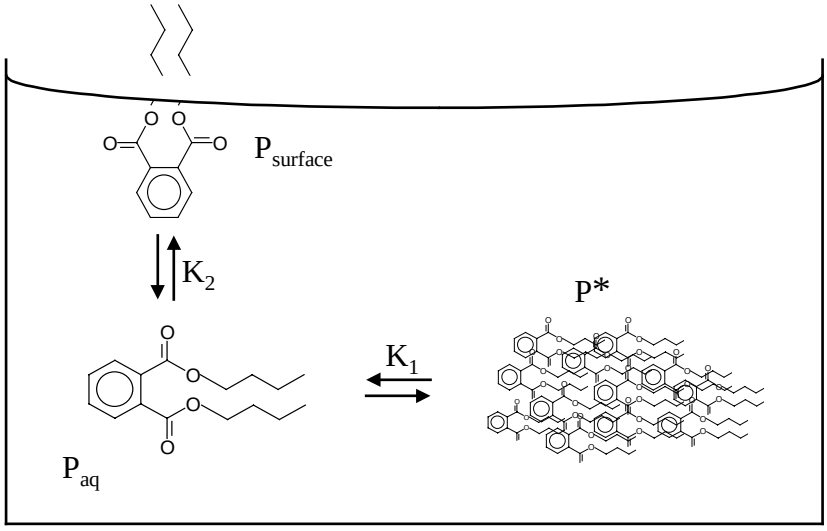
QSARs in Environmental Risk Assessment

-Interpretation and validation of SAR/QSAR based on multivariate data analysis



Marianne Thomsen
Department of Environmental Chemistry
National Environmental Research Institute
DK-4000 Roskilde, Denmark

&

Department of Life Science and Chemistry
Roskilde University
DK-4000 Roskilde, Denmark

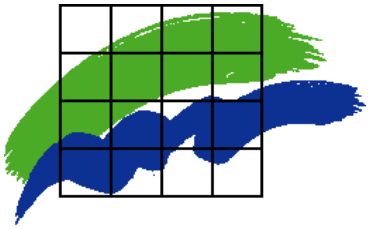
The figure on the front cover is a model representation of phthalates (DnBP) in equilibria in bulk and surface aqueous solution. The partition of phthalates between a liquid phthalate phase and the bulk water phase system is described by the equilibrium, $P^* \rightleftarrows P_{(aq)}$*, and, further, the partition of phthalates between the bulk water phase and the surface,* $P_{(aq)} \rightleftarrows P_{surface}$.

QSARs in Environmental Risk Assessment

-Interpretation and validation of SAR/QSAR based on multivariate data analysis

by

Marianne Thomsen



Department of Environmental Chemistry
National Environmental Research Institute
DK-4000 Roskilde, Denmark

&

Department of Life Science and Chemistry
Roskilde University
DK-4000 Roskilde, Denmark

Abstract

Environmental Risk assessment (ERA) of the more than 100,000 chemical compounds on the European Inventory of Existing Chemicals (EINECS List) is a tasks beyond human, technical and economical resources. An important tool to overcome this hurdle is to use Quantitative Structure-Activity Relationships (QSAR). QSARs are models that quantify endpoints, e.g. physico-chemical properties as well as fixed toxicity parameters, as a function of inherent molecular property descriptors. As such, these are multi-pollutant models that may be used for supplying data of identified hazardous pollutants, where experimental measurements are missing. In this way QSARs may be the link to overcome the backlog with respects to the number of chemicals, which are to be assessed by the EU member states.

In this dissertation, the limitations in the use of QSARs for predicting endpoints, such as partitioning coefficients between different natural occurring phases and fixed toxicity endpoints, for use in ERA is investigated. Furthermore, the potential of multivariate SAR and QSAR for increasing the knowledge of significant structural and electronic intrinsic molecular properties explaining the variations in endpoint values is investigated.

To overcome the limitations in the application of QSARs for supplying data for ERA, as well as for gaining knowledge concerning mechanisms and significant parameters determining the potential hazards of environmental pollutants, an elucidation of uncertainties and unknown parameters which affect the measured endpoint is needed. Simple endpoints such as the aqueous solubility and octanol-water partition coefficients show significant variations between experimental standard methods and specific experimental conditions in the measured system. In this respects, quality assurance, or preprocessing, of the data used for calibrating QSARs, as well as process understanding with respect to the measured system, are shown be crucial for the predictive power of QSARs.

In this dissertation different aspects with respect to the development of scientific valid QSAR are identified: 1) The informational content of the empirical versus the non-empirical and quantum-chemical descriptors has been evaluated, 2) The performance of QSARs based on simple linear regression (LR) and partial least square regression (PLS) have been investigated, and 3) The importance of the quality of data, as well as understanding of experimental/environmental measured systems, to be modelled have been elucidated.

The present ERA concept, as well as the paradigm of QSARs, are based on substance specific properties only and do not include any effects from variations in the nature and characteristics of the natural phases, e.g. soil, sediment and water, in which the pollutants occur. This dissertation focuses on inconsistency and uncertainties in measured endpoints, which result in additional unknown parameters included in the calibration of QSARs. Through investigations of the additional non-quantified uncertainties, or known but not included background data, the quality of data used in the development of QSARs is shown to be critical for the robustness and validity of QSARs. Main aspects are shown to explain the variability in endpoint data found in the literature. These are 1) significant influences of background data, i.e., environmental or experimental parameters such as pH and temperature, 2) the presence of dissolved organic matter (DOM), and 3) the thermodynamic equilibrium description of the pollutant and phases of one and multi-phase systems, when quantifying physico-chemical properties of organic hydrophobic substances.

Through the use of classical statistics as well as multivariate data analysis, the quality of data, interpretations of informational content and model performances of QSARs have been evaluated. Furthermore, the influence of environmental parameters, e.g. pH, temperature, solutions vs. mixtures, and dissolved organic matter (DOM), on the model performance of QSARs have been analysed.

The most critical aspects with respect to the development of scientific valid QSARs seems not to be the model concepts, but high uncertainties and inconsistency in the data used for calibrating QSARs. Concepts of how to overcome the critical aspects and thus make substantial improvements in the applicability of QSARs are proposed.

Supervisors:

Director of Research Department, Dr. Scient., *Lars Carlsen*
and Senior Scientist, Ph.D., *Peter B. Sørensen*,
Department of Environmental Chemistry,
National Environmental Research Institute
Associate Professor, Ph.D., *Søren Hvidt*
and Professor, Dr. Scient., *Poul Erik Hansen*,
Department of Life Science and Chemistry,
Roskilde University
DK-4000 Roskilde, Denmark

Resumé

En miljørisikovurdering (MRV) af de mere end 100,000 kemiske stoffer på den europæiske opgørelse over eksisterende stoffer (EINECS listen) rækker ud over de tilrådighedværende menneskelige, tekniske og økonomiske ressourcer. Et vigtigt supplerende redskab til brug i miljørisikovurderingen af de mange kemiske stoffer er derfor kvantitative struktur-aktivitets relationer (QSAR). QSAR er modeller der kvantificerer såkaldte endpoints, f.eks. fysisk-kemiske egenskaber og toksicitets parametre, som en funktion af kvantitative variable, der beskriver molekylære elektroniske og strukturelle egenskaber. QSARs kan således beskrives som værende multi-stof specifikke modeller, som med fordel kan anvendes til at levere data for identificerede miljøproblematiske stoffer, hvor der er mangler i det eksperimentelle datagrundlag. På denne måde kan QSAR blive nøglen til at få bugt med puklen af risikovurderinger af de mange stoffer, som EU-landede har forpligtet sig til at vurdere miljørisikoen af.

Denne afhandling indeholder en undersøgelse af de begrænsninger, der ligger i brugen af QSAR til at estimere endpoints, som f.eks. ligevægtsfordelingskoefficienter mellem forskellige naturligt forekommende faser såvel som toksiske egenskaber til brug i miljørisikovurderingsprocessen. Endvidere behandles potentialet i at anvende SAR og QSARs baseret på kemometriske data-analyseteknikker til en maksimal ekstraktion af information, og dermed viden omkring betydningen de individuelle elektroniske og strukturelle beskrivende variable i kvantificeringen af endpoints.

En undersøgelse af usikkerheder og ukendte parametre forbundet med de eksperimentelle data er nødvendig, for at overkomme de begrænsninger, der er i anvendelsen af QSARs. Estimering af usikkerheder er således afgørende for at sænke støjniveauet i QSARs. Hermed øges signifikansen af den ekstraherede information og viden omkring bagvedliggende mekanismer, der forklarer de modellerede potentielle endpoints til identifikation af risikoen af miljøproblematiske stoffer. Variansen på eksperimentelle data, bestemt ved forskellige metoder og eksperimentelle system parametre, som f.eks. vandopløseligheder og fordelingskoefficienter er høj. Derfor er kvalitets sikring, eller statistisk behandling af rådata, med henblik på udeladelse af ekstreme data vigtig. Ligeledes udvælgelse af konsistente data mht. eksperimentelle system parametre helt central før data anvendes til kalibrering af QSARs.

I afhandlingen belyses forskellige aspekter, der er centrale for udviklingen af videnskabeligt pålidelige QSARs: 1) Informationsindholdet i QSARs baseret på empiriske henholdsvis non-empiriske og kvantekemiske deskriptorere, 2) modelpræstationer af QSARs baseret på simple lineær regression samt PLS-regression, og 3) vigtigheden af kvaliteten af data, såvel som forståelse af eksperimentelle samt miljøparametre i de målte systemer.

Nuværende MRV koncepter, såvel som QSAR paradigmet, er baseret på stofspecifikke egenskaber alene. Disse inkluderer ikke variationer i baggrundsdata, såsom betydningen af variationer i naturlige systemers iboende egenskaber og karakteristika på det målte endpoint. Baggrundsdata er typisk variationer i jord og sediment karakteristika, samt pH og indholdet af opløst organisk materiale i vandige systemer. Usikkerheder samt inkonsistens i eksperimentelt bestemte data resulterer i, at additionelle ukendte parametre bliver inkluderet i kalibreringen af QSARs. Undersøgelse af de additionelle parametre, dvs. usikkerheder på data, og herunder baggrundsdata, viser at disse er kritiske for robustheden samt prediktionsevnen af QSARs.

Hovedforklaringen på de høje usikkerheder på eksperimentelt bestemte data fundet i litteraturen er: 1) Observerede signifikante variationer i baggrundsdata som pH og temperatur, 2) tilstedeværelse af

opløst organisk materiale samt 3) den thermodynamiske beskrivelse af de organiske miljøproblematiske stoffer samt den valgte fasebeskrivelse af multi-komponent systemer i relation til kvantificeringen af stoffers fysisk-kemiske egenskaber, miljø parametre og toksicitet.

Kvaliteten af data, tolkninger af informationsindholdet i modellerne, og præstationsevnen af QSAR modeller evalueres via brug af klassisk statistik og kemometriske metodikker. Ligeledes analyseres eksperimentelle system parametres indflydelse på model robusthed og prediktionsevne.

De mest kritiske aspekter i relation til udviklingen af prediktive, robuste QSARs er ikke de statistiske model koncepter, ej heller et spørgsmål om valget af empiriske eller ikke empiriske beskrivende variable, men derimod inkonsistens og usikkerheder på de eksperimentelt bestemte data som QSAR-kalibreringen baseres på. I afhandlingen introduceres metoder til at kvantificere og overkomme datakvalitetsproblemet i forbindelse med kalibreringsprocessen af QSARs, hvilket er grundlaget for en forøgelse af anvendeligheden af QSAR i miljørisikovurderingen af kemiske stoffer.

Preface

This dissertation is submitted as part of the requirements for obtaining the Danish Ph.D. degree. The work was supported by a grant from the Danish Research Academy and the National Environmental Research Institute (NERI), Department of Environmental Chemistry. The work was performed at Roskilde University, Department of Life Science and Chemistry, from September 1997 to May 2001 under supervision of Associate Professor, Ph.D., Søren Hvidt and Professor, Dr. Scient., Poul Erik Hansen, the Department of Environmental Chemistry and from NERI, Director of Research Department, Dr. Scient., Lars Carlsen. Lars Carlsen, now Professor in environmental chemistry at Roskilde University was substituted by Senior Scientist, Ph.D., Peter Borgen Sørensen through the last eight months of my Ph.D studies.

The Dissertation includes seven chapters each covering important aspects of the use of QSARs as a tool for assessing the fate and effects of chemicals within the natural environment. The subjects of the individual chapters are shortly outlined below.

Chapter 1 is a short description of the status and problems with respect to the assessment of fate and effects of environmental pollutants. The major limitations for the development of predictive QSARs for use in ERA are introduced.

Chapter 2 gives an overview of the basic principles of risk assessment as well a general description of the purposes of QSARs at a generic EU risks assessment level. In addition, an introduction to the endpoint-studies included in this dissertation is presented.

Chapter 3 is a description of the foundation for the development of scientific valid QSAR models, through a description of the basic assumptions for relating the inherent properties of molecular structures to the molecular activity in a given biotic or abiotic medium. Furthermore, a scheme for how to control and understand the level and type of uncertainties and/or errors in calibration data is presented.

Chapter 4 is a description of the dissolution process of hydrophobic substances, e.g. the phthalates, in an aqueous bulk phase based on classical thermodynamic. The difficulties in direct measurements of true unimeric solubilities of hydrophobic compounds are discussed with reference to the OECD guideline for measuring aqueous solubilities. A method for indirect measurement of the aqueous solubility, the surface tension method, is presented. The influence of the activity of phthalates in the aqueous media, as defined by classical thermodynamics, on measured bioconcentration factors is elucidated.

Chapter 5 is focussing on the complexity of natural organic matter and the influence of the heterogeneity of fixed phase matrices on the standard deviation on QSAR predicted logK_{oc} values for the PAHs and N, S and O hetero-analogues of the PAHs. In addition, the significance of the heterogeneity and variability in the inherent properties of dissolved organic matter (DOM) of difference type and origin is investigated with respect to the unique number status of equilibrium partitioning coefficients. The significance of the inherent properties of DOM in explaining the high variabilities in partitioning values for organic pollutants to DOM is elucidated. An evaluation of the amount and type of extractable information from models based on PLS and simple regression analysis is performed.

Chapter 6 is discussion and conclusion of the main aspect included in the dissertation and Chapter 7 perspectives including further concepts of how to improve the current use of QSARs in risk assessment.

Acknowledgements

Numerous people have inspired and contributed with invaluable help during the last three years. First of all I want to thank the Department of Environmental Chemistry, NERI, for introducing me to the multidisciplinarity of Environmental Chemistry. I want to thank all the people I have been working with through the projects: *Environmentally Hazardously Chemicals* and *Models for Risk Assessment of Chemicals* for fruitful and inspiring discussions concerning different aspects of the complexity of fate modelling, experimental measurements, and data quality analysis.

I want to thank Lars Carlsen for our many stimulating discussions, his enthusiasm, support and encouragement throughout the three years, and Søren Hvidt for his patience through our discussions concerning the thermodynamics of aqueous solutions and colloidal science, and his never failing friendship. Poul Erik Hansen has helped me with the preprocessing and evaluation of NMR analysis through his great expertise and fruitful discussions.

During the last three years I have spent several weeks and days together with Professor Hans-Rolf Schulten, Institute of Soil Science at the University of Rostock, who taught me the universality of life and of natural organic matter, in his very enthusiastic and outstanding way. And I think he managed to convince me of his modelling concepts, even though his very first reply to my response to his average model of dissolved organic matter was: ”The Queen is amused, but not impressed!” I want to thank Professor, Hans-Rolf Schulten, for interesting discussions and exceptional hospitality, especially during the last months of my pregnancy with Sofie.

Through the last year of my project a special thank to Peter Borgen Sørensen for his excellent ability to find a balance between supervising and supporting, and to both Peter Borgen Sørensen and Patrik Fauser for introducing me to the capabilities of the interplay between different kinds of modelling concepts. My deepest regards to Søren Hvidt, Peter Borgen Sørensen and Patrik Fauser for invaluable support during the last year of my Ph.D.-studies.

A teacher at the University of Copenhagen once said to me: ”Don’t be so sad about your many detailed questions Marianne. Don’t you see that you are putting together a jigsaw, while the other students are building with toy bricks**”.** My very special thanks to the friends and colleagues who helped and challenged me to find the missing bricks of my own jigsaw in science and life, when things were difficult.

Above all, my family and friends, who supported me through periods of ups and downs.

Marianne Thomsen
Roskilde, May 2001

Major Concepts, Acronyms and Symbols

ANOVA: ANalysis Of Variances

BCF: Bioconcentration factor

Biomarker: Shortened form of "biological marker"; any alteration in cells or biochemical processes that can be measured in a biological system or sample.

BzBP: benzyl butyl phthalate

Chemical Domain: Phrase used for specifying the predictive capability of QSAR models. Chemical domains of models for estimating properties are class specific, i.e. aromatic/aliphatic esters, phenols, aliphatic amines, or the may be more general and can be used for estimating the properties of hydrophobic compounds.

Chemometrics: a chemical discipline that use applied mathematical and statistical methods to design and select optimal measurement procedures and experiments, and provide maximum chemical information by analysing chemical data.

Classical statistics: in its broadest sense, the science of how to draw conclusions on the basis of observed phenomena accompanied by uncertainties and/or errors.

DEHP: di-(2-ethyl-hexyl) phthalate (for structure cf. Appendix D)

DiDP: di-iso-decyl phthalate (for structure cf. Appendix D)

DiNP: di-iso-nonyl phthalate (for structure cf. Appendix D)

DnBP: di-*n*-butyl phthalate (for structure cf. Appendix D)

DOM: Dissolved Organic Matter (model of an average structure of a DOM macromolecule shown in Figure 15)

EM: Empirical, two-dimensional, descriptors

Endpoint: QSAR estimated key parameter for use in fate or effect assessment, e.g. equilibrium partitioning coefficients between two phases.

Effect assessment: (phrase used analogous Dose-Response assessment) At generic risk assessment level, the transformation of dose-response measurements to fixed endpoint toxicity values, i.e. $L(E)C_{50}$ (Lethal or Effect (e.g. reproduction) Concentration at which 50% of the test population is affected) NOEC (No Observable Effect Concentration) obtained through standard OECD guidelines.

Environmental parameters: Variations in the microenvironment surrounding the molecule such as the DOM, temperature and ionic strength, which show impact on the measured endpoint.

EUSES: (European Union System for the Evaluation of Substances). Computer program consisting of several modules and models, which form the basis of a decision support system, designed for evaluating the risk of substances to humans and the environment[10-13]. The guideline is not legally binding and other methods or approaches may be used by the member states if they are more appropriate, provided that they are scientifically justified and compatible with the general principles laid down in Directive 93/67/EEC[9]. EUSES consist of three modules, which are release estimations, effect assessment and risk characterisation. The exposure assessment is the most complex part of EUSES and consists of a number of compartment models for estimating the direct and indirect human exposure[11-13, 76, 80, 81]. The module for evaluating the release of chemicals into the environment is composed of direct and indirect emission estimates. Direct emission to air, industrial soil, water and wastewater are estimated from use patterns, tonnage and substance properties. Indirect emissions are based on SimpleTreat[81] which is a model for calculating the fate of chemicals in sewage treatment plants, i.e. degradation, evaporation to air, sorption to sewage sludge which are emitted through sludge amendment on agricultural soils[81] and discharge into rivers[80]. Exposure assessment comprise regional models, e.g. SimpleBox[77], for calculating background concentrations, local distribution models for calculating concentrations close to emitters. In addition, one food chain model to calculate the exposure to aquatic and terrestrial and one for estimating indirect exposure to humans[80]. All together a number of model packages are linked together deriving at estimated environmental, i.e. indirect human, exposure concentrations. The direct and indirect human exposure assessment is combined with the effect assessment module in a final Risk Characterisation module (cf. Chapter 2, Figure 1 and 2).

Experimental parameters: secondary parameters, besides the variation in molecular inherent properties, which affect the measured endpoint, e.g. variations in temperature and DOM. (cf. *environmental parameters*). In addition, non-dilute concentration levels of pollutants in systems for measuring partitioning coefficients for which the secondary parameter determining the equilibrium partitioning is the activity coefficient (cf. Appendices A, B and Equations 2.1 and 2.1.).

Fate assessment: (phrase used analogous to Exposure assessment) Modelled and measuring of the processes, e.g. partitioning, transport, degradation and bioaccumulation, which determine the exposure concentration of chemicals within the natural compartments air, terrestrial and aquatic environment. Exposure concentrations are modelled as a function of the fate processes.

Hazard Identification: an evaluation of the potential effects and concerns related to the intrinsic properties of the substance.

LR: Linear Regression

NEM-QC: Non-Empirical and quantumchemical, three-dimensional, descriptors.

PAH: Polycyclic Aromatic Hydrocarbons

PCA: Principal Component Analysis

PCB: Poly-Chlorinated Biphenyls

*PE*C: Predicted Environmental Concentration. Predicted concentration of a substance, which is likely to be found in the environment. It is defined for each environmental compartment (air, water, soil, biota, etc.). The deterministic PEC values may be evaluated through e.g. Monte Carlo simulations or probabilistic uncertainty analysis.

PNEC: Predicted No Effect Concentration. Predicted concentration of a substance below which adverse effects in an environmental compartment of concern are not expected to occur. The PNEC may be calculated by applying an assessment factor to the values resulting from tests on organisms (LD_{50}, LC_{50}, EC_{50}, *NOEL(C)*, *LOEL(C)* or other appropriate methods, e.g. probabilistic uncertainty analysis.

PLS: Partial Least Squares

Property-logK_{ow} regression: Simple linear regression models with logK_{ow} as explanatory variable quantifying various other equilibrium partitioning coefficients (properties) such as logK_{oc}, BCF and Henry's law constant as well as fixed toxicity endpoints. In multivariate QSARs based on the inherent molecular structural and electronic properties quantified through EM or NEM-QC descriptors.

QSAR: Quantitative structure-activity relationship, represent mathematical models relating the observed properties (activities) of chemicals to descriptors of their structure. The molecular structure may be quantified by various descriptors such as molecular surface or physico-chemical properties like 1-octanol/water partition coefficient.

Risks Assessment: Process comprising four steps, i.e., Hazard Identification, Dose-Response Assessment (cf. Effect assessment), Exposure Assessment (cf. Fate assessment) and Risk Characterisation.

Risk Characterisation: Evaluation of RARs (risk characterisation ratios), i.e. matrix-specific PEC/PNEC values (cf. PEC and PNEC) for the natural phases soils, sediments, aqueous media, and air.

Contents

List of papers

Publications/manuscripts, which form the basis of this dissertation, is given below, and will be referred to by the Roman numerals I-VI:

I. M. Thomsen, S. Hvidt and L. Carlsen (2001). Solubility of phthalates revisited. Environmental implications. In, *Handbook on QSARs for predicting Environmental Fate of Chemicals* (J.D. Walker, Ed.). Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA, in press.

II. M. Thomsen, L. Carlsen and S. Hvidt (2001). *Solubilities and surface activities of phthalates investigated by surface tension measurements*, Environmental Toxicology and Chemistry, 20, 127-132.

III. M. Thomsen and L. Carlsen (2001). *Evaluation of Empirical contra non-empirical Descriptors.* SAR and QSAR in Environmental Chemistry, in press.

IV. M. Thomsen, P. Lassen, S. Dobel, P.E. Hansen, L. Carlsen and B.B. Mogensen (2001). *Characterisation of humic substances of different origin: A multivariate approach for qualifying the latent properties of dissolved organic matter (DOM).* Submitted to SAR and QSAR in Environmental Chemistry.

V. M. Thomsen, P. Lassen, S. Dobel, B.B. Mogensen, L. Carlsen and P.E. Hansen (2001). *Inverse QSAR for modelling the sorption of esfenvalerate to dissolved organic matter (DOM). A multivariate approach.* Submitted to SAR and QSAR in Environmental Chemistry.

VI. P. Fauser, M. Thomsen (2001). *Sensitivity analysis of calculated exposure concentrations and dissipation of DEHP in a topsoil compartment – The influence of the third phase effect and Dissolved Organic Matter (DOM)*. Submitted to The Science of the Total Environment.

Additional publications/manuscripts which are less central for the topic of this dissertation:

i. J. Vikelsøe, M. Thomsen and E. Johansen (1998). *Sources of phthalates and nonylphenols in municipal waste water. A study in a local environment.* Department of Environmental Chemistry, NERI Technical Report No. 225, National Environmental Research Institute, Roskilde, 52 pages.

ii. J. Vikelsøe, M. Thomsen and L. Carlsen (1998). Kilder til og forekomst af nonylphenoler og phthalater i spildevand og jord. In, *Diffuse forureninger på landbrugs- og byjord*, Akademiet for Tekniske Videnskaber (ATV).

iii. M. Thomsen and L. Carlsen (1998). *Phthalater i miljøet. Opløselighed, sorption og transport.* Department of Environmental National Environmental Research Institute, NERI Technical Report No. 249, National Environmental Research Institute, Roskilde, 120 pages.

iv. P.B. Sørensen, B.B. Mogensen, L. Carlsen and M. Thomsen (1998). The role of uncertainty in Hasse diagram ranking. In, *Order theoretical tools in environmental sciences* (R. Brüggemann, Ed.) Berichte des IGB, Heft 6, Sonderheft I. Institut für Gewässerökologie und Binnenfischerie im Forschungsverbund Berlin e.V., Berlin, 71-84.

v. M. Thomsen, A.G. Rasmussen and L. Carlsen (1998). *SAR/QSAR approaches to solubility, partitioning and sorption of phthalates,* Chemosphere, 38, 2613-2624.

vi. J. Vikelsøe, M. Thomsen, E. Johansen and L. Carlsen (1999). *Phthalates and nonylphenols in soil. A field study of different soil profiles.* Department of Environmental Chemistry, NERI Technical Report No. 268, National Environmental Research Institute, Roskilde, 128 pages.

vii. P.B. Sørensen, B.B. Mogensen, L. Carlsen and M. Thomsen (2000). *The influence of partial order ranking from input parameter uncertainty. Definition of a robustness parameter*, Chemosphere, 41, 595-601.

viii. L. Carlsen, P.B. Sørensen and M. Thomsen (2000). Estimation of octanol-water distribution coefficients using partial order technique. In, *Order Theoretical Tools in Environmental Sciences* (P.B. Sørensen, L. Carlsen, B.B. Mogensen, R. Brüggemann, B. Luther, S. Pudenz, U. Simon, E. Halfon, T. Bittner, K. Voigt, G. Welzl, and F. Rediske, Eds.), Proceeding of the Second Workshop October 21st, 1999 in Roskilde, Denmark, NERI Technical Report No. 318, 105-115.

ix. J.A. Jakobsen, J. Strand, S. Foverskov, G. Pritzl, P. Lassen, M. Thomsen, K. Toudal and H. Larsen (2000). *Antibegroningsmidler og PAH i havbunden i Vadehavet samt i Esbjerg og Rømø havne*, NERI Technical Report, National Environmental Research Institute, Roskilde, 69 pages.

x. L. Carlsen, M. Thomsen, S. Dobel, P. Lassen, B.B. Mogensen and P. E. Hansen. (2000). The interaction between esfenvalerate and humic substances of different origin. In, *Humic substances. Versatile components of plants, soil and water* (E.A. Ghabbour and G. Davies, Eds.), Royal Society of Chemistry, Cambridge, Special Publication No. 259, 177-189.

