QSAR study for predicting the ecotoxicity of NADES towards *Aliivibrio fischeri*. Exploring the use of mixing rules.

Beatriz Giner[a], Carlos Lafuente[b], David Lapeña[a], Diego Errazquin[a], Laura Lomba[a]

[a]Facultad de Ciencias de la Salud, Universidad San Jorge. Campus Universitario, Autov A23 km 299, 50830. Villanueva de Gállego Zaragoza. Spain

[b]Departamento Química Física, Facultad de Ciencias, Universidad de Zaragoza, 50009, Zaragoza, Spain

*Corresponding author: Beatriz Giner, e-mail: bginer@usj.es, phone: 0034976060100

HIGHLIGHTS:

A QSAR model for predicting the ecotoxicity of NADES has been obtained.

Aliivibrio fischeri is the biomodel selected to obtain the mathematical model.

Several molecular descriptors have been used as independent variables.

Different mixing rules have been explored to construct the independent variables.

ABSTRACT:

(Eco)toxicological information of Natural Deep Eutectic Solvents (NADES) is scarce and thus, Quantitative Structure Activity Relationship (QSAR) models are an important tool to tackle the prediction of toxicity in this case. For that reason, in this manuscript a new QSAR model for predicting the ecotoxicity of NADES towards *Aliivibrio fischeri* biomodel, using mixing rules, is proposed. The main advantage of the method is that the individual components of the mixtures are molecular modeled and then, a mixing rule is used, which simplifies the process. For developing the model, a total of 11 descriptors for each component have been used: acidity constant, partition coefficient, Van der Waals volume, Van der Waals surface area, topological polar surface area, solvent accessible surface area, minimum projection area, maximum projection area, minimum projection radius, maximum projection radius and molecular weight. The final obtained model includes topological polar surface area and acidity constant, mechanistically interpreted as the ability to transport through biological membranes and the negative severe effect of the pH in the toxicity and biological response of *Aliivibrio fischeri* bacteria. The OECD Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships has been followed to develop the mathematical model.

35 KEYWORDS: QSAR, NADES, *Aliivibrio fischeri*, multiple regression model,
36 ecotoxicity.

1. Introduction

Currently, the Natural Deep Eutectic Solvents (NADES) are considered as alternatives to ionic liquids due to their special characteristics. Firstly, NADES are formed from sustainable raw materials of natural origin (carbohydrates, organic acids and esters, natural salts or aminoacids, among others). Furthermore, the synthesis process is easy and cheap (Lapeña et al., 2018; Liu et al., 2018) and their ability of solubilisation is quite high (Lores et al., 2017; Tang et al., 2016). All of this interesting features have made that NADES can be chosen as possible green solvents already available (Dai et al., 2013; Paiva et al., 2014).

The NADES are mixtures that contain, at least, one substance acting as a hydrogen bond donor (HBD) and another that is an acceptor (HBA). The mixture process is determined by the formation of a H-bond network that produces a severe decreasing in the melting temperature, that makes the mixture liquid at room temperature.

The safety of the NADES in terms of toxicity has been studied. However there is still a field to explore and information provided in the bibliography is scarce and disjointed. For instance, the cytotoxicity of several NADES containing choline chloride and sugars or other natural components, has been evaluated in some cell lines such as PC3, A375, HepG2, HT29, MCF-7, OKF6 or H314 (Hayyan et al., 2015; Hayyan et al., 2016; Mbous et al., 2017). The endpoints analysed and the experimental conditions used in each case depend on the cell line and equipment used which makes difficult to extrapolate results.

The toxic information of NADES for the environment is even scarcer, with no data regarding the toxic effect on crustaceans or algae. However, the ecotoxicity of NADES towards the marine bacteria *A. fischeri* has been explored in a greater extent (de Morais et al., 2015; Macario et al., 2018; Ventura et al., 2014), probably due to the ease

of use, availability and cost-effectiveness of the technique (Abbas et al., 2018), being the main advantage the homogeneity of the methodology and uniformity in the explanation of the results.

On the other hand, Quantitative Structure Activity Relationship (QSAR) methodology is a valuable tool to tackle the prediction of toxicity in case experimental data is not accessible. The QSAR methods use mathematical functions for relating the molecular characteristics and properties of the chemicals under study with their biological activities, including toxic effects (Ghaedi, 2015; Levet et al., 2016; Perales et al., 2017b; Yousefinejad and Hemmateenejad, 2015; Zuriaga et al., 2018). Thus, QSAR models could be quite useful for achieving the green and rational design of NADES. However, NADES are complex mixtures and developing QSAR models for multiple-component mixtures is not common. In fact, as far as we know, there is only one attempt of modelling the toxic behaviour of NADES towards the cell line HEK-293 using QSAR methodology (Ahmadi et al., 2018). In this case, how the composition of NADES has been included in the model is not been made explicit. In any way, QSAR studies for multicomponent mixtures are relatively infrequent owing to their complexity: multiple possibilities of combinations and compositions which make the study difficult to approach and manage from the point of view of homogeneity, uniformity and universality of the results obtained.

