A quantitative AOP of mitochondrial toxicity based on data from three cell lines

Tebby, C.; Gao, W.; Delp, J.; Carta, G.; Stel, W. van der; Leist, M.; … ; Bois, F.Y.

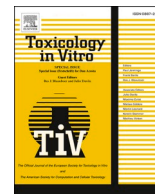# A quantitative AOP of mitochondrial toxicity based on data from three cell lines

Cleo Tebby [a],[*], Wang Gao [a],[b], Johannes Delp [c],[d], Giada Carta [e], Wanda van der Stel [f], Marcel Leist [c], Paul Jennings [e], Bob van de Water [f], Frederic Y. Bois [a],[g]

[a] Experimental Toxicology and Modeling (TEAM) Unit, Ineris, rue Jacques Taffanel, 60550 Verneuil-en-Halatte, France
[b] Applied Mathematics Laboratory (LMAC EA 2222), Université Technologique de Compiègne (UTC), 57 avenue de Landshut, 60203 Compiègne Cedex, France
[c] In Vitro Toxicology and Biomedicine, Department inaugurated by the Doerenkamp-Zbinden Foundation, University of Konstanz, Universitaetsstr. 10, 78464, Konstanz, Germany
[d] Cooperative Doctorate College InViTe, University of Konstanz, Konstanz, Germany
[e] Division of Molecular and Computational Toxicology, Amsterdam Institute for Molecules, Medicines and Systems, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[f] Division of Drug Discovery and Safety, Leiden Academic Centre of Drug Research, Leiden University, Leiden, the Netherlands
[g] Certara, Simcyp, Sheffield, United Kingdom

ARTICLE INFO

ABSTRACT

Adverse Outcome Pathways (AOPs) are increasingly used to support the integration of *in vitro* data in hazard assessment for chemicals. Quantitative AOPs (qAOPs) use mathematical models to describe the relationship between key events (KEs). In this paper, data obtained in three cell lines, LHUMES, HepG2 and RPTEC/TERT1, using similar experimental protocols, was used to calibrate a qAOP of mitochondrial toxicity for two chemicals, rotenone and deguelin. The objectives were to determine whether the same qAOP could be used for the three cell types, and to test chemical-independence by cross-validation with a dataset obtained on eight other chemicals in LHUMES cells. Repeating the calibration approach for both chemicals in three cell lines highlighted various practical difficulties. Even when the same readouts of KEs are measured, the mathematical functions used to describe the key event relationships may not be the same. Cross-validation in LHUMES cells was attempted by estimating chemical-specific potency at the molecular initiating events and using the rest of the calibrated qAOP to predict downstream KEs: toxicity of azoxystrobin, carboxine, mepronil and thifluzamide was underestimated.

Selection of most relevant readouts and accurate characterization of the molecular initiating event for cross-validation are critical when designing *in vitro* experiments targeted at calibrating qAOPs.

1. Introduction

Adverse outcome pathways (AOPs) have become an organizing framework to facilitate the development and integration of alternative test methods for assessing hazard of chemicals to human health and the environment. A dedicated program is currently running under the auspices of the Organisation for Economic Co-operation and Development (OECD) (Organisation for Economic Co-operation and Development (OECD), 2021). AOPs are intended to represent critical *in vivo* phenomena by capturing the essential elements of the perturbations of a biological system with a simple linear model (Organisation for Economic Co-operation and Development (OECD), 2016). Practically, an AOP is a chemical-independent description of a linear or branching path

from a molecular initiating event (MIE) to an eventual adverse outcome (AO) at the organism or population level (Villeneuve et al., 2014). In between, there can be any number of intermediate critical and measurable key events (KEs) connected by key event relationships (KERs) (LaLone et al., 2017; Organisation for Economic Co-operation and Development (OECD), 2018)

AOPs are useful for hazard assessment as they describe mechanism of action. AOPs can support the development of integrated testing strategies (ITS) and their application in risk assessment (Leist et al., 2017; Vinken, 2013). In case of ITS building, the data generated by alternative methods (*i.e., in silico, in vitro*), when combined with existing animal data, are used and assessed by means of a fixed data interpretation procedure (Sachana and Leinala, 2017). For this purpose, quantitative
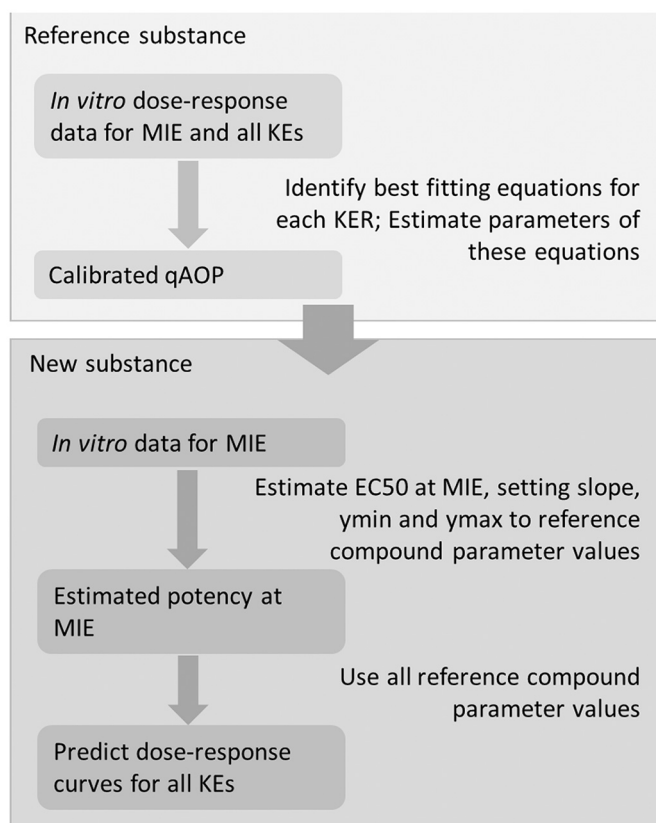

Fig. 1. Workflow for calibrating the qAOP with a reference chemical and applying it to a new chemical.

AOPs (qAOPs) that provide dose-response and time-course predictions (Conolly et al., 2017; Spinu et al., 2019) are likely to be more valuable for ITS construction than qualitative AOPs.

qAOPs are built by modeling key event relationships of a qualitative AOP with mathematical equations, based on available data, including specially produced experimental data (Zgheib et al., 2019) or data collected from the literature (Burgoon et al., 2020; Foran et al., 2019). Several methods have been used to quantify AOPs, based either on categorical Bayesian decision trees or networks (Jaworska et al., 2015; Rovida et al., 2015), or on continuous models ranging from dose response models (Hassan et al., 2017), up to systems toxicology models (Conolly et al., 2017). Categorical Bayesian decision trees or networks predict hazard in the form of probabilities of being non-toxic, weakly toxic, strongly toxic *etc.* The underlying distributions are multinomial and continuous test responses are discretized to fit in the multinomial framework. Such decision trees have already been applied to AOPs in the area of skin sensitization (Jaworska et al., 2015). They facilitate potency assessment for classification purposes and inform hazard characterization in a semi-quantitative way. A larger number of continuous qAOPs models have been developed using dose-response models (Hassan et al., 2017; Zgheib et al., 2019), dynamic Bayesian networks (Zgheib et al., 2019), systems biology or systems toxicology models (Conolly et al., 2017; Zgheib et al., 2019). qAOPs need to be calibrated by fitting the KER to data, setting parameters to their best estimates, or, in a Bayesian approach, to their posterior distributions. KER parameter estimates can be obtained by estimating each dose-response relationship separately, or by fitting the entire dataset simultaneously (Villeneuve et al., 2014). In theory, once the data for a single chemical has been modelled by a qAOP, the set of equations developed should help better understand how KEs are related. It should also be usable to predict responses at concentration levels or for chemicals that have not been tested (chemical-independence) (Perkins et al., 2019). However, we have little evidence

that the continuous qAOPs developed so far are chemical-independent.

Mitotoxicity-induced cell death has been described by an AOP entitled "Inhibition of the mitochondrial complex I of nigro-striatal neurons leads to parkinsonian motor deficits" in the AOPWiki (https://aopwiki.org/aops/3). Its molecular initiating event (MIE) is the binding of an inhibitor to NADH-ubiquinone oxidoreductase (complex I), which leads successively to inhibition of Complex I, mitochondrial dysfunction, impaired proteostasis, a combination of neuroinflammation and degeneration of DA neurons of the nigrostriatal pathway, and finally to Parkinsonian motor deficits. The AOP was developed using two stressors, rotenone and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP/MPP+).

The objectives of this research were to:

- Check whether a quantitative version of the above mitotoxicity AOP, calibrated using *in vitro* experimental data on rotenone and deguelin in Lund human mesencephalic (LHUMES) neuronal cells, HepG2 liver cells, and RPTEC/TERT1 renal cells, could be developed using the same structural equations in the three cell types (cell-type independence) and for both chemicals (chemical-independence).
- Further check chemical-independence by predictive cross-validation with a MIE activation and AO dataset obtained on eight other chemicals (antimycin A, azoxystrobin, carboxine, fenpyroximate, mepronil, pyraclostrobin, pyrimidifen, and thifluzamide) in LHUMES cells.
- Identify challenges and practical difficulties in development and calibration of qAOPs, in order to a) help increase relevance of design of experiments aimed at calibrating and validating qAOPs, and b) provide working examples of quantitative models of KERs can be combined to form a qAOP.

2. Material and methods

The workflow followed for qAOP calibration and subsequent cross-validation is shown in Fig. 1.

2.1. Structure of the AOP

The mitotoxicity-induced cell death AOP is based on the first parts of the "Inhibition of the mitochondrial complex I of nigro-striatal neurons leads to parkinsonian motor deficits" AOP described in AOPWiki (https://aopwiki.org/aops/3). We investigated mitochondrial complex I inhibition-induced mitotoxicity *in vitro* in neuronal cells, renal cells and hepatocytes. Therefore, we adapted the existing AOP so that the KEs were representative of events observed in the three cell types (*e.g.* DA neurons degeneration was translated as cell death for liver and renal cells). The various *in vitro* readouts were then mapped to the KEs of this adapted AOP (Fig. 2). For technical reasons, the readouts for the various KEs were not exactly the same for the three cell types. A larger set of readouts was obtained for the kidney and liver cells compared to neurons. While mitochondrial respiration was measured in neuronal cells, mitochondrial dysfunction *per se* was not assessed. The actual neuronal qAOP therefore included a reduced set of KEs. Conversely, the MIE was not measured in kidney cells. See Table 1 and precise AOP structures for each type of cell in Fig. S1, S2, and S3 in Supplemental Materials.

2.2. Data

2.2.1. In vitro data in hepatic cells

Effects of rotenone and deguelin in HepG2 cells were quantified based on four readouts: Mitochondrial respiration (in intact or permeabilized cells), mitochondrial membrane potential, lactate concentrations and resazurin conversion. All data were expressed as percentage of DMSO control.

