Improving the Predicting Power of Partial Order Based QSARs through Linear Extensions

Lars Carlsen,*,† Dorte B. Lerche,‡,§ and Peter B. Sørensen‡

Department of Environment, Technology and Social Studies, Roskilde University, P.O. Box 260, DK-4000 Roskilde, Denmark, Department of Policy Analysis, National Environmental Research Institute, Frederiksborgvej 399, DK-4000 Roskilde, Denmark, and Department of Chemistry, H. C. Ørsted Institute, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

Received September 7, 2001

Partial order theory (POT) is an attractive and operationally simple method that allows ordering of compounds, based on selected structural and/or electronic descriptors (modeled order), or based on their end points, e.g., solubility (experimental order). If the modeled order resembles the experimental order, compounds that are not experimentally investigated can be assigned a position in the model that eventually might lead to a prediction of an end-point value. However, in the application of POT in quantitative structure−activity relationship modeling, only the compounds directly comparable to the noninvestigated compounds are applied. To explore the possibilities of improving the methodology, the theory is extended by application of the so-called linear extensions of the model order. The study show that partial ordering combined with linear extensions appears as a promising tool providing probability distribution curves in the range of possible end-point values for compounds not being experimentally investigated.

INTRODUCTION

The development of quantitative structure−activity relationships (QSARs)[1] often relies heavily on the application of statistical methods such as multilinear regression (MLR) or principal component analysis/partial least-squares regression (PCA/PLS). These methods, in turn, are based on a series of assumptions, e.g., normal distributions and linearity. As an alternative and more simplified approach, partial order theory (POT)[2] has been introduced as a complement to traditional QSAR methodologies.[3,4] Partial order theory does not require specific functional relationships between the single descriptors, e.g., electronic or structural characteristics and the physicochemical or toxicity characteristics (end points) such as solubility or EC50 values.

Initially, in studies on environmental risk assessment POT has been applied when ordering chemical compounds according to their environmental impact.[2a,5,6] In POT-based QSARs the compounds are ordered twice. First the compounds are ordered according to a given end point, based on selected descriptors characterizing their structural and/or electronic nature (modeled order). This order can then be compared to an order based on the same end points found experimentally (experimental order). If the modeled order closely corresponds to the experimental order, other compounds, not being experimentally investigated, can be assigned a position in the model order and hereby obtain an end-point value based on end-point values for experimentally investigated compounds.

In the original studies on the application of POT for QSAR modeling,[3] only the compounds directly comparable to the unknown compounds are applied. In this study the possibilities of applying additional relationships inherent in the partial order are explored. Thus, the new methodological development comes from the study of the so-called linear extensions. A linear extension is a total order, where all compounds are compared to each other and that exhibits no ordering conflicts with the ordering in the partial order.[7,8]

In principle a study of the total set of linear extensions will allow determination of the most probable interval of end-point values. In every linear extension, an unknown compound will be directly comparable to other compounds that is not necessarily the case in the partial order. This is expected to lead to a probabilistic and more specific estimation of the end point of interest.

Unfortunately, in the case of partial order sets including more than 20−25 compounds, it is not practically possible to identify the total set of linear extensions. This is due to the extensive combinatorial demand. However, a newly developed method to analyze partial orders that are too large to be analyzed using the total set of linear extensions[9] is applied in the present study on QSAR modeling. The basic principle is to apply a randomly chosen fraction (subset) of all possible linear extensions. Thus, the present study elucidates, based on an illustrative example, the applicability of partial ordering and linear extensions as a tool for QSAR modeling.

PARTIAL ORDER THEORY

The theory of partial ordering with special focus on the applicability for QSAR studies has been presented in previous papers.[3,4,10]

* Corresponding author phone: +45 4674 2568; fax: +45 4674 3041; e-mail: LC@ruc.dk.
† Roskilde University.
‡ National Environmental Research Institute.
§ University of Copenhagen.

