HAL
open science

Bioinformatic and modelling approaches for a system-level understanding of oxidative stress toxicity

Elias Zgheib

To cite this version:

Elias Zgheib. Bioinformatic and modelling approaches for a system-level understanding of oxidative stress toxicity. Quantitative Methods [q-bio.QM]. Université de Technologie de Compiègne, 2018. English. NNT : 2018COMP2464 . tel-02088169

HAL Id: tel-02088169

https://theses.hal.science/tel-02088169v1

Submitted on 2 Apr 2019

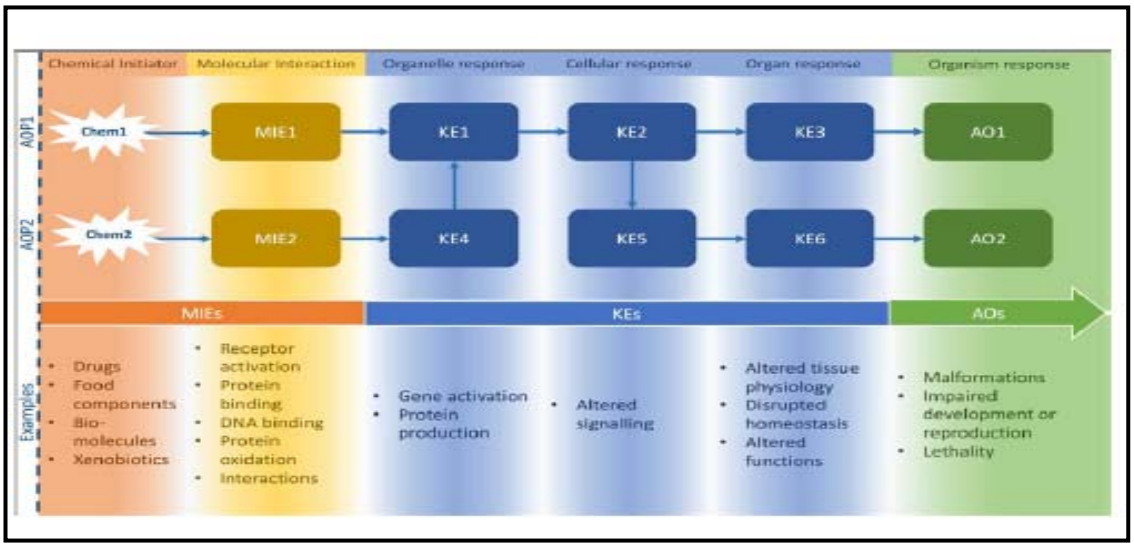
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Par **Elias ZGHEIB**

Bioinformatic and modelling approaches for a system-level understanding of oxidative stress toxicity

Thèse présentée pour l'obtention du grade de Docteur de l'UTC



Soutenue le 18 décembre 2018
Spécialité : Bio-ingénierie et Mathématiques Appliquées : Unité de Recherche Biomécanique et Bio-ingénierie (UMR-7338)

D2464

BIOINFORMATIC AND MODELLING APPROACHES FOR A SYSTEM-LEVEL UNDERSTANDING OF OXIDATIVE STRESS TOXICITY

A THESIS SUBMITTED TO THE

UNIVERSITE DE TECHNOLOGIE DE COMPIEGNE

SORBONNE UNIVERSITES

LABORATOIRE DE BIO-MECANIQUE ET BIOINGENIERIE

UMR CNRS 7338 – BMBI

18TH OF DECEMBER 2018

For the degree of Doctor

Spécialité : Bio-ingénierie et Mathématiques Appliquées

Elias ZGHEIB

SUPERVISED BY

Prof. Frédéric Y. BOIS

JURY MEMBERS

Mme.	Karine AUDOUZE	Rapporteur
Mr.	Vincent FROMION	Rapporteur
Mme.	Cécile LEGALLAIS	Examiner
Mr.	Maxime CULOT	Examiner
Mr.	Frédéric Y. BOIS	Supervisor

TABLE OF CONTENTS

Table of Contents **2**
Acknowledgements **5**
List of Abbreviations **7**
List of Figures **9**
List of Tables **14**
1 Introduction **16**
2 Bibliography **24**
2.1 Toxicology 24
2.1.1 Definition of Toxicity 24
2.1.2 Predictive Toxicology: Prevention 24
2.1.3 Birth of Toxicology 25
2.1.4 Limitations of Traditional Toxicology 25
2.1.5 A Paradigm Shift in Toxicology 26
2.1.6 Modern Toxicology 33
2.2 Biology Underlying Toxicology 38
2.2.1 Oxidative Stress, Nrf2 and some Associated Pathways 39
2.2.2 Systems Biology – SB 45
2.2.3 Adverse Outcome Pathways – AOP 47
2.3 Mathematical Considerations 49
2.3.1 Ordinary Differential Equations – ODE – Systems 49
2.3.2 Michaelis-Menten – MM – Kinetics 51
2.3.3 The Hill Equation 53
2.3.4 Bayesian Statistical Tools 57
2.3.5 Model's Calibration 61
3 Construction of Systems Biology Model of Nrf2 Control of Oxidative Stress **64**
3.1 Starting Models 64
3.1.1 The model of 'Hamon *et al.* (2014)' 64
3.1.2 The model of 'Geenen *et al.* (2012) and Reed *et al.* (2008)' 65
3.2 Methods 66
3.2.1 Remodelling Hamon's model 66
3.2.2 Assembling two models 71
3.3 Results 74
4 SB and other Tools for the Development of quantitative AOPs **78**
4.1 Study Context 78
4.2 Methods 80

4.2.1 Experimental data 80
4.2.2 Chronic Kidney Disease – CKD – AOP 81
4.2.3 Dose-Response based qAOP 82
4.2.4 Bayesian Network – BN – qAOP 83
4.2.5 The Systems Biology – SB – Model 85
4.2.6 Parameter Estimation 87
4.2.7 Uncertainty propagation 89
4.2.8 Software 89
4.3 Results 90
4.3.1 Dose-Response based qAOP Model 90
4.3.2 Bayesian Network – BN – qAOP Model 93
4.3.3 System biology – SB – Model 95
4.4 Discussion 101
4.5 Conclusion 107
5 Investigation of Nrf2, AhR and ATF4 Activation in Toxicogenomic Databases 109
5.1 The General Approach 109
5.2 Material and Methods 111
5.2.1 Generation of Target Gene Lists 111
5.2.2 Construction of a Chemical-Effects Transcriptomics Database 112
5.2.3 Data Sources 114
5.2.4 Bioinformatics Methods 116
5.2.5 Pathway's Signature-Based Prioritization of Chemicals 120
5.3 Results 123
5.3.1 Pathways' Global Signatures 123
5.3.2 Pathways' Stratified Signatures in Liver 127
5.3.3 Human Liver Category: Pathway's Signature-Based Prioritization of Chemicals 132
5.4 Discussion 136
5.1 Conclusion 139
6 Thesis Summary and Conclusion 140
Dissemination Activities 144
References 145
7 Appendix – Supplementary Material 162
7.1 Supplementary Information for Chapter 3 162
7.2 Supplementary Information for Chapter 4 170
7.2.1 Experimental Data 170
7.2.2 Statistical Dose-Response based qAOP Model 175
7.2.3 Bayesian Network qAOP – Node to node relationships 182

7.2.4 SB Model Validation 188
7.2.5 Effectopedia Implementation 192
7.3 Supplementary Information for Chapter 5 194
7.4 The Nrf2 SB Model Code 198
Abstract 217
Résumé 217

ACKNOWLEDGEMENTS

First and foremost I want to thank my supervisor Frédéric Y. Bois for his time and support. He was always ready to help and give constructive advice when I needed; and when I was behind in work or skills, he showed patience and kindness. I am grateful for his fundamental and uplifting contribution to my professional life. I owe him a lot in this area, it has been an honor to be his PhD student.

I thank Karine Audouze and Vincent Fromion for accepting to be my thesis rapporteurs and Cécile Legallais and Maxime Culot for being part of the jury.

The research leading to the results mentioned in this thesis has received support from the Innovative Medicines Initiative Joint Undertaking (IMIJU) under grant agreement number 115439, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. This manuscript reflects only the author's views and neither the IMI JU nor EFPIA nor the European Commission are liable for any use that may be made of the information contained therein.

For their work and contribution to the administrative part of the thesis, I thank Salima Aaras-Andaloussi, Catherine Marque, Catherine Lacourt, Alexandra Cousin, Stéphanie Rossard, Joelle Chaumette from UTC.

I thank members of METO (Models for Ecotoxicology and Toxicology) research unit of INERIS: senior scientists (Céline Brochot, Rémy Beaudouin, Enrico Mombelli, Cléo Tebby and Florence Zeman) and PhD students (Claudio-Ileana Cappelli, Marc Codaccioni, Viviane David, Paul Quindroit, Wang Gao, Ségolène Siméon and Delphine). Even if I was not officially a part of METO, they did all the efforts possible to make me feel like one of them and make my visits to INERIS pleasant.

I thank Paul Jennings for receiving me in his laboratory in the Medical University of Innsbruck (Austria) back in fall 2015. I also thank Alice Limonciel and Anja Wilmes who helped me a lot in my work in Innsbruck. Paul, Alice and Anja actively participated to the analysis, writing and correction of an important part of two journal articles related to the thesis.

I thank all people who stood by my side and shared with me cheerful and difficult moments of my extra-professional time during these four years. Their presence and support contributed directly or indirectly to the success of this project: In Lebanon (Jennifer Daou, Valérie Daou, Christian Khoury, Samir Khouzam, Eliane Makhoul, Habib, Céline & Melissa Safi, Myriam Saliba), in Compiègne (Sally Al Kantar, Mohammad-Ali Assaad, Nancy Chaaya, Jean-Baptiste Gachignard, Rita Maalouf, Jacqueline Maximilian, Hugues Richard de Monti, Halim Tannous), in Paris region (Ariane & Bertrand Chevalier-Chantepie, Arnauld de Genouillac, Philippe de Maistre, Digóenes Flima, Andreas Garshol Sørensen, Mildred-Jeanne and Manfred Hawran, Timothée Jolivet, Clément Lescat, Christophe Roger), in Lille (Marthe & Rémi Dujardin, Patrick Gonçalves, Brice Laureau, Anne-Claire Lozier, Nicolas Maury, Agathe Purcell, Camille & Paul) and Stéphanie & Maxime Dechelette.

Most of all, I thank my parents Kamil & Salwa without whom nothing of all of this would have been possible. I would like to thank my dearest sisters and brothers-in-law for they love and support, Clara & Doumit Sfeir and Nancy & Jean Tabet. Jean Tabet was the first person to encourage me to go forward, embark in a thesis project and discover the world of research.

Success and progress are to be celebrated with people we love, or in their memory –

Julian A. Khoury (1993 – 2014).

LIST OF ABBREVIATIONS

AhR: aryl hydrocarbon receptor

AO: adverse outcome

AOP: adverse outcome pathway

ARE: antioxidant response element

ARNT: AhR nuclear translocator

ATF4: activating transcription factor 4

ATP: adenosine triphosphate

BN: Bayesian network

CAC: chemical activation capacity

ChIP: chromatin immunoprecipitation

CKD: chronic kidney disease

CYP: cytochrome P450

DBN: dynamic Bayesian network

DCF: 2',7'-dichlorofluorescein

carboxy-DCF: 6-carboxy-2',7'-dichlorofluorescein

DRE: dioxin response element

ER: endoplasmic reticulum

FC: fold changes

GCL: glutamate cysteine ligase

GCLC: glutamate cysteine ligase catalytic subunit

GCLM: glutamate cysteine ligase modifier subunit

γGC: gamma-glutamyl-cysteine

GPX: glutathione peroxidase

GR: glutathione reductase

GS: glutathione synthetase

GSH: glutathione

GST: glutathione S-transferase

HTS: high-throughput screening

ITS: integrative testing strategy

$KBrO_3$: potassium bromate

KE: key event

Keap1: kelch-like-ECH-associated protein 1

KER: key events relationship

MCMC: Markov chain Monte Carlo

MIE: molecular initiating event

MM: Michaelis-Menten

MRP: multidrug-resistance protein

Nrf2: nuclear factor (erythroid-derived 2)-like 2

nucNrf2: nuclear Nrf2

ODE: ordinary differential equation

OECD: organization for economic cooperation and development

PERK: RNA (PKR)-like endoplasmic reticulum kinase

qAOP: quantitative adverse outcome pathway

RFU: relative fluorescence units

ROS: reactive oxygen species

SB: systems biology

TF: transcription factor

nucX-AhR: the nuclear complex formed by the association of a xenobiotic X and AhR

QIVIVE: quantitative *in vitro-in vivo* extrapolation

LIST OF FIGURES

Figure 1. Schematic representation of two theoretical interacting AOPs. Through a timeline, different sections correspond to AOP levels (boxes represent the events, some examples are available in the lower part; arrows correspond to KERs).

Figure 2. Timeline illustrating the birth and development of toxicology from first *in vivo* experiments by Paracelsus up to HTS initiatives in the European Union and the United States of America (Zgheib *et al.*, 2017).

Figure 3. Standard Microplates of **(A)** 96-, **(B)** 384-, and **(C)** 1536-well formats respectively.

Figure 4. Schematic representation of the Nrf2 signaling pathway in basal unstressed condition as well as under its activation by oxidative (or electrophilic) stress (Taguchi *et al.*, 2011).

Figure 5. Plot of enzymatic reaction's velocity v against substrates concentration [S] in 10 different cases for Hill's coefficient α gradually increasing from 1 (hyperbolic: Michaelis-Menten case) to 10 (all other curves (2 to 10) are S-shaped) (Duke, Modeling Cooperativity).

Figure 6. Prior, likelihood and posterior distributions for θ. The 'posterior inference' is a formal compromise between the 'observed evidence' (likelihood), summarizing the 'prior distribution' of the data alone (Bayesian Analysis for a Logistic Regression Model - MATLAB & Simulink Example).

Figure 7. Schematic overview of the assembled SB model. This model covers both transcriptional and biochemical aspects of GSH synthesis and metabolism and its control by the Nrf2-Keap1 signaling pathway. The blue compartment is cytosol and the red one is nucleus. Blue arrows show reactant(s):product(s) exchange during biochemical or transport reactions, and red arrows indicate enzymatic catalysis (diamond heads) or gene transcription (round heads). In the nucleus, red boxes represent genes and arrows indicate gene activation. Names

of genes are in orange, of mRNA are in green, of enzymes are in purple, of other proteins and metabolites in blue and of extracellular constants in yellow.

Figure 8. Venn diagram showing the contribution (overlapping areas) of different source models (*i.e.* Hamon *et al.* (2014) in green, Geenen *et al.* (2012a) in pink and Reed *et al.* (2008) in orange) to our final assembled SB model (in blue) describing the control of oxidative stress by the Nrf2-Keap1 signaling pathway. This diagram also shows the parts of each model that were left out (non-overlapping areas). Two more genes (*i.e.*, *SRXN1* and *HMOX1*) that are often used as activation markers for Nrf2 pathway were added to the model (yellow).

Figure 9. γGC and GSH synthesis reactions according to Geenen *et al.* (2012a) (left) and to Hamon *et al.* (2014) (right). [Cys = cysteine, Glut = glutamate, glc and r-GC = gamma-glutamyl-cysteine; other acronyms are explained in the 'List of abbreviations'].

Figure 10. MCMC curve fitting of *GCLC* mRNA (example of gene activated by one single TF) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (red dots) and Hill-based (black curve) SB models.

Figure 11. MCMC curve fitting of *MRP* mRNA (example of gene activated by two TFs) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (colored dots) and Hill-based (colored curves) SB models. nucNrf2 dose increase is operated over time (every 400,000 seconds) and nucX-AhR dose is displayed on different curves (0 (red), 0.5 (orange), 1 (green), 10 (blue) and 100 (magenta) zeptomols of nucX-AhR).

Figure 12. A CKD AOP diagram. KERs are represented by arrows.

Figure 13. Structure of the DBN qAOP for CKD. $KBrO_3$ concentration and the GSH readout do not vary with time, while the DCF and lactate readouts were observed at different time intervals. The arrows indicate probabilistic dependencies.

Figure 14. $KBrO_3$ and DCF specific reactions of the SB model. Other abbreviations: extGSH is extra-cellular glutathione; cytGSH: cytosolic glutathione; extGSSG: extra-cellular

oxidized glutathione; cytGSSG: cytosolic oxidized glutathione. Reactions are represented by red circles: **a.** the oxidation of extGSH by $KBrO_3$; **b.** oxidation of carboxy-DCF by ROS; **c.** DCF bleaching; **d.** oxidation of carboxy-DCF by $KBrO_3$; **e.** oxidation of cytGSH by $KBrO_3$.

Figure 15. Fit of the $KBrO_3$ - GSH data (circles; each color represents one of the replicates) using the three qAOP models developed. The black line corresponds to the empirical model (equation 4.1). The best fit (solid line) is shown along with 20 additional random fits (gray), showing the uncertainty of the model predictions. The black dashed line represents the best fit obtained the DBN qAOP. The red line shows the best fit for the SB model.

Figure 16. Fit (top row) and predictions (bottom row) of the dose-response based qAOP for the DCF (measured in RFU) (left) and lactate (right) readouts. The best fit surfaces (gray) are plotted along with all individual data (colored dots). The predicted chemical-independent relationships (in red) for GSH - time - DCF, or GSH - time - lactate were obtained by inversion of the qAOP equations (see 'Supplementary Material 7.2.2'). The maximum posterior parameter values given in **Table S5** were used to draw the figures.

Figure 17. Fit (top row) and predictions (bottom row) of the DBN qAOP for the DCF (measured in RFU) (left) and lactate (right) readouts. The best fit surfaces (gray) are plotted along with the data mean (black dots) and all individual data (colored dots). The predicted chemical-independent relationships (in red) are shown for GSH - time - DCF and GSH - time - lactate. The maximum posterior parameter values given in **Table S7** were used to draw the figures.

Figure 18. Best fits of SB model (gray surfaces) to the DCF RFU data (colored dots), for different levels of complexity: **(A)** action of $KBrO_3$ on external GSH and formation of DCF by ROS; **(B)** same as A with the addition of DCF bleaching; **(C)** same as B with the addition of a direct formation of DCF by $KBrO_3$; **(D)** same as B, but with the addition of an action of $KBrO_3$ on internal GSH.

Figure 19. Venn diagram of the number of genes of each of the three studied pathways (AhR, Nrf2 and ATF4) and their overlapping zones, included in the analysis.

Figure 20. Methods summarizing workflow

Figure 21. Geometric representation of chemical specificity and potency for the Nrf2 and AhR pathways. *K* represents a chemical and its coordinates are ($CAC_{AhR,\,K}$, $CAC_{Nrf2,\,K}$). *K* also defines the vector $\overrightarrow{\boldsymbol{OK}}$ linking the origin *O* (0, 0) to point *K*. The absolute value of the cosine of the angle α between $\overrightarrow{\boldsymbol{OK}}$ and a pathway's axis can be used to measure the specificity of a chemical for the given pathway (the smaller α, the more specific the chemical). On the other hand, the overall activation potency of a chemical increases proportionally with the length of $\overrightarrow{\boldsymbol{OK}}$. Points *A*, *B* and *C* represent three other chemicals with different specificities and potencies for pathways' activation (see text).

Figure 22. Venn diagram of the number of genes per pathway's global signatures and names of genes of overlapping zones.

Figure 23. Network representation of AhR, Nrf2 and ATF4 pathway signatures and their overlapping zones.

Figure 24. Venn diagrams of the number of genes per pathway's stratified signatures and names of genes of overlapping zones. Categories: **(A)** All liver data, **(B)** Rat Liver *in vitro* data, **(C)** Rat Liver *in vivo* data, **(D)** Human Liver *in vitro* data. *Refers to genes that were known to be part of the same overlapping zone according to **Table S11** lists. White is the color of gene names that appear in an overlapping zone of only one of the four categories studied, and black is the color of gene names that appear in more than one category (two, three or four).

Figure 25. Distribution of chemicals by potency (Y-axis: module $\|\overrightarrow{\boldsymbol{OK}}\|$ of the vector linking the origin O(0,0) to the chemical's point in a 3D space) and specificity to the AhR pathway (X-axis: the absolute value of the $|\boldsymbol{cos}(\boldsymbol{\alpha})|$ of the angle between $\overrightarrow{\boldsymbol{OK}}$ and the AhR axis in a 3 D space). Chemicals are represented by their rank in the alphabetically ordered list.

Chemicals that are both strong (horizontal blue dashed line: $\|\overrightarrow{\boldsymbol{OK}}\| > \mathbf{0.5}$) and ATF4 specific (vertical blue dashed line: $\boldsymbol{cos}(\boldsymbol{\alpha}) = \frac{1}{\sqrt{3}}$) are in red and their names are listed in the legend on the right.

Figure 26. Distribution of chemicals by potency (Y-axis: module $\|\overrightarrow{\boldsymbol{OK}}\|$ of the vector linking the origin O(0,0) to the chemical's point in a 3D space) and specificity to the Nrf2 pathway (X-axis: the absolute value of the $|\boldsymbol{cos}(\boldsymbol{\alpha})|$ of the angle between $\overrightarrow{\boldsymbol{OK}}$ and the Nrf2 axis in a 3D space). Chemicals are represented by their rank in the alphabetically ordered list. The only chemical that is both strong (horizontal blue dashed line: $\|\overrightarrow{\boldsymbol{OK}}\| > \mathbf{0.5}$) and Nrf2 specific (vertical blue dashed line: $\boldsymbol{cos}(\boldsymbol{\alpha}) = \frac{1}{\sqrt{3}}$) Sulindac, is in red and it is listed in the legend on the right.

Figure 27. Distribution of chemicals by potency (Y-axis: module $\|\overrightarrow{\boldsymbol{OK}}\|$ of the vector linking the origin O(0,0) to the chemical's point in a 3D space) and specificity to the ATF4 pathway (X-axis: the absolute value of the $|\boldsymbol{cos}(\boldsymbol{\alpha})|$ of the angle between $\overrightarrow{\boldsymbol{OK}}$ and the ATF4 axis in a 3D space). Chemicals are represented by their rank in the alphabetically ordered list. Chemicals that are both strong (horizontal blue dashed line: $\|\overrightarrow{\boldsymbol{OK}}\| > \mathbf{0.5}$) and ATF4 specific (vertical blue dashed line: $\boldsymbol{cos}(\boldsymbol{\alpha}) = \frac{1}{\sqrt{3}}$) are in red and their names are listed in the legend on the right.

LIST OF TABLES

Table 1. The 36 partners of the StemBANCC project listed in alphabetical order after the names of the two leaders: F. Hoffmann-La Roche Ltd and University of Oxford.

Table 2. Toxicity-testing options defined by the 'Toxicity testing in the 21st century: A vision and a strategy' report (National Research Council, 2007) in order to enhance the paradigm shift in toxicity research (Zgheib *et al.*, 2017).

Table 3. Virtual exposure scheme applied on both Hamon's (old) and Hill-based (new) SB models to perform MCMC curve fitting and establish equivalency between them. Genes that are activated by a single TF (*i.e.*, *CYP*, *GS*, *GCLC* and *GCLM*) were exposed to five doses (one dose per time-point) ranging from 0 to 100 zeptomol doses of their TF (*i.e*, nucNrf2 or nucX-AhR). Genes that are activated by both TFs (*i.e.*, *Nrf2*, *GST*, *GPX* and *MRP*) were exposed to five different and separate combinations of doses per time-point (25 combinations are possible). All exposures are in zeptomol.

Table 4. Hill parameter values (maximum posterior values) for gene transcription in the SB assembled model of the Nrf2 control of oxidative stress. These values were obtained by MCMC simulations. Since calibration was performed with virtual data, we were not interested in the mean and the standard deviation of the distributions (not mentioned).

Table 5. Prior distributions of the parameters of the SB qAOP calibrated with the DCF data.

Table 6. Summary of the posterior distribution of the five SB model parameters describing the action of $KBrO_3$ on the formation of DCF. The best parameterization (setting $k_{GSHc,KBrO3}$ at zero) is presented.

Table 7. Assessment of the SB model fit to the $KBrO_3$ - time - DCF data using various criteria and for increasing model complexity. The various steps explain the main text of 'Methods 4.2.5'. **Step 1** is omitted since it does not require DCF data (parameter $k_{GSHe,KBrO3,}$

quantifying the action of $KBrO_3$ on extra-cellular GSH, was independently calibrated from the GSH data and set to its maximum likelihood value in all cases). The other parameters were introduced as follows: **Step 2**: action of KBrO3 on external GSH and formation of DCF by ROS (parameter $k_{DCF,ROS}$); **Step 3**: adding DCF bleaching (parameter k_{bl}); **Step 4a**: adding a direct formation of DCF by $KBrO_3$ (parameter $k_{DCF,KBrO3}$); **Step 4b**: same as step 3, plus adding an action of $KBrO_3$ on internal GSH (parameter $k_{GSHc,KBrO3}$); **Step 5**: all parameters added.

Table 8. Number of chemicals used in each experimental category.

Table 9. Chosen pathway specific chemical through the dataset.

Table 10. Number of conditions (chemicals, concentrations, time-points) tested per category.

Table 11. Pathway's global signatures for AhR, Nrf2 and ATF4 pathways and the signatures of their overlapping zones for all available data. Gray background indicates genes that appear in the signature of the pathway from previous studies (**Table S11**) and confirmed here. Non-grayed out values are novel allocations from this analysis.

Table 12. AhR, Nrf2 and ATF4 pathways' signatures stratified in liver data and by all liver data sub-categories ('Rat Liver *in vitro*' data, 'Rat Liver *in vivo*' data and 'Human Liver *in vitro*' data).

1 INTRODUCTION

A xenobiotic is an extrinsic chemical that is foreign to a certain living organism and its metabolism (Croom, 2012). Xenobiotics, either natural or artificially conceived, can be components of daily life's ordinary objects (*e.g.* clothes, food, drugs, jewels, paintings, skincare products, plastic cups, pesticides *etc*.). Upon exposure, interactions between xenobiotics and biomolecules may elicit a perturbation in local biology and impair critical physiological functions of the organism. In fact, for some xenobiotics (*e.g.* pharmaceuticals), despite the strictly regulated toxicological control they undergo, unexpected adverse reactions may emerge leading to their failed licensing or even post-licensing withdrawal from market (Geenen *et al.*, 2012). Thus, potential toxic impact of xenobiotics on human health is becoming of major clinical and socio-economic concern.

Toxicology can be defined as the science that examines the negative biological repercussions of xenobiotics on l iving organisms (Gundert-Remy *et al.*, 2015). The main societal goal of toxicology is to develop reliable predictions of the human health impact of exposures to chemicals even before such events occur (Pelkonen, 2010). However, traditional toxicology, either *in vivo* or partially *in vitro*, has multiple limitations: high cost, low productivity, ethically equivocal protocols *etc*. (Zgheib *et al.*, 2017). Furthermore, new understanding of biology shows more and more that the mechanisms that underlie toxicity are complex and involve multiple biological processes and pathways (Liu *et al.*, 2011; Park *et al.*, 2000). Considering traditional toxicology's limitations and the complex underlying biological reality, does toxicology today have real chances to become a predictive science? If yes, through which channels would it be possible?

‘Systems biology’ (SB) is a relatively new discipline that provides a framework for investigating the interactions between the separate parts of biological systems in order to understand their functioning and detect any new emergent properties (Geenen *et al.*, 2012). By integrating data concerning molecules and their interactions into an understanding of network behavior, SB provides insights into underlying mechanisms and basis of susceptibility to xenobiotics (Waters and Fostel, 2004) and creates a holistic view of biological systems (Chandra, 2009).

To handle and analyze complex biological systems and complex networks, Goelzer *et al.* (2008) showed how they can be broken down into sets of elementary functional modules. In the same spirit, signaling pathways and ‘adverse outcome pathways’ (AOP) are new emerging concepts that suggest broadening the toxicology framework to a system-wide level (Vinken, 2013) and help in the design of complex biology network models (Wittwehr *et al.*, 2017) by summarizing them into more tractable components (Edwards *et al.*, 2015). Practically, an AOP is a chemical-independent description of a linear path from a ‘molecular initiating event’ (MIE) to an eventual ‘adverse outcome’ (AO) at the organism or population level. In between, there can be any number of intermediate critical and measurable ‘key events’ (KEs) connected by ‘key events relationships’ (KERs). In typical AOP diagrams, KEs are represented by boxes and KERs by single one-directional arrows connecting them. (Allen *et al.*, 2014; Ankley *et al.*, 2010; Edwards *et al.*, 2015; LaLone *et al.*, 2017; Villeneuve *et al.*, 2014). **Figure 1** shows a schematic representation of two interacting AOPs: Boxes represent important events of an AOP (MIE, KEs or AO) with examples of each, and arrows are KERs.

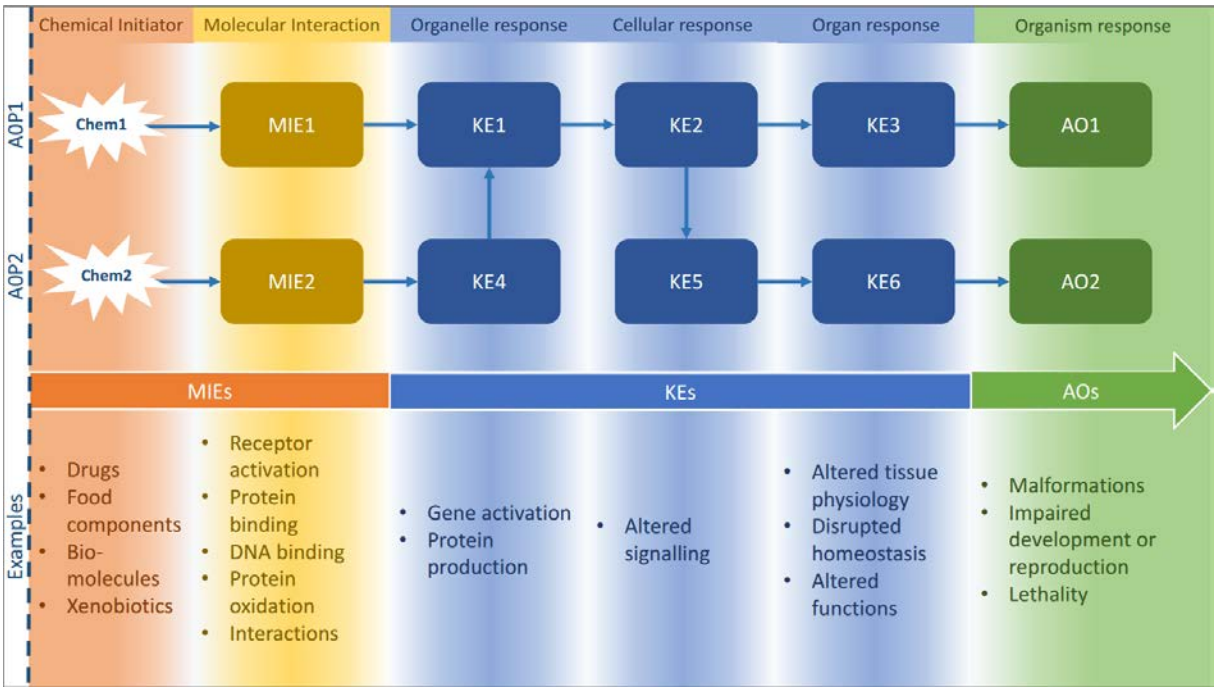


Figure 1. Schematic representation of two theoretical interacting AOPs. Through a timeline, different sections correspond to AOP levels (boxes represent the events, some examples are available in the lower part; arrows correspond to KERs).

AOPs and SB are some of the tools that can assist toxicology in moving from being a descriptive activity to becoming a more predictive mechanistic science (Materi and Wishart, 2007). For this purpose, AOPs and SB may either be used separately or combined. For example, a SB model can become a primary node, somewhere between a MIE and a KE in an AOP, setting the foundation for considering higher order questions of adaptive or compensatory responses and cross-talks among various pathways (Ankley *et al.*, 2010). The theme of this doctoral thesis is the combination of these two approaches for safety assessment of chemicals.

The StemBANCC[1] Project (2012-2018) was to develop an accessible and sustainable bio-bank of high quality well characterized patient-derived induced pluripotent stem cells lines that should speed up the drug development process and make therapies more adapted to specific human patients. Part of StemBANCC effort was devoted to demonstrating the use of such cells for drug safety research. StemBANCC was a five years European research project that started

[1] http://stembancc.org/ [Accessed October 24th, 2018]

in October 2012 and finished in March 2018. It was a collaboration between researchers from pharmaceutical companies, academic institutions and small and medium enterprises (**Table 1**). StemBANCC academic researchers received funding from the European Union's Innovative Medicines Initiative[2]. Pharmaceutical companies involved provided in-kind contributions. The characterization of induced pluripotent stem cells in terms of genetic, protein, and metabolic profiles, with the help of bioinformatics and SB models, was also an important part of StemBANCC.

Within StemBANCC, our group was in charge of modeling in drug safety aspects. Our contribution to the project consisted mainly in defining and implementing mathematical models of transport and cellular effects of tested molecules. This PhD work focuses on the development of a SB model for a major toxicity pathway: the 'nuclear factor (erythroid-derived 2)-like 2' (Nrf2) pathway. The Nrf2 pathway is a very important adaptive response to oxidative stress (Andrews *et al.*, 1993). Oxidative stress, linked to the over-production of 'reactive oxygen species' (ROS), is a major cause of chemical-induced injury and associated chronic diseases (*e.g.* cancer, Parkinson's disease *etc.*) (Kong *et al.*, 2014). Implicated in xenobiotics' metabolism and transport, Nrf2 contributes to and modulates ROS scavenging by 'glutathione' (GSH) (Leclerc *et al.*, 2014).

Our SB model was to be calibrated with induced pluripotent stem cells experimental data from StemBANCC partners. Having not received in time induced pluripotent stem cells data from the StemBANCC consortium, we have finally been constrained to calibrate our model with data produced with "ordinary" kidney *in vitro* human cells (RPTEC/TERT1) from a StemBANCC partner, the Medical University of Innsbruck (Prof. Paul Jennings, now based in Vrije Universiteit Amsterdam).

[2] https://www.imi.europa.eu/ [Accessed October 24th, 2018]

Apart the introduction, the present document is presented in four sections followed by a conclusion. First, Bibliography, is a literature review of each of the three aspects of the project: (*i*) toxicology (definition, history and transition to modern toxicology), (*ii*) biological context (oxidative stress, Nrf2 pathway, system-level approaches (SB and AOPs) to study biology) and (*iii*) computational tools used. The next section describes the building of a SB model (of the Nrf2 control of oxidative stress) for the development of a quantitative AOP. Then, in the following section, the SB model we conceived is calibrated and compared to two other mathematical approaches to quantitative AOPs. Finally, the last section, published as Zgheib *et al.* (2018), is a transcriptomic-based analysis of the cross-talks between Nrf2 and two other toxicity pathways: the 'activating transcription factor 4' (ATF4) branch of the unfolded protein response and the dioxin response *i.e.* 'aryl hydrocarbon receptor' (AhR) pathway.

The works of this doctoral thesis resulted in two published articles, a third paper that is currently in press and three posters. The first article, a literature review of 'high-throughput methods for toxicology and health risk assessment', was published in the 'Environnement Risque Santé' journal (Zgheib *et al.*, 2017). The SB model constructed in 'chapter 3' was presented in two posters (StemBANCC general assembly and steer committee meetings). The analysis performed in 'chapter 4' is currently in submission as a journal article. Finally, 'chapter 5', the product of the work accomplished during the scientific visit to the laboratory of Prof. Paul Jennings (Medical University of Innsbruck, StemBANCC partner), was published in the 'Frontiers in Genetics' journal (impact factor 4.151) (Zgheib *et al.*, 2018).

NB: In this document, to be distinguished from protein names, gene names are italicized.

Table 1. The 36 partners of the StemBANCC project listed in alphabetical order after the names of the two leaders: F. Hoffmann-La Roche Ltd and University of Oxford.

	Institute Name	*City*	*Country*	*Logo*
Leader	*F. Hoffmann-La Roche Ltd*	Basel	Switzerland	Roche
Leader	*University of Oxford*	Oxford	United Kingdom	UNIVERSITY OF OXFORD
	AbbVie Deutschland GmbH	Wiesbaden - Delkenheim	Germany	abbvie
	AstraZeneca	Södertälje	Sweden	AstraZeneca
	Boehringer Ingelheim International GmbH	Ingelheim	Germany	Boehringer Ingelheim
	Charité Universitätsmedizin	Berlin	Germany	CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN
	Concentris Research Management	Fürstenfeldbruck	Germany	concentris research management
	Eli Lilly	Basingstoke	United Kingdom	Lilly Answers That Matter.
	Gurdon Institute, University of Cambridge	Cambridge	United Kingdom	Gurdon INSTITUTE
	Helmholtz Zentrum München	Neuherberg	Germany	HelmholtzZentrum münchen German Research Center for Environmental Health
	Hannover Medical School	Hannover	Germany	lebao

Innsbruck Medical University	Innsbruck	Austria	
Institut National de la Santé et de la Recherche Médicale	Paris	France	
Institut National de l'Environnement Industriel et des Risques	Verneuil-en-Halatte	France	
Janssen Research & Development	Beerse	Belgium	
King's College London	London	United Kingdom	
Linköping University	Linköping	Sweden	
Medical Research Council - Functional Genomics Unit	Swindon	United Kingdom	
Merck Serono	Darmstadt	Germany	
Natural and Medical Sciences Institute	Reutlingen	Germany	
Novo Nordisk AS	Bagsvaerd	Denmark	
Orion Corporation	Espoo	Finland	
Pfizer Limited	Kent	United Kingdom	
Region Hovedstaden Glostrup Hospital	Hillerod	Denmark	

Sanofi-Aventis Recherche & Développement	Chilly-Mazarin	France	
Tel Aviv University	Tel Aviv	Israel	
The Hebrew University of Jerusalem	Jerusalem	Israel	
Univercell-Biosolutions	Toulouse	France	
University College London	London	United Kingdom	
Université de Genève	Genève	Switzerland	
Université de Lausanne	Lausanne	Switzerland	
Université de Technologie de Compiègne	Compiègne	France	
University of Birmingham	Birmingham	United Kingdom	
University of Edinburgh	Edinburgh	United Kingdom	
University of Luebeck	Luebeck	Germany	
University of New Castle	New Castle upon Tyne	United Kingdom	

2 BIBLIOGRAPHY

2.1 TOXICOLOGY

2.1.1 Definition of Toxicity

In certain conditions, a xenobiotic may induce perturbation in local biology and impair critical physiological functions of the organism (Hooper *et al.*, 2013). The organism's homeostatic defense against such chemical effects includes many biological processes from metabolic biotransformation, to cellular trans-membrane transport and activation of immune responses (Geenen *et al.*, 2012). Toxicity occurs when physiological homeostatic regulatory processes are lost or deactivated, and/or when defense mechanisms are overwhelmed and are no longer efficient and sufficient for protection (Aschauer *et al.*, 2015).

2.1.2 Predictive Toxicology: Prevention

The importance of toxicology in our days is relative to the amplitude of uncertainty and lack of information about toxicity of new and existing xenobiotics. Gathering appropriate knowledge, specific tools and various techniques, toxicology aims to spot harmful exposures, to assess their risk and to understand the mechanism of their toxicity in order to better prevent them. Prevention is possible when the toxic potential of an exposure is evaluated and accurately predicted even before the exposure occurs (Pelkonen, 2010).

2.1.3 Birth of Toxicology

Historically, experimental observations of toxicity, first described by Paracelsus *ca.* 1534, were re-framed into proper test methods during the 20th century (Trevan, 1927). Those methods mainly consisted in measuring adverse health outcomes in homogeneous animal groups at lethal or near-lethal doses and extrapolating them empirically to potentially estimate safe doses in humans (Bhattacharya *et al.*, 2011). Since the 1940s, the basic, mainly animal-based, experimental protocols for assessing the effects of chemicals on health have changed little (Shukla *et al.*, 2010).

2.1.4 Limitations of Traditional Toxicology

Whereas that traditional approach to toxicology has provided very important results through a century so far, it is still costly and resource-intensive (Zhu *et al.*, 2014). In numbers, the global yearly expenses on animal experimentation reach about €10 billion, 20% of which for toxicology alone, sacrificing 100 million animals worldwide every year (Hartung, 2009). Moreover, animal studies are low-throughput, too slow to screen the more than 80,000 chemicals already commercialized, for which little toxicity information exists (Taboureau and Audouze, 2017), and the new chemical entities reaching the market every year (National Toxicology Program, 2004). In addition, animal to human transposition is not always reliable and is affected by many uncertainties. We are not 70 kg rats: basal metabolic rates and metabolic pathways are among the major species-specific differences making inter-species transposition difficult and imprecise (Kongsbak *et al.*, 2014; Rangarajan and Weinberg, 2003). Besides, the extrapolation from the high-dose effects to low-dose responses is very difficult to validate. Finally, standardized animal tests make it difficult to take into account metabolic differences between different age groups and inter-subject variability in human populations (Szymański *et al.*, 2011), even though progress has been recently made in that area (Zgheib *et al.*, 2017).

2.1.5 A Paradigm Shift in Toxicology

The aforementioned hurdles created pressure to develop human-cell-based models. A need for a paradigm shift in toxicology started to emerge around 1980 (Rowan, 1983). The 3R's principle of replacement, reduction and refinement (Russell and Burch, 1959) had not gotten much echo in toxicology until that moment, at which scientific and technological advances, financial, ethical and legislative imperatives converged. Advances in molecular biology, cell biology (with stem cells technologies (Kitambi and Chandrasekar, 2011)), bioinformatics, SB and computational toxicology, introduced innovative methods less animal-based and with a higher-throughput productivity (Cotgreave, 2011). This new capacity to perform rapid examination of thousands of single agents or complex mixtures per day at relevant exposure levels, and the tools that make it possible, are named 'high-throughput screening' (HTS) (National Research Council, 2007). HTS *in vitro* assays using human cells allow the investigation of toxic effects in humans from different life stages and ethnicities (Inglese *et al.*, 2006). With the support of computational mathematical methods, HTS has the potential to largely improve the human health risk assessment of xenobiotics (Bois, 2009; Krewski *et al.*, 2009).

However, toxicological research did not evolve by virtue of innovation alone. Several initiatives from the European Union and the United States of America ran in the same direction, pushing for change since the beginning of the 21st century (Zhu *et al.*, 2014) (**Figure 2**). We focus next on those efforts, noting that Japan has also followed the trend a bit later (Omoe, 2006).

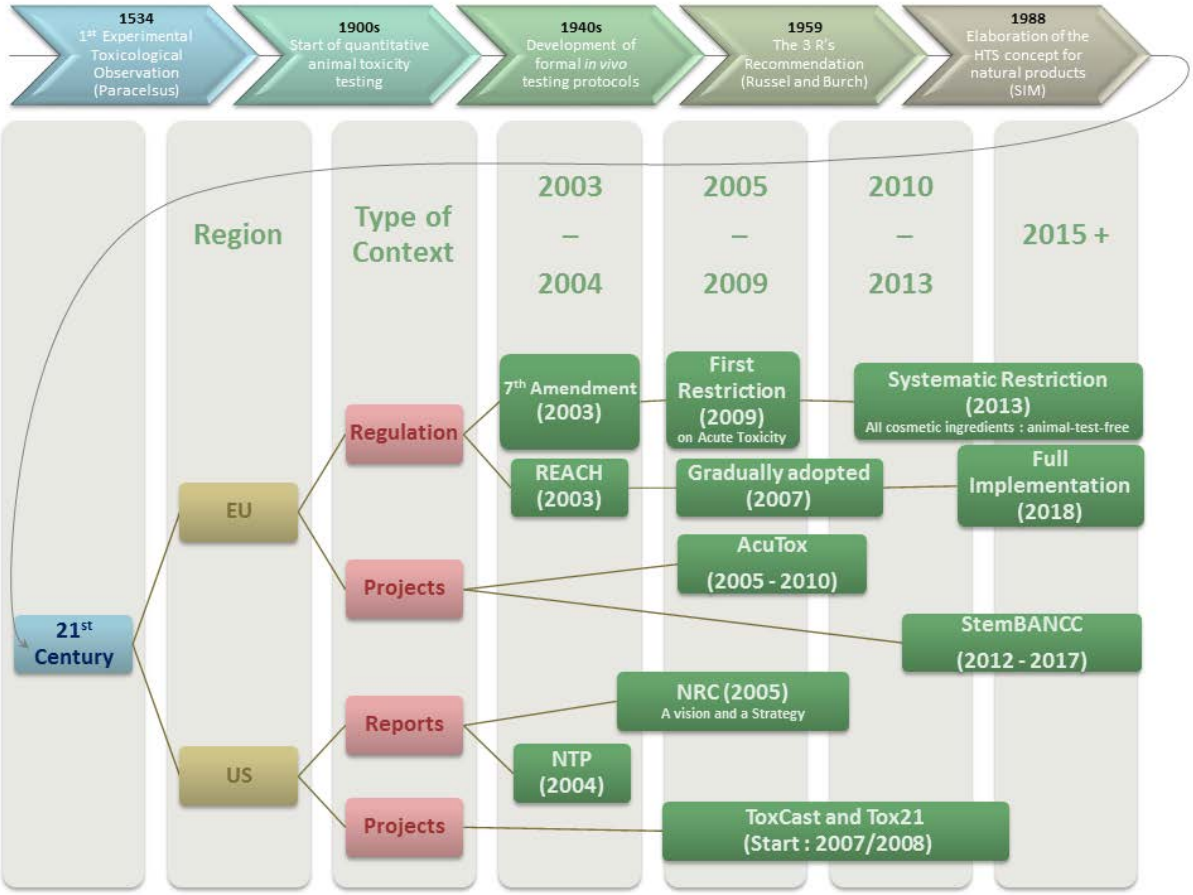


Figure 2. Timeline illustrating the birth and development of toxicology from first *in vivo* experiments by Paracelsus up to HTS initiatives in the European Union and the United States of America (Zgheib *et al.*, 2017).

❖ *Regulatory and Scientific Initiatives in the European Union*

o The 7th Amendment to the Cosmetics Directive

On January 15th 2003, the European 7th Amendment (2003/15/EC) to the Cosmetics Directive (76/768/EEC) restricted the use of animals in all cosmetic testing (Seidle and Stephens, 2009). It also set a time frame for the development of eventually validated alternative methods for toxicity testing (Pauwels and Rogiers, 2004). In 2009, a first restriction on acute toxicity animal-based testing took effect (Bhattacharya *et al.*, 2011). By 2013, by European law, all new cosmetic ingredients intended for the European market had to be animal-test-free. That legislation has become a motor of change, and pushed for the development of eventually validated alternative testing strategies (Hartung, 2011).

o REACH Regulation: The Registration, Evaluation, Authorization and Restriction of Chemicals

Adopted by the European Commission in 2003, and implemented in 2007, the REACH regulation established a local regulatory framework for the safety assessment of chemicals produced or imported in quantities greater than one ton per year (Foth and Hayes, 2008). It calls for the development of computational and experimental *in vitro* testing methods, integrated toxicity testing strategies, keeping *in vivo* experiments as a last resort. That comprehensive program aimed at evaluating the risks of more than 30,000 synthetic chemicals already in use in Europe by June 2018 (van Vliet, 2011). By this deadline only 20,000 chemicals were evaluated.

o European Union Scientific Research Projects

European actions have not only been legislative or regulatory. The FP7[3] and Horizon 2020[4] research programs have accompanied legislation consistently by pushing for the development of corresponding knowledge and technologies. The European Union has funded and launched many large-scale projects with different themes: ACuteTox Project[5] in acute toxicity alternative testing, Scrtox[6] Project and StemBANCC[7] Project in stem cell technology, COSMOS[8] in computational modeling, NOTOX[9] in SB, the SEURAT-1[10] cluster and EU-ToxRisk[11] in predictive toxicology *etc*.

❖ *Reports, Programs and Other Initiatives in the US*

o The National Toxicology Program Road-Map

Aware of the above-mentioned development, the National Toxicology Program proposed in 2004 a road map for the future of toxicology testing entitled 'A national toxicology program for the 21st century' (National Toxicology Program, 2004), which called for a shift from observational methods towards more predictive, target-specific and mechanism-based alternative assays. It also placed the emphasis on tools like physiologically based pharmacokinetic modeling and quantitative structure-activity relationships to better support quantitative risk assessment. In 2005, the National Toxicology Program initiated a collaboration with the National Chemical Genomics Center to develop chemical libraries and HTS assays (Inglese *et al.*, 2006; Shukla *et al.*, 2010).

---

[3] https://ec.europa.eu/research/fp7/index_en.cfm [Accessed October 24th, 2018]
[4] http://www.horizon2020.gouv.fr/ [Accessed October 24th, 2018]
[5] http://www.acutetox.eu/ [Accessed October 24th, 2018]
[6] http://www.scrtox.eu/ [Accessed October 24th, 2018]
[7] http://stembancc.org/ [Accessed October 24th, 2018]
[8] http://www.cosmostox.eu/ [Accessed October 24th, 2018]
[9] http://www.notox-sb.eu/ [Accessed October 24th, 2018]
[10] http://www.seurat-1.eu/ [Accessed October 24th, 2018]
[11] http://www.eu-toxrisk.eu/ [Accessed October 24th, 2018]

- ### ToxCast Program: The Toxicity Forecaster of the Environmental Protection Agency

ToxCast is a multi-year research program launched in 2007 by the Environmental Protection Agency to run automated HTS *in vitro* assays and computational analyses for prioritizing further toxicity assessments of chemicals (Dix *et al.*, 2007). It is based on bioactivity profiling of chemicals and screening changes in cells or proteins' activity after exposure, with the ambition of picking out "remarkable" toxicity off the mass of data accumulated. Another goal is to establish causal links between eventual exposures and effects on biological pathways and targets (Environmental Protection Agency, 2007). Obviously, the latter calls for the development of high throughput exposure, toxicodynamic and toxicokinetic models (Judson *et al.*, 2014).

- ### Toxicity Testing in the 21st Century: A vision and a Strategy

In 2005, the National Research Council report entitled 'Toxicity testing in the 21st century: A vision and a strategy' proposed to government, academia, and industry, a paradigm shift in toxicology through the application of emerging disciplines and technologies (omics, SB, computational modeling, *etc.*) (Kavlock *et al.*, 2007; National Research Council, 2007). The proposed approach advocates heavier use of mechanistically informative *in vitro* assays to study how chemicals interact with cellular response networks and turn them into toxicity pathways (Raunio, 2011). The report considers four options for toxicity testing summarized in

Tox21: Toxicology testing in the 21st Century

Toxicology testing in the 21st Century (Tox 21) is another collaborative testing and evaluation program that was established in 2008 via a Memorandum of Understanding between the National Toxicology Program, the National Chemical Genomics Center, and the Environmental Protection Agency, later joined by the US Food and Drug Administration. Tox21's chemical library contains over 8,000 chemicals of different kinds (*e.g.*, pesticides,

marketed pharmaceuticals, food additives, industrial chemicals, cosmetic ingredients, chemicals found in household products and clothes *etc.*) (Schmidt, 2009).

Table 2: *option 1* represents the *statu quo* and primarily relies on animal-based *in vivo* tests and *option 2* takes into consideration the available information on the substance studied and its mechanisms of action, and is already operational. The remaining two options respond to the 'National Research Council' calls at two different degrees: the extreme *option 4* calls for an *in vivo*-free strategy (as envisioned in the legislation of the European Union for cosmetics' ingredients), while the intermediate *option 3* leaves open the possibility of using animal-based tests in complementarity to innovative mechanistic approaches (Carmichael *et al.*, 2006).

❖ *Tox21: Toxicology testing in the 21st Century*

Toxicology testing in the 21st Century (Tox 21)[12] is another collaborative testing and evaluation program that was established in 2008 via a Memorandum of Understanding between the National Toxicology Program, the National Chemical Genomics Center, and the Environmental Protection Agency, later joined by the US Food and Drug Administration. Tox21's chemical library contains over 8,000 chemicals of different kinds (*e.g.*, pesticides, marketed pharmaceuticals, food additives, industrial chemicals, cosmetic ingredients, chemicals found in household products and clothes *etc.*) (Schmidt, 2009).

Table 2. Toxicity-testing options defined by the 'Toxicity testing in the 21st century: A vision and a strategy' report (National Research Council, 2007) in order to enhance the paradigm shift in toxicity research (Zgheib *et al.*, 2017).

Criteria	**Option 1** ***in vivo***	**Option 2** ***Tiered in vivo***	**Option 3** ***in vivo / in vitro***	**Option 4** ***in vitro***
Biology	Animal	Animal	Mostly Human	Mostly Human
Concentrations used	High	High	Multiple	Multiple
Throughput	Low	Low	Medium and High	High

[12] https://www.epa.gov/chemical-research/toxicology-testing-21st-century-tox21 [Accessed October 24th, 2018]

Quantity of animals used	High	Low	Low	None
Possibility of *in silico* screens	None	Limited	Possible	Yes

2.1.6 Modern Toxicology

The field of toxicology has significantly evolved as we have seen above, with the progressive introduction of *in vitro* (expansion of toxicological databases substantially) and *in silico* methods (fine-tuning of computational methods), so that the latter now appear feasible and highly suitable (Kongsbak *et al.*, 2014; Taboureau and Audouze, 2017). Considering the importance of this progress, we can start talking about HTS in toxicology. HTS tissue models have been developed at the interface between biotechnology, biomaterial engineering, bioinformatics and medical sciences. HTS has both qualitative and quantitative advantages. Quantitatively, HTS can be defined as the set of screening techniques that can be scaled up to test libraries of molecules at a rate exceeding thousands of structures daily in a concentration-response format using standardized protocols (Judson *et al.*, 2013; Kavlock *et al.*, 2007; Zhu *et al.*, 2014). Qualitatively, a distinct advantage of HTS is its ability to test complex mixtures, combine experimental conditions and end-points to develop extensive dose-response relationships for different pathways across large concentration ranges for different exposure schedules (Astashkina *et al.*, 2012; Boekelheide and Andersen, 2010).

Many elements contribute to the establishment of this modern approach to study toxicology. In this section, we will evoke four of the pillars of this emergent field: robotics, induced pluripotent stem cells, omics and bioinformatics.

❖ *Robotics*

If HTS is possible, that is due to the rigorous robotic spotting technologies, the miniaturization of the assay vial (*i.e.*, micro-plates) and automation (Rangarajan and Weinberg, 2003). The capacity of the micro-plates has significantly increased with time. From 96-well plates, originally used in virology (Feng *et al.*, 2007), to 384- and 1,536-micro-well plates currently in use (Inglese *et al.*, 2007), the equipment has been gradually improved to test more molecules and concentrations (van Vliet, 2011). The volume of the wells in a micro-plate has also decreased, down to volumes as low as 2µL (Mayr and Fuerst, 2008) (**Figure 3**).

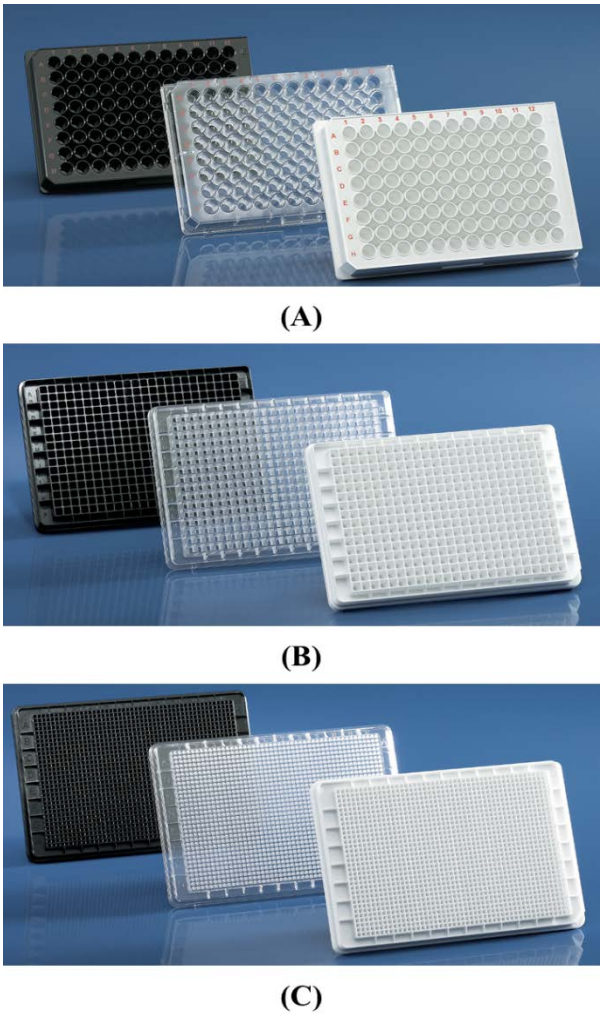


Figure 3. Standard Microplates of **(A)** 96-, **(B)** 384-, and **(C)** 1536-well formats respectively[13].

[13] https://www.wellplate.com/standard-microplates/ [Accessed October 24th, 2018]

❖ *Induced Pluripotent Stem Cells*

Many features make induced pluripotent stem cells attractive for toxicity screening. Other than their uniform physiology and donor-specific genetic profile, they have unlimited self-renewal potential and are pluripotent (and therefore differentiable into various cell types such as hepatocytes, cardiomyocytes, neurons *etc*.) Human stem cells can be derived from embryonic cultures (isolated in the inner cell mass of the blastocyst (Bongso *et al.*, 1994)), adult tissues (*e.g.*, bone marrow (Pittenger, 2008), skin (Fernandes *et al.*, 2009), liver (Gaudio *et al.*, 2009), umbilical cord blood (Moon *et al.*, 2008), and brain (Clarke *et al.*, 2000) *etc*.), or through genetic reprogramming of easily accessible cells (*e.g.*, skin fibroblasts, renal epithelial cells shed in urine *etc*.) into induced pluripotent stem cells (O'Malley *et al.*, 2009). Although embryonic stem cells have a higher degree of pluripotency than induced pluripotent stem cells, they continue to be subject of ethical debates. Furthermore, the difficulty of inducing a reliable and efficient differentiation of all cells in one culture remains a major limitation of these techniques (Menasché, 2011), but progress is being made to alleviate that problem.

❖ *Omics (Transcriptomics, Proteomics, Metabolomics) and Biomarkers*

In traditional toxicology, cell count and lactate dehydrogenase activity in the culture medium were at some point the only cytotoxicity endpoints measurable *in vitro* (Blaauboer, 2008). Nowadays, different cell death pathways are known and their activation can be followed using many cellular biomarkers (van Vliet, 2011). Omic technologies are aimed primarily at the universal detection of biomarkers, either they are genes (genomics), mRNA (transcriptomics), proteins (proteomics) or metabolites (metabolomics), in a specific biological sample in a non-targeted and non-biased manner (Horgan and Kenny, 2011). The use of omics profiling contributes to a better understanding of toxicology due to the considerable size of datasets it provides and its capacity of discovery of new more specific biomarkers. The amount of data generated by various omics technologies contributes to a better understanding of a drug's (and other chemical's) safety profile (Gautier *et al*., 2013).

❖ *Bioinformatics*

These days, it is not more difficult to measure the activity of a whole genome than it is that of a single gene, or even to sequence the genomes of thousands of micro-organisms or hundreds of human beings. Microscopy now offers extremely high resolution so thousands of single cells and multitude of parameters can be analyzed in parallel for each patient. Confronted with this flood of data, biologists are often at a loss because experimental planning and analysis methods need to be adapted accordingly (Systems Biology at University of Lyon — BioSyL)[14]. Using the increasingly large amount of biological and chemical data available and combining it with bioinformatics has become a promising approach permitting a chemical safety assessment across multiple scales of complexity from molecular to cellular and system levels in human health (Gautier *et al.*, 2013).

[14] www.biosyl.org [Accessed October 24th, 2018]

The National Research Council works have clearly shown that extrapolated results obtained by even the most advanced *in vitro* methods require the development of specific mathematical models. In analogy with the commonly used '*in vitro*' and '*in vivo*', the term '*in silico*' describes an analysis performed on a computer (Raunio, 2011). In toxicology, *in silico* techniques, also called 'computational toxicology', form a sub-discipline that uses computer and mathematical models to understand and predict the physio-pathological mechanisms of toxicity and their ultimate outcome as adverse effects (Cohen Hubal *et al.*, 2010; Collins *et al.*, 2008). *In silico* experiments can be carried out to test the effects of perturbations on the system and to identify the processes that control the system. Some of these experiments may only be feasible using a computer and most of them are faster and cheaper *in silico* than *in vitro*. Such "dry experiments" (computational modelling) may generate new hypotheses about the system, which can then be tested experimentally in "wet experiments" (laboratory experimentation). Computational toxicology offers remarkable possibilities by allowing the analysis of a large number of chemicals and biological interactions, yet more proof-of-concept studies are needed to demonstrate its added value and make it fully adopted by risk assessors and regulators.

2.2 BIOLOGY UNDERLYING TOXICOLOGY

In their review, Gautier *et al.* (2013) considered that studying a drug action and protein's function in a global physiological environment may better inform us on the chemical's toxicity. Network-level approach studies phenomena in their small details whilst linking them to a wider setting of interactions with their surroundings. In our work for example, the genesis of oxidative stress by xenobiotics and the dynamics of its control by the Nrf2 pathway, are described by a SB model that can be a node in a wider framework: for example in a 'chronic kidney disease' (CKD) AOP.

In this section we first cover the biological context of our mathematical models: the oxidative stress, the Nrf2-GSH response to oxidative stress, and other associated pathways (*i.e.*, AhR and ATF4). Then, we will present in details system-level approaches used (*i.e.*, SB and AOP).

2.2.1 Oxidative Stress, Nrf2 and some Associated Pathways

❖ *Cellular Metabolism of Xenobiotics*

Inside the cell, an important homeostatic intracellular system of xenobiotic metabolizing enzymes families controls the intracellular levels of xenobiotics and their metabolites (Zhang *et al.*, 2009). Some of these enzymes metabolize the xenobiotic via various reactions and some others conjugate it to other entities to detoxify it (Xu *et al.*, 2005). Practically, upon the entry of a xenobiotic to the cell, this homeostatic system is triggered when xenosensor receptor molecules (*e.g.*, AhR, constitutive androstane receptor, or pregnane X receptor *etc.*) are activated. This induces metabolizing enzymes (*e.g.*, 'cytochrome P450' (CYP)) that may transform the xenobiotic into an intermediate metabolites (Nebert and Dalton, 2006). Then, the parent compound or its metabolites may activate 'transcription factors' (TFs) (*e.g.*, Nrf2, ATF4 *etc.*) to induce so-called detoxifying enzymes (*e.g.*, 'glutathione-S-transferase' (GST), 'glutamate cysteine ligase' (GCL) *etc.*) that catalyze a set of conjugation reactions that add hydrophilic conjugates to it. Finally, metabolites may be exported to the extracellular compartment by membrane-residing transporters (*e.g.*, 'multidrug-resistance protein' (MRP) *etc.*). The role of this system is to control the amounts of xenobiotics and their metabolites that can accumulate in the cell, hopefully restricting their downstream toxicity (Zhang *et al.*, 2009). The Nrf2 signaling pathway, like many other intracellular toxicity pathways, follows the aforementioned scheme.

❖ *Oxidative Stress*

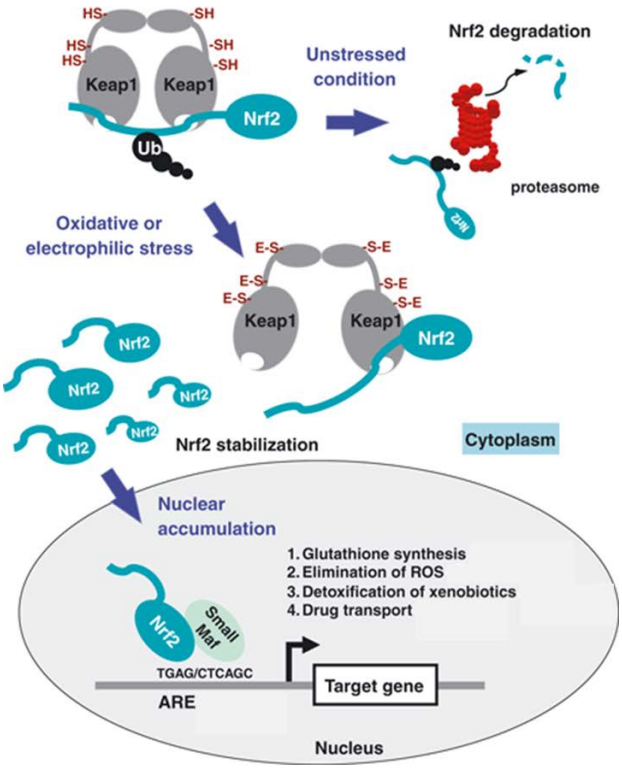
Under physiological conditions, the metabolism of oxygen and nitric oxide generates reactive byproducts (*e.g.*, hydrogen peroxide, superoxide anion, *etc.*), *aka* free radicals, that are gathered under the terms of ROS and reactive nitrogen species (*e.g.*, nitrite, nitrate, peroxynitrite *etc.*) respectively (Halliwell *et al.*, 1992). The bioactivation of xenobiotics into reactive electrophilic metabolites undergoing redox cycling, is another source of free radicals (Zhang *et al.*, 2009). These free radicals can react with DNA, protein, and lipids. Oxidative stress occurs when accumulation of intracellular ROS and reactive nitrogen species in a cell becomes uncontrolled due to the imbalance between their intracellular formation and removal from the cell (Himmelstein *et al.*, 2004). Since the exposures tested here are nitrogen-free chemicals, only ROS will be used in this thesis to refer to oxidative stress.

Oxidative stress is a major cause of chemical-induced injury and associated chronic and degenerative diseases (*e.g.*, cancer, Parkinson's disease, arthritis, aging, autoimmune disorders, and cardiovascular diseases *etc.*) (Kong *et al.*, 2014; Pham-Huy *et al.*, 2008; Taguchi *et al.*, 2011). However, several mechanisms can be put in place to counteract oxidative stress. First, the endogenous cellular enzymatic defense system (*e.g.*, superoxide dismutase, catalase, 'glutathione peroxidase' (GPX), peroxiredoxins, 'glutathione reductase' (GR), sulfiredoxin, GST *etc.*) (Reddy, 2008). Second, other non-enzymatic antioxidants mainly acquired by food and supplementation (*e.g.*, vitamin C (L-ascorbate), vitamin A, vitamin E *etc.*) quench ROS levels and thereby play key roles in modulating oxidative stress (Kohen and Nyska, 2002).

❖ *Nrf2 Control of GSH Synthesis and Oxidative Stress*

Nrf2, a basic leucine zipper TF, is a cytoplasm based protein of 605 amino-acids in six domains (noted Neh1 through Neh6) (Itoh *et al.*, 1999). Neh2 is a regulatory domain that, under basal conditions, interacts with the cytoskeleton-associated 'kelch-like-ECH-associated protein 1' (Keap1) of 624 amino acids of three domains. In the absence of oxidative stress, Nrf2 being constantly trapped by Keap1, is targeted by the ubiquitin ligase complex for ubiquitination. Ubiquitination, allowing Nrf2 degradation in the proteasome (Kobayashi *et al.*, 2004), keeps Nrf2's half-life very short (~10 minutes). Keap1, being rich of cysteine, reacts with ROS due to high electrophilicity (Deshmukh *et al.*, 2017). Upon oxidative stress, reactivity of ROS with Keap1 cysteine improves which increases the oxidation of Keap1 (Kaspar *et al.*, 2009). This conformational change in Keap1 lowers the ubiquitination of Nrf2, and thus makes Keap1's binding to Nrf2 less favorable (Villeneuve *et al.*, 2010). Once its cytoplasmic retention mechanism by Keap1 is inactivated, Nrf2 translocates to the nucleus (Huang *et al.*, 2000) where it binds to small proteins called 'Maf' to form 'Nrf2-Maf' heterodimers (Nguyen *et al.*, 2000). Nrf2-Maf binds to the 'antioxidant response element' (ARE) in the promoter region of several genes (*e.g.*, 'glutamate cysteine ligase catalytic subunit' (*GCLC*), 'glutamate cysteine ligase modifier subunit' (*GCLM*), 'glutathione synthetase' (*GS*), *GPX*, and *MRP etc.* (Kaspar *et al.*, 2009)) to up-regulate their expression in response to a variety of stimuli. GS, GCLC and GCLM enzymes are involved in GSH synthesis and recycling, GPX contributes to its metabolism and ROS scavenging by GSH, and finally MRP helps eliminate its metabolites (Andrews *et al.*, 1993; Jennings *et al.*, 2013).

By serving as a substrate for antioxidant enzymes in redox cycles, GSH protects cells against electrophilic compounds and reactive metabolites by undergoing rapid oxidation and regeneration to maintain the intracellular redox status. However, under strong oxidative stress, such Nrf2-mediated detoxification processes consume GSH in a faster rate than its regeneration. GSH depletion makes cells more susceptible to oxidative stress which may damage DNA or impair cell viability. For a better visualization of the Nrf2 signaling pathway, we propose a schematic representation (**Figure 4**) of its behavior under both conditions: presence and absence of oxidative stress (Taguchi *et al.*, 2011),.