Mitochondrial respiration: Decrease in mitochondrial respiration was and measured using an Agilent® Seahorse OCR equipment.
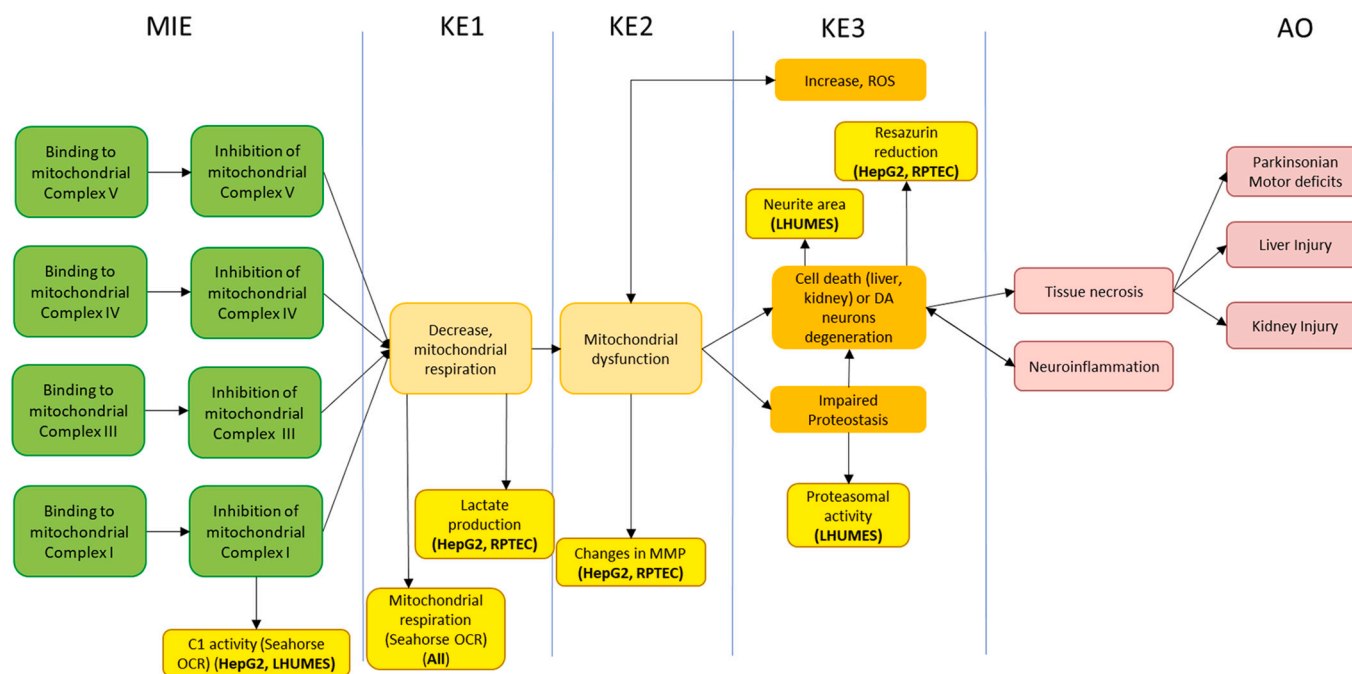
Fig. 2. Complex I inhibition-induced mitotoxicity AOP for dopaminergic (DA) neurons, kidney and liver. The readouts are in yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
List of the mitotoxicity AOP readouts studied. All are expressed in percentage of DMSO control.

Readout abbreviation	Description	Position in AOP	Cell lines
CI	Complex I specific mitochondrial respiration	MIE	Neuronal, hepatic
LC	Lactate concentration	KE1	Hepatic, renal
MR	Mitochondrial respiration	KE1	All
PT	Proteasomal activity	KE2	Neuronal
MMP	Mitochondrial membrane potential	KE2	Hepatic, renal
NR	Neurite Area	KE3	Neuronal
RZ	Resazurin reduction	KE3	Hepatic, renal

Mitochondrial respiration is based on the oxygen consumption rate and is assessed directly (within 15–30 min) after exposure. Several features can be extracted from this measurement, including basal and maximal respiration. The full method description can be found in (van der Stel et al., 2020). As the aim was a comparison of the effect of chemicals on a single cell line, the maximal respiration was used as a readout rather than the difference between basal and maximal respiration.

Lactate concentrations: Lactate production was measured using a colorimetric assay capturing the conversion of pyruvate to lactate by lactate dehydrogenase (LDH) enzyme. The presence of lactate was monitored in the supernatant after 24 h of chemical exposure. The data and method description can be found in (van der Stel et al., 2020).

Mitochondrial membrane potential: Mitochondrial dysfunction was quantified based on the loss of mitochondrial membrane potential (MMP) using the potential sensitive Rhodamine123 dye. The data and method description can be found in (van der Stel et al., 2020).

Resazurin concentration: Cells' viability was measured based on their capacity to reduce resazurin after 24 h chemical exposure. The reduction of resazurin to resorufin results in the increase of fluorescence signal around 580 nm. This increase in resorufin was measured after 24 h chemical exposure. The data and method description can be found in (van der Stel et al., 2020).

2.2.2. In vitro data in renal cells

Methods were as for hepatocytes, except that inhibition of mitochondrial Complex I, the MIE, was not measured in RPTEC/TERT1 cells. Lactate production and resazurin reduction were measured in three (rotenone) or four (deguelin) experiments with two technical replicates each. MMP was measured in three experiments with two technical replicates each.

2.2.3. In vitro data in neuronal cells

Complex I inhibition and mitochondrial respiration were measured as in HepG2 cells. Two readouts were specific to the AOP applied to neuronal cells: impaired proteostasis was measured *via* proteasomal activity, neuronal degeneration was measured *via* an estimation of their neurites' area.

Complex I activity was measured using in proliferating LHUMES cells. Mitochondrial respiration and proteasomal activity were measured using the same cells at a stage of neurite growth (day 3 of differentiation).
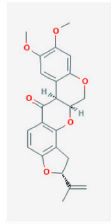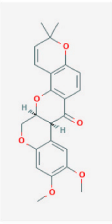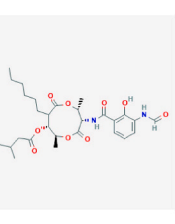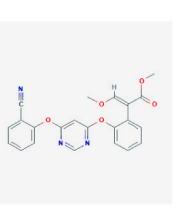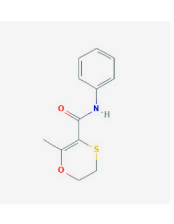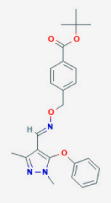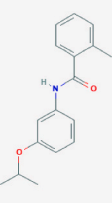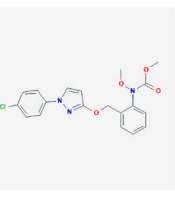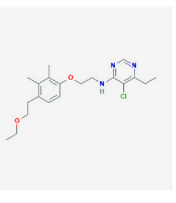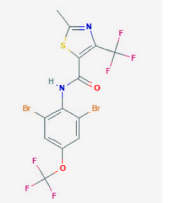
The proteasomal function of cells was assessed at 24 h after toxicant exposure by a fluorogenic substrate that increases in fluorescence when the proteasome is active (Delp et al., 2021).

Neurite area was measured using LHUMES cells at a stage of neurite growth (day 2 of differentiation). The neurite areas (which serves as indirect measurement of neuronal interconnectivity) of stained differentiating neurons, as well as cellular viability are measured simultaneously at 24 h after toxicant exposure using high content imaging. The processes of neurite outgrowth and cell death are measured.

2.2.4. Tested chemicals

Rotenone, deguelin, antimycin A, azoxystrobin, carboxine, fenpyroximate, mepronil, pyraclostrobin, pyrimidifen, and thifluzamide were obtained from Merck and distributed to the different sites by the JRC (Ispra, Italy). Their chemical structures are shown in Table 2. The test concentrations are given in see Tables S1 to S7 of the Supplemental Materials. Chemical purity was greater than 95% for rotenone, 97% for thifluzamide and 98% for other chemicals except for antimycin A (undetermined).

Table 2

Chemical structure of the chemicals tested on neurons for the development of the complex I inhibition-induced mitotoxicity AOP.

Rotenone	Deguelin	Antimycin	Azoxystrobin	Carboxine
Fenpyroximate	Mepronil	Pyraclostrobin	Pyrimidifen	Thifluzamide

Fig. 3. Graphical representation of the three KER functions: A) log-logistic, B) 1st order linear, C) asymptotic with horizontal shift.

Table 3

List of models used to represent KERs in each type of cell. The seven KE readouts are coded as: CI (complex I activity), LC (lactate production), MR (mitochondrial respiration), MMP (mitochondrial membrane potential), PR (proteasomal activity), and NR (neurite area), RZ (resazurin level).

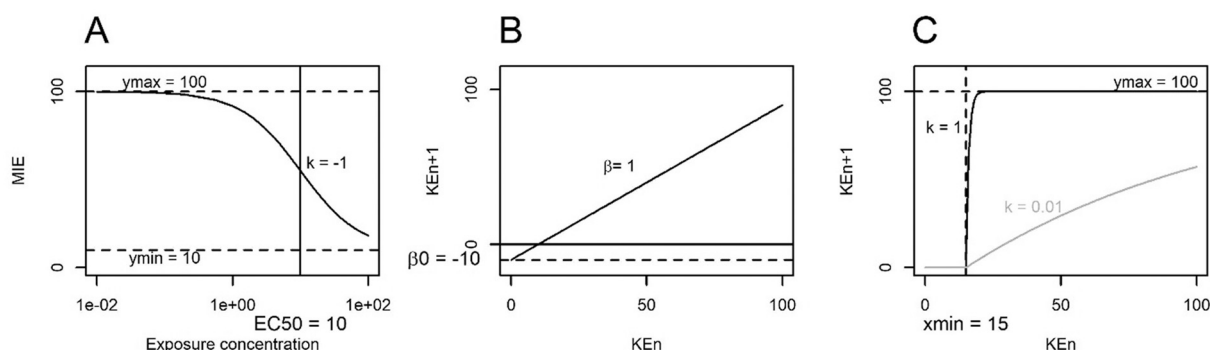| KER	Neuronal cells	Liver cells	Renal cells
Dose → CI	Log-logistic [a]	Log-logistic	
Dose → LC			Log-logistic
CI → MR	Linear [b]	Asymptotic [c]	
CI → LC		Linear	
Dose → logMR			Log-logistic
MR → PR	Asymptotic		
MR × PR → NR	Asymptotic [d]		
MR + LC → logMMP		Linear [e]	
logMR + LC → logMMP			Linear [f]
logMMP → RZ		Log-logistic	Log-logistic

[a] See Eq. 1; [b] See Eq. 2; [c] See Eq. 3; [d] See Eq. 6; [e] See Eq. 4; [f] See Eq. 5.

2.3. Calibration of quantitative AOPs

2.3.1. Models of KER

Each KER was modelled using one of the three functions defined hereafter and represented in Fig. 3. Model A is a 4-parameter log-logistic function f (Eq.1), which is a sigmoid when x is on a logarithmic scale,

where y_{max} is the response at high dose of KE_x levels, y_{min} is the response at low dose levels, k is a slope parameter, and $EC50$ is the inflexion point of the sigmoid on a log scale:

$$f(KE_x) = y_{min} + \frac{y_{max} - y_{min}}{1 + exp(-k(log(KE_x) - log(EC50)))}$$ (1)

Model B is a linear function g (Eq. 2) with slope β and intercept β_0.

$$g(KE_x) = \beta \times KE_x + \beta_0$$ (2)

Model C is a regression function which increases and has a horizontal asymptote and a shift on the x-axis (Eq. 3), where y_{max} is the response at high KE_x levels, k is a slope parameter and x_{min} is a KE_x threshold below which the next KE is equal to 0.

$$h(KE_x) = y_{max}(1 - e^{-k(KE_x - x_{min})}) \text{ if } KE_x > x_{min}, 0 \text{ otherwise.}$$ (3)

The general behaviour of those functions is illustrated in Fig. 3.