10.1021/ci010380n CCC: $22.00
Published on Web 05/15/2002

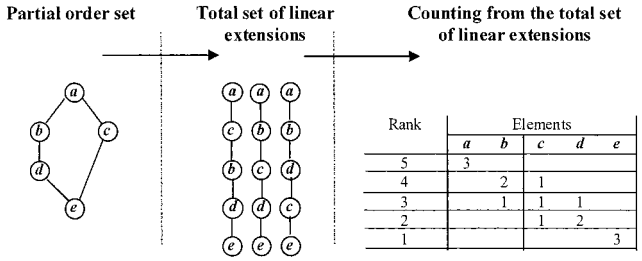

Figure 1. The principle of identification and interpretation of linear extensions from partially ordered sets (for explanation, see text).

In brief, partial order theory (POT) appears as an extremely simple principle, which a priori includes "≤" as the only mathematical relation. If a system is considered that can be described by a series of descriptors p_i, a given compound A, characterized by the descriptors p_i(A), can be compared to another compound B, characterized by the descriptors p_i(B), through comparison of the single descriptors, respectively. Thus, compound A will be ordered higher than compound B, i.e., B ≤ A, if at least one descriptor for A is higher than the corresponding descriptor for B and no descriptor for A is lower than the corresponding descriptor for B. In mathematical terms this can be expressed as

$$\mathrm{B} \le \mathrm{A} \Leftrightarrow p_i(\mathrm{B}) \le p_i(\mathrm{A}) \text{ for } \textit{all } i \qquad (1)$$

Obviously, if all descriptors for A are equal to the corresponding descriptors for B, i.e., p_i(B) = p_i(A) for all *i*, the two compounds will have identical order and will be considered as equivalent. It follows that if A ≤ B and B ≤ C, then A ≤ C. If no order can be established between A and B, these compounds are denoted as incomparable; i.e., they cannot be assigned a mutual order.

In POT—in contrast to standard multidimensional statistical analysis—neither assumptions about linearity nor any assumptions about distribution properties are made. POT may be considered as a parameter-free method. Thus, all descriptors are of equal importance and designations, i.e., "high" and "low" are attributed based on the actual value of the single descriptors only. The graphical representation of the partial ordering (cf. Figure 1) is often given in a so-called Hasse diagram.[2,11]

LINEAR EXTENSIONS

Generally, so-called linear extensions have been used to establish the most probable linear order.[6,9] A linear extension is a total (linear) order, where all the ordering done in the partial ordering is reproduced in the linear order.[7,8] In mathematical terms, a linear extension is the image of an order-preserving map.[12] The principle is illustrated in Figure 1.

Linear extensions are constructed from the partially ordered set. If the number of linear extensions is counted where a selected compound obtains a specific linear order, the probability for the selected compound on the specific order is given. The method takes in a transparent way all information inherent in the partial order into account. As illustrated in Figure 1, it can be seen that compound **a** is related to the highest order only since **a** is placed above all other compounds in the partial order set. On the other hand, compound **c** is equally related to three ordering levels.

Table 1. Calculated Total Rank Probabilities Using Random Selection[9] [a]

	compounds				
rank	**a**	**b**	**c**	**d**	**e**
5	1.00 (1.00)	0	0	0	0
4	0	0.63 (0.67)	0.37 (0.33)	0	0
3	0	0.37 (0.33)	0.26 (0.33)	0.37 (0.33)	0
2	0	0	0.37 (0.33)	0.63 (0.67)	0
1	0	0	0	0	1.00 (1.00)

[a] The values given in parentheses are the corresponding probabilities calculated based on all linear extensions.

Although it appears possible to construct computer programs that identify all possible linear extensions in a partially ordered set,[7] Figure 1 indicates that the actual number of possible linear extensions is closely related to characteristics of the partially ordered set. Thus, the number of comparisons, and consequently the number of compounds in the set, plays the dominant role. It therefore follows that the number of linear extensions will increase dramatically as the number of compounds increases. By applying the combinatorial rule, the upper limit of the number of linear extensions will tend to depend on the faculty of the number of compounds (*N*!, where *N* is the number of compounds). Therefore, in practice it is not possible to identify all possible linear extensions for larger partially ordered sets even by using the most powerful computers.

