Use of prior knowledge in biological systems modelling
Reshetova, P.

Citation for published version (APA):
Reshetova, P. V. (2017). Use of prior knowledge in biological systems modelling

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3
Using Petri nets for experimental design in a multi-organ elimination pathway

Genistein is a soy metabolite with estrogenic activity that may result in (un)favourable effects on human health. The elucidation of the mechanisms through which food additives such as genistein exert their beneficiary effects is a major challenge for the food industry. A better understanding of the genistein elimination pathway could shed light on such mechanisms. We developed a Petri net model that represents this multi-organ elimination pathway and which assists in the design of future experiments. Using this model we show that metabolic profiles solely measured in venous blood are not sufficient to uniquely parameterize the model. Based on simulations we suggest two solutions that provide better results: parameterize the model using gut epithelium profiles or add additional biological constraints in the model.

3.1. Introduction
Genistein is a soy metabolite with estrogenic activity that may result in (un)favourable effects on human health (for a review see [75]). The elucidation of the mechanisms through which food additives such as genistein exert their beneficiary effects is a major challenge for the food industry. A better understanding of the genistein elimination pathway could shed light on such mechanisms. Parts of this pathway are hosted by specific organs (including the small intestine, gut, liver, and kidney). Metabolite degradation products travel between these organs and eventually are secreted through the gut or kidney. Although many nutrikinetics studies have been conducted to explore the genistein multi-compartment elimination pathway in human and animal models, relatively few of its details are known and, consequently, the precise metabolic pathways and routes remain to be established. In this work we therefore do not consider detailed metabolic reactions involved in the elimination pathway (which are largely unknown) but focus on the routes of degradation products through the network of compartments (the involved organs and blood).

Mathematical modelling helps to gain a more detailed understanding of the genistein elimination pathway but this requires a model describing this system in sufficient
detail. However, Petri nets are able to use incomplete and/or imprecise information to reconstruct system’s behaviour. Petri nets developed by Carl Adam Petri provide a generic approach for modelling of concurrent systems [15]. A Petri net is a bipartite graph with two types of nodes - places and transitions. In biological applications, places generally represent biological entities such as molecules, genes and enzymes. Places contain tokens that reflect, for example, metabolite concentrations or gene expression levels. Transitions represent relations between biological entities such as enzymatic reactions or metabolite transport. A Petri net simulation results in a time-dependent redistribution of tokens reflecting system dynamics. Simulation of the Petri net model implies that we select and “fire” a specific transition resulting in tokens being moved from one place to the next. The firing rules define which transition fires and the number of tokens subsequently transferred. This, together with the topology of a Petri net model, results in a qualitative representation of the system's dynamics.

Petri nets have become a popular tool for studying biological networks such as metabolic networks [76]. The review paper of Baldan et al explains how metabolic pathways have been represented and modelled with Petri nets [3]. The authors also discuss various ways to use Petri nets for modelling network topology structures such as negative feed-back loops and inhibition. Modelling network topology with Petri nets has been shown to give qualitative biologically relevant insights about the dynamics of biological systems [77, 78]. A quantitative analysis of biological network dynamics required further extension of Petri nets in a way similar to mathematical modelling using ordinary differential equations (e.g., [20]). Such Petri nets require knowledge of kinetic parameters. However, it has been shown that network dynamics might be determined by using only network topology [79, 80]. For example, Ruths and co-workers [81] assumed that network connectivity is the most significant determinant of the signal propagation and that discarding kinetic parameters from the model still results in model outcome that agrees with experimental data in the majority of cases. A similar example, involved the use of Fuzzy Logic to reconstruct the topology of a cell signalling network from gene expression data in silico [19].

Petri net models can also assist in designing new wet-lab experiments to further characterize the system under investigation. In this paper we demonstrate how a Petri net model representing the human genistein multi-organ elimination network fits this purpose (Figure 3.1). This model describes various routes involved in the elimination of genistein after dietary exposure to this compound. Each transition of metabolites within or between organs is associated with a fraction (F) that indirectly represents its relative flux. The network topology and fractions define a model configuration and allow the simulation of time-resolved metabolite relative concentration profiles for different organs given an amount of genistein input G(I) administered to an individual (Figure 3.2A). The challenge, however, is to estimate fractions from measured profiles. In this work we use concentration profiles from LC-MS venous blood measurements obtained in a nutrikinetics study in which healthy volunteers were exposed to dietary genistein [82]. The estimation of fractions is challenging because given current domain knowledge and available data, the genistein elimination pathway is insufficiently constrained and, therefore, ambiguous in terms of model configurations (i.e., sets of fractions) being in agreement with experimental data. Therefore, we used our model as an experimental
design tool to investigate which additional information (data or prior knowledge) would be required to further constrain the system to allow accurate estimations of the fractions. In particular, we investigated if additional metabolite profiles and additional constrains (e.g., fixing fractions associated with excretion transitions) would result in better parameter estimates.

To answer these experimental design questions, we used simulated annealing (SA) to estimate fractions from experimental or simulated metabolite profiles.