2.3.2. KER model selection

The qAOPs for rotenone and deguelin were calibrated in a stepwise approach, starting by plotting the first KER (relationship between MIE and KE1) to determine visually the most appropriate dose-response model, calibrating the first KER, and predicting KE1 at the concentration levels where KE2 was measured. These predictions were used to plot the second KER (between predicted KE1 and observed KE2) to determine

Fig. 4. Predicted and observed dose-response relationships in the mitotoxicity qAOP in neurons for each readout (complex I inhibition, mitochondrial respiration inhibition, proteasomal activity, neurite area. Red line: predictions obtained with the maximum posterior parameter values. Grey lines: predictions obtained with 30 random parameter vectors drawn from their joint posterior distribution. Dots: experimental data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

visually the most appropriate model for the second KER. The first two KERs were then calibrated together and the predictions (for the MIE, KE1, and KE2) were checked against the observations *etc.*

In cases where several models could be used, the final models were selected by comparing the results of leave-one-out cross-validation (Vehtari et al., 2017) using Pareto-smoothed importance sampling (PSIS) by calculating the pointwise likelihood of the data given each of the tested models.

Table 4

Parameter estimates of the neuronal cells mitotoxity qAOP (maximum posterior values and 95% credibility interval) for rotenone and deguelin.

KER	Function	Parameter	Rotenone	Deguelin
Concentration → CI	Log-logistic (Eq. 1)	y_{minCI} (%)	8.24 [0.345; 13.6]	9.09 [0.176; 12.1]
		y_{maxCI} (%)	103 [95.8; 117]	102 [93.6; 112]
		k_{CI}	−1.13 [−1.23; −0.783]	−1.1 [−1.37; −0.722]
		$EC50$ (nM)	18.8 [12.5; 29.9]	62.3 [48.9; 109]
		σ_{CI}	6.54 [4.77; 13.6]	7.15 [4.95; 14.7]
CI → MR	Linear (Eq. 2)	β_{MR}	0.883 [0.715; 1.17]	0.876 [0.736; 1.08]
		β_{0MR}	8.31 [−7.04; 20.2]	5.84 [−7.31; 14.5]
		σ_{MR}	19.1 [13.5; 28.7]	10.9 [8.8; 18.1]
MR → PR	Asymptotic (Eq. 3)	y_{maxPR} (%)	95.1 [91.4; 103]	99.5 [95.3; 104]
		k_{PR}	6.29 [0.463; 8.67]	2.31 [0.37; 6.19]
		x_{minPR} (%)	15.5 [2.91; 23.6]	13.7 [0.65; 17.2]
		σ_{PR}	13.6 [11.1; 18.5]	8.92 [7.56; 12.3]]
MR × PR → NR	Asymptotic (Eq. 7)	k_{NRMR}	0.0773 [0.043; 0.101]	0.856 [0.191; 1.6]
		$x_{minNRMR}$ (%)	12.4 [0.356; 21.2]	13.6 [0.647; 17.1]
		k_{NRPR}	0.0837 [0.0537; 0.927]	0.058 [0.0362; 0.693]
		$x_{minNRPR}$ (%)	37.3 [2.89; 59.2]	35.7 [2.6; 55.1]
		σ_{NR}	11.7 [10.5; 14.3]	11.9 [10.9; 14.1]

2.3.3. Models for multiple readouts

In the liver and renal qAOPs, lactate production (LC) and decreased mitochondrial respiration (MR) were readouts for the same KE, decrease in mitochondrial respiration (KE1). The two readouts were expressed as percent absolute change to control and averaged when entered in the linear relationship with KE2 (Eq. 4). In the renal qAOP, MR was log-transformed (Eq. 5). The transformation into percent change to control ensures that changes occur in the same direction (in the raw data, lactate concentration increases with dose whereas resazurin decreases), and avoids negative values.

$$KE_2 = g\left(\frac{LC_{control}/LC + MR/MR_{control}}{2} \times 100\right) \tag{4}$$

$$KE_2 = g\left(\frac{LC_{control}/LC + logMR/logMR_{control}}{2} \times 100\right) \tag{5}$$

In Eq. 4 and Eq. 5, *LC* is the measured lactate production (increasing with dose), $LC_{control}$ is the mean lactate production in controls estimated by the model, *MR* is the measured mitochondrial respiration (decreasing with dose), and $MR_{control}$ is the mean mitochondrial respiration in controls estimated by the model.

In the neuronal cells qAOP, two KEs lead to KE3: mitochondrial dysfunction (readout *MR*) causes impaired proteostasis (readout *PR*) and neuron degeneration (readout *NR*), while impaired proteostasis also causes neuron degeneration. This AOP is therefore not strictly linear and *NR* was modelled as the product of two functions of *MR* and *PR* (Eq. 6) under the assumption that both KEs acted independently on it:

$$KE_3 = h\left(\frac{MR}{MR_{control}}\right) \times h\left(\frac{PR}{PR_{control}}\right) \tag{6}$$

Which can be rewritten as in Eq. 7:

$$KE_3 = y_{maxNRMR}\left(1 - e^{-k_{NRMR}\left(\frac{MR}{MR_{control}} - x_{minNRMR}\right)}\right)$$

$$\times y_{maxNRPR}\left(1 - e^{-k_{NRPR}\left(\frac{PR}{PR_{control}} - x_{minNRPR}\right)}\right) \text{ if } MR > x_{minNRMR} \text{ and } PR > x_{minNRPR}, 0 \text{ otherwise} \tag{7}$$

Where $y_{maxNRMR} = y_{maxNRPR} = 1$, k_{NRMR} and k_{NRPR} are the slope parameters.
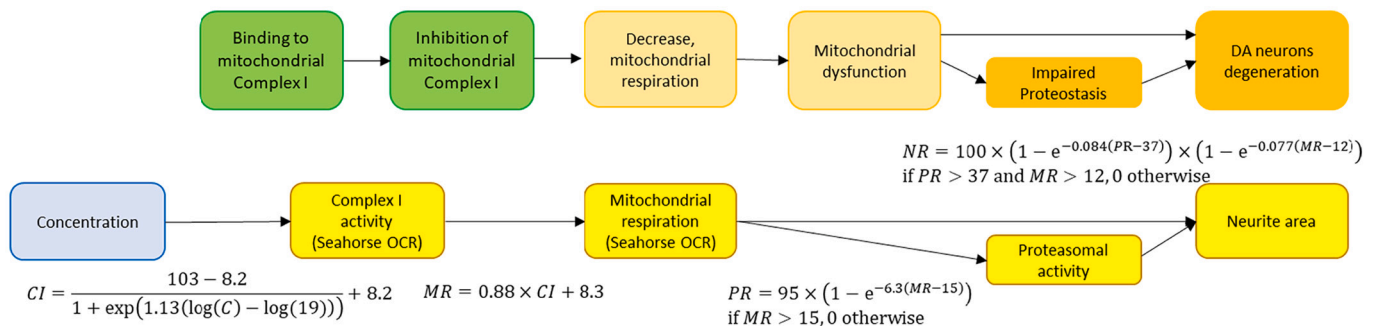
Fig. 5. Equations for the qAOP calibrated in neuronal cells for rotenone. The parameter values in the equations are the MPV estimates. All readouts are expressed in percentage (%) and concentration *C* is nanomolar (nM).
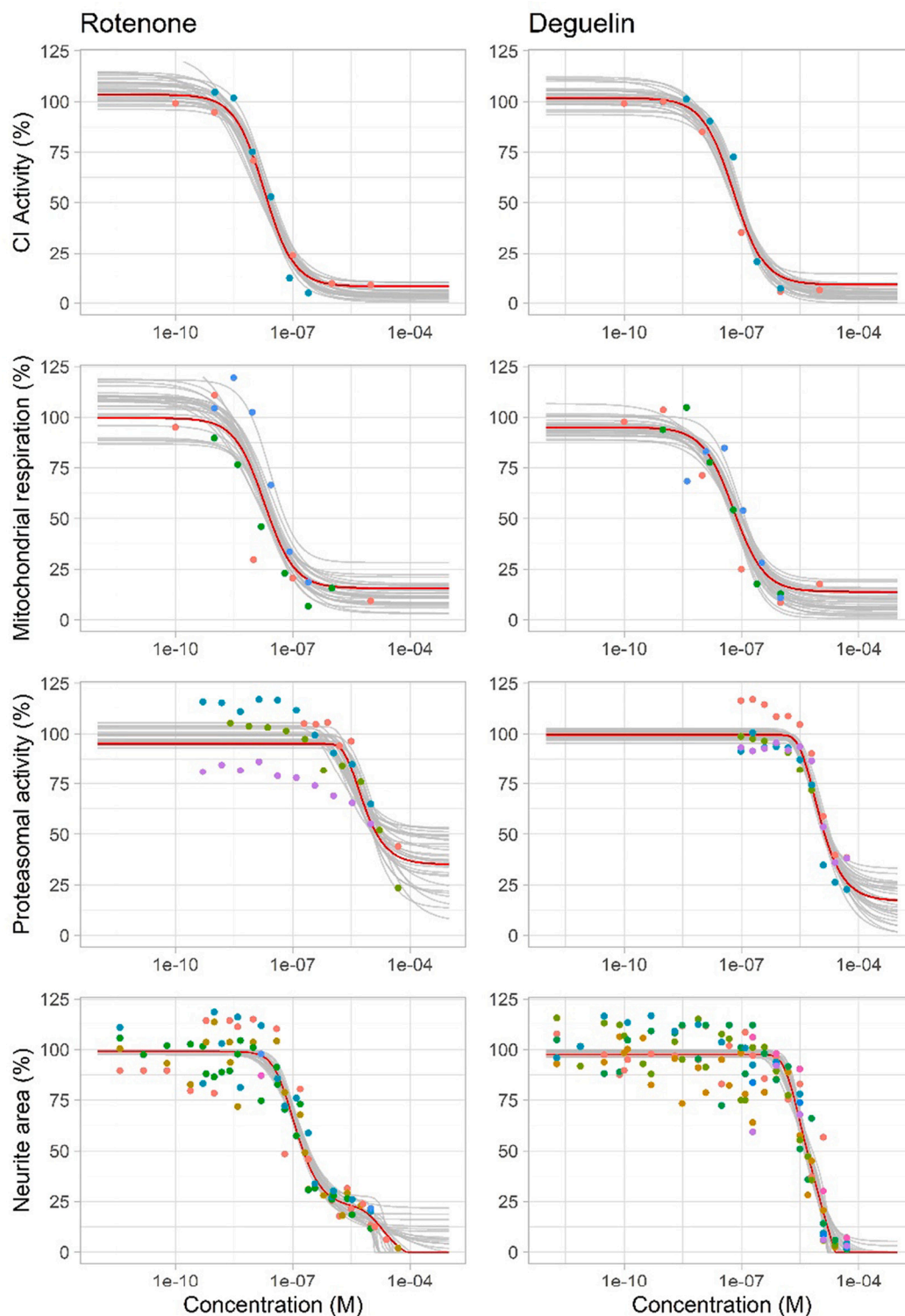
Fig. 6. Predicted and observed dose-response relationships in the mitotoxicity qAOP in liver cells for each readout (complex I inhibition, mitochondrial respiration inhibition, proteasomal activity, neurite area. Red line: predictions obtained with the maximum posterior vector. Grey lines: predictions obtained with 30 random parameter vectors drawn from their joint posterior distribution. Dots: experimental data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Parameter estimates in the liver cells mitotoxity qAOP (maximum posterior values and 95% posterior distribution).