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

xii. H.-R. Schulten, M. Thomsen and L. Carlsen (2001). *Humic complexes of di-ethyl phthalates: molecular modelling of the sorption process,* Chemosphere, 43, in press.

xiii. L. Carlsen, M. Thomsen, P.B. Sørensen and R. Brüggemann (2001*). QSAR's based on Partial Order Ranking*, SAR and QSAR in Environmental Chemistry, in press.

xiv. L. Carlsen, P.B. Sørensen, J. Vikelsøe, P. Fauser and M. Thomsen (2001). *On the fate of Xenobiotics. The Roskilde region as case story.* Department of Environmental Chemistry, NERI Technical Report, National Environmental Research Institute, Roskilde, in press.

xv. J. Vikelsøe, M. Thomsen, L. Carlsen and E. Johansen (2001). *Persistent Organic Pollutants in Soil, Sludge and Sediment - A Multianalytical Field Study of Selected Organic Chlorinated and Brominated Compounds.* Department of Environmental Chemistry, NERI Technical Report, National Environmental Research Institute, Roskilde, in press.

xvi. J. Vikelsøe, M. Thomsen and L. Carlsen (2001). *Phthalates and nonylphenols in profiles of differently dressed soils*, submittet for publication in The Science of the Total Environment.

xvii. M. Thomsen, P. Fauser and P.B. Sørensen (2001). *Gaps in knowledge within QSAR´s in relation to risk assessment - Case study on PAHs and Pesticide*. Department of Environmental Chemistry, NERI Technical Report No. x, National Environmental Research Institute, Roskilde, in prep.

xviii. R. Brüggemann, S. Prudenz, L. Carlsen, P.B. Sørensen, M. Thomsen and R.K. Mishra (2001). *The use of Hasse diagrams as a potential approach for inverse QSAR*. SAR and QSAR in Environmental Research, 11, 473-487.

1.Introduction

Environmental pollutants are continuously being emitted from anthropogenic activities in large amounts. They move throughout the ecosystem by a variety of processes and are eventually degraded or accumulated somewhere in the environment. The chemical diversity of pollutants and the variations in behaviour, both with respect to transport, degradation, accumulation (fate) and toxicity (effects) has introduced the field of risk assessment. As a consequence, this has become an essential and necessary discipline in the legislation for controlling the use and release of chemical substances to ensure that they pose negligible risk to the health of humans, wildlife, and the larger ecosystem of which we are only a small part.

To assess the potential risk of chemicals on humans and ecosystems, it is essential to develop a quantitative description of the pathways of contaminant transport and the resulting exposure combined with an estimate of the potential effects on the exposed organisms. More than 100,000 chemicals are included in the EU list of existing chemicals and the fate and toxicological profile data of many of these substances are incomplete[1-4]. New chemicals are continuously being introduced to the environment either directly (primary emission) or indirectly as an unintended result of the use. Furthermore, a large number of substances are formed due to subsequent transformations, e.g., as results of combustion[5-7]. Assessment of all these chemicals based on experimental studies requires tremendous efforts and costs, and such a task is presently beyond existing human, technical, and economic resources. Therefore risk assessment needs to be based on an interdisciplinary approach involving monitoring of exposure concentrations in the natural environment, laboratory scale fate and effect studies and predictive models designed to establish a quantitative link between sources, exposure levels, and risk of effects of potential hazardous pollutants. This will allow for a quantitative description of acceptable risk instead of the management based hazard identification, and no risk acceptance for which the present solution is substitution[8].

Mathematical models, that are able to rank the potential hazards of chemicals by including sources, exposure levels as well as risks of effects, are crucial for keeping the individual pollutant sources at acceptably low levels. This to avoid the problems of unacceptable contamination from excessive sources on one hand, and uneconomic, unnecessary regulations on the other. Such balanced regulation is best effected through full and quantitative information about the substance fate and effects.

Since 1981 a harmonised notification system for new chemical substances has been in force in all Member States of the European Union, which requires a priori risk assessment of new and notified substances[9]. At the present time a generic risk assessment technology at the EU level is described in detail in the technical guidance document (TGD). The TGD includes a detailed description of how to assess the risk of new and existing chemicals, and is presented in four separate parts, i.e., *Risk assessment for human health* (part I), *Environmental risk assessment* (ERA, part II), *Use of (Quantitative) Structure Activity Relationships ((Q)SARs)* (part III) and *Emission Scenario Documents* (part IV)[10-13]. Still, however, many chemicals undergo a satisfactory risk assessment only when the effects have been identified or observed in the environment.

At present, the EU-directive on hazardous substances includes a list of approximately 140 priority pollutants[14-17], which are to be risk assessed by the member states. The work started in 1993[15] and at present time only a limited number of compounds have been evaluated according to the EU risk

assessment scheme. Among these are five EU Risk assessment reports on the phthalates D*n*BP, BzBP, DEHP, D*i*NP and D*i*DP at their finalisation[18-22]. The explanation for the limited outcome seems to be gaps in knowledge on key parameters, in QSAR called endpoints, which are essential for the risk assessment of pollutants to humans and the environment, as well as data of high uncertainty/ambiguity. A way to improve the existing procedures is to optimise the interplay between the generic risk assessment and more fundamental scientific methods. This must be done with the purpose to continuously increase the understanding of the key environmental factors as well as pollutant properties that determine the exposure and potential hazards of pollutants.

One of the methods to exploit in order to advance the science of multi-pollutant risk assessment is Quantitative Structure-Activity Relationships (QSARs)[e.g.23-29]. QSARs quantifies significant relations between the variations in molecular inherent properties of the pollutants and variations in given endpoints, which have been characterised as endpoints for risk assessment. QSARs are used for estimating endpoints for use in compartment modelling for predicting exposure concentrations where experimental data is missing. Furthermore, QSARs are used for predicting fixed toxicity endpoint derived from dose-response measurements, as well as for classifying chemicals according to their mode of action [e.g. 12,30,31].

Endpoints are, e.g., partitioning between the environmental phases such as water and soil, solubility, bioconcentration, and measured exposure concentrations causing, e.g., 50 % of a population to be affected by exposure to a given environmental pollutant[12,32,33]. However, simple endpoint values such as the aqueous solubility and octanol-water partition coefficients show significant variations between experimental standard methods and specific experimental conditions in the measured system[I,iii,v,34-39]. Therefore, the quality of data seems to be a major hindrance for the development of robust and predictive QSARs. With respect to the data quality, classical statistical quality analysis of endpoint values prior to the calibration and validation of QSAR models is used to test the significance of variations in endpoint data with respect to uncertainties between methods and laboratories. Due to the centrality principle of classical statistics, however, the use of mean endpoint values, excluding outliers, as calibration data in QSARs may lead to erroneous results[40-42]. This is due to the presence of significant biases in results. Biases occur between results obtained from different analytical methods, and further due to inappropriate concentration levels of the compound of interest in experimental systems for measuring physicochemical properties. If the assumptions upon which QSAR models are based, are not obeyed with respect to experimental conditions of measured data, then the answers, i.e. information extracted from the models, can be wrong. Additionally, with respect to biological data, underlying factors such as pH, temperature and the composition of heterogeneous mixtures affect the measured endpoint value[I, II, IV, V, v, x, 43-65]. Provided that data are available, experimental background data must be included in the model analysis. The influence of experimental parameters must be investigated, and data classified according to significant variations in experimental conditions if present. If data obey the minimum requirement of QSAR, i.e. that a description at molecular unit level applies to the conditions of the measured system, then separate models for each characteristic system parameter can be developed.

The goal of this dissertation is to set up a scheme for how to improve the ability of QSARs to supply data, where experimental results are missing, conflicting and/or have high uncertainty. The objectives are evaluation of the foundation for the development of scientifically valid QSARs with special focus on experimental background data and data quality evaluations. Through the use of simple linear regression and pattern recognition (or multivariate data analysis) approaches for developing QSARs, the influence of data quality on the model performance is investigated. In addition, the applicability of property-logK_{ow} regression models, used at generic EU risk assessment level[12], is critically evaluated with focus on process understanding and data quality.

2. QSARs for support in Environmental Risk Assessment

2.1 Basic Principles of Environmental Risk Assessment

Environmental Risk Assessment (ERA) can be divided into four major steps: Hazard Identification, Dose-Response (effect) Assessment, Exposure (fate) Assessment, and Risk Characterisation. Hazard Identification is the first step in the process of carrying out effect assessment in both the USEPA[66] and the EU[10-13] risk assessment scheme. This step is followed by an estimation of Dose-Response Assessment for each Identified Hazard. The third and final step is Risk Characterisation, where the measured Dose-Response relationships for each Identified Hazard are compared with monitored or predicted exposure concentration levels. Risk assessment is carried out for all three natural compartments, i.e. the terrestrial environment, the aquatic environment and air. In addition hazard identification and hazard assessment[67] concepts are used for classification and labelling of chemical substances through so-called risk and safety phrases[67-69].

One way of presenting the basic steps of risk assessment is illustrated in Figure 1.

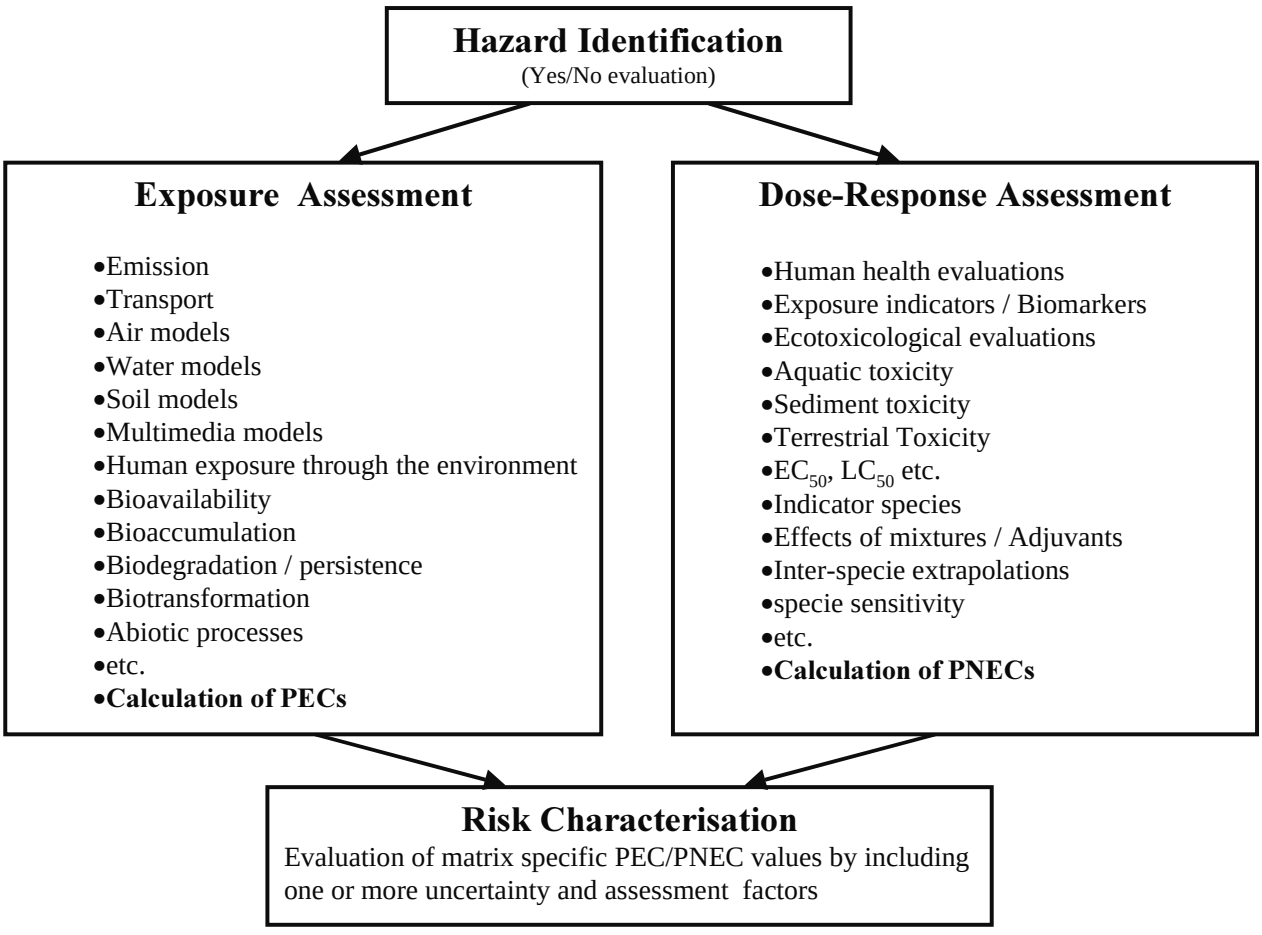


Figure 1 *Basic elements in Risk Assessment. Through Hazard Identification, the potential effects and concerns of a substance are evaluated with respect to the intrinsic properties of the substance. Through risk characterisation, the bioavailable exposure concentrations in all compartments, i.e. air, terrestrial and aquatic environments are assessed and compared to measured dose-response curves. This is done for each identified hazard. PEC is the Predicted Environmental Concentration calculated from compartment models[11]. Input data include monitoring data as well as QSAR estimated data. The PEC values are evaluated based on substance-specific uncertainties only[70-73]. PNEC is the Predicted No Effect Concentration and includes the so-called assessment factor/-s, which account for extrapolation between species[11,70,71]. The final step is evaluation of so-called Risk Characterisation Ratios (RAR), PEC/PNEC, which in term are matrix specific, e.g. for soil, sediment and the aquatic environment.*

Hazard Identification should be understood as an evaluation of the potential effects and concerns related to the intrinsic properties of the substance[69]. Furthermore, through a continuous review of the classification of substances, i.e. hazard assessment, class memberships for new and existing substances are derived[11-28-29].

In Dose-Response Assessment (also called effect assessment), chemicals that are suspected to be harmful are tested in vivo and in vitro on different species. The main goal is the Prediction of No-Effect Concentrations, PNECs, for humans, terrestrial and aquatic organisms. This is done through acute and chronic dose-response measurements, which are transformed to fixed toxicity endpoints that are considered to be appropriate as a quantitative estimation of the intrinsic toxicity of the substance. Fixed toxicity endpoints ($L(E)C_{50}$) used in effect assessment are typically lethal concentrations (LC_{50}), or effect concentrations (EC_{50}) which cause 50 % of the test species to be affected. Furthermore, long-term effects are evaluated through the so-called No-Observable-Effect-Concentration, *NOEC*, e.g. sub-lethal effects such as reduction in reproduction, inhibited growth and/or other adverse effects[32]. *PNEC* values are calculated for the most sensitive of a selected list of organisms according to appropriate OECD guidelines[69], or through the measurement of biomarkers combined with specie-interpolations and specie sensitivity analysis[11,70,71].

Exposure assessment[11,72] (also called fate assessment) is based on monitoring data[i, ii, vi, ix, xv, xvi] and mathematical descriptions of the key processes determining the degradation, transport and distribution of individual substances within and between compartments, i.e. sediment, soil, water and air[xiv,VI]. The objective of environmental fate modelling is to estimate predicted environmental concentration PEC[11,71,72] in the natural phases of soils, sediments, air, waters and biota. In each compartment the partitioning between the different phases and the processes taking place in these phases form the basis for the calculation of PEC.

In addition to the list in Figure 1 there are additional elements such as production, use and emissions, which are treated specifically in Substance Flow Analyses[74]. These elements represent the natural starting point in a fate assessment[69,72,75]. The evaluation of PEC values only includes substance-specific uncertainties, e.g. partitioning coefficients, degradation rates and emission factors. Uncertainties of environmental parameters, such as the organic matter content in soil, dilution factors, temperature, bacterial population, are not included in the guidelines for probabilistic risk assessment[71-73].

The overall risk assessment combines the derived PNEC and the PEC values in a resulting risk characterisation ratio, the PEC/PNEC-ratio, which is an estimate of whether the chemical presents a risk to man and/or the environment[9,69]. The assessment factor/-s that is/are included account for unknown knowledge with respect to inter-specie sensitivity, i.e. variations in sensitivity between species, are purely empirical[71]. Furthermore, no uncertainties due to the model concept of assuming well-mixed compartments (cf. SimpleBox, Section 2.2) are included in the calculation of PEC[71,VI].

At present there is a tendency for decision-makers to base their judgement of chemicals on Hazard Identification and no-risk acceptance, for which the present solution is a prohibition against the use, or substitution, of the chemical[8]. The no-risk acceptance may seem inappropriate, when examples like, e.g. addition of methyl *t*-butyl ether (MTBE) to gasoline, can occur and remind us that introduction of non-assessed chemicals can cause new, not foreseen, problems. Efforts must therefore be made in the future with respect to optimising the relationship between molecular properties of chemicals with endpoint parameters for use in exposure and dose-response assessment in a preventive risk assessment approach. Furthermore environmental parameters and system characteristics need to be included[I,II,IV,V,VI].

2.2 Conventional QSARs in Exposure and Dose-Response Assessment

The EU risk assessment is supported with the computer program EUSES (European Union System for the Evaluation of Substances). EUSES comprises a number of model packages[72] that aim to:

1. Predict intermedia concentration ratios for the purpose of harmonisation of environmental quality objectives for air, water, sediment and soil.

2. Predict exposure concentrations in the environment for the purpose of evaluation of chemicals.

EUSES is based on the principles outlined in the Technical Guidance Document for New and Existing Substances[10-13] (cf. List of Concepts, Acronyms and Symbols, p. viii). With respect to the fate of chemicals within the natural environment, an important application in EUSES is SimpleBox[72], which is a multi-media fate model based on the concepts formulated by Mackay (1991)[75]. The environment is modelled as consisting of a set of well-mixed, homogeneous compartments, i.e. air, two water compartments, sediment, three soil compartments and two vegetation compartments in regional, continental and global scales[72]. Emission rates, distribution coefficients and rate constants are used for modelling transport and transformation, and performing steady-state and non-steady-state computations, deriving PEC values for specific environmental situations. The final result is predicted environmental distributions of chemical substances within the natural environment.

Compartment models, such as EUSES, are based on input data such as measured equilibrium partitioning coefficients between the different natural occurring phases within and between compartments as well as degradation coefficients and e.g. sorption/desorption rates. A central issue of QSARs is to quantify these properties for substances where data are missing as well as to validate the measured, e.g., equilibrium partitioning coefficients of the substances between the different phases, i.e. air, soil/sediment and water. Therefore, in order to improve environmental risk assessment models there is a strong need for the development of validated and predictive QSARs.

Four purposes for the use of QSARs are listed in the TGD for the assessment of new notified and existing chemicals[12]. These are:

1. *to assist in data evaluation;*

2. *to contribute to the decision making process on whether further testing is necessary to clarify an endpoint of concern and, if further testing is needed, to optimise the testing strategies, where appropriate;*

3. *establishing (input) parameters, which are necessary to conduct the exposure and/or effect assessment;*

4. *Identifying effects, which may be of potential concern on which test data are not available.*

The role of QSAR can thus be illustrated as in Figure 2.

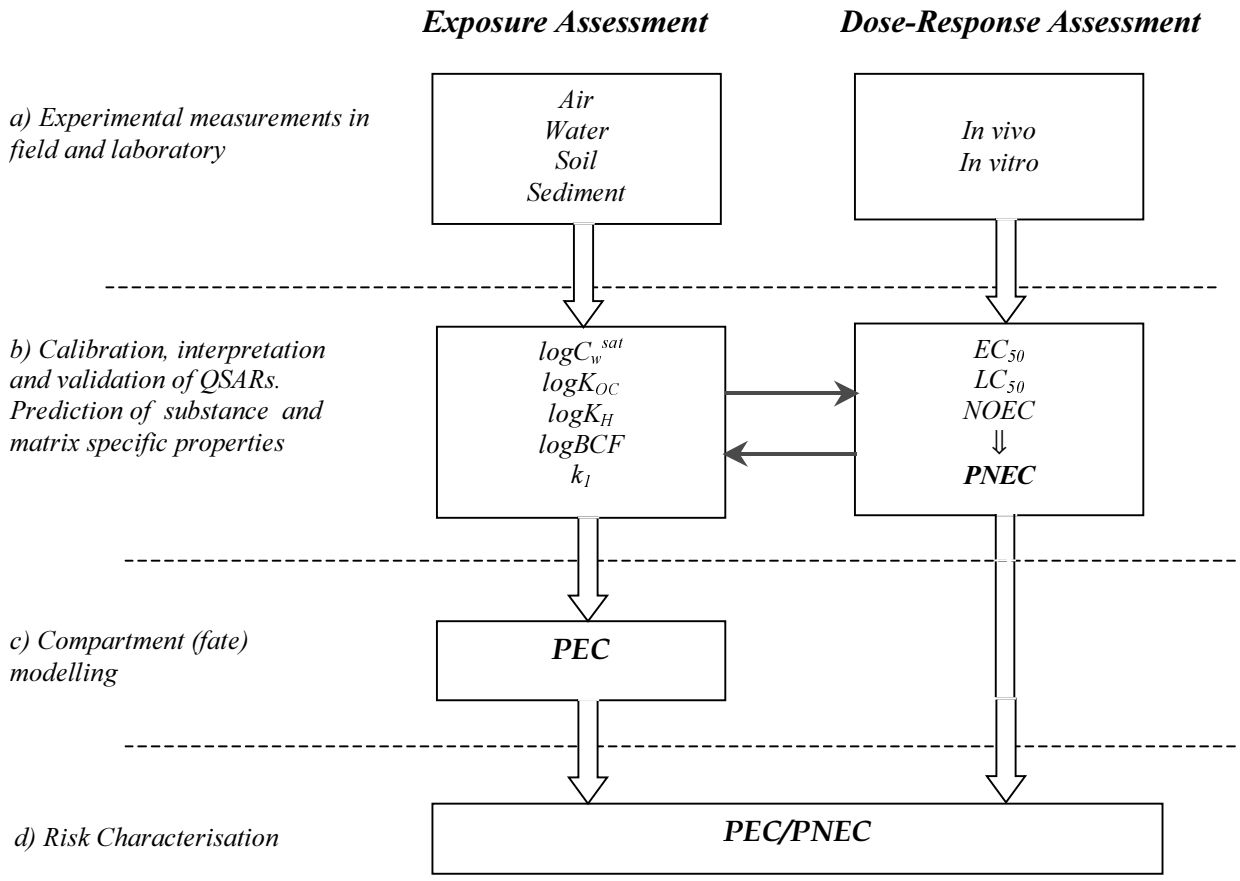


Figure 2 *Progress for Exposure and Dose-Response Assessment. a) Experimental measurements of chemical substances in biotic and abiotic phases in the real environment (monitoring data) as well as at laboratory level, b) Data validation, interpretation and development of QSARs for predicting endpoints of new or existing chemicals, c) Compartment modelling of the indirect human exposure of chemicals, i.e. fate modelling, d) Risk Characterisation.*

QSARs for estimating *n*-octanol water partitioning (logK_{ow}), soil and sediment sorption (logK_{oc}), biodegradation, photolysis in the atmosphere, hydrolysis (rate constants or half-lives in the individual media) Henry's law constant (logK_H), log$L(E)C_{50}$ and log$NOEC$ are implemented in the risk assessment of chemical substances in the present TGD[10-13], which is currently under revision. In general QSAR estimated endpoints are used as input data where data are unreliable or missing[11,12].

With respect to Exposure Assessment (cf. Figure1 and 2) in individual compartments, QSARs are used to estimate physicochemical properties, which affect the fate and behaviour of chemicals in the environment, e.g. solubility in water and partition coefficient between water/sediment, water/soil and water/natural lipids[42,76-96]. QSARs for estimating the solubility of chemicals has received minor attention in the TGD[12], whereas logK_{ow}, is used as a measure of the hydrophobicity, and as reference system, in class specific linear regression models for estimating, e.g., logK_{oc}[e.g.76,79] and logBCF [e.g.84,85,91-96].

Conventionally, the logK_{oc} value has been considered the main key parameter determining the mobility and bioavailability of environmental pollutants in the aquatic and terrestrial compartment system. This is due to the fact that mobility and transport of pollutants is determined by partitioning to fixed as well as suspended organic matter, which increases the persistence, but decreases the bioavailability of chemicals within the environment[e.g.43,45,57,62,102,103,125]. The simple logK_{oc}-logK_{ow} relationships seem to be restricted to non-polar chemicals for which the sorption is driven mainly by hydrophobic effects in the aqueous bulk phase, as well as hydrophobic interaction between the

sorbate molecules and sorbent phase[12,79]. Pollutants with more polar groups may interact with the solid phase via more specific interactions, i.e. analogous to toxicological responses caused by specific mode-of-action. In the EU guideline for QSARs in risk assessment, it is stated that in the absence of data for sorption, the sorption to soil and sediment, logK_d, is calculated from logK_{oc}, which again is derived by regression to octanol-water partitioning[12,127]. There are 19 class specific linear logK_{oc}-logK_{ow} regression equations are implemented in the TGD[12].

