

Structure–Fate Relationships of Organic Chemicals Derived from the Software Packages E4CHEM and WHASSE

Rainer Brüggemann*

Department Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries,
Müggelseedamm 310, D-12587 Berlin, Germany

Guillermo Restrepo

Laboratorio de Química Teórica, Universidad de Pamplona, Pamplona, Colombia

Kristina Voigt

GSF-National Research Center for Environment and Health, Institute of Biomathematics and Biometry,
Ingolstädter Landstr. 1, D-85764 Neuherberg, Germany

Received September 26, 2005

A risk assessment of chemicals is to be performed on the basis of the model EUSES, developed by the Commission of the European Union. The model package E4CHEM (Exposure Estimation for Potentially Ecotoxic Environmental Chemicals), developed in 1984–1992, is presented and applied in this paper as a model which allows a model-supported evaluation of chemicals. E4CHEM consequently does not have the wide applicability and technical comfort of the more recently developed model EUSES. The simulation models of E4CHEM characterize the chemical behavior in the environment by many aspects. Hence, there is a need to condense all of these aspects to get a clear impression of what will be the fate of the chemicals. Starting with the already published concept of exposure maps, we will discuss how partial orders may be helpful in establishing generalized structure–fate relationships. The software WHASSE is applied.

1. RISK ASSESSMENT OF CHEMICALS

Risk assessment of chemicals^{1,2} is often performed by means of the E. U. model EUSES.^{2–4} EUSES is a harmonized quantitative risk assessment tool for chemicals. However, EUSES calls for an extensive data sampling for each chemical. Note, however, that Verdonck and co-workers⁴ boiled down EUSES to only a three-chemical parameter model. The software package E4CHEM (Exposure Estimation for Potentially Ecotoxic Environmental Chemicals) was developed during the years 1984–1992.⁷ E4CHEM consists of a system of modules that describes the chemical's behavior in different environmental targets and by different stages of data availability. In comparison to EUSES, the model package E4CHEM does not contain many features such as the estimation of the chemical's behavior in wastewater treatment plants or the estimation of pathways to humans in order to estimate health risks. On the other hand, E4CHEM has modules such as DTEST, giving a high degree of automatic estimation of needed chemical properties, or EXWAT, describing the behavior of chemicals in rivers, which we want to apply. However, E4CHEM does not have the technical comfort shown by EUSES.

Transport, the partitioning of a chemical among different targets, and reactivity are competitive processes. Hence, the substance properties must be combined in an appropriate way in order to extract adequate information about chemicals. There are different possibilities: One may consider the chemical as a potentially multimedia one, which can be

partitioned in a non-negligible amount either into soils, sediments, water, or air simultaneously, or one can consider specified single-media models and discuss what amount of the chemical remains in this specific medium or its probability of being transferred to other ones. Here, besides EXTND (the multimedia model of chemicals in closed systems),⁷ the surface-water model EXWAT (which is one of the modules of the software package E4CHEM)^{5–7} is introduced. EXWAT couples a few substance properties in order to map these to a few descriptors about the environmental behavior. The model EXTND allows a very clear presentation of structure–fate relationships but can only be applied in a very restricted way⁵ (see below). A more general approach is provided by the concept of partial orders (partial order: see, for example, refs 5, 8–10). Hence, this paper is organized as follows: After a brief description of E4CHEM, the module EXTND is explained. The kind of structure–fate relationship derived from EXTND is very useful; however, it cannot be generalized, and therefore, another theoretical idea must be introduced. First, as a leading example, the model EXWAT is introduced; then steps toward generalized structure–fate relationships are briefly discussed. Finally, an application of EXWAT is demonstrated for 19 chemicals monitored in the German river Main. A critical discussion concludes the paper.

2. SOFTWARE PACKAGE E4CHEM

2.1. Overview. We show in Figure 1 the general idea of E4CHEM. The package E4CHEM consists of four main parts: (1) DTEST,^{5,11} the supporting module which fills up

* Corresponding author tel.: 49-30-64181666; e-mail: brg@igb-berlin.de.

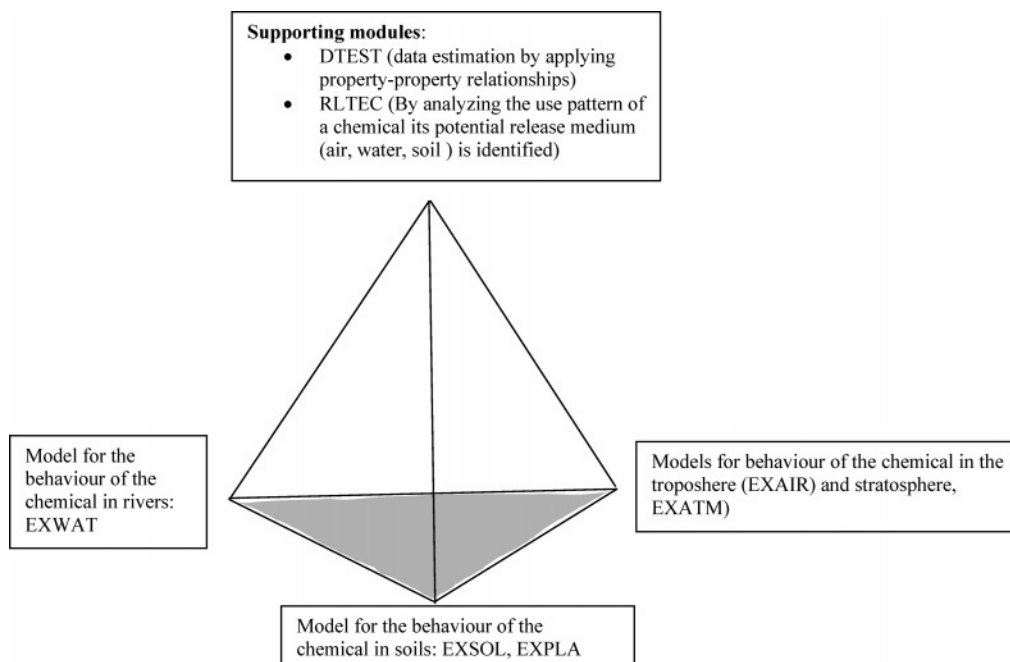


Figure 1. Layout of the model E4CHEM^{5–7}.

the chemical property list; (2) RLTEC,^{5,12} which estimates the potentially affected media through the use of the chemical's pattern; (3) single-medium/-target simulation models (EXWAT,^{13,14} EXAIR,¹⁵ EXATM,¹⁶ EXSOL,¹⁷ and EXPLA),^{18,19} which respectively study the behavior of a chemical in rivers, the troposphere, the stratosphere, soil, and plants (models written at the corners of the basis triangle; Figure 1); and (4) multimedia models, which describe chemicals having no dominating medium or target. These models could be located in Figure 1 in the shaded basis triangle area (not shown) and are named EXTND⁵ and EXFUG. EXFUG is a fugacity model (level III).²⁰ Besides EXTND and EXFUG, there is EXINT, which couples the models EXSOL, EXWAT, and EXAIR.²¹

2.2. Structure–Fate Relationships based on EXTND.

As already mentioned, EXTND is a simple partitioning model. This model, describing the partition of chemicals in a thermodynamically closed system, is useful to determine which environmental medium or target (i.e., water, soil, or air) will mainly be affected by a chemical.