KER	Function	Parameter	Rotenone	Deguelin
Conc → CI	Log-logistic (Eq. 1)	y_{minCI} (%)	13.8 [3; 23.9]	22.4 [12.3; 27.6]
		y_{maxCI} (%)	117 [103; 152]	95.4 [75.9; 116]
		k_{CI}	−0.725 [−0.979; −0.515]	−1.07 [−1.35; −0.744]
		$EC50_{CI}$ (nM)	16.3 [5.28; 29.2]	60.5 [31.6; 197]
		σ_{CI}	8.52 [6.12; 15.5]	5.36 [4.26; 13.6]
CI → LC	Linear (Eq. 2)	β_{LC}	−0.274 [−0.385; −0.13]	−0.404 [−0.63; −0.277]
		β_{0LC}	136 [128; 144]	134 [125; 143]
		σ_{LC}	12.1 [9.84; 16.9]	9.17 [7.44; 13]
CI → MR	Asymptotic (Eq. 3)	y_{maxMR} (%)	125 [98.3; 191]	88.7 [77.9; 167]
		k_{MR}	0.0178 [0.0077; 0.0344]	0.0643 [0.0114; 0.128]
		x_{minMR} (%)	11.3 [0.806; 21.2]	20.1 [7.15; 26.1]
		σ_{MR}	13.8 [11.8; 20.2]	17.9 [12.9; 22.6]
MR + LC → logMMP	Linear (Eq. 4)	β_{MMP}	0.0783 [0.0642; 0.115]	0.0729 [0.0617; 0.11]
		β_{0MMP}	−2.65 [−4.49; −1.76]	−2.63 [−4.07; −1.57]
		σ_{MMP}	0.423 [0.363; 0.539]	0.473 [0.382; 0.556]
logMMP → RZ	Log-logistic (Eq. 1)	$ymin_{RZ}$ (%)	19.1 [1.56; 67.8]	35.4 [1.74; 78.2]
		$ymax_{RZ}$ (%)	96.6 [94; 100]	103 [99; 134]
		k_{RZ}	3.13 [1.45; 9.02]	0.792 [0.226; 3.26]
		$EC50_{RZ}$ (%)	1.60 [1.15; 3.20]	1.12 [0.236; 12.3]
		σ_{RZ}	4.94 [4.27; 7.48]	5.05 [3.92; 6.91]

2.3.4. Statistical modeling

The qAOP developed can be understood as Bayesian networks (BNs) with continuous random variables. Modeling KEs as continuous variables increases relevance of the qAOP for risk assessment since ultimately the AO is quantitatively related to the exposure concentration level. In typical AOP diagrams, KEs are represented by boxes and KERs by single one-directional arrows connecting them. The path linking the various KEs should not form loops (feedback or feed-forward loops between two consecutive KEs can simply be indicated by a symbol and included in the KER). Thus, according to graph theory, in the absence of loops, AOPs are acyclic directed graphs (Pavlopoulos et al., 2011). BNs use acyclic directed graphs as their organizational structure (Oates and Mukherjee, 2012). The links between their nodes correspond to statistical dependencies.

As explained above, the model was calibrated in a stepwise approach to identify each KER model sequentially. At each step, the model was calibrated using four Hamiltonian MCMC (HMCMC) run in parallel with at least 10,000 iterations (up to 50,000 for the whole AOP), keeping the last 5000 iterations. A total of 17 parameters (4 KERs) were estimated in the neuronal and renal cell qAOPs; 20 parameters (5 KERs) were estimated in the liver cell qAOP. Prior distributions of parameters were uniform or log-uniform with wide bounds, as reported in the Supplemental Materials, section 4.1. The data likelihood of each point was calculated according to Eq. 8.

$$L(\theta) \sim \prod_i Normal\_df(CI_i|f(conc), \sigma_{CI}) \times \prod_j Normal\_df(MR_j|g(CI), \sigma_{MR})$$
$$\times \prod_k Normal\_df(PR_k|h(MR), \sigma_{PR}) \times \prod_l Normal\_df\left(NR_l\middle|\left(h\left(\frac{MR}{MR_{control}}\right)\right.\right.$$
$$\left.\left. \times h\left(\frac{PR}{PR_{control}}\right)\right), \sigma_{NR}\right) \tag{8}$$

Where Θ is the vector of parameters, CI_i are the data for Complex I inhibition $i \in \{1,...,n_i\}$, MR_i are the data for mitochondrial respiration $j \in \{1,...,n_j\}$, PR_k are the data for proteostasis $k \in \{1,...,n_k\}$, NR_l are the data for neurite area $l \in \{1,...,n_l\}$, $g(CI)$ is the predicted value of CI at the concentrations where MR was measured.

$$L(\theta) \sim \prod_i \prod_j Normal\_df\left(KE_{i,j}|KER_{i-1,i}(KE_{i-1}), \sigma_{KEi}\right) \tag{9}$$

where.

where $KE_{i,j}$ is the jth observation of KE_i, $KER_{i-1,i}$ is the KER describing the relationship between KE_{i-1} and KE_i and is one of the functions described in Eqs. 1 to 6. A different standard deviation was estimated for each KE. σ_{CI}, σ_{LC}, σ_{MR}, σ_{PR}, σ_{NR} σ_{MMP}, σ_{RZ}; measurement error was assumed to be normally distributed, except for MR in renal cells and for MMP, where it was assumed to be log-normally distributed.

Throughout the AOPs, the error structure that may have arisen from reproduction of the experiments and replicates within experiment was not taken into account: pointwise errors were considered independent, including in the case of lactate and resazurin production, which were measured during the same assays.

Parameter estimates in the qAOP calibrated for rotenone and deguelin were compared by identifying overlap in the 95% credibility interval.

2.4. Predictions based on calibrated qAOPs

The predictive performance of the neuronal qAOP calibrated with rotenone and deguelin data was checked on eight other chemicals (antimycin A, azoxystrobin, carboxine, fenpyroximate, mepronil, pyraclostrobin, pyrimidifen, and thifluzamide) by predicting all KEs including neurite area, given the MIE (Fig. 2). In this validation dataset, complex I inhibition and mitochondrial respiration were only measured at one concentration level for each chemical, between 12.5 and 100 μM, whereas proteasomal activity and neurite area were measured at over 10 concentration levels over at least a 1000-fold range of concentration levels. First, the EC50's of the MIE were estimated for each chemical using Eq. 1. The values of the maximum, minimum and slope of those dose-response, and all other KER parameters were set to those of the reference chemical (either rotenone or deguelin).

2.5. Software

Calculations were performed using R version 3.6.1 (R Core Team, 2019) with packages *loo* (Vehtari et al., 2020), *tidyverse* (Wickham et al., 2019), and *rstan* (Stan Development Team, 2020), which called the *Stan* HMCMC software (Carpenter et al., 2017). An example of code for the calibration of the liver qAOP for deguelin is provided in Supplementary Materials, section 6, and working examples for rotenone in liver and neuronal cells are provided at DOI https://doi.org/10.5281/zenodo.5549495.

3. Results

3.1. Calibration of qAOPs

The qAOP were successfully calibrated for the three cell lines by choosing appropriate models for the KERs and calibrating these models
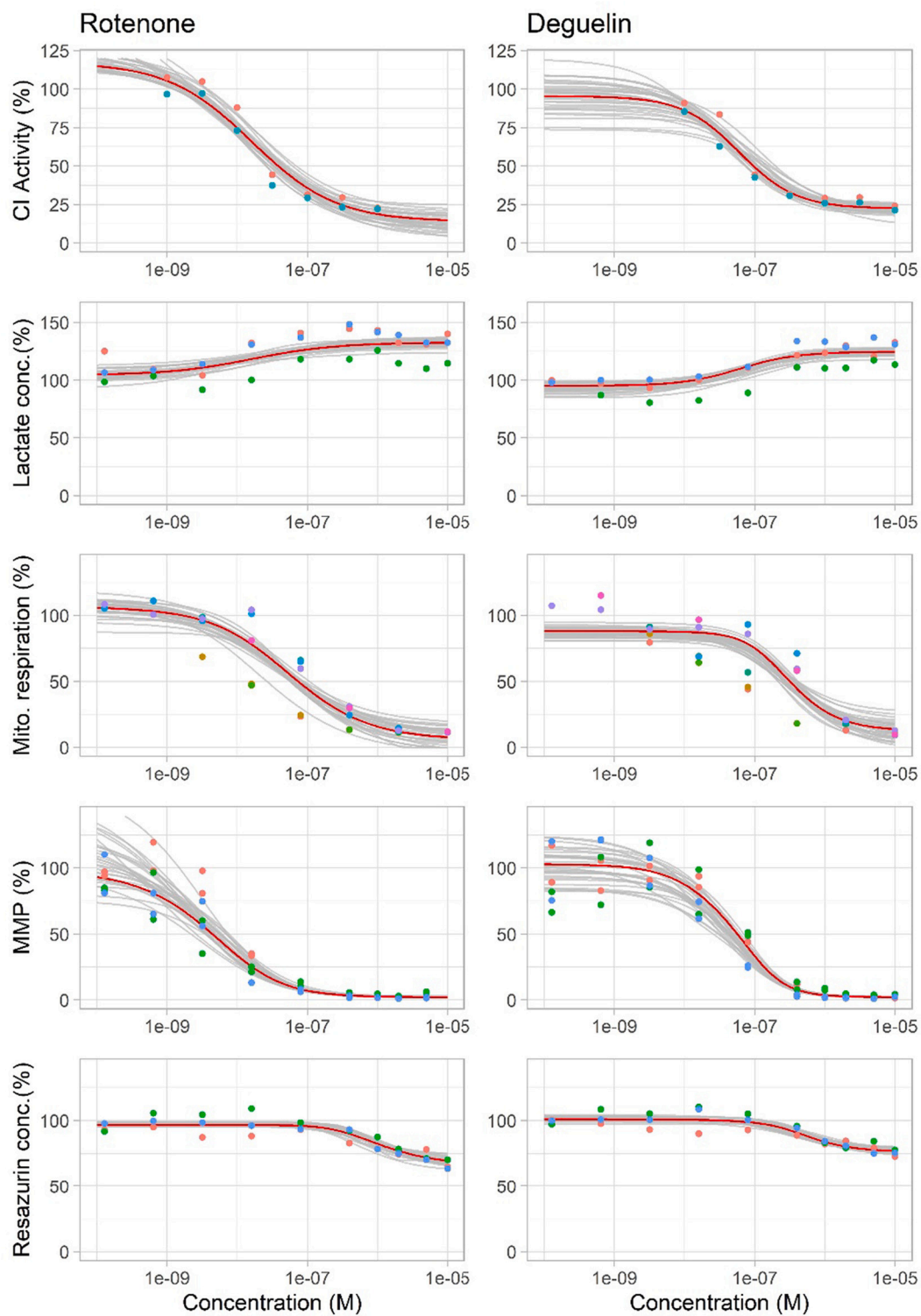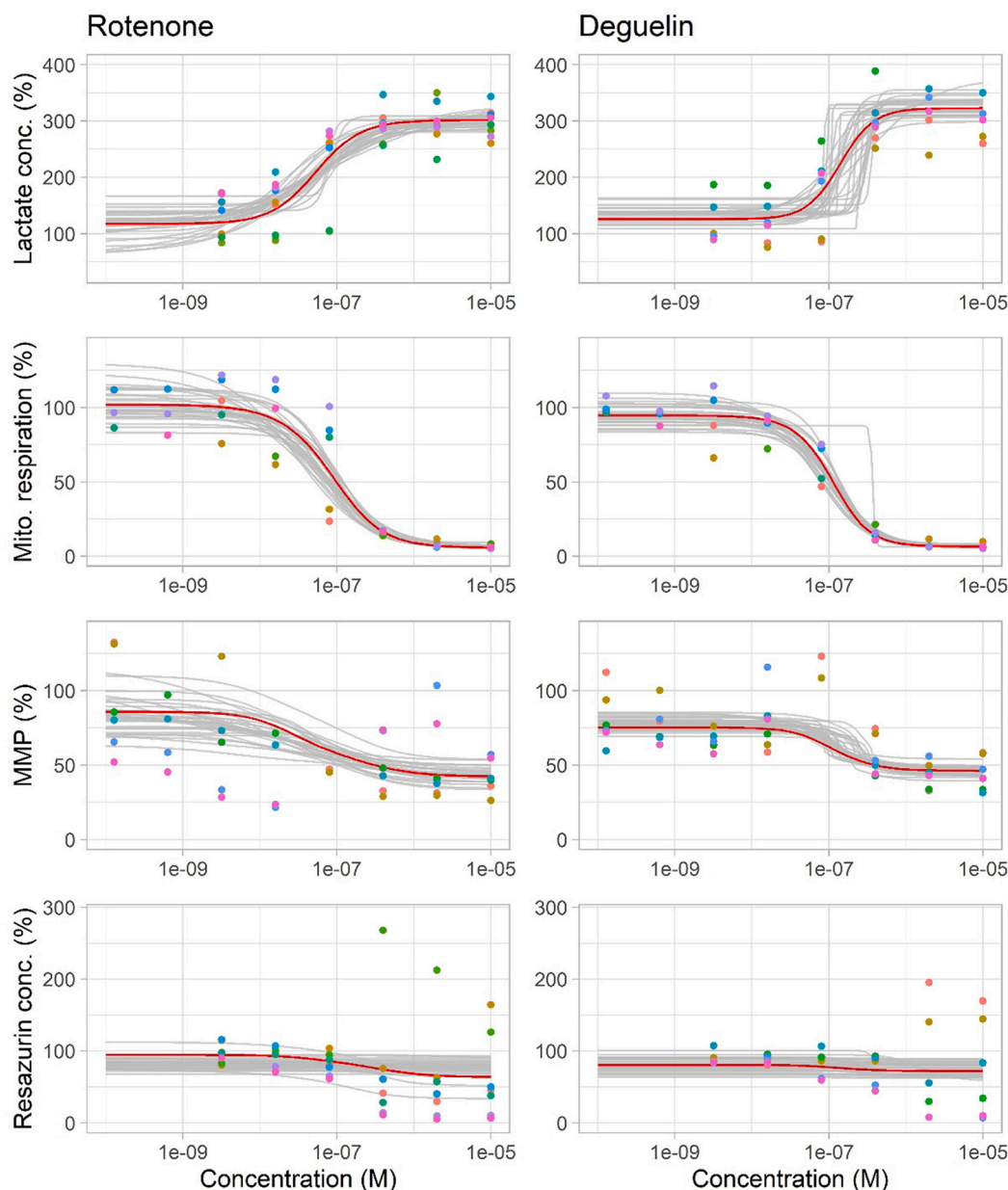
Fig. 7. Predicted and observed dose-response relationships in the mitotoxicity qAOP in renal cells for each readout (complex I inhibition, mitochondrial respiration inhibition, proteasomal activity, neurite area. Red line: predictions obtained with the maximum posterior vector. Grey lines: predictions obtained with 30 random parameter vectors drawn from their joint posterior distribution. Dots: experimental data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