With respect to Dose-Response Assessment (cf. Figure 1 and 2), logEC_{50} in used as a measure of the intrinsic toxicity of chemical substances, whereas log$NOEC$–logK_{ow} relationships are used in the evaluation of PNECs. The chemical domains of the regression models are restricted to chemical classes, e.g. esters, phenols and amines. Furthermore, few models based on classification of the chemical compounds into their mode-of-action; i.e. polar and non-polar narcosis, are included in the TGD[12,25,30,31,97-101]. No models for the prediction specific receptor-mediated responses[98] are available in the TGD.

LogK_{ow} has been used in innumerable papers as a key parameter describing the fate and effects of chemicals[e.g.41,79,88,89,96,104,105,139]. This is substantiated by the fact the TGD only includes QSARs based on class specific property-logK_{ow} regressions, and not QSARs based on relations between quantified molecular properties as described in Chapter 3[III, IV, V]. The linearity in property-logK_{ow} regression models is restricted to logK_{ow} values below approximately 6[12]. However, for logBCF-logK_{ow} relationships, three models for estimating bioconcentration of chemicals are included in the TGD, which are linear, bilinear and non-linear, respectively[12,42,92-96,105]. The significance of the non-linear relationships is however not validated, and is therefore based on strictly empirical fitting to inconsistent experimental data (cf. Chapter 4, section 4.2 and Appendix J)[42].

The data available for use in model calibration are of varying quality (cf. Appendix D). The nature of the biotic as well as the abiotic phase varies in the individual experiments. Factors such as specie type and age of the biotic phase, as well as, e.g., characteristics and content of natural organic matter of fixed abiotic phases, may influence the measured endpoint (cf. Chapter 5 and Appendix I). Furthermore system parameters such as temperature, pH, ionic strength and the presence of DOM vary between experiments (cf. Chapter 5, Appendices F and I).

2.3 Important substance specific endpoints treated in this study

The multimedia models are based on the assumption that there exists both an equilibrium distribution and homogeneous mixing in each compartment, e.g. water. The multi-media compartment models can be described as consisting of linked, single medium models, which simulate the physical and chemical processes that drive the transport of chemicals across air/water, air/soil, and water/soil interfaces. In such studies the equilibrium assumptions are not implicit but are derived through solubility considerations in each specific phase. Mackay (1991)[75] has a less restrictive way of expressing it, than described in this dissertation, stating that partitioning between octanol and water is determined by the aqueous solubility, as the solubility of organic compounds in octanol have a narrow range of approximately 0.2 to 2 mol/L[75]. The definition of the equilibrium partitioning coefficient (cf. Appendix B), as well as the definition of the true water solubility (cf. Appendix A and Chapter 4) for specific substances are essential, since these represent the starting point in the calculation of the transport, transformation and distribution processes in the environment. Especially for hydrophobic organic chemicals a correct definition and measurement of the true solubility[I,II,37,38] (cf. Chapter 4) and endpoint measurements for these substances have shown conflicting results[iii,v,38] (cf. Appendix C, D, E and F), which affect the effectiveness and reliability of measurements with respect to the risk assessment procedure. This dissertation includes different aspects of homogeneous and heterogeneous mixtures versus dilute solutions, with respect

to the thermodynamic definition of activity, as well as their influence on key processes as is illustrated in Figure 3.

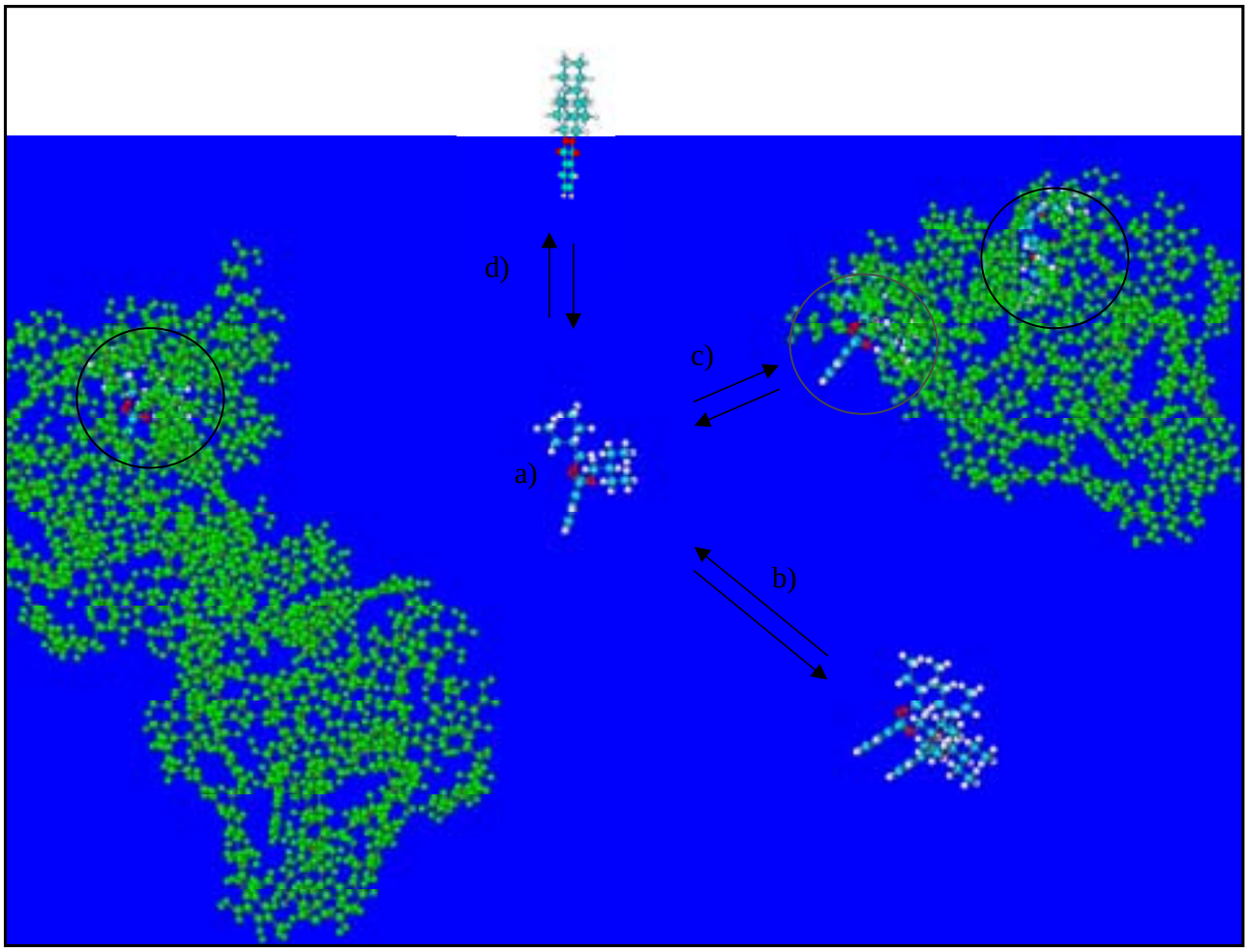


Figure 3 *Exposure Assessment model system illustrating important processes that are included in this work (model structures of dissolved organic matter (DOM), by permission from H.-R. Schulten[106]) . a) Aqueous solubility (Chapter 4, Appendices A and D), b) Partitioning between true aqueous phase and emulsion phase (Chapter 4, Appendices B, D and E), c) Partitioning between true aqueous phase and a colloidal third phase or DOM; black circle illustrating absorption, red circle illustrating adsorption (Chapter 5, Appendix B), d) Partitioning between true aqueous phase and air/water interface (Chapter 4).*

Conventional compartment modelling will normally only take relatively few main compartments into account, e.g. partitioning of a substance between sediment, pure water phase and air. QSAR estimated parameters would in this case be the organic carbon normalised partitioning coefficient, logK_{OC}, and Henrys Law constant, logK_H. However, the natural aqueous compartment, as well as sediment and soil pore water, of every natural system includes natural organic matter. This implies the presence of a third phase effect, which impacts the fate and effect of chemicals in the environmental to an extent that has not yet been fully elucidated. The mobile fraction of natural organic matter is denoted DOM (dissolved organic matter) and its size and structural characteristics depend on, e.g. the concentration level at which it occurs in the aqueous media[IV, V, iii].

The inclusion of DOM, a mobile organic matter phase, is essential with respect to the fate of chemical compounds in the environment. However, due to the assumptions of homogeneous well-mixed phases, EUSES is not able to perform risk assessment on heterogeneous systems[39,69], as it does not consider spatial distribution of the chemical, e.g., sorption at interfaces of compounds of surfactant nature[107-110]. By modelling fate for a system constrained by a pure water phase the adsorption to fixed organic material in soil and sediment, implies a simple relationship, i.e. that a

large sorption capacity leads to low mobility and reduced bioavailability. In reality the presence of DOM may influence the fate and effects of pollutants simply by increasing the apparent solubility as illustrated in Figure 3[V,45,49,54,56,57]. Through a change in the activity of the pollutant in aqueous bulk phase, and consequently important processes such as uptake by organisms, microbial degradation, partitioning to abiotic and biological matrices change. Therefore, the activity in the aqueous bulk phase is a central parameter with respect to endpoints within exposure as well as dose-response assessment.

The simple property-logK_{ow} models are based on the use of octanol phase as the descriptor for the bulk properties of homogeneous lipid phases of biological matrices and the organic content of the solid matrices sediment and soil (cf. section 4.2 and 5.1 and Appendix B). The aqueous phase is included in most natural partitioning processes in the environment as is illustrated in Figure 3. Therefore, the aqueous solubility of environmental pollutants may in some cases be an overlooked parameter of great importance with respect to processes determining the fate and effects of chemicals within the natural environment (cf. paper I, II, VI). Processes to consider include:

- Mobility, through infiltration of chemicals in soil columns, molecular diffusion, surface runoff, transport in continental surface waters and from continental surface waters to the sea.

- Microbial degradation.

- Bioavailability and bioaccumulation.

- Partitioning to different organic liquid and solid phases within and between the different phases air, water and sediment/soil.

Within the area of environmental chemistry partitioning coefficients are preferred before the solubility in linear relations for use in environmental risk assessment. The most critical aspect, however, are the use of partitioning coefficients defined as concentration ratios of pollutant in the respective phases as unique values. If the assumption used for describing a given system is wrong (cf. section 3.1.1 and Appendices A and B), then the answer extracted from data and models are wrong.

Equilibrium partitioning coefficients, quantified as the ratio of the molar concentration of component *i* in a given abiotic or biotic phase and the aqueous bulk phase, is only valid based on the assumption of dilute solutions (Appendix B, Equation B12 and B13). This means that the uniqueness of the equilibrium partitioning coefficients valid when expressed

$$K_{eq} = \lim_{C_i \to 0} \frac{C_i^m}{C_i^{aq}} \tag{2.1}$$

where C_i^m is the concentration of the solute in the non-aqueous medium. When expressed as Equation 2.1, it becomes apparent that the analytical limitation may be a major reason for the limited availability of experimental data on hydrophobic chemicals of low aqueous solubility for which direct measurements are difficult. In general, increasing variability in endpoint data is observed for compounds of increasing hydrophobicity, e.g., as the relative uncertainty in measurements increases at decreasing concentration level (cf. Appendices D and F).

In every case where infinite dilution does not prevail (cf. Equation 4.1 and Appendix B) Equation 2.1 is *a priori* not applicable, i.e. in this case the equilibrium is correctly described through

$$K_{eq} = \frac{a_i^m}{a_i^{aq}} = \frac{\gamma_i^m C_i^m}{\gamma_i^{aq} C_i^{aq}} \tag{2.2}$$

where a_i^{aq} is the activity of component *i* in the aqueous phase and a_i^m the activity of component *i* in a nonpolar phase *m*. In cases where the activity coefficient of component *i* does not equal one in the nonpolar and aqueous phase, respectively, additional parameter/-s, i.e. the activity coefficients, are determining the measured endpoint values (cf. Equation 3.1). The partitioning coefficient will in the case of non-dilute solutions depend on the concentration of the component *i* in the individual experimental system. Equilibrium partitioning is in this case not a unique number, as it depends on, e.g., the degree of association – dissociation properties of the solute.

Data derived from systems, which do not obey the criteria of dilute solutions (Equation 2.1) are critical when used for calibration of QSARs as well as predicted and measured PEC/PNEC ratios. A concentration dependent activity, in its thermodynamic sense, may be a significant factor for explaining the non-linearity in property-logK_{ow} models, at logK_{ow} values above 6 as illustrated in the Chapter 4, section 4.2.

3. Principles and Methodologies in QSARs

The paradigm of QSAR is that variation in structurally and electronic inherent properties of molecular similar compounds, reflects the variation in a given biological or physicochemical activity. The aim of QSAR is therefore to describe the interactions between chemicals and biological / natural ecosystems through a quantification of specific structural characteristics of the molecule in question. The two main objectives for the development of QSARs, at generic EU risk assessment level and a more fundamental scientifically level, may be stated as follows

1. Development of predictive and robust QSARs, with a specified chemical domain, for estimating endpoints for use in risk assessment within areas of environmental chemistry, ecotoxicology and human health.

2. SAR/QSARs as an informative tool concerning the mechanisms causing a given endpoint, i.e. informative by extracting significant patterns in descriptors related to the measured endpoint. This may be obtained through the use of pattern recognition (PARC), which allows for the elucidation of significant similarities-dissimilarities in molecular structural and or electronic properties, as well as correlation patterns between descriptors and their explanatory significance with respect to the intrinsic endpoint value.

3.1 Fundamental aspects

The behaviour of organic compounds, with respect to environmental partitioning and transport processes within and between different phases, as well as unwanted toxicological responses of living organisms, is related to the inherent molecular properties of the compound in question. For approximately 150 years scientists have been investigating this phenomenon quantitatively[111,112]. In general, models for estimating key fate parameters and toxicological responses have been based on three properties governing molecular activity, i.e., hydrophobic, electronic and structural inherent properties of the compounds

$$Endpoint = f(P_{hydrophobicity}, P_{electronic}, P_{structural}, P_x) + e \quad (3.1)$$

The hydrophobicity, $P_{hydrophobicity}$, is related to the individual compound affinity for partitioning between aqueous and abiotic or biotic phases, e.g. a biological membrane. Structural, $P_{structural}$, and electronic, $P_{electronic}$, molecular inherent properties have been related to the ability to, e.g., pass through the membrane, and bind to a receptor or specific sorption site[30,97]. The parameter, $P_{x,}$ accounts for underlying known or unknown effects, which influence the measured endpoint. These may be experimental system parameters and/or environmental parameters, e.g. variation in temperature or concentration of DOM, which affects the measured endpoint.

Background data on measured endpoints may therefore contribute significantly to the explanation in the variation of the endpoint (cf. Chapters 4 and 5). This is why the model residuals or noise, *e*, which is the variance not explained by the model, must be considered. The presence of unknown, e.g. systematic, non-quantifiable variation in data, increases the noise, *e*, and may result in underestimated models.

In QSARs for estimating ecotoxicity, mechanistic model concepts have developed where the hydrophobicity is related to the individual compound affinity for partition to a biological membrane. Steric effects have been related to the ability of the compound to pass through a membrane or to bind to a specific receptor site. These are the basic molecular features for classifying the biological response towards chemicals according to their mode of action[27,30]. The models are successful, especially at in vitro level where the complexity of the route of exposure has been eliminated[113].

In environmental fate studies, hydrophobicity is similarly related to non-specific interactions, i.e. the theory of a partition-like sorption process to soil organic matter driven by the hydrophobic effects[iii] (cf. paper II), which are expected to correlate to the size or logK_{ow} of the molecule. In analogue to the mechanistic model concept described above, the electronic and structural intrinsic molecular descriptors, quantifies site-specific sorbate-sorbent interactions[xii].

The hydrophobicity parameter is often quantified through logK_{ow}, while the steric and electronic properties are quantified by molecular empirical or quantum-chemical descriptors (cf. Section 3.2.3 and Appendix G). Due to the generally increased uncertainty in experimental data for hydrophobic compounds, molecular descriptors such as solvent accessible surface area and volume have, in some cases, been used as alternatives to logK_{ow}.

3.1.1 Bulk properties and endpoints of specific electronic or steric requirements

The basic assumption behind attempts to relate molecular structural and electronic descriptors to i.e. biological activity requires that the compound is present at a concentration level, in which the activity coefficient can be assumed to be equal to 1 (cf. Appendix A). This is analogous to defining the molecule in question as being present at unimeric concentration levels in the modelled phases, e.g. lipid, bulk water, sediment and soil pore water. Furthermore, equilibrium systems are normally assumed. However, many processes require a certain amount of activation energy in order to occur, and may therefore be kinetically controlled and proceed at very low rates at certain system conditions, even though the process is thermodynamically favourable. Such systems may often be described using assumption about steady-state processes under non-equilibrium conditions. The Gibbs Free Energies of the states during a general reaction are shown in Figure 4

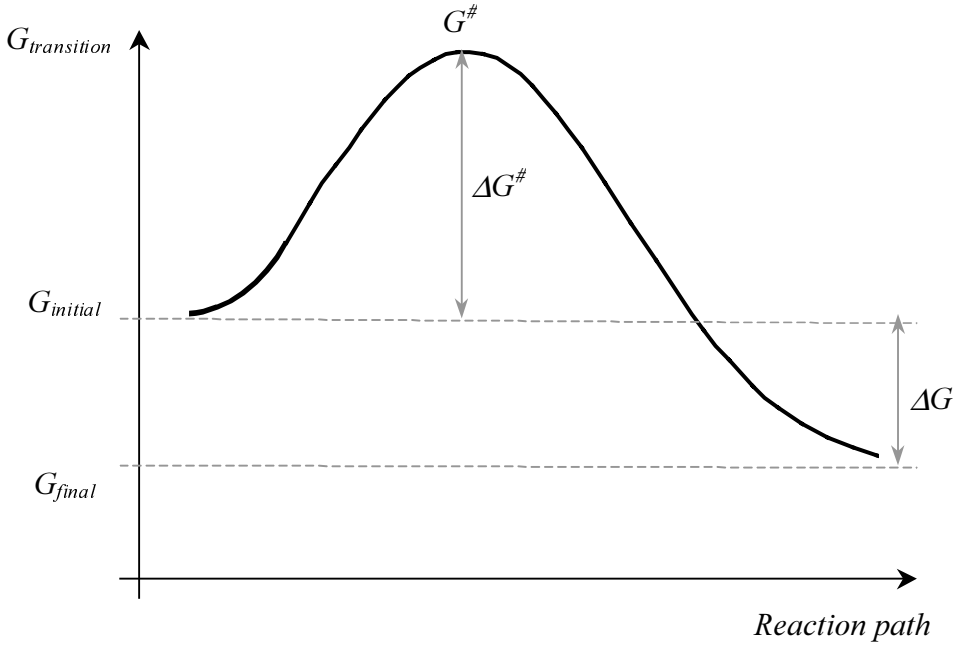


Figure 4 *Diagrams illustrating the Gibbs free energy of a compound, during a binding process to e.g. a membrane. $G_{initial}$ is the Gibbs energy at the initial state, and G_{final}, the minimal Gibbs free energy, i.e. of equilibrium partitioning between a bulk miscible, e.g., aqueous phase and the hydrophobic phase of a living organism. An explanation for the observed decrease in the bioconcentration factor for hydrophobic compounds has been suggested to be due to steric hindrance with respect to penetration of the biological membrane[23,42]. This explanation suggests the presence of an energy barrier, illustrated by the free energy required for attainment of some transition state, $\Delta G^{\#}$, with respect to the penetration of the membrane.*

Processes characterised by thermodynamic control are path-independent (cf. Appendices A and B), and at constant system conditions, variations in endpoints can be expected to be modelled nicely as a function of molecular inherent structural and electronic properties. In kinetically controlled reactions however, the rate-determining step is controlled by the free energy of activation for the transition state. This has to be quantified from the electronic and steric molecular properties, e.g. the Hammett equation[104,122]. However, neither the principles of thermodynamic nor the theories of reaction rates predicts or requires that there should be a linear relation as in the Hammett equation. This is the reason for the exception of chemicals acting by specific modes of action compared to the general paradigm of QSARs, i.e., chemicals of similar structure may act differently. Classification of chemicals, which act by specific modes of action should be understood as a classification into chemically similar groups with respect to, e.g., some sub-structural property which is central for the transition state of kinetically controlled reactions. In a kinetically controlled reaction the chemical domain of the models is restricted to "similar" substructural properties explaining variation in the transition state energy due to small variations in substructurally and/or electronic molecular intrinsic properties. The equilibrium and steady-state considerations are important in relation to the validity and interpretation of experimental data, which are used for developing QSAR models for predictions of measured endpoint values.

3.2 Model development process

Through the development, optimisation and validation of QSARs, it is possible to extract structure-activity relations (SAR) that provide insight into the role and nature of specific organic molecular structures with respect to a given endpoint[II, III, V]. Through the evaluation of the significance of

different sub-structures and functional groups within the molecule with respect to interactions with the specific microenvironment under investigation it is possible to classify chemicals according to their mode of action, i.e. mechanisms of sorption to organic matter or non-polar and polar narcosis. Provided that the assumptions upon which the models are based are true, the developed QSARs are able to provide information concerning the mechanism of action with respect to given endpoint.

In relation to the general exposure and dose-response assessment, as illustrated in Figure 2, the specific disciplines in QSARs based on multivariate data analysis (cf. Fig. 2 step c) are according to Figure 5.

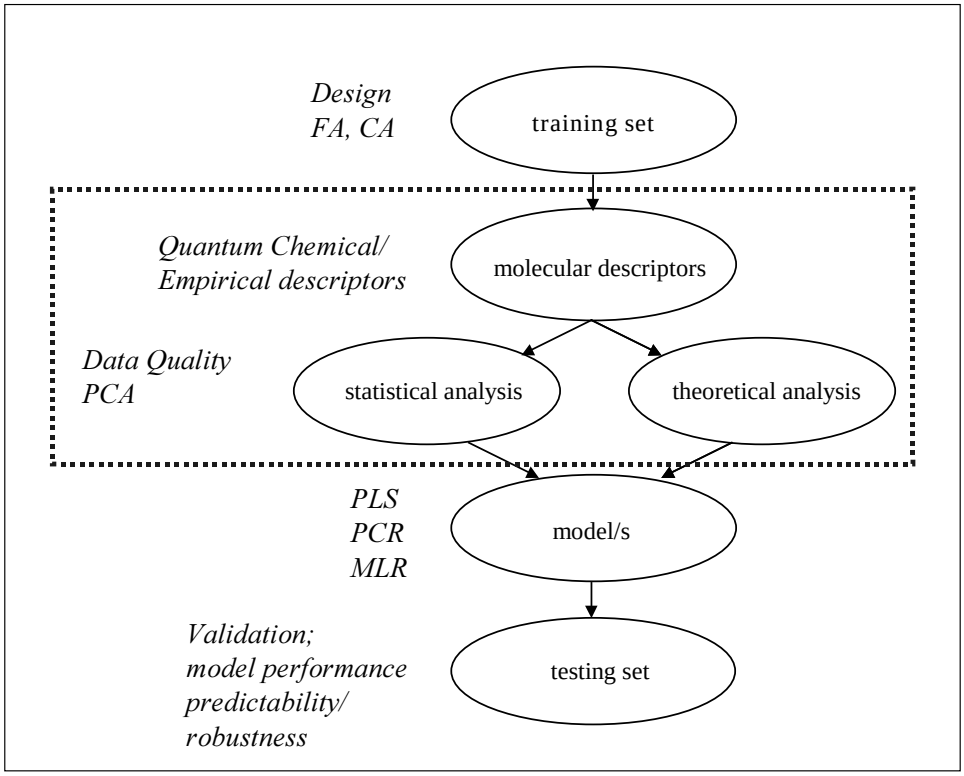


Figure 5 *Model development process for QSARs based on multivariate data analysis, cf. Fig. 2, step b. CA: Cluster analysis, FA: Factor analysis, PCA: Principal Component Analysis, PLS: Partial-Least-Squares regression, PCR: Principal Component Regression, MLR: Multi Linear Regression. The test and training sets are derived from the experimental raw data, cf. Fig. 2 step a. Data Preprocessing, marked by a blue box, comprises analysis of background data and measured endpoint data by analysis of variance and/or pattern recognition, e.g. CA and PCA for the elucidation of the presence of significant additional parameters (c.f. Equation 3.1). The selection of a training set for calibrating the model is based on PCA and may include the specific endpoint to be modelled*[xvii]*. The training set is to be selected to cover the chemical domain of the model, which is defined, e.g., as class-specific (aromatic esters, i.e. phthalates), sub-general (e.g. substituted phenols) or general (hydrophobic/hydrophilic). Through model validation on an independent testing set, or by cross-validation, the robustness and predictability of the model is evaluated.*

Through quantification and homogeneous spanning of the chemical property space[III,xvii] (cf. Appendix G) a test set of chemical compounds may be selected. The test set used for calibrating the model is selected to span chemical property space through experimental design[xvii]. For thermodynamically controlled reactions, and chemicals acting by non-specific mode of action, the test set may be selected based on molecular descriptors alone (cf. Appendix G). However, in the case kinetically controlled reactions, or chemicals acting by specific mode of action, the endpoint data need to be included in the experimental design for selecting the test set[xvii].

Prior to model calibration, data preprocessing of exposure or fixed toxicity endpoints, by variance analysis and pattern recognition need to be performed as described in section 3.2.1. Validation of the model through the training set is the last step of the model developing process as shown in Figure 5.

3.2.1 Data preprocessing

Preprocessing of data is an often-overlooked step, or even non-addressed steps in the process of developing QSAR models. However, before developing QSARs it is importance to know the type and level of uncertainty on measured endpoints and this is obtained through preprocessing, which may consist of several steps depending on the nature, amount and quality of data. Preprocessing may as such comprise analysis of variance homogeneity in data [III,V], as well as hypothesis testing of endpoint data (cf. Equation 3.2 to 3.4).

Unfortunately, homogeneity of variances in endpoint data between compounds almost never exists. However, the variances on the substance specific endpoint values have to be acceptable, i.e. low and homogeneous[III]. Dependent on the requirements and purpose of the model, data can be pre-processed at several levels.