In EXTND, the partitioning of a chemical between water and air, on the one hand, and between water and soil or suspended matter (“solids”), on the other hand, is assumed. By mass balance and chemical equilibria, one can derive mass fractions x_i of a chemical in air, water, and (for example) soil: $0 \leq x_i \leq 1$, ($i = 1, 2$, and 3). Besides environmental parameters, gathered in a tuple p , x_i is a function of the partitioning coefficients $K_{\text{water,solids}}$ and K_{AW} . For hydrophobic chemicals, the partitioning coefficient $K_{\text{water,solids}}$ can be related to the basic parameter K_{OW} (n -octanol–water partitioning coefficient). K_{AW} (air–water partitioning coefficient) is related to the Bunsen coefficient or to the Henry law coefficient (HLC).

$$x_i = f_i(K_{\text{OW}}, K_{\text{AW}}, p) \quad (i = 1, 2, \text{ and } 3) \quad (1)$$

Additionally, there is another equation that expresses the mass balance; that is, the sum of the three x_i values must be

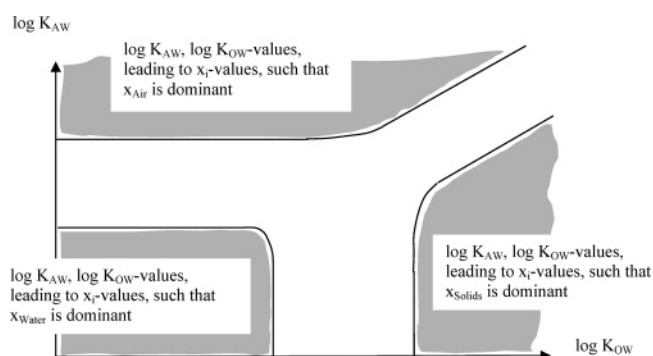


Figure 2. Schematic presentation of an “exposure map”.

1. Equivalently, the equation system (eq 1) can also be rearranged as follows:

$$\begin{aligned} K_{\text{AW}} &= f_1'(x_1, K_{\text{OW}}, p) \\ K_{\text{AW}} &= f_2'(x_2, K_{\text{OW}}, p) \\ K_{\text{AW}} &= f_3'(x_3, K_{\text{OW}}, p) \end{aligned} \quad (1a)$$

Given a certain value of x_i , there will be a dependency of K_{AW} on K_{OW} , which is different according to the medium we are interested in. It is more comfortable to use logarithmic quantities; hence, we arrive at

$$\log K_{\text{AW}} = g_i(\log K_{\text{OW}}, x_i, p) \quad (i = 1, 2, \text{ and } 3) \quad (1b)$$

In a plane ($\log K_{\text{AW}}$ as ordinate and $\log K_{\text{OW}}$ as abscissa, “ g plane”), $\log K_{\text{AW}} = g_i(x_i, \log K_{\text{OW}})$ appears as a graph for x_i ($i=1, 2$, and 3 ; “ g curves”). When, for example, 90% of the mass fraction is taken in air, in water, and in soil, three g curves appear, separating regions in the $\log K_{\text{AW}}$ – $\log K_{\text{OW}}$ plane. (Figure 2).

In Figure 2, the coordinates ($-\infty < \log K_{\text{OW}} < +\infty$) and ($-\infty < \log K_{\text{AW}} < +\infty$) span a plane, the “ g plane”. Substances characterized by numerical values of $\log K_{\text{OW}}$ and $\log K_{\text{AW}}$ can be located as “points” in the g plane.

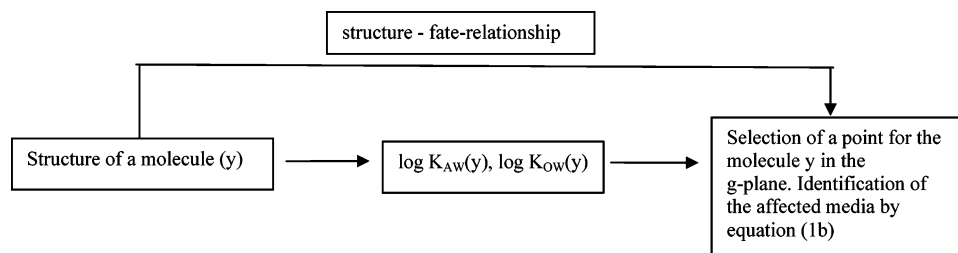


Figure 3. Scheme of the application of EXTND to derive structure–fate relationships.

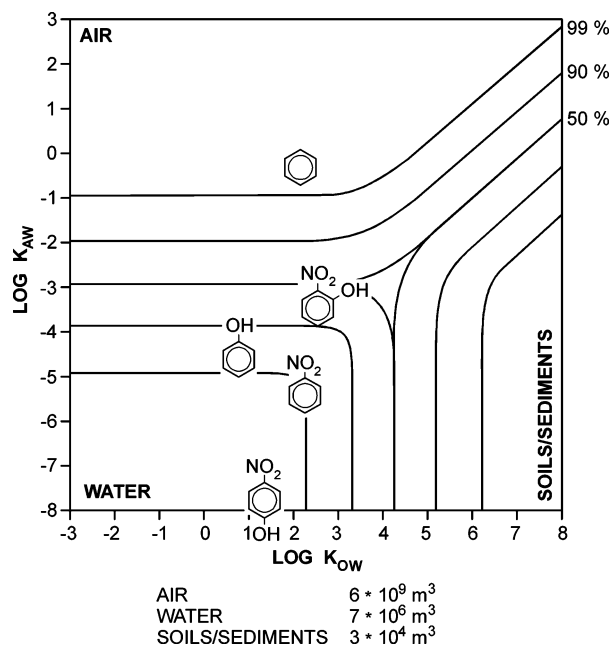


Figure 4. Structure–fate relationships for five benzene derivatives: benzene, phenol, nitrobenzene, ortho-nitrophenol, and para-nitrophenol.

Depending on their position in one of the four fields [three gray ones and the remaining two- or three-media area (blank)], one can derive the potential fate of a chemical. If the g curves are known, which depend on environmental parameters, the pair ($\log K_{AW}$ and $\log K_{OW}$) is decisive for the fate of the chemical. In many cases, the structure of a chemical and the values of $\log K_{AW}$ and $\log K_{OW}$ are related. We summarize this idea in Figure 3.

