

Editorial

Partial Order in Chemistry

Abstract: In chemistry, there is broad use of graphs and graph-theoretical invariants derived from them. Less known are directed graphs, as e.g., motivated in reaction networks. Here, the focus is on directed graphs derived from order theory. General approaches based on the structure of molecules as well as those derived from multivariate aspects in evaluation are discussed.

Keywords: Graph theory, order in chemistry.

INTRODUCTION

Ordering can be seen as one of the basic steps in chemical research. Indeed, in many cases, a first step in developing a theory is to deliver results in sequences or series of chemical entities. Even if we look back into the history of chemistry, one finds sequences, for example sequences of the “value” of metals. The historical sequence of metals’ value differs not too much from our modern concept of the electrochemical series of metals. Discrepancies appear, because in history, the evaluation of metals was also influenced by their rarity and technical usefulness. The periodic system of elements is a highlight of the idea to order elements according to their properties, and in this issue, the reader gets an exciting insight concerning how to extend the idea of a periodic system to molecules. Often, the final outcome of research is in form of a linear order, where any item of interest can be compared with all the others. However, this may be a too high levelled result! It implies that we know all the relations among the objects perfectly so that all comparisons can really be done. More often, we do not know enough about the relations among the objects, and then, a partial order may be the adequate tool to organize the manifold results. The idea behind partial order is that we are accepting not to know all relevant information and consequently the translation into a digraph represents (“models”) our status of knowledge (see El-Basil in this issue for more about mathematical tools to model in chemistry). Whether or not the concepts of partial order can be helpful in motivating further results as hypotheses generator depends on many facts, and this issue of *Combinatorial Chemistry & High Throughput Screening* demonstrates the successful use of thinking in order theoretical concepts.

OVERVIEW ABOUT THE PAPERS

Following the concepts of D.J. Klein, we can divide the papers in three groups:

1. *A priori* partial orders, the structure of molecules implies a partial order, from which properties may be derived.
2. *A posteriori* partial orders, measurable properties characterize chemicals as a multivariate entity. This point of view is mainly applied in environmental chemistry in the context of evaluation of chemicals.
3. *A posterior* partial order can also be used to predict new properties. As in the first group, QSAR (Quantitative Structure Activity Relations in a broad sense, also including non-biological properties) are based on partial order. This time, however, the basis is a set of other more easily available chemical properties.

The special issue of *Combinatorial Chemistry & High Throughput Screening* is consequently organized following this crude grouping.

First Group

In the paper of Hefferlin, we follow him along paths in a botanical garden, where still many other paths are not opened, and we only see sections of that garden. Which subset of small molecules, Hefferlin is discussing, follows which ordering entity? In his exciting contribution, he shows us ways to detect periodicities among small molecules by use of elegantly constructed diagrams, of matrices and of group theory. However, there is no partial order! Perhaps, we still need more research to establish partial orders, which in turn are helpful for us. I think this is an important point. Not in all cases, where we see that there is a partial order we can draw from this insight useful results! Indeed we could draw Hasse diagrams in some of the cases. However, up to now (sic) no deep conclusion could be drawn from them. Nevertheless, Hefferlin speaks of a mathematical procedure, “analogous to the determination of configurations for two or three electrons using Clebsch-Gordan coefficients”. However, Young diagrams and their partial order are useful tools in determination of Clebsch-Gordan coefficients. Hence, future research may indeed relate Young diagrams with the periodic system of small molecules!

A deep insight into mathematical modeling in chemistry is given by El-Basil. In his review, he shows how discrete mathematics, i.e. combinatorics, graph theory, partial order and group theory can be helpful in modeling chemical facts. El-Basil shows what the general idea is: Construct from molecules a graph, characterize this graph by graph theoretical invariants, make use of the Muirhead and Karamata theorems, and establish an order. He demonstrates that observations follow this order. It is interesting to see that the partial order which is of great help in predictions of properties of aromatic compounds can be considered as a red line because partial orders appear everywhere: Indeed partial order becomes just prominent because it was (and is) used as a tool to analyze subset-set-relations of any mathematical structure based on a ground set. So one may expect that the powerful tool of (graph theoretical) Mark tables can also be related to partial order.

