Posetic Quantitative Superstructure/Activity Relationships (QSSARs) for Chlorobenzenes

Teodora Ivanciuc,*,† Ovidiu Ivanciuc,‡ and Douglas J. Klein*,†

Department of Marine Sciences, Texas A&M University, Galveston, Texas 77551, and Department of Human Biology Chemistry and Genetics, University of Texas Medical Branch, Galveston, Texas 77555

Received April 14, 2005

As a result of the widespread industrial use of polychlorinated hydrocarbons, they have accumulated in nearly all types of environmental compartments, especially in aquatic systems. Particularly, chloroaromatics are among the most undesirable industrial effluents because of their persistence and toxicity. To predict chlorobenzene (CB) toxicities, we make use of a novel scheme that looks beyond simple molecular structure to the manner in which such a structure embeds in an overall reaction network. Thence, a resultant modeling gives a quantitative superstructure/activity relationship (QSSAR) with the (chloro-substitution) reaction network viewed mathematically as a partially ordered set (or poset). Different numerical fittings to the overall poset lead to different QSSAR models, of which we investigate three: average poset, cluster expansion, and splinoid poset QSSAR models for the CBs' toxicities against various species (*Poecilia reticulata*, *Pimephales promelas*, *Daphnia magna*, *Rana japonica*, etc). Excellent results are obtained for all QSSAR toxicity models. On the basis of the poset reaction diagram, all three of these QSSAR models reflect, in distinct ways, the topology of the network that describes the interconversion of chemical species. Although in the majority of investigated datasets all poset QSSAR models give very good predictions, in some cases, they complement each other. These differences show that more reliable predictions can be obtained by using a consensus prediction that combines data from the three posetic models.

1. INTRODUCTION

The large number of organic pollutants is a direct result of the increasing use of toxic chemicals, such as herbicides, fungicides, pesticides, industrial solvents, or petroleum products. The environmental hazards of organic pollutants can be assessed by collecting information on their distribution and persistence among various compartments of the natural environment[1]. As a response to the growing concern over the potential hazards of environmental pollution, the Environmental Protection Agency (EPA) (http://www.epa.gov/) has labeled several chemicals as "potentially hazardous" and has initiated investigations of their toxicities. Chlorinated compounds are widely used in the chemical industry as cooling agents, additives, and solvents; for insecticides and herbicides; for the synthesis of PVC and other polymers; and for other organic syntheses. Their environmental toxicity and persistence ranks them among the most undesirable industrial effluents. For example, polychlorinated monoaromatic hydrocarbons have accumulated in many environmental compartments, especially in aquatic systems.[2−10]

As a result of the large number of chemical compounds that are released in the environment coupled with the significant costs of testing the toxicity of these chemicals against a high number of species, there is considerable interest in the prediction of toxicological activities from molecular structure. Various quantitative structure/activity relationship (QSAR) models have been developed to predict the toxicity using various structural descriptors, taken as simple "topological indices",[10−19] more computationally expensive quantum chemical descriptors,[6,20−26] or relevant experimentally measured properties.[5,27−31] Also, there are similarity-based approaches[32,33] making comparisons between pairs of structures either in terms of maximal common substructures[34−37] or in terms of the values of a list of considered descriptors.[38−41]

Here, we use another predictive scheme based on similarity comparisons to the corresponding activities of related structures. But unlike these other similarity schemes, the comparisons we make are with reference to a "superstructure" containing a special collection of structures. The special superstructure here considered is a substitution-reaction network, which for the chlorination of the benzene skeleton appears as in Figure 1. Starting from an unsubstituted compound, substituents are progressively introduced one after another, with earlier substituents fixed at their different possible positions. When substituents are chlorine atoms, the diagram starts with benzene at the top and ends with hexachlorobenzene at the bottom, while all the different patterns of substitution occur in between. The arrows indicate the hierarchic generation of the different patterns of more substituted compounds from the different patterns of less substituted ones. From this diagram, it is easy to recognize that from *para*-chlorobenzene, only a single trichlorobenzene is obtained, whereas from *ortho*-chlorobenzene, two trichlorobenzenes are obtained, and from *meta*-chlorobenzene, all three trichlorobenzenes may be obtained.

* Corresponding author e-mail: ivanciuc@netscape.net (T.I.) and kleind@tamug.edu (D.J.K.).
† Texas A&M University.
‡ University of Texas Medical Branch.

10.1021/ci0501342 CCC: $30.25
Published on Web 06/28/2005

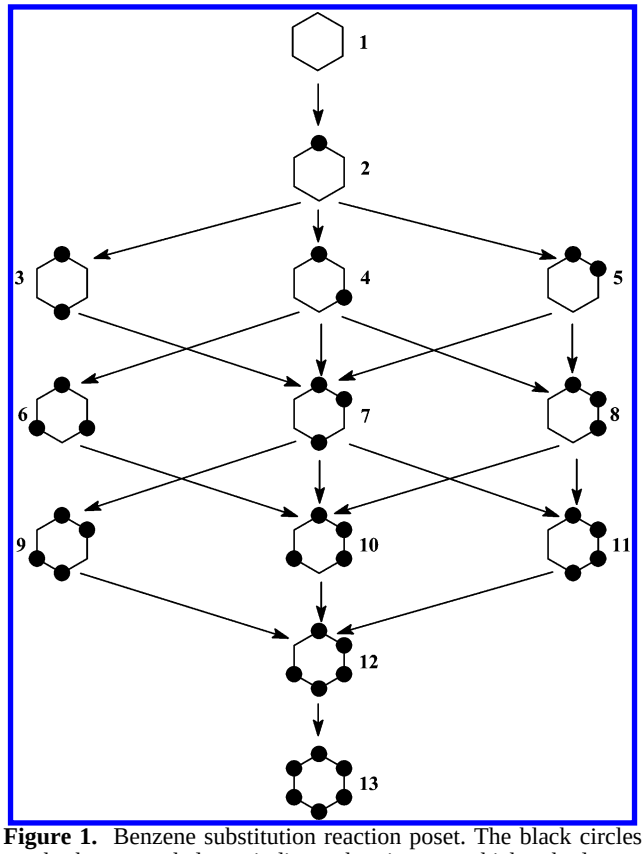

Figure 1. Benzene substitution reaction poset. The black circles on the benzene skeleton indicate the sites on which a hydrogen atom has been replaced by a chlorine atom.

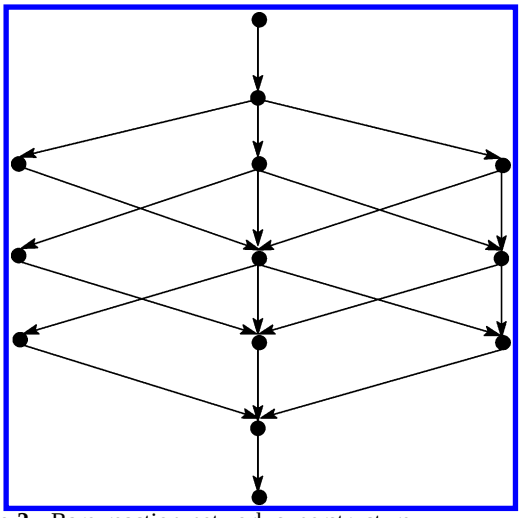

Figure 2. Bare reaction-network superstructure.

Clearly, Figure 1 incorporates a relation between molecular structure and reaction network. Correlations of molecular properties are virtually always addressed in terms of molecular structure without reference to the reaction network. Thus, it is natural to explore the complementary approach developing correlations of properties using the reaction network without reference to the molecular structure. That is, the network of Figure 2, stripped of explicit reference to molecular structure, is a natural candidate for study to correlate chemical properties. Granted success in this endeavor, and thereby coming to understand the use of such networks as in Figure 2, it would then be natural to explore the simultaneous use of molecular structure and the reaction network (or superstructure).