As an example, EXTND is applied to a series of substituted benzenes (Figure 4). Besides the molecular formula and relative molar mass, the input data are (a) chemical properties K_{OW} and K_{AW} and (b) environmental properties such as pH, temperature, organic carbon content, and soil porosity.

According to the $\log K_{AW}$ and $\log K_{OW}$ values, benzene is located in the upper part of the g plane, where the mole fraction of air $\geq 99\%$. Hence, benzene is an “air chemical”. Substitution by NO_2 or OH “pushes” the substance to a lower point of the g plane, namely, into that region where accumulation in water prevails [$x_{\text{Water}}(\text{Ph}-\text{NO}_2) \approx 99\%$]. The substitution of both in the para position shifts the point to the bottom of the g plane because both dipoles lead to a favorable substance–water interaction. If, however, both substituents are in an ortho position, then obviously the dipoles interact mainly intramolecularly and the attractive forces to the water molecules are reduced. Consequently, the molecule ortho-nitrophenol tends to be a multimedia

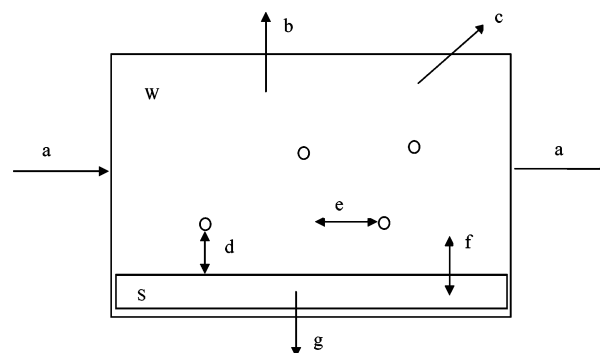


Figure 5. One compartment of EXWAT.

substance. A similar “exposure map” is explained in ref 16, where chlorinated benzenes and anilines were studied.

2.3. Exposure Model EXWAT. We mentioned above that substance data alone cannot be used to calculate the hazard due to exposure. Instead, a deterministic mathematical exposure model is needed, by which the environmental parameters are coupled with substance parameters in the correct way. The mathematical basis of such a model is the differential mass balance, that is

$$dc/dt = \text{Input} - \text{Output} \quad (2)$$

If different targets are to be considered, eq 2 is not sufficient. In EXWAT, for example, each target, sediment on the one hand and water body of the surface water on the other hand, gets its own differential equation. Hence, instead of eq 2, a river segment is described by two differential equations. After the concentration in the outflow of one river segment (compartment) is determined, the inputs of the downstream section can be calculated. Hence, a profile of the behavior of a chemical in a river can be derived. However, for the structure–fate relationship, it is sufficient to study just one segment, where all relevant processes are adequately described. Figure 5 shows that segment.

Each modeled segment of a river consists of two compartments, the water body, W , and the sediment, S . There is an inflow (a) with an upstream concentration of the chemical and an outflow (a) with the resulting concentration due to the processes within the compartment. In the water body, suspended material is transported (small circles) which can be deposited or resuspended (d). It is assumed that the dissolved chemical is in equilibrium with its sorbed form on the suspended material (e). By dispersive forces, the dissolved chemical enters the interstitial water (f). The processes of volatilization (b), degradation (c) (not specified according to the speciation of the chemical), and sediment burial (g) are considered as sinks. That is, no metabolites are considered.

The model EXWAT couples all substance parameters (such as K_{OW} , K_{AW} , and degradation rates) and environmental data (such as concentrations of the suspended solids, temperature, pH, depth of the active sediment zone, organic carbon content, porosities, water discharge, wind velocity, deposition rate of the suspended matter, and geometrical data of the river).

Hence, one can deduce fate descriptors (if the input of chemicals and the temporal behavior of the environmental system is ignored) from the solution of equation systems like eq 1. In the case of EXWAT, four fluxes are derived:

- sedimentation (hydrophobicity of the chemical, properties of the sediments and the river)
- volatilization (Henry law coefficient and properties of the river)
- mineralization flux (depending on chemical's partitioning)
- advection (downstream transport of the chemical).

These four fluxes have the dimension of $[M/T]$. Because the mineralization flux does not have the same orientation as the other fluxes (a high mineralization flux implies a loss of pollution, whereas the other three fluxes imply an accumulation in different targets), the evaluation is based on (i) sedimentation (geoaccumulation), (ii) volatilization (loading of the river bank environment), (iii) persistence, and (iv) advection (downstream pollution, hazards for the sea). In summary, because EXWAT delivers four descriptors, each of them encompassing valuable information, concepts of partial order theory can be applied.

3. TOWARD A GENERALIZATION OF STRUCTURE–FATE RELATIONS

3.1. Drawbacks of the Exposure Map. Generalizations of the exposure map, based on EXTND, are obvious. As long as two leading substance properties are known, which at the same time are sensitive to molecular structures, and as long as the presumptions of a partitioning model are fulfilled (closed system and constant environmental parameters), one may deduce exposure maps. To generalize, in all cases where it is possible to relate environmental concentrations (expressing the fate) to two substance properties, one may derive g planes. Two crucial points remain: (1) in the more generalized concept, one needs some limiting values in order to decide whether a chemical will be an “ i th-medium chemical”, and (2) the restriction that the g plane can only relate two parameters is uncomfortable.

3.2. Partial Order as a Basis for Generalized Structure–Fate Relationships. *Definition 1.* A characteristic property indicating the i th fate is called a “descriptor” and is denoted by D_i . Such descriptors may be the accumulation into an environmental section, or the transformation efficiency to toxic metabolites, or a residence time, or any other kind of model result expressing the behavior of a chemical. For example, Mackay and Neely introduced characteristic times in his fugacity level II model.²⁰

Definition 2. The set of chemicals is called G . G is the “ground set”, which is to be supplied by a binary algebraic relation “ \leq ”.