In a recent study the most probable linear order has been predicted based on a fraction of all possible linear extensions.[9] Hence, random linear extensions are chosen by focusing on the incomparable pairs. Using Figure 1 as an example, two incomparable pairs are present (i.e., **c**−**b** and **d**−**c**). In a first step one of these is randomly chosen, e.g., the pair **d**−**c** for the sake of illustration. In a second step a subsequent random decision on the mutual order of these two compounds is taken. If this decision, e.g., yields that **d** > **c** the compounds are consequently ordered below compound **b** in order to maintain the original partial order **b** > **d**. Thus, the eventual linear order will in this example be **a** > **b** > **d** > **c** > **e** (cf. Figure 1). To obtain the actual probability of the possible orderings of a given compound, this procedure is repeated 2000 times. The estimated ordering probabilities as well as the corresponding values based on all linear extensions are given in Table 1. This type of Monte Carlo Simulation typically allows us to obtain the total order probabilities within an uncertainty of 20%−25% compared to the ordering probability found by using all linear extensions.[9] For larger systems the uncertainty seems to decrease.

ASSIGNMENT OF END-POINT VALUES TO UNKNOWN COMPOUNDS

In a previous paper[3] we have reported on POT-based QSARs for the estimation of solubilities and octanol−water partitioning as an illustrative example of the versatile nature of the technique. The predicting power of POT-based QSARs was discussed and shall be only briefly summarized in the following.

For the compounds placed within the "ordering net" as displayed in Figure 2 (compounds **b**, **c**, and **d**), a verification of the model applicability can be obtained by using the partial

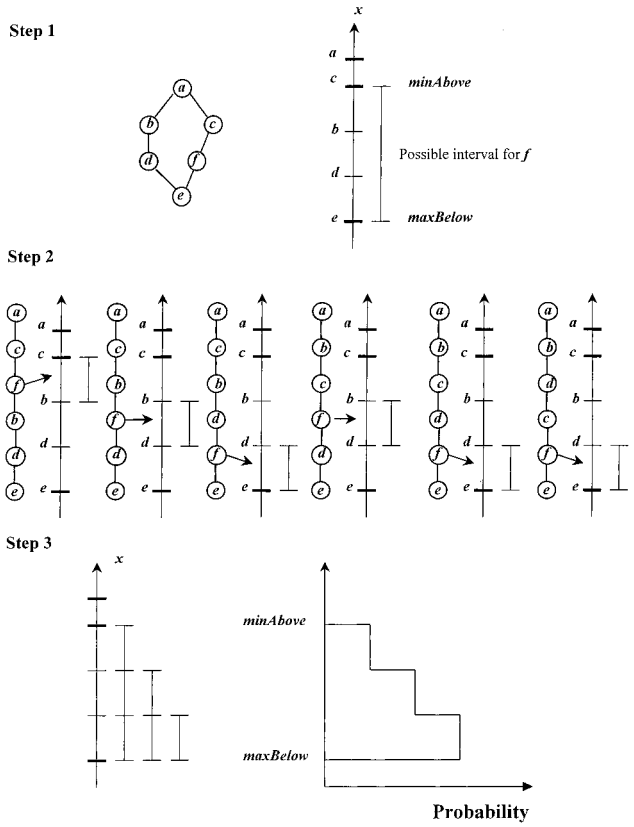

Figure 2. The principle of deriving the probability function based on linear extensions from partially ordered set (for explanation, see text).

order model to predict end points, as e.g. solubilities.[3] The compounds located in the top (maximals; **a** in Figure 2) and bottom (minimals; **e** in Figure 2) layers, respectively, are excluded in this context. The predicted values for a given compound X (ValueX) can be obtained by simple arithmetic means between the lowest value of the comparable compounds ordered above X (minAbove) and the highest value of the comparable compounds ordered below X (maxBelow).

$$\text{ValueX} = (\text{minAbove} + \text{maxBelow})/2 \qquad (2)$$

The predicted values are compared to the corresponding experimentally derived end-point values.

The uncertainty of the ValueX is equally distributed in the interval $\pm 0.5|\text{minAbove} - \text{maxBelow}|$.[3]

In certain cases the distance between the maxBelow and the minAbove values obviously is too large to obtain precise predictions of ValueX. Or alternatively, X is located as a maximal or minimal, with the original approach leading only to a lower or upper limit for ValueX, respectively. In these cases the applicability of the linear extension approach in theory allows us in the first case to estimate the probability of a narrower range and in the second case to estimate the probability of a certain upper or lower limit of ValueX, respectively. Prediction of end points for maximals and minimals is not pursued in the present study.