Thence, here quantitative superstructure/activity relationship (QSSAR) models are explored to make fits to this network of Figure 2 and so predict missing activities via interpolation, or extrapolation.

Many reaction networks, and in particular, substitution-reaction networks such as that described here, have[42−44] the mathematical structure of a partially ordered set (or poset). That is, first, if structure α leads to structure β via (successive) substitution, then α and β are distinct and ordered $\alpha \rightsquigarrow \beta$; second, if $\alpha \rightsquigarrow \beta$ and $\beta \rightsquigarrow \gamma$, then $\alpha \rightsquigarrow \gamma$. Thence, the QSSAR models are to be fittings to such posets, with three such models available: a average poset model,[45] a splinoid-poset model,[46] and a cluster expansion model.[47−50] Here, each of these QSSAR models is applied for chlorobenzene (CB) toxicities of several species (*Poecilia reticulata*, *Pimephales promelas*, *Daphnia magna*, *Rana japonica*, etc).

2. EXPERIMENTAL DATA AND USES OF CHLOROBENZENES

2.1. Principal Uses of CBs. CBs are omnipresent in the environment with all 12 isomers (one mono-, three di-, three tri-, three tetra-, one penta-, and one hexachlorobenzene) reported to occur in coastal marine sediments, freshwater lake sediments, soils, sewage sludge, wastewater, groundwater, rivers, and Herring gulls and their eggs in the Great Lakes.[51−60] CBs may accumulate in the food chain because of their lipophilic character and slow degradation in the environment.

CBs are metabolized by various microorganisms involved in the global cycle of transforming halogenated organic compounds. The less chlorinated benzenes are amenable to oxidative degradation, and organisms that catalyze their aerobic degradation have been isolated. Tetrachlorobenzenes, pentachlorobenzene, and hexachlorobenzene have a higher resistance to aerobic degradation but are susceptible to anaerobic dechlorination.[61]

Monochlorobenzene is used primarily as a solvent, degreasing agent, and chemical intermediate in various syntheses. Acute inhalation exposure of animals to monochlorobenzene produces narcosis, restlessness, tremors, and muscle spasms. Dichlorobenzenes are widely used for industrial and domestic purposes such as a moth repellent, heat transfer medium, and toilet deodorant (1,4-dichlorobenzene) and as paint removers, engine cleaners, deinking solvents, fuel additives, coolants, and a sewer and septic tank cleaner (1,2-dichlorobenzene). 1,2-Dichlorobenzene is also used as a solvent in the production of azo-dyes, pharmaceuticals, and pesticides. 1,2,4-Trichlorobenzene is used as a coolant in electrical equipment and in glass tempering. Trichlorobenzenes are likely to accumulate in soil and water, and the vapors of 1,2,4-trichlorobenzene are eye and respiratory tract irritants. Tetrachlorobenzenes have minor industrial use, mainly as fire retardants. The principal use of 1,2,4,5-tetrachlorobenzene is in the synthesis of the herbicide 2,4,5-T. Pentachlorobenzene has no significant industrial uses, but it is found as an impurity in the pesticide pentachloronitrobenzene. Hexachlorobenzene is a hydrophobic pollutant included in the priority pollutant lists of the European Union (EU) and the U.S. Environmental Protection Agency. It has been widely used as a fungicidal dressing of seed grains. On the basis of animal studies that have reported cancer of

Table 1. Chlorobenzene (CB) Toxicity Datasets for Various Species

dataset		property[a]	ratio[b]	references
1	log 1/LC_{50} (μmol/l)	the lethal concn. of CBs that reduces the fish population by 50% for guppy (*Poecilia reticulata*)	12/13	*Chemosphere* **1992**, *25*, 471.
2	log 1/LC_{50} (mol/l)	the lethal concn. of CBs that reduces the fish population by 50% for guppy (*Poecilia reticulata*)	11/13	*SAR QSAR Environ. Res*. **2003**, *14*, 113. *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 559.
3	log 1/LC_{50} (mmol/l)	the lethal concn. of CBs that reduces the fish population by 50% for guppy (*Poecilia reticulata*) after 7−14 d exposure	12/13	*Chemosphere* **1983**, *12*, 1421.
4	log 1/LC_{50} (μmol/l)	the lethal concn. of CBs that reduces the fish population by 50% for guppy (*Poecilia reticulata*) after 14 d exposure	12/13	*J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1162. *Toxicology* **1981**, *19*, 209.
5	log 1/LC_{50} (mol/l)	aqueous lethal concn. for CBs that causes 50% mortality in the fathead minnow (*Pimephales promelas*) after 96 h exposure	11/13	*Quant. Struct.-Act. Relat*. **2000**, *19*, 3. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 885. *THEOCHEM* **2002**, *578*, 129.
6	log 1/LC_{50} (mmol/l)	aqueous lethal concn. for CBs that causes 50% mortality in the fathead minnow (*Pimephales promelas*) after 96 h exposure	9/13	*Environ. Sci. Technol*. **2004,** *38*, 3659.
7	log 1/EC_{50} (mol/l)	the effective concn. of CBs that causes 50% mortality in the fathead minnow (*Pimephales promelas*) after 96 h exposure	9/13	*Chemosphere* **1993**, *26*, 1971. *Quant. Struct.-Act. Relat*. **1998**, *17*, 131. *Quant. Struct.-Act. Relat*. **2000**, *19*, 3.
8	log 1/EC_{50} (mol/l)	the effective concn. of CBs that reduces the population by 50% for brine shrimp (*Artemia spp)* after 15 min exposure	7/13	*Chemosphere* **1993**, *26*, 1971.
9	log1/LC_{50} (mol/l)	aqueous lethal concn. of CBs that reduces the population by 50% for brine shrimp (*Artemia spp)* after 24h exposure	7/13	*Chemosphere* **1993**, *26*, 1971.
10	log 1/IC_{50} (mol/l)	24 h immobilization concentrations (IC_{50}) of CBs to *Daphnia magna*	7/13	*Quant. Struct.-Act. Relat*. **2000**, *19*, 3. *Quant. Struct.-Act. Relat*. **1998**, *17*, 131.
11	log 1/IC_{50} (μmol/l)	48 h immobilization concentrations (IC_{50}) of CBs to *Daphnia magna*	13/13	*Aquat. Toxicol*. **1988**, *12*, 33.
12	log 1/EC_{50} (mmol/l)	the effective concn. of CBs that reduces 50% of the primary productivity of a freshwater green alga (*Ankistrodesmus falcatus*)	13/13	*Chemosphere* **1984**, *13*, 991.
13	log 1/EC_{50} (mmol/l)	the effective concn. of CBs that reduces 50% of grow inhibition to algae (*Chlorophyceans*) after 96 h exposure	9/13	*Ecotoxicol. Environ. Saf.* **2001**, *49*, 206.
14	log 1/LC_{50} (mol/l)	acute lethal concn. of CBs to *Rana japonica* tadpoles after 12 h exposure	6/13	*Chemosphere* **2003**, *53*, 963.
15	log BCF	bioconcentration factor to fish	13/13	*Chemosphere* **2000**, *41*, 1675. *QSAR Comb. Sci*. **2003**, *22*, 374.
16	log BCF	bioconcentration factor to fish	12/13	*Chemosphere* **1999**, *39*, 987.
17	log BCF	bioconcentration factor to fish	13/13	*Chemosphere* **1988**, *17*, 21. *Environ. Sci. Techn*ol. **1974**, *8*, 1113.
18	log BCF	bioconcentration factor for *Oncorhynchus mykiss*	11/13	*Chemosphere* **1996**, *33*, 1047.

[a] LC_{50} (U) denotes the concentration (concn.) in units U for 50% lethality. [b] Ratio between the number of experimental data over the total number of data (compounds; vertices in poset).

the liver, thyroid, and kidney from exposure to hexachlorobenzene, the EPA has classified hexachlorobenzene as a probable human carcinogen.[62]

2.2. Experimental Data. An exhaustive literature survey has been carried out to collect data for the toxicity of CBs to different species. When available, we have used in our QSSAR models experimental data originating from a single laboratory, and when not possible, a careful choice was made considering the literature. The collection of toxicity data for CBs used in this study is presented in Table 1.

