MODELING AQUATIC TOXICITY THROUGH CHROMATOGRAPHIC SYSTEMS

Alejandro Fernández-Pumarega[1], Susana Amézqueta[1], Sandra Farré[1], Laura Muñoz-Pascual[1], Michael H. Abraham[2], Elisabet Fuguet[1,3], Martí Rosés[1*]

[1] Departament de Química Analítica, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1-11, 08028, Barcelona, Spain.

[2] Department of Chemistry, University College London, 20 Gordon Steet, London WC1H 0AJ, UK.

[3] Serra Húnter Programme. Generalitat de Catalunya. Spain

* Correspondence:

Prof. Martí Rosés. E-mail: marti.roses@ub.edu

Departament de Química Analítica, Facultat de Química, Universitat de Barcelona

c/ Martí i Franquès 1-11, 08028, Barcelona, Spain

Phone: (+34) 93 403 92 75

Fax: (+34) 93 402 12 33

Running title. Modeling aquatic toxicity through chromatographic systems.

ABSTRACT

Environmental risk assessment requires information about the toxicity of the growing number of chemical products coming from different origins that can contaminate water and become toxicants to aquatic species or other living beings via the trophic chain. Direct toxicity measurements using sensitive aquatic species can be carried out but they may become expensive and ethically questionable. Literature refers to the use of chromatographic measurements that correlate to the toxic effect of a compound over a specific aquatic species as an alternative to get toxicity information.

In this work, we have studied the similarity in the response of the toxicity to different species and we have selected eight representative aquatic species (including tadpoles, fish, water fleas, protozoan and bacteria) with known non-specific toxicity to chemical substances. Next, we have selected four chromatographic systems offering good perspectives for surrogation of the eight selected aquatic systems, and thus prediction of toxicity from the chromatographic measurement. Then toxicity has been correlated to the chromatographic retention factor. Satisfactory correlation results have been obtained to emulate toxicity in five of the selected aquatic species through some of the chromatographic systems. Other aquatic species with similar characteristics to these five representative ones could also be emulated by using the same chromatographic systems. The final aim of this study is to model chemical products toxicity to aquatic species by means of chromatographic systems to reduce *in vivo* testing.

KEYWORDS

Aquatic toxicity, solvation parameter model, micellar electrokinetic chromatography, high-performance liquid chromatography, surrogate systems, protozoan, fish, bacteria, tadpoles, water fleas

INTRODUCTION

Aquatic environments receive directly and indirectly chemical substances that may result in toxicity to their inhabitants. There are several protocols and analytical methods to determine the toxicity of these chemicals to aquatic species[1]. While *in vivo* experiments provide reliable measurements, they often require expensive, long, and complex procedures. The US Environmental Protection Agency and the European Chemicals Agency promote the use of alternative methods to avoid unnecessary animal testing [2,3]. Models based on *in silico* methods or other predictive models based on physicochemical properties measurements can be used as alternatives [4].

A popular quantitative structure-property relationships (QSPR) model to estimate aquatic toxicity and other biochemical properties is the solvation parameter model (SPM) proposed by Abraham [5]. The following equation that includes five different molecular descriptors is used to model the solvation that a neutral solute undergoes in a biphasic system.

$$\log SP = c + eE + sS + aA + bB + vV \qquad \text{[Eq. 1]}$$

Here, SP is the dependent solute property in a given partitioning system, i.e. equilibrium constant or some other free energy related property such as a lethal dose. The E, S, A, B and V independent variables are the solute descriptors proposed by Abraham. E represents the excess molar refraction, S is the solute dipolarity/polarizability, A and B are the solute's effective hydrogen-bond acidity and hydrogen-bond basicity, respectively, and V is McGowan's solute volume. The coefficients of the equation are characteristic of the biphasic system and reflect the difference of the two phases in properties complementary to the corresponding solute property. For any system, the coefficients of this equation can be obtained by multiple linear regression analysis between the log SP values acquired for an appropriate group of solutes and their descriptor values. Equations based on the SPM have been used to characterize many biological systems that depend on the solutes partition into two phases, *i.e.* an aqueous solution and a biological membrane. Literature proposes equations based on this model to estimate the toxicity of chemical substances to tadpoles, fish, water fleas, protozoan and

bacteria [6–11]. Furthermore, more than one hundred physicochemical systems, mainly based on liquid-liquid partition and chromatographic and electrophoretic partition systems, have been characterized using this model[12–18].

Characterizing biological and physicochemical systems using the same model (the SPM in this work) makes them comparable, since similar partitioning systems will have similar coefficients. To compare similarity between toxicity systems and physicochemical systems (mainly chromatographic and electrophoretic) the *d* distance parameter on the SPM coefficients can be calculated. Dendrogram and Principal Components Analysis (PCA) plots lead to a visual representation of the systems closeness. In addition, the precision of the correlation between the toxicity parameter and the physicochemical parameter can also be estimated. This precision depends on the errors of the biological and physicochemical models and the systems dissimilarity [16,19–21]. Those physicochemical systems closer to the biological one (with smallest *d* or closest in the dendrogram and principal components space plots) and with highest estimated correlation precision will probably best emulate the toxicity parameter. To this end, the physicochemical property (mainly the retention factor in chromatographic and electrophoretic systems) is determined and a correlation with the biological property is carried out for a series of representative compounds. If a good correlation is established between the properties of these two different systems, the biological property of a new chemical compound can be predicted by measuring the corresponding physicochemical property. The main advantage of this approach over QSPR studies is that it is not necessary to know the molecular descriptor values of the new compound such as in the SPM model. Furthermore, the use of chromatographic and electrophoretic measurements for prediction of biological properties is of main interest due the high level of automatization, speed of analysis, low cost, and high reproducibility of these techniques.

Previous works on aquatic toxicology have shown that a micellar electrokinetic chromatography (MEKC) system based on sodium taurocholate (STC) micelles and chromatographic measurements using an immobilized artificial membrane (IAM) column are able to predict the neutral organic

substances toxicity to Fathead minnow (FM) fish [22] and *Rana* tadpoles (RT) [23]. The aim of this work is to check if one of these chromatographic systems or others based on HPLC (high performance liquid chromatography) or MEKC could be used to model in a more general way the toxicity to aquatic species. In this sense, some representative biological systems have been selected and the ability of some representative and promising physicochemical systems to emulate the toxicity of chemical compounds to these aquatic species have been evaluated. Those biological systems ruled by similar toxicity mechanisms could be putatively emulated using the same physicochemical system.

EXPERIMENTAL SECTION

Equipment

HPLC measurements were done using a 10A series chromatograph from Shimadzu (Kyoto, Japan) equipped with a quaternary pump and a diode array detector and fitted with either a Symmetry C18 column (15 cm × 4.6 mm i.d., 5 μm particle size) (Waters, Milford, MA, US) preceded by the corresponding guard cartridge (1 cm), or an IAM.DD 2 immobilized artificial membrane column (10 cm × 4.6 mm i.d., 12 μm particle size) (Regis Technologies, Morton Grove, IL, US).