Thus, the aim of this work is to develop a new, easy and accessible QSAR methodology for evaluating the (eco)toxicity of NADES using mixing rules (Wang et al., 2018) that allow include any composition by the molecular modelling of individual components of the NADES. In this case, we have selected the biomodel *A. fischeri* because there is a sufficient number of data to develop the study. Furthermore, it must be said that we have followed the guidelines of the ENV/JM/MONO(2007)2 OECD

Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models (2014).

2. Material and Methods

The experimental procedure is schematically shown in Figure 1.

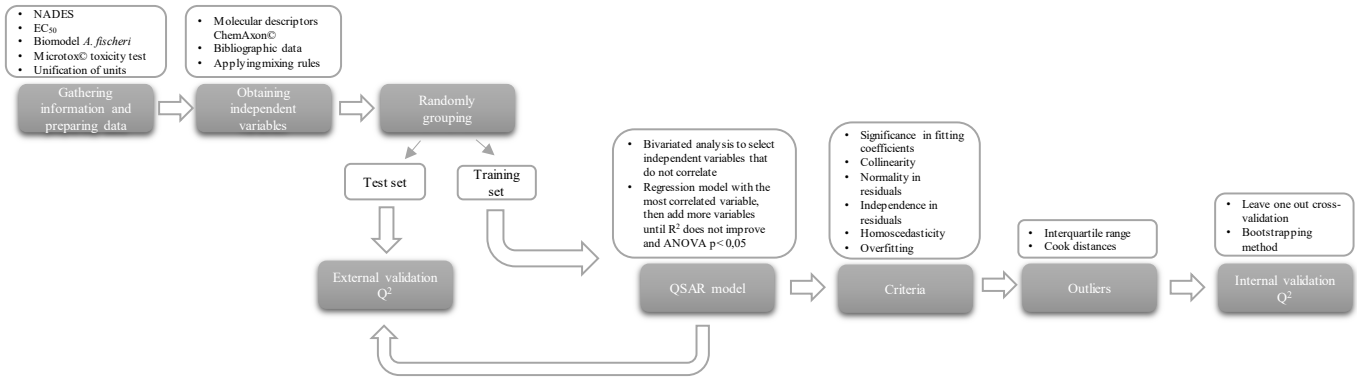


Figure 1: Scheme of the different steps of the experimental procedure.

2.1 Input data gathering, training set and test set

The input ecotoxicological data used has been selected used the following criteria:

- Chemicals are those classified as NADES
- EC_{50} towards *A. fischeri* for the NADES are available. Experimental measuring method is Microtox© toxicity test, exposure time 30 minutes, no acidity correction made. Input EC_{50} units have been transformed in mol/L (de Morais et al., 2015; Macario et al., 2018; Ventura et al., 2014)

A total of 16 different NADES with different combinations of two (A or B) components (choline, acetic acid, lactic acid, citric acid, glycolic acid, ethylene glycol, glycerol, urea, 1,2-propanediol, 1-propanol and cholinium, dihydrogen citrate, butanoate, propanoate, salicylate, bitartrate, dihydrogen phosphate anions) has been selected. The final number of cases gathered are $n = 42$. The mathematical model has been constructed using as dependent variable *log* EC_{50} obtained from the bibliographic data towards *A. fischeri* bacteria. A total of 11 descriptors for each component, $D_{A,B}$, of the mixture have

been used: acidity constant (*pKa*), Van der Waals volume (*VdWV*), Van der Waals surface area (*VdWSA*), topological polar surface area (*TPSA*), solvent accessible surface area (*SASA*), minimum projection area (*MINPA*), maximum projection area (*MAXPA*), minimum projection radius (*MINPR*), maximum projection radius (*MAXPR*), molecular weight (*MW*) and partition coefficient (*log P*). All molecular descriptors have been obtained and/or calculated form the following references: ChemAxon Confluence 5.9.11 Copyright © 2003 - 2013 Atlassian Corporation Pty Ltd., ACD/LABS, Chemspider and (Willighagen et al., 2006).

To build the training set (data used to get the QSAR model) and test set (used for the external validation), the total input data (n = 42) was randomly separated into a training set of 38 compounds (90%) and a test set of 4 compounds (10%).