$D_i = f_i[p, q(x)]$ (p : tuple of environmental parameters, $q(x)$: tuple of properties of substance x). The D_i values of a

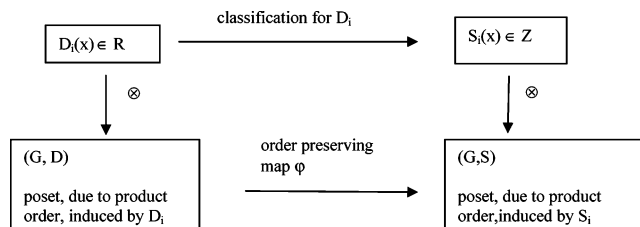


Figure 6. Classification and order-preserving maps. \mathbb{R} : the set of real numbers. \mathbb{Z} : the set of integers.

chemical x belong to a tuple $D(x)$, and $D(x)$ characterizes the fate of a chemical x with respect to the environmental system considered. By $D_i(x) \leq D_i(y)$ for all i , the set G gets a structure,²² namely, that of a partially ordered set. The order relation is well-known²² and is called a product- or component-wise order.¹⁰ In most mathematical textbooks, the symbol (G, \leq) for a partially ordered set is used. Brüggemann and co-workers have introduced the notation (G, IB) [IB: “information base” (i.e., the set of properties from which the product order is generated)].²³ Here, we will use (G, D) to indicate the dependence of the poset on the fate characterizing tuple D . If $D_i(x) < D_i(y)$ for some indices i and $D_i(x) > D_i(y)$ for some other indices, then x and y are “incomparable”, denoted as $x \parallel y$. The graphical representation of posets by Hasse diagrams finds more and more interest in the literature; a recent example of Hasse diagrams and their use in quantitative superstructure–activity relationships was published in this journal.²⁴ A Hasse diagram can always be constructed. In this paper, our focus is to show how Hasse diagrams may be used to represent structure–fate relationships rather than to support decisions or evaluations by linear orders, for example, by the construction of averaged ranks, see, for example, ref 25.

To use Hasse diagrams as a graphical representation of structure–fate relationships, the role of the scale level of the quantities D_i must be clarified. Most often, the quantities D_i are continuous quantities (such as outcomes from a simulation model). The immediate use of the D_i values to express order relations often leads to very confusing Hasse diagrams because even insignificant numerical differences will affect the order relations. It is very hard to explain in general and in strict terms what is a “nonrelevant numerical difference”. From a technical point of view, several strategies are useful and have already been published: cluster analysis²⁶ or a classification into discrete intervals.²⁷ A promising attempt was made by Restrepo and co-workers,²⁸ who introduced the concept of topological chemical spaces based on the concept of a neighborhood for each chemical depending on the values of its descriptors. Its use for structure–fate relationships will be outlined in another paper. Classifications into discrete increasingly labeled intervals can be considered as preserving maps, which we call φ :

$$\varphi(D) = S \quad S \in Z \text{ (set of integers) for all } x \in G, \\ \varphi \text{ order preserving (3)}$$

By eq 3, we come to a representation, as shown in Figure 6. We call S the tuple of scores of x , and we keep open which specific nature is behind φ . In many cases, different φ 's arise just from different classification schemes. For more details, see Brüggemann and Bartel.²⁷

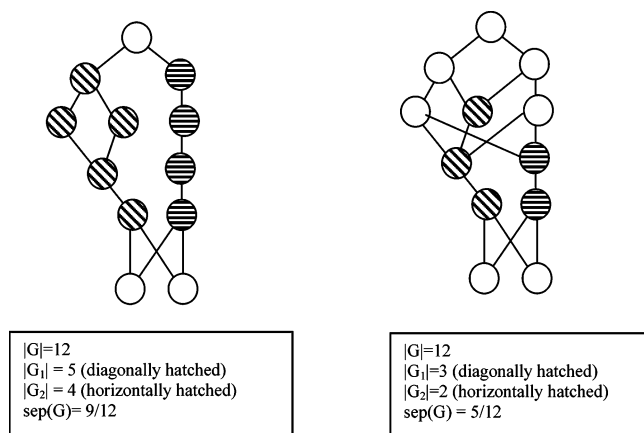


Figure 7. Concept of $\text{sep}[G, \varphi(D)]$, explained by two examples. (Note that, here, we use G_i instead of T_i because, here, the equivalence relation \mathcal{R} is not relevant.)

The main point is that the use of different order-preserving mappings φ gives us the freedom to find—if possible—clearly structured Hasse diagrams. By a different processing of the descriptor values, maintaining the order relations found by D alone, a set of posets, X , results, which is defined as follows:

$$X = \{[G, \varphi(D)], \varphi \text{ order preserving}\} \quad (4)$$

That is, the set X consists of all posets, based on the same ground set G but resulting from different φ 's (see also, section 3.3.1), which are applied to D . From the set X , one poset can be selected which is “suitable” to represent structure–fate relations. It is still not clear what we mean by “suitable”. A direct definition is hardly possible. Instead, we intend to define characteristic numbers, which help us to quantify to what degree a Hasse diagram can be understood as being suitable to represent structure–fate relationships.

3.3. Characterizing the Degree of a Hasse Diagram To Represent Structure–Fate Relationships. *3.3.1. Order-Theoretical Constraints.* The mapping φ may generate equivalence classes.²⁹ Hence, both the set G and the quotient set $G/\mathcal{R}(\varphi)$ with $\mathcal{R}(\varphi)$ defined by eq 5 are needed.

$$(x \mathcal{R}(\varphi) y) \Leftrightarrow \varphi[D(x)] = \varphi[D(y)] \quad (5)$$

To avoid cumbersome notations, we will proceed as follows: (1) If order relations are to be counted, then we consider the partial order of a transversal set³⁰ of representatives $T = T\{\varphi[D(G)]\}$ taken from the equivalence classes due to the equivalence relation \mathcal{R} (eq 5). Indeed, we actually examine partially ordered sets $\{T, \varphi[D(G)]\}$. (2) If elements are to be counted, for example, if chemicals are counted because of their structural characteristics, then we consider not only the representatives but also their equivalent elements.²⁹ In technical terms, given an element $x \in T$, then the set G_x of all elements of G is to be found, whose canonical map $c(G_x) = x$. For example, $G = \{a, b, c, d, e, f\}$ and $G/\mathcal{R} = \{K_1, K_2, K_3\}$, where $K_1 = \{a, b, c\}$, $K_2 = \{d, e\}$, and $K_3 = \{f\}$. One out of six possible realizations of T is $T = \{a, d, f\}$. Take the element “ a ”; then, $G_a = \{a, b, c\}$, because $c(a) = K_1$. It is useful to relate subsets of G to subsets of T_i by applying the reverse canonical map c^{-1} . Continuing the example above, $T_1 = \{a, d\}$; then, $c^{-1}(T_1) = \{a, b, c, d, e\}$.

In more detail, $c^{-1}(T_1) = \{a, b, c\} \cup \{d, e\}$. Equivalence classes are disjoint sets; hence, we identify $c^{-1}(T_1)$ with $\{a, b, c, d, e\}$.

We are now able to define the main precondition (pre1) for Hasse diagrams to be representations of structure–fate relationships.

pre1. There must be subsets T_i ($i = 1, 2, \dots, t$) with the following properties:

$$|T_i| > 1 \quad (6)$$

$$T_i \cap T_j = \emptyset \text{ for } i \neq j \quad (7)$$

$$x \parallel y \text{ for all } x \in T_i \text{ and all } y \in T_j; \quad (8)$$

$$\left| \bigcup_{i=1}^t c^{-1}(T_i) \right| / |G| > \delta \quad (9)$$

Remarks. In this paper, the T_i values are just found by optical inspection of the Hasse diagrams.