When using data from databases or from the literature, a minimum requirement to data is significance testing with respect to variances on compound specific endpoint values[III]. If the variance on substance specific endpoint values is not significantly lower than the variance between substance specific endpoint values, then background data should form the basis for analysis of variance inhomogeneities. The influence of background data, i.e. experimental or environmental parameters (cf. Equation 3.1), on the variance on endpoint values is elucidated through simple variance analysis or pattern recognition, e.g. by including background data in an initial PCA/PLS analysis (c.f. Appendix J) [e.g. 114,115]. The elimination of noise due to variations in experimental parameters as well as biases between methods can only be eliminated if the variance criterias are fulfilled. Hypothesis testing can be carried out through F-testing of individual hypotheses, e.g.

$$\sigma^2_{i,endpoint} < \sigma^2_{i,endpoint/bewteen-i} \tag{3.2}$$

The variance on all substance specific endpoint values, $\sigma^2_{i,endpoint}$, should be significant lower than the variance between substance specific endpoint values, $\sigma^2_{i,endpoint/between-i}$.

$$\sigma^2_{i,endpoint/method} < \sigma^2_{i,endpoint} \tag{3.3}$$

Biases in results obtained from different methods are evaluated through a comparison of the variances on each method with the variance on the average endpoint. Substance specific data obtained from methods of large uncertainty ranges are eliminated from the data set, e.g. the shake-flask method (cf. Appendix D)[38].

$$\sigma^2_{i,endpoint} < \sigma^2_{i,endpoint/system-parameters} \tag{3.4}$$

Finally the presence of significant system parameters should be evaluated, e.g., such as matrix effects due to the presence of DOM, pH and temperature and interspecies differences.

Variance inhomogeneity in endpoint data between compounds seldom occurs. Not in data found in the literature (from different laboratories and different analytical methods) and nor from laboratory specific data (within laboratory, within and/or between methods). Variance analysis and/or pattern recognition are important tools for testing the significance of additional parameter (cf. Equation 3.1) with respect to the nature and quality of endpoint data.

As the PLS algorithm is based on extraction of systematic and significant variance patterns between the endpoint and descriptors including the standard deviation on endpoint values as explanatory variables, should therefore only increase the noise but optimal show no explanatory significance. Provided that the endpoint data used for calibrating the model is precise and accurate, descriptors with no systematic variation related to the endpoint will have no explanatory significance, i.e. identified as having insignificant loading weights[116].

The influence of standard deviations on calibration data can be evaluated by, e.g., including these in the initial data analysis to secure the significance of the explanatory latent variables, i.e. as a control variable with respect to the causality of the model. If the standard deviations on calibration data are perfectly homogeneous, which is never the case, then the loading weight of the control variable is zero. If however there is a small systematic increase in the standard deviation by increasing hydrophobicity of the objects, i.e. compounds, then the control variable may be used to secure the significance of the principal components included in the model (cf. Appendix J).

3.2.2 SAR and QSAR based on PCA, PLS and MLR or simple LRs

QSAR based on multiple linear regression (MLR) requires normally distributed, independent and 100 % relevant descriptors. This means that every descriptor, i.e. x-variable, applied in the QSAR model is assumed 100% relevant for the explanation of the "cause" of the measured endpoint. Such a situation is difficult to find. Therefore, methods such as PLS where the variation in x-variables which have predictive power for the endpoint variable y is extracted, into so-called latent variables, seems more appropriate to use.

The multivariate methods like partial-least-squares regression (PLS) and principal component analysis (PCA) are based on ortogonalisation of a large number of descriptors by linear combination and projection of inter-correlated X-descriptors onto hyper-planes spanned by few latent variables. The major advantage of using PLS and PCA techniques in model development is that such approaches are able to handle a large number of X-variables, which are correlated in different ways and to different degree[e.g.115,116,117].

PLS is a regression technique for regressing Y onto X, where X is an $n*k$ matrix (n is the number of chemical compounds included in the model and k is the number of descriptors). In QSAR we only deal with one Y-variable, and therefore Y is an $n*1$ matrix. The method is especially suitable when the descriptors of X are intercorrelated.

The PCA approach is an attempt to explain the structure of the variation in the X-data by projection of the original variables in the m-dimensional space onto a lower dimensional hyperspace spanned by few numbers of latent variables called principal components. In QSAR the X-matrix is generally mean centered and scaled before PCA is applied. The methodology of PCA is to decompose the X-data matrix into the following bilinear form:

$$\mathbf{X} = \sum_{a=1}^{A} \left(\mathbf{t}_a * \mathbf{p}_a^T \right) + \mathbf{E}_A \qquad (3.5)$$

where $\mathbf{t_a}$´s comprise the score values of the n compounds in the hyper-planes spanned by a significant number principal components $\mathbf{p_a}$´s. The correct number of principal components can be determined from a number of stopping criteria, with cross-validation being one of the most often used techniques. The Y matrix can be similarly decomposed into

$$\mathbf{Y} = \sum_{a=1}^{A} \left(\mathbf{u}_a * \mathbf{q}_a^T \right) + \mathbf{F}_A \tag{3.6}$$

Performing a regression of **u** onto **t** corresponds to Principal Component Regression. PLS follows a similar procedure except that the decomposition of X and Y are performed simultaneously in an iterative manner with the purpose of extracting most relevant variation in X with respect to Y. The latent principal component, i.e. vectors (**t** and **u**), depend both on the **X** and the **Y** spaces and are related through an inner relationship $\mathbf{u}_a = \mathbf{b}_a \mathbf{t}_a + \mathbf{e}_a$, where $\mathbf{b_a}$ is the least squares regression coefficient and $\mathbf{e_a}$ is the residual[e.g.115,116,117].

3.2.3 Steps in development of multivariate QSARs

The concept of the development of QSAR models based on multivariate data analysis can be described through four steps as stated below[116,118].

1. *Quantification of the chemical property space through quantitative descriptions of the chemical structure*

   When developing QSARs the starting point is the chemical structures. In QSAR modelling an optimum span of the chemical property space is obtained by including structures representing variations in all structural subparts of the molecules. If groupings are present within the X-space, separate models are developed for each group. However, as described in paper III, in some cases groupings may be eliminated by identification and excluding class-specific variables. The basic philosophy of QSARs is that an observed variation of a given biological or physico-chemical property is caused by an analogue variation in the latent properties of the chemical model system. The chemical model system comprises a quantitative description of the molecular structural and electronic properties. Two dominating trends among QSAR experts have appeared, one developing QSARs based on the simple two-dimensional empirical (EM) descriptors and the second building QSAR models based on the three-dimensional non-empirical and quantum chemical (NEM-QC) descriptors.

   Empirical Descriptors *references as cited in III*
   The empirical descriptors used comprise simple molecular connectivity indices (MCIs), ${}^{p}\chi_0$-${}^{p}\chi_6$, ${}^{c}\chi_3$, ${}^{pc}\chi_4$, quantifying variations in structural characteristics. Furthermore valence MCIs, ${}^{p}\chi^{v}_0$-${}^{p}\chi^{v}_{10}$, ${}^{c}\chi^{v}_3$, ${}^{pc}\chi^{v}_5$ and the sum of electrotopological state indices quantifying information on structural and electronic properties of the molecules. Lastly the valence and simple MCI difference descriptors, $d\chi_n = {}^{p}\chi_n - {}^{p}\chi^{v}_n$ for n=0 to 6, quantifying for strictly electronic properties. The empirical descriptors are calculated using the programme Molconn-Z[119].

   Quantum-Chemical and Non-Empirical Descriptors *references as cited in III*
   The quantum chemical descriptors used comprise energies of frontier orbitals, E_{HOMO} and E_{LUMO}, and the second lowest and second highest MO energies, E_{NHOMO} and E_{NLUMO}. Furthermore the hardness of the molecules, denoted hardness ($[-E_{HOMO}+E_{LUMO}]/2$), the electronegativity, denoted E_{neg} ($[-E_{HOMO}-E_{LUMO}]/2$), heat of formation, dH_f, and the Debye dipole moment, denoted dipolm. In addition the polarisability, solvent accessible area and volume, A_{sas} and V_{sav} and van der Waals

molecular surface area and volume, A_{vdW} and V_{vdW}, respectively, are calculated from the Chemplus software included in the HyperChem software[120].

The EM descriptors are very simple and quickly calculated[III,v], while the latter are obtained from more complicated quantum chemical calculations based on semi-empirical AM1 geometry-optimisations, i.e. energy-minimisations, of the three-dimensional molecular conformation, using the unrestricted Hartree-Fock method[120]. An evaluation of the model performance parameters of QSARs based on the EM and the NEM-QC descriptors have been assessed with respect to the complexity and quality of the endpoint. The assessment was based on a case study on PCBs and no significant difference in model performance parameters was observed[III].

In general, the more descriptors included in QSAR models based on multivariate data analysis, the better is the structural characterisation of the chemical property space, and the robustness of the model increases.

2. *Multivariate design for selecting series of test compounds[116]*

When developing empirical models, experimental and statistical design is crucial for the chemical domain, and predictive power of the resulting model. An ideal chemical property space would include descriptors quantifying all possible variations in electronic and structural factors of molecules to span the whole domain in which any prediction will be performed. If compounds are selected according to a statistical design where all structural factors are varied simultaneously, we have a better chance of extracting relevant information concerning the mechanism causing a given endpoint.

3. *Measuring and/or evaluating ecotoxicological or physico-chemical activities raw data*

Patterns in biological or physico-chemical activities are best analysed by including several different measurements in the respective test systems. Different types of measurement will increase the amount and types of information concerning the e.g. biological activity pattern. In spite of the ability to handle uncertainty in data, handling of data of different quality and uncertainty is required when using endpoint data found in the literature for calibrating the models. Due to the fact that results of different experimental methods and laboratories may vary significantly, especially the measurements of the activities of very hydrophobic chemicals may be followed by high uncertainties[38,iii]. Raw data obtained from the literature should therefore always be subject to analysis of homogeneity in variance and the uncertainty; noise in data should be quantified to assure that the increase in explained variance is significantly greater than the noise in data as described in section 3.2.1.

4. *Step IV Mathematical modelling of the relation between the structure and activity of chemicals*

The success of QSARs based on multivariate projection (PCA and PLS) methods is due to the ability to include many inter-correlated as well as non-correlated variables, the non-sensitivity to imprecision and non-relevant descriptors. These aspects allow for the investigation of unknown relations, and pattern recognitions as the projection methodology are based on extraction of systematic variations in data. Model developments are extended through calibration and cross-validation and if possible model validation with an external validation set, i.e. training set (cf. Figure 5). As such the model development process consists of two steps 1) model derivation and interpretation (upper five modules of Figure 5) and 2) model validation and use (lowest module of Figure 5).

4. Aqueous Solubility

According to the OECD guideline no. 105 for the testing of chemicals, the aqueous solubility of a substance is the saturation mass concentration of the substance in water at a given temperature, expressed in mass of solute per volume of solution[155]. The guideline suggests two methods, i.e., the shake flask-method for substances with solubilities above 1 mg/L, and column elution methods for solutes with an aqueous solubility below 1 mg/L. These methods are not suitable for measuring unimeric solubilities for hydrophobic compounds, for which microcolloids are formed by exceeding the saturation point in the bulk aqueous phase. Microemulsions are colloidal particles, with a dimensional range of 10^{-9} to 10^{-6} m, and in the lower particle size range, i.e. micro-colloidal nm size range the particles are to small to be observed by light scattering[110].

The main reason, for the overestimated solubility as well as various partitioning coefficients for the phthalates is due to the similar densities of phthalates and water[I,II]. The problems encountered when determining equilibrium partition coefficients and solubilities of phthalates have been noted by several authors[II,iii]. As a result several methods have been developed with the purpose of avoiding the formation of a microemulsions in the aqueous phase[35,37].

4.1 Homogeneous true solutions or heterogeneous mixtures

A system containing two or more compounds is called a mixture. When one of the components, called the solvent, is present in large amounts compared with the other compounds, called the solute, then the mixture is called a solution. A solution is a single phase consisting of a mixture on a molecular level of a solvent and solute(s). When a mixture or a solution system consists of more than one phase, the system is heterogeneous.

4.1.1 Dissolution process of Phthalates in aqueous media

Due to the limited aqueous solubility of phthalates, the ideal dilute solution description (cf. Appendix A) is convenient to use for the dissolution of phthalates in the aqueous bulk phase[I, II]. At equilibrium, $[PAE]_{liq} \rightleftarrows [PAE]_{aq}$, the standard Gibbs Free Energy for the dissolution of phthalates is described:

$$\Delta G^{\circ} = \mu_p^{*} - \mu_p^{\circ} = -RT \cdot \ln K_{eq} = -RT \cdot \ln \frac{a_p^{aq}}{a_p^{liq}} \tag{4.1}$$

where μ_p^{*} is the standard chemical potential of pure phthalate, and for the solute the standard state, μ_p°, at infinite dilute solution. The liquid phthalate phase (∗), cf. Figure 6, is practically immiscible with water and the activity, a_p^{liq}, equals unity (cf. Appendix B). It is assumed that the activity coefficient, γ_p^{aq} of the solutes in the aqueous bulk is unity within the unimeric concentration range of the individual phthalates (cf. Appendix B). Equation 4.1 then reduces to

$$\Delta G^{o} = -RT \ln C_p^{liq} \tag{4.2}$$

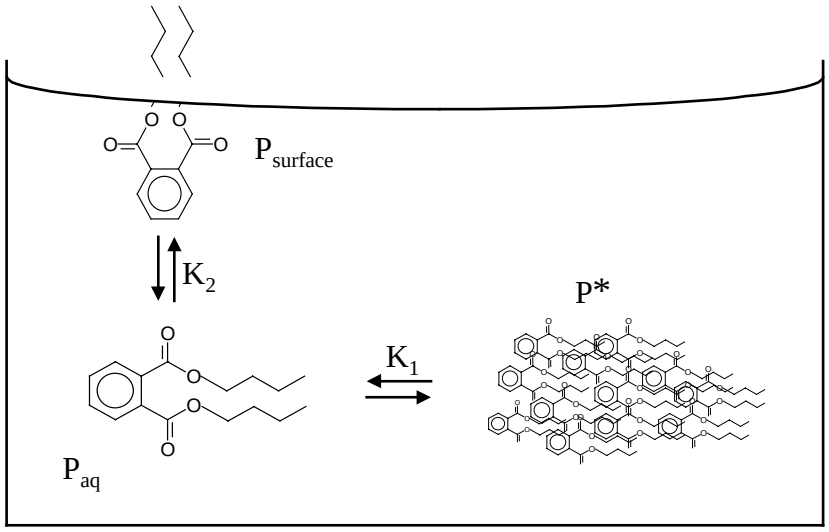


Figure 6 *Two-phase system consisting of phthalate microemulsion phase, P^*, dispersed in the continuous phase, i.e. aqueous phase. The partitioning of phthalate molecules between the pure phthalate phase, P^*, and dissolved in the aqueous phase, $P_{(aq)}$, are described by K_1. A second partition process of unimeric phthalate molecules between the bulk water and the air-water interface is described by K_2.*

Figure 6 illustrates an example of a heterogeneous system consisting of a thoroughly mixed pure liquid phthalate and water, i.e. as in the preliminary OECD-solubility test (cf. Appendix D and E), at a ratio that causes of two-phase system to form. In this system, the process of spontaneous mixing, through the partition coefficients, K_1 and K_2, continues until the Gibbs free energy, at constant temperature and pressure reaches its minimal value.

At t=0, the concentration of phthalates in bulk water will be insignificant, i.e. $C_{i,aq} \approx 0$. The chemical potential of phthalate molecules in the pure liquid state is high, and phthalate molecules, n_i, will diffuse from the pure phthalate phase and out in the bulk water. Partitioning to the surface will occur until the chemical potential in both phases and at the interface are equal, i.e. $d\mu_i=0$ and K_1 and K_2 are at equilibrium. Changes in the chemical potentials of phthalate in the individual phases may be described through the equation[110]

$$\sum_i n_i^{aq} d\mu_i^{aq} + \sum_i n_i^{liq} d\mu_i^{liq} + \sum_i n_i^{s} d\mu_i^{s} + A_s d\gamma = 0 \tag{4.3}$$

where $d\gamma$ is the change in the surface tension of the air-water interface of area, A_s, caused by the presence of component *i*. By equating the expression against Gibbs-Duhem equation (cf. Appendix B) the bulk phase terms are eliminated and Equation 4.3 reduced to the so-called Gibbs adsorption equation

$$\begin{aligned} &\sum_i n_i^{s} d\mu_i^{s} + A_s d\gamma = 0 \Leftrightarrow \\ &d\gamma = \sum_i \frac{n_i^{s}}{A_s} d\mu_i^{s} \end{aligned} \tag{4.4}$$

This equation relates the change in surface tension, γ, to the surface excess concentration of component *i* through the number of moles, n_i^s, per surface area, A_s, and the change chemical potential of at the interface. In the systems illustrated in Figure 6, two components are present at the surface, i.e. water and phthalate molecules, the composition determined by the bulk water concentration of phthalates through the partition coefficient, K_2. By proper definition of the reference point in the surface[110], the change in surface tension, γ, is directly measurable and quantified through

$$d\gamma = \Gamma_p^s d\mu_p \tag{4.5}$$

The change in chemical potential of phthalates dissolved in the bulk water phase is described through

$$d\mu_i^{aq} = RTd\ln a_i^{aq} = RTd\ln x_i^{aq} + RTd\ln\gamma_i^{aq} \tag{4.6}$$

The activity coefficient, γ_i^{aq}, in a dilute solution of phthalate in the bulk water, is unity (cf. Appendix A), and the last term on the right side of Equation 4.6 equals zero. For equilibrium partitioning transfer of solute molecules from the dilute solution to the air-water interface, $d\mu_p^s = d\mu_p^{aq}$, and

$$\Gamma_p^s = -\frac{d\gamma}{RTd\ln C_p^{aq}} \tag{4.7}$$

For adsorption from dilute solutions, the activity coefficient equals unity, and by combining Equation 4.5 and 4.7, replacing the activity in bulk water by the molar concentration, the change in surface tension is described through a change in bulk water concentration

$$\Gamma_p^s = -\frac{1}{\ln 10 \cdot RT}\frac{\partial\gamma_p^s}{\partial\log C_p^{aq}} \tag{4.8}$$

The measured surface tension as a function of bulk water concentration is illustrated for DEP and DnBP in Figure 7. The surface tension exhibits two distinct linear concentration regions. The high slope region is caused by free energy transfer to the surface through the partitioning process, K_2 in Figure 6, and continues until no further partitioning to the surface occurs. At the intercept between the steepest first part and the flat second part of the curve, the concentration of unimeric phthalates in the bulk phase has reached its saturation point, and a two-phase system takes form (cf. K_1 in Figure 6). By this method it is possible to discriminate the borderline between a true homogeneous solution and a heterogeneous mixture.

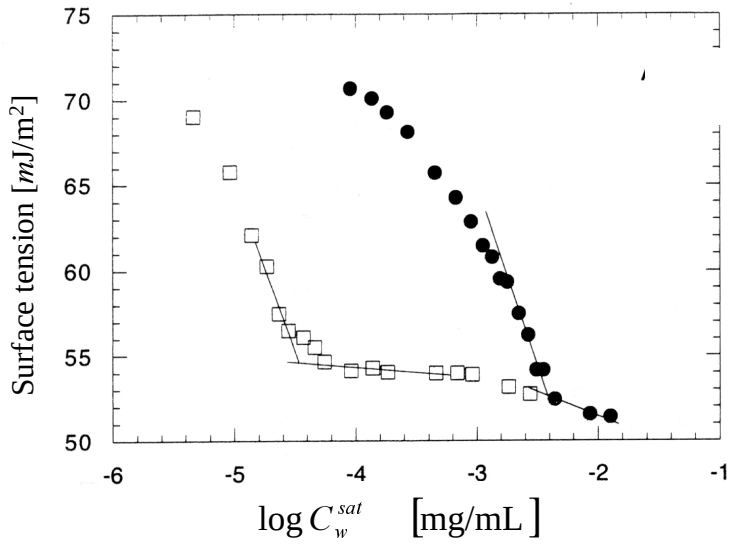


Figure 7 *Surface tension of phthalates, DEP (•) and DnBP () , as function of bulk concentration at 25 °C. The unimeric solubilities are determined as the concentration where there is a significant change of slope,* $\partial\gamma/\partial\log C$.

The surface tension method is expected to work well for all hydrophobic compounds[110]. The reason being that the solubility of these compounds are determined by the hydrophobic effect, i.e. they are expelled from the bulk water phase and thereby present at interfaces.

The solubility of phthalates was furthermore investigated by isothermal micro-calorimetry, which is a method for measuring changes in enthalpy[II] upon changes in mixture composition. However the solubility of the phthalates are too low to be measured by this method as the lowest possible injection volume exceeds the unimeric solubility.

The attractive forces between hydrophobic compounds are weak van der Waals forces, while strong cohesive forces exist between water molecules throughout the aqueous bulk phase. This difference in nature of nonpolar solute-solute, solute-solvent and interaction of the water molecules are the driving force for the hydrophobic effect. Ionic or highly polar substances tend to be easily soluble in water since solute and solvent interact to compensate for the disruption of the strong intermolecular forces between water molecules, but for hydrophobic molecules only weak to weak van der Waals solvent-solute attractions exist upon dissolution of non-polar solutes. The dissolution occurs through the formation of a solvation shell of ordered water molecules [104,153]. The hydrophobic effect can be expressed as the tendency for water to exclude hydrophobic molecules from the water phase, thereby minimising the disruptive effects on the hydrogen bond network between water molecules[107,153]. The phthalates and the PCBs are among a wide range of chemical classes of hydrophobic compounds for which only weak van der Waals solute-solvent and solute-solute interactions prevail in the aqueous continuous phase. Both classes of compounds are liquids at room temperature, and the increased variation in measurements upon increasing hydrophobicity (cf. Appendix F) is properly due to the formation of clusters within the bulk water phase, i.e. thereby minimising the solvent accessible volume. The formation of clusters and the presence of solute-solute interactions are the explanation for overestimated solubilities and underestimated partitioning coefficients, as illustrated in section 4.2.

4.2 Simple linear relationship for estimating the bioconcentration of phthalates

As described in Chapter 2, section 2.2, the equilibrium partitioning of organic pollutants to a variety of abiotic as well as biotic phases has been modelled through simple linear regressions using the octanol-water partitioning coefficient the as explanatory variable. A number of investigations concerning the non-linearity of the log*BCF*-logK_{ow} relationship, when going from non-polar to highly hydrophobic, i.e. lipophilic, compounds exist. Several explanations focussing on limiting solubilities, kinetics, as well as limiting bioconcentration due to molecular size restrictions has been suggested as possible explanations for the log*BCF* increasing less than proportional to logK_{ow} for logK_{ow} values above 6 [94,95,105].

Due to the generally increased uncertainty in experimental data for hydrophobic compounds, molecular descriptors such as molecular weight (Mw), solvent accessible surface area or volume has, in some cases, been used as alternatives to logK_{ow}. The bioconcentration factor, log*BCF*, as a function of molecular weight of the solute, using classical basic statistics for calculating the mean and standard deviation on the measured BCF values, is illustrated in Figure 9.

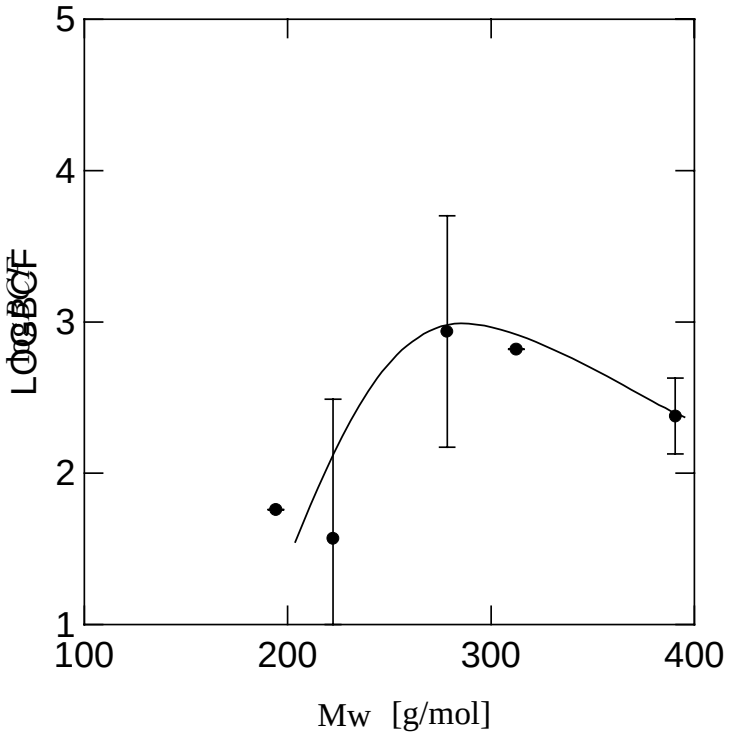


Figure 9 *Bioconcentration factors, logBCF, as a function of molecular weight of the phthalates, indicating that size is the limiting factor with respect to bioconcentration in aquatic organisms[42]. (cf. Chapter 3, section 3.1.1).*

Immediately, Figure 9 indicates that molecular size is a limiting factor with respect to the bioconcentration of high molecular weight compounds. The model is an example of a wrong answer to a questing based on data, where the data do not obey the assumption upon which the model was developed. However, by expanding the data background, an important additional parameter (cf. Equation 3.1) is elucidated, which explains the first observed parabolic relation between molecular weight and log*BCF*. This is illustrated through Figure 10.

