Prediction of Noninteractive Mixture Toxicity of Organic Compounds Based on a Fuzzy Set Method

Mulaisho Mwense,[†] Xue Zhong Wang,*[,†] Frances V. Buontempo,[†] Nigel Horan,[‡] Anita Young,[§] and Daniel Osborn[||]

Department of Chemical Engineering and School of Civil Engineering, The University of Leeds, Leeds LS2 9JT, U.K., Brixham Environmental Laboratory, AstraZeneca UK Ltd., Freshwater Quarry, Brixham, Devon TQ5 8BA, U.K., and Centre of Ecology and Hydrology, Monks Wood, Huntingdon PE28 2LS, U.K.

Received February 16, 2004

Current methods for the prediction of mixture toxicity have shown to be valid for mixtures that conform to some assumptions that were ideally formulated for mixtures comprising constituents exhibiting either completely similar or dissimilar mechanisms of action. Approaches are needed that predict the toxicity of mixtures representative of real environmental occurrences i.e., those comprising constituents of mixed similar and dissimilar compounds and therefore are more complex. In this paper such a methodology is proposed which uses molecular descriptors and fuzzy set theory to characterize the degree of similarity and dissimilarity of mixture constituents, integrates the concentration addition and independent action models, and therefore is called INFCIM (INtegrated Fuzzy Concentration addition − Independent action Model). INFCIM is tested in two case studies using toxicity data of four mixtures, and its performance is compared against those of both concentration addition and independent action models. Mixture 1 consists of 18 *s*-triazines acting on green freshwater algae *scenedemus vacuolatus*. Mixture 2 comprises 16 acting constituents tested on *scenedemus vacuolatus*. Both mixtures inhibit reproduction in the biological assays. There are 10 quinolone compounds in mixture 3 and 16 phenol derivative compounds in mixture 4 all causing long-term inhibition of bioluminescence in the marine bacterium *Vibrio fischeri*. It was shown that INFCIM performed comparably or better than the best performing existing model in the original studies for all the mixtures tested.

1. INTRODUCTION

Worldwide, more than 16.7 million distinct organic and inorganic chemicals are known, of which 70 000 are in commercial production.[1] Additionally, modern combinatorial chemistry techniques have led to the production of vast libraries of new chemicals at a very rapid rate. Modern speciality chemical manufacturers achieve their competitiveness through rapid introduction of new products on a regular basis. It is estimated that in the U.K. alone, more than 10 000 new chemicals are evaluated annually for possible production. In evaluating possible production of new chemicals, the environmental impact has in recent years overtaken energy and equipment, becoming the dominant cost factor. Toxicity is considered as one of the most important variables characterizing the hazardous effect of a chemical. The traditional way of assessing the toxic risk of compounds is by laboratory tests using biological assays. There are several drawbacks with this approach. First it is very expensive. Second it is time-consuming. In addition such tests using animals have little public support.

A promising technique that can minimize, if not completely replace laboratory test is predicting toxicity using quantitative structure−activity relationship (QSAR) models. There are now commercial QSAR software systems available for toxicity prediction such as TOPKAT,[2] DEREK expert systems,[3] and MultiCase.[4] These systems use statistical techniques, neural networks, and expert systems to relate molecular structural, physicochemical, and biological descriptors to endpoint toxicity measures. However, they only predict toxicity of pure substances, not mixtures, though the toxicants are rarely discharged as pure compounds.

Models exist for the prediction of mixture toxicity[5−7] that rely on full qualitative and quantitative information of the mixture constituents such as constituent molar compositions, concentration response profiles, and methods for determining the type of joint action within the mixture. They have shown to be applicable to mixtures that satisfy the assumption that the constituents of the mixture act with either completely similar or dissimilar acting mechanisms. Unfortunately, mixtures of the aquatic environments rarely satisfy such ideal conditions.[8] Of common occurrence in such aquatic environments are mixture constituents with mixed similar and dissimilar actions. No models or approaches are currently available that can predict a priori the toxicity of mixtures of such characteristics, though in industrial practice, some indices are calculated for such mixtures that give a useful insight as to whether the mixture would result in an additive, more than additive, or less than additive toxicity for noninteractive mixtures. For these reasons, an approach capable of determining mixture toxicity a priori regardless

* Corresponding author phone: +44 113 343 2427; fax: +44 113 343 2405; e-mail: x.z.wang@leeds.ac.uk.
[†] Department of Chemical Engineering, The University of Leeds.
[‡] School of Civil Engineering, The University of Leeds.
[§] AstraZeneca UK Ltd.
[||] Centre of Ecology and Hydrology.

10.1021/ci0499368 CCC: $27.50
Published on Web 07/20/2004

of whether mixture constituents possess similar, dissimilar, or mixed similar and dissimilar acting mechanisms is needed.

The paper is organized as follows. Section 2 briefly reviews the existing models for predicting mixture toxicity. Section 3 presents the framework of the proposed approach. The case study mixtures used are introduced in section 4. Section 5 describes in greater detail the method for characterizing similarity and dissimilarity of compounds in a mixture using descriptors and fuzzy membership functions. Then the INFCIM model is presented in section 6 which makes use of the fuzzy characterization to calculate the toxicity of a mixture. Results of applying the INFCIM model to the case study mixtures are presented in section 7, and conclusions are made in section 8.

2. EXISTING MODELS FOR PREDICTING MIXTURE TOXICITY

Existing models for mixture toxicity prediction were developed for mixtures with characteristics that meet the standard two-valued logic criteria, i.e., for mixtures comprising constituents with either completely similar or dissimilar mechanisms of action. For mixtures comprising similarly acting constituents, the overall effect can be assessed using the concept of concentration addition, while independent action can be used for assessing the overall effect of mixtures comprising dissimilarly acting constituents.

Concentration addition is based on the Loewe additivity reference model. This model assumes mixture constituents to contribute to the overall toxicity in proportion to their dose. In using this model, a constituent can be expressed in terms of another but adjusted to account for differences in their relative toxic potency. This means that constituents can be replaced by equally potent compounds with no change in the overall mixture effect. Under this model, even constituents present below zero effect concentrations contribute to the overall mixture effect.[9−13] The alternative concept, independent action, is based on the Bliss independence reference model which assumes mixture constituents to not cooperate physically, chemically, nor biologically when eliciting a toxic response[9,10,13−16] i.e., mixture constituents are assumed to act on completely different physiological system by different mechanisms. Constituents present below zero effect concentrations are assumed to not contribute to the overall mixture effect.

Both models have been tested for their general applicability in previous studies.[8,12,17−28] More often than not, these models are applied in a more flexible manner than suggested by their definitions i.e., the strict interpretations of similarity and dissimilarity with respect to the mechanisms of action are not usually adhered to.[29] The main conclusion from these studies has been that both models predict well the type of mixtures they are ideally developed for. However, if given a mixture with mixed similar and dissimilar actions it has not be tested and nor is it clear and unambiguous which of the two models would be selected a priori to model the toxicity of such a mixture. In the studies cited above, concentration addition is slightly more favored for general mixture toxicity modeling because of its tendency to overestimate the toxic risk posed for mixtures that do not meet the assumptions from which it is developed. It is also easier to carry out a toxicity assessment using this model requiring only effect concentrations of the constituents of the mixture effect desired as opposed to full concentration response curves of the constituents. The fact that an adverse effect can be elicited from the combination of mixture constituents below environmental safe levels (such as NOEC; no observed effect concentrations)[17,19,21,23] also make this concept more favorable.

Several other approaches based on Loewe additivity and Bliss independence other than the concepts of concentration addition and independent action exist for the prediction of mixture toxic effects. These include the median effect approach, isobolographic approach, effect summation, and fractional effect summation. Each of these approaches, despite being applicable in certain situations, have immense limitations that prevent them as candidates for the best general reference model for predicting mixture effects.[30] Detailed reviews of these approaches together with their limitations for general mixture toxicity modeling are presented by Greco et al.,[30] Cassee et al.,[10] and Berenbaum.[31]

Clear from all reviewed studies is that neither prediction concept can give unambiguous mechanistic explanation for the combined action assessment of mixtures representative of those occurring in aquatic environments. The choice of either concentration addition or independent action as the best reference model for general toxicity assessment of environmental mixtures is an ongoing debate. Since the concentration addition model tends to give higher toxicity predictions than independent action, it was recommended by some people that concentration addition model should be used. This idea is sometimes called the precautionary principle. However, using the precautionary principle could lead to significant overestimations of the toxic risk posed. The focus should then be on the development of an approach that enables the combined action assessment of mixtures characteristic of those present in real aquatic environments.