3. POSETIC METHODOLOGY

3.1. Posetic Applications in General. Partially ordered sets (or posets) have a long standing in mathematics,[63] with Birkhoff saying, "The world around us abounds with examples of partly ordered sets." They have been advocated as of very general utility in chemistry,[42,43] with a brief listing of the many chemical applications given in ref 44. And also, especially Brüggemann and co-workers[64−70] have advocated their use as an attractive way of handling complex information within the environmental area, and the idea has gained other advocates.[70−74] Here, the use of reaction-network posets is a development from ref 49 and follows most especially ref 45.

Again, but more formally, a poset consists of a set P with a relation $\rightsquigarrow$, which satisfies two conditions: first, for α, $\beta \in$ P and $\alpha \rightsquigarrow \beta \Rightarrow$ not $\beta \rightsquigarrow \alpha$; and second, for α and β, $\gamma \in$ P, $\alpha \rightsquigarrow \beta$, and $\beta \rightsquigarrow \gamma \Rightarrow \alpha \rightsquigarrow \gamma$. Here, our set P consists of chemical compounds and the ordering $\alpha \rightsquigarrow \beta$ means that β is obtainable from α after some (nonzero) number of chlorinations. The relation that allows either $\alpha \rightsquigarrow \beta$ or $\alpha = \beta$ is denoted $\alpha \succcurlyeq \beta$, and the relation where $\alpha \rightsquigarrow \beta$ without any intervening members of P is denoted $\alpha \rightarrow \beta$, and in mathematical language, one says α covers β.

3.2. QSSAR Models. Here, the general idea is to utilize the reaction network (or superstructure) as a whole to fit activities (or properties) for the set of individual nodes (each corresponding to a molecular structure). The case of the substitution-reaction network for CBs provides a nice test bed for such proposed models in that frequently activities for many members of the network are known. Here, our QSSAR models are evaluated by comparison of their leave-one-out (LOO) cross-validation statistics: correlation coefficient r and standard deviation s. Our simplest (average poset) model is especially complicit with the LOO prediction, and we here also adopt this cross-validation technique for our splinoid poset and cluster expansion models in order to make comparable predictions. Following our earlier work,[45] introducing and testing the simple average poset QSSAR model, we deal with the chlorination substitution network for benzene. We next briefly describe each of the three posetic QSSAR models.

3.3. Average Poset Model. The average poset method[45] starts from the Hasse diagram of the substitution reaction in which each node represents a chemical compound (Figure 1) that can be obtained from the chemicals from the previous level that send incoming arrows and generates the compounds from the next level that receive the outgoing arrows. Each vertex in the Hasse diagram may be viewed to contain the property value for the corresponding compound. The predicted value $X(B)_{\text{pred}}$ for a property X and a compound B is computed as the average of two averages, namely, the average of experimental values $X(A)_{\text{exp}}$ for all compounds A from the previous level that send incoming arrows toward B and the average of experimental values $X(C)_{\text{exp}}$ for all compounds C from the next level that receive outgoing arrows from B. The experimental property values must be available for all diagram positions adjacent to B. For example, in Figure 3, we present the reaction poset diagram for dataset 1 in Table 1, in which each vertex (compound) has attached the experimental value for log $1/LC_{50}$ for the guppy (*Poecilia reticulata*). The average poset log $1/LC_{50}$ predicted value for 1,2,4-Cl_3Bz is computed with the formula

$$\log 1/LC_{50}(1,2,4\text{-}Cl_3Bz) = \frac{1}{2}\left\{\frac{1}{3}[\log 1/LC_{50}(1,4\text{-}Cl_2Bz) + \log 1/LC_{50}(1,3\text{-}Cl_2Bz) + \log 1/LC_{50}(1,2\text{-}Cl_2Bz)] + \frac{1}{3}[\log 1/LC_{50}(1,2,4,5\text{-}Cl_4Bz) + \log 1/LC_{50}(1,2,3,5\text{-}Cl_4Bz) + \log 1/LC_{50}(1,2,3,4\text{-}Cl_4Bz)]\right\} = \frac{1}{2}\left\{\frac{1}{3}[3.444 + 3.525 + 3.433] + \frac{1}{3}[4.604 + 4.658 + 4.635]\right\} = 4.050$$

Incidentally, for this compound, the experimental and predicted values are identical. As one can see from this example, the properties computed with the average poset method are parameter-free predictions, and the statistical indices obtained are equivalent with LOO statistics.

This scheme is limited to interpolation, though one might imagine extending it to take into account next-nearest neighbor (or more distant) positions in the reaction network. If such were done, something more akin to extrapolation would arise and one might be able to more readily deal with cases where neighbor values are not known. Indeed, it may be argued that the next QSSAR model represents such an extension, in its full generality. Conversely, the splinoid poset method of the next section can be arranged upon specialization to give the average poset method. That is, if the splinoid poset method is constrained to deal solely with the subposet consisting of an element $\xi \in P$ that is immediately connected to ξ, then the average poset result is obtained. Of course, one may imagine many further averaging schemes, but those of this and the next model utilize natural choices for the minimal and maximal amounts of information used within the splinoid fitting procedure. Some discussion of other averaging procedures is found in refs 69, 70, and 75, as well as our earlier use indicated in ref 45.

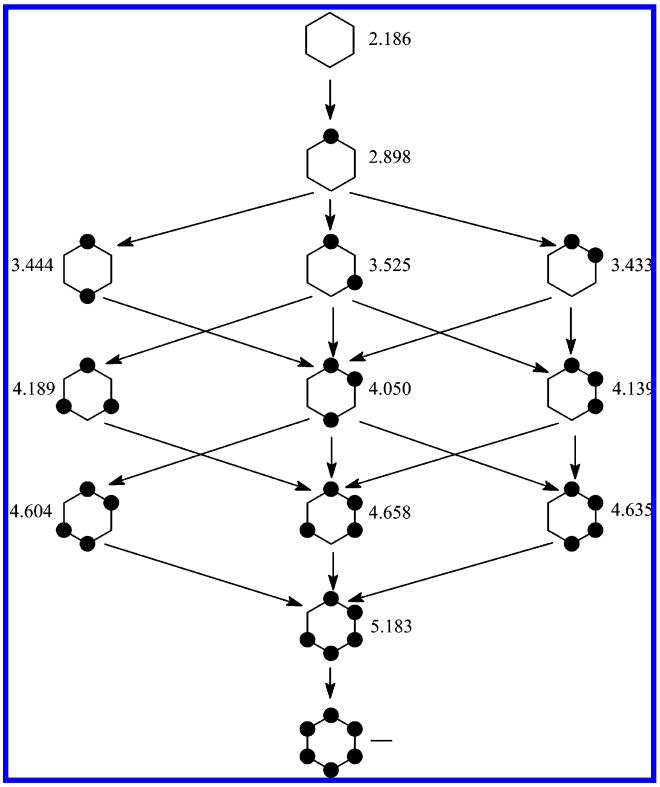

Figure 3. Posetic presentation of CBs' acute aquatic toxicity to the guppy, *Poecilia reticulata*, (log $1/LC_{50}$, [μmol/l], dataset 1).