Nrf2: Nuclear Factor (erythroid – derived 2) – like 2

Keap1: Kelch-like-ECH-associated protein 1

ARE: Antioxidant Response Element

Ub: Ubiquitine

Figure 4. Schematic representation of the Nrf2 signaling pathway in basal unstressed condition as well as under its activation by oxidative (or electrophilic) stress (Taguchi *et al.*, 2011).

❖ *Other Associated Pathways*

Nrf2 is one of the important pathways that can be activated upon exposure to xenobiotics like oxidants. Nrf2 control of GSH synthesis, metabolism and transport, is an adaptive defense response of the cell to oxidative stress. This makes Nrf2 a central signaling pathway to be studied. However, in the modern understanding of biology, a pathway is never isolated. Thus to better locate Nrf2 in the toxicological panorama, we have studied, in 'chapter 5', its interactions and cross-talks with two other toxicity pathways here presented: AhR and ATF4.

o Aryl hydrocarbon Receptor Pathway - AhR

AhR is a ligand-activated TF that controls the transcription of a wide range of genes involved in the synthesis of certain key xenobiotic- and drug-metabolizing enzymes mainly belonging to the *CYP* family genes, (*e.g.*, *CYP1A1*, *CYP1B1* and *CYP1A2 etc.*) implicated in the metabolism of endogenous and exogenous substrates. Like Nrf2, AhR is a cytoplasm-based molecule trapped in a complex (Petrulis and Perdew, 2002). Upon ligand (xenobiotic) binding, the AhR TF shuttles into the nucleus where it dimerizes with the 'AhR nuclear translocator' (ARNT) and binds to so-called xenobiotic-responsive elements (*i.e.*, 'dioxin response element' (DRE)) in the promoter region of some oxidative stress related genes to stimulate their expression (Haarmann-Stemmann *et al.*, 2012).

o Activating Transcription Factor 4 Pathway – ATF4

ATF4 is another protein and TF involved in the regulation of an Nrf2 target, the 'heme oxygenase' gene, linked to the adaptive response to oxidative stress (He *et al.*, 2001). ATF4 is a major branch of the unfolded protein response and is activated in response to endoplasmic reticulum (ER) disturbances or proteotoxicity where unfolded proteins accumulate in the ER and compete with an important sensing protein named 'RNA (PKR)-like ER kinase' (PERK) for the inhibitory protein BiP (Hetz, 2012). Activated PERK phosphorylates the eIF2α (eukaryotic translation initiation factor 2 α) which inhibits general protein translation while inducing AT4 translation. ATF4 in turn binds to the CARE consensus sequence and drives transcription of genes involved in amino acid synthesis, amino acid transport and aminoacyl-tRNA synthesis (Jennings *et al.*, 2012).

2.2.2 Systems Biology – SB

SB is a discipline that encompasses the relationship between the "science of the whole system" (physiology) and the "science of the individual components" (molecular biology). SB has provided a framework for investigating the interactions between the separate parts of a biological system in order to understand its functioning (Geenen *et al.*, 2012). A typical SB approach combines holism and reductionism. While the reductionist approach would provide detailed information about properties of the small entities of a system under artificial conditions where they are more or less uncoupled, the holistic approach tests these entities as they are embedded in the living system in a more natural and realistic setting. Nevertheless, in the holistic approach, detailed and high quality data is much harder to obtain and analyze (Klipp *et al.,* 2010).

The strength of the SB approach tackles the complexity of biological systems and their dynamic behavior at every relevant organizational level (from molecules, cells and organs through organisms and ecosystems). The interconnection between different cellular processes, such as metabolism and genetic regulation, reflects the importance of the holistic approach introduced by the SB paradigm. Although most cellular components have been studied individually, the behavior of the cell emerges at the network-level and requires an integrative analysis (Machado *et al.*, 2011). Considering all (or most) of the components of a system simultaneously and not separately makes possible the identification and study of new emergent properties of the system. Emergent properties are functional properties not present within the individual components of the system and only arise when system components interact among each other. A common example to illustrate this is the interaction between hydrogen and oxygen to make water: the resulting change in properties is unpredictable if only the individual properties of hydrogen and oxygen are known (Aderem, 2005).

To study emergent properties, SB uses many computational and experimental tools and skills of various disciplines (Geenen *et al.*, 2012). Intrinsic to SB is its interdisciplinary nature consisting in coupling different levels of information (*e.g.*, experimental results, mathematical models, statistical tools *etc.*) in order to develop predictive models of the biological behavior (Systems Biology at University of Lyon — BioSyL)[15]. In this logic, incorporation of omics data streams for building improved SB models (Cramer *et al.*, 2011; Zhang *et al.*, 2010) contributes to a better understanding of the data and an improved prediction ability of the models (Hamon *et al.*, 2014; Quignot and Bois, 2013; Tan *et al.*, 2009). However, it is not only data that is involved; the study of a living system relies on a multitude of parameters (*e.g.*, half-life, diffusion speed, affinity *etc.*) that cannot all be measured experimentally.

In order to make computational model predictions precise and develop a reliable scientific understanding, it is necessary to integrate experiments in a spiral of iterative cycles of validation/falsification with computational modeling, simulation and theory (Westerhoff and Kell, 2007). The modeling methodology is bottom up, inserting kinetic equations for all molecular processes and then integrating these to predict network behavior around the physiological state (Geenen *et al.*, 2013). The emergent properties produced by this process become the hypotheses to be confirmed in "wet experiments" as explained previously. Thus, SB experiments are hypothesis-generating, using holistic approaches where no hypothesis is known or prescribed but all data are acquired and analyzed to define a hypothesis that can be further tested (Horgan and Kenny, 2011). In summary, in SB, modeling is not the final goal, but it is a tool to increase understanding of the system, to develop more directed experiments and, finally, allow predictions.

[15] www.biosyl.org [Accessed October 24th, 2018]

2.2.3 Adverse Outcome Pathways – AOP

Xenobiotics, beyond their target sites, can perturb a whole balanced equilibrium of complex intracellular system of pathways, to achieve their toxicity. The key for a more general view of toxicity schemes, is in understanding the different networks and pathways involved, their respective contribution to random outcomes as well as their potential interactions and cross-talks (Liu *et al.*, 2011; Park *et al.*, 2000). This kind of approach permits a better understanding of the system, elucidates emergent properties and opens the door for a genuine investigation of what happens behind the scenes, and therefore makes of toxicology a predictive science (Materi and Wishart, 2007; Zgheib *et al.*, 2017).

In the same line, the AOP, an ecotoxicology emerging concept, has rapidly drawn the attention of industries and regulatory agencies alike (Groh and Tollefsen, 2015). 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). AOPs are intended "to outline and capture existing knowledge concerning the biologically plausible and empirically supported foundations for predicting apical toxicity from mechanistic data"(OECD, 2016).

As mentioned above, an AOP portrays a linear pathway from one MIE to one eventual AO (Allen *et al.*, 2014; Ankley *et al.*, 2010; Villeneuve *et al.*, 2014). AOP diagram networks are modular structures having KEs and KERs as fundamental units. In the traditional AOP diagram, KEs are represented by boxes whereas KERs are represented as the arrows connecting a pair of upstream and downstream boxes. In a graph theory context, KEs represent nodes and KERs represent edges (Pavlopoulos *et al.*, 2011). Organization of knowledge into AOP frameworks can help in the design of complex biology network models (Wittwehr *et al.*, 2017). A powerful tool to progress in this direction and reduce uncertainty would be assembling

different AOPs and sharing data to identify and fill in the data gaps like in the AOP knowledge base[16] (OECD, 2016), the AOP-Wiki[17] (a user-friendly, open-source interface), and other knowledge repositories (Groh and Tollefsen, 2015). The modularity of KEs and KERs we mentioned earlier, offers efficiency in updating and constructing AOP: any modifications made to those KEs or KERs descriptions in an AOP knowledge repository are automatically updated for all AOPs that included that KE or KER (Villeneuve *et al.*, 2014).

The contribution of AOPs to predictive toxicology starts with chemical grouping or classification (OECD, 2013), priority setting for further testing, and hazard identification. On the longer-term, AOPs can be part of 'integrated approaches to testing and assessment' (Conolly *et al.*, 2017) or 'integrated testing strategies' for regulatory decision making (Edwards *et al.*, 2015; Vinken, 2013). For this purpose, quantitative AOPs (qAOPs) need to be developed to provide dose-response, and time-course predictions (Conolly *et al.*, 2017). Parameter values of qAOPs can often be obtained directly from literature or targeted experimental work, or indirectly by optimizing the fit of model predictions to data (Villeneuve *et al.*, 2014). In addition, statistical analyses can be used to evaluate alternative model formulations and simplifications through identification of correlated variables or parameters (Friend and Schadt, 2014; Li and Vu, 2013; Rodriguez-Fernandez *et al.*, 2013).

[16] www.aopkb.org [Accessed October 24th , 2018]
[17] www.aopwiki.org [Accessed October 24th, 2018]

2.3 MATHEMATICAL CONSIDERATIONS

Due to the diversity of phenomena that occur in living organisms, many formalisms can be used to model biological systems (*e.g.*, 'Boolean networks', 'Bayesian networks', 'Petri nets', 'process algebras', 'constraint-based models', 'differential equations', 'rule-based models', 'interacting state machines', 'cellular automata', 'agent-based models' *etc.*) Some of these mathematical tools and methods (*i.e.* 'ordinary differential equations', 'Bayesian networks') were used through this thesis. These techniques and other common equations (*i.e.*, Michaelis-Menten kinetics, Hill's equation) that were used for the development of the computational mathematical models conceptualized in this thesis, are described in this section.

2.3.1 Ordinary Differential Equations – ODE – Systems

Differential equations describe the rate of change of continuous variables. They are typically used for modeling dynamical systems in several areas like SB (Machado *et al.*, 2011). In SB, the network of chemical reactions happening among the different biomolecules (*e.g.*, genes, proteins, lipids, metabolites *etc.*) and xenobiotics present in a predefined compartment, can be described by systems of nonlinear 'ordinary differential equations' (ODEs). Practically, ODEs are used to describe the variation of the amount of species in the modeled system as a function of time (Machado *et al.*, 2011). In this kind of systems, each equation corresponds to the chemical reaction producing or consuming the concerned molecule through time (Geenen *et al.*, 2012). The goal is that the ODEs based model captures most of the available kinetic information regarding the system. However, building ODE models requires insight into the reaction mechanisms to select the appropriate rate laws to define the model structure and the associated kinetic equations. Then, the unknown model parameters are estimated using fitting of experimental data (Hasdemir *et al.*, 2015).

ODEs is a well-understood formalism, fast, mathematically robust, rigorous and adaptable (De Jong, 2002; Kitano, 2002; Orton *et al.*, 2005). For these advantages and others, the ODEs methodology of translating biochemical reactions into mathematics and then integrating them over time using numerical methods, became a privileged tool of many SB research projects (*e.g.*, metabolic pathways (Ideker *et al.*, 2001), mitosis in yeast (Tyson, 1991), genetic regulatory circuits (Elowitz and Leibler, 2000), *etc.*) including in toxicology.

Other types of differential equations exist but will not be detailed in this manuscript (*e.g.*, 'stochastic differential equations', 'partial differential equations', 'Piecewise-linear differential equations' *etc.*).

2.3.2 Michaelis-Menten – MM – Kinetics

Enzymatic reactions are generally composed of two basic steps. First, the reversible binding of a substrate molecule *S* to an enzyme *E* in order to form the complex *ES* (binding and unbinding are defined by the k_b (measuring unit: L.mol^{-1}.s^{-1}) and k_u (measuring unit: s^{-1}) rates respectively). In the second step, by the k_{cat} (measuring unit: s^{-1}) rate parameter, the catalysis of *ES* releases the enzyme *E* and generates the product *P*. In both cases, either *ES* is catalyzed into *E* and *P* or unbound to restore reactants *E* and *S*, *E* is free again to associate with other substrate molecules (Reuveni *et al.*, 2014).

$$E + S \rightleftarrows ES \xrightarrow{k_{cat}} E + P$$

While the first step is a quasi-instantaneous equilibrium, we consider that the second part of the equation, that is irreversible, is the limiting step. Thus the reaction's velocity *v* (measuring unit: mol.L^{-1}.s^{-1}) depends on k_{cat} and the concentration of [*ES*] only, (in the equations of this section, square brackets refer to a molecule's molar concentration (measuring unit: mol.L^{-1})) as shown in equation 2.1

$$v = k_{cat} \cdot [ES] \qquad (2.1)$$

Equations 2.2 and 2.3 give the kinetics of *ES* formation v_1 and elimination v_2 respectively (for both, measuring unit: mol.L^{-1}.s^{-1}):

$$v_1 = k_b \cdot [E] \cdot [S] \qquad (2.2)$$

$$v_2 = (k_u + k_{cat}) \cdot [ES] \qquad (2.3)$$

During the stationary equilibrium phase, [*ES*] is stable, which means that its formation and elimination are equal (see equation 2.4):

$$v_1 = v_2 \Leftrightarrow k_b \cdot [E] \cdot [S] = (k_u + k_{cat}) \cdot [ES] \Leftrightarrow \frac{[E]}{[ES]} = \frac{k_u + k_{cat}}{k_b \cdot [S]} \qquad (2.4)$$

According to mass conservation law, the total quantity of the enzyme E_T is constant and equal to the sum of its bound ES and unbound E fractions. Based on this assumption, in equation 2.5 we express ES in function of E and E_T.

$$[E_T] = [ES] + [E] = [ES] \cdot \left(1 + \frac{[E]}{[ES]}\right) \Leftrightarrow [ES] = \frac{[E_T]}{1+\frac{[E]}{[ES]}} \qquad (2.5)$$

Replacing $\frac{[E]}{[ES]}$ in equation 2.5 by its expression from equation 2.4 gives equation 2.6:

$$[ES] = \frac{[E_T]}{1+\frac{k_u+k_{cat}}{k_b \cdot [S]}} \qquad (2.6)$$

Finally, to obtain the 'Michaelis-Menten' (MM) equation (2.7), $[ES]$ in equation 2.1 should be written under its expression obtained in equation 2.6:

$$v = k_{cat} \cdot \frac{[E_T]}{1+\frac{k_u+k_{cat}}{k_b \cdot [S]}} = \frac{k_{cat} \cdot [E_T] \cdot [S]}{[S]+\frac{k_u+k_{cat}}{k_b}} \Leftrightarrow v = \frac{V_{max} \cdot [S]}{K_m+[S]} \qquad (2.7)$$

In the MM equation, $V_{max} = k_{cat} \cdot [E_T]$ (measuring unit: mol.L^{-1}.s^{-1}) is the maximal enzymatic velocity attained when the binding sites of the enzymes are saturated at high $[S]$, and $K_m = \frac{k_u+k_{cat}}{k_b}$ (measuring unit: mol.L^{-1}) is the so-called 'Michaelis constant' that is interpreted as the substrate concentration at which enzymatic velocity attains half its maximal value. This MM reaction scheme, linking enzymatic velocity v to the substrate concentration, has been applied to the analysis of enzymatic kinetics, for over a century and continues nowadays to be an important reference in different scientific fields like biochemistry, pharmacology and physiology.

2.3.3 The Hill Equation

MM kinetics applies well to a single molecule *S* binding (enzymatic) reaction, but things get more complicated when additional molecules try to associate with the enzyme *E*. In fact, binding of one molecule of *S* at one site may alter the affinity of the enzyme *E* (or any macromolecule: receptor, transporter *etc.*) for other new substrates and hence regulates their binding rate. The property behind this phenomenon is called the cooperative binding or 'cooperativity'. 'Cooperativity' is positive when the binding of one molecule of *S* increases *E*'s affinity for other substrates, and negative when this affinity is decreased. However, if this is not the case and *E*'s affinity is not changed, binding of different substrates *S* is completely independent and thus is considered non-cooperative (Weiss, 1997). While non-cooperative binding can be modeled by the MM equation (Alon, 2007), the other cases require different kinetics. Graphically, by plotting *v* against [*S*], we obtain a sigmoidal S-shaped curve when biding is cooperative and hyperbolic when it is not (**Figure 5**).

Considering the multiple binding patterns reaction where *n* molecules of *S* bind to the same macromolecule *E* forming an *ES* complex, the equilibrium that takes place can be represented as follows:

$$E + nS \rightleftharpoons ES_n$$

On equilibrium, applying the law of mass action permits to write a K_d-dependent expression of $[ES_n]$ in equation 2.8; K_d (measuring unit: $(\text{mol.L}^{-1})^n$) being the ratio of k_u (measuring unit: s^{-1}) to k_b (measuring unit: $(\text{L.mol}^{-1})^n.\text{s}^{-1}$) (Atkins, 1973):

$$k_u \cdot [ES_n] = k_b \cdot [E][S]^n \Leftrightarrow \frac{[E][S]^n}{[ES_n]} = \frac{k_u}{k_b} = K_d \Leftrightarrow [ES_n] = \frac{[E][S]^n}{K_d} \qquad (2.8)$$

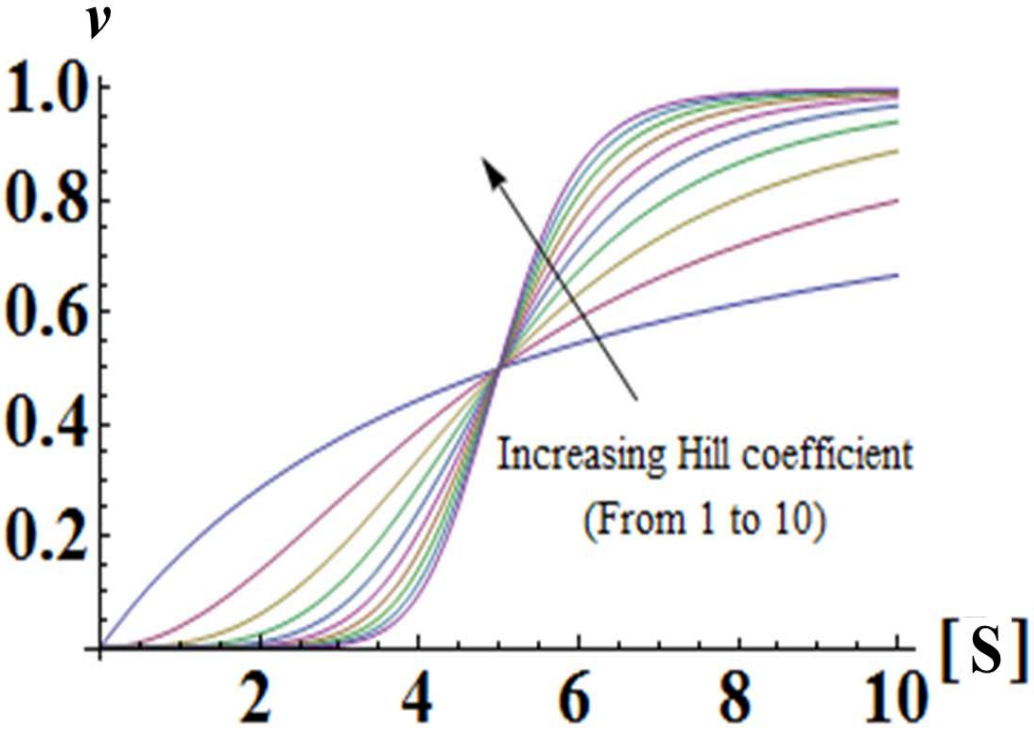


Figure 5. Plot of enzymatic reaction's velocity *v* against substrates concentration [*S*] in 10 different cases for Hill's coefficient *α* gradually increasing from 1 (hyperbolic: Michaelis-Menten case) to 10 (all other curves (2 to 10) are S-shaped) (Duke, Modeling Cooperativity)[18].

[18] http://2013.igem.org/wiki/index.php?title=Team:Duke/Modeling/Cooperativity&oldid=215310 [Accessed October 24th, 2018]

Equation 2.9 gives for E, the ratio r of the bound portion the enzyme $[ES_n]$ over its total amount available. Equation 2.10 is what we find by substituting $[ES_n]$ in equation 2.9 by its expression from equation 2.8.

$$r = \frac{[ES_n]}{[ES_n]+[E]} \qquad (2.9)$$

$$r = \frac{[S]^n}{K_d+[S]^n} \qquad (2.10)$$

Back in 1910, Archibald Vivian Hill (1886–1977) proposed is a sigmoid shaped quantitative model of oxygen binding to hemoglobin that took his name: the 'Hill equation' (equation 2.11) (Gesztelyi *et al.*, 2012; Stefan and Le Novère, 2013). Later on, and not mentioned in the works of Hill, (Clark, 1926) and (McLean, 1938) pointed out the strong connection between the aforementioned laws of equilibrium (equations 2.8, 2.9 and 2.10) and the 'Hill equation' (equation 2.11). Then, on the rational basis of the receptor occupancy theory (Ariens, 1954), the 'Hill equation' was proposed by Wagner (1968) to be used for the analysis of the nonlinear drug concentration-effect mechanisms (Csajka and Verotta, 2006; Mager *et al.*, 2003) (*e.g.*, the renal uptake of aminoglycosides, the tubule glomerular feedback in kidney (Rougier *et al.*, 2003), ligand binding in voltage-dependent ion channels (Haynes *et al.*, 1986) *etc.*).

$$P_{sat} = 100\frac{{P_{O_2}}^{\alpha}}{K_d+{P_{O_2}}^{\alpha}} \qquad (2.11)$$

The original 'Hill equation' (equation 2.11) was developed to quantify the percentage of hemoglobin saturation with oxygen P_{sat} based on t he partial pressure of oxygen P_{O_2} (measuring unit: mmHg). This equation was then rewritten as a rational function applied to concentrations (equation 2.12). In this currently used version of the 'Hill equation', *f* represents the fraction of *E*'s saturated binding sites (in analogy with the P_{sat} percentage of equation 2.11 and α is the 'Hill coefficient' that represents the degree of 'cooperativity'.

$$f = \frac{v}{V_{max}} = \frac{[S]^{\alpha}}{K_d + [S]^{\alpha}} = \frac{[S]^{\alpha}}{{K_m}^{\alpha} + [S]^{\alpha}} \qquad (2.12)$$

'Cooperativity' is positive when $\alpha > 1$, negative when $\alpha < 1$ and binding is independent and non-cooperative if $\alpha = 1$. In this last case Hill equation is equivalent to the MM kinetics. Mathematically, the MM equation is a special case of the Hill equation.

Despite the perfect analogy between the two models, α and *n* should not be mixed up. The first, the degree of 'cooperativity' α, can be a decimal and tell us about post-binding *E*'s affinity to associate with other substrate molecules. The second, the number *n* of *S* molecules bound to the enzyme *E*, is a stoichiometry indicator and can only take integer values. However, it has been shown that the 'Hill coefficient' α is a correct estimate for the number of binding sites *n* in some cases like the positive cooperative binding case ($\alpha > 1$) (Weiss, 1997). This two-fold interpretation of the same model represents the theoretical shift in the approach towards the 'Hill equation', from a mechanistic explanatory model (with *n*) to a descriptive curve-fitting model (with α).

2.3.4 Bayesian Statistical Tools

❖ *The Bayes Theorem*

Bayesian analysis proceeds by inferring about (hidden) "causes" on t he basis of (observed) "effects" *i.e.*, on the basis of data. Being probabilistic in essence, Bayesian analysis derives its inferences in the form of probability distributions for the variables it seeks to identify. Such distributions, called the 'posterior distribution', are obtained by combining what is already known (*i.e.*, the 'prior distribution') to what is experimentally observed about those variables (*i.e.*, data, 'observed evidence'); and is therefore a compromise of the two (Gill *et al.*, 2005; Holleman and Simel, 1997) (**Figure 6**). The values of model parameters, if not precisely known, are also described by probability distributions and are treated as "random variables" (Bois, 2012).

The Bayes theorem, first described by Thomas Bayes in 1774 (Bayes and Mr. Price, 1763; Stigler, 1986), consists in updating knowledge of the phenomenon on t he basis of observations, in three steps. The first step consists in defining the background knowledge/belief about variables, gathering information and setting up the 'prior distribution' (Bois, 2012).This initial probability estimate is often based on scientific literature and/or previous experiences (Phelps and Levitt, 2004) and can be precise (for common physiological values for example) or vaguely approximated (*e.g.*, in case of diffusion velocity of a given chemical between two compartments).

In the second step, appropriate experiments are performed, needed data are collected and 'observed evidence' is set. Finally, in the third step, using the Bayes formula, the 'prior distribution' is updated on the basis of the 'observed evidence' (*i.e.*, experimental data) (Bois, 2012).

Thomas Bayes' idea was to simply apply the definition of conditional probabilities to these inferences. By definition the conditional *probability* of an event *A*, given event *B*, is as follows:

$$P(A|B) = \frac{P(A,B)}{P(B)} \qquad (2.13)$$

Where $P(A,B)$ denotes the joint probability that both A and B occur, and $P(B)$ the probability that B occurs, regardless of A. That definition applies to probabilities, but also, more generally, to probability distributions, be they discrete or continuous density functions.

If experimental data obtained (*i.e.*, 'observed evidence') is different from what we expect (*i.e.*, 'prior distribution'), we usually want to infer about the parameter values susceptible to have led to such observations. That requires computing $P(\theta|y)$, the 'posterior distribution' of all model's parameters, θ, given the data y (*i.e.*, posterior to collecting y). Applying equation 2.13, we simply obtain:

$$P(y|\theta).P(\theta) = P(y,\theta) = P(\theta|y).P(y) \Rightarrow P(\theta|y) = \frac{P(y|\theta).P(\theta)}{P(y)} \qquad (2.14)$$

In conclusion, the Bayes theorem states that the probability distribution of the unknowns given the data at hand are proportional to the 'prior distribution' $P(\theta)$ of those unknowns times the 'data likelihood', $P(y|\theta)$, which depends on the model. The term P(y) is called the prior predictive probability of the data. Since the data are considered fixed numerical values, P(y) can be considered as a normalization constant. The posterior parameters' distribution summarizes what is known about θ after collecting the data y and the remaining uncertainty about it. It is obtained by "updating" the prior $P(\theta)$ using the data likelihood (equation 2.14), and this updating is a simple multiplication (Bois, 2012).

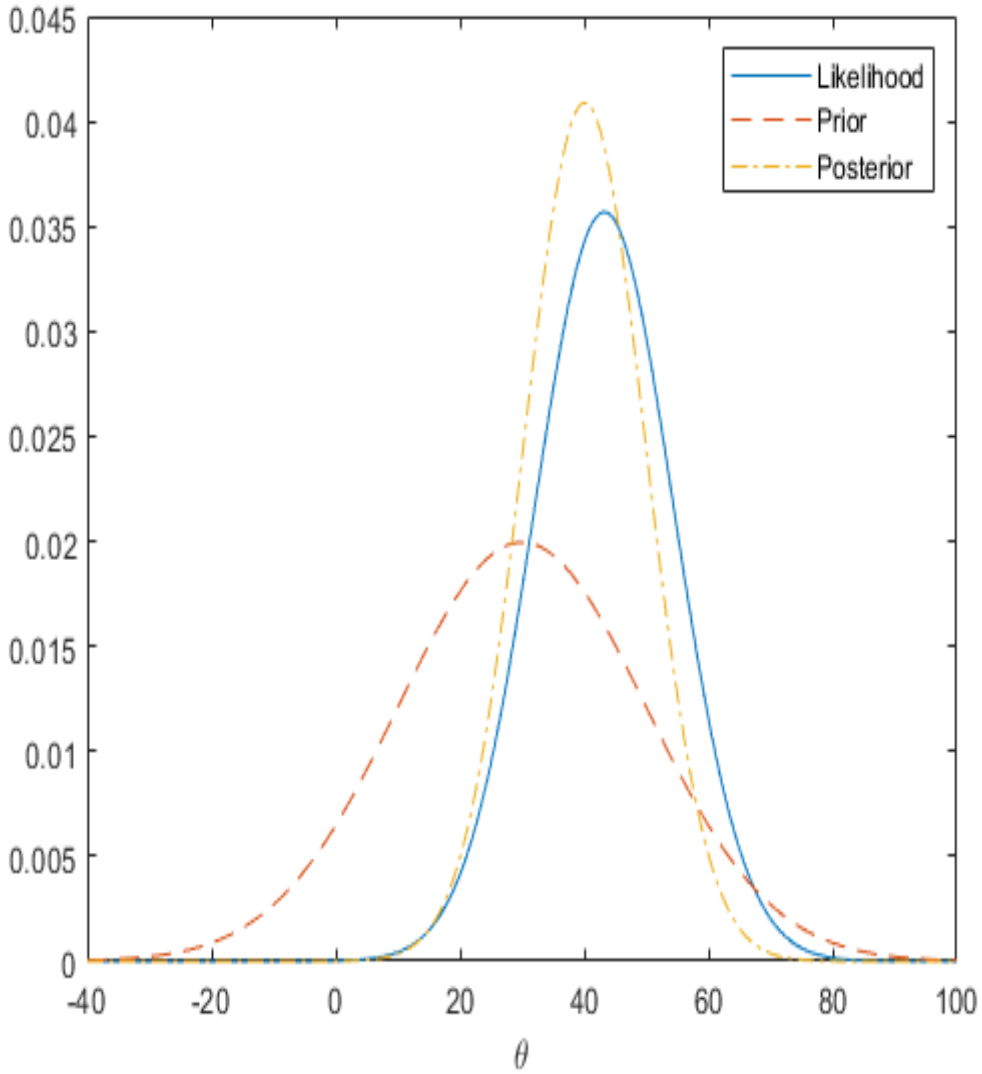


Figure 6. Prior, likelihood and posterior distributions for θ. The 'posterior inference' is a formal compromise between the 'observed evidence' (likelihood), summarizing the 'prior distribution' of the data alone (Bayesian Analysis for a Logistic Regression Model - MATLAB & Simulink Example)[19].

[19] https://www.mathworks.com/help/stats/examples/bayesian-analysis-for-a-logistic-regression-model.html [Accessed October 24th, 2018]

❖ *Bayesian Network – BN*

A Bayesian network (BN) is a probabilistic model whose underlying structure is a graph (equivalently, a network) where each node represents a variable of the problem (*i.e.*, for an AOP: chemical substance, MIE, KEs and AO), and each arc between two nodes represents a direct dependency (ideally, a causal relationship) (Pearl, 1988). Within such a BN, a probabilistic relationship (specifically, a component of a conditional distribution function) is defined by each arc linking two variables. For example, if an arc joins variables A and B, a relationship such as "A is distributed normally around $k \times B$, with a variance equal to s^2" has to be defined. As a result, every node of the network has a probability distribution conditioned by other network variables. This implies that a variable cannot depend upon itself, even indirectly, and therefore cycles are not allowed in BNs. Evidence on a set of nodes (for example, measurement of some KEs) updates the probability distributions of all their dependent nodes (Jaworska *et al.*, 2013). Learning a BN from data means that one searches for those dependencies (and associated distributions) between variables that best explain the data. On the other hand, calibrating a BN implies estimating the parameters of the distribution functions that link variables.

However, standard BNs do not provide a direct mechanism for representing temporal dependencies. In cases where the data time evolution is progressive rather than instantaneous, it is natural to use a dynamic BN (DBN) to integrate those data (Kjærulff and Madsen, 2008). DBNs, typically, replicate an underlying structure at several (discrete) times corresponding to measurement time points. Each node of a given time slice may depend on nodes in the previous time slice and on nodes in the same time slice (Pavlovic, 1999). In this way, the value of a node at time t_i may depend on its own value at time t_{i-1}, without introducing a loop in the graph.

2.3.5 Model's Calibration

Bayesian model's calibration is the estimation of the (joint) posterior distribution of the values of a model's parameters. If the model is checked, then we can perform model validation. Validation goes beyond checking and allows to verify if the model will correctly predict, even outside of the data range. It consists in verifying the adequacy of predictions of new data and then to check the plausibility of the model for the purpose for which it will be used. As Bayesian calibration allows to fit the data, it can also adjust all the parameters and therefore plot the estimation of metabolism rate.

For many years, Bayesian statistics was essentially restricted to very simple models like conjugate models where the mathematical form of the prior and likelihood are jointly chosen to ensure that the posterior may be evaluated with ease. Numerical integration methods based on analytic approximations were developed in 70s and 80s of the last century with some success, but a revolutionary change occurred in the early 1990s with the adoption of "indirect methods" that draw random samples from the 'posterior distribution' without needing a closed-form of the distribution to sample from. A large number of such algorithms exists (*e.g.*, Gibbs sampling Markov chain Monte Carlo *etc.*) (Gilks *et al.*, 1996). In these methods, widely used nowadays, the *a posteriori* distribution integrates *a priori* information and experimental data in order to represent the "updated" knowledge about parameters. Model's calibration is the Bayesian estimation of this *a posteriori* and of the value a model's parameters. Bayesian calibration of a model starts by defining, for each parameter, the *a priori* distribution reflecting the knowledge we have about concerned parameters, even before the beginning of data collection and observation (van de Schoot *et al.*, 2014). In the following paragraphs, the Monte Carlo method and the Markov chain Monte Carlo (MCMC) algorithm will be presented.

❖ *The Monte Carlo Method*

Simple Monte Carlo simulations are based on s uccessive random and independent samples from a given distribution. Any 'posterior distribution' (and its properties: mean, variance, quantiles *etc.*) may be approximated by taking a very large random sample of realizations of θ from $p(\theta|y)$. Samples from the posterior can be generated in several ways, without exact knowledge of the analytical form of $p(\theta|y)$. Direct methods include rejection sampling, which generates independent proposals for θ, and accepts them at a probability proportional to the desired posterior. Importance sampling can also be used by appropriately weighting independent samples from a user-chosen distribution on θ, properties of the posterior $p(\theta|y)$ can be estimated (Spiegelhalter and Rice, 2009). Realizations from the posterior used in Monte Carlo methods need not be independent, or generated directly. When more powerful MCMC methods are used.

❖ *Markov chain Monte Carlo (MCMC) method*

The MCMC simulation algorithm is a widely used indirect method for models' parameters calibration. MCMC is an iterative procedure (Kruschke, 2011, 2013).

The MCMC sample of each draw is and conditioned by the precedent iteration, hence the appellation "Markov chain" because the new value depends partly on the previous. Then, a ratio of probabilities between the two last draws is calculated, to determine if the new candidate θ' is selected or not. The algorithm continues sample proposed values and accepts or rejects them, according to the value of the calculated ratio, as long as the user wish (Bois, 2012). After a sufficient number of draws, the simulated chain converges in probability towards a prescribed joint density of model parameters, for example towards their 'posterior distribution' (Bois, 2012). Practically, it is common to simulate two, three (or more) chains for the calibration of the same parameter(s) with the same likelihood, each time beginning from a different starting point. All simulated chains are run for a certain (typically large, >1000) number of draws until the convergence of all chains approximately obtained (Gelman and Rubin, 1992). It is then possible to estimate empirically the *a posteriori* distributions of model parameters, for example by computing its quantiles and moments.

The posterior density forms the basis for evaluating the quality of model fit, comparing different hypotheses about parameter values, and choosing the parameter values for which the model best fits the data.

3 CONSTRUCTION OF SYSTEMS BIOLOGY MODEL OF NRF2 CONTROL OF OXIDATIVE STRESS

3.1 STARTING MODELS

GSH being a key element in the physiological defense mechanism of the organism against oxidative stress. Understanding the implication of GSH in ROS scavenging is primordial to study toxicity of oxidants. Controlling the transcription of genes coding for the synthesis of enzymes involved in the GSH cycle, Nrf2 orchestrates an important part of the GSH defense response. To model the Nrf2 signaling pathway, we have merged two SB models. The first, conceived by Hamon *et al.* (2014), highlights the contribution of Nrf2 to the GSH response to oxidative stress. The second is a simplification of the model of Reed *et al.* (2008), was developed by Geenen *et al.* (2012) and describes the synthesis, the metabolism and the transport of GSH under oxidative stress.

3.1.1 The model of 'Hamon *et al.* (2014)'

In 2014, Hamon *et al.* published a SB model offering an interesting description of the Nrf2 signaling pathway and its interactions with the AhR pathway, its auto-induction as well as of how it controls GSH synthesis and the transport of its metabolites. This model parametrized to simulate the exposure of human kidney RPTEC/TERT1 cells to cyclosporine A. The validation of this model was completed by a quantitative *in vivo-in vitro* extrapolation (QIVIVE) (Hamon *et al.*, 2015). In 'Supplementary Material 7.1', **Figure S1** shows a schematic representation of this model.

3.1.2 The model of 'Geenen *et al.* (2012) and Reed *et al.* (2008)'

In 2012, Geenen *et al.* proposed a SB model of GSH synthesis inspired by the work of Reed *et al. (*2008).

Reed *et al.* tried to explore GSH's metabolism using a mathematical model including the one-carbone-metabolism, the trans-sulfuration cycle, the folate cycle, the synthesis and the metabolism of GSH. That model contained four compartments (*i.e.*, mitochondria, cytosol and nucleus within cells and the extracellular environment) and was based on pr operties and regulation of key enzymes of oxidative stress. The works of Reed *et al.* can be used to simulate observed metabolic profiles of some diseases and compare them to clinical data. A schematic representation of this model is presented in **Figure S2** of 'Supplementary Material 7.1'.

Geenen *et al.* (2012a) have significantly modified the model of Reed by simplifying the folate cycle and limiting it to three equations, by adding two biomarkers (*i.e.*, 5-oxoproline and ophthalmic acid) and by adapting the model to the detoxification of specific xenobiotics (in particular, paracetamol). All modifications brought by Geenen shouldn't affect the initial steady state of the model. Please refer to **Figure S3** in 'Supplementary Material 7.1' to see the schematic representation of that model. Some of Geenen's model parameter values were found in literature and others were simply adjusted to metabolites concentrations at steady state within the physiological limits of liver metabolism. That model was used to study the oxidative stress with the SB approach following exposure to xenobiotics, using GSH and 5-oxoproline and ophthalmic acid as biomarkers.

3.2 METHODS

3.2.1 Remodelling Hamon's model

Despite the general outlook on the Nrf2 pathway that it offers, the model of Hamon *et al.* (2014) has two limitations: first, modelling of transcription and translation is too complicated, and second, GSH synthesis is over-simplified.

In Hamon's model, two gene activator schemes are available: either the xenobiotic X binds to AhR to form an activator complex that we named nucX-AhR, or, under increasing amounts of ROS, a part of the trapped cytoplasmic Nrf2 dissociates from Keap1 to travel to the nucleus (*i.e.*, nucNrf2) and activates its target genes. Hamon's model details the transcription and translation of eight genes: *CYP*, *GS*, *GCLC*, *GCLM*, *Nrf2*, *GST*, *GPX* and *MRP*. These genes split into two categories: those activated by only one TF (either X-AhR (*e.g.*, *CYP*) or Nrf2 (*e.g.*, *GS*, *GCLC* and *GCLM*)), and those activated by both TFs (*e.g.*, *Nrf2*, *GST*, *GPX*, and *MRP*).

To describe the transcription and translation of each of these TFs' targets, the model incorporates the following steps: a binding-unbinding equilibrium between each of the gene's activators and their specific genetic receptor, transcription induction by the activator-receptor complex, followed by translation and mRNA degradation. In the nuclear (gray) compartment of **Figure S1** in 'Supplementary Material 7.1', all the steps of this cascade of reactions are illustrated. Application of the same process to each of the eight genes results in a large number of state variables (51) and parameters (78), with a cascade of mostly linear differential equations resulting in a complex system of equations hard to integrate (some reactions are extremely fast).

❖ *Hill-based model for transcription and translation*

To simplify that part of the model, we modeled transcription control cascades according to the 'Hill equation' in order to have a single equation per gene. For genes controlled by one activator (*i.e.*, TF) x_a we obtain equation 3.1:

$$\frac{\partial(mRNA)}{\partial t} = k_0 + \frac{V_{max} \cdot {x_a}^n}{{k_m}^n + {x_a}^n} - k_{deg} \cdot mRNA \qquad (3.1)$$

and for genes that are controlled by two TFs x_a and x_b (*i.e.*, nucNrf2 and nucX-AhR) we obtain equation 3.2:

$$\frac{\partial(mRNA)}{\partial t} = k_0 + \frac{Vmax_{Nrf2} \cdot Nrf2^n}{km^n_{Nrf2} + Nrf2^n} + \frac{Vmax_{AhR} \cdot AhR^n}{km^n_{AhR} + AhR^n} - \frac{Vmax_{Nrf-AhR} \cdot Nrf2^n \cdot AhR^n}{\left(km^n_{Nrf2} + Nrf2^n\right)\left(km^n_{AhR} + AhR^n\right)} - k_{deg} \cdot mRNA$$

where *mRNA* represents the quantity of produced mRNA (in zeptomols) and δmRNA/δt is its derivative with respect to time, k_0 is the basal transcription rate under zero exposure, and k_{deg} is the mRNA degradation rate. In equation 3.2, *i.e.* in the case where two TFs can contribute to transcription of a single gene, we consider the additive contribution of each regulator separately (referred to by the subscripts *a* and *b* in equation 3.2) while subtracting the overlap, following the models proposed by Alon (2007) for multi-dimensional input functions that integrate more than one TF. The nuclear (red) compartment of **Figure 7** shows how transcription-translation cascade of reactions was simplified and reduced to one equation per gene, regardless if is activated by one or two TFs.

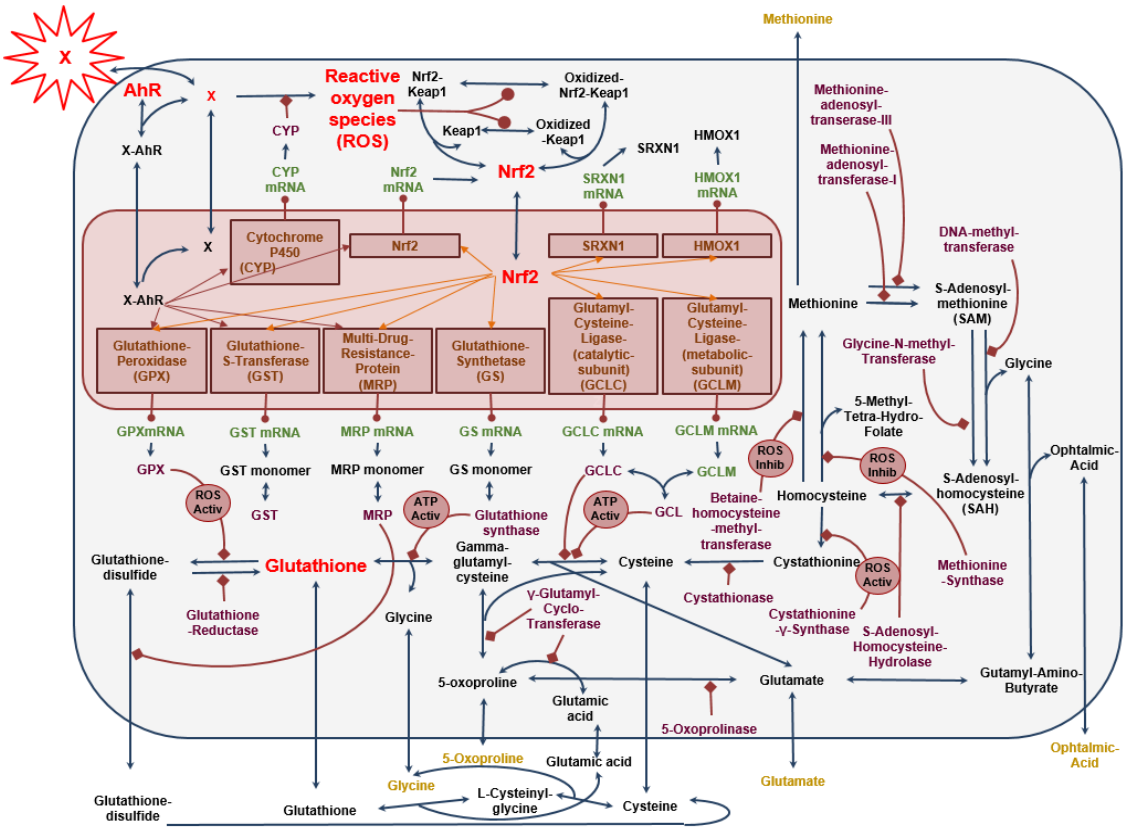


Figure 7. Schematic overview of the assembled SB model. This model covers both transcriptional and biochemical aspects of GSH synthesis and metabolism and its control by the Nrf2-Keap1 signaling pathway. The blue compartment is cytosol and the red one is nucleus. Blue arrows show reactant(s):product(s) exchange during biochemical or transport reactions, and red arrows indicate enzymatic catalysis (diamond heads) or gene transcription (round heads). In the nucleus, red boxes represent genes and arrows indicate gene activation. Names of genes are in orange, of mRNA are in green, of enzymes are in purple, of other proteins and metabolites in blue and of extracellular constants in yellow.

❖ *Calibration protocol*

We wanted our simplified model to behave as closely as possible as the original Hamon's model. In order to find appropriate values for the Hill parameters (for the transcription of each gene: k_0, V_{max} and k_m) we simulated virtual data that covers all dose range combinations with substantial transcription. Thus, both models (the new and the original) were run with a number of incremental doses of TFs (*i.e.*, nucNrf2 and/or nucX-AhR) between zero and saturation level during a time period that ensures reaching a stable equilibrium between exposures. **Table 3** summarizes the protocol used: starting from zero, every 400,000 seconds, an incremental dose of the TF(s) was added. For genes that are under control of both TFs, all possible combinations of concentrations of the two TFs were considered. For equations 3.1 and 3.2, MCMC simulations were applied to find parameter values for which the curve of the new Hill-based model fits best the curve of Hamon's model.

❖ *Software*

The Hill-based SB model was simulated and calibrated with the *GNU MCSim* software, version 5.6.6 (Bois, 2009a). For all genes and parameters, two MCMC chains were run in parallel for 10,000 iterations and convergence was checked on the last 9,000 iterations. All fitting plots were created with R, version 3.4.4 (R Development Core Team, 2013).

Table 3. Virtual exposure scheme applied on both Hamon's (old) and Hill-based (new) SB models to perform MCMC curve fitting and establish equivalency between them. Genes that are activated by a single TF (*i.e.*, *CYP*, *GS*, *GCLC* and *GCLM*) were exposed to five doses (one dose per time-point) ranging from 0 to 100 zeptomol doses of their TF (*i.e*, nucNrf2 or nucX-AhR). Genes that are activated by both TFs (*i.e.*, *Nrf2*, *GST*, *GPX* and *MRP*) were exposed to five different and separate combinations of doses per time-point (25 combinations are possible). All exposures are in zeptomol.

Genes	**Exposure starting times (seconds)**				
	0	400,000	800,000	1,200,000	1,600,000
CYP	**0** nucX-AhR	**0.5** nucX-AhR	**1** nucX-AhR	**10** nucX-AhR	**100** nucX-AhR
GS **GCLC** **GCLM**	**0** nucNrf2	**0.5** nucNrf2	**1** nucNrf2	**10** nucNrf2	**100** nucNrf2
Nrf2 **GST** **GPX** **MRP**	**0** nucX-AhR + **0** nucNrf2	**0.5** nucX-AhR + **0** nucNrf2	**1** nucX-AhR + **0** nucNrf2	**10** nucX-AhR + **0** nucNrf2	**100** nucX-AhR + **0** nucNrf2
	0 nucX-AhR + **0.5** nucNrf2	**0.5** nucX-AhR + **0.5** nucNrf2	**1** nucX-AhR + **0.5** nucNrf2	**10** nucX-AhR + **0.5** nucNrf2	**100** nucX-AhR + **0.5** nucNrf2
	0 nucX-AhR + **1** nucNrf2	**0.5** nucX-AhR + **1** nucNrf2	**1** nucX-AhR + **1** nucNrf2	**10** nucX-AhR + **1** nucNrf2	**100** nucX-AhR + **1** nucNrf2
	0 nucX-AhR + **10** nucNrf2	**0.5** nucX-AhR + **10** nucNrf2	**1** nucX-AhR + **10** nucNrf2	**10** nucX-AhR + **10** nucNrf2	**100** nucX-AhR + **10** nucNrf2
	0 nucX-AhR + **100** nucNrf2	**0.5** nucX-AhR + **100** nucNrf2	**1** nucX-AhR + **100** nucNrf2	**10** nucX-AhR + **100** nucNrf2	**100** nucX-AhR + **100** nucNrf2

3.2.2 Assembling two models

In order to better study the transcriptional regulations of the GSH pathway by the Nrf2-Keap1 signaling cascade, we have merged the Nrf2 pathway model developed by Hamon *et al.* (2014) with the GSH synthesis and metabolism model proposed by Geenen *et al.* (2012a). In fact, GSH synthesis and response to oxidative stress was much more developed and detailed in the model of Geenen. The link between the two models is that the transcription part of Hamon's Nrf2-Keap1 model codes for the synthesis of key enzymes of GSH synthesis in Geenen's model. Even though GSH synthesis was much more developed and detailed in Geenen's model, the added value of Hamon's version was the elaboration of the role of 'adenosine triphosphate' (ATP) and energy uptake in the process. Other than that, the only changes we made to Geenen's model were the definitive suppression of the folate cycle and the application to the metabolism of paracetamol. Finally, we added two extra genes (*i.e.*, *HMOX1* and *SRXN1*) which are often used as activation markers for Nrf2 pathway (**Figure 8**).

Assembling those two models was a multi-step process that started with the deep understanding of the functioning and specificities of each of the two models and then by spotting the common points between them. Next, the fusion of the two models required a rigorous work of homogenization of names and symbols of all participating elements (*i.e.*, state variables, reaction names, parameters, constants, volumes, exposure molecule(s), *etc.*) between the two models. Some differences between the two models emerged at this stage. For instance, the 'gamma-glutamyl-cysteine' (γGC) enzyme was named 'glc' in Geenen's model and 'r-GC' in Hamon's. In Geenen's model, the synthesis of γGC was catalyzed by the enzyme 'glutamyl cysteine synthetase' when the same reaction in Hamon's model was catalyzed by GCL and GCLC, and consumes ATP (**Figure 9**). For this reaction, ATP and the action of GCL and GCLC from Hamon were taken into account as an added value to the equation, and integrated to GSH synthesis according to Geenen's model. **Figure 7** is a schematic representation of the final

(assembled) SB model we constructed, showing all that happens between the entry of the xenobiotic X to the cell and the GSH cycle (*i.e.*, synthesis, oxidation and export), passing by the nuclear transcription of genes coding for key enzymes. The full code of this SB model is given in 'Supplementary Material 7.4'.

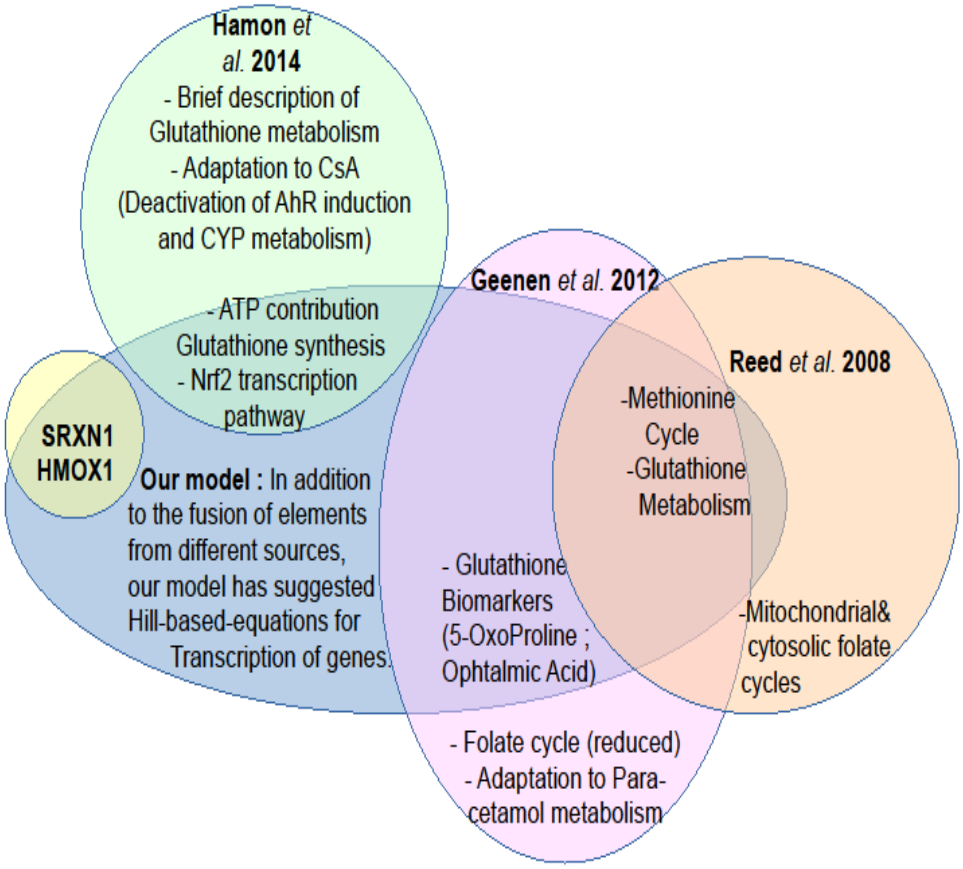


Figure 8. Venn diagram showing the contribution (overlapping areas) of different source models (*i.e.* Hamon *et al.* (2014) in green, Geenen *et al.* (2012a) in pink and Reed *et al.* (2008) in orange) to our final assembled SB model (in blue) describing the control of oxidative stress by the Nrf2-Keap1 signaling pathway. This diagram also shows the parts of each model that were left out (non-overlapping areas). Two more genes (*i.e.*, *SRXN1* and *HMOX1*) that are often used as activation markers for Nrf2 pathway were added to the model (yellow).

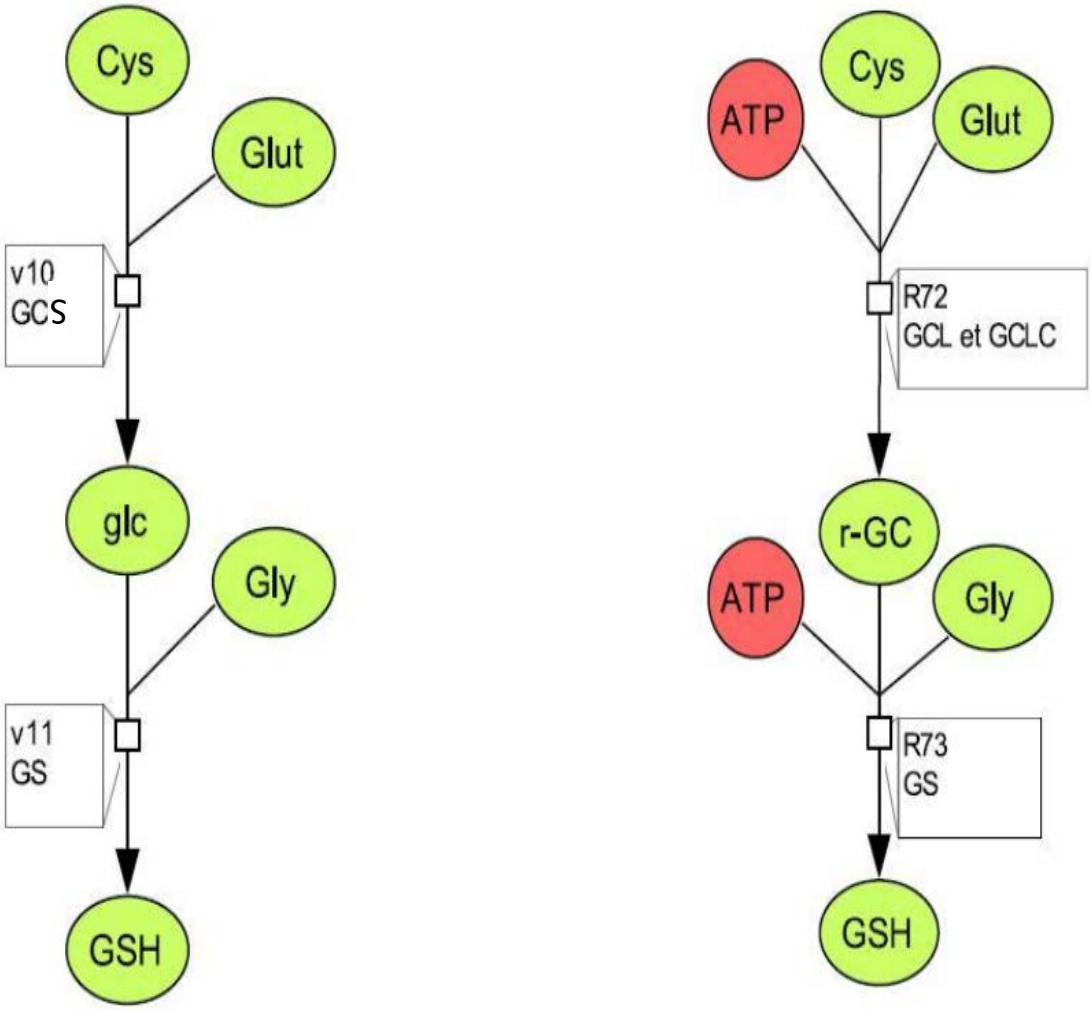


Figure 9. γGC and GSH synthesis reactions according to Geenen *et al.* (2012a) (left) and to Hamon *et al.* (2014) (right). [Cys = cysteine, Glut = glutamate, glc and r-GC = gamma-glutamyl-cysteine; other acronyms are explained in the 'List of abbreviations'].

3.3 RESULTS

Hill parameter values obtained by MCMC simulations for all eight genes are listed in **Table 4**. These parameters were used to plot the curve fitting graph for each gene, in order to check the equivalency of both versions of transcription model: Hamon's model (old version) and Hill-based model (the new version). In this section, we have shown one example to illustrate each of the two cases we have: *GCLC* (**Figure 10**) for genes that are under the effect of one single TF (either nucNrf2 or nucX-AhR) and *MRP* (**Figure 11**) for genes that are activated by both TFs (nucNrf2 and nucX-AhR). The rest of the graphs are presented in 'Supplementary Material 7.1': *CYP* (**Figure S4**), *GCLM* (**Figure S5**), *GS* (**Figure S6**), *GST* and *GPX* (**Figure S7**) and *Nrf2* (**Figure S8**).

For graphs of the genes that are activated by a single TF (*i.e.*, *CYP*, *GCLC*, *GCLM* and *GS*), ten data-points generated with Hamon's model are displayed (red dots), and the results generated by the Hill-based model are represented by a black curve. The figures (**Figure 10**, **Figure S4**, **Figure S5** and **Figure S6**) display the amount of mRNA (in zeptomol) in the cytosol, for each of the two versions of the model, through the timeline, following the exposures as described in the protocol of **Table 3**. As we can see, the black curves pass through all red dots of all four figures. This shows that the two versions are equivalent for these genes and can be interchangeable. For graphs of the genes that are activated by two TFs (*i.e.*, *CYP*, *GCLC*, *GCLM* and *GS*), ten data-points generated with Hamon's model, for each dose of nucX-Ahr (five different colored curves), are displayed (colored dots). Five nucNrf2 doses are added through the timeline: refer to the experimental protocol of **Table 3**. The figures (**Figure 11**, **Figure S7** and **Figure S8**) display the amount of mRNA (in zeptomol) in the cytosol, for each of the two versions of the model, through the timeline. In these cases, the fit is not as good as for genes that are activated by a single TF, but it is still acceptable, since the error between the curves and the dots remains small. So basically, a much simpler Hill model can successfully

replace a cascade of differential equations of the original Hamon's model. The new model replaces 78 parameters and 46 differential equations by 8 Hill's equations and a total of 30 parameters.

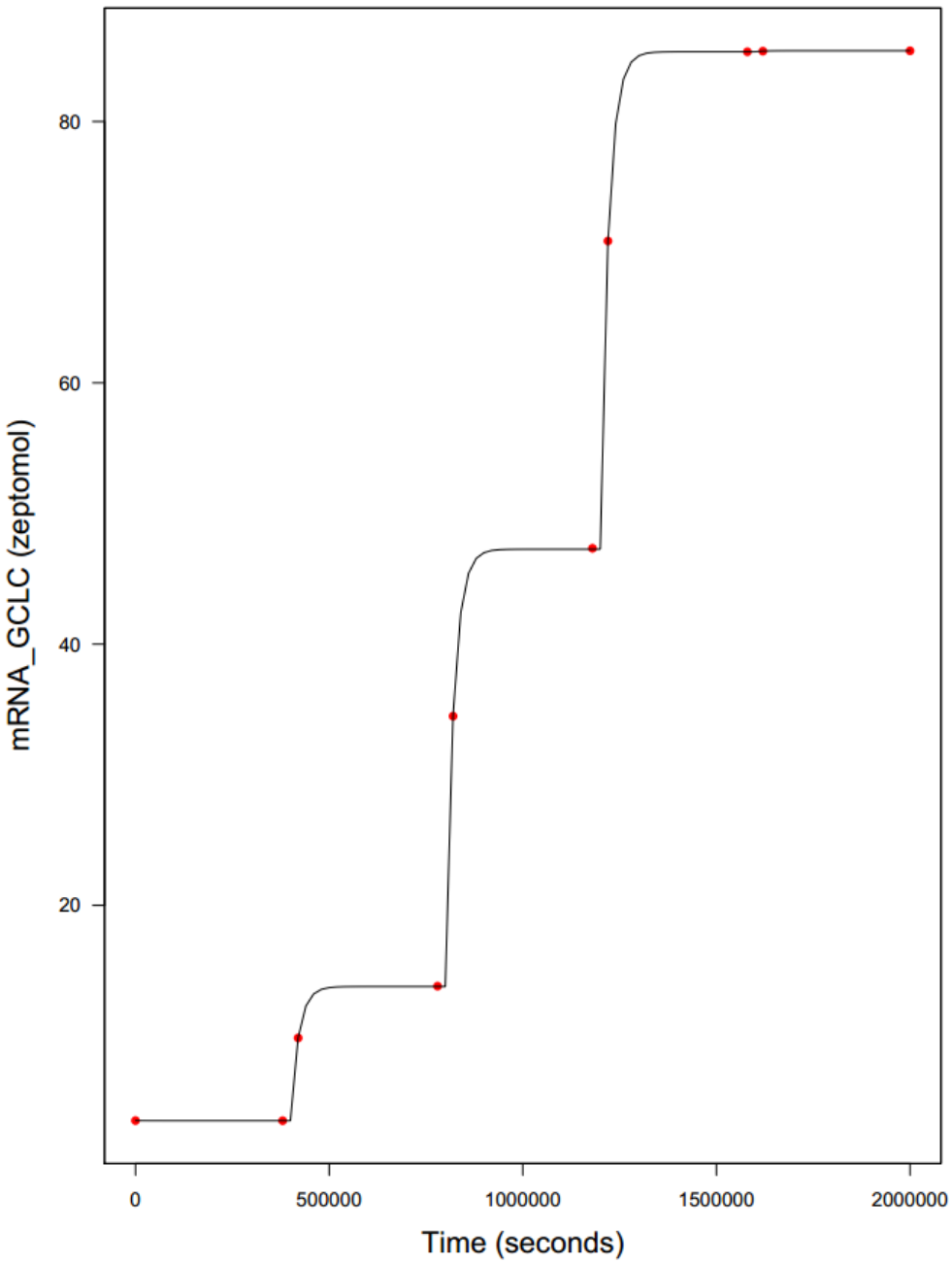


Figure 10. MCMC curve fitting of *GCLC* mRNA (example of gene activated by one single TF) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (red dots) and Hill-based (black curve) SB models.

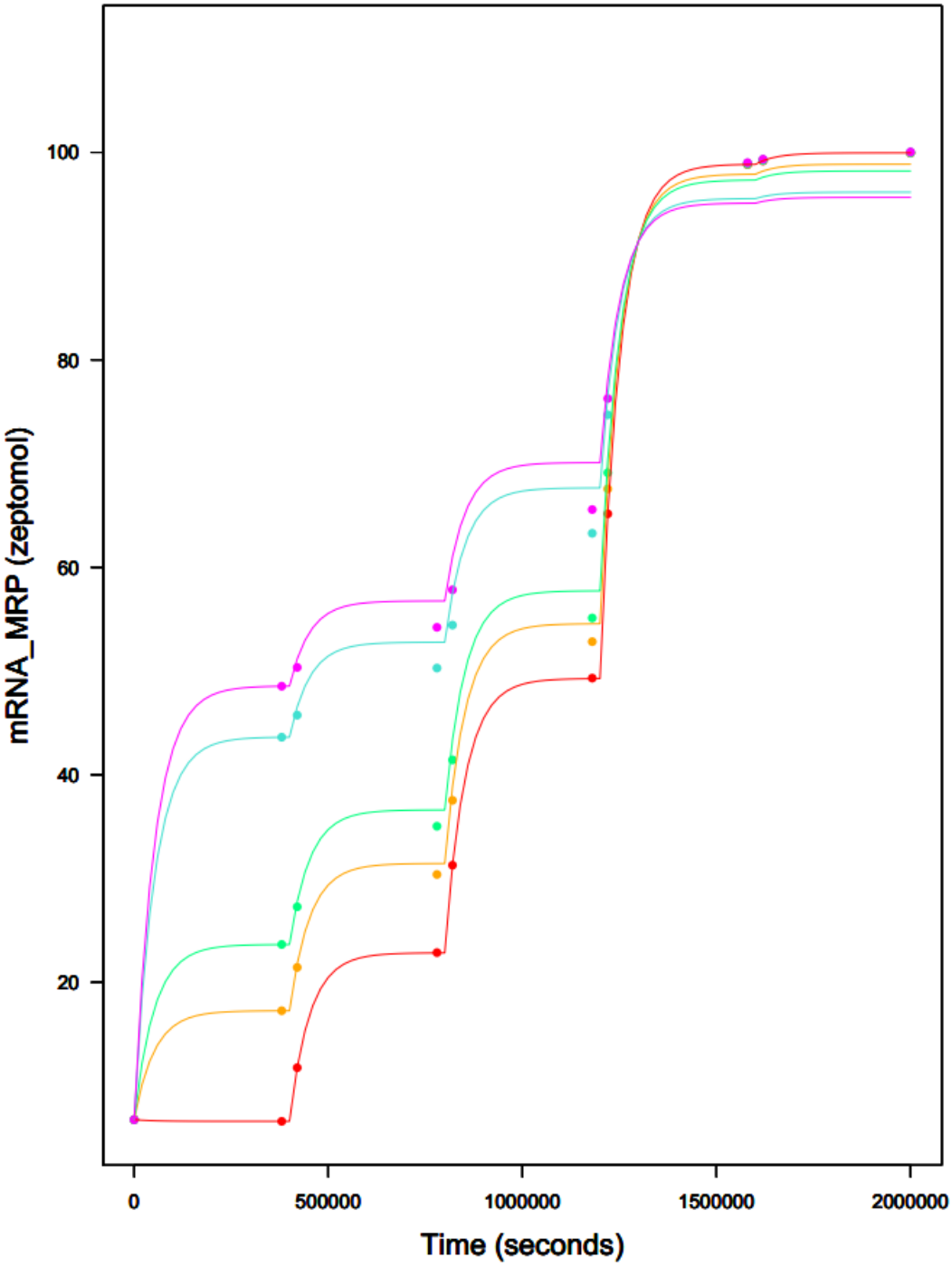


Figure 11. MCMC curve fitting of *MRP* mRNA (example of gene activated by two TFs) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (colored dots) and Hill-based (colored curves) SB models. nucNrf2 dose increase is operated over time (every 400,000 seconds) and nucX-AhR dose is displayed on different curves (0 (red), 0.5 (orange), 1 (green), 10 (blue) and 100 (magenta) zeptomols of nucX-AhR).

Table 4. Hill parameter values (maximum posterior values) for gene transcription in the SB assembled model of the Nrf2 control of oxidative stress. These values were obtained by MCMC simulations. Since calibration was performed with virtual data, we were not interested in the mean and the standard deviation of the distributions (not mentioned).

Gene	Hill Parameter	Value	Measuring unit
CYP	k_0	8.91E-5	zeptomol/s
	V_{max}	4.24E-6	µM/s
	k_m	9.76E-3	µM
Nrf2	k_0	2.73E-6	zeptomol/s
	$V_{max,a}$	2.71E-8	µM/s
	$V_{max,b}$	2.37E-8	µM/s
	$V_{max,ab}$	2.51E-8	µM/s
	$k_{m,a}$	1.50E-3	µM
	$k_{m,b}$	1.20E-3	µM
GS	k_0	1.21E-4	zeptomol/s
	V_{max}	5.47E-6	µM/s
	k_m	4.30E-3	µM
GCLC	k_0	1.70E-4	zeptomol/s
	V_{max}	9.89E-6	µM/s
	k_m	2.39E-3	µM
GCLM	k_0	9.17E-5	zeptomol/s
	V_{max}	1.28E-5	µM/s
	k_m	3.83E-3	µM
GST and **GPX**	k_0	4.92E-5	zeptomol/s
	$V_{max,a}$	1.17E-6	µM/s
	$V_{max,b}$	1.62E-7	µM/s
	$V_{max,ab}$	1.87E-7	µM/s
	$k_{m,a}$	3.18E-3	µM
	$k_{m,b}$	3.11E-3	µM
MRP	k_0	1.27E-4	zeptomol/s
	$V_{max,a}$	4.50E-6	µM/s
	$V_{max,b}$	2.06E-6	µM/s
	$V_{max,ab}$	2.27E-6	µM/s
	$k_{m,a}$	2.72E-3	µM
	$k_{m,b}$	3.75E-3	µM

4 SB AND OTHER TOOLS FOR THE DEVELOPMENT OF QUANTITATIVE AOPS

4.1 STUDY CONTEXT

AOPs, the organizing principle for all OECD's testing are practically a chemical-independent description of a linear path from a MIE to an eventual AO at the organism or population level, as already earlier mentioned in the Bibliography 'sub-chapter 2.2.3' (Villeneuve *et al.*, 2014) (see **Figure 12**). Between the MIE and the AO, there can be any number of intermediate critical and measurable KEs connected by KERs (LaLone *et al.*, 2017).