MEKC measurements were done using the CE capillary electrophoresis system from Agilent Technologies (Santa Clara, CA, US) equipped with a diode array detector. The fused-silica capillary (40 cm effective length, 50 μm i.d.) was obtained from Composite Metal Services Ltd (Shipley, UK).

Reagents

Methanol (HPLC-grade), hydrochloric acid (25 % in water), sodium hydroxide (>99%), sodium dihydrogenphosphate monohydrate (>99%), disodium hydrogenphosphate (>99%) and sodium dodecyl sulfate (SDS, >99%) were from Merck (Darmstadt, Germany). Acetonitrile (HPLC grade) was from VWR International (West Chester, Pennsylvania, US). Taurocholic acid sodium salt monohydrate (STC, 98%) was from Acros Organics (Geel, Belgium) and Brij 35 was from Scharlab

123 (Sentmenat, Spain). Tetradecyltrimethylammonium bromide (TTAB, >98%) and dodecanophenone
124 (98%) were from Sigma-Aldrich (St. Louis, MO, US). Water was purified by a Milli-Q plus system
125 from Millipore (Bedford, MA, US), with a resistivity of 18.2 MΩ cm.
126 Tested substances were reagent grade or better and obtained from several manufacturers (Merck,
127 Sigma-Aldrich, Carlo Erba (Milano, Italy), Baker (Center Valley, PA, US), Panreac (Castellar del
128 Vallès, Spain), Thermo Fisher Scientific (Waltham, MA, US), and Scharlab).
129
130 Analysis by HPLC
131 Tested substances were solved and analyzed as described elsewhere [23]. The detection wavelength was
132 214 nm.
133 The HPLC retention factor (*k*), was calculated according to Eq. 2.
134 $$k = \frac{t_R - t_0}{t_0} \qquad \text{[Eq. 2]}$$
135 where t_R corresponds to the solute retention time and t_0 is the column hold-up time determined by an
136 aqueous potassium bromide solution.
137
138 Analysis by MEKC
139 The target compounds were analyzed using three different pseudostationary phases: a 50 mM solution
140 of taurocholic acid sodium salt monohydrate in 20 mM phosphate aqueous buffer adjusted to pH 7.0;
141 a 20 mM solution of TTAB in 20 mM phosphate aqueous buffer adjusted to pH 7.0; and a mixture of
142 surfactants that consists of 50 mM SDS and 10 mM Brij 35 in 20 mM phosphate aqueous buffer
143 adjusted to pH 7.0.
144 Tested substances were solved and analyzed as described elsewhere [23]. In the case of TTAB, solutes
145 and dodecanophenone (micellar marker) were prepared at 200 mg L^{-1}.
146 The MEKC retention factor (*k*), was calculated according to Eq.3.
147 $$k = \frac{t_m - t_{eof}}{\left(1 - \frac{t_m}{t_{mc}}\right) t_{eof}} \qquad \text{[Eq. 3]}$$

where t_m is the solute migration time, t_{eof} corresponds to the migration time of methanol or acetonitrile (electroosmotic flow markers), and t_{mc} is the migration time of dodecanophenone (micelle marker).

Comparison of biological and physicochemical systems

The SPM normalized coefficients[22] of both kind of systems have been used as input data to calculate the *d* distance and/or plot of the corresponding PCA and dendrogram. These approaches are based on simple and fast calculations and allow handling with a high number of data at once. They provide information about the systems similarity and are very adequate to choose representative biological systems and to do a first selection of the physicochemical systems that can better emulate the toxicity ones [16,20]. Another strategy to select the more promising physicochemical systems is to estimate the precision of the correlation between the toxicity parameter and the physicochemical parameter [21]. This kind of estimation is more laborious, so it has been performed only for those systems preselected with the previous approaches. Finally, after a general evaluation of all comparison approaches and other technical considerations, those physicochemical systems with similar characteristics to the representative biological ones have been selected for ongoing experimental tests.

d distance parameter

The similarity between two systems both characterized by means of the SPM can be measured through the *d* distance parameter [16]. Considering the coefficients of any system as a vector in a five-dimensional space, the *d* parameter measures the distance between the normalized unitary vectors of a pair of systems. Thus, the *d* distance provides a measure of the similarity between the two considered systems: the smaller *d* is, the closer the two systems are. In previous works, we assumed that distances below 0.25 indicate that the two compared systems are quite similar [22,23].

Dendrogram plot

The dendrogram is a diagram plotted using a hierarchical clustering algorithm that shows the distances between pairs of sequentially merged classes. In this work, the distances between each pair of classes (toxicity and physicochemical systems) have been calculated through the normalized coefficients of the SPM equations. Clustering has been performed using the *d* distance parameter (straight-line distance) and the weighted-linkage method (the distance between two groups is defined as the weighted average distance). Those systems located nearer in the dendrogram plot will have more similar chemical characteristics.

Principal components analysis

PCA is a chemometric tool used to transform the input data in a multivariate space (normalized SPM equation coefficients) to a new multivariate space (principal components (PCs) space) whose axes are uncorrelated and rotated with respect to the original space. The number of PCs is equal to the number of original variables and the first PCs are those that more explain the system variance. The main PCs plot (PC2 *vs* PC1 scores centered plot) using the normalized coefficients of the SPM equations of the different toxicity and physicochemical systems as the input data distributes the different systems in the new chemical space, so that systems with similar characteristics are close in the scores plot. In this work, the PCA analysis helps to visualize the physicochemical space. However, in some cases, the simplification leads to a loss of information.

Estimation of the precision of the correlation between a biological and a physicochemical system

In this work the physicochemical systems used are based on chromatography, so in order to estimate the precision of biological-chromatographic correlations (Eq. 4), the approach described elsewhere [22] has been used.

$$\log SP_{bio} = q + p \log SP_{chrom} \qquad \text{[Eq. 4]}$$

Here, SP_{bio} is the solute biological property, SP_{chrom} is the solute chromatographic property (in this case, the chromatographic or electrophoretic retention factor), and q and p are the ordinate and slope of the correlation, respectively.

In short, the correlation precision (SD_{corr}^2) can be considered as the sum of three different contributions to the variance of the correlation: the biological data precision ($\sigma_{bio}^2 \approx SD_{bio}^2$), the chromatographic data precision ($\sigma_{chrom}^2 \approx p \times SD_{chrom}^2$) and the error due to the dissimilarity between the correlated systems (SD_d^2). SD_{bio} and SD_{chrom} values are obtained from the respective standard deviations of the SPM characterizations. In order to know p and also SD_d^2 the biological property and the chromatographic property are calculated through their SPM equations and solutes' descriptors. In this way, SD_{bio} and SD_{chrom} are zero. The slope of the correlation of these calculated values provides p, and the SD of the correlation can be entirely attributed to the dissimilarity between both systems.