The crucial conditions are given by eqs 8 and 9. Equation 8 demands that elements in T_i and T_j are not connected by directed edges. T_i and T_j are separated. Equation 9 means that the union of the disjoint sets $c^{-1}(T_i)$ should be meaningful, in order to be relevant. The quantity δ is related to the density of G/\mathcal{R} ,³¹ which will be more closely examined in a forthcoming paper. Intuitively, δ should be at least >0.5 ; however, we need more practical experience in order to give an appropriate range for δ .

A Hasse diagram represented by a rooted directed tree is, thus, a good candidate for being a structure–fate representation. The branches of the tree may be considered as T_i .

In general, a poset $[T, \varphi(D)]$ will not be decomposed into subsets which are completely separated. Hence, it is useful to count the number of elements of completely and mutually separated subsets T_i and compare this number with the number of all elements in G (note that, here, we count the equivalent elements explicitly). We call τ the family of the subsets T_i :

$$\tau = \{T_1, T_2, \dots\}$$

Therefore, the expression on the left-hand side of eq 9 is called the effectiveness of complete separation, $\text{sep}(G, \tau)$. Explicitly:

$$\text{sep}(G, \tau) = \left| \bigcup_{i=1}^t c^{-1}(T_i) \right| / |G| \quad (10)$$

The larger the quantity $\text{sep}(G, \tau)$ is, the better-suited the members of the family τ are for representing possible structure–fate relations, because large values in $\text{sep}(G, \tau)$ mean that the subsets T_i are the dominating parts of G . A notational remark may be useful: The effectiveness of separation depends on the family of subsets T_i selected. Note that this quantity also depends on the specific poset, which we take from the set X by means of φ . Therefore, we also write $\text{sep}[G, \varphi(D)]$, or even simply $\text{sep}(G)$, if the selection from the set X is of no importance.

Figure 7 may be a useful example. If in the set X of posets no set can be found that fulfills the requirements of pre1,

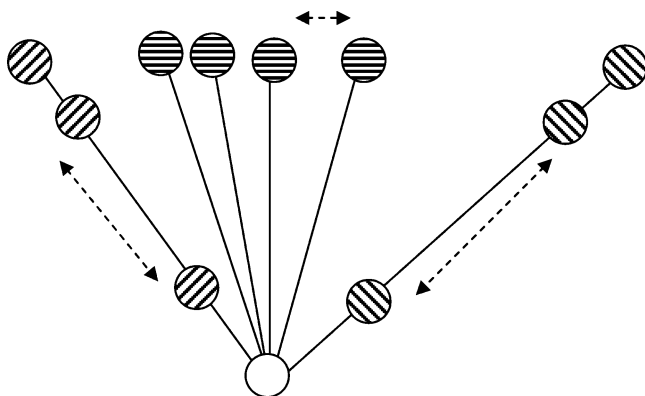


Figure 8. Hasse diagram fulfils prec1. Two branches are well-separated and contain many elements. The bundle represented by horizontal hatched circles may or may not represent relevant parts for a structure–fate relationship.

then within the environmental system, the used model, the used descriptors D_i , and the mappings φ there is no structure–fate relationship. Perhaps intrinsically such a relation does not exist under the above-mentioned conditions. One reason for this failing is that the descriptors may not be sensitive to structural changes.

Now, assume that an appropriate element of X was found. For example, a complete antichain would trivially lead to $\text{sep}(G) = 1$. Therefore, we need another parameter, describing the degree of nontriviality, NT, with respect to realized comparability relations (notation: $x \perp y$, meaning $x \leq y$ or $x \geq y$).

Let us consider $T_i \subseteq G$ and

$$C(T_i) = |\{(x_1, x_2), x_1 < x_2, x_1, x_2 \in T_i\}| \quad (11)$$

We assume that the elements of T_i formed a chain. Then, the number of all comparabilities is

$$\text{Ch}(T_i) = |T_i|(|T_i| - 1)/2 \quad (12)$$

To exclude well-separated but trivial posets, like antichains, we define

$$\text{NT}(G, T_i) = \frac{C(T_i)}{\text{Ch}(T_i)} \quad (13)$$

One may wonder why one does not simply count the elements of T_i but the comparable pairs as well. The reason is best explained by a simple Hasse diagram (Figure 8).

Figure 8 shows that it is meaningful not to consider all parts of a Hasse diagram simultaneously. The antichain, represented by the horizontal hatched circles, might represent a common structure of the chemical compounds, but a causal relation with the descriptor values can hardly be established. It should be possible to relate the structural characteristic with an intensity of some descriptor constellations. If T_i were a chain and a common structure is found for all elements belonging to the chain, then obviously the structures can be ordered. Hence, we get more additional insight into the structure–fate relationships the more “chainy” the subset T_i is.

In the case of Figure 7, in the left-hand diagram, $C(T_1) = 9$, $\text{Ch}(T_1) = 10$, and $\text{NT}(G, T_1) = 9/10$.

3.3.2. Counting the Chemical’s Structures. Within the subsets of G (note that we are back at the set of all elements), there are compounds bearing different structures. We will not discuss how we can define similar or dissimilar structures. Instead, we keep this question open and assume that it can be decided whether a common structure is found or not and that we are able just to find a priori common structures.

Let $|[c^{-1}(T_i, \text{st})]|$ be the number of compounds in $c^{-1}(T_i)$ which bear the structure “st”; then, we introduce

$$d(T_i, \text{st}) = \frac{c^{-1}(T_i, \text{st})}{c^{-1}(T_i)} \quad (14)$$

The quotient of the number of elements belonging to T_i and bearing structure st and of all elements belonging to T_i expresses the degree to which the structure st is associated with a well-separated part (see prec1) of the poset. So that there will be no confusion, we also write $d(i, \text{st})$.

The condition to find from X , the set of all posets, that poset which is suitable for structure–fate relations can now be given as follows:

Find that φ that
fulfils prec1 and that
maximizes $\text{sep}(G)$, NT, and $d(i, \text{st})$ (15)

Clearly, without an appropriate computer code, this task is not manageable. Different φ ’s are to be defined, the Hasse diagrams constructed, and the precondition (prec1) tested; then, an appropriate family t must be found, and the values $\text{sep}[G, \varphi(D)]$, $\text{NT}(T_i)$, and $d(i, \text{st})$ must be determined. Because more than two subsets of G are possible, let there be k subsets which are order-theoretically separated; then, a full documentation of a Hasse diagram representing a structure–fate relationship would be made by writing down all $k(k - 1)/2$ triples. Instead, here, a heuristic methodology is chosen: Let us perform one of the possible classifications, check the resulting Hasse diagram, and characterize it by these three numbers (see section 4).