In the long term, AOPs can support the development of 'integrated testing strategies' and their application in risk assessment (Leist *et al.*, 2017; Vinken, 2013). In case of 'integrated testing strategies' 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 (OECD, 2016; Sachana and Leinala, 2017). For this purpose, quantitative AOPs that provide dose-response and time-course predictions (Conolly *et al.*, 2017) are likely to be more valuable for 'integrated testing strategies' construction than qualitative AOPs. Parameter values for a qAOP can be either obtained from legacy data or new targeted experimental work, or by optimizing the fit of model predictions to data (Villeneuve *et al.*, 2014). So far, the few qAOPs published use either empirical dose-response models to quantify KERs (*e.g.*, (Hassan *et al.*, 2017), or are based on an underlying SB model (*e.g.* Conolly *et al*., 2017). In canonical linear AOP diagrams, KEs are represented by boxes and KERs by arrows connecting them, without cycles (**Figure 12**). The path linking the various KEs should not form loops (feedback of feed-forward loops between two consecutive KEs can simply be indicated by a symbol). Thus, according to graph theory, AOPs are acyclic directed graphs (Pavlopoulos *et al.*, 2011). which are the underlying structure of BNs (Oates and Mukherjee, 2012). The links between their nodes correspond to simple statistical dependencies.

Thus, BNs can be viewed as an intermediate approach between empirical models and SB models. They have already been applied to AOPs in the area of skin sensitization to facilitate potency assessment for classification purposes and to support hazard characterization in a semi-quantitative way (Jaworska *et al.*, 2015; Rovida *et al.*, 2015). Here, we further demonstrate the application of DBNs to AOP quantification.

The work described in this chapter is intended to validate the SB model of the Nrf2 control of oxidative stress described in 'chapter 3' using experimental data and to compare its behavior to two other models (a statistical dose-response relationships model and a DBN model). This study will be soon submitted to publication.

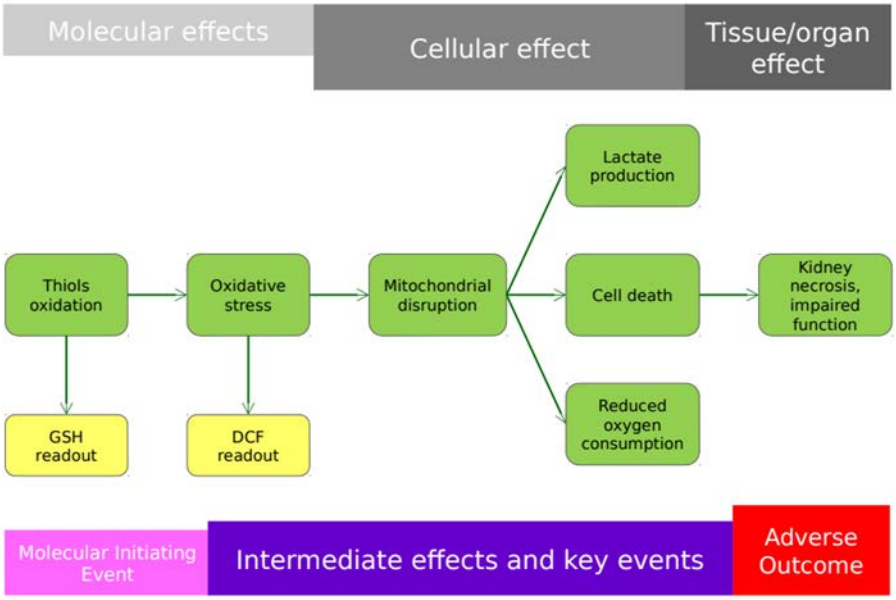


Figure 12. A CKD AOP diagram. KERs are represented by arrows.

4.2 METHODS

4.2.1 Experimental data

Thiol oxidation following exposure to various concentrations of potassium bromate ($KBrO_3$) (control, 0.375, 0.75, 1.5. 3, a nd 6 mM) (see **Table S1** in 'Supplementary Material 7.2.1') was measured through GSH depletion in a cel l-free environment. Depletion was measured after 1 hour, using the luminescence-based GSH-Glo kit from Promega (V6912), according to manufacturer's instructions, as described in Limonciel *et al.* (2012)

Oxidative stress was read out by oxidation of the cell permeant reagent 6-carboxy-2',7'-dichlorofluorescein diacetate (carboxy-DCFDA). After diffusion into cells, carboxy-DCFDA is deacetylated by cellular esterases to 6-carboxy-2',7'-dichlorofluorescein (carboxy-DCF), which remains trapped in the cell and is oxidized by hydroxyl, peroxyl radicals and other ROS to 2',7'-dichlorofluorescein (DCF), which is highly fluorescent. RPTEC/TERT1 cells were grown as described by Aschauer *et al.* (2013) and exposed to various concentrations of $KBrO_3$ (control, 0.75, 1.5. 3, a nd 6 mM) as described by Limonciel *et al.* (2012)) (see **Table S2** in 'Supplementary Material 7.2.1'). Briefly, cells were grown and matured into a mature monolayer in 96-well cell culture plates kept at 37°C / 5% CO_2 and were fed fresh medium 24 hours before chemical exposure. Cells were incubated with 40µM carboxy-DCFDA (Invitrogen) 4 hours before washing out the excess extracellular dye and starting exposure to $KBrO_3$ dissolved in culture medium. DCF production was measured over time (approximately every 15 m inutes, up t o 24 hour s) as relative fluorescence units (RFU) by fluorescence spectroscopy using over time using a Tecan Pro M200 microplate reader.

Supernatant lactate per time is a measure of glycolytic rate (inversion of glucose consumption rate). Increased glycolysis can be due to a decrease in mitochondrial respiration, an increase in energy demand, or alteration in pathways involved in glycolysis regulation (*e.g.*, HIF1 alpha (positive), or p53 (negative). The culture medium, with the given $KBrO_3$

concentrations was changed every day after an aliquot was taken for lactate measurement. An absorbance-based assay described in Limonciel *et al.* (2011)) (see **Table S3** in 'Supplementary Material 7.2.1').

4.2.2 Chronic Kidney Disease – CKD – AOP

The proposed AOP (**Figure 12**) links thiol oxidation to CKD via oxidative and mitochondrial stress. Within the nephron, the proximal tubule is especially susceptible to injury from oxidative chemicals, as they can cause mitochondrial damage, which in turn can result in impairment of active and secondary transport, as well as in cell death. CKD is characterized by a progressive loss of renal function, the onset of which is initiated and or accelerated by other factors such as diabetes, high blood pressure or exposure to nephrotoxic chemicals (Aschauer *et al.*, 2015; Staples and Wong, 2010). Given its high energy demand for active transport, the nephron proximal tubule is especially susceptible to injury from oxidative chemicals, as they can cause mitochondrial damage (Kong *et al.*, 2014). Here, we analyze the AOP until the initiation of cell death following induction of oxidative stress, since our analysis is based on *in vitro* data obtained in human proximal tubule (RPTEC/TERT1) cells exposed to $KBrO_3$. The link from cell death to kidney function impairment thus cannot be modeled based on the available data and we will focus on a set of core early KEs leading to proximal tubule damage.

4.2.3 Dose-Response based qAOP

In the empirical dose-response approach, dose(-time)-response equations were fitted to data on the effect of $KBrO_3$ on GSH, DCF and lactate. With such data, linking chemical exposures to KEs, the corresponding equations need to be mathematically inverted to obtain chemical-independent KERs. Only the exposure to MIE relationship can be used as is.

In the work detailed in 'Supplementary Material 7.2.2', the article's co-authors Frédéric Y. BOIS (thesis director) and Cléo TEBBY have created a dose-response based qAOP model of three equations 4.1, 4.2 and 4.3. First, equation 4.1, with a modified exponential decrease, establishes the relationship between the concentration of $KBrO_3$ (C_{KBrO3}) and the percentage of GSH (Pct_{GSH}) remaining *in vitro* after one hour representing the MIE. The parameters used in equation 4.1 are the GSH degradation rate constant k, and power b (which increases the degradation rate if b>1).

$$Pct_{GSH} = 100 \times exp\left(-k \cdot C_{KBrO_3}^{b}\right) \qquad (4.1)$$

Then, equation 4.2 models the relationship between C_{KBrO3}, time t and Q_{DCF} (representing the amount of oxidative stress). Its parameters are A (baseline response), B (maximum increase above baseline), δ (maximum increase modulation by dose), k_d (dose coefficient), k_t (time coefficient).

$$Q_{DCF} = A + B \cdot \left(1 + \delta - exp\left(-k_d \cdot C_{KBrO_3}\right)\right)\left(1 - exp(-k_t \cdot t)\right) \qquad (4.2)$$

Finally, the polynomial equation 4.3 that models the C_{KBrO3} - time - lactate concentration (C_{lac}) relationship, fits the data adequately.

$$C_{lac} = a + bC_{KbrO_3} + (c + eC_{KbrO3})t + (d + fC_{KbrO3})t^2 \qquad (4.3)$$

4.2.4 Bayesian Network – BN – qAOP

The AOP shown on **Figure 12** can be taken as a BN structure. Here, we do not need to learn our BN structure, but we need to calibrate it. Given that we have dose-time-response data on DCF and lactate, and that their time evolution is progressive rather than instantaneous, it is natural to use a DBN to integrate those data (Kjærulff and Madsen, 2008). **Figure 13** shows the DBN that the article's co-author Wang GAO constructed to quantify the CKD AOP. In this figure, the DCF readout at a given time depends on its previous value (indeed, in the *in vitro* system DCF accumulates with time). The same applies to the lactate concentration. There are also some instantaneous or constant dependencies: We considered that C_{KBrO3} was constant with time throughout the experiments (note that this is an approximation, but we have no information on the kinetics of $KBrO_3$ in the *in vitro* system). The thiol depletion readout (GSH level remaining after 1 hour) is simply an indicator of $KBrO_3$ potency and was also taken to be constant.

In the work of Wang GAO that is explained in 'Supplementary Material 7.2.3', the developed DBN model includes three equations 4.4, 4.5 and 4.6. In equation 4.4, describing the dependence of observed Pct_{GSH} on C_{KBrO3}, k_{GSH} represents the depletion rate constant and σ^2_{GSH} the variance.

$$Pct_{GSH} \sim Normal(100 \times exp(-k_{GSH} \cdot C_{KBrO3}), \sigma^2_{GSH}) \qquad (4.4)$$

The conditional distribution of Q_{DCF} observations at a given time t, given Pct_{GSH} and the Q_{DCF} observation at the previous time t-h is given by is an extension of the standard DBN model in which $Pct_{GHS,t}$ influences the equilibrium value ($E_{DCF,t}$) for $Q_{DCF,t}$ to which it converges over time at exponential dampening rate ν (equations 4.5 and 4.6):

$$Q_{DCF,t} \sim Normal\left(E_{DCF,t} - \left[E_{DCF,t} - Q_{DCF,t-h}\right] \cdot e^{-\nu_{DCF}h}, \sqrt{\frac{1-e^{-\nu_{DCF}h}}{1-e^{-\nu_{DCF}}}} \cdot \sigma^2_{DCF}\right) \quad (4.5)$$

$$E_{DCF,t} = \beta_{0,DCF} + \beta_{DCF}\, Pct_{GSH} \quad (4.6)$$

Where $E_{DCF,t}$ is the equilibrium value of Q_{DCF} (a linear function of $Pct_{GSH,t}$ at time t), h is the (positive) time interval between two consecutive observations, ν_{DCF} (positive), $\beta_{0,DCF}$, β_{DCF}, and variance σ^2_{DCF} are parameters to estimate.

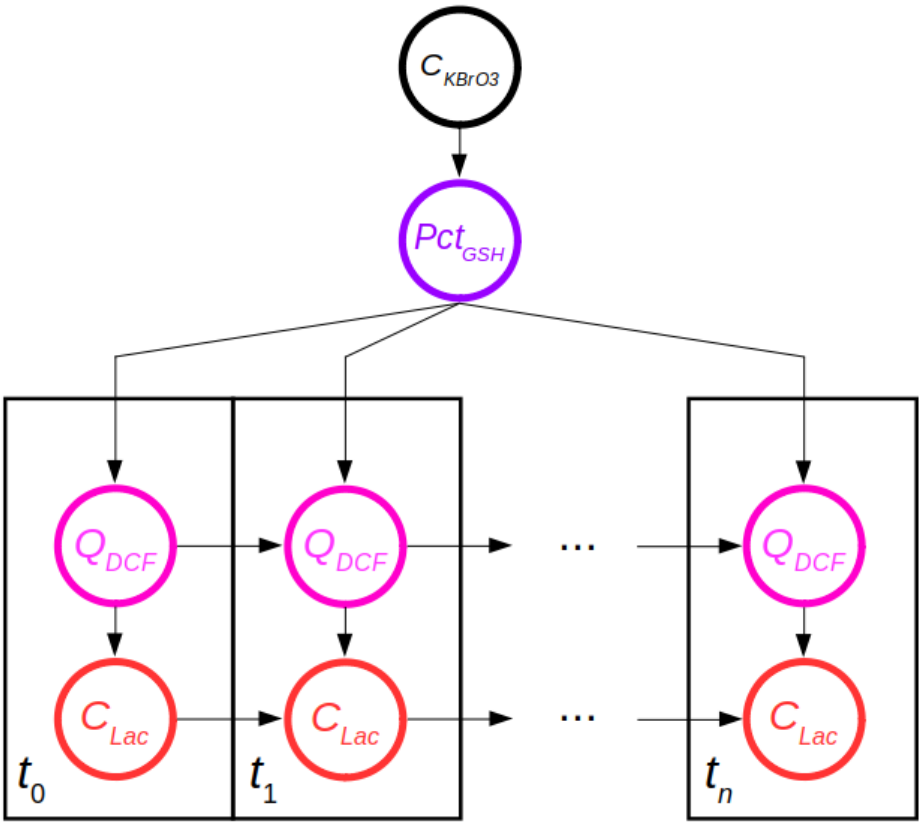


Figure 13. Structure of the DBN qAOP for CKD. $KBrO_3$ concentration and the GSH readout do not vary with time, while the DCF and lactate readouts were observed at different time intervals. The arrows indicate probabilistic dependencies.

4.2.5 The Systems Biology – SB – Model

We used the SB model developed in 'chapter 3' to analyze of the oxidative stress (DCF) data. The model does not describe lactate formation and hence we did not consider the lactate data in this approach. As mentioned before, this SB model focuses on control of the oxidative stress by Nrf2 and GSH, one of the major toxicity pathway studied in systems toxicology (Geenen *et al.*, 2012a; Hamon *et al.*, 2014; Jennings *et al.*, 2013). Therefore, we used it only to study the relationship between $KBrO_3$ exposure, time, and DCF fluorescence in detail.

Upon oxidative stress, when the intra-cellular level of ROS exceeds the capacity of this defense system to replenish GSH through new synthesis, GSH depletion occurs and ROS are left free to cause extensive cellular damage, cell death, nephron attrition and CKD.

Figure 7 shows the assembled SB model we developed to study the transcriptional regulation of the GSH pathway by the Nrf2 - Keap1 complex, which merges variants of the with the Hamon *et al.* (2014) model for RPTEC/TERT1 cells and a model developed by Geenen *et al.* (2012a).

In order to calibrate the model with the experimental data on $KBrO_3$ effect on GSH and DCF, we added several first order reactions to the model (**Figure 14**): (a) Action of $KBrO_3$ on extra-cellular GSH, with parameter $k_{GSHe,KBrO3}$; (b) Formation of DCF from carboxy-DCF by ROS-mediated oxidation, parameter $k_{DCF,ROS}$; (c) Bleaching of DCF, parameter k_{bl} ; (d) Formation of DCF from carboxy-DCF by direct action of $KBrO_3$, parameter $k_{DCF,KBrO3}$; (e) Action of $KBrO_3$ on intra-cellular GSH, parameter $k_{GSHc,KBrO3}$ (this parameter is multiplied by $k_{GSHe,KBrO3}$ to yield the reaction rate constant, and is in fact the ratio of the external to internal reaction rate constants).

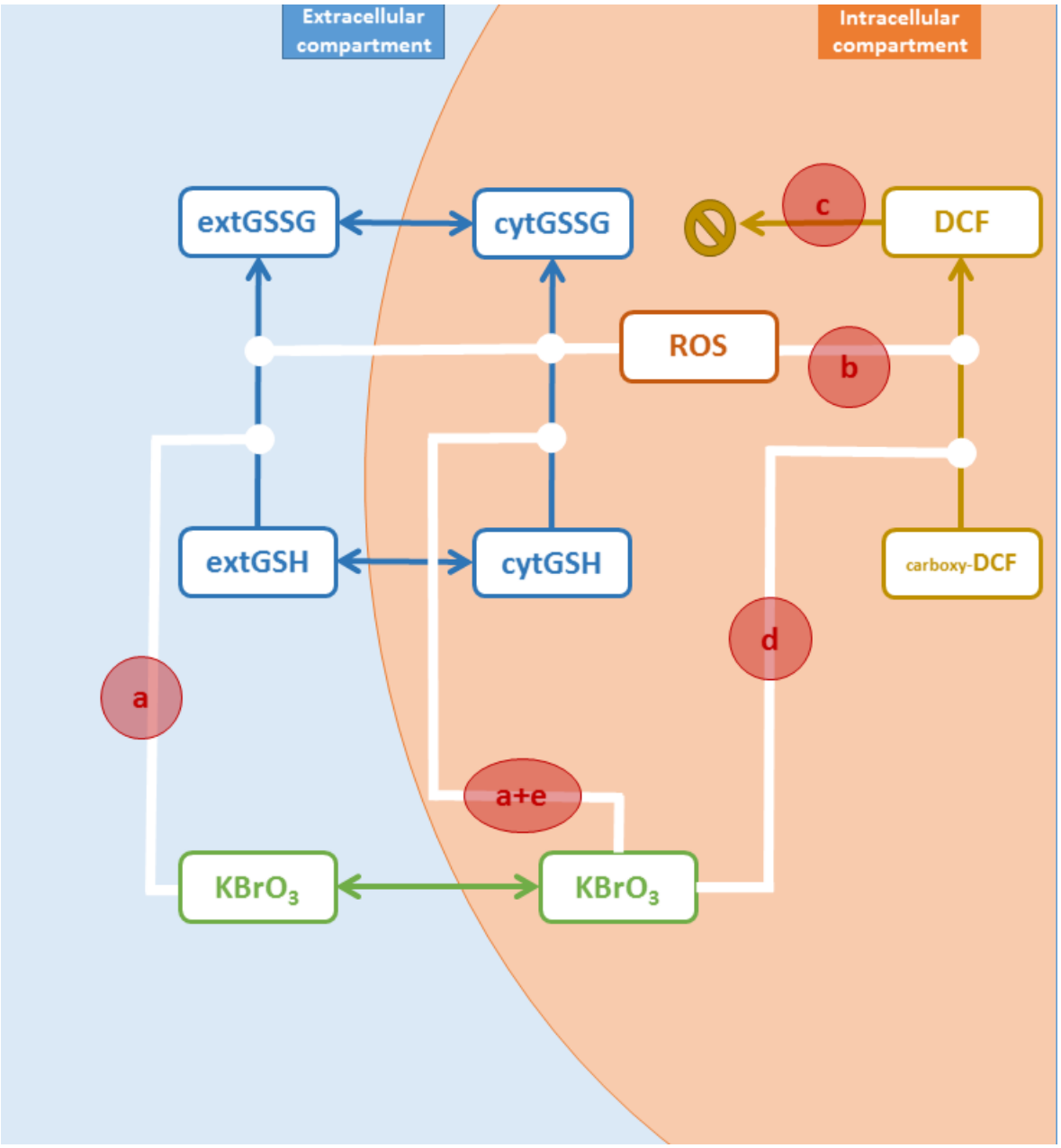


Figure 14. $KBrO_3$ and DCF specific reactions of the SB model. Other abbreviations: extGSH is extra-cellular glutathione; cytGSH: cytosolic glutathione; extGSSG: extra-cellular oxidized glutathione; cytGSSG: cytosolic oxidized glutathione. Reactions are represented by red circles: **a.** the oxidation of extGSH by $KBrO_3$; **b.** oxidation of carboxy-DCF by ROS; **c.** DCF bleaching; **d.** oxidation of carboxy-DCF by $KBrO_3$; **e.** oxidation of cytGSH by $KBrO_3$.

4.2.6 Parameter Estimation

Parameter calibrations for the three types of qAOPs investigated were done in a Bayesian statistical framework, using MCMC simulations (Bernillon and Bois, 2000; Bois, 2012), or Hamiltonian MCMC (Girolami and Calderhead, 2011). Basically, for each parameter to calibrate, a prior distribution summarizing existing knowledge was updated on the basis of the likelihood of the current data to yield a posterior distribution. Those distributions were obtained by random sampling from several simulated Markov chains. The convergence of the simulated chains was checked using the R_{hat} criterion of Gelman and Rubin (1992).

The complexity of the various qAOP models differed and slightly different sampling strategies were used. For parameters estimation of the dose-response based model and for the DBN model, please refer to **Table S4**, **Table S6** and the explanation in 'Supplementary Material 7.2.2' and '7.2.3'.

For the SB model, parameter calibration was done by Metropolis-Hastings MCMC with *GNU MCSim* (Bois, 2009a). Three Markov chains of 30,000 iterations were run in parallel, keeping the last 5,000 iterations. For each estimated parameter, non-informative uniform prior distributions were used (see **Table 5**).

The data likelihood is clearly separated from the structural equations. To calibrate the model with our experimental data on the effect of $KBrO_3$ on GSH and DCF, we proceeded step by step, increasing the complexity of the model by introducing reactions according to the following schedule:

1. Action of $KBrO_3$ on extra-cellular GSH (parameter $k_{GSHe,KBrO3}$), on the basis of the $KBrO_3$ - GSH cell-free experiment data; $k_{GSHe,KBrO3}$ was held at its maximum posterior value in the subsequent steps.

2. Action of $KBrO_3$ on extra-cellular GSH (parameter $k_{GSHe,KBrO3}$) and formation of DCF by ROS-mediated oxidation ($k_{DCF,ROS}$): this is a minimal model for explaining the $KBrO_3$ - time - DCF data.

3. Adding bleaching of DCF (k_{bl})

4. Adding the direct formation of DCF by $KbrO_3$ ($k_{DCF,KBrO3}$) (step *4a*) or the action of $KBrO_3$ on intra-cellular GSH ($k_{GSHc,KBrO3}$) (step *4b*)

5. All of the above.

Table 5. Prior distributions of the parameters of the SB qAOP calibrated with the DCF data.

Parameter	Units	Prior distribution
$k_{GSHe,KBrO3}$	$(\mu M.s)^{-1}$	Uniform (0, 10^{-6})
$k_{DCF,ROS}$	$(zmol.s)^{-1}$	Uniform (0, 10^{-6})
k_{bl}	s^{-1}	Uniform (0, 10^{-4})
$k_{DCF,KBrO3}$	$(\mu M.s)^{-1}$	Uniform (0, 10^{-8})
$k_{GSHc,KBrO3}$	-	Uniform (0, 3)
σ_{DCF}	RFU	Normal (1, 0.2) truncated to [1.01, 2]

To compare the eventual improvement in fit brought by those various model refinements we used various measures of model fit to the data: the data log-likelihood, the residual GSD (geometric standard deviation), the AIC (Akaike information criterion) (twice the difference between the number of parameters and the data log-likelihood), the BIC (Bayesian information criterion), and the DIC (Deviance information criterion) (Gelman *et al.*, 2004).

4.2.7 Uncertainty propagation

The output of MCMC simulations is a sample of parameter sets (or parameter vectors) drawn from their joint distribution. Those sets of parameter values were used to rerun the corresponding model to make predictions for unobserved values. This is a type of Monte Carlo simulations in which the MCMC sampler acts as a random parameter values generator. We obtained distributions of predicted values that reflect the uncertainty of all parameter values.

4.2.8 Software

The dose-response based qAOP and the SB model were simulated and calibrated with the *GNU MCSim* software, version 5.6.6 (Bois, 2009a)[20]. The BN qAOP model was simulated and calibrated using *Stan* (Carpenter *et al.*, 2017)[21]. All plots were created with *R*, version 3.4.4 (R Development Core Team, 2013)[22]. Effectopedia[23] version 1.2.51 (OECD, 2016) was used for implementation of the qAOP on internet (for Effectopedia, please check 'Supplementary Material 7.2.5'). Effectopedia is an OECD guideline-compliant software tool that aims to gather experimental data and models in a unified representation, so that they can be compared, further analyzed, and used for hazard and risk assessment purposes (OECD, 2017).

[20] https://www.gnu.org/software/mcsim/ [Accessed October 24th, 2018]
[21] http://mc-stan.org/ [Accessed October 24th, 2018]
[22] https://cran.r-project.org/ [Accessed October 24th, 2018]
[23] https://www.effectopedia.org/ [Accessed October 24th, 2018]

4.3 RESULTS

4.3.1 Dose-Response based qAOP Model

The empirical dose response models given by equations 4.1, 4.2 and 4.3 described the $KBrO_3$ - GSH, $KBrO_3$ - time - DCF, and $KBrO_3$ - time - lactate relationships reasonably well (see **Figure 15** and **Figure 16**, top row). Equivalent 2D representations of the time course of DCF and lactate at the various $KBrO_3$ concentrations are given in 'Supplementary material 7.2.2' **Figure S9** and **Figure S10**, respectively. The uncertainty of the model predictions is low for GSH (**Figure 15**), and it amounts to about 0.5% to 1.5% for DCF and 5% to 12% for lactate (this cannot be usefully represented on **Figure 16** for reasons of readability). Residual uncertainty (an estimate of measurement error) is about 22% for GSH, 20% for DCF and 30% for lactate. **Table S5** in 'Supplementary Material 7.2.2' summarizes the 'posterior distributions' of the parameter values obtained by Bayesian calibration.

By inversion of the empirical models, we can deduce the relationship between GSH, time, and DCF or GSH, time, and lactate production (**Figure 16**, bottom row). These relationships should, in theory, be independent of the thiol reactive chemical. They can be used to make predictions, including full parametric uncertainty propagation since we used a Bayesian statistical framework for parameter inference. For example, a chemical dose causing 80% reduction of GSH after 1 hour (*i.e.*, 20% GSH left), in the test conditions described in the 'Methods 4.2', should lead to a lactate concentration of 4.6 ± 0.3 [4.1, 5.1] mM (mean, SD, 5 and 95 percentiles) after 3 days of exposure.

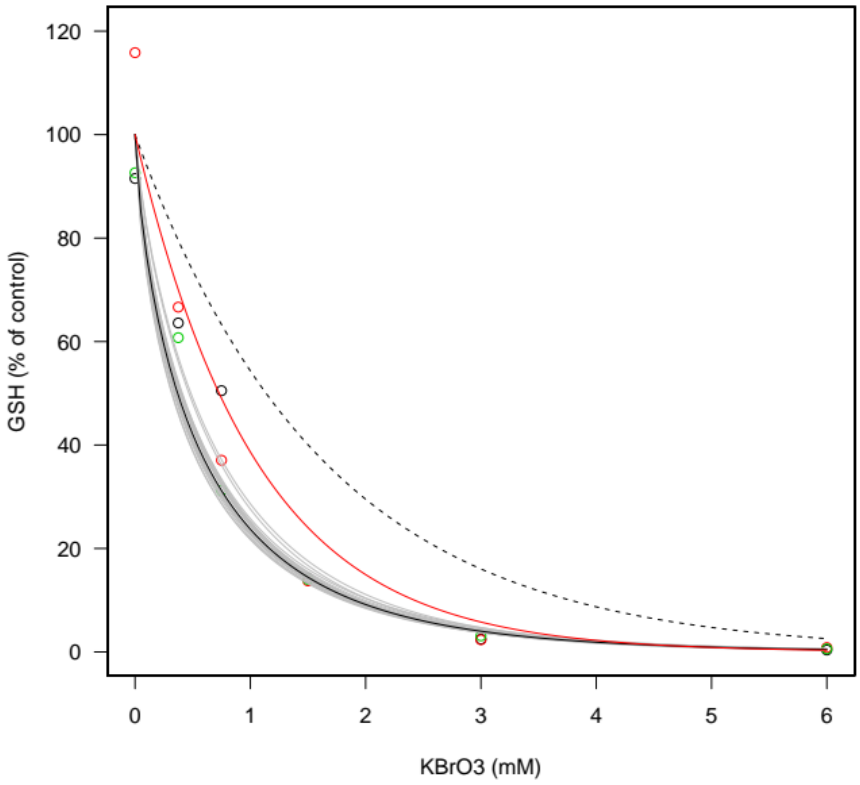


Figure 15. Fit of the $KBrO_3$ - GSH data (circles; each color represents one of the replicates) using the three qAOP models developed. The black line corresponds to the empirical model (equation 4.1). The best fit (solid line) is shown along with 20 additional random fits (gray), showing the uncertainty of the model predictions. The black dashed line represents the best fit obtained the DBN qAOP. The red line shows the best fit for the SB model.

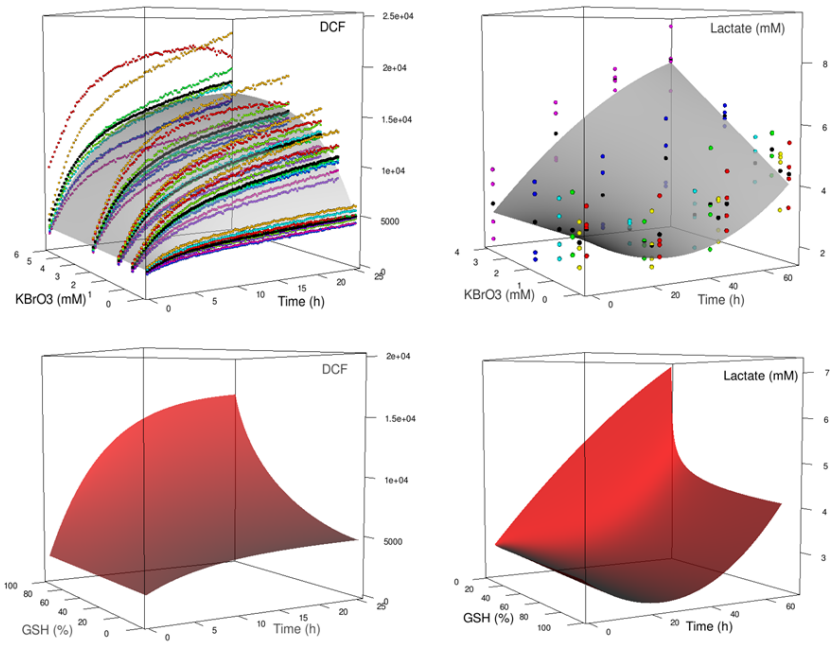


Figure 16. Fit (top row) and predictions (bottom row) of the dose-response based qAOP for the DCF (measured in RFU) (left) and lactate (right) readouts. The best fit surfaces (gray) are plotted along with all individual data (colored dots). The predicted chemical-independent relationships (in red) for GSH - time - DCF, or GSH - time - lactate were obtained by inversion of the qAOP equations (see 'Supplementary Material 7.2.2'). The maximum posterior parameter values given in **Table S5** were used to draw the figures.

4.3.2 Bayesian Network – BN – qAOP Model

The fit of the DBN qAOP to GSH, DCF, and lactate data is shown on **Figure 15** and **Figure 17**. Equivalent 2D representations are given in 'Supplementary material 7.2.3' **Figure S11** and **Figure S12**. The fits for GSH and DCF are less good than those of the empirical models. The fit to the lactate data (**Figure 17**) looks very different for the DBN model, because the DBN model takes into account the change of medium every 24 hours. Note that all parameters of the DBN model are estimated together, so that modeling error are spread over the overall dataset. Also, the model uses linear relationship between nodes, except for the link $KBrO_3$ - GSH. Residual uncertainty (an estimate of measurement error) is about 50% for GSH, 25% for DCF and 10% for lactate. The error model, however, is different (normally distributed residuals, rather than log-normally distributed as in the empirical model). **Table S7** in 'Supplementary Material 7.2.3' summarizes the 'posterior distributions' of the parameter estimates obtained. The model parameters have some physical interpretation: Parameter v controls the speed at which plateaus are reached in **Figure 17**. The β parameters condition the height of the plateaus. However, there is a subtle interplay between convergence speed, plateau level, time and dose, as can be seen on **Figure 15**. All parameters are significantly different from zero.

The DBN qAOP model does not need mathematical inversion to produce chemical-independent predictions of the levels of DCF and lactate as a function of GSH depletion and time, because they can be directly simulated (**Figure 17**, bottom row). The resulting relationship for DCF is quite similar to that obtained with the previous qAOP (except for the linearity of the GSH - DCF relationship). However, the GSH - lactate relationship is very different, even though constant exposures to $KBrO_3$ are simulated in both cases (the simulation is now considering a single medium change at time point zero). Note that lactate starts at zero to reach a plateau in about three days. The relationship between GSH and lactate is predicted to be linear

by the DBN model, instead of being strongly nonlinear in the empirical qAOP. As before, predictions with uncertainty estimates can be easily made. For example, the DBN qAOP predicts that a chemical dose causing 80% reduction of GSH after 1 hour (*i.e.*, 20% GSH left), leads to a lactate concentration of 5.8 ± 0.4 [5.2, 6.5] mM (mean, SD, 5 and 95 percentiles) after 3 days of exposure. This is significantly different from the prediction of the empirical qAOP.

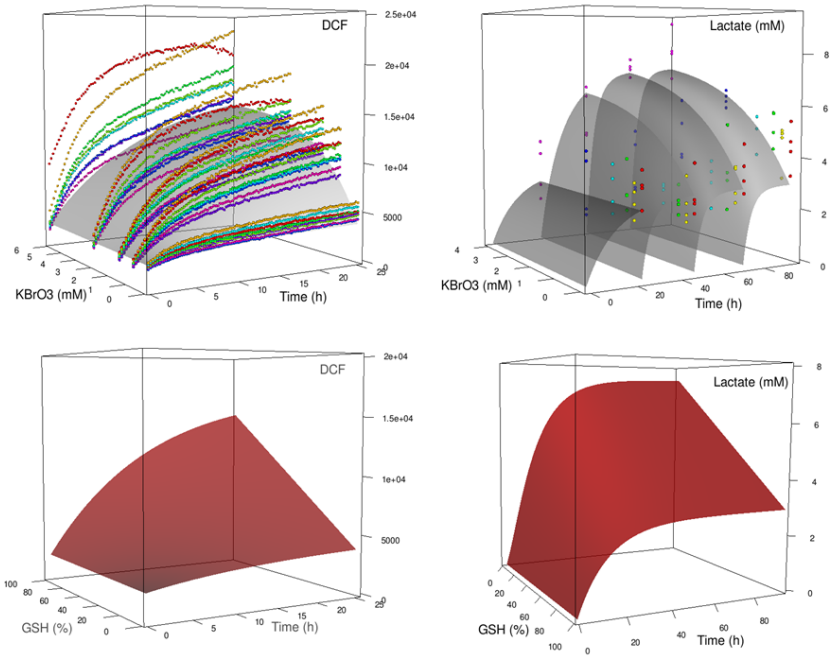


Figure 17. Fit (top row) and predictions (bottom row) of the DBN qAOP for the DCF (measured in RFU) (left) and lactate (right) readouts. The best fit surfaces (gray) are plotted along with the data mean (black dots) and all individual data (colored dots). The predicted chemical-independent relationships (in red) are shown for GSH - time - DCF and GSH - time - lactate. The maximum posterior parameter values given in **Table S7** were used to draw the figures.

4.3.3 System biology – SB – Model

The fit of the SB model to the GSH data (calibration step 1) is show on **Figure 15** (red line). It is better than the fit of the DBN qAOP (residual uncertainty for the GSH data is about 40%), despite the fact that both use the same decreasing exponential relationship between $KBrO_3$ and GSH. However, the $k_{GSHe,KBrO3}$ parameter was calibrated to the data independently of the other parameters and its fit is not constrained by the other data. The fits obtained for the $KBrO_3$ - time - DCF data at the various model calibration steps (parameters were re-calibrated at each step) are shown on **Figure 18**. Equivalent 2D representations are given in 'Supplementary material 7.2.4' **Figure S13** to **S16**. Measures of the quality of fit are given in

Table 7. Note that the model takes into account the 4 hours of cells pre-incubation with carboxy-DCFDA, and simulation time starts therefore before exposure to $KBrO_3$ (which is defined to occur at time point zero). During that period of time, ROS already starts forming DCF, explaining the relatively high level of fluorescence at time point zero. At step 2, with just a depletion of extra-cellular GSH by $KBrO_3$ and the formation of DCF by ROS the model is unable to explain the data (**Figure 18A**). The depletion of extra-cellular GSH has only a minor effect on the intra-cellular GSH level (see 'Supplementary Material 7.2.4' **Figure S13**). Therefore, only background cellular ROS produces DCF, at a constant rate, and the accumulation of DCF is predicted to be linear (according to the experimental protocol carboxy-DCF is expected to be in excess, and not depleted). Allowing DCF bleaching offers an explanation for the leveling off of the DCF fluorescence, yet the effect of $KBrO_3$ is still not explained satisfactorily and the data fit is very poor (Step 3, **Figure 18B**). Adding the possibility that $KBrO_3$ directly oxidizes DCF improves the fit markedly (Step 4a, **Figure 18C**), and the residual error σ_{DCF} goes down to about 20% (see **Table 6**). However, the effect of $KBrO_3$ is linear, which is not exactly what the data shows. Instead of a direct oxidation of DCF by $KBrO_3$, we tested the possibility that $KBrO_3$ acts on intra-cellular GSH (Step 4b, **Figure 18D**). This has a clear effect on DCF production is clearly seen, but is it extremely nonlinear and does not lead to a reasonable fit to the data. Finally, in step 5, we put all the above parameters in the model and re-calibrated them. This did not lead to improvement compared to step 4a (see **Table 7**), and the effect of $KBrO_3$ on intra-cellular GSH was estimated to be nearly absent (data not shown).

Table 6 lists the best value (maximum posterior), the mean, the standard deviation and the confidence interval [2.5 percentile, 9.75 percentile] of each of the parameters calibrated at step 4a (yielding the best and most parsimonious model). The values of the parameters directly related to DCF do not have an explicit biological interpretation because DCF is measured in

RFU (which should be proportional to concentration, but with an unknown proportionality constant). Note that the DCF bleaching rate constant corresponds to a half life of about 6 hours. The SB model can also be used to make predictions, with full uncertainty propagation. For example, a 4 mM concentration of $KBrO_3$ is predicted to lead to a DCF fluorescence of 16600 ± 250 [16200, 17100] RFU (mean, SD, 5 and 95 percentiles) after 24 hours.

Table 6. Summary of the posterior distribution of the five SB model parameters describing the action of $KBrO_3$ on the formation of DCF. The best parameterization (setting $k_{GSHc,KBrO3}$ at zero) is presented.

Parameter	**Units**	**Maximum posterior**	**mean (SD) [2.5pctile, 97.5pctile]**
$k_{GSHe,KBrO3}$	$(\mu M.s)^{-1}$	2.65×10^{-7}	$2.65\times10^{-7} \pm 8.45\times10^{-9}$ [2.48×10^{-7}, 2.81×10^{-7}]
$k_{DCF,ROS}$	$(zmol.s)^{-1}$	1.20×10^{-7}	$1.21\times10^{-7} \pm 3.2\times10^{-9}$ [1.14×10^{-7}, 1.27×10^{-7}]
k_{bl}	s^{-1}	3.50×10^{-5}	$3.50\times10^{-5} \pm 1.4\times10^{-6}$ [3.23×10^{-5}, 3.77×10^{-5}]
$k_{DCF,KBrO3}$	$(\mu M.s)^{-1}$	1.22×10^{-9}	$1.22\times10^{-9} \pm 4.5\times10^{-11}$ [1.13×10^{-9}, 1.30×10^{-9}]
$k_{GSHc,KBrO3}$	-	0	0
σ_{DCF}	RFU	1.20	$1.20 \pm 6.8\times10^{-3}$ [1.18, 1.21]

Table 7. Assessment of the SB model fit to the $KBrO_3$ - time - DCF data using various criteria and for increasing model complexity. The various steps explain the main text of 'Methods 4.2.5'. **Step 1** is omitted since it does not require DCF data (parameter $k_{GSHe,KBrO3}$, quantifying the action of $KBrO_3$ on extra-cellular GSH, was independently calibrated from the GSH data and set to its maximum likelihood value in all cases). The other parameters were introduced as follows: **Step 2**: action of KBrO3 on external GSH and formation of DCF by ROS (parameter $k_{DCF,ROS}$); **Step 3**: adding DCF bleaching (parameter k_{bl}); **Step 4a**: adding a direct formation of DCF by $KBrO_3$ (parameter $k_{DCF,KBrO3}$); **Step 4b**: same as step 3, plus adding an action of $KBrO_3$ on internal GSH (parameter $k_{GSHc,KBrO3}$); **Step 5**: all parameters added.

Step	Maximum log-likelihood	Residual error (GSD)*	AIC	BIC	DIC
2	-4981	1.58	9967	9975	9969
3	-4919s	1.51	9843	9856	9844
4a	-4480	1.20	8969	8986	8969
4b	-4755	1.35	9518	9535	9518
5	-4480	1.20	8970	8992	8971

* **GSD:** best estimate of the geometric standard deviation (the coefficient of variation equals approximately 100×(GSD - 1).

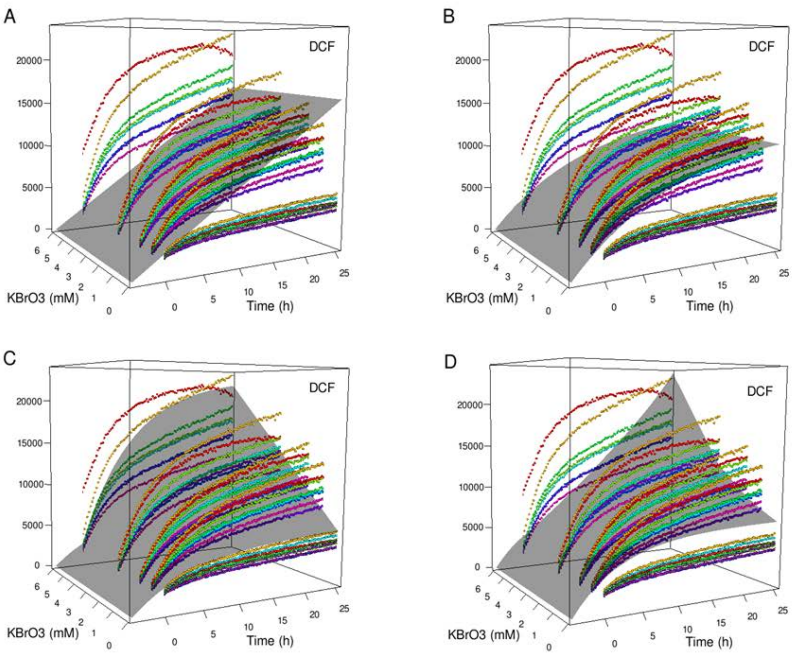


Figure 18. Best fits of SB model (gray surfaces) to the DCF RFU data (colored dots), for different levels of complexity: **(A)** action of $KBrO_3$ on external GSH and formation of DCF by ROS; **(B)** same as A with the addition of DCF bleaching; **(C)** same as B with the addition of a direct formation of DCF by $KBrO_3$; **(D)** same as B, but with the addition of an action of $KBrO_3$ on internal GSH.

4.4 DISCUSSION

In this paper, we explored various options for quantifying an AOP and deriving chemical independent KERs. Quantitative AOPs have been previously described (Conolly *et al.*, 2017; Hassan *et al.*, 2017), but here, we strove for a rigorous statistical treatment of the data and parametric inference. That is particularly important for a co rrect quantification of the uncertainties associated with predictions made when extrapolating to humans, for example. For this purpose, we used MCMC simulations in a Bayesian framework (Bernillon and Bois, 2000). We also considered dose-time-response data, which significantly complicates the problem. Very few off-the-shelf software provide adequate tools and models for such data, despite the fact that time is a key variable in qAOPs. Actually, while spatial structure is clearly apparent in AOP schemata (from molecules to cells, to tissues *etc.*), time is probably as important, but implicit: the time scale of molecular reactions is typically of the order of seconds, cells respond on a time scale of hours, tissues in a matter of days, and the whole body can take years to be significantly affected due to inbuilt redundancies in biology. This is particularly true for renal disease as humans have a large renal functional reserve and ill health is only apparent when the functional reserve is breached, but the time phenomenon is likely to be relevant for many, if not all, AOPs. This mix of time scales implies extrapolations in time from one KE to the next, which in the absence of obvious simplifying assumptions (steady-state *etc.*) requires the introduction of time and dose in the KERs.

The simplified AOP we used is not an OECD approved one, and we deliberately focused on a short sequence of KEs to demonstrate what can be achieved with different modeling approaches. The link to cell death and the subsequent link to kidney function impairment have not been included in our models given the absence of data on these downstream KEs.

Another important time-related consideration is obviously the kinetics of exposure to stressors. For QIVIVE, or in general for risk assessment, qAOPs can be linked with

pharmacokinetic models, but only if they are time-consistent. The predictions of a qAOP considering only dose, with the implicit assumption that processes are instantaneous or have fixed delays represents a simplification of realistic time-varying exposures. Kinetics of exposure should also be considered during the development of qAOPs, because *in vitro* cellular concentrations of test chemicals are usually different from the nominal exposure medium concentration and change with time (Fisher *et al.*, submitted). Nevertheless, in the absence of *in vitro* kinetic data on $KBrO_3$ concentrations, we considered the nominal $KBrO_3$ concentrations to be an adequate measure of (constant) exposure.

For the dose-response based qAOP, we used purely empirical models, *i.e.*, simple models that adequately "fit" the data. Given the probably infinite number of such models, we did not attempt to find the "best" model, so the question of model choice and uncertainty associated with it is certainly relevant. Thus, despite the good fits obtained, such models and the resulting qAOP should typically *not* be used outside the time and dose domains in which the data were gathered. In such an approach, the data were also taken at "face value". For example, the fact that a four-hour pre-incubation period of cells with carboxy-DCF led to a non-zero DCF fluorescence just after exposure to $KBrO_3$ was not taken into account, despite the fact that it provides information on the background rate of ROS formation. More importantly, the fact that medium was changed every day and that medium lactate concentrations were therefore zero immediately after that time was not modeled. It would have been difficult to empirically model the (more correct) dose-time response obtained with BNs (**Figure 17**) and we therefore limited the complexity of the empirical models. Furthermore, to obtain a correct statistical inference and at chemical-independent KERs, we resorted mathematical inversion of the $KBrO_3$ - time - response models fitted to the data. This was indispensable, pending direct observation of ROS - time - lactate (or DCF - time - lactate) data, for example. However, inversion poses constraints on the form and complexity of the KERs that can be used.

In summary, of the various qAOP approaches employed, the empirical qAOP was the simplest and fastest to obtain. It described the data rather well, from a naive "goodness of fit" point of view. The universal nature of these models in their Effectopedia implementation allows them to be reused, expanding on the idea of shared KE and KERs. However, a correct propagation of uncertainties along the chain of KERs, as done here, requires some mathematical and statistical sophistication (function inversion and Bayesian statistical inference), not provided by most software packages. Simply chaining dose-response relationships (that is, using the best predictions for one KE as input to the next KER, as it is often done) does not account for uncertainties in the "independent" variable at each step. In that case, uncertainty is not properly propagated through the AOP. The choice of models for KERs is arbitrary and does not offer mechanistic insight in the process. Moreover, their parameters do not have a biological interpretation (like the coefficients of a polynomial equation) and cannot be obtained by other means (*e.g.*, QSAR models, specific experiments, *etc.*). Accounting for model uncertainty would further complicate matters. Finally, the domain of application of empirical qAOPs is strictly limited to the data range and strongly depends on the relevance of the experimental protocol to the actual disease process. Their extrapolation is perilous.

The DBN qAOP we propose here is, to our knowledge, the first attempt to use such a model for a continuous dose-time-response predictive model. The work of Jaworska *et al.* (2010, 2013 and 2015) pioneered the application of BNs for qualitative (*i.e.*, hazard) assessment of chemicals and here we aim to extend this towards risk assessment with qAOPs. BNs are intermediate between empirical models (the KERs are usually simple linear links) and SB models (the whole set of KERs is modeled jointly and the links can represent cause-effect relationships). To accommodate the time variable of the data, we use in fact a special DBN – a straightforward extension of BNs – where time enters the KERs. (D)BN modeling is in a way simpler than the empirical dose-response qAOP proposed above, because *i.* the same basic KER

formula is used for each link, and *ii*. they can handle uncertainty in a flexible, unified, and statistically homogeneous framework. With this model, we obtained a fairly good representation of the data, and successfully modeled (*cf*. **Figure 17**) a fairly complex time-dose-relationship for the lactate readout. The end-results differed visually from those of the dose-response qAOP, because in our DBN the KER links for DCF and lactate are linearly related to GSH levels. We are currently working on nonlinear extensions of the DBN model. Finally, in a realistic risk assessment framework, pharmacokinetics *in vitro* or *in vivo* should be accounted for. This would add its own set of additional complexities, but it is possible to couple them with DBN models, either by pre-computing the value of the dose nodes in the DBN with a pharmacokinetic model, or by extending the DBN to simulate the pharmacokinetic data available.

Overall, (D)BN qAOPs offer an automatic or standardized way to develop semi-empirical qAOPs, while tuning the complexity of the KERs. They can nicely describe complex time dependencies. However, the software for parameterizing such models (*e.g.*, *GNU MCSim*, or *Stan*) require a mastering of their syntax for model building and fitting. The largest constraint for (D)BNs concerns the design of the experiments needed to develop the qAOP. The same doses and observation times should be used as much as possible. Otherwise, statistical imputation has to be used *a posteriori* to obtain a uniform dose and time schedules across experiments, and the statistical estimation problem is likely to become overwhelming. From an experimental point of view, however, it might not be feasible to observe the different KEs with the same time frame. Some events might happen in seconds (binding), days (cellular responses) or months or more (organ responses). This is because some events happen in seconds (binding), and others in days (cellular responses) or months or years (organ responses). In such cases, it might be possible to simplify time dependencies by separating time scales, *i.e.*, by considering some effects to be instantaneous in comparison to others.

The SB model we developed addresses only part of the CKD AOP, but probably the most important one: the link between GSH, oxidative stress and the formation of fluorescent DCF. The model describes in detail the sensing and control of oxidative stress by the Nrf2 pathway. It is quite complex, with 57 differential equations and 335 parameters. However, since it has been already parameterized for RPTEC/TERT1 cells, only the five parameters specific of the $KBrO_3$ and DCF reactions were calibrated with the data. We essentially found that a reasonable fit could be obtained if $KBrO_3$ acts directly on DCF, and that DCF bleaches significantly with time. We also found that modeling the pre-incubation period gives important information about the cellular background rate of oxidative stress. Such informative modeling is easy to do with a mechanistic model. The non-linearity of the effect of $KBrO_3$ is not well explained by a first-order reaction, but we did not want to introduce *ad hoc* equations or further hypotheses, because the mismatch already allows to arrive at the following point of discussion: According to our SB model, neither action on extra-cellular nor on intra-cellular GSH can explain the DCF data. This questions the naive application of the GSH readout as a measure of $KBrO_3$ effect in this AOP. While it is well accepted that thiol depletion can induce oxidative stress, the model suggests that this may not be the main mechanism of action of $KbrO_3$ in the readout test. Thus $KBrO_3$ may *not* be well suited to quantify our AOP, which also calls into question the results obtained with the other two models. However, we cannot exclude that the SB model is misleading us, because the parameter may not have been calibrated perfectly, and we cannot assess the overall uncertainty in the predictions of that model.

In terms of pros and cons, SB models have a huge advantage: They force us to think mechanistically about the data, asking which biochemical reactions could explain them. With some statistical sophistication, this allows us to formally check whether the data are compatible with our hypotheses. Aspects like time, dose, and spatial organization (at the organelle, cell, or tissue level) can be seamlessly integrated through the use of differential equations. SB models

can also simulate particular details of the experimental protocols and background cellular processes, and that improves our understanding of the biology and of the tests themselves. They can also naturally integrate pharmacokinetic models, since they are built from the same principles and same mathematical objects. However, those models are complex to develop. They demand specialized software for computation, and many data for parameter estimation. In fact, the amount of data required is very large, so that SB models may never be completely validated, leaving some uncertainty about the correctness of their predictions. Therefore, such complicated SB models could be seen as investment for the future rather than a quick answer to urgent questions.

4.5 CONCLUSION

The three approaches tested have different advantages. Dose-response based qAOPs may seem the easiest to develop at first sight, but they have very limited extrapolation and explanation power. BNs are in fact easier to develop, once the technology is mastered, but they impose either strong constraints on e xperimental design (fixed dosing and observation schedules) or require complex statistical treatment (imputation). SB models are more complex to develop, but one can strive for parsimony, as when we simplified the gene regulation part of our model. Importantly, they offer insight in the data collection and biology that the other approaches cannot afford. In any case, the three approaches we presented can all fully propagate uncertainty about qAOP predictions, which is essential for proper risk assessment. The contrasted results we obtained demonstrate that the choice of approach is not neutral. They also emphasize the importance of data collection:

- On *in vitro* kinetics, to understand and take into account the fate of the chemicals in the test system;
- On the baseline behavior of the cells, in the absence of chemical exposure. To this purpose, the experimental raw data be delivered to the modelers without pre-processing such as the normalization to background values. For example, if such normalization had been applied to our DCF data we would have lost important information on the background ROS production. Correcting for background erases a large part of the essential mechanistic understanding of an AOP. AOPs are as much about the underlying biology than about the effects of stressors;
- From different readouts, to select the most relevant one for the underlying KE or to better understand a complex KE (such as oxidative stress);
- On other chemicals to check whether the parameterized KERs are robust and really chemical-independent.

To avoid pitfalls in qAOP development, we propose to take at least 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. At the moment dose-response based modeling is the simplest, but we hope that we can contribute to a more wide-spread dissemination of DBNs in this area. In this spirit, one of the goals of the Effectopedia platform is to facilitate the creation of qAOPs by integrating and comparing the results brought by various modeling approaches.

5 INVESTIGATION OF NRF2, AHR AND ATF4 ACTIVATION IN TOXICOGENOMIC DATABASES

5.1 THE GENERAL APPROACH

Many transcriptionally activated pathways are intimately involved in responses to chemical induced perturbations and toxicological outcomes (Jennings *et al.*, 2013). These pathways may be independent, correlated and partially or fully overlapping.

To this end, we investigated the segregation of the genes belonging to the three following transcriptionally regulated pathways: the dioxin response or AhR pathway, the Nrf2 pathway that regulates the response to oxidative stress and the ATF4 branch of the unfolded protein response. While these pathways have specific non-overlapping activation mechanisms and specific non-overlapping DNA binding elements reviewed in (Jennings *et al.*, 2013), they also have overlapping downstream target genes. Adding to this complexity, converging toxicological mechanisms may lead to co-activation. Measuring their activation using transcriptomic approaches has great potential in increasing mechanistic understanding of chemical perturbations and to develop better prediction tools (Aschauer *et al.*, 2015; Limonciel *et al.*, 2015). In addition, such an approach could be used for biological read across. We precisely aim to investigate the dynamics of the interactions between these three pathways from toxicogenomic data in order to define the signature of each of them.

However, there is still a knowledge gap pertaining to the interplay between the Nrf2, AhR, and ATF4 pathways. It is known that several of their downstream targets have promotor sequences for more than one of these TF. For example, *NQO1* is driven by both AhR and Nrf2. Also, it is likely that the pathways may cooperate in redressing certain homeostatic perturbations. For example, we have shown that Nrf2 and ATF4 cooperate on the level of GSH, where ATF4 promotes the uptake of GSH amino acid building blocks including glutamine and

cysteine and promotes glutamate production via induction of asparagine synthetase. Nrf2 in turn through induction of GCL and GS produces new GSH (Wilmes *et al.*, 2013).

Very little is known about species differences, tissue specificity, chemical specificity, or other subtleties in the activation of these pathways. To investigate this further, we performed a transcriptomic analysis of large and medium size toxicogenomic datasets from the European Union's 6th and 7th framework projects carcinoGENOMICS (Vinken *et al.*, 2008) and Predict-IV (Mueller *et al.*, 2015), as well as from TG-GATEs (Igarashi *et al.*, 2015). Within these studies, we also identified some potentially useful specific TFs of the pathways investigated. $KBrO_3$ and Phorone have been used to experimentally activate Nrf2. $KBrO_3$ is an oxidizing agent causing ROS injury and oxidative stress induced DNA damage (Ballmaier and Epe, 1995; Limonciel *et al.*, 2012). In a recent study, Limonciel *et al.* (2018) showed that $KBrO_3$ activated the Nrf2 response without activation of the ATF4 response. Phorone can similarly activate Nrf2 due to GSH depletion (Iannone *et al.*, 1990; Oguro *et al.*, 1996; Younes *et al.*, 1986). Tunicamycin is a prototypical activator of the unfolded protein response (including the ATF4 branch) by causing an accumulation of misfolded glycoproteins in the ER (Oslowski and Urano, 2011). More specifically, Tunicamycin inhibits the N-glycosylation of newly formed proteins by the *DPAGT1* gene, leading to an interruption in glycoprotein production (Bassik and Kampmann, 2011). Benzo(a)pyrene and Omeprazole have been used to activate AhR. Benzo(a)pyrene is a polycyclic aromatic hydrocarbon and a prototypical AhR agonist (Nebert *et al.*, 2004). Omeprazole, a proton pump inhibitor (Howden, 1991) is also an AhR activator (Jin *et al.*, 2012).

The aim of the study was to investigate potential codependences of ATF4, Nrf2 and/or AhR, to develop a signature panel for each pathway and to develop a chemical activity scoring system, for chemical grouping. This study was recently (October 2018) published in the Frontiers in Genetics scientific journal (Zgheib *et al.*, 2018).

5.2 MATERIAL AND METHODS

The toxicogenomic datasets from the three projects (carcinoGENOMICS, Predict-IV and TG-GATEs) that were obtained in various experimental conditions (in human and rat *in vitro* liver and kidney models and rat *in vivo*, with bolus administration and with repeated doses), were combined and consolidated where overlaps between datasets existed. A bioinformatic analysis was performed to refine pathways' signatures and to create chemical activation capacity scores to classify chemicals by their potency and selectivity of activation of each pathway. With some refinement, such an approach may improve chemical safety classification and allow biological read across on a pathway level.

5.2.1 Generation of Target Gene Lists

For each of the three TF of interest (AhR, Nrf2, and ATF4), the following three search strategies, from the works of (Limonciel *et al.*, 2015), were applied in PubMed to retrieve TF target genes: (i) search for TF name and Chromatin Immunoprecipitation (ChIP)-sequencing, or ChIP-microarray studies, (ii) search for TF name and TF-specific response element and 'Electrophilic Mobility Shift Assay' or ChIP studies, and (iii) search for TF name and TF-specific DNA response element and name of a target gene known. In the first tier of this strategy, high-throughput sequencing datasets were retrieved, which provided extensive lists of genes shown to have the TF bind in their promoter region. In the second tier, lower throughput investigations were included, providing target genes that were more deeply investigated in the article with proven TF binding of the promoter region. These first two tiers provided an unbiased source of target genes that was completed in the third tier with manually added target genes for which at least one study showed binding of the TF in their promoter region.

PubMed searches were performed on 24.11.2014 for Nrf2 and 17.12.2014 for ATF4 and AhR. Gene lists are reported in **Table S8** in 'Supplementary Material 7.3' and are illustrated in **Figure 19**.

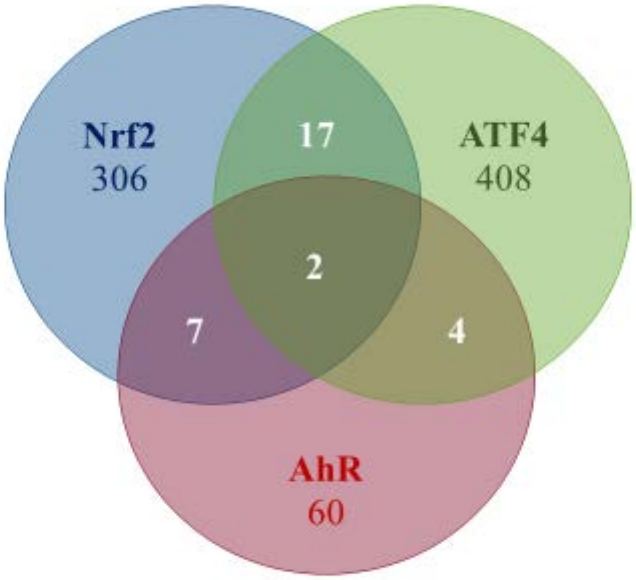


Figure 19. Venn diagram of the number of genes of each of the three studied pathways (AhR, Nrf2 and ATF4) and their overlapping zones, included in the analysis.

5.2.2 Construction of a Chemical-Effects Transcriptomics Database

As stated before, the database of chemical-induced transcriptomic changes comes from three projects: carcinoGENOMICS (Vinken *et al.*, 2008), Predict-IV (Mueller *et al.*, 2015) and TG-GATEs (Igarashi *et al.*, 2015). In carcinoGENOMICS, human and rat kidney cells were exposed to bolus concentrations of up to 31 chemicals in *in vitro* settings for up to 72 hours. In Predict-IV, human kidney cells and liver cells from human and rat were exposed daily *in vitro* for up to 14 days to up to 22 chemicals. Up to 171 chemicals from TG-GATEs were tested in various rat *in vivo* and *in vitro* systems, with various treating regimes. **Table 8** summarizes this and shows the 211 chemicals tested and dispatched in different categories of one or more of the three projects. **Table S9** in 'Supplementary Material 7.3' presents the exhaustive lists of chemicals by category.

Table 8. Number of chemicals used in each experimental category.

Project	Species	Tissue	Setting	Mode	Time-points	Number of chemicals	Notes
All dataset [211]*							(1-2)
Carcino-GENOMICS [31]	Human	Kidney	*in vitro*	Bolus	6h, 24h, 72h	30	(3-4)
	Rat	Kidney	*in vitro*	Bolus	6h, 24h, 72h	15	
PREDICT-IV [22]	Human	Kidney	*in vitro*	Repeated doses	1d, 3d, 14d	12	(5-6)
	Human and Rat	Liver	*in vitro*	Repeated doses	1d, 3d, 14d	11	(7)
TG-GATEs [171]	Human	Liver	*in vitro*	Bolus	2h, 8h, 24h	160	(8)
	Rat	Liver	*in vitro*	Bolus	2h, 8h, 24h	145	(9)
		Liver	*in vivo*	Bolus	3h, 6h, 9h, 24h	158	(10-11)
		Liver	*in vivo*	Repeated doses	4d, 8d, 15d, 29d	143	-
		Kidney	*in vivo*	Bolus	3h, 6h, 9h, 24h	41	(12)
		Kidney	*in vivo*	Repeated doses	4d, 8d, 15d, 29d	41	

(1) Number of chemicals assayed in at least one of the three source projects.

(2) ***Cyclosporine A*** *is the only chemical that was used in the three projects.* ***Cyclosporine A*** *appears in every single experimental category and sub-category (except carcinoGENOMICS's Rat tests).*

(3) *In carcinoGENOMICS, all 15 chemicals tested on rat cells, except one (**Dimethylnitrosamine**), were also tested on human cells.*

(4) *Beside **Cyclosporine A**, and five of the chemicals that appear in TG-GATEs as well, all chemicals are specific to carcinoGENOMICS (**2-Nitrofluorene** and **N-nitrosomorpholine** (TG-GATEs "Human liver in vitro bolus" and "Rat liver in vivo bolus"); and **Diclofenac**, **Nifedipine** and **Tolbutamide** (all liver categories of TG-GATEs)).*

(5) *The 12 chemicals tested on kidney cells and the 11 tested on liver cells in PREDICT-IV are distinct; Only **Cyclosporine A** is presented in these two categories.*

(6) *Among the chemicals tested on kidney cells in PREDICT-IV, only **Cisplatin** appears elsewhere (in TG-GATEs rat tests).*

(7) *Among the chemicals tested on liver cells in PREDICT-IV, only **Acetaminophen** and **Valproic acid** appear in all TG-GATEs categories; **Amiodarone**, **Chlorpromazine**, **Fenofibrate**, **Ibuprofen** and **Metformin** were tested on liver cells of TG-GATEs, and **Rosiglitazone** as well (except in "Rat liver in vitro bolus").*

(8) *In TG-GATEs, five chemicals were tested on human cells only (**HGF**, **IL1beta**, **IL6**, **INFalpha**, **Nefazodone** and **TGFbeta1**) and six others on animal categories only (**Carboplatin**, **Cephalotin**, **Cisplatin**, **Gentamicin**, **TNFalpha** and **Trimethadione**).*

(9) *Five chemicals appear in liver in vitro bolus categories only (human and rat): **Alpidem**, **Buspirone**, **Clozapine**, **Nefazodone** and **Venlafaxine**.*

(10) ***3-Methylcholantrene**, **Bortezomib**, **Gefitinib**, **Imatinib** and **Puromycin** appear in the "Rat liver in vivo bolus" category exclusively.*

(11) ***2-Nitrofluorene**, **Aflatoxin B1**, **Dexamethasone**, **N-methyl-N-nitrosourea** and **TNF** are common to TG-GATEs' "Human" and "Rat liver in vivo bolus" categories and were not tested in other conditions.*

(12) *The 41 chemicals that are used for TG-GATEs kidney in vivo testing are the same for both modes (bolus and repeated doses) and are common for all other categories (exceptions: **Gentamicin**, **Carboplatin**, **Cephalotin**, **Cisplatin**, **Desmopressin acetate**, **Amphotricine B** and **Acetamide**).*

* *The number between brackets refers to the number of chemicals per project*

5.2.3 Data Sources

The carcinoGENOMICs and Predict-IV data are publicly accessible on the diXa database (diXa Data Warehouse) hosted by The European Bioinformatics Institute[24]. In carcinoGENOMICS, *in vitro* renal cell experiments were performed using the human cell lines RPTEC/TERT1 (human, telomerase transfected) and NRK-52E (rat). The study no. is DIXA-003. Differentiated cell cultures were exposed to a single bolus of non or low cytotoxic (<IC10) concentration of chemical for 6, 24, or 72 hours before lysis in TRIZOL, RNA purification and transcriptomic analysis on Affymetrix microarrays as described (Limonciel *et al.*, 2012). Affymetrix Human Genome U133 Plus 2.0 GeneChIP arrays were used for human samples and Rat Genome 230 2.0 GeneChIP for rat samples. Normalization quality controls, including scaling factors, average intensities, present calls, background intensities, noise and raw Q-values were within acceptable limits for all chips. Hybridization controls were identified on all chips and yielded the expected increases in intensities. All subsequent analyses were based on normalized expression values generated using the MAS5 normalization algorithm. It is noted that RMA or GCRMA normalization would have been preferred. Normalized data was imported into GeneSpring (Agilent) to identify log2 fold change (FC) values for selected genes.

Within PREDICT-IV, *in vitro* testing of nephrotoxic and hepatotoxic compounds were performed on R PTEC/TERT1 cells (renal model), primary human hepatocytes, and rat hepatocytes (PHH and PRH, respectively). The study no. on the diXa database is DIXA-095. Differentiated cell cultures were exposed daily to a high (≤10% cell death) or low concentration of chemical for 1, 3 or 14 da ys, as described (Aschauer *et al.*, 2015; Crean *et al.*, 2015; Limonciel *et al.*, 2015; Wilmes *et al.*, 2013, 2014). Transcriptomic analysis was carried out on Illumina® HT 12 v4 B eadChip arrays for kidney and PHH human samples, except RPTEC/TERT1 exposed to CsA (HT 12 v3 chips). PRH samples were analyzed with Illumina®

[24] http://wwwdev.ebi.ac.uk/fg/dixa/index.html [Accessed October 24th, 2018]

RatRef-12 v1 BeadChIP arrays. Results were normalized by quantile normalization and expressed as log2 fold over time-matched control. Where several probes existed for a given gene, the probe with the highest variation across the dataset was selected.