2.2 Mixing rules

To construct the independent variables, several mixing rules have been proposed using composition, $X_{A,B}$ and descriptors, $D_{A,B}$. The mixing rules (R_1-R_8) proposed in this work are listed in Table 1:

$R_1 = \lvert D_A + D_B \rvert$	$R_5 = \log \lvert D_A + D_B \rvert$
$R_2 = \lvert D_A \cdot X_A + D_B \cdot X_B \rvert$	$R_6 = \log \lvert D_A \cdot X_A + D_B \cdot X_B \rvert$
$R_3 = \lvert (D_A \cdot X_A + D_B \cdot X_B) \rvert^{0.5}$	$R_7 = \log \left(\lvert (D_A \cdot X_A + D_B \cdot X_B) \rvert^{0.5} \right)$
$R_4 = (D_A \cdot D_A \cdot X_A + D_B \cdot D_B \cdot X_B)^{0.5}$	$R_8 = \log \left[(D_A \cdot D_A \cdot X_A + D_B \cdot D_B \cdot X_B)^{0.5} \right]$

Table 1: Proposed mixing rules. $D_{A,B}$ denotes the descriptor used for each component of the NADES (A or B) and $X_{A,B}$ denotes the mole fraction of each NADES.

129 *2.3 QSAR, mathematical correlations and statistics*

130 The literature describes QSAR mathematical models for describing the toxic
131 behaviour of any biomodel and endpoint using different numbers and natures of
132 descriptors. Some of them depend exclusively on one variable, mainly lipophilicity
133 (Mazzatorta et al., 2004), assuming a baseline toxicity mechanism essentially governed
134 by *log P*. However, it is also frequent to use molecular descriptors and/or physicochemical
135 properties in order to improve the models (Chandana and Bijay kumar, 2018; Garcia et
136 al., 2015; Levet et al., 2013; Perales et al., 2017a). In these cases, multiple linear
137 regression models (MLR) are also normally used.

138 In this manuscript, we have chosen this option and the QSAR study has consisted of
139 getting a MLR to explain the dependent variable (*log* EC_{50}) in molar units, using the
140 independent variables described above.

141 Due to the large number of possible independent variables (88 different
142 possibilities, combinations of 8 mixing rules and 11 descriptors), a selection process has
143 been made. To begin, to check the correlations between the dependent variable *log* EC_{50}
144 and the independent variables, a bivariate analysis has been performed. The Pearson
145 correlation coefficient has been used to assess the correlation; a null hypothesis equal to
146 zero was contrasted, and an alpha level of 0.05 was used to reject or accept the null
147 hypothesis. Then, a selection of independent variables for the QSAR model was made on
148 the basis of the following criterion: there is no correlation among independent variables.

149 Afterwards, the training set data has been used for getting a multiple regression
150 mathematical model. R^2 has been used to assess the goodness of the model; $R^2 > 0.8$ for
151 *in vivo* data can be considered as good (Kubinyi, 1993). Firstly, a simple regression model
152 was obtained with the most correlated variable. The rest of the variables were added to
153 this model to construct a multiple regression model if the following criteria were met:

- Coefficients of the new variables were significant, with a p value less than 0.05.
- Independent variables were no collinear. Collinearity has been evaluated using FIV (criteria $FIV < 10$) and tolerance with criteria $tol > 0.2$.
- Residuals were normal. The normality was checked by means of the Shapiro-Wilk method (a significance level higher than 0.05 indicated normality in residuals)
- Residuals were independent. The runs test has been used to check the independence of residuals using the mode and median. The selecting criteria for independence in residuals was that significance of the mode and median was lower than 0.05.
- The model showed homoscedasticity. The homogeneity of variance was checked visually with plots of the standardized residuals versus standardized predicted values.
- There was no overfitting. Adjusted R^2 was similar to the goodness of the model R^2.
- Better possible R^2. The model was finished if the R^2 did not improve and if there were no more significant variables available, with the significance of the ANOVA of the model less than 0.05.

Finally, with the aim of detect influential observations in the final model and to take them into consideration, a search of outliers has been also made. The interquartile range rule using a multiplier of 1.5 and Cook distances (criteria $D_i > 1$) has been used for identifying possible outliers in the dependant variables of the final model.

Internal validation of the model has been tested using two methods. The first one is the leave-one-out (LOO) cross-validation (Hawkins et al., 2003) in which the toxicity value of each chemical has been predicted by the regression obtained mathematical function without using the input data of that chemical. The second one is the

bootstrapping (BOO) method (Wehrens et al., 2000) that consists of the simulation of the model if the cases were selected randomly.

External validation (EXT) was performed using the test set.

The goodness of the model Q^2 has been calculated for LOO, BOO and EXT validation with the following equation:

$$Q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (1)$$

where y_i is the experimental *log* EC_{50} independent variable value for the *i-th* case, $\hat{y}_i$ is the predicted value for the *i-th* case estimated by using the mathematical regression model without using this *i-th* case and the $\bar{y}_i$ (independent variable) is the average experimental *log* EC_{50} value of the training set.

The statistical analysis was carried out with SPSS 21.0 (licence Universidad San Jorge).