3. THE FRAMEWORK OF THE INFCIM MODEL

To overcome some of the limitations of the existing models, a new approach called integrated concentration addition-independent action modeling (INFCIM) is proposed. The underlying assumption for this new approach is that the toxicity of a mixture is due to both the similarity and dissimilarity of constituents within that mixture. This approach is illustrated using Figure 1 and comprises various elements that are discussed further in the subsequent sections. In this section, only an overview of it is presented.

The method comprises the following steps:

(1) For a mixture of n components, derive the dose or concentration response curves (CRCs) for all the components and the mixture $\boldsymbol{M}$ at a given composition $\boldsymbol{P}$.

(2) Calculate the descriptors for each compound using molecular modeling. It is important that descriptors employed in QSAR models for toxicity prediction should be included.

(3) Calculate intermolecular distances using the descriptors and then use fuzzy membership functions to calculate binary similarity and dissimilarity between mixture components.

(4) Calculate the overall mixture similarity and dissimilarity measures based on the similarity and dissimilarity values of molecular pairs.

(5) The CRC for the mixture $\boldsymbol{M}$ at the given composition $\boldsymbol{P}$ is used to optimize the selection of fuzzy membership

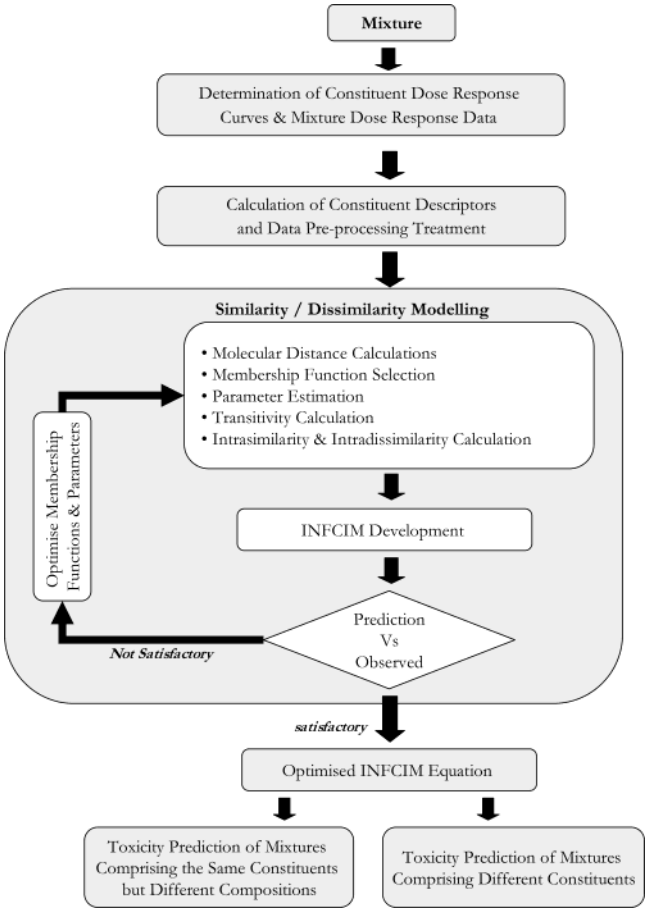

Figure 1. Overview of the INFCIM approach for determining toxicity of mixtures.

functions and tune the membership function parameters, as shown in the loop of Figure 1. The optimized fuzzy membership functions and parameter values can be used for future prediction of toxicity of mixtures of the same constituents but with different compositions.

(6) If only the CRCs for single compounds are available, but not for the mixture, then the fuzzy membership functions and parameter values obtained from a different mixture can be used as long as the toxicity end point is the same for both mixtures, as will be demonstrated in the case studies.

The last two points are potentially very beneficial to chemical manufacturers, because they mean that the toxicity of a mixture, e.g. effluents, just needs to be tested once. The INFCIM model can then be used to predict the toxicity of subsequent mixtures produced from changed operating conditions.

4. CASE STUDY MIXTURES

Since the case studies will be frequently used to help the illustration of the proposed mixture toxicity approach, we describe the case study mixtures here before the more detailed introduction of the method in the next section. Due to the obvious difficulties, mixture toxicity data are rarely available in the literature. We have compiled four mixtures which are organized into two case studies. The first case study includes two mixtures causing toxicity to green freshwater algae *scenedemus vacuolatus* and the second case study has two mixtures causing toxicity to the marine bacterium *Vibrio fischeri*. The four sets of data have previously been investigated in the literature to compare the performance of concentration addition and independent action toxicity models. An advantage of using these data is therefore that the current INFCIM model can be compared with concentration addition and independent action models. The mixtures used are summarized in Table 1.

Mixture 1 comprises 18 *s*-triazines acting on green freshwater algae *scenedemus vacuolatus*.[22] *s*-Triazines are specific inhibitors of the photosynthetic electron transport system. The 18 *s*-triazines are considered as following a similar mechanism of action. Both existing models were used to estimate the toxicity of this mixture, and as expected concentration addition modeled the mixture toxicity better than independent action. Mixture 2 consists of 16 dissimilar acting constituents.[23] This mixture was tested on *scenedemus vacuolatus* to study the combined action behavior of the constituents. Like for mixture 1, inhibition of reproduction was the toxicity parameter. As expected, independent action modeled the mixture toxicity very well and better than concentration addition.

Mixtures 3 comprised 10 quinolone compounds with a similar primary mechanism of action (inhibition of bacterial gyrases)[21] for causing long-term inhibition of bioluminescence in the marine bacterium *Vibrio fischeri*. Mixture 4 comprised 16 phenol derivative compounds known to be uncouplers of oxidative phosphorylation.[17] Like for the quinolone mixture, the toxicity parameter used was the inhibition of bioluminescence to *Vibrio fischeri*. Concentration addition expectedly modeled the toxicity of both mixtures better than independent action for all mixture ratios tested. Detailed descriptions of the preparations of the test substances and biological cultures together with concentration response curve analyses for the mixture constituents and the mixtures are given in the cited publications.

5. CHARACTERIZATION OF THE DEGREE OF SIMILARITY AND DISSIMILARITY OF COMPOUNDS IN A MIXTURE USING FUZZY MEMBERSHIP FUNCTIONS

5.1. Fuzzy Set Membership Functions. Classical set theory imposes a strict membership of an object to a set; that is, objects either belong or do not belong to a set. Fuzzy set theory was developed to depart from this two valued logic approach. A fuzzy set is defined by its objects and their respective grades of membership to the set. The grade of membership of an object in the fuzzy set is given by a defined fuzzy membership function. The value of the grade of membership of an object can range from 0 to 1. Figure 2 depicts fuzzy membership functions defining the fuzzy concepts “L” (Low), “M” (Medium), and “H” (High). The dotted lines show when the numerical value of the variable is 22.0, it is converted to a fuzzy value “H” with a fuzzy membership value of 0.7, i.e., (H, 0.7).

The most important feature of fuzzy set theory is the ability to express in numerical format the impression that stems from a grouping of elements into classes that do not have sharply defined boundaries. Fuzzy set theory has been found to be a powerful mathematical tool in artificial intelligence especially in the areas of knowledge representation and designing fuzzy expert systems and neural networks, qualitative reasoning, and pattern recognition. It has also been widely

Table 1: Summary of Mixtures Used in This Study[a]