The TG-GATEs datasets comprised *in vivo* rat data from liver and kidney tissue, as well as data from *in vitro* primary rat and human hepatocyte cultures, after a single administration of chemical and repeat dosing **Table 8**. CEL files were downloaded from the Open TG-GATEs[25] database of the Toxicogenomics Project and Toxicogenomics Informatics Project under CC Attribution-Share Alike 2.1 Japan. Probe annotation for the primary human hepatocyte data was performed using the hthgu133pluspmhsentrezg.db package version 17.1.0 and probe mapping was performed with hthgu133pluspmhsentrezgcdf downloaded from NuGO. Probe annotation for the rat data was performed using the rat2302rnentrezg.db package version 19.0.0 and probe mapping was performed with the rat2302rnentrezgcdf package version 19.0.0 downloaded from NuGO (R/Bioconductor support libraries)[26]. These mappings summarize the corresponding probes to a single probe set per gene. Probe-wise background correction (Robust Multi-Array Average expression measure), between-array normalization within each treatment group (quantile normalization) and probe set summaries (median polish algorithm) were calculated with the RMA function of the Affy package (Affy package, version 1.38.1) (Irizarry *et al.*, 2003). The normalized data were statistically analyzed for differential gene expression using a linear model with coefficients for each experimental group within a treatment group (Wolfinger *et al.*, 2001). A contrast analysis was applied to compare each exposure with the corresponding vehicle control. For hypothesis testing the moderated t-statistics by empirical Bayes moderation was used followed by an implementation of the multiple testing correction of Hochberg and Benjamini (1990) using the LIMMA package (Smyth *et al.*, 2005).

[25] https://dbarchive.biosciencedbc.jp/en/open-tggates/desc.html [Accessed October 24th, 2018]
[26] http://nmg-r.bioinformatics.nl/NuGO_R.html [Accessed October 24th, 2018]

All interspecies gene conversions where done using the provided human gene symbols which were converted to human or rat gene identifiers using the online conversion tool of bioDBnet (Biological Database Network)[27].

Altogether, the collected data concern 804 genes from the 857 genes identified in PubMed as targets of AhR, Nrf2 and ATF4. The 53 target genes that are not covered with data from any of the three projects were excluded from this study. These genes are listed in the last row of **Table S8** in 'Supplementary Material 7.3'.

5.2.4 Bioinformatics Methods

❖ *Data Selection*

The heterogeneity of the sources of information of our database widens its coverage and strengthens its capacity to represent multiple conditions. However, this richness makes the database's structure complex. To simplify the analysis without losing potentially important information, we focused on conditions providing the best background to study the three pathways individually. The effects observed following exposure to a chemical could vary greatly depending on exposure duration. Exposures lasting more than 24 hours tend to cause mixed stress responses that make it difficult to delineate the activation of specific molecular pathways and the initial mechanisms of toxicity of chemicals. These conditions could be a potential source of noise for the analysis and were thus excluded. Excluding all data obtained after 24 hours reduced the dataset from 7,042 to 4,685 testing conditions. We chose not to eliminate the early kidney *in vivo* time points (at 3 and 6 hours), even though they may be more reflective of background levels in case of slow absorption of the chemical administered.

[27] https://biodbnet-abcc.ncifcrf.gov/ [Accessed October 24th, 2018]

❖ *Pathway Specific Chemicals*

In order to distribute the genes to pathways and pathway overlapping zones, log2 genes FC were ranked in decreasing order and examined on reduced datasets containing conditions relative to pathway specific activators. We define a pathway specific activator as a chemical where the mode of action is known, that the mode of action activates the specific pathways and that this mode of action is not expected to activate the other pathways under investigation. Thus, at relatively short exposures, to relatively low concentrations these chemicals will only act on their specific target. It is however possible at higher concentrations or longer time exposure, other targets will be affected due to increasing toxicity. As shown in **Table 8**, some chemicals were not tested in all categories and tissue types. Thus, it was not possible to find pathway specific activators able to cover the entire database. **Table 9** shows the coverage of the datasets by the pathway specific activators selected as reference for analysis. Although none of the toxicogenomic databases analyzed here were designed to specifically address any of our three pathways of interest, most datasets included at least one chemical that could be considered as a specific pathway activator. Two specific chemicals were selected for AhR (Benzo(a)pyrene and Omeprazole) and Nrf2 ($KBrO_3$ and Phorone) and one for ATF4 (Tunicamycin). However, within 'Rat Kidney *in vivo*' category, no Nrf2 specific chemicals were found, and for all kidney data no ATF4 specific chemical were found either.

Table 9. Chosen pathway specific chemical through the dataset.

Pathway	Species	Kidney		Liver	
		in vitro	*in vivo*	*in vitro*	*in vivo*
AhR	Human	Benzo(a)pyrene		Omeprazole	
	Rat				
Nrf2	Human	$KBrO_3$		Phorone	
	Rat		-		
ATF4	Human	-		Tunicamycin	
	Rat				

❖ *Construction of Pathway Signatures*

For each of the pathway specific chemicals, all testing conditions were selected. For every gene, the mean of log2(FC) throughout all those conditions was calculated, to form the average activation value of each gene by each of the pathway specific activator. For AhR and Nrf2, the two average activation values obtained (one for each of the pathway specific activator) were themselves averaged. Genes were then sorted in decreasing order of average activation values per pathway. It is important to note that, since the expression of some genes can be inhibited (down regulated) by some chemicals or in certain conditions, some of the average activation values were negative. In order to select the most sensitive genes for each pathway, we computed the mean (μ) and the standard deviation (σ) of the genes' average activation values in each list. A pathways signature was formed by the genes whose average activation values were greater than $\mu + 2\sigma$ or smaller than $\mu - 2\sigma$ for this pathway. Genes appearing in the signature of more than one pathway were set apart in "overlapping signatures."

Furthermore, we stratified signatures by original databases' categories ('Rat liver cells *in vitro*', 'Rat liver cells *in vivo*', 'Human liver cells *in vitro*' *etc.*) (which correspond to primary cells), to check if there would be any species-specific or *in vitro*/*in vivo* differences among signatures. We chose to work only with liver data since more data were available for liver (602 conditions in kidney *vs.* 4,083 tested in liver, see **Table 10**).

Table 10. Number of conditions (chemicals, concentrations, time-points) tested per category.

Species	**Kidney**		**Liver**		**TOTAL**
	in vitro	*in vivo*	*in vitro*	*in vivo*	
Human	85	0	963	0	1048
Rat	30	487	1282	1838	3637
TOTAL	602		4083		4685

Following the same procedure as above, we constructed pathway signatures for AhR, Nrf2, and ATF4 in each of the following liver categories: (a) 'Rat liver cells *in vitro*', (b) 'Rat liver cells *in vivo*', and (c) 'Human liver cells *in vitro*'.

In all cases, general or stratified, some genes were excluded for having no data on effect of the chosen pathway specific chemicals. A list of those genes appears in **Table S10** in 'Supplementary Material 7.3'. A summary of the above-described protocols and the following procedures of Methods are presented in the workflow of Methods summarizing workflow **Figure 20**.

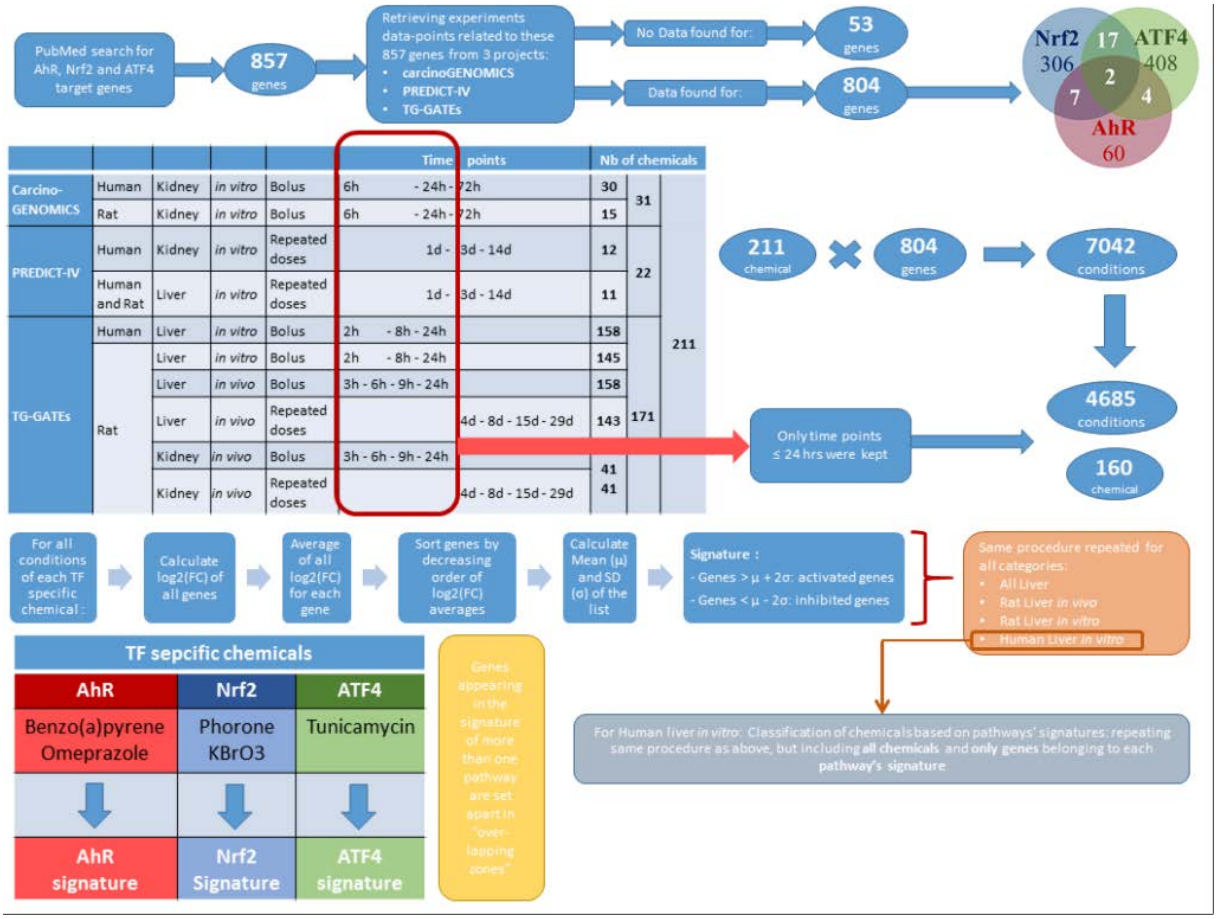


Figure 20. Methods summarizing workflow

5.2.5 Pathway's Signature-Based Prioritization of Chemicals

Among the three liver categories where signatures were stratified, we chose to focus on the 'Human liver cells *in vitro*' sub-category exclusively since the ultimate goal of our toxicity pathways' analyses and models is risk assessment of human cells' exposure to xenobiotics. We considered only the genes belonging to the signature of each of the three pathways, but not their overlapping zones. This selection of experimental category and genes reduces the number of studied chemicals from 211 to 160 for the lack of data on the rest of chemicals in this section. Then, for each of the 160 chemicals investigated, we averaged log2(FC) of the pathway signature genes over experimental conditions. Therefore, for each of the three pathways, we obtained a 'chemical activation capacity' (CAC) value per chemical. This value reflects how strongly a chemical can activate a given toxicity pathway. Those CAC can be negative for chemicals inhibiting the majority of the genes of a pathway. We used CAC to estimate the pathway's selectivity of chemicals as well as the importance of their impact. Each chemical can be considered as a point having three CAC as coordinates in a 3-dimensional space which axes correspond to a given pathway. Let us consider a chemical *K* that has a point in a bi-dimensional graph where the X-axis corresponds to AhR and the Y-axis to Nrf2. In this graph, *K*'s coordinates would be: ($CAC_{AhR,\,K}$, $CAC_{Nrf2,\,K}$), see **Figure 21**. *K* also defines the vector $\overrightarrow{\boldsymbol{OK}}$ linking the origin *O* (0, 0) to the point *K*.

The specificity of a chemical for a given pathway can be measured by the proximity of its point *K* to the axis representing that pathway. Proximity can be mathematically evaluated by the absolute value of the cosine of the angle (α) between the pathway's axis and $\overrightarrow{\boldsymbol{OK}}$. The more *K* is specific to AhR, the closer it is to the AhR's axis, the smaller α is, and the bigger $\mathbf{cos}\ (\boldsymbol{\alpha})$. In theory, in a 3-dimensional space, a point is closer to an axis than to the two others when its $\mathbf{cos}\ (\boldsymbol{\alpha})$ with this axis is greater than $\frac{\mathbf{1}}{\sqrt{\mathbf{3}}}$. Thus, the value of 0.57735 ($\frac{\mathbf{1}}{\sqrt{\mathbf{3}}}$) was chosen as a cut-off point for $\mathbf{cos}\ (\boldsymbol{\alpha})$. On the other hand, the activation potency of a chemical proportionally

increases with the module of the vector $\overrightarrow{\boldsymbol{OK}}$ vector noted $\|\overrightarrow{\boldsymbol{OK}}\|$ (the distance between the origin and the chemical's point). The value of 0.5 was chosen as a cut-off point for $\|\overrightarrow{\boldsymbol{OK}}\|$. For instance, chemicals *A* and *B* in **Figure 21** are both quite specific of Nrf2, but *A*'s activation potency is relatively limited compared to *B*'s ($\|\overrightarrow{\boldsymbol{OA}}\| < \|\overrightarrow{\boldsymbol{OB}}\|$).

Similarly, even though *C* seems to have a greater activation potency than *A* and *B* (greater module), it is equidistant to both axes and therefore is not specific of any of the two pathways. The same logic applies for a 3-dimensional space, adding one extra axis for the ATF4 pathway.

In our signature-based classification of chemicals, for each pathway, after applying the chosen cut-off points, we sorted chemicals by the result of the product $\mathbf{cos}\ (\boldsymbol{\alpha}) \times \|\overrightarrow{\boldsymbol{OK}}\|$. Thus, chemicals which are both pathway specific (high **cos (α)**) and potent (high $\|\overrightarrow{\boldsymbol{OK}}\|$) show up first in our lists.

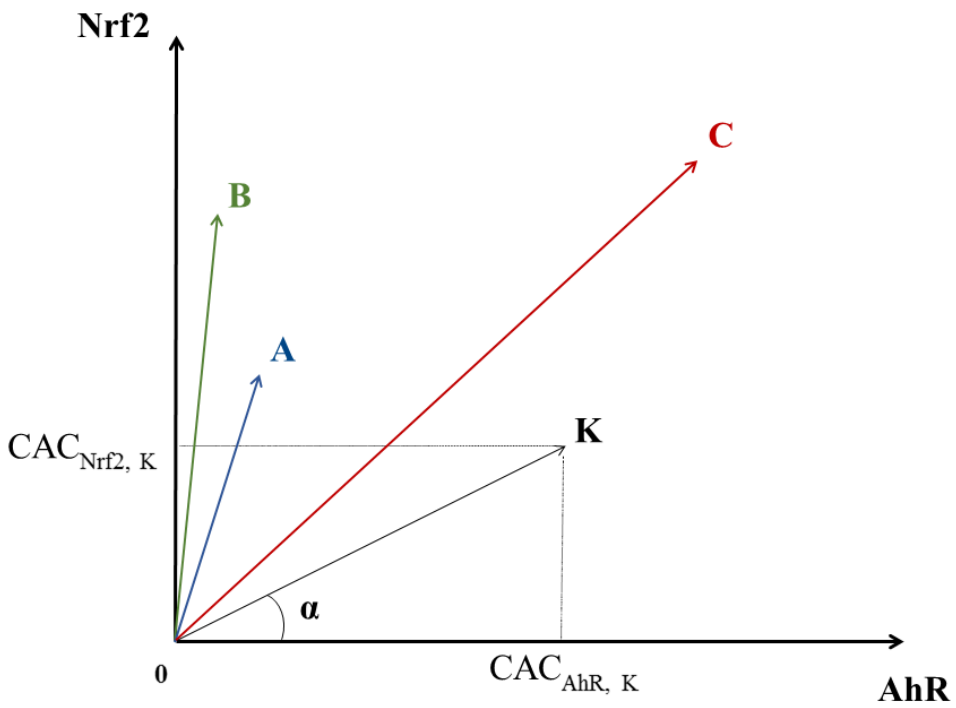


Figure 21. Geometric representation of chemical specificity and potency for the Nrf2 and AhR pathways. *K* represents a chemical and its coordinates are ($CAC_{AhR, K}$, $CAC_{Nrf2, K}$). *K* also defines the vector $\overrightarrow{OK}$ linking the origin *O* (0, 0) to point *K*. The absolute value of the cosine of the angle *α* between $\overrightarrow{OK}$ and a pathway's axis can be used to measure the specificity of a chemical for the given pathway (the smaller *α*, the more specific the chemical). On the other hand, the overall activation potency of a chemical increases proportionally with the length of $\overrightarrow{OK}$. Points *A*, *B* and *C* represent three other chemicals with different specificities and potencies for pathways' activation (see text).

5.3 RESULTS

A visual depiction of the workflow is provided in **Figure 20**.

5.3.1 Pathways' Global Signatures

Pathway's signatures defined on the basis of the whole data set are listed in **Table 11**. Each signature has two parts: 'Activated genes' (those having positive log2(FC) averages and are greater than $\mu + 2\sigma$) and 'Inhibited genes' (those having negative log2(FC) averages and are smaller than $\mu - 2\sigma$); The two parts are merged in one in the overlapping signatures. In all lists, genes are sorted by the decreasing absolute value of the genes' log2(FC) averages. The number of genes in the obtained pathway's signature was 24 for AhR, 27 for Nrf2 and 30 for ATF4. In each pathway, at least half (12 for AhR, 15 for Nrf2 and 19 for ATF4) were 'Activated genes'. The *a priori* pathway is the one for which the gene has come up in PubMed searches; **Table 11** shows that most of activated genes were *a priori* suspected to belong to the target pathway (for example: *CYP1A1*, *RUNX2*, and *CYP1A2* were known to be activated by AhR, *HMOX1* and *SRXN1* by Nrf2 and *DDIT3* and *HERPUD1* by ATF4; those genes are highlighted in gray) while this wasn't the case of the 'Inhibited genes' part of the lists. **Figure 22** shows the overlapping zones. Among the five genes that are in the AhR-Nrf2 overlapping zone (*NQO1*, *DLGAP5*, *CFTR*, *RAB39B* and *GSTA1*), only *NQO1* is a mainly activated gene while this was the case of most seven genes of the Nrf2-ATF4 overlapping zone (*ATF3*, *SLC7A11*, *TRIB3*, *CABC1*, *GDF15*) with two exceptions (*CCL2* has negative averages for both pathways and *KCNT2* for Nrf2). *CYP1B1* is the only mutual gene for AhR (strong activation) and ATF4 (inhibition) and TPX2 is the only mutual gene for all three pathways (inhibition). **Figure 23** shows a network representation of the three signatures and their overlapping zones.

Table 11. Pathway's global signatures for AhR, Nrf2 and ATF4 pathways and the signatures of their overlapping zones for all available data. Gray background indicates genes that appear in the signature of the pathway from previous studies (**Table S11**) and confirmed here. Non-grayed out values are novel allocations from this analysis.

	AhR Signature			Nrf2 Signature			ATF4 Signature		
	Genes (1)	**log_2(FC) averages (2)**	***A priori* pathway (3)**	**Genes**	**log_2(FC) averages**	***A priori* pathway**	**Genes**	**log_2(FC) averages**	***A priori* pathway**
Activated genes	*CYP1A1*	4.35	AhR	*HMOX1*	1.12	Nrf2	*DDIT3*	1.59	ATF4
	DLL1	1.36	AhR	*SRXN1*	0.97	ATF4 Nrf2	*TSLP*	1.51	ATF4
	RUNX2	1.03	AhR	*MAFF*	0.78	AhR Nrf2	*AKNA*	1.30	ATF4
	SLC16A9	0.92	Nrf2	*OSGIN1*	0.67	Nrf2	*HERPUD1*	1.23	ATF4
	FAM65C	0.79	AhR	*DUSP5*	0.66	ATF4	*SLC1A4*	1.15	ATF4
	FLRT1	0.78	ATF4	*TXNRD1*	0.63	ATF4	*IL23A*	1.05	ATF4
	FIBIN	0.77	ATF4	*GCLC*	0.60	ATF4	*CHAC1*	0.99	ATF4
	TIPARP	0.73	AhR	*PPP1R15A*	0.57	ATF4	*FGF21*	0.95	ATF4
	CYP1A2	0.69	AhR	*GCLM*	0.57	Nrf2	*HSPA5*	0.94	ATF4
	ASB3	0.67	Nrf2	*HSPA1B*	0.56	Nrf2	*NUPR1*	0.94	ATF4
	PDE1A	0.66	ATF4	*FBXO30*	0.55	ATF4	*GTPBP2*	0.91	ATF4
	PBX1	0.64	Nrf2	*GSTP1*	0.53	Nrf2	*PDIA4*	0.87	Nrf2
				PHGDH	0.46	Nrf2	*FAM129A*	0.87	ATF4
				TMEFF2	0.46	ATF4	*LONP1*	0.80	ATF4
				RUNX3	0.46	Nrf2	*VNN3*	0.78	ATF4
							SESN2	0.75	ATF4
							MTHFD2	0.73	ATF4
							PYCR1	0.72	ATF4
							BACH1	0.68	Nrf2
Inhibited genes	*SLC1A7*	-1.57	ATF4	*TMEM189*	-1.48	ATF4	*COCH*	-1.25	Nrf2
	PSG5	-1.43	AhR	*NREP*	-0.99	ATF4	*SNAI2*	-1.20	ATF4
	PRKAR2B	-1.23	Nrf2	*KIFC1*	-0.79	ATF4	*INSIG1*	-1.02	Nrf2
	SOAT2	-0.80	ATF4	*DLX2*	-0.78	Nrf2	*AKR1B10*	-0.96	Nrf2
	DAAM2	-0.78	Nrf2	*BMF*	-0.73	ATF4	*PMAIP1*	-0.88	Nrf2
	WDR63	-0.70	AhR	*TGFB2*	-0.72	ATF4	*ANGPTL4*	-0.87	ATF4
	FAM69A	-0.68	Nrf2	*DDC*	-0.71	Nrf2	*SNRNP35*	-0.77	ATF4
	CDH11	-0.67	Nrf2	*GLI2*	-0.71	ATF4	*SERPINE1*	-0.68	Nrf2
	LCN2	-0.66	ATF4	*AURKB*	-0.69	ATF4	*PRC1*	-0.65	Nrf2
	PLA2G4A	-0.66	Nrf2	*NEDD9*	-0.67	ATF4	*LMCD1*	-0.64	AhR
	CXCL5	-0.64	Nrf2	*TFPI*	-0.65	ATF4	*LBH*	-0.61	Nrf2
	WISP1	-0.62	ATF4	*OSMR*	-0.59	Nrf2			

	AhR-Nrf2 Overlapping signature			Nrf2-ATF4 Overlapping signature		
	Genes	**AhR log_2(FC) averages**	**Nrf2 log_2(FC) averages**	**Genes**	**Nrf2 log_2(FC) averages**	**ATF4 Log2 FC average**
Activated or Inhibited genes	*NQO1*	0.7	0.83	*ATF3*	0.73	0.90
	DLGAP5	-0.64	-0.56	*SLC7A11*	0.70	0.69
	CFTR	-0.69	-0.73	*TRIB3*	0.70	1.02
	RAB39B	-0.92	-0.52	*CABC1*	0.56	2.90
	GSTA1	-1.43	-0.83	*GDF15*	0.48	0.80
				CCL2	-0.61	-1.28
				KCNT2	-0.9	0.76

	AhR-ATF4 Overlapping signature			AhR-Nrf2-ATF4 Overlapping signature			
Activated or Inhibited genes	**Genes**	**AhR log_2(FC) averages**	**ATF4 log_2(FC) averages**	**Genes**	**AhR Log2 FC average**	**Nrf2 log_2(FC) averages**	**ATF4 log_2(FC) averages**
	CYP1B1	3.56	-0.63	*TPX2*	-0.75	-0.8	-2.38

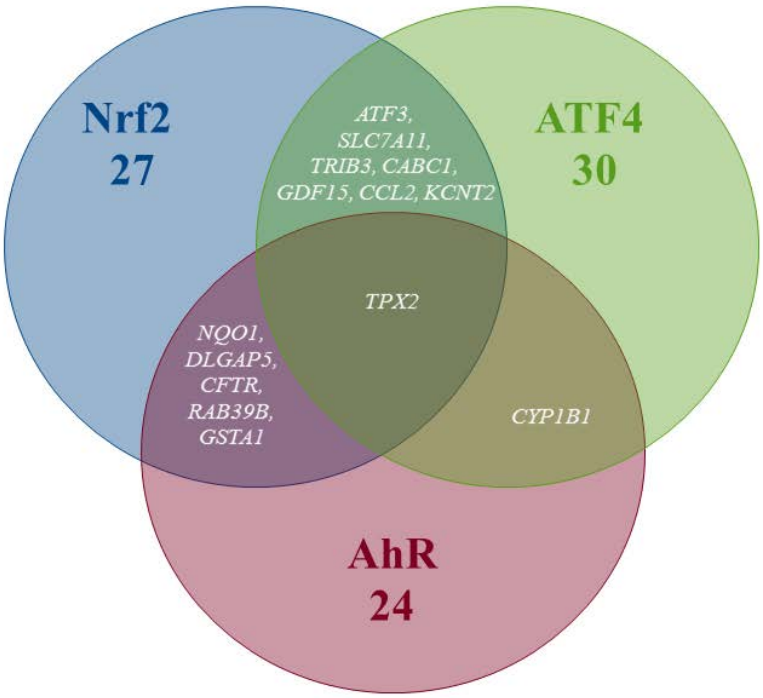


Figure 22. Venn diagram of the number of genes per pathway's global signatures and names of genes of overlapping zones.

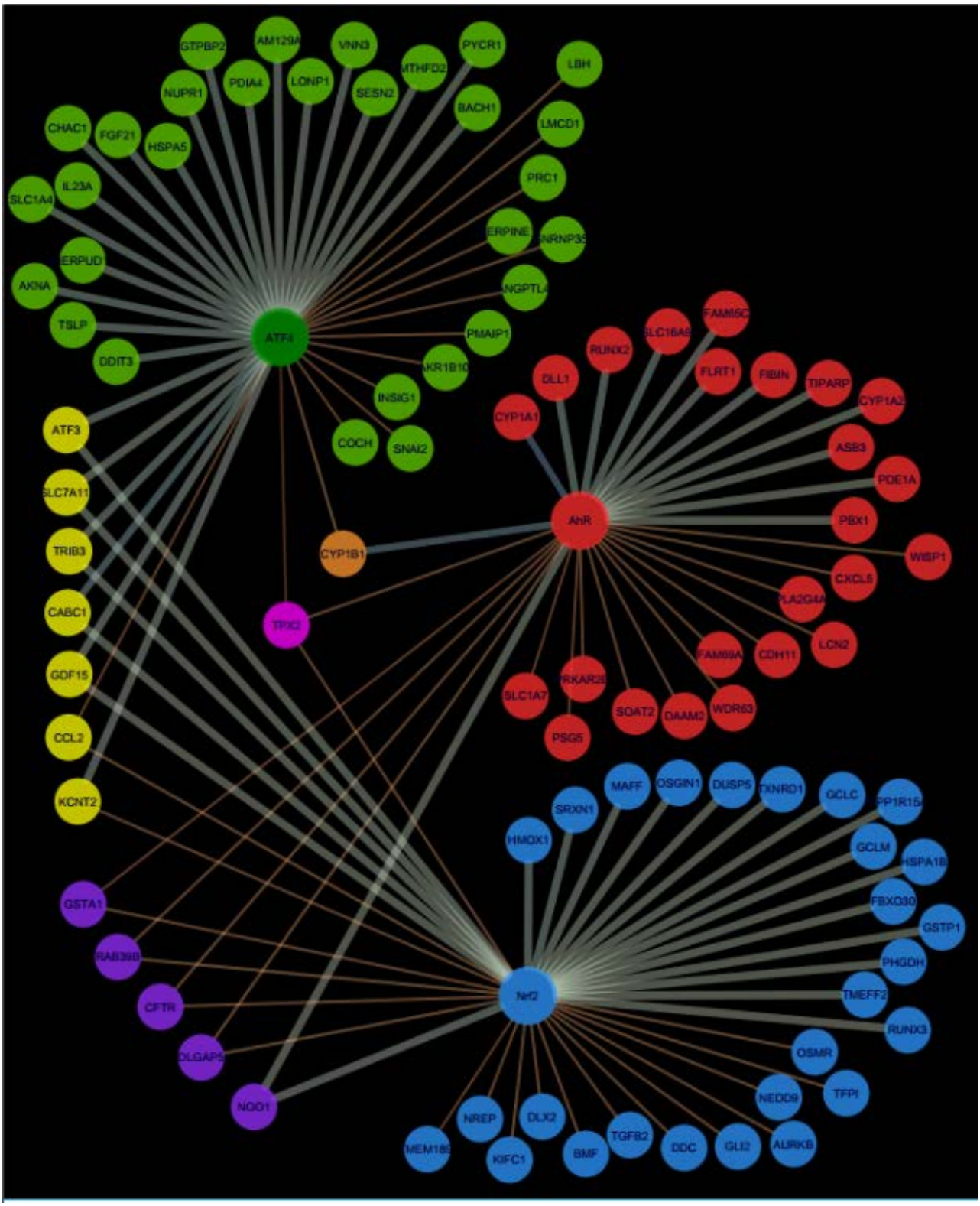


Figure 23. Network representation of AhR, Nrf2 and ATF4 pathway signatures and their overlapping zones.

5.3.2 Pathways' Stratified Signatures in Liver

❖ *The Three Main Pathways' Stratified Signatures in Liver*

Table 12 shows the stratified signatures in liver of each pathway in four columns (categories): each containing the genes' names and their log2(FC) averages. Genes that appear in more than one column are highlighted in gray and empty lines were left in order to display those genes on the same line in all the categories where they appear. Genes of the first column, sorted by the decreasing absolute values of their log2(FC) averages, appear first, followed by genes appearing in more than one category but not the first column and then the rest of the genes sorted by the decreasing absolute values of their log2(FC) averages as well.

❖ *AhR Stratified Signatures*

Table 12 shows that *CYP1A1* is clearly, by far the most activated gene in this pathway. Three other genes appear in the AhR signature in more than one column: *CYP1A2* everywhere except 'Rat liver *in vitro*', *TIPARP* everywhere except 'Rat liver *in vivo*' and *ABCC4* shows up in these two categories only. 'Rat liver *in vitro*' AhR signature is completed by five additional genes, 'Rat liver *in vivo*' by one more and 'Human liver *in vitro*' by three.

❖ *Nrf2 Stratified Signatures*

Nrf2 signatures are bigger: 22 genes in the all liver data signature, 28 for 'Rat Liver *in vitro*' and 15 for each of 'Rat Liver *in vivo*' and 'Human Liver *in vitro*'. Around two third of those genes are "Activated genes" and the rest have negative log2(FC) averages. *MAFF*, *SLC3A2*, *OSGIN2* are among the 'Activated genes' that appear in three out of the four categories we are studying. Other important genes show up i n two columns (*HSPA1B*, *PPP1R15A*, and *GCLC*) and some, in only one (*SRXN1* in 'Rat Liver *in vitro*' and *HMOX1* in 'Rat Liver *in vivo*'). The values of the 'Rat liver *in vivo*' are also higher than the 'Rat liver *in vitro*' and 'Human liver *in vitro*' categories.

Table 12. AhR, Nrf2 and ATF4 pathways' signatures stratified in liver data and by all liver data sub-categories ('Rat Liver *in vitro*' data, 'Rat Liver *in vivo*' data and 'Human Liver *in vitro*' data).

AhR signatures								
	All liver data		**Rat liver *in vitro***		**Rat liver *in vivo***		**Human liver *in vitro***	
	Genes	**Log_2(FC) averages**	Genes	**Log_2(FC) averages**	Genes	**Log_2(FC) averages**	Genes	**Log_2(FC) averages**
Activated genes	*CYP1A1*	4.55	*CYP1A1*	1.30	*CYP1A1*	6.86	*CYP1A1*	4.72
	CYP1A2	1.47			*CYP1A2*	1.71	*CYP1A2*	2.44
	TIPARP	0.64	*TIPARP*	0.40			*TIPARP*	1.21
			ABCC4	0.25	*ABCC4*	0.97		
			IL1R1	0.24	*HTATIP2*	1.19	*CYP1B1*	3.49
			TAF15	0.22			*SLC20A1*	0.78
Inhibited genes			*PRKAR2B*	-0.20			*KCNT2*	-0.60
			ANXA1	-0.18				
			ANGPTL4	-0.17				
Nrf2 signatures								
	All liver data		**Rat liver *in vitro***		**Rat liver *in vivo***		**Human liver *in vitro***	
	Genes	**Log_2(FC) averages**	Genes	**Log_2(FC) averages**	Genes	**Log_2(FC) averages**	Genes	**Log_2(FC) averages**
Activated genes	*MAFF*	1.42	*MAFF*	0.67	*MAFF*	2.37		
	FBXO30	0.92					*FBXO30*	0.35
	HSPA1B	0.82	*HSPA1B*	0.37			*HSPA1B*	0.63
	PPP1R15A	0.77			*PPP1R15A*	1.16		
	GSTP1	0.67			*GSTP1*	1.24		
	GCLC	0.66	*GCLC*	0.35				
	PSAT1	0.64			*PSAT1*	1.54		
	DUSP5	0.62	*DUSP5*	0.64				
	SLC3A2	0.60			*SLC3A2*	1.09	*SLC3A2*	0.40
	OSGIN1	0.58			*OSGIN1*	0.91	*OSGIN1*	0.42
	SLC6A9	0.57			*SLC6A9*	1.06		
	SLC20A1	0.52	*SLC20A1*	0.41				
	ABCC3	0.52			*ABCC3*	1.00		
			YPEL5	0.47			*YPEL5*	0.37
			CPT1A	0.38			*CPT1A*	0.36
	ASNS	0.75	*SRXN1*	0.66	*HMOX1*	2.03	*ATF5*	0.37
	PHGDH	0.55	*PHLDA1*	0.53	*SLC7A11*	1.74	*AP5Z1*	0.35
	PLA2G12A	0.50	*TXNRD1*	0.41	*GDF15*	1.30		
	SLC7A1	0.48	*ABCC2*	0.39	*BTG2*	0.89		
			PIR	0.34				
			FLVCR2	0.33				
			GSR	0.33				
			GABARAPL1	0.33				
			AGPAT9	0.57				
			TBCEL	0.48				
			MMD	0.33				
Inhibited genes							*MMD*	-0.4
	LCN2	-0.45			*LCN2*	-0.97		
			TGFB2	-0.34			*TGFB2*	-0.44
	MID1IP1	-0.48	*TNFAIP2*	-0.44	*BMF*	-0.88	*ALDH1A1*	-0.61

	IL33	-0.46	*VASN*	-0.39	*DHRS7*	-0.69	*DDC*	-0.42
	NREP	-0.45	*AURKB*	-0.38			*DUT*	-0.35
	SERPINB9	-0.42	*RAB32*	-0.36			*IFIT3*	-0.33
			CD36	-0.36			*UGT1A6*	-0.32
			DCN	-0.34				
			CTSC	-0.34				
			LBH	-0.32				
			CXCL3	-0.32				
ATF4 signatures								
	All liver data		**Rat liver *in vitro***		**Rat liver *in vivo***		**Human liver *in vitro***	
	Genes	**Log$_2$(FC) averages**	Genes	**Log$_2$(FC) averages**	Genes	**Log$_2$(FC) averages**	Genes	**Log$_2$(FC) averages**
Activated genes	*TSLP*	1.51					*TSLP*	1.51
	AKNA	1.30					*AKNA*	1.30
	HERPUD1	1.23	*HERPUD1*	1.28	*HERPUD1*	0.61	*HERPUD1*	2.39
	IL23A	1.05	*IL23A*	1.69			*IL23A*	1.86
	HSPA5	0.94					*HSPA5*	3.28
	GTPBP2	0.91	*GTPBP2*	1.12			*GTPBP2*	1.89
	PDIA4	0.87	*PDIA4*	0.92			*PDIA4*	2.18
	FAM129A	0.87					*FAM129A*	2.92
	PYCR1	0.72	*PYCR1*	0.91				
			CHAC1	1.40	*CHAC1*	0.50		
			KLF15	0.81	*KLF15*	0.43		
	SLC1A4	1.15	*TRIB3*	1.12	*HES1*	0.57	*FIBIN*	2.72
	NUPR1	0.94	*BCAT2*	0.97	*USP2*	0.55	*LCN2*	1.91
	LONP1	0.80	*ARHGEF2*	0.93	*ENC1*	0.48	*CTH*	1.62
	VNN3	0.78	*CASP4*	0.84	*TSC22D3*	0.44	*NFE2L1*	1.2
	SESN2	0.75	*KLF4*	0.82	*DDIT4*	0.39		
	BACH1	0.68	*BET1*	0.82	*SLC38A2*	0.38		
			WARS	0.80	*IP6K2*	0.62		
			PCK2	0.73				
			SLC25A33	0.71				
			SLC7A5	0.71				
			ACOT2	0.83				
			MANEA	0.75				
Inhibited genes	*PRC1*	-0.65	*PRC1*	-0.61				
	LMCD1	-0.64	*LMCD1*	-0.80			*LMCD1*	-1.73
	LBH	-0.61					*LBH*	-2.56
	SNAI2	-1.20	*DPYSL2*	-0.98	*FOXA2*	-0.61	*FRMD6*	-1.52
	AKR1B10	-0.96	*DUSP6*	-0.97	*ABCG2*	-0.49	*SLC39A10*	-1.35
	PMAIP1	-0.88	*IFIT3*	-0.72	*NEDD9*	-0.43	*GPNMB*	-1.26
	SNRNP35	-0.77	*EMILIN1*	-0.69	*TMEM159*	-0.37	*ANKRD1*	-1.16
	SERPINE1	-0.68	*FCER1G*	-0.65			*PHLDA1*	-1.16
			SQRDL	-0.61				
			IFI44	-0.61				

❖ *ATF4 Stratified Signatures*

ATF4 signatures size is similar to Nrf2's signatures with a comparable proportion of activated genes: 23 genes in the all liver data signature, 28 for 'Rat liver *in vitro*' and 14 for each of 'Rat liver *in vivo*' and 19 for 'Human liver *in vitro*'. *HERPUD1* is an important gene in this pathway; it is part of the signature of every single category we are examining and exhibits values as high as 2.39 in 'Human Liver *in vitro*' (among the highest in ATF4 signatures). Other genes also are present in the majority of the categories: *IL23A*, *GTPBP2*, and *PDIA4*. It is noteworthy that the ATF4 signature of 'Rat Liver *in vivo*' results don't have a lot in common with the other three categories and its log2(FC) averages are lower than the rest (the highest value is 0.61 for *HERPUD1*).

❖ *The Overlapping Zones Stratified Signatures*

Figure 24 shows that the AhR-ATF4 overlapping zone is the least populated (four genes maximum in all liver data, no genes for 'Rat Liver *in vivo*' and two genes in the two other categories). The number of genes in the AhR-Nrf2 overlapping signatures ranges from four to eight, with many typical key Nrf2 genes (*NQO1*, *SRXN1*, *HMOX1*, *TXNRD1*, and *GCLM*) appearing in more than one category. The Nrf2-ATF4 overlapping signatures contain six to eleven genes (*DDIT3*, *ATF3*, and *CHAC1* are among the repetitive genes). Finally, *TRIB3*, *FGF21*, *GDF15*, *SLC7A11*, and *TPX2* are in the signature of the zone mutual to all three pathways for at least two of the four categories studied.

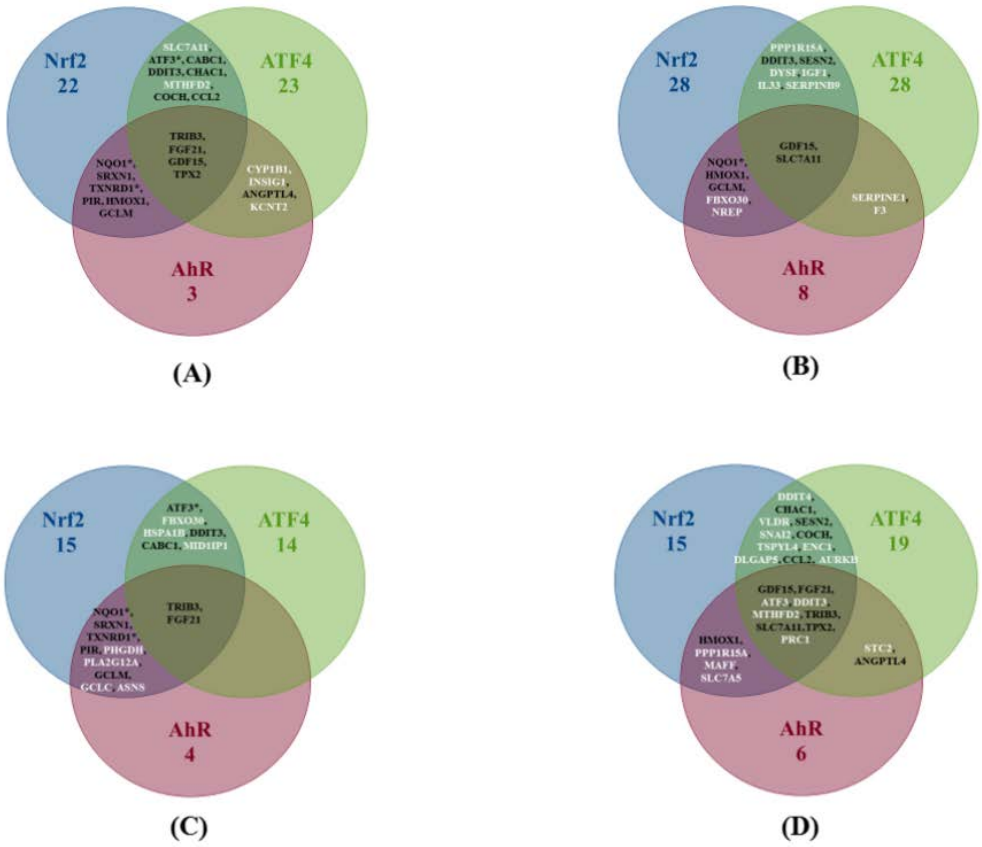


Figure 24. Venn diagrams of the number of genes per pathway's stratified signatures and names of genes of overlapping zones. Categories: **(A)** All liver data, **(B)** Rat Liver *in vitro* data, **(C)** Rat Liver *in vivo* data, **(D)** Human Liver *in vitro* data. *Refers to genes that were known to be part of the same overlapping zone according to **Table S11** lists. White is the color of gene names that appear in an overlapping zone of only one of the four categories studied, and black is the color of gene names that appear in more than one category (two, three or four).

5.3.3 Human Liver Category: Pathway's Signature-Based Prioritization of Chemicals

Figure 25, **Figure 26** and **Figure 27** plot the 160 chemicals' vector modules *vs.* the absolute value of **cos ($\boldsymbol{\alpha}$)**, which represents the pathway activation scores of chemicals that activate each pathway both selectively and strongly. Chemicals are represented by a number that corresponds to their rank in the alphabetically ordered list. The blue dashed lines mark the vertical ($\mathbf{cos}(\boldsymbol{\alpha}) = \frac{\mathbf{1}}{\sqrt{\mathbf{3}}}$) and horizontal ($\|\overrightarrow{\boldsymbol{OK}}\| = \mathbf{0.5}$) limits we set.

The number chemicals that are off these limits is 34 for AhR, one for Nrf2 and four for ATF4; these chemicals are in red and their names are listed in the legend on the right by the order of the decreased values of the product result $\mathbf{cos}(\boldsymbol{\alpha}) \times \|\overrightarrow{\boldsymbol{OK}}\|$. As we can see in these figures' legends, 'pathway specific activators' show up first in the lists of AhR (Omeprazole) and ATF4 (Tunicamycin), but do not appear at all in the list of Nrf2 (Phorone).

The annotation of chemicals for **Figure 25**, **Figure 26** and **Figure 27** is presented in **Table S11** in 'Supplementary Material 7.3'.

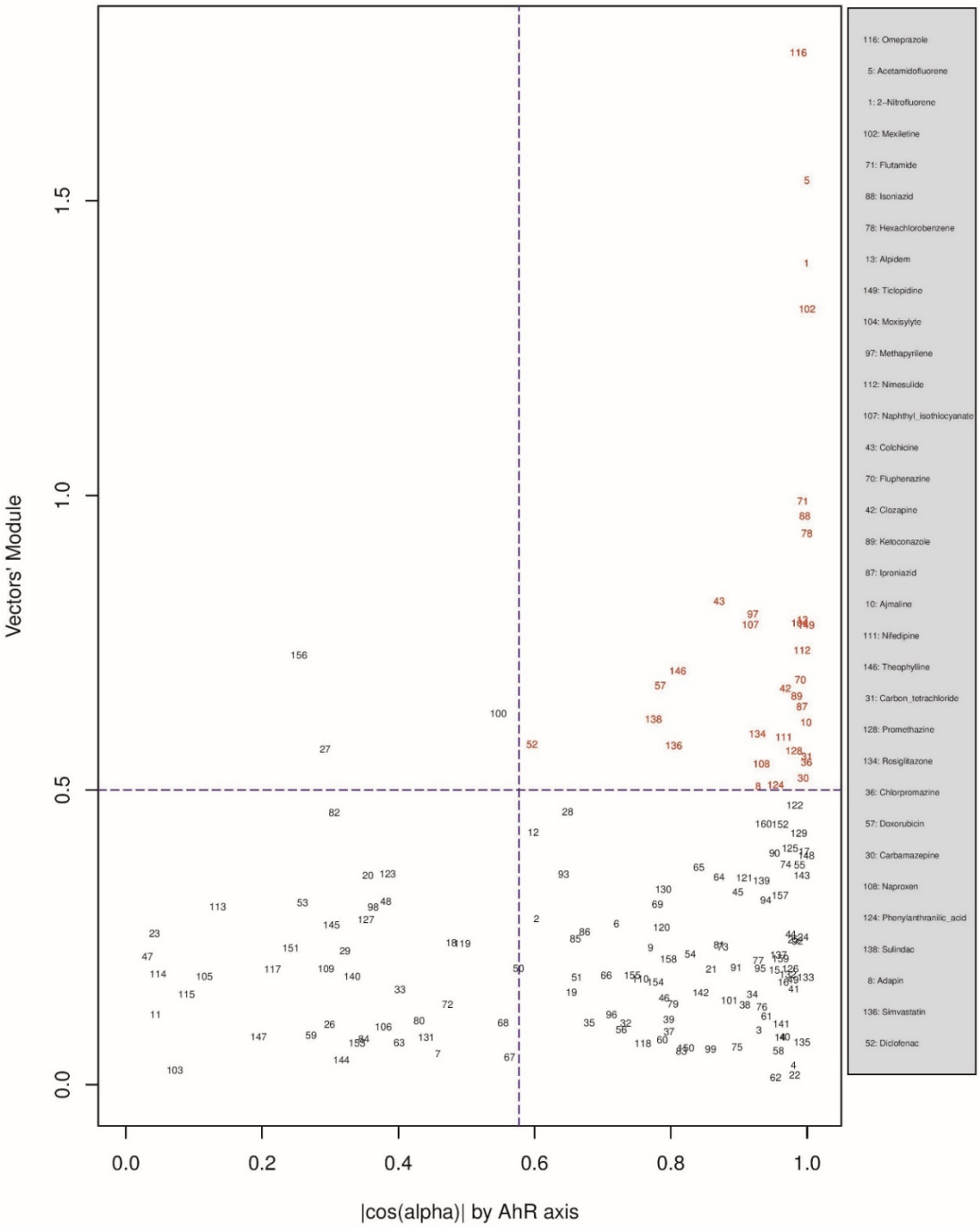


Figure 25. Distribution of chemicals by potency (Y-axis: module $\|\overrightarrow{OK}\|$ of the vector linking the origin O(0,0) to the chemical's point in a 3D space) and specificity to the AhR pathway (X-axis: the absolute value of the $|\boldsymbol{cos}(\boldsymbol{\alpha})|$ of the angle between $\overrightarrow{OK}$ and the AhR axis in a 3D space). Chemicals are represented by their rank in the alphabetically ordered list.

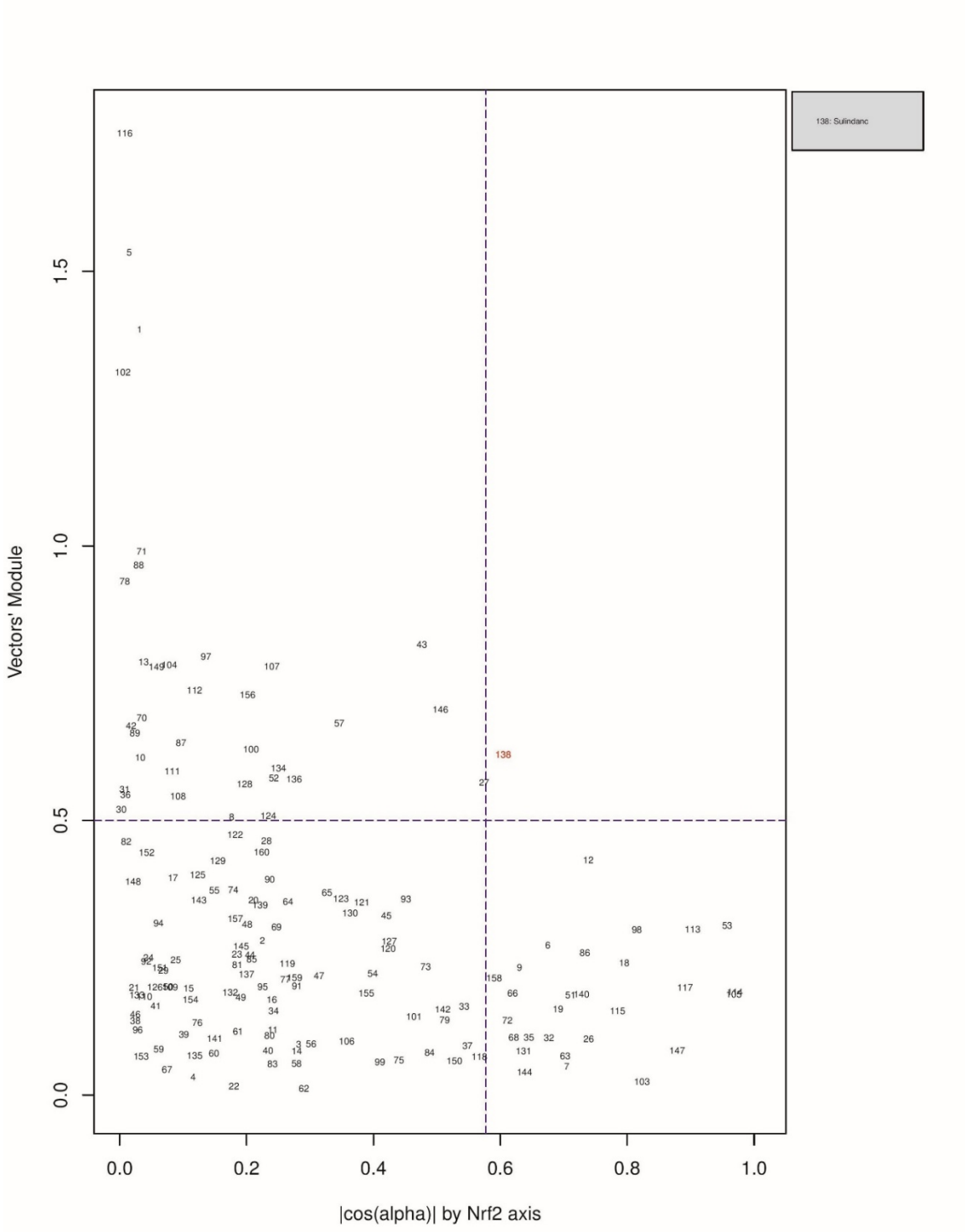


Figure 26. Distribution of chemicals by potency (Y-axis: module $\|\overrightarrow{OK}\|$ of the vector linking the origin O(0,0) to the chemical's point in a 3D space) and specificity to the Nrf2 pathway (X-axis: the absolute value of the $|\boldsymbol{cos}(\boldsymbol{\alpha})|$ of the angle between $\overrightarrow{OK}$ and the Nrf2 axis in a 3D space). Chemicals are represented by their rank in the alphabetically ordered list. The only chemical that is both strong (horizontal blue dashed line: $\|\overrightarrow{OK}\| > \mathbf{0.5}$) and Nrf2 specific (vertical blue dashed line: $\boldsymbol{cos}(\boldsymbol{\alpha}) = \frac{1}{\sqrt{3}}$) Sulindac, is in red and it is listed in the legend on the right.

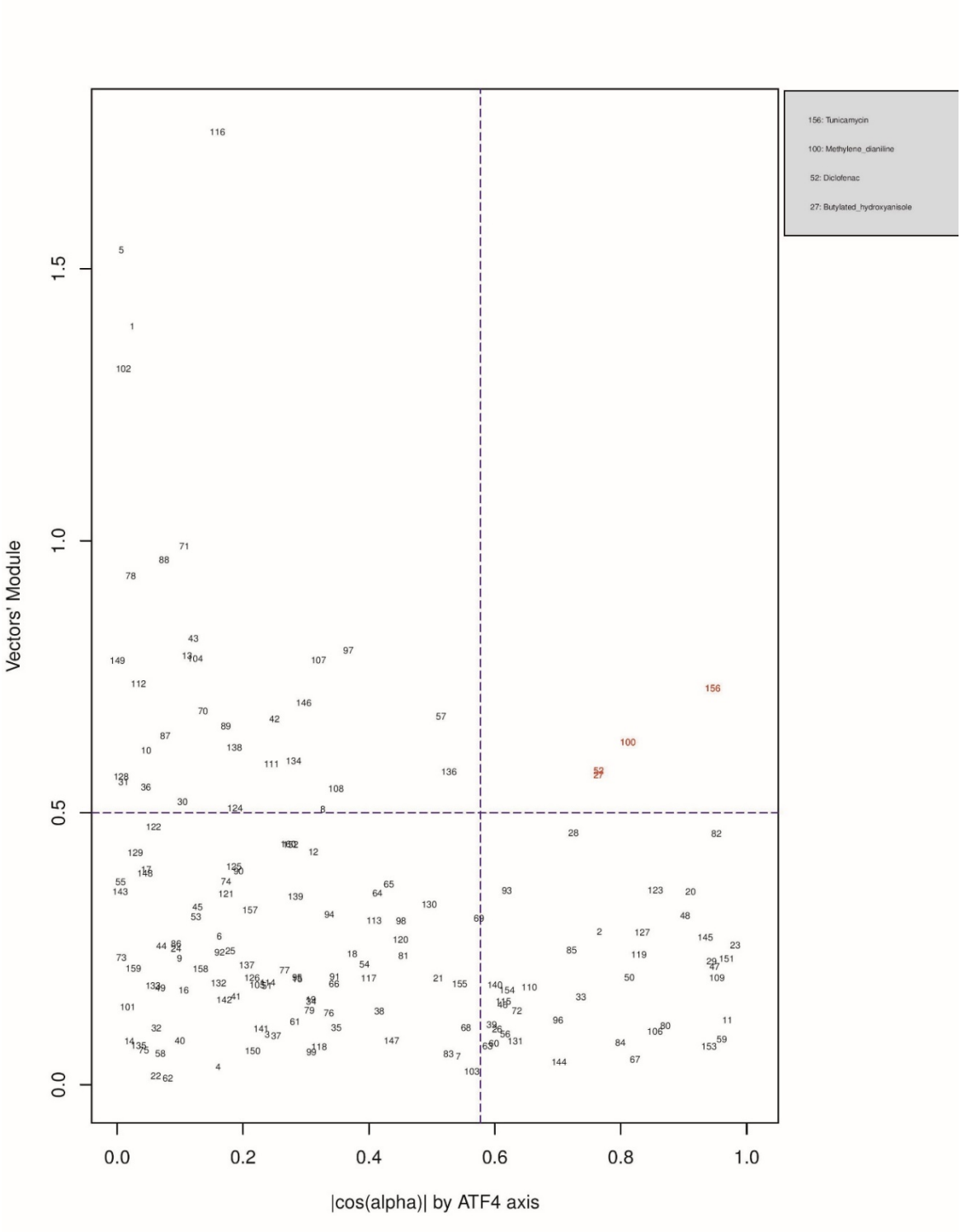


Figure 27. Distribution of chemicals by potency (Y-axis: module $\|\overrightarrow{OK}\|$ of the vector linking the origin O(0,0) to the chemical's point in a 3D space) and specificity to the ATF4 pathway (X-axis: the absolute value of the $|\boldsymbol{cos}(\boldsymbol{\alpha})|$ of the angle between $\overrightarrow{OK}$ and the ATF4 axis in a 3D space). Chemicals are represented by their rank in the alphabetically ordered list. Chemicals that are both strong (horizontal blue dashed line: $\|\overrightarrow{OK}\| > \mathbf{0.5}$) and ATF4 specific (vertical blue dashed line: $\boldsymbol{cos}(\boldsymbol{\alpha}) = \frac{1}{\sqrt{3}}$) are in red and their names are listed in the legend on the right.

5.4 DISCUSSION

Nrf2, ATF4 and AhR are important TFs in toxicological contexts and have well described downstream gene targets (Jennings *et al.*, 2013). Each of these TF have distinct unrelated upstream activation points, unique gene targets, but also have direct (*i.e.*, via multiple upstream promoter regions) and likely indirect overlaps on some specific gene targets. The AhR protein is a cytosolic protein receptor, where activation via chemical ligand binding causes nuclear translocation, DNA binding to it consensus sequence and RNA transcription. Several toxic compounds including dioxin-like compounds activate AhR. The TF Nrf2 is liberated from its cytosolic inhibitor Keap1, where the latter is sensitive to electrophiles and ROS. The TF ATF4 is activated via PERK, where PERK is activated when its inhibitor BiP, dissociates from PERK to bind unfolded proteins. All sorts of ER disturbances can cause an increase in unfolded proteins.

Using multiple toxicogenomic databases, we investigated the most appropriate activators of these three pathways, where it is expected that the chemical does not directly activate the other two pathways. These compounds were, Benzo(a)pyrene and Omeprazole for AhR, $KBrO_3$ and Phorone for Nrf2 and Tunicamycin for ATF4. All conditions up to and including 24 hours were pooled to generate a list of genes allocated to the three pathways (**Table 11**). This list confirmed the majority of a *priori* literature based information of 'Activated genes' (*i.e.*, upregulated). Although some genes were now reallocated to different pathways. The overlap with 'Inhibited genes' (*i.e.*, down regulated), was much poorer. This is too be expected as TF activated gene down regulation is much more complex and is often due to competition for auxiliary transcription facilitating proteins. Cytochrome P450 1A was the central element of the AhR pathway: *CYP1A1* is the most prominent gene of this pathway, regardless of the experimental category, followed by *CYP1A2*. These findings are similar to previous investigations and have been implemented in a systems biology model (Hamon *et al.*,

2014). For the Nrf2 pathway, the prototypical Nrf2 genes (*HMOX1*, *SRXN1* and *GCLM*) appear in the Nrf2 signature of all datasets, but also in the AhR-Nrf2 overlapping signature for most liver categories. This may reflect the fact that several AhR agonists are themselves metabolized to reactive chemicals via AhR dependent CYP expression. For example Benzo(a)pyrene is a substrate of the CYP1 sub family of cytochrome P450 enzymes, and it promotes its own metabolism to reactive epoxide and quinone products (Gelboin, 1980). These metabolic products can lead to oxidative stress and to an activation of the Nrf2 pathway as part of a second line of responses (Burchiel and Luster, 2001). The only activated gene that appears in the ATF4 signature of each of the three studied categories is *HERPUD1*. In most cases, *HERPUD1* also had the highest log2(FC) averages. Overlapping zones show an interaction between AhR and Nrf2, between Nrf2 and ATF4, but a very limited or non-existent interaction between AhR and ATF4 pathways.

We have used the exclusive pathway genes to create pathway CAC scores. The CAC reflects both specificity for the pathway ($\boldsymbol{cos}\ (\boldsymbol{\alpha})$) and the activation potency$\|\overrightarrow{\boldsymbol{OK}}\|$. CAC scores were generated for 160 chemicals using the TG-GATEs liver data. For ATF4, Tunicamycin, Methylene dianiline, Diclofenac and Butylated hydroxyanisole were ranked highest, in that order. Tunicamycin was used as a s pecific ATF4 specific activator. Both Diclofenac and Butylated hydroxyanisole have previously been demonstrated to positive modulate the ATF4 pathway (Afonyushkin *et al.*, 2010; Fredriksson *et al.*, 2014). The molecular mechanism for Methylene dianiline has not been fully elucidated and this evidence would suggest an ER disturbance and/or proteotoxic mechanism. For AhR, 34 chemicals were considered positive by CAC scores. Omeprazole was ranked highest, followed by Acetamidofluorene, 2-Nitrofluorene, Mexiletine, Flutamide, Isoniazid and Hexachlorobenzene. Many of the 34 chemicals have not been previously linked with AhR, but several are. These include, Hexachlorobenzene (de Tomaso Portaz *et al.*, 2015; Randi *et al.*,

2008), Ketoconazole (Novotna *et al.*, 2014), Clozapine (Donohoe *et al.*, 2008), and Doxorubicin (Volkova *et al.*, 2011). Fluphenazine has not been established as a ligand for the AhR, its structure – a halogenated aromatic ring system – closely matches the motif involved in binding to this receptor (Donohoe *et al.*, 2008). In a recent study we have demonstrated that Isoniazid induced *CYP1A1* in HepaRG cells, which is a potential indicator of AhR activation (Limonciel *et al.*, 2018). Only Sulindac from the 160 was ranked as active using the CAC selection criteria, which may seem surprising given the frequency of oxidative injury in liver toxicities. Although Butylated hydroxyanisole was marginal. The reason for a lack of Nrf2 activation prediction might be simply due to the fact that none of the 160 compounds, including the positive compound Phorone cause an Nrf2 response in the liver within the first 24 hours. Another possibility is that removing the overlapping genes has weakened the ability to pick up this pathway. Indeed, this is a weakness in the overall strategy as it is difficult to determine in such data sets if the pathways themselves are co-regulated since there are several gene overlaps amongst the pathways.

5.1 CONCLUSION

The size of the data set, its multiple sources, abundancy of compounds, concentrations and time of exposures, *in vitro* and *in vivo*, different organs are both a blessing and a curse. On the one hand, it is generally an advantage to have as broad as data set as possible, but the different sizes and focuses of the individual data sets/studies meant we needed to reduce the data to the lowest denomination. Another major issue was the low abundance of well described pathway activators. Despite these issues we have made some interesting observations and have developed a method to quantify a chemical's capacity to activate one three pathways.

We uncovered variations in AhR, ATF4 and Nrf2 signatures across tissues, compounds, species and *in vivo* vs *in vitro*. Some of these alterations are likely to be linked to pharmacokinetics, including distribution and metabolism, others may be linked to tissue specific regulation of these pathways. W hile some genes were very variable across experimental conditions, some were extremely robust, for example *CYP1A1* in the AhR pathway and *HERPUD1* in the ATF4 pathway. Some genes swing between a pathway's specific signature and overlapping zones for example *GCLC* between Nrf2 and AhR-Nrf2. Others are regularly on overlapping signatures for example *TPX2* and *TRIB3*. However, it is not possible with this type of analysis to delineate whether these overlaps are solely on a gene level or also on the pathway level.

The CAC score system developed, based on $\mathbf{cos}(\boldsymbol{\alpha}) \times \|\overrightarrow{\boldsymbol{OK}}\|$, can be used to quantify a chemical's specificity and potency to selectively activate one of these pathways. However, future work will be required to validate and optimize the gene signatures utilized.

6 THESIS SUMMARY AND CONCLUSION

Either among industries or research institutions, the reductionist approach to toxicology research and risk assessment remains predominant nowadays. While this approach has undeniably contributed to the progress of science for decades, it is progressively showing its weaknesses when it comes to studying multifactorial situations. With the development of modern experimental techniques, of bioinformatics tools and of omics, applying holistic system-level approaches (*i.e.*, SB, AOP, DBN *etc.*) to toxicology is becoming manageable, possible and even necessary.

More and more we discover that the mechanisms that underlie toxicity are complex and involve multiple biological processes and pathways. The key for a more general view of toxicity schemes is in understanding the different networks and pathways involved, their respective contribution to random outcomes as well as their potential interactions and cross-talks.

In toxicology, as in other fields, mathematical models are useful to gain insights into the governing principles of experimental observations, as well as to predict the behavior of a system in various situations. The challenge is to conceive tools and models able to reflect the complexity of interconnecting networks and pathways constituting a biological system.

In the introduction of this thesis (chapter 1), I described how the toxicology approach to date was leaving important questions surrounding the Nrf2 control oxidative stress unanswered. In addition, the question of a potential predictive and mechanistic vocation of toxicology was considered. In line with this reflection and expectation, a combination of SB and AOPs tools was suggested. In 'chapter 2' (Bibliography) the available published information covering the three facets of this subject (*i.e.*, toxicology, biological context and mathematical tools) was gathered and presented.

In order to uncover the mechanisms at play, we have elaborated in the 'chapter 3' a SB model of the role of the Nrf2 toxicity pathway in the control of oxidative stress. Our model of

the Nrf2 signaling pathway is a fusion of two complementary models: the first describes the synthesis, the metabolism and the transport of GSH under oxidative stress, and the second highlights the contribution of Nrf2 to the GSH response to oxidative stress. The latter was improved by remodeling the transcription/translation process of its genes using the Hill's model equation.

In 'chapter 4', using appropriate experimental data (*i.e.*, GSH, DCF and lactate levels following the exposure of RPTEC/TERT1 cells to of $KBrO_3$ for different doses and time-points) and statistical procedures (*i.e.*, MCMC simulations in a Bayesian framework), our SB model was calibrated, evaluated and compared to two other computational models (*i.e.*, an empirical dose-response statistical model and a DBN model). These three methods were explored as options for quantifying an AOP and deriving chemical independent KERs with rigorous statistical treatment of the data and parametric inference. While the "easy-to-develop" dose-response based qAOPs have a very limited extrapolation and explanation power and do not offer mechanistic insight, DBNs are in fact easier to develop, once the technology is mastered, but they either impose strong constraints on experimental design or require complex statistical treatment. Developing SB models is more complex, but they offer insight in the data collection and biology that the other approaches cannot afford.

Finally, in 'chapter 5' we studied the potential interactions of the Nrf2 pathway with two other signaling pathways (*i.e.*, AhR and ATF4) using multiple databases. This analysis pointed out the important codependences between the three pathways. Concerning the interactions with the AhR pathway, the results confirm the adequacy of inclusion of nucX-AhR as a co-TF for some genes in the Nrf2 SB model and encourage us to consider a hypothetical nucX-AhR activation of other prototypical Nrf2 genes of our model (*e.g.*, *HMOX1*, *SRXN1* and *GCLM*). In addition, these results open the door for testing a possible association of the ATF4 pathway (partially at least) to our SB model in the future. Moreover, uncovering variations of

the pathways signatures across different testing conditions (*i.e.*, tissues, compounds, species and *in vivo* versus *in vit*ro), this analysis improves the adaptability of our Nrf2 SB model and prepares it for a quantitative *in vitro in vivo* extrapolation and integration in a larger network setting.

One remarkable strength of the SB model is that it forces us to think mechanistically about new hypotheses and check whether they are compatible with the data. Emergent properties are actually the product of the integrating of these computational models with experiments in a spiral of iterative cycles of validation/falsification, simulation and theory. In our work for example, prediction and emergent properties could be confirmed, if some of the findings (*i.e.*, that a reasonable fit could be obtained if $KBrO_3$ acts directly on DCF, and that DCF bleaches significantly with time *etc.*) are validated in future experiments. Another importance of the SB approach is that it can fully propagate correct quantification of uncertainty associated with predictions, which is essential for proper risk assessment. Finally, SB models can naturally integrate pharmacokinetic models, since they are built from the same principles and same mathematical objects.

However, the work presented in this thesis shows that the use of SB is not easy and needs to mature. SB models, even though they provide a quite complete outlook of the biological systems and their components, they remain data-*hungry* and their development and calibration are time-consuming. Therefore, such complicated SB models could be seen as investment for the future rather than a quick answer to urgent questions. For an optimized calibration, it is very important that the generation of needed data be the fruit of experimental protocols that were elaborated by collective efforts including contributions of different research units participating to the conception and validation of the model. The problem is that, often, omics data produced are not specifically intended for SB model calibration and do not converge with the needs and expectations of the researchers working on the SB model development and

validation. Another limitation of this thesis is that at present the model quality is insufficient to claim that the precise calculations described here lead to reliable results. It was not possible to validate the whole model as we were unable to measure all the metabolites in the pathway. Also, the model does not always predict the experimental data, which suggests that there are additional reactions or regulations that need to be included in the model.

