Interspecies Correlations for Predicting the Acute Toxicity of Xenobiotics

James Devillers, Pascal Pandard, Eric Thybaud, and Anne Merle

Abstract LD50 tests on rat and mouse are commonly used to express the relative hazard associated with the acute toxicity of new and existing substances. These tests are expensive, time consuming, and actively fought by Animal Rightists. Consequently, there is a need to find alternative methods. If the design of QSAR models can be used as surrogate, the search for interspecies correlations also represents a valuable alternative to the classical mammalian laboratory tests. In this chapter, the different toxicity $= f$ (ecotoxicity) models available in the literature were first critically analyzed. In a second step, a strong bibliographical investigation was performed to collect oral, intraperitoneal, and intravenous rat and mouse LD50 data for a large collection of structurally diverse chemicals. In the meantime, EC50 data on *Vibrio fischeri* (Microtox^{TM} test) and *Daphnia magna* were also retrieved from literature. Numerous oral, intraperitoneal, and intravenous rat and mouse toxicity models were derived using *Vibrio fischeri* and *Daphnia magna* as independent variables alone or together through a stepwise regression analysis. Most of the models on *Daphnia magna* were totally new and some of them presented acceptable quality. They outperformed the Microtox^{TM} models. The usefulness of the 1-octanol/water partition coefficient (log P) as additional independent variable was also tested. The interest of nonlinear statistical tools for deriving toxicity $= f$ (ecotoxicity) models was also experienced.

Keywords Interspecies correlation · Mammalian toxicity · *Vibrio fischeri* · *Daphnia magna* · Regression analysis · Nonlinear methods

1 Introduction

The debate surrounding the use of animals for research and testing can be traced back to the eighteenth century when the English utilitarian philosopher and jurist, Jeremy Bentham (1748–1832), who is at the origin of the word "deontology,"

J. Devillers (✉)
CTIS, 3 Chemin de la Gravière, 69140 Rillieux La Pape, France
e-mail: j.devillers@ctis.fr

focused attention on animal rights and on their capacity to suffer. In 1831, the well-known English neurophysiologist Marshall Hall (1790–1857) clearly formulated five basic principles to better control the use of animals in experiments and to take into account their suffering [1]. These principles are at the origin of the UK regulation concerning animal experiments. In fact, surprisingly, before 1986, legislation on the protection of animals used in research and testing existed in only a limited number of European countries. Indeed, in 1986, for the first time, the European Parliament adopted legislation aiming at the protection of laboratory animals. Thus, after many years of discussions, the Council of Europe approved regulations on the protection of vertebrate animals used for experimental and other scientific purposes (Convention ETS 123) [2]. In 1986, Directive 86/609/EEC [3], based on the Convention ETS 123 but more concise and restrictive, was also adopted. This Directive contains provisions on the housing and care of laboratory animals, the education and training of persons manipulating animals, the use of non-wild animals, and more generally the promotion of alternative methods to reduce the number of animals, especially the vertebrates used in the laboratories [4]. While the EU Member States are bound to implement the provisions of Directive 86/609/EEC via their national legislation, Convention ETS 123 takes effect only when ratified by a Member State [4,5]. It is noteworthy that due to the development of sciences, especially in the field of biomedicine and also the evolution of the mentalities, in 2002, the European Commission was called by the European Parliament to prepare a proposal for a revision of Directive 86/609/EEC [4]. This work is currently in its final stage. Broadly speaking the revised Directive will reinforce the well known three R's (reduction, replacement, refinement) pioneered by Russell and Burch [6].

Despite regulatory guidelines aiming at reducing the use of animals in experiments, the total number of animals used for experimental and other scientific purposes in 2005 in the 25 Member States was about 12 million. Rodents together with rabbits represented almost 78% of the total and mice were by far the most commonly used species covering 53% of the total use, followed by rats with 19% [7]. More than 60% of animals were used in research and development for human and veterinary medicine, dentistry, and in fundamental biology studies. Production and quality control of products and devices in human medicine, veterinary medicine, and dentistry required the use of 15.3% of the total number of animals reported in 2005. Toxicological and other safety evaluations represented 8% of the total number of animals used for experimental purposes. This represents a significant decrease with 2002 where this percentage was equal to 9.9 [7]. Undoubtedly, this tendency will continue in the future with Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) [8], the new European Community Regulation on chemicals and their safe use entered into force on June 1, 2007. The Article 13, entitled "General requirements for generation of information on intrinsic properties of substances" stresses that regarding "human toxicity, information shall be generated whenever possible by means other than vertebrate animal tests, through the use of alternative methods, for example, in vitro methods or qualitative or quantitative structure–activity relationship models or from information from structurally related substances (grouping or read-across)".

This evolution of the regulation of chemicals prompted us to evaluate the interest of the interspecies correlations for estimating the acute toxicity of chemicals to rats and mice, which are widely used for testing the acute toxicity of new and existing substances. In a first step, literature was investigated to retrieve equations in the general form $LD50_{tox} = f(LC50_{ecotox}$ or $EC50_{ecotox})$ where LC50 refers to the concentration inducing 50% of mortality among the tested population and the EC50 stands for the effective concentration required to induce a 50% effect in the tested organisms, in both cases, in comparison with a control. These different models were critically analyzed. In a second step, attempts were made to derive new equations allowing the prediction of the acute toxicity of chemicals to rat and mouse from LC50 and EC50 data obtained on invertebrates.

2 Bibliographical Survey

2.1 Methodological Framework

Bibliographical investigations were made in journals, books, and reports as well as in bibliographical and factual databases. In that case, a Boolean search was made from the following keywords connected by the logical operators AND, OR, and NOT:

– Correlation, relationship, comparison, intercomparison
– Model, predictive, prediction, in vivo
– Species, interspecies, toxicity, ecotoxicity
– Invertebrate, Daphnid, alga, earthworm, nematode, Microtox[TM], *Vibrio fischeri*, *Daphnia*, *Tetrahymena pyriformis*, *Eisenia fetida*
– Vertebrate, rat, mammal, mouse, mice, human

Only the original publications including regression equations with their statistical parameters were selected. In addition, only the in vivo/in vivo correlations were considered, the in vitro/in vivo and in vitro/in vitro correlations being voluntarily eliminated from the present study. Here, the term in vitro only refers to animal and human cell lines.

2.2 Correlations of Ecotoxicity Test Data with Rat or Mouse LD50 Data

2.2.1 Correlations of Bacteria Test Data with Rat or Mouse LD50 Data

The standard Microtox[TM] test involving the bioluminescent bacterium *Vibrio fischeri*, formerly known as *Photobacterium phosphoreum*, is a commonly used ecotoxicological bioassay whose EC50 values (inhibition of bioluminescence),

generally recorded after 5, 15 or 30 min, have been correlated to EC50 and LC50 values of numerous nonmammalian species [9]. Conversely, the number of papers aiming at correlating Microtox^{TM} data with rat or mouse acute toxicity data is very limited.

In 1992, Fort [10] proposed two regression equations allowing the prediction of *Vibrio fischeri* (*V.f.*) EC50 values from oral and intravenous (i.v.) mouse LD50 values, (1) and (2). The EC50 and LD 50 values were expressed in mg/l and mg/kg, respectively.

$$\log (\text{EC50 } V.f.) = 0.55 \log (\text{LD50 Mouse oral}) - 0.13, \tag{1}$$
$$n = 123, r = 0.29, p = 0.0012.$$
$$\log (\text{EC50 } V.f.) = 1.6 \log (\text{LD50 Mouse i.v.}) - 1.8, \tag{2}$$
$$n = 51, r = 0.73, p < 0.0001.$$

Although a weak correlation was obtained with the oral LD50 data, a more interesting relationship was recorded with the intravenous LD50 data but the size of the training set was about twice less important.

Kaiser and coworkers [11] tried to extend these results from larger datasets and by considering the oral, i.v., and intraperitoneal (i.p.) routes of exposure for rat and mouse. The EC50 and LD50 values were expressed in mmol/l and mmol/kg, respectively. This yielded the design of six equations (3)–(8).

$$\log (1/\text{LD50 Rat oral}) = 0.20 \log (1/\text{EC50 } V.f.) - 0.96, \tag{3}$$
$$n = 471, r = 0.35, \text{se} = 0.74.$$
$$\log (1/\text{LD50 Mouse oral}) = 0.20 \log (1/\text{EC50 } V.f.) - 0.86, \tag{4}$$
$$n = 344, r = 0.35, \text{se} = 0.72.$$
$$\log (1/\text{LD50 Rat i.p.}) = 0.29 \log (1/\text{EC50 } V.f.) - 0.48, \tag{5}$$
$$n = 195, r = 0.48, \text{se} = 0.82.$$
$$\log (1/\text{LD50 Mouse i.p.}) = 0.25 \log (1/\text{EC50 } V.f.) - 0.49, \tag{6}$$
$$n = 378, r = 0.43, \text{se} = 0.70.$$
$$\log (1/\text{LD50 Rat i.v.}) = 0.40 \log (1/\text{EC50 } V.f.) - 0.25, \tag{7}$$
$$n = 54, r = 0.73, \text{se} = 0.79.$$
$$\log (1/\text{LD50 Mouse i.v.}) = 0.35 \log (1/\text{EC50 } V.f.) - 0.30, \tag{8}$$
$$n = 165, r = 0.68, \text{se} = 0.61.$$

Before outlier removal, the oral route entry regressions for rat (3) and mouse (4) present the same slope and correlation coefficient. Moreover the intercepts and standard error of estimates (se) are very close. Inspection of (5)–(8) shows that for the i.p. and i.v. routes of exposure, a rather good similarity also exists between the regressions for rat and mouse. This similarity between the two mammalian species for the same exposure route prompted the authors to extend the datasets by means of (9)–(11).

$$\log (1/\text{LD50 Rat oral}) = 0.97 \log (1/\text{LD50 Mouse oral}) - 0.04, \tag{9}$$
$$n = 330, \ r = 0.94, \ r^2 = 0.88, \ \text{se} = 0.30.$$
$$\log (1/\text{LD50 Rat i.p.}) = 1.02 \log (1/\text{LD50 Mouse i.p.}) - 0.02, \tag{10}$$
$$n = 162, \ r = 0.96, \ r^2 = 0.92, \ \text{se} = 0.28.$$
$$\log (1/\text{LD50 Rat i.v.}) = 0.99 \log (1/\text{LD50 Mouse i.v.}) - 0.10, \tag{11}$$
$$n = 41, \ r = 0.97, \ r^2 = 0.94, \ \text{se} = 0.29.$$

Using the extended rat datasets for each of the oral, i.p., and i.v. exposure routes, linear regressions were then determined vs. the corresponding Microtox$^{\text{TM}}$ data yielding (12)–(14). Deletion of outliers allowed the increase of the statistical parameters of the models, (15)–(17).

$$\log (1/\text{LD50 Rat oral}) = 0.19 \log (1/\text{EC50 } V. f.) - 0.95, \tag{12}$$
$$n = 531, \ r = 0.33, \ \text{se} = 0.72, \ F = 63.4.$$
$$\log (1/\text{LD50 Rat i.p.}) = 0.25 \log (1/\text{EC50 } V. f.) - 0.50, \tag{13}$$
$$n = 427, \ r = 0.43, \ \text{se} = 0.70, \ F = 95.7.$$
$$\log (1/\text{LD50 Rat i.v.}) = 0.35 \log (1/\text{EC50 } V. f.) - 0.20, \tag{14}$$
$$n = 180, \ r = 0.66, \ \text{se} = 0.65, \ F = 139.7.$$
$$\log (1/\text{LD50 Rat oral}) = 0.20 \log (1/\text{EC50 } V. f.) - 1.03, \tag{15}$$
$$n = 506, \ r = 0.41, \ \text{se} = 0.59, \ F = 102.2.$$
$$\log (1/\text{LD50 Rat i.p.}) = 0.26 \log (1/\text{EC50 } V. f.) - 0.57, \tag{16}$$
$$n = 406, \ r = 0.51, \ \text{se} = 0.59, \ F = 141.5.$$
$$\log (1/\text{LD50 Rat i.v.}) = 0.36 \log (1/\text{EC50 } V. f.) - 0.26, \tag{17}$$
$$n = 171, \ r = 0.75, \ \text{se} = 0.52, \ F = 219.3.$$

Inspection of (12)–(17) shows that the data ranges and slopes of the regressions are unaffected by the outlier removal but that the correlation coefficients and more important the F tests and standard errors of the estimates are much improved.