Beside the relation between mathematical structures and their ordering, there is in chemistry a more direct way to define partial orders, which finally can be helpful for prediction: The key to this is the graph theoretical approach, in which chemical entities are considered as vertices and which relates the vertices following some specific rules. Here, the chemical entities are atoms or small groups, and the graph theoretical relation is the chemical bond. A natural way to obtain a partial order is then to examine subgraph-graph relations, as is demonstrated for example by Klein. Another natural way is to consider graph theoretical invariants and to discuss for example the branching degree, which in turn is related to the analysis of Young diagrams. One may also start with orders of chemical fragments (as Randić has shown) or even establishing an order (based on their structure) among molecules. In any case, the crucial step is then to define a relation among those chemical entities which also has a predictive power. An interesting example we can see in the paper of Klein *et al.* who introduce in their review the concept of Quantitative Super Structure Activity Relationships. It gives us the very chemical idea that molecules should not be considered in isolation but in relation to each other. Many more examples of directed graphs, which can be interpreted as graphs of partial order are discussed in the paper of Klein *et al.* (with many helpful additional references). Klein *et al.* stress the idea of a partial order, which is derived from appropriate “neighbor-relationships”, similar to how the periodic system of elements was established. We may call this kind of partial orders *a priori* orders because they are derived from chemical structural properties.

Second Group

In environmental chemistry, the point of view is changed: Instead of directly relating the structure of molecules with a partial order, one is starting from a data matrix, which describes a set of molecules by a set of properties (either measured or calculated) and from that a partial order is derived (following Klein *et al.* this kind of partial order can be called *a posteriori* partial order). In most cases, the aim is to support a decision about chemicals with respect to their environmental hazard potential. Here, the use of partial order under the aspect of incomplete knowledge is especially fair. We do not know deterministically what the final hazard is exerted to the environment and humans, although we know some properties describing diverse aspects of the environmental hazard. The conventional way is to replace the gaps in deterministic knowledge by subjective preferences. Partial order just avoids this controversially discussed (and not really deterministically resolvable) step. In the review by Brüggemann and Voigt, some basic facts about partial order derived from given properties are described and illuminated by many toy examples.

Clearly the question arises, whether or not a probable linear order can be derived. This is indeed possible. The mathematical way to get a probable total order is a typical combinatorial approach: First, one has to find the set of those linear orders which do not contradict the original partial order. Second, based on this set of linear orders, one can derive average ranks and other useful terms within a setting of probability. In the worst case of n mutually incomparable objects we would then obtain $n!$ linear orders because no permutation can contradict the partial order of n incomparable objects. Any increasing knowledge about the objects will enrich the partial order, i.e. will allow comparing some objects with some others. Correspondingly, the number of linear orders, consistent with the partial order is more and more reduced and the measures of uncertainty, like the variance are decreasing, indicating the increasing sharpness in the results. Hence, the set of linear orders consistent with a partial order (called the set of linear extensions) allows the discussion of evaluation in the setting of probability. So far the theory! However, the estimation of averaged ranks is not yet feasible with the computer programs currently existing. Therefore, there is an urgent need for alternatives.

There is indeed a method, which relates partial order to its “big brother” the theory of lattices. The paper of De Loof *et al.* deepens the concept of linear extensions and relates the derivation of “statistical” quantities (as mentioned above) with the lattice of order ideals. In some sense, it continues the elementary paper of Brüggemann and Voigt and introduces new and important concepts. The actual question how to enrich partial orders in order to be helpful in evaluation procedures is enlightened by the detailed description how to obtain average ranks taking the lattice of order ideals as granted. The example is taken from the other side of environmental chemistry: The chemicals themselves are not evaluated, but geographical regions which are polluted due to short- and long-distance transport of chemicals.

As stressed in the paper of Brüggemann and Voigt, partially ordered sets (abbr: posets) derived from data matrices may be considered as one of the many tools in multivariate statistics. Hence, the idea arises pretty naturally whether it is possible to simplify the evaluation analysis by posets, just by thinking of the properties, which span the data matrix as being a part of a hierarchy. Some properties are illuminating one aspect, some others another aspect - perhaps not as important as the first one. How can we combine these two posets, and how can we infer different preferences, without at the same time introducing too much bias by using weights and selecting special functions to combine the attributes? In the paper of Rademaker *et al.*, the algebraic point of view (shown in Brüggemann, Voigt) is elegantly extended, and methods are shown regarding how we can model the preferences in a most objective way. Clearly, the word, “combination”, as used above must be well defined in order to give it a precise meaning. Rademaker *et al.* explain various possibilities for combining different posets, starting with a simple intersection procedure and ending with a comfortable hierarchical union process. Like the example shown in De Loof *et al.*, the pollution of geographical units is selected.

An alternative to the elegant approach of Rademaker *et al.*, albeit inferring somewhat more bias into the evaluation, is the Method of Evaluation by Order Theory, METEOR. This procedure is described in the review by Voigt and Brüggemann. The heart of METEOR is the assumption that the attributes characterizing, e.g., chemicals or polluted areas, may be linearly combined in a step-by-step procedure. This procedure is in contrast to conventional decision support methods, where all the weights expressing preferences among the attributes have to be found simultaneously. To find weights in conventional decision support systems is in fact a very difficult procedure. Imagine a geographical region that is to be evaluated for more than 10

different aspects of pollution, then, one has to get a feeling for the role of the second digit after the decimal point! The step-by-step procedure intends to avoid this difficulty. In the contribution of Voigt and Brüggemann, pharmaceuticals are selected as the example and evaluated with respect to their data availability in data bases, as this data availability is the crucial step in all further steps of evaluation.