based on the dose-response relationships measured for each readout.

Table 3 lists the functions used to model the KERs in each cell line. None of the KERs are comparable between all three cell lines due to differences in readouts available. However, the concentration-response relationships for the MIE were always modelled as loglogistic, thus inducing sigmoidal concentration-response relationships in subsequent KERs, which were, in all cases, linear between the first few KEs and with more complex KERs models between the last KEs. MR was log-transformed in the renal cell qAOP. MMP, in the liver and renal cell AOPs, was log-transformed.

The predicted concentration-response relationships resulting from the calibration of the KER are represented for each cell line in the following sections. The predicted and observed KERs are provided in SI: observed KERs are obtained by predicting the KE_{n-1} at the concentration levels where KE_n was measured and plotting these values against the

observations of KE_n.

3.1.1. Neuronal cells

The fit of the model after calibration with the rotenone and deguelin data on complex I inhibition, mitochondrial respiration inhibition, and impairment of proteostasis is shown in Fig. 4. In all cases the fits are acceptable: all data points are within a 3-fold factor of the maximum likelihood prediction, all but 5 are within a 2-fold factor. The shape of the model for the last KE, NR, with an intermediate plateau for rotenone, is due to the combined effects of the two precursor KEs: Impaired proteostasis compounds the effect of mitochondrial respiration inhibition at high effect levels.

Table 4 gives a summary of the posterior parameter distributions obtained by MCMC sampling for rotenone and deguelin in neuronal cells. The overlap of parameter distributions for both rotenone and

Table 6

Parameter estimates in the renal cells mitotoxicity qAOP in (maximum posterior values and 95% interval of posterior distribution).

KER	Function	Parameter	Rotenone	Deguelin
Conc. → LC	Loglogistic (Eq. 1)	$ymin_{CI}$ (%)	117 [56.4; 156]	126 [104; 164]
		$ymax_{CI}$ (%)	302 [279; 341]	323 [292; 351]
		k_{CI}	1.48 [0.463; 9.48]	2 [1.3; 88]
		$EC50_{CI}$ (nM)	53.1 [18.6; 114]	140 [85.9; 374]
		σ_{CI}	42.2 [36.4; 56]	46.4 [43.3; 67.3]
Conc. → logMR	Loglogistic (Eq. 1)	$ymin_{MR}$ (%)	1.73 [1.48; 2.12]	1.88 [1.65; 2.05]
		$ymax_{MR}$ (%)	4.62 [4.42; 4.81]	4.55 [4.42; 4.68]
		k_{MR}	−1.18 [−1.93; −0.742]	−1.63 [−2.38; −1.04]
		$EC50_{MR}$ (nM)	256 [149; 366]	231 [184; 331]
		σ_{MR}	0.311 [0.243; 0.41]	0.193 [0.169; 0.295]
logMR + LC → logMMP	Linear (Eq. 1)	β_{MMP}	0.0114 [0.00449; 0.0188]	0.00816 [0.0065; 0.0122]
		β_{0MMP}	3.31 [2.72; 3.82]	3.5 [3.17; 3.65]
		σ_{MMP}	0.543 [0.474; 0.728]	0.235 [0.193; 0.301]
logMMP → RZ	Loglogistic (Eq. 1)	$ymin_{RZ}$ (%)	21.5 [2.07; 79.7]	21.1 [2.54; 82.4]
		$ymax_{RZ}$ (%)	97 [67.3; 135]	108 [70.6; 135]
		k_{RZ}	4.43 [0.119; 73.1]	0.884 [0.116; 76.8]
		$EC50_{RZ}$ (%)	40.2 [10.6; 58.2]	31.3 [10.5; 57.2]
		σ_{RZ}	46.6 [42.9; 66.9]	40.1 [34; 54.9]

Table 7

Maximum posterior values and 95% interval of posterior distribution of the EC$_{50}$ for each chemical when the other parameters of the concentration-response relationship ($ymin_{CI}$, $ymax_{CI}$, k_{CI}) are set to the rotenone values.

Chemical	Estimated EC$_{50CI}$ (μM)
Antimycin A	0.0219 [6.06e-05; 10.3]
Azoxystrobin	26.7 [0.00148; 52.6]
Carboxine	360 [219; 19,200]
Fenpyroximate	0.074 [6.18e-05; 11.4]
Mepronil	149 [98.1; 2290]
Pyraclostrobin	6.06 [6.27e-05; 12.2]
Pyrimidifen	0.123 [6.25e-05; 13.5]
Thifluzamide	468 [224; 19,300]

deguelin is strong, apart from the potencies of the chemicals ($EC50_{CI}$) and the slope of the relationship between *PR* and *NR* (k_{NRPR}). Since there is a difference in shape of the KER in the final step of the qAOP (k_{NRPR}), the difference in potency observed at the MIE is not the same as at NR, the final outcome of the AOP. Mitochondrial respiration inhibition appears to be more determinant in neurite area reduction with rotenone than with deguelin.

Modelled KERs are shown in the Supplemental Material (Fig. S16). For both rotenone and deguelin, the shape of the modelled KER implied that, at the lowest tested concentration levels, decrease in neurite area (NR) was caused by MR mostly, since k_{NRPR} was much greater than k_{PR} whereas at higher concentration levels, PR started to cause NR as well. The qAOP equations calibrated for rotenone are represented in Fig. 5 (see Fig. S19 for deguelin), based on the maximum posterior values

(MPV), which are the most likely or equivalently best fit estimates given the uniform priors used.

3.1.2. Liver cells

The mean responses of each KE in liver cells were accurately predicted for both rotenone and deguelin (Fig. 6) (all but one data points were within a 3-fold factor of the maximum likelihood prediction, and all but seven were within a 2-fold factor). KEs up to MMP have similar EC50 levels. Strong non-linearity of the KER occurs between MMP and RZ (Fig. S17): the EC50 is around 100-fold higher for RZ than for MMP.

Parameter distributions overlapped between rotenone and deguelin (Table 5), apart from the EC50 of the MIE concentration-response, which points only to a difference in potency between the two chemicals.

3.1.3. Renal cells

The predicted dose response relationships of each readout for rotenone and deguelin are shown in Fig. 7. MR data was log transformed before analysis since this improved the fit according to LOO validation. Variability was larger than in the same readouts measured in liver cells, for example the estimate of σ_{CI} was less than 10% in liver cells and over 40% in renal cells. This resulted in a larger number of data points that were not well predicted: 10 data points were not within a three-fold factor of the predictions that maximized likelihood, 21 were not within a two-fold factor. In particular, measurements of resazurin levels were highly variable: high levels measured in some replicates at high doses implied that resazurin levels were not modelled as decreasing at high doses (Fig. 7). All parameter distributions overlapped strongly between rotenone and deguelin apart from the EC50 of the concentration-LC relationship (Table 6). The uncertainty on $ymax_{RZ}$ and $xmin_{RZ}$ is likely due to the fact that the high effect plateau of RZ is not well characterized by the data, and results in uncertainty on the KER between MMP and RZ and a high value of σ_{RZ} (see also Fig. S18). The high values of σ_{RZ} and σ_{CI} reflect the fact that most RZ and CI measurements do not reach low levels.

3.2. Predictions based on calibrated qAOPs

As a predictive check, the qAOP calibrated with rotenone or deguelin data in neuronal cells was used to predict the concentration-AO relationships for antimycin A, azoxystrobin, carboxine, fenpyroximate, mepronil, pyraclostrobin, pyrimidifen, and thifluzamide.

The estimated EC$_{50}$ of the MIE (complex I inhibition) concentration-response relationship are reported in Table 7. These were obtained by setting $ymax_{CI}$, $ymin_{CI}$, and k_{CI} to the values obtained with rotenone (see Table S11 for EC$_{50CI}$ obtained by setting $ymax_{CI}$, $ymin_{CI}$, and k_{CI} to the values obtained with deguelin). When the observed responses were either low or high (close to zero or full complex I inhibition), the EC50 could not be accurately estimated, which is the case for antimycin A, fenpyroximate, pyraclostrobin and pyrimidifen. Calibrated dose responses for the MIE are represented in Fig. 8 (see Fig. S26 for those based on the deguelin AOP).