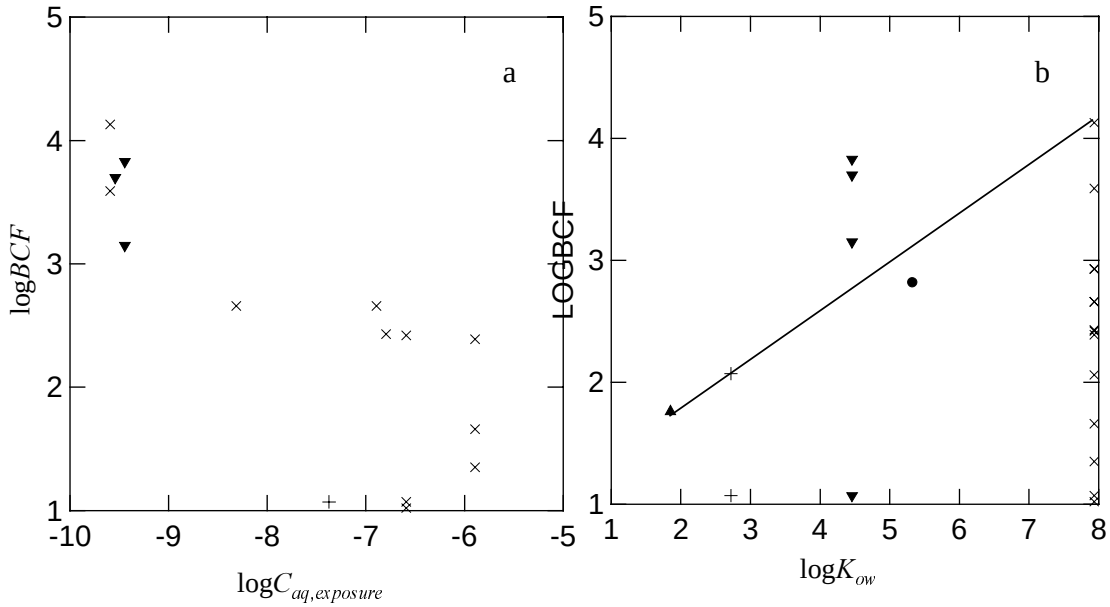


Figure 10 *Bioconcentration factors for DMP (▲), DEP (+), DnBP (▼), BBzP (●), and DEHP (×) in fish as function of a) aqueous exposure concentration, and b) UNIFAC calculated octanol-water partitioning coefficients (cf. Appendix C). a) The aqueous exposure concentrations in the experimental setup for measuring the bioconcentration factor of DEHP is above the aqueous solubility of –7.4 in 8 out of 11 investigations found in the literature.*

The aqueous exposure concentration is a significant factor with respect to the observed variability in log*BCF* values. As observed in Figure 10a, the bioconcentration factor decreases with increasing exposure concentration of DEHP (×). The aqueous solubility of DEHP is –7.4 given in log units[I, II], and thereby the exposure concentration exceeds the unimeric solubility in the majority of the measured BCF values. For the phthalates the limited solubility as well as the formation of micro-emulsion by exceeding the unimeric solubility is the main reason for the observed non-linearity in the log*BCF*-logK_{ow} relationship. The first assumed parabolic relationship between log*BCF* and logK_{ow} shows the importance of concentration levels of the test substance with respect to the dilute solution description, which is required for equilibrium partitioning estimations according to Equation 2.2. Limitations in bioconcentration factors are in this case not due to a molecular cut-off value as proposed in the literature[105], but due to a lowered activity in the aqueous bulk phase and the formation of a third phase of micro-emulsions in the bulk phase (cf. Figure 3 and 6).

Large bioconcentration factor is not to be considered as a significant factor with respect to hazard assessment (cf. Figure 1), as the actual limiting factor is the solubility, e.g. 0.017 ppm for DEHP. A hazard that should be considered in future risk assessment is the presence of organic compounds forming colloidal dispersion or emulsions at concentration levels above the unimeric solubility. This is due to the fact that the property of self-assembling units can not be quantified through molecular structure-activity relationships (cf. Chapter 3 and Appendix E). The migration behaviour is changed[VI], and adverse effects may be difficult to quantify and control. In this respect, it should be mentioned that a major fraction of the toxicity measurements for the phthalates have been carried out at concentration levels above the unimeric solubility[20,21,22], which makes the process of risk assessment difficult due to inconsistency in data. Even though the effects of the high molecular weight phthalates have been assessed to pose no hazard, long term effects as well as the adjuvant effects of the phthalates needs yet to be elucidated.

5. Effect of Environmental Parameters

Environmental abiotic and biotic phases are in most cases far from the homogeneous well-mixed phase approximation. Environmental properties of importance are the temperature of the system[II], soil characteristics, ionic strength, and third phase effects such as the presence of dissolved organic matter (DOM) in the terrestrial and aquatic environments[IV,V]. Therefore, to derive endpoint parameters of importance for the real compartment systems it is necessary to

1) *identify environmental parameters which determine the activity of pollutants with respect to their respective phases*

2) *to classify the environmental compartments according to these parameters*

The significance of environmental compartment parameters with respect to the fate of environmental pollutants is weakly elucidated. Matrix characteristics such as the presence of DOM in the aqueous bulk phase, i.e. the third phase effect, has shown to affect the fate and effects of organic compounds in the environment[as cited in V]. Thus QSARs that produce endpoint estimates that reflect the influence of environmental properties, i.e. matrix specific properties, as well as substance specific properties with respect to a given activity will increase the amount of obtainable information useful for risk assessment of environmental pollutants.

5.1 Sorption to natural organic matter (NOM)

A main topic within the area of environmental chemistry is the structural characteristics of NOM in water and soil. There are numerous reports on investigations of the kinetics and thermodynamics of sorption mechanisms for different classes of xenobiotics to NOM which deal with experimental as well as modelling approaches for elucidating the mechanisms by which xenobiotics interact with NOM. Thus, significant efforts by several investigators have resulted in knowledge of the possible mechanisms of intermolecular and intramolecular interactions within the NOM structure as well as interactions between xenobiotics and NOM[e.g.V,VI,126,127,128,129,121]. NOM consists of different types of sorption sites. At low concentration of solute in the aqueous phase, the solute will show highest affinity towards the more energetically favourable adsorption sites. By increasing the concentration of solute in the aqueous phase, the energy barrier for sorption may change, e.g. increase. Such a relationship between the concentration of pollutant sorbed to NOM as a function of the concentration of pollutant in the aqueous phase has been described through the Freundlich equation for experimentally determined equilibrium isotherms[iii,104]

$$C_i^{NOM} = K \cdot \left(C_i^{aq}\right)^n \tag{5.1}$$

where C_i^{NOM} is the concentration of the compound *i* sorbed to NOM, C_i^{aq} is the concentration of pollutant dissolved in the aqueous bulk phase, K is the Freundlich constant and *n* a factor defining the shape of C_i^{NOM} a function of C_i^{aq}.

In Equation 5.1, n<1 describes a situation where the energy barrier for sorption increases by increasing solute concentration. This is the typically observed shape of the isotherm found in the

literature, i.e. a non-linear function of C_i^{NOM} versus C_i^{aq}, characterised by linear sorption at low solute concentration and decreasing sorption asymptotic approaching a constant value by increasing C_i^{aq}.

Isotherms for which n>1 is explained by initial adsorption leading to a modification of the surface, which results in a lowering of the energy barrier for sorption. This is typically observed for detergents[104,xiii].

In its natural logaritmic form, Equation 5.1, may be expressed as

$$\ln C_i^{NOM} = \ln K + n \ln C_i^{aq} \tag{5.2}$$

For the case of n=1, Equation 5.1 is in accordance with the dilute solution description of thermodynamics (cf. Appendix A, Equation A12 and A13), and in this case the activity is correctly expressed in molar concentration. However, the Freundlich equation violates the rules of thermodynamics in cases where n≠1. By rearranging the equilibrium partitioning coefficient equation (cf. Equation 2.2), an expression most similar to the Freundlich equation in 5.2 is

$$\ln C_i^{NOM} = \ln K + \ln \frac{\gamma_i^{aq}}{\gamma_i^{NOM}} + \ln C_i^{aq} \tag{5.3}$$

The non-linearity is hereby not described by the empirical Freundlich constant *n*. Instead the non-linearity is a function of the activity coefficients of the partitioning pollutant, *i*, in the respective phases. Equation 5.2 is equal to Equation 5.3 for systems, which can be described through the solution description, i.e. n = 1, and only in cases where the activity coefficient equals unity in both phases (cf. Appendix B and Equation 2.1).

The above description is yet another illustration of unfavourable system descriptions in relation to the paradigm of QSARs, as it causes the inconsistencies in the data to increase due to the presence of non-quantifiable additional parameters (cf. Chapter 3, Equation 3.1, Chapter 4.2 and Appendix D).

5.2 QSARs for prediction of sorption of PAHs and hetero-substituted PAHs to organic matter

Based on a study in the literature, logK_{oc} values for 65 PAHs, i.e. 7 S-hetero analogue, 10 O-hetero analogues, 50 N-hetero analogues and 52 unsubstituted PAHs, were collected. In addition a number of alkyl- and nitro- and hydroxy- substituted PAH was included (cf. Appendix G). The range in experimental system parameters was as follows: ionic strength from 0.001 to 0.1 M, pH from 4 to 11, equilibrium time from 1 to 11 days, and matrix type varying from clay, silt, sand, fulvic acid, humic acid, soil and sediment. This kind of data collection is not unusual. Analysis of variance showed that the variances in compound specific organic carbon normalised partitioning coefficients, neglecting the influence of background data, were significant higher than variances in partitioning coefficients classified according to e.g. matrix type[xvii]. Due to missing data only models classified into the matrix types sediment and humic acid, respectively, could be investigated.

5.2.1 Simple linear logK_{oc}- logK_{ow} relationships classified according to matrix type

Data preprocessing of raw data was performed through F and t- tests (cf. Appendix H, Table H1-H3)[xvii]. The simple logK_{oc}-logK_{ow} models for the matrices humic acids (HA) and sediments are shown in Figure 11.

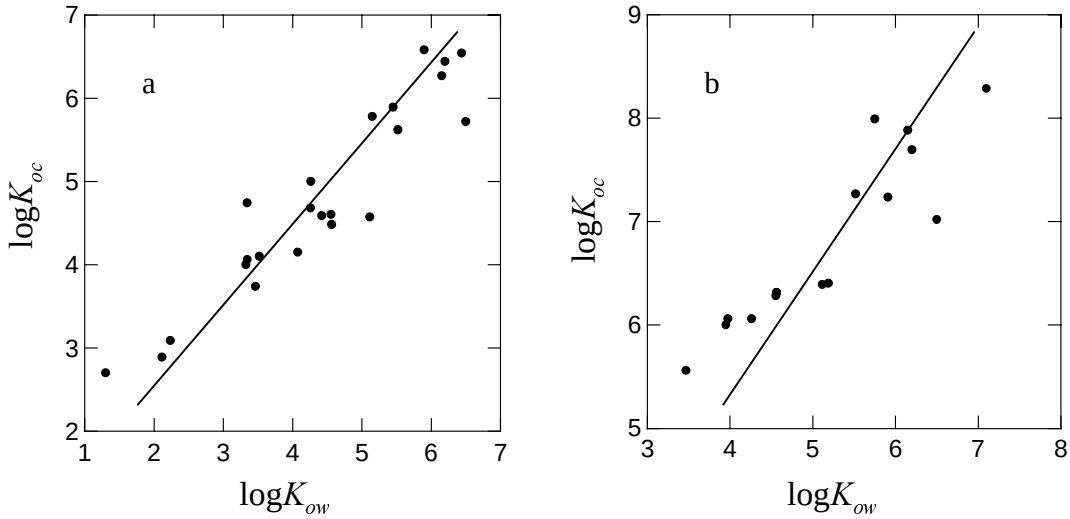


Figure 11 *LogK$_{oc}$ versus logK$_{ow}$ for different matrix types a) HA of different origin and b) organic carbon normalised partitioning coefficients for sediments.*

The slope of both regression models is approximately 0.74, while the intercept differs by approximately 1.5 log units. The model performance parameters of the simple linear regressions for the two matrix types, and by ignoring environmental parameters, are given in Table 5.1.

Table 5.1 *Model performance parameters for matrix specific simple QSARs for estimating the sorption of PAHs to organic matter.*

matrix type	*n*	*a*	*b*	R^2	R^2_{adj}	*SEE*	*CVS*	*CVI*	*F-ratio*
humic acid	23	0.758	1.458	0.89	0.89	0.29	0.21	0.57	170
Sediment	16	0.731	3.021	0.84	0.82	0.35	0.18	0.98	71
non-specified	49	0.826	1.480	0.585	0.576	0.884	0.21	1.01	66

n: number of samples, a: regression coefficients, b: intercept, R^2: squared correlation coefficient, R^2_{adj}: the adjusted squared correlation coefficient, SEE: standard error of estimate, CVS: confidence interval of the slope (P=0.005), CVI: confidence interval of the intercept (P=0.005), F-ratio: hypothesis testing of zero slope at p value (P)<0.0005. The last row shows the results of the linear regression model by neglecting classification according to matrix type.

As shown in Table 5.1, model performance parameters are increased significantly by classifying data according to the environmental parameters, i.e. in this case matrix type. Considering the complexity of natural organic carbon (NOM)[106, as cited in V], the success of this rather simplistic approach is most probably due to the hydrophobic effect[e.g.104,153] driving the equilibrium partitioning of hydrophobic compounds between the organic matter and aqueous bulk phase. Variations in the inherent properties between sediments, as well as between humic acids may contribute to the standard errors of estimates.

By using K_{ow} as the predictor variable for the partitioning to NOM, only strictly hydrophobic interactions are assumed to be driving the sorption process. The chemical domain of the linear relations are restricted to hydrophobic compounds, and relies on the assumption dilute solutions

where the activity can be quantified by the molar concentrations in the respective phases (cf. Appendixes A and B).

In analogy to the explanation used in log*BCF*-logK_{ow} relationships[42], the significant differences in the intercepts may explained through a significant difference in the nature of the sorbent, as the chemical class is the same for the two models[131,132,134,138]. Such explanations should however be considered with cautions, dependent on the confidence interval of the intercept and slope.

5.2.2 PLS-QSARs for predicting logK_{oc} values classified according to matrix type

To investigate and extract more detailed information with respect to the effect of differences in the characteristics of matrices on the model performance, QSARs based on PLS regression were performed. The presence of quantifiable differences in descriptors significance patterns with respect to logK_{oc} would explain the differences in intercepts of sediment contra HA regression models given in Table 5.1.

Results of a variance analysis of the raw data showed lowest mean variance in compounds specific logK_{oc} values for the matrix denoted humic acid (HA). No significant difference in variance of compound specific endpoint data for the matrix denoted sediment and the variance of endpoint data not classified according to matrix origin was observed[xvii].

Descriptors included in the models are energies of frontier orbitals, E_{HOMO} and E_{LUMO}, the second lowest and second highest MO energies, E_{NHOMO} and $E_{NLUMO,}$ the hardness of the molecules, denoted hardness ([-E_{HOMO}+E_{LUMO}]/2), the electronegativity, denoted E_{neg} ([-E_{HOMO}-E_{LUMO}]/2), the Debye di-pole moment, denoted dipolm, solvent accessible area and volume, A_{sas} and V_{sav} and van der Waals molecular surface area and volume, A_{vdW} and V_{vdW}, respectively. Non-significant, noisy descriptors, i.e. the heat of formation, dH_f, the hydration energy, E_{hyd}, the polarisability and refractivity, were excluded from the model.

In Table 5.2, the model performance parameters for the individual models classified according to matrix type are given. In analogue to Table 5.1 a model of undefined matrices was included in order to evaluate the ignorance of background data, i.e. environmental parameters.

Table 5.2 *Model performance parameters for PLS-QSARs for estimating logK_{oc} values classified according to matrix notation given in original works found in the literature.*

Endpoint	*Matrix*	*n*	*n PCs*	Q^2	R^2	*RMSEC*	*RMSEP*
LogK_{oc}	HA	27	2	0.95	0.97	0.273	0.339
LogK_{oc}	Sediment	20	1	0.82	0.86	0.527	0.610
LogK_{oc}*	Undefined	65	3	0.74	0.80	0.837	0.979

**undefined matrices: 3 fulvic acids, 2 sand, 3 silt, 1 clay and 9 undefined matrices*
n: number of compounds included in the models
n PCs: number of significant principal components
Q^2: squared cross-validated correlation coefficient
R^2: squared correlation coefficient for the calibrated model
RMSEC: root mean square error of calibration
RMSEP: root mean square error of predictions

The chemical domain of all models comprises PAHs with different numbers of condensed ring systems, as well as different hetero-analogous and hetero-substituted PAHs (cf. Appendix G)[xvii]. In Figure 12 to 14 the results of HA classified matrix (cf. Table 5.2.) is illustrated and discussed.

Structures exhibiting high leverage, i.e. Perylene and 2-hydroxy quinoline, in the model plane spanned by the principal components *PC1* and *PC2* was excluded from the model. The original variable space of the HA matrix specific model is projected into a two-dimensional space. In *PC1* 65% of the variance in X is used for explaining 88 % of the variation in logK_{oc}. In *PC2*, 17% variance in X is used for explaining 6% of the variation in logK_{oc}. In the discussion of the Figure 12 and 13, the loading vectors and score vectors are mentioned as *PC1* and *PC2*. When used with respect to Figure 12 the PCs are loading vectors and with respect Figure 13 score vectors (cf. Section 3.2.2).

In Figure 12 the loading weights of each original variable with respect to the logK_{oc} loading are shown.

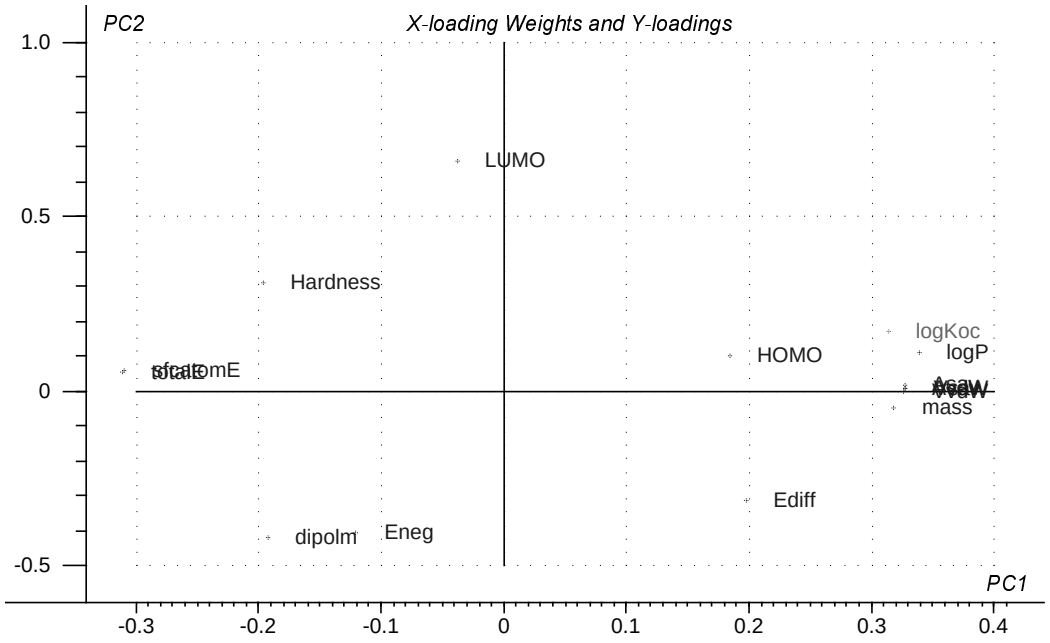


Figure 12 *logK_{oc} loading and X-loading weights showing the correlation patterns, and significance, of the individual descriptors with respect to logK_{oc} and each other. As seen from PC1, logK_{ow} (denoted logP) is strongly correlated to logK_{oc}. In addition logP is strongly correlated to the mass, area and volume descriptors. LogK_{oc} is inversely related to total energy as well as the self-consistent field energy descriptors. Only 6% of the variance in logK_{oc} in explained in PC2, by use of 17% explained X-variance.*

The first principal component, *PC1*, quantifies the variation in size and hydrophobicity of the molecular structures, which increases from left to right, and are inversely correlated to the total energy and SFC atomic energy descriptors. The second principal component, *PC2*, quantifies the variation in the inherent electronic properties of the molecular structures. Correlation patterns in variables with respect to patterns in score values of the individual PAHs are discussed.

In Figure 13 which shows a bi-plot of scores of the compounds and loading weights of original variables are shown.

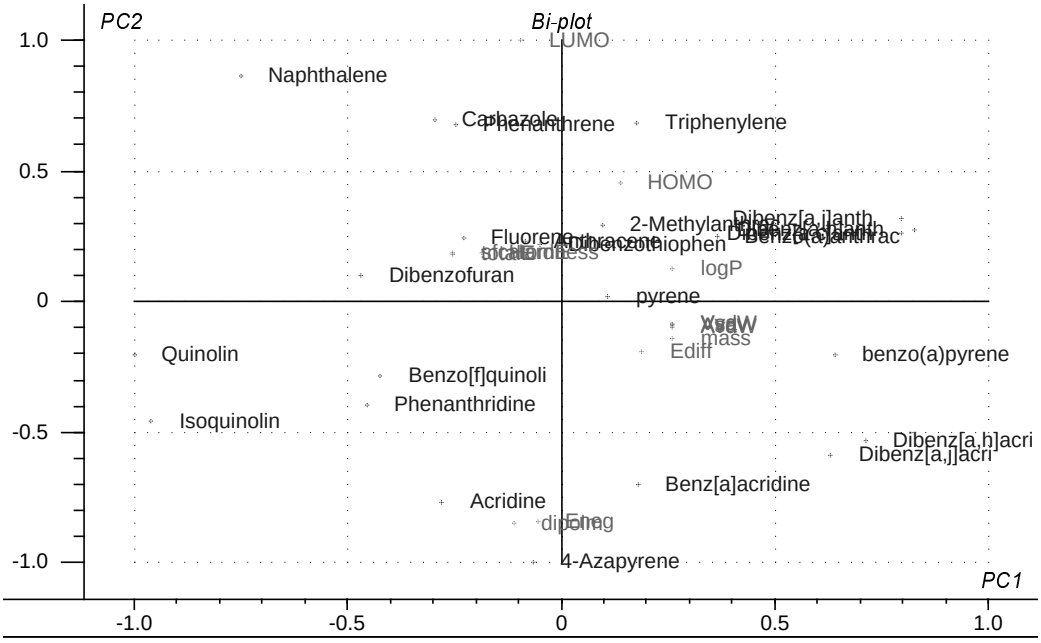


Figure 13 *Bi-plot of compound scores and variable loading weights in PC2 as function of PC1. 66% explained X-variance in PC1 is used to explain 88% of the variance in Y, i.e. logK_{oc}, 17% explained X-variance is used for explaining 6% Y-variance in PC2.*

The latent size and electronic variables *PC1* and *PC2*, respectively, explains the patterns in score values nicely. The high and negative score values of naphthalene and the N- analogues Quinoline and Isoquinoline in *PC1*, and at the same time a high range in score values in *PC2* reflects similar size but significant variations in electronic inherent properties determining the affinity for sorption to HA.

Similarly, e.g. high and positive score values of the Dibenz(a,h)- and Dibenz(a,j)- anthracenes and acridines analogues in *PC1*, showed a significant span in score values of the same compounds in *PC2*, again indicating that variations in electronic properties explains the variation in logK_{oc}.

With the exception of e.g. Carbazole, there is a general trend for all of the N- substituted or hetero analogues of the PAHs to have high negative score values in *PC2* (cf. Appendix G).

As illustrated in Figure 12 and 13 more detailed information concerning dominating variations in molecular inherent properties determining variations in endpoint values are extracted from models based on PLS[III,V]. The calibrated and cross-validated model predicted logK_{oc} values versus measured logK_{oc} values are shown in Figure 14.

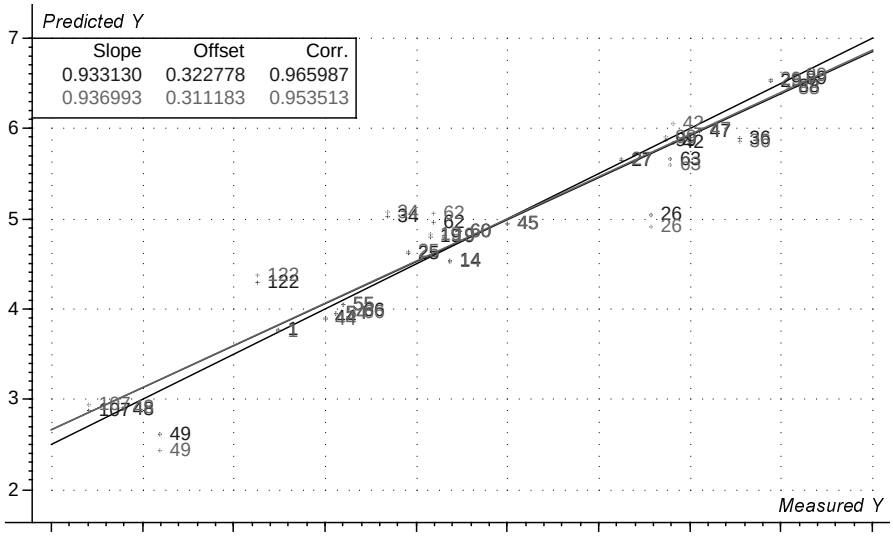


Figure 14 *Predicted versus average measured logK_{oc} values. The squared correlation coefficient of the model is 0.97 and the cross-validated square is 0.95.*

The standard deviation of predicted compared to average measured logK_{oc} values for all PAHs included in this study are given in Appendix I[xvii].

Due to the insignificance of *PC2*, the QSAR was tested by including the standard deviation in model, which showed that the variations in standard deviation on the individual compounds may be the a reason for the low explained variance in *PC2* (cf. Appendix I). Due to noise in calibration data the significance of *PC2* and conclusion concerning specific interactions between organic matter and the partitioning solute, i.e. PAH should be drawn with caution. Predictions based on the PLS-QSAR are given in Appendix I.

The most significant observations based on the available data quality are differences in sorption affinities of the PAHs to organic matter denoted HA and organic matter in sediment. This difference in explained variance is properly is simply due to the fact that PAHs show relatively large affinity for sorption to the inorganic fraction. Normalising to organic carbon therefore results in overestimated logK_{oc} values[131,132,138,144].

Furthermore, a possible explanation for the large variance in measured compound specific partitioning values to organic matter of different origin may be due to variations in sorbent characteristics. This aspect is elucidated in the following section.