Data analysis

PCA and dendrogram plots were performed with Matlab package from MathWorks (Natick, MA, USA). Excel 2010 from Microsoft (Redmond, WA, US) was used for data calculations and multiple linear regression analyses.

Substances' pK_a values and Abraham descriptors were obtained from Percepta software version 2014 from ACD/Labs (Toronto, Canada).

RESULTS AND DISCUSSION

Similarity of biological systems

Twenty-one biological systems related to toxicity to different aquatic species have been considered in the present study. They are tadpoles (RT: *Rana* tadpoles), fish (FM: fathead minnow (*Pimephales*

223 *promelas*); GP: guppy (*Poecilia reticulata*); BG: bluegill (*Lepomis macrochirus*); GO: golden orfe
224 (*Leuciscus idus melanotus*); GF: goldfish (*Carassius auratus*); MK48: medaka high-eyes 48 h;
225 MK96: medaka high-eyes 96 h (*Oryzias latipes*)), water fleas (DM24: *Daphnia magna* 24 h; DM48:
226 *Daphnia magna* 48 h; CD: *Ceriodaphnia dubia*; DP: *Daphnia pulex*), protozoan (TP: *Tetrahymena*
227 *pyriformis*; SA: *Spirostomum ambiguum*; ES: *Entosiphon sulcantum*; UP: *Uronema parduczi*; CP:
228 *Chilomonas paramecium*) and bacteria (PP: *Pseudomonas putida*; PG: *Porphyromonas gingivalis*;
229 SR: *Selenomonas artemidis*; SS: *Streptococcus sobrinus*). All of them have been characterized
230 through the SPM and results are presented in Table S-1 of the supplementary material.
231 In general, hydrogen-bond basicity and solute volume are the major factors that influence the
232 compounds toxic action. Solutes with high hydrogen-bond basicity and low volume will result less
233 toxic to aquatic species.
234 In order to show the similarity of these toxicity systems, the normalized SPM coefficients of the
235 systems have been analyzed according to a dendrogram of *d* distances and a PCA of the normalized
236 coefficients. These plots are represented in Figure 1 and they lead to the same conclusions about
237 similarity. The dendrogram shows that at the *d*-levels of 0.35-0.5, all systems are clustered together
238 except UP and GF. These two systems are radically different from the other ones (also shown in the
239 PCA plot) probably because of the big differences on e and b coefficients from the ones of the other
240 systems, and also on s coefficient for GF. Therefore, they will not be further considered for similarity.
241 At the 0.3 *d*-level, the cluster divides in two different clusters of 10 and 9 systems. The PCA plot
242 shows that the two clusters are differentiated by the PC1 value. The first cluster contains systems with
243 negative PC1 value and includes all bacteria and tadpole, and several fish and protozoan systems. The
244 second cluster is composed of all water fleas and several fish and protozoan systems, which have
245 positive PC1 values.
246 At the 0.25 *d*-level, CD separates from the cluster of positive PC1, whereas the cluster of negative
247 PC1 divides into three different clusters. One cluster is formed by GO, ES, MK96 and CP which have

248 negative PC2 values. SR, MK48, SS and PG form another cluster with positive PC2. The third cluster
249 is formed with RT and PP, which have positive and negative PC2, respectively, but very similar and
250 close to 0 negative PC1. However, RT and PP separates according to PC2 at *d*-level of 0.23. Notice,
251 that the main clustering factor is PC1 which explains 62% of the variance, whereas PC2, explaining
252 only 22% of variance, has a minor effect on clustering.
253 In previous works [22,23], a *d* distance of 0.25 or less was considered as adequate for surrogating
254 biological systems by chromatographic ones. Following this criteria, we can consider similar all
255 systems in the same cluster at 0.25 *d*-level and select one or several representative systems of each
256 cluster, in general the ones that have larger number of chemical compounds with known toxicity data.
257 Thus, GO fish and CP protozoa have been selected as representatives of the cluster including GO and
258 MK96 fish, and CP and ES protozoan. SS bacteria is representative of MK48 fish, and SR, SS and
259 PG bacteria cluster. RT tadpole and PP bacteria have been taken both as representatives of their
260 cluster. FM fish, TP protozoa and DM24 water flea are representative of last cluster composed of
261 FM, GP and BG fish, TP and SA protozoa, and DP, DM48, and DM24 water fleas. In all, we have
262 selected systems belonging to the five different species (tadpoles, fish, water fleas, protozoan and
263 bacteria), in each of the four main clusters.
264
265 Selection of physicochemical systems to emulate toxicity of neutral organic compounds
266
267 The SPM has been applied to nearly one hundred physicochemical systems including solvent
268 partition, HPLC (high-performance liquid chromatography), MLC (micellar liquid chromatography),
269 MELC (microemulsion liquid chromatography), MEKC (micellar electrokinetic chromatography),
270 MEEKC (microemulsion electrokinetic chromatography), LEKC (liposome electrokinetic
271 chromatography) and poly-EKC (polymeric electrokinetic chromatography) systems [12–17]. Similar to

the case of the biological systems, the main factors that drive solute partition are the magnitude of the coefficients v and b.

Among these physicochemical systems, eleven (coefficients detailed in Table S-2 of the supplementary material) commonly used to surrogate biological systems and that showed *d* values below 0.25 to at least one of the selected toxicity biological systems were selected (Table 1). The selection comprises the octanol-water partition system (O/W), two liquid chromatography systems, five MEKC systems, one microemulsion electrokinetic chromatography MEEKC system, one LEKC system and one EKC system based on a polymeric surfactant.

In addition, the variance of the final correlation ($SD_{corr\ cal}^2$) between aquatic toxicity data and the physicochemical property data (either the partition coefficient or the retention factor) of the selected systems was estimated. The detailed calculations are given in Table S-3 of the supplementary material and final results are shown in Table 2.

According to the results of Tables 1 and 2, the cluster of GO and CP (also including MK96 and ES) is best surrogated by TTAB, followed by SLN (sodium *N*-lauroylsarcosinate), STC and SDS-Brij 35 systems, which show the shortest predicted *d* distances and correlation variances to them. The TTAB is also clearly the system showing best *d* distance and variance to SS, and thus expected to emulate well MK48, PG and SR of the same cluster. This is not surprising because these two clusters are very close on the dendrogram of Figure 1. The toxicity cluster of RT and PP is well surrogated by STC, SDS-Brij 35 and SLN. Finally, SLN, SDS-Brij 35 and DPPG-DPPC (dipalmitoylphosphatidyl glycerol - dipalmitoylphosphatidyl choline) are the best systems to surrogate the cluster with FM, DM24 and TP (in addition to DM48, GP, DP, SA and BG).

