Toxicology in Vitro xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

Selection of test methods to be included in a testing strategy to predict acute oral toxicity: An approach based on statistical analysis of data collected in phase 1 of the ACuteTox project

A. Kinsner-Ovaskainen [a,*], P. Prieto [a], S. Stanzel [b], A. Kopp-Schneider [b]

[a] Institute for Health and Consumer Protection (IHCP), Joint Research Centre, European Commission, Via Fermi 2749, 21027 Ispra, VA, Italy
[b] Department of Biostatistics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, D-69009 Heidelberg, Germany

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Acute oral toxicity
LD_{50}
Testing strategy
In vitro
Concentration–response analysis
Classification and regression trees

ABSTRACT

More than 50 different *in vitro* and *in silico* methods assessing specific organ- and system-toxicity, such as haemato-, neuro-, nephro- and hepatotoxicity, as well as intestinal absorption, distribution and metabolism, have been used in the first phase of the ACuteTox project to test a common set of 57 chemicals. This paper describes the methods used for statistical evaluation of concentration–response data collected for each of the endpoint assays, and for the development of a testing strategy applicable for acute toxicity classification of chemicals based on the achieved results of the concentration–response analysis. A final list of *in vitro* test methods considered to be promising candidates for building blocks of the testing strategy is presented. Only these selected test methods were further investigated in the prevalidation phase of the project. The test methods were chosen according to their reproducibility and reliability and most importantly, according to their potential to classify chemicals into the official EU CLP acute oral toxicity categories. The potential of the test methods to correctly classify the chemicals was assessed by Classification and Regression Trees (CART) analysis.

1. Introduction

In the last two decades the scientific community has looked for alternatives to replace *in vivo* testing for acute oral toxicity (Clemedson and Ekwall, 1999; Halle, 2003; Anon 2006). Despite all the research efforts worldwide, to date cytotoxicity assays are recognised only as additional tests that can be used to estimate the initial dose for acute oral systemic toxicity tests *in vivo* (Anon, 2006; OECD, 2010).

Currently, acute oral toxicity is assessed in rats in accordance with the OECD Test Guidelines 420 (Fixed Dose Procedure, FDP; OECD, 2001a), 423 (Acute Toxic Class Method, ATC; OECD, 2001b), 425 (Up and Down Procedure, UDP; OECD, 2001c). One of the main drivers for conducting these acute oral toxicity studies is classification and labelling. Substances are categorised according to their potential hazards and the dose required to cause toxicity (Creton et al., 2010; Seidle et al., 2010). The NICEATM/ECVAM validation study (Anon, 2006) assessed, for the first time and among other objectives, the capability of *in vitro* neutral red uptake (NRU) cytotoxicity tests to predict the official acute oral toxicity categories according to the global harmonised system (GHS; UN, 2011). With regard to this particular objective, the study showed that the overall accuracy of the 3T3/NRU cytotoxicity assay to correctly predict the five GHS acute oral toxicity categories and the unclassified category was rather poor, around 30% (Anon, 2006). This joint validation study started as a follow up of an international workshop held in 2000, where the implementation of *in vitro* basal cytotoxicity assays in regulatory screening testing strategies was reviewed (Anon, 2001). One of the recommendations made at the workshop was to further develop, optimise, and validate *in vitro* test methods with focus on target organ specificity and on mechanistic factors such as absorption, distribution, metabolism, and excretion, which act to modulate lethality of xenobiotic response. These aspects were further discussed in 2003 at an ECVAM workshop on acute toxicity (Gennari et al., 2004) during which the strategies to replace *in vivo* acute systemic toxicity testing were addressed in more detail. The recommendations of the ECVAM workshop served as basis for the ACuteTox Project funded by the EU 6th Framework Programme for Research (FP6) in 2005. The ultimate goal of the ACuteTox Project was to design, to optimise and to further prevalidate a non-animal testing strategy for classification of chemicals into the official EU CLP acute oral toxicity categories using solely *in vitro* and *in silico* methods (Anon, 2008).

* Corresponding author. Current address: Nanobiosciences Unit, Institute for Health and Consumer Protection (IHCP), European Commission Joint Research Centre, Via Fermi 2749, 21027 Ispra, VA, Italy. Tel.: +39 0332 789246; fax: +39 0332 785787.
E-mail address: agnieszka.kinsner@jrc.ec.europa.eu (A. Kinsner-Ovaskainen).

0887-2333/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.tiv.2012.11.010

The first phase of the project aimed at identification of suitable *in vitro* and *in silico* methods to be used as building blocks for the testing strategy. This phase included the compilation and evaluation of high quality *in vivo* oral rat acute toxicity data for comparative analyses, and the *de novo* generation of an *in vitro/in silico* database including a large number of endpoint assays assessing biokinetics, metabolism and target organ toxicity (liver, central nervous system, kidney). Moreover, innovative tools (e.g. cytomics) and cellular systems for anticipating animal and human toxicity were explored.

The selection of promising *in vitro* and *in silico* methods from the total number of 53 test methods examined (AXLR8, 2010) was performed on the basis of an in depth statistical analysis of a large dataset generated during the first phase of the project for a training set of 57 common chemicals that were tested with all test methods under investigation. To ensure that the analysis was carried out in an objective and consistent way, it was performed independently from the testing laboratories, by the Department of Biostatistics at the German Cancer Research Center (DKFZ).