The Williams plot was used to visualize the applicability domain of the model (standardized residuals versus leverage) (Netzeva et al., 2005). The leverage value of each chemical from the training set was calculated using the Hat matrix (H) (influence matrix):

$$H = X(X_T X)^{-1} X_T \quad (2)$$

where X and X_T are the independent variable matrix and the transposed matrix.

Two limits were stabilized in the Williams plot; standardized residuals warning value (three standard deviation units, 3δ) and leverage warning value, h^*, calculated as follows:

$$h^* = \frac{3p}{n} \quad (3)$$

where $p = k + 1$, being k the number of independent variables used in the model, and n is the number of chemicals in the training set.

3. Results and discussion

In table I (Supplementary Material), the *n* cases of NADES under study, with their codes, composition and EC_{50} values towards *A. fischeri* are shown. In Table II a-d, the input matrix containing the dependent variable log EC_{50} and all the possible independent variables (combinations of 8 mixing rules and 11 descriptors) for the *n* cases is shown.

The randomly selection of the training and test set was performed. Test set was formed with cases 4, 10, 16 and 34. Rest of cases formed the training set.

The first step is the selection of the independent variables using a bivariate analysis to check if there is correlation with the dependent variable and there is no correlation among them.

Then according to the methodology described earlier, the final MLR model selected is the following:

$$\log EC_{50} = -5{,}596\, R_7\_TPSA + 0{,}145\, R_2\_pKa$$

$$(R^2 = 0{,}858;\ \text{Adjusted } R^2 = 0{,}850;\ \ \text{ANOVA F} = 103,\ \text{p value} < 0.001) \qquad (4)$$

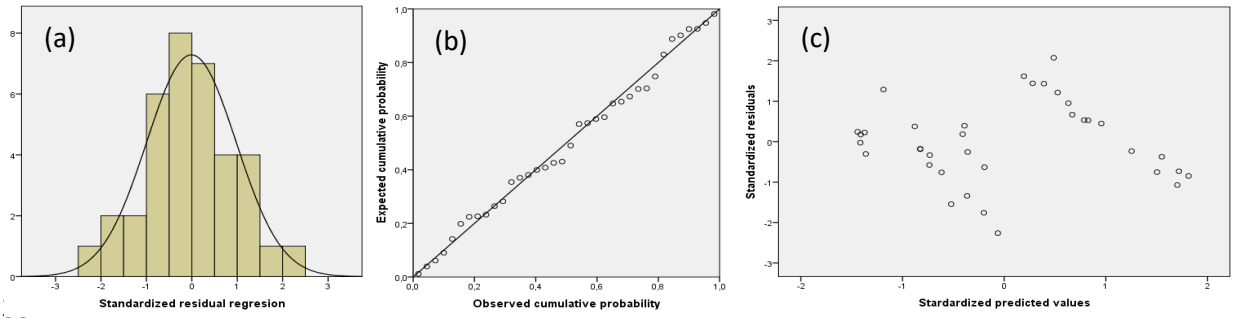
Statistical details of the model can be found in Table 2.

Table 2. Statistical details of the QSAR multiple regression model obtained (Equation 4).

Model QSAR	Coefficients		p
	B	Typical error	
R_7_TPSA	-5,596	0,454	<0,001
R_2_pKa	-0,145	0,019	<0,001
Model BOO	**Coefficients**		**p**
	B	Typical error	
R_7_TPSA	-5,596	0,350	0,001
R_2_pKa	-0,145	0,015	0,001

The parameter used to evaluate the goodness of the model is R^2 = 0,858. In this case, R^2 = 0,850, which can be considered quite enough, taking into account the different structures and composition of the NADES under study. The adjusted R^2 is 0,948, which is similar to the obtained R^2, showing that there is no overfitting. Collinearity evaluation test showed a value *FIV* = 4,957 and *tol* =0,202, which confirmed that the independent

227 variables used in the final model are not correlated. The Shapiro-Wilk method applied to
228 test the normality in residuals showed a significance of 0,980, thus residuals can be
229 considered normal. In Figure 2, the histogram (a) and the normal p-p chart for the
230 residuals (b) are also shown, for illustrating the normality of residuals. The runs test
231 confirmed that the residuals can be considered independent (significance of the mode and
232 median was > 0,001 in the runs test). The homoscedasticity of the model has been checked
233 and showed in Figure 2 (c).
234



2

236 Figure 2: Histogram of frequency of standardized residuals (a), normal p-p chart for
237 checking normality in residuals (b) and homoscedasticity visual checking: standardized
238 residuals vs. standardized predicted values (c).