3.2. Results

Fraction estimation from simulated reference profiles for all places.

We first explored if fractions can be correctly estimated based on concentration reference profiles simulated for all places in our model (Figure 3.1). Consequently, we configured the Petri net with all fractions arbitrarily set to 0.5 except for fractions F2, F30 and F31 which were set to 0.34, 0.33 and 0.33 respectively (Figure 3.2 (A)). This fulfills the requirement that the fractions corresponding to outgoing transitions of a single place sum to one, and no preference for a specific transition is assumed (see Materials and Methods section). Then, we executed ten simulations with 1000 input tokens for place G(I). A simulation is terminated when all tokens left the Petri net. During simulations we recorded the number of tokens in each place to obtain the concentration reference profiles. Subsequently, using these simulated reference profiles we reconstruct the fractions through simulated annealing (Figure 3.2 (B)). The results from these reconstructions show that some fractions are precisely estimated (e.g., F1 and F7 were estimated within 2% of their true value), while other fractions showed large variability (e.g., F17, F16 deviated 90% of their true value; Appendix Figure 3.12). Note that all boxplots presented in this paper are sorted according to their estimation variances.

We defined three classes of relative estimation errors, which only can be calculated for estimates based on simulated reference profiles since the true fractions underlying the experimental data are unknown. A transition was classified as “determinant” if the relative estimation error was less than 10%. The “moderate” and “flexible” classes correspond to estimation errors of 10-25%, and >25% respectively. Although simulated annealing may converge to sub-optimal solutions (fractions), we decided not to remove one or more runs with high(er) estimation errors (i.e. outliers in the boxplots) since this may lead to a biased result and, moreover, since this is also not possible for results based on experimental data due to unknown estimation errors.

To facilitate comparison with experimental data we defined three variance classes (low, medium, high, Appendix Figure 3.10). These classes are based on visual inspection of estimation variances observed from experimental data (Appendix Figure 3.11). Figure 3.3 shows a visual representation of the estimation errors and variances in the context of our model. This simulation experiment shows that even if concentration profiles are available for all places, only a part of the fractions can be determined with sufficient accuracy. For example, fractions F17/F16, F25/F26, F15/F18, and F24/F27 have the estimation error up to 90% in 10 optimization runs, while fractions F1/F7 have the maximum estimation error of 2% (Appendix Figure 3.12). Inspecting the concentration profiles that are produced from the model we observe that these are close to the simulated
Figure 3.1: **Petri net model of the human genistein multi-compartment elimination pathway.** This model includes three metabolites (G - genistein, GG; genistein-7-glucuronide; S - genistein-7-glucuronide-4-sulphate) that travel within and between six compartments (organs, blood). G(I) is the input place, which is set to 1000 tokens at the start of a simulation and represents the amount of genistein administered to an individual. Each transition is associated with a fraction (F) indirectly representing a relative flux. Associated fractions corresponding to outgoing transitions originating from the same place are shown in the same color. Outgoing transitions associated with places without other outgoing transitions are shown in black. Six transitions (F7, F17, F26, F8, F18 and F27) represent the excretion of metabolites from the gut or kidney.

reference profile (Figure 3.4). Compared to the venous blood profiles the gut epithelium profiles are closer to the reference profiles due to intrinsic constraints between the various gut epithelium transitions. However, overall we conclude that additional model constraints are required to improve the estimation of the fractions.

**Fraction estimation from simulated and experimental reference profiles for venous blood places.**

Our experimental data comprised only three metabolite profiles (genistein, genistein-7-glucuronide, genistein-7-glucuronide-4-sulphate) measured in venous blood. Given our previous simulations, these profiles are not expected to provide sufficient information to estimate all fractions in the network with high accuracy. To confirm this, we simulated reference profiles for these three metabolites only. Fractions in the model were set either as 0.5 or 0.33 as previously. Subsequently, we estimated all fractions in the Petri net model from these three profiles. Indeed, the simulation shows that the fractions for none of the transitions could be precisely determined (Figure 3.5 and Appendix Figure 3.13). Compared to the previous results there is a clear shift to higher estimation errors (Appendix Figure 3.11). Despite the low accuracy of the estimated fractions, the model metabolite profiles for the venous blood places show a good approximation to the reference profile (Appendix Figure 3.14). The model profiles for, for example, the gut places now show much higher variability since these were not constrained by gut
Figure 3.2: Simulation of reference profiles and estimation of fractions from reference profiles. (A) The Petri net model is used to simulate reference concentration profiles by manually defining a model configuration (set of fractions) and, subsequently, executing the Petri Net. The number of transitions that is required to remove all tokens from the model is set to 36 hours. (B) Simulated annealing (SA) is used to find fractions that generate model profiles that reproduce the reference profiles (simulated or experimental) from selected places. The Petri net is configured by randomly initializing the fractions. Subsequently the model is executed to generate simulated model profiles from the trial fractions. These profiles are compared to selected reference profiles by calculating of the root square error (RSE). Based on the comparison, the trial fractions are adjusted. This process continues until convergence to a set of optimal fractions, representing a specific network configuration, is reached.