2.2.2   Correlations of Protozoan Test Data with Rat or Mouse LD50 Data

Collections of chemicals have been tested on the freshwater ciliate protozoan *Tetrahymena pyriformis* (*T.p.*) but surprisingly, their use in the design of toxicity = *f* (ecotoxicity) models is very limited. Thus, Sauvant et al. [12] evaluated the effects of $BaCl_2$ salt, $CdCl_2$, $CoCl_2$, $CrCl_3$, $CuCl_2$, $FeCl_3$, GeO_2, $HgCl_2$, $MnCl_2$, $NbCl_5$, $Pb(NO_3)_2$, $SbCl_3$, $SnCl_4$, $TiCl_4$, $VOSO_4$, and $ZnCl_2$ on the growth rate of *T. pyriformis*. IC50s (inhibitory concentration 50%) were expressed in mmol/l and correlated with corresponding rat oral LD50 values (mmol/kg) retrieved from literature, yielding (18).

$$\log(\text{LD50 Rat oral}) = 0.571\log(\text{IC50 }T.p.) + 0.389, \qquad (18)$$
$$n = 16,\ r = 0.463,\ p = 0.07.$$

2.2.3 Correlations of Rotifer Test Data with Rat or Mouse LD50 Data

The 24-h LC50 values of the first ten chemicals of the multicentre evaluation of in vitro cytotoxicity (MEIC) program were tested against the estuarine rotifer *Brachionus plicatilis* (*B.p.*) and the freshwater rotifer *Brachionus calyciflorus* (*B.c.*) [13]. These chemicals were the following: paracetamol (CAS RN 103-90-2), acetylsalicylic acid (CAS RN 50-78-2), ferrous sulfate heptahydrate (CAS RN 7782-63-0), amitriptyline HCl (CAS RN 549-18-8), isopropanol (CAS RN 67-63-0), ethanol (CAS RN 64-17-5), methanol (CAS RN 67-56-1), ethylene glycol (CAS RN 107-21-1), diazepam (CAS RN 439-14-5), and digoxin (CAS RN 20830-75-5). The acute toxicity data, expressed in μmol/l, were compared by regression analysis with oral LD50 (μmol/kg) in rat, mouse, and man (HLD = human oral lethal dose). Diazepam and digoxin were excluded from the regressions (19)–(24) because their 24-h LC50 values on the two rotifers were only determined as $>35{,}100$ μmol/l and $>12{,}800$ μmol/l, respectively. Even if the r^2 values of (19)–(24) are high, the interest of these models is very limited due the nature and limited number of chemicals.

$$\log(\text{LD50 Mouse oral}) = 0.48\log(\text{LC50 }B.p.) + 2.08, \qquad (19)$$
$$n = 8,\ r^2 = 0.86.$$
$$\log(\text{LD50 Rat oral}) = 0.48\log(\text{LC50 }B.p.) + 2.12, \qquad (20)$$
$$n = 8,\ r^2 = 0.92.$$
$$\log(\text{LD50 HLD oral}) = 0.44\log(\text{LC50 }B.p.) + 1.94, \qquad (21)$$
$$n = 8,\ r^2 = 0.83.$$
$$\log(\text{LD50 Mouse oral}) = 0.43\log(\text{LC50 }B.c.) + 2.40, \qquad (22)$$
$$n = 8,\ r^2 = 0.81.$$
$$\log(\text{LD50 Rat oral}) = 0.42\log(\text{LC50 }B.c.) + 2.44, \qquad (23)$$
$$n = 8,\ r^2 = 0.88.$$
$$\log(\text{LD50 HLD oral}) = 0.40\log(\text{LC50 }B.c.) + 2.22, \qquad (24)$$
$$n = 8,\ r^2 = 0.81.$$

2.2.4 Correlations of Crustacean Test Data with Rat or Mouse LD50 Data

The water flea *Daphnia magna* (*D.m.*) is one of the most widely used invertebrates in freshwater aquatic toxicology. The criterion of acute toxicity determined with this organism is the effective concentration yielding the complete immobilization of 50% of the population of *Daphnia* after 24 or 48 h of exposure (24-h or 48-h EC50). To be considered as immobilized, the animals have to be unable to swim after a gentle agitation of the test vessel. Different authors have tried to correlate

LD50 values recorded in rat and/or mouse to EC50 values obtained on *D.m.* Thus, Khangarot and Ray [14] tested various organic and inorganic chemicals on young *D.m.* and the obtained 48-h EC50 values (mg/l) were correlated to oral rat and mouse LD50 values (mg/kg) retrieved from literature, (25) and (26). Even if the correlation coefficients of (25) and (26) are very high, the interest of these two models is reduced due to limited size of their training set.

$$\text{LD50 (Rat oral)} = 2.056 \, \text{EC50} \, (D.m.) + 776.2, \tag{25}$$
$$n = 13, \ r = 0.992.$$
$$\text{LD50 (Mouse oral)} = 1.020 \, \text{EC50} \, (D.m.) + 312.94, \tag{26}$$
$$n = 10, \ r = 0.991.$$

Interestingly, Enslein and coworkers [15] tried to increase the performances of a simple LD50 vs. EC50 model (27) by introducing molecular descriptors in the equation from a stepwise regression analysis. This yielded a new model (28) showing better statistics. In both equations, LC50 and LD50 values were expressed in mmol/l and mmol/kg, respectively.

$$\log (\text{LD50 Rat}) = f(\log \text{LC50} \ D.m.), \tag{27}$$
$$n = 147, \ r^2 = 0.53, \ s = 0.53.$$
$$\log (\text{LD50 Rat}) = 0.287 \, D.m. - 0.520 \, \text{aryl nitro} + 0.362 \, \text{DIFPAT5}$$
$$+ \, 0.328 \, \text{Nb electron releasing groups on a benzene ring} - 0.496 \, \text{Ring perimeter}$$
$$- \, 0.608 \, \text{NH}_2, \ \text{NH or 3-branched aliphatic amine}$$
$$- \, 0.408 \, \text{aryl alcohol} - 0.619 \, \text{methylene diphenyl linkage}$$
$$- \, 0.826 \, \text{aliphatic ether} + 0.337 \, \text{primary aliphatic hydroxyl}$$
$$- \, 0.568 \, \text{any carbamate} - 0.487 \, \text{pentane fragment}$$
$$- \, 0.279 \, \text{propane/propene fragment} + 3.415, \tag{28}$$
$$n = 147, \ r^2 = 0.75, \ s = 0.40.$$

Inverse relationships (e.g., (29) and (30)) were also proposed by these authors [16] but only the model specifically designed for cholinesterase-inhibiting compounds was interesting (30). In (30), MW is the molecular weight and $^2\chi^v$ is the valence path molecular connectivity index of second order.

$$\log (1/\text{EC50} \ D.m.) = f \log (1/\text{LD50 Rat oral}), \tag{29}$$
$$n = 182, \ r^2 = 0.452, \ s = 1.116.$$
$$\log (1/\text{EC50} \ D.m.) = 0.738 \log (1/\text{LD50 Rat oral})$$
$$+ \, 6.399 \, \text{MW} - 0.147 \, ^2\chi^v - 9.29, \tag{30}$$
$$n = 12, \ r^2 = 0.80, \ s = 0.432, \ F = 10.66.$$

Calleja and Persoone [13] also tested the first ten chemicals of the MEIC program against the halophytic anostracan *Artemia salina* (*A.s.*) and the freshwater

anostracan *Streptocephalus proboscideus* (*S.p.*). The LC50s recorded after 24 h of exposure and expressed in μmol/l were compared by regression analysis to oral LD50 (μmol/kg) in rat, mouse, and man (HLD). Equations at eight chemicals (i.e., (31), (32), (35), (37), (38), (41)) were established without digoxin and diazepam, while for the equations at nine chemicals (i.e., (33), (34), (36), (39), (40), (42)) only the former compound was excluded. As previously indicated, the interest of these equations is rather limited.

$$\log (\text{LD50 Mouse oral}) = 0.54 \log (\text{LC50 } A.s.) + 1.86, \tag{31}$$
$$n = 8, \ r^2 = 0.89.$$

$$\log (\text{LD50 Rat oral}) = 0.51 \log (\text{LC50 } A.s.) + 2.01, \tag{32}$$
$$n = 8, \ r^2 = 0.87.$$

$$\log (\text{LD50 Mouse oral}) = 0.56 \log (\text{LC50 } A.s.) + 1.76, \tag{33}$$
$$n = 9, \ r^2 = 0.90.$$

$$\log (\text{LD50 Rat oral}) = 0.49 \log (\text{LC50 } A.s.) + 2.09, \tag{34}$$
$$n = 9, \ r^2 = 0.87.$$

$$\log (\text{LD50 HLD oral}) = 0.45 \log (\text{LC50 } A.s.) + 1.82, \tag{35}$$
$$n = 8, \ r^2 = 0.80.$$

$$\log (\text{HLD oral}) = 0.54 \log (\text{LC50 } A.s.) + 1.44, \tag{36}$$
$$n = 9, \ r^2 = 0.80.$$

$$\log (\text{LD50 Mouse oral}) = 0.49 \log (\text{LC50 } S.p.) + 2.30, \tag{37}$$
$$n = 8, \ r^2 = 0.98.$$

$$\log (\text{LD50 Rat oral}) = 0.41 \log (\text{LC50 } S.p.) + 2.63, \tag{38}$$
$$n = 8, \ r^2 = 0.75.$$

$$\log (\text{LD50 Mouse oral}) = 0.51 \log (\text{LC50 } S.p.) + 2.12, \tag{39}$$
$$n = 9, \ r^2 = 0.94.$$

$$\log (\text{LD50 Rat oral}) = 0.42 \log (\text{LC50 } S.p.) + 2.56, \tag{40}$$
$$n = 9, \ r^2 = 0.77.$$

$$\log (\text{LD50 HLD}) = 0.44 \log (\text{LC50 } S.p.) + 2.15, \tag{41}$$
$$n = 8, \ r^2 = 0.94.$$

$$\log (\text{HLD oral}) = 0.50 \log (\text{LC50 } S.p.) + 1.81, \tag{42}$$
$$n = 9, \ r^2 = 0.82.$$

2.2.5 Correlations of Fish Test Data with Rat or Mouse LD50 Data

The relative vulnerability of most of the fish species to pollutants and the ecological importance of these organisms in the functioning of the ecosystems have contributed to their selection as surrogates to assess the aquatic ecotoxicity of chemicals yielding

the production of collections of acute toxicity data for all the kinds of compounds susceptible to contaminate the environment. Moreover, due to the taxonomical position of these organisms, numerous equations have been proposed for predicting the acute toxicity of chemicals to rat or mouse from fish LC50s.