5.2. Significance of Dissolved Organic Matter (DOM) as environmental parameter

In the terrestrial compartment system the vertical movement of dissolved organic matter (DOM) is a natural part of the pedogenetic process, and the mobility of DOM has been shown to depend on the nature of the immobile phase. As such Al and Fe hydrous oxides, clay minerals such as kaolinite, and soil organic matter (SOM) are major constituents controlling the retention of DOM in soils[146]. A major fraction of DOM in water, sediment and soil water consists of humic substances. However the extent to which DOM occurs in different environmental samples is seldom monitored, as focus is on specific pollutants. DOM is the major organic constituent in natural waters and show impact on the fate and effects of pollutants within the environment by changing the activity of the solutes

in the aqueous phase[59,148]. Therefore a simultaneous monitoring and investigation of the occurrence and the nature of DOM is crucial with respect to assessment of exposure concentrations as well as toxicity in natural compartments. As DOM serves as transport vehicle for environmental pollutants knowledge concerning the parameters controlling the transport of pollutants is a needed through an elucidation of the diversity and nature of the natural DOM.

5.2.1 Heterogeneous mixtures

DOM consists of humic macromolecules with multifaceted properties. Humic materials contain both hydrophobic and hydrophilic binding sites, exhibit large surface area and comprise complex chemical structures. DOM of aquatic origin usually has a high content of fulvic acids (FA), while DOM of terrestrial origin generally comprise a major fraction of humic acids (HA). Non-fractionated humic materials from the natural environment contain both types and are called humic substances (HS). As such, humic materials are rather diverse in size as well as in properties depending on origin and age. Furthermore their shape and compactness depend on environmental factors such as pH and ionic strength. A proposed average model of a humic acid molecule is given in Figure 15.

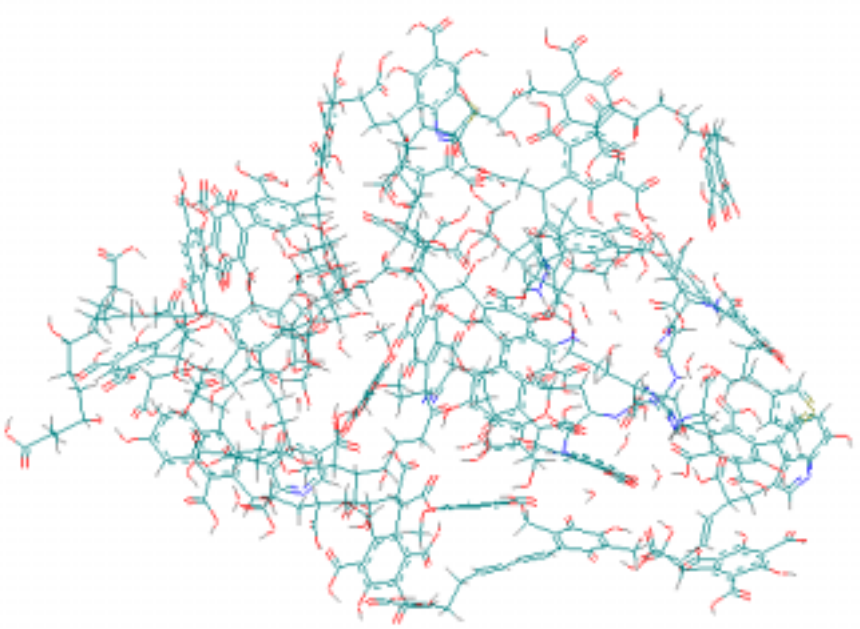


Figure 15 *Picture of an average monomeric DOM macromolecule detailed described, constructed and optimised by Schulten (1997)*[106].

As shown in Figure 15, humic macromolecules are heterogeneous structures that comprise hydrophobic voids as well as more hydrophilic domains. The structural characteristics, as well as associations within and between humic macromolecules, depend on temperature, ionic strength and concentration. Different theories concerning the sorption capacity of humic materials with respect to the partitioning of pollutants have been described. At low concentrations of DOM, e.g. HA, in the aqueous bulk, the system may be described as an aqueous solution of humic monomeric macromolecules and the pollutant, i.e. the solute. Such a system is described as a one-phase system and the effects of DOM is considered a co-solute effect, i.e. equilibrium complexation between the DOM macromolecules and pollutants within the bulk water prevails[V].

However, at higher concentration of DOM, the character of the system changes, depending on the degree of aggregation of the macromolecular humic molecules in bulk water. Colloidal systems are however not true solutions, but instead a dispersion of aggregated particles within the bulk water

phase[107,110]. There is no distinct definition of when to characterise macromolecular system as a two-phase system, i.e. a dispersion of an organic phase in water, as it depends on the inherent properties of the individual DOMs. However, at higher concentration the size of the organic macromolecular colloids increases, and the system changes from being a solution to being a two-phase systems. This aspect is important as an additional constraint on the application of Equation 5.1, i.e. dilute conditions of pollutant in all phases, is the presence of a two-phase system.

5.2.2 Testing the homogeneous phase description of DOM

As illustrated in Figure 15 organic matter may not be adequately described through a bulk octanol-phase. To gain a deeper understanding of the interactions between pollutants and organic matter, this aspect has been investigated. To "bypass" the number of unknown parameters when dealing with humic materials, an inverse QSAR concept was adopted, by selecting one pollutant and measuring the equilibrium partitioning to DOMs of different origin.

The first step is to investigate and evaluate the presence of significant quantifiable differences in the characteristics of the environmental parameter DOM. Secondary to evaluate if the environmental parameter affects the activity of the pollutant to a degree of significance.

First step was carried out by characterisation of inherent chemical properties of eight different dissolved organic matters (DOMs) originating from the natural compartments. Soil, surface- and groundwater were included. Differences in the inherent properties of DOM, were investigated by different spectroscopic methods[IV,x], and the similarities-dissimilarities investigated by use of principal component analysis are shown in Figure 16.

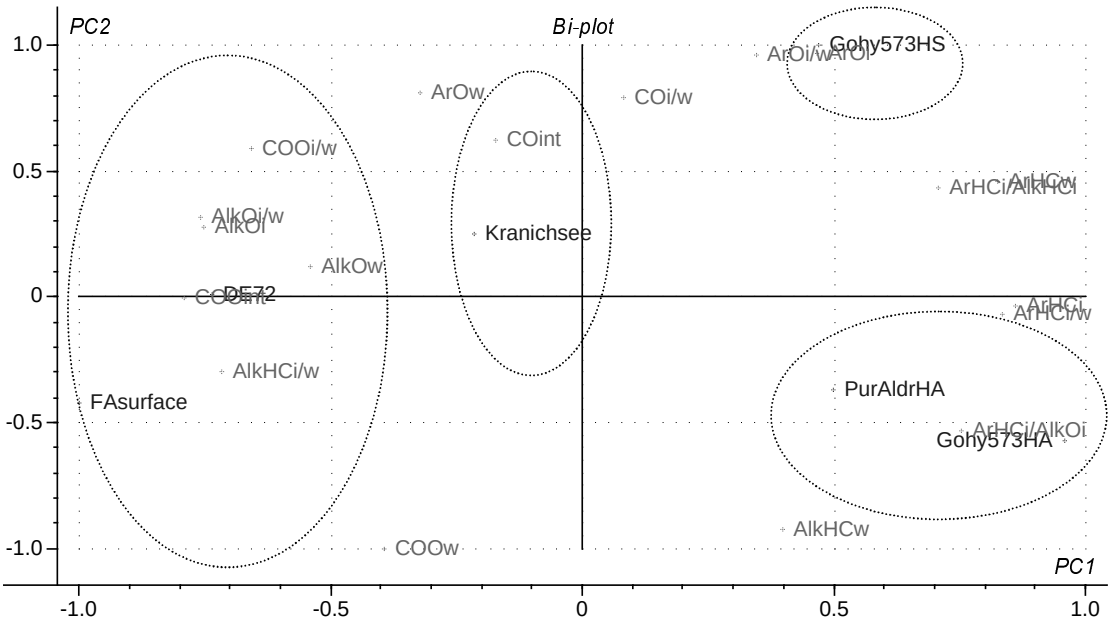


Figure 16 *Bi-plot of loadings and scores, showing PC_1 on the ordinate and PC_2 on the abscissa. PC_1 explains 47 %, and PC2 29 % of the variation in X. The score values indicate four groups, a high aliphatic high polar group (FA-surface and DE72), a high aromaticity and low polarity group (Purified Aldrich HA and Gohy-573-HA-(H^+)II), a high aromatic and more specific high content of O-substituted aromatics (Gohy-573-HS-(H^+)II)[IV].*

By linear combination of the original variables, the multi-dimensional space is reduced to a two-dimensional space, thus explaining 76 % of the variance in the original X space. Significant variabilities in the inherent properties of DOM are quantified, and explained through the correlation patterns of the original variables and score values of the individual humic materials. In the

horizontal direction, i.e. PC_1, the aromaticity decreases from left to right. In the vertical direction, PC_2 reflects decreasing contents of ketonic and O-substituted aromatic groups as one goes downward in the Figure. In addition to grouping into FAs, HAs and HSs, the score of Kranichsee HA indicates a group of properties intermediate to that of FA and HA characterised by a significant degree of aromaticity and high polarity. The robustness of the models based on descriptors types derived from the different spectroscopic methods was low due the groupings and limited number of samples, but still significant dissimilarities between humic materials of different type and origin has been quantified by use of conventional analytical methods[IV].

The sorption of the pyrethroid, esfenvalerate, to DOM of different origin showed significant variation in K_{DOM} as function of the inherent properties of DOM at a concentration level of 100 ppm. Therefore, a classification based on uniform DOM characteristics is required to improve the degree of information concerning the mechanisms of interactions in relation to measured equilibrium values. The inverse QSAR for estimating the sorbent specific logK_{DOM} values for esfenvalerate is shown in Figure 17.

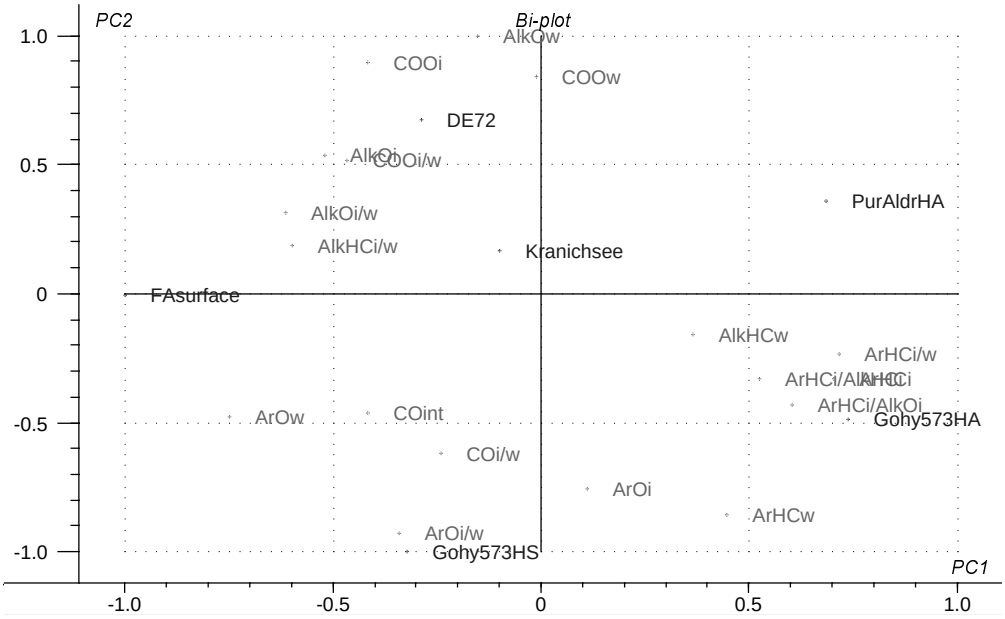


Figure 17 *The PLS model shows a homogeneous spanned X-space consisting of FA-surface, DE72, Gohy-573-HS-(H$^+$)II, Kranichsee HA, Purified Aldrich HA and Gohy-573-HA-(H$^+$)II. The first principal component, PC1, explains 78 %, and the second principal component, PC2, 16 % of the variation in K_{DOM}.*

In Figure 17, it is illustrated that the partitioning to DOM increases with increasing un-substituted or C-substituted aromaticity, quantified by the descriptors $ArHC_i$ and $ArHC_i/AlkHC_i$. The logK_{DOM} value is inversely related to the aliphatic carbon shape descriptors, $AlkHC_{i/w}$, and carbohydrate, $AlkO_{i/w}$, descriptors. Furthermore a significant inverse relation to the width descriptors, ArO_w and $ArO_{i/w}$ is observed. In *PC2*, the descriptors COO_i, $AlkO_w$ and COO_w have high positive loading weights, whereas the descriptors ArO_i, $ArO_{i/w}$ and $ArHC_w$ have high negative loading weights. The partitioning of esfenvalerate to DOM is inversely related to the O-substituted aromaticity shape and width descriptors in the third quadrant.

By fitting the X-matrix to logK_{DOM} the weighting of the original descriptor variables changes compared to the PCA model based exclusively on the DOM property descriptors. By comparing the spanning of the X-space in the PCA model (Figure 16)[IV] and the PLS model (Figure 17)[V], it is observed that there is a more homogeneous spanning of the X-space when fitted to logK_{DOM}. This

results in an elimination of samples of high leverages, and the robustness of the model by cross-validation is increased.

No simple relations seem to be able to predict the partitioning of pesticides. In addition to the decrease in bioavailability, DOM may also be an additional factor with respect to the problems in variabilities in measured degradation times. Temperature, moisture, ionic strength and pH are fairly easy to control in experimental systems for measuring degradation times. In this regard the high uncertainties in measured degradation times may partly be due to the binding to DOM. By use of the inverse QSAR concept it is possible to quantify equilibrium partitioning of esfenvalerate to DOM of varying properties. The approach is promising with respect to deriving correction factors with respect to exposure concentration, i.e. bioavailability, mobility as well as degradability, e.g. by classification of DOM into classes of similar properties and performing traditional QSARs (cf. Chapter 7).

6. Discussion and Conclusions

In this dissertation the following important aspects concerning the basis and application of scientific valid QSAR have been addressed:

- SAR and QSARs based on simple linear regression and PCA/PLS
- Data quality in relation to model performance and the development of causal models
- Data validation by including a pre-processing step in the development process for QSARs
- Process understanding, system descriptions and experimental/environmental parameter
- Informational content of empirical vs. non-empirical and quantum-chemical descriptors
- Pattern recognition for evaluating the presence of systematic noise in data, and well as for the elucidation and classification of significant system parameters

Limitations and advantages in the use and informational content of SAR and QSARs based on simple linear regression and PCA/PLS have been addressed. The two model approaches support each other well. The linear regression approach seems promising with respect to high-through-put screening analysis, while the PCA/PLS approach allows for more detailed information. The advantage of QSARs based on PLS is the ability to include many different and correlated descriptors. This allows for the extraction of information with respect to the correlation patterns between descriptors as well as the significance of the individual descriptor-endpoint relationships in a model. All correlation patterns being visually illustrated, and latent variables explain the similarity-dissimilarity in activity of the individual molecular structures included in the model. Critical points for both model approaches, however, seems to be the quality of the measured endpoint data as well as the influence of environmental parameters. Data of low quality is most pronounced for the hydrophobic chemicals, possibly due to analytical limitations with respect to detection of compounds of low solubility. However, the most critical point seems to be that the activity of the chemical, within the phases of the experimental system, is a function of the concentration and activity coefficient of the chemical in the respective phases[II].

Significant systematic variations in the standard deviation of measured endpoint data found in the scientific literature exist. In addition, it is not unusual that the variations in endpoint data from different sources are significantly higher than the between-compound variation. The majority of QSAR studies published in the literature do not include a data pre-processing step, and as such the probability of chance correlations increases. These aspects are critical with respect to the use of QSARs, as the robustness and causality of the models are restricted by the quality of experimental calibration data.

Based on the on combined data analysis[v,iii], SAR/QSAR investigations[v] and experimentally well designed measurements of the dissolution process of the phthalates[I,II] it has been shown that data quality and process understanding is crucial for the development of scientific valid QSARs. Process understanding and increase of the data quality used for calibration QSAR models can be obtained through a combination of classical statistics[iii, v, III, V] and PARC techniques[III, IV,V].

The activity of a chemical compound may depend on several combined parameters. First of all, the activity, in its thermodynamic definition, changes with the bulk concentration of the chemical compounds at concentration levels, which do not obey the requirement of the dilute solution description. This means that even for constant system parameters, the measured endpoint may vary with the experimental concentration level. Therefore, empirical non-linear QSARs, such as the bi-linear and parabolic logBCF-logK_{ow} relationships, discussed in Chapter 4 and Appendix J, need to be based on solid investigations of sources and patterns in the inhomogeneities in the standard deviations of compound specific data, upon which the models are built. Pre-processing of simple endpoint data, such as the aqueous solubility, reveals that significant bias in literature data exists. Therefore, the use of classical statistical analysis of variance should be used with caution due to the centrality principle. As the hypothesis of variance homogeneity of all compound endpoints within a given data set is not valid, the standard deviations in endpoint data should be included in the initial PARC analysis, e.g. PCA as illustrated in Chapter 5.

In addition to the requirements of dilute solution conditions, additional system parameters, as well as uncertainties in endpoint data should be included in a data pre-processing step.

This is emphasised by several examples in this dissertation listed below:

a) The influence of exposure concentration on bioconcentration in aquatic organisms
b) The PARC screening analysis of pesticide solubilities including system parameters pH, and temperature, as well as uncertainty in endpoint data classified according to the environmental parameters
c) The insignificance of the second principal component in the PLS-QSAR for estimating the partitioning of a heterogeneous class of PAHs constituting N, S, O-PAH analogous, due to noise in data.
d) The inverse QSAR for predicting sorption of esfenvalerate to DOM of different type and origin.
e) Identification of class specific empirical descriptors, exemplified for the PCBs.

Significant influence of data quality, process understanding and environmental parameters have been shown to be the main aspects with respect to the QSAR paradigm. However, due to missing or inconsistent data, a validation of the influence of environmental parameters is not possible in all cases, and the data is primarily used for eliminating outliers.

When monitoring the occurrence of pollutants in environmental matrices, it is important to know what is actually being measured. This dissertation has focussed on data quality and process understanding in relation to measurements of environmental risk assessment endpoints. Serious inconsistency problems in endpoint data have been investigated, which may lead to large errors with respect to calculation of risk, i.e. PEC/PNEC ratios. For hydrophobic compounds, the problem of inconsistency in data seems to be due to the characterisation of the aqueous phase, which depends on concentration levels exceeding the unimeric solubility of these compounds. The aqueous phase participates in almost all processes in the terrestrial and aquatic environments. The problems described for aqueous solutions will therefore influence partitioning as well as biotic and abiotic processes, and is thus a critical parameter with respect to risk assessment of xenobiotics in the environment.

7. Perspectives

The exposure concentration within the natural compartment systems is dependent on the fate processes. Fate processes are not only determined by compound specific parameters, but also system specific, or environmental, parameters. In most or all QSAR studies found in the literature no investigations include background data in the model analysis. Furthermore, the EUSES concept is based on homogeneous mixing and does only include compound specific endpoint values. This dissertation has shown that environmental parameters are crucial with respect to fate modelling.

Additionally solutions may be mixtures or homogeneous true solutions, which have distinct characteristics that determine the activity of pollutants in experimental systems as well as in the natural environment. Therefore, it is of utmost importance that system specific and/or environmental conditions are well defined and similar in exposure and dose-response measurements. Through a combination of theoretical considerations and use of proper analytical methods the characteristics of the measured systems need to be included in the SAR/QSAR as well as compartment modelling process.

Fate processes, such as sorption to fixed (FOM), and mobile, i.e. dissolved organic matter (DOM) determines the mobility as well as bioavailability of organic pollutants. DOM is omnipresent throughout the aqueous compartment systems. However, the influence of DOM on the migration behaviour of pollutants within the environment is only sparsely elucidated, although a number of investigations dealing with the inherent properties of natural organic matter of terrestrial and aquatic origin exist. Surface runoff, as well as infiltration of DOM in soil profiles, are well-known processes, but still these environmental parameters are not included in the EUSES for calculating the exposure concentrations within the natural compartments. In addition to influencing pollutant mobility, DOM also has a pronounced effect on the bioavailability, which is a most important property determining the dose-response relationship.

A sensitivity analysis of a topsoil compartment model describing the fate of DEHP clearly illustrates the need for an increase in the data quality and process understanding[VI]. The results show that the effect of DOM or particulate organic matter is a central parameter for the assessment of the fate as well as effects of pollutants within the environment, and should therefore be analysed in combined well designed experiments as indicated by the red arrows in Figure 2. To account for the presence of DOM correction factors in relation to fixed toxicity endpoints as well as fate parameters may be needed in future risk assessment[VI]. Such correction factors, and/or classification of compartments according to significant variations in environmental parameters, are able to evaluate the effect of the presence of mixtures and/or colloidal dispersions within bulk phases. One way of quantifying the effect DOM is through Equations 7.1 and 7.2 below

$$K_{OC}^{eff} = K_{FOC} + \alpha \cdot C_{DOM} \cdot f_{DOM}^{K_{oc}} \cdot f_{xenob}^{K_{oc}} \tag{7.1}$$

$$L(E)C_{50}^{eff} = L(E)C_{50} + \alpha \cdot C_{DOM} \cdot f_{DOM}^{L(E)C_{50}} \cdot f_{xenob}^{L(E)C_{50}} \tag{7.2}$$

where K_{OC}^{eff} and $L(E)C_{50}^{eff}$ express the effective sorption of specific pollutants to organic carbon and effective effect concentration. These values are quantified as functions of the conventionally established parameters, K_{FOC} and $L(E)C_{50}$, that are defined for systems comprising a dissolved, or dispersed, organic matter phase as well as fixed sorbents, e.g. soil organic matter. C_{DOM} is the

concentration of dissolved organic matter. The inherent properties of DOM are quantified by, e.g., PCA, which forms the basis for deriving the quality indices for DOM, f_{DOM}, through the latent variables, i.e. PCs[IV]. QSARs for estimating the equilibrium partitioning coefficients for pollutants are expressed in the correction factors $f_{xenob}^{K_{OC}}$ and $f_{xenob}^{L(E)C_{50}}$, quantifying the change in activity of pollutants in the aqueous bulk phase. The constant α may take the value +1 or –1, α.

There is a need not only to monitor the occurrence of specific environmental hazardous pollutants, but also to monitor the simultaneous occurrence of DOM, when analysing samples taken in the aqueous and terrestrial environment. The reason for this is the still increasing possibility of matrix effect not only caused by the presence of DOM, but also from possible co-solute effects in aqueous samples. Measuring the amount of simultaneously occurrence of DOM in environmental samples will provide information concerning an expected key parameter within fate and risk assessment in relation to pollutants and their transportation into the aquatic environment. With respect to the fate of pollutants this would make it possible to elucidate the effects of DOM on the transport patterns of different classes of pollutants as functions of time, and space and presence of simultaneously occurring components.

Analogously, correction parameters, as described in Equations 7.1 and 7.2, can in principle be applied in the TGD model design for process parameters describing transport, degradation and partitioning between phases. In this way it is possible to include environmental parameters that influence the fate and effects of environmental pollutants in aquatic, soil and sediment matrices.

Literature cited

1 The International Uniform Chemical Information Database (IUCLID), 1996.

2 European Inventory of Existing Chemicals (EINECS) List. (http://ecb.ei.jrc.it/existing-chemicals/).

3 The 1997 OECD List of High Production Volume Chemicals. Environment Directorate. Organisation for Economic co-operation and development, Paris 1997. Internet Edition, http://www.oecd.org/ehs/ehsmono/HPV97.doc.

4 Pölloth, C., Mangelsdorf, I. (1997) *Commentary on the Application of (Q)SAR to the toxicological evaluation of existing chemicals*. Chemosphere 35, 2525-2542

5 Blumenstock, M., Zimmermann, R., Schramm, K.-W., Kettrup, A. (2000) *Influence of combustion conditions on the PCDD/F-, PCB-, PCBz- and PAH-concentrations in the post-combustion chamber of a waste incineration pilot plant*. Chemosphere 40, 987-993.

6 Durlak, S. K., Biswas, P., Shi, J., Bernard, M. J. (1998) *Characterization of polycyclic hydrocarbon particulate and gaseous emmissions from polystyrene combustion.* Environmental Toxicology and Chemistry 32, 2301-2307.

7 Zimmermann, R., van Vaeck, L., Davidovic, M., Beckmann, M., Adams, F. (2000) *Analysis of Polycyclic Aromatic hydrocarbons (PAH) adsorbed on soot particles by Fourier transform laser microprobe mass Spectrometry (FT LMMS): variation of the PAH patterns at different positions in the combustion chamber of an incineration plant*. Environ. Sci. Technol. 34 (22), 4780-4788.

8 Commission of the European Communities. White Paper-Strategy for a future Chemicals Policy. Bruxelles, 2001. Internet Edition, http://www.europa.eu.int-comm-environment-chemicals-whitepaper.htm.

9 EC European Commission (1993) Commission Directive 93/67/EEC of 20 Juli 1993 laying down the principles for assessment of risks to man and the environment of substances notified in accordance with Council Directive 67/548/EEC. Official Journal L 227, 08/09/1993, pp. 9-18.

10 EC European Commission (1996) *Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New and Notified Substances and Commission Directive (EC) No. 1488/94 on Risk Assessment for Existing Substances. 1996. Luxembourg: Office for Official Publications of the European Communities. Part I. General Introduction, Risk Assessment for Human Health*, ISBN 92-827-8011-2.

11 EC European Commission (1996) *Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New and Notified Substances and Commission Directive (EC) No. 1488/94 on Risk Assessment for Existing Substances. 1996. Luxembourg: Office for Official Publications of the European Communities. Part II. Environmental Risk Assessment,* ISBN 92-827-8012-0.

12 EC European Commission (1996) *Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New and Notified Substances and Commission Directive (EC) No. 1488/94 on Risk Assessment for Existing Substances. 1996. Luxembourg: Office for Official Publications of the European Communities. Part III. Use of QSAR, Use Categories,* Risk Assessment Format, CR-48-96-003-En-C, ISBN 92-827-8013-9.