Next we estimated all fractions in the model from the experimental metabolite profiles of genistein, genistein-7-glucuronide, and genistein-7-glucuronide-4-sulphate measured in venous blood. In agreement with the results from the simulated data most of the fraction estimations show high or moderate variance (Figure 3.6, Appendix Figure 3.10). Results are even slightly worse compared to the previous results based on simulated profiles as a result of more noisy data. Even fractions F5 and F15 associated with the blood places could not be determined with high accuracy. Inspecting the model profiles we observe that the model profiles approximate the measured genistein and genistein-7-glucuronide to a lesser extent than the experimental profile for genistein-7-glucuronide-4-sulphate (Appendix Figure 3.15). Surprisingly, however, the estimation variance associated with transition F24 and F27 is low. Although we do not know the true fractions underlying our experimental data it seems that these two transitions take on more extreme fraction values (median values of 0.95 and 0.06 respectively; Appendix Figure 3.16) providing a more stringent constraint for our model compared to fractions that are closer to 0.5.

To test if a model initialized with more extreme fractions than the values we used so
Figure 3.3: Presentation of relative estimation errors (box outline thickness) and estimation variances (box colour) based on simulated reference profiles for all places (filled blue circles). Green and orange correspond to low and medium estimation variances respectively. Boxes with thick outline represent determinant fractions (errors < 10%). Dashed boxes represent moderate fractions (errors between 10 and 25%). Boxes with thin outline represent flexible fractions (errors > 25%). Note that most low variance estimates correspond to 'determinant' fractions.

far (0.5 and 0.33) would provide more accurate estimates we performed another simulation. We initialized the fractions according to the median values obtained from the experimental data (Appendix Table 3.2). Figure 3.7 shows that there is no improvement of the estimations (Appendix Figure 3.17).

Fraction estimation from simulated reference profiles for gut and liver places.

Simulations based on reference profiles of all places showed that fractions associated with gut epithelium and liver places were determined with higher accuracy (Figure 3.3). We therefore asked if gut epithelium and liver measurements would provide more accurate fraction estimates compared to using three reference profiles from venous blood. Since gut and liver biopsy data were not available we generated reference profiles for three metabolites associated with determinant transitions, i.e., genistein in gut epithelium, and genistein and genistein glucuronide in liver. Fractions in our model were set either as 0.5 or 0.33 as previously. Results from the subsequent simulations indeed show that, in contrast to results obtained from simulated venous blood reference profiles, estimations from gut/liver profiles do not only allow to accurately determine several fractions associated with the reference profiles but also result in a better estimation of other fractions in the network (Figure 3.8, Appendix Figure 3.18). The overall distribution of estimation errors (Appendix Figure 3.11) confirms this improvement.
Figure 3.4: **Selected metabolite profiles generated from model shown in Figure 3.3.** The thick red line represents the simulated reference profile. Overall, the profiles generated from the 10 models produced by simulated annealing runs resemble the reference. Compared to the venous blood profiles, the profiles for gut epithelium are closer to the reference profile.

**Inclusion of other constraints.**
We performed a simulation to investigate if additional biological constraints would improve fraction estimations. Since it is possible to measure metabolites in urine it might be possible to constrain kidney fluxes which correspond to fractions F8, F18, F27 in our model. To test that we fixed the fractions F5/F8, F15/F18, F24/F27 to 0.5 and did not optimize them with simulated annealing. The simulation shows that the additional constraints essentially improve the estimation of the rest of the fractions (Figure 3.9 and Appendix Figures 3.19 and 3.20). Three fractions F2/F30/F31 have low variance and F31 is classified as moderate fraction. Variance also has decreased for most of the fractions and is classified as medium variance class.
Figure 3.5: Presentation of relative estimation errors and estimation variances (box colour) based on simulated reference profiles from only three venous blood places (G, GG, S; filled blue circles). Orange boxes correspond to medium estimation variances (no low and high variance class estimates). Based on the relative estimation errors all transitions were classified as ‘flexible’ (errors > 25%).

Figure 3.6: Presentation of estimation variances (box colour) based on experimental reference profiles for three venous blood places (G, GG, S; filled circles). Green, orange and red correspond to low, medium and high estimation variances respectively.
Figure 3.7: Estimated fractions from **Presentation of estimation variances (box colour) based on simulated reference profiles initialized with extreme fractions.** Reference profiles for three venous blood places were used (G, GG, S; filled circles). Orange and red correspond to medium and high estimation variances respectively.

Figure 3.8: **Presentation of relative estimation errors (box outline thickness) and estimation variances (box colour) based on simulated reference profiles for three places (gut epithelium G, liver G, liver GG; filled blue circles).** Green, orange and red correspond to low, medium and high estimation variances respectively. Boxes with thick outline represent determinant fractions (errors < 10%). Dashed boxes represent moderate fractions (errors between 10 and 25%). Boxes with thin outline represent flexible fractions (errors > 25%).