case study 1						case study 2					
mixture 1[22]			mixture 2[23]			mixture 3[21]			mixture 4[17]		
	EC_{50}[b] (nmol/L)	EC_1[b] (nmol/L)		EC_{50}[b] (μmol/L)	EC_1[b] (μmol/L)		EC_{50}[b] (nmol/L)	EC_1[b] (nmol/L)		EC_{50}[b] (μmol/L)	EC_1[b] (μmol/L)
ametryn	15.6	0.7	norflurazon	0.0203	0.00564	ofloxacin	37.6	2.6	FCCP[e]	0.57	0.09
terbutryn	32.3	3.2	aclonifen	0.0297	0.00351	lomefloxacin	63.6	4.58	CCCP	0.81	0.1
desmetryn	36.8	3.2	DTMAC[c]	0.0302	0.00788	norfloxacin	68.9	16.3	4-phenylazophenol	9.88	0.05
simetryn	47.8	1.9	terbuthylazinee	0.0693	0.00572	flumequine	72.8	8.78	dinoterb	12.2	1.51
dimethametryn	51.5	6.6	metazachlor	0.168	0.00540	oxolinic acid	87.7	13.1	pentachlorophenol	12.98	3.34
prometryn	51.8	1.7	8-azaguanine	0.381	0.00044	enoxacin	154	12.4	2,4-dinitro-1-naphthol	15.44	3.3
methoprotryne	69.1	5.2	paraquat dichloride	0.781	0.215	piromidic acid	421	30	dinoserb	16.4	1.29
terbuthylazine	69.3	5.7	CCCP[d]	1.24	0.105	cinoxacin	446	15.5	2,3,4-trichlorophenol	17.25	1.98
dipropetryn	80.1	3.6	azaserine	1.78	0.00256	nalidixic acid	861	162	2,3,5-trichlorophenol	19.05	8.15
terbumeton	103	4.1	kresoxim-methyl	2.04	0.161	pipemidic acid	3360	718	2,4,6-trichlorophenol	24.38	1.3
cyanazine	113	6.9	triadimenol	3.46	0.319				2,6-dinitro-4-methyl-phenol	29.76	0.98
sebuthylazine	133	8.6	metsulfuron-methyl	4.02	0.0430				2,4-dinitrophenol	39.24	0.05
secbumeton	168	1.5	fenfuram	4.32	0.0784				4,6-dinitro-*o*-cresol	58.95	3.2
atrazine	180	1.7	chloramphenicol	32.5	2.75				2,6-dinitrophenol	81.1	6.78
simazine	282	3.2	nalidixic acid	98.7	5.01				3,4-dinitrophenol	87.54	3.38
propazine	310	2.4	metalaxyl	205	6.08				2,3,6-trinitrophenol	90.6	1.23
atraton	378	6.9									
prometon	547	22									

[a] Mixtures 1, 3, and 4 comprise similarly acting constituents, while mixture 2 comprises dissimilarly acting constituents. [b] EC_{50}, EC_1: concentrations estimated to provoke an effect of 50% and 1%. [c] DTMAC: dodecyltrimethylammonium-chloride. [d] CCCP: carbonyl cyanide-*m*-chlorophenyl-hydrazone. [e] FCCP: carbonyl cyanide-*p*-trifluoro-methoxyphenyl-hydrazone.

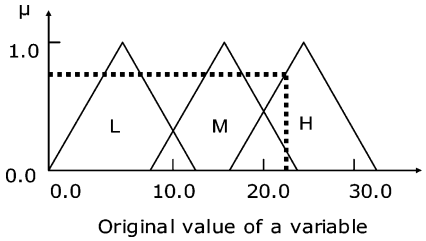

Figure 2. An example showing when the original value of a variable is 22.0, it takes a fuzzy value of "H" (High) with a fuzzy membership value μ of 0.7, i.e., (H, 0.7).

studied in developing intelligent fuzzy logic control and optimization systems in engineering.

The terms "similar" and "dissimilar" in describing the mechanisms of actions of compounds in a mixture are also fuzzy concepts. It naturally leads to the idea of quantifying the similarity and dissimilarity of compounds in a mixture using fuzzy membership functions. The detailed method for characterizing the similarity and dissimilarity of compounds in a mixture will be introduced in section 5.3. Here we continue the mathematical discussion on the fuzzy membership functions. Fuzzy membership functions can have any shape, but the most common are triangular, trapezoidal, and bell. In some applications, selection of the shape of fuzzy membership functions may not be very important. However, in some other applications such as in the current study, proper selection of the membership function can significantly improve the modeling performance. The determination of the best optimized fuzzy membership function is still a research topic in fuzzy mathematics, and there are sophisticated methods proposed in the literature for their automatic determination such as the differential evolution algorithms,[32] genetic algorithms, threshold methods, and the hillclimbing methodologies, as reviewed in greater detail by Blazquez et al.[33] These methods select fuzzy membership functions and/or tune function parameters to optimize the modeling accuracy. However it is generally accepted that they are often too complicated and time-consuming to be of practical use. As will be described later, in this study we did not use these automatic methods for the fuzzy membership determination.

5.2. Similarity and Molecular Descriptors. The qualitative assessment of compound similarity employed when using existing mixture toxicity prediction models is often based on a loosened definition of mechanism of action that could be mistaken for the definition of mode of action. This intermittent lack of clarity between these two definitions[34] could lead to different similarity assessments and hence the use of the wrong mixture toxicity prediction model. In real environmental mixtures, this assessment is further complicated by compounds possessing multiple toxic mechanisms of action leading to a common response such as death. An objective quantitative assessment of similarity and dissimilarity is thus desirable that relies less on the rare, lengthy, and costly to determine mechanisms of action.

Quantitative structure activity relationships (QSAR) can be used to help determine similarity between compounds. They have been studied extensively to predict toxicity endpoints of single compounds using molecular descriptors.[35−52] The underlying assumption for QSAR toxicity models is that compounds possessing similar structural, physicochemical, and biological properties or descriptors lead to similar biological action. The descriptors used for QSAR modeling can be broadly classified into empirical and theoretical descriptors. Empirical descriptors are derived from experimentation, whereas theoretical descriptors are derived from molecular simulation.[53,54] Theoretical descriptors are thus faster to compute, more readily available, and are applicable to not yet synthesized compounds. Quantum chemical descriptors and connectivity indices are examples of such descriptors. Detailed descriptions of these descriptors are presented in several handbooks and reviews.[53−55] Commercial software packages can be used to calculate these descriptors;

these include TSAR and VAMP for 2-, 3-D, and quantum mechanic descriptors,[2] DRAGON for 0−3D descriptors,[56] and MOPAC[57] and AMPAC[58] for calculating semiempirical quantum chemical descriptors.

The mechanistic meaning behind the use of empirical descriptors is better understood than for theoretical descriptors in general.[59] Nevertheless, theoretical descriptors such as the quantum chemical descriptors (e.g. E_{LUMO}) and some topological indices have been extensively studied in QSAR and similarity analysis to ease their mechanistic interpretation and as possible alternatives to the more traditional empirical descriptors.[60−67]

A consequence of the large number of descriptors that can be calculated for a particular molecule is that it is difficult to determine the best set of descriptors for toxicology applications. This is because different classes of descriptors encode different classes of information.[68] Hence, descriptors that fail to characterize a particular activity do so solely because they do not cover the important part of the characterization of structural features important for the activity considered and vice versa.[60] The best set or most representative set of descriptors can thus be viewed as those comprising descriptors from several classes. However, the selection of descriptors from different classes is still a difficult task. Descriptors cannot be eliminated unless they are shown to not model the activity being considered. This is not as big a problem when modeling congeneric compounds where a parent functional group exists. Where the compound set is heterogeneous or noncongeneric, the absence of a parent functional group results in larger descriptor sets for which selection of representative descriptors becomes more dependent on statistical analysis.[69] Livingstone,[68] Bajorath,[70] and Benigni et al.[69] suggested the use of different types of descriptors chosen after analysis of the information content of a larger set using techniques such as principal component analysis, procrusters analysis, and canonical correlation. Practical considerations such as the availability of descriptors and the number of times descriptors have been used previously are also part of the selection process.[59]

In this study, the molecular descriptor software system DRAGON was used to calculate 1497 descriptors. DRAGON descriptors were used previously in the literature for predictive toxicology studies.[63,71−74] A full list of the descriptors used can be obtained from the authors of DRAGON.[55,56] The inputs to DRAGON were geometrically optimized molecules saved as HIN files in HyperChem release 7.[75] The number of descriptors was reduced to 1208 and 1183 for case study 1 and 2, respectively, after the removal of variables that have constant values. No further reduction of descriptor numbers was carried for reasons mentioned earlier. Though quite obvious, it is still worth mentioning that the proposed similarity assessment in this paper does not rely on the use of all the descriptors. Descriptors known to be relevant for describing the toxicity in this study could have been used on their own were such knowledge available.

5.3. Fuzzy Characterization of Similarity and Dissimilarity. The degree of similarity for the compounds in a given mixture can be either characterized by the distances between molecule pairs or calculated using a clustering approach such as the fuzzy *c*-means clustering technique,[76] which computes the distance of each compound from the cluster center. In this study, the former method is used. In other words we first calculate the Euclidean distances between all pairs of molecules. The distances between molecule pairs are then used to calculate the degrees of similarity and dissimilarity using two fuzzy membership functions, one for similarity and the other for dissimilarity. This results in similarity and dissimilarity fuzzy sets from which an overall measure of the similarity and dissimilarity can be calculated (vide infra). Several other metrics instead of Euclidean distance could have been used for calculating molecular distances such as the Mahalanobis and Tanimoto coefficients. However, the Euclidean distance is frequently adopted in chemometric analysis and was therefore the method of choice in this study.