239 The interquartile range rule and the Cook distances criteria were met in all cases,
240 showing that there are no influential values for the chosen independent variables.
241 Regarding to the nature of the independent variables used to construct the model (*TPSA*
242 and *pKa*), a mechanistic interpretation of the relationship between the toxic effect and the
243 molecular descriptors and properties that describe the behaviour, can be given. As
244 mentioned before, one of the most significant descriptors in the MLR models involving
245 pure and non-ionic species is the partition coefficient, *log P* (Garcia et al., 2015; Mutalib
246 and Ben Ghanem, 2017; Perales et al., 2017a; Zuriaga et al., 2018) assuming a baseline
247 toxicity governed by bioavailability. The property *log P* is usually related with the

capacity of molecules to cross the cellular membranes (Katritzky et al., 2009). Mostly, the adjustable coefficients of the model associated with the descriptor *log P* are negative (Levet et al., 2016) and this is related to chemical bioaccumulation and bioavailability, which are preferential in lipophilic molecules (promotion of transmembrane passage). However, although *log P* can be also found as behavioural descriptor for mixtures and ionic species (Kusumahastuti et al., 2019) it is not quite usual in that cases and some other structural molecular and topological descriptors are commonly used in the QSAR development of the mathematical model (Ma et al., 2015; Roy et al., 2015; Yan et al., 2015). In this study, *log P* does not correlate with *log* EC_{50} in any of the mixing rules described. However, there is a strong relation of the ecotoxic effect towards *A. fischeri* with *TPSA* in the form of R_7 mixing rule. The descriptor *TPSA* (also known as Polar Surface Area defined as the summation of areas of polar atoms in the molecule) has been related with passive molecular cell membrane transport (Ma et al., 2015) and has been previously used as independent variable in QSAR studies for predicting the toxicity (Cassotti et al., 2014; Jiang et al., 2010). Thus, *TPSA* in our model can be considered as a reflection of the ability of molecules to transport through biological membranes. The other independent variable obtained in the model is *pKa* in the form of R_2 mixing rule. This can be explained taking into account the nature of the biomodel used. The optimal pH range for the culturing medium of these bacteria is 6-8,5 (2007). It has been previously demonstrated the severe effect of the pH in the toxicity and biological response of *A. fischeri* bacteria (Berzinskiene and Travkina, 2003). The experimental methodology used for obtaining the raw data utilized for developing the model (de Morais et al., 2015; Macario et al., 2018; Ventura et al., 2014) does not explicit the pH control of the samples used for the exposure of toxicants to the bacteria. Besides, there is a positive correlation between the toxic effect towards *A. fischeri* (raw EC_{50} values obtained from bibliography)

and the proportion of an organic acid in the mixtures. Thus, the variable *pKa* seems to be an adequate descriptor in this case in the. Both mixing rules selected, R_2 and R_7 take into account the composition of the components of the NADES.

The Williams plot (Figure 3) was used to visualize the applicability domain of the model. The warning leverage value limit is $h^* = 0{,}237$. All of the chemicals of the training set and test set are within this h^* limit and three standard deviation units, 3δ. At this point, it should be remarked that ecotoxicity values provided by this QSAR model will take into account the toxic effect of pH.

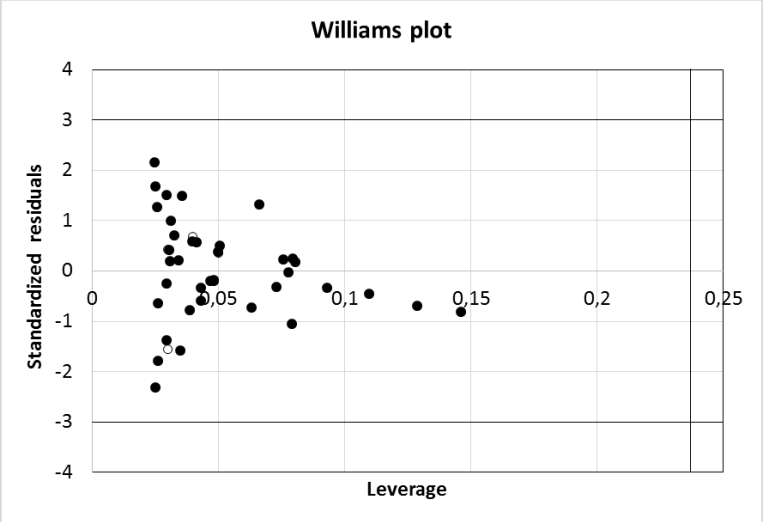


Figure 3: Williams plot. Standardized residuals versus leverage for training ● and test ○ set data. Limits $h^* = 0{,}237$ and 3δ are marked.

Experimental $log\ EC_{50}$ values were compared to those obtained with the proposed model in order to visualize the prediction ability of the model. In Figure 4 (a), plots of the experimental values vs. those calculated with the obtained model (Equation 4) are shown and in Figure 5, absolute residual values are also given for each mixture.