Figure 3.9: Presentation of estimation variances (box colour) based on simulated reference profiles for all three venous blood places (G, GG, S; filled circles) and constrained outgoing routes. Green, orange and red correspond to low, medium and high estimation variances respectively. Dashed boxes represent moderate fractions (errors between 10 and 25%). Boxes with thin outline represent flexible fractions (errors > 25%).
3.3. Discussion

Petri nets have developed into a popular tool to model biological networks to gain qualitative and quantitative understanding of its properties in the absence of information about kinetic parameters and stoichiometry. Lack of this information excludes use of methods such as kinetic models (e.g., [83]) or Flux Balance Analysis (FBA)[84], which assumes steady state while our model represents a perturbation of the steady state, i.e., healthy volunteers take genistein which is then secreted from the body. Dynamic FBA [85] would be an alternative for non-steady state systems but still requires network stoichiometry as well as concentration profiles for all metabolites in organs considered. In the current work we showed how network models can be used as an experimental design approach. In particular, we focused on the human multi-organ genistein elimination pathway which is of importance to gain more understanding of the (beneficial) health effects of genistein breakdown products. We used a standard Petri net approach but implemented a specific firing to ensure that experimental concentration profiles can be modelled. We integrated the Petri net model with simulated annealing to estimate the relative contributions (fractions) of the various paths in this network. Implicitly, our Petri net assumes first order kinetics. Each transition depends on only a single first-order metabolite. Alternative or future descriptions of the elimination pathway may therefore require changes in the Petri net model. However, the principles (and likely the results) shown in this paper remain valid.

The experimental profiles show a slight delay in onset (Appendix Figure 3.15). This is modelled in a natural way by the current Petri net. The parameterisation of the Petri net model imposes a similar delay on the tokens through the network. In principle, this network could also be modelled by delay differential equations but this requires the inclusion of additional parameters to account for these delays. Consequently, the estimation of the fractions from simulated or experimental data becomes more difficult (data not shown).

In the current work we used simulations to investigate to what extent the fractions of the genistein elimination pathway could be estimated from metabolic reference profiles from all or selected places. This was motivated by the observation that nutrikinetic studies are often based on metabolic measurements of easily accessible fluids such as blood and urine to gain additional insights in the underlying pathways. All presented results are based on ten optimization runs (i.e., ten sets of estimated fractions) and provide a worst case scenario since we did not leave out runs with high estimation errors and/or large variability between the estimations, which occasionally happens in global optimization problems. Overall, we observe that fractions in our model are mostly non-identifiable due to lack of sufficient constraints from data or prior knowledge. Increasing the number of SA iterations or using a local optimization method [86] following SA did not improve the results. This is a common problem in such studies [87, 40, 41].

To investigate if the estimation accuracy of a fraction correlates with the distance to the Petri net input G(I) or the distance to the closest output place we plotted this distances against the three accuracy classes (determinant, moderate, flexible; Appendix Figure 3.21). The figure reveals a relation between the distance to the input place and the assigned class. Particularly, places directly near the input place G(I) are all determinant. This indicates that fluxes in the beginning of mass distribution influence the con-
centration profiles the most. Also the figure shows that all fractions with low estimation accuracy (Flexible; F7, F17, F26, F8, F18, F27) are next to an output place or further away from the input G(I). In contrast, moderate fractions do not show an obvious relation to the distance. Although it is interesting to see this behavior it does not directly provide guidelines to improve the estimation. Since the input place is fixed this excludes the possibility to change the distance between the input and the fraction to be estimated. This is also true for the output places. However, as we already argued, additional constraints on the outgoing transitions may improve the estimation accuracy (e.g., Figure 3.9).

As expected our results show that none of the fractions could precisely be estimated from simulated blood profiles of genistein, genistein-7-glucuronide, and genistein-7-glucuronide-4-sulphate. Fractions were non-identifiable (errors $>10\%$). In addition, there was considerable variability between the SA runs. Estimations from experimental blood profiles showed even worse results as a result more noisy data although, surprisingly, two fractions with more extreme values were estimated with higher precision. Based on simulations using metabolite profiles for all places we showed that still only 9 of 21 fractions could be correctly estimated (error $<10\%$). These estimates, as well as few others, also showed lower variability between the ten optimization runs. Most of the low-error estimates were associated with gut epithelium and liver genistein and genistein-7-glucuronide. Subsequent simulations showed that these three profiles are sufficient to estimate some of the gut/liver associated fractions within an error of 10%. Moreover, the results also revealed an overall decrease in estimation variances and estimation errors of all fractions in the network compared to using venous blood profiles.

Another way to improve results would be to incorporate additional biological constraints into the model. We showed that constraints on outgoing routes (presumably it is possible to know elimination flux of the compounds through kidney) would significantly decrease estimation variance overall and would allow to obtain more reliable estimates for route F31. In addition to use the Petri net as an experimental design tool we hope that our approach inspires new research towards the modelling of multi-compartment networks.