4. RESULTS

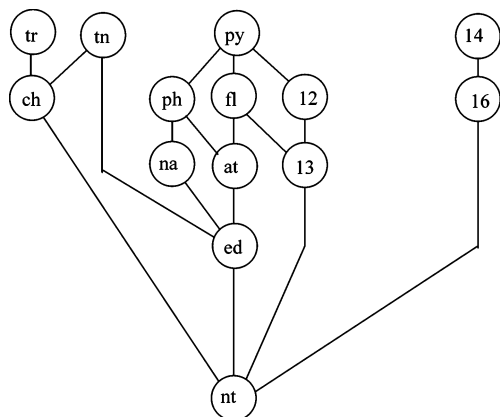
4.1. Data Handling, That Is, Finding an Appropriate Mapping φ .

The model EXWAT was applied to 19 chemicals which were monitored in the river Main, Germany. Hence, the set G consists of 19 chemicals. Four descriptors are derived from the model EXWAT, namely, volatilization, sedimentation, persistence, and advection. Hence, set IB contains four descriptors. All in all, we arrive at a 19×4 data matrix. As a preprocessing step toward an evaluation of the chemicals, a classification was performed in order to avoid irrelevant data differences. In contrast to the use of models for a prognosis, this classification can be rather coarse because the main characteristics of the chemicals are of interest (for details, see ref 32). The descriptors were considered as random variables, where the values from the 19 chemicals are realizations by which a distribution density function can be estimated. With the exception of the persistence, quartiles were calculated. If a realization falls in the first quartile, the class-score gets the value 1; in the second quartile, the class score gets the value 2, and so forth. As orientation rules we consider increasing the values of D_i

Table 1. Classification Results of the Descriptors^a

ID	name	id	volat.	sedim.	persist.	advect.
1	naphthalene	na	3	2	2	3
2	phenanthrene	ph	3	2	2	4
3	pyrene	py	3	3	2	4
4	fluoranthene	fl	2	3	2	4
5	nitrilotriacetic acid (NTA)	nt	1	1	0	1
6	ethylenediaminetetraacetic acid (EDTA)	ed	1	1	1	3
7	chloroform	ch	4	1	2	2
8	tetrachloromethane	tt	4	1	2	3
9	trichloroethane	tn	4	1	2	3
10	trichloroethylene	tr	4	2	2	2
11	tetrachloroethylene (PER)	pe	3	2	2	3
12	PCB 28	12	3	3	2	2
13	PCB 52	13	2	3	2	2
14	PCB 101	14	2	4	2	1
15	PCB 138	15	2	4	2	1
16	PCB 153	16	1	4	2	1
17	PCB 180	17	1	4	2	1
18	atrazine	at	1	2	2	4
19	nonylphenol	no	2	3	2	2

^a Explanation of abbreviations: id = identification of chemical; volat. = volatilization; sedim. = sediment; persist. = persistence; advect. = advection.

**Figure 9.** Hasse diagram, based on the data of Table 1 and the order relation, eq 3. Only representatives of equivalence classes are shown.

(and of S_i) as an indication of an increased hazard level. Thus, the following Table 1 results: the transition from the descriptors to the scores in Table 1 is the result of one φ . It is now a question of how the resulting digraph (i.e., Hasse diagram) can be considered as a structure–fate relationship.

4.2. Generalized Structure–Fate Relationship. Each row of Table 1 represents a chemical by the values of four scores. These four scores are examined with respect to eq 3. Instead of a boring table of \leq relations, a Hasse diagram is drawn (Figure 9).

With the data handling procedures of the equivalence relation, \mathcal{R} , equality in the values of the scores arises, according to eq 5. Hence, the Hasse diagram shown in Figure 9 represents the partially ordered transversal set $T = \{\text{tr, tn, ch, py, ph, fl, 12, na, at, 13, ed, nt, 14, 16}\}$. The precondition (pre1) is fulfilled because, by visual inspection, three subposets can be defined: $T_1 = \{\text{tr, tn, ch}\}$, $T_2 = \{\text{py, ph, fl, 12, 13, na, at}\}$, and $T_3 = \{14, 16\}$.

The nontrivial equivalence classes and the selected representatives are shown in Table 2.

Table 2. Nontrivial (More Than One Element) Equivalence Classes Due to the Data Handling Procedure^a

equivalence class	representative
tn, tt	tn
na, pe	na
14, 15	14
13, no	13
16, 17	16

^a The representatives of nontrivial equivalence classes are given in bold.

Table 3. Structure Subposet Matrix Derived from the Hasse Diagram of Figure 9^a

	Structure CCl, like 	Structure PAH, like 	Structure PCB, like
T_1	4/4	0/4	0/4
T_2	1/9	4/9	2/9
T_3	0/4	0/4	4/4

^a The entries of the matrix are the $d(i, st)$ indices.

As a preparatory step to calculating the d index (eq 14), we calculate the three $c^{-1}(T_i)$ sets:

$$c^{-1}(T_1) = c^{-1}(\{\text{tr, tn, ch}\}) = G_1 = \{\text{tr, tn, ch, tt}\}$$

$$c^{-1}(T_2) = c^{-1}(\{\text{py, ph, fl, 12, 13, na, at}\}) = G_2 = \{\text{py, fl, na, ph, pe, no, at, 12, 13}\}$$

$$c^{-1}(T_3) = c^{-1}(\{14, 16\}) = G_3 = \{14, 15, 16, 17\}$$

We associate with the three G_i sets the following structures: G_1 , the structural characteristic, st, is chlorinated compounds with one or two C atoms (abbr.: CCl); G_2 , the structural characteristic, st', is polyaromatic hydrocarbons (PAH); and G_3 , the structural characteristic, st'', is polychlorinated biphenyls (PCB).

Note that this setting is considered as a priori information. With further research, we should be able to automatically find those structures which optimize the triple of parameters, sep, NT, and d .

Now, the characterizing numbers should clarify the assumption of how far this Hasse diagram can be considered as a representation of structure–fate relationships:

$$\text{sep}(G, \{T_1, T_2, T_3\}) = 17/19 = 0.895$$

$$\text{NT}(T_1) = 2/3 = 0.66$$

$$\text{NT}(T_2) = 11/21 = 0.52$$

$$\text{NT}(T_3) = 2/2 = 1$$

The level of nontriviality in T_2 is rather low, because of many $x \parallel y$ relations in T_2 which persist after discretization by the mapping of φ . Because three T_i subsets and three structures are to be considered, we present the index d as a matrix (Table 3). This matrix could be further evaluated; however, we leave this aspect to a forthcoming paper.

The structural separation between $c^{-1}(T_1)$ and $c^{-1}(T_3)$ is very good, whereas the degree of mixing in the subset $c^{-1}(T_2)$ (two PCBs, two other chemicals which do not belong to any of the three structures, st, st', or st'', considered here) prevents a better result in terms of the d values (see discussion).