In addition to the aforementioned suggested improvements of our SB model, this work points to several directions for future research. After merging with adequate pharmacokinetic models for quantitative *in vitro in vivo* extrapolation, the application of SB tools developed here to toxicology has the unique opportunity to provide network insights into underlying mechanisms and basis of susceptibility to xenobiotics. First, using this SB model to evaluate exposures to mixtures of chemicals is a supplementary step towards a better modelling of biological and environmental realities. Second, by integrating individual-specific data to the model, it may be possible to better understand inter-individual differences in susceptibility to adverse effect of xenobiotics. Finally, on the longer-term, SB models and AOPs can be part of 'integrated approaches to testing and assessment' or 'integrated testing strategies' for regulatory decision making.

DISSEMINATION ACTIVITIES

Zgheib E, Limonciel A, Jiang X, Wilmes A, Wink S, van de Water B, Kopp-Schneider A, Bois FY & Jennings P. **2018**. Investigation of Nrf2, AhR and ATF4 Activation in Toxicogenomic Databases. Front Genet 9; doi:10.3389/fgene.2018.00429.

Zgheib E, Bechaux C, Crepet A, Mombelli E & Bois FY. **2017.** High-throughput methods for toxicology and health risk assessment. Environnement Risques Santé 44–58; doi:10.1684/ers.2016.0943.

Zgheib E, Gao W, Limonciel A, Aladjov H, Yang H, Tebby C, Gayraud G, Jennings P, Sachana M, Beltman JB & Bois FY. **2019**. Application of three approaches for quantitative AOP development to renal toxicity. Computational Toxicity (in press).

REFERENCES

ACuteTox Project Available at: http://www.acutetox.eu/ [Accessed October 24, 2018].

Aderem, A. (2005). Systems biology: its practice and challenges. *Cell* 121, 511–513. doi:10.1016/j.cell.2005.04.020.

Afonyushkin, T., Oskolkova, O. V., Philippova, M., Resink, T. J., Erne, P., Binder, B. R., et al. (2010). Oxidized Phospholipids Regulate Expression of ATF4 and VEGF in Endothelial Cells via NRF2-Dependent Mechanism: Novel Point of Convergence Between Electrophilic and Unfolded Protein Stress Pathways. *Arterioscler. Thromb. Vasc. Biol.* 30, 1007–1013. doi:10.1161/ATVBAHA.110.204354.

Allen, T. E. H., Goodman, J. M., Gutsell, S., and Russell, P. J. (2014). Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. *Chem. Res. Toxicol.* 27, 2100–2112. doi:10.1021/tx500345j.

Alon, U. (2007). *An introduction to systems biology: design principles of biological circuits*. Boca Raton, FL: Chapman & Hall/CRC.

Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P., and Orkin, S. H. (1993). Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* 362, 722–728. doi:10.1038/362722a0.

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., et al. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* 29, 730–741. doi:10.1002/etc.34.

AOP knowledge base Available at: https://aopkb.oecd.org/ [Accessed October 24, 2018].

AOP-Wiki Available at: https://aopwiki.org/ [Accessed October 24, 2018].

Ariens, E. J. (1954). Affinity and intrinsic activity in the theory of competitive inhibition. I. Problems and theory. *Arch. Int. Pharmacodyn. Ther.* 99, 32–49.

Aschauer, L., Gruber, L. N., Pfaller, W., Limonciel, A., Athersuch, T. J., Cavill, R., et al. (2013). Delineation of the key aspects in the regulation of epithelial monolayer formation. *Mol. Cell. Biol.* 33, 2535–2550. doi:10.1128/MCB.01435-12.

Aschauer, L., Limonciel, A., Wilmes, A., Stanzel, S., Kopp-Schneider, A., Hewitt, P., et al. (2015). Application of RPTEC/TERT1 cells for investigation of repeat dose nephrotoxicity: A transcriptomic study. *Toxicol. In Vitro* 30, 106–116. doi:10.1016/j.tiv.2014.10.005.

Astashkina, A., Mann, B., and Grainger, D. W. (2012). A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. *Pharmacol. Ther.* 134, 82–106. doi:10.1016/j.pharmthera.2012.01.001.

Atkins, G. L. (1973). A simple digital-computer program for estimating the parameters of the hill equation. *Eur. J. Biochem.* 33, 175–180.

Ballmaier, D., and Epe, B. (1995). Oxidative DNA damage induced by potassium bromate under cell-free conditions and in mammalian cells. *Carcinogenesis* 16, 335–342. doi:10.1093/carcin/16.2.335.

Bassik, M. C., and Kampmann, M. (2011). Knocking out the door to tunicamycin entry. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11731–11732. doi:10.1073/pnas.1109035108.

Bayes, and Mr. Price (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philos. Trans. 1683-1775* 53, 370–418.

Bayesian Analysis for a Logistic Regression Model - MATLAB & Simulink Example Available at: https://www.mathworks.com/help/stats/examples/bayesian-analysis-for-a-logistic-regression-model.html [Accessed October 24, 2018].

Bernillon, P., and Bois, F. Y. (2000). Statistical issues in toxicokinetic modeling: a Bayesian perspective. *Environ. Health Perspect.* 108 (suppl. 5), 883–893.

Bhattacharya, S., Zhang, Q., Carmichael, P. L., Boekelheide, K., and Andersen, M. E. (2011). Toxicity testing in the 21st century: defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. *Plos One* 6, e20887. doi:10.1371/journal.pone.0020887.

bioDBnet - Biological Database Network Available at: https://biodbnet-abcc.ncifcrf.gov/ [Accessed October 24, 2018].

Blaauboer, B. J. (2008). The contribution of in vitro toxicity data in hazard and risk assessment: Current limitations and future perspectives. *Toxicol. Lett.* 180, 81–84. doi:10.1016/j.toxlet.2008.05.008.

Boekelheide, K., and Andersen, M. E. (2010). A mechanistic redefinition of adverse effects - a key step in the toxicity testing paradigm shift. *ALTEX* 27, 243–252. doi:none.

Bois, F. (2012). “Bayesian inference,” in *Computational Toxicology vol. II*, eds. B. Reisfeld and A. N. Mayeno (New-York: Humana Press), 597–636.

Bois, F. Y. (2009a). GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. *Bioinformatics* 25, 1453–1454.

Bois, F. Y. (2009b). Physiologically-based modelling and prediction of drug interactions. *Basic Clin. Pharmacol. Toxicol.* 106, 154–161.

Bongso, A., Fong, C. Y., Ng, S. C., and Ratnam, S. (1994). Isolation and culture of inner cell mass cells from human blastocysts. *Hum. Reprod.* 9, 2110–2117. doi:none.

Burchiel, S. W., and Luster, M. I. (2001). Signaling by environmental polycyclic aromatic hydrocarbons in human lymphocytes. *Clin. Immunol. Orlando Fla* 98, 2–10. doi:10.1006/clim.2000.4934.

Carmichael, N. G., Barton, H. A., Boobis, A. R., Cooper, R. L., Dellarco, V. L., Doerrer, N. G., et al. (2006). Agricultural chemical safety assessment: A multisector approach to the modernization of human safety requirements. *Crit. Rev. Toxicol.* 36, 1–7.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). *Stan*: a probabilistic programming language. *J. Stat. Softw.* 76. doi:10.18637/jss.v076.i01.

Chandra, N. (2009). Computational systems approach for drug target discovery. *Expert Opin. Drug Discov.* 4, 1221–1236. doi:10.1517/17460440903380422.

Clark, A. J. (1926). The reaction between acetyl choline and muscle cells: Part II. *J. Physiol.* 64, 123–143.

Clarke, D. L., Johansson, C. B., Wilbertz, J., Veress, B., Nilsson, E., Karlström, H., et al. (2000). Generalized potential of adult neural stem cells. *Science* 288, 1660–1663. doi:10.1126/science.288.5471.1660.

Cohen Hubal, E. A., Richard, A. M., Shah, I., Gallagher, J., Kavlock, R. J., Blancato, J., et al. (2010). Exposure science and the U.S. EPA National Center for Computational Toxicology. *J. Expo. Sci. Environ. Epidemiol.* 20, 231–236. doi:10.1038/jes.2008.70.

Collins, F. S., Gray, G. M., and Bucher, J. R. (2008). TOXICOLOGY: Transforming Environmental Health Protection. *Science* 319, 906–907. doi:10.1126/science.1154619.

Conolly, R. B., Ankley, G. T., Cheng, W., Mayo, M. L., Miller, D. H., Perkins, E. J., et al. (2017). Quantitative Adverse Outcome Pathways and Their Application to Predictive Toxicology. *Environ. Sci. Technol.* 51, 4661–4672. doi:10.1021/acs.est.6b06230.

COSMOS - Integrated In Silico Models for the Prediction of Human Repeated Dose Toxicity of COSMetics to Optimise Safety Available at: http://www.cosmostox.eu/home/welcome/ [Accessed October 24, 2018].

Cotgreave, I. A. (2011). How can stem cell technologies be applied to replace animal use in toxicity testing? *Toxicol. Lett.* 205, S29. doi:10.1016/j.toxlet.2011.05.116.

Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., and Shinozaki, K. (2011). Effects of abiotic stress on pl ants: a systems biology perspective. *BMC Plant Biol.* 11, 163. doi:10.1186/1471-2229-11-163.

Crean, D., Bellwon, P., Aschauer, L., Limonciel, A., Moenks, K., Hewitt, P., et al. (2015). Development of an in vitro renal epithelial disease state model for xenobiotic toxicity testing. *Toxicol. Vitro Int. J. Publ. Assoc. BIBRA* 30, 128–137. doi:10.1016/j.tiv.2014.11.015.

Croom, E. (2012). Metabolism of xenobiotics of human environments. *Prog. Mol. Biol. Transl. Sci.* 112, 31–88. doi:10.1016/B978-0-12-415813-9.00003-9.

Csajka, C., and Verotta, D. (2006). Pharmacokinetic-pharmacodynamic modelling: History and perspectives. *J. Pharmacokinet. Pharmacodyn.* 33, 227–279.

De Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* 9, 67–103.

de Tomaso Portaz, A. C., Caimi, G. R., Sánchez, M., Chiappini, F., Randi, A. S., Kleiman de Pisarev, D. L., et al. (2015). Hexachlorobenzene induces cell proliferation, and aryl

hydrocarbon receptor expression (AhR) in rat liver preneoplastic foci, and in the human hepatoma cell line HepG2. AhR is a mediator of ERK1/2 signaling, and cell cycle regulation in HCB-treated HepG2 cells. *Toxicology* 336, 36–47. doi:10.1016/j.tox.2015.07.013.

Deshmukh, P., Unni, S., Krishnappa, G., and Padmanabhan, B. (2017). The Keap1-Nrf2 pathway: promising therapeutic target to counteract ROS-mediated damage in cancers and neurodegenerative diseases. *Biophys. Rev.* 9, 41–56. doi:10.1007/s12551-016-0244-4.

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., and Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* 95, 5–12. doi:10.1093/toxsci/kfl103.

diXa Data Warehouse Available at: http://wwwdev.ebi.ac.uk/fg/dixa/index.html [Accessed October 24, 2018].

Donohoe, D. R., Weeks, K., Aamodt, E. J., and Dwyer, D. S. (2008). Antipsychotic drugs alter neuronal development including ALM neuroblast migration and PLM axonal outgrowth in Caenorhabditis elegans. *Int. J. Dev. Neurosci. Off. J. Int. Soc. Dev. Neurosci.* 26, 371–380. doi:10.1016/j.ijdevneu.2007.08.021.

Duke, Modeling Cooperativity Available at: http://2013.igem.org/wiki/index.php?title=Team:Duke/Modeling/Cooperativity&oldid=215310 [Accessed October 24, 2018].

Edwards, S. W., Tan, Y.-M., Villeneuve, D. L., Meek, M. E., and McQueen, C. A. (2015). Adverse Outcome Pathways--Organizing Toxicological Information to Improve Decision Making. *J. Pharmacol. Exp. Ther.* 356, 170–181. doi:10.1124/jpet.115.228239.

Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi:10.1038/35002125.

Environmental Protection Agency (2007). ToxCast^{TM} program: predicting hazard, characterizing toxicity pathways, and prioritizing the toxicity testing of environmental chemicals. Washington, DC: EPA Available at: http://www.epa.gov/ncct/toxcast/index.html.

EU-ToxRisk – An Integrated European 'Flagship' Programme Driving Mechanism-based Toxicity Testing and Risk Assessment for the 21st century Available at: http://www.eu-toxrisk.eu/ [Accessed October 24, 2018].

Feng, B. Y., Simeonov, A., Jadhav, A., Babaoglu, K., Inglese, J., Shoichet, B. K., et al. (2007). A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* 50, 2385–2390. doi:10.1021/jm061317y.

Fernandes, T. G., Diogo, M. M., Clark, D. S., Dordick, J. S., and Cabral, J. M. S. (2009). High-throughput cellular microarray platforms: applications in drug discovery, toxicology and stem cell research. *Trends Biotechnol.* 27, 342–349. doi:10.1016/j.tibtech.2009.02.009.

Fisher, C., Siméon, S., Jamei, M., Gardner, I., Bois, F. Y., and Gardner, I. (in press). VIVD: virtual *in vitro* distribution model for the mechanistic prediction of intracellular concentrations of chemicals in *in vitro* toxicity assays. *Toxicol. In Vitro*.

Foth, H., and Hayes, A. (2008). Background of REACH in EU regulations on evaluation of chemicals. *Hum. Exp. Toxicol.* 27, 443–461. doi:10.1177/0960327108092296.

FP7 Available at: https://ec.europa.eu/research/fp7/index_en.cfm [Accessed October 24, 2018].

Fredriksson, L., Wink, S., Herpers, B., Benedetti, G., Hadi, M., de Bont, H., et al. (2014). Drug-induced endoplasmic reticulum and oxidative stress responses independently sensitize toward TNFα-mediated hepatotoxicity. *Toxicol. Sci. Off. J. Soc. Toxicol.* 140, 144–159. doi:10.1093/toxsci/kfu072.

Friend, S. H., and Schadt, E. E. (2014). Clues from the resilient. *Science* 344, 970–972. doi:10.1126/science.1255648.

Gaudio, E., Carpino, G., Cardinale, V., Franchitto, A., Onori, P., and Alvaro, D. (2009). New insights into liver stem cells. *Dig. Liver Dis.* 41, 455–462. doi:10.1016/j.dld.2009.03.009.

Gautier, L., Taboureau, O., and Audouze, K. (2013). The effect of network biology on drug toxicology. *Expert Opin. Drug Metab. Toxicol.* 9, 1409–1418. doi:10.1517/17425255.2013.820704.

Geenen, S., du Preez, F. B., Reed, M., Frederik Nijhout, H., Gerry Kenna, J., Wilson, I. D., et al. (2012a). A mathematical modelling approach to assessing the reliability of biomarkers of glutathione metabolism. *Eur. J. Pharm. Sci.* 46, 233–243. doi:10.1016/j.ejps.2011.08.017.

Geenen, S., du Preez, F. B., Snoep, J. L., Foster, A. J., Sarda, S., Kenna, J. G., et al. (2013). Glutathione metabolism modeling: A mechanism for liver drug-robustness and a new biomarker strategy. *Biochim. Biophys. Acta BBA - Gen. Subj.* 1830, 4943–4959. doi:10.1016/j.bbagen.2013.04.014.

Geenen, S., Taylor, P. N., Snoep, J. L., Wilson, I. D., Kenna, J. G., and Westerhoff, H. V. (2012b). Systems biology tools for toxicology. *Arch. Toxicol.* 86, 1251–1271.

Gelboin, H. V. (1980). Benzo[alpha]pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes. *Physiol. Rev.* 60, 1107–1166. doi:10.1152/physrev.1980.60.4.1107.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis, 2nd Edition*. London: Chapman & Hall.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* 7, 457–511.

Gesztelyi, R., Zsuga, J., Kemeny-Beke, A., Varga, B., Juhasz, B., and Tosaki, A. (2012). The Hill equation and the origin of quantitative pharmacology. *Arch. Hist. Exact Sci.* 66, 427–438. doi:10.1007/s00407-012-0098-5.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Gill, C. J., Sabin, L., and Schmid, C. H. (2005). Why clinicians are natural bayesians. *BMJ* 330, 1080–1083. doi:10.1136/bmj.330.7499.1080.

Girolami, M., and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B* 73, 123–214. doi:10.1111/j.1467-9868.2010.00765.x.

Goelzer, A., Bekkal Brikci, F., Martin-Verstraete, I., Noirot, P., Bessières, P., Aymerich, S., et al. (2008). Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of Bacillus subtilis. *BMC Syst. Biol.* 2, 20. doi:10.1186/1752-0509-2-20.

Groh, K. J., and Tollefsen, K. E. (2015). *The Challenge* : Adverse outcome pathways in research and regulation-Current status and future perspectives. *Environ. Toxicol. Chem.* 34, 1935–1937. doi:10.1002/etc.3042.

Gundert-Remy, U., Barth, H., Bürkle, A., Degen, G. H., and Landsiedel, R. (2015). Toxicology: a discipline in need of academic anchoring—the point of view of the German Society of Toxicology. *Arch. Toxicol.* 89, 1881–1893. doi:10.1007/s00204-015-1577-7.

Haarmann-Stemmann, T., Aarbakke, J., Fritsche, E., and Krutmann, J. (2012). The AhR–Nrf2 Pathway in Keratinocytes: On the Road to Chemoprevention? *J. Invest. Dermatol.* 132, 7–9. doi:10.1038/jid.2011.359.

Halliwell, B., Gutteridge, J. M., and Cross, C. E. (1992). Free radicals, antioxidants, and human disease: where are we now? *J. Lab. Clin. Med.* 119, 598–620.

Hamon, J., Jennings, P., and Bois, F. Y. (2014). Systems biology modeling of omics data: effect of cyclosporine a on the Nrf2 pathway in human renal cells. *BMC Syst. Biol.* 8, 76. doi:10.1186/1752-0509-8-76.

Hamon, J., Renner, M., Jamei, M., Lukas, A., Kopp-Schneider, A., and Bois, F. Y. (2015). Quantitative in vitro to in vivo extrapolation of tissues toxicity. *Toxicol. In Vitro*. doi:10.1016/j.tiv.2015.01.011.

Hartung, T. (2009). A Toxicology for the 21st Century--Mapping the Road Ahead. *Toxicol. Sci.* 109, 18–23. doi:10.1093/toxsci/kfp059.

Hartung, T. (2011). From alternative methods to a new toxicology. *Eur. J. Pharm. Biopharm.* 77, 338–349. doi:10.1016/j.ejpb.2010.12.027.

Hasdemir, D., Hoefsloot, H. C. J., and Smilde, A. K. (2015). Validation and selection of ODE based systems biology models: how to arrive at more reliable decisions. *BMC Syst. Biol.* 9, 32. doi:10.1186/s12918-015-0180-0.

Hassan, I., El-Masri, H., Kosian, P. A., Ford, J., Degitz, S. J., and 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. doi:10.1093/toxsci/kfx163.

Haynes, L. W., Kay, A. R., and Yau, K. W. (1986). Single cyclic GMP-activated channel activity in excised patches of rod outer segment membrane. *Nature* 321, 66–70. doi:10.1038/321066a0.

He, C. H., Gong, P., Hu, B., Stewart, D., Choi, M. E., Choi, A. M., et al. (2001). Identification of activating transcription factor 4 (ATF4) as an Nrf2-interacting protein. Implication for heme oxygenase-1 gene regulation. *J. Biol. Chem.* 276, 20858–20865. doi:10.1074/jbc.M101198200.

Hetz, C. (2012). The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat. Rev. Mol. Cell Biol.* 13, 89–102. doi:10.1038/nrm3270.

Himmelstein, M. W., Carpenter, S. C., Evans, M. V., Hinderliter, P. M., and Kenyon, E. M. (2004). Kinetic modeling of beta-chloroprene metabolism: II. The application of physiologically based modeling for cancer dose response analysis. *Toxicol. Sci. Off. J. Soc. Toxicol.* 79, 28–37. doi:10.1093/toxsci/kfh096.

Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Stat. Med.* 9, 811–818.

Holleman, D. R., and Simel, D. L. (1997). Quantitative assessments from the clinical examination. How should clinicians integrate the numerous results? *J. Gen. Intern. Med.* 12, 165–171.

Hooper, M. J., Ankley, G. T., Cristol, D. A., Maryoung, L. A., Noyes, P. D., and Pinkerton, K. E. (2013). Interactions between chemical and climate stressors: A role for mechanistic toxicology in assessing climate change risks. *Environ. Toxicol. Chem.* 32, 32–48. doi:10.1002/etc.2043.

Horgan, R. P., and Kenny, L. C. (2011). 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics: The Obstetrician & Gynaecologist. *Obstet. Gynaecol.* 13, 189–195. doi:10.1576/toag.13.3.189.27672.

Horizon 2020 Available at: http://www.horizon2020.gouv.fr/ [Accessed October 24, 2018].

Howden, C. W. (1991). Clinical pharmacology of omeprazole. *Clin. Pharmacokinet.* 20, 38–49. doi:10.2165/00003088-199120010-00003.

Huang, H. C., Nguyen, T., and Pickett, C. B. (2000). Regulation of the antioxidant response element by protein kinase C-mediated phosphorylation of NF-E2-related factor 2. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12475–12480. doi:10.1073/pnas.220418997.

Iannone, A., Tomasi, A., Vannini, V., and Swartz, H. M. (1990). Metabolism of nitroxide spin labels in subcellular fractions of rat liver. II. Reduction in the cytosol. *Biochim. Biophys. Acta* 1034, 290–293.

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., et al. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934. doi:10.1126/science.292.5518.929.

Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., et al. (2015). Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 43, D921-927. doi:10.1093/nar/gku955.

Inglese, J., Auld, D. S., Jadhav, A., Johnson, R. L., Simeonov, A., Yasgar, A., et al. (2006). Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci.* 103, 11473–11478. doi:10.1073/pnas.0604348103.

Inglese, J., Johnson, R. L., Simeonov, A., Xia, M., Zheng, W., Austin, C. P., et al. (2007). High-throughput screening assays for the identification of chemical probes. *Nat. Chem. Biol.* 3, 466–479. doi:10.1038/nchembio.2007.17.

Innovative Medicines Initiative *IMI Innov. Med. Initiat.* Available at: http://www.imi.europa.eu/ [Accessed October 24, 2018].

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostat. Oxf. Engl.* 4, 249–264. doi:10.1093/biostatistics/4.2.249.

Itoh, K., Wakabayashi, N., Katoh, Y., Ishii, T., Igarashi, K., Engel, J. D., et al. (1999). Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev.* 13, 76–86.

Jaworska, J., Dancik, Y., Kern, P., Gerberick, F., and Natsch, A. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice: Integrated testing strategy to assess skin sensitization potency. *J. Appl. Toxicol.*, n/a-n/a. doi:10.1002/jat.2869.

Jaworska, J., Gabbert, S., and Aldenberg, T. (2010). Towards optimization of chemical testing under REACH: a Bayesian network approach to Integrated Testing Strategies. *Regul. Toxicol. Pharmacol.* 57, 157–167. doi:10.1016/j.yrtph.2010.02.003.

Jaworska, J. S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., and 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. doi:10.1007/s00204-015-1634-2.

Jennings, P., Limonciel, A., Felice, L., and Leonard, M. (2012). An overview of transcriptional regulation in response to toxicological insult. *Arch. Toxicol.*, 1–24. doi:10.1007/s00204-012-0919-y.

Jennings, P., Limonciel, A., Felice, L., and Leonard, M. O. (2013). An overview of transcriptional regulation in response to toxicological insult. *Arch. Toxicol.* 87, 49–72. doi:10.1007/s00204-012-0919-y.

Jin, U.-H., Lee, S., and Safe, S. (2012). Aryl hydrocarbon receptor (AHR)-active pharmaceuticals are selective AHR modulators in MDA-MB-468 and BT474 breast cancer cells. *J. Pharmacol. Exp. Ther.* 343, 333–341. doi:10.1124/jpet.112.195339.

Judson, R., Houck, K., Martin, M., Knudsen, T., Thomas, R. S., Sipes, N., et al. (2014). *In Vitro* and Modelling Approaches to Risk Assessment from the U.S. Environmental Protection

Agency ToxCast Programme. *Basic Clin. Pharmacol. Toxicol.* 115, 69–76. doi:10.1111/bcpt.12239.

Judson, R. S., Kavlock, R. J., Martin, M., Reif, D., Houck, K., Knudsen, T., et al. (2013). Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX* 30, 51–56. doi:10.14573/altex.2013.1.051.

Kaspar, J. W., Niture, S. K., and Jaiswal, A. K. (2009). Nrf2:INrf2 (Keap1) signaling in oxidative stress. *Free Radic. Biol. Med.* 47, 1304–1309. doi:10.1016/j.freeradbiomed.2009.07.035.

Kavlock, R. J., Ankley, G., Blancato, J., Breen, M., Conolly, R., Dix, D., et al. (2007). Computational Toxicology--A State of the Science Mini Review. *Toxicol. Sci.* 103, 14–27. doi:10.1093/toxsci/kfm297.

Kitambi, S. S., and Chandrasekar, G. (2011). Stem cells: a model for screening, discovery and development of drugs. *Stem Cells Cloning Adv. Appl.* 4, 51–59. doi:10.2147/SCCAA.S16417.

Kitano, H. (2002). Computational systems biology. *Nature* 420, 206–210. doi:10.1038/nature01254.

Kjærulff, U. B., and Madsen, A. L. (2008). *Bayesian Networks and Influence Diagrams*. New York, NY: Springer New York doi:10.1007/978-0-387-74101-7.

Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., and Herwig, R. eds. (2010). *Systems biology: a textbook*. 1. reprint. Weinheim: Wiley-VCH.

Kobayashi, A., Kang, M.-I., Okawa, H., Ohtsuji, M., Zenke, Y., Chiba, T., et al. (2004). Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. *Mol. Cell. Biol.* 24, 7130–7139. doi:10.1128/MCB.24.16.7130-7139.2004.

Kohen, R., and Nyska, A. (2002). Oxidation of biological systems: oxidative stress phenomena, antioxidants, redox reactions, and methods for their quantification. *Toxicol. Pathol.* 30, 620–650. doi:10.1080/01926230290166724.

Kong, Y., Trabucco, S. E., and Zhang, H. (2014). "Oxidative Stress, Mitochondrial Dysfunction and the Mitochondria Theory of Aging," in *Interdisciplinary Topics in Gerontology*, eds. L. Robert and T. Fulop (Basel: S. KARGER AG), 86–107. Available at: http://www.karger.com?doi=10.1159/000358901 [Accessed October 24, 2018].

Kongsbak, K., Vinggaard, A. M., Hadrup, N., and Audouze, K. (2014). A computational approach to mechanistic and predictive toxicology of pesticides. *ALTEX - Altern. Anim. Exp.* 31, 11–22. doi:10.14573/altex.1304241.

Krewski, D., Andersen, M. E., Mantus, E., and Zeise, L. (2009). Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal.* 29, 474–479. doi:10.1111/j.1539-6924.2008.01150.x.

Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspect. Psychol. Sci.* 6, 299–312. doi:10.1177/1745691611406925.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* 142, 573–603. doi:10.1037/a0029146.

LaLone, C. A., Ankley, G. T., Belanger, S. E., Embry, M. R., Hodges, G., Knapen, D., et al. (2017). Advancing the adverse outcome pathway framework-An international horizon scanning approach. *Environ. Toxicol. Chem.* 36, 1411–1421. doi:10.1002/etc.3805.

Leclerc, E., Hamon, J., Legendre, A., and Bois, F. (2014). Integration of pharmacokinetic and NRF2 system biology models to describe reactive oxygen species production and subsequent glutathione depletion in liver microfluidic biochips after flutamide exposure. *Toxicol. In Vitro*.

Leist, M., Ghallab, A., Graepel, R., Marchan, R., Hassan, R., Bennekou, S. H., et al. (2017). Adverse outcome pathways: opportunities, limitations and open questions. *Arch. Toxicol.* 91, 3477–3505. doi:10.1007/s00204-017-2045-3.

Li, P., and Vu, Q. D. (2013). Identification of parameter correlations for parameter estimation in dynamic biological models. *BMC Syst. Biol.* 7, 91. doi:10.1186/1752-0509-7-91.

Limonciel, A., Aschauer, L., Wilmes, A., Prajczer, S., Leonard, M. O., Pfaller, W., et al. (2011). Lactate is an ideal non-invasive marker for evaluating temporal alterations in cell stress and toxicity in repeat dose testing regimes. *Toxicol. In Vitro* 25, 1855–1862. doi:10.1007/s00204-012-0897-0.

Limonciel, A., Ates, G., Carta, G., Wilmes, A., Watzele, M., Shepard, P. J., et al. (2018). Comparison of base-line and chemical-induced transcriptomic responses in HepaRG and RPTEC/TERT1 cells using TempO-Seq. *Arch. Toxicol.* 92, 2517–2531. doi:10.1007/s00204-018-2256-2.

Limonciel, A., Moenks, K., Stanzel, S., Truisi, G. L., Parmentier, C., Aschauer, L., et al. (2015). Transcriptomics hit the target: Monitoring of ligand-activated and stress response pathways for chemical testing. *Toxicol. Vitro Int. J. Publ. Assoc. BIBRA* 30, 7–18. doi:10.1016/j.tiv.2014.12.011.

Limonciel, A., Wilmes, A., Aschauer, L., Radford, R., Bloch, K. M., McMorrow, T., et al. (2012). Oxidative stress induced by potassium bromate exposure results in altered tight junction protein expression in renal proximal tubule cells. *Arch. Toxicol.* 86, 1741–1751. doi:10.1007/s00204-012-0897-0.

Liu, Z., Shi, Q., Ding, D., Kelly, R., Fang, H., and Tong, W. (2011). Translating clinical findings into knowledge in drug safety evaluation--drug induced liver injury prediction system (DILIps). *PLoS Comput. Biol.* 7, e1002310. doi:10.1371/journal.pcbi.1002310.

Machado, D., Costa, R. S., Rocha, M., Ferreira, E. C., Tidor, B., and Rocha, I. (2011). Modeling formalisms in systems biology. *AMB Express* 1, 45. doi:10.1186/2191-0855-1-45.

Mager, D. E., Wyska, E., and Jusko, W. J. (2003). Diversity of mechanism-based pharmacodynamic models. *Drug Metab. Dispos.* 31, 510–518.

Materi, W., and Wishart, D. S. (2007). Computational systems biology in drug discovery and development: methods and applications. *Drug Discov. Today* 12, 295–303. doi:10.1016/j.drudis.2007.02.013.

Mayr, L. M., and Fuerst, P. (2008). The Future of High-Throughput Screening. *J. Biomol. Screen.* 13, 443–448. doi:10.1177/1087057108319644.

McLean, F. C. (1938). APPLICATION OF THE LAW OF CHEMICAL EQUILIBRIUM (LAW OF MASS ACTION) TO BIOLOGICAL PROBLEMS. *Physiol. Rev.* 18, 495–523. doi:10.1152/physrev.1938.18.4.495.

Menasché, P. (2011). Stem cell therapy for chronic heart failure: lessons from a 15-year experience. *C. R. Biol.* 334, 489–496. doi:10.1016/j.crvi.2011.03.006.

Moon, Y. J., Lee, M. W., Yoon, H. H., Yang, M. S., Jang, I. K., Lee, J. E., et al. (2008). Hepatic differentiation of cord blood-derived multipotent progenitor cells (MPCs) in vitro. *Cell Biol. Int.* 32, 1293–1301. doi:10.1016/j.cellbi.2008.07.017.

Mueller, S. O., Dekant, W., Jennings, P., Testai, E., and Bois, F. (2015). Comprehensive summary--Predict-IV: A systems toxicology approach to improve pharmaceutical drug safety testing. *Toxicol. Vitro Int. J. Publ. Assoc. BIBRA* 30, 4–6. doi:10.1016/j.tiv.2014.09.016.

National Research Council (2007). Toxicity Testing in the 21st Century: A Vision and A Strategy. Washington, DC: National Academy Press Available at: http://www.nap.edu/openbook.php?record_id=11970&page=R15 [Accessed October 24, 2018].

National Toxicology Program (2004). A National Toxicology Program for the 21st Century : A Roadmap for the Future. Washington, DC: NTP Available at: http://ntp.niehs.nih.gov/ntp/main_pages/ntpvision.pdf [Accessed March 17, 2015].

Nebert, D. W., and Dalton, T. P. (2006). The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nat. Rev. Cancer* 6, 947–960. doi:10.1038/nrc2015.

Nebert, D. W., Dalton, T. P., Okey, A. B., and Gonzalez, F. J. (2004). Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J. Biol. Chem.* 279, 23847–23850. doi:10.1074/jbc.R400004200.

Nguyen, T., Huang, H. C., and Pickett, C. B. (2000). Transcriptional regulation of the antioxidant response element. Activation by Nrf2 and repression by MafK. *J. Biol. Chem.* 275, 15466–15473. doi:10.1074/jbc.M000361200.

NOTOX Available at: http://www.notox-sb.eu/ [Accessed October 24, 2018].

Novotna, A., Korhonova, M., Bartonkova, I., Soshilov, A. A., Denison, M. S., Bogdanova, K., et al. (2014). Enantiospecific effects of ketoconazole on aryl hydrocarbon receptor. *PloS One* 9, e101832. doi:10.1371/journal.pone.0101832.

NuGO R/Bioconductor support libraries Available at: http://nmg-r.bioinformatics.nl/NuGO_R.html [Accessed October 24, 2018].

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

OECD (2013). Guidance document on developing and assessing adverse outcome pathways.

OECD, 2016 (2016). "Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways," in Series on Testing and Assessment. (Environment, Health and Safety Publications). Available at: http://aopkb.org/common/AOP_Handbook.pdf [Accessed October 25, 2016].

Oguro, T., Hayashi, M., Numazawa, S., Asakawa, K., and Yoshida, T. (1996). Heme oxygenase-1 gene expression by a glutathione depletor, phorone, mediated through AP-1 activation in rats. *Biochem. Biophys. Res. Commun.* 221, 259–265. doi:10.1006/bbrc.1996.0583.

O'Malley, J., Woltjen, K., and Kaji, K. (2009). New strategies to generate induced pluripotent stem cells. *Curr. Opin. Biotechnol.* 20, 516–521. doi:10.1016/j.copbio.2009.09.005.

Omoe, H. (2006). Recent Trends in Animal experimentation in Japan — On the Revision and Implementation of the Law for the Humane Treatment and Management of Animals — . *Trends Q. Rev.* 21, 13–31.

Open TG-GATEs Available at: https://dbarchive.biosciencedbc.jp/en/open-tggates/desc.html [Accessed October 24, 2018].

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 Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)67&doclanguage=en.

Organisation for Economic Co-operation and Development (OECD) (2017). Guidance Document on Developing and Assessing Adverse Outcome Pathways. Second Edition. Series on T esting and Assessment, No. 184, ENV/JM/MONO(2013)6. OECD Environment, Health and Safety Publications, Paris, France Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en.

Orton, R. J., Sturm, O. E., Vyshemirsky, V., Calder, M., Gilbert, D. R., and Kolch, W. (2005). Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. *Biochem. J.* 392, 249–261. doi:10.1042/BJ20050908.

Oslowski, C. M., and Urano, F. (2011). Measuring ER stress and the unfolded protein response using mammalian tissue culture system. *Methods Enzymol.* 490, 71–92. doi:10.1016/B978-0-12-385114-7.00004-0.

Park, B. K., Kitteringham, N. R., Powell, H., and Pirmohamed, M. (2000). Advances in molecular toxicology-towards understanding idiosyncratic drug toxicity. *Toxicology* 153, 39–60.

Pauwels, M., and Rogiers, V. (2004). Safety evaluation of cosmetics in the EU. *Toxicol. Lett.* 151, 7–17. doi:10.1016/j.toxlet.2004.01.026.

Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., et al. (2011). Using graph theory to analyze biological networks. *BioData Min.* 4. doi:10.1186/1756-0381-4-10.

Pavlovic, V. I. (1999). Dynamic Bayesian Networks for Information Fusion with Applications to Human-computer Interfaces.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, Calif: Morgan Kaufmann Publishers.

Pelkonen, O. (2010). Predictive toxicity: grand challenges. *Front. Pharmacol.* 1. doi:10.3389/fphar.2010.00003.

Petrulis, J. R., and Perdew, G. H. (2002). The role of chaperone proteins in the aryl hydrocarbon receptor core complex. *Chem. Biol. Interact.* 141, 25–40.

Pham-Huy, L. A., He, H., and Pham-Huy, C. (2008). Free radicals, antioxidants in disease and health. *Int. J. Biomed. Sci. IJBS* 4, 89–96.

Phelps, M. A., and Levitt, M. A. (2004). Pretest probability estimates: a pitfall to the clinical utility of evidence-based medicine? *Acad. Emerg. Med. Off. J. Soc. Acad. Emerg. Med.* 11, 692–694.

Pittenger, M. F. (2008). Mesenchymal stem cells from adult bone marrow. *Methods Mol. Biol.* 449, 27–44. doi:10.1007/978-1-60327-169-1_2.

Quignot, N., and Bois, F. Y. (2013). A Computational Model to Predict Rat Ovarian Steroid Secretion from In Vitro Experiments with Endocrine Disruptors. *PLoS ONE* 8, e53891. doi:10.1371/journal.pone.0053891.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing Available at: http://www.R-project.org.

Randi, A. S., Sanchez, M. S., Alvarez, L., Cardozo, J., Pontillo, C., and Kleiman de Pisarev, D. L. (2008). Hexachlorobenzene triggers AhR translocation to the nucleus, c-Src activation and EGFR transactivation in rat liver. *Toxicol. Lett.* 177, 116–122. doi:10.1016/j.toxlet.2008.01.003.

Rangarajan, A., and Weinberg, R. A. (2003). Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat. Rev. Cancer* 3, 952–959. doi:10.1038/nrc1235.

Raunio, H. (2011). In Silico Toxicology – Non-Testing Methods. *Front. Pharmacol.* 2. doi:10.3389/fphar.2011.00033.

Reddy, S. P. (2008). The antioxidant response element and oxidative stress modifiers in airway diseases. *Curr. Mol. Med.* 8, 376–383.

Reed, M. C., Thomas, R. L., Pavisic, J., James, S. J., Ulrich, C. M., and Nijhout, H. F. (2008). A mathematical model of glutathione metabolism. *Theor. Biol. Med. Model.* 5, 8. doi:10.1186/1742-4682-5-8.

Reuveni, S., Urbakh, M., and Klafter, J. (2014). Role of substrate unbinding in Michaelis-Menten enzymatic reactions. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4391–4396. doi:10.1073/pnas.1318122111.

Rodriguez-Fernandez, M., Rehberg, M., Kremling, A., and Banga, J. R. (2013). Simultaneous model discrimination and parameter estimation in dynamic models of cellular systems. *BMC Syst. Biol.* 7, 76. doi:10.1186/1752-0509-7-76.

Rougier, F., Claude, D., Maurin, M., Sedoglavic, A., Ducher, M., Corvaisier, S., et al. (2003). Aminoglycoside nephrotoxicity: modeling, simulation, and control. *Antimicrob. Agents Chemother.* 47, 1010–1016.

Rovida, C., Alépée, N., Api, A. M., Basketter, D. A., Bois, F. Y., Caloni, F., et al. (2015). Integrated testing strategies (ITS) for safety assessment. *Altern. Lab. Anim.* 32, 25–40. doi:http://dx.doi.org/10.14573/altex.1411011.

Rowan, A. N. (1983). "Alternatives: Interaction between science and animal welfare.," in *Product safety evaluation* (New York: Mary Ann Liebert), 113–133. Available at: https://habricentral.org/resources/27259/about [Accessed October 24, 2018].

Russell, W. M. S., and Burch, R. L. (1959). *The principles of humane experimental technique*. Special ed. Methuen, London.

Sachana, M., and 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. Vitro Toxicol.* 3, 227–233. doi:10.1089/aivt.2017.0013.

Schmidt, C. W. (2009). TOX 21: new dimensions of toxicity testing. *Environ. Health Perspect.* 117, A348-353.

Scrtox Project Available at: http://www.scrtox.eu/ [Accessed October 24, 2018].

Seidle, T., and Stephens, M. L. (2009). Bringing toxicology into the 21st century: a global call to action. *Toxicol. In Vitro* 23, 1576–1579. doi:10.1016/j.tiv.2009.06.012.

SEURAT-1 - Towards the Replacement of in vivo Repeated Dose Systemic Toxicity Testing Available at: http://www.seurat-1.eu/ [Accessed October 24, 2018].

Shukla, S. J., Huang, R., Austin, C. P., and Xia, M. (2010). The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. *Drug Discov. Today* 15, 997–1007. doi:10.1016/j.drudis.2010.07.007.

Smyth, G. K., Michaud, J., and Scott, H. S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinforma. Oxf. Engl.* 21, 2067–2075. doi:10.1093/bioinformatics/bti270.

Spiegelhalter, D., and Rice, K. (2009). Bayesian statistics. *Scholarpedia* 4, 5230. doi:10.4249/scholarpedia.5230.

Standard Microplates - 96-,384-, and 1536-well format *WellPlate.com*. Available at: https://www.wellplate.com/standard-microplates/ [Accessed October 24, 2018].

Staples, A., and Wong, C. (2010). Risk factors for progression of chronic kidney disease. *Curr. Opin. Pediatr.* 22, 161–169. doi:10.1097/MOP.0b013e328336ebb0.

Stefan, M. I., and Le Novère, N. (2013). Cooperative binding. *PLoS Comput. Biol.* 9, e1003106. doi:10.1371/journal.pcbi.1003106.

StemBANCC Project Available at: http://stembancc.org/ [Accessed October 24, 2018].

Stigler, S. M. (1986). Laplace's 1774 Memoir on Inverse Probability. *Stat. Sci.* 1, 359–363. doi:10.1214/ss/1177013620.

Systems Biology at University of Lyon — BioSyL Available at: http://www.biosyl.org/ [Accessed October 24, 2018].

Szymański, P., Markowicz, M., and Mikiciuk-Olasik, E. (2011). Adaptation of High-Throughput Screening in Drug Discovery—Toxicological Screening Tests. *Int. J. Mol. Sci.* 13, 427–452. doi:10.3390/ijms13010427.

Taboureau, O., and Audouze, K. (2017). Human environmental disease network: A computational model to assess toxicology of contaminants. *ALTEX - Altern. Anim. Exp.* 34, 289–300. doi:10.14573/altex.1607201.

Taguchi, K., Motohashi, H., and Yamamoto, M. (2011). Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution: Molecular mechanisms of the Keap1-Nrf2 pathway. *Genes Cells* 16, 123–140. doi:10.1111/j.1365-2443.2010.01473.x.

Tan, T. W., Lim, S. J., Khan, A. M., and Ranganathan, S. (2009). A proposed minimum skill set for university graduates to meet the informatics needs and challenges of the "-omics" era. *BMC Genomics* 10 Suppl 3, S36. doi:10.1186/1471-2164-10-S3-S36.

Trevan, J. W. (1927). The Error of Determination of Toxicity. *Proc. R. Soc. B Biol. Sci.* 101, 483–514. doi:10.1098/rspb.1927.0030.

Tyson, J. J. (1991). Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7328–7332.

US EPA, O. Toxicology Testing in the 21st Century (Tox21). *Toxicol. Test. 21st Century Tox21*. Available at: https://www.epa.gov/chemical-research/toxicology-testing-21st-century-tox21 [Accessed October 24, 2018].

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., and van Aken, M. A. G. (2014). A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Dev.* 85, 842–860. doi:10.1111/cdev.12169.

van Vliet, E. (2011). Current standing and future prospects for the technologies proposed to transform toxicity testing in the 21st century. *ALTEX* 28, 17–44. doi:10.14573/altex.2011.1.017.

Villeneuve, D. L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T. H., LaLone, C. A., et al. (2014). Adverse Outcome Pathway (AOP) Development I: Strategies and Principles. *Toxicol. Sci.* 142, 312–320. doi:10.1093/toxsci/kfu199.

Villeneuve, N. F., Lau, A., and Zhang, D. D. (2010). Regulation of the Nrf2-Keap1 antioxidant response by the ubiquitin proteasome system: an insight into cullin-ring ubiquitin ligases. *Antioxid. Redox Signal.* 13, 1699–1712. doi:10.1089/ars.2010.3211.

Vinken, M. (2013). The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicology* 312, 158–165. doi:10.1016/j.tox.2013.08.011.

Vinken, M., Doktorova, T., Ellinger-Ziegelbauer, H., Ahr, H.-J., Lock, E., Carmichael, P., et al. (2008). The carcinoGENOMICS project: critical selection of model compounds for the development of omics-based in vitro carcinogenicity screening assays. *Mutat. Res.* 659, 202–210. doi:10.1016/j.mrrev.2008.04.006.

Volkova, M., Palmeri, M., Russell, K. S., and Russell, R. R. (2011). Activation of the aryl hydrocarbon receptor by doxorubicin mediates cytoprotective effects in the heart. *Cardiovasc. Res.* 90, 305–314. doi:10.1093/cvr/cvr007.

Wagner, J. G. (1968). Kinetics of pharmacologic response. I. Proposed relationships between response and drug concentration in the intact animal and man. *J. Theor. Biol.* 20, 173–201.

Waters, M. D., and Fostel, J. M. (2004). Toxicogenomics and systems toxicology: aims and prospects. *Nat. Rev. Genet.* 5, 936–948. doi:10.1038/nrg1493.

Weiss, J. N. (1997). The Hill equation revisited: uses and misuses. *FASEB J.* 11, 835–841. doi:10.1096/fasebj.11.11.9285481.

Westerhoff, and Kell (2007). *The methodologies of systems biology*. Boogerd F.C. Hofmeyr J.-H.S. and Westerhoff H.V, B. F. J., editor: Systems Biology - Philosophical Foundations.

Wilmes, A., Bielow, C., Ranninger, C., Bellwon, P., Aschauer, L., Limonciel, A., et al. (2014). Mechanism of cisplatin proximal tubule toxicity revealed by integrating transcriptomics, proteomics, metabolomics and biokinetics. *Toxicol. Vitro Int. J. Publ. Assoc. BIBRA*. doi:10.1016/j.tiv.2014.10.006.

Wilmes, A., Limonciel, A., Aschauer, L., Moenks, K., Bielow, C., Leonard, M. O., et al. (2013). Application of integrated transcriptomic, proteomic and metabolomic profiling for the delineation of mechanisms of drug induced cell stress. *J. Proteomics* 79, 180–194. doi:10.1016/j.jprot.2012.11.022.

Wittwehr, C., Aladjov, H., Ankley, G., Byrne, H. J., de Knecht, J., Heinzle, E., et al. (2017). How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology. *Toxicol. Sci.* 155, 326–336. doi:10.1093/toxsci/kfw207.

Wolfinger, R. D., Gibson, G., Wolfinger, E. D., Bennett, L., Hamadeh, H., Bushel, P., et al. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* 8, 625–637. doi:10.1089/106652701753307520.

Xu, C., Li, C. Y.-T., and Kong, A.-N. T. (2005). Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Arch. Pharm. Res.* 28, 249–268.

Younes, M., Sharma, S. C., and Siegers, C. P. (1986). Glutathione depletion by phorone organ specificity and effect on hepatic microsomal mixed-function oxidase system. *Drug Chem. Toxicol.* 9, 67–73. doi:10.3109/01480548609042831.

Zgheib, E., Bechaux, C., Crepet, A., Mombelli, E., and Bois, F. Y. (2017). High-throughput methods for toxicology and health risk assessment. *Environ. Risques Santé*, 44–58. doi:10.1684/ers.2016.0943.

Zgheib, E., Limonciel, A., Jiang, X., Wilmes, A., Wink, S., van de Water, B., et al. (2018). Investigation of Nrf2, AhR and ATF4 Activation in Toxicogenomic Databases. *Front. Genet.* 9. doi:10.3389/fgene.2018.00429.

Zhang, Q., Pi, J., Woods, C. G., and Andersen, M. E. (2009). Phase I to II cross-induction of xenobiotic metabolizing enzymes: a feedforward control mechanism for potential hormetic responses. *Toxicol. Appl. Pharmacol.* 237, 345:356.

Zhang, W., Li, F., and Nie, L. (2010). Integrating multiple “omics” analysis for microbial biology: application and methodologies. *Microbiol. Read. Engl.* 156, 287–301. doi:10.1099/mic.0.034793-0.

Zhu, H., Zhang, J., Kim, M. T., Boison, A., Sedykh, A., and Moran, K. (2014). Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. *Chem. Res. Toxicol.* 27, 1643–1651. doi:10.1021/tx500145h.

7 APPENDIX – SUPPLEMENTARY MATERIAL

7.1 SUPPLEMENTARY INFORMATION FOR CHAPTER 3

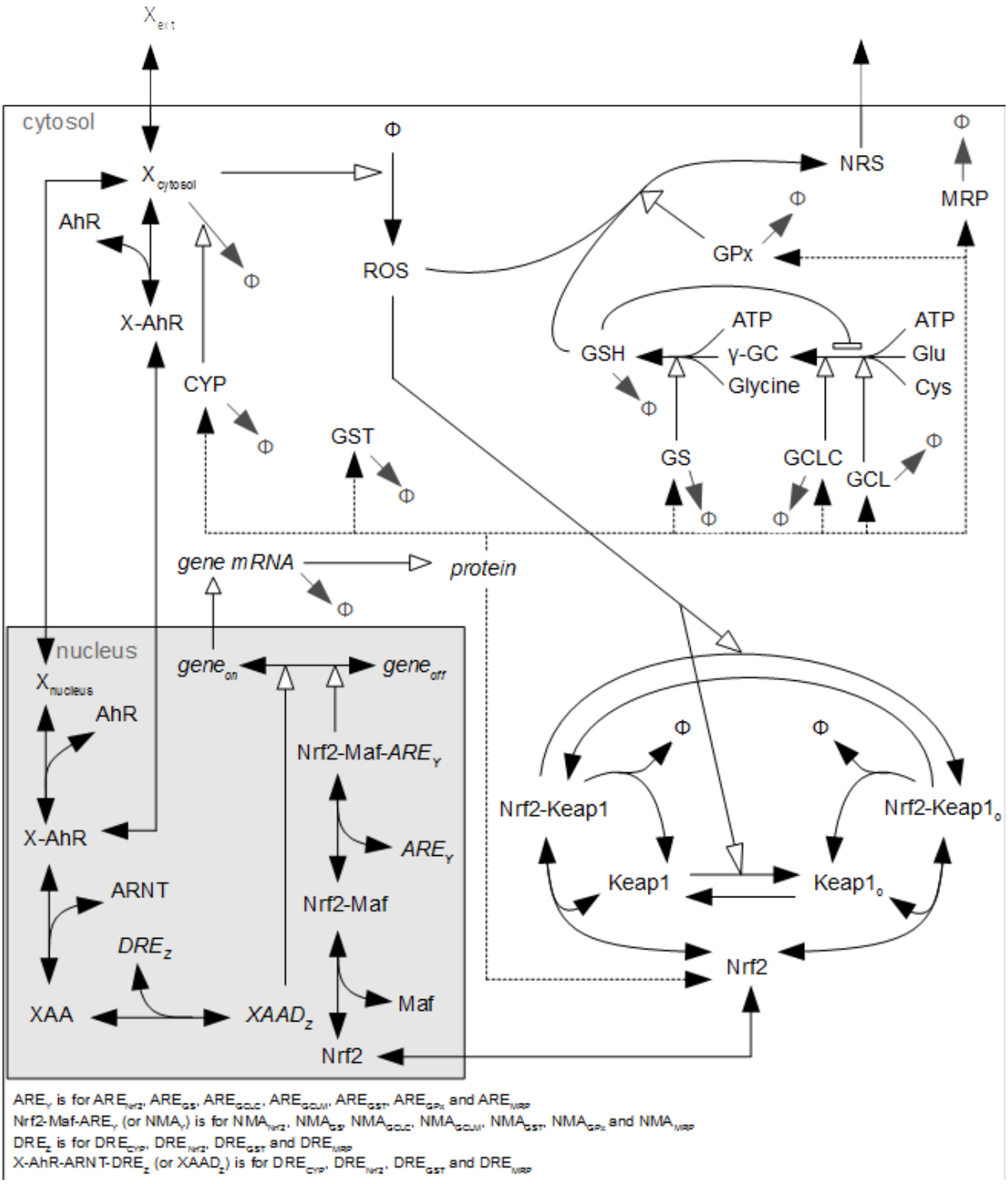


Figure S1. Schematic representation of the SB model of the Nrf2 signaling pathway by Hamon *et al.* (2014).

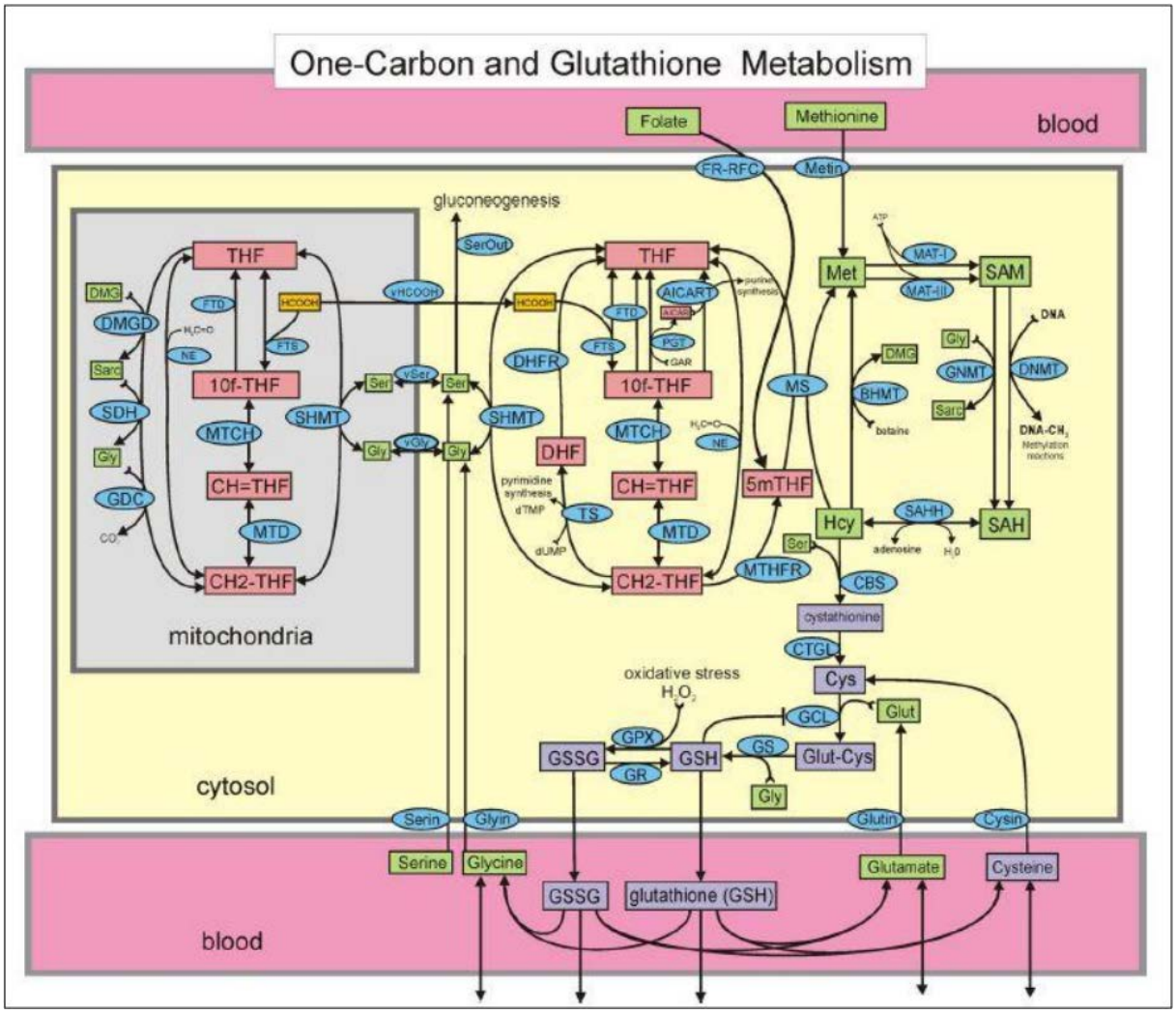


Figure S2. Schematic representation of the SB model of the GSH metabolism pathway by Reed *et al.* (2008).

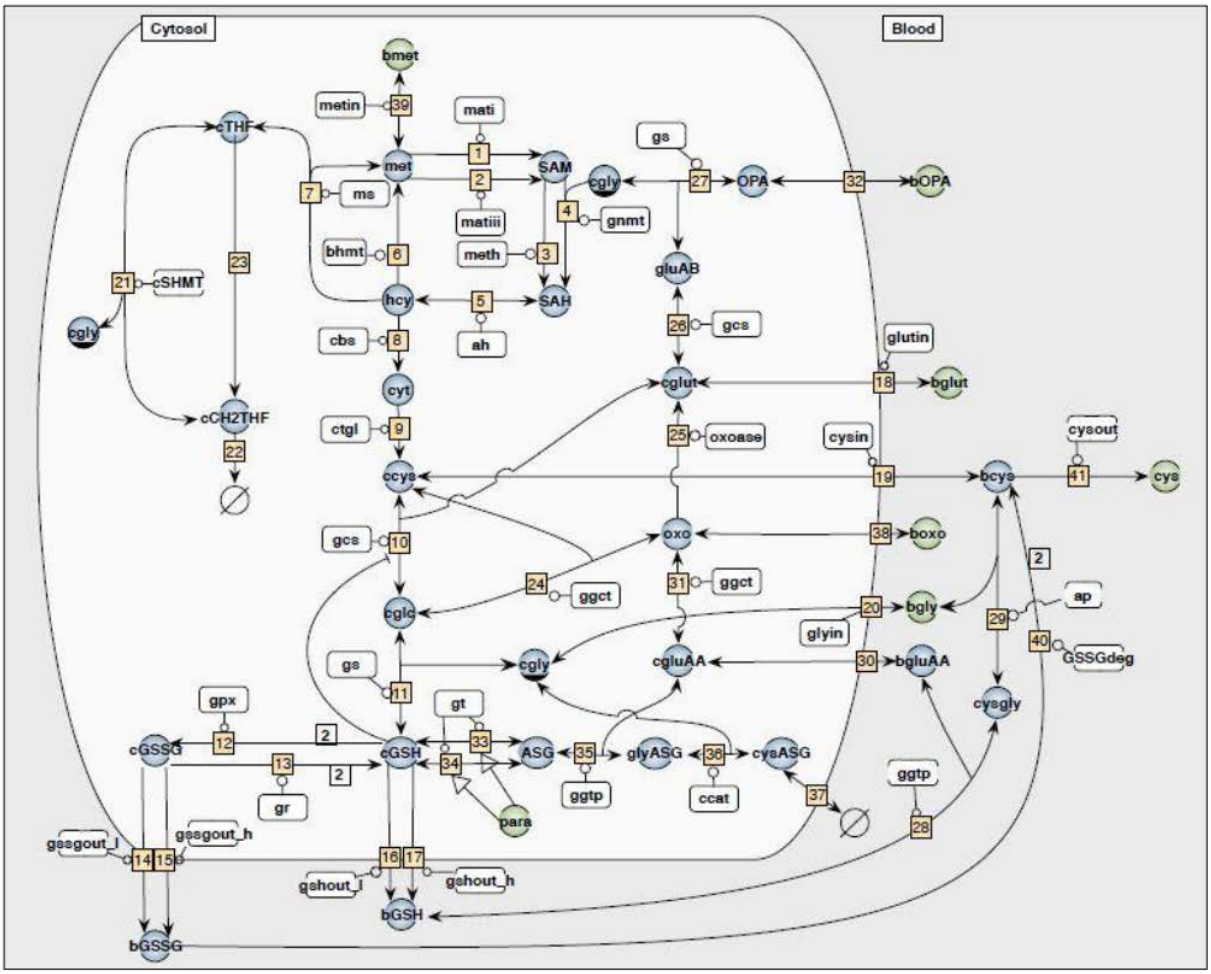


Figure S3. Schematic representation of the SB model of the GSH metabolism pathway by Geenen *et al.* (2012).

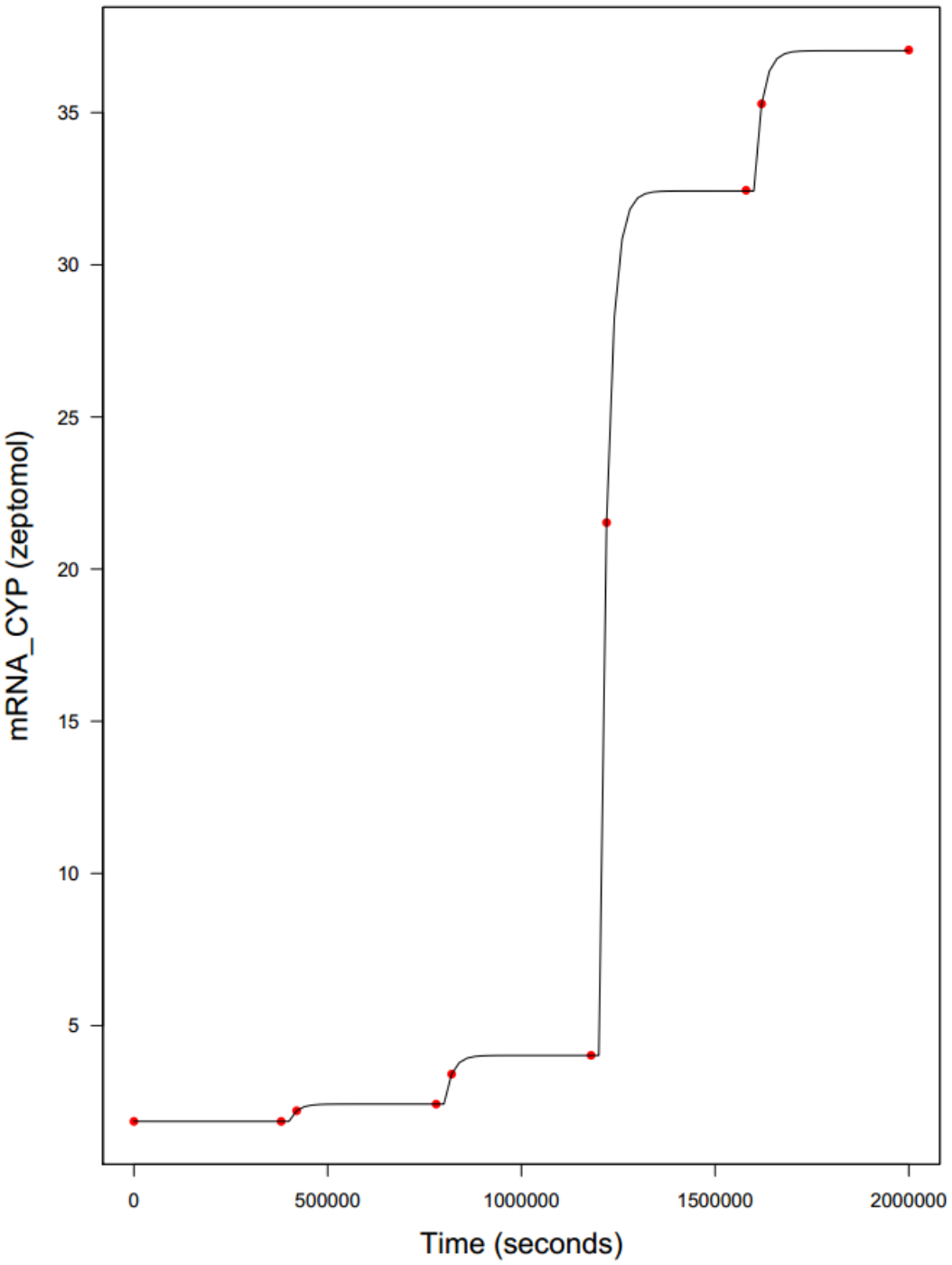


Figure S4. MCMC curve fitting of *CYP* mRNA (example of gene activated by one single activator) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (red dots) and Hill-based (black curve) SB models.

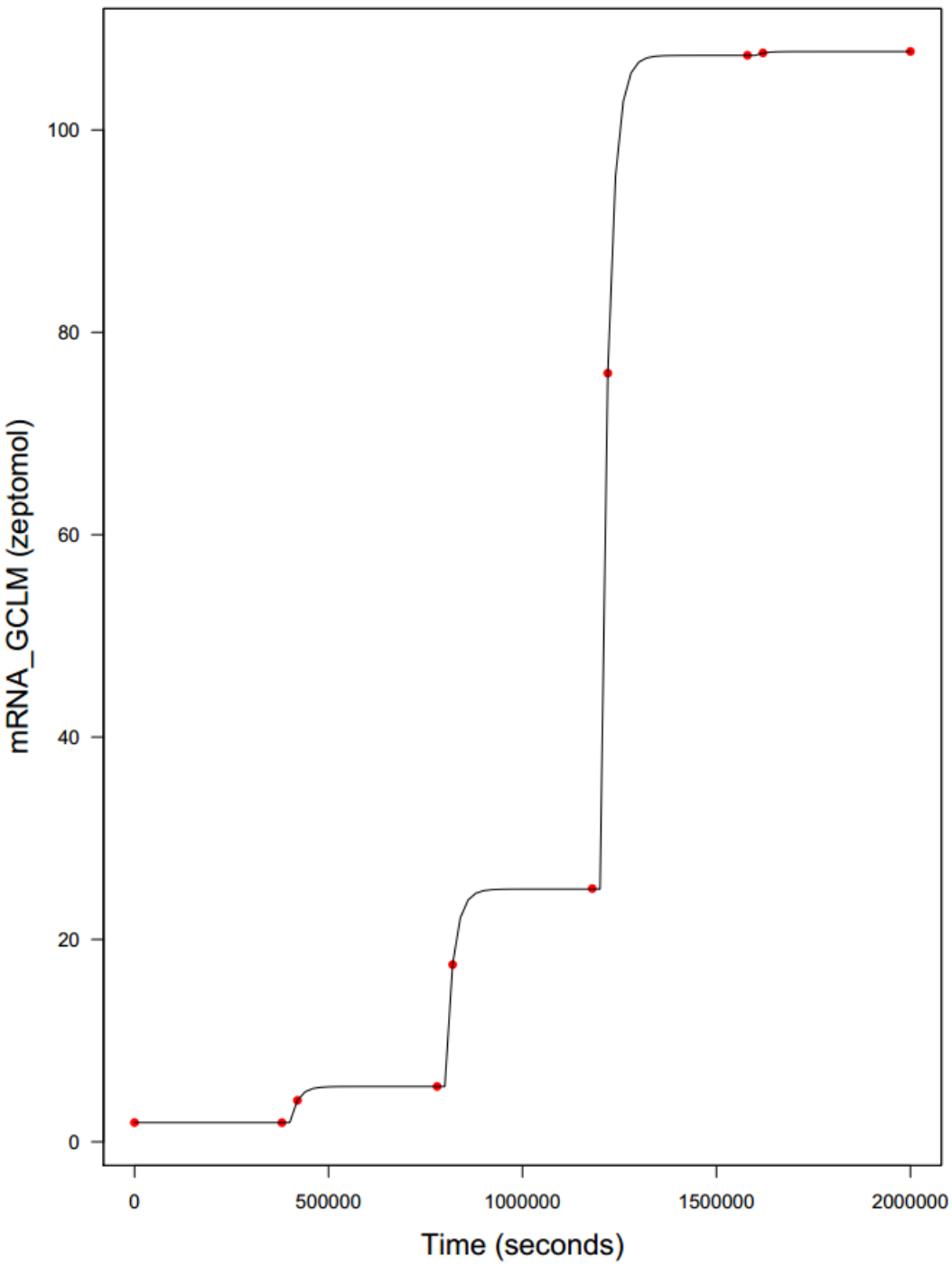


Figure S5. MCMC curve fitting of *GCLM* mRNA (example of gene activated by one single activator) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (red dots) and Hill-based (black curve) SB models.