### 3.4. Conclusion

We demonstrated how Petri net simulations help gaining insight in the definition of new experiments required to allow more precise fraction estimations for the genistein elimination pathway. Unfortunately, the results from our simulation are not easily validated since this will require data from liver and other organs that is difficult to obtain from healthy individuals in nurtikinetik studies. However, future data from model organisms may come to rescue. Our approach can be used for experimental design for similar (multi-organ) metabolite elimination pathways. Although other types of pathways (e.g., signalling) require different (sometimes existing) Petri net implementations, the presented strategy is generic.
3.5. Materials and Methods

Experimental data
Venous blood metabolic profiles from 12 healthy individuals were obtained in a recent nutrikinetics study [82]. One week before and during the experiment these volunteers followed a diet with low level of polyphenols. Each individual took one genistein tablet, containing 30mg of genistein (G(I) in Figure 3.1). Plasma samples were collected at seven time points: 30 minutes before the intake of the genistein tablet, and 3, 5, 9, 16, 24, 36 hours after the intake. Genistein, genistein-7-glucuronid, and genistein-7-glucuronid-4-sulphate (included in our model) were measured in venous blood through targeted mass spectrometry. Due to large variability between the measured profiles of these individuals we used data from only a single individual (Table 3.1).

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>0h</th>
<th>3h</th>
<th>5h</th>
<th>9h</th>
<th>16h</th>
<th>24h</th>
<th>36h</th>
</tr>
</thead>
<tbody>
<tr>
<td>genistein</td>
<td>7499</td>
<td>14615</td>
<td>47356</td>
<td>111118</td>
<td>55609</td>
<td>18977</td>
<td>0</td>
</tr>
<tr>
<td>genistein-7-glucuronid</td>
<td>0</td>
<td>1331</td>
<td>12464</td>
<td>2638</td>
<td>8241</td>
<td>2046</td>
<td>2160</td>
</tr>
<tr>
<td>genistein-7-glucuronid-4-sulphate</td>
<td>0</td>
<td>7540</td>
<td>47884</td>
<td>48228</td>
<td>59803</td>
<td>45840</td>
<td>11498</td>
</tr>
</tbody>
</table>

A Petri net model of genistein elimination pathway
In a basic Petri net a transition fires at the moment a token is present in the associated place. However, these basic (sometimes referred to as original or time-less) Petri nets cannot reproduce bell-shaped time-resolved metabolite profiles (Appendix Figure 3.15) because tokens (molecules) will be immediately transferred to subsequent places preventing molecules to accumulate in a specific place. To overcome this limitation various extensions have been developed in the past, such as Stochastic Petri nets [16], Time Petri Nets [17], Hybrid Functional Petri Nets [18]. However, these approaches are specifically tailored towards signaling networks and require kinetic or other physical/chemical parameters of the system, which are not available for the genistein elimination pathway.

In the current work we use a standard Petri net approach but defined a firing rule that allows places (representing metabolites in compartments) to reproduce metabolite profiles measured by LC-MS experiments in venous blood. Consequently, we can also use the model to simulate such profiles. Our Petri net model (Figure 3.1) was designed with information from domain experts and literature. Places reflect metabolites in specific compartments (organs, blood), and tokens reflect the amount of metabolite molecules as a measure for relative concentration. Transitions reflect degradation (F30, F31), elimination (F7, F17, F26, F8, F18, F27) or transport (other transitions) fluxes within or between these compartments. Fluxes, represented as fractions, associated with these transitions are estimated from experimental or simulated metabolite reference profiles using a global optimization method (simulated annealing; [88]). A flux is not represented by an absolute quantity but as the fraction of molecules that flows from one place to the next. If a place is associated with multiple outgoing transitions (e.g., F3 and F11 for G in liver) then fractions associated to these transitions sum to one. These fractions are
used as probabilities (see ‘firing rules’ below) for the selection of a specific transition that will be ‘fired’ in a simulation step. Therefore, by definition, each fraction assumes a value between 0 and 1 and the sum of fractions of transitions that leave the same place should sum to one. Thus, fractions F3 and F11 for the ‘G’ place in liver will be F3=0.5 and F11=0.5 if no preference for a specific transition is assumed. Consequently, the fractions represent relative contributions of transitions in our compartment network. Eight places are associated with only a single outgoing transition. The corresponding fractions (F4, F6, F12, F14, F19, F21, F23, F28) are by definition fixed to one and are not estimated in the simulated annealing runs. Our Petri net consists of 31 transitions and 19 places distributed over six compartments (organs and blood). The model represents an open and non-steady-state system. Genistein G(I) enters the elimination pathway at the first day of our simulation and the simulation stops (return to steady-state) when all genistein products (tokens) have disappeared from the Petri net.