Relationships between rat oral LD50 (mmol/kg) and *Lepomis macrochirus* (*L.m.*) and *Pimephales promelas* (*P.p.*) 96-h LC50 (μmol/l) values obtained from the US water quality criteria documents for 47 priority pollutants, including nine organochlorine pesticides were examined by Janardan et al. [17]. Interestingly, these authors also tried to derive regression equations on data obtained from uniform protocol studies for fish [18] and male and female rats [19]. It is noteworthy that this second set only included chlorinated, organophosphorus, and carbamate pesticides. The inverse correlations were also considered because the regression analysis used in this study, which considered separate error terms for the *x* and *y* variables, provided different statistics for them. Thus, 20 different models were proposed by Janardan et al. [17] (43)–(62).

$$\log (\text{LD50 Rat}) = 0.43 \log (\text{LC50 } L.m.) - 0.056, \tag{43}$$
$n = 44, \ r = 0.74$ (priority pollutants).

$$\log (\text{LD50 Rat male}) = 0.47 \log (\text{LC50 } L.m.) - 0.272, \tag{44}$$
$n = 48, \ r = 0.73$ (priority pollutants + pesticides).

$$\log (\text{LD50 Rat female}) = 0.49 \log (\text{LC50 } L.m.) - 0.313, \tag{45}$$
$n = 45, \ r = 0.75$ (priority pollutants + pesticides).

$$\log (\text{LD50 Rat male}) = 0.46 \log (\text{LC50 } L.m.) + 0.125, \tag{46}$$
$n = 12, \ r = 0.76$ (chlorinated pesticides).

$$\log (\text{LD50 Rat female}) = 0.66 \log (\text{LC50 } L.m.) + 0.345, \tag{47}$$
$n = 11, \ r = 0.92$ (chlorinated pesticides).

$$\log (\text{LC50 } L.m.) = 1.21 \log (\text{LD50 Rat}) + 0.539, \tag{48}$$
$n = 44, \ r = 0.71$ (priority pollutants).

$$\log (\text{LC50 } L.m.) = 1.04 \log (\text{LD50 Rat male}) + 0.428, \tag{49}$$
$n = 48, \ r = 0.66$ (priority pollutants + pesticides).

$$\log (\text{LC50 } L.m.) = 1.04 \log (\text{LD50 Rat female}) + 0.492, \tag{50}$$
$n = 45, \ r = 0.68$ (priority pollutants + pesticides).

$$\log (\text{LC50 } L.m.) = 1.45 \log (\text{LD50 Rat male}) - 0.639, \tag{51}$$
$n = 12, \ r = 0.88$ (chlorinated pesticides).

$$\log (\text{LC50 } L.m.) = 1.51 \log (\text{LD50 Rat female}) - 0.521, \tag{52}$$
$n = 11, \ r = 0.999$ (chlorinated pesticides).

$$\log (\text{LD50 Rat}) = 0.35 \log (\text{LC50 } P.p.) - 0.161, \tag{53}$$
$n = 38, \ r = 0.63$ (priority pollutants).

$$\log (\text{LD50 Rat male}) = 0.33 \log (\text{LC50 } P.p.) - 0.34, \tag{54}$$
$n = 28, \ r = 0.58$ (priority pollutants + pesticides).

$$\log(\text{LD50 Rat female}) = 0.36 \log(\text{LC50 } P.p.) - 0.259, \tag{55}$$
$$n = 25, \ r = 0.67 \ (\text{priority pollutants} + \text{pesticides}).$$

$$\log(\text{LD50 Rat male}) = 0.59 \log(\text{LC50 } P.p.) + 0.192, \tag{56}$$
$$n = 9, \ r = 0.999 \ (\text{chlorinated pesticides}).$$

$$\log(\text{LD50 Rat female}) = 0.28 \log(\text{LC50 } P.p.) + 0.380, \tag{57}$$
$$n = 8, \ r = 0.999 \ (\text{chlorinated pesticides}).$$

$$\log(\text{LC50 } P.p.) = 1.37 \log(\text{LD50 Rat}) + 0.799, \tag{58}$$
$$n = 38, \ r = 0.77 \ (\text{priority pollutants}).$$

$$\log(\text{LC50 } P.p.) = 1.15 \log(\text{LD50 Rat male}) + 0.820, \tag{59}$$
$$n = 28, \ r = 0.65 \ (\text{priority pollutants} + \text{pesticides}).$$

$$\log(\text{LC50 } P.p.) = 1.53 \log(\text{LD50 Rat female}) + 0.689, \tag{60}$$
$$n = 25, \ r = 0.83 \ (\text{priority pollutants} + \text{pesticides}).$$

$$\log(\text{LC50 } P.p.) = 1.70 \log(\text{LD50 Rat male}) - 0.326, \tag{61}$$
$$n = 9, \ r = 0.98 \ (\text{chlorinated pesticides}).$$

$$\log(\text{LC50 } P.p.) = 1.29 \log(\text{LD50 Rat female}) - 0.490, \tag{62}$$
$$n = 8, \ r = 0.96 \ (\text{chlorinated pesticides}).$$

Significant relationships between species were obtained for the priority pollutants, priority pollutants plus pesticides, and chlorinated pesticides. Conversely, the authors did not find an acceptable correlation when only the organophosphate and carbamate pesticide toxicities were compared between rat and fish. In the same way, no significant relationships were obtained between the rat and fishes over all classes of pesticides (equations missing). From the scatter in plots for the priority pollutants, Janardan et al. [17] deducted that fish were relatively more sensitive (LC50/LD50 < 1) than rats for substances with an LD50 < 1 mmol/kg (rat) and less sensitive (LC50/LD50 > 1) for substances with an LD50 > 1 mmol/kg. From the regression models, they showed that the two fish species presented about the same sensitivity to the priority pollutants but bluegill (*L.m.*) was less sensitive than fathead minnow (*P.p.*) to pesticides.

Kaiser and coworker [20] proposed a rat vs. fathead minnow model (*P.p.*) with a larger domain of application (63). The performance of the model increased with a three parameter equation also including *V.f.* EC50 and 1-octanol/water partition coefficient (log *P*) data (64).

$$\log(1/\text{LD50 Rat}) = 0.36 \log(1/\text{LC50 } P.p.) - 1.16, \tag{63}$$
$$n = 91, \ r^2 = 0.34.$$

$$\log(1/\text{LD50 Rat}) = f(\log 1/\text{EC50 } V.f., \ \log 1/\text{LC50 } P.p., \ \log \text{P}), \tag{64}$$
$$n = 91, \ r^2 = 0.41.$$

Hodson [21] compared the toxicity of industrial chemicals to *Oncorhynchus mykiss* (formerly *Salmo gairdneri*), as shown by i.p. injections (i.p. LD50), oral dosing (oral LD50), and aqueous exposure (LC50), with published values for i.p. LD50s and oral LD50s of mice and rats. Prior correlation analysis, the toxicity data were expressed on a millimolar basis. Twenty equations were obtained (65)–(84). When mouse and rat oral LD50s are compared with fish i.p. LD50s, the correlation coefficients are equal to 0.807 and 0.897, respectively (65) and (66) but when the comparison is made between i.p. LD50s, (67) and (68), r is improved to 0.936 and 0.933, respectively. Despite small sample sizes, there is a strong relationship between fish oral LD50s and rat ($r = 0.827$) and mouse ($r = 0.914$) i.p. LD50s, (69) and (70). Conversely, the rat and mouse oral LD50s are not strongly related to fish oral LD50s, (71) and (72).

$$\log (\text{LD50 Mouse oral}) = 0.8046 \log (\text{LD50 } O.m. \text{ i.p.}) + 0.4267, \tag{65}$$
$$n = 13, \ r = 0.807, \ p < 0.05.$$

$$\log (\text{LD50 Rat oral}) = 0.9429 \log (\text{LD50 } O.m. \text{ i.p.}) + 0.1495, \tag{66}$$
$$n = 25, \ r = 0.8971, \ p < 0.05.$$

$$\log (\text{LD50 Mouse i.p.}) = 0.8288 \log (\text{LD50 } O.m. \text{ i.p.}) - 0.1831, \tag{67}$$
$$n = 12, \ r = 0.936, \ p < 0.05.$$

$$\log (\text{LD50 Rat i.p.}) = 1.0051 \log (\text{LD50 } O.m. \text{ i.p.}) - 0.1693, \tag{68}$$
$$n = 16, \ r = 0.933, \ p < 0.05.$$

$$\log (\text{LD50 Rat i.p.}) = 1.4080 \log (\text{LD50 } O.m. \text{ oral}) - 0.5262, \tag{69}$$
$$n = 6, \ r = 0.827, \ p < 0.05.$$

$$\log (\text{LD50 Mouse i.p.}) = 0.8264 \log (\text{LD50 } O.m. \text{ oral}) - 0.2479, \tag{70}$$
$$n = 7, \ r = 0.914, \ p < 0.05.$$

$$\log (\text{LD50 Mouse oral}) = 0.6831 \log (\text{LD50 } O.m. \text{ oral}) + 0.3048, \tag{71}$$
$$n = 7, \ r = 0.657.$$

$$\log (\text{LD50 Rat oral}) = 0.8255 \log (\text{LD50 } O.m. \text{ oral}) + 0.1606, \tag{72}$$
$$n = 9, \ r = 0.588.$$

There is considerably more variation in comparisons of i.p. LD50s to fish LC50s (73)–(75). The best relationship between i.p. LD50s and LC50s is for rat (74) while the poorest is for mouse, (73).

$$\log (\text{LD50 Mouse i.p.}) = 0.7709 \log (\text{LC50 } O.m.) + 1.1008, \tag{73}$$
$$n = 8, \ r = 0.48.$$

$$\log (\text{LD50 Rat i.p.}) = 0.5891 \log (\text{LC50 } O.m.) + 1.2136, \tag{74}$$
$$n = 11, \ r = 0.83, \ p < 0.05.$$

$$\log (\text{LD50 } O.m. \text{ i.p.}) = 0.8883 \log (\text{LC50 } O.m.) + 1.4608, \tag{75}$$
$$n = 13, \ r = 0.60, \ p < 0.05.$$

$$\log (\text{LD50 Mouse oral}) = 0.7576 \log (\text{LC50 } O.m.) + 1.6616, \qquad (76)$$
$$n = 10, \ r = 0.19.$$
$$\log (\text{LD50 Rat oral}) = 0.7086 \log (\text{LC50 } O.m.) + 1.6553, \qquad (77)$$
$$n = 15, \ r = 0.62, \ p < 0.05.$$
$$\log (\text{LD50 } O.m. \text{ oral}) = 0.797 \log (\text{LC50 } O.m.) + 1.5216, \qquad (78)$$
$$n = 9, \ r = 0.58.$$

An attempt was also made by Hodson [21] to relate fish and mammal oral and i.p. LD50s to fish LC50s amended by the octanol/water partition coefficient (P). The assumption was that P could correct differences in toxicity due to the effect of partitioning of chemicals on uptake and toxicity during aqueous exposure. Six new regression equations were produced (79)–(84).

$$\log (\text{LD50 } O.m. \text{ i.p.}) = 2.464 \log (\text{LC50 } O.m. \times P) - 0.499, \qquad (79)$$
$$n = 11, \ r = 0.32.$$
$$\log (\text{LD50 Rat i.p.}) = -2.128 \log (\text{LC50 } O.m. \times P) + 0.619, \qquad (80)$$
$$n = 9, \ r = 0.01.$$
$$\log (\text{LD50 Mouse i.p.}) = 1.601 \log (\text{LC50 } O.m. \times P) - 0.172, \qquad (81)$$
$$n = 7, \ r = 0.21.$$
$$\log (\text{LD50 } O.m. \text{ oral}) = 1.8661 \log (\text{LC50 } O.m. \times P) + 0.129, \qquad (82)$$
$$n = 7, \ r = 0.45.$$
$$\log (\text{LD50 Rat oral}) = 1.9138 \log (\text{LC50 } O.m. \times P) + 0.089, \qquad (83)$$
$$n = 13, \ r = 0.53.$$
$$\log (\text{LD50 Mouse oral}) = 1.7397 \log (\text{LC50 } O.m. \times P) + 0.193, \qquad (84)$$
$$n = 8, \ r = 0.62.$$

Inspection of (73)–(84) shows that an improvement was only noted for the equations dealing with mouse oral LD50s (i.e., (76) vs. (84)).