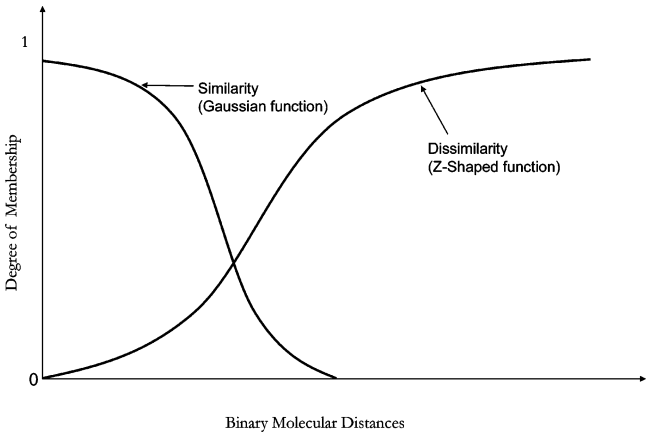

Figure 3. Use of membership functions to describe intermolecular similarity and dissimilarity. Small intermolecular distances correspond to large similarities and low dissimilarities and vice versa.

The proper choice of the fuzzy membership functions can improve the performance of the INFCIM model. The Gaussian membership function (equation 1) was used to describe the degree of similarity between mixture constituents, whereas the Z-membership function (eq 2) used to describe the degree of dissimilarity between the constituents. The two functions are shown in Figure 3.

$$y = e^{(-(x-c)^2/2\sigma^2)} \tag{1}$$

$$y = 1 - \begin{cases} 1 & \text{if} \quad x < x_1 \\ 1 - 2\left(\dfrac{(x - x_1)}{(x_1 - x_0)}\right)^2 & \text{if} \quad x_1 \le x < \dfrac{x_1 + x_0}{2} \\ 2\left(\dfrac{(x_0 - x)}{(x_1 - x_0)}\right)^2 & \text{if} \quad \dfrac{x_1 + x_0}{2} \le x < x_0 \\ 0 & \text{if} \quad x > x_0 \end{cases} \tag{2}$$

In these equations, y is the degree of membership, x is the binary molecular distance, and c the mean), σ is the standard deviation, and x_1 and x_0 are the start and end points of the slope part of the Z-shaped membership function that are the adjustable membership function parameters. These two membership functions have two parameters each that can be adjusted. In this study, "c" in eq 1 and x_1 in eq 2 are set to zero so that a binary molecular distance of zero corresponds to a similarity of 1 and a dissimilarity of 0. Therefore, only one parameter for each membership function needs to be optimized (σ for the Gaussian function and x_0 for the Z-shaped function). As a result, we can use only one

set of data to train and optimize the membership functions, although more can lead to better results. The meaning of the word training here has a major difference from training PLS or neural network based QSAR models for toxicity prediction of pure compounds. In here, only one set of data about the mixture is needed, because only one parameter needs to be optimized through training, while in training QSAR models for toxicity prediction of pure compounds, a large number of training data cases are needed in order to obtain a meaningful model due to the large number of parameters such as neural network weights that have to be determined.

However, the calculated similarity and dissimilarity fuzzy sets (matrices) have to fulfill three properties; reflexivity, symmetry, and transitivity.[77] Given a matrix **X** with elements $x_{1,1}$, ..., $x_{n,n}$, a matrix is reflexive if $x_{i,j} \geq 0$ and $x_{i,j} = 1$ if $i = j$, symmetric if $x_{i,j} = x_{j,i}$ and transitive if $x_{i,j} \geq \vee_k (x_{i,k} \wedge x_{j,k})$, where $\wedge$ and $\vee$ stand for the *min* and *max* operation. A matrix can be easily checked for whether it fulfils the reflexive and symmetric conditions. However, the fuzzy transitive condition is not necessarily satisfied by similarity and dissimilarity matrices. The transitive closures of both these matrices have to be computed so that for any three objects *i*, *j*, and *k*, if object *i* is similar to object *k* and object *k* is similar to object *j*, then object *i* is similar to object *j*. This condition is indispensable for clustering based on fuzzy similarity matrices. Once the transitive closures have been calculated, the bootstrapped mean estimate for each matrix (fuzzy set) are calculated to give an indication of the overall degree of similarity (*intrasimilarity*, α_{sim}) and dissimilarity (*intradissimilarity*, α_{dis}) within the mixture.

6. THE INFCIM MODEL

Following the quantification of the degrees of similarity and dissimilarity within a mixture, the INtegrated Fuzzy Concentration addition-Independent action Modeling (INFCIM) equation (eq 3) can be used to calculate mixture toxicity.

$$EC_{x,mix} = \omega_A \cdot (CA) + \omega_B \cdot (IA) \tag{3}$$

In this equation, $EC_{x,mix}$ is the mixture effect concentration, and coefficients ω_A and ω_B are weightings for the concentration additive and independent action contributions. The weightings are calculated using molecular descriptors and fuzzy membership functions modeling. In practice, they are written in terms of the *intrasimilarity* (α_{sim}) and *intradissimilarity* (α_{dis}). Equation 3 can be rewritten in terms of these newly defined parameters as eq 4. The final values for these parameters are only obtained after the selected membership functions have been optimized so that they give *intrasimilarity* and *intradissimilarity* parameters that lead to a good prediction of the mixture toxicity $EC_{x,mix}$.

$$EC_{x,mix} = \frac{1}{\alpha_{sim}}\left[\frac{1}{\sum_{i=1}^{n}\left(\frac{p_i}{EC_{x,i}}\right)}\right] + \alpha_{dis} \cdot [1 - \prod_{i=1}^{n}(1 - E(c_i))] \tag{4}$$

In this equation, the weightings for the concentration additive and independent active contributions are represented by the reciprocal of the overall *intrasimilarity* ($1/\alpha_{sim}$) and the *intradissimilarity* (α_{dis}), respectively. The reason for placing the *intrasimilarity* (α_{sim}) in the denominator is to set an upper toxicity bound for noninteractive mixtures. This is in accordance to observations from mixture toxicity studies which suggest concentration addition to predict the highest toxicity of noninteractive aquatic mixtures. The equations in the square brackets are that for describing the expected effect according to concentration addition and independent action. In these equations, $EC_{x,i}$ is the effect concentration of a mixture constituent, p_i is the mole fraction of a mixture constituent, x is the percentage effect, $E(c_i)$ is the effect of a mixture constituent at concentrations c_i, and $EC_{x,mix}$ is the mixture effect concentration.

Prerequisites for using INFCIM are knowledge of the full concentration response profiles and molar compositions of mixture constituents. The main feature of this approach is the loop starting with carefully selected membership functions and parameters from which the overall *intrasimilarity* (α_{sim}) and *intradissimilarity* (α_{dis}) are estimated and ending with whether the resulting mixture toxicity predictions are satisfactory or not. If the predictions are not satisfactory, the parameters for the membership functions are updated, and the loop is initiated once again. This process is continued until the best predictions for the mixtures being considered are obtained. However it is important to reiterate that the fuzzy membership functions and parameters can be used to predict toxicity of completely different mixtures as long as the end points are the same.

7. RESULTS

This section presents the results of applying the INFCIM model to the case study mixtures. Model development was achieved using the steps outlined in Figure 1. Once the model was created, mixtures not used in the development were tested for their toxicity.

7.1. Case Study 1. Out of the 1497 descriptors calculated, 1208 were used in this case study after removal of descriptors with constant values. These descriptors were then scaled by their variances before calculating intermolecular distances and subsequently intermolecular similarities and dissimilarities.

In case study 1, two of the mixtures were used to optimize parameters for similarity and dissimilarity membership functions (training), and the remaining two mixtures were used for testing the performance of the resulting INFCIM model. Figure 4a−d shows results of the first case study carried out to mixtures eliciting an effect to green freshwater algae. The two training set mixtures used to determine the similarity and dissimilarity membership function parameters were the EC_1 ratio mixtures (Figure 4a,c). These results were achieved with $\mu = 210$ for the Gaussian similarity membership function and $x_0 = 200$ for the Z-shaped dissimilarity function. These parameters were obtained by manual optimization (trial and error) as opposed to using more elaborate methodologies outlined earlier. As is evident, this did not seemingly reduce the performance of the INFCIM training. Figure 7a shows the prediction error between INFCIM and the experimental results for the *s*-triazine training mixture.