The distribution of tokens over all places at time point \( t \) is called the ‘marking’ of the Petri net model and is denoted by \( m_t \). The marking of the initial state of simulation \((t = 0)\) is called the initial marking and is denoted by \( m_0 \). Since the number of G(I) tokens for the initial marking affects the reproducibility of Petri net executions we first determined the optimal number of tokens. We executed 100 Petri net for 20, 50, 100, 1 000, 10 000, 50 000, 100 000 and 1 000 000 input tokens for G(I). Note that the Petri net is executed 100 times due to the probabilistic nature of the firing rules (see below). For each set of 100 models and each place we determined the mean profile and, subsequently, calculated the root square error (RSE) as a measure for the variance of 100 profiles. Since the RSE depends on the magnitude of the generated profiles we normalized the RSE by the initial marking. Based on the results (Appendix Figure 3.22) we choose 1000 tokens for the initial marking in our simulation experiments. A larger number of tokens did not significantly decrease the variability between Petri net executions but it significantly increased computation time.

Firing rules
A Petri net firing rule determines when and how many tokens are transferred from one place to the next. To allow the Petri net to generate bell-shaped metabolic profiles for each place we implemented a two-stage firing rule. The first stage comprises a probabilistic selection of a place based on the number of tokens in each place. The probability for a place being selected is \( \frac{n}{n_{total}} \), where \( n \) is the number of tokens in a place and \( n_{total} \) is the total number of tokens still in the model (which may be less than the initial number of tokens in G(I)). This selection ensures that places with a low number of tokens have less chance of being selected and, consequently, have more time to acquire additional tokens. If a single transition is associated with the selected place then this transition will always fire. If multiple transitions are associated then, as a second stage, one of the outgoing transitions is probabilistically selected to be fired. This selection is made based on the fractions assigned to these transitions. Firing of a transition will always move one token to the next place.

Execution of the Petri net involves repeated rounds of selection and firing and stops when all input tokens (G(I)=1000) have left the model through one of the outgoing transitions (elimination through gut lumen or kidneys). The precise number \( (N_t) \) of transitions that will fire during a simulation depends on the probabilistic process of place
and transition selection. By defining the first firing event to occur at \( t=0 \) hours and event \( N_t \) to occur at 36 hours (the last time point from the experiment profiles), we define the correspondence between Petri net execution steps (firings) and the measurement time (Figure 3.2).

**Parameterization of the Petri net model**

Given experimental or simulated metabolite reference profiles the challenge is to find the Petri net configuration (set of fractions) that reproduces these profiles. We approached this as a global optimization problem [89] and used simulated annealing [88] as the optimization method of choice (Figure 3.2B). Simulated annealing is an iterative procedure dedicated to find a set of parameters (fractions) that minimizes a chosen error function. In our application the difference between reference profiles and profiles produced by simulated annealing are minimized.

Simulated annealing starts with randomly initializing the fractions constrained to the condition that the sum of the fractions associated with outgoing transitions of the same place sum to one. Subsequently, the Petri net is executed 20 times to generate 20 token profiles from which average profiles are calculated for each place. These simulated annealing trial profiles (S) are compared to the model reference profiles (R) of all or selected places through the calculation of an error value based on the root square error (RSE):

\[
RSE = \sum_{p=1}^{P} \sum_{t=1}^{T} \sqrt{(S_{tp} - R_{tp})^2} \tag{3.1}
\]

where the first sum runs over all selected \( P \) places and the second sum runs over seven time points \( T \).

To compare simulated annealing trial profiles (S) to experimental data we have to modify the error function. The experimental data represents relative concentrations measured by mass spectroscopy. Therefore peak heights among metabolites are not directly comparable. Using this data we were more interested in reproducing peak positions and peak shapes than peak heights. Accordingly we modified Equation 3.1 in such a way that it used relative concentrations and compared the shape of simulated and experimental profiles. To be exact, \( S_{tp} \) and \( R_{tp} \) became relative concentrations \( X_{tp,rel} \) and were calculated by Equation 3.2

\[
X_{tp,rel} = \frac{X_{tp}}{\sum_{t=1}^{T} (X_{tp})} \tag{3.2}
\]

Once an initial error (RSE) value is determined, a new set of trial fractions is generated according to:

\[
f_{new_i} = f_{old_i} + T \times N_i(0,0.1) \tag{3.3}
\]

where \( f_{new_i} \) and \( f_{old_i} \) are the new and the previous values of fraction i. \( T \) is the SA temperature (see below) and \( N_i \) is a random number drawn from a normal distribution with mean=0 and standard deviation=0.1. After generation of the new fractions, the fractions of outgoing transitions corresponding to the same place are re-scaled such that their sum is one.
Given the new set of trial fractions the Petri net is again executed and the error calculated. If, compared to the previous iteration, the error value decreased then the new set of fractions is an improvement and accepted. Otherwise, the new fractions are accepted with a probability calculated by the Boltzmann criterion.

