Bull Environ Contam Toxicol (2011) 87:494–498
DOI 10.1007/s00128-011-0393-x

A Linear Model to Predict Chronic Effects of Chemicals on *Daphnia magna*

Enrico Mombelli · Alexandre R. R. Pery

Received: 31 January 2011 / Accepted: 25 August 2011 / Published online: 10 September 2011

Abstract Chronic toxicity data for *Daphnia magna* are information requirements in the context of regulations on chemical safety. This paper proposes a linear model for the prediction of chemically-induced effects on the reproductive output of *D. magna*. This model is based on data retrieved from the Japanese Ministry of Environment database and it predicts chronic effects as a function of acute toxicity data. The proposed model proved to be able to predict chronic toxicities for chemicals not used in the training set. Our results suggest that experiments involving chronic exposure to chemicals could be reduced thanks to the proposed model.

Keywords *Daphnia magna* · Chronic toxicity · Acute toxicity · Predictive model · OECD (Q)SAR Toolbox

Daphnia magna is a crustacean in the order Cladocera that is widely used for conducting acute and chronic toxicological testing because of its ecological importance and responsiveness to a wide range of chemical stressors (Adema 1978). Moreover, it is the preferred species for short and long-term toxicological testing on invertebrates in the framework of the European regulation REACH (EC 2006) (Registration, Evaluation, Authorization and Restriction of Chemical substances) on chemical safety (Henegar et al. 2011). In the context of this regulation, long-term aquatic toxicity studies on *D. magna* are a mandatory requirement for substances manufactured or imported in quantities of 100 tons or more, but these studies have also to be considered for lower tonnage levels if the substance is poorly water soluble (EC 2006).

The assessment of long-term toxicity is demanding in terms of time and resources, and the possibility of predicting chronic EC_{50} values without the necessity of performing these time-consuming experiments would be an advantage in economic and operational terms. This paper is therefore focused on the statistical modeling of chemically-induced chronic toxicity effects on *D. magna* in order to quantify the impact that chemicals have on the reproduction of the crustacean. In the model proposed in this paper, long-term effects on reproduction are predicted as a function of acute toxicity data (immobilization test) whose determination takes only 48 h (OECD 1984) instead of the 21 days required by the standard reproduction test (OECD 1998).

The only works we found relative to the extrapolation to chronic effects in *D. magna* were performed by Santojanni et al. (1998), who acknowledged the robustness of the relationship between body length at the 7th and 15th day and fecundity. Our work extends these observations further by proposing statistical models that can directly predict EC_{50} for chronic effects as a function of EC_{50} for acute effects. To our knowledge, there are no available models for the direct prediction of EC_{50} chronic toxicity values for *D. magna* and the results discussed in this paper can provide a practical and useful tool for the assessment of the toxicity of chemicals to aquatic species. It is also important to point out that, according to the current state of knowledge on daphnids, mechanistic links between acute effects on mortality/immobilization and chronic effects on

Electronic supplementary material The online version of this article (doi:10.1007/s00128-011-0393-x) contains supplementary material, which is available to authorized users.

E. Mombelli (✉) · A. R. R. Pery
Unité Modèles Pour l'Ecotoxicologie et la Toxicologie (METO), Institut National de l'Environnement Industriel et des Risques (INERIS), Verneuil en Halatte, France
e-mail: enrico.mombelli@ineris.fr

reproduction have not been established yet. Nevertheless, this article highlights that for the adopted database there is a statistically sound and predictive linear relationship between the two toxicities.

Materials and Methods

Chronic and acute toxicity EC_{50} values were retrieved from the website of the Japanese MoE (2010). Data therein reported were obtained by adopting the OECD technical guidance 202 for acute toxicity tests (OECD 1984) and the OECD technical guidance 211 for chronic toxicity tests (OECD 1998). Some of the tests performed before 2002 were conducted using a dispersant but it was impossible to identify the concerned chemicals on the basis of existing information. Before undergoing statistical analysis, acute and chronic toxicities were converted into decimal logarithm of the reciprocal of EC_{50} values (expressed in mmol/L).