3.4. Splinoid Poset Model. The chloro-substitution network of benzene is represented here as a Hasse diagram H(P) (Figure 2), which mathematically represents a finite poset P. An oriented edge in the Hasse diagram here represents the transition between a chemical compound α with n chlorine atoms to one β with $n + 1$ chlorine atoms and is denoted by $\alpha \rightarrow \beta$, and we attach a real variable $x_{\alpha\rightarrow\beta}$ ranging from 0 to 1, in proceeding from α to β. When formulating the QSSAR model[46] for a property X, we consider cubic spline polynomials on the oriented edges $\alpha \rightarrow \beta$ of the Hasse diagram H(P):

$$f_{\alpha\rightarrow\beta}(x_{\alpha\rightarrow\beta}) = a_{\alpha\rightarrow\beta}x_{\alpha\rightarrow\beta}^{\ 3} + b_{\alpha\rightarrow\beta}x_{\alpha\rightarrow\beta}^{\ 2} + c_{\alpha\rightarrow\beta}x_{\alpha\rightarrow\beta} + d_{\alpha\rightarrow\beta}$$

with $a_{\alpha\rightarrow\beta}$, $b_{\alpha\rightarrow\beta}$, $c_{\alpha\rightarrow\beta}$, and $d_{\alpha\rightarrow\beta}$ as constants. Each vertex α of H(P) or P is identified by a value a_α and a slope b_β. The splinoid poset QSSAR model is generated on the basis of known values of the property X for a subset of the chemical compounds, namely, for vertices $\alpha \in K \subseteq P$. The splinoid fit

is such that, first, the cubic splines match values a_α at the nodes $\alpha \in P$, with the a_α for $\alpha \in K$ equal to the known property values; second, the slopes b_α through the nodes match; and third, a relevant total "curvature" is minimized. With the splinoid QSSAR determined for the vertices from K, one can predict the property values for the remaining chemical compounds that do not have an experimental value for the property X, compounds that form the set Y of vertices $\alpha \notin K$. An algorithm results for predicting the values of X for the set Y of chemical compounds. Let **A** denote the adjacency matrix of the Hasse diagram H(P), and let **S** denote the oriented adjacency matrix of H(P), where

$$S_{\alpha\beta} = \begin{cases} 1 & \text{if } \beta \rightarrow \alpha \\ -1 & \text{if } \alpha \rightarrow \beta \\ 0 & \text{otherwise} \end{cases}$$

The in-degree on vertex $\alpha \in P$ is denoted by $d_{\rightarrow\alpha}$, and the out-degree on vertex $\alpha \in P$ is denoted by $d_{\alpha\rightarrow}$. On the basis of this notation, we introduce the following two diagonal matrices:

$$\mathbf{D} = \mathrm{diag}[d_{\alpha\rightarrow} - d_{\rightarrow\alpha}]$$

$$\Delta = \mathrm{diag}[d_{\alpha\rightarrow} + d_{\rightarrow\alpha}]$$

We define the matrices **U** (the $|Y| \times |P|$ submatrix of the unity matrix **I**, with rows indexed by the elements of Y) and **K** (the $|K| \times |P|$ submatrix of the unity matrix **I**, with rows indexed by the elements of K) and the derived matrix

$$\mathbf{M} = 2(\Delta - \mathbf{A}) - 3(\mathbf{D} - \mathbf{S})(\mathbf{A} + 2\Delta)^{-1}(\mathbf{D} + \mathbf{S})$$

Then, the vector $\vec{u}$ that contains the unknown values α_2 of the property is computed from the vector $\vec{k}$ of known values via

$$\vec{u} = -(\mathbf{UMU}^t)^{-1}(\mathbf{UMK}^t)\vec{k}$$

For a few different reaction networks, we have studied the matrix $\mathbf{UMU}^t$, and it appears, in practice, to be invertible regardless of how sparse the "known" data is in the network up to the point that very few (≤ 2) known data are available. The a, b, c, and d coefficients appearing in the spline polynomials f do not explicitly appear in our splinoid formula for $\vec{u}$, but they are complicit in the derivation of this formula for $\vec{u}$. If so desired, these coefficients could be computed, as indicated in ref 46. The present formula gives $\vec{u}$ in terms of the poset and, thence, completes the splinoid QSSAR algorithm, which turns out to give a robust model in accommodating the missing values for several compounds (which may possibly even be adjacent). This is a significant advantage of the splinoid model, which uses the topology of the Hasse diagram to generate a response web for the investigated property. To be comparable with the results from the other poset QSSAR models, we have used the splinoid model in the LOO cross-validation procedure.

3.5. Cluster Expansion Model. Formal cluster expansions, in general, re-express a scalar function (or property) for the different members of a poset in terms of related functions focusing more strongly on earlier members of the poset. Much of the formal theory is described by Rota[76] for general posets, and its chemical application in the case that the partial ordering is the subgraph partial ordering is described in refs 47−50. Generally, for a scalar property X defined on the members of a poset P (with partial ordering $\rightsquigarrow$), one may expand X for $\alpha \in P$, as

$$X(\alpha) = \sum_{\beta}^{\succcurlyeq\alpha} f(\beta,\alpha)\, X_f(\beta)$$

where the sum goes over all $\beta \succcurlyeq \alpha$ and $f(\beta,\alpha)$ is a *cluster* function that maps pairs of members of P onto real numbers, with $f(\beta,\alpha) = 0$ whenever β is not $\rightsquigarrow \alpha$, and is such that $f(\alpha,\alpha) \neq 0$. Further, $X_f(\beta)$ is an f *transform* property depending on X and the cluster function f. Conveniently, this cluster expansion may be truncated to a limited sequence of nonzero cluster approximants and so applied whenever the earlier terms offer a good approximation of the property X.

For our reaction-network posets, we choose[50] $f(\beta,\alpha)$ to be the number of ways in which substitution pattern α occurs as a subset of substitution pattern β. This, thence, proceeds somewhat beyond the initial aim of considering purely network-dependent methods independent of molecular structure, but still, it seams worthwhile to investigate, as we do here.

For the poset diagram of CBs, we have truncated the cluster expansion model to X_f contributions from the chlorine atoms [$a \equiv X_f(2)$] and corrections for two chlorine atoms situated in *ortho* [$b \equiv X_f(3)$], *meta* [$c \equiv X_f(4)$], and *para* [$d \equiv X_f(5)$] positions. This truncated cluster expansion model proved to be able to model the properties of CBs. In each series of QSSAR models, benzene was considered as a reference structure; that is, the property values are shifted so that X(benzene) = 0. The set of $X_f(\beta)$ parameters can be computed by a least-squares procedure based on a subset of molecules or by inversion from small systems. The present cluster expansion then yields

$$X(\mathbf{2}) = a$$
$$X(\mathbf{3}) = 2a + d$$
$$X(\mathbf{4}) = 2a + c$$
$$X(\mathbf{5}) = 2a + b$$
$$X(\mathbf{6}) = 3a + 3c$$
$$X(\mathbf{7}) = 3a + b + c + d$$
$$X(\mathbf{8}) = 3a + 2b + c$$
$$X(\mathbf{9}) = 4a + 2b + 2c + 2d$$
$$X(\mathbf{10}) = 4a + 2b + 3c + d$$
$$X(\mathbf{11}) = 4a + 3b + 2c + d$$
$$X(\mathbf{12}) = 5a + 4b + 4c + 2d$$
$$X(\mathbf{13}) = 6a + 6b + 6c + 3d$$

when all higher order cluster terms (e, f, g, ...) are neglected. On the basis of available experimental data, the cluster expansion parameters were computed with a least-squares procedure.

All models were tested in a LOO cross-validation procedure, to obtain results comparable with those from the other poset QSSAR models.