\[ e^{-\frac{\Delta E}{T}} > U(0, 1) \]  

where \( \Delta E \) is the difference between the previous and current error value (positive if the error increased), \( T \) is the temperature, and \( U \) is a random number drawn from a uniform distribution. The probability for accepting fractions that increase the RSE depends on the magnitude \( \Delta E \) of this increase, and the temperature. Temperature at the start of a simulation run is generally set to a high value such that trial fractions that increase the RSE can be accepted to allow escapes from a local minima. During the optimization procedure the temperature gradually decreases according to a geometric cooling scheme \( (T_{new} = \alpha \times T_{old}) \) \( (\alpha = 0.9) \) decreasing the acceptance probability of trial fractions that increase the error. We chose the initial temperature such that 90% of new trial fractions were accepted. The temperature was lowered after generating 2500 new sets of fractions or after accepting 250 sets. This whole procedure of parameter generation, error function evaluation, and selection was repeated until convergence (less than 5% of the new parameter sets are accepted). The set of fractions corresponding to the lowest error value is reported and used to configure the final model.

We used SA Toolbox for Optimization (SATFO) [90]. Petri net simulation program was implemented in C++. SATFO and Petri net simulation program are publicly available at http://www.bdagroup.nl/content/Downloads/software/software.php.

**Competing interests**
The authors declare that they have no competing interests.

**Author’s contributions**
PR performed simulations. PR and AHCK performed data analysis and wrote the paper. All authors contributed to the study design, read and approved the final manuscript.

**Acknowledgements**
We are grateful to Prof. Dr. Stanley Brul, University of Amsterdam and Dr. Igor Bendik-Falconnier, DSM Nutritional Products for a constructive discussion on the model of genistein elimination pathway. We thank Dr. Igor Bendik-Falconnier and Dr. Ric de Vos (Plant Research International, Wageningen) for providing us the experimental data. We also would like to thank Prof. Jaap Heringa and Dr. Anton Feenstra from Centre for Integrative Bioinformatics, Free University, Amsterdam for a helpful discussion on Petri nets theory and assistance in the implementation of Petri net simulation program.

This research is funded by The Netherlands Bioinformatics Center (NBIC).
3.6. Appendix

Figure 3.10: Variance of estimated fractions from experimental reference profiles for three venous blood places (G, GG, S). Variances are based on 10 repeated simulated annealing runs. The fractions associated with the 21 transitions (x-axes) are sorted according to estimation variances (y-axes). To facilitate comparison between different simulations we defined three broad variance classes based on visual inspection of this plot: low variance (< 0.01), medium variance (> 0.01 and <= 0.06), and high variance (> 0.06).
Figure 3.11: Distribution of estimation errors of fractions for all transitions in our model. Red: estimations errors based on simulated profiles for all places. Green: estimations errors based on simulated profiles three venous blood profiles. Blue: estimations errors based on simulated profiles associated with gut and liver metabolites (G gut epithelium, G liver, and GG liver). Dashed lines indicate 10% and 25% error defining three classes of transitions. The distributions show that availability of metabolite profiles for all places give a shift to smaller estimation errors. If only three reference profiles are available (3 venous blood places or 3 gut/liver profiles) then the gut/liver profiles give a slight advantage to estimate all fractions in the model.
Figure 3.12: **Estimated fractions from simulated reference profiles for all places.** Each box plot shows the fractions of 10 SA runs (dots) corresponding to the 21 transitions in our model. Fractions in this model were set of 0.5 except for fraction F2, F31 and F30 which were set to 0.33. The boxes define the 25% and 75% quartiles. The box line is the median value. The whiskers represent the 1.5 inter-quartile range. The transitions are sorted according to the variance of the fractions. The 13 transitions left to the red line correspond to the low variance class. The other boxes correspond to the medium variance class. No transitions were classified in the high variance class.

Figure 3.13: Estimated fractions from simulated reference profiles for three venous blood places (G, GG, S). Each box plot shows the fractions of 10 SA runs (dots) corresponding to the 21 transitions in our model. Grey boxes represent fractions corresponding to transitions associated with blood. Fractions in this model were set of 0.5 except for fraction F2, F31 and F30 which were set to 0.33. The boxes define the 25% and 75% quartiles. The middle box line is the median value. The whiskers represent the 1.5 inter-quartile range. The transitions are sorted according to the variance of the fractions. None of the transitions were classified to the low variance class.