13 EC European Commission (1996) *Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New and Notified Substances and Commission Directive (EC) No. 1488/94 on Risk Assessment for Existing Substances. 1996. Luxembourg: Office for Official Publications of the European Communities. Part IV. Emission Scenario Documents*, CR-48-96-003-En-C, ISBN 92-827-8013-9.

14 EC European Commission (1994) Commission Regulation (EC) No 1179/94 of 25 May 1994 concerning the first list of priority substances as foreseen under Council Regulation (EEC) No 793/93, (http://ecb.ei.jrc.it/).
15 EC European Commission (1995) Commission Regulation (EC) No 2268/95 of 27 September 1995 concerning the second list of priority substances as foreseen under Council Regulation (EEC) No 793/93, (http://ecb.ei.jrc.it/).
16 EC European Commission (1997) *Commission Regulation (EC) No 143/97 of 27 Januar 1997 concerning the third list of priority substances as foreseen under Council Regulation* (EEC) No 793/93, (http://ecb.ei.jrc.it/).
17 EC European Commission (2001) *Commission Regulation (EC) No xxx concerning the third list of priority substances as foreseen under Council Regulatio*n (EEC) No 793/93. Draft, (http://ecb.ei.jrc.it/).
18 *Risk assessment. Di-butyl phthalate*, CAS-No.: 84-74-2, EINECS-No.: 201-557-4, Final version, 17 October 2000.
19 *Risk assessment. Benzyl butyl phthalate*, CAS-No.: 85-68-7, EINECS-No.: 201-622-7, Chapter 1 General Substance Information, Chapter 2 General Information On Exposure, Chapter 3 Environment, Draft of December 2000.
20 *Risk assessment. Bis(2-ethylhexyl) phthalate*, CAS-No.: 117-81-7, EINECS-No.: 204-211-0, Draft May 2000.
21 *Risk assessment. 1,2-Benzenedicarboxylic acid, di-C9-11-branched alkyl esters, C10-rich and di-''isodecyl''phthalate,* CAS-No.: 68515-49-1 and CAS-No.: 26761-40-0, EINECS-No.: 271-091-4 and EINECS-No.: 247-977-1, Draft of November 2000.
22 *Risk assessment. 1,2-Benzenedicarboxylic acid, di-C8-10-branched alkyl esters, C9-rich and di-''isononyl'' phthalate,* CAS-No.: 68515-48-0 and CAS-No.: 28553-12-0, EINECS-No.: 271-090-9 and EINECS-No.: 249-079-5, Draft of August 2000.
23 Karcher,W., Devillers, J. (Eds.) (1992) *Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology*. Kluwer Academic Publishers, London, 470 pages.
24 Eriksson, L., Johansson, E. (1996) *Multivariate design and modelling in QSAR*. Chemom. Intell. Lab. Syst. 34, 1-19.
25 Eriksson, L., Hermens, J. L. M., Verhaar, H. J. M., Wold, S. (1995) *Multivariate analysis of aquatic toxicity data with PLS*. Aqua. Sci. 57/3, 1015-1621.
26 Cramer, R. D., Bunce, J. D., Patterson, D. E., Frank, I. E. (1988) *Crossvalidation, Boottrapping, and Partial Least Squares compared with Multiple Regression in conventional QSAR studies*. QSAR 7, 18-25.
27 Hermens, J. (1995) *Assesment of QSARs for predicting fate and effects of chemicals in the environment: an international european project*. SAR QSAR Environm. Res. 3, 223-236.
28 Guidance for the use of structure-activity relationships (SAR) in the HPV chemicals programme.
29 The use of Structure-activity Relationships (SAR) in the High Production Volume Chemicals Challenge Program. Chemical Right-to-Know Initiative, US EPA, Office of Pollution Prevention and Toxics. (http://www.epa.gov/chemrtk/sarfinl1.htm).
30 Verhaar, H. J., Ramos, E. U., Hermens, J. L. (1996) *Classifying Environmental Pollutants. 2: Separation of Class 1 (Baseline Toxicity) and Class 2 ("Polar Narcosis") type Compounds based on Chemical Descriptors*. J. Chemometrics 10, 149-162.
31 Verhaar, H. J., Solb, J., Speksnijder, J., van Leeuwen, C. J., Hermens, J. M. (2000) *Classifying environmental pollutants: Part 3. External validation of the classification system*. Chemosphere 40, 875-883.
32 Elmegaard, N., Jagers op Akkerhuis, G. A. J. M. (2000) *Safety Factors in Pesticide Risk Assessment.* Ministry of Environment and Energy, National Environmental Research Institute, Department of Terrestrial Ecology. NERI Technical Report. Map 325.

33 Juberg, D.R., Kava, R., Lukachko, A., Ponirovskaya, Y. (1999) *Traces of Environmental Chemicals in the Human Body: Are they a risk to health?*, The American Council on Science and Health, NY (http://www.acsh.org).
34 Cousins, I., Mackay, D. (2000) *Correlating the physical-chemical properties of phthalate esters using the "three solubility" approach.* Chemosphere 41, 1389-1399.
35 DeBruijn, J., Busser, F., Seinen, W., Hermens, J. (1989) *Determination of Octanol/Water Partition Coefficients for Hydrophobic Organic Chemicals with the "slow-stirring" method.* Environmental Toxicology and Chemistry 8, 449-512.
36 Brooke, D., Nielsen, I., de Brujin, J., Hermens, H. (1990) *An interlaboratory evaluation of the stir-flask method for the determination of octanol-water partition coefficients (log P_{ow})*. Chemosphere 21, 119-133.
37 Ellington, J. J. (1999) *Octanol/water partition coefficients and water solubilities of phthalates esters.* J. Chem. Eng. Data 44, 1414-1418.
38 Staples, C. A., Peterson, D. R., Parkerton, T. F., Adams, W. J. (1997) *The environmental fate of phthalate esters: a literature review.* Chemosphere 35, 667-749.
39 McAuliffe, C. (1969) Solubility in water of normal C9 and c10 alkane hydrocarbons. Science 163, 478-479.
40 Slob, W. (1994) *Uncertainty Analysis in Multiplicative Models*. Risk Anal. 14, 571-576.
41 Hansen, B.G., Paya-Perez, A. B., Rahman, M., Larsen, B. R. (1999) *QSAR's for K_{ow} and K_{oc} of PCB congeners: A critical Examination of Data, Assumptions and Statistical Approaches.* Chemosphere 39, 2209-2228.
42 Nendza, M. (1991) *QSARs of bioconcentration: validity assessment of logP_{ow}/logBCF correlations. Bioaccumulation in aquatic systems*, Nagel, R, 43-66.
43 Haitzer, M., Höss, S., Trauspurger, W., Steinberg, C. (1998) *Effects of dissolved organic matter (DOM) on the bioconcentration of organic chemicals in aquatic organisms - A review.* Chemosphere 37, 1335-1362.
44 Gordon, J. E., Thorne, R. L. (1967) *Salt effects on non-electrolyte activity coefficients in mixed aqueous electrolyte solutions-II. Artificial and natural sea waters.* Geochimica et Cosmochimica Acta 31, 2433-2443.
45 Boehm, P. D., Quinn, J. G. (1973) *Solubilization of hydrocarbons by the dissolved organic matter in sea water.* Geochimica et Cosmochimica Acta 37, 2459-2477.
46 Sutton, C., Calder, J. A. (1975) Solubility of alkylbenzenes in distilled water and seawater at 25,0 degrees. J. Chem. Eng. Data 20, 320-323.
47 Rao, P. S. C., Hornsby, A. G. T., Kilcrease, D. P., Nkedi-Kizza, P. (1985) *Sorption and transport of hydrophobic organic chemicals in aqueous and mixed solvent systems: model development and preliminary evaluation*. J. Environ. Qual. 14, 376-382.
48 Yalkowsky, S. H., Amidon, G. L., Zografi, G., Flynn, G. L. (1975) *Solubility of nonelectrolytes in polar solvents III: Alkyl p-aminobenzoates in polar and mixed solvents*. J. Pharm. Sci. 64, 48-52.
49 Lassen, P., Carlsen, L. (1999) *The effect of humic acids on the water solubility and water - organic carbon partitioning of fluorene and its NSO-heteroanalogues: Carbazole, dibenzofuran and dibenzothiophene*. Chemosphere 38, 2959-2968.
50 Burgess, R. M., McKinney, R. A., Brown, W. A. (1996) *Enrichment of marine sediment colloids with polychlorinated biphenyls: Trends resulting form PCB solubility and chlorination.* Environ. Sci. Technol. 30, 2556-2566.
51 Landrum, P. F., Nihart, S. R., Eadle, B. J., Gardner, W. S. (1998) *Reverse-phase separation method for determining pollutant binding to Aldrich Humic acid and dissolved organic carbon of natural waters.* Environ. Sci. Technol. 18, 187-192.
52 Wauchope, D. R., Getzen, F. W. (1972) *Temperature dependence of solubilities in water and heats of fusion of solid aromatic hydrocarbons.* J. Chem. Eng. Data 17, 38-41.
53 Li, A., Doucette, W. J., Andren, A. W. (1992) *Solubility of polychlorinated biphenyls in binary water/organic solvent systems.* Chemosphere 24, 1347-1360.

54 Schulze, M., Wilkes, H., Vereecken, H. (1999) *ct determination of hydrophobic organic compounds in aqueous solution in the presence of dissolved organic carbon by high-performance liquid chromatography*. Chemosphere 39, 2365-2374.
55 Chien, Y.-Y., Bleam, W. F. (1997) *Fluorine-19 nuclear magnetic resonance study of Atrazine in humic and sodium dodecyl sulfate micelles swollen by polar and nonpolar solvents.* Langmuir 13, 5283-5288.
56 Eadie, B. J., Morehead, N. R., Landrum, P. (1990) *Three-phase partitioning of hydrophobic organic compounds in great lakes waters*. Chemosphere 20, 161-178.
57 Morehead, N. R., Eadie, B. J., Lake, B., Landrum, P. F., Berner, D. (1986) The sorption of PAH onto dissolved organic matter in lake Michigan Waters. Chemosphere 15, 403-412.
58 Geyer, H., Viswanathan, R., Freitag, D., Korte, F. (1981) *Relation between water solubility of organic chemicals and their bioaccumulation by the alga Chlorella*. Chemosphere 10, 1307-1313.
59 Kukkonen, J., Oikari, A. (1991) *Bioavailability of Organic Pollutants in Boreal Waters with Varying Levels of Dissolved Organic Material.* Wat. Res. 25, 455-463.
60 Wildi, E., Nagel, R., Steinberg, C.E.W. (1994) *Effects of pH on the bioconcentration of pyrene in the larval midge, Chironomus riparius.* Wat. Res. 28(12), 2553-2559.
61 Fahl, G. M., Kreeft, L., Altenburger, R., Faust, M., Boedeker, W., Grimme, L. H. (1995) *pH-dependent sorption bioconcentration and algae toxicity of sulfonylurea herbicides.* Aquatic Toxicol. 31, 175-187.
62 Haitzer, M., Höss, S., Traunspurger, W., Steinberg, C. (1999) *Relationship between concentration of dissolved organic matter (DOM) and the effect of DOM on the bioconcentration of benzo(a)pyrene.* Aquatic Toxicol. 45, 147-158.
63 Steinberg, C. E. W., Haitzer, M., Brüggemann, R., Perminova, I. V., Yashchenko, N. Y., Petrosyan, V. S. (2000) *Towards a quantitative structure activity relationship (QSAR) of dissolved humic substances as detoxifying agents in freshwaters.* Internat. Rev. Hydrobiol. 85, 253-266.
64 Boese, B. L., Ozretich, R. J., Lamberson, J. O., Swartz, R. C., Cole, F. A., Pelletier, J., Jones, J. (1999) *Toxicity and phototoxicity of mixtures of highly lipophilic PAH compounds on marine sediment: can the sum-PAH model be extrapolated?* Arch.Environ.Contam.Toxicol. 36, 270-280.
65 Landrum, P. F., Eadie, B. J., Faust, W. R. (1991) *Toxicokinetics and toxicity of a mixture of sediment-associated polycyclic aromatic hydrocarbons to the Amphipod Diporeia sp.* Environ. Toxicol. Chem. 10, 35-46.
66 USEPA. *Guidelines for ecological risk assessment*. EPA/630/R-95/002F. Washington, DC:EPA, 1998.
67 ECETOC. *Environmental hazard assessment of substances,* Technical report No. 51. Brussels: ECETOC, 1992.
68 EC. *Classification, packaging and labelling of dangerous substances in the European Union.* Luxembourg: Office for Official Publications of the European Communities, 1997.
69 Trapp, S., Schwartz, S. (2000) Proposal to overcome limitations in the EU chemical risk assessment scheme. Chemosphere 41, 965-971.
70 Jager, T., den Hollander, H.A., Janssen,G.B., van der Poel,P., Rikken, M.G.J., Vermeire, T.G. (2000) *Probabilistic risk assessment for new and existing chemicals: Sample calculations.* National Institute of Public Health and the Environment (RIVM), Report 679102049, Bilthoven.
71 Jager, T., Rikken, M. G. J., van der Poel, P. (1997) *Uncertainty analysis of EUSES: Improving risk management by probabilistic risk assessment.* National Institute of Public Health and the Environment (RIVM), Report 679102039, Bilthoven.
72 Brandes, L.J., den Hollander, H., ven de Meent, D. (1996) *SimpleBox 2.0: a nested multimedia fate model for evaluating the environmental fate of chemicals,* Report no.

719101029. National Institute of Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, the Netherlands.

73 Jager, D. T., Slob, W. (1995) *Uncertainty Analysis of the Uniform System for the Evaluation of Substances (USES)*. National Institute of Public Health and the Environment (RIVM), Report 679102027, Bilthoven.

74 *Paradigm for Substance Flow Analyses – Guide for SFA's Carried Out for the Danish EPA* (2000) Danish Environmental Protection Agency. Draft.

75 Mackay, D. (1991) *Multimedia Environmental Models – The Fugacity Approach.* Lewis Publishers.

76 Kamlet, M. J., Doherty, R. M., Abraham, M. H., Marcus,Y., Taft, R. W. (1988) *Linear Solvation Energy Relationships. 46. An Improeved Equation for correlation and prediction of octanol/water partition coefficients of organic nonelectrolytes (Including Strong Hydrogen Bond Donor Solutes)*. J. Phys. Chem. 92, 5244-5255.

77 McDuffie, B. (1981) *Estimation of Octanol/water Partitioning Coefficients for Organic Pollutants using reverse-phase HPLC*. Chemosphere 10, 73-83.

78 Hawker, D. W., Connell, D. W. (1988) *Octanol-water partitioning coefficients of polychlorinated biphenyl congeners.* Environ. Sci. Technol. 22, 382-387.

79 Sabljic', A., Güsten, H., Verhaar, H., Hermens, J. (1995) *QSAR modelling of soil sorption. Improvements and systematics of log K_{oc} versus logK_{ow} correlations*. Chemosphere 31, 4489-4514.

80 Bintein, S., Devillers, J. (1994) *QSAR for organic chemical sorption in soils and sediments.* Chemosphere 28, 1171-1188.

81 Sabljic', A., Protic', M. (1982) *Relationship between molecular connetivity indices and soil sorption coefficients of polycyclic aromatic hydrocarbons.* Bull. Environm. Contam. Toxicol. 28, 162-165.

82 Tao, S., Lu, X. (1999) *Estimation of organic carbon normalized sorption coefficient (Koc) for soils by topological indices and polarity factors.* Chemosphere 39, 2019-2034.

83 Damborskì, J. (1996) *A mechanistic Approach to deriving Quantitative Structure-Activity Relationship models for Microbial Degradation of Organic Compounds.* SAR QSAR Environn. Res. 5, 27-36.

84 Lu, X., Tao, S., Dawson, R. W. (1999) *Prediction of fish bioconcentration factors of nonpolar organic pollutants based on molecular connectivity indices*. Chemosphere 39, 987-999.

85 Lu, X., Tao, S., Hu, H., Dawson, R. W. (2000) *Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors.* Chemosphere 41, 1675-1688.

86 Lange, A. W., Vormann, K. (1995) *Experiences with the application of QSAR in the routine of the notification procedure.* SAR QSAR Environn. Res. 3, 171-177.

87 Dunnivant, F.M., Elzerman, A. W. (1988) *Aqueous solubility and Henry´s law constant data for PCB congeners for evaluation of quantitative structure-property relationships (QSPRs).* Chemosphere 17, 525-541.

88 Krop, H. B., van Velzen, M. J. M., Parsons, J. R., Govers, H. A. J. (1997) *n-octanol-water partition coefficients, aqueous solubilities and Henrys law constants of fatty acid esters.* Chemosphere 34, 107-119.

89 Miller, M. M., Waslk, S. P., Huang, G. L., Shiu, W. Y., Mackay, D. (1985) *Relationships between octanol-water partition coefficient and aqeous solubility.* Environ. Sci. Technol. 19, 522-529.

90 Dowdy, D. L., McKone, E. T. (1997) *Predicting Plant Uptake of Organic Chemmicals from Soil or Air using Octanol/water and octanol/air Partition ratios and Molecular Connectivity Index.* Environ. Toxicol. Chem. 16, 2448-2456.

91 Mackay, D. (1982) *Correlation of bioconcentration factors.* Environ. Sci. Technol. 16, 274-278.

92 Verhaar, H. J. M., Jong, J. D., Hermens, J. L. M. (1999) *Modeling the Bioconcentration of Organic Compounds by Fish: A Novel Approach.* Environ. Sci. Technol. 33, 4069-4072.
93 Connell, D. W., Hawker, D. W. (1988) *Use of polynomial expressions to describe the bioconcentration of hydrophobic chemicals by fish.* Ecotox. Environ. Safety 16, 242-257.
94 Hawker, D. W., Connell, D. W. (1986) Bioconcentration of lipophillic compounds by som aquatic organisms. Ecotox. Environ. Safety 11, 184-197.
95 Hawker, D. W., Connell, D. W. (1985) *Relationship between partition coefficient, uptake rate constant, clearance rate constant and time to equilibrium for bioaccumulation.* Chemosphere 14, 1205-1219.
96 Meylan, W. M., Howard, P. H., Boethling, R. S., Aronson, D. (1999) *Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partitioning coefficient.* Environ. Toxicol. Chem. 18, 664-672.
97 Legierse, K. C. H. M., Verhaar, H. J. M., Vaaes, W. H. J., De Bruijn, J. H. M., Hermens, J. L. M. (1999) *Analysis of the Time-Dependent Acute Aquatic Toxicity of Organophosphorus Pesticides: The Critical Target Occupation Model.* Environ. Sci. Technol. 33, 917-925.
98 Roberts, D. W., Marshall, S. J. (1995) *Application of hydrophobicity parameters to prediction of the acute aquatic toxicity of commercial surfactant mixtures.* SAR QSAR Environn. Res. 4, 167-176.
99 Schultz, T. W., Moulton, B. A. (1985) *Structure-Activity relationships for Nitrogen-containing aromatic molecules.* Environ. Toxicol. Chem. 4, 353-359.
100 Zhao, Y. H., Cronin, M. T. D., Dearden, J. C. (1998) *Quantitative Structure-Activity Relationship of chemicals acting by non-ploar narcosis-theoretical considerations.* QSAR 17, 131-138.
101 Hansch, C., Fujita, T. (1995) *Classical and Three-dimensional QSAR in Agrochemistry.* ACS Symposium Seires 606 ed. American Chemical Society, Washington, DC. 342 pages.
102 Müller, M., Kördel, W. (1996) *Comparison of screening methods for estimation of adsorption coefficients on soil.* Chemosphere 32, 2493-2504.
103 Karickhoff, S. W. (1981) *Semi-empirical estimation of sorption of hydrophobic pollutants on naturel sediments and soils.* Chemosphere 10, 833-846.
104 Schwarzenbach, R. P., Gschwend, P. M., Imboden, D. M. (1993) *Environmental organic chemistry.* Wiley-Interscience, New York.
105 Bintein, S., Devillers, J. (1993) *Nonlinear dependence of fish bioconcentration on n-octanol/water partition coefficient.* SAR QSAR Environn. Res. 1, 29-39.
106 Schulten, H.-R. (1999) *Analytical Pyrolysis and Computational Chemistry of Aquatic Humic Substances and Dissolved Organic Matter.* J. Anal. Appl. Pyrol., 49, 385-415.
107 Hunter, J. R. (1995) Foundations of colloid science. Vol. I, Clarendon Press, Oxford
108 Ran, Y., Fu, J. M., Sheng, G., Beckett, R., Hart, B., T. (2000) *Fractionation and composition of colloidal and suspended particulate materials in rivers.* Chemosphere 41, 33-43.
109 Mato, Y., Isobe, T., Takada, H., Kanehiro, H., Ohtake, C., Kaminuma, T. (2001) *Plastic Resin Pellets as a tranport medium for toxic chemicals in the marine environment.* Environ. Sci. Technol. 35, 318-324.
110 Hiemenz, P. C., Rajagopalan, R. (1997) *Principles of Colloid and Surface Chemistry. Third Edition, Revised and Expanded.* Marcel Dekker, Inc.
111 Mabey, W., Mill, T. (1978) *Critical review of organic compounds in water under environmental conditions.* J. Phys. Chem. Ref.data 7, 383-415.
112 Hansch, C., Quinlan, J. E., Lawrence, G. L. (1968) *The LFER between Partition Coefficients and the aqueous Solubility of Organic Liquids.* J. Org. Chem. 33, 347-350.
113 Van der Burght, A. S. A. M., Tysklind, M., Andersson, P. L., Horbach, G. J., van der Berg, M. (2000) *Structure dependent induction of CYP1A by polyhlorinated biphenyls in hepatocytes of male castrated pigs.* Chemosphere 41, 1697-1708.
114 Miller, J. C., Miller, J. N. (Eds.) (1988) *Statistics for analytical chemistry.* 2nd ed. Ellis Horwood, Chichester. 227 pages.

115 Höskuldsson, A. (1996). *Prediction Methods in Science and Technology.* Thor Publishing, Denmark, p. 405.
116 CAMO ASA (1998) *The Unscrambler 7.01,* Oslo, Norway.
117 Wold, S., Sjöström, M. (1998) *Chemometrics and its Roots in Physical Organic Chemistry*. Acta Chem. Scan. 52, 517-523.
118 Skagerberg, B. (1989) *Principal Properties in Design and Structural Description in QSAR*. Ph.D. Dissertation, Research Group for Chemometrics, Department of Organic Chemistry, University of Umeå, S-901 87 Umeå. 55 p.
119 Hall Associates Consulting (1999) *Molconn-Z 3.50,* Quincy, MA, U.S.A.
120 Hypercube Inc. (1995) *Hyperchem 4.5 for SGI*, Gainsville, FL, U.S.A.
121 Borisov, M. D., Graber, E. R. (1998) *Organic Compound Sorption Enthalpy and Sorption Mechanisms in Soil Organic Matter.* J. Environ. Qual. 27, 312-317.
122 Carey, F. A., Sundberg, R. J. (1991) *Advanced Organic Chemistry – Part A: Structure and Mechanisms.* Third Edition. Plenum Press, New York and London.
123 Uhle, M. E., Chin, Y.-P., Aiken, G., Mcknight, D. M. (1999) *Binding of polychlorinated biphenyls to aquatic humic substances: the role of substrate and sorbate properties on partitioning.* Environ. Sci. Technol. 33, 2715-2718.
124 Gombar, V. K., Enslein, K. (1996) *Assessment of n-Octanol/Water Partition Coefficient: When Is the Assessment Reliable?* J. Chem. Inf. Comput. Sci. 36, 1127-1134.
125 Karickhoff, S. W. (1984) *Organic pollutant sorption in aquatic systems.* J. Hyd. Eng. 110, 707-733.
126 Gauthier, T. D., Seltz, W. R., Grant, C. L. (1987) *Effects of Structural and Compositional Variations of Dissolved Organic Humic Materials on Pyrene K_{oc} Values.* Environ. Sci. Technol. 21, 243-248.
127 Di Toro, D. M. (1985) *A particle interaction model of reversible organic chemical sorption.* Chemosphere 14, 1503-1538.
128 Iglesias-Jiménez, E., Poveda, E., Sanches-Camazano, M. (1997) *Effect of the nature exogenous organic matter on pesticide sorption by the soil.* Arch.Environ.Contam.Toxicol. 33, 117-124.
129 Merkelbach, R. C. M., Leeuwangh, P., Deneer, J. (1993) *The role of hydrophobicity in the association of polychlorinated biphenyl congeners with humic acid.* Sci. Total Environ., 409-415.
130 Schlautman, M. A., Morgan, J. J. (1993) *Effects of aqueous chemistry on the binding of polycyclic aromatic hydrocarbons by dissolved humic materials.* Environ. Sci. Technol. 27, 961-969.
131 Broholm, M. M., Broholm, K., Arvin, E. (1999) *Sorption of heterocyclic compounds on natural clayey till.* J. Contam. Hydrol. 39, 183-200.
132 Banwart, W. L., Hassett, J. J., Wood, S.G., Means, J. C. (1982) *Sorption of nitrogen-heterocyclic compounds by soils and sediments.* Soil Sci. 133, 42-47.
133 Szabo, G., Bulman, R. A. (1994) *Comparison of adsorption coefficient (Koc) for soils and HPLC retention factors of aromatic hydrocarbons using a chemically immobilized humic acid column in RP-HPLC.* J. Liq. Chrom. 17, 2593-2604.
134 Chin, Y.-P., Aiken, G. R., Danielsen, K. M. (1997) *Binding of Pyrene to Aquatic and Commercial Humic Substances: The Role of Molecular Weight and Aromaticity.* Environ. Sci. Technol. 31, 1630-1635.
135 Torrents, A., Jayasundera, S., Schmidt, W. J. (1997) *Influence of the Polarity of Organic Matter on the Sorption of Acetamide Pesticides.* J. Agri. Food Chem 45, 3320-3325.
136 Hassett, J. J., Means, J. C., Banwart, W. L., Wood, S. G. (1980) *Sorption properties of sediments and energy related pollutants.* EPA. 600. Map 3-80-041.
137 Chiou, C. T., Peters, L. J., Freed, V. H. (1979) *A physical concept of soil-water equilibria for nonionic organic compounds.* Science 206, 831-832.