Table 2. Correlation Coefficients *r* and Standard Deviations *s* of the Three QSSAR Models (average poset, cluster expansion, and splinoid poset) for the CB-Toxicity Datasets from Table 1

		average poset		cluster expansion		splinoid poset	
dataset	property	*r*	*s*	*r*	*s*	*r*	*s*
1	log $1/LC_{50}$	0.997	0.051	0.999	0.040	0.997	0.065
2	log $1/LC_{50}$	0.951	0.132	0.975	0.183	0.969	0.198
3	log $1/LC_{50}$	0.969	0.148	0.990	0.126	0.982	0.170
4	log $1/LC_{50}$	0.961	0.164	0.984	0.160	0.978	0.189
5	log $1/LC_{50}$	0.959	0.117	0.977	0.194	0.969	0.263
6	log $1/LC_{50}$	0.975	0.133	0.904	0.469	0.958	0.356
7	log $1/EC_{50}$	0.781	0.230	0.922	0.362	0.931	0.423
8	log $1/EC_{50}$	0.958	0.145	0.968	0.118	0.976	0.105
9	log$1/LC_{50}$	0.985	0.078	0.976	0.174	0.990	0.106
10	log $1/IC_{50}$	0.709	0.184	0.954	0.166	0.783	0.268
11	log $1/IC_{50}$	0.959	0.189	0.986	0.162	0.969	0.269
12	log $1/EC_{50}$	0.963	0.163	0.949	0.260	0.953	0.213
13	log $1/EC_{50}$	0.972	0.283	0.654	0.791	0.927	0.384
14	log $1/LC_{50}$	0.769	0.090	0.842	0.296	0.984	0.105
15	log BCF	0.967	0.190	0.967	0.261	0.976	0.194
16	log BCF	0.991	0.102	0.995	0.096	0.991	0.132
17	log BCF	0.934	0.225	0.934	0.307	0.927	0.341
18	log BCF	0.977	0.108	0.975	0.145	0.982	0.129

4. RESULTS AND DISCUSSION

We have collected from the literature 18 datasets for the CB toxicity against various species (Table 1), and we use them to compare the predictive power of the three posetic QSSAR models: average poset, splinoid poset, and cluster expansion. The LOO cross-validation statistics (correlation coefficient *r* and standard deviation *s*) are presented in Table 2, and representative plots of the experimental versus the predicted CB toxicity are given in Figures 4−6.

An overview of all statistical indices presented in Table 2 shows that posetic QSSAR models generally give good and very good correlations. In many cases, all *r* values are higher than 0.95 (datasets 1−5, 8, 9, 11, 15, 16, and 18), and in all other cases, at least one *r* value is higher than 0.9. In general, the prediction statistics indicate that these methods are robust and can reliably predict toxicity data for CBs. It is important to stress that these posetic algorithms complement each other. This cannot be inferred from the cases where all three methods give equally good predictions but can for the rare situations when one or another of the methods has poor performance. A comparative analysis of the results from Table 2 shows that, occasionally, each posetic QSSAR method makes a poor prediction. When this happens, the other posetic models give good predictions, indicating that there is not a single posetic algorithm that is always best for the toxicity datasets investigated here.

Although all three methods are based on the same poset reaction diagram, the hierarchical information regarding the structure of the chemicals is encoded into different algorithms, and this is apparent from the variation in the predictive power for several datasets. The average poset (AP) method has poor predictions for dataset 7 (log $1/EC_{50}$ for fathead minnow, with $r_{AP} = 0.781$, $r_{CE} = 0.922$, and $r_{SP} = 0.931$), but both cluster expansion (CE) and splinoid poset (SP) have reasonable predictions. We also have an example

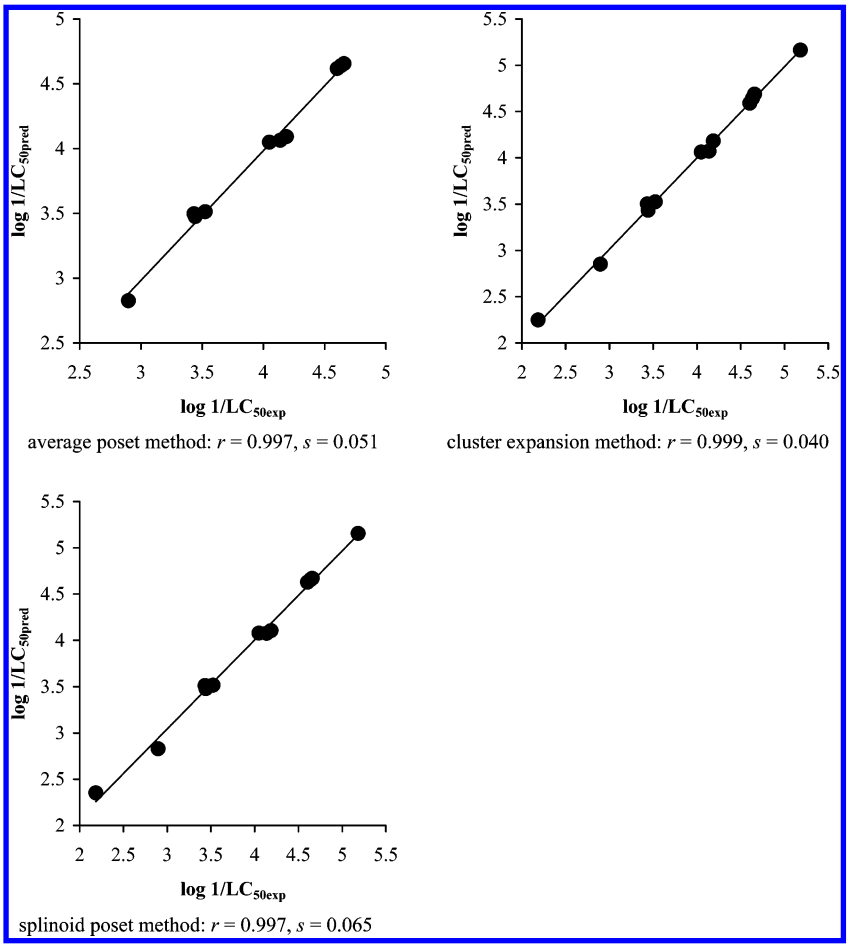

Figure 4. Plot of experimental vs predicted acute aquatic toxicity for the guppy (*Poecilia reticulata*), with the three posetic QSSAR models (log $1/LC_{50}$ [μmol/l], dataset 1 from Table 1).

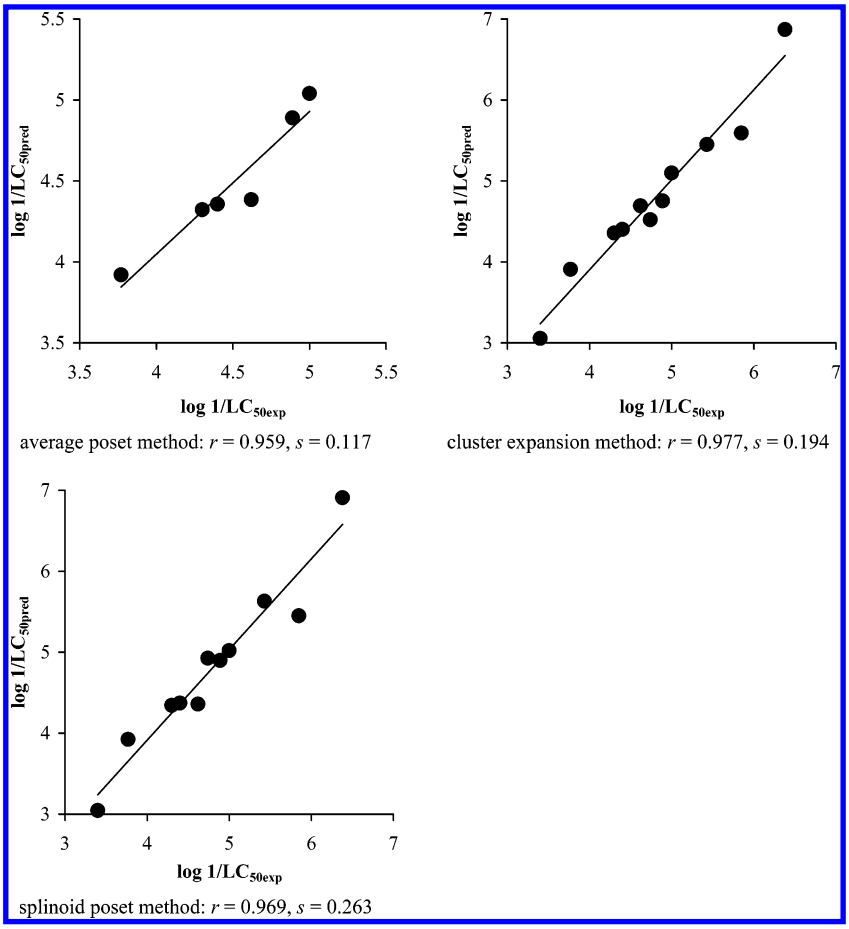

Figure 5. Plot of experimental vs predicted acute aquatic toxicity for the fathead minnow (*Pimephales promelas*), with the three posetic QSSAR models (log $1/LC_{50}$ [mol/l], dataset 5 from Table 1).