Only data relative to discrete organic chemicals were considered during our analysis and data on inorganic substances and polymers were discarded. Chemicals characterized by a ratio between chronic and acute $EC_{50} > 1$ were also not considered when deriving our model. Statistical analyses were carried out thanks to the free software environment R v2.12.0 (R Development Core Team 2010) and the following packages: robustbase (for robust regressions) and gvlma (for the global validation of linear models assumptions). The aquatic toxicity classifications mentioned in the paper (e.g., anilines, neutral organics) were obtained by applying the ECOSAR aquatic toxicity profiler integrated within the OECD (Q)SAR Toolbox v2.0 (OECD 2010). Values for the partition coefficient logP between *n*-octanol and water and water solubility (logS) were estimated thanks to the free online software (http://www.vcclab.org/) ALOGPS 2.1 (Tetko et al. 2005). Log *D* estimations were carried out thanks to the online software (http://www.archemcalc.com/sparc/) SPARC (Hilal et al. 1995).

Results and Discussion

Acute Toxicity as a Predictor Variable

Acute toxicity values (48 h EC50) and chronic toxicity values (21 days EC50) for 244 chemicals (Table 1 in supporting information) were retrieved from the MoE database (MoE 2010). Seventy-two chemicals were included in an external test set that was not used in the development of the linear regression model. This external set served only as an independent evaluation set for the assessment of the predictive performance of the regression equation. The chemicals composing the external test set were chosen in a way to cover the whole toxicity range of the entire dataset while also adequately representing its structural heterogeneity. The remaining 172 chemicals (training set) were used to derive a model for the prediction of chronic effects on *D. magna* as a function of acute toxicity data obtained on the same crustacean.

In order to check for the presence of outliers, a robust regression was carried out according to the procedure described in (Varmuza and Filzmoser 2009). The plot showing the robust standardized residuals versus the robust Mahalanobis distances highlighted eight outliers in the *y*-variable (i.e., chronic toxicity) that lie beyond the two horizontal dotted lines at ± 2.5 that separate regular observations from outliers (Fig. 1). According to the ECOSAR profiler of the OECD QSAR Toolbox, seven of these data points are anilines (chemicals 5, 19, 72, 95, 155 and 158, 172 listed in the supplementary material together with all the other chemicals hereafter cited) and one of them (chemical 62) is a neutral organic. All of these outliers are characterized by high values of the ratio between acute and chronic EC_{50}. Indeed, several published studies (Dom et al. 2010; Neuwoehner et al. 2010) indicate that anilines are particularly toxic towards daphnids by means of a specific mode of action involving the steric accessibility of the NH_2 group (Ramos et al. 2002). Three data points are also outlying in the *x*-space (i.e., acute toxicity) while being inliers for the *y*-variable. These objects are characterized by a high leverage but they stabilize the

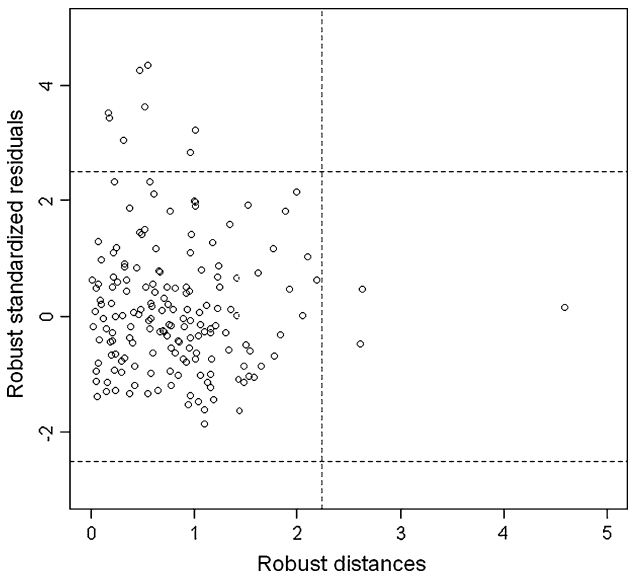

Fig. 1 Diagnostic plot for the robust regression on *D. magna* data. Data beyond the *horizontal dotted lines* are outliers in the *y*-space and were excluded from the model

regression line because they are along the direction of the linear trend of the data (Varmuza and Filzmoser 2009) and therefore they were not excluded from the model.