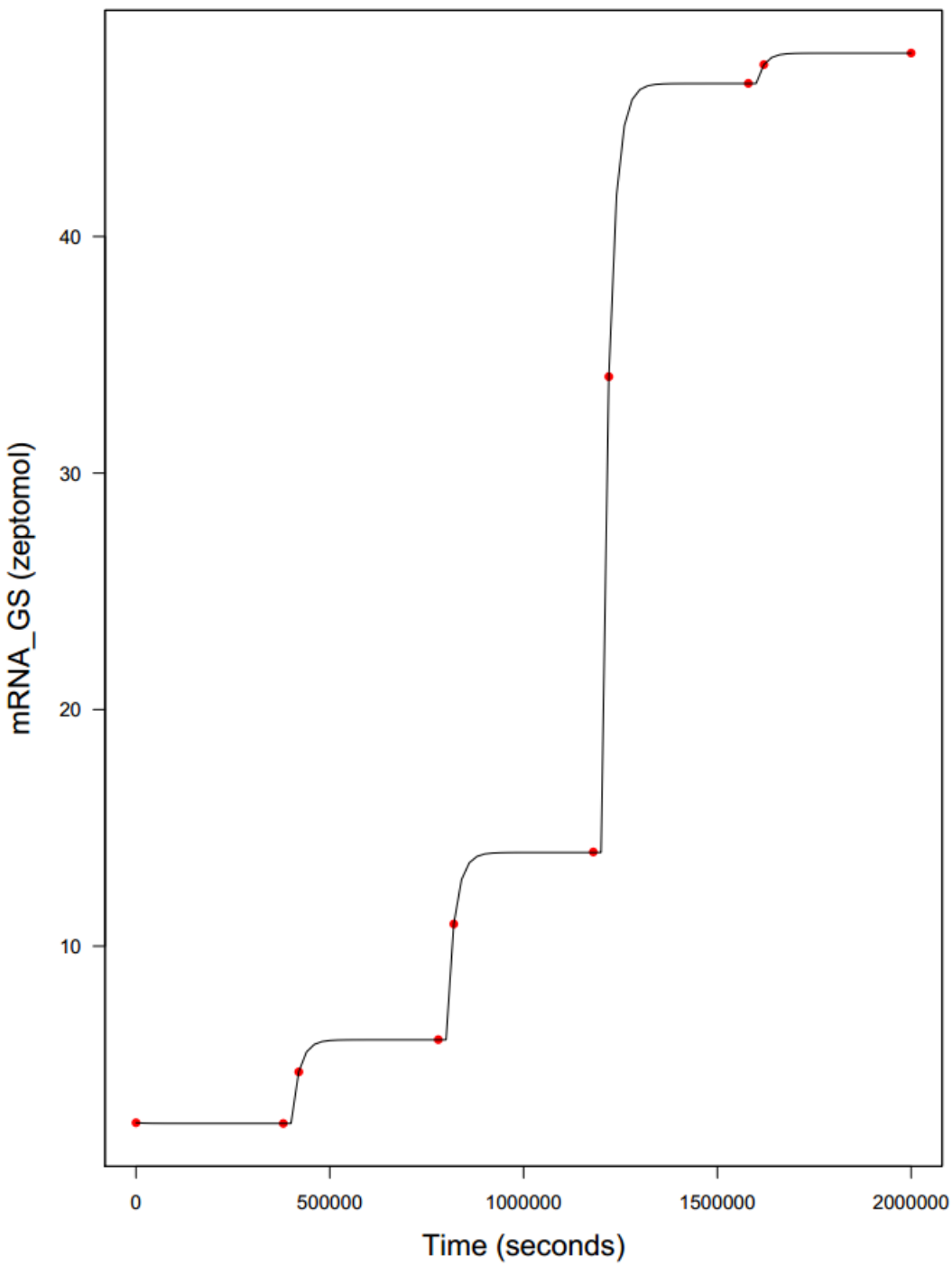
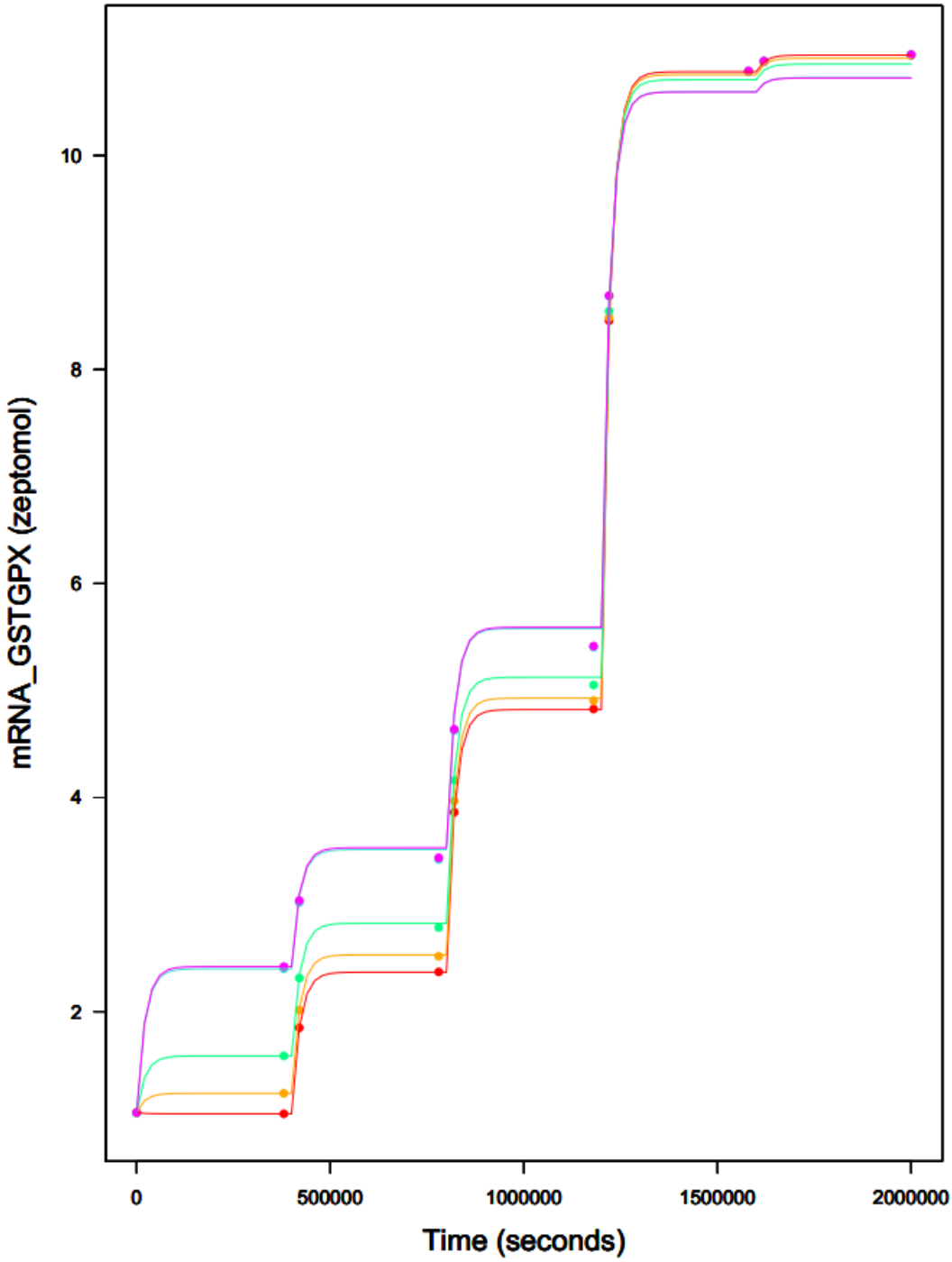


Figure S6. MCMC curve fitting of *GS* mRNA (example of gene activated by one single activator) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (red dots) and Hill-based (black curve) SB models.



Ta

Figure S7. MCMC curve fitting of *GST* and *GPX* mRNA (example of gene activated by two activators) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (coloured dots) and Hill-based (coloured curves) SB models. nucNrf2 dose increase is operated over time (every 400,000 seconds) and nucX-AhR dose is displayed on different curves (0 (red), 0.5 (orange), 1 (green), 10 (blue) and 100 (magenta) zeptomols of nucX-AhR).

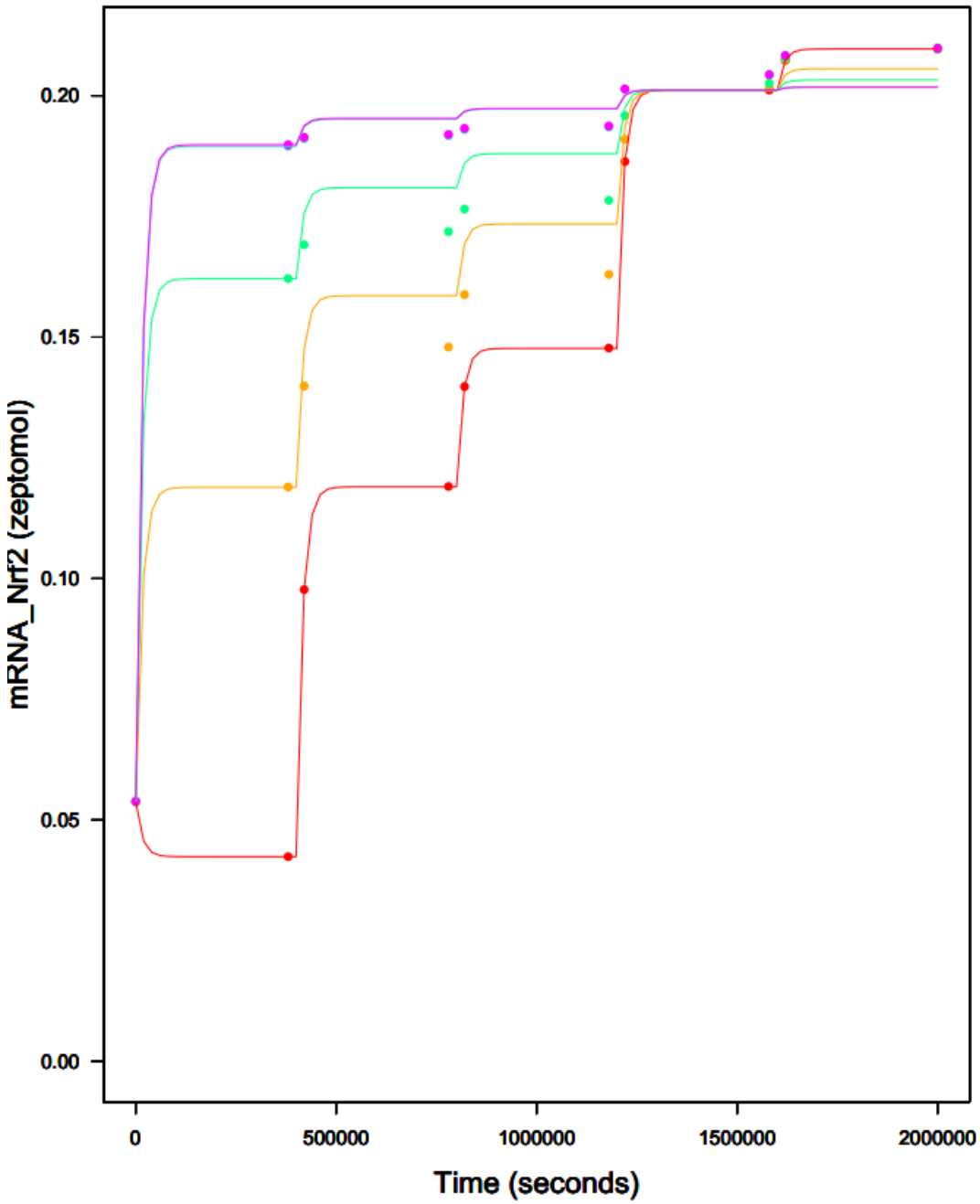


Figure S8. MCMC curve fitting of Nrf2 mRNA (example of gene activated by two activators) rate equivalency by time according to virtual exposure scheme presented in **Table 3** applied on both Hamon's (coloured dots) and Hill-based (coloured curves) SB models. nucNrf2 dose increase is operated over time (every 400,000 seconds) and nucX-AhR dose is displayed on different curves (0 (red), 0.5 (orange), 1 (green), 10 (blue) and 100 (magenta) zeptomols of nucX-AhR).

7.2 SUPPLEMENTARY INFORMATION FOR CHAPTER 4

The latest version of the article summarizing the study described in chapter 4, as well as the computational code of the constructed SB model and its corresponding input file, are submitted in three attached files under the names of 'Tools_qAOP_dev_Zgheib_etal.pdf', 'v7.11_Nrf2_GSH_KBrO3.model' and 'v7.11_Nrf2_GSH_KBrO3.in' respectively.

7.2.1 Experimental Data

Table S1. *In vitro* GSH depletion data used for the qAOP calibration.

KBrO3 concentration (mM)	Experiment	GSH (percent of control)
0	1	91.52
0.375	1	63.59
0.75	1	50.50
1.5	1	14.28
3	1	2.477
6	1	0.377
0	2	115.8
0.375	2	66.65
0.75	2	37.05
1.5	2	13.73
3	2	2.355
6	2	0.841
0	3	92.62
0.375	3	60.75
0.75	3	31.21
1.5	3	14.00
3	3	3.115
6	3	0.598

Table S2. *In vitro* DCF fluorescence data used for the qAOP calibration. Time is in hours, DCF fluorescence is in arbitrary relative fluorescence units (RFU). Eight experiments were performed at each $KBrO_3$ dose level.

Time	Control ($KBrO_3$ = 0)								0.75 mM $KBrO_3$								1.5 mM $KBrO_3$								3 mM $KBrO_3$								6 mM $KBrO_3$							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
0.000	2121	2441	2042	1978	2189	1826	1952	1897	2195	2144	1868	1808	1768	1734	1470	1643	2298	2223	2078	1843	1881	1629	1524	1696	2876	3041	2120	1880	1791	1723	1716	1606	8877	3631	2404	2220	1936	1733	1716	1522
0.267	2336	2570	2125	2092	2371	1962	2076	1988	2534	2488	2160	2075	2042	2002	1740	1887	2758	2668	2503	2223	2320	1991	1808	2018	3461	3644	2654	2364	2341	2209	2257	2028	9446	4261	2945	2813	2571	2255	2205	2014
0.517	2439	2689	2274	2244	2521	2063	2167	2070	2842	2897	2526	2406	2305	2306	1922	2168	3157	3094	2911	2619	2755	2310	2156	2426	4002	4178	3167	2815	2762	2654	2733	2431	9984	4956	3489	3474	3102	2772	2739	2502
0.767	2604	2809	2361	2348	2634	2118	2273	2123	3176	3235	2801	2698	2613	2612	2171	2433	3609	3527	3307	3013	3114	2650	2501	2740	4529	4767	3626	3324	3266	3146	3135	2795	10510	5720	3996	4091	3685	3225	3234	2909
1.000	2707	2982	2456	2472	2779	2183	2401	2212	3520	3576	3106	2898	2871	2842	2346	2667	3961	3964	3755	3367	3482	2945	2764	3078	5004	5309	4142	3763	3642	3478	3576	3234	11020	6333	4624	4672	4204	3762	3655	3308
1.267	2751	3128	2559	2511	2770	2213	2442	2286	3840	3945	3393	3177	3095	3059	2581	2898	4352	4456	4076	3703	3805	3290	3038	3423	5477	5795	4599	4230	4079	3894	3983	3557	11470	7059	5107	5226	4712	4160	4149	3656
1.517	2940	3205	2569	2655	2903	2350	2495	2364	3912	4193	3650	3405	3390	3296	2819	3126	4668	4856	4361	4054	4239	3551	3276	3739	5941	6392	4964	4508	4543	4226	4339	3913	11911	7614	5514	5709	5192	4661	4475	4102
1.70	2986	3386	2670	2689	3018	2352	2612	2447	4256	4442	3862	3588	3582	3506	2984	3258	4994	5166	4751	4225	4453	3835	3517	3780	6302	6700	5260	4836	4759	4562	4644	4128	12510	8062	5931	6229	5553	5031	4870	4383
1.95	3071	3402	2766	2851	3087	2415	2666	2498	4599	4684	4071	3818	3793	3736	3125	3388	5310	5483	5074	4524	4718	4032	3748	4092	6747	7127	5806	5120	5244	4769	5003	4425	12743	8744	6314	6581	6067	5338	5168	4696
2.20	3089	3464	2834	2852	3195	2451	2783	2544	4794	5084	4368	4043	3879	3914	3315	3618	5644	5726	5165	4794	4948	4265	3972	4362	7236	7621	6173	5506	5577	5117	5344	4741	13303	9318	6861	7216	6371	5684	5528	5011
2.45	3165	3603	2860	2968	3317	2531	2851	2648	5098	5296	4603	4293	4164	4104	3436	3797	5986	6070	5700	5112	5292	4462	4096	4602	7579	7980	6512	5910	5977	5498	5714	4922	13991	9973	7221	7697	6873	6064	5825	5220
2.70	3220	3650	2892	3033	3296	2567	2908	2708	5295	5520	4853	4322	4256	4140	3587	3936	6178	6433	5871	5386	5555	4673	4421	4830	7958	8451	6916	6159	6155	5759	5971	5195	14257	10483	7534	8047	7233	6382	6288	5589
2.95	3296	3728	2988	3070	3484	2614	2930	2672	5407	5797	4963	4658	4558	4364	3794	4056	6529	6607	6115	5614	5699	4926	4623	5023	8315	8726	7202	6478	6605	5929	6252	5396	14797	10911	7986	8452	7655	6652	6490	5734
3.20	3357	3905	2997	3132	3493	2667	3001	2783	5730	6021	5165	4797	4701	4539	3831	4149	6751	6951	6404	5848	5977	5010	4750	5249	8603	9060	7492	6779	6715	6055	6491	5641	14889	11393	8241	8799	7864	6969	6803	6120
3.45	3289	3821	3099	3195	3583	2749	3014	2795	5795	6334	5351	4901	4886	4712	4036	4269	7013	7245	6561	5923	6246	5271	4992	5337	8825	9570	7812	6974	6883	6515	6646	5879	15493	11723	8635	9157	8143	7219	7078	6204
3.70	3430	3978	3146	3221	3670	2775	3142	2855	6124	6459	5508	5144	5012	4846	4123	4474	7340	7577	6798	6224	6425	5382	5072	5572	9294	9786	8020	7281	7395	6765	7008	6068	15886	12175	8838	9305	8511	7488	7211	6534
3.95	3452	4017	3156	3259	3581	2801	3086	2831	6246	6661	5748	5246	5128	5010	4176	4579	7495	7686	6975	6353	6525	5597	5274	5673	9467	10072	8219	7466	7495	7028	7136	6268	15989	12569	9281	9772	8816	7725	7575	6722
4.20	3508	4041	3257	3375	3684	2813	3207	2874	6373	6609	5782	5367	5315	5190	4246	4726	7629	7911	7269	6528	6698	5758	5440	5936	9721	10287	8607	7584	7756	7154	7264	6406	16313	12958	9555	10010	9025	7916	7899	6828
4.45	3601	4121	3264	3387	3796	2837	3210	2864	6577	7051	6038	5592	5376	5248	4493	4840	7903	8147	7477	6648	6947	5835	5519	6114	10059	10529	8695	7962	8018	7359	7590	6519	16528	13211	9731	10408	9243	8087	7988	7084
4.70	3673	4214	3266	3445	3827	2920	3233	2926	6813	7128	6186	5662	5568	5404	4567	5037	7945	8278	7657	6899	7122	6048	5728	6277	10170	10914	9040	8066	8051	7403	7802	6711	16738	13475	9950	10531	9502	8437	8220	7232
4.95	3722	4123	3325	3394	3844	2946	3287	2973	6978	7237	6269	5777	5645	5494	4607	5036	8263	8531	7810	7018	7291	6222	5853	6304	10346	11162	9139	8329	8335	7540	7981	6945	17063	13793	10155	10910	9713	8571	8417	7498
5.20	3718	4263	3379	3471	3902	3003	3299	3028	7115	7444	6383	5938	5799	5573	4838	5170	8366	8710	7862	7164	7479	6340	5824	6518	10632	11421	9327	8413	8561	7908	8073	7074	17407	13945	10310	11141	10036	8738	8588	7573
5.45	3715	4286	3403	3551	3916	2939	3321	3060	7202	7592	6477	5991	5946	5614	4879	5220	8656	8883	8191	7273	7601	6409	5940	6604	10935	11545	9572	8555	8665	8032	8279	7147	17618	14379	10626	11197	10296	8934	8833	7700
5.70	3702	4225	3345	3596	3953	2967	3405	3055	7380	7771	6618	6159	5960	5884	4906	5359	8795	9044	8334	7490	7701	6554	6136	6718	11034	11758	9743	8839	8904	8060	8411	7291	17656	14444	10933	11387	10407	9142	8929	7948
5.95	3805	4339	3475	3637	4028	3045	3411	3052	7460	7645	6763	6252	6168	5961	5074	5455	8910	9281	8469	7627	7843	6673	6342	6883	11328	11881	9945	8906	8841	8330	8579	7440	17763	14811	11016	11670	10627	9243	9182	8028
6.20	3805	4474	3411	3602	4054	3040	3431	3147	7629	8055	6909	6308	6185	6033	5030	5558	8967	9367	8482	7764	7938	6748	6359	6969	11458	12166	10142	9050	9226	8454	8745	7601	18203	14980	11151	11927	10730	9475	9185	8078
6.45	3932	4456	3470	3641	4087	3045	3448	3125	7678	8217	7093	6548	6289	6135	5244	5618	9317	9630	8759	7920	8056	6898	6404	6934	11691	12264	10380	9199	9295	8538	8944	7659	18196	15099	11284	12149	10790	9505	9491	8245
6.70	3944	4435	3554	3633	4082	3090	3459	3175	7938	8375	7234	6605	6363	6373	5330	5704	9291	9472	8949	8043	8238	6958	6572	7169	11751	12434	10500	9332	9398	8655	8999	7875	18527	15288	11404	12255	11034	9873	9545	8350

6.95	4024	4461	3505	3746	4123	3076	3490	3191	8016	8464	7214	6620	6496	6308	5295	5743	9426	9660	9024	8111	8348	7111	6722	7298	11777	12631	10492	9498	9475	8785	9156	7845	18360	15501	11616	12405	11293	9810	9558	8511
7.20	4002	4562	3516	3733	4229	3122	3523	3161	7999	8609	7345	6765	6599	6310	5390	5875	9612	9963	9157	8279	8444	7099	6763	7364	12109	12796	10467	9614	9715	8921	9245	8076	18769	15739	11812	12684	11465	10004	9814	8692
7.45	4013	4648	3601	3745	4263	3158	3516	3185	8180	8688	7413	6853	6649	6467	5467	5781	9711	10076	9304	8413	8649	7376	6839	7477	12229	13022	10885	9768	9808	8997	9371	8087	18887	16139	12125	12975	11524	10088	10082	8697
7.70	4130	4593	3574	3837	4202	3181	3598	3251	8275	8787	7553	6919	6750	6503	5544	5942	9941	10240	9341	8464	8629	7383	6902	7609	12514	13235	11009	9792	9934	9255	9469	8224	19048	16071	12127	12979	11712	10259	10086	8812
7.95	4129	4576	3616	3844	4251	3171	3644	3236	8433	8842	7694	6935	6883	6661	5642	6050	10062	10366	9653	8594	8825	7513	7040	7671	12634	13360	11298	9980	10015	9312	9719	8371	19217	16462	12119	13044	11958	10454	10208	8984
8.20	4099	4642	3619	3798	4256	3219	3642	3255	8346	8842	7820	7266	6943	6773	5732	6173	10070	10584	9637	8651	8856	7648	7128	7833	12769	13524	11339	10076	10249	9493	9701	8252	19345	16446	12393	13143	12005	10487	10400	9036
8.45	4175	4672	3685	3856	4321	3268	3669	3341	8639	9174	7854	7275	6925	6940	5794	6093	10033	10597	9650	8784	9067	7642	7170	7853	12664	13750	11242	10287	10314	9415	9943	8443	19505	16690	12267	13385	12132	10463	10475	9161
8.70	4214	4701	3712	3914	4330	3279	3665	3339	8758	9165	7928	7184	7074	6839	5753	6290	10334	10882	9805	8942	9193	7838	7312	7927	12793	13739	11632	10521	10478	9675	10040	8546	19662	16960	12659	13470	12217	10567	10551	8965
8.95	4207	4711	3699	3894	4364	3276	3596	3381	8827	9398	7990	7380	7049	6962	5800	6349	10348	10847	9843	8949	9178	7785	7187	8071	13173	13900	11540	10481	10465	9700	10149	8743	19750	17013	12731	13754	12323	10746	10727	9329
9.20	4218	4774	3692	3964	4376	3298	3671	3317	8829	9203	8099	7326	7167	6934	5872	6488	10668	10895	10111	9008	9317	7906	7362	8027	12923	13990	11741	10637	10661	9826	10153	8728	19843	17091	12856	13767	12386	10956	10727	9349
9.45	4246	4776	3724	3926	4453	3307	3676	3344	8937	9528	8049	7466	7263	7097	5923	6452	10435	11060	9960	9206	9455	7871	7461	8145	13196	14048	11845	10755	10522	10053	10284	8810	19836	17485	12976	13811	12505	10990	10739	9366
9.70	4175	4824	3724	3982	4445	3315	3785	3380	9043	9557	8252	7590	7413	7191	6063	6425	10638	11058	9959	9224	9414	7975	7529	8221	13450	14388	12004	10811	10869	10096	10248	8785	19911	17475	13158	13992	12677	11098	11136	9456
9.95	4210	4962	3760	4022	4416	3331	3789	3443	9107	9733	8334	7694	7464	7193	6118	6658	10814	11218	10299	9365	9505	8036	7550	8381	13328	14466	11880	10767	10955	10129	10441	8986	20065	17701	12938	14016	12793	11145	11076	9583
10.20	4242	4919	3786	4030	4545	3313	3797	3423	9328	9854	8454	7586	7463	7400	6136	6602	10752	11403	10198	9364	9599	8207	7615	8267	13653	14563	12075	10822	11051	10123	10507	9011	20293	17549	13219	14175	13050	11172	11139	9401
10.45	4286	4845	3754	4042	4519	3371	3782	3442	9312	9838	8462	7684	7502	7394	5976	6673	10915	11239	10298	9497	9714	8089	7696	8409	13590	14638	12077	10991	11158	10223	10588	8984	19984	17749	13395	14288	13020	11221	11193	9599
10.70	4308	4879	3899	4064	4584	3397	3801	3513	9291	9848	8591	7693	7617	7241	6262	6739	11074	11560	10475	9598	9807	8277	7875	8443	13893	14837	12341	11068	11047	10236	10701	9211	20252	18076	13499	14346	13103	11397	11284	9813
10.95	4333	4919	3904	4127	4608	3406	3843	3453	9273	10089	8650	7669	7621	7470	6325	6773	11143	11572	10525	9681	9883	8395	7821	8505	13970	14716	12388	11105	11243	10318	10824	9109	20302	17938	13602	14557	13277	11418	11474	9840
11.20	4312	4958	3780	4136	4687	3401	3883	3462	9493	10041	8760	7923	7718	7622	6348	6719	11281	11718	10539	9654	9809	8294	7927	8662	13843	14915	12511	11110	11335	10478	10788	9435	20660	18150	13677	14633	13514	11568	11475	9916
11.45	4389	5041	3930	4119	4602	3419	3844	3466	9586	10202	8744	8031	7722	7553	6374	6784	11346	11742	10671	9709	10090	8351	7928	8882	14086	15116	12339	11043	11466	10526	10750	9339	20414	18283	13656	14685	13371	11534	11462	9856
11.70	4354	4950	3856	4172	4647	3501	3853	3571	9637	10240	8851	8012	7861	7687	6353	7009	11335	11747	10645	9907	10091	8610	8017	8820	14034	15030	12530	11480	11474	10711	11000	9474	20503	18256	13804	14813	13710	11742	11593	10048
11.95	4347	5101	3981	4201	4616	3482	3875	3546	9569	10320	8839	8026	7818	7635	6545	7033	11488	11934	10869	9817	10341	8691	7974	8835	14097	15171	12766	11617	11589	10699	11076	9543	20935	18502	13878	14857	13593	11719	11636	10073
12.20	4405	4972	3921	4067	4686	3444	3843	3531	9717	10353	8917	8110	7905	7750	6458	7179	11397	11862	10900	9948	10283	8591	8066	8893	14159	15296	12795	11533	11628	10678	11188	9634	20660	18586	13912	15102	13791	11865	11636	10158
12.45	4412	5084	3932	4145	4712	3472	3884	3535	9833	10487	9068	8209	8080	7813	6522	7148	11725	12144	10943	10117	10173	8720	8162	8974	14022	15539	12717	11509	11655	10897	11238	9546	20840	18625	14067	15043	13701	11796	11682	10179
12.70	4441	5146	3946	4168	4674	3412	3896	3472	9569	10214	8954	8293	8049	7854	6598	7103	11680	12124	11136	9912	10410	8754	8161	8955	14393	15342	12724	11679	11725	10885	11255	9761	20852	19062	14156	15122	13790	12006	11798	10251
12.95	4501	5086	4015	4299	4725	3497	3941	3559	9735	10625	9155	8440	8071	7985	6768	7121	11988	12227	11402	10240	10433	8941	8396	9127	14485	15430	13034	11783	11771	11196	11391	9792	21075	19173	14331	15282	13976	12184	12093	10383
13.20	4368	5182	3973	4294	4759	3484	3943	3521	9991	10661	9126	8254	8206	8050	6741	7248	11849	12283	11081	10134	10527	8899	8305	9078	14266	15601	13031	11882	11873	11016	11406	9780	20893	19181	14402	15460	14038	11980	12028	10077
13.45	4547	5239	4043	4204	4729	3579	3988	3634	10117	10924	9305	8525	8333	8158	6653	7362	12252	12557	11314	10321	10744	9096	8315	9290	14608	15969	13053	12074	12123	11215	11391	10017	21163	19375	14373	15669	14088	12269	12265	10470
13.70	4551	5200	4020	4237	4767	3624	4033	3628	10056	10679	9228	8461	8244	8160	6822	7334	11953	12463	11449	10140	10506	9109	8500	9330	14709	15845	13116	12043	12013	11070	11488	9943	21116	19204	14409	15663	14170	12187	12203	10492
13.95	4560	5208	3984	4274	4845	3529	4024	3641	10211	10869	9479	8595	8423	8279	6859	7370	12219	12602	11449	10295	10691	9143	8273	9175	14461	15854	13109	12039	12174	11313	11685	10050	21217	19714	14394	15696	14299	12347	12261	10558
14.20	4452	5210	4103	4241	4884	3569	4002	3659	10332	10931	9268	8564	8328	8218	6843	7480	12518	12673	11585	10393	10745	9088	8504	9327	14755	15882	13218	12120	12134	11265	11715	9990	21339	19490	14728	15827	14494	12509	12258	10471
14.45	4509	5299	4010	4259	4771	3639	4003	3661	10344	11082	9581	8576	8486	8280	6871	7593	12420	12626	11593	10518	10773	9072	8494	9351	14409	15959	13375	12025	12355	11277	11765	10114	21348	19722	14709	15765	14411	12647	12498	10665
14.70	4570	5363	4120	4234	4873	3572	4047	3661	10270	10956	9516	8620	8444	8280	6866	7390	12340	12713	11540	10605	10909	9124	8699	9555	14911	16090	13579	12138	12382	11332	11901	10145	21302	19919	14927	15849	14551	12682	12500	10655
14.95	4588	5325	4102	4426	4944	3621	4034	3651	10233	11100	9599	8780	8482	8344	6948	7607	12321	12653	11781	10555	11015	9370	8642	9481	14879	16096	13619	12375	12449	11573	11886	10262	21170	19790	14839	16172	14540	12634	12515	10751
15.20	4643	5284	4103	4253	4947	3665	4095	3703	10405	10989	9664	8698	8521	8336	7015	7638	12470	12899	11608	10668	11014	9319	8755	9721	14902	16342	13474	12295	12461	11514	11860	10175	21115	19860	15118	16106	14625	12858	12699	10858
15.45	4621	5387	4022	4268	4904	3639	4124	3700	10549	11193	9750	8826	8357	8504	6977	7763	12407	13057	11583	10597	10997	9356	8725	9624	14996	15999	13611	12445	12641	11555	11925	10334	21444	19920	15156	16299	14777	13052	12695	10732

15.70	4602	5415	4126	4341	4959	3640	4133	3756	10611	11268	9783	8880	8625	8552	7087	7750	12771	13068	11792	10786	11169	9369	8857	9789	15072	16440	13809	12573	12657	11602	12149	10289	21519	20380	15199	16539	14904	13065	12842	10881
15.95	4626	5401	4128	4417	4950	3741	4141	3714	10555	11417	9824	8951	8596	8565	7065	7772	12696	13023	11982	10732	11323	9533	8874	9642	15028	16469	13734	12471	12577	11701	12136	10274	21492	20336	15172	16426	14865	13070	12825	10916
16.20	4679	5459	4163	4455	4993	3728	4115	3745	10520	11527	9725	8931	8748	8655	7216	7811	12926	13013	11890	10821	11253	9541	8973	9921	15123	16541	13838	12538	12740	11765	12153	10411	21549	20574	15357	16620	15105	13238	13014	10977
16.45	4679	5428	4196	4477	4971	3707	4111	3731	10610	11377	10038	9035	8836	8625	7197	7984	12931	13344	12136	10900	11152	9534	8960	9917	15018	16941	14062	12665	12791	11651	12269	10500	21388	20315	15517	16714	15127	13250	13144	11017
16.70	4697	5392	4224	4448	5031	3622	4212	3828	10657	11663	9995	9072	8808	8660	7206	7863	12828	13206	12064	10902	11430	9685	8977	9937	15133	16767	13859	12716	12752	11782	12315	10544	21640	20597	15462	16513	15111	13191	13234	11011
16.95	4807	5378	4162	4470	4985	3756	4216	3834	10803	11404	10100	9084	8793	8718	7226	7967	13132	13519	12226	11087	11395	9707	9029	9958	15321	16666	13974	12585	12793	11850	12355	10659	21461	20679	15645	16906	15350	13315	13366	11177
17.20	4828	5449	4168	4474	5119	3748	4217	3807	10805	11663	10039	9132	8975	8803	7274	7993	13141	13478	12112	11131	11390	9563	9075	9858	15460	16522	14057	12824	12872	11994	12316	10586	21713	20844	15669	16953	15272	13285	13319	11023
17.45	4688	5419	4265	4477	4985	3751	4228	3920	10856	11656	10190	9254	8816	8877	7404	8027	12795	13523	12166	11205	11548	9726	8954	10105	15319	16918	14260	13007	12991	12052	12453	10608	21763	20734	15745	16691	15512	13475	13385	11188
17.70	4800	5517	4246	4453	5115	3767	4288	3829	11010	11676	10054	9225	9072	8775	7392	8016	12909	13471	12172	11129	11566	9800	9209	10044	15328	16955	14276	12869	12929	12008	12495	10709	21652	20670	15723	16734	15593	13273	13591	11412
17.95	4717	5574	4295	4554	5067	3729	4300	3820	10988	11829	10155	9327	8975	9079	7504	8067	13190	13614	12356	11148	11702	9597	9275	10033	15406	16895	14136	13088	13011	12146	12538	10483	21740	21251	15671	17023	15470	13530	13611	11268
18.20	4767	5544	4264	4519	5059	3759	4311	3862	10880	11979	10159	9339	9062	8952	7542	8170	12931	13684	12351	11154	11651	9921	9278	10164	15337	17131	14114	13083	13252	12140	12545	10759	21733	21012	15848	17267	15550	13551	13691	11319
18.45	4814	5593	4205	4582	5029	3773	4370	3811	11118	11957	10439	9387	9190	9114	7517	8184	13018	13737	12579	11233	11835	9909	9342	10397	15566	17344	14662	13151	13234	12407	12733	10754	21897	21373	15979	17183	15659	13591	13788	11441
18.70	4736	5633	4197	4606	5131	3806	4247	3920	10993	12055	10349	9457	9124	8954	7416	8233	13330	13974	12563	11267	11733	9915	9343	10389	15421	17231	14124	13138	13228	12274	12618	10786	21878	21226	16092	17222	15884	13508	13760	11374
18.95	4813	5622	4273	4603	5212	3801	4419	3934	11078	12010	10511	9560	9207	9120	7569	8166	13300	14016	12633	11427	11853	10058	9441	10390	15439	17253	14426	13410	13226	12383	12783	10953	21829	21320	16054	17393	15784	13491	13687	11581
19.20	4858	5649	4363	4618	5228	3891	4384	3929	11223	12159	10455	9521	9277	9235	7714	8321	13370	14198	12620	11533	11935	10145	9471	10464	15566	17382	14655	13307	13439	12389	12990	11015	22097	21514	16288	17494	15888	13983	14036	11787
19.45	4891	5607	4306	4636	5258	3889	4366	3986	11245	12234	10466	9594	9340	9222	7639	8351	13481	14020	12492	11597	12057	10098	9584	10431	15533	17603	14637	13331	13527	12312	12884	10950	21936	21506	16145	17282	15719	14015	13834	11681
19.70	4765	5580	4322	4586	5182	3889	4330	3951	11275	12395	10711	9630	9365	9320	7749	8498	13411	14147	12834	11614	12055	10035	9711	10509	15470	17622	14631	13356	13434	12477	12993	10912	22207	21771	16470	17533	16076	14120	13916	11802
19.95	4866	5589	4331	4580	5298	3882	4436	3925	11404	12280	10602	9585	9474	9142	7520	8329	13391	14126	12865	11506	12186	10133	9606	10588	15785	17512	14707	13505	13478	12603	12903	11035	22011	21557	16159	17497	16042	14083	14142	11829
20.20	4840	5697	4412	4679	5313	3852	4434	3917	11247	12259	10746	9750	9539	9375	7787	8479	13231	14236	12866	11473	11951	10345	9658	10609	15532	17537	14830	13390	13611	12399	13024	10989	21722	21780	16348	17619	16008	14153	14244	11936
20.45	4815	5615	4421	4786	5325	3871	4457	3943	11205	12460	10757	9533	9307	9445	7742	8490	13451	14324	12783	11460	12222	10344	9698	10728	15626	17886	14680	13545	13613	12480	12849	10933	21833	21848	16090	17670	16196	14082	14210	11780
20.70	4886	5723	4403	4822	5263	4017	4439	3973	11257	12502	10951	9762	9564	9461	7733	8603	13368	14389	13012	11794	12249	10412	9777	10662	15726	17450	14696	13653	13804	12547	13329	11236	21947	21825	16611	18001	16409	14156	14182	11997
20.95	4899	5820	4384	4775	5260	3898	4436	4010	11228	12585	10832	9751	9388	9333	7869	8455	13532	14430	13139	11652	12195	10204	9704	10669	15781	17847	15012	13177	13688	12704	13246	11089	21898	22134	16511	17928	16298	14312	14388	11935
21.20	4894	5682	4415	4732	5313	3935	4557	4020	11444	12608	10922	9759	9600	9480	7910	8628	13548	14313	13101	11751	12303	10484	9717	10792	15430	17925	14938	13506	13948	12717	13154	11145	21756	21916	16537	17978	16386	14263	14411	12057
21.45	4873	5776	4385	4778	5327	4019	4537	4035	11240	12665	10921	9916	9577	9376	7888	8653	13586	14582	13217	11846	12411	10336	9983	10770	15757	17999	15065	13792	13879	12821	13365	11272	21766	22268	16630	18153	16618	14467	14554	12143
21.70	4974	5856	4369	4808	5396	4011	4594	3973	11438	12752	10978	10025	9645	9680	7869	8837	13674	14595	13289	11972	12485	10482	9771	10904	15788	18117	14905	13750	13933	12773	13325	11407	21983	22171	16845	18164	16586	14587	14792	12256
21.95	4885	5780	4436	4805	5318	3987	4530	4065	11339	12842	10898	9906	9656	9688	7887	8703	13830	14734	13205	11971	12381	10409	9984	10954	15639	18038	15095	13598	13901	12848	13403	11406	21890	22330	16930	18090	16593	14713	14659	12210
22.20	4915	5778	4537	4854	5348	4050	4542	4126	11539	12935	11156	9888	9798	9713	7895	8793	13816	14696	13389	12106	12529	10713	9998	10879	15573	17960	15284	14079	13961	12931	13353	11532	21752	22626	17083	18471	16731	14669	14868	12204
22.45	4917	5798	4444	4859	5364	4024	4520	4033	11241	12845	11015	9883	9668	9525	8028	8691	13807	14771	13427	12070	12506	10495	10083	10984	15538	18283	15244	13911	13840	12851	13394	11373	21668	22542	17073	18438	16779	14683	14819	12263
22.70	4959	5811	4483	4987	5345	4034	4550	4063	11506	12855	11185	9979	9844	9744	8098	8777	13893	14943	13235	11888	12442	10609	9986	11087	15927	18318	15233	13984	14177	13151	13405	11424	21840	22900	17053	18628	16859	14770	14875	12385
22.95	4946	5910	4512	4846	5459	4042	4597	4038	11533	12877	11197	10127	9903	9750	8089	8896	13909	14842	13193	12105	12564	10725	10129	11040	15633	17986	15261	13808	14020	13085	13648	11407	21490	22669	17068	18861	16735	14811	14856	12366
23.20	4955	5902	4462	4871	5380	4007	4608	4068	11631	12845	11367	10242	9959	9892	8131	8891	13959	14618	13434	12160	12754	10644	10035	11156	15679	18356	15292	13805	14157	13095	13450	11397	21073	22877	17143	18785	16860	14844	15017	12451
23.45	4981	5851	4461	4955	5458	4090	4634	4098	11665	13110	11333	10210	9901	9880	8160	8935	14009	15047	13412	12011	12672	10749	10164	11222	15994	18283	15575	14079	14354	13198	13685	11408	21539	23061	17350	18569	16990	14802	15030	12376
23.70	4943	5909	4482	4979	5507	4067	4632	4154	11669	13045	11276	10190	9972	9780	8132	8835	14088	14935	13400	12154	12773	10820	10152	11142	15857	18335	15398	13974	14268	13235	13698	11383	21308	22996	17166	18841	16893	14843	15159	12533
23.95	4987	5868	4532	4918	5506	4107	4635	4110	11682	13020	11459	10193	10029	10056	8186	9028	14040	15078	13425	12016	12676	10948	10189	11314	15832	18485	15494	14339	14307	13316	13739	11486	21595	22892	17399	18767	17061	15021	15245	12489
24.20	4954	5800	4578	5043	5528	4128	4663	4173	11600	13142	11543	10378	10064	10059	8353	8910	13885	15238	13522	12400	12877	10792	10206	11346	15919	18503	15510	14131	14583	13178	13881	11639	21397	23215	17544	18955	17311	15192	15249	12614

24.45	4939	5871	4597	4956	5506	4153	4659	4207	11802	13303	11363	10327	10041	10065	8300	9140	14156	14884	13606	12289	12852	10770	10286	11480	15740	17994	15427	14050	14571	13302	13688	11532	21274	23214	17551	18869	17051	15089	15318	12454
24.70	4870	5868	4584	4950	5469	4097	4739	4154	11626	13192	11273	10321	10129	9990	8407	9093	14095	15366	13727	12303	12917	10892	10319	11413	15662	18668	15558	14384	14472	13331	13985	11623	21173	23265	17553	18954	17184	14997	15435	12669
24.95	5024	5886	4654	4983	5520	4110	4770	4183	11736	13247	11599	10586	10154	9978	8122	9168	13981	15367	13858	12422	13088	11035	10347	11535	15621	18651	15547	14338	14534	13430	13930	11708	20844	23457	17699	19144	17400	15135	15484	12505
25.20	5043	6038	4613	5017	5537	4158	4775	4205	11904	13401	11670	10620	10115	10163	8325	9213	14161	15500	14025	12432	13105	11108	10369	11514	15644	18813	15771	14470	14560	13457	14116	11740	20997	23651	17814	19294	17375	15351	15404	12774
25.45	4988	5941	4717	5058	5500	4232	4799	4217	11762	13235	11736	10477	10307	10076	8459	9271	14181	15491	13787	12602	13041	11120	10447	11437	15901	18795	15742	14659	14632	13776	14157	11811	20770	23677	17656	19404	17163	15436	15669	12794

Table S3. *In vitro* lactate concentration data used for the qAOP calibration. Four experiments were performed at each $KBrO_3$ dose level.

Day	**4 mM $KBrO_3$**				**2 mM $KBrO_3$**				**1 mM $KBrO_3$**				**0.5 mM $KBrO_3$**				**0.25 mM $KBrO_3$**				**Control ($KBrO_3$ = 0)**			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
0	2.329	3.677	4.198	1.713	1.896	3.805	4.198	1.669	2.491	3.362	3.913	2.039	2.874	4.22	2.329	2.894	2.039	3.074	2.622	3.403	2.473	3.935	3.382	3.279
1	4.467	6.105	6.505	4.377	2.473	4.332	4.879	2.584	2.913	2.776	2.993	2.258	2.51	3.115	1.959	2.364	1.773	2.473	2.039	3.508	2.173	4.044	2.679	2.547
2	6.887	7.38	7.249	7.696	3.827	5.113	6.055	3.698	2.894	3.341	3.762	1.864	3.115	5.278	1.975	3.424	3.634	2.679	2.311	3.698	3.217	4.626	2.737	3.784
3	9.315	7.963	6.81	8.071	5.956	6.179	6.43	5.687	3.892	4.949	5.445	3.892	4.879	5.638	4.22	5.663	4.995	4.879	3.115	4.717	4.603	5.397	3.258	4.22

7.2.2 Statistical Dose-Response based qAOP Model

In the empirical dose-response approach, dose(-time)-response equations were fitted to data on the effect of $KBrO_3$ on GSH, DCF and lactate. With such data, linking chemical exposures to KEs, the corresponding equations need to be mathematically inverted to obtain chemical-independent KERs. Only the exposure to MIE relationship can be used as is. For example, if we have three data sets for the activity at dose D_X of chemical X on each KE of an AOP, we need to fit three dose-response equations:

$$KE_1 = f(D_X) \qquad (7.1)$$

$$KE_2 = g(D_X) \qquad (7.2)$$

$$KE_3 = h(D_X) \qquad (7.3)$$

The relationship between KE_1 and D_X is given directly by equation 7.1. However, the relationship between KE_1 and KE_2 needs to be derived from equations 7.1 and 7.2:

$$KE_2 = g(D_X) = g\left(f^{-1}(KE_1)\right) \qquad (7.4)$$

Where f^{-1} denotes the inverse of function f. Similarly, for the relationship between KE_3 and KE_2 we have:

$$KE_3 = h(D_X) = h\left(g^{-1}(KE_2)\right) \qquad (7.5)$$

For dose-time-response relationships, the principle is the same, with time as an extra variable in the above functions. However, in some cases the function may not be monotonic and therefore will not be invertible.

The relationship between the concentration of $KBrO_3$ (C_{KBrO3}) and the percentage of GSH (Pct_{GSH}) remaining *in vitro* after one hour, representing the MIE, was modeled with a modified exponential decrease equation (equation 7.6):

$$Pct_{GSH} = 100 \times exp\left(-k \cdot C_{KBrO_3}^{b}\right) \qquad (7.6)$$

Its parameters are the GSH degradation rate constant k, and power b (which increases the degradation rate if b>1).

The inverse of equation 7.6 is the equation 7.7:

$$C_{KBrO_3} = \left(\frac{log(100)-log(Pct_{GSH})}{k}\right)^{1/b} \qquad (7.7)$$

The relationship between C_{KBrO3}, time t and Q_{DCF} (representing the amount of oxidative stress) was modeled empirically by equation 7.8:

$$Q_{DCF} = A + B \cdot \left(1 + \delta - exp\left(-k_d \cdot C_{KBrO_3}\right)\right)\left(1 - exp(-k_t \cdot t)\right) \qquad (7.8)$$

Its parameters are A (baseline response), B (maximum increase above baseline), δ (maximum increase modulation by dose), k_d (dose coefficient), k_t (time coefficient).

The solution of equation 7.8 for C_{KBrO3} is:

$$C_{KBrO3} = log\left(\left(1 + \delta - \frac{Q_{DCF}-A}{B\cdot(1-exp(-k_t\cdot t))}\right)^{-k_d}\right) \qquad (7.9)$$

Replacing C_{KBrO3} in equation 7.8 by the expression given in equation 7.7, we obtain the following KER between Pct_{GSH} and Q_{DCF}.

$$Q_{DCF} = A + B \cdot \left(1 + \delta - exp\left(-k_d \cdot \left(\frac{log(100)-log(Pct_{GSH})}{k}\right)^{1/b}\right)\right)(1 - exp(-k_t \cdot t)) \qquad (7.10)$$

To model the C_{KBrO3} – time - lactate concentration (C_{lac}) relationship, we used a polynomial equation which adequately fitted the data:

$$C_{lac} = a + bC_{KbrO_3} + (c + eC_{KbrO3})t + (d + fC_{KbrO3})t^2 \qquad (7.11)$$

If we replace C_{KBrO3} in equation 7.11 by the value given in equation 7.8, the relationship between Q_{DCF}, time and C_{lac} becomes:

$$C_{Lac} = a + b \cdot log\left(\left(1 + \delta - \frac{Q_{DCF} - A}{B \cdot (1 - exp(-k_t \cdot t))}\right)^{-k_d}\right) + \left[c + e \cdot log\left(\left(1 + \delta - \frac{Q_{DCF} - A}{B \cdot (1 - exp(-k_t \cdot t))}\right)^{-k_d}\right)\right] t + \left[d + f \cdot log\left(\left(1 + \delta - \frac{Q_{DCF} - A}{B \cdot (1 - exp(-k_t \cdot t))}\right)^{-k_d}\right)\right] t^2 \quad (7.12)$$

A relationship (even more complex) between GSH and lactate concentration could be obtained by replacing Q_{DCF} by Pct_{GSH}, using equation 7.10.

For parameter estimation, a Metropolis-Hastings MCMC algorithm was used, as implemented in the *GNU MCSim* software (Bois, 2009a). Two Markov chains of 50,000 iterations were run in parallel, keeping one in four of the last 40,000 iterations. For each estimated parameter, non-informative uniform prior distributions were used (note that the boundaries of those prior distributions were never reached) (see **Table S4**). As usually done for measurements at different concentrations, the data were considered to be log-normally distributed with geometric means given by the corresponding model predictions and geometric standard deviations (σ_{GSH}, σ_{DCF}, and σ_{lac}), sampled from half-normal distributions (with *a priori* about 5%, 20% and 20% precision respectively, see **Table S4**). Note that in this qAOP, the statistical model (*i.e.*, the likelihood of the data) is clearly separated from the structural equations.

Table S4. Prior parameter distributions for the dose-response based qAOP.

Parameter	Units	Prior distribution
$KBrO_3$-GSH model		
k	$1/mM^b$	Uniform (0, 3)
b	-	Uniform (0.3, 1.5)
σ_{GSH}	%	Normal (1, 0.05) truncated to [1, 2]
$KBrO_3$-time-DCF model		
A	RFU	Uniform (0, 5000)
B	RFU	Uniform (10000, 20000)
δ	-	Uniform (0.05, 0.5)
k_d	1/mM	Uniform (0.5, 1.5)
k_t	1/h	Uniform (0.05, 0.5)
σ_{DCF}	RFU	Normal (1, 0.2) truncated to [1.01, 2]
$KBrO_3$-time-lactate model		
a	mM	Uniform (1, 5)
b	-	Uniform (-1, 1);
c	mM/h	Uniform (-0.1, 0)
d	mM/h^2	Uniform (0, 0.01)
e	1/h	Uniform (0, 0.1)
f	$1/h^2$	Uniform (-0.001, 0)
σ_{lac}	mM	Normal (1, 0.2) truncated to [1, 2]

Table S5. Summary of the posterior parameter distributions for the dose-response based qAOP fitted to GSH, DCF and lactate data.

Parameter	Units	Maximum posterior value	mean (SD) [2.5pctile, 97.5pctile]
$KBrO_3$-GSH model			
k	$1/mM^b$	1.44	1.44 ± 0.06 [1.32, 1.56]
b	-	0.73	0.73 ± 0.028 [0.68, 0.79]
σ_{GSH}	%	1.22	1.22 ± 0.022 [1.18, 1.27]
$KBrO_3$-time-DCF model			
A	RFU	2100	2100 ± 33 [2000, 2200]
B	RFU	12500	12500 ± 210 [12200, 12800]
δ	-	0.21	$2.1\times10^{-1} \pm 5.3\times10^{-3}$ [0.2, 0.22]
k_d	1/mM	0.62	$6.2\times10^{-1} \pm 1.7\times10^{-2}$ [0.6, 0.65]
k_t	1/h	0.14	$0.14 \pm 6.7\times10^{-3}$ [0.13, 0.15]
σ_{DCF}	RFU	1.19	1.19 ± 0.0022 [1.18, 1.19]
$KBrO_3$-time-lactate model			
a	mM	2.9	2.8 ± 0.22 [2.4, 3.2]
b	-	-6.2×10^{-2}	$-5.0\times10^{-3} \pm 0.11$ [-0.18, 0.18]
c	mM/h	-0.057	$-5.5\times10^{-2} \pm 0.015$ [-0.080, -0.030]
d	mM/h^2	1.0×10^{-3}	$0.001 \pm 2.2\times10^{-4}$ [6.5×10^{-4}, 0.0013]
e	1/h	0.041	$0.040 \pm 9.6\times10^{-3}$ [0.023, 0.056]
f	$1/h^2$	-3.8×10^{-4}	$-3.7\times10^{-4} \pm 1.5\times10^{-4}$ [-6.1×10^{-4}, -1.2×10^{-4}]
σ_{lac}	mM	1.27	1.28 ± 0.026 [1.24, 1.34]

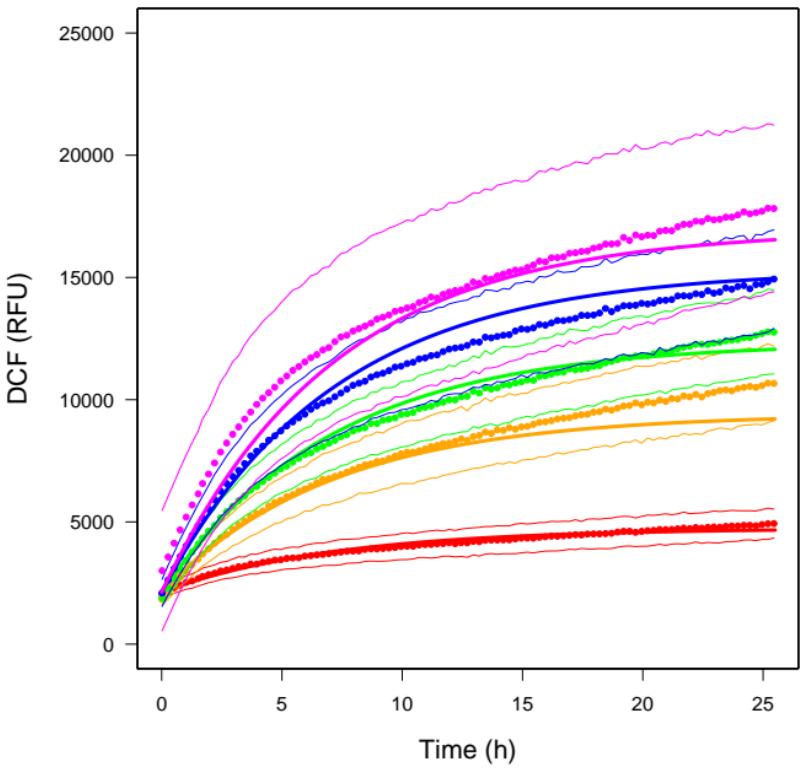


Figure S9: Best fit of the dose-response based qAOP (equations 4.2 and 7.8) to the $KBrO_3$ - time - DCF data. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **orange**: 0.75 mM; **green**: 1.5 mM; **blue**: 3 mM; **magenta**: 6 mM. The best fit curves (thick lines) are plotted along with the mean of eight DCF measurements (dots). The thin lines correspond to +/- one measurement SD around the mean.

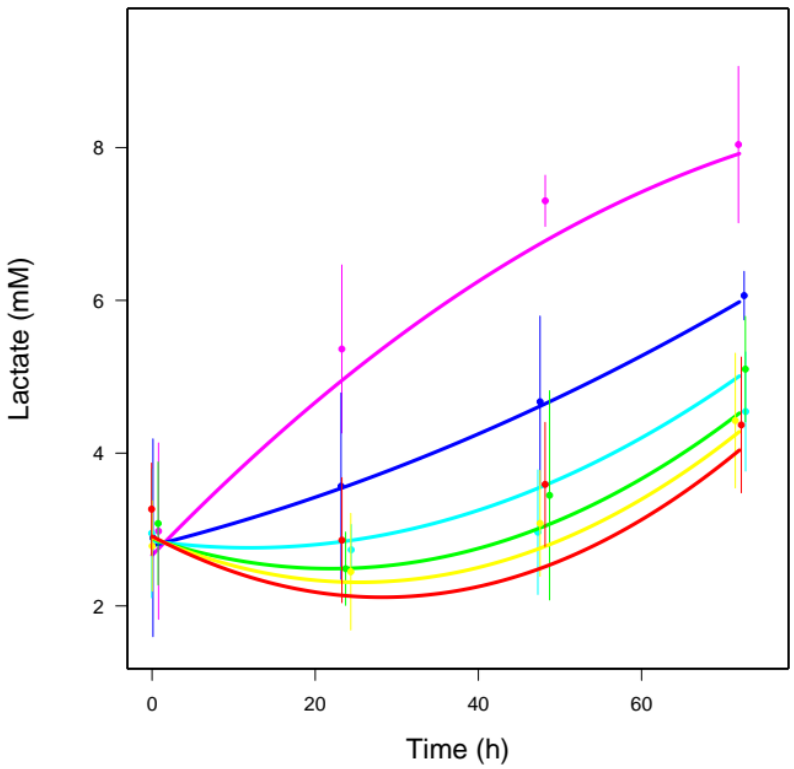


Figure S10: Best fit of the dose-response based qAOP (equations 4.3 and 7.11) to the $KBrO_3$-time-lactate data. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **yellow**: 0.25 mM; **green**: 0.5 mM; **light blue**: 1 mM; **dark blue**: 2 mM; **magenta**: 4 mM. The best fit curves (thick lines) are plotted along with the mean of four lactate measurements (dots). The error bars correspond to +/- one measurement SD. Measurement times have been jittered a bit to increase readability.

7.2.3 Bayesian Network qAOP – Node to node relationships

For the dependence of observed Pct_{GSH} on C_{KBrO3} we use a simplified probabilistic version the dose-response based qAOP (*cf.* equation 7.6):

$$Pct_{GSH} \sim Normal(100 \times exp(-k_{GSH} \cdot C_{KBrO3}), \sigma^2_{GSH}) \qquad (7.13)$$

With depletion rate constant k_{GSH} and variance σ^2_{GSH}. Note that for simplicity we set parameter b to 1.

The conditional distribution of Q_{DCF} observations at a given time t, given Pct_{GSH} and the Q_{DCF} observation at the previous time $t - h$ is given by an extension of the standard DBN model in which $Pct_{GHS,t}$ influences the equilibrium value ($E_{DCF,t}$) for $Q_{DCF,t}$ to which it converges over time at exponential dampening rate ν:

$$Q_{DCF,t} \sim Normal\left(E_{DCF,t} - \left[E_{DCF,t} - Q_{DCF,t-h}\right] \cdot e^{-\nu_{DCF}h}, \sqrt{\frac{1-e^{-\nu_{DCF}h}}{1-e^{-\nu_{DCF}}}} \cdot \sigma^2_{DCF}\right) \qquad (7.14)$$

$$E_{DCF,t} = \beta_{0,DCF} + \beta_{DCF}\, Pct_{GSH} \qquad (7.15)$$

where $E_{DCF,t}$ is the equilibrium value of Q_{DCF} (a linear function of $Pct_{GSH,t}$ at time t), h is the (positive) time interval between two consecutive observations, ν_{DCF} (positive), $\beta_{0,DCF}$, β_{DCF}, and variance σ^2_{DCF} are parameters to estimate. The DCF RFU value at time zero, carboxy-DCF_0, was not measured, but should be different from zero given the 4-hour pre-treatment phase of the protocol and was therefore also estimated. Positive values of ν and h ensure that $e^{-\nu h}$ is comprised between 0 and 1.

A similar relationship was used for lactate, replacing $Q_{DCF,t}$ by $C_{lactate,t}$, and Pct_{GHS} by $Q_{DCF,t}$ in equations 7.14 and 7.15. Given the recurrent experimental change of medium during the experiment, lactate concentration was set to zero at the start of the experiment and reset to that value every 24 hours.

For the parameter estimation, posterior parameter distributions were obtained by Hamiltonian MCMC, using the *Stan* software (Carpenter *et al.*, 2017). Three simulated Markov chains were run in parallel for 12,000 i terations, keeping the last 6,000 iterations. Non-informative uniform prior distributions were used for each parameter except for the parameters in the DCF - time - lactate portion of the model where weakly informative Gaussian priors were used to stabilize inference (see **Table S6**). In this qAOP model, the data likelihood is embedded in the model formulation. There is one clear constraint for this model: time and exposure conditions *must* match for all the variables entering a particular node to node relationship. For example, lactate was measured every 24 hours and depends on DCF, which was measured every 15 minutes, but for different $KBrO_3$ concentrations. Therefore we need to statistically "impute" (randomly draw from their conditional distribution) the expected DCF values at the concentrations used in the lactate experiment. Note that the alternative solution of describing the DCF dynamics only at time points zero and 24 hours would discard most of the DCF data and is thus unsatisfactory. To reduce the number of data points to be imputed, we chose to use only one in four DCF data points (one per hour), thus imputing 432 missing lactate data values.

Table S6. Prior parameter distributions of the DBN qAOP.

Parameter	Units	Prior distribution
$KBrO_3$-GSH link		
k_{GSH}	1/mM	Normal (1.1, 2)
σ_{GSH}	%	Uniform (0, ∞)
GSH-DCF link		
DCF_0	RFU	Uniform (0, 3000)
$\beta_{0,DCF}$	RFU/%	Uniform (0, ∞)
β_{DCF}	RFU/%	Uniform (-∞, ∞)
ν_{DCF}	1/h	Uniform (0, 1)
σ_{DCF}	RFU	Uniform (0, ∞)
DCF-lactate link		
$\beta_{0,lac}$	mM	Normal (1, 10) truncated to [0, ∞[
β_{lac}	RFU/mM	Normal (0, 0.01) truncated to [0, ∞[
ν_{lac}	1/h	Normal (1, 0.1) truncated to [0, 1]
σ_{lac}	mM	Normal (1, 10) truncated to [0, ∞[

Table S7: Summary of posterior parameter distributions of the DBN qAOP fitted to GSH, DCF and lactate data.

Parameter	Units	Maximum posterior value	mean (SD) [5pctile, 95pctile]
$KBrO_3$-GSH link			
k_{GSH}	1/mM	0.61	0.75 ± 0.18 [0.48, 1.1]
σ_{GSH}	%	18	15 ± 6.3 [6.9, 27]
GSH-DCF link			
DCF_0	RFU	2160	2160 ± 20 [2130, 2190]
$\beta_{0,DCF}$	RFU	1.89×10^4	$1.9\times10^4 \pm 1.1\times10^3$ [1.7×10^4, 2.1×10^4]
β_{DCF}	RFU/%	-117	-130 ± 9.8 [-148, -117]
ν_{DCF}	1/h	0.0783	0.10 ± 0.011 [8.2×10^{-2}, 0.12]
σ_{DCF}	RFU	906	890 ± 10 [880, 910]
DCF-lactate link			
$\beta_{0,lac}$	mM	9.68×10^{-3}	$1.7 \pm 3.9\times10^{-1}$ [1.05, 2.3]
β_{lac}	RFU/mM	4.05×10^{-4}	$2.5\times10^{-4} \pm 3.7\times10^{-5}$ [1.95×10^{-4}, 3.2×10^{-4}]
ν_{lac}	1/h	0.267	0.35 ± 0.064 [0.25, 0.46]
σ_{lac}	mM	0.185	0.64 ± 0.097 [0.48, 0.78]

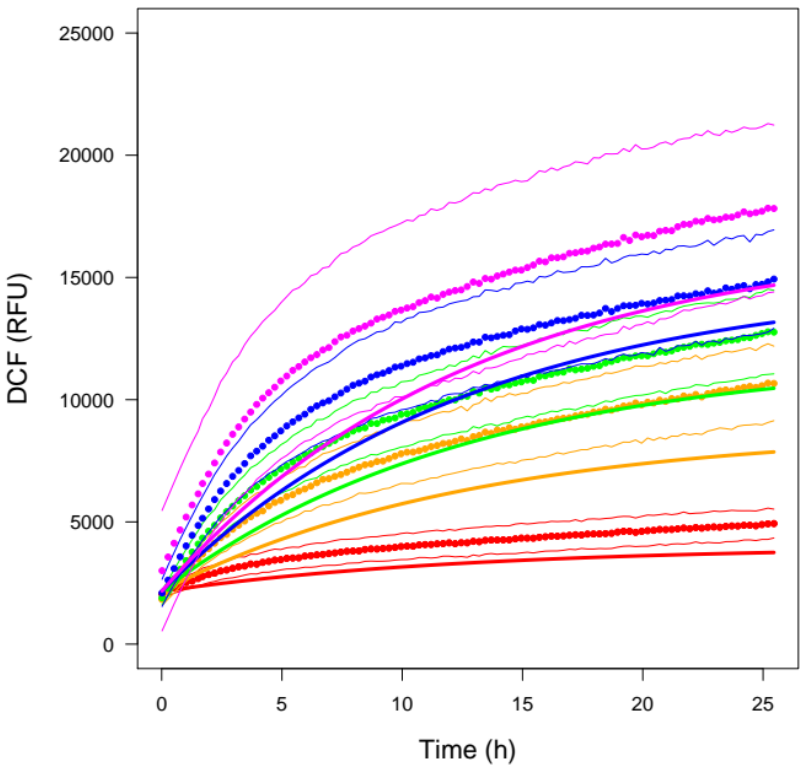


Figure S11: Best fit of the DBN qAOP to the $KBrO_3$ - time - DCF data. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **orange**: 0.75 mM; **green**: 1.5 mM; **blue**: 3 mM; **magenta**: 6 mM. The best fit curves (thick lines) are plotted along with the mean of eight DCF measurements (dots). The thin lines correspond to +/- one measurement SD around the mean.

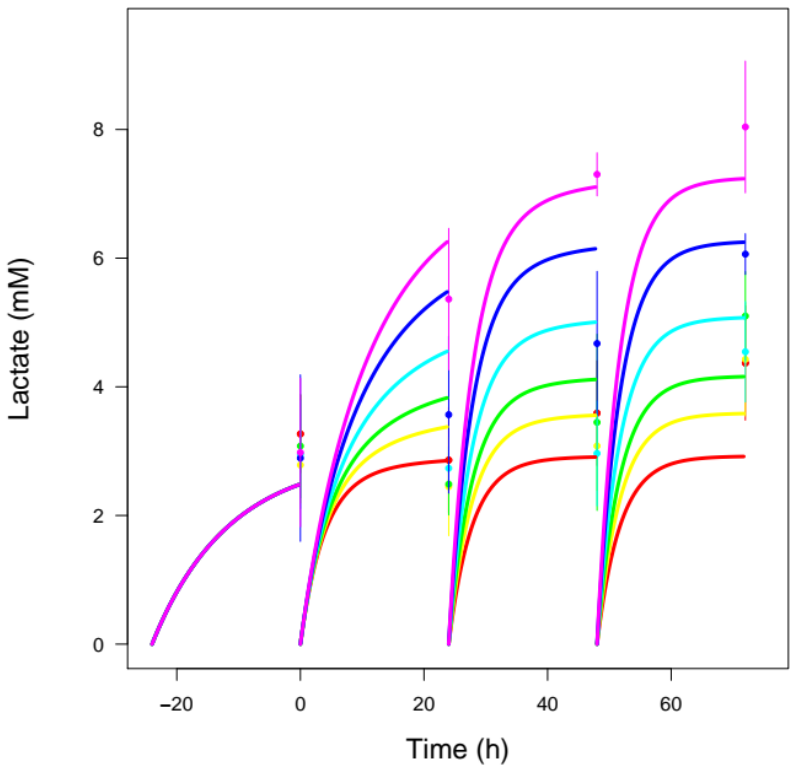


Figure S12: Best fit of the DBN qAOP to the $KBrO_3$ - time - lactate data. Simulations start one day before exposure to $KBrO_3$, is simulated. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **yellow**: 0.25 mM; **green**: 0.5 mM; **light blue**: 1 mM; **dark blue**: 2 mM; **magenta**: 4 mM. The best fit curves (thick lines) are plotted along with the mean of four lactate measurements (dots). The error bars correspond to +/- one measurement SD. Measurement times have been jittered a bit to increase readability.