Third Group

If data for chemicals are at hand, a natural step is to relate data with the structure. The detection of relations between molecular structure and molecular properties is the field of Quantitative Structure Activity Relationships. Here the concept, "activity", is thought of as encompassing all macroscopic properties of substances. In the contribution of Klein *et al.* (and to some degree in that of El Basil), it is shown how we can relate structure and (macroscopic) properties of molecules by *a priori* posets. Another way to establish property-property-relationships is shown in the contributions of Duchowicz and Castro and of Carlsen, where a posteriori posets are analyzed.

Duchowicz and Castro give a general overview and discuss conventional statistical methods, too. Starting from the problem of assessing chemicals with respect to their environmental hazard, they give a good introduction into some concepts of partial order and introduce several technical terms which are used in that scientific field. A special focus of their paper is the role of uncertainty in partial order application. Within the context of the QSAR and the problem how to select molecular descriptors for ranking, they explain the concept of genetic algorithms. In section 4, they describe in detail the use of partial order for prediction of environmental relevant chemical properties and illuminate this by many real life examples. There is a close relation to the contribution of Klein *et al.*, when the question of extrapolation is discussed.

For the readers of this journal, it may be of special interest that Duchowicz and Castro are also explaining the use of partial order for searching of pharmacophores. They describe the approach of Randić, where the mutagenic potential of nitrosamines is investigated.

Besides applications of QSAR in the framework of partial order the contribution of Carlsen gives a broad overview about methods to enrich the partial order, which is recently one of the main topics of research. He explains two methods, one is the hierarchical partial order ranking which is based on averaged ranks taken from contextual similar attributes. The averaged ranks themselves are now used as new attributes in order to obtain the next (and in general) enriched partial order. This method can be iterated and is well designed for a hierarchy of criteria. The other one he calls, "accumulating partial order ranking" (APOR), which aggregates certain attributes and constructs a new partial order. He shows that this method can support decisions about "the toxicity" of substances and could therefore be a valuable tool for decision makers, especially if a "battery of tests" is used, where some tests aim at the same effect but are based on different methodologies.

The special issue of *Combinatorial Chemistry and High Throughput Screening* will appear in two parts because of redactional, technical reasons:

First Part: (CCHTS Vol. 11 - No. 9)

Hefferlin

El-Basil

Klein *et al.*

De Loof *et al.*

Rademaker *et al.*

Brüggemann and Voigt

Second Part: (CCHTS Vol. 11 - No. 10)

Voigt and Brüggemann

Duchowicz and Castro

Carlsen

General Articles:

Alcalde

Zheng

Andrea

Chen

Sotelo

DISCUSSION

This special issue of *Combinatorial Chemistry & High Throughput Screening* should be considered as an appetizer. Many very interesting applications or theoretical developments are not covered in this issue (for example the application of posets in visualizing experimental designs). So, what should the reader take home as a message, after having this issue in his/her hands?

Partial order theory can be helpful in four aspects:

1. It is a simple mathematical structure based on subset-set-relations and appears everywhere, where a mathematical structure (for example that of groups) is to be analyzed and subsets are to be discussed.
2. Partial orders appear in counting problems, and are therefore often the appropriate combinatorial tool. The use of the famous Möbius-function should be mentioned in that context.
3. Partial order is a “chemical relation”. On the one hand, it represents in a very natural way the concept of subgraph-graph (substructure-structure) relations, where the way how to construct a graph from a molecule may be different. For example, benzenoid systems may be better represented by Gutman trees, whereas in other cases the best representation is just to associate atoms with vertices and bonds with edges (i.e. molecular graphs). On the other hand, partial order is helpful to order chemical entities when there is not a full knowledge on them available.
4. Multivariate posets: One way to obtain multivariate posets is just to select properties of the chemicals and to compare them directly. Another way is to associate with molecules sequences of numbers (for example sequences of topological indices of a certain type). After ordering any sequence, a set of partial sums can be obtained, and as in the case above, these partial sums can now be considered as new attributes, and based on them an order relation is found.

Looking at all the contributions of this issue, the aspects 1, 3 and 4 can be detected in a more or less detailed manner. In any case, the basic point is that of a comparison. Comparison by what? Answering this question is the basic step. I think that just the idea of looking of the ordering quantity is marvelously represented in the first paper by Hefferlin. Therefore let us start with it!

Rainer Brüggemann

Leibniz-Institute of Freshwater Ecology and Inland Fisheries
Müggelseedamm 310
12587 Berlin
Germany
E-mail: brg_home@web.de