The first part of the statistical analysis consisted of concentration–response analysis as well as of the assessment of within-assay variability and of bivariate association of results obtained for different endpoint assays. Moreover, results obtained from concentration–response analysis were transformed to rat LD_{50} values using a formula based on estimation of oral intestinal absorption. In a second approach, a data transformation algorithm considering blood brain barrier (BBB) passage was applied to the concentrations of the compounds that were tested with the neurotoxicity endpoint assays. Subsequently, classification analysis was carried out to select test methods considered to be promising candidates for building blocks of the proposed testing strategy, by quantification of their potential to correctly classify chemicals into the official EU CLP acute oral toxicity categories. Univariate and multivariate classification analysis was conducted by application of Classification and Regression Trees (CART) to the summary values obtained from concentration–response analysis, as well as to the values obtained from the two kinetics transformations described above.

In this paper, we present and discuss in detail the results of the statistical analysis performed for the data collected in the first phase of the ACuteTox Project, the training phase. At the end of this statistical analysis 11 test methods and 2 neural network models were selected as candidate building blocks for the non-animal testing strategy. Univariate and multivariate CART analysis results, data quality, within-assay variability, along with biological considerations and cost arguments, were taken into account for the test method selection. Only these methods have been further evaluated in the second phase of the ACuteTox Project, the prevalidation study (Prieto et al., this issue).

2. Materials and methods

2.1. In vitro and in silico methods

During the first phase of the project, and based on a preliminary assessment done at the level of each Work Package, 23 test methods (a total of 60 endpoint assays) were identified among all *in vitro* and *in silico* methods evaluated (AXLR8, 2010). These endpoint assays were used to test 57 test chemicals. The analysis of the generated data is presented here.

2.2. Test chemicals

An original list of 97 reference chemicals was created during the first phase of the ACuteTox Project as part of the activities carried out in Work Package 1 (Hoffmann et al., 2010; Clothier et al., 2008). From this list, 57 test chemicals (Table 1) were identified and tested in the 60 endpoint assays. Results from these experiments were included in the statistical data analysis presented in this paper. The final list of 57 chemicals included the 16 chemicals selected at the kick off meeting of the project, 24 chemicals nominated by the Work Packages according to their research needs, e.g. target organ toxicities and biokinetic modelling, and additional 17 chemicals identified by the Management Board from the original list.

2.3. Data extraction

Each test chemical was tested with every endpoint assay in one or several concentration–response experiments. Raw data generated in those experiments were stored in Microsoft Excel files (a total of about 10,000 files) and uploaded in the on-line database Acutoxbase (Kinsner-Ovaskainen et al., 2009).

Programme-based automated statistical analysis was carried out to evaluate the concentration–response data. Analysing data in this automated way requires that the data are available in one standardized data format across all the concentration–response experiments (Stanzel et al., this issue). In contrast, different data formats were used for data storage by the partners performing the concentration–response experiments. To replace these heterogeneous data formats by one standardized data format containing all the relevant information (e.g. endpoint assay name, chemical name, experimental ID, lab ID, concentration–response data), an automated data extraction routine was designed and applied to all the raw data files (Stanzel et al., this issue). Data quality was checked by visual inspection of response variability, especially by assessment of control response variability.

2.4. Statistical data analysis

2.4.1. Concentration–response experiments

Concentration–response analysis was performed separately for every concentration–response experiment. In each run of a single experiment, a response value (potentially normalised to mean control response) was measured in dependence of the tested concentration level. The aim of the statistical evaluation of the concentration–response data was the computation of a characteristic value for every experiment. Mostly, the characteristic value of interest was the EC_{50} value. In some instances, the EC_{20} value was to be reported instead or in addition. For some endpoint assays, instead of computation of the EC_{50}, the Lowest Observed Effect Concentration (LOEC) was desired. Throughout the paper EC_x is used to denote EC_{50} or EC_{20}. In some cases, concentration–response relationships were always decreasing, then EC_x is denoted as IC_x. If both directions of concentration–response relationships are possible, the more general notation $IC_x|EC_x$ is used, which indicates that an IC_x is computed in case of a decreasing concentration–response relationship and an EC_x in case of an increasing concentration–response relationship.

2.4.2. $IC_x|EC_x$ estimation

A two-step data analysis approach was applied for $IC_x|EC_x$ estimation. If three or more observations were available per concentration, one-way ANOVA and post hoc Dunnett contrast testing (Dunnett, 1955) of the contrast 'control vs. maximum concentration tested' were carried out in step one of the approach to assess whether concentration has a consistent effect on response. No consistent effect was concluded if (a) ANOVA failed to demonstrate a global effect of concentration on response ($p > 0.05$; Fig. 1A) or (b) a global but inconsistent effect was revealed by ANOVA ($p \leqslant 0.05$) followed by post hoc Dunnett contrast test ($p > 0.05$; Fig. 1B). Note that no Dunnett contrast testing was conducted for

Please cite this article in press as: Kinsner-Ovaskainen, A., et al. Selection of test methods to be included in a testing strategy to predict acute oral toxicity: An approach based on statistical analysis of data collected in phase 1 of the ACuteTox project. Toxicol. in Vitro (2012), http://dx.doi.org/10.1016/j.tiv.2012.11.010