The eight outliers in the *y*-space were excluded from the training set together with all the aromatic amines and the robust regression was repeated on the remaining 152 chemicals. The analysis of the plot of the robust standardized residuals versus the robust Mahalanobis distances identified 2-Acetoacetylaminotoluene (chemical 149) as the only outlier in the *y*-space. This chemical is an amide that can be hydrolyzed to an aromatic amine and it was therefore also excluded from all the subsequent analysis. The final training set (151 chemicals) did not contain any aromatic amines or aromatic amides that can be hydrolyzed to aromatic amines.

A linear fit was performed by ordinary least squares (OLS) on this final training set and the linear model assumptions (linearity, normality of residuals and homoscedasticity) were met at a level of significance of 0.05. Since the OLS regression estimator is the most precise estimator among all the unbiased estimators, this linear regression on 151 data points was adopted as a model (Fig. 2). Toxicity parameters estimated after the year 2002 (black dots in Fig. 2) and before the year 2002 (white dots in Fig. 2) confirm to same linear trend without any major differences indicating that the potential effect of a dispersant (potentially used for some tests performed before 2002) does not have an impact on the linear relationship between chronic and acute toxicities. The estimated regression parameters are reported in Table 1. The null hypothesis that the estimated parameters are equal at 0 can be firmly rejected as the *t* values clearly indicate.

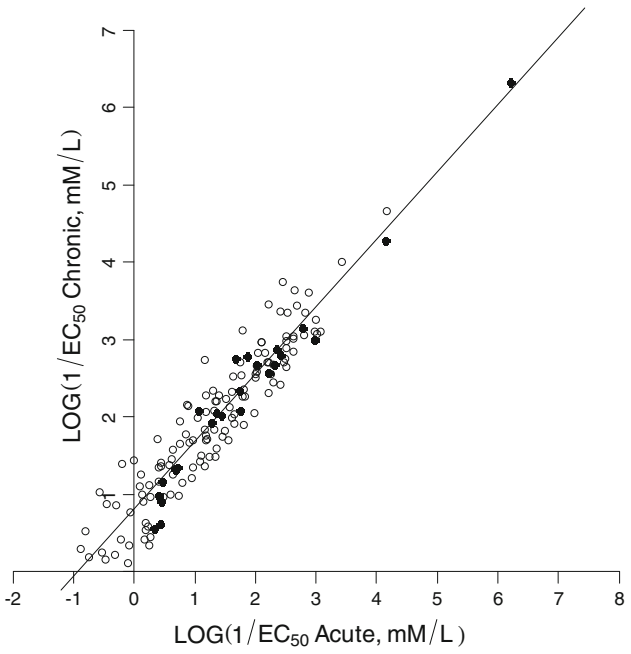

Fig. 2 Linear regression by OLS relating chronic toxicity to acute toxicity for *D. magna*. The *black dots* indicate data obtained after the year 2002

Table 1 Estimated parameters of the linear model ($R^2 = 0.89$) relating chronic toxicity to acute toxicity for *D. magna*

	Estimate	SE	*t* value	Pr (>|*t*|)
Intercept	0.80651	0.04341	18.58	<2e−16
Slope	0.86520	0.02457	35.21	<2e−16

Table 2 Validation parameters from Golbraikh et al. (2003)

R^2	R_0^2	$R_0'^2$	K
0.82	0.74	0.82	0.97

A model is regarded as adequate if $(R^2 - R_0^2)/R^2 < 0.1$, $0.85 < k < 1.15$ and $|R_0^2 - R_0'^2| < 0.3$, where R^2 is the coefficient of determination between the predicted and observed toxicities, R_0^2 is the coefficient of determination between the predicted versus the observed toxicities with the Y intercept set to zero, $R_0'^2$ is the coefficient of determination of the observed versus predicted toxicities with the *Y* intercept set to zero, *K* is the slope of the predicted versus observed toxicities through the origin. The model detailed in this article satisfies all these requirements

The anilines that were present within the original external test set were removed and the predictivity of the model was checked on the remaining 62 chemicals that were not used during the calibration of the linear fit. The squared correlation coefficient between predicted versus observed values for the external test set was equal at 0.82 and the median absolute error of prediction was equal at 0.28 indicating the good predictivity of the model. Eleven chemicals were characterized by an absolute error of prediction >0.5 and among them only two chemicals (CAS RN 3380-34-5 and 101-83-7, Table 1 in Supporting Information) were characterized by an absolute error >1. For six chemicals belonging to this subset, toxicities were under-predicted and for 5 chemicals they were over-predicted. From a structural point of view, these eleven chemicals are very heterogeneous and their high absolute error cannot be explained on the basis of some specific structural features.