Let us now assume that the end point of the compounds **a**, ***b***, ***c***, ***d***, and **e** have been determined experimentally (cf. Figure 1). We introduce an compound **f** that based on the selected descriptors is ordered between compounds **c** and **e**; however, it has an unknown end point C_f. The latter can be estimated to be between those of **c** and **e**, respectively[3] (cf. Figure 2). According to the original approach, the predicted value for C_f would be equal to

$$C_f = (C_c + C_e)/2 \qquad (3)$$

The principle of using linear extensions to possibly improve the predicting power of the above original approach is outlined in Figure 2.

In a first step the interval maxBelow − minAbove is calculated as previously described.[3,10] In the second step all linear extensions, or in the case of larger systems a fraction of all linear extensions (vide infra), are formed including compound **f**. For each of the linear extensions the subinterval for **f**, within the interval maxBelow − minAbove, is calculated in a way that the value for **f** most likely will coincide with the actual ordering in the specific linear extension. In the third and final step the individual subintervals are counted and summed. The result forms the probability for the end point of **f** to be within the individual subintervals. Eventually this forms a stepwise probability distribution within the interval maxBelow − minAbove.

RESULTS AND DISCUSSION

In the simple example given in the above section (Figure 2), it is seen that the **f** value most probably will be found in the lower end of the interval of maxBelow and minAbove.

From the graphical representation, the so-called Hasse diagram, it is clear that compound ***f*** is not comparable to the compounds **b** and **d**. However, the relationship between **f** and **b** and **d**, respectively, is not equal. Compound **b** can be claimed to exhibit a tendency to be higher ordered than compound **f**, which is obvious from the linear extensions (Figure 2). This is also indicated in the simple diagram (Figure 2) as compound **b** has two compounds under and one above while **f** has one compound below and two above. Thus, there is strong evidence for compound **b** to ordered "high" compared to compound **f** in agreement with the calculated high probability for the end-point value of **f** to be in the lower end of the maxBelow and minAbove interval.

As an illustrative example, we have selected a series of non-hydrogen bond donor compounds, which have previously been studied using MLR-based QSARs.[13] Thus, solubilities[13] for a series of compounds exhibiting rather different structural and electronic characteristics were retrieved. In the present study we have ordered the compounds under investigation by applying the linear solvation energy relationship (LSER) descriptors, i.e., the volume (Vi/100), the polarity (π^*), and the hydrogen bond basicity (β).[13]

In our previous study a total of 46 compounds were ordered with respect to solubilities.[3] In the present study, for which the objective is the verification of the modeling approach, seven of these compounds, for which no experimental solubilities were available, were excluded. In addition, for the 10 compounds that are either maximals or minimals probability distributions were not achievable. In Figure 3A the comparison from the original study between the experimentally derived solubilities and the corresponding values predicted by the model are displayed as derived according to eq 2. Noting that maximals and minimals are excluded from the figure, a total 29 compounds are included in the evaluation. In Figure 3B the deviation of the predicted solubilities, also derived according to eq 2, from the

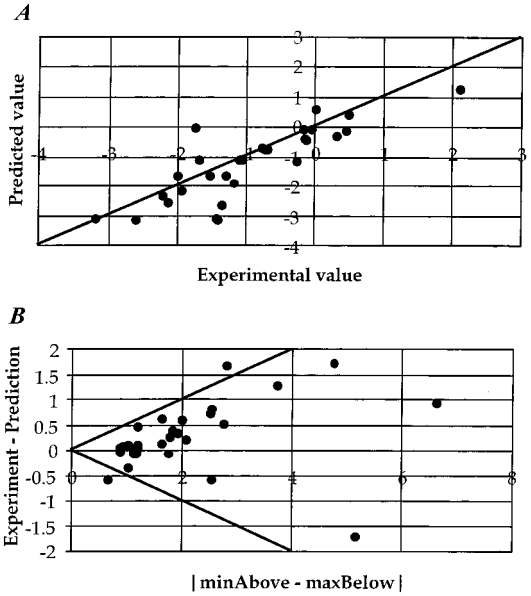

Figure 3. (A) Experimental vs predicted logarithmic solubilities. (B) Deviation of the predicted (using eq 2) solubilities from the experimentally derived values as a function of the distance between the values of minAbove and maxBelow.