From the above reasoning, it can be deduced that among the electrophoretic systems, TTAB, on one hand, and STC, SLN, SDS-Brij 35 and DDPG-DPPC, on the other hand, are the best ones to emulate aquatic toxicity, whereas SDS MECK and SDS MEECK do not perform as well. It is noteworthy that the classical and widely used O/W partition and C18 and IAM HPLC systems are not expected to

provide the best correlations for any of the aquatic toxicity systems. In fact, only IAM seems to be suitable for aquatic systems of the cluster of SS and the cluster of FM, DM24 and TP. These conclusions can be graphically observed in Figure 2, which presents the dendrogram and PCA plots of the joint aquatic toxicity and physicochemical systems. At 0.25 *d*-level, there are four clusters. One cluster is only formed by physicochemical systems (O/W, RP18 (C_{18} reverse phase HPLC column), SDS MEECK and SDS MEKC), without any toxicity system, and placed in the upper left side corner of PCA plot. Another cluster in the central part of the PCA plot includes the toxicity systems of RT, FM, TP and DM24, and the physicochemical systems of STC, SLN, SDS-Brij 35, AGESS (dodecane allyl glycidyl ether sulfite-modified siloxane), DPPG-DPPC and even IAM. IAM is placed in one side of the cluster, close to the O/W cluster. TTAB and SS are in another cluster. Finally, there is one cluster with only toxicity systems (PP, GO and CP), but TTAB is the physicochemical system closest to this cluster. In fact, the central position of TTAB among all the selected toxicity systems suggests that it can surrogate well all the aquatic toxicity systems, confirmed by the good *d* distances and predicted variances for TTAB presented in Tables 1 and 2. Also SLN, STC and SDS-Brij 35 are also close to TTAB, in the central part of PC plot, and may surrogate well many aquatic toxicity systems. According to the results obtained through the different comparison tools and our previous experience on aquatic toxicity modeling [22,23], three MEKC systems (STC, TTAB and SDS-Brij 35) and one HPLC system (IAM) have been selected to correlate their experimental values against the ones of the biological systems. STC and SDS-Brij 35 micelles are anionic surfactants whereas TTAB belongs to cationic surfactants class. SLN has been discarded because it has a high UV-Vis absorbance and may interfere with the tested substances detection. The other SDS-based systems are not expected to model toxicity as well as the ones selected. DPPG-DPPC system, which uses liposomes as pseudo-stationary phase, has not been selected because liposomes are not as easy to prepare and manageable as the other systems present in the same cluster. As regard HPLC systems, the one based on an IAM column is the one that can better model the aquatic toxicity; thus, it has been included in the experimental study.

322

323 Selection of the solutes to be tested

324

325 Nearly 500 substances with known toxicity values to at least one of the eight selected aquatic species
326 have been considered in this work[7,9–11,24–33]. A PCA analysis of the available solutes has been done
327 according to their SPM molecular descriptors[14]. In this way, compounds are distributed in the scores
328 plot according to their physicochemical properties. Four criteria have been followed to select the
329 compounds that will further be analyzed in the chosen chromatographic systems: firstly, substances
330 must cover all the PCA chemical space to assure a set of compounds that are chemically diverse (this
331 is, their descriptors must be representative of the chemical space[14]); secondly, they must be neutral at
332 the working pH, since all the SPM equations compared stand only for neutral compounds; third,
333 selected compounds must have chromophore groups due to detection requirements; and finally, a
334 minimum of 10 solutes for each species must be selected. According to these criteria, a selection of
335 152 compounds of known toxicity data was made (Table S-4 of the supplementary material). Figure
336 S-1 of the supplementary material shows the final distribution of the selected compounds, all of the
337 graphic area is covered with the selected solutes.

338

339 Evaluation of the performance of chromatographic systems to estimate nonspecific toxicity to aquatic
340 species

341

342 After selection of the compounds, their retention factors were determined in the physicochemical
343 systems (Table S-5 of the supplementary material). Then, a regression analysis of experimental
344 toxicity property logarithm values *vs* retention factor logarithm value was done, according to (Eq. 4).
345 In case of RT, literature reports narcosis data of different RT species, most of them belonging to *R.*
346 *temporaria* and *R. japonica* ones. Similarly to previous studies [9], a flag descriptor (I_{jap}) was introduced

into the global equation to obtain more accurate predictions. Therefore, toxicity to RT was correlated through the following equation:

$$\log C_{nar} = q + p \log k + i I_{jap} \qquad \text{[Eq. 5]}$$

The indicator variable I_{jap} is set to 1 when the toxicological property has been evaluated on *R. japonica* species and to 0 otherwise (*R. temporaria*).

Table 3 includes the regression parameters of all of the systems evaluated. As already shown on a previous research work, toxicity to RT can be modeled using any of the three systems in its cluster (STC, SDS-Brij 35 and IAM). TTAB prediction is not so good. In the case of FM, the system that best can surrogate the toxicity is STC, followed by IAM. SDS-Brij 35 and TTAB can also be used but they show weaker prediction abilities. Toxicity to DM24 and TP, in the same cluster as RT and FM, can be modeled well using either IAM or SDS-Brij 35; STC is not a so precise system. TTAB shows good prediction ability for TP but not so good for DM24.

Initial predictions pointed out that TTAB would be the best system to emulate SS. We also found that SDS-Brij 35 and STC could be alternatives. These results agree quite well with the predictions based on Tables 1 and 2 and Figure 2. SDS-Brij 35 and STC were in the neighbor cluster to that of SS, the *d* distance was close to 0.25 and the $SD_{corr\ cal}^2$ value was similar to that of TTAB.

TTAB was not clustered to GO, CP and PP systems but was in the neighborhood and would be the best of the tested systems to emulate the toxicity to these aquatic species. In fact, this system has moderate ability to emulate GO, CP and PP toxicity. As expected, R^2 and F are not as high as in the case of SS because these three systems do not have the same level of similarity as SS to TTAB. Further efforts have to be done to find suitable chromatographic systems to model the toxicity to PP. In the case of GO and CP, the TTAB system seems promising but the number of tested solutes should be increased to confirm its ability to model toxicity.