Delistraty et al. [22] examined acute toxicity relationships over several exposure routes in rainbow trout ($O.m.$) and rats. An initial database of 217 chemicals (126 pesticides and 91 nonpesticides) was constituted. 1-octanol/water partition coefficient ($\log P$) values for the organic molecules were also retrieved from literature. LC50 and LD50 values were expressed in mmol/l and mmol/kg, respectively. The authors showed that the stratification of the data into pesticides and nonpesticides did not particularly improve predictions of trout LC50s from rat oral LD50s (85)–(87). Addition of $\log P$ in the model (88) increased the r and r^2 values but it is noteworthy that the number of chemicals used to derive the model was lower (i.e., 213 vs. 145).

$$\log (\text{LC50 } O.m.) = 0.722 \log (\text{LD50 Rat oral}) - 2.16, \qquad (85)$$
$$n = 213, \ r = 0.512, \ r^2 = 0.262.$$

$$\log (\text{LC50 } O.m.) = 0.476 \log (\text{LD50 Rat oral}) - 2.42, \tag{86}$$
$$n = 125, \ r = 0.380, \ r^2 = 0.144 \text{ (pesticides)}.$$
$$\log (\text{LC50 } O.m.) = 0.925 \log (\text{LD50 Rat oral}) - 1.98, \tag{87}$$
$$n = 88, \ r = 0.540, \ r^2 = 0.292 \text{ (nonpesticides)}.$$
$$\log (\text{LC50 } O.m.) = 0.644 \log (\text{LD50 Rat oral}) - 0.463 \log P - 0.953, \tag{88}$$
$$n = 145, \ r = 0.729, \ r^2 = 0.531.$$

Trout LC50 values were also predicted from rat LD50 data with regressions matched on exposure route. Statistically significant models were obtained for the three routes of exposure (89)–(91). Addition of log P in the models (92)–(94) did not improve the models, except for the i.p. route, (94).

$$\log (\text{LC50 } O.m. \text{ oral}) = 0.918 \log (\text{LD50 Rat oral}) + 0.153, \tag{89}$$
$$n = 27, \ r = 0.907, \ r^2 = 0.823.$$
$$\log (\text{LC50 } O.m. \text{ dermal}) = 0.794 \log (\text{LD50 Rat dermal}) + 0.384, \tag{90}$$
$$n = 11, \ r = 0.914, \ r^2 = 0.835.$$
$$\log (\text{LC50 } O.m. \text{ i.p.}) = 0.852 \log (\text{LD50 Rat i.p.}) + 0.355, \tag{91}$$
$$n = 13, \ r = 0.761, \ r^2 = 0.579.$$
$$\log (\text{LC50 } O.m. \text{ oral}) = 0.970 \log (\text{LD50 Rat oral}) \tag{92}$$
$$+ 0.050 \log P + 0.035,$$
$$n = 25, \ r = 0.904, \ r^2 = 0.817.$$
$$\log (\text{LC50 } O.m. \text{ dermal}) = 0.890 \log (\text{LD50 Rat dermal}) \tag{93}$$
$$+ 0.079 \log P + 0.094,$$
$$n = 8, \ r = 0.755, \ r^2 = 0.570.$$
$$\log (\text{LC50 } O.m. \text{ i.p.}) = 0.463 \log (\text{LD50 Rat i.p.}) \tag{94}$$
$$- 0.324 \log P + 1.08,$$
$$n = 13, \ r = 0.912, \ r^2 = 0.832.$$

Models for predicting trout LC50s from rat inhalation (inh) LD50s were also designed by Delistraty [23] (95)–(98). Toxicity data were expressed in mmol/l, ppmw (parts per million by weight), or ppmv (parts per million by volume). Addition of molecular descriptors only slightly increased the performances of the best one parameter equation (i.e., (96) vs. (99)).

$$\log (\text{LC50 } O.m. \text{ mmol/l}) = 0.953 \log (\text{LD50 Rat inh mmol/l}) + 0.235, \tag{95}$$
$$n = 60, \ r = 0.678, \ r^2 = 0.459.$$
$$\log (\text{LC50 } O.m. \text{ mmol/l}) = 0.955 \log (\text{LCT50 Rat inh mmol-h/l}) - 0.126, \tag{96}$$
$$n = 46, \ r = 0.745, \ r^2 = 0.556.$$
$$\log (\text{LC50 } O.m. \text{ ppmw}) = 0.899 \log (\text{LD50 Rat inh ppmv}) - 1.46, \tag{97}$$
$$n = 15, \ r = 0.592, \ r^2 = 0.350.$$

$$\log (\text{LC50 } O.m. \text{ ppmw}) = 1.16 \log (\text{LCT50 Rat inh ppmv-h}) - 3.22, \quad (98)$$
$$n = 11, \ r = 0.747, \ r^2 = 0.558.$$
$$\log (\text{LC50 } O.m. \text{ mmol/l}) = 0.725 \log (\text{LCT50 Rat inh mmol-h/l})$$
$$-3.19 \log \text{MW} - 0.266 \log \text{VP} + 0.263 \log S \text{ (mmol/l)} + 5.72, \quad (99)$$
$$n = 38, \ r = 0.873, \ r^2 = 0.763.$$

2.3 Main Characteristics of the Published Models

This literature survey clearly reveals a limited number of toxicity $= f$ (ecotoxicity) models. The available correlations between toxicological and ecotoxicological endpoints only deal with a limited number of species as well as reduced sets of chemicals. The interspecies correlations are established considering rodents (rat and mouse) and aquatic species (mainly fish and bacteria but also crustaceans and rotifers).

Generally, models are designed for predicting mammalian toxicity from aquatic toxicity data but the converse is also found. The 99 correlation equations collected from literature were only derived from acute toxicity data on pure chemicals (LD50s, EC50s, LC50s). No interspecies relationships were investigated using chronic or sublethal effects, probably due to the lack of such data. The toxicity data are always retrieved from literature, while the ecotoxicity data can be obtained from experiments [12–14].

Several exposure routes were considered for mammalian species (i.e., oral, dermal, intraperitoneal, intravenous, inhalation) leading to the development of specific predictive models presenting different qualities. Thus, for example, Kaiser et al. [11] found that the interspecies relationships between *Vibrio fischeri*, rat, and mouse increased significantly from oral, to intraperitoneal, and to intravenous data. Regarding the rats and mice, generally no distinction is made between males and females in the modeling processes.

Most of the models for predicting mammalian toxicity from aquatic toxicity data were designed from simple linear regression analysis. However, it is noteworthy that some authors successfully included molecular descriptors in their equation, especially the 1-octanol/water partition coefficient ($\log P$). Interestingly, Kaiser and Esterby [20] established a predictive model for rat toxicity using the results of tests performed on *Vibrio fischeri, Pimephales promelas*, and from $\log P$.

Despite some significant correlations, it appears that the toxicity $= f$ (ecotoxicity) models found in the literature cannot be used in practice. Indeed, most of them were established from a limited number of chemicals. Thus, more than 70% of the models found in the literature were derived from less than 50 chemicals. Moreover, it is important to note that chemicals are very often eliminated before or during the regression processes without clear justifications.

This prompted us to develop new models focused on the prediction of rat and mouse LD50s from invertebrate EC50s or LC50s. The selected species were *Daphnia magna* and *Vibrio fischeri* because these organisms are widely used for assessing the hazard of chemicals and hence, collections of EC50s for these organisms are available in the literature.

3 Design of New Toxicity = f (Ecotoxicity) Models

3.1 Data Sources, Notations, and Treatments

LD50s for rats and mice were retrieved from CD-ROMs (e.g., Merck Index, ECDIN, IUCLID) and data banks such as MSDS (http://physchem.ox.ac.uk/MSDS/) or Extonet (http://extoxnet.orst.edu/pips/ghindex.html) but also directly from scientific articles, books, and reports. Scripts in Python were written to navigate into this wealth of toxicological information and to structure and gather the most interesting one. CAS RNs were also retrieved for all the collected chemicals to eliminate the problem of compounds indexed twice or more with different names. This allowed us to eliminate a little bit less than 2,000 LD50s. At the end of the refining process, the toxicological database included about 23,000 rat and mouse oral, i.p., and i.v. LD50s for more than 7,000 organic and inorganic chemicals.

The same strategy was adopted for collecting EC50 values for *Daphnia magna*. Regarding *Vibrio fischeri*, a different approach was used. All the data included in the book of Kaiser and Devillers [24] were first gathered. Thus 1,800 EC50s for 1,290 organic and inorganic molecules and their corresponding CAS RN were collected and structured via Python scripts. This database was then completed from online bibliographic searching in ScienceDirect, Medline, and Google. This allowed us to retrieve 150 additional EC50s corresponding to 110 new molecules. After retrieval of the missing CAS RNs and the removal of duplicates, 82 molecules with their EC50s on *Vibrio fischeri* were added to the initial Microtox™ database.

It is important to note that water solubility data were also collected when available to validate the ecotoxicological data.

For both types of data, the results were not averaged when different values were gathered for the same endpoint and chemical. In that case, the most reliable data were selected. Reliability was mainly based on the existence of test protocols but also on peer review exercises made by experts.

Furthermore, for modeling purposes, it was decided to convert all the (eco)toxicity data into log ($1/C$, C in mmol/kg or mmol/l).

Because sometimes the literature survey showed that the 1-octanol/water partition coefficient (log P) yielded interesting results when introduced as additional variable in the toxicity = f (ecotoxicity) models, it was also decided to consider this parameter in the modeling process. All the log P values were calculated from the KowWin v. 1.67 program [25].

3.2 Linear Regressions of Rat and Mouse LD50s vs. Microtox™ 5-, 15-, and 30-min EC50s

Because Kaiser et al. [11] obtained rather significant correlations between Microtox™ EC50 data and oral, i.p., and i.v. rat and mouse LD50 data, in a first step, an attempt was made to at least confirm their results. However, while Kaiser et al. [11] did not differentiate the time of exposure for *Vibrio fischeri* in their modeling strategy, it was decided to derive different Microtox™ models for the data recorded after 5, 15, and/or 30 min of exposure and the oral, i.p., and i.v. rat and mouse LD50 data. The Rv. 2.3.1 program written in *R* and freely available from the CRAN library, was used for deriving the different toxicity $= f$ (ecotoxicity) models.