Table 2. Experimental and Estimated (Mean, cf. Ref 3) Data for Solubilities for Selected Molecules together with Respective maxBelow and minAbove Values

no.	compound	log *Sw*(exp)	maxBelow	minAbove	log *Sw*(estd)
6	*c*-C_6H_{12}	−3.18	−3.97	−2.22	−3.10
9	$CHCl_3$	−1.12	−1.53	−0.65	−1.09
10	CCl_4	−2.22	−3.18	−1.53	−2.36
12	*t*-CHCl=CHCl	−1.19	−3.18	−0.65	−1.92
14	CHCl=CCl_2	−1.95	−3.18	−1.12	−2.15
15	CH_3CCl_3	−2.00	−2.14	−1.12	−1.63
16	CH_2ClCH_2Cl	−1.05	−1.53	−0.65	−1.09
17	$CHCl_2CHCl_2$	−1.75	−2.61	2.55	−0.03
18	CCl_3CHCl_2	−2.61	−4.52	−1.75	−3.14
19	C_3H_7Cl	−1.53	−2.14	−1.12	−1.63
20	C_4H_9Cl	−2.14	−3.18	−2.00	−2.59
22	$(C_2H_5)_3N$	−0.26	−4.52	2.11	−1.21
24	$(C_2H_5)_2O$	−0.13	−1.44	0.49	−0.475
25	$(n\text{-}C_3H_7)_2O$	−1.44	−4.52	−1.70	−3.11
26	$(i\text{-}C_3H_7)_2O$	−1.70	−1.44	−0.78	−1.11
27	CH_3COOCH_3	0.46	−1.12	0.88	−0.012
28	$CH_3COOC_2H_5$	−0.05	−0.68	0.49	−0.10
29	$CH_3COOC_3H_7$-*n*	−0.74	−1.36	−0.18	−0.77
30	$CH_3COOC_4H_9$-*n*	−1.36	−4.52	−0.78	−2.65
31	$CH_3CH_2COOC_2H_5$	−0.68	−1.36	−0.18	−0.77
34	CH_3CH_2CN	0.33	−1.12	0.53	−0.30
36	C_2H_5–CO−CH_3	0.49	−0.05	0.88	0.42
37	$n\text{-}C_3H_7$–CO−CH_3	−0.18	−0.68	0.49	−0.10
38	$n\text{-}C_4H_9$–CO−CH_3	−0.78	−1.30	−0.18	−0.74
39	$n\text{-}C_5H_{11}$–CO−CH_3	−1.43	−5.54	−0.78	−3.16
40	cyclohexanone	0.01	−0.68	1.86	0.59
41	$n\text{-}C_3H_7CH$=O	−0.15	−1.30	0.49	−0.41
42	$n\text{-}C_5H_{11}CH$=O	−1.3	−2.61	−0.78	−1.70
44	CH_3−CO−$N(CH_3)_2$	2.11	0.01	2.55	1.28

experimentally derived values are displayed as a function of the difference minAbove − maxBelow. The corresponding data representing the 29 compounds can be found in Table 2.

Figure 3B shows that the predictability of the method rapidly decreases for increased minAbove−maxBelow distances.

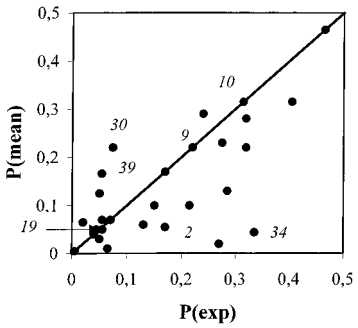

Figure 4. Probability, as derived by the POT using the linear extension approach, at the experimentally derived solubilities, *P*(exp), vs that at the mean values (given by (minAbove + maxBelow)/2), *P*(mean), excluding minimals and maximals.