The distributions of all other parameters were then set to the posterior distributions obtained for rotenone and for deguelin separately. Predicted dose-responses for neurite area obtained with the qAOP calibrated for rotenone and chemical-specific EC$_{50CI}$ are represented in Fig. 9. Supplemental Material Figs. S24 and S25 show the intermediate KEs. Based on the MPV, all predictions were within a 3-fold factor for complex I inhibition which suggests the values of $ymax_{CI}$, $ymin_{CI}$, k_{CI} obtained for rotenone are suitable for other chemicals too. Predictions were not always accurate for the chemicals for which an EC50 of the MIE dose-response could be estimated (azoxystrobin, carboxine, mepronil and thifluzamide). The decrease in neurite area occurred at concentration levels approximately 10 times smaller than those predicted to azoxystrobin, carboxine and mepronil, and 1000 times smaller than those predicted for thifluzamide. For the other chemicals, neurite area can be predicted by the qAOP, although with large uncertainty: the data
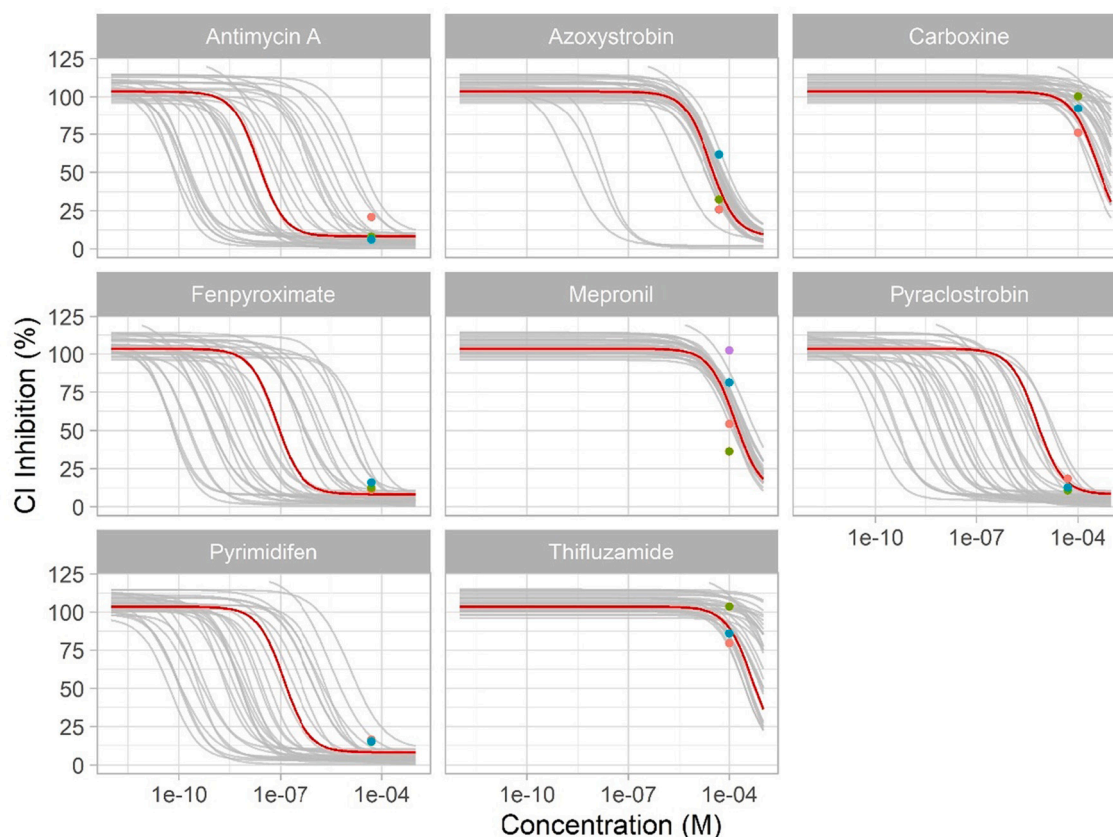
Fig. 8. Modelled dose-responses of complex I inhibition for each validation chemical, when (y_{minCI}, y_{maxCI}, k_{CI}) are set to the rotenone values. Red line uses the chemical-specific EC50$_{CI}$ MPV and other qAOP parameters set to the rotenone MPV. Grey lines: 30 samples from the posterior distribution of chemical-specific EC50$_{CI}$ each combined with one sample from the posterior distribution of the other qAOP parameters calibrated on rotenone. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measured at the MIE only provide a limited amount of information. The entire AOP is represented for two chemicals in Fig. 10.

Supplemental Material Figs. S27, S28, S29 show all predictions obtained using the qAOP calibrated for deguelin and chemical-specific EC50$_{CI}$. The results do not differ markedly from the results obtained with the rotenone-calibrated qAOP.

4. Discussion

The AOP for complex I inhibition-induced mitotoxicity was successfully modelled and calibrated using three cell lines for rotenone and deguelin, based on an extensive dataset of readout dose-responses of each KE. Comparison of the results obtained for rotenone and deguelin suggested chemical independence to a certain extent. However, application of the calibrated qAOP in neuronal cells to other chemicals showed that, in cases where the MIE response could be modelled, neurite area decreased at concentrations smaller than predicted by the qAOP.

4.1. Experimental data

The experimental data used to calibrate the mitotoxicity qAOP was obtained as part of a case study of the EU-ToxRisk project. The data for each of the three cell types was produced in one laboratory within a short time span, which can be considered to be the ideal type of data for calibrating qAOPs and provides an opportunity to model KERs precisely, since variability was reduced to measurement and replicate effects (Spinu, 2021). For technical reasons, the readouts for the various KEs were not identical for the three cell types and all readouts were not measured during the same assay. This adds inter-assay and inter-

readouts variability and complicates quantitative analysis. The Bayesian methods used here can integrate variability and uncertainty in the data into predicted uncertainty on outcomes which are useful in risk assessment (Spinu, 2021) but sources of variability which are poorly understood and quantified cannot be taken into account. Modeling would benefit from all readouts being measured during the same experiment, for example in homogeneous systems like micro-chips (Ramme et al., 2019). However, differences in time scales between KEs and, even more so, between readouts of KEs, may prevent easy analysis of experimental data.

The qAOP developed in this paper is based only on molecular and cellular events; relationships with organism-level events may occur at very different time-scales (Spinu, 2021). In case *in vivo* data is included as the adverse outcome, one cannot expect uncertainty to be any smaller than the variability and uncertainty observed in *in vivo* studies, where confidence intervals around reference doses can reach several orders of magnitude (Ly Pham et al., 2020).

4.2. Model building

Model specification is the main challenge in calibrating quantitative AOPs, as the model should describe correctly the mean response. PSIS diagnostics based on data likelihood did not reveal any severe model miss-specification, which suggests that the choice of three simple models for KER that we suggest (log-logistic, linear, asymptotic) was sufficient for this dataset. Modeling attempts with other KER functions even showed that some freedom is allowed regarding the choice of link functions. In liver cells for example, between complex I inhibition (CI) and MR, the linear function was preferred over the asymptotic
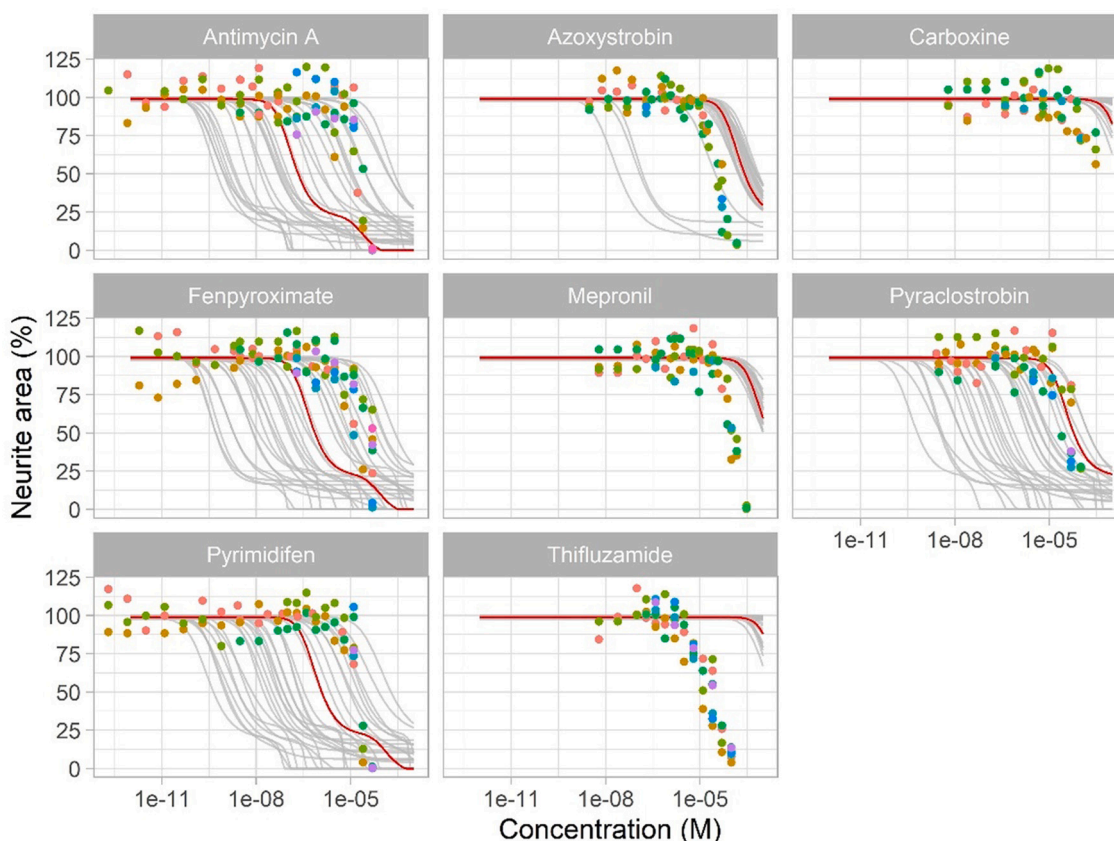
Fig. 9. Predicted dose-responses of neurite area for each validation chemical, when (y_{minCI}, y_{maxCI}, k_{CI}) are set to the rotenone values. Red line uses the MPV for chemical-specific EC50 and qAOP parameters. Grey lines: 30 samples from the posterior distribution of EC50 each combined with a sample from the posterior distribution of the qAOP calibrated on rotenone. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

regression for the sake of parsimony and taking into account the PSIS diagnostics.

However, in some cases such as multiple readouts, strong non-linearities of the KER, separating branches in the AOP, modeling the KER or how the various KERs must be combined can be complex. Strong non-linearity of the KER occurs when the EC50 is vastly different between two adjacent KEs. For example, in the relationship between MMP and RZ in liver cells (Fig. 11), although strong effects were observed on MMP (responses close to 0), RZ was only mildly decreased, at higher concentration levels, in particular for rotenone. As a result, the most-complex of our three KER models had to be used, with a steep slope.

Multiple readouts for one KE represent another challenge in modeling. In liver and renal cells, decrease in mitochondrial respiration had two readouts (LC and MR): consequently, MPP was modelled as a function of the sum of both readouts, assuming that they provide similar information on the same underlying biological event. The qAOP in liver cells were also separately calibrated without MR data and without LC data (results not shown), resulting in slightly larger prediction intervals, and in some cases, increased data likelihood. However, with our approach, the AOP could even be simplified by leaving out KEs, as a response-response relationship would then be defined between two non-adjacent KEs. Including all available information provides a better understanding of the mechanism of action, quantified relationships between new dose-response data obtained at any level of the AOP and a potential outcome, and in the present case, allows us to further study chemical and cell-type independence. Defining the KE may not be critical compared to identification of reliable, widely-used biomarkers as read-outs.

In the AOP for neuronal cells, one KE is directly linked to two of the

subsequent KEs: the AO (NR) is affected directly by two readouts (MR and PR). We assumed that the effect on NR resulted from the multiplication of the effects of MR and PR, a standard assumption of independent action in mixture dose-response modeling. Modeling NR as a weighted sum of MR and PR required a larger number of parameters without any significant improvement in adjustment (results not shown).

While the original AOP dissociates altered mitochondrial respiration and mitochondrial dysfunction, the latter was not measured in neuronal cells, although it might have helped better understand the specificity of neurons' response to rotenone. Furthermore, additional data characterizing the MIE dose response in the validation set might have reduced the mismatch between observations and predictions: only the EC_{50} was estimated because CI was measured at only one concentration.