Interspecies regressions between mouse and rat LD50 values were first derived for the three routes of exposure yielding (100)–(105).

$$\log (1/\text{LD50 Rat oral}) = 1.01 \log (1/\text{LD50 Mouse oral}), \tag{100}$$
$$n = 633, \; r^2 = 0.89, \; s = 0.29, \; F = 5{,}288.$$

$$\log (1/\text{LD50 Mouse oral}) = 0.88 \log (1/\text{LD50 Rat oral}) - 0.07, \tag{101}$$
$$n = 633, \; r^2 = 0.89, \; s = 0.27, \; F = 5{,}288.$$

$$\log (1/\text{LD50 Rat i.p.}) = 0.95 \log (1/\text{LD50 Mouse i.p.}) - 0.01, \tag{102}$$
$$n = 306, \; r^2 = 0.91, \; s = 0.28, \; F = 3{,}183.$$

$$\log (1/\text{LD50 Mouse i.p.}) = 0.96 \log (1/\text{LD50 Rat i.p.}) - 0.01, \tag{103}$$
$$n = 306, \; r^2 = 0.91, \; s = 0.28, \; F = 3{,}183.$$

$$\log (1/\text{LD50 Rat i.v.}) = 0.99 \log (1/\text{LD50 Mouse i.v.}) + 0.04, \tag{104}$$
$$n = 145, \; r^2 = 0.95, \; s = 0.27, \; F = 2{,}593.$$

$$\log (1/\text{LD50 Mouse i.v.}) = 0.96 \log (1/\text{LD50 Rat i.v.}) - 0.02, \tag{105}$$
$$n = 145, \; r^2 = 0.95, \; s = 0.26, \; F = 2{,}593.$$

In Kaiser et al. [11], mouse toxicity data were only used as independent variables. Nevertheless, comparison of (9)–(11) with (100), (102), and (104) clearly shows that the latter group of models outperforms the former having better statistics and presenting a much larger domain of application.

Rat and mouse oral LD50s were correlated to *V.f.* EC50 values recorded after 5, 15, and 30 min of exposure. The corresponding equations are given below.

$$\log (1/\text{LD50 Rat oral}) = 0.24 \log (1/\text{EC50 } V.f. - 5 \text{ min}) - 0.97, \tag{106}$$
$$n = 339, \; r^2 = 0.15, \; s = 0.75, \; F = 60.1.$$

$$\log (1/\text{LD50 Rat oral}) = 0.22 \log (1/\text{EC50 } V.f. - 15 \text{ min}) - 1.01, \tag{107}$$
$$n = 297, \; r^2 = 0.14, \; s = 0.71, \; F = 48.8.$$

$$\log (1/\text{LD50 Rat oral}) = 0.24 \log (1/\text{EC50 } V.f. - 30 \text{ min}) - 1.01, \tag{108}$$
$$n = 272, \; r^2 = 0.17, \; s = 0.67, \; F = 53.5.$$

$$\log (1/\text{LD50 Mouse oral}) = 0.20 \log (1/\text{EC50 } V.f. - 5 \text{ min}) - 0.93, \quad (109)$$
$$n = 251, \ r^2 = 0.12, \ s = 0.74, \ F = 34.0.$$
$$\log (1/\text{LD50 Mouse oral}) = 0.21 \log (1/\text{EC50 } V.f. - 15 \text{ min}) - 1.03, \quad (110)$$
$$n = 222, \ r^2 = 0.15, \ s = 0.67, \ F = 39.7$$
$$\log (1/\text{LD50 Mouse oral}) = 0.19 \log (1/\text{EC50 } V.f. - 30 \text{ min}) - 0.93, \quad (111)$$
$$n = 209, \ r^2 = 0.11, \ s = 0.69, \ F = 26.3.$$

Equations (106)–(111) show rather poor statistical parameter values but it is important to note that no outlier removal was performed to follow the modeling strategy adopted by Kaiser et al. [11]. Otherwise, the intercepts and slopes of these equations do not differ significantly of those obtained by Kaiser et al. [11] for the oral route entry models for the rat and mouse, (3) and (4).

In the same way, rat and mouse i.p. LD50s were correlated to $V.f.$ EC50 values recorded after 5, 15, and 30 min of exposure yielding (112)–(117).

$$\log (1/\text{LD50 Rat i.p.}) = 0.31 \log (1/\text{EC50 } V.f. - 5 \text{ min}) - 0.65, \quad (112)$$
$$n = 142, \ r^2 = 0.26, \ s = 0.77, \ F = 48.7.$$
$$\log (1/\text{LD50 Rat i.p.}) = 0.31 \log (1/\text{EC50 } V.f. - 15 \text{ min}) - 0.63, \quad (113)$$
$$n = 122, \ r^2 = 0.26, \ s = 0.77, \ F = 41.4.$$
$$\log (1/\text{LD50 Rat i.p.}) = 0.35 \log (1/\text{EC50 } V.f. - 30 \text{ min}) - 0.70, \quad (114)$$
$$n = 126, \ r^2 = 0.31, \ s = 0.74, \ F = 55.7.$$
$$\log (1/\text{LD50 Mouse i.p.}) = 0.29 \log (1/\text{EC50 } V.f. - 5 \text{ min}) - 0.54, \quad (115)$$
$$n = 216, \ r^2 = 0.24, \ s = 0.71, \ F = 66.1.$$
$$\log (1/\text{LD50 Mouse i.p.}) = 0.30 \log (1/\text{EC50 } V.f. - 15 \text{ min}) - 0.60, \quad (116)$$
$$n = 187, \ r^2 = 0.25, \ s = 0.69, \ F = 63.2.$$
$$\log (1/\text{LD50 Mouse i.p.}) = 0.30 \log (1/\text{EC50 } V.f. - 30 \text{ min}) - 0.59, \quad (117)$$
$$n = 169, \ r^2 = 0.26, \ s = 0.66, \ F = 57.3.$$

Equations (112)–(117) present better statistical parameter values than (106)–(111) but they were derived from fewer learning sets. The same tendency was observed when rat and mouse i.v. LD50s were correlated to $V.f.$ EC50 values recorded after 5, 15, and 30 minutes (118)–(123).

$$\log (1/\text{LD50 Rat i.v.}) = 0.43 \log (1/\text{EC50 } V.f. - 5 \text{ min}) - 0.24, \quad (118)$$
$$n = 44, \ r^2 = 0.55, \ s = 0.74, \ F = 51.0.$$
$$\log (1/\text{LD50 Rat i.v.}) = 0.41 \log (1/\text{EC50 } V.f. - 15 \text{ min}) - 0.44, \quad (119)$$
$$n = 30, \ r^2 = 0.65, \ s = 0.63, \ F = 51.7.$$
$$\log (1/\text{LD50 Rat i.v.}) = 0.44 \log (1/\text{EC50 } V.f. - 30 \text{ min}) - 0.38, \quad (120)$$
$$n = 29, \ r^2 = 0.70, \ s = 0.62, \ F = 62.4.$$

$$\log (1/\text{LD50 Mouse i.v.}) = 0.41 \log (1/\text{EC50 } V.\,f. - 5 \text{ min}) - 0.36, \quad (121)$$
$$n = 79,\ r^2 = 0.50,\ s = 0.64,\ F = 78.4.$$

$$\log (1/\text{LD50 Mouse i.v.}) = 0.41 \log (1/\text{EC50 } V.\,f. - 15 \text{ min}) - 0.46, \quad (122)$$
$$n = 70,\ r^2 = 0.56,\ s = 0.60,\ F = 86.8.$$

$$\log (1/\text{LD50 Mouse i.v.}) = 0.39 \log (1/\text{EC50 } V.\,f. - 30 \text{ min}) - 0.39, \quad (123)$$
$$n = 57,\ r^2 = 0.66,\ s = 0.51,\ F = 106.$$

The slopes and intercepts of (106)–(123) appear rather similar for the same route of exposure. This similarity increases when, within each route of exposure, the equations are matched according to the time of exposure (i.e., 5, 15, 30 min). There is a significant increase in the regression slopes in the order oral < intraperitoneal < intravenous exposure. As stressed by Kaiser et al. [11], the change in the slopes results from the corresponding decrease in metabolic degradation with a decreasing requirement for cell membrane diffusion and resulting higher efficacy of the intravenous route relative to the oral exposure. Correlations with rats always outperform those with mice except when the rat and mouse oral LD50s are correlated to the Microtox$^{\text{TM}}$ 15-min EC50s (i.e., (107) vs. (110)). This is surprising due to the high level of correlation that exists between the two mammalian species (100)–(105).

While Kaiser et al. [11] did not distinguish the time of exposure with *Vibrio fischeri*, in the present study, different equations were produced with the Microtox$^{\text{TM}}$ data recorded after 5, 15, and 30 min of exposure. This difference of strategy with Kaiser et al. [11] cannot explain alone the difference of size of the learning sets between the two studies, especially if we consider that our databases were larger than those of Kaiser et al. [11]. This claim can be easily verified when we compare the size of the learning sets used in both studies for deriving the correlations between rat and mouse LD50 values. For the oral, i.p., and i.v. routes of exposure the learning sets used by Kaiser et al. [11] included 330, 162, and 41 chemicals, respectively (9)–(11). In the present study, the same sets included 633 (100), 306 (102), and 145 molecules (104), respectively.

This difference in the number of chemicals used in the linear regressions of rat and mouse LD50s vs. Microtox$^{\text{TM}}$ EC50s should be also explained by the fact that Kaiser et al. [11] did not take into account the hydrosolubility values of the chemicals in the selection of their ecotoxicity data. Thus, for example, inspection of Table 4 (page 1,604) of their paper shows that they selected a value of 1.41 (in log $1/C$ mmol/l) for the EC50 of p, p'-DDT against *Vibrio fischeri*. This corresponds to an EC50 value equal to 13.39 mg/l while the hydrosolubility of this chemical at 15 °C is only 0.017 mg/l [26].

In addition, while in the present study, the EC50 values only determined with superior or inferior limits were discarded from the database, this rule was not adopted by Kaiser et al. [11]. Thus, for example, an intensive bibliographical search on the Microtox$^{\text{TM}}$ toxicity of mitomycin C only provided one reference showing that the EC50 values of this chemical against *Vibrio fischeri* after 5, 10, 15, and 20 min of exposure were < 16, <16.1, <15.2, and <13.7 mg/l, respectively [27]. Surprisingly, Kaiser et al. [11] selected a value of 1.39 (in log $1/C$ mmol/l), which is equivalent to 13.7 mg/l (see Table 4, page 1,604).

Consequently, even if the toxicity $= f$ (ecotoxicity) models of the present study were derived from fewer training sets than those of Kaiser et al. [11], they present better foundations.

Inspection of our Microtox^{TM} database showed that while some chemicals were characterized by 5-, 15-, and 30-min EC50s, for others, the EC50 values were only available for one or two times of exposure. This prompted us to first derive equations allowing the prediction of 30-min EC50s from 5 and 15-min EC50s and then to use the observed and calculated Microtox^{TM} 30-min EC50s for computing new toxicity $= f$ (ecotoxicity) models from larger training sets. This work is presented in the next section. It is noteworthy that all the calculations were made with Statistica ver. 6 (StatSoft, Paris).