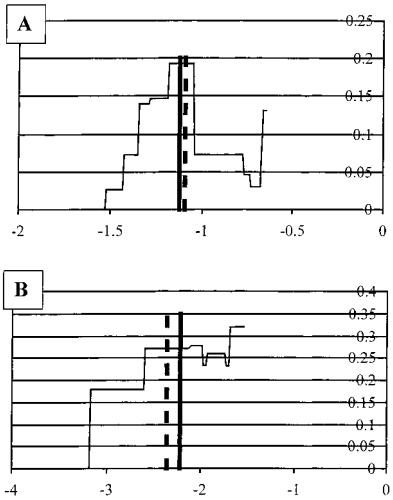

Figure 5. Probability as function of solubility of (A) trichloromethane (**9**) and (B) tetrachloromethane (**10**). The solubility range was limited by the respective minAbove and maxBelow values.[3] The solid and dotted vertical lines correspond to the experimentally derived solubility and the mean value (given by (minAbove + maxBelow)/2), respectively.

To validate the potential of applying the linear extension approach to obtain an improved predictability of partial order based QSARs, the probability that single compounds have solubilities within the range limited by the respective minAbove and maxBelow values were calculated. In Figure 4 the calculated probabilities for the compounds to exhibit the experimentally derived solubility, *P*(exp), are compared to the probabilities corresponding to the originally obtained mean values, *P*(mean), given by eq 2, i.e., (minAbove + maxBelow)/2.[3] The probabilities, *P*(exp) and *P*(mean), are derived as the height of the probability distribution curves (cf. Figures 5−8) at the solubility values corresponding to the experimental solubilities and the mean values, respectively. These two values are compared assuming that if the probability for the experimentally derived value, *P*(exp), is higher than the probability for the mean value, *P*(mean), the probability distribution curve may be able to improve the prediction of end-point values relative to the situation where only the mean value is applied.

It can be seen that the probabilities at the experimentally derived solubilities, *P*(exp), are virtually equal to or higher than the probabilities at the mean value, *P*(mean). For example, compounds **9** (trichloromethane) and **10** (tetrachloromethane) display equal probabilities corresponding to

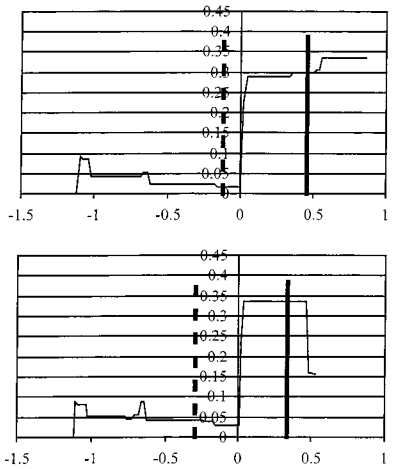

Figure 6. Probability as function of solubility of (A) methyl acetate (**27**) and (B) cyanoethane (**34**). The solubility range was limited by the respective minAbove and maxBelow values.[3] The solid and dotted vertical lines correspond to the experimentally derived solubility and the mean value (given by (minAbove + maxBelow)/2), respectively.

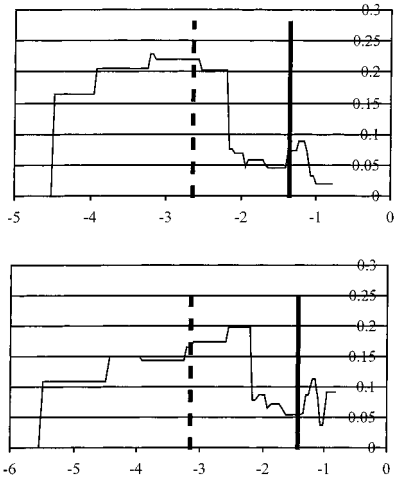

Figure 7. Probability as function of solubility of (A) *n*-butyl acetate (**30**) and (B) 2-heptanone (**39**). The solubility range was limited by the respective minAbove and maxBelow values.[3] The solid and dotted vertical lines correspond to the experimentally derived solubility and the mean value (given by (minAbove + maxBelow)/2), respectively.