138 Wu, S.-C., Gschwend, P. M. (1986) *Sorption kinetics of hydrophobic organic compounds to naturel sediments and soils.* Environemtal Science and Technology 20, 717-725.
139 Wasik, S. P., Miller, M. M., Tewari, Y. B., May, W. E., Sonnefield, W. J., DeVoe, H., Zoller, W. H. (1983) *Determination of the vapor pressure, aqueous solubility, and octanol/water partition coefficient of hydrophobic substances by coupled generator column/liquid chromatographic methods. In: Residue Reviews.* Residues of Pesticides and other Contaminants in the Total Environment. Vol. 85. (Eds: Gunther,Francis A; Gunther,JD) Springer-Verlag, NY, 29-55.
140 Seip, H. M., Alstad, J., Carlberg, G. E., Martinsen, K., Skaane, R. (1986) *Measurement of mobility of organic compounds in soil.* Sci. Total Environ. 50, 87-101.
141 Maxin, C. R., Knabner-Kögel, I. (1995) *Partitioning of polyaromatic hydrocarbons (PAH)to water-soluble soil organic matter.* Eu. J. Soil Sci. 46, 193-204.
142 Suzuki, T., Kondo, H., Yaguchi, K., Maki, T., Suga, T. (1998) *Estimation of Leachability and Persistence of Pesticides at Golf Courses from Point-Source Monitoring and Model To Predict Pesticide Leaching to Groundwater*. Environ. Sci. Technol. 32, 920-929.
143 Zimmerman, L. R., Thurman, E. M., Bastian, K. C. (2000) *Detection of Persistent organic pollutants in the Mississippi Delta using semipermeable membrane devices.* Sci. Total Environ. 248, 169-179.
144 Jonker, M. T. O., Smedes, F. (2000) *Preferential sorption of planar contaminants in sediments from Lake Ketelmeer, the Netherlands.* Environ. Sci. Technol. 34, 1620-1626.
145 Helweg, C., Nielsen, T., Hansen, P. E. (1997) *Determination of octanol-water partition coefficients of polar polycyclic aromatic compounds (N-PAC) by high performance liquid chromatography.* Chemosphere 8, 1673-1684.
146 Kaiser, K., Haumaier, L., Zech, W. (2000) *The sorption of organic matter in soils as affected by the nature of soil carbon.* Soil Sci. 165, 305-313.
147 Tiller, C. L., Jones, K. D. (1997) *Effects of dissolved oxygen and light exposure on determination of Koc values for PAHs using fluorescence quenching.* Environ. Sci. Technol. 31, 424-429.
148 Axelman, J., Broman, D. N. (1997) *Field measurements of PCB partitioning between water and planktonic organisms: Influence of growth, particle size, and solute-solvent interactions.* Environ. Sci. Technol. 31, 665-669.
149 Perminova, I. V., Grechishcheva, N. Y., Petrosyan, V. S. (1999) *Relationships between structure and binding affinity of humic substances for polycyclic aromatic hydrocarbons: Relevance of molecular descriptors.* Environ. Sci. Technol. 33, 3781-3787.
150 Huang, W., Young, T. M., Schlautman, M. A., Yu, H., Weber Jr., W. J. (1997) *A distributed reactivity model for sorption by soils and sediments. 9. General isotherm nonlinearity and applicability of the dual reactive domain model.* Environ. Sci. Technol. 31, 1703-1710.
151 Young, T. M., Weber Jr., W. J. (1997) *A distributed reactivity model for sorption by soils and sediments. 7. Enthalpy and polarity effects on desorption under supercritical fluid conditions.* Environ. Sci. Technol. 31, 1692-1696.
152 de Voogt, P., Wegener, J. W. M., Klamer, J. C., van Zijl, G. A., Govers, H. (1988) *Prediction of environmental fate and effects of heteroatomic polycyclic aromatics by QSAR: The position of n-Octanol/water partition coefficients.* Biomed. Environ. Sci. 1, 194-209.
153 Tanford, C. (1980) *The hydrophobic effect: Formation of micelles and biological membranes.* Ed. Duke, J.B., A Wiley Interscience Publication, John Wiley & Sons, 1980, NY.
154 Nendza, M., Jückel, H., Müller, M. (1993) *Estimation of exposure and ecotoxicity related parameters by computer based structure-property and structure-activity relationships.* Toxicol. Environ. Chem. 40, 57-69.
155 OECD guideline for the testing of chemicals. *Water solubility.* Guideline no. 105.
156 Hansen, H. K., Rasmussen, P (1991) *Vapor-Liquid Equilibria by UNIFAC Group Contribution. 5 Revision and Extension.* Ing. Eng. Chem. Res. 30, pp. 2352-2356.

157 Chen, F., Holten-Andersen, J., Tyle, H. (1993) *New developments of the UNIFAC model for environmental application,* Chemosphere, 26, pp. 1325-1354.

Appendix A

Mixtures and Solutions

The molar fraction of the individual pollutants, x_i, is small with respect to the molar fraction of the bulk water, x_w, in natural environmental aqueous compartments. In such systems any organic pollutant will the minor component, i.e. $x_w >> x_i$. Therefore, with respect to the pollutant contamination, most aqueous compartments are properly described as dilute aqueous solutions.

Gibbs free energy, G, of any system in equilibrium is a state function defined as

$$G = H - TS \tag{A1}$$

where H is the enthalpy, T, the absolute temperature, and, S, the entropy of the system. For an equilibrium system of constant T and P, the Gibbs free energy is minimal. The Gibbs free energy of system containing j components is determined by the number of moles, n_i, and chemical potential, μ_i, of the individual components i through the expression

$$G = \sum_{i=1}^{j} \mu_i n_i \tag{A2}$$

and the differential form of G with respect to μ_i and n_i for a total of j components is then expressed as

$$dG = \sum_{i=1}^{j} \mu_i dn_i + \sum_{i=1}^{j} n_i d\mu_i \tag{A3}$$

For any spontaneous process the change in Gibbs free energy, at constant T and P, is negative, i.e. $dG_{T,P} < 0$, and described as function of T, P and n_i, the differential of G is given by

$$dG = VdP - SdT + \sum_{i=1}^{j} \mu_i dn_i \tag{A4}$$

Equation A4 is called the fundamental equation of thermodynamics, which at constant T and P reduces to

$$dG_{T,P} = \sum_{i=1}^{j} \mu_i dn_i \tag{A5}$$

For spontaneous processes, the change in Gibbs free energy is negative, i.e. the component i moves from high chemical potential to low chemical potential. The chemical potential of component i, in any given phase, is described through

$$\mu_i = \mu_i^{standard} + RT \ln a_i \tag{A6}$$

where the standard state of component *i*, $\mu_i^{standard}$, depends on the type of system, i.e. the relative amounts of components *i*, which is to be described.

Mixture description

In the description of a mixture, the standard state is defined as the pure component, μ_i^*. The chemical potential of component *i* in a mixture is defined as

$$\mu_i = \mu_i^* + RT \ln a_i \tag{A8}$$

where μ_i^*, is the standard chemical potential of the pure component. The activity is defined as

$$a_i = \gamma_i x_i \tag{A9}$$

where γ_i is the activity coefficient, and x_i the mol fraction of component *i.* The activity coefficients express the deviation from ideality. In the limiting case of Equation A9, all components obey Raoult´s law, and we define the activity coefficient by

$$\gamma_i \to 1 \quad \text{as} \quad x_i \to 1 \tag{A10}$$

So the activity for a component *i* of a mixture, as well as for the solvent component of a solution, can be quantified by replacing the activity with the mol fraction of *i*

$$\mu_i = \mu_i^* + RT \ln x_i \tag{A11}$$

As most environmental systems consist on one major component, e.g. an organic or aqueous phases, whereas the components of environmental concern, i.e. the pollutants, are hopefully present at dilute solution conditions. For such systems another description is useful, and this is *the solution description.*

Solution description

In the solution description the standard state for the solute is defined at infinite dilution, μ_i°. Infinite dilution is a hypothetical state, where each solute molecule is surrounded by pure solvent, e.g. water, and se nothing than but solvent molecules[110]. The chemical potential of the solvent is in this situation given by Equations A9 and A10. The chemical potential of a component of interests, namely the solute, is the standard chemical potential, μ_i°, defined at infinite dilution

$$\mu_i^\circ = \lim_{x \to 0} (\mu_i - RT \ln x_i) \tag{A12}$$

and the standard state for the dilute solute activity is therefore

$$\gamma_i \to 1 \quad \text{as} \quad x_i \to 0 \tag{A13}$$

In Equation A12 component *i* is an infinite dilute solute, i.e. the activity coefficient approaches unity as the mol fraction of solvent, e.g. water, approaches one. The solution description corresponds to the conditions of pollutants in most environmental compartments. For these systems, as expressed in Equations A12 and A13, the molar fraction of solute can be replaced by, e.g., molar concentrations by proper definition of $\mu_i^{\circ 110}$.

Appendix B

Multicomponent systems and equilibrium partitioning

In the literature there are often confusion concerning the choice of state of the standard chemical potential for components of mixtures and solutions. In the following text we will use the dilute solution standard state for the solute as defined by Equation A12 in Appendix A.

Natural, e.g. aqueous, media, are multicomponent systems, may include several phases and several solutes. Such systems could be, e.g., 1) solute in aqueous media partitioning to soil and dissolved organic matter, respectively or 2) solute in a simple aqueous media partitioning into an aquatic organism consisting of several phases. At equilibrium state the chemical potential of the partitioning solute will be equal in each phase within the aqueous media including the aqueous continuous phase. This implies that the equilibrium chemical potential of the

individual components, *i*, in a multicomponent system consisting of *J* phases is obeying the expression

$$\mu_i^I = \mu_i^{II} ..... = \mu_i^J \tag{B1}$$

If we have a system consisting of a single solute partitioned between two immiscible (or only partially miscible) phases, which constitutes a closed system, i.e. heat and work are exchanged with the surrounding, but the solute never leaves the liquid-liquid system. The two phases, within the closed system, are open systems and the solute can be transferred from a polar phase, *I*, to a non-polar phase, *II*. The partition process continues until G is minimal. If we assume that the two-phase equilibrium, i.e. the mutual saturation of phase *I* in phase *II* and phase *II* in phase *I*, do not change by the presence of the solute, then the equilibrium state for the solute is described as

$$\left(\frac{\partial G_I}{\partial n_i}\right)_{T,P,n_I,n_{II}} = \left(\frac{\partial G_{II}}{\partial n_i}\right)_{T,P,n_I,n_{II}} \tag{B2}$$

or simply

$$dG = (\mu_i^I - \mu_i^{II})dn_i = 0 \tag{B3}$$

If we assume that the conditions for the partitioning solute in every phase is a dilute solution state, then the chemical potential of the solute is given by

$$\mu_i = \mu_i^\circ + RT \ln C_i \gamma_i \tag{B4}$$

where C_i is the molar concentration of the solute in a given phase, the standard chemical potential being defined at the infinite dilute solution state (cf. Equations A12 and A13, Appendix A).

At equilibrium the chemical of the solute are equal in each phase (cf. Equation A1) and the standard Gibbs free energy for equilibrium partitioning derived as follows

$$\begin{aligned} &\mu_i^I = \mu_i^{II} \Rightarrow \\ &\mu_i^{\circ,I} + RT \ln \gamma_i^I C_i^I = \mu_i^{\circ,II} + RT \ln \gamma_i^{II} C_i^{II} \end{aligned} \quad \text{(B5)}$$

by assuming ideal dilute solution state in both phases, the activity coefficient γ_i is unity and the equilibrium partitioning of the solute from the polar phase the non-polar phase.

$$\begin{aligned} &\mu_i^{\circ,II} - \mu_i^{\circ,I} = -\Delta\mu_i^\circ = RT \ln \frac{C_i^{II}}{C_i^I} \Rightarrow \\ &\Delta\mu_i^\circ = \Delta G_i^\circ = -RT \ln K_{eq} \end{aligned} \quad \text{(B6)}$$

In the case of the polar phase, *I*, being water, and the non-polar phase, *II*, octanol Equation B6 can be expressed as

$$\Delta G^\circ = -RT \ln \frac{C_i^{octanol}}{C_i^{water}} = -RT \ln K_{ow} \quad \text{(B7)}$$

This is how the equilibrium partitioning coefficient is defined for ideal dilute two-phase systems for which K_{ow} can be quantified as the molar concentration ratio of component *i* in two slightly miscible or immiscible phases.

Equilibrium partitioning description according to, e.g., Schwarzenbach et al.[104] and Mackay et al.[91]

In works of Schwarzenbach and Mackay et al. the standard state for the solute is the pure organic compound, and in this case the activity coefficient of a hydrophobic compound in dilute aqueous solution is high. In this case the activity can not be replaced by molar concentrations, and the equilibrium partitioning, in Equation B5, is therefore expressed as

$$\begin{aligned} &\mu_i^{*,II} - \mu_i^{*,I} = -\Delta\mu_i^* = RT \ln \frac{\gamma_i^{II} C_i^{II}}{\gamma_i^I C_i^I} \Rightarrow \\ &\Delta\mu_i^* = \Delta G_i^* = 0 = -RT \ln K_{ow} \end{aligned} \quad \text{(B8)}$$

As the standard chemical potential is the pure compound for both phases, i.e. $\Delta\mu_i^* = \mu_i^{*,II} - \mu_i^{*,I} = 0$, and the lower expression in B8 seems nonsense, i.e. indicates that K_{ow} equals one for all components *i*. In Schwarzenbach et al. an apparent partition constant is therefore derived through the expression

$$\mu_i^{II} - \mu_i^I = \Delta G = RT \ln \gamma_i^I x_i^I - RT \ln \gamma_i^{II} x_i^{II} \quad \text{(B9)}$$

At non-equilibrium a net flow of component *i* from the phase of highest chemical potential to the phase of lowest chemical potential will occur until equilibrium where $\mu_i^I = \mu_i^{II}$. The right term of Equation B9 is rearranged to

$$RT \ln \frac{x_i^{II}}{x_i^I} = -\left(RT \ln \gamma_i^{II} - RT \ln \gamma_i^I\right) \Rightarrow$$

$$K'_{eq} = \frac{x_i^{II}}{x_i^I} = \exp\left[\frac{-\Delta g^e}{RT}\right] \qquad \text{(B10)}$$

the apparent partition constant, $K'_{eq,}$ based on the mixture description in Appendix A, quantifies the relative abundance of component *i* in the two phases at equilibrium. The activity coefficient quantifies the additional free energy in a non-ideal mixture caused by the presence of the component *i*, i.e. the term $RT \ln \gamma_i$, which is the partial molar excess free energy, g_i^e of component *i*[104]. For non-polar or hydrophobic compounds, the activity coefficient is assumed close to one for the non-polar phase and as such the partitioning is in most cases mainly determined by the non-ideality due to dissimilar solute:solvent interactions in the aqueous bulk phase. The term apparent partitioning coefficient is to be found in the above expression, as the contribution from the non-ideal left-hand term of Equation B10 is an additional parameter (cf. Chapter 3, Equation 3.1) in experimental measurements of apparent partitioning coefficients. In the description of Schwarzenbach et al., the partitioning coefficient is the redefined by replacing the molar fractions by molar concentration, as this is the most common way of expressing the apparent partitioning coefficient.

Furthermore, by use of Amagat´s law, $V_{mix} = \sum_i x_i V_i$, stating that the molar volume of the mixture is the sum of the product of molar fraction and partial molar volume for each component *i*, Equation B10 is expressed

$$\ln K_{eq} = \ln \frac{C_i^{II}}{C_i^I} = \ln \frac{x_i^{II} / V_{mix,II}}{x_i^I / V_{mix,I}} = \ln \frac{\gamma_i^I / V_{mix,II}}{\gamma_i^{II} / V_{mix,I}} \qquad \text{(B11)}$$

The molar volume of the octanol, $V_{mix,II}$ and aqueous phase, $V_{mix,I}$, as expressed in Equation B11, is approximately equal to the molar volume of the solvent, i.e. octanol, V_{II}, and water, V_I, at low aqueous solute concentrations[91,104]. Furthermore, it is assumed that the compounds mix with no or insignificant change in volume. For the organic phase due similar to specific molecule:molecule attractions or repulsions, and for the aqueous phase due to the insignificant solubility of hydrophobic compounds.

In Equation B11 the molar fractions is replaced by molar concentration which is confusing as this is only valid in the dilute solution description (cf. Appendix A, Equations A12 and A13). After this the molar concentration is set equal to the molar fraction divided by the molar volume of the solvents. This way of interchanging between molar fractions and molar concentrations are conflicting with respect to the definition of the standard chemical potential of mixtures versus dilute solutions.

The Mackay[91] description of the partition coefficient between two immiscible phases

In the description of Mackay et al., the starting point is

$$x_i^I \gamma_i^I = x_i^{II} \gamma_i^{II} \tag{B12}$$

as derived from Equation B8, and in accordance with the mixture description (cf. Appendix A, Equations A8 to A10). They use the same approximation, that the mol fraction of the solute in each phase *I* and *II* may be expressed as a product of the molar concentration of the solute, C_i, and the molar volume of the solvent phase, V , and Equation B12 can be written as

$$C_i^{II} V_{II} \ \gamma_i^{II} = C_i^I V_I \ \gamma_i^I \tag{B13}$$

The partition coefficient between the two phases can now be quantified as

$$K_{eq} = \frac{C_i^{II}}{C_i^I} = \frac{x_i^{II} / V_{II}}{x_i^I / V_I} = \frac{V_I \ \gamma_i^I}{V_{II} \ \gamma_i^{II}} \tag{B14}$$

Equation B14 is analogous to Equation B11, and if the two partially miscible liquids were water and octanol the equation would look like

$$K_{ow} = \frac{C_i^o}{C_i^w} = \frac{x_i^o / V_o}{x_i^w / V_w} = \frac{V_w \gamma_i^w}{V_o \gamma_i^o} \tag{B15}$$

If the solute was partitioning between an aqueous phase and different phases of an organism (fish), then the amount of the partitioning solute, *i,* in the organism could be quantified by:

$$C_i^{fish} = \sum v_m C_i^m = \sum \left(\frac{v_m}{V_m \gamma_i^m} \right) \tag{B16}$$

where v_m is the volume fraction of the different organic phases, *m*, within the organism and V_m the molar volume of the different liquid phases within the organism. At equilibrium, in analogues to Equation B15, the partitioning of a component *i* between an aqueous phase and a fish is expressed

$$BCF = \frac{C_{fish}}{C_w} = V_w \gamma_i^w \sum \left(\frac{v_m}{V_m \gamma_i^m} \right) \tag{B17}$$

Based on this description the equilibrium property-property relationship between log*BCF* and logK_{ow} may work as long as 1) the biological matrix population, i.e. fish, have similar phase composition (v_n), and 2) as long at the ratio between significant phase activity coefficients, γ_i^{liq} , and the activity coefficient in octanol, γ_i^o , is constant!

These restrictions are critical with respect to log*BCF*-logK_{ow} relations for polar compounds. For non-polar compounds which have high affinity for partitioning to the lipid phase, however, the activity coefficient in the lipid phases are near 1, and due to the non-specific molecular:molecular

interactions the system is less sensitive to small interspecies changes in composition. Based on the mixture description the correlation between logBCF and logK_{ow} is in the latter situation controlled mainly by the activity coefficient of the solute in the aqueous phase. This may explain the linear relation between K_{ow} and BCF in cases where the relation is linear.

The complexity of the above description seems inappropriate, when compared to the uncertainty in measurements. The use of apparent partitioning coefficients, $K_{eq} = \frac{C_i^{II}}{C_i^{I}}$, is most critical, and contribute to the high variations in partitioning values found in the literature for other two-phase systems as well. This due to the acceptance of aqueous concentration levels, at which solute:solute interaction determines the measured partitioning coefficients, i.e. is significantly dependent on the activity coefficient.

Appendix C

UNIFAC calculated unimeric solubilities of the phthalates

UNIFAC (UNIversal Functional group Activity Coefficient) is a thermodynamic model that can be used to predict activity coefficients of non-electrolytes in a medium. The method is based on the mixture description, i.e. the standard chemical potential of all components is the pure component state, as described in Appendices A and B. The model is based on a group-contribution concept, i.e. mixtures are considered to consist, not of molecules, but as a mixture of fragments of molecules. By quantifying the molecules through common fragment properties, the model is able to predict activity coefficients for a vast number of molecules

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \tag{C1}$$

where γ_i is the activity coefficient for the component *i*. The combinatorial term, γ_i^C, depends on the volume and surface area of each molecule to represent the entropy contribution, and the residual term, γ_i^R is mainly controlled by the energetic interaction parameter among different groups[v,156]. The accuracy of group contribution methods remains controversial because of their interaction parameters. These are derived by calibrating the model against experimental vapour-liquid equilibrium data of a so-called consistency data set, where available data obeys the Gibbs-Duhem equation[157]. In the work of Chen et al.[157] it is stated that predictions may be limited to the infinite concentration range, as the accuracy and errors of predictions increases significantly for hydrophobic compounds with logK_{ow} values above 6. In the case of hydrophobic chemicals the solubility is calculated as the inverse of the infinite dilute solution activity coefficient, as described in a number of papers[as cited in v]. The model were used, in paper v, for calculating the aqueous solubility and octanol-water partitioning of phthalates for which data are given in Table C1.

***Table C1**. UNIFAC derived aqueous solubilities, in molar concentrations, and octanol-water partitioning coefficients and error factors relative to the respective experimental data for selected phthalates[v].*

Acronym	cas.no.	$^1\gamma^w$	$^1\gamma^o$	logK_{ow}	logC_w^{sat}	^{2}EF(C_w^{sat})	^{2}EF(K_{ow})
DMP	131-11-3	2.3E+03	4.91	1.85	-1.60	0.05	0.02
DEP	84-66-2	1.5E+04	4.19	2.72	-2.42	-0.06	-0.04
DAP	131-17-9	1.5E+05	5.79	3.61	-3.45	-0.02	0.63
DiP(3)P	605-45-8	1.0E+05	3.87	3.59	-3.26	-0.38	0.76
DnP(3)P	131-16-8	9.9E+04	3.86	3.59	-3.25	0.15	-0.05
DiBP	84-69-5	7.2E+05	3.74	4.46	-4.11	0.013	-0.02
DnBP	84-74-2	7.1E+05	3.73	4.46	-4.11	0.29	0.09
DnP(5)P	131-18-0	5.2E+06	3.71	5.33	-4.98	0.86	0.48
DnH(6)P	84-75-3	3.9E+07	3.78	6.20	-5.85	0.29	-0.1
DEHP	117-81-7	2.4E+09	4.12	7.94	-7.63	-1.26	0.88
DnOP	117-84-0	2.3E+09	4.11	7.94	-7.63	-1.49	0.95
DnDP	84-77-5	1.4E+11	4.63	9.67	-9.42	-3.24	
DnTP	119-06-2	7.3E+13	5.76	12.28	-12.12	-5.93	3.88
BzBP	85-68-7	1.4E+07	9.79	5.33	-5.39	-0.21	0.74
DCHP	84-61-7	2.2E+07	5.23	5.80	-5.60	-2.97	0.9
DboEP	117-83-9	2.9E+06	0.71	5.79	-4.71	-0.66	

[*] γ^w is the activity coefficient of the compounds in the aqueous phase at infinite dilution, γ^o is the activity coefficient of the compound in the octanol phase.

$^2EF(C_w^{sat})=log(C_w^{sat,pred.}/C_w^{sat,exp.})$ and $EF(K_{ow})=log(K_{ow}^{pred.}/K_{ow}^{exp.})$

The UNIFAC-predicted versus average measured solubility and octanol-water partitioning coefficients are shown in Figure C1 and C2.

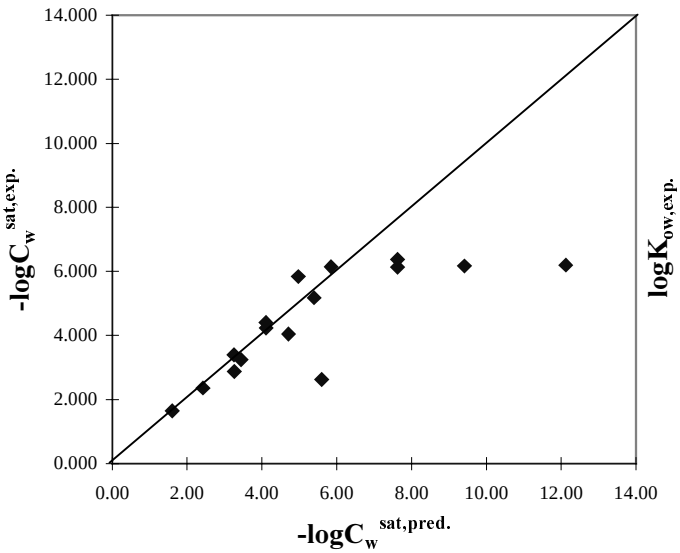


***Figure C1** Experimental solubilities vs.UNIFAC derived values.*

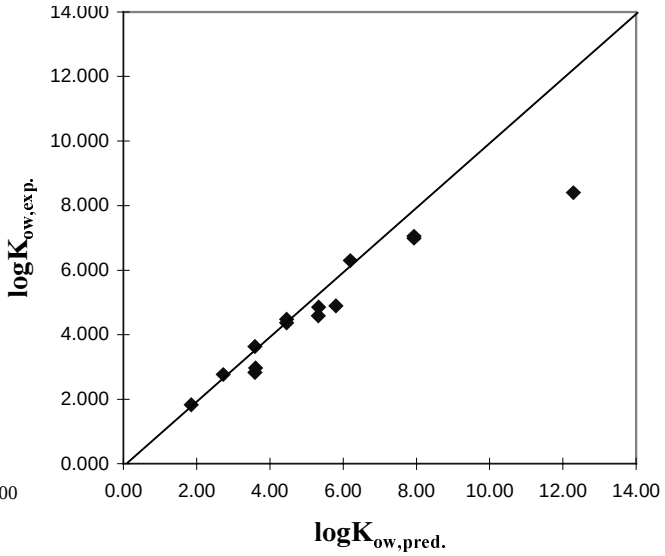