The statistical validity of the proposed model is also supported by the analysis recommended by (Golbraikh et al. 2003) with respect to the external validation of QSAR models (Table 2). Moreover, the means of the ratios in the training set between chronic and acute toxicities for the actual experimental data (mean = 6.2; SD = 6.8) and for the values predicted by the model (mean = 4.4; SD = 1.5) are not significantly different as indicated by a Wilcoxon test at the 0.05 significance level.

Table 3 Chemical classes (as defined by the ECOSAR aquatic toxicity profiler) defining the applicability domain of the linear model relating chronic toxicity to acute toxicity for *D. magna*

Acid halides	Carbamate esters	Thiocarbamates, mono
Acid moiety|aliphatic amines	Dinitrobenzenes	Thiols and mercaptans
Acid moiety|neutral organics	Dinitrophenols|phenols	Thiophenes
Acid moiety|phenols	Epoxides, mono|methacrylates	Thioureas
Acid moiety|thiols and mercaptans	Esters	Vinyl/allyl alcohols
Acrylates	Esters, monothiophosphates	Vinyl/allyl aldehydes
Acrylates|aliphatic amines	Esters|esters (phosphate)	Vinyl/allyl amines
Aldehydes (mono)	Esters|peroxy acids	Vinyl/allyl halides
Aldehydes (mono)|phenols	Esters|phenols	Vinyl/allyl ketones
Aldehydes (poly)	Esters|vinyl/allyl esters	Vinyl/allyl nitriles
Aliphatic amines	Halo alcohols	
Aliphatic amines|benzyl amines	Halo ethers	
Aliphatic amines|esters	Hydrazines	
Aliphatic amines|methacrylates	Imides|thioureas	
Amides	Methacrylates	
Amides|hydrazines	Neutral organics	
Amides|limides	Peroxy acids	
Benzyl alcohols	Phenols	
Benzyl halides	Phenols, poly	
Bromoalkanes	Phthalonitriles	

For chemicals belonging to chemical classes other than those reported in the table the model could yield unreliable predictions

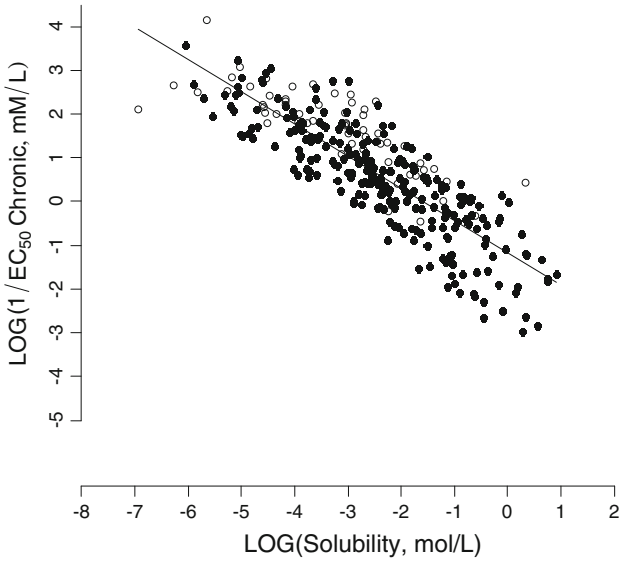

Fig. 3 Linear regression by OLS relating acute toxicity and water solubility for the base surface narcotics of the training set (*white dots*) and the fish acute toxicity data (*black dots*) extracted from Mackay et al. (2009)

The applicability domain (i.e., the range of chemical structures for which the model is applicable) can be defined on the basis of the ECOSAR chemical classes. Indeed, the model could yield unreliable predictions for chemical classes other than those reported in Table 3. Aromatic amides that could be hydrolyzed to the corresponding anilines should also be excluded from the applicability domain of the model.