3.3 Linear Regressions of Rat and Mouse LD50s vs. Microtox^{TM} 30-min* EC50s

The two models allowing the prediction of Microtox^{TM} 30-min EC50 values from 5-min EC50s or 15-min ECs are given below, (124) and (125). They are highly statistically significant. Inspection of these models let to suppose that no difference exist between the EC50 data recorded after 5, 15, or 30 min of exposure. Although it is true for chemicals, it is totally wrong for others [24, 28]. Undoubtedly, the best strategy would consist in the design of specific models for encoding these different particularities but for the sake of simplicity we decided not to do so.

$$\log 1/\text{EC50} - 30 \text{ min} = \log 1/\text{EC50} - 5 \text{ min} + 0.03, \qquad (124)$$
$$n = 951, \ r = 0.98, \ s = 0.22, \ F = 22,785, \ p < 10^{-5}.$$
$$\log 1/\text{EC50} - 30 \text{ min} = \log 1/\text{EC50} - 15 \text{ min} - 0.01, \qquad (125)$$
$$n = 903, \ r = 0.996, \ s = 0.1, \ F = 108,400, \ p < 10^{-5}.$$

In the models, the *V.f.* variable being constituted of observed and approximated Microtox^{TM} 30-min EC50 values, it is spotted by an asterisk to avoid confusions with the previous models.

$$\log (1/\text{LD50 Rat oral}) = 0.25 \log (1/\text{EC50 } V.f. - 30 \text{ min}^*) - 1.00, \qquad (126)$$
$$n = 407, \ r = 0.46, \ s = 0.66, \ F = 106, \ p < 10^{-5}.$$
$$\log (1/\text{LD50 Mouse oral}) = 0.23 \log (1/\text{EC50 } V.f. - 30 \text{ min}^*) - 0.98, \qquad (127)$$
$$n = 297, \ r = 0.43, \ s = 0.69, \ F = 67.3, \ p < 10^{-5}.$$
$$\log (1/\text{LD50 Rat i.p.}) = 0.33 \log (1/\text{EC50 } V.f. - 30 \text{ min}^*) - 0.77, \qquad (128)$$
$$n = 159, \ r = 0.65, \ s = 0.61, \ F = 117, \ p < 10^{-5}.$$
$$\log (1/\text{LD50 Mouse i.p.}) = 0.32 \log (1/\text{EC50 } V.f. - 30 \text{ min}^*) - 0.60, \qquad (129)$$
$$n = 239, \ r = 0.60, \ s = 0.61, \ F = 132, \ p < 10^{-5}.$$

$$\log (1/\text{LD50 Rat i.v.}) = 0.42\,(\log 1/\text{EC50}\ V.f. - 30\,\text{min}^*) - 0.29, \qquad (130)$$
$$n = 49,\ r = 0.79,\ s = 0.71,\ F = 75.2,\ p < 10^{-5}.$$

$$\log (1/\text{LD50 Mouse i.v.}) = 0.42 \log (1/\text{EC50}\ V.f. - 30\,\text{min}^*) - 0.44, \qquad (131)$$
$$n = 92,\ r = 0.84,\ s = 0.50,\ F = 207,\ p < 10^{-5}.$$

Equations (126)–(131) significantly outperform (106)–(123). This is mainly due to the removal of some outliers. In the first step of this study, the goal was mainly to confirm or infirm the results obtained by Kaiser et al. [11], and hence it was necessary to follow at best their methodology, which first consisted in considering the whole datasets without outlier removal. In a second step, also in agreement with the strategy used by Kaiser et al. [11], an attempt was made to optimize a little bit the equations. Most of the eliminated outliers were inorganic chemicals for which it was difficult to know whether the Microtox$^{\text{TM}}$ EC50 values were reported to the element, the salt, etc. Inspection of (126)–(131) shows that the slopes of the models are similar for the same route of exposition in rats and mice.

3.4 Linear Regressions of Rat and Mouse LD50s vs. Daphnia magna 48-h* EC50s

The database on *Daphnia magna* including EC50 values recorded after 24 and 48 h of exposure, a regression equation was first computed to convert the 24-h EC50s into 48-h EC50s, (132).

$$\log 1/\text{EC50} - 48\,\text{h} = 0.99 \log (1/\text{EC50} - 24\,\text{h}) + 0.29, \qquad (132)$$
$$n = 258,\ r = 0.97,\ s = 0.40,\ F = 4{,}769.$$

Equation (132) presents a high predictive power as well as a large domain of application, which is clearly shown in Fig. 1.

The models allowing the prediction of LD50s in rats and mice after oral, i.p., and i.v. absorption are presented below. Because the models include observed and calculated *Daphnia magna* 48-h EC50 values from (132), an asterisk is used to characterize the independent variable.

$$\log (1/\text{LD50 Rat oral}) = 0.30 \log (1/\text{EC50}\ D.m. - 48\,\text{h}^*) - 1.13, \qquad (133)$$
$$n = 588,\ r = 0.66,\ s = 0.63,\ F = 448,\ p < 10^{-5}.$$

$$\log (1/\text{LD50 Mouse oral}) = 0.28 \log (1/\text{EC50}\ D.m. - 48\,\text{h}^*) - 1.06, \qquad (134)$$
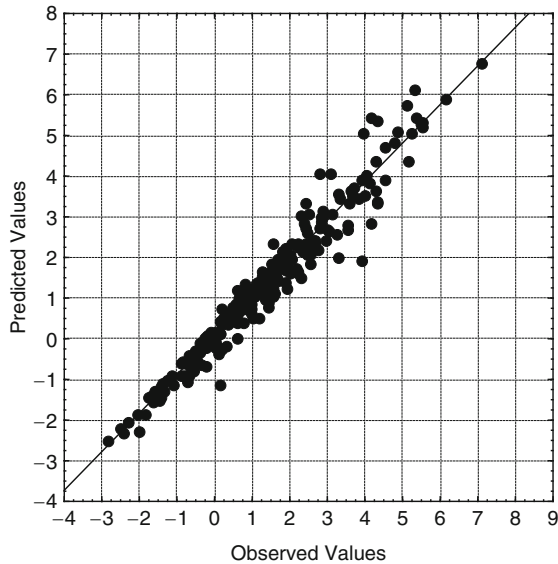$$n = 374,\ r = 0.64,\ s = 0.64,\ F = 254,\ p < 10^{-5}.$$

$$\log (1/\text{LD50 Rat i.p.}) = 0.35 \log (1/\text{EC50}\ D.m. - 48\,\text{h}^*) - 0.72, \qquad (135)$$
$$n = 191,\ r = 0.75,\ s = 0.62,\ F = 237,\ p < 10^{-5}.$$

$$\log (1/\text{LD50 Mouse i.p.}) = 0.33 \log (1/\text{EC50}\ D.m. - 48\,\text{h}^*) - 0.62, \qquad (136)$$
$$n = 261,\ r = 0.69,\ s = 0.62,\ F = 241,\ p < 10^{-5}.$$

Fig. 1 Observed vs. calculated *Daphnia* 48-h EC50 values ($-\log$ (mmol/l)) from model (132)

$$\log (1/\text{LD50 Rat i.v.}) = 0.43 \log (1/\text{EC50 } D.m. - 48\,\text{h}^*) - 0.28, \quad (137)$$
$$n = 61, \ r = 0.87, \ s = 0.62, \ F = 181, \ p < 10^{-5}.$$
$$\log (1/\text{LD50 Mouse i.v.}) = 0.38 \log (1/\text{EC50 D.m.} - 48\,\text{h}^*) - 0.39, \quad (138)$$
$$n = 108, \ r = 0.79, \ s = 0.61, \ F = 175, \ p < 10^{-5}.$$

The observed vs. calculated LD50s from (133) to (138) are displayed in Figs. 2–7.

Inspection of (126)–(138) shows that it is preferable to predict rat and mouse oral and i.p. LD50s and rat i.v. LD50s from EC50s obtained from *Daphnia magna* instead of *Vibrio fischeri*, while it is the converse regarding the intravenous route of exposure in mouse.

From these results, it was interesting to test whether the use of *Vibrio fischeri* and *Daphnia magna* as independent variables in the rat and mouse regression equations improved their predictive power. The obtained results are presented in the next section.

3.5 Linear Regressions of Rat and Mouse LD50s vs. Microtox™ 30-min* EC50s + Daphnia magna 48-h* EC50s

The confrontation of the oral LD50s on rat and mouse with the EC50 values for *Vibrio fischeri* and *Daphnia magna* did not yield statistically significant two parameter equations. Conversely, statistically valid models were obtained with the i.p. LD50s, (139) and (140).
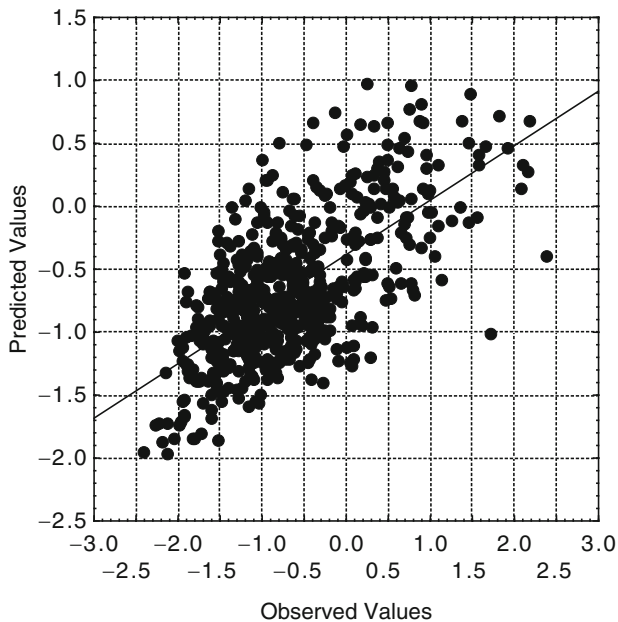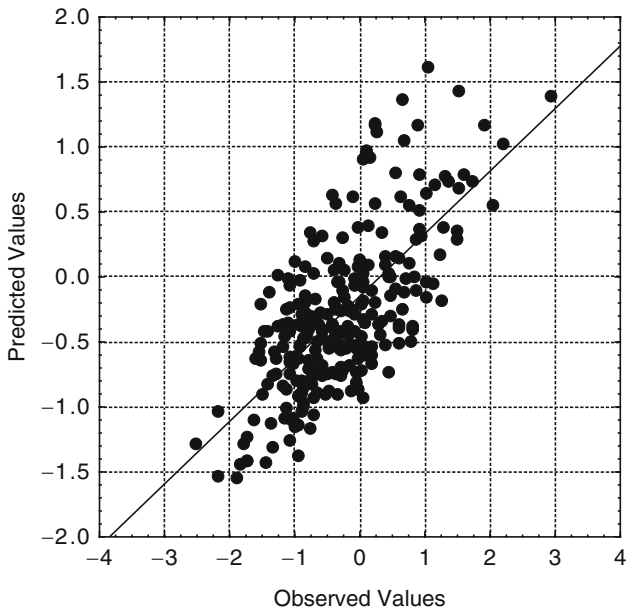
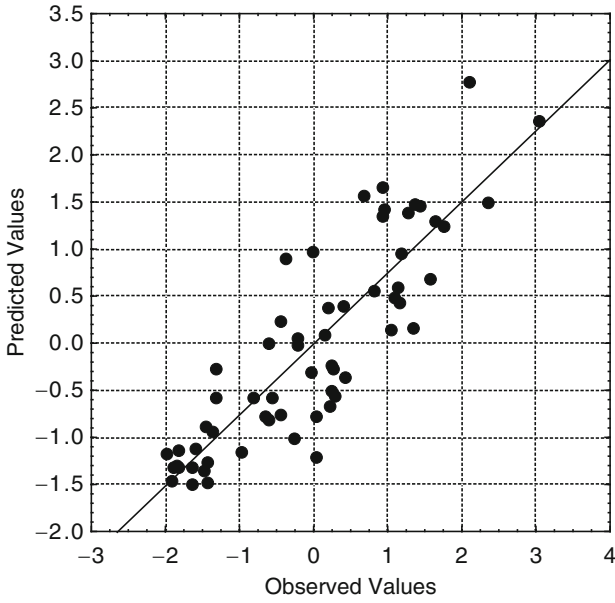**Fig. 2** Observed vs. calculated rat oral LD50 values (− log (mmol/kg)) from model (133)
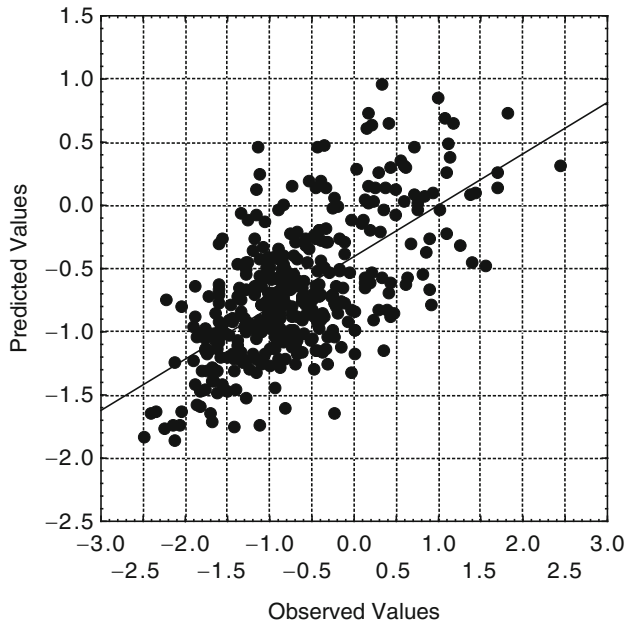
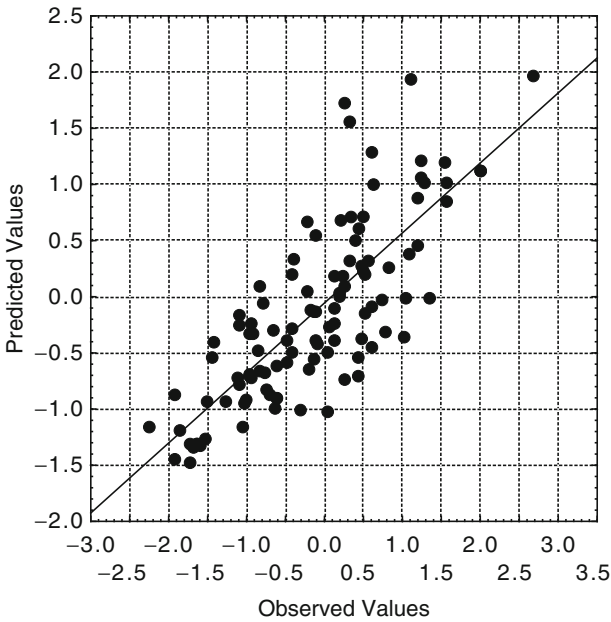**Fig. 3** Observed vs. calculated mouse oral LD50 values (− log (mmol/kg)) from model (134)
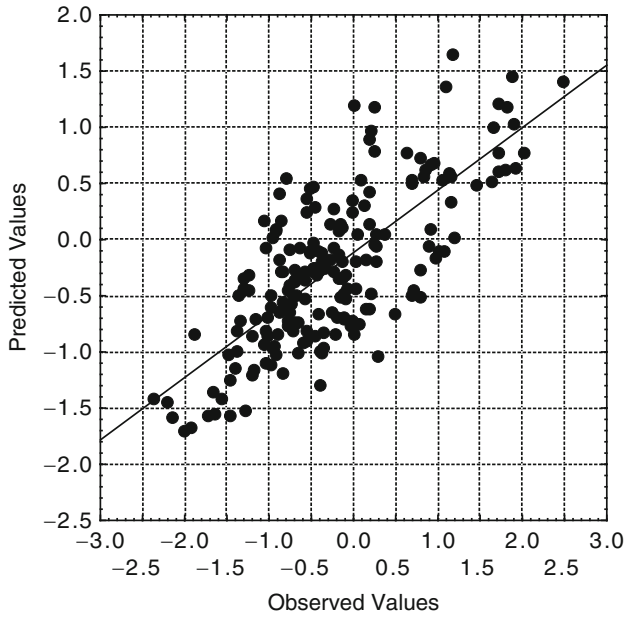
Fig. 4 Observed vs. calculated rat intraperitoneal LD50 values (−log (mmol/kg)) from model (135)