372

373 Validation of the best chromatographic models

374 Those systems showing the best correlation parameters have been validated to prove their robustness
375 and prediction ability. The selected models have been internally and externally validated in order to
376 check their robustness and predictive ability, respectively. To perform the model's validation, the set
377 of solutes was divided into a training set (around 2/3 of the compounds) and a test set (around 1/3 of
378 the compounds). To ensure all type of compounds were included in both the training and test sets,
379 this selection was done on the basis of a PCA with solute SPM descriptors that represents the chemical
380 space (see Figure S-1 of the Supplementary material). For the internal validation, the models were
381 established again, but only with the solutes of the training set. Table 4 shows the correlation
382 parameters obtained and also the toxicity range evaluated for each of the systems. Equations'
383 coefficients are similar to those of the models with all solutes (Table 3), which is indicative of the
384 robustness of the models. Adequate determination coefficients, standard deviations and F values were
385 obtained. Furthermore, an additional parameter, the leave-multiple-out cross-validation coefficient
386 was calculated. This coefficient was higher than 0.6 in all cases, which also points out the robustness
387 of the selected systems [34].
388 Finally, the external validation was performed. A regression was done between the experimental
389 toxicity and the one predicted through the training set equations for the compounds of the test set.
390 Table 5 shows the correlation parameters and statistics, including the leave-multiple-out cross-
391 validation coefficient. According to statistics, all models considered show good prediction ability: the
392 slopes of the trend lines are not significantly different from unity and the intercepts from zero at 95%
393 confidence level by students t-test; the variances (SD^2) are of the same order of that of the biological
394 data, the determination coefficients (R^2) are above 0.6 and similar to $R_{adjusted}^2$; the correlation cross-
395 validation coefficients ($QLMO^2$) are above 0.5; and Fisher's F parameter is significant.
396

CONCLUSION

The similarity analysis from *d* distances between different biological systems used to predict aquatic toxicity shows that most systems provide similar toxicity information. The same methodology together with estimation of the variance of the toxicity-chromatographic retention prediction show that some chromatographic systems can surrogate aquatic toxicity measurement systems. Thus, the toxicity of chemicals to a particular aquatic species can be easily estimated from the chromatographic retention of the chemicals in the surrogating chromatographic systems. The similarity analysis shows that the chromatographic systems that are able to best emulate the aquatic toxicology to eight representative species (*Rana* tadpoles, Fathead minnow, Golden orfe, *Dapnia Magna 24h*, *Tetrahymena Pyriformis, Chilomonas paramecium Pseudomonas putida* and *Streptococus sobrinus*) are the micellar electrokinetic chromatographic systems with micellar pseudoestationary phases of STC, a mixture of SDS and Brij 35, or TTAB. The similarity analysis also shows that other biological systems ruled by similar toxicity mechanisms (such as Guppy, Bluegill, Medaka high-eyes, *Daphnia magna* 48 h, *Ceriodaphnia dubia*, *Daphnia pulex*, *Spirostomum ambiguum*, *Porphyromonas gingivalis* or *Selenomonas artemidis*) can be surrogated using the same chromatographic systems.

SUPPORTING INFORMATION

Supporting information file includes:

Table S-1. Coefficients, statistics and normalized coefficients of the SPM characterization of systems related to toxicity to aquatic species.

Table S-2. Coefficients, statistics and normalized coefficients for physicochemical systems characterized through the SPM.

Table S-3. Contributions that determine the overall variance ($SD_{corr\ cal}^2$) in the correlations between aquatic species toxicity data and chromatographic or partitioning data of the considered physicochemical systems.

Table S-4. Experimental toxicity data for the selected solutes.

Table S-5. Solutes retention factor logarithm under the different chromatographic conditions.

Figure S-1. PCA scores plot of solutes with known toxicity data and the projection of the solute selection made in this work.

Figure S-2. Plots of biological property logarithm vs. experimental retention factor logarithm of the physicochemical systems with the best regressions. Solid lines are the plot of the regression equation.

ACKNOWLEDGEMENTS

Financial support from the Ministerio de Economía y Competitividad from the Spanish Government (CTQ2014-56253-P) and the Catalan Government (2014SGR277) is acknowledged. AFP wishes to thank the University of Barcelona for his APIF PhD fellowship.

CONFLICT OF INTEREST

The authors declare no competing financial interest.

REFERENCES

(1) Ankley, G. T.; Villeneuve, D. L. *Aquat. Toxicol.* **2006**, *78*, 91–102.

(2) ECHA (European Chemicals Agency). What about animal testing? http://echa.europa.eu/chemicals-in-our-life/animal-testing-under-reach (accessed Apr 11, 2016).

(3) EPA. Process for evaluating & implementing alternative approaches to traditional in vivo acute toxicity studies for FIFRA regulatory use https://www.epa.gov/sites/production/files/2016-03/documents/final_alternative_test_method_guidance_2-4-16.pdf (accessed Jan 31, 2017).

(4) Poole, C. F.; Ariyasena, T. C.; Lenca, N. *J. Chromatogr. A* **2013**, *1317*, 85–104.

(5) Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73.

(6) Abraham, M. H.; Rafols, C. *J. Chem. Soc. Perkin Trans. 2* **1995**, No. 10, 1843.

(7) Hoover, K. R.; Acree, W. E.; Abraham, M. H. *Chem. Res. Toxicol.* **2005**, *18*, 1497–1505.

(8) Bowen, K. R.; Flanagan, K. B.; Acree, W. E.; Abraham, M. H. *Sci. Total Environ.* **2006**, *369*, 109–118.

(9) Bowen, K. R.; Flanagan, K. B.; Acree, W. E.; Abraham, M. H.; Rafols, C. *Sci. Total Environ.* **2006**, *371*, 99–109.

(10) Mintz, C.; Acree, W. E.; Abraham, M. H. *QSAR Comb. Sci.* **2006**, *25*, 912–920.

(11) Hoover, K. R.; Flanagan, K. B.; Acree Jr., W. E.; Abraham, M. H. *Sect. Title Toxicol.* **2007**, *6*, 165–174.

(12) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1085–1100.

(13) Trone, M. D.; Khaledi, M. G. *Electrophoresis* **2000**, *21*, 2390–2396.

(14) Fuguet, E.; Ràfols, C.; Bosch, E.; Abraham, M. H.; Rosés, M. *J. Chromatogr. A* **2002**, *942*, 237–248.

(15) Schulte, S.; Palmer, C. P. *Electrophoresis* **2003**, *24*, 978–983.

(16) Lázaro, E.; Ràfols, C.; Abraham, M. H.; Rosés, M. *J. Med. Chem.* **2006**, *49*, 4861–4870.

(17) Liu, J.; Sun, J.; Wang, Y.; Liu, X.; Sun, Y.; Xu, H.; He, Z. *J. Chromatogr. A* **2007**, *1164*, 129–138.

(18) Kipka, U.; Di Toro, D. M. **2009**, *28*, 1429–1438.

(19) Fuguet, E.; Ràfols, C.; Bosch, E.; Abraham, M. H.; Rosés, M. *Electrophoresis* **2006**, *27*,

1900–1914.

(20) Castillo-Garit, J. a.; Marrero-Ponce, Y.; Escobar, J.; Torrens, F.; Rotondo, R. *Chemosphere* **2008**, *73*, 415–427.

(21) Hidalgo-Rodríguez, M.; Fuguet, E.; Ràfols, C.; Rosés, M. *Anal. Chem.* **2010**, *82*, 10236–10245.

(22) Hidalgo-Rodríguez, M.; Fuguet, E.; Ràfols, C.; Rosés, M. *Anal. Chem.* **2012**, *84*, 3446–3452.