The statistical component of models fitted to data depends on how the data has been normalized. The data was expressed as percentage of control in each replicate. Prior parameter distributions are therefore unit-independent and easier to define, and modeling replicate effects is not as critical. However, choosing a measurement error model for normalized data is not straightforward. For example, the ratio of two normally distributed measurements is not normally distributed and could be modelled using Chi-square, gamma, or Rayleigh distributions. However, as the choice of equations and accurate modeling of the mean response was our main concern, a standard Gaussian likelihood was selected. Replicate effects are still visible, for example with PR activity, and could be modelled in a multilevel statistical framework (Wilson et al., 2014).
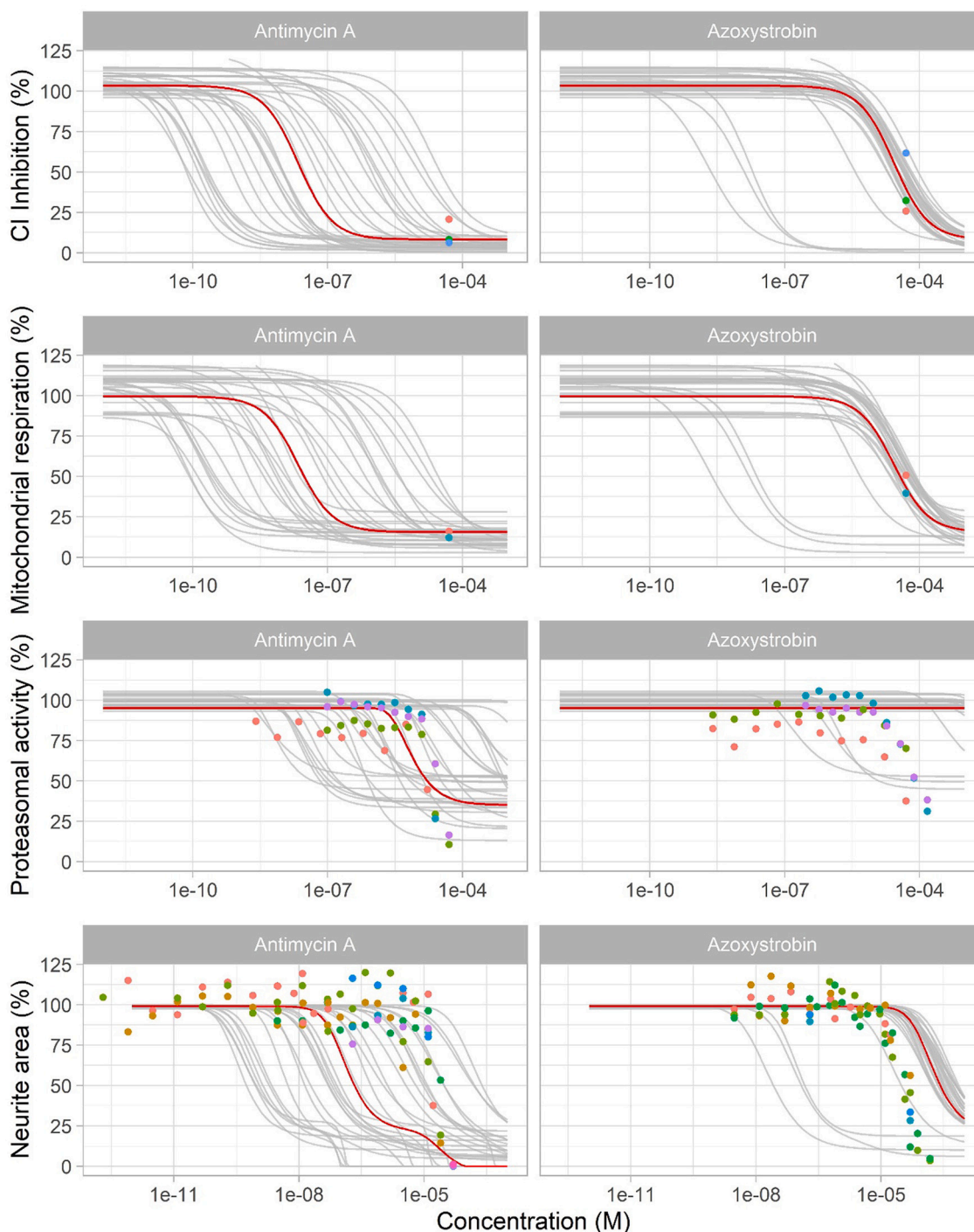
Fig. 10. Predicted dose-responses of each KE for antimycin A and azoxystrobine, when (y_{minCI}, y_{maxCI}, k_{CI}) are set to the rotenone values. Red line uses the chemical-specific EC50$_{CI}$ MPV and other qAOP parameters set to the rotenone MPV. Grey lines: 30 samples from the posterior distribution of chemical-specific EC50$_{CI}$ each combined with one sample from the posterior distribution of the other qAOP parameters calibrated on rotenone. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Are the qAOPs built chemical-independent?

KER forms were the same and parameter estimates were similar for rotenone and deguelin within a given cell type. In the liver and neuronal cells qAOPs, small differences in potency were modelled between the two chemicals at different steps of the AOP. These differences in curvature of some KERs may be attributable to uncertainties in the data. As expected, the main, consistent, difference between the two chemicals was the EC_{50} for the MIE, rotenone being more potent than deguelin.

However, in neuronal cells, the particular shape of the dose-response curves for neurite area (Fig. 4) and the very different parameter estimates between rotenone and deguelin (Table 4) suggest a difference in the interplay between mitochondrial respiration (MR), proteasomal activity (PR), which are concurrent processes toward neurite area. With rotenone, it seems that, as MR shuts down, cell death increases progressively and when MR is low enough, impaired proteostasis gives abruptly a final blow to cell survival (Fig. 4). With deguelin, neurite area decreases at higher effect levels compared to MR. Additional
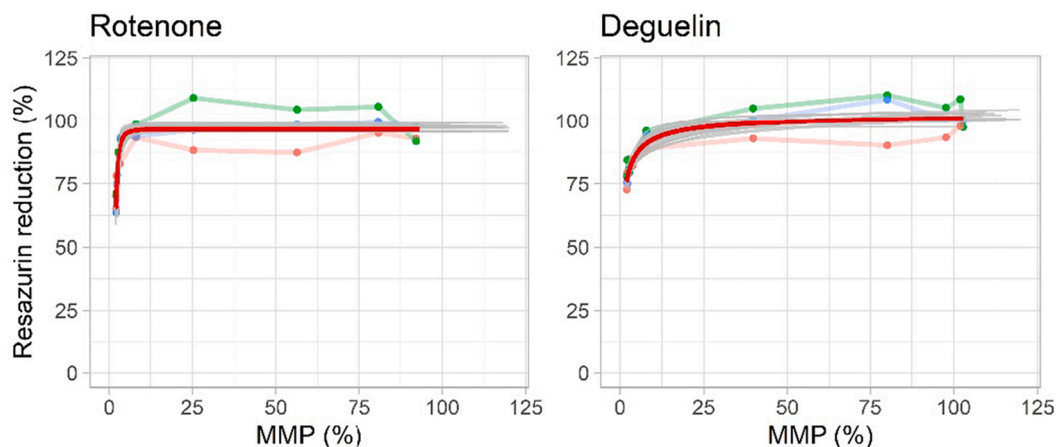
Fig. 11. Predicted and observed relationships between MMP and resazurin in the mitotoxicity qAOP in liver cells, for rotenone and deguelin. Red line: predictions obtained with the maximum posterior vector. Grey lines: predictions obtained with 30 random parameter vectors drawn from their joint posterior distribution. Dots: observations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mechanisms of action for rotenone might help explain the difference between deguelin and rotenone actions toward decreased MR.

Predictivity of the neuronal cells qAOP, calibrated with either rotenone or deguelin, was checked in a validation exercise with data on eight other chemicals: if the qAOP is chemical-independent, inputting chemical-specific MIE activation potencies in the qAOP should lead to a correct prediction of the AO. However, for four of these chemicals (antimycin A, fenpyroximate, pyraclostrobin and pyrimidifen), potency at activating the MIE was uncertain since only an upper bound could be reliably evaluated. Consequently, predicted effects on MR and PR were acceptable but highly uncertain, and the predicted neurotoxicity, even more uncertain, tended to overestimate their toxicity. In a risk assessment context, those chemicals would be flagged as suspect and begging for more data collection about MIE activation. The MIE of the four remaining chemicals (azoxystrobin, carboxine, mepronil, and thifluzamide) was relatively precisely estimated, and their effect on MR was well predicted. However, their effect on PR and the AO were largely underestimated, with over-confidence. In a risk assessment context, these chemicals might be wrongly classified as safe. Lack of validity of the calibrated qAOP in azoxystrobin, carboxine, mepronil, and thifluzamide may be due to differences in mode of action or related to the fact that deguelin and rotenone, the reference chemicals, were active at exposure levels several orders of magnitude lower.

4.4. Are the qAOP built cell-type independent?

The models used for the same KERs across cell-lines were mostly the same, except that, in renal cells, MR had to be log-transformed. The parameters of the calibrated KERs, however, differed between cell types due to differences in KEs studied, differences in readouts, and differences in overall potency, which could reflect differences in organ or cell-type sensitivity or *in vitro* kinetics. The same readouts were measured in the liver and renal cells, apart from CI which was not measured in renal cells. The measurements, in particular LC and RZ, displayed more uncertainty in renal cells, and caused difficulties in modeling RZ. In liver or neuronal cells, the same readout of the MIE was measured, and the potencies were similar. This indicates that, for either rotenone or deguelin, the liver cell qAOP could be used to predict neuronal toxicity, or *vice versa*. This may not work with the kidney cell qAOP, and in liver and kidney cells cell-type independence was not investigated at this stage.

Since rotenone and deguelin are neurotoxicants, neuronal cells could be expected to be more sensitive than liver or kidney cells at equal internal exposure levels. Indeed, although MMP was a sensitive readout in liver and kidney cells (and not measured in neuronal cells), the final

readout in liver and kidney cells, RZ, was not as sensitive as neurite area, NR. However, internal exposure levels depend on the toxicokinetics of the chemicals and may differ between organs and chemicals.

4.5. Perspectives

Risk assessment could benefit from the combination of the qAOP with an exposure model relating external exposure scenarios to internal exposure levels in the target organ cells. Although qAOP based solely on *in vitro* data do not provide readily usable output for risk assessment, quantitative understanding of the KER increase confidence in AOP (Organisation for Economic Co-operation and Development (OECD), 2018). However, a validation process for qAOPs should be defined, in particular for the definition of their domain of validity. Furthermore, even without modeling additional, *in vivo*, data, PBPK models could be used to predict internal concentration levels and historical *in vivo* dose-response data could then be used to check the overall link between cytotoxicity at the organ level and adverse outcome (El-Masri et al., 2016).

A qAOP can also help designing experiments relevant to use in risk assessment. For example, our results point to the importance of characterization of the potency at inducing the MIE. At the same time, quantitative information should be gathered quite uniformly along the chain of KEs, but this is difficult to design *a priori*, given the differences in readouts and methods involved. Added-value of information methods (Hammitt and Shlyakhter, 1999; Yokota et al., 2004) could be applied to a preliminary qAOP to help inform efficient designs. There is a large statistical scientific literature on causality and intervention design using Bayesian networks, and it might be worth exploring its applicability to the experimental analysis of complex toxicity mechanisms.

