addition, such models also offer a unique possibility for further mechanistic analyses such as investigations of drug-drug interactions occurring simultaneously in different tissues.

Acknowledgments

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

Authorship Contributions

Participated in research design: Meyer, Schnecker, Ludewig, Kuepfer, and Lippert.

Performed data analysis: Meyer, Schnecker, Ludewig, and Kuepfer.

Wrote or contributed to the writing of the manuscript: Meyer, Schnecker, Kuepfer, and Lippert.

References


Address correspondence to: Dr. Lars Kuepfer, Systems Biology and Computational Solutions, Bayer Technology Services GmbH, Building 9115, 51368 Leverkusen, Germany. E-mail: lars.kuepfer@bayer.com