(dataset 10) where both AP and SP have low quality predictions (log $1/IC_{50}$ for *Daphnia magna*, with $r_{AP} = 0.709$, $r_{SP} = 0.783$, and $r_{CE} = 0.954$), but CE gives good predictions. These occasional poor performances are the exception in the cases here investigated, but they point to differences between the three algorithms. The average poset is a local nonparametric method, the cluster expansion is a global parametric method, and the splinoid poset method is an interpolation method applied to the reaction diagram points. These differences show that more reliable predictions can be obtained by using a consensus prediction that combines data from the three posetic models.

The toxicity prediction results collected in Table 2 support the concept of using these three posetic algorithms as a battery of QSSAR models, which increases the chances of good predictions compared with each method taken individually. Besides the common derivation from the posetic reaction diagram, another characteristic is the absence of structural descriptors in these QSSAR models. The local or global topology of the poset reaction diagram is the only input information that is used to predict the CB toxicities.

In our effort to have a comprehensive test of the poset QSSAR, we have collected datasets with a high coverage of the vertices in the Hasse diagram. Some datasets are similar, being determined for the same organism, but were reported in different publications and have small numerical differences. We have included them in order to evaluate the poset QSSAR stability for data originating from different laboratories. The first four datasets in Table 1 represent log $1/LC_{50}$ data for the guppy (*Poecilia reticulata*), determined by different groups in slightly different conditions. The prediction results from Table 2 are very good for all three poset QSSAR models and all four datasets, showing that the predictions are not affected by the provenance of the experimental data. In Figure 4, we give the plots of the experimental versus predicted CB toxicities for dataset 1, showing the good correlation and small residuals for all CBs.

Datasets 5−7 represent CB toxicities against the fathead minnow (*Pimephales promelas*), from different sources but similar conditions. With the exception of r for the average poset QSSAR in dataset 7, all other correlation coefficients are higher than 0.9, and dataset 5 has the highest predictions (see Figure 5 for the experimental vs predicted CB toxicities plots). The CB toxicities for brine shrimp (*Artemia* spp.) occur in datasets 8 and 9, and all poset QSSARs have high prediction statistics. The immobilization concentrations of CBs to *Daphnia magna* (datasets 10 and 11) are predicted with very good accuracy with two exceptions: the average and splinoid poset QSSARs for dataset 10. CB toxicities for algae (datasets 12 and 13) have very good statistics. The CB toxicities for tadpoles (dataset 14) are best predicted with the splinoid poset QSSAR, whereas the other methods give poor results.

We have also included four bioconcentration factor (BCF) datasets for fish (datasets 15−18), and all of the models have excellent predictive power. In Figure 6, we give the plots of the experimental versus the predicted CB BCFs for dataset

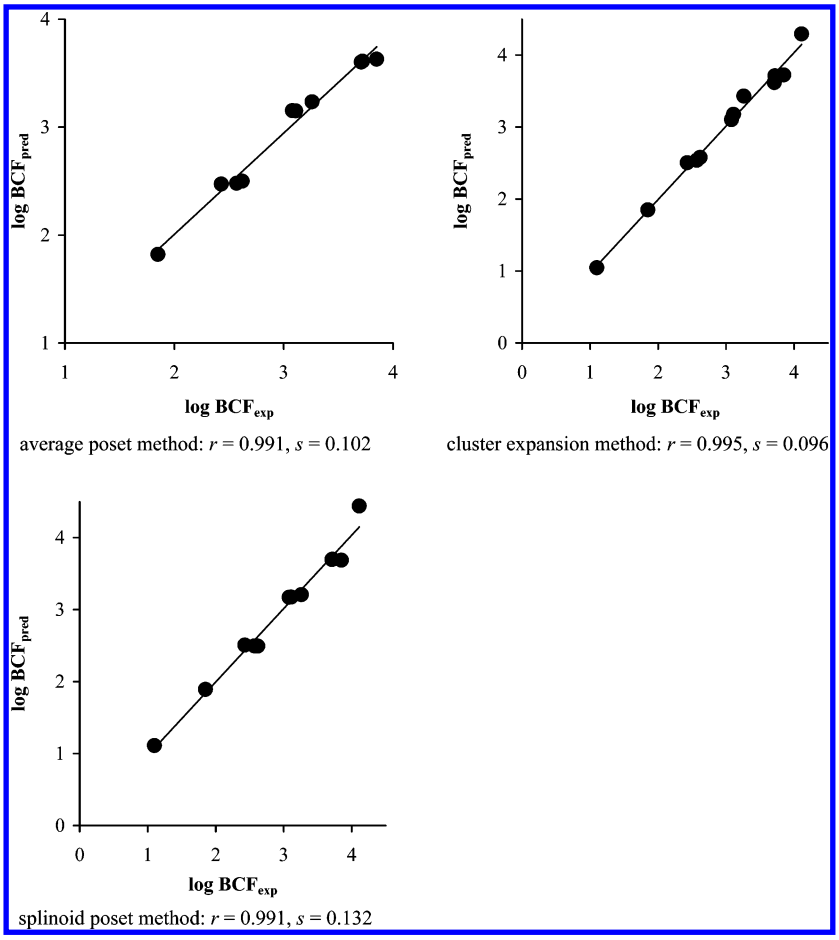

Figure 6. Plot of experimental vs predicted CB bioconcentration factors for fish, with the three posetic QSSAR models (log BCF, dataset 16 from Table 1).

16, showing the good correlation and small residuals for all CBs.

5. CONCLUSIONS

As a result of their widespread industrial and domestic use, organic chemicals often find their way into the environment. Substituted benzenes are among the most prevalent industrial organic chemicals, and their toxicity to fish and other aquatic organisms is used to model their effects on both humans and the environment. CBs are omnipresent in the environment, with all 12 isomers reported to occur in coastal marine sediments, freshwater lake sediments, soils, sewage sludge, wastewater, groundwater, and rivers. Because of their lipophilic character and slow degradation in the environment, they may accumulate in the food chain.

To predict the CBs' toxicity, we compared the predictive power of three poset-derived QSSAR models developed in our group, namely, the average poset,[45] cluster expansion,[50] and splinoid poset models.[46] These QSSAR models were purposely developed for modeling properties in collections of compounds that can be formally derived by substitution from a parent skeleton, such as benzene, but also naphthalene, or biphenyl. On the basis of the poset reaction diagram, all three of these QSSAR models reflect, in distinct ways, the topology of the network that describes the interconversion of chemical species. The average poset method is a local nonparametric method, the cluster expansion method is a parametric method, and the splinoid poset method is a global interpolation method. When the cluster expansion method is used in our current least-squares fitting procedure, it is a global method.

The predicting ability of the models has been demonstrated for 18 datasets representing various species (*Poecilia reticulata*, *Pimephales promelas*, *Daphnia magna*, *Rana japonica*, etc). The prediction results indicate that, within certain limits, the precision of predicted toxicities of CBs to different species is quite acceptable for risk assessment purposes. Although in the majority of investigated datasets all poset QSSAR models give very good predictions, in some cases, they complement each other. These differences show that more reliable predictions can be obtained by using a consensus prediction that combines data from the three posetic models.

The success here as well as in our earlier work[45] focused on properties indicates promise in application to more complex networks (superstructures), such as those that we are now beginning to treat.