5. Conclusion

We have developed and demonstrated the use of Bayesian networks (BNs) to AOP quantification. In our hands, BNs required manual determination of the KERs, like in standard dose-response models. The choice of KER model at each step required careful observation of the data, in an iterative procedure. Moreover, it is not always clear which model is best when several readouts of the same KE are available, or when the AOP is strongly nonlinear. Selecting the most biologically relevant readout of each KE may help to simplify qAOP modeling, which could also be improved by computer-assisted determination of the best KER models to use.

To avoid pitfalls in qAOP development, we suggest taking two approaches in parallel: first, a mechanistic modeling path, able to help test

hypotheses, design experiments and deeply understand the results; second, because we cannot always wait to have a fully mechanistic model developed, a lighter statistical approach, such as dose-response based modeling or Bayesian networks.

With the chemicals and cell types studied here, it is not clear whether a qAOP calibrated with data from a particular type of cell could be used to predict accurately the toxicity in another type of cells. The qAOPs for neuronal cells, hepatocytes and kidney cells showed only small differences, but we developed them only for two chemicals and cannot make a general claim of cell-type independence. As to chemical-independence, the validation exercise first shows that it is important to characterize the concentration-response relationship of the MIE accurately, as it is a major input into a qAOP. Ideally, activation of the MIE should be simple and quick to assess; its dose-response relationship should be well documented experimentally. Second, the under-prediction of toxicity, and over-confidence, observed for half of the validation chemicals suggests that one should be cautious when using a qAOP parameterized with only one or a small set of chemicals, and that qAOPs may not be fully chemical-independent if they do not include all the possible mechanisms of action relevant to activation of the AO.

Funding

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tiv.2022.105345.

References

Burgoon, L.D., Angrish, M., Garcia-Reyero, N., Pollesch, N., Zupanic, A., Perkins, E., 2020. Predicting the probability that a chemical causes steatosis using adverse outcome pathway Bayesian networks (AOPBNs). Risk Anal. 40, 512–523. https://doi.org/10.1111/risa.13423.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. *Stan*: a probabilistic programming language. J. Stat. Softw. 76 https://doi.org/10.18637/jss.v076.i01.

Conolly, R.B., Ankley, G.T., Cheng, W., Mayo, M.L., Miller, D.H., Perkins, E.J., Villeneuve, D.L., Watanabe, K.H., 2017. Quantitative adverse outcome pathways and their application to predictive toxicology. Environ. Sci. Technol. 51, 4661–4672. https://doi.org/10.1021/acs.est.6b06230.

Delp, J., Cediel-Ulloa, A., Suciu, I., Kranaster, P., van Vugt-Lussenburg, B.M., Munic Kos, V., van der Stel, W., Carta, G., Bennekou, S.H., Jennings, P., van de Water, B., Forsby, A., Leist, M., 2021. Neurotoxicity and underlying cellular changes of 21 mitochondrial respiratory chain inhibitors. Arch. Toxicol. 95, 591–615. https://doi.org/10.1007/s00204-020-02970-5.

El-Masri, H., Kleinstreuer, N., Hines, R.N., Adams, L., Tal, T., Isaacs, K., Wetmore, B.A., Tan, Y.-M., 2016. Integration of life-stage physiologically based pharmacokinetic models with adverse outcome pathways and environmental exposure models to screen for environmental hazards. Toxicol. Sci. Off. J. Soc. Toxicol. 152, 230–243. https://doi.org/10.1093/toxsci/kfw082.

Foran, C.M., Rycroft, T., Keisler, J., Perkins, E.J., Linkov, I., Garcia-Reyero, N., 2019. A modular approach for assembly of quantitative adverse outcome pathways. ALTEX 36, 353–362. https://doi.org/10.14573/altex.1810181.

Hammitt, J.K., Shlyakhter, A.I., 1999. The expected value of information and the probability of surprise. Risk Anal. 19, 135–152.

Hassan, I., El-Masri, H., Kosian, P.A., Ford, J., Degitz, S.J., Gilbert, M.E., 2017. Neurodevelopment and thyroid hormone synthesis inhibition in the rat: quantitative understanding within the adverse outcome pathway framework. Toxicol. Sci. 160, 57–73. https://doi.org/10.1093/toxsci/kfx163.

Jaworska, J.S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., Miyazawa, M., 2015. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a

decision support system for quantitative weight of evidence and adaptive testing strategy. Arch. Toxicol. 89, 2355–2383. https://doi.org/10.1007/s00204-015-1634-2.

LaLone, C.A., Ankley, G.T., Belanger, S.E., Embry, M.R., Hodges, G., Knapen, D., Munn, S., Perkins, E.J., Rudd, M.A., Villeneuve, D.L., Whelan, M., Willett, C., Zhang, X., Hecker, M., 2017. Advancing the adverse outcome pathway framework-an international horizon scanning approach. Environ. Toxicol. Chem. 36, 1411–1421. https://doi.org/10.1002/etc.3805.

Leist, M., Ghallab, A., Graepel, R., Marchan, R., Hassan, R., Bennekou, S.H., Limonciel, A., Vinken, M., Schildknecht, S., Waldmann, T., Danen, E., van Ravenzwaay, B., Kamp, H., Gardner, I., Godoy, P., Bois, F.Y., Braeuning, A., Reif, R., Oesch, F., Drasdo, D., Höhme, S., Schwarz, M., Hartung, T., Braunbeck, T., Beltman, J., Vrieling, H., Sanz, F., Forsby, A., Gadaleta, D., Fisher, C., Kelm, J., Fluri, D., Ecker, G., Zdrazil, B., Terron, A., Jennings, P., van der Burg, B., Dooley, S., Meijer, A.H., Willighagen, E., Martens, M., Evelo, C., Mombelli, E., Taboureau, O., Mantovani, A., Hardy, B., Koch, B., Escher, S., van Thriel, C., Cadenas, C., Kroese, D., van de Water, B., Hengstler, J.G., 2017. Adverse outcome pathways: opportunities, limitations and open questions. Arch. Toxicol. 91, 3477–3505. https://doi.org/10.1007/s00204-017-2045-3.

Ly Pham, L., Watford, S., Pradeep, P., Martin, M.T., Thomas, R., Judson, R., Setzer, R.W., Paul Friedman, K., 2020. Variability in in vivo studies: defining the upper limit of performance for predictions of systemic effect levels. Comput Toxicol 15, 1–100126. https://doi.org/10.1016/j.comtox.2020.100126.

Oates, C.J., Mukherjee, S., 2012. Network inference and biological dynamics. Ann. Appl. Stat. 6, 1209–1235. https://doi.org/10.1214/11-AOAS532.

Organisation for Economic Co-operation and Development (OECD), 2016. Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). Series on Testing and Assessment, No. 260, ENV/JM/MONO(2016)67. OECD Environment, Health and Safety Publications, Paris, France.

Organisation for Economic Co-operation and Development (OECD), 2018. Users' Handbook Supplement to the Guidance Document for Developing and Assessing AOPs, Second edition. Series on Testing and Assessment, No. 233, ENV/JM/MONO (2016)12. OECD Environment, Health and Safety Publications, Paris, France.

Organisation for Economic Co-operation and Development (OECD), 2021. Adverse Outcome Pathways, Molecular Screening and Toxicogenomics - OECD - Web Page [WWW Document]. URL. https://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm.

Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N., Soldatos, T.G., Kossida, S., Aerts, J., Schneider, R., Bagos, P.G., 2011. Using graph theory to analyze biological networks. BioData Mining 4. https://doi.org/10.1186/1756-0381-4-10.

Perkins, E.J., Ashauer, R., Burgoon, L., Conolly, R., Landesmann, B., Mackay, C., Murphy, C.A., Pollesch, N., Wheeler, J.R., Zupanic, A., Scholz, S., 2019. Building and Applying Quantitative Adverse Outcome Pathway Models for Chemical Hazard and Risk Assessment. Environ. Toxicol. Chem. 38 (9), 1850–1865.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramme, A.P., Koenig, L., Hasenberg, T., Schwenk, C., Magauer, C., Faust, D., Lorenz, A. K., Krebs, A.-C., Drewell, C., Schirrmann, K., Vladetic, A., Lin, G.-C., Pabinger, S., Neuhaus, W., Bois, F., Lauster, R., Marx, U., Dehne, E.-M., 2019. Autologous induced pluripotent stem cell-derived four-organ-chip. Future Sci. OA 5, FSO413. https://doi.org/10.2144/fsoa-2019-0065.

Rovida, C., Alépée, N., Api, A.M., Basketter, D.A., Bois, F.Y., Caloni, F., Corsini, E., Daneshian, M., Eskes, C., Ezendam, J., Fuchs, H., Hayden, P., Hegele-Hartung, C., Hoffmann, S., Hubesch, B., Jacobs, M.N., Jaworska, J., Kleensang, A., Kleinstreuer, N., Lalko, J., Landsiedel, R., Lebreux, F., Luechtefeld, T., Locatelli, M., Mehling, A., Natsch, A., Pitchford, J.W., Prater, D., Prieto, P., Schepky, A., Schüürmann, G., Smirnova, L., Toole, C., van Vliet, E., Weisensee, D., Hartung, T., 2015. Integrated testing strategies (ITS) for safety assessment. Altern. Lab. Anim 32, 25–40. https://doi.org/10.14573/altex.1411011.

Sachana, M., Leinala, E., 2017. Approaching chemical safety assessment through application of integrated approaches to testing and assessment: combining mechanistic information derived from adverse outcome pathways and alternative methods. Appl. In Vitro Toxicol. 3, 227–233. https://doi.org/10.1089/aivt.2017.0013.

Spinu, N., 2021. Modelling of Quantitative Adverse Outcome Pathways (Doctoral). Liverpool John Moores University. https://doi.org/10.24377/LJMU.t.00015012.

Spinu, N., Bal-Price, A., Cronin, M.T.D., Enoch, S.J., Madden, J.C., Worth, A.P., 2019. Development and analysis of an adverse outcome pathway network for human neurotoxicity. Arch. Toxicol. 93, 2759–2772. https://doi.org/10.1007/s00204-019-02551-1.

Stan Development Team, 2020. RStan: the R interface to Stan.

van der Stel, W., Carta, G., Eakins, J., Darici, S., Delp, J., Forsby, A., Bennekou, S.H., Gardner, I., Leist, M., Danen, E.H.J., Walker, P., van de Water, B., Jennings, P., 2020. Multiparametric assessment of mitochondrial respiratory inhibition in HepG2 and RPTEC/TERT1 cells using a panel of mitochondrial targeting agrochemicals. Arch. Toxicol. 94, 2707–2729. https://doi.org/10.1007/s00204-020-02792-5.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., Gelman, A., 2020. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models.

Vehtari, A., Gelman, A., Gabry, J., 2017. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.". Statist. Comput. 27 (5), 1413–1432.

Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C. A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M., 2014. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol. Sci. 142, 312–320. https://doi.org/10.1093/toxsci/kfu199.

Vinken, M., 2013. The adverse outcome pathway concept: a pragmatic tool in toxicology. Toxicology 312, 158–165. https://doi.org/10.1016/j.tox.2013.08.011.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686.

Wilson, A., Reif, D.M., Reich, B.J., 2014. Hierarchical dose-response modeling for high-throughput toxicity screening of environmental chemicals. Biometrics 70, 237–246. https://doi.org/10.1111/biom.12114.

Yokota, F., Gray, G., Hammitt, J.K., Thompson, K.M., 2004. Tiered chemical testing: a value of information approach. Risk Anal. 24, 1625–1639.

Zgheib, E., Gao, W., Limonciel, A., Aladjov, H., Yang, H., Tebby, C., Gayraud, G., Jennings, P., Sachana, M., Beltman, J.B., Bois, F.Y., 2019. Application of three approaches for quantitative AOP development to renal toxicity. Comput. Toxicol. 11, 1–13. https://doi.org/10.1016/j.comtox.2019.02.001.