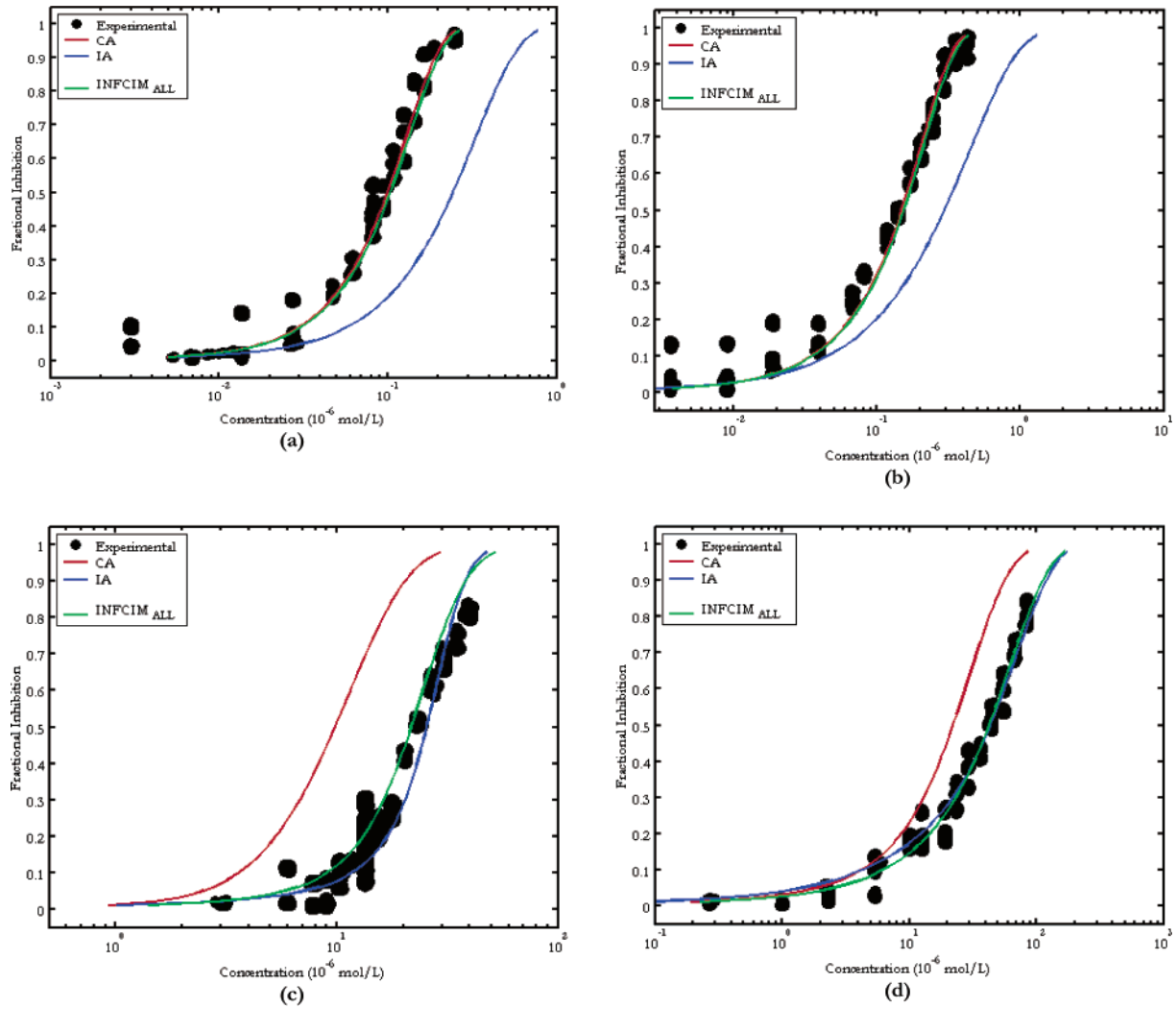

Figure 4. Comparison of concentration addition, independent action, and INFCIM predictions against observed algal toxicity for an *s*-triazine mixture in ratios of its EC_1 (a) and EC_{50} values (b), and for a mixture of dissimilarly acting constituents in ratios of its EC_1 (c) and EC_{50} (d) values using the same similarity and dissimilarity membership functions and parameters to represent mixture constituents characterized by all calculated DRAGON descriptors.

This was less than 18% for all effect levels tested (5−90%) and was typical of the training performance in both case studies.

Using these optimized similarity and dissimilarity membership functions, INFCIM excellently modeled the toxicity of the test mixtures (Figure 4b,d). The difference in EC_{50} values between the experimental data (42.5 μmol/L) and INFCIM prediction (42.3 μmol/L) was less than 1% for the mixture of dissimilarly acting constituents (Figure 4d). This compares very well with the best predicting model in the original study, independent action (40.9 μmol/L), and is better than that of concentration addition. The prediction errors between INFCIM and the experimental results was less than 10% for all effect levels tested (5−90%); see Figure 7b. Prediction errors were greater in the lower effect regions ranging between 10 and 20%. Similar results are observed for the *s*-triazine mixture (Figure 4b). Here, the difference in EC_{50} values between the experimental data (145 nmol/L) and INFCIM prediction (154 nmol/L) was less than 15%. The performance of INFCIM is made more significant when it is taken into consideration that one of the mixtures used in this case study has completely similarly acting constituents, while the other has completely dissimilarly acting mixture constituents. This shows the better general applicability of INFCIM over concentration addition and independent action.

7.2. Case Study 2. The same approach was employed for the second case study. 1183 variance scaled descriptors were used to represent compounds after removal of variables with constant values.

The performance of INFCIM training using EC_{50} and EC_1 ratio mixtures (Figure 5a,d) was once again excellent; matching the performance seen in the previous case study. The range of prediction errors for 5−90% effect levels were all within an order of magnitude. Predicted EC_{50} values for these training mixtures were both within 16% of the experimental data (32.9 μmol/L for the phenol derivative mixture and 0.8315 μmol/L for the quinolone mixture). These results were achieved with $\mu = 170$ for the Gaussian similarity membership function and $x_0 = 600$ for the Z-shaped dissimilarity function.

Using these optimized parameters for similarity and dissimilarity membership functions INFCIM yields excellent results for the test mixtures (Figure 5b,c). The difference in EC_{50} values between INFCIM predictions and experimental data was approximately 11% for both the quinolone mixture (0.582 μmol/L) and phenol derivative mixture (21.35 μmol/