ACKNOWLEDGMENT

The authors acknowledge the support (through Grant BD-0894) of the Welch Foundation of Houston, Texas.

REFERENCES AND NOTES

(1) Solbé, J.; Mark, U.; Buyle, B.; Guhl, W.; Hutchinson, T.; Kloepper-Sams, P.; Länge, R.; Munk, R.; Scholz, N.; Bontinck, W.; Niessen,

H. Analysis of the ECETOC aquatic toxicity (EAT) database I − general introduction. *Chemosphere* **1998**, *36*, 99−113.
(2) McCarty, L. S.; Mackay, D. Enhancing ecotoxicological modeling and assessment. *Environ. Sci. Technol.* **1993**, *27*, 1718−1728.
(3) Dalecw, G. M.; Larson, R. E.; Gingerich, W. H. Acute and chronic toxicity studies with monochlorobenzene in rainbow trout. *Aquat. Toxicol*. **1982**, *2*, 127−142.
(4) Blum, D. J. W.; Speece, R. E. Determining chemical toxicity of aquatic species. *Environ. Sci. Technol.* **1990**, *24*, 284−293.
(5) Dearden, J. C.; Cronin, M. T. D.; Dobbs, A. J. Quantitative structure−activity relationships as a tool to assess the comparative toxicity of organic chemicals. *Chemosphere* **1995**, *31*, 2521−2528.
(6) Warne, M. A.; Osborn, D.; Lindon, J. C. J.; Nicholson, K. Quantitative structure-toxicity relationships for halogenated substituted-benzenes to *Vibrio fischeri*, using atom-based semiempirical molecular-orbital descriptors. *Chemosphere* **1999**, *38*, 3357−3382.
(7) Burden, F. R.; Winkler, D. A. A quantitative structure−activity relationships model for the acute toxicity of substituted benzenes to *Tetrahymena pyriformis* using bayesian-regularized neural networks. *Chem. Res. Toxicol.* **2000**, *13*, 436−440.
(8) 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*, 1284−1295.
(9) Dyer, S. D.; White-Hull, C. E.; Shephard, B. K. Assessments of chemical mixtures via toxicity reference values overpredict hazard to Ohio fish communities. *Environ. Sci. Technol*. **2000**, *34*, 2518−2524.
(10) Ren, S. Modeling the toxicity of aromatic compounds to *Tetrahymena pyriformis*: The response surface methodology with nonlinear methods. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1679−1687.
(11) Lu, X.; Tao, S.; Hu, H.; Dawson, R. W. Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors. *Chemosphere* **2000**, *41*, 1675−1688.
(12) Gramatica, P.; Papa, E. QSAR modeling of bioconcentration factor by theoretical molecular descriptors. *QSAR Comb. Sci*. **2003**, *22*, 374−385.
(13) Martin, T. M.; Young, D. M. Prediction of the acute toxicity (96-h LC_{50}) of organic compounds to the fathead minnow (*Pimephales promelas*) using a group contribution method. *Chem. Res. Toxicol*. **2001**, *14*, 1378−1385.
(14) Seward, J. R.; Hamblen, E. L.; Schultz, T. W. Regression comparisons of *Tetrahymena pyriformis* and *Poecilia reticulata* toxicity. *Chemosphere* **2002**, *47*, 93−101.
(15) Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D. QSAR for dihydrofolate reductase inhibitors with molecular graph structural descriptors. *THEOCHEM* **2002**, *582*, 39−51.
(16) Freidig, A. P.; Hermens, J. L. M. Narcosis and chemical reactivity QSARs for acute fish toxicity. *Quant. Struct.-Act. Relat.* **2001**, *19*, 547−553.
(17) Tao, S.; Hu, H.; Lu, X.; Dawson, R. W.; Xu, F. Fragment constant method for prediction of fish bioconcentration factors of nonpolar chemicals. *Chemosphere* **2000**, *41*, 1563−1568.
(18) 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*, 131−138.
(19) Devillers, J.; Balaban, A. T. *Topological indices and related descriptors in QSAR and QSPAR*; Gordon & Breach: Langhorne, PA, 1999; p 811.
(20) Ivanciuc, O. Support vector machine identification of the aquatic toxicity mechanism of organic compounds. *Internet Electron. J. Mol. Des.* **2002**, *1*, 157−172. http://www.biochempress.com.
(21) Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D. 3D quantitative structure activity relationships with CoRSA. Comparative receptor surface analysis. Application to calcium channel agonists. *Analusis* **2000**, *28*, 637−642.
(22) Basak, S. C.; Balasubramanian, K.; Gute, B. D.; Mills, D.; Gorczynska, A.; Roszak, S.; Prediction of cellular toxicity of halocarbons from computed chemodescriptors: A hierarchical QSAR approach. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1103−1109.
(23) Stanton, D. T. On the physical interpretation of QSAR models. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1423−1433.
(24) Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D. Comparative receptor surface analysis (CoRSA) model for calcium channel antagonists. *SAR QSAR Environ. Res.* **2001**, *12*, 93−111.
(25) Ivanciuc, O.; Ivanciuc, T.; Filip, P. A.; Cabrol-Bass, D. Comparative receptor surface analysis (CoRSA) of calcium channel agonist activity of 1,4-dihydro-2,6-dimethyl-3-nitro-4-pyridyl-5-pyridinecarboxylates. *Rev. Roum. Chim*. **2001**, *46*, 69−74.
(26) Ivanciuc, T.; Ivanciuc, O. Bone resorption inhibition quantitative structure−activity relationships for aryl-substituted bisphosphonates. *Internet Electron. J. Mol. Des.* **2003**, *2*, 403−412. http://www.biochempress.com.
(27) Richards, W. G. *Quantum Pharmacology*; Butterworths: London, 1977; 1983.
(28) Mekenyan, O. G.; Veith, G. D. Relationships between descriptors for hydrophobicity and soft electrophilicity in predicting toxicity. *SAR QSAR Environ. Res.* **1993**, *1*, 335−344.
(29) Hermens, J.; Busser, F.; Leeuwangh, P.; Musch, A. Quantitative structure−activity relationships and mixture toxicity of organic chemicals in *Photobacterium phosphoreum*: the Microtox test. *Ecotoxicol. Environ. Saf.* **1985**, *9*, 17−25.
(30) Thumm, W.; Brüggemann, R.; Freitag, D.; Kettrup, A. Toxicity tests with luminescent photobacteria and quantitative structure activity relationships for nitroparaffins. *Chemosphere* **1992**, *24*, 1835−1843.
(31) Hansch, C.; Leo, A., Exploring QSAR. *Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995; p 580.
(32) Willett, P. *Similarity and clustering in chemical information systems*; Research Studies Press: Letchworth, Hertfordshire, England, 1989.
(33) Johnson, M. A.; Maggiora, G. M. *Concepts and applications of molecular similarity*. John Wiley & Sons: New York, 1990; p 393.
(34) McGreagor, J. J.; Willett, P. Use of a maximum common subgraph algorithm in the automatic identification of ostensible bond changes occurring in chemical reactions. *J. Chem. Inf. Comput. Sci*. **1981**, *21*, 137−140.
(35) Chartrand, G.; Heria, H.; Johnson, M. A. A representation of chemical transformations by labeled graphs. *J. Math. Chem*. **1993**, *12*, 59−80.
(36) Skvortsova, M. I.; Baskin, I. I.; Stankevich, I. V.; Palyulin, V. A.; Zefirov, N. S. Molecular similarity. I. Analytical description of the set of graph similarity measures. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 785−790.
(37) Jørgensen, A. M. M.; Pedersen, J. T. Structural diversity of small molecule libraries. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 338−345.
(38) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem*. **2002**, *45*, 4350−4358.
(39) Kearsley, S. K.; Sallamack, S.; Fluder, E. M.; Andose, J. D.; Mosley, R. T.; Sheridan, R. P. Chemical similarity using physiochemical property descriptors. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 118−127.
(40) Delaney, J. S. Assessing the ability of chemical similarity measures to discriminate between active and inactive compounds. *Mol. Diversity* **1996**, *1*, 217−222.
(41) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 983−996.
(42) Klein, D. J. Similarity and dissimilarity in posets. *J. Math. Chem*. **1995**, *18*, 321−348.
(43) Klein, D. J.; Babić, D. Partial orderings in chemistry. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 656−671.
(44) Klein, D. J. Prolegomenon on partial orderings in Chemistry. *Match* **2000**, *42*, 7−21.
(45) Ivanciuc, T.; Klein, D. J. Parameter-free structure−property correlation via progressive reaction posets for substituted benzenes. *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 610−617.
(46) Došlić, T.; Klein, D. J. Splinoid interpolation on finite posets. *J. Comput. Appl. Math.* **2005**, *177*, 175−185.
(47) Klein, D. J. Chemical graph-theoretic cluster expansions. *Int. J. Quantum Chem., Quantum Chem. Symp*. **1986**, *20*, 153−171.
(48) Schmalz, T. G.; Živković, T.; Klein, D. J. Cluster expansion of the Hückel molecular energy for acyclics: Applications to pi resonance theory. *MATH/ CHEM./COMP. 1987, Proc. Int. Course Conf. Interfaces Math., Chem., Comput. Sci, 1987* **1987**, *54*, 173−190.
(49) Klein, D. J.; Schmalz, T. G.; Bytautas, L. Chemical sub-structural cluster expansions for molecular properties. *SAR QSAR Environ. Res*. **1999**, *10*, 131−156.
(50) Ivanciuc, T.; Klein, D. J.; Ivanciuc, O. Posetic cluster expansion for substitution-reaction diagrams and its application to cyclobutane. *J. Math. Chem.* submitted for publication.
(51) Fishbein, L. Potential halogenated industrial carcinogenic and mutagenic chemicals: IV. Halogenated aryl derivatives. *Sci. Total Environ*. **1979**, *11*, 259−278.
(52) Galassi, S.; Vighi, M. Testing toxicity of volatile substances with algae. *Chemosphere* **1981**, *10*, 1123−1126.
(53) Hallett, D. J.; Norstrom, R. J.; Onuska, F. I.; Comba, M. E. Incidence of chlorinated benzenes and chlorinated ethylenes in Lake Ontario Herring Gulls. *Chemosphere* **1982**, *11*, 277−285.
(54) Calamari, D.; Galassi, S.; Setti, F.; Vighi, M. Toxicity of selected chlorobenzenes to aquatic organisms. *Chemosphere* **1983**, *12*, 253−262.
(55) Ryan, J. A.; Bell, R. M.; Davidson, J. M.; O'Connor, G. A. Plant uptake of nonionic organic chemicals from soils. *Chemosphere* **1988**, *17*, 2299−2323.