5. SUMMARY AND DISCUSSION

The only way to correctly combine substance properties is to use deterministic simulation models, by which a deterministic coupling of environmental and substance properties results. Here, the software package E4CHEM was briefly discussed. The partitioning model EXTND, which is part of E4CHEM, allows a graphical representation, called an “exposure map”, which can be considered as a structure–fate relationship. The idea of structure–fate relationships needs a generalization. To exemplify this, the module EXWAT is explained. As an output of the simulation model EXWAT, four descriptors are generated, which are the basis for an analysis by partial order. The result of applying the component order is a digraph which exhibits a characteristic *fate* structure. Clearly, without a structure of the Hasse diagram as shown here, one would never get the idea of applying this kind of graph as a structure–fate relationship. Some critical numbers are developed, to examine the quality of the φ function. Here, the precondition, that nontrivial subsets of T can be found, is obviously given. The NT numbers are quite low, the main reason for which was already explained. Mainly, the high degree of incomparabilities in the subposet T_2 leads to the low NT number. The conclusion is that, obviously, the PAH group is not as homogeneous in its descriptor values as the general evidence might suggest. It is clear that the PAHs represent a transitional behavior between volatilization and sedimentation. Therefore, there is a greater differentiation, leading to more incomparabilities. Taking into account that structure–fate relations cannot be discussed without checking the environmental system, it is striking that the T_2 group is characterized by a high degree of advection. Hence, the conclusion is that a medium tendency to accumulate to sediments may, nevertheless, lead to some degree of sorption to suspended matter, which underlies advective transport. Therefore, the Hasse diagram shows some switching behavior. Very strong sorption leads to an accumulation in sediments and reduces the amount of transport downstream. This is valid for PCBs 14, 15, 16, and 17. The other PCBs and the PAHs considered in T_2 are not deposited in the sediments, and they don't have as high a degree of volatilization as the short chained compounds, so they must have a considerable tendency of being transported downstream.

The d numbers are lowered and are distributed over all three structures, st, st', and st'', if the subset $c^{-1}(T_2)$ is to be examined. Why? The reason is just the same as discussed above. The transient character from being a “water chemical”, being volatilized, or being sedimented is documented by the high variety of structures. PCBs, nonylphenol, atrazine, and tetrachloroethylene (pe) belong to this group too.

We can check this by selecting for T_2 a reduced set, which contains only the PAHs. $T_2' = \{\text{py, ph, fl, na}\}$. Now the set of indices—sep, NT, and d —evolves as follows:

$$\text{sep}(G) = 13/19 = 0.68$$

$$\text{NT}(T_2) = 4/6 = 0.67$$

$$d(T_2, \text{st}) = 0, \quad d(T_2, \text{st}') = 5/5, \quad \text{and} \quad d(T_2, \text{st}'') = 0$$

The NT and d indices show improved values. However, the price is a reduced significance, because $\text{sep}(G)$ was seriously affected by reducing T_2 to T_2' .

So far, the Hasse diagram, shown in Figure 9, is at least partially a good representation of structure–fate relationships in the environmental system “river” and for a set of compounds, which was monitored in the river (river Main, Bavaria/Germany). In Brüggemann et al.,⁸ the exposure map was also applied to some chemicals found in the river Main. It turned out that the PAHs and the PCBs occupy the same field in the g plane, because of their substance properties. The Hasse diagram shows, however, some degree of separation, which must be the result of the specific environmental system, as explained above.

This paper is a first attempt to establish structure–fate relations by partial orders. The future research must lead to an automatic optimization procedure to find good triplets of sep, NT, and d values. There is clearly a conflict, because reducing the content of the subsets will improve the NT and d results but reduce the values for sep. One may find a new set of indices either by providing better definitions or by appropriately combining the three indices NT, d , and sep. This must be checked in further research. There are still many theoretical questions to be solved. Certainly a major one is to clarify how likely it is that a Hasse diagram gets such a structure by chance if the attributes D_i are randomly given. First attempts to clarify this question are discussed in Sørensen et al.³³ In that context, the concept of randomly generated graphs as originated by Erdős and Renyi³⁴ will be of relevance. Furthermore, it should be checked whether the concept of substitutional patterns in posets³⁵ can be successfully applied for the posetic representation of structure–fate relationships. Finally, from a technical–statistical point of view, it is tempting to analyze the structure–subposet matrix with respect to its “purity”. How far do substructures in digraphs represent chemical structures?

Turning back to the starting point of the paper, the risk/hazard assessment aspect, any simulation model of chemical behavior in the environment, even a very specialized one like, for example, that of Fauser et al.,³⁶ could be applied. Descriptors and a diverse set of chemicals should be selected. By introducing an order relation among the descriptor tuples, one aspect of risk assessment may be fulfilled as far as the choice of descriptors and their orientation describe an environmental hazard. Here, in this paper, we want to establish another, additional step, not only the aspect of hazard assessment but also which chemical structures are related to different aspects of hazards. We conclude that there are three structures; two of them are related to processes such as volatilization and sedimentation, whereas one structure, the structure of aromatic hydrocarbons, is related to both processes.

ACKNOWLEDGMENT

We thank the reviewers for valuable suggestions to improve the paper.