(23) Fernández-Pumarega, A.; Amézqueta, S.; Fuguet, E.; Rosés, M. *J. Chromatogr. A* **2015**, *1418*, 167–176.

(24) Overton, E. *Studien über die Narkose, zurgleich eing Bertrag zur allgemeinen Pharmakologie.*; Jena, Gustav Fischer, 1901.

(25) Huang, H.; Wang, X.; Ou, W.; Zhao, J.; Shao, Y.; Wang, L. *Chemosphere* **2003**, *53*, 963–970.

(26) Wang, X.; Dong, Y.; Xu, S.; Wang, L.; Han, S. *Bull. Environ. Contam. Toxicol.* **2000**, *64*, 859–865.

(27) Overton, E. *Studies on Narcosis*; Lipnick, R. L., Ed.; Chapman and Hall: London, 1991.

(28) Meyer, K.; Hemmi, H. *Biochem. Z.* **1935**, *277*, 39–71.

(29) Kita, Y.; Bennett, L. J.; Miller, K. W. *Biochim. Biophys. Acta - Biomembr.* **1981**, *647*, 130–139.

(30) Juhnke, I.; Luedemann, D. *Z. Wasser Abwasser Forsch* **1978**, *11*, 161–164.

(31) Bringmann, G.; Kühn, R. *Water Res.* **1980**, *14*, 231–241.

(32) *Aquatic toxicology and environmental fate*; Suter, G., Lewis, M., Eds.; ASTM: Philadelphia, 2007.

(33) Dobbins, L. L.; Usenko, S.; Brain, R. A.; Brooks, B. W. *Environ. Toxicol. Chem.* **2009**, *28*, 2744–2753.

(34) Roy, K. *Expert Opin. Drug Discov.* **2007**, *2*, 1567–1577.

TABLES

Table 1

d distance values in the correlations between aquatic species toxicity data and chromatographic or partitioning data of the considered physicochemical and biological systems.

System		O/W	RP18	IAM	SDS MEKC	SLN	STC	TTAB	SDS-Brij 35	SDS MEEKC	DPPG-DPPC	AGESS
Tadpoles	RT	0.121	0.223	0.143	0.173	0.105	0.09	0.178	0.109	0.155	0.198	0.103
Fish	FM	0.198	0.259	0.186	0.347	0.162	0.183	0.246	0.128	0.176	0.095	0.118
	GO	0.391	0.480	0.360	0.417	0.246	0.306	0.271	0.298	0.395	0.282	0.290
Water fleas	DM24	0.267	0.310	0.258	0.402	0.224	0.240	0.307	0.201	0.238	0.158	0.190
Protozoan	TP	0.205	0.277	0.208	0.310	0.121	0.156	0.238	0.131	0.186	0.123	0.107
	CP	0.380	0.448	0.343	0.371	0.280	0.270	0.262	0.293	0.396	0.327	0.300
Bacteria	PP	0.295	0.329	0.315	0.238	0.242	0.172	0.316	0.252	0.295	0.320	0.244
	SS	0.313	0.424	0.258	0.344	0.217	0.274	0.123	0.246	0.352	0.270	0.257

503 **Table 2**

504 Calculated precision ($SD_{corrcal}^2$) values in the correlations between aquatic species toxicity data and

505 chromatographic or partitioning data of the considered physicochemical and biological systems.

System		O/W	RP18	IAM	SDS MEKC	SLN	STC	TTAB	SDS-Brij 35	SDS MEEKC	DPPG-DPPC	AGESS
Tadpoles	RT	0.206	0.317	0.201	0.167	0.177	0.168	0.172	0.095	0.224	0.162	-
Fish	FM	0.183	0.250	0.145	0.251	0.123	0.140	0.188	0.114	0.183	0.094	-
	GO	0.351	0.408	0.318	0.288	0.245	0.204	0.205	0.241	0.356	0.246	-
Water fleas	DM24	0.248	0.342	0.181	0.244	0.125	0.135	0.144	0.127	0.252	0.117	-
Protozoan	TP	0.125	0.180	0.087	0.168	0.069	0.074	0.112	0.066	0.132	0.057	-
	CP	0.509	0.617	0.383	0.500	0.297	0.303	0.227	0.307	0.529	0.279	-
Bacteria	PP	0.292	0.353	0.242	0.239	0.157	0.144	0.169	0.176	0.297	0.182	-
	SS	0.150	0.174	0.137	0.138	0.112	0.114	0.107	0.120	0.154	0.125	-

506

507

Table 3

Experimental evaluation of the performance of the studied chromatographic systems to emulate toxicity of organic compounds to eight aquatic species (standard deviations in brackets). The biological systems have been grouped according to the clustering in the dendrogram (Fig. 2b).

		q (SD_q)	p (SD_p)	i(SD_i)	$SD_{correxp}{}^2$	n	R^2	F	$n_{outliers}$
R. tadpoles	IAM	3.10 (0.06)	2.67 (0.13)	0.21 (0.08)	0.080	60	0.900	255	5
	STC	3.21 (0.05)	1.24 (0.06)	0.08 (0.07)	0.065	56	0.897	231	4
	SDS-Brij 35	2.40 (0.05)	1.00 (0.05)	0.10 (0.08)	0.067	57	0.896	232	3
	TTAB	2.66 (0.08)	0.86 (0.09)	0.19 (0.12)	0.162	54	0.691	57	1
F. minnow	IAM	3.55 (0.05)	2.26 (0.14)	-	0.152	63	0.815	268	3
	STC	3.77 (0.04)	0.93 (0.06)	-	0.072	54	0.840	273	5
	SDS-Brij 35	2.71 (0.11)	1.18 (0.10)	-	0.296	41	0.763	126	1
	TTAB	3.24 (0.06)	0.97 (0.07)	-	0.155	58	0.784	203	1
D. magna (24 h)	IAM	3.58 (0.07)	2.74 (0.16)	-	0.197	47	0.865	289	1
	STC	4.02 (0.07)	0.86 (0.10)	-	0.129	40	0.679	80	1
	SDS-Brij 35	2.61 (0.12)	1.42 (0.12)	-	0.237	32	0.819	135	1
	TTAB	3.24 (0.10)	1.16 (0.11)	-	0.216	37	0.774	120	1
T. pyriformis	IAM	0.02 (0.05)	2.08 (0.12)	-	0.129	61	0.825	277	3
	STC	0.28 (0.04)	0.85 (0.06)	-	0.099	53	0.784	185	1
	SDS-Brij 35	-0.71 (0.04)	1.09 (0.05)	-	0.054	45	0.930	571	5
	TTAB	-0.28 (0.04)	0.92 (0.05)	-	0.071	55	0.862	331	4
S. sobrinus	IAM	-0.93(0.07)	1.75 (0.17)	-	0.113	28	0.811	112	0
	STC	-0.84 (0.06)	1.11 (0.08)	-	0.096	30	0.875	195	0
	SDS-Brij 35	-1.50(0.08)	0.77(0.06)	-	0.055	24	0.879	160	2

