# Automated workflows for modelling chemical fate, kinetics and toxicity

José Vicente Sala Benito [a], Alicia Paini [a,*], Andrea-Nicole Richarz [b], Thorsten Meinl [c], Michael R. Berthold [d], Mark T.D. Cronin [b], Andrew P. Worth [a]

[a] *Chemical Safety and Alternative Methods Unit, EURL ECVAM, Directorate F – Health, Consumers and Reference Materials, Joint Research Centre, European Commission, Ispra, Italy*
[b] *Liverpool John Moores University, School of Pharmacy and Biomolecular Sciences, Byrom Street, Liverpool L3 3AF, UK*
[c] *KNIME.com AG, Zurich, Switzerland*
[d] *Universität Konstanz, Fachbereich Informatik und Informationswissenschaft, Box 712, 78457 Konstanz, Germany*

## ARTICLE INFO

## ABSTRACT

Automation is universal in today's society, from operating equipment such as machinery, in factory processes, to self-parking automobile systems. While these examples show the efficiency and effectiveness of automated mechanical processes, automated procedures that support the chemical risk assessment process are still in their infancy. Future human safety assessments will rely increasingly on the use of automated models, such as physiologically based kinetic (PBK) and dynamic models and the virtual cell based assay (VCBA). These biologically-based models will be coupled with chemistry-based prediction models that also automate the generation of key input parameters such as physicochemical properties. The development of automated software tools is an important step in harmonising and expediting the chemical safety assessment process. In this study, we illustrate how the KNIME Analytics Platform can be used to provide a user-friendly graphical interface for these biokinetic models, such as PBK models and VCBA, which simulates the fate of chemicals in vivo within the body and in vitro test systems respectively.

## 1. Introduction

Future human safety assessments will rely increasingly on the use of multi-scale models, such as physiologically-based kinetic/dynamic (PBK/D) models and virtual cell based assay (VCBA) models to calculate internal concentrations and perform extrapolations such as in vitro to in vivo extrapolation (IVIVE). These biologically-based models will be coupled with chemistry-based prediction models that also automate the generation of key input parameters such as physicochemical properties. The implementation and integration of such models within an automated and user-friendly computational platform will make them more easily accessible and applicable to support the chemical risk assessment process.

The development of computational tools and predictive models to support the safety assessment of chemicals, in particular cosmetics-related substances, was the goal of the EU COSMOS project (http://www.cosmostox.eu/), with the aim to making the models publicly available in a user-friendly format. The models included prediction models for specific effects based on structure-activity relationships, absorption models to support extrapolation from oral to dermal exposure, especially relevant for cosmetics substances, as well as the biokinetics models described here.

The VCBA can be used to support (i) the design of in vitro High Throughput Screening (HTS) experiments; (ii) hazard identification (based on acute systemic toxicity); and (iii) QIVIVE approaches to help risk assessment decision making. The VCBA is a mathematical model that was built using the R language and was applied to study the toxicological effects of chemicals on cells (Zaldivar et al., 2010, 2012; Zaldivar Comenges et al., 2011, in press). The model consists of differential equations whose solution allows the calculation over time of the dissolved concentration of a chemical in a well plate as well as the internal concentration in the cells. The VCBA model (Zaldivar Comenges et al., in press) consists of four inter connected models, describing (i) a fate and transport of the time-dependent chemical concentration in the medium and in the headspace; (ii) partitioning in the cell, assuming an instant partitioning of the chemical to water, lipid, and protein within the cell after uptake; (iii) the four stage of cell cycle, (cell growth and division); and (iv) cell dynamics: toxicity and effect. In addition the model takes into account the experimental conditions (i.e. well shape).

Similarly, PBK models mimic the distribution of a chemical in the body. These models represent the body as interconnected compartments, describing the organs, and with inter-compartment fluxes described by differential equations. The level of complexity of these

models depends on the intended application and available biological information: by predicting concentration and time profile curves ($C_{max}$ and AUC), these models can support the submission of drugs to be evaluated by medical agencies; they can be used to support chemical risk assessments; as well as informing experimental design. In the last decade there has been an increase in development of PBK models, because of more accurate simulations of in vivo adsorption, distribution, metabolism and excretion processes compared to the classical PK models (Gajewska et al., 2015). For example, 11 chemical specific PBK models were developed within the EU COSMOS project (Bois et al., in press; Teng et al., 2015; Gajewska et al., 2014, 2015). Furthermore, an initiative to develop a database of existing PBK models is currently ongoing at US EPA (Lu et al., 2016). This database provides useful information on how the models were built. Furthermore, several commercial tools are available to program (Matlab, R language, Berkeley Madonna, etc.), build and use (SimCyp, gastroplus/ADMET, PkSim, ACSL/X, etc.) as well as tools that can be used without a commercial licence: MERLIN EXPO (Ciffroy et al., 2016; Suciu et al., 2016), MEgen and Rvis, (Loizou and Hogg, 2011) COSMOS (Bois et al., in press).