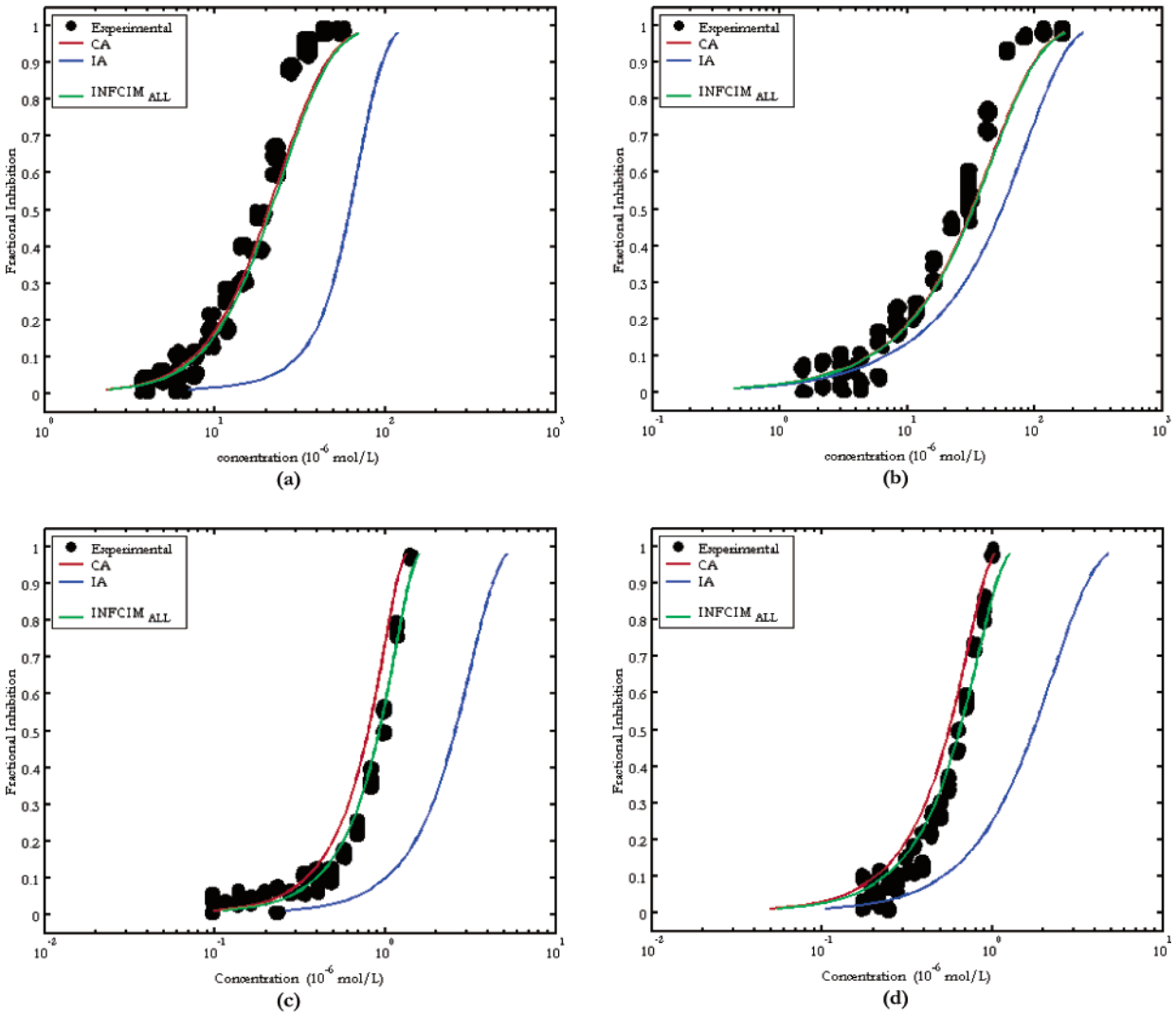

Figure 5. Comparison of concentration addition, independent action and INFCIM predictions against observed *Vibrio fischeri* toxicity for a phenol-derivative mixture in ratios of its EC_1 (a) and EC_{50} values (b), and for a quinolone mixture in ratios of its EC_1 (c) and EC_{50} (d) values using the same similarity and dissimilarity membership functions and parameters to represent mixture constituents characterized by all calculated DRAGON descriptors.

L). The range of difference in predictions was once again within 1 order of magnitude for the quinolone mixture (<1%−<30%) and the phenol derivative mixture for 5−90% effect levels. INFCIM modeled the quinolone mixture data slightly better than in the original study (Figure 5c) and matched the performance of concentration addition in the phenol derivative mixture study (Figure 5b). Independent action was overwhelmingly outperformed in both tests.

Thus, INFCIM has shown that valid predictions for mixtures of completely similarly or completely dissimilarly acting constituents can be made using a common set of membership function parameters that describe molecular similarity and dissimilarity. Thus INFCIM shows a better general applicability than both existing models for predicting mixture toxicity.

7.3. Discussion. The predictability of INFCIM has been demonstrated in both case studies. Figure 4a−d corresponding to case study 1 demonstrates the excellent modeling performance of INFCIM to mixtures at different mole ratios eliciting a toxic effect to green freshwater algae *scenedemus vacuolatus*. The mixtures of completely similar or dissimilar acting constituents were investigated in this case study.

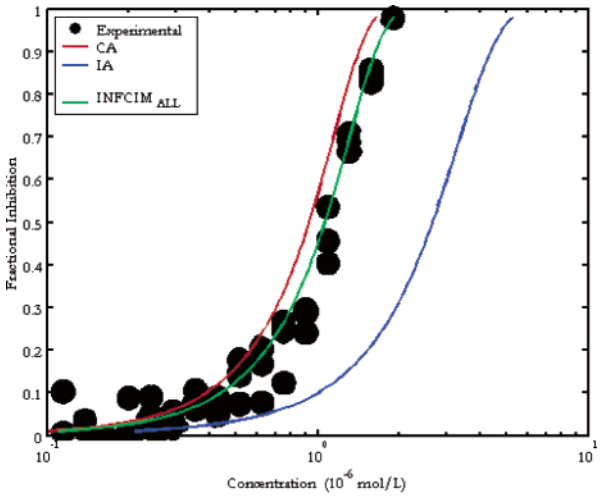

Figure 6. Comparison of concentration addition, independent action and INFCIM predictions against observed *Vibrio fischeri* toxicity for a quinolone mixture in ratios of its NOEC values using the same similarity and dissimilarity membership functions and parameters to represent mixture constituents characterized by all calculated DRAGON descriptors.

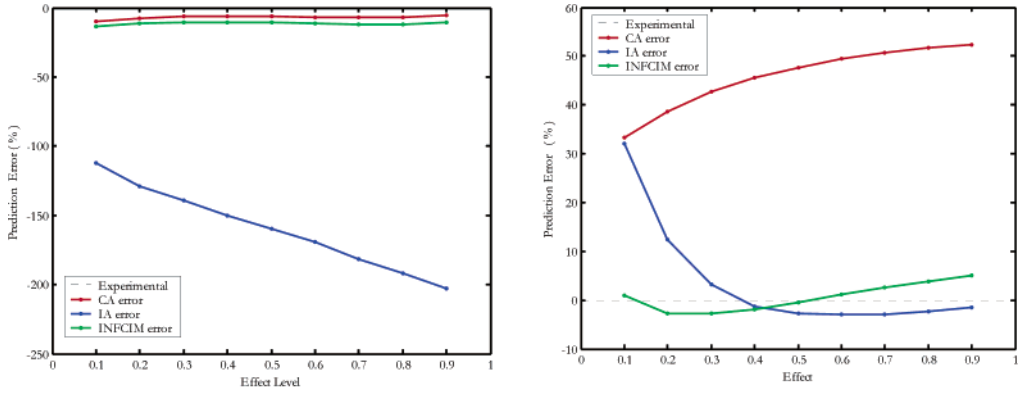

Figure 7. Comparison of the CA, IA, and INFCIM errors of prediction at different effect levels for case study 1. Lines above the experimental dotted line ($y=0$) show overpredictions (underpredictions below the line), and the magnitude of the error is represented by the distance from this line to the prediction line at each level. Comparisons of prediction errors for the *s*-triazine mixture in ratios of EC_1 values are shown in (a), and (b) illustrates those for the mixture of dissimilarly acting agents in ratios of the EC_{50} values.

INFCIM also showed good toxicity modeling performance for mixtures eliciting an effect to the marine bacterium *Vibrio fischeri* (Figure 5a−d).

The use of this approach is best captured in the toxicity modeling of the quinolone mixture. A comparison of predictions by INFCIM, concentration addition, and independent action are shown in Figure 5a,b. It was concluded in the original study that concentration addition best modeled this mixture albeit with a slight overestimation of the toxicity. This observation is similar to results from a study by Hermens et al.[25] on the toxicity of supposedly similarly acting compounds to *Daphnia magna*. In the case of the quinolone mixture, the slight overestimation of mixture toxicity was put down to the molecular mechanism of action not being entirely identical as originally thought. In fact, it was acknowledged that additional mechanisms of action have been proposed for the inhibition of bacterial gyrases. Alternatively, it could be said that there exists a significant degree of dissimilarity within this mixture that could not be captured accurately by concentration addition modeling on its own. INFCIM is able to capture this dissimilarity, in addition to the similarity, resulting in the better modeling of the mixture toxicity. In other words, unlike both existing models for mixture toxicity prediction, INFCIM is capable of taking into account more successfully the differences in the underlying mechanism of action or presence of secondary mechanisms of action that in turn affect the overall similarity and dissimilarity within the mixture. This is shown to be the case for all quinolone mixture ratios tested.

The INFCIM approach proposed in this study also assumes concentration addition to predict the highest mixture toxicity of a noninteractive mixture. This criterion was set using observations from mixture toxicity studies.[19,21−24,28] This leads to interesting observations in case study 2 depicted by Figure 5a,b. Here, concentration addition is found to model the mixture toxicity data very well in the low effect regions but with a slight suggestion of underestimation in the higher regions. INFCIM predictions match those of concentration addition, the best performing model reported in the original study; hence the same conclusions can be drawn about its performance. The phenol derivatives constituting the mixture are ascribed an uncoupling mechanism of action to *Vibrio fischeri*. However, the notion of uncoupling of oxidative phosphorylation as a specific mechanism of action of chemicals in prokaryotic cells has been challenged[17] citing slight evidence of interaction. Hence the toxicity of a mixture comprising phenol derivatives could be slightly more than that predicted by concentration addition. This is similar to toxicity results of a binary mixture of diclofenac and ibruprofen where it was concluded that the excess toxicity was due to the lack of a common specific toxic action.[78] The INFCIM approach proposed is purely for noninteractive mixtures and hence would not capture any excess toxicity caused by the interactive behavior of constituents.

INFCIM also performed very well when modeling mixtures with constituents at or below their NOEC values (Figure 6). This is consistent with results from several studies[8,21,27] and the view that severe toxic effects can result from the joint action of single compounds present at "safe" environmental levels.[28,79] Henceforth, we support the view that NOECs are unsuitable approximations of safe environmental concentrations.