Figure 3.14: Selected metabolite profiles generated from model shown in Figure 3.5. The thick red line represents the simulated reference profiles for G, GG and S in venous blood. Overall, the profiles generated from the 10 models produced by simulated annealing runs resemble the reference. The profiles for gut epithelium show more variability since these were not constrained by reference profiles for gut places.
Figure 3.15: **Metabolite profiles generated for venous blood places (G, GG and S) from model shown in Figure 3.6.** The thick red line represents the experimental data for a single individual. Larger variance is observed between the genistein and genistein-7-glucuronide model profiles. Profiles for genistein-7-glucuronide-4-sulphate show less variability and approximate the experimental profile to a larger extent.
Figure 3.16: Estimated fractions from experimental reference profiles measured for three venous blood places (G, GG, S). Each box plot shows the fractions of 10 SA runs (dots) corresponding to the 21 transitions in our model. Grey boxes represent fractions corresponding to transitions associated with venous blood. The boxes define the 25% and 75% quartiles. The middle box line is the median value. The whiskers represent the 1.5 interquartile range. The transitions are sorted according to the variance of the fractions. Transition T24 and T27 were classified as low variance and are associated with more extreme fraction values (median values of 0.95 and 0.06 respectively).
Figure 3.17: **Estimated fractions from simulated reference profiles initialized with extreme fractions and reference profiles measured for three venous blood places (G, GG, S).** Each box plot shows the fractions of 10 SA runs (dots) corresponding to the 21 transitions in our model. Grey boxes represent fractions corresponding to transitions associated with venous blood. The boxes define the 25% and 75% quartiles. The middle box line is the median value. The whiskers represent the 1.5 inter-quartile range. The transitions are sorted according to the variance of the fractions. The 6 transitions right to the red line correspond to the high variance class. The other boxes correspond to the medium variance class. TODO: the red line should be drawn between F13 and F16.
Figure 3.18: Estimated fractions from simulated reference profiles for three places (gut epithelium G, liver G, liver GG). Each box plot shows the fractions of 10 SA runs (dots) corresponding to the 21 transitions in our model. Grey boxes represent fractions (F1, F2, F31, F30, F11, F3, F13, F20) associated with the three gut/liver places. Fractions in this model were set of 0.5 except for fraction F2, F31 and F30 which were set to 0.33. The boxes define the 25% and 75% quartiles. The box line is the median value. The whiskers represent the 1.5 inter-quartile range. The transitions are sorted according to the variance of the fractions. The 5 transitions left to the red line correspond to the low variance class. The other boxes correspond to the medium or high variance class.
Figure 3.19: Estimated fractions from simulated reference profiles for all three venous blood places (G, GG, S) and constrained outgoing routes. Each box plot shows the fractions of 10 SA runs (dots) corresponding to the 21 transitions in our model. Fractions F5/F8, F15/F18, F24/F27 were set to 0.5. Other fractions were estimated by simulated annealing. The boxes define the 25% and 75% quartiles. The middle box line is the median value. The whiskers represent the 1.5 inter-quartile range. The transitions are sorted according to the variance of the fractions.

Figure 3.20: Variance of estimated fractions from simulated reference profiles for all three venous blood places (G, GG, S) and constrained outgoing routes. Variances are based on 10 repeated simulated annealing runs. Fractions F5/F8, F15/F18, F24/F27 were set to 0.5. Other fractions (x-axes) were estimated by simulated annealing and are sorted according to estimation variances (y-axes).
Figure 3.21: Correlation between fraction estimation error classes (D: determinant, M: moderate, F: Flexible) and distance of corresponding transition to input place $G(I)$ (A) or the closest output place (B). The distance is defined as the minimum number of transitions between the input or the closest output place and the transition for which accuracy of the fraction was estimated. The figure shows the estimated accuracies for the 21 fractions in the Petri Net, which were estimated from the simulation represented in Figure 3.3.

Figure 3.22: Deviation of RSE mean. Petri net simulations were performed 100 times for each initial marking. The deviation did not decrease from the initial marking of 1000 and therefore we chose it for further simulations.
Table 3.2: **Median fraction values.** Fractions were obtained from estimations based on experimental reference profiles for three venous blood places (corresponding to Figure 3.6 and Appendix Figure 3.15). Rounded values were used to initialize the model and generate reference profiles to test a set of extreme fractions (optimization correspond to Figure 3.7).

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>0.81</td>
</tr>
<tr>
<td>F7</td>
<td>0.19</td>
</tr>
<tr>
<td>F2</td>
<td>0.19</td>
</tr>
<tr>
<td>F30</td>
<td>0.31</td>
</tr>
<tr>
<td>F31</td>
<td>0.50</td>
</tr>
<tr>
<td>F11</td>
<td>0.34</td>
</tr>
<tr>
<td>F3</td>
<td>0.66</td>
</tr>
<tr>
<td>F5</td>
<td>0.83</td>
</tr>
<tr>
<td>F8</td>
<td>0.17</td>
</tr>
<tr>
<td>F17</td>
<td>0.81</td>
</tr>
<tr>
<td>F16</td>
<td>0.19</td>
</tr>
<tr>
<td>F13</td>
<td>0.87</td>
</tr>
<tr>
<td>F20</td>
<td>0.13</td>
</tr>
<tr>
<td>F18</td>
<td>0.79</td>
</tr>
<tr>
<td>F15</td>
<td>0.21</td>
</tr>
<tr>
<td>F26</td>
<td>0.11</td>
</tr>
<tr>
<td>F25</td>
<td>0.89</td>
</tr>
<tr>
<td>F22</td>
<td>0.74</td>
</tr>
<tr>
<td>F29</td>
<td>0.26</td>
</tr>
<tr>
<td>F24</td>
<td>0.95</td>
</tr>
<tr>
<td>F27</td>
<td>0.05</td>
</tr>
</tbody>
</table>