The coupling of a PBK model, describing the kinetics, and the VCBA, describing the dynamic effects (that can be adverse or beneficial) will give rise to a model that can be applied to extrapolate from an in vitro concentration to the external dose of exposure. In vitro to in vivo extrapolation (IVIVE) is a useful approach in the prioritization of chemical testing in chemical risk assessment (Kramer et al., 2015; Allen et al., in preparation). To better exploit the utility of IVIVE, there is a need to develop automated tools that perform the extrapolation in a simple and fast way.

As described above, these models consist of many parameters that are feeding the equations, physicochemical, cell line and experimental characteristics. The use of such models entails a need to manage a large and diverse set of data, from parameters to in vitro dose response curves analysis. The best way is to implement good data management practice and develop automated workflows. The KNIME Analytics Platform (Berthold et al., 2009; http://www.knime.org/) is a free, user-friendly graphical workbench for data analytics including data management, data transformation, investigation, visualization and reporting. The COSMOS biokinetic models have been implemented in two KNIME versions, so that they can be executed either locally on a desktop computer (following installation of the KNIME Analytics Platform) or directly online from a web browser using a KNIME WebPortal (http://knimewebportal.cosmostox.eu). With a view to promoting end-user acceptance of these KNIME-based workflows, bespoke documentation is available (see http://www.cosmostox.eu/what/webtutorials/). Fig. 1 describes schematically how the workflows for the VCBA, PBK models, and IVIVE approach were set up, with an input zone indicating which input parameters are needed to run the simulation per type of workflow, the core model and output zone, with description of which output the model provides per workflow. The optimal interface should allow untrained users to use these workflows in an intuitive way. In this study, we illustrate how the KNIME workflows for the PBK model, the VCBA, and IVIVE approach were built and how they can be used as a user-friendly graphical interface to predict different model simulations and outcomes. The development of automated software tools is an important step in harmonising and expediting the chemical safety assessment process.

## 2. Methodology

We implemented the PBK models, VCBA, and the IVIVE approaches as open source platforms using KNIME (version 3.0) and R programs (which both are freely available). The KNIME Analytics Platform is a user-friendly graphical workbench for data analysis (http://www.knime.org/) and R is a language and environment for statistical computing and graphics (http://www.r-project.org/). KNIME consists of a series of pieces of program code called nodes that can be connected in such

way that the input of one node is the output of the previous one. Each node has a dialog in which the user can configure the operation of the node.

### 2.1. Implementation of the virtual cell-based assay model in KNIME

This VCBA KNIME implementation includes three important features:

- Several operational modes: single exposure, repeated exposure, parameter optimization and optimization check. These are further described below.
- Three separate zones: input, model, and output.
- Can be run on the KNIME Server (online) or in the KNIME Analytics Platform as locally installed version in a desktop computer.

In the following, we will introduce and explain the most significant nodes that build up the KNIME workflow as represented in Fig. 2.

The automated VCBA KNIME workflow can be run in several modes. The *Single Exposure* mode of operation describes the process for a single exposure of the chemical to the cells. The present version can be run for several chemical concentrations. From one run to the next all parameters are reset to initial values except the initial chemical concentration. The other mode is *Repeated Exposure*, which describes the effect of several cell exposures of a single chemical concentration. From one run to the next, the last values of chemical concentration inside the cell and the number of cells are considered as the initial values for the next exposure, whereas the initial chemical concentration remains the same. The toxicological effect of cell exposure to chemicals in the VCBA is described by two parameters: the No Effect Concentration (NEC) and kt (where t is exposure time, and k is a characteristic of the chemical and cell line). The values of NEC and kt are then optimised by comparing the simulated ($Viability_{sim}$) and experimental ($Viability_{exp}$) cell viabilities and by minimizing the following error function:

$$error = \sum_{i=1}^{n^o\ exp} \left( Viability_{exp} - Viability_{sim} \right)^2 \tag{1}$$

The *optimization check* gives us a graphical interface which is used to validate the values of NEC and kt. Paired values of chemical concentration and viability are introduced in a *List Box Input* node, while the initial values of NEC and kt are in two *Double Input* nodes. The Optimization Check functionality was separated from the Parameter Optimization to enable available values of NEC and kt to be entered directly without the need to run the time-consuming Parameter Optimization Mode every time.

The building blocks of the three zones are explained in more detail in the following section.

1) The **Input Zone** is used to feed the model with the necessary input data. It consists of a series of nodes where files are uploaded to input and select data for chemicals and cell lines:

- Data related to the cell line such as duration, mortality, mass, volume in each cell stage, cell duration (see Zaldivar Comenges et al., in press).
- Data related to the organic compound. By uploading a file containing the information on the parameters for the organic compound to be run. The chemical data needed to run simulation are chemical name (Chem name), CAS number (cas), logKow, molar volume, atomic diffusion, MW, water degradation, air degradation, and Henry's law constant (see Zaldivar Comenges et al., in press).
- Toxicological data: the NEC and kt for each chemical and cell line (see Zaldivar Comenges et al., in press).
- Experimental set-up data: The Tissue Culture (TC) plates node is a table that contains information on the technical specifications of the plates in the High Throughput Screening (HTS) as well as non-HTS experiments. These data are used to compute properties related to fate