8. FINAL REMARKS

The proposed INFCIM (INtegrated Fuzzy Concentration addition − Independent action Modeling) approach for toxicity prediction of noninteractive mixtures has been demonstrated to perform comparably or better than the best performing existing model for all the four mixtures presented in the two case studies, regardless of mechanism of action of mixture compounds being similar or dissimilar.

The INFCIM approach can be regarded as a QSAR (quantitative structure−activity relationship) approach for mixtures toxicity prediction. QSAR models for single compounds using multivariate data analysis techniques such as PLS and neural networks correlate the structural, physicochemical, and biological descriptors to a toxicity end point. The fact that there are many parameters to be determined in a PLS or neural network model means that a large amount of training data is needed for model development. The difficulty with applying the same method to a mixture is obvious due to the impracticality of testing many compositions of mixtures. The INFCIM method proposed in this paper uses an equation with only a few parameters that need to be determined (for the fuzzy membership functions used

in the paper, only two parameters need to be determined) and can be determined using only one set of mixture data, i.e., the concentration response curve of the mixture for one given composition. The INFCIM model can then be used to predict toxicity of the mixture at any other composition. The case studies in this paper have gone one step further: the trained model has shown to predict with great accuracy the toxicity of mixtures of completely different constituents as long as the toxicity end-point is the same.

Like in single compound QSAR models, molecular descriptors are employed in INFCIM. However, they are used to measure the similarity and dissimilarity between molecules as opposed to correlating them to toxicity values.

Since there is no well accepted conclusion on what constitutes the most appropriate descriptors correlating to toxicity, we have used all the descriptors calculated by the molecular modeling software system DRAGON in this work. In the future, it is worth investigating the use of some data compression techniques such as principal component analysis to preprocess the data before fuzzy characterization of similarity and dissimilarity. In addition, fuzzy membership functions and their parameters were selected and determined manually. Clearly, automatic approaches for the selection of fuzzy membership functions and optimization of parameters should be investigated in future. Furthermore, the approach has to be tested with more mixture data sets and its performance further compared to existing models. Finally, integration of the approach with process simulation tools could provide a means for selecting toxicity minimization strategies, e.g. at process stream level or at the end of the pipe.

ACKNOWLEDGMENT

The first author is funded by a Keyworth PhD scholarship of Keyworth School of Manufacturing and Information Systems of the University of Leeds. The third author is funded by EPSRC Crystal Faraday Partnership on Green Technology, Brixham Environmental Laboratory of Astra-Zeneca, and the Centre of Ecology and Hydrology. The authors would like to thank Daniel Oliver and Mark Earl of Umetrics for contributing their data analysis tool SIMCA-P 8.0 and for providing advice on toxicity modeling using multivariate data analysis methods and QSAR development.

REFERENCES AND NOTES

(1) Smith, D. A. Assessing toxicology quickly and efficiently. *Chem. Eng.* **2000**, *107*(4), 125−128.
(2) Acceryls, *TOPKAT and TSAR*, www.accelrys.com.
(3) DEREK, *www.chem.leeds.ac.uk/luk/derek/*.
(4) MultiCase, *www.multicase.com*.
(5) Feron, V. J.; Groten, J. P. Toxicological evaluation of chemical mixtures. *Food Chem. Toxicol.* **2002**, *40*(6), 825−839.
(6) Feron, V. J.; Cassee, F. R.; Groten, J. P. Toxicology of chemical mixtures: International perspective. *Environ. Health Perspect.* **1998**, *106*, 1281−1289.
(7) ECETOC, *Aquatic Toxicity of Mixtures.* Technical Report, 2001. No. 80: pp 1−64.
(8) Walter, H.; et al. Mixture toxicity of priority pollutants at No Observed Effect Concentrations (NOECs). *Ecotoxicology* **2002**, *11*, 299−310.
(9) Broderius, S. J. Modelling the joint toxicity of xenobiotics to aquatic organisms: Basic concepts and approaches. *Aquatic Toxicol. Risk Assess.* **1991**, *14*, 107−127.
(10) Cassee, F. R.; et al. Toxicological evaluation and risk assessment of chemical mixtures. *Crit. Rev. Toxicol.* **1998**, *28*(1), 73−101.
(11) Groten, J. P.; Feron, V. J.; Suhnel, J. Toxicology of simple and complex mixtures. *Trends Pharmacol. Sci.* **2001**, *22*(6), 316−322.
(12) Pearce, J.; et al. Predicting microbial toxicity of nonuniform multicomponent mixtures of organic chemicals. *J. Environ. Eng.-Asce* **1997**, *123*(4), 329−334.
(13) Bliss, C. I. The toxicity of poisons applied jointly. *Annals Appl. Biol.* **1939**, *26*, 585−615.
(14) Hall, E.; et al. Toxicity of organic chemicals and their mixtures to activated sludge microorganisms. *J. Environ. Eng.-Asce* **1996**, *122*(5), 424−429.
(15) Nirmalakhandan, N.; et al. Additivity in microbial toxicity of nonuniform mixtures of organic chemicals. *Ecotoxicol. Environ. Safety* **1997**, *37*(1), 97−102.
(16) Groten, J. P. Mixtures and interactions. *Food Chem. Toxicol.* **2000**, *38*, S65−S71.
(17) Altenburger, R.; et al. Predictability of the toxicity of multiple chemical mixtures to *Vibrio fischeri*: Mixtures composed of similarly acting chemicals. *Environ. Toxicol. Chem.* **2000**, *19*(9), 2341−2347.
(18) Altenburger, R.; et al. Regulations for combined effects of pollutants: Consequences from risk assessment in aquatic toxicology. *Food Chem. Toxicol.* **1996**, *34*(11−12), 1155−1157.
(19) Backhaus, T.; et al. Predictability of the toxicity of a multiple mixture of dissimilarly acting chemicals to *Vibrio fischeri*. *Environ. Toxicol. Chemistry* **2000**, *19*(9), 2348−2356.
(20) Backhaus, T.; Grimme, L. H. The toxicity of antibiotic agents to the luminescent bacterium *Vibrio fischeri*. *Chemosphere* **1999**, *38*(14), 3291−3301.
(21) Backhaus, T.; Scholze, M.; Grimme, L. H. The single substance and mixture toxicity of quinolones to the bioluminescent bacterium *Vibrio fischeri*. *Aquatic Toxicol.* **2000**, *49*(1−2), 49−61.
(22) Faust, M.; et al. Predicting the joint algal toxicity of multicomponent *s*-triazine mixtures at low-effect concentrations of individual toxicants. *Aquatic Toxicol.* **2001**, *56*, 13−32.
(23) Faust, M.; et al. Joint algal toxicity of 16 dissimilarly acting chemicals is predictable by the concept of independent action. *Aquatic Toxicol.* **2003**, *63*(1), 43−63.
(24) Faust, M.; et al. Predictive assessment of the aquatic toxicity of multiple chemical mixtures. *J. Environ. Qual.* **2000**, *29*(4), 1063−1068.
(25) Hermens, J.; et al. Joint Effects of a Mixture of 14 Chemicals on Mortality and Inhibition of Reproduction of Daphnia-Magna. *Aquatic Toxicol.* **1984**, *5*(4), 315−322.
(26) Rajapakse, N.; Silva, E.; Kortenkamp, A. Combining xenoestrogens at levels below individual No-observed- effect concentrations dramatically enhances steroid hormone action. *Environ. Health Perspect.* **2002**, *110*(9), 917−921.
(27) Silva, E.; Rajapakse, N.; Kortenkamp, A. Something from "nothing" − Eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. *Environ. Sci. Technol.* **2002**, *36*(8), 1751−1756.
(28) Vighi, M.; et al. Water quality objectives for mixtures of toxic chemicals: problems and perspectives. *Ecotoxicol. Environ. Safety* **2003**, *54*(2), 139−150.
(29) Poch, G. Additivity and independence. Basic considerations and new aspects. *Archives Complex Environ. Studies* **1996**, *8*(1−2), 25−31.
(30) Greco, W. R.; Bravo, G.; Parsons, J. C. The search for synergy: A critical review from a response surface perspective. *Pharmacol. Rev.* **1995**, *47*(2), 331−385.
(31) Berebaum, M. C. The expected effect of a combination of agents: the general solution. *J. Theor. Biol.* **1985**, *114*, 413−431.
(32) Chang, C. S.; Xu, D. Y.; Quek, H. B. Pareto-optimal set based multiobjective tuning of fuzzy automatic train operation for mass transit system. *IEE Proc.-Electric Power Applications* **1999**, *146*(5), 577−583.
(33) Marin-Blazquez, J. G.; Shen, Q.; Tuson, A. Tuning fuzzy membership functions with neighbourhood search techniques: a comparative study. http://www.inf.ed.ac.uk/people/staff/Qiang_Shen.html.
(34) Escher, B. I.; Hermens, J. L. M. Modes of action in ecotoxicology: Their role in body burdens, species sensitivity, QSARs, and mixture effects. *Environ. Sci. Technol.* **2002**, *36*(20), 4201−4217.
(35) Cronin, M. T. D.; Schultz, T. W. Validation of *Vibrio fisheri* acute toxicity data: mechanism of action-based QSARs for nonpolar narcotics and polar narcotic phenols. *Sci. Total Environ.* **1997**, *204*(1), 75−88.
(36) Dearden, J. C.; et al. QSAR studies of compounds acting by polar and nonpolar narcosis: An examination of the role of polarisability and hydrogen bonding. *Quant. Struct.−Act. Relat.* **2000**, *19*(1), 3−9.
(37) Konemann, H. Quantitative Structure−Activity-Relationships in Fish Toxicity Studies. 1. Relationship for 50 Industrial Pollutants. *Toxicology* **1981**, *19*(3), 209−221.
(38) Ramos, E. U.; et al. Quantitative structure−activity relationships for the aquatic toxicity of polar and nonpolar narcotic pollutants. *J. Chem. Inf. Comput. Sci.* **1998**, *38*(5), 845−852.