	TTAB	-1.57 (0.08)	0.82 (0.07)	-	0.055	23	0.871	142	2
G. orfe	IAM	3.29 (0.15)	2.30 (0.42)	-	0.364	19	0.641	30	0
	STC	3.49 (0.11)	0.38 (0.20)	-	0.153	13	0.247	4	1
	SDS-Brij 35	2.51 (0.15)	1.20 (0.20)	-	0.402	23	0.638	37	0
	TTAB	2.93 (0.11)	1.20 (0.16)	-	0.194	18	0.782	57	1
C. paramecium	IAM	3.46 (0.18)	1.62 (0.58)	-	0.401	17	0.342	8	0
	STC	3.89 (0.21)	1.22 (0.34)	-	0.277	15	0.501	13	0
	SDS-Brij 35	2.93 (0.15)	1.01 (0.26)	-	0.379	19	0.474	15	0
	TTAB	3.16 (0.14)	1.10 (0.27)	-	0.323	17	0.518	16	0
P. putida	IAM	3.56 (0.11)	2.14 (0.32)	-	0.315	33	0.594	45	0
	STC	3.81 (0.12)	0.67 (0.19)	-	0.259	25	0.357	13	0
	SDS-Brij 35	2.98 (0.10)	0.96 (0.13)	-	0.309	34	0.617	52	0
	TTAB	3.41 (0.09)	1.02 (0.15)	-	0.209	26	0.666	48	1

513

514

515

516 **Table 4**

517 Correlation parameters and statistics of the internal validation.

518

Biological system	Physicochemical System		q (SD_q)	p (SD_p)	i_{jap} ($SD_{i,jap}$)	SD_{corr}^2	N	R^2	F	Q_{LMO}^2	Toxicity range evaluated
RT	27	STC	3.23(0.07)	1.29 (0.09)	0.07 (0.11)	0.075	31	0.893	117	0.90	1.6 - 5.3 [a]
	29	SDS-Brij 35	2.44(0.07)	1.05(0.07)	-0.05(0.12)	0.083	32	0.904	137	0.90	0.8 - 4.4 [b]
FM	27	STC	3.81(0.05)	0.99(0.08)		0.071	31	0.849	163	0.85	2.4 - 5.3 [b]
DM24	24	IAM	3.53(0.08)	2.94(0.18)		0.175	32	0.900	269	0.90	1.1 - 5.3 [b]
	29	SDS-Brij 35	2.57(0.15)	1.43(0.16)		0.250	21	0.812	82	0.85	1.1 - 4.8 [b]
TP	29	SDS-Brij 35	-0.73(0.05)	1.08(0.06)		0.047	27	0.929	328	0.93	-1.8 - 1.6 [c]
SS	28	TTAB	-1.61(0.07)	0.85(0.06)		0.026	16	0.943	233	0.94	-1.6 - 0.5 [d]

519 [a] -log C_{nar} (C_{nar}: narcosis concentration), [b] -log LC50 (LC50: median lethal concentration, 50 %), [c] -log IGC50 (IGC50: median
520 inhibitory growth concentration, 50 %), [d] -log MIC (MIC: minimum inhibitory concentration towards bacterial growth)

521

Table 5

Correlation parameters and statistics of the external validation.

Biological system	Physicochemical System	q (SD_q)	p (SD_p)	SD_{corr}^2	n	R^2	R_{adj}^2	F	Q_{LMO}^2
RT	27 STC	0.05(0.20)	0.99(0.07)	0.06	25	0.90	0.90	213	0.92
	29 SDS-Brij 35	0.30(0.23)	0.91(0.07)	0.05	25	0.86	0.86	147	0.92
FM	27 STC	0.24(0.34)	0.96(0.09)	0.08	23	0.85	0.84	116	0.86
DM24	24 IAM	-0.21(0.61)	1.03(0.15)	0.30	15	0.79	0.78	49	0.81
	29 SDS-Brij 35	0.43(0.46)	0.85(0.13)	0.22	11	0.83	0.82	46	0.83
TP	29 SDS-Brij 35	-0.08(0.06)	0.92(0.06)	0.07	17	0.93	0.93	201	0.94
SS	28 TTAB	-0.19(0.22)	0.80(0.24)	0.14	7	0.70	0.64	12	0.77

FIGURE CAPTIONS

Figure 1

a) Dendrogram plot of the biological systems included in Table 1: tadpoles (T), fish (F), water fleas (W), protozoan (P) and bacteria (B). Selected systems are shown in boldface.

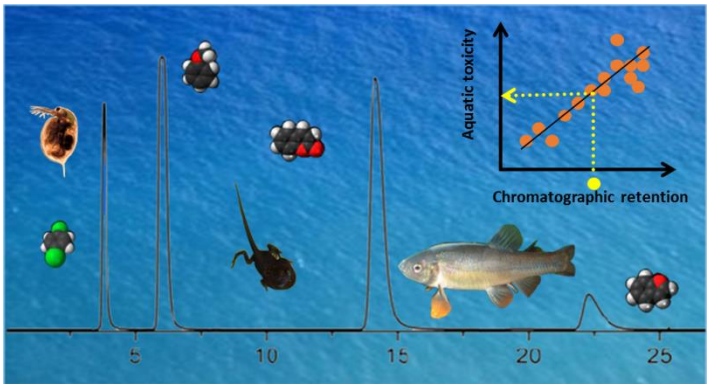
b) PCA scores plot of the biological systems included in Table 1. Selected systems are shown in dark grey.