Fig. 5 Observed vs. calculated mouse intraperitoneal LD50 values (−log (mmol/kg)) from model (136)

Fig. 6 Observed vs. calculated rat intravenous LD50 values (− log (mmol/kg)) from model (137)

Fig. 7 Observed vs. calculated mouse intravenous LD50 values (− log (mmol/kg)) from model (138)

$$\log (1/\text{LD50 Rat i.p.}) = 0.13 \log (1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$+ 0.25 \log (1/\text{EC50 } V.f. - 30\,\text{min}^*) - 0.87, \quad (139)$$
$$n = 99, \ r = 0.75, \ s = 0.55, \ F = 61.0, \ p < 10^{-5}.$$

$$\log (1/\text{LD50 Mouse i.p.}) = 0.23 \log (1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$+ 0.14 \log (1/\text{EC50 } V.f. - 30\,\text{min}^*) - 0.75, \quad (140)$$
$$n = 128, \ r = 0.73, \ s = 0.55, F = 71.4, \ p < 10^{-5}.$$

Equation (139) presents a better correlation coefficient and standard error than (128), which only includes *Vibrio fischeri* as independent variable but the former model was obtained from 99 chemicals while the latter was derived from 159 compounds. In the same way, while (139) shows slightly better statistics than (135) with only *Daphnia magna* as independent variable, the size of its training set is about twice less important (i.e., 99 vs. 191).

Equation (140) outperforms the corresponding univariate regression equations (i.e. (129) and (136)) but again there are important differences in the size of the training sets (i.e. 128 vs. 239 and 261).

Although the confrontation of the rat intravenous LD50s with the EC50 values for *Vibrio fischeri* and *Daphnia magna* did not yield a statistically significant two parameter equation, an interesting model was obtained with the mouse data (141).

$$\log (1/\text{LD50 Mouse i.v.}) = 0.14 \log (1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$+ 0.33 \log (1/\text{EC50 } V.f. - 30\,\text{min}^*) - 0.58, \quad (141)$$
$$n = 51, \ r = 0.89, \ s = 0.47, \ F = 92.3, \ p < 10^{-5}.$$

Again, (141) outperforms (131) and (138) but having a lower training set, its domain of application is also less important.

It is interesting to note that in (139)–(141), *Daphnia magna* and *Vibrio fischeri* contribute positively for predicting rat and mouse LD50s.

Because the 1-octanol/water partition coefficient ($\log P$) seemed yield interesting results when introduced as additional variable in the toxicity $= f$ (ecotoxicity) models [20, 22], it was also decided to consider this important physicochemical parameter as additional independent variable in the models. The results obtained with this descriptor of the hydrophobicity of chemicals are presented in the next section.

3.6 Introduction of log P in the Regressions of Rat and Mouse LD50s vs. Vibrio and Daphnia EC50s

A stepwise regression analysis was first used to correlate rat and mouse LD50 data with Microtox$^{\text{TM}}$ 30-min EC50 or daphnid 48-h EC50 data, and $\log P$ values calculated from the KowWin v. 1.67 program [25].

Regarding the oral and intravenous LD50 data, only a two parameter equation was obtained for the oral toxicity on rat (142) the others being not statistically significant.

$$\log (1/\text{LD50 Rat oral}) = 0.32 \log (1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$- 0.03 \log P - 1.16, \tag{142}$$
$$n = 478,\ r = 0.71,\ s = 0.54,\ F = 247,\ p < 10^{-5}.$$

Even if the contribution of the 1-octanol/water partition coefficient ($\log P$) in (142) is low, the introduction of this hydrophobic parameter in the model increases its quality. This is clearly shown when the statistical parameters of (133) and (142) are compared as well as the scatterplots of the LD50s obtained from both models (Figs. 2 and 8).
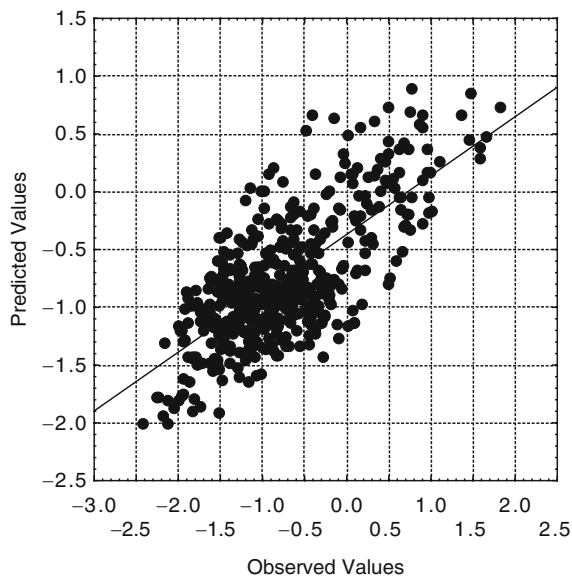
Four equations (143)–(146) were successfully computed for predicting the i.p. toxicity of chemicals to rat and mouse from *Vibrio fischeri* or *Daphnia magna* EC50 data and $\log P$.

$$\log (1/\text{LD50 Rat i.p.}) = 0.36 \log (1/\text{EC50 } V.f. - 30\,\text{min}^*)$$
$$- 0.1 \log P - 0.67, \tag{143}$$
$$n = 149,\ r = 0.64,\ s = 0.58,\ F = 51.0,\ p < 10^{-5}.$$
$$\log (1/\text{LD50 Rat i.p.}) = 0.40 \log (1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$- 0.05 \log P - 0.75, \tag{144}$$
$$n = 150,\ r = 0.81,\ s = 0.55,\ F = 138,\ p < 10^{-5}.$$

Fig. 8 Observed vs. calculated rat oral LD50 values ($-\log$ (mmol/kg)) from model (142)

$$\log(1/\text{LD50 Mouse i.p.}) = 0.34 \log(1/\text{EC50 } V.f. - 30\,\text{min}^*)$$
$$- 0.07 \log P - 0.55, \qquad (145)$$
$$n = 227, \; r = 0.59, \; s = 0.58, \; F = 59.5, \; p < 10^{-5}.$$

$$\log(1/\text{LD50 Mouse i.p.}) = 0.37 \log(1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$- 0.07 \log P - 0.66, \qquad (146)$$
$$n = 202, \; r = 0.72, \; s = 0.59, \; F = 108, \; p < 10^{-5}.$$

The influence of $\log P$ in (143)–(146) is very limited. Furthermore, while the introduction of $\log P$ in the models with *Daphnia magna* slightly increases their quality, this is the converse regarding *Vibrio fischeri*.

Last, a stepwise regression analysis used to correlate LD50 data with Microtox$^{\text{TM}}$ 30-min EC50 data, daphnid 48-h EC50 data and $\log P$ values only yielded satisfying results for the rat and mouse intraperitoneous toxicity data, the former model (147) outperforming the latter, (148).

$$\log(1/\text{LD50 Rat i.p.}) = 0.21 \log(1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$+ 0.24 \log(1/\text{EC50 } V.f. - 30\,\text{min}^*)$$
$$- 0.15 \log P - 0.72, \qquad (147)$$
$$n = 91, \; r = 0.80, \; s = 0.46, \; F = 49.9, \; p < 10^{-5}.$$

$$\log(1/\text{LD50 Mouse i.p.}) = 0.25 \log(1/\text{EC50 } D.m. - 48\,\text{h}^*)$$
$$+ 0.16 \log(1/\text{EC50 } V.f. - 30\,\text{min}^*)$$
$$- 0.13 \log P - 0.63, \qquad (148)$$
$$n = 120, \; r = 0.74, \; s = 0.48, \; F = 47.8, \; p < 10^{-5}.$$

Equations (147) and (148) outperform (139) and (140) without $\log P$. Inspection of these equations shows the *Daphnia magna* and *Vibrio fischeri* contribute positively to the toxicity, the values of their coefficients in the regression equations being only slightly different. Conversely, in both equations, $\log P$ contributes negatively to the toxicity. The observed and calculated i.p. LD50s with these two models are displayed in Figs. 9 and 10.