REFERENCES AND NOTES

- (1) Proposal for a Regulation of the European Parliament and of the Council Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency and Amending Directive 1999/45/EC and Regulation (EC) {on Persistent Organic Pollutants}. Proposal for a Directive of the European Parliament and of the Council Amending Council Directive 67/548/EEC in order to Adapt It to Regulation (EC) of the European Parliament and of the Council Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals, COM 2003 644 Final. <http://europa.eu.int/eur-lex/en/com/pdf/2003/act0644en03/1.pdf> (accessed mmm yyyy).
- (2) Attias, L.; Boccardi, P.; Boeije, G.; Brooke, D.; de Bruijn, J.; Comber, M.; Dolan, B.; Fischer, S.; Heinemeyer, G.; Koch, V.; Lijzen, J.; Müller, B.; Murray-Smith, R.; Rikken, M.; Tadeo, J.; Vermeire, T. European union system for the evaluation of substances: the second version. *Chemosphere* **2005**, *59*, 473–485.
- (3) The European Chemicals Bureau Home Page. <http://ecb.jrc.it/> (accessed mmm yyyy).
- (4) Verdonck, F. A. M.; Boeije, G.; Vandenberghe, V.; Comber, M.; de Wolf, W.; Feijtel, T. C. J.; Holt, M.; Koch, V.; Lecloux, A.; Siebel-Sauer, A.; Vanrolleghem, P. A. A rule-based screening environmental risk assessment tool derived from EUSES. *Chemosphere* **2005**, *58*, 1169–1176.
- (5) Brüggemann, R.; Drescher-Kaden, U. *Einführung in die Modellgestützte Bewertung von Umwelchemikalien—Datenabschätzung, Ausbreitung, Verhalten, Wirkung und Bewertung*; Springer-Verlag: Berlin, 2003; pp 1–519.
- (6) Brüggemann, R.; Restrepo, G.; Voigt, K. Towards an Evaluation of Chemicals. *WSEAS Trans. Inf. Sci. Appl.* **2005**, *8*, 1023–1033.
- (7) Matthies, M.; Brüggemann, R.; Münzer, B.; Schernewski, G.; Trapp, S. Exposure and Ecotoxicity Estimation for Environmental Chemicals (E4CHEM): Application of Fate Models for Surface Waters and Soils. *Ecol. Modell.* **1989**, *47*, 115–130.
- (8) Davey, B. A.; Priestley H. A. *Introduction to Lattices and Order*; Cambridge University Press: Cambridge, U. K., 1990.
- (9) Schröder, B. S. W. *Ordered Sets—an Introduction*; Birkhäuser: Boston, MA, 2003.
- (10) Brüggemann, R.; Halfon, E.; Welzl, G.; Voigt, K.; Steinberg, C. Applying the Concept of Partially Ordered Sets on the Ranking of Near-Shore Sediments by a Battery of Tests. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 918–925.
- (11) Brüggemann, R.; Altschuh, J.; Matthies, M. X. In *Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Toxicology and Chemistry*; Devillers, J., Karcher, W., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; pp 197–211.
- (12) Brüggemann, R.; Drescher-Kaden, U.; Münzer, B. *E4CHEM, a Simulation Program for the Fate of Chemicals in the Environment*; GSF—Forschungszentrum für Umwelt und Gesundheit: Neuherberg, Germany, 1996; GSF-Bericht 2/96.
- (13) Trapp, S.; Brüggemann, R.; Münzer, S. Estimation of Releases into Rivers with the Steady–State Surface Water Model EXWAT Using Dichloromethane. *Ecotoxicol. Environ. Saf.* **1989**, *19*, 72–80.
- (14) Brüggemann, R.; Trapp, S. Release and Fate Modelling of Highly Volatile Solvents in the River Main. *Chemosphere* **1988**, *17*, 2029–2041.
- (15) Matthies, M.; Trenkle, R. Variability of exposure estimations for hazardous chemicals released to the environment. *Chemosphere* **1988**, *17*, 1471–1486.
- (16) Matthies, M.; Brüggemann, R.; Trenkle, R. In *Proceedings of the Workshop on Environmental Modelling for Priority Setting among Existing Chemicals*; GSF/PUC, Ed.; Ecomed: Landsberg/Lech, Germany, 1986; pp 211–252.
- (17) Behrendt, H.; Matthies, M.; Gildemeister, H.; Görlitz, G. Leaching and Transformation of Glufosinate-Ammonium and its Main Metabolite in a Layered Soil Column. *Environ. Toxicol. Chem.* **1990**, *9*, 541–549.
- (18) Trapp, S.; Matthies, M.; Scheunert, I.; Topp, E. M. Modeling the Bioconcentration of Organic Chemicals in Plants. *Environ. Sci. Technol.* **1990**, *24*, 1246–1252.
- (19) Trapp, S. Modeling uptake into roots and subsequent translocation of neutral and ionisable organic compounds. *Pest Manage. Sci.* **2000**, *56*, 767–778.
- (20) Neely, W. B.; Mackay, D. In *Modelling the Fate of Chemicals in the Aquatic Environment*; Maki, A. W., Cairns J. J., Eds.; Ann Arbor Science: Ann Arbor, MI, 1982; pp 127–143.
- (21) Matthies, M.; Brüggemann, R.; Trenkle, R. In *Informatikanwendungen im Umweltbereich*; Jaeschke, A., Page, B., Eds.; Springer-Verlag: Heidelberg, Germany, 1988; pp 189–199.
- (22) Potter, M. *Set Theory and its Philosophy*; Oxford: New York, 2004; pp 72, 103–112.
- (23) Brüggemann, R.; Schwaiger, J.; Negele, R. D. Applying Hasse diagram technique for the evaluation of toxicological fish tests. *Chemosphere* **1995**, *30* (9), 1767–1780.
- (24) Ivanciuc, T.; Ivanciuc, O.; Klein D. J. Posetic Quantitative Superstructure/Activity Relationships (QSSARs) for Chlorobenzenes. *J. Chem. Inf. Model.* **2005**, *45*, 870–879.
- (25) Brüggemann, R.; Simon, U.; Mey, S. Estimation of averaged ranks by extended local partial order models. *MATCH* **2005**, *54*, 489–518.
- (26) Luther, B.; Brüggemann, R.; Pudenz, S. An Approach to Combine Cluster Analysis with Order Theoretical Tools in Problems of Environmental Pollution. *MATCH* **2000**, *42*, 119–143.
- (27) Brüggemann, R.; Bartel, H.-G. A Theoretical Concept to Rank Environmentally Significant Chemicals. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 211–217.
- (28) Restrepo, G.; Mesa, H.; Llanos, E. J.; Villaveces, J. L. Topological Study of the Periodic System. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 68–75.
- (29) Restrepo, G.; Brüggemann, R. Ranking regions through cluster analysis and posets. *WSEAS Trans. Inf. Sci. Appl.* **2005**, *7*, 1023–1033.
- (30) Jungnickel, D. *Transversaltheorie*; Akademische Verlagsgesellschaft Geest & Portig KG: Leipzig, Germany, 1982; pp 1–192.
- (31) Seidman, S. B. Network Structure and Minimum Degree. *Social Networks* **1983**, *5*, 269–287.
- (32) Brüggemann, R.; Behrendt, H.; Seiler, K. P.; Trapp, S.; Voigt, K. Einsatz der Hasse-Diagrammtechnik zum Vergleich von Chemikalien hinsichtlich ihrer Umweltgefährdungspotentiale. *DGM* **1994**, *38*, 122–127.
- (33) Sørensen, P. B.; Brüggemann, R.; Thomsen, M.; Lerche, D. Applications of multidimensional rank-correlation. *MATCH* **2005**, *54*, 643–670.
- (34) Erdős, P.; Renyi, A. On the Evolution of Random Graphs. *Publ. Math. Inst. Hungarian Acad. Sci.* **1960**, *5A*, 17–61.
- (35) Möhring, R. H.; Rademacher, F. J. Substitution Decomposition for Discrete Structures and Connections with Combinatorial Optimization. *Ann. Discrete Math.* **1984**, *19*, 257–356.
- (36) Fauser, P.; Vikelsoe, J.; Sørensen, P. P.; Carlsen, L. Phthalates, nonylphenols and LAS in an alternately operated wastewater treatment plant – fate modelling based on measured concentrations in wastewater and sludge. *Water Res.* **2003**, *37*, 1288–1295.

CI050424I