Figure 2

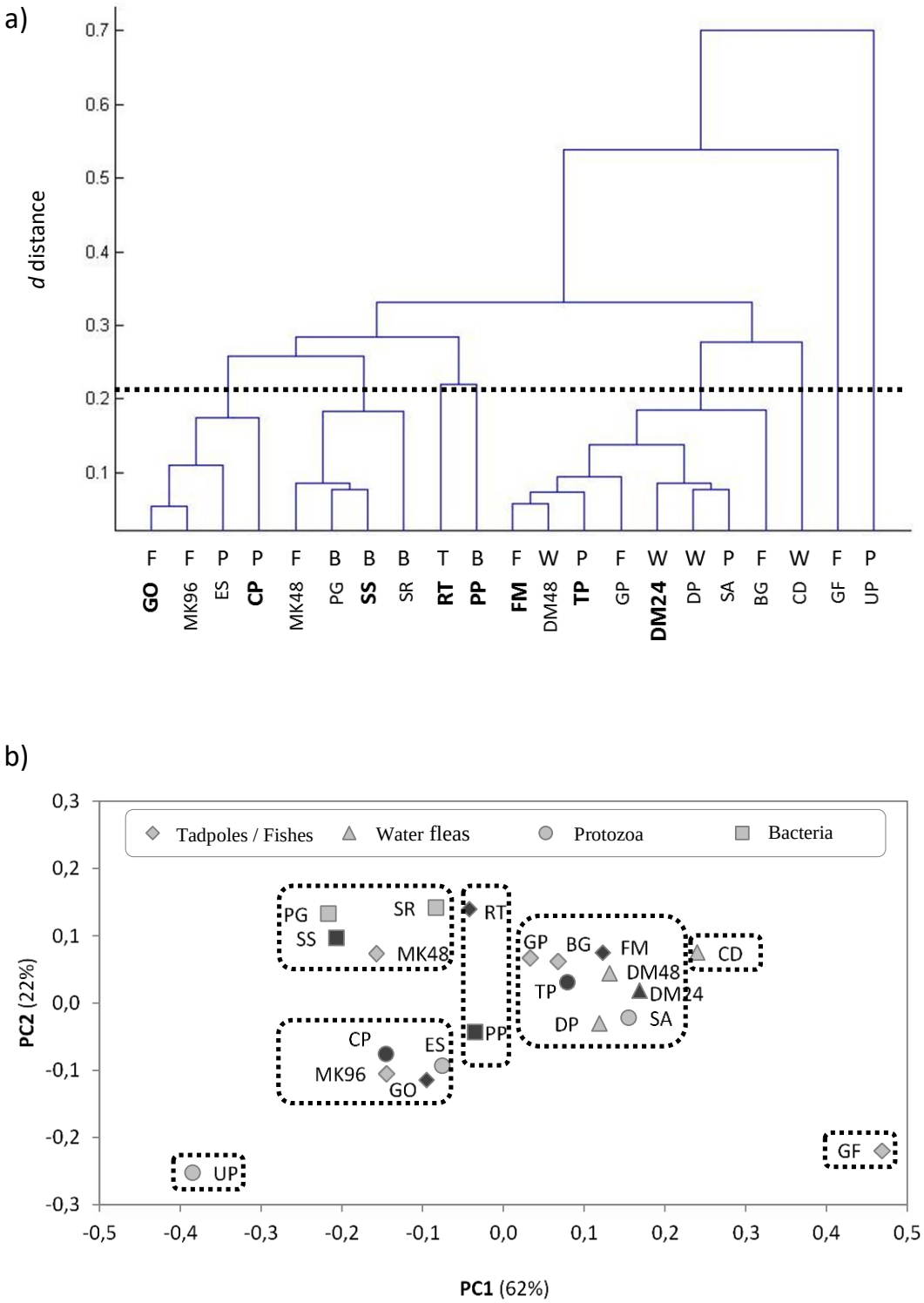
a) Dendrogram plot of the biological and physicochemical systems included in Table 1. Selected systems are shown in boldface.

b) PCA scores plot of the eight biological systems (circle) and the eleven physicochemical systems (diamond) evaluated in this work. The five final selected physicochemical systems are shown in dark grey.

FOR TOC ONLY



563 **Figure 1**

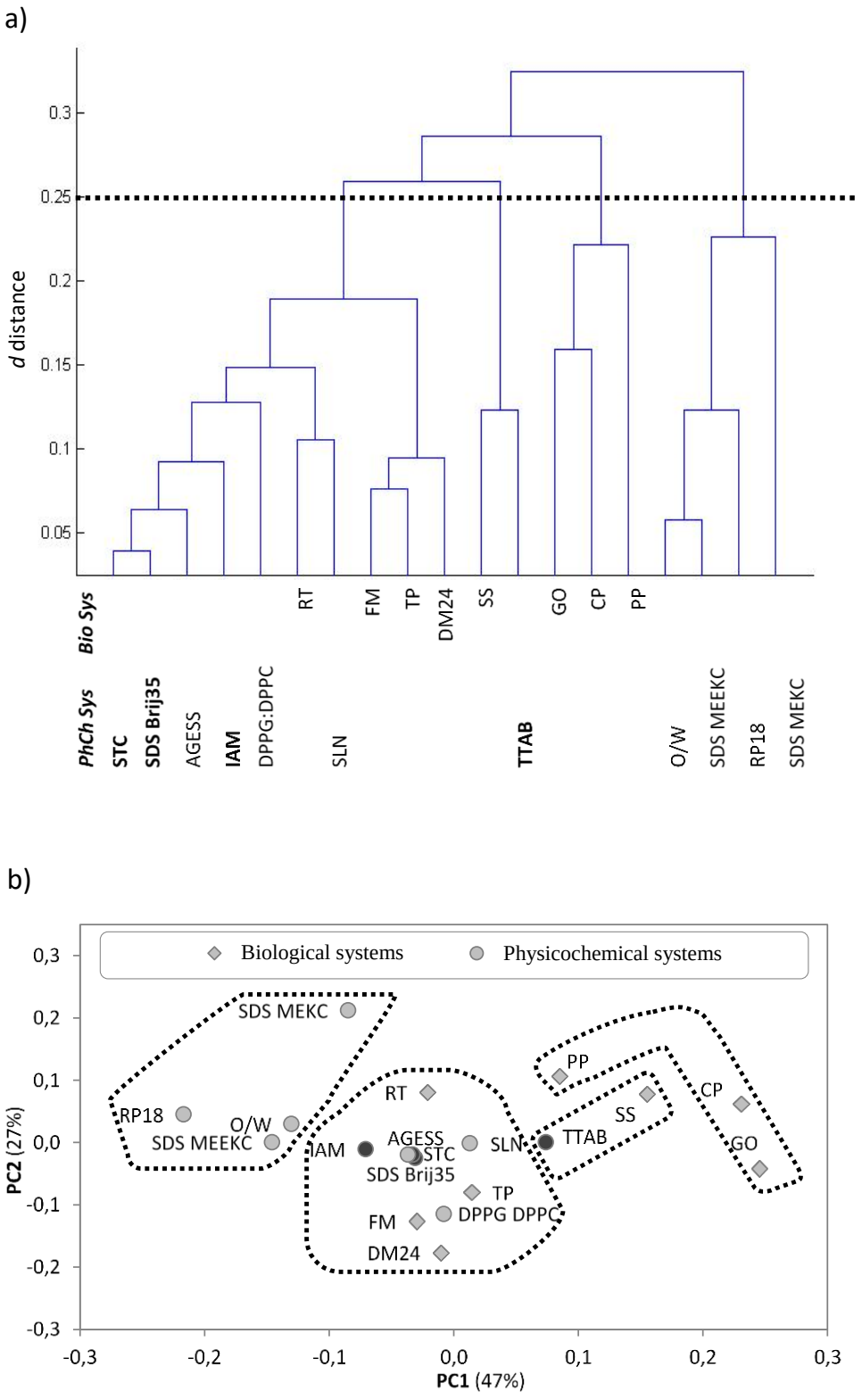


564

565

566

567 **Figure 2**



568