3.7 Linear vs. Nonlinear Interspecies Toxicity Modeling

Because numerous QSAR studies have shown that the artificial neural networks (ANNs) outperformed the classical linear methods to find complex relationships between the structure of the molecules and their biological activity [28, 29], see also Chap. 1), an attempt was made to test the usefulness of these nonlinear tools for designing toxicity $= f$ (ecotoxicity) models. The ANN the most suited for this kind of study was a three layer perceptron (TLP) [30]. A TLP requiring at least three neurons in the input layer and two neurons in the hidden layer to be used reasonably, only (147) and (148) were concerned by the comparison exercise.
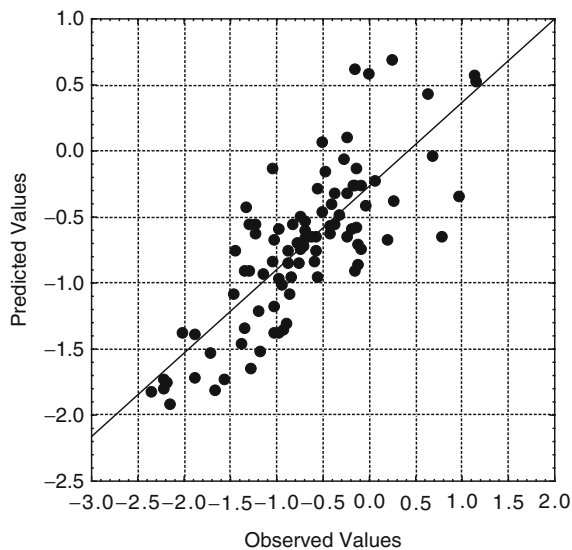
Fig. 9 Observed vs. calculated rat intraperitoneal LD50 values ($-$log (mmol/kg)) from model (147)
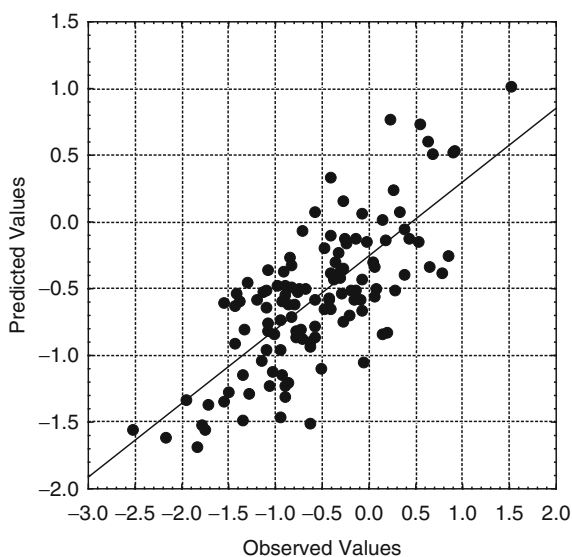
Fig. 10 Observed vs. calculated mouse intraperitoneal LD50 values ($-$log (mmol/kg)) from model (148)

Thus, a TLP was used for predicting i.p. rat and mouse LD50 data from Microtox^{TM} 30-min EC50, daphnid 48-h EC50, and log P data. Ten percent of the data used for deriving (147) and (148) were randomly selected for constituting the external testing sets, the remaining data being used as learning sets to train a 3/2/1

TLP. A classical min/max transformation was used to scale the data. All the calculations were performed with the Statistica ANN software (Statsoft, Paris). Different learning algorithms (back-propagation, conjugate gradient descent, Levenberg-Marquardt, quick-prop), activation and transfer functions were tested. Despite more than 7,500 runs for each endpoint by randomly varying the composition of the training and testing sets as well as the architecture of the ANN, it was impossible to obtain significantly better results for the testing sets than those obtained with a regression analysis performed from the same chemical data sets. The results were the same or at least slightly better than those recorded with the stepwise regression analysis. Obviously, the results obtained with the learning sets were always better than those recorded with regression analysis but it is not surprising, the TLPs being powerful learning devices.

It is noteworthy that no attempts were made to increase the number of neurons in the hidden layer.

Because the support vector machines (SVMs) [31] have shown their interest for classification problems and more recently to correlate data of various origins, an attempt was also made to test their usefulness on the different learning and testing sets previously used with the ANNs. The e1071 program, written in R and freely available from the CRAN library, was employed for deriving the different toxicity $= f$ (ecotoxicity) models. Unfortunately, due to different constraints, only a limited number of runs was performed. They provided better results on the external testing sets than the TLP but additional investigations are absolutely necessary to correctly estimate the performances and interest of SVM for predicting LD50s from ecotoxicity data.

Last, different nonlinear regression analyses were tested by using Statistica v. 6 as well as CurveExpert v. 1.3. The most interesting regression analysis presented the following form: $y = a(1 - e^{-bx})$. However, it was impossible to obtain significantly better results for the testing sets than those obtained with a regression analysis performed under the same conditions. Consequently, it was not justified to select these nonlinear equations as final models.

4 Conclusion

Surprisingly, there is a limited number of models available in the literature for predicting the acute toxicity of chemicals to rats and mice from ecotoxicity test data. Furthermore, when available, most of the models were derived from limited datasets. Thus about 70% of the models found in the literature were obtained from less than 50 chemicals. Moreover, it is important to note that chemicals are very often eliminated before or during the regression processes without clear justifications. Consequently, despite some significant correlations, these models cannot be used in practice. This prompted us to develop new models focused on the prediction of rat and mouse LD50s from invertebrate EC50s or LC50s. The selected species were *Daphnia magna* and *Vibrio fischeri* because these organisms are widely used

for assessing the hazard of chemicals and hence, collections of EC50s for these organisms are available in the literature.

Consequently, in a first step, a strong bibliographical investigation was performed to collect oral, i.p., and i.v. rat and mouse LD50 data. In the meantime, EC50 data for *Vibrio fischeri* (Microtox^{TM} test) and *Daphnia magna* were also retrieved from literature. Python scripts were used to structure and format the toxicity and ecotoxicity data to facilitate their statistical manipulation.

A collection of oral, i.p., and i.v. rat and mouse toxicity models was derived using *Vibrio fischeri* and *Daphnia magna* as independent variables alone or together through a stepwise regression analysis. Most of the models on *Daphnia magna* are totally new. They outperform those obtained with *Vibrio fischeri*. The usefulness of the 1-octanol/water partition coefficient (log P) as additional independent variable was also tested. When included in the models, its contribution is always negative and generally marginal, except in the case of the three parametric equations including *Daphnia magna* and *Vibrio fischeri* for the prediction of rat and mouse intraperitoneal LD50s.

The interest of nonlinear statistical tools for deriving toxicity $= f$ (ecotoxicity) models was also experienced. The results obtained with a three-layer perceptron and different nonlinear regressions were disappointing. The SVMs seemed to yield more interesting results but more investigations should be necessary to see whether they are more suited than classical regression analysis for deriving toxicity $= f$ (ecotoxicity) models.

References

1. Gauthier C, Griffin G (2005) Using animals in research, testing and teaching. Rev Sci Tech Off Int Epiz 24: 735–745
2. Council of Europe (CoE) (1986) European Convention for the protection of vertebrate animals used for experimental and other scientific purposes. Strasbourg, 18.III.1986. CoE, Strasbourg
3. European Community (1986) Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. Off J Eur Communities L 358: 1–28
4. van Zutphen B (2007) Legislation of animal use – Developments in Europe. AATEX 14: 805–809
5. Bayvel ACD (2005) The use of animals in agriculture and science: Historical context, international considerations and future direction. Rev Sci Tech Off Int Epiz 24: 791–797
6. Russell WMS, Burch RL (1959) The principles of humane experimental technique. Methuen & Co. Ltd, London (altweb.jhsph.edu/publications/humane_exp/het-toc.htm (accessed on 11 January 2008)
7. European Community (2007) Fifth report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. Brussels, 5.11.2007, COMM(2007) 675 final
8. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals

(REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC

9. Kaiser KLE (1998) Correlations of *Vibrio fischeri* bacteria test data with bioassay data for other organisms. Environ Health Perspect 106(Suppl 2): 583–591

10. Fort FL (1992) Correlation of Microtox EC_{50} with mouse LD_{50}. In Vitro Toxicol 5: 73–82

11. Kaiser KLE, McKinnon MB, Fort FL (1994) Interspecies toxicity correlations of rat, mouse and *Photobacterium phosphoreum*. Environ Toxicol Chem 13: 1599–1606

12. Sauvant MP, Pepin D, Bohatier J, Groliere CA, Guillot J (1997) Toxicity assessment of 16 inorganic environmental pollutants by six bioassays. Ecotoxicol Environ Safety 37: 131–140

13. Calleja MC, Persoone G (1992) Cyst-based toxicity tests. IV. The Potential of ecotoxicological tests for the prediction of acute toxicity in man as evaluated on the first ten chemicals of the MEIC programme. ATLA 20: 396–405

14. Khangarot BS, Ray PK (1988) The confirmation of a mammalian poison classification using a water flea (*Daphnia magna*) screening method. Arch Hydrobiol 113: 447–455

15. Enslein K, Tuzzeo TM, Borgstedt HH, Blake BW, Hart JB (1987) Prediction of rat oral LD50 from *Daphnia magna* LC50 and chemical structure. In Kaiser KLE (ed) QSAR in Environmental Toxicology – II, Reidel Publishing Company, Dordrecht

16. Enslein K, Tuzzeo TM, Blake BW, Hart JB, Landis WG (1989) Prediction of *Daphnia magna* EC50 values from rat oral LD50 and structural parameters. In Suter GW, Lewis MA (eds) Aquatic Toxicology and Environmental Fate, Vol. 11, ASTM STP 1007, Philadelphia

17. Janardan SK, Olson CK, Schaeffer DJ (1984) Quantitative comparisons of acute toxicity of organic chemicals to rat and fish. Ecotoxicol Environ Safety 8: 531–539

18. Johnson WW, Finley MT (1980) Handbook of acute toxicity of chemicals to fish and aquatic invertebrates. Resource Publication 135. US Department of the Interior, Fish and Wildlife Service, Washington, DC

19. Gaines TB (1969) Acute toxicity of pesticides. Toxicol Appl Pharmacol 14: 515–534

20. Kaiser KLE, Esterby SR (1991) Regression and cluster analysis of the acute toxicity of 267 chemicals to six species of biota and the octanol/water partition coefficient. Sci Total Environ 109/110: 499–514

21. Hodson PV (1985) A comparison of the acute toxicity of chemicals to fish, rats and mice. J Appl Toxicol 5: 220–226

22. Delistraty D, Taylor B, Anderson R (1998) Comparisons of acute toxicity of selected chemicals to rainbow trout and rats. Ecotoxicol Environ Safety 39: 195–200

23. Delistraty D (2000) Acute toxicity to rats and trout with a focus on inhalation and aquatic exposures. Ecotoxicol Environ Safety 46: 225–233

24. Kaiser KLE, Devillers J (1994) Ecotoxicity of chemicals to *Photobacterium phosphoreum*. Gordon and Breach Science Publishers, Reading

25. Meylan WM, Howard PH (1995) Atom/fragment contribution method for estimating octanol–water partition coefficients. J Parm Sci 84: 83–92

26. Biggar JW, Riggs RL (1974) Apparent solubility of organochlorine insecticides in water at various temperatures. Hilgardia 42: 383–391

27. Yates IE (1985) Differential sensitivity to mutagens by *Photobacterium phosphoreum*. J Microbiol Meth 3: 171–180

28. Devillers J, Domine D (1999) A noncongeneric model for predicting toxicity of organic molecules to *Vibrio fischeri*. SAR QSAR Environ Re. 10: 61–70

29. Devillers J (1996) Neural networks in QSAR and drug design. Academic Press, London

30. Eberhart RC, Dobbins RW (1990) Neural network PC tools. A practical guide, Academic Press, San Diego

31. Vapnik V (1995) The nature of statistical learning theory. Springer-Verlag, Berlin