(39) Ramos, E. U.; et al. Acute toxicity of polar narcotics to three aquatic species (*Daphnia magna*, *Poecilia reticulata* and *Lymnaea stagnalis*) and its relation to hydrophobicity. *Chemosphere* **1998**, *37*(4), 633−650.
(40) Vaes, W. H. J.; et al. Understanding and estimating membrane/water partition coefficients: Approaches to derive quantitative structure property relationships. *Chem. Res. Toxicol.* **1998**, *11*(8), 847−854.
(41) Vaes, W. H. J.; et al. Acute toxicity of nonpolar versus polar narcosis: Is there a difference? *Environ. Toxicol. Chem.* **1998**, *17*(7), 1380−1384.
(42) Verhaar, H. J. M.; Vanleeuwen, C. J.; Hermens, J. L. M. Classifying Environmental-Pollutants. 1. Structure−Activity- Relationships for Prediction of Aquatic Toxicity. *Chemosphere* **1992**, *25*(4), 471−491.
(43) Verhaar, H. J. M.; Ramos, E. U.; Hermens, J. L. M. Classifying Environ. pollutants 0.2. Separation of class 1 (baseline toxicity) and class 2 ('polar narcosis') type compounds based on chemical descriptors. *J. Chemometrics* **1996**, *10*(2), 149−162.
(44) Zhao, Y. H.; Cronin, M. T. D.; Dearden, J. C. Quantitative structure−activity relationships of chemicals acting by nonpolar narcosis − Theoretical considerations. *Quant. Struct.−Act. Relat.* **1998**, *17*(2), 131−138.
(45) Ren, S. J.; Schultz, T. W. Identifying the mechanism of aquatic toxicity of selected compounds by hydrophobicity and electrophilicity descriptors. *Toxicol. Lett.* **2002**, *129*(1−2), 151−160.
(46) Gonzalez-Arjona, D.; Lopez-Perez, G.; Gonzalez, A. G. Nonlinear QSAR modeling by using multilayer perceptron feedforward neural networks trained by back-propagation. *Talanta* **2002**, *56*(1), 79−90.
(47) Wang, X.-D.; Yin, C.-S.; Wang, L.-S. Structure activity relationships and response surface analysis of nitroaromatics toxicity to the yeast (saccharomyces cerevisiae). *Chemosphere* **2002**, *46*, 1045−1051.
(48) Wang, X. Z.; et al. Acute toxicity of substituted phenols to rana japonica tadpoles and mechanism based quantitative structure activity relationship (QSAR) study. *Chemosphere* **2001**, *44*, 447−455.
(49) Patel, H.; Berge, W.; Cronin, M. T. D. Quantitative structure activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. *Chemosphere* **2002**, *48*, 603−613.
(50) Gini, G.; et al. Predictive carcinogenicity: A model for aromatic compounds, with nitrogen-containing substituents, based on molecular descriptors using an artificial neural network. *J. Chem. Inf. Comput. Sci.* **1999**, *39*(6), 1076−1080.
(51) Cronin, M. T. D.; Schultz, T. W. Development of quantitative structure−activity relationships for the toxicity of aromatic compounds to Tetrahymena pyriformis: Comparative assessment of the methodologies. *Chem. Res. Toxicol.* **2001**, *14*(9), 1284−1295.
(52) Benigni, R.; et al. Quantitative structure−activity relationships of mutagenic and carcinogenic aromatic amines. *Chem. Rev.* **2000**, *100*(10), 3697−3714.
(53) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*(3), 1027−1043.
(54) Karelson, M. *Molecular Descriptors in QSAR/QSPR*; 2000; p 448.
(55) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Mannhold, R., Kubinyi, H., Timmerman, H. Eds.; 2000; p 667.
(56) Todeschini, R.; et al. *DRAGON 3.0 Web Version*; 2003. www.disat-.unimib.it/chm/.
(57) Schrodinger, *MOPAC*, www.schrodinger.com.
(58) Semichem, *CODESSA and AMPAC*, www.codessa.com.
(59) Cronin, M. T. D.; Schultz, T. W. Pitfalls in QSAR. *J. Mol. Struct. (THEOCHEM)* **2003**, *622*, 39−51.
(60) Randic, M. On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.
(61) Maran, U.; Karelson, M.; Katritzky, A. R. A comprehensive QSAR treatment of the genotoxicity of heteroaromatic and aromatic amines. *Quant. Struct.−Act. Relat.* **1999**, *18*, 3−10.
(62) Katritzky, A. R.; Tatham, D. B.; Maran, U. Theoretical Descriptors for the Correlation of Aquatic Toxicity of Environmental Pollutants by Quantitative Structure-Toxicity Relationships. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1162−1176.
(63) Feng, J.; et al. Predictive toxicology: benchmarking molecular descriptors and statistical methods. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1463−1470.
(64) Basak, S. C.; Bertelsen, S.; Grunwald, G. D. Application of graph theoretical parameters in quantifying molecular similarity and structure−activity relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 270−276.
(65) Basak, S. C.; Grunwald, G. D. Molecular similarity and estimation of molecular properties. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 366−372.
(66) Basak, S. C.; et al. Quantitative molecular similarity methods in the property/toxicity estimation of chemicals: a comparison of arbitrary versus tailored similarity spaces. *J. Mol. Struct. (THEOCHEM)* **2003**, *622*, 127−145.
(67) Xu, L.; Yang, J.-A.; Wu, Y.-P. Effective descriptions of molecular structures and the quantitative structure−activity relationship studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 602−606.
(68) Livingstone, D. J. The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 195−209.
(69) Benigni, R.; et al. On the equivalence between different descriptors of molecules: value for computational approaches. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 575−578.
(70) Bajorath, J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 233−245.
(71) Consonni, V.; et al. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*(3), 693−705.
(72) Di Marzio, W.; et al. Traditional versus WHIM molecular descriptors in QSAR approaches applied to fish toxicity studies. *Chemosphere* **2001**, *44*(3), 401−406.
(73) Gramatica, P.; et al. QSAR approach for the selection of congeneric compounds with a similar toxicological mode of action. *Chemosphere* **2001**, *42*(8), 873−883.
(74) Gramatica, P.; Di Guardo, A. Screening of pesticides for environmental partitioning tendency. *Chemosphere* **2002**, *47*(9), 947−956.
(75) Hypercube, *HyerChem*. 2002, www.hypercube.com.
(76) Yeung, D. S.; Wang, X. Z. Improving performance of similarity-based clustering by feature weight learning. *IEEE Trans. Pattern Analysis Mach. Intell.* **2002**, *24*(4), 556−561.
(77) Yeung, D. S.; Wang, X. Z. Using a neuro-fuzzy technique to improve the clustering based on similarity. *IEEE Trans. Pattern Analysis Mach. Intell.* **2000**.
(78) Cleuvers, M. Aquatic ecotoxicity of pharmaceuticals including the assessment of combination effects. *Toxicol. Lett.* **2003**, *142*, 185−194.
(79) Kortenkamp, A.; R. Altenburger Approaches to assessing combination effects of oestrogenic environmental pollutants. *Sci. Total Environ.* **1999**, *233*(1−3), 131−140.

CI0499368