(a)

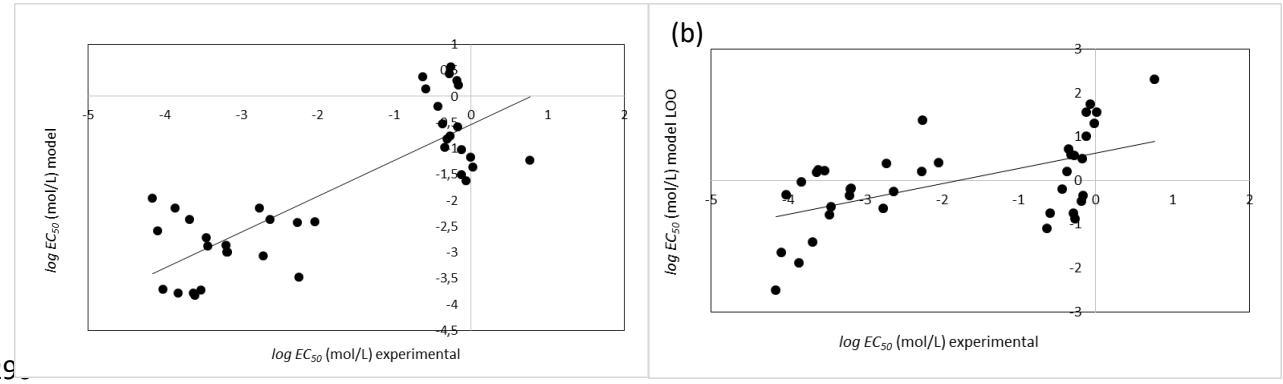


Figure 4: (a) Plots of experimental data vs. predicted values of *Log EC_{50}* as calculated through the obtained model (Equation 4). Solid line: linear regression between experimental data and predicted values, for comparing. (b) Plots of the experimental data vs. those predicted by the LOO model. Solid line: linear regression between experimental data and predicted values, for comparing.

There is no a clear trend with regard to the NADES whose toxic behaviour can be better described with the model and predictions depend not only on the chemicals involved but also on the composition of each component. The case 15, ChClCA 11 is the mixture that is better predicted, with the smallest deviation (absolute value of residual < 0,04) while the case 30, ChClPA 14 is the NADES with highest deviation (absolute residual value > 2). It is worth mentioning that mixtures containing propionic acid show high residual values in general. Furthermore, the deviations increase with the composition of propionic acid.

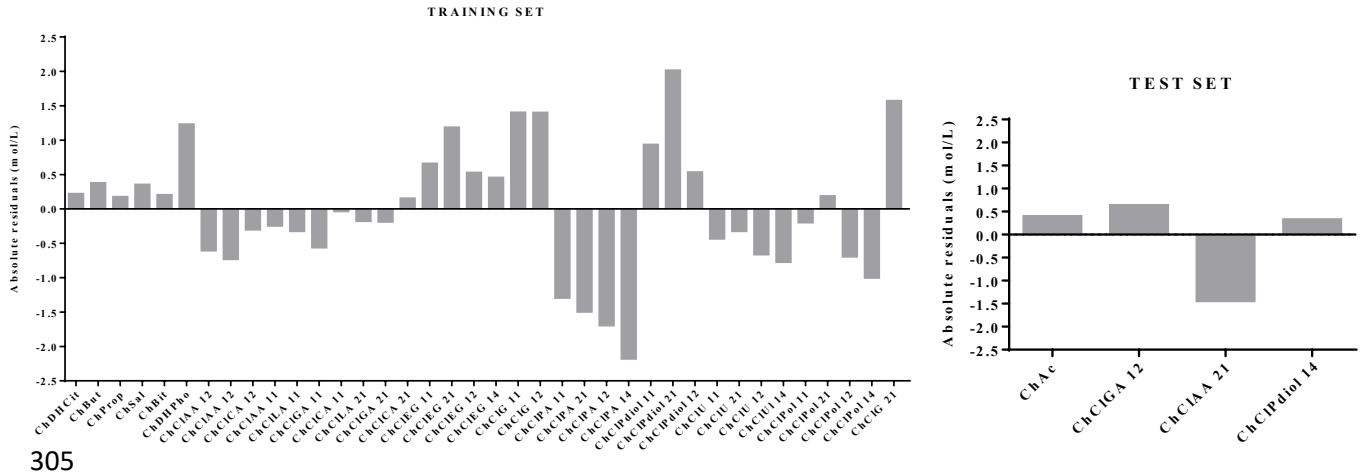


Figure 5: Absolute residuals calculated for the training set and test set.

Internal validations have been carried out by means of the LOO and BOO methods. LOO results are graphically represented in Figure 4 (b), being the calculated Q^2_{LOO}= 0,601. This value, although low, is still reasonable for a biological response and the difference between Q^2_{LOO} and R^2 *is* lower than 0.3 units.

BOO internal validation results are shown in Table 3, with the same adjustable coefficient values for the multiple linear regression as in our model and Q^2_{BOO} = 0,601.