The range of acute toxicities that the model can reliably cover goes from −0.9 to 4.5 on a logarithmic scale [Log(1/EC_{50}), EC_{50} expressed in mM/L]. Moreover, because of the important role that log *P* and *D* have in modulating ecotoxicological effects, the applicability domain of the model should be limited to chemicals characterized by a log *P* comprised between 0.4 and 5.9 (as predicted by the online software ALOGSP 2.1) and a log *D* (for ionizable chemicals) comprised between −5.3 and 5.7 (as predicted by the online calculator SPARC).

From a mechanistic point of view, it is interesting to observe that when toxicity values for base surface narcotics (according to the OASIS acute aquatic toxicity profiler of the OECD (Q)SAR Toolbox) are plotted against water solubility together with data for baseline fish acute toxicity (Mackay et al. 2009), both data sets follow the same linear trend (Fig. 3). This result is consistent with a direct linear relationship between increasing lethal exposure concentrations and increasing solubility in the exposure media regardless of the organism as Mackay et al. (2009) previously pointed out.

Acknowledgments This work was supported by the National Research Agency (ANR) within the project AMORE (Contract number: 2009 CESA 15 01) and by the French Ministry in charge of Ecology and Sustainable Development, within the framework of Programmes 189 and 190.

References

Adema DM (1978) *Daphnia magna* as a test animal in acute and chronic toxicity tests. Hydrobiologia 59:125–134

Dom N, Knapen D, Benoot D, Nobels I, Blust R (2010) Aquatic multi-species acute toxicity of (chlorinated) anilines: experimental versus predicted data. Chemosphere 81:177–186

EC (2006) Regulation 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, The European parliament and the council of the European union. Available at http://www.eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_396/l_39620061230en00010849.pdf

Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A (2003) Rational selection of training and test sets for the development of validated QSAR models. J Comput Aid Mol Des 7:241–253

Henegar A, Mombelli E, Pandard P, Pery ARR (2011) What can be learnt from an ecotoxicity database in the framework of the REACH regulation? Sci Total Environ 409:489–494

Hilal SH, Karickhoff SW, Carreira LA (1995) A rigorous test for SPARC's chemical reactivity models: estimation of more than 4,300 ionization pK(a)s. Quant Struct Act Relat 14:348–355

Mackay D, Arnot JA, Petkova EP, Wallace KB, Call DJ, Brooke LT, Veith GD (2009) The physicochemical basis of QSARs for baseline toxicity. SAR QSAR Environ Res 20:393–414

MoE (2010) Japan ecotoxicity tests data (version dated March 2010). Available at http://www.env.go.jp/chemi/sesaku/02e.pdf

Neuwoehner J, Zilberman T, Fenner K, Escher BI (2010) QSAR-analysis and mixture toxicity as diagnostic tools: influence of degradation on the toxicity and mode of action of diuron in algae and daphnids. Aquat Toxicol 97:58–67

OECD (1984) *Daphnia* sp. acute immobilisation test and reproduction test, part 1: 24 h EC50 acute immobilisation test available at http://www.oecd.org/dataoecd/17/21/1948249.pdf

OECD (1998) *Daphnia magna* reproduction test available at http://www.oecd.org/dataoecd/17/63/1948277.pdf

OECD (2010) OECD (Q)SAR application Toolbox. Software available from the OECD website http://www.oecd.org/document/54/0,3343,en_2649_34379_42923638_1_1_1_1,00.html

R Development Core Team (2010) R: a language and environment for statistical computing, Vienna, Austria. R Development Core Team, Vienna. Available at http://www.R-project.org

Ramos EU, Vaal MA, Hermens JLM (2002) Interspecies sensitivity in the aquatic toxicity of aromatic amines. Environ Toxicol Pharm 11:149–158

Santojanni A, Gorbi G, Sartore F (1998) Prediction of fecundity in chronic toxicity tests on *Daphnia magna*. Wat Res 32:3146–3156

Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV (2005) Virtual computational chemistry laboratory: design and description. J Comput Aid Mol Des 19:453–463

Varmuza K, Filzmoser P (2009) Introduction to multivariate statistical analysis in chemometrics. CRC Press, Boca Raton