(56) Oliver, B. G.; Nicol, K. D. Chlorobenzenes in sediments, water, and selected fish from Lakes Superior, Huron, Erie, and Ontario. *Environ. Sci. Technol.* **1982**, *16*, 532−536.
(57) Jan, J.; Malnerščič, S. Chlorinated benzene residues in fish in Slovenia (Yugoslavia). *Bull. Environ. Contam. Toxicol.* **1980**, *24*, 824−827.
(58) Rogers, H. R.; Campbell, J. A.; Crathorne, B.; Dobbs, A. J. The occurrence of chlorobenzenes and permethrins in twelve U.K. sewage sludges. *Water Res*. **1989**, *23*, 913−921.
(59) Belfroid, A.; Vanweze, A.; Sikkenk, M.; Vangestel, K.; Seinen, W.; Hermens, J. The toxicokinetic behavior of chlorobenzenes in Earthworms (*Eisenia andrei*): Experiments in water. *Ecotoxicol. Environ. Saf*. **1993**, *25*, 154−165.
(60) Scheunert, I.; Topp, E.; Attar, A.; Korte, U. F. Pathways of chlorobenzenes in plants and their correlation with *n*-octanol/water partition coefficients. *Ecotoxicol. Environ. Saf.* **1994**, *27*, 90−104.
(61) Wu, Q.; Milliken, C. E.; Meier, G. P.; Watts, J. E. M.; Sowers, K. R.; May, H. D. Dechlorination of chlorobenzenes by a culture containing bacterium DF-1, a PCB dechlorinating microorganism. *Environ. Sci. Technol*. **2002**, *36*, 3290−3294.
(62) Muñoz, M. J.; Ramos, C.; Tarazona, J. V. Bioaccumulation and toxicity of hexachlorobenzene in *Chlorella vulgaris* and *Daphnia magna*. *Aquat. Toxicol*. **1996**, *35*, 211−220.
(63) Birkhoff, G. *Lattice theory*, third edition; American Mathematical Society Colloquium Publications: Providence, R. I., 1967; Vol. XXV, p 418.
(64) Brüggemann, R.; Münzer, B.; Halfon, E. An algebraic/graphical tool to compare ecosystems with respect to their pollution—the German river "Elbe" as an example—I: Hasse-diagrams. *Chemosphere* **1994**, *28*, 863−872.
(65) Münzer, B.; Brüggemann, R.; Halfon, E. An algebraic/graphical tool to compare ecosystems with respect to their pollution II: Comparative regional analysis. *Chemosphere* **1994**, *28*, 873−879.
(66) Brüggemann, R.; Schwaiger, J.; Negele, R. D. Applying Hasse diagram technique for the evaluation of toxicological fish tests. *Chemosphere* **1995**, *30*, 1767−1780.
(67) Brüggemann, R.; Bartel, H. G. A theoretical concept to rank environmentally significant chemicals. *J. Chem. Inf. Comput. Sci*. **1999,** *39*, 211−217.
(68) Pudenz, S.; Brüggemann, R.; Luther, B.; Kaune, A.; Kreimes, K. An algebraic/graphical tool to compare ecosystems with respect to their pollution V: Cluster analysis and Hasse diagrams. *Chemosphere* **2000**, *40*, 1373−1382.
(69) Brüggemann, R.; Pundez, S.; Carlsen, L.; Sørensen, P. B.; Thomsen, M.; Mishra, R. K. The use of Hasse diagrams as a potential approach for inverse QSAR. *SAR QSAR Environ. Res.* **2001**, *11*, 473−487.
(70) Carlsen, L.; Sørensen, P. B.; Thomsen, M.; Brüggemann, R. QSAR's based on partial order ranking. *SAR QSAR Environ. Res.* **2002**, *13*, 153−165.
(71) Lerche, D.; Sørensen, P. B.; Larsen, H. L.; Carlsen, L.; Nielsen, O. J. Comparison of the combined monitoring-based and modelling-based priority setting scheme with partial order theory and random linear extensions for ranking of chemical substances. *Chemosphere* **2002**, *49*, 637−649.
(72) Lerche, D.; Sørensen, P. B. Evaluation of the ranking probabilities for partial orders based on random linear extensions. *Chemosphere* **2003**, *53*, 981−992.
(73) Lerche, D.; Matsuzaki, S. Y.; Sørensen, P. B.; Carlsen, L.; Nielsen, O. J. Ranking of chemical substances based on the Japanese pollutant release and transfer register using partial order theory and random linear extensions. *Chemosphere* **2004**, *55*, 1005−1025.
(74) Sørensen, P. B.; Mogensen, B. B.; Carlsen, L.; Thomsen, M. The influence on partial order ranking from input parameter uncertainty: Definition of a robustness parameter. *Chemosphere* **2000**, *41*, 595−601.
(75) Klein, D. J.; Bytautas, L. Directed reaction graphs as posets. *Match* **2000**, *42*, 261−289.
(76) Rota, G. On the foundation of combinatorial theory I, Theory of Möbius functions. *Z. Wahr. Verw. Gebiete* **1964**, *2*, 340−368.

CI0501342