Finally, test set (cases 4, 10, 16 and 34) was used for external validation purposes. The predicted values were compared with the experimental values with Q^2_{EXT} = 0,621, even better than for the training set. As mentioned before, all of the test set values are within the applicability domain. Obtained residuals (Figure 5) for the test set are quite low except for the case 16, containing acetic acid (ChClAA21).

4. Conclusions

In this manuscript a new, easy and available way of obtaining QSAR mathematical models for predicting the (eco)toxicicty of complex mixtures as NADES has been obtained. The method is based on the use of several mixing rules that contains the information of the characteristics and composition of the individual components. For

developing the QSAR study, only the individual components of the mixtures have been modelled, which a clear advantage is since simplifies the model and reduces time of modelling.

In this case, we have obtained a model for the prediction of the ecotoxic behaviour of NADES towards the marine bacteria *A. fischeri*. The final obtained model comprised two independent variables that encoded information about bioavailability, R_7_TPSA and toxic effect of pH, R_2_pKa and the composition of each of the component in the NADES. These two independent variables are enough for describing the observed behaviour and there is no overfitting, one of the problems that many previous studies suffer from. The obtained model meets all the criteria proposed by the OECD guidance document for QSAR methodology. The applicability domain is also provided and thus, the model proposed in this work can be used in a future taking into the limitations given. Finally, the predicted EC_{50} values will take into account the toxic effect of pH.

Acknowledgements

PLATON research group acknowledges financial support from Gobierno de Aragón and Fondo Social Europeo "Construyendo Europa desde Aragón". Additionally, we thank the EEE53 SL. business groups, Pinares de Venecia División Energética and Brial (ENATICA), for their support. Both business groups are committed to sustainable developments through environmental respect.

References

Water quality - Determination of the inhibitory effect of water samples on the light emission of Vibrio fischeri (Luminescent bacteria test). UNE-EN-ISO 11348-3. In: UNE-EN-ISO, (Ed.), 11348-3, 2007.

Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. Vol. 69. OECD, OECD Series on Testing and Assessment, No. 69, OECD Publishing, Paris., 2014.

Abbas, M., et al., 2018. Vibrio fischeri bioluminescence inhibition assay for ecotoxicity assessment: A review. Science of the Total Environment. 626**,** 1295-1309.

Ahmadi, R., et al., 2018. Assessment of cytotoxicity of choline chloride-based natural deep eutectic solvents against human HEK-293 cells: A QSAR analysis. Chemosphere. 209**,** 831-838.

Berzinskiene, J., Travkina, T., 2003. pH-dependence of the effects of substituted phenols on Vibrio fischeri cells. Fresenius Environmental Bulletin. 12**,** 914-918.

Cassotti, M., et al., 2014. Prediction of Acute Aquatic Toxicity Toward Daphnia magna by using the GA-kNN Method. Atla-Alternatives to Laboratory Animals. 42**,** 31-41.

Chandana, A., Bijay kumar, M., 2018. Quantitative Structure-Activity Relationships of Aquatic Narcosis: A Review. Current Computer-Aided Drug Design: CAD. 14**,** 7-28.

Dai, Y. T., et al., 2013. Natural deep eutectic solvents as new potential media for green technology. Analytica Chimica Acta. 766**,** 61-68.

de Morais, P., et al., 2015. Ecotoxicity of Cholinium-Based Deep Eutectic Solvents. Acs Sustainable Chemistry & Engineering. 3**,** 3398-3404.

Garcia, J. I., et al., 2015. Ecotoxicity studies of glycerol ethers in Vibrio fischeri: checking the environmental impact of glycerol-derived solvents. Green Chemistry. 17**,** 4326-4333.

Ghaedi, A., 2015. Predicting the cytotoxicity of ionic liquids using QSAR model based on SMILES optimal descriptors. Journal of Molecular Liquids. 208**,** 269-279.

374 Hawkins, D. M., et al., 2003. Assessing model fit by cross-validation. Journal of Chemical
375 Information and Computer Sciences. 43**,** 579-586.
376 Hayyan, M., et al., 2015. In Vitro and In Vivo Toxicity Profiling of Ammonium-Based
377 Deep Eutectic Solvents. Plos One. 10.
378 Hayyan, M., et al., 2016. Natural deep eutectic solvents: cytotoxic profile. Springerplus.
379 5.
380 Jiang, L., et al., 2010. Toxicity Prediction of Antibiotics on Luminescent Bacteria,
381 Photobacterium phosphoreum, Based on Their Quantitative Structure-Activity
382 Relationship Models. Bulletin of Environmental Contamination and Toxicology.
383 85**,** 550-555.
384 Katritzky, A. R., et al., 2009. Quantitative Structure-Activity Relationship (QSAR)
385 Modeling of EC50 of Aquatic Toxicities for Daphnia magna. Journal of
386 Toxicology and Environmental Health-Part a-Current Issues. 72**,** 1181-1190.
387 Kubinyi, H., 1993. QSAR Hansch analysis and related approaches. VCH, Weinheim;
388 New York; Basel; Cambridge; Tokyo.
389 Kusumahastuti, D. K. A., et al., 2019. Toxicity profiling of 24 L-phenylalanine derived
390 ionic liquids based on pyridinium, imidazolium and cholinium cations and
391 varying alkyl chains using rapid screening Vibrio fischeri bioassay.
392 Ecotoxicology and Environmental Safety. 172**,** 556-565.
393 Lapeña, D., et al., 2018. The NADES Glyceline as a Potential Green Solvent: A
394 Comprehensive Study of its Thermophysical Properties and Effect of Water
395 Inclusion. Journal of Chemical Thermodynamics. 128, 164-172.
396 Levet, A., et al., 2013. Quantitative structure-activity relationship to predict acute fish
397 toxicity of organic solvents. Chemosphere. 93**,** 1094-1103.