7.2.4 SB Model Validation

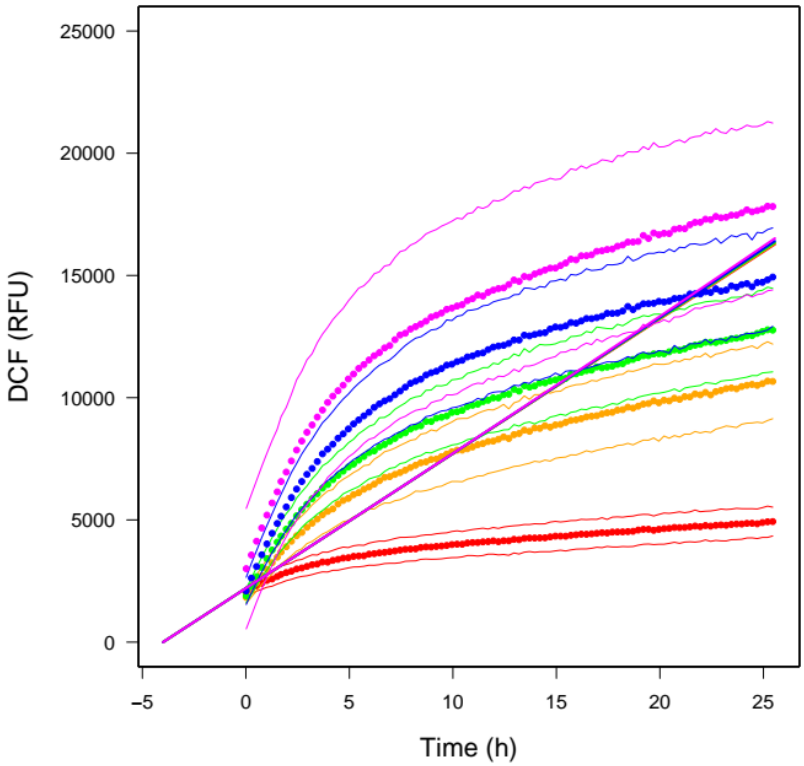


Figure S13: Fit of the SB model (with action of $KBrO_3$ on external GSH and formation of DCF by ROS) to the $KBrO_3$ - time - DCF data. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **orange**: 0.75 mM; **green**: 1.5 mM; **blue**: 3 mM; **magenta**: 6 mM. The maximum posterior fit curves (thick lines) are plotted along with the mean of eight DCF measurements (dots). Note the very faint effect of dose (the five fit curves are not exactly superimposed). The thin lines correspond to +/- one measurement SD around the mean.

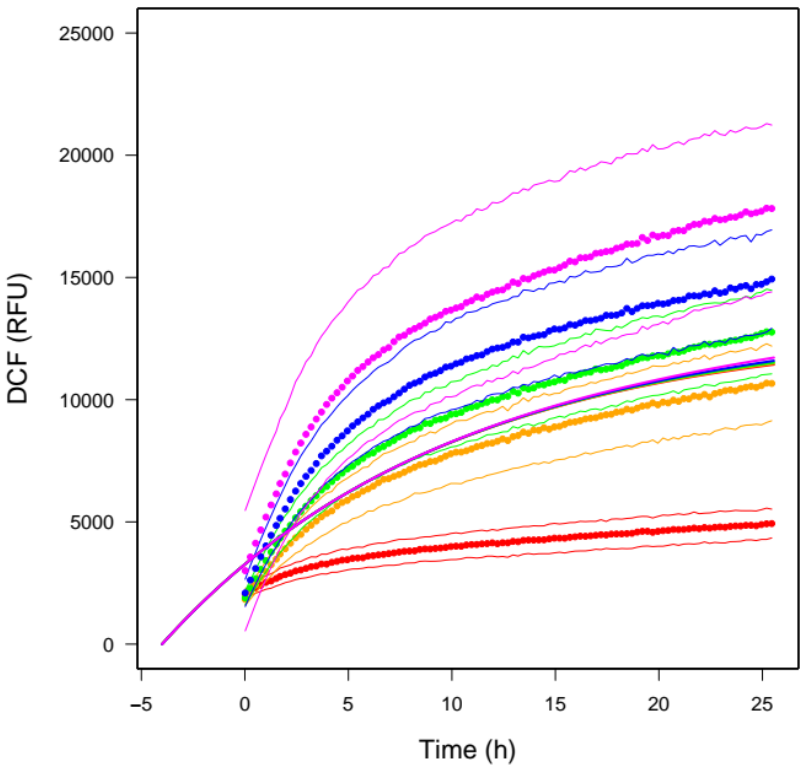


Figure S14: Fit of the SB model (with action of $KBrO_3$ on external GSH, formation of DCF by ROS, and DCF bleaching) to the $KBrO_3$ - time - DCF data. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **orange**: 0.75 mM; **green**: 1.5 mM; **blue**: 3 mM; **magenta**: 6 mM. The maximum posterior fit curves (thick lines) are plotted along with the mean of eight DCF measurements (dots). Note the very faint effect of dose (the five fit curves are not exactly superimposed). The thin lines correspond to +/- one measurement SD around the mean.

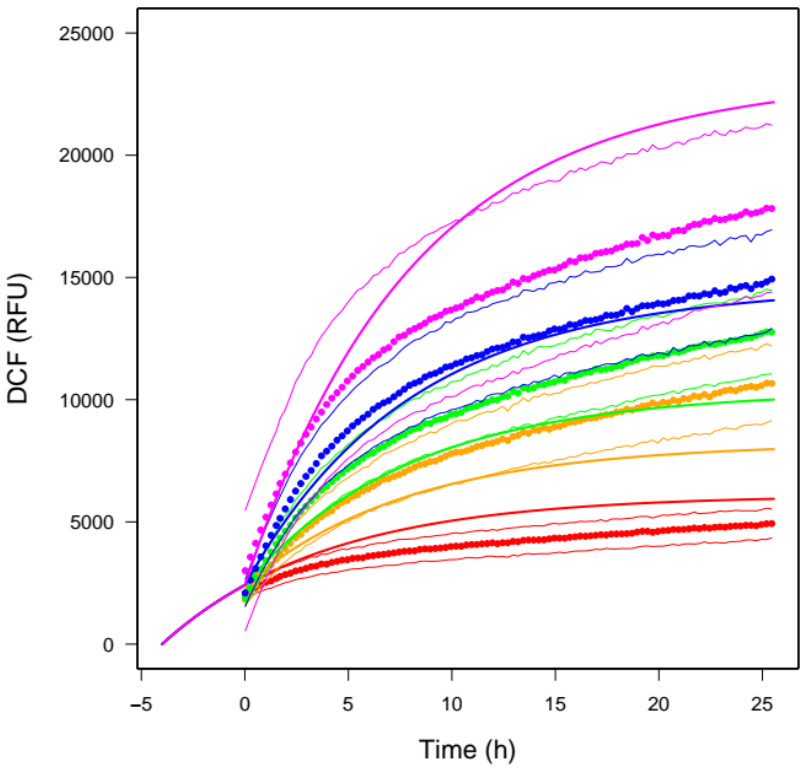


Figure S15: Fit of the best SB model to the $KBrO_3$ - time - DCF data. The model includes action of $KBrO_3$ on external GSH, formation of DCF by ROS and $KBrO_3$, and DCF bleaching. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **orange**: 0.75 mM; **green**: 1.5 mM; **blue**: 3 mM; **magenta**: 6 mM. The maximum posterior fit curves (thick lines) are plotted along with the mean of eight DCF measurements (dots). The thin lines correspond to +/- one measurement SD around the mean.

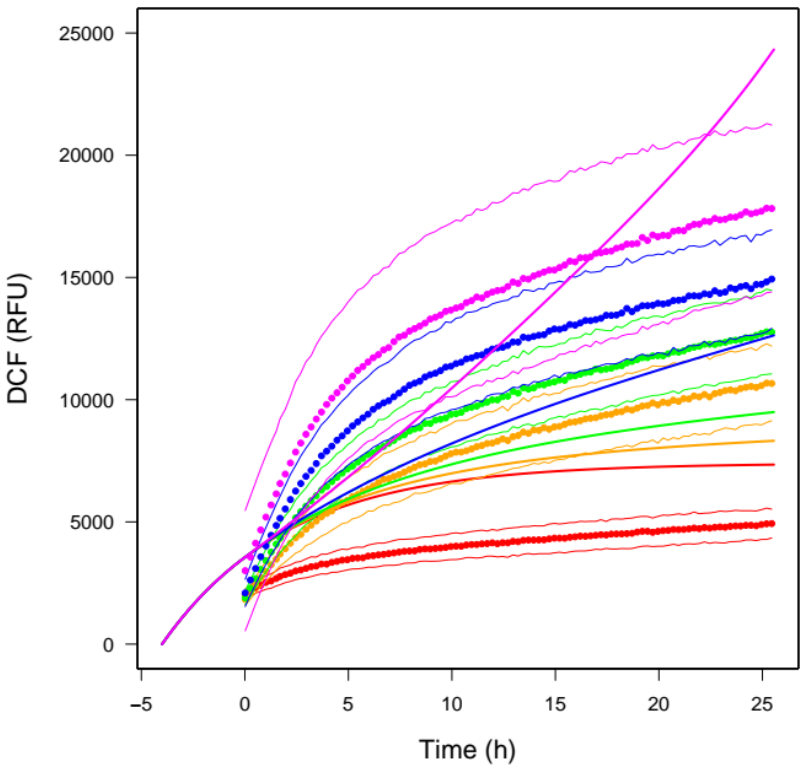


Figure S16: Fit of the SB model (with action of $KBrO_3$ on external and internal GSH, formation of DCF by ROS, and DCF bleaching) to the $KBrO_3$ - time - DCF data. The colors correspond to the various $KBrO_3$ exposure concentrations: **red**: 0; **orange**: 0.75 mM; **green**: 1.5 mM; **blue**: 3 mM; **magenta**: 6 mM. The maximum posterior fit curves (thick lines) are plotted along with the mean of eight DCF measurements (dots). The thin lines correspond to +/- one measurement SD around the mean.

7.2.5 Effectopedia Implementation

Effectopedia provides a graphical user interface to build an AOP diagram, which in turn gives easy access to relevant descriptions, data and models. In addition to a qualitative description of the AOP, Effectopedia provides structure for representation of test methods, collected data and executable models implemented in the supported programming languages (R, MATLAB, Java). Effectopedia was used to create several iterations of the AOP diagram. Initially, the sequence of KEs included relevant feedback mechanisms or parallel processes (branches). However, in the following step of identification of measurement methods, some of these events did not have a separate method of observation and were therefore combined into a single KE. Other events were determined to be modification factors rather than being causally related to the AO and were removed from the pathway diagram. The current version of the AOP diagram implemented in Effectopedia is shown on the **Figure S17**. Each of the elements in the diagram can be expanded and details can be added to their description. Models were implemented in R and their source code contributed to the description of the *in silico* models, allowing other users to execute them with the same and/or different data and model parameters.

Effectopedia implementation of both BN and SB models faces similar challenges, of which the most important is matching the internal structure of the models to the conceptual structure provided by the AOP. Currently, Effectopedia allows "global models" in which one BN or SB model can cover several KEs. Such models need to have specific outputs matching the AOP KEs. A problem in that approach is the derivation of reusable KERs. If the global model contains complex time or variable dependencies between non-adjacent KEs, they need to be explicitly represented in the AOP as feedbacks, feed-forwards or modifying factors. However, extracting such dependencies is non-trivial. Alternatively, the AOP could be re-designed if the global model indicates that some tightly coupled KEs can be merged.

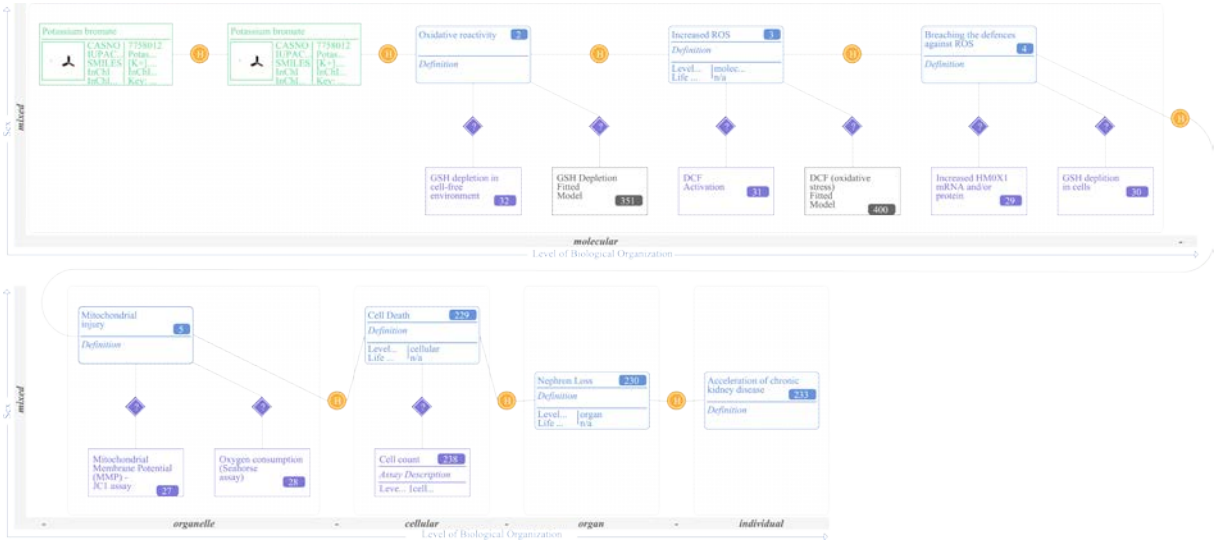


Figure S17: Diagram of renal qAOP (with *in silico* models and test data) exported form Effectopedia (broken into two segments for readability purposes). The diagram starts with extracellular $KBrO_3$ (first green box) which is transported into cells (second green box). The orange link between the two green boxes represents the transport across the cell membrane and be described with a toxicokinetic model. Intra-cellular $KBrO_3$ is then connected to the MIE (blue box ID2) 'Oxidative reactivity'. There is one *in vitro* test method (purple box ID32) and one *in silico* model (gray box ID351) that can be used to measure/estimate the MIE. The MIE is followed by a sequence of KEs (blue boxes ID3-ID5, ID229-230) leading to the AO (blue box ID233). Orange circles between KEs represent KERs. KERs can include multiple executable response-response models in their description. Purple rhomboids between test methods and KEs represent test-response mappings which describe how measured results can be interpreted or transformed to reflect the *in vivo* context of the KE. Experimental data for 'GSH depletion in the cell-free environment' (box ID32), 'DCF Activation' (box ID31), and 'increased lactate production' (box ID26) were entered into Effectopedia. The same data were used for fitting models described in 'GSH depletion Fitted Model' (box ID351), 'DCF (oxidative stress) Fitted Model' (box ID400)'.

7.3 SUPPLEMENTARY INFORMATION FOR CHAPTER 5

Table S8: Target genes generated by the PubMed searches for AhR, Nrf2, ATF4 pathways.

Pathway (Number of genes)	**Genes** (804)											
AhR (60)	AHRR	ALDH1A3	ALS2CL	ATXN1	BATF	BCL3	BMF	CACNA1D	CYP1A1	CYP1A2	CYP1B1	DDX24
	DHX37	DLL1	DLX1	DNMBP	DPP9	EDC3	ELF4	EREG	FAM32A	FAM65C	FLVCR2	FOSL2
	FREM2	IGF1R	LAMA3	LMCD1	MAPRE2	NEDD9	NIN	NPY1R	PITPNM2	PLEC	PRPS1	PSG5
	PYGL	RND1	RRP12	RUNX1	RUNX2	SAMD12	SAT1	SECTM1	SIPA1L2	SLC27A2	SLC2A11	STRBP
	TFAP2A	TH	TIPARP	TMEM45B	TPCN1	TRAFD1	TRUB2	USP3	VDR	VIPR1	VTCN1	WDR63
Nrf2 (306)	ABCB6	ABCC1	ABCC2	ABCC3	ABCC5	ABHD4	ACAP2	ADAM23	ADCY1	ADO	AGPAT9	AHR
	AIFM2	AKAP7	AKR1B10	AKR1C1	AKR1C2	AKR1C3	ALDH1A1	ALS2	AMBP	ANKRD44	ANXA1	AOX1
	AP5Z1	APPL2	ARHGAP18	ARHGDIB	ASB3	ATF4	ATG16L2	AURKB	B3GNT2	BACH1	BBS9	BCL2
	BEND6	BET1	BLVRB	BMP8A	BMPER	BMPR1A	BTG2	C12orf29	C16orf72	C8orf4	C9orf3	CABC1
	CAPN7	CBARA1	CBR3	CCDC104	CCDC109B	CCDC53	CCDC90B	CD27	CD302	CD36	CDC2L6	CDH11
	CDKN2B	CES1	CFTR	CHPT1	CLCN1	CLIP4	CLTC	COCH	COL24A1	COL4A1	COLEC12	
		COMMD6										
	CORO7	COX6C	CPNE8	CPT1A	CTSB	CTSC	CTSO	CTTN	CXCL5	DAAM2	DAGLB	DCN
	DENND4C	DHX40	DLGAP5	DUSP22	DUSP5	ECM1	EDA2R	EMC7	ENC1	ENTPD5	EPHX1	EXOC7
	FAM102A	FAM167A	FAM210B	FAM214A	FAM69A	FBF1	FBXL3	FBXO30	FECH	FKBPL	FNBP1	FOPNL
	FOXA2	FRMD6	FTH1	FUK	FZD6	GABARAPL1	GAS1	GCLC	GCLM	GCNT3	GDF15	GDI2
	GLI2	GLO1	GPNMB	GRIP2	GRM1	GSR	GSTA1	GSTA4	GSTM1	GSTM2	GSTM3	GSTP1
	GVIN1	HACE1	HBP1	HECA	HES1	HINT3	HIST1H1C	HMHB1	HRASLS2	HRSP12	HSPA1B	IDH1
	IDS	IFIT3	IGF1	IGF2R	IL33	IL6	INSIG1	INSIG2	IRF4	ITGB2	JAG1	
		KCNAB1										
	KDELR2	KEAP1	KIF26B	KLHDC8A	KRAS	LAPTM4B	LAYN	LBH	LMF1	LNX1	LPL	
		LRRC4C										
	LTA	LYPD6B	LYPLAL1	MACF1	MAFF	MAFG	MAGOHB	MAMDC2	MAP3K9	MAPK8	MATN2	ME1
	MEGF9	MEIS1	MFSD9	MGST2	MLYCD	MMD	MRPL14	MRPL33	MSC	MT2A	NCAPD2	NETO2
	NPNT	NQO2	NREP	OAT	OIT3	OSGIN1	P2RY10	PAIP1	PBX1	PCDH7	PDDC1	PDGFC
	PDIA4	PGAP1	PHLDA1	PION	PIP5K1C	PIR	PLA2G4A	PLN	PMAIP1	PPP1R12B	PQLC3	PRC1
	PRDM1	PRDX1	PRKAR2B	PRKCD	PRR13	PSMB5	PTCD1	PTGR1	PTTG1IP	RAB31	RAB32	
		RAB33A										
	RFFL	RGS10	RND3	RNF121	RNF141	RNF216	RPL10A	RPS6	RXRA	SARM1	SCAMP1	SCYL1
	SDCCAG8	SDPR	SEC63	SEMA3E	SERPINE1	SH3RF2	SH3TC1	SIM2	SLC12A8	SLC16A6	SLC16A9	
	SLC39A10											
	SLC44A3	SLMAP	SNX1	SNX13	SNX22	SNX4	SOAT1	SOCS5	SOD1	SOD2	SQRDL	SRP9
	SRXN1	SSR3	SUB1	SYNPO2	TALDO1	TBC1D23	TBCEL	TBXAS1	TFE3	TFPI	TGFB2	TINAG
	TKT	TMEFF2	TMEM159	TMEM206	TMEM63A	TMEM64	TNFSF9	TRIM69	TSPAN3	TXLNB	TXN	TXNDC5
	UBE2V2	UGT1A1	UGT1A6	UGT2B7	UNKL	VASN	VPS8	VSTM4	WAPAL	WASF1	WDR81	WDSUB1
	WNT5A	YAF2	ZC3H11A	ZDHHC20	ZFAND2B	ZMAT4						
ATF4 (408)	AARS	ACACA	ACAD8	ACOT2	ACOX2	ADM2	AFF1	AGAP1	AKAP2	AKNA	ALDH18A1	
		ALDH1L2										
	ALKBH5	AMACR	ANAPC1	ANGPTL4	ANGPTL6	ANK2	ANKRD1	ANKRD11	ANP32B	APBB2	APOBEC1	APOE
	ARHGEF2	ARID5B	ARPC5L	ASCC2	ASNS	ATF5	ATF6	ATL2	ATP6V1G1	ATXN2L	B4GALNT2	B9D2
	BCAR1	BCAT1	BCAT2	BCMO1	BDNF	BGLAP	BHLHE22	BNIP1	BOLA1	BTF3L4	CA9	CABP1
	CALCRL	CAMSAP1	CARS	CASC5	CASP4	CAST	CCL2	CCL7	CCT8L1	CD276	CD74	CDH24
	CDSN	CEBPB	CEBPG	CHAC1	CLCN3	CLIC4	CNOT1	CNPY2	COASY	COL18A1	CP	CRLS1
	CRYBG3	CTH	CUL2	CXADR	CXCL3	CXCR5	CYB5R1	DAAM1	DDIT3	DDR2	DEGS2	DHRS7
	DLEU7	DNAJA3	DNAJB12	DPF2	DPYSL2	DTNBP1	DUSP6	DUT	DYSF	ECHS1	EDEM3	EIF1
	EIF2S2	EIF3C	EIF4EBP1	EIF4G2	EIF5	ELMSAN1	EMILIN1	EPO	EPRS	ERLIN1	ERO1L	ETF1
	ETS1	F3	FADS3	FAM119A	FAM129A	FAM159A	FAM175A	FAM175B	FAM188B	FAM26F	FAM83F	
		FAM96A										
	FBXL2	FBXO11	FBXO31	FCER1G	FGF19	FGF21	FIBIN	FLAD1	FLNC	FLRT1	FNDC7	GARS
	GHITM	GNG5	GNPNAT1	GOT1	GPAM	GPATCH3	GPR85	GPT2	GPX4	GRB10	GSK3A	GSTO1
	GTPBP2	GTPBP4	H2AFZ	HAS2	HAX1	HDAC8	HERPUD1	HFE	HHIPL1	HSPA5	HTRA1	IARS
	ICK	IER2	IFI44	IFT172	IHH	IL1R1	IL2	IL23A	INCENP	INPP4B	INPP5B	IP6K2
	IPMK	ITGB5	ITGB7	ITIH1	JAGN1	JDP2	JHDM1D	KBTBD5	KCNT2	KDM5A	KDM6B	KHNYN
	KIF13B	KIFC1	KLF15	KLF4	KLHDC10	KRTCAP2	LARS	LCN2	LEPREL1	LEPROTL1	LGALS3	
		LGALS8										
	LHFPL2	LMBRD1	LMO4	LONP1	LRRK1	LSG1	LSM14A	LTBP3	MACROD1	MANEA	MARS	MAT2B
	MAZ	MCL1	METTL9	MFF	MID1IP1	MMP19	MRAS	MRPL13	MRPL24	MRPL54	MTHFD1L	
		MTHFD2										
	MTHFR	MTM1	MYCBP2	MYOM1	NAAA	NARS	NCL	NDRG1	NDUFA4L2	NF1	NFE2L1	NFKB1
	NFU1	NIPBL	NME2	NOS2	NOSIP	NRBF2	NRIP2	NUPR1	OPTC	ORMDL3	OSMR	OTUB2
	PACSIN2	PAQR3	PAX8	PCK2	PDAP1	PDE1A	PDE4DIP	PER2	PER3	PFDN1	PHF10	PHF21A
	PHGDH	PHYHD1	PIK3C2G	PKMYT1	PLA2G12A	PLD1	PLEKHH3	PNRC2	POP5	PPP1R15A	PPP1R15B	
		PPP2R5A										
	PRDM15	PRKG2	PRSS35	PRSS36	PSEN1	PTBP1	PTGS2	PTN	PTPN12	PTPN21	PTPRS	PVRL2
	PYCR1	RAB39B	RAD21	RAI1	RAN	RBM39	RBM4	RBM9	RCC2	RD3	RELN	
		RHBDD1										
	RHOQ	RNF114	RPL13A	RPL7	RPRD2	RPS6KA2	RSBN1	RUNX3	RUVBL2	SAMD8	SARS	SERPINB9
	SERPINF1	SERTAD2	SESN2	SH3BP5L	SHMT2	SLC16A14	SLC19A2	SLC1A4	SLC1A5	SLC1A7	SLC20A1	
	SLC25A28											
	SLC25A33	SLC25A39	SLC35B4	SLC38A2	SLC38A7	SLC41A3	SLC6A9	SLC7A1	SLC7A11	SLC9A9	SLIT1	SLTM
	SNAI2	SNRNP35	SNX12	SNX33	SOAT2	SPATS2	SPEN	SRPK1	SRSF1	STARD5	STC2	STEAP1
	STK40	SURF6	SYNJ2	TAB2	TAF15	TAF5	TAGLN2	TARS	TBC1D9B	TBPL1	TET3	TGIF1
	TLCD2	TMEM11	TMEM154	TMEM189	TMEM74	TMIGD1	TNFAIP2	TNFAIP6	TNKS	TNPO1	TNRC6C	TOR3A
	TPX2	TRAF3IP2	TRIB3	TRIM35	TRIM66	TSC22D3	TSLP	TSPAN5	TSPYL4	TTC23	TTC9C	UBA3
	UBE2L3	UBE2W	UBP1	UBR2	UBR4	UHRF1BP1	UQCRC2	UQCRQ	USP11	USP2	USP36	USP6NL
	VAC14	VARS	VLDLR	VNN3	VPS54	WARS	WBP4	WDR27	WISP1	WWP2	WWTR1	XPOT
	YARS	YLPM1	YWHAG	ZAK	ZBTB38	ZBTB7B	ZBTB7C	ZC3H18	ZC3HAV1	ZFC3H1	ZNF268	ZYG11B
AhR Nrf2 (7)	ABCG2	DLX2	MCOLN2	NFE2L2	NQO1	PSPC1	TXNRD1					
AhR ATF4 (4)	CRISPLD2	DDIT4	SLC7A5	UBE2G2								
Nrf2 ATF4 (17)	ABCC4	ATF3	CLN5	CR1L	DDC	DHX57	GOLIM4	HMOX1	HTATIP2	MDFIC	PSAT1	SCPEP1
	SEMA6D	SPRED1	SQSTM1	TBCE	YPEL5							
AhR Nrf2 ATF4 (2)	SLC3A2	TMTC2										
Genes with no data* (53)	**Nrf2** (16) **:** AI646023, AKR1A4, BC067068, BCLX, CYP2A5, GSPDX, H2-T10, IQWD1, NDP52, PREI4, PRX1, SNF1LK2, SPEER3, SPEER4A, TXN1, ZFP51.											
	ATF4 (36) **:** ABCB1A, ACAA1A, B230217C12RIK, BB014433, BC019943, BC055111, CYP2AB1, D48WG0951E, DEB1, DEFB43, DNAIC2, GCFC1, GM129, GM5627, GM5820, GM5867, H2-Q7, HIST2H4, IFITM6, JMJD3, LIPIN2, MTAP4, PGPEP1L, RNASET2B, SAMD4, SERPINA3C, SLCO1A5, TREM3, UGT1A6A, VEGF, ZFP238, ZFP365, ZFP598, ZFP608, ZFP623, ZFP708.											
	Nrf2 ATF4 (1) **:** B230315N10RIK.											
	* No data are available for these genes in the carcinoGENOMICS, PREDICT-IV, or TG-GATEs databases.											

Table S9: Exhaustive list of chemicals used in each experimental category.

Carcino-GENOMICS	Human	Kidney	*in vitro*	Bolus	(30)	1-Amino-2,4-dibromoanthraquinone, 2-Nitrofluorene, 4-Acetylaminofluorene, Aristolochic acid, Benzo(a)pyrene, Benzoin, Benzyl alcohol, Bromodichloromethane, Chlorothalonil, Citrinin, Clonidine, Cyclosporine A, D-mannitol, DCVC, Diclofenac, DL-Menthol, Fumonisin-B1, Hydroquinone, Lead-(II)-acetate-trihydrate, Monuron, N-Ethyl-N-(2-hydroxyethyl)nitrosamine, N-Nitrosomorpholine, Nifedipine, Nitrilotriacetic acid, Ochratoxin A, Potassium bromate, Streptozotocin, Tolbutamide, Triclosan (Igarsen), Tris(2,3-dibromopropyl)phosphate.
	Rat	Kidney	*in vitro*	Bolus	(15)	2-Nitrofluorene, Aristolochic acid, Benzo(a)pyrene, Bromodichloromethane, Chlorothalonil, Clonidine, D-Mannitol, DCVC, Diclofenac, Dimethylnitrosamine, Monuron, Nifedipine, Ochratoxin A, Potassium bromate, Tolbutamide.
PREDICT -IV	Human	Kidney	*in vitro*	Repeat	(12)	Adefovir dipivoxil, Adefovir dipivoxil-hypoxia, Cadmium chloride, Chloroacetaldehyde, Cidofovir, Cisplatin, Clodronate, Cyclosporine A, Hypoxia, Ifosfamide, Zoladronate-hypoxia, Zoledronate.
	Human and Rat	Liver	*in vitro*	Repeat	(11)	Acetaminophen, Amiodarone, Chlorpromazine, Cyclosporine A, EMD335823, Fenofibrate, Ibuprofen, Metformin, Rosiglitazone, Troglitazone, Valproic acid.
TG -GATEs	Human	Liver	*in vitro*	Bolus	(160)	2-Nitrofluorene, 2,4-dinitrophenol, Acarbose, Acetamide, Acetamidofluorene, Acetaminophen, Acetazolamide, Adapin, Aflatoxin B1, Ajmaline, Allopurinol, Allyl alcohol, Alpidem, Amiodarone, Amitriptyline, Amphotericin B, Aspirin, Azathioprine, Bendazac, Benzbromarone, Benziodarone, Bromobenzene, Bromoethylamine, Bucetin, Buspirone, Buthionine sulfoximine, Butylated hydroxyanisole, Caffeine, Captopril, Carbamazepine, Carbon tetrachloride, Chloramphenicol, Chlormadinone, Chlormezanone, Chlorpheniramine, Chlorpromazine, Chlorpropamide, Cimetidine, Ciprofloxacin, Clofibrate, Clomipramine, Clozapine, Colchicine, Coumarin, Cycloheximide, Cyclophosphamide, Cyclosporine A, Danazol, Dantrolene, Dexamethasone, Diazepam, Diclofenac, Diethylmaleate, Diltiazem, Disopyramide, Disulfiram, Doxorubicin, EMD335823, Enalapril, Erythromycin, Ethambutol, Ethanol, Ethinylestradiol, Ethionamide, Ethionine, Etoposide, Famotidine, Fenofibrate, Fluoxetine, Fluphenazine, Flutamide, Furosemide, Galactosamine, Gemfibrozil, Glibenclamide, Griseofulvin, Haloperidol, Hexachlorobenzene, HGF, Hydroxyzine, Ibuprofen, IL1beta, IL6, Imipramine, Indomethacin, INFalpha, Iproniazid, Isoniazid, Ketoconazole, Labetalol, Lomustine, Lornoxicam, LPS, Mefenamic acid, Meloxicam, Metformin, Methapyrilene, Methimazole, Methyldopa, Methylene dianiline, Methyltestosterone, Mexiletine, Monocrotaline, Moxisylyte, N-methyl-N-nitrosourea, N-nitrosomorpholine, Naphthyl isothiocyanate, Naproxen, Nefazodone, Nicotinic acid, Nifedipine, Nimesulide, Nitrofurantoin, Nitrofurazone, Nitrosodiethylamine, Omeprazole, Papaverine, Pemoline, Penicillamine, Perhexiline, Phalloidin, Phenacetin, Phenobarbital, Phenylanthranilic acid, Phenylbutazone, Phenytoin, Phorone, Promethazine, Propranolol, Propylthiouracil, Quinidine, Ranitidine, Rifampicin, Rosiglitazone, Rotenone, Simvastatin, Sulfasalazine, Sulindac, Sulpiride, Tacrine, Tamoxifen, Tannic acid, Terbinafine, Tetracycline, TGFbeta1, Theophylline, Thioacetamide, Thioridazine, Ticlopidine, Tiopronin, TNF, Tolbutamide, Triazolam, Trimethadione, Troglitazone, Tunicamycin, Valproic acid, Venlafaxine, Vitamin A, WY-14643.
	Rat	Liver	*in vitro*	Bolus	(145)	Acarbose, Acetamidofluorene, Acetaminophen, Acetazolamide, Adapin, Ajmaline, Allopurinol, Allyl alcohol, Alpidem, Amiodarone, Amitriptyline, Aspirin, Azathioprine, Bendazac, Benzbromarone, Benziodarone, Bromobenzene, Bromoethylamine, Bucetin, Buspirone, Buthionine sulfoximine, Caffeine, Captopril, Carbamazepine, Carbon tetrachloride, Carboplatin, Cephalothin, Chloramphenicol, Chlormadinone, Chlormezanone, Chlorpheniramine, Chlorpromazine, Chlorpropamide, Cimetidine, Ciprofloxacin, Cisplatin, Clofibrate, Clomipramine, Clozapine, Colchicine, Coumarin, Cycloheximide, Cyclophosphamide, Cyclosporine A, Danazol, Dantrolene, Diazepam, Diclofenac, Diethylmaleate, Diltiazem, Disopyramide, Disulfiram, Doxorubicin, Enalapril, Erythromycin, Ethambutol, Ethanol, Ethinylestradiol, Ethionamide, Ethionine, Etoposide, Famotidine, Fenofibrate, Fluphenazine, Flutamide, Furosemide, Galactosamine, Gemfibrozil, Gentamicin, Glibenclamide, Griseofulvin, Haloperidol, Hexachlorobenzene, Hydroxyzine, Ibuprofen, Imipramine, Indomethacin, Iproniazid, Isoniazid, Ketoconazole, Labetalol, Lomustine, Lornoxicam, LPS, Mefenamic acid, Meloxicam, Metformin, Methapyrilene, Methimazole, Methyldopa, Methyltestosterone, Mexiletine, Monocrotaline, Moxisylyte, Naphthyl isothiocyanate, Naproxen, Nefazodone, Nicotinic acid, Nifedipine, Nimesulide, Nitrofurantoin, Nitrofurazone, Nitrosodiethylamine, Omeprazole, Papaverine, Pemoline, Penicillamine, Perhexiline, Phalloidin, Phenacetin, Phenobarbital, Phenylanthranilic acid, Phenylbutazone, Phenytoin, Phorone, Promethazine, Propylthiouracil, Puromycin, Quinidine, Ranitidine, Rifampicin, Simvastatin, Sulfasalazine, Sulindac, Sulpiride, Tacrine, Tamoxifen, Tannic acid, Terbinafine, Tetracycline, Theophylline, Thioacetamide, Thioridazine, Ticlopidine, Tiopronin, TNFalpha, Tolbutamide, Triamterene, Triazolam, Trimethadione, Tunicamycin, Valproic acid, Venlafaxine, Vitamin A, WY-14643.
	Rat	Liver	*in vivo*	Bolus	(158)	2-Nitrofluorene, 2,4-dinitrophenol, 3-Methylcholanthrene, Acarbose, Acetamidofluorene, Acetaminophen, Acetazolamide, Adapin, Aflatoxin B1, Ajmaline, Allopurinol, Allyl alcohol, Alpidem, Amiodarone, Amitriptyline, Amphotericin B, Aspirin, Azathioprine, Bendazac, Benzbromarone, Benziodarone, Bortezomib, Bromobenzene, Bromoethylamine, Bucetin, Buthionine sulfoximine, Butylated hydroxyanisole, Caffeine, Captopril, Carbamazepine, Carbon tetrachloride, Chloramphenicol, Chlormadinone, Chlormezanone, Chlorpheniramine, Chlorpromazine, Chlorpropamide, Cimetidine, Ciprofloxacin, Clofibrate, Clomipramine, Colchicine, Coumarin, Cycloheximide, Cyclophosphamide, Cyclosporine A, Danazol, Dantrolene, Dexamethasone, Diazepam, Diclofenac, Diethylmaleate, Diltiazem, Disopyramide, Disulfiram, Doxorubicin, Enalapril, Erythromycin, Ethambutol, Ethanol, Ethinylestradiol, Ethionamide, Ethionine, Etoposide, Famotidine, Fenofibrate, Fluoxetine, Fluphenazine, Flutamide, Furosemide, Galactosamine, Gefitinib, Gemfibrozil, Gentamicin, Glibenclamide, Griseofulvin, Haloperidol, Hexachlorobenzene, Hydroxyzine, Ibuprofen, Imatinib, Imipramine, Indomethacin, Iproniazid, Isoniazid, Ketoconazole, Labetalol, Lomustine, Lornoxicam, LPS, Mefenamic acid, Meloxicam, Metformin, Methapyrilene, Methimazole, Methyldopa, Methylene dianiline, Methyltestosterone, Mexiletine, Monocrotaline, Moxisylyte, N-methyl-N-nitrosourea, N-nitrosomorpholine, Naphthyl isothiocyanate, Naproxen, Nicotinic acid, Nifedipine, Nimesulide, Nitrofurantoin, Nitrofurazone, Nitrosodiethylamine, Omeprazole, Papaverine, Pemoline, Penicillamine, Perhexiline, Phalloidin, Phenacetin, Phenobarbital, Phenylanthranilic acid, Phenylbutazone, Phenytoin, Phorone, Promethazine, Propranolol, Propylthiouracil, Puromycin, Quinidine, Ranitidine, Rifampicin, Rosiglitazone, Rotenone, Simvastatin, Sulfasalazine, Sulindac, Sulpiride, Tacrine, Tamoxifen, Tannic acid, Terbinafine, Tetracycline, Theophylline, Thioacetamide, Thioridazine, Ticlopidine, Tiopronin, TNF, Tolbutamide, Triamterene, Triazolam, Trimethadione, Tunicamycin, Valproic acid, Vitamin A, WY-14643.
	Rat	Liver	*in vivo*	Repeat	(143)	2,4-dinitrophenol, Acarbose, Acetamide, Acetamidofluorene, Acetaminophen, Acetazolamide, Adapin, Ajmaline, Allopurinol, Allyl alcohol, Amiodarone, Amitriptyline, Amphotericin B, Aspirin, Azathioprine, Bendazac, Benzbromarone, Benziodarone, Bromobenzene, Bromoethylamine, Bucetin, Butylated hydroxyanisole, Caffeine, Captopril, Carbamazepine, Carbon tetrachloride, Carboplatin, Cephalothin, Chloramphenicol, Chlormadinone, Chlormezanone, Chlorpheniramine, Chlorpromazine, Chlorpropamide, Cholesterol sodium cholate 1to4, Cimetidine, Ciprofloxacin, Cisplatin, Clofibrate, Clomipramine, Colchicine, Coumarin, Cyclophosphamide, Cyclosporine A, Danazol, Dantrolene, Desmopressin acetate, Diazepam, Diclofenac, Diltiazem, Disopyramide, Disulfiram, Doxorubicin, Enalapril, Erythromycin, Ethambutol, Ethanol, Ethinylestradiol, Ethionamide, Ethionine, Etoposide,, Famotidine, Fenofibrate, Fluoxetine, Fluphenazine, Flutamide, Furosemide, Gemfibrozil, Gentamicin, Glibenclamide, Griseofulvin, Haloperidol, Hexachlorobenzene, Hydroxyzine, Ibuprofen, Imipramine, Indomethacin, Iproniazid, Isoniazid, Ketoconazole, Labetalol, Lomustine, Lornoxicam, Mefenamic acid, Meloxicam, Metformin, Methapyrilene, Methimazole, Methyldopa, Methylene dianiline, Methyltestosterone, Mexiletine, Monocrotaline, Moxisylyte, Naphthyl isothiocyanate, Naproxen, Nicotinic acid, Nifedipine, Nimesulide, Nitrofurantoin, Nitrofurazone, Nitrosodiethylamine, Omeprazole, Papaverine, Pemoline, Penicillamine, Perhexiline, Phalloidin, Phenacetin, Phenobarbital, Phenylanthranilic acid, Phenylbutazone, Phenytoin, Promethazine, Propranolol, Propylthiouracil, Puromycin, Quinidine, Ranitidine, Rifampicin, Rosiglitazone, Rotenone, Simvastatin, Sulfasalazine, Sulindac, Sulpiride, Tacrine, Tamoxifen, Tannic acid, Terbinafine, Tetracycline, Theophylline, Thioacetamide, Thioridazine, Ticlopidine, Tiopronin, Tolbutamide, Triamterene, Triazolam, Trimethadione, Valproic acid, Vitamin A, WY-14643.
	Rat	Kidney	*in vivo*	Bolus and Repeat	(41)	Acetaminophen, Acetazolamide, Allopurinol, Allyl alcohol, Amphotericin B, Bromobenzene, Bromoethylamine, Bucetin, Caffeine, Captopril, Carboplatin, Cephalothin, Ciprofloxacin, Cisplatin, Clofibrate, Cyclophosphamide, Cyclosporine A, Desmopressin acetate, Doxorubicin, Enalapril, Erythromycin, Ethinylestradiol, Ethionine, Gentamicin, Hexachlorobenzene, Imipramine, Indomethacin, Ketoconazole, Lomustine, Methyltestosterone, Monocrotaline, Nitrofurantoin, Omeprazole, Phenacetin, Phenylanthranilic acid, Phenylbutazone, Puromycin, Rifampicin, Thioacetamide, Triamterene, Valproic acid.

Table S10: Genes removed for lack of data for them concerning "receptor specific chemicals" in particular categories.

Category (Number of genes)	**Removed Genes**										
All database (25)	BCMO1	C12orf29	C16orf72	C8orf4	C9orf3	CBARA1	CCDC104	CCT8L1	CDC2L6	CES1	DLEU7
	GSTM1	FAM119A GVIN1	JHDM1D	KBTBD5	LEPREL1	NME2	PION	PRR13	RBM9	RPL10A	RPL7
	UGT2B7	UGT1A1									
Rat liver *in vitro* *(164)**	ABCC1	ADM2	ADO	AKAP7	AKNA	AKR1B10	AKR1C2	ALDH18A1	ALDH1A3	ALDH1L2	ALKBH5
	ASB3	ANK2 B4GALNT2	B9D2	BACH1	BCMO1	BEND6	BLVRB	BTF3L4	C12orf29	C16orf72	C8orf4
	CABC1	C9orf3 CBARA1	CCDC104	CCDC109B	CCL2	CCT8L1	CD27	CDC2L6	CDH24	CDSN	CES1
	COCH	CLIP4 CORO7	CRYBG3	CTSO	CTTN	CXCL5	CYB5R1	CYP1B1	DAAM2	DENND4C	DLEU7
	ECHS1	DLX2 EDA2R	EIF1	EXOC7	FAM119A	FAM159A	FAM65C	FAM69A	FBXL2	FNDC7	FREM2
	GLI2	FTH1 GPATCH3	GSTA1	GSTA4	GSTM1	GSTM3	GTPBP4	GVIN1	H2AFZ	HAX1	HDAC8
	HMHB1	HIST1H1C HRASLS2	IDS	IFT172	INPP5B	INSIG1	IRF4	JHDM1D	KBTBD5	KCNT2	LAYN
	LONP1	LEPREL1 LPL	LTA	MAGOHB	MAMDC2	MAPK8	MAZ	MEGF9	MFF	MSC	MT2A
	NME2	NCL NOS2	NPNT	NRBF2	NUPR1	OIT3	PAQR3	PDDC1	PGAP1	PHF10	PION
	PMAIP1	PLA2G4A POP5	PPP2R5A	PRDX1	PRR13	PRSS35	PSG5	PTBP1	PTGS2	RAB39B	RBM39
	RPL10A	RBM9 RPL13A	RPL7	RPRD2	RPS6	RUNX2	RXRA	SDCCAG8	SH3RF2	SLC16A14	SLC1A4
	SLC2A11	SLC1A7 SLC44A3	SLMAP	SLTM	SNAI2	SNRNP35	SNX1	SNX22	SOAT2	SPEN	SURF6
	TLCD2	TARS TMEFF2	TMEM74	TMTC2	TPX2	TRIM69	TSLP	UBE2L3	UBE2V2	UBE2W	UGT1A1
	UGT2B7	UGT1A6 VIPR1	VNN3	WDR27	WDR63	YLPM1	ZFC3H1	ZNF268			
Rat liver *in vivo* *(196)***	ABCC1	ACAP2	ADAM23	ADCY1	ADM2	ADO	AFF1	AGPAT9	AKAP7	AKNA	AKR1B10
	ALDH18A1	AKR1C2 ALDH1A3	ALDH1L2	ALKBH5	ALS2	ANK2	ASB3	B4GALNT2	B9D2	BACH1	BCMO1
	BLVRB	BEND6 BTF3L4	C12orf29	C16orf72	C8orf4	C9orf3	CASC5	CBARA1	CCDC104	CCL2	CCT8L1
	CDC2L6	CD27 CDH24	CDSN	CES1	CLIP4	COCH	COL24A1	CRYBG3	CTSO	CTTN	CYB5R1
	DAAM2	CYP1B1 DENND4C	DLEU7	DLX2	DUT	ECHS1	EDA2R	EDEM3	EIF1	ELF4	EXOC7
	FAM159A	FAM119A FAM65C	FAM69A	FBXL2	FLRT1	FNDC7	FREM2	FTH1	GLI2	GPATCH3	GSK3A
	GSTA4	GSTA1 GSTM1	GSTM3	GTPBP4	GVIN1	H2AFZ	HAX1	HDAC8	HIST1H1C	HMHB1	HRASLS2
	IDS	IDH1 IFT172	INPP5B	INSIG1	IRF4	JHDM1D	KBTBD5	KCNT2	KIF13B	KLHDC10	KRAS
	LEPREL1	LAYN LONP1	LPL	LRRK1	MAGOHB	MAMDC2	MANEA	MAPK8	MAZ	MEGF9	MFF
	MT2A	MSC MTM1	NCL	NME2	NOS2	NPNT	NRBF2	NREP	NUPR1	OIT3	ORMDL3
	PAQR3	OSMR PCDH7	PDDC1	PGAP1	PHF10	PION	PLA2G4A	PMAIP1	POP5	PPP2R5A	PRDX1
	PRSS35	PRR13 PSG5	PTBP1	PTGS2	PYCR1	RAB39B	RBM39	RBM9	RPL10A	RPL13A	RPL7
	RPS6	RPRD2 RUNX2	RXRA	SDCCAG8	SEMA3E	SH3RF2	SLC16A14	SLC16A9	SLC1A4	SLC1A7	SLC2A11
	SLMAP	SLC44A3 SLTM	SNAI2	SNRNP35	SNX1	SNX22	SOAT2	SPEN	SPRED1	SRPK1	SUB1
	TAB2	SURF6 TARS	TBCEL	TLCD2	TMEFF2	TMEM154	TMEM74	TMTC2	TNKS	TPX2	TRIM69
	UBE2L3	TSLP UBE2V2	UBE2W	UGT1A1	UGT1A6	UGT2B7	VIPR1	VNN3	WDR27	WDR63	YLPM1
	ZC3HAV1	ZBTB38 ZDHHC20	ZFC3H1	ZNF268							
Human liver *in vitro* *(44)****	ACOT2	AKAP2	AMACR	ANAPC1	ATF4	BCMO1	BGLAP	C12orf29	C16orf72	C8orf4	C9orf3
	CBARA1	CABC1 CCDC104	CCT8L1	CD302	CDC2L6	CES1	DLEU7	EIF3C	FAM119A	FAM188B	FBF1
	GSTM2	GSTM1 GVIN1	HHIPL1	JHDM1D	KBTBD5	LEPREL1	NME2	PAIP1	PION	PRR13	PTCD1
	RBM9	RAN RFFL	RPL10A	RPL7	TMIGD1	TXNDC5	UGT1A1	UGT2B7			

** All genes that were removed from the "Rat liver in vitro" category were removed from "Rat liver in vivo" category as well except five:* ***CABC1***, ***CCDC109B***, ***CORO7***, ***CXCL5*** *and* ***LTA***.

*** All genes that were removed from the "Rat liver in vivo" category were removed from "Rat liver in vitro" category as well except 37:* **ACAP2**, **ADAM23**, **ADCY1**, **AFF1**, **AGPAT9**, **ALS2**, **CASC5**, **COL24A1**, **DUT**, **EDEM3**, **ELF4**, **FLRT1**, **GSK3A**, **IDH1**, **KIF13B**, **KLHDC10**, **KRAS**, **LRRK1**, **MANEA**, **MTM1**, **NREP**, **ORMDL3**, **OSMR**, **PCDH7**, **PYCR1**, **SEMA3E**, **SLC16A9**, **SPRED1**, **SRPK1**, **SUB1**, **TAB2**, **TBCEL**, **TMEM154**, **TNKS**, **ZBTB38**, **ZC3HAV1** *and* ***ZDHHC20***.

****All genes that were had no rat data (in vitro or in vivo) had no human data neither (exception:* ***CABC1*** *removed from "Rat Liver in vivo" but not from human data), but the opposite is not always true: 18 of the genes that were removed for lack of human data, were kept for both rat categories (****ACOT2****,* ***AKAP2****,* ***AMACR****,* ***ANAPC1****,* ***ATF4****,* ***BGLAP****,* ***CD302****,* ***EIF3C****,* ***FAM188B****,* ***FBF1****,* ***GSTM2****,* ***HHIPL1****,* ***PAIP1****,* ***PTCD1****,* ***RAN****,* ***RFFL****,* ***TMIGD1*** *and* ***TXNDC5****).*

Table S11: Annotation of chemicals for **Figure 25**, **Figure 26** and **Figure 27**.

7	Chemical
1	2-Nitrofluorene
2	2-4-dinitrophenol
3	Acarbose
4	Acetamide
5	Acetamidofluorene
6	Acetaminophen
7	Acetazolamide
8	Adapin
9	Aflatoxin B1
10	Ajmaline
11	Allopurinol
12	Allyl alcohol
13	Alpidem
14	Amiodarone
15	Amitriptyline
16	Amphotericin B
17	Aspirin
18	Azathioprine
19	Bendazac
20	Benzbromarone
21	Benziodarone
22	Bromobenzene
23	Bromoethylamine
24	Bucetin
25	Buspirone
26	Buthionine sulfoximine
27	Butylated hydroxyanisole
28	Caffeine
29	Captopril
30	Carbamazepine
31	Carbon tetrachloride
32	Chloramphenicol
33	Chlormadinone
34	Chlormezanone
35	Chlorpheniramine
36	Chlorpromazine
37	Chlorpropamide
38	Cimetidine
39	Ciprofloxacin
40	Clofibrate

Nb	Chemical
41	Clomipramine
42	Clozapine
43	Colchicine
44	Coumarin
45	Cycloheximide
46	Cyclophosphamide
47	Cyclosporine A
48	Danazol
49	Dantrolene
50	Dexamethasone
51	Diazepam
52	Diclofenac
53	Diethylmaleate
54	Diltiazem
55	Disopyramide
56	Disulfiram
57	Doxorubicin
58	EMD335823
59	Enalapril
60	Erythromycin
61	Ethambutol
62	Ethanol
63	Ethinylestradiol
64	Ethionamide
65	Ethionine
66	Etoposide
67	Famotidine
68	Fenofibrate
69	Fluoxetine
70	Fluphenazine
71	Flutamide
72	Furosemide
73	Galactosamine
74	Gemfibrozil
75	Glibenclamide
76	Griseofulvin
77	Haloperidol
78	Hexachlorobenzene
79	HGF
80	Hydroxyzine

Nb	Chemical
81	Ibuprofen
82	IL1beta
83	IL6
84	Imipramine
85	Indomethacin
86	INFalpha
87	Iproniazid
88	Isoniazid
89	Ketoconazole
90	Labetalol
91	Lomustine
92	Lornoxicam
93	LPS
94	Mefenamic acid
95	Meloxicam
96	Metformin
97	Methapyrilene
98	Methimazole
99	Methyldopa
100	Methylene dianiline
101	Methyltestosterone
102	Mexiletine
103	Monocrotaline
104	Moxisylyte
105	N-methyl-N-nitrosourea
106	N-nitrosomorpholine
107	Naphthyl isothiocyanate
108	Naproxen
109	Nefazodone
110	Nicotinic_acid
111	Nifedipine
112	Nimesulide
113	Nitrofurantoin
114	Nitrofurazone
115	Nitrosodiethylamine
116	**Omeprazole**
117	Papaverine
118	Pemoline
119	Penicillamine
120	Perhexiline

Nb	Chemical
121	Phalloidin
122	Phenacetin
123	Phenobarbital
124	Phenylanthranilic acid
125	Phenylbutazone
126	Phenytoin
127	**Phorone**
128	Promethazine
129	Propranolol
130	Propylthiouracil
131	Quinidine
132	Ranitidine
133	Rifampicin
134	Rosiglitazone
135	Rotenone
136	Simvastatin
137	Sulfasalazine
138	Sulindac
139	Sulpiride
140	Tacrine
141	Tamoxifen
142	Tannic_acid
143	Terbinafine
144	Tetracycline
145	TGFbeta1
146	Theophylline
147	Thioacetamide
148	Thioridazine
149	Ticlopidine
150	Tiopronin
151	TNF
152	Tolbutamide
153	Triazolam
154	Trimethadione
155	Troglitazone
156	**Tunicamycin**
157	Valproic acid
158	Venlafaxine
159	Vitamin A
160	WY-14643