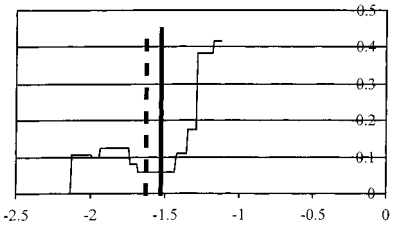

Figure 8. Probability as function of solubility of 1-chloropropane (**19**). The solubility range was limited by the respective minAbove and maxBelow values.[3] The solid and dotted vertical lines correspond to the experimentally derived solubility and the mean value (given by (minAbove + maxBelow)/2), respectively.

the experimental values and the mean values (P(exp) ≈ P(mean)). In the case of, e.g., compounds **27** (methyl acetate) and **34** (cyanoethane) the probabilities corresponding to the experimental values are significantly higher than those corresponding to the mean values (P(exp) ≫ P(mean)). This demonstrates that the correct solubilities are more likely to be located in areas of high-calculated probabilities than in the areas with low probabilities. However, it should be noted that in exceptional cases, as displayed, e.g., by the compounds **30** (*n*-butyl acetate) and **39** (2-heptanone), the reverse trend is observed (P(exp) ≪ P(mean)). Conclusions drawn based on the position of the compounds in the P(exp) vs P(mean) diagram (Figure 4) shall be discussed in the following based on the above-mentioned selected examples.

Compounds being located at or close to the line in Figure 4, e.g., compounds **9** (trichloromethane) and **10** (tetrachloromethane), apparently display generally equal probabilities at the experimentally derived solubilities and the solubilities obtained according to eq 2. In Figure 5 two typical examples are depicted. The figure displays the probability functions of trichloromethane (Figure 5A) and tetrachloromethane (Figure 5B), respectively, within the minAbove − maxBelow interval.

The calculated probability functions appear to be rather broad, and are centered around the arithmetic means of the solubility range covered. Further, it is noted that the probabilities in the central range are relatively high. In these cases the application of the linear extension approach will not lead to further improvement of the predicting power of the method. Selecting a predicted solubility in the area displaying the highest probability will lead to a value close to the true value.

In the cases where compounds are located significantly below the line in Figure 4, e.g., compounds **27** (methyl acetate) and **34** (cyanoethane), a different picture develops as is depicted in Figure 6. In these cases the probabilities at the experimentally derived solubilities apparently are significantly higher than those corresponding to the mean values.

In the two examples shown in Figure 6 it can be noted that the probability functions are dominated by two different ranges located in each end of the solubility range. Thus, we find a low probability range actually including the mean value and a high probability range located in one end of the range of the experimentally derived solubilities. Thus, in these cases the application of the linear extension approach obviously will improve the predictability if the predicted solubilities are selected in the high probability range.

A third case covers the exceptions where the experimentally derived solubilities are found in a low probability range, whereas the mean values are found in a high probability range, e.g., compounds **30** (*n*-butyl acetate) and **39** (2-heptanone) (Figure 7).

In these cases it is obvious that the ability to predict unknown solubilities will not be improved using the linear extension approach. On the other hand, selecting solubilities to be predicted in the high probability range will not cause a reduction in the predicting power of the method compared to the original approach.[3] Thus, it is still suggested to select predicted solubilities in the high probability range.

Special attention should be devoted to the area in Figure 4 where both the experimentally derived solubilities and the mean values are located in low probability ranges, e.g., compound **19** (1-chloropropane). In Figure 8 it can be seen that both values are located in a valley in the probability function between two high probability ranges.

In this case, predicting the solubility to be in the high probability range, e.g., equal to log $Sw = -1.25$, will be a reduction of the predictability compared to the original method.[3] However, the range of possible solubility values is already quite narrow, so compared to the experimentally derived solubility equal to log $Sw = -1.53$, it is in this case still an acceptable prediction. On this basis it is suggested that also in these cases predictions should be made in the high probability ranges.

It should be noted that the present study is a first attempt to apply linear extensions to improve the predictability of partial order based QSARs. Further improvements of the method aiming at the prediction of a precise value, e.g., by introducing the 50 percentile, are in progress.

CONCLUSIONS

The present study, using aqueous solubilities of a variety of organic compounds as an illustrative example, suggests an approach based on the application of linear extensions to improve the predicting power of partial order based QSARs.

The present study has been aiming at a verification of the modeling methodology using a well investigated data set that is nicely handled using conventional QSAR methods.[13] However, the strength of the partial order based QSARs including the use of linear extensions is to be looked for in the nonmetric nature of partial order theory. Thus, the method is able to handle cases not being subject to, e.g., statistical methods such as multilinear regression.