Levet, A., et al., 2016. Acute aquatic toxicity of organic solvents modeled by QSARs. Journal of Molecular Modeling. 22.

Liu, Y., et al., 2018. Natural Deep Eutectic Solvents: Properties, Applications, and Perspectives. Journal of Natural Products. 81**,** 679-690.

Lores, H., et al., 2017. Natural deep eutectic solvents in combination with ultrasonic energy as a green approach for solubilisation of proteins: application to gluten determination by immunoassay. Talanta. 162**,** 453-459.

Ma, S. Y., et al., 2015. Predicting the ecotoxicity of ionic liquids towards Vibrio fischeri using genetic function approximation and least squares support vector machine. Journal of Hazardous Materials. 283**,** 591-598.

Macario, I. P. E., et al., 2018. The antagonist and synergist potential of cholinium-based deep eutectic solvents. Ecotoxicology and Environmental Safety. 165**,** 597-602.

Mazzatorta, P., et al., 2004. QSAR in Ecotoxicity: An Overview of Modern Classification Techniques. J. Chem. Inf. Comput. Sci. Journal of Chemical Information and Computer Sciences. 44**,** 105-112.

Mbous, Y. P., et al., 2017. Unraveling the cytotoxicity and metabolic pathways of binary natural deep eutectic solvent systems. Scientific Reports. 7.

Mutalib, M. I. A., Ben Ghanem, O., 2017. Ecotoxicity of Ionic Liquids Towards Vibrio fischeri: Experimental and QSAR Studies. Progress and Developments in Ionic Liquids. 429-449.

Netzeva, T. I., et al., 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships.

Paiva, A., et al., 2014. Natural Deep Eutectic Solvents - Solvents for the 21st Century. Acs Sustainable Chemistry & Engineering. 2**,** 1063-1071.

Perales, E., et al., 2017a. Ecotoxicity and QSAR studies of glycerol ethers in Daphnia magna. Chemosphere. 183, 277-285.

Perales, E., et al., 2017b. Ecotoxicity and QSAR studies of glycerol ethers in Daphnia magna. Chemosphere. 183, 277-285.

Roy, K., et al., 2015. Predictive QSAR modelling of algal toxicity of ionic liquids and its interspecies correlation with Daphnia toxicity. Environmental Science and Pollution Research. 22, 6634-6641.

Tang, N., et al., 2016. Solubilities of Three Flavonoids in Different Natural Deep Eutectic Solvents at T = (288.15 to 328.15) K. Journal of Chemical and Engineering Data. 61, 4203-4208.

Ventura, S. P. M., et al., 2014. Ecotoxicity analysis of cholinium-based ionic liquids to Vibrio fischeri marine bacteria. Ecotoxicology and Environmental Safety. 102, 48-54.

Wang, Y. et al., 2018. Distributive structure-properties relationship for flash point of multiple components mixture. Fluid Phase Equilibria. 474, 1-5.

Wehrens, R., et al., 2000. The bootstrap: a tutorial. Chemometrics and Intelligent Laboratory Systems. 54, 35-52.

Willighagen, E. L., et al., 2006. On the use of H-1 and C-13 1D NMR spectra as QSPR descriptors. Journal of Chemical Information and Modeling. 46, 487-494.

Yan, F., et al., 2015. Topological study on the toxicity of ionic liquids on Vibrio fischeri by the quantitative structure-activity relationship method. Journal of Hazardous Materials. 286, 410-415.

Yousefinejad, S., Hemmateenejad, B., 2015. Chemometrics tools in QSAR/QSPR studies: A historical perspective. Chemometrics and Intelligent Laboratory Systems. 149, 177-204.

Zuriaga, E., et al., 2018. Exploring the Usefulness of Key Green Physicochemical Properties: Quantitative Structure-Activity Relationship for Solvents From Biomass. Environmental Toxicology and Chemistry. 37**,** 1014-1023.