7.4 THE NRF2 SB MODEL CODE

```
# ------------------------------------------------------------------------------
# v7.11_Nrf2_GSH_KBrO4
#
# UNITS :
# ===============
#  Quantity in zeptomol (1 zeptomol = 1.00E-21 mol )
#  Volume in microm3 (1 microm3 = 1.00E-15 L )
#  Concentration in microM (1 microM = 1.00E-06 mol/L
#    = E-21/E-15 mol/L
#    = zmol/microm3 )
#  Time in s (seconds)
#
#
# PREFIXES :
# ===============
#  cyt   in Cytosol
#  ext   in Medium
#  nuc   in Nucleus
#  ss    Steady State Value
#  vitro in vitro KBrO3-GSH sub-model
# ------------------------------------------------------------------------------


# STATES :
# ===============

 States = {

 DCF,    # 2',7'- DiChloroFluorescein
 DCFDA,  # 2',7'- DiChloroFluoresceinDiAcetate

 cytAhR, # Cytosolic Arylhydrocarbon Receptor
 nucAhR, # Nucleic Arylhydrocarbon Receptor
 CYP1A1, # Cytochrome P450 1A1
 CYP1A1mRNA, # mRNA of CYP1A1
 Cystathionine,
 cytCysteine, # Cytosolic Cysteine
 extCysteine, # in Medium Cysteine
 GammaGlutamylCysteine,
 GCL, # Glutamate-Cysteine Ligase
 GCLC, # Glutamate-Cysteine Ligase, Catalytic subunit
 GCLCmRNA, # mRNA of GCLC
 GCLM, # Glutamate-Cysteine Ligase, Modifier subunit
 GCLMmRNA, # mRNA of GCLM
 cytGlutamate, # Cytosolic Glutamate
 cytGlutamicAminoAcid, # Cytosolic Glutamic Amino Acid
 extGlutamicAminoAcid, # in Medium Glutamic Amino Acid
 GlutamylAminoButyrate,
 cytGlycine, # Cytosolic Glycine
 GPX, # Glutathione Peroxidase
 GPXmRNA, # mRNA of GPX
 GS, # Glutathione Synthetase
 cytGSH, # Cytosolic Glutathione
 extGSH, # in Medium Glutathione

 vitroGSH, # to simulation Alice KBrO3 data

 GSmono, # Monomer of GS
 GSmRNA, # mRNA of GS
 cytGSSG, # Cytosolic Glutathione Disulfide (Reduced Glutathione)
 extGSSG, # in Medium Glutathione Disulfide (Reduced Glutathione)
 GST, # Glutathione S-Transferase
 GSTmono, # Monomer of GST
 GSTmRNA, # mRNA of GST
 Homocysteine,
 HMOX1, # Heme Oxygenase (decycling) 1
 HMOX1mRNA, # mRNA of HMOX1
 Keap1, # Kelch-Like ECH-associated protein 1
 Keap1o, # Oxidized form of Kelch-Like ECH-associated protein 1
 extLCysteinylGlycine, # in Medium L-Cysteinyl-Glycine
 cytMethionine, # Cytosolic Methionin
 MRP, # Multi-drug Resistance-associated Protein
```

```
 MRPmono, # Monomer of MRP
 MRPmRNA, # mRNA of MRP
 cytNrf2, # Cytosolic Nuclear Factor Erythroid 2 Like 2
 nucNrf2, # Nucleic Nuclear Factor Erythroid 2 Like 2
 Nrf2Keap1, # Nrf2-Keap1 Complex
 Nrf2Keap1o, # Nrf2-Keap1 Complex (Oxidized form)
 Nrf2mRNA, # mRNA of Nrf2 gene
 cytOphtalmicAcid, # Cytosolic Ophtalmic Acid
 cyt5Oxoproline, # Cytosolic 5-Oxoproline
 ROS, # Reactive Oxygen Species
 SAH, # S-Adenosyl-Homocysteine
 SAM, # S-Adenosyl-Methionine
 SRXN1, # Sulfiredoxin 1
 SRXN1mRNA, # mRNA of SRXN1
 cytXAhR, # Cytosolic X-AhR Complex
 nucXAhR }; # Nucleic X-AhR Complex

 Inputs = {KBrO3        }; # Concentration : Unit : microM (= zeptomol / microm3 )

 Outputs = {
 Synthesis_GammaGlutamylCysteine_GCL_term,
 Synthesis_GammaGlutamylCysteine_GCLC_term,
 Synthesis_GammaGlutamylCysteine,
 Synthesis1_cyt5Oxoproline,
 Synthesis_cytGSH,
 perc_vitroGSH};


# CONSTANTS :
# ===============
 # microm3

cytVolume = 2005.89; # Volume of Cytosol
nucVolume = 200.589; # Volume of Nucleus
extVolume = 1.43E+6; # Volume of Medium
totVolume = 1432206; # Total Volume

 # were State Variables under boundary conditions
 # zeptomol

 ATP = 10029450; # Adenosine TriPhosphate
 AminoButyrate = 20058.9; # 2-Amino-Butyrate
 Betaine = 100294.5;
extGlutamate = 8.58E+7; # in Medium Glutamate
extGlycine = 1.86E+9; # in Medium Glycine
 ssGSSG = 133812.9; # Glutathione Disulfide Steady State value
extMethionine = 4.29E+7; # in Medium Methionine
 MethylTetraHydroFolate= 8926.21; # 5-Methyl-Tetra-Hydro-Folate
 NADPH = 100294.5; # Nicotinamide Adenine Dinucleotide Phosphate
extOphtalmicAcid = 1.43E+6; # in Medium Ophtalmic Acid
ext5Oxoproline = 1.43E+6; # in Medium 5-Oxoproline
 ssROS = 20.0589; # ROS Steady State Value
 Serine = 1129316;


# PARAMETERS :
# ===============

# Eurotox_AOP KBrO3_DCF experiment parameters

k_kbro3    ; # 1/(microM.s) # GSH oxidation by KBrO3
k_direct   ; # 1/(microM.s) # Direct action of KBrO3 on DCFDA
k_oxdcf    ; # 1/(zmol.s)   # Oxidation of DFCDA by ROS
k_e_oxdcf  ; # 1/ s         # DCF bleaching
b_kbro3_in ; # -            # Intracellular action of KBrO3 on GSH

# Enzymatic Reactions Parameters are destined to Concentrations

# ROS Production and Basal level
Background_ProductionRate_ROS = 0.02; # microM / s

# Nrf2 Retention by Keap1 Cycle
Oxidation_Keap1 = 0.26; # 1 /(microM.s)
Reduction_Keap1o = 0.1; # 1 / s
Oxidation_Nrf2Keap1 = 0.26; # 1 /(microM.s)
Reduction_Nrf2Keap1o = 0.1; # 1 / s
DegradationRate_Nrf2Keap1 = 0.014; # 1 / s
```

```
DegradationRate_Nrf2Keap1o = 1.00E-4; # 1 / s
unBind_Nrf2Keap1 = 0.02; # 1 / s
Bind_cytNrf2_Keap1 = 2; # 1 /(microM.s)
unBind_Nrf2Keap1o = 0.02; # 1 / s
Bind_cytNrf2_Keap1o = 2; # 1 /(microM.s)

# XAhR Complex Synthesis in Cytosol and in Nucleus
unBind_cytXAhR = 0; # 1 / s
Bind_cytX_cytAhR = 0; # 1 /(microM.s)
unBind_nucXAhR = 0; # 1 / s
Bind_nucX_nucAhR = 0; # 1 /(microM.s)

# Transport of XAhR and Nrf2 between Cytosol and Nucleus
nuc_to_cyt_XAhR = 0.01; # 1 / s
cyt_to_nuc_XAhR = 0.1; # 1 / s
nuc_to_cyt_Nrf2 = 0.01; # 1 / s
cyt_to_nuc_Nrf2 = 0.02; # 1 / s

# Nrf2 Degradation in Nucleus
DegradationRate_nucNrf2 = 1.00E-4; # 1 / s

# CYP1A1 gene Transcription and CYP1A1 mRNA Translation
Transcription_CYP1A1_basal = 8.907817E-5; # zeptomol / s
XAhR_Transcription_CYP1A1_hill = 2; # -
XAhR_Transcription_CYP1A1_km = 0.0097641; # microM
XAhR_Transcription_CYP1A1_vmax = 4.23714E-6; # microM / s

DegradationRate_CYP1A1mRNA = 4.81E-5; # 1 / s
TranslationRate_CYP1A1mRNA = 0.0417; # 1 / s
DegradationRate_CYP1A1 = 3.85E-5; # 1 / s

# Nrf2 gene Transcription and Nrf2 mRNA Translation
Transcription_Nrf2_basal = 2.72546E-6; # zeptomol / s
Nrf2_Transcription_Nrf2_hill = 1; # -
Nrf2_Transcription_Nrf2_km = 0.00149636; # microM
Nrf2_Transcription_Nrf2_vmax = 2.70548E-8; # microM / s
XAhR_Transcription_Nrf2_hill = 2; # -
XAhR_Transcription_Nrf2_km = 0.00120393; # microM
XAhR_Transcription_Nrf2_vmax = 2.37072E-8; # microM / s
Mixed_Transcription_Nrf2_vmax = 2.51270E-8; # microM / s

DegradationRate_Nrf2mRNA = 6.43E-5; # 1 / s
TranslationRate_Nrf2mRNA = 0.0417; # 1 / s
DegradationRate_cytNrf2 = 1.00E-4; # 1 / s

# GS gene Transcription and GS mRNA Translation
Transcription_GS_basal = 0.0001208762; # zeptomol / s
Nrf2_Transcription_GS_hill = 2; # -
Nrf2_Transcription_GS_km = 0.00429446; # microM
Nrf2_Transcription_GS_vmax = 5.46569E-6; # microM / s

DegradationRate_GSmRNA = 4.83E-5; # 1 / s
TranslationRate_GSmRNA = 0.0417; # 1 / s
DegradationRate_GSmono = 3.86E-5; # 1 / s
unBind_GS = 0.02; # 1 / s
Bind_two_GSmono = 0.2; # 1 /(microM.s)
DegradationRate_GS = 1.93E-5; # 1 / s

# GCLC gene Transcription and GCLC mRNA Translation
Transcription_GCLC_basal = 0.0001695692; # zeptomol / s
Nrf2_Transcription_GCLC_hill = 3; # -
Nrf2_Transcription_GCLC_km = 0.00238843; # microM
Nrf2_Transcription_GCLC_vmax = 9.8907E-6; # microM / s

DegradationRate_GCLCmRNA = 4.83E-5; # 1 / s
TranslationRate_GCLCmRNA = 0.0417; # 1 / s
DegradationRate_GCLC = 3.86E-5; # 1 / s

# GCLM gene Transcription and GCLM mRNA Translation
Transcription_GCLM_basal = 9.166869E-5; # zeptomol / s
Nrf2_Transcription_GCLM_hill = 3; # -
Nrf2_Transcription_GCLM_km = 0.00382673; # microM
Nrf2_Transcription_GCLM_vmax = 1.27808E-5; # microM / s

DegradationRate_GCLMmRNA = 4.83E-5; # 1 / s
TranslationRate_GCLMmRNA = 0.0417; # 1 / s
DegradationRate_GCLM = 3.86E-5; # 1 / s
```

```
# GCL Synthesis and Degradation
unBind_GCL = 0.02; # 1 / s
Bind_GCLC_GCLM = 0.02; # 1 /(microM.s)
DegradationRate_GCL = 3.86E-5; # 1 / s

# GST gene Transcription and GST mRNA Translation
Transcription_GST_basal = 4.918909E-5; # zeptomol / s
Nrf2_Transcription_GST_hill = 2; # -
Nrf2_Transcription_GST_km = 0.00318146; # microM
Nrf2_Transcription_GST_vmax = 1.16523E-6; # microM / s
XAhR_Transcription_GST_hill = 2; # -
XAhR_Transcription_GST_km = 0.00310782; # microM
XAhR_Transcription_GST_vmax = 1.62064E-7; # microM / s
Mixed_Transcription_GST_vmax = 1.87320E-7; # microM / s

DegradationRate_GSTmRNA = 4.71E-5; # 1 / s
TranslationRate_GSTmRNA = 0.0417; # 1 / s
DegradationRate_GSTmono = 1.29E-4; # 1 / s
unBind_GST = 0.02; # 1 / s
Bind_two_GSTmono = 0.2; # 1 /(microM.s)
DegradationRate_GST = 1.29E-5; # 1 / s

# GPX gene Transcription and GPX mRNA Translation
Transcription_GPX_basal = 4.918909E-5; # zeptomol / s
Nrf2_Transcription_GPX_km = 0.00318146; # microM
Nrf2_Transcription_GPX_hill = 2; # -
Nrf2_Transcription_GPX_vmax = 1.16523E-6; # microM / s
XAhR_Transcription_GPX_km = 0.00310782; # microM
XAhR_Transcription_GPX_hill = 2; # -
XAhR_Transcription_GPX_vmax = 1.62064E-7; # microM / s
Mixed_Transcription_GPX_vmax = 1.87320E-7; # microM / s

DegradationRate_GPXmRNA = 4.71E-5; # 1 / s
TranslationRate_GPXmRNA = 0.0417; # 1 / s
DegradationRate_GPX = 1.29E-5; # 1 / s

# MRP gene Transcription and MRP mRNA Translation
Transcription_MRP_basal = 0.0001274065; # zeptomol / s
Nrf2_Transcription_MRP_km = 0.00272391; # microM
Nrf2_Transcription_MRP_hill = 2; # -
Nrf2_Transcription_MRP_vmax = 4.50481E-6; # microM / s
XAhR_Transcription_MRP_km = 0.00374616; # microM
XAhR_Transcription_MRP_hill = 1; # -
XAhR_Transcription_MRP_vmax = 2.05536E-6; # microM / s
Mixed_Transcription_MRP_vmax = 2.26571E-6; # microM / s

DegradationRate_MRPmRNA = 1.93E-5; # 1 / s
TranslationRate_MRPmRNA = 0.0417; # 1 / s
DegradationRate_MRPmono = 1.93E-5; # 1 / s
unBind_MRP = 0.02; # 1 / s
Bind_two_MRPmono = 0.01; # 1 /(microM.s)
DegradationRate_MRP = 7.15E-6; # 1 / s

# HMOX1 gene Transcription and HMOX1 mRNA Translation
Transcription_HMOX1_basal = 0.0001208762; # zeptomol / s
Nrf2_Transcription_HMOX1_km = 0.00429446; # microM
Nrf2_Transcription_HMOX1_hill = 2; # -
Nrf2_Transcription_HMOX1_vmax = 5.46569E-6; # microM / s

DegradationRate_HMOX1mRNA = 4.83E-5; # 1 / s
TranslationRate_HMOX1mRNA = 0.0417; # 1 / s
DegradationRate_HMOX1 = 1.93E-5; # 1 / s

# SRXN1 gene Transcription and SRXN1 mRNA Translation
Transcription_SRXN1_basal = 0.0001208762; # zeptomol / s
Nrf2_Transcription_SRXN1_km = 0.00429446; # microM
Nrf2_Transcription_SRXN1_hill = 2; # -
Nrf2_Transcription_SRXN1_vmax = 5.46569E-6; # microM / s

DegradationRate_SRXN1mRNA = 4.83E-5; # 1 / s
TranslationRate_SRXN1mRNA = 0.0417; # 1 / s
DegradationRate_SRXN1 = 1.93E-5; # 1 / s

# Transport of Methionine between Exterior and Cytosol
cyt_to_ext_Methionine = 2.78E-4; # 1 / s
ext_to_cyt_Methionine_km = 150; # microM
```

```
ext_to_cyt_Methionine_vmax = 0.230043; # microM / s

# S-Adenosyl-Methionine (SAM) Synthesis

# Enzyme1: Methionine-Adenosyl-Transferase-I : urn:miriam:ec-code:2.5.1.6
Synthesis1_SAM_inhib_GSSG = 2140; # microM
Synthesis1_SAM_inhib_SAM = 50; # microM
Synthesis1_SAM_km_Methionine = 41; # microM
Synthesis1_SAM_vmax = 0.16576; # microM / s

# Enzyme2: Methionine-Adenosyl-Transferase-III : urn:miriam:ec-code:2.5.1.6
Synthesis2_SAM_activ_Methionine_coef = 1.21; # -
Synthesis2_SAM_activ_SAM_coef = 2; # -
Synthesis2_SAM_activ_SAM_EC50 = 360; # microM
Synthesis2_SAM_activ_SAM_Top = 7.2; # -
Synthesis2_SAM_inhib_GSSG = 4030; # microM
Synthesis2_SAM_km_Methionine = 300; # microM
Synthesis2_SAM_vmax = 0.06467; # microM / s

# S-Adenosyl-Homocysteine (SAH) Synthesis

# Enzyme1: DNA-Methyl-Transferase : urn:miriam:ec-code:2.1.1.72
Synthesis1_SAH_inhib_SAH = 1.4; # microM
Synthesis1_SAH_km = 1.4; # microM
Synthesis1_SAH_vmax = 0.05664; # microM / s

# Enzyme2: Glycine-N-Methyl-Transferase : urn:miriam:ec-code:2.1.1.20
Synthesis2_SAH_inhib_SAH = 18; # microM
Synthesis2_SAH_km_Glycine = 130; # microM
Synthesis2_SAH_km_SAM = 32; # microM
Synthesis2_SAH_vmax = 0.15876; # microM / s

# S-Adenosyl-Homocysteine (SAH) Hydrolysis
# Enzyme : S-Adenosyl-Homocysteine-Hydrolase : urn:miriam:ec-code:3.3.1.1
Hydrolysis_SAH_equilibrium = 0.0602; # -
Hydrolysis_SAH_km_Homocysteine = 150; # microM
Hydrolysis_SAH_km_SAH = 6.5; # microM
Hydrolysis_SAH_vmax = 0.7843; # microM / s

# Methionine Synthesis

# Enzyme1: BetaineHomocysteineMethylTransferase: urn:miriam:ec-code:2.1.1.5
Synthesis1_Methionine_inhib_ROS = 0.01; # microM
Synthesis1_Methionine_km_Betaine = 100; # microM
Synthesis1_Methionine_km_Homocysteine = 12; # microM
Synthesis1_Methionine_vmax = 0.56869; # microM / s

# Enzyme2: Methionine-Synthase : urn:miriam:ec-code:2.1.1.13
Synthesis2_Methionine_inhib_ROS = 0.01; # microM
Synthesis2_Methionine_km_Homocysteine = 1; # microM
Synthesis2_Methionine_km_MethylTetraHydroFolate = 25; # microM
Synthesis2_Methionine_vmax = 0.086; # microM / s

# Cystathionine Synthesis
# Enzyme : Cystathionine-Beta-Synthase : urn:miriam:ec-code:4.2.1.22
Synthesis_Cystathionine_activ_SAM_SAH = 30; # microM
Synthesis_Cystathionine_activ_SAM_SAH_coef = 2; # -
Synthesis_Cystathionine_activ_SAM_SAH_Top = 1.086; # -
Synthesis_Cystathionine_activ_ROS = 0.035; # microM
Synthesis_Cystathionine_km_Homocysteine = 1000; # microM
Synthesis_Cystathionine_km_Serine = 2000; # microM
Synthesis_Cystathionine_vmax = 128.8875; # microM / s

# Cystathionine Hydrolysis
# Enzyme : Cystathionase : urn:miriam:ec-code:4.4.1.1
Hydrolysis_Cystathionine_km = 500; # microM
Hydrolysis_Cystathionine_vmax = 0.398669; # microM / s

# Transport of Cysteine and Glutamic Amino Acid between Exterior and Cytosol

ClearanceRate_extCysteine = 1.40E-4; # 1 / s
cyt_to_ext_Cysteine = 1.105E-4; # 1 / s
ext_to_cyt_Cysteine_km = 2100; # microM
ext_to_cyt_Cysteine_vmax = 0.384561; # microM / s

cyt_to_ext_GlutamicAminoAcid_equilibrium = 13.8755; # -
cyt_to_ext_GlutamicAminoAcid = 1000; # microM
```

```
ext_to_cyt_GlutamicAminoAcid_vmax = 27.18925; # microM / s
ext_to_cyt_GlutamicAminoAcid_km = 1000; # microM

# 5-Oxoproline Synthesis
# Enzyme : Gamma-Glutamyl-Cyclo-Transferase : urn:miriam:ec-code:2.3.2.4
# Source1: Gamma-Glutamyl-Cysteine
Synthesis1_cyt5Oxoproline_equilibrium = 2.80831; # microM
Synthesis1_cyt5Oxoproline_km_Cysteine = 2177.76; # microM
Synthesis1_cyt5Oxoproline_km_GammaGlutamylCysteine= 7.9799; # microM
Synthesis1_cyt5Oxoproline_km_5Oxoproline = 10002.5; # microM
Synthesis1_cyt5Oxoproline_vmax = 16.6223; # microM / s

# Enzyme : Gamma-Glutamyl-Cyclo-Transferase : urn:miriam:ec-code:2.3.2.4
# Source2: Glutamic-Amino-Acid
Synthesis2_cyt5Oxoproline_equilibrium = 4400.48; # -
Synthesis2_cyt5Oxoproline_km_GlutamicAminoAcid = 2200; # microM
Synthesis2_cyt5Oxoproline_km_5Oxoproline = 10002.5; # microM
Synthesis2_cyt5Oxoproline_vmax = 16.6223; # microM / s

# 5-Oxoproline Transport between Exterior and Cytosol
cyt_to_ext_5Oxoproline = 1.036E-4; # 1 / s

# 5-Oxoproline Hydrolysis
# Enzyme : 5-Oxoprolinase : urn:miriam:ec-code:3.5.2.9
Hydrolysis_cyt5Oxoproline_km_Glutamate = 1.18025; # microM
Hydrolysis_cyt5Oxoproline_km_5Oxoproline = 5.00484; # microM
Hydrolysis_cyt5Oxoproline_vmax = 235.258; # microM / s

# Transport of Glutamate between Exterior and Cytosol
cyt_to_ext_Glutamate = 3.6923E-7; # 1 / s
ext_to_cyt_Glutamate_km = 300; # microM
ext_to_cyt_Glutamate_vmax = 0.021314; # microM / s

# Gamma-Glutamyl-Cysteine and Glutamyl-Amino-Butyrate Synthesis
# Enzyme : Glutamyl-Cysteine-Ligase(GCL & GCLC): urn:miriam:ec-code:6.3.2.2
# Parameters expressed by enzyme name
# ATP parameters concern Gamma-Glutamyl-Cysteine Synthesis exclusively

freeGCL_ATP_inhib_GSH = 6500; # microM
boundGCL_ATP_inhib_GSH = 3900; # microM
GCL_km_ATP = 870; # microM

GCL_catalytic = 8.2; # 1 / s
GCL_disso_GammaGlutamylCysteine = 300; # microM
GCL_equilibrium = 0.002366; # 1 / microM
GCL_inhib_GSH = 8200; # microM
GCL_km_AminoButyrate = 2300; # microM
GCL_km_Cysteine = 100; # microM
GCL_km_Glutamate = 1900; # microM
GCL_km_GlutamylAminoButyrate = 10000; # microM

freeGCLC_ATP_inhib_GSH = 1300; # microM
boundGCLC_ATP_inhib_GSH = 400; # microM
GCLC_km_ATP = 5000; # microM

GCLC_catalytic = 1.9; # 1 / s
GCLC_disso_GammaGlutamylCysteine= 300; # microM
GCLC_equilibrium = 0.0024; # 1 / microM
GCLC_inhib_GSH = 8200; # microM
GCLC_km_AminoButyrate = 2300; # microM
GCLC_km_Cysteine = 100; # microM
GCLC_km_Glutamate = 1900; # microM
GCLC_km_GlutamylAminoButyrate = 10000; # microM

# Transport of Glycine between Exterior and Cytosol
cyt_to_ext_Glycine = 2.15E-5; # 1 / s
ext_to_cyt_Glycine_km = 150; # microM
ext_to_cyt_Glycine_vmax = 0.089136; # microM / s

# Glutathione (GSH) and Ophtalmic Acid Synthesis
# Enzyme : Glutathione-Synthetase (GS) : urn:miriam:ec-code:6.3.2.3
# Parameters expressed by enzyme name
# ATP parameters concern Glutathione (GSH) Synthesis exclusively
# GS_equilibrium are reaction specific (we have 2 different values)

GS_km_ATP = 70000; # microM
```

```
GS_catalytic = 6.5; # 1 / s
GS_disso_GSH = 30; # microM
GS_km_GammaGlutamylCysteine = 22; # microM
GS_km_GlutamylAminoButyrate = 200; # microM
GS_km_Glycine = 300; # microM
GS_km_OphtalmicAcid = 100; # microM

GS_equilibrium_GSH = 0.217774; # 1 / microM
GS_equilibrium_OphtalmicAcid = 0.002203; # 1 / microM

# Transport of Ophtalmic Acid between Exterior and Cytosol
cyt_to_ext_OphtalmicAcid = 0.25026; # 1 / s

# Export of Glutathione from Cytosol to Exterior
cyt_to_ext_highaffinity_GSH_km = 150; # microM
cyt_to_ext_highaffinity_GSH_vmax= 0.019; # microM / s
cyt_to_ext_lowaffinity_GSH_hill = 3; # -
cyt_to_ext_lowaffinity_GSH_km = 3000; # microM
cyt_to_ext_lowaffinity_GSH_vmax = 0.099; # microM / s

# Glutathione Hydrolysis in Blood
# Enzyme : Gamma-Glutamyl-Trans-Peptidase : urn:miriam:ec-code:2.3.2.2
Hydrolysis_extGSH_km_GlutamicAminoAcid = 979.802; # microM
Hydrolysis_extGSH_km_GSH = 670; # microM
Hydrolysis_extGSH_km_LCysteinylGlycine = 1090; # microM
Hydrolysis_extGSH_equilibrium = 99915.6; # microM
Hydrolysis_extGSH_vmax = 2.42926; # microM / s

# LCysteinylGlycine Hydrolysis
# Enzyme : Amino-Peptidase : urn:miriam:ec-code:3.4.11.2
Hydrolysis_extLCysteinylGlycine_km_Cysteine = 9988.69; # microM
Hydrolysis_extLCysteinylGlycine_km_LCysteinylGlycine= 2500; # microM
Hydrolysis_extLCysteinylGlycine_equlibrium = 98.3765; # -
Hydrolysis_extLCysteinylGlycine_vmax = 40.387; # microM / s

# Glutathione Oxidation
# Enzyme : Glutathione-Peroxidase : urn:miriam:ec-code:1.11.1.9
Oxidation_cytGSH_catalytic = 0.46935; # 1 / s
Oxidation_cytGSH_km_GSH = 1330; # microM
Oxidation_cytGSH_km_ROS = 0.09; # microM
Oxidation_cytGSH_coef_GSH = 2; # -

# Glutathione Disulfide Reduction
# Enzyme : Glutathione-Reductase : urn:miriam:ec-code:1.8.1.7
Reduction_cytGSSG_km_GSSG = 107; # microM
Reduction_cytGSSG_km_NADPH = 10.4; # microM
Reduction_cytGSSG_vmax = 0.04; # microM / s

# Export of Glutathione Disulfide from Cytosol to Exterior
cyt_to_ext_highaffinity_GSSG_activ_ROS = 0.01; # microM
cyt_to_ext_highaffinity_GSSG_km = 1250; # microM
cyt_to_ext_highaffinity_GSSG_vmax = 5.83E-5; # microM / s
cyt_to_ext_lowaffinity_GSSG_activ_ROS = 0.01; # microM
cyt_to_ext_lowaffinity_GSSG_km = 7710; # microM
cyt_to_ext_lowaffinity_GSSG_vmax = 0.001116; # microM / s

# Glutathione Disulfide Hydrolysis in Blood
Hydrolysis_extGSSG_vmax = 0.002839; # 1 / s

# Initial state values (zmol)

cytAhR_0 = 40.1178;
nucAhR_0 = 0;
CYP1A1_0 = 2005.86;
CYP1A1mRNA_0 = 1.85194;
Cystathionine_0 = 82072;
cytCysteine_0 = 123348;
extCysteine_0 = 432066;
GammaGlutamylCysteine_0 = 138662;
GCL_0 = 1164.2;
GCLC_0 = 2634.69;
GCLCmRNA_0 = 3.51648;
GCLM_0 = 888.065;
GCLMmRNA_0 = 1.8997;
cytGlutamate_0 = 1.91895E6;
cytGlutamicAminoAcid_0 = 6212.14;
extGlutamicAminoAcid_0 = 1.55258E6;
```

```
GlutamylAminoButyrate_0 = 1820.31;
cytGlycine_0 = 2.47981E6;
GPX_0 = 3400.22;
GPXmRNA_0 = 1.05187;
GS_0 = 2078.05;
cytGSH_0 = 2.4717E6;
extGSH_0 = 12934.7;
vitroGSH_0 = 1;      # to simulate Alice KBrO3 data
GSmono_0 = 645.938;
GSmRNA_0 = 2.52148;
cytGSSG_0 = 326079;
extGSSG_0 = 22.0471;
GST_0 = 358.627;
GSTmono_0 = 268.296;
GSTmRNA_0 = 1.05187;
Homocysteine_0 = 2646.54;
HMOX1_0 = 5447.97;
HMOX1mRNA_0 = 2.52148;
Keap1_0 = 38.8604;
Keap1o_0 = 1.10841;
extLCysteinylGlycine_0 = 7291.82;
cytMethionine_0 = 59056;
MRP_0 = 10691.3;
MRPmono_0 = 6550.31;
MRPmRNA_0 = 6.698;
cytNrf2_0 = 0.12548;
nucNrf2_0 = 0.0248475;
Nrf2Keap1_0 = 0.144365;
Nrf2Keap1o_0 =0.00458466;
Nrf2mRNA_0 = 0.0488393;
cytOphtalmicAcid_0 = 2029.27;
cyt5Oxoproline_0 = 6265.33;
ROS_0 = 22.0145;
SAH_0 = 48279.1;
SAM_0 = 50307.8;
SRXN1_0 = 5447.97;
SRXN1mRNA_0 = 2.52148;
cytXAhR_0 = 0;
nucXAhR_0 = 0;


# fold changes for initial state values

FC_GPX7mRNA = 1;
FC_GCLCmRNA = 1;
FC_GCLMmRNA = 1;
FC_NFE2L2mRNA = 1;
FC_GSSmRNA = 1;
FC_ABCC1mRNA = 1;


# statistical parameters
sd1;
sd2;
sd3;
sd4;
sd5;

# INITIALIZE:
# ===========

Initialize {

 # Initial state values (zeptomol)

 cytAhR = cytAhR_0;
 nucAhR = nucAhR_0;
 CYP1A1 = CYP1A1_0;
 CYP1A1mRNA = CYP1A1mRNA_0;
 Cystathionine = Cystathionine_0;
 cytCysteine = cytCysteine_0;
 extCysteine = extCysteine_0;
 GammaGlutamylCysteine = GammaGlutamylCysteine_0;
 GCL = GCL_0;
 GCLC = GCLC_0;
 GCLCmRNA = GCLCmRNA_0 * FC_GCLCmRNA; # modulable
 GCLM = GCLM_0;
```

```
 GCLMmRNA = GCLMmRNA_0 * FC_GCLMmRNA; # modulable
 cytGlutamate = cytGlutamate_0;
 cytGlutamicAminoAcid = cytGlutamicAminoAcid_0;
 extGlutamicAminoAcid = extGlutamicAminoAcid_0;
 GlutamylAminoButyrate = GlutamylAminoButyrate_0;
 cytGlycine = cytGlycine_0;
 GPX = GPX_0;
 GPXmRNA = GPXmRNA_0 * FC_GPX7mRNA; # modulable
 GS = GS_0;
 cytGSH = cytGSH_0;
 extGSH = extGSH_0;

 vitroGSH = vitroGSH_0;

 GSmono = GSmono_0;
 GSmRNA = GSmRNA_0 * FC_GSSmRNA; # modulable
 cytGSSG = cytGSSG_0;
 extGSSG = extGSSG_0;
 GST = GST_0;
 GSTmono = GSTmono_0;
 GSTmRNA = GSTmRNA_0;
 Homocysteine = Homocysteine_0;
 HMOX1 = HMOX1_0;
 HMOX1mRNA = HMOX1mRNA_0;
 Keap1 = Keap1_0;
 Keap1o = Keap1o_0;
 extLCysteinylGlycine = extLCysteinylGlycine_0;
 cytMethionine = cytMethionine_0;
 MRP = MRP_0;
 MRPmono = MRPmono_0;
 MRPmRNA = MRPmRNA_0 * FC_ABCC1mRNA; # modulable
 cytNrf2 = cytNrf2_0;
 nucNrf2 = nucNrf2_0;
 Nrf2Keap1 = Nrf2Keap1_0;
 Nrf2Keap1o = Nrf2Keap1o_0;
 Nrf2mRNA = Nrf2mRNA_0 * FC_NFE2L2mRNA; # modulable
 cytOphtalmicAcid = cytOphtalmicAcid_0;
 cyt5Oxoproline = cyt5Oxoproline_0;
 ROS = ROS_0;
 SAH = SAH_0;
 SAM = SAM_0;
 SRXN1 = SRXN1_0;
 SRXN1mRNA = SRXN1mRNA_0;
 cytXAhR = cytXAhR_0;
 nucXAhR = nucXAhR_0;

} # End of Initialize


# DYNAMICS:
# =========

# Reaction Rates are "Mass / Time" (Unit: zeptomol/s)
# for all state variables except for those
# starting with ext and wall : "Concentration / Time" (Unit: microM /s)

Dynamics {

 # Eurotox_AOP KBrO3_DCF experiment dynamics
 KBrO3_Action_cytGSH      = b_kbro3_in * k_kbro3    * KBrO3 *   cytGSH ;
 KBrO3_Action_extGSH      =              k_kbro3    * KBrO3 *   extGSH ;
 KBrO3_Action_vitroGSH    =              k_kbro3    * KBrO3 * vitroGSH ;
 KBrO3_DirectAction_DCFDA =              k_direct   * KBrO3 *    DCFDA ;
 KBrO3_Oxidation_DCFDA    =              k_oxdcf    * ROS   *    DCFDA ;
 DCF_Bleaching            =              k_e_oxdcf  * DCF              ;

 Background_Production_ROS = cytVolume
 * Background_ProductionRate_ROS;

 # Nrf2 Retention by Keap1 Cycle
 Redox_Keap1 = cytVolume
 * (Oxidation_Keap1 * (Keap1 / cytVolume )
  * (ROS / cytVolume )
 - Reduction_Keap1o * (Keap1o / cytVolume ));
 Redox_Nrf2Keap1 = cytVolume
 * (Oxidation_Nrf2Keap1 * (Nrf2Keap1 / cytVolume )
  * (ROS / cytVolume )
```

```
- Reduction_Nrf2Keap1o * (Nrf2Keap1o / cytVolume ));
Degradation_Nrf2Keap1 = cytVolume
* (DegradationRate_Nrf2Keap1 * (Nrf2Keap1 / cytVolume ));
Degradation_Nrf2Keap1o = cytVolume
* (DegradationRate_Nrf2Keap1o * (Nrf2Keap1o / cytVolume ));
Complex_Nrf2Keap1 = cytVolume
* (Bind_cytNrf2_Keap1 * (cytNrf2 / cytVolume )
 * (Keap1 / cytVolume )
- unBind_Nrf2Keap1 * (Nrf2Keap1 / cytVolume ));
Complex_Nrf2Keap1o = cytVolume
* (Bind_cytNrf2_Keap1o * (cytNrf2 / cytVolume )
 * (Keap1o / cytVolume )
- unBind_Nrf2Keap1o * (Nrf2Keap1o / cytVolume ));

# XAhR Complex Synthesis in Cytosol and in Nucleus
Complex_cytXAhR = 0;
Complex_nucXAhR = 0;

# Transport of XAhR and Nrf2 between Cytosol and Nucleus
CytNuc_Transport_XAhR = nucVolume
* (cyt_to_nuc_XAhR * (cytXAhR / cytVolume )
- nuc_to_cyt_XAhR * (nucXAhR / nucVolume ));
CytNuc_Transport_Nrf2 = nucVolume
* (cyt_to_nuc_Nrf2 * (cytNrf2 / cytVolume )
- nuc_to_cyt_Nrf2 * (nucNrf2 / nucVolume ));

# Nrf2 Degradation in Nucleus
Degradation_nucNrf2 = nucVolume
* (DegradationRate_nucNrf2 * (nucNrf2 / nucVolume ));

# CYP1A1 gene Transcription and CYP1A1 mRNA Translation
Transcription_CYP1A1 =
Transcription_CYP1A1_basal + nucVolume
* ((XAhR_Transcription_CYP1A1_vmax * (pow((nucXAhR / nucVolume ),
XAhR_Transcription_CYP1A1_hill ))) /((pow
(XAhR_Transcription_CYP1A1_km, XAhR_Transcription_CYP1A1_hill ))
+ (pow((nucXAhR / nucVolume ), XAhR_Transcription_CYP1A1_hill))));
Degradation_CYP1A1mRNA = cytVolume
* (DegradationRate_CYP1A1mRNA * (CYP1A1mRNA / cytVolume ));
Translation_CYP1A1mRNA = cytVolume
* (TranslationRate_CYP1A1mRNA * (CYP1A1mRNA / cytVolume ));
Degradation_CYP1A1 = cytVolume
* (DegradationRate_CYP1A1 * (CYP1A1 / cytVolume ));

# Nrf2 gene Transcription and Nrf2 mRNA Translation
Transcription_Nrf2 = Transcription_Nrf2_basal + nucVolume
*(((Nrf2_Transcription_Nrf2_vmax * (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_Nrf2_hill))) /((pow
(Nrf2_Transcription_Nrf2_km, Nrf2_Transcription_Nrf2_hill))
+ (pow((nucNrf2 / nucVolume), Nrf2_Transcription_Nrf2_hill))))
+ ((XAhR_Transcription_Nrf2_vmax * (pow((nucXAhR / nucVolume),
XAhR_Transcription_Nrf2_hill))) /((pow
(XAhR_Transcription_Nrf2_km, XAhR_Transcription_Nrf2_hill))
+ (pow((nucXAhR / nucVolume), XAhR_Transcription_Nrf2_hill))))
- ((Mixed_Transcription_Nrf2_vmax * (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_Nrf2_hill)) * (pow((nucXAhR / nucVolume),
XAhR_Transcription_Nrf2_hill))) /(((pow(Nrf2_Transcription_Nrf2_km,
Nrf2_Transcription_Nrf2_hill)) + (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_Nrf2_hill))) *((pow (XAhR_Transcription_Nrf2_km,
XAhR_Transcription_Nrf2_hill)) + (pow((nucXAhR / nucVolume),
XAhR_Transcription_Nrf2_hill))))));
Degradation_Nrf2mRNA = cytVolume
* (DegradationRate_Nrf2mRNA * (Nrf2mRNA / cytVolume ));
Translation_Nrf2mRNA = cytVolume
* (TranslationRate_Nrf2mRNA * (Nrf2mRNA / cytVolume ));
Degradation_cytNrf2 = cytVolume
* (DegradationRate_cytNrf2 * (cytNrf2 / cytVolume ));

# GS gene Transcription and GS mRNA Translation
Transcription_GS =
Transcription_GS_basal + nucVolume
* ((Nrf2_Transcription_GS_vmax * (pow((nucNrf2 / nucVolume ),
Nrf2_Transcription_GS_hill ))) /((pow
(Nrf2_Transcription_GS_km, Nrf2_Transcription_GS_hill ))
+ (pow((nucNrf2 / nucVolume ), Nrf2_Transcription_GS_hill))));
Degradation_GSmRNA = cytVolume
* (DegradationRate_GSmRNA * (GSmRNA / cytVolume ));
```

```
Translation_GSmRNA = cytVolume
* (TranslationRate_GSmRNA * (GSmRNA / cytVolume ));
Degradation_GSmono = cytVolume
* (DegradationRate_GSmono * (GSmono / cytVolume ));
Complex_GS = cytVolume
* (Bind_two_GSmono * (GSmono / cytVolume )
 * (GSmono / cytVolume )
- unBind_GS * (GS / cytVolume ));
Degradation_GS = cytVolume
* (DegradationRate_GS * (GS / cytVolume ));

# GCLC gene Transcription and GCLC mRNA Translation
Transcription_GCLC =
Transcription_GCLC_basal + nucVolume
* ((Nrf2_Transcription_GCLC_vmax * (pow((nucNrf2 / nucVolume ),
Nrf2_Transcription_GCLC_hill ))) /((pow
(Nrf2_Transcription_GCLC_km, Nrf2_Transcription_GCLC_hill ))
+ (pow((nucNrf2 / nucVolume ), Nrf2_Transcription_GCLC_hill ))));
Degradation_GCLCmRNA = cytVolume
* (DegradationRate_GCLCmRNA * (GCLCmRNA / cytVolume ));
Translation_GCLCmRNA = cytVolume
* (TranslationRate_GCLCmRNA * (GCLCmRNA / cytVolume ));
Degradation_GCLC = cytVolume
* (DegradationRate_GCLC * (GCLC / cytVolume ));

# GCLM gene Transcription and GCLM mRNA Translation
Transcription_GCLM =
Transcription_GCLM_basal + nucVolume
* ((Nrf2_Transcription_GCLM_vmax * (pow((nucNrf2 / nucVolume ),
Nrf2_Transcription_GCLM_hill ))) /((pow
(Nrf2_Transcription_GCLM_km, Nrf2_Transcription_GCLM_hill ))
+ (pow((nucNrf2 / nucVolume ), Nrf2_Transcription_GCLM_hill ))));
Degradation_GCLMmRNA = cytVolume
* (DegradationRate_GCLMmRNA * (GCLMmRNA / cytVolume ));
Translation_GCLMmRNA = cytVolume
* (TranslationRate_GCLMmRNA * (GCLMmRNA / cytVolume ));
Degradation_GCLM = cytVolume
* (DegradationRate_GCLM * (GCLM / cytVolume ));

# GCL Synthesis and Degradation
Complex_GCL = cytVolume
* (Bind_GCLC_GCLM * (GCLC / cytVolume )
 * (GCLM / cytVolume )
- unBind_GCL * (GCL / cytVolume ));
Degradation_GCL = cytVolume
* (DegradationRate_GCL * (GCL / cytVolume ));

# GST gene Transcription and GST mRNA Translation
Transcription_GST = Transcription_GST_basal + nucVolume
*(((Nrf2_Transcription_GST_vmax * (pow((nucNrf2/ nucVolume),
Nrf2_Transcription_GST_hill))) /((pow
(Nrf2_Transcription_GST_km, Nrf2_Transcription_GST_hill))
+ (pow((nucNrf2 / nucVolume), Nrf2_Transcription_GST_hill))))
+ ((XAhR_Transcription_GST_vmax* (pow((nucXAhR / nucVolume),
XAhR_Transcription_GST_hill))) /((pow
(XAhR_Transcription_GST_km, XAhR_Transcription_GST_hill))
+ (pow((nucXAhR / nucVolume), XAhR_Transcription_GST_hill))))
- ((Mixed_Transcription_GST_vmax* (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_GST_hill)) * (pow((nucXAhR / nucVolume),
XAhR_Transcription_GST_hill))) /(((pow(Nrf2_Transcription_GST_km,
Nrf2_Transcription_GST_hill)) + (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_GST_hill))) *((pow(XAhR_Transcription_GST_km,
XAhR_Transcription_GST_hill)) + (pow((nucXAhR / nucVolume),
XAhR_Transcription_GST_hill))))));
Degradation_GSTmRNA = cytVolume
* (DegradationRate_GSTmRNA * (GSTmRNA / cytVolume ));
Translation_GSTmRNA = cytVolume
* (TranslationRate_GSTmRNA * (GSTmRNA / cytVolume ));
Degradation_GSTmono = cytVolume
* (DegradationRate_GSTmono * (GSTmono / cytVolume ));
Complex_GST = cytVolume
* (Bind_two_GSTmono * (GSTmono / cytVolume )
 * (GSTmono / cytVolume )
- unBind_GST * (GST / cytVolume ));
Degradation_GST = cytVolume
* (DegradationRate_GST * (GST / cytVolume ));
```

```
# GPX gene Transcription and GPX mRNA Translation
Transcription_GPX= Transcription_GPX_basal+ nucVolume
*(((Nrf2_Transcription_GPX_vmax* (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_GPX_hill))) /((pow
(Nrf2_Transcription_GPX_km, Nrf2_Transcription_GPX_hill))
+ (pow((nucNrf2 / nucVolume), Nrf2_Transcription_GPX_hill))))
+ ((XAhR_Transcription_GPX_vmax* (pow((nucXAhR / nucVolume),
XAhR_Transcription_GPX_hill))) /((pow
(XAhR_Transcription_GPX_km, XAhR_Transcription_GPX_hill))
+ (pow((nucXAhR / nucVolume), XAhR_Transcription_GPX_hill))))
- ((Mixed_Transcription_GPX_vmax* (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_GPX_hill)) * (pow((nucXAhR / nucVolume),
XAhR_Transcription_GPX_hill))) /((((pow(Nrf2_Transcription_GPX_km,
Nrf2_Transcription_GPX_hill)) + (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_GPX_hill))) *((pow (XAhR_Transcription_GPX_km,
XAhR_Transcription_GPX_hill)) + (pow((nucXAhR / nucVolume),
XAhR_Transcription_GPX_hill))))));
Degradation_GPXmRNA = cytVolume *
(DegradationRate_GPXmRNA * (GPXmRNA / cytVolume ));
Translation_GPXmRNA = cytVolume *
(TranslationRate_GPXmRNA * (GPXmRNA / cytVolume ));
Degradation_GPX = cytVolume *
(DegradationRate_GPX * (GPX / cytVolume ));

# MRP gene Transcription and MRP mRNA Translation
Transcription_MRP=Transcription_MRP_basal+ nucVolume
*(((Nrf2_Transcription_MRP_vmax* (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_MRP_hill))) /((pow
(Nrf2_Transcription_MRP_km, Nrf2_Transcription_MRP_hill))
+ (pow((nucNrf2 / nucVolume), Nrf2_Transcription_MRP_hill))))
+ ((XAhR_Transcription_MRP_vmax* (pow((nucXAhR / nucVolume),
XAhR_Transcription_MRP_hill))) /((pow
(XAhR_Transcription_MRP_km, XAhR_Transcription_MRP_hill))
+ (pow((nucXAhR / nucVolume), XAhR_Transcription_MRP_hill))))
- ((Mixed_Transcription_MRP_vmax* (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_MRP_hill)) * (pow((nucXAhR / nucVolume),
XAhR_Transcription_MRP_hill))) /((((pow(Nrf2_Transcription_MRP_km,
Nrf2_Transcription_MRP_hill)) + (pow((nucNrf2 / nucVolume),
Nrf2_Transcription_MRP_hill))) *((pow (XAhR_Transcription_MRP_km,
XAhR_Transcription_MRP_hill)) + (pow((nucXAhR / nucVolume),
XAhR_Transcription_MRP_hill))))));
Degradation_MRPmRNA = cytVolume
* (DegradationRate_MRPmRNA * (MRPmRNA / cytVolume ));
Translation_MRPmRNA = cytVolume
* (TranslationRate_MRPmRNA * (MRPmRNA / cytVolume ));
Degradation_MRPmono = cytVolume
* (DegradationRate_MRPmono * (MRPmono / cytVolume ));
Complex_MRP = cytVolume
* (Bind_two_MRPmono * (MRPmono / cytVolume )
 * (MRPmono / cytVolume )
- unBind_MRP * (MRP / cytVolume ));
Degradation_MRP = cytVolume
* (DegradationRate_MRP * (MRP / cytVolume ));

# HMOX1 gene Transcription and HMOX1 mRNA Translation
Transcription_HMOX1 =
Transcription_HMOX1_basal + nucVolume
* ((Nrf2_Transcription_HMOX1_vmax * (pow((nucNrf2 / nucVolume ),
Nrf2_Transcription_HMOX1_hill ))) /((pow
(Nrf2_Transcription_HMOX1_km, Nrf2_Transcription_HMOX1_hill ))
+ (pow((nucNrf2 / nucVolume ), Nrf2_Transcription_HMOX1_hill))));
Degradation_HMOX1mRNA = cytVolume
* (DegradationRate_HMOX1mRNA * (HMOX1mRNA / cytVolume ));
Translation_HMOX1mRNA = cytVolume
* (TranslationRate_HMOX1mRNA * (HMOX1mRNA / cytVolume ));
Degradation_HMOX1 = cytVolume
* (DegradationRate_HMOX1 * (HMOX1 / cytVolume ));

# SRXN1 gene Transcription and SRXN1 mRNA Translation
Transcription_SRXN1 =
Transcription_SRXN1_basal + nucVolume
* ((Nrf2_Transcription_SRXN1_vmax * (pow((nucNrf2 / nucVolume ),
Nrf2_Transcription_SRXN1_hill ))) /((pow
(Nrf2_Transcription_SRXN1_km, Nrf2_Transcription_SRXN1_hill ))
+ (pow((nucNrf2 / nucVolume ), Nrf2_Transcription_SRXN1_hill))));
Degradation_SRXN1mRNA = cytVolume
* (DegradationRate_SRXN1mRNA * (SRXN1mRNA / cytVolume ));
```

```
Translation_SRXN1mRNA = cytVolume
* (TranslationRate_SRXN1mRNA * (SRXN1mRNA / cytVolume ));
Degradation_SRXN1 = cytVolume
* (DegradationRate_SRXN1 * (SRXN1 / cytVolume ));

# Transport of Methionine between Exterior and Cytosol
ExtCyt_Transport_Methionine = cytVolume
*(((ext_to_cyt_Methionine_vmax * (extMethionine / extVolume ))
/ (ext_to_cyt_Methionine_km + (extMethionine / extVolume )))
- (cyt_to_ext_Methionine * (cytMethionine / cytVolume )));

# S-Adenosyl-Methionine (SAM) Synthesis

# Enzyme1: Methionine-Adenosyl-Transferase-I : urn:miriam:ec-code:2.5.1.6
Synthesis1_SAM = cytVolume
*(((Synthesis1_SAM_vmax * (cytMethionine / cytVolume ))
/ (Synthesis1_SAM_km_Methionine + (cytMethionine / cytVolume )
+ ((Synthesis1_SAM_km_Methionine * (SAM / cytVolume ))
/ Synthesis1_SAM_inhib_SAM ))) *
((Synthesis1_SAM_inhib_GSSG + ( ssGSSG / cytVolume ))
/ (Synthesis1_SAM_inhib_GSSG + (cytGSSG / cytVolume ))));

# Enzyme2: Methionine-Adenosyl-Transferase-III : urn:miriam:ec-code:2.5.1.6
Synthesis2_SAM = cytVolume
* (((Synthesis2_SAM_vmax * (pow((cytMethionine / cytVolume ),
Synthesis2_SAM_activ_Methionine_coef)))/((pow
(Synthesis2_SAM_km_Methionine, Synthesis2_SAM_activ_Methionine_coef ))
+ (pow((cytMethionine / cytVolume),Synthesis2_SAM_activ_Methionine_coef))))
* (1+ ((Synthesis2_SAM_activ_SAM_Top * (pow((SAM / cytVolume ),
Synthesis2_SAM_activ_SAM_coef ))) / ((pow
(Synthesis2_SAM_activ_SAM_EC50, Synthesis2_SAM_activ_SAM_coef ))
+ (pow((SAM / cytVolume),Synthesis2_SAM_activ_SAM_coef )))))
* ((Synthesis2_SAM_inhib_GSSG + ( ssGSSG / cytVolume ))
/ (Synthesis2_SAM_inhib_GSSG + (cytGSSG / cytVolume ))));

# S-Adenosyl-Homocysteine (SAH) Synthesis

# Enzyme1: DNA-Methyl-Transferase : urn:miriam:ec-code:2.1.1.72
Synthesis1_SAH = cytVolume
* ((Synthesis1_SAH_vmax * (SAM / cytVolume ))
/ (Synthesis1_SAH_km * (1 +
 ((SAH / cytVolume )
/ Synthesis1_SAH_inhib_SAH )) + (SAM / cytVolume )));

# Enzyme2: Glycine-N-Methyl-Transferase : urn:miriam:ec-code:2.1.1.20
Synthesis2_SAH = cytVolume
* ((Synthesis2_SAH_vmax * (cytGlycine / cytVolume )
 * (SAM / cytVolume ))
/ ((Synthesis2_SAH_km_Glycine + (cytGlycine / cytVolume ))
* (Synthesis2_SAH_km_SAM + (SAM / cytVolume ))
* (1  + (SAH / cytVolume )
/ Synthesis2_SAH_inhib_SAH  )));

# S-Adenosyl-Homocysteine (SAH) Hydrolysis
# Enzyme : S-Adenosyl-Homocysteine-Hydrolase : urn:miriam:ec-code:3.3.1.1
Hydrolysis_SAH = cytVolume
* ((Hydrolysis_SAH_vmax * ((SAH / cytVolume )
- ((Homocysteine / cytVolume ) / Hydrolysis_SAH_equilibrium )))
/ (Hydrolysis_SAH_km_SAH + (SAH / cytVolume )
+ ((Hydrolysis_SAH_km_SAH * (Homocysteine / cytVolume ))
/ Hydrolysis_SAH_km_Homocysteine  )));

# Methionine Synthesis

# Enzyme1: BetaineHomocysteineMethylTransferase: urn:miriam:ec-code:2.1.1.5
Synthesis1_Methionine = cytVolume
*(((Synthesis1_Methionine_vmax * (Betaine / cytVolume )
 * (Homocysteine / cytVolume ))
/ ((Synthesis1_Methionine_km_Homocysteine + (Homocysteine / cytVolume ))
* (Synthesis1_Methionine_km_Betaine + (Betaine / cytVolume ))))
* ((Synthesis1_Methionine_inhib_ROS + (ssROS / cytVolume ))
/ (Synthesis1_Methionine_inhib_ROS + ( ROS / cytVolume ))));

# Enzyme2: Methionine-Synthase : urn:miriam:ec-code:2.1.1.13
Synthesis2_Methionine = cytVolume
*(((Synthesis2_Methionine_vmax * (Homocysteine / cytVolume )
* (MethylTetraHydroFolate/ cytVolume ))
```

```
/ ((Synthesis2_Methionine_km_Homocysteine + (Homocysteine / cytVolume ))
* (Synthesis2_Methionine_km_MethylTetraHydroFolate
+ (MethylTetraHydroFolate/ cytVolume))))
* ((Synthesis2_Methionine_inhib_ROS + (ssROS / cytVolume ))
/ (Synthesis2_Methionine_inhib_ROS + ( ROS / cytVolume ))));

# Cystathionine Synthesis
# Enzyme : Cystathionine-Beta-Synthase : urn:miriam:ec-code:4.2.1.22
Synthesis_Cystathionine = cytVolume
*(((Synthesis_Cystathionine_vmax * (Serine / cytVolume )
 * (Homocysteine / cytVolume ))
/ ((Synthesis_Cystathionine_km_Homocysteine
+ (Homocysteine / cytVolume ))
* (Synthesis_Cystathionine_km_Serine + (Serine / cytVolume ))))
* ((Synthesis_Cystathionine_activ_SAM_SAH_Top
* (pow(((SAM / cytVolume ) + (SAH / cytVolume )),
Synthesis_Cystathionine_activ_SAM_SAH_coef ))) /((pow
(Synthesis_Cystathionine_activ_SAM_SAH,
Synthesis_Cystathionine_activ_SAM_SAH_coef ))
+ (pow(((SAM / cytVolume ) + (SAH / cytVolume )),
Synthesis_Cystathionine_activ_SAM_SAH_coef ))))
* ((Synthesis_Cystathionine_activ_ROS + (ROS / cytVolume ))
/ (Synthesis_Cystathionine_activ_ROS + (ssROS / cytVolume ))));

# Cystathionine Hydrolysis
# Enzyme : Cystathionase : urn:miriam:ec-code:4.4.1.1
Hydrolysis_Cystathionine = cytVolume
* ((Hydrolysis_Cystathionine_vmax * (Cystathionine / cytVolume ))
/ (Hydrolysis_Cystathionine_km + (Cystathionine / cytVolume )));

# Transport of Cysteine and Glutamic Amino Acid between Exterior and Cytosol
ExtCyt_Transport_Cysteine = cytVolume
*(((ext_to_cyt_Cysteine_vmax * (extCysteine / extVolume ))
/ (ext_to_cyt_Cysteine_km + (extCysteine / extVolume )))
- (cyt_to_ext_Cysteine * (cytCysteine / cytVolume )));
Clearance_extCysteine = extVolume
* (ClearanceRate_extCysteine * (extCysteine / extVolume ));

ExtCyt_Transport_GlutamicAminoAcid = cytVolume
* ((ext_to_cyt_GlutamicAminoAcid_vmax *
(extGlutamicAminoAcid / extVolume ) * ( 1 - (
(cytGlutamicAminoAcid / cytVolume )
/ (cyt_to_ext_GlutamicAminoAcid_equilibrium
* ((extGlutamicAminoAcid /extVolume))))))/ (ext_to_cyt_GlutamicAminoAcid_km
* (1 + ((extGlutamicAminoAcid/extVolume) / ext_to_cyt_GlutamicAminoAcid_km)
+ ((cytGlutamicAminoAcid / cytVolume ) / cyt_to_ext_GlutamicAminoAcid
  ))));

# 5-Oxoproline Synthesis
# Enzyme : Gamma-Glutamyl-Cyclo-Transferase : urn:miriam:ec-code:2.3.2.4
# Source1: Gamma-Glutamyl-Cysteine
Synthesis1_cyt5Oxoproline = cytVolume
* ((Synthesis1_cyt5Oxoproline_vmax
* (GammaGlutamylCysteine / cytVolume )
* (1 -(((cytCysteine / cytVolume ) * (cyt5Oxoproline/ cytVolume ))
/ (Synthesis1_cyt5Oxoproline_equilibrium
* (GammaGlutamylCysteine / cytVolume )  ))))
/ (Synthesis1_cyt5Oxoproline_km_GammaGlutamylCysteine
* (1 +((cytCysteine / cytVolume )
/ Synthesis1_cyt5Oxoproline_km_Cysteine  )
+ ((GammaGlutamylCysteine / cytVolume )
/ Synthesis1_cyt5Oxoproline_km_GammaGlutamylCysteine )
+ ((cyt5Oxoproline / cytVolume )
/ Synthesis1_cyt5Oxoproline_km_5Oxoproline  )
+(((cytCysteine / cytVolume ) * (cyt5Oxoproline/ cytVolume ))
/ (Synthesis1_cyt5Oxoproline_km_Cysteine
* Synthesis1_cyt5Oxoproline_km_5Oxoproline )))));

# Enzyme : Gamma-Glutamyl-Cyclo-Transferase : urn:miriam:ec-code:2.3.2.4
# Source2: Glutamic-Amino-Acid
Synthesis2_cyt5Oxoproline = cytVolume
* ((Synthesis2_cyt5Oxoproline_vmax *
(cytGlutamicAminoAcid / cytVolume ) * ( 1 -
((cyt5Oxoproline / cytVolume )
/ (Synthesis2_cyt5Oxoproline_equilibrium
* (cytGlutamicAminoAcid / cytVolume )  ))))
/ (Synthesis2_cyt5Oxoproline_km_GlutamicAminoAcid * ( 1 +
```

```
((cytGlutamicAminoAcid / cytVolume )
/ Synthesis2_cyt5Oxoproline_km_GlutamicAminoAcid )
+ ((cyt5Oxoproline / cytVolume )
/ Synthesis2_cyt5Oxoproline_km_5Oxoproline ))));

# 5-Oxoproline Transport between Exterior and Cytosol
ExtCyt_Transport_5Oxoproline = cytVolume
* (cyt_to_ext_5Oxoproline * ((cyt5Oxoproline/ cytVolume )
 - (ext5Oxoproline/ extVolume )));

# 5-Oxoproline Hydrolysis
# Enzyme : 5-Oxoprolinase : urn:miriam:ec-code:3.5.2.9
Hydrolysis_cyt5Oxoproline = cytVolume
* ((Hydrolysis_cyt5Oxoproline_vmax * (cyt5Oxoproline/ cytVolume ))
/ (Hydrolysis_cyt5Oxoproline_km_5Oxoproline
+ (cyt5Oxoproline / cytVolume ) + ((cytGlutamate / cytVolume )
* (Hydrolysis_cyt5Oxoproline_km_5Oxoproline
/ Hydrolysis_cyt5Oxoproline_km_Glutamate ))));

# Transport of Glutamate between Exterior and Cytosol
ExtCyt_Transport_Glutamate = cytVolume
*(((ext_to_cyt_Glutamate_vmax * (extGlutamate / extVolume ))
/ (ext_to_cyt_Glutamate_km + (extGlutamate / extVolume )))
- (cyt_to_ext_Glutamate * (cytGlutamate / cytVolume )));

# Gamma-Glutamyl-Cysteine and Glutamyl-Amino-Butyrate Synthesis
# Enzyme : Glutamyl-Cysteine-Ligase(GCL & GCLC): urn:miriam:ec-code:6.3.2.2
# Parameters expressed by enzyme name
# ATP parameters concern Gamma-Glutamyl-Cysteine Synthesis exclusively

# Gamma-Glutamyl-Cysteine Synthesis
GCL_term_ATP =
((ATP / cytVolume ) / (GCL_km_ATP
* (1 +((cytGSH / cytVolume ) / freeGCL_ATP_inhib_GSH ))
+ ((ATP / cytVolume )
* (1 +((cytGSH / cytVolume ) / boundGCL_ATP_inhib_GSH )))));
Synthesis_GammaGlutamylCysteine_GCL_term = cytVolume
* ((GCL_catalytic * (GCL / cytVolume )
* (cytCysteine / cytVolume ) * (cytGlutamate / cytVolume )
* GCL_term_ATP * (1 -
((GammaGlutamylCysteine / cytVolume ) / (GCL_equilibrium
* (cytCysteine / cytVolume ) * (cytGlutamate / cytVolume )))))
/ (GCL_km_Cysteine * GCL_km_Glutamate
* (1 +((AminoButyrate / cytVolume ) / GCL_km_AminoButyrate )
+ ((cytCysteine / cytVolume ) / GCL_km_Cysteine)) * ( 1 +
((GammaGlutamylCysteine / cytVolume ) / GCL_disso_GammaGlutamylCysteine)
+ ((cytGlutamate / cytVolume ) / GCL_km_Glutamate )
+ ((cytGSH / cytVolume ) / GCL_inhib_GSH )
+ ((GlutamylAminoButyrate / cytVolume ) / GCL_km_GlutamylAminoButyrate))));
GCLC_term_ATP =
((ATP / cytVolume ) / (GCLC_km_ATP
* (1 +((cytGSH / cytVolume ) / freeGCLC_ATP_inhib_GSH ))
+ ((ATP / cytVolume )
* (1 +((cytGSH / cytVolume ) / boundGCLC_ATP_inhib_GSH )))));
Synthesis_GammaGlutamylCysteine_GCLC_term = cytVolume
* ((GCLC_catalytic * (GCLC / cytVolume )
* (cytCysteine / cytVolume ) * (cytGlutamate / cytVolume )
* GCLC_term_ATP * (1 -
((GammaGlutamylCysteine / cytVolume ) / (GCLC_equilibrium
* (cytCysteine / cytVolume ) * (cytGlutamate / cytVolume )))))
/ (GCLC_km_Cysteine * GCLC_km_Glutamate
* (1 +((AminoButyrate / cytVolume ) / GCLC_km_AminoButyrate )
+ ((cytCysteine / cytVolume ) / GCLC_km_Cysteine)) * ( 1 +
((GammaGlutamylCysteine / cytVolume ) / GCLC_disso_GammaGlutamylCysteine)
+ ((cytGlutamate / cytVolume ) / GCLC_km_Glutamate )
+ ((cytGSH / cytVolume ) / GCLC_inhib_GSH )
+ ((GlutamylAminoButyrate / cytVolume ) / GCLC_km_GlutamylAminoButyrate))));
Synthesis_GammaGlutamylCysteine =
Synthesis_GammaGlutamylCysteine_GCL_term
+ Synthesis_GammaGlutamylCysteine_GCLC_term;

# Glutamyl-Amino-Butyrate Synthesis
Synthesis_GlutamylAminoButyrate_GCL_term = cytVolume
* ((GCL_catalytic * (GCL / cytVolume )
* (AminoButyrate / cytVolume ) * (cytGlutamate / cytVolume )
* (1 - (1 / ( GCL_equilibrium * (cytGlutamate / cytVolume )))))
/ (GCL_km_AminoButyrate * GCL_km_Glutamate
```

```
* (1 +((AminoButyrate / cytVolume ) / GCL_km_AminoButyrate )
+ ((cytCysteine / cytVolume ) / GCL_km_Cysteine)) * ( 1 +
((GammaGlutamylCysteine / cytVolume ) / GCL_disso_GammaGlutamylCysteine)
+ ((cytGlutamate / cytVolume ) / GCL_km_Glutamate )
+ ((cytGSH / cytVolume ) / GCL_inhib_GSH )
+ ((GlutamylAminoButyrate / cytVolume ) / GCL_km_GlutamylAminoButyrate))));
Synthesis_GlutamylAminoButyrate_GCLC_term = cytVolume
* ((GCLC_catalytic * (GCLC / cytVolume )
* (AminoButyrate / cytVolume ) * (cytGlutamate / cytVolume )
* (1 - (1 / ( GCLC_equilibrium * (cytGlutamate / cytVolume )))))
/ (GCLC_km_AminoButyrate * GCLC_km_Glutamate
* (1 +((AminoButyrate / cytVolume ) / GCLC_km_AminoButyrate )
+ ((cytCysteine / cytVolume ) / GCLC_km_Cysteine)) * ( 1 +
((GammaGlutamylCysteine / cytVolume ) / GCLC_disso_GammaGlutamylCysteine)
+ ((cytGlutamate / cytVolume ) / GCLC_km_Glutamate )
+ ((cytGSH / cytVolume ) / GCLC_inhib_GSH )
+ ((GlutamylAminoButyrate / cytVolume ) / GCLC_km_GlutamylAminoButyrate))));
Synthesis_GlutamylAminoButyrate =
Synthesis_GlutamylAminoButyrate_GCL_term
+ Synthesis_GlutamylAminoButyrate_GCLC_term;

# Transport of Glycine between Exterior and Cytosol
ExtCyt_Transport_Glycine = cytVolume
*(((ext_to_cyt_Glycine_vmax * (extGlycine / extVolume ))
/ (ext_to_cyt_Glycine_km + (extGlycine / extVolume )))
- (cyt_to_ext_Glycine * (cytGlycine / cytVolume )));

# Glutathione (GSH) and Ophtalmic Acid Synthesis
# Enzyme : Glutathione-Synthetase (GS) : urn:miriam:ec-code:6.3.2.3
# Parameters expressed by enzyme name
# ATP parameters concern Glutathione (GSH) Synthesis exclusively
# GS_equilibrium are reaction specific (we have 2 different values)

Synthesis_cytGSH = cytVolume
* ((GS_catalytic * (GS / cytVolume )
* (GammaGlutamylCysteine / cytVolume ) * (cytGlycine / cytVolume )
* (1 -((cytGSH / cytVolume )
/ (GS_equilibrium_GSH * (cytGlycine / cytVolume )
* (GammaGlutamylCysteine / cytVolume)))))
/ (GS_km_GammaGlutamylCysteine * GS_km_Glycine
* (1 +((cytGlycine / cytVolume ) / GS_km_Glycine )
+ ((GlutamylAminoButyrate / cytVolume ) /
GS_km_GlutamylAminoButyrate )) * (1 +
((GammaGlutamylCysteine / cytVolume ) / GS_km_GammaGlutamylCysteine )
+ ((cytGSH / cytVolume ) / GS_disso_GSH )
+ ((cytOphtalmicAcid / cytVolume )
/ GS_km_OphtalmicAcid ))) * (ATP / cytVolume )
/ (GS_km_ATP + (ATP / cytVolume )));

Synthesis_cytOphtalmicAcid = cytVolume
* ((GS_catalytic * (GS / cytVolume )
* (GlutamylAminoButyrate / cytVolume ) * (cytGlycine / cytVolume )
* (1 -((cytOphtalmicAcid / cytVolume )
/ (GS_equilibrium_OphtalmicAcid * (cytGlycine / cytVolume )
* (GlutamylAminoButyrate / cytVolume)))))
/ (GS_km_GammaGlutamylCysteine * GS_km_Glycine
* (1 +((cytGlycine / cytVolume ) / GS_km_Glycine )
+ ((GlutamylAminoButyrate / cytVolume ) /
GS_km_GlutamylAminoButyrate )) * (1 +
((GammaGlutamylCysteine / cytVolume ) / GS_km_GammaGlutamylCysteine )
+ ((cytGSH / cytVolume ) / GS_disso_GSH )
+ ((cytOphtalmicAcid / cytVolume ) / GS_km_OphtalmicAcid ))));

# Transport of Ophtalmic Acid between Exterior and Cytosol
ExtCyt_Transport_OphtalmicAcid = cytVolume
* (cyt_to_ext_OphtalmicAcid
* ((cytOphtalmicAcid / cytVolume  )
- (extOphtalmicAcid / extVolume  )));

# Export of Glutathione from Cytosol to Exterior
ExtCyt_highaffinityTransport_GSH = cytVolume
* ((cyt_to_ext_highaffinity_GSH_vmax * (cytGSH / cytVolume ))
/ (cyt_to_ext_highaffinity_GSH_km + (cytGSH / cytVolume )));
ExtCyt_lowaffinityTransport_GSH = cytVolume
* ((cyt_to_ext_lowaffinity_GSH_vmax *
pow((cytGSH / cytVolume ),
cyt_to_ext_lowaffinity_GSH_hill )) / (pow
```

```
(cyt_to_ext_lowaffinity_GSH_km,
cyt_to_ext_lowaffinity_GSH_hill ) + pow
((cytGSH / cytVolume ),
cyt_to_ext_lowaffinity_GSH_hill  )));

# Glutathione Hydrolysis in Blood
# Enzyme : Gamma-Glutamyl-Trans-Peptidase : urn:miriam:ec-code:2.3.2.2
Hydrolysis_extGSH = extVolume
* ((Hydrolysis_extGSH_vmax * (extGSH / extVolume )
* (1 -(((extGlutamicAminoAcid/extVolume) * (extLCysteinylGlycine/extVolume))
/ (Hydrolysis_extGSH_equilibrium * (extGSH / extVolume )))))
/ (Hydrolysis_extGSH_km_GSH *
(1 +((extGlutamicAminoAcid / extVolume)
/ Hydrolysis_extGSH_km_GlutamicAminoAcid  )
+ ((extGSH / extVolume ) / Hydrolysis_extGSH_km_GSH )
+ ((extLCysteinylGlycine / extVolume )
/ Hydrolysis_extGSH_km_LCysteinylGlycine  )
+(((extGlutamicAminoAcid / extVolume ) * (extLCysteinylGlycine/extVolume))
/ (Hydrolysis_extGSH_km_GlutamicAminoAcid
* Hydrolysis_extGSH_km_LCysteinylGlycine )))));

# LCysteinylGlycine Hydrolysis
# Enzyme : Amino-Peptidase : urn:miriam:ec-code:3.4.11.2
Hydrolysis_extLCysteinylGlycine = extVolume
* ((Hydrolysis_extLCysteinylGlycine_vmax
* (1 -((extCysteine / extVolume )
/ (Hydrolysis_extLCysteinylGlycine_equlibrium
* (extLCysteinylGlycine / extVolume)))) * (extLCysteinylGlycine/extVolume))
/ (Hydrolysis_extLCysteinylGlycine_km_LCysteinylGlycine
* (1 +((extCysteine / extVolume )
/ Hydrolysis_extLCysteinylGlycine_km_Cysteine )
+ ((extLCysteinylGlycine / extVolume )
/ Hydrolysis_extLCysteinylGlycine_km_LCysteinylGlycine ))));

# Glutathione Oxidation
# Enzyme : Glutathione-Peroxidase : urn:miriam:ec-code:1.11.1.9
Oxidation_cytGSH = cytVolume
* ((Oxidation_cytGSH_catalytic * (ROS / cytVolume )
 * (GPX / cytVolume )
* pow((cytGSH / cytVolume ), Oxidation_cytGSH_coef_GSH ))
/ ((Oxidation_cytGSH_km_ROS + (ROS / cytVolume ))
* pow((Oxidation_cytGSH_km_GSH + (cytGSH / cytVolume ))
, Oxidation_cytGSH_coef_GSH )));

# Glutathione Disulfide Reduction
# Enzyme : Glutathione-Reductase : urn:miriam:ec-code:1.8.1.7
Reduction_cytGSSG = cytVolume
* ((Reduction_cytGSSG_vmax * (cytGSSG / cytVolume )
 * (NADPH / cytVolume ))
/ ((Reduction_cytGSSG_km_NADPH + (NADPH / cytVolume ))
* (Reduction_cytGSSG_km_GSSG + (cytGSSG / cytVolume ))));

# Export of Glutathione Disulfide from Cytosol to Exterior
ExtCyt_highaffinityTransport_GSSG = cytVolume
* ((cyt_to_ext_highaffinity_GSSG_vmax * (( ROS / cytVolume )
+ cyt_to_ext_highaffinity_GSSG_activ_ROS)* (cytGSSG / cytVolume ))
 /(((ssROS / cytVolume )
+ cyt_to_ext_highaffinity_GSSG_activ_ROS)*((cytGSSG / cytVolume )
+ cyt_to_ext_highaffinity_GSSG_km  )));
ExtCyt_lowaffinityTransport_GSSG = cytVolume
* ((cyt_to_ext_lowaffinity_GSSG_vmax * (( ROS / cytVolume )
+ cyt_to_ext_lowaffinity_GSSG_activ_ROS)* (cytGSSG / cytVolume ))
 /(((ssROS / cytVolume )
+ cyt_to_ext_lowaffinity_GSSG_activ_ROS)* ((cytGSSG / cytVolume )
+ cyt_to_ext_lowaffinity_GSSG_km  )));

# Glutathione Disulfide Hydrolysis in Blood
Hydrolysis_extGSSG = extVolume
* (Hydrolysis_extGSSG_vmax * (extGSSG / extVolume ));


# Differentials by State Variable
# --------------------------------
dt(cytAhR) =  - Complex_cytXAhR;
dt(nucAhR) =  - Complex_nucXAhR;
dt(CYP1A1) = + Translation_CYP1A1mRNA - Degradation_CYP1A1;
dt(CYP1A1mRNA) = + Transcription_CYP1A1 - Degradation_CYP1A1mRNA;
```

```
dt(Cystathionine) = + Synthesis_Cystathionine - Hydrolysis_Cystathionine;
dt(cytCysteine) = + ExtCyt_Transport_Cysteine
+ Hydrolysis_Cystathionine
+ Synthesis1_cyt5Oxoproline - Synthesis_GammaGlutamylCysteine;
dt(extCysteine) = + Hydrolysis_extLCysteinylGlycine
- ExtCyt_Transport_Cysteine
+ 2 * Hydrolysis_extGSSG - Clearance_extCysteine;
dt(DCF)   = + KBrO3_DirectAction_DCFDA + KBrO3_Oxidation_DCFDA - DCF_Bleaching ;
dt(DCFDA) = 0 ;
dt(GammaGlutamylCysteine) = + Synthesis_GammaGlutamylCysteine - Synthesis_cytGSH
  - Synthesis1_cyt5Oxoproline;
dt(GCL) = + Complex_GCL - Degradation_GCL;
dt(GCLC) = + Translation_GCLCmRNA - Complex_GCL
  - Degradation_GCLC;
dt(GCLCmRNA) = + Transcription_GCLC - Degradation_GCLCmRNA;
dt(GCLM) = + Translation_GCLMmRNA - Complex_GCL
  - Degradation_GCLM;
dt(GCLMmRNA) = + Transcription_GCLM - Degradation_GCLMmRNA;
dt(cytGlutamate) = + Hydrolysis_cyt5Oxoproline - Synthesis_GammaGlutamylCysteine
+ ExtCyt_Transport_Glutamate - Synthesis_GlutamylAminoButyrate;
dt(cytGlutamicAminoAcid) = + ExtCyt_Transport_GlutamicAminoAcid - Synthesis2_cyt5Oxoproline;
dt(extGlutamicAminoAcid) = + Hydrolysis_extGSH
- ExtCyt_Transport_GlutamicAminoAcid;
dt(GlutamylAminoButyrate) = + Synthesis_GlutamylAminoButyrate - Synthesis_cytOphtalmicAcid;
dt(cytGlycine) = + ExtCyt_Transport_Glycine - Synthesis2_SAH
  - Synthesis_cytGSH
 - Synthesis_cytOphtalmicAcid;

dt(GPX) = + Translation_GPXmRNA - Degradation_GPX;

dt(GPXmRNA) = + Transcription_GPX - Degradation_GPXmRNA;

dt(GS) = + Complex_GS - Degradation_GS;

dt(cytGSH)   = + Synthesis_cytGSH - 2 * Oxidation_cytGSH + 2 * Reduction_cytGSSG
                         - ExtCyt_highaffinityTransport_GSH - ExtCyt_lowaffinityTransport_GSH
               - KBrO3_Action_cytGSH ;

dt(extGSH)   = + ExtCyt_highaffinityTransport_GSH + ExtCyt_lowaffinityTransport_GSH
               - Hydrolysis_extGSH - KBrO3_Action_extGSH ;

dt(vitroGSH) = - KBrO3_Action_vitroGSH ; # in vitro GSH sub-model

dt(GSmono) = + Translation_GSmRNA - 2 * Complex_GS
  - Degradation_GSmono;

dt(GSmRNA) = + Transcription_GS - Degradation_GSmRNA;

dt(cytGSSG) = + Oxidation_cytGSH - Reduction_cytGSSG
  - ExtCyt_highaffinityTransport_GSSG
  - ExtCyt_lowaffinityTransport_GSSG;

dt(extGSSG) = + ExtCyt_highaffinityTransport_GSSG
 + ExtCyt_lowaffinityTransport_GSSG
 - Hydrolysis_extGSSG;


dt(GST) = + Complex_GST - Degradation_GST;

dt(GSTmono) = + Translation_GSTmRNA - 2 * Complex_GST
  - Degradation_GSTmono;

dt(GSTmRNA) = + Transcription_GST - Degradation_GSTmRNA;
dt(Homocysteine) = + Hydrolysis_SAH - Synthesis1_Methionine
  - Synthesis2_Methionine
 - Synthesis_Cystathionine;
dt(HMOX1) = + Translation_HMOX1mRNA - Degradation_HMOX1;
dt(HMOX1mRNA) = + Transcription_HMOX1 - Degradation_HMOX1mRNA;
dt(Keap1) = + Degradation_Nrf2Keap1 - Complex_Nrf2Keap1
  - Redox_Keap1;
dt(Keap1o) = + Degradation_Nrf2Keap1o - Complex_Nrf2Keap1o
+ Redox_Keap1;

dt(extLCysteinylGlycine) = + Hydrolysis_extGSH - Hydrolysis_extLCysteinylGlycine;
dt(cytMethionine) = + Synthesis1_Methionine - Synthesis1_SAM
+ Synthesis2_Methionine - Synthesis2_SAM
+ ExtCyt_Transport_Methionine;
```

```
 dt(MRP) = + Complex_MRP - Degradation_MRP;
 dt(MRPmono) = + Translation_MRPmRNA - 2 * Complex_MRP
   - Degradation_MRPmono;
 dt(MRPmRNA) = + Transcription_MRP - Degradation_MRPmRNA;
 dt(cytNrf2) = + Translation_Nrf2mRNA - Degradation_cytNrf2
 - Complex_Nrf2Keap1o
 - Complex_Nrf2Keap1
 - CytNuc_Transport_Nrf2;
 dt(nucNrf2) = + CytNuc_Transport_Nrf2 - Degradation_nucNrf2;
 dt(Nrf2Keap1) = + Complex_Nrf2Keap1 - Redox_Nrf2Keap1
   - Degradation_Nrf2Keap1;
 dt(Nrf2Keap1o) = + Redox_Nrf2Keap1 + Complex_Nrf2Keap1o
   - Degradation_Nrf2Keap1o;
 dt(Nrf2mRNA) = + Transcription_Nrf2 - Degradation_Nrf2mRNA;
 dt(cytOphtalmicAcid) = + Synthesis_cytOphtalmicAcid - ExtCyt_Transport_OphtalmicAcid;
 dt(cyt5Oxoproline) = + Synthesis1_cyt5Oxoproline - Hydrolysis_cyt5Oxoproline
 + Synthesis2_cyt5Oxoproline - ExtCyt_Transport_5Oxoproline;
 dt(ROS) = + Background_Production_ROS
  - Redox_Keap1
  - Redox_Nrf2Keap1
  - Oxidation_cytGSH;

dt(SAH) = + Synthesis1_SAH
 + Synthesis2_SAH - Hydrolysis_SAH;
 dt(SAM) = + Synthesis1_SAM - Synthesis1_SAH
 + Synthesis2_SAM - Synthesis2_SAH;
 dt(SRXN1) = + Translation_SRXN1mRNA - Degradation_SRXN1;
 dt(SRXN1mRNA) = + Transcription_SRXN1 - Degradation_SRXN1mRNA;
 dt(cytXAhR) = + Complex_cytXAhR - CytNuc_Transport_XAhR;
 dt(nucXAhR) = + Complex_nucXAhR
 + CytNuc_Transport_XAhR;

} # End of Dynamics

CalcOutputs {
  perc_vitroGSH = (vitroGSH / vitroGSH_0 ) * 100   ;
}

End.
```

ABSTRACT

New understanding of biology shows more and more that the mechanisms that underlie toxicity are complex and involve multiple biological processes and pathways. Adverse outcome pathways (AOPs) and systems biology (SB) can be appropriate tools for studying toxicology at this level of complexity. This PhD thesis focuses on the elaboration of a SB model of the role of the Nrf2 pathway in the control of oxidative stress. The model's calibration with experimental data (obtained with RPTEC/TERT1 renal cells exposed to various doses of potassium bromate) comprised several rounds of hypotheses stating/verification, through which new reactions were progressively added to the model. Some of these new hypotheses (*e.g.*, direct action of potassium bromate on DCF, bleaching of DCF with time, *etc.*) could be confirmed by future experiments. Considered in a wider framework, this SB model was then evaluated and compared to two other computational models (*i.e.*, an empirical dose-response statistical model and a dynamic Bayesian model) for the quantification of a 'chronic kidney disease' AOP. All parameter calibrations were done by MCMC simulations with the *GNU MCSim* software with a quantification of uncertainties associated with predictions. Even though the SB model was indeed complex to conceive, it o ffers insight in biology that the other approaches could not afford. In addition, using multiple toxicogenomic databases; interactions and cross-talks of the Nrf2 pathway with two other toxicity pathways (*i.e.*, AhR and ATF4) were examined. The results of this last analysis suggest adding new AhR contribution to the control of some of the Nrf2 genes in our SB model (*e.g.*, *HMOX1*, *SRXN1* and *GCLM*), and integrating in it description of the ATF4 pathway (partially at least). Despites their complexity, precise SB models are precious investments for future developments in predictive toxicology.

Keywords: toxicology, Nrf2, oxidative stress, systems biology, adverse outcome pathways, bioinformatics, toxicogenomics.

RESUME

Avec les nouvelles avancées en biologie et toxicologie, on constate de plus en plus la complexité des mécanismes et le grand nombre de voies de toxicité. Les concepts de 'biologie systémique' (SB) et de 'voies des effets indésirables' (adverse outcome pathway, AOP) pourraient être des outils appropriés pour l'étude de la toxicologie à ces niveaux de complexité élevés. Le point central du travail de cette thèse est le développement d'un modèle de SB du rôle de la voie de signalisation Nrf2 dans le contrôle du stress oxydant. Pour la calibration de ce modèle avec des données expérimentales (exposition des cellules rénales RPTEC/TERT1 à différentes doses de bromate de potassium), plusieurs cycles de proposition/vérification d'hypothèses ont progressivement contribué à l'ajout de nouvelles réactions. Ces nouvelles hypothèses (par exemple : action directe du bromate de potassium sur le DCF, atténuation de la fluorescence du DCF avec le temps, *etc.*) devraient être confirmées par de futures expérimentations. Ce modèle de SB a été ensuite utilisé pour la quantification d'un AOP de l'insuffisance rénale chronique et comparé à deux autres approches: l'utilisation de modèles statistiques empiriques et celle d'un réseau Bayésien dynamique. Les calibrations des paramètres ont été effectuées par chaînes de Markov simulées MCMC avec le logiciel *GNU MCSim* avec une quantification des incertitudes associées aux prédictions. Même si la mise au point du modèle SB a été une tâche complexe, la compréhension de la biologie qu'offre ce modèle n'est pas accessible aux deux autres approches. Nous avons aussi évalué les interactions entre Nrf2 et deux autres voies de toxicité, AhR et ATF4, dans le cadre d'une analyse utilisant des données de toxico-génomique provenant de trois projets différents. Les résultats de cette dernière analyse suggèrent d'ajouter au modèle SB de Nrf2 la co-activation par AhR de plusieurs gènes (par exemple, *HMOX1*, *SRXN1* et *GCLM*) ainsi que d'associer (au moins partiellement) à ce modèle la voie ATF4. Malgré leur complexité, les modèles SB constituent un investissement intéressant pour le développement de la toxicologie prédictive.

Mots-Clés: toxicologie, Nrf2, stress oxydant, biologie systémique, voies des effets indésirables, bio-informatique, toxico-génomique.