It has been demonstrated that probability distribution curves for the possible solubilities can be derived based on the application of linear extensions. The study suggests that, when predicting solubilities, selecting values in high probability ranges will be an improvement of the method compared to the original approach, where predicted solubilities were chosen as arithmetic means.[3] This can be related to the fact that the majority of the compounds investigated in the present study appear to exhibit an equal or higher probability for the experimental value than for the arithmetic means as displayed in Figure 4.

REFERENCES AND NOTES

(1) (a) Johnson, M. A.; Maggiora, G. M. *Concepts and applications of molecular similarity*; Wiley: New York, 1990. (b) Hermens, J. L. M. Quantitative Structure−Activity Relationships of environmental pollutants. In *Handbook of Environmental Chemistry*; Hutzinger, O., Ed.; Springer: Berlin, 1989; Vol. 2E, pp 111−162. (c) Hermens, J. L. M.; Verhaar, H. J. M. QSAR in environmental toxicology and chemistry: recent developments. In *Classical and 3D-QSAR in Agrochemistry and Toxicology*; (d) *ACS Symposium Series*; Hansch, C., Fujita, J., Eds.; American Chemical Society: Washington, DC, 1996. (e) *Quantitative Structure−Activity Relationships in Environmental Sciences-VII*; Chen, F., Schüürmann, G., Eds.; Soc. Environ. Toxicol. Chem.: Pensacola, FL, 1997.

(2) (a) Halfon, E.; Reggiani, M. G. On the ranking of chemicals for environmental hazard. *Environ. Sci. Technol.* **1986**, *20*, 1173−1179. (b) Brüggemann, R.; Halfon, E.; Welzl, G.; Voigt, K.; Steinberg, C. E. W. Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests. *J. Chem. Inf. Comput. Sci.* **2000**, *41*, 918−925. (c) Brüggemann, R.; Halfon, E.; Bücherl, C. *Theoretical base of the program "Hasse"*; GSF-Bericht 20/95; Neuherberg; 1995.

(3) Carlsen, L.; Sørensen, P. B.; Thomsen, M. Partial order ranking based QSAR's: Estimation of solubilities and octanol−water partitioning. *Chemosphere* **2001**, *43*, 295−302.

(4) Brüggemann, Pudenz, 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.

(5) Halfon, E.; Galani, S.; Brüggemann, R.; Provini, A. Selection of priority properties to assess environmental hazards of pesticides. *Chemosphere* **1966**, *33*, 1543−1562.

(6) Brüggemann, R.; Bücherl, C.; Pudenz, S.; Steinberg, C. E. W. Application of the concept of partial order on comparative evaluation of environmental chemicals. *Acta Hydrochim. Hydrobiol.* **1999**, *23*, 170−178.

(7) Fishburn, P. C. On the family of linear extensions of a partial order. *J. Combinat. Theory* **1974**, *17*, 240−243.

(8) Graham, R. L. Linear Extensions of Partial Orders and the FKG Inequality. In *Ordered Sets*; Rival, I., Ed.; 1982; pp 213−236.

(9) Sørensen, P. B.; Lerche, D. B.; Carlsen, L.; Brüggemann, R. Statistically approach for estimating the total set of linear orders. A possible way for analysing larger partial order sets. *Order Theoretical Tools in Enviromental Science*; Pudenz, S., Ed.; in press.

(10) Carlsen, L.; Sørensen, P. B.; Thomsen, M.; Brüggemann, R. QSAR's Based on Partial Order Ranking. *SAR QSAR Environ. Res.*, in press.

(11) Hasse, H. *Über die Klassenzahl abelscher Zahlkörper*; Akademie Verlag: Berlin, 1952.

(12) Davey, B. A.; Priestley, H. A. *Introduction to Lattices and Order*; Cambridge University Press: Cambridge, 1990.

(13) Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Carr, P. W.; Doherty, R. F.; Taft, R. W. Important differences between aqueous solubility relationships for aliphatic and aromatic solutes. *J. Phys. Chem.* **1987**, *91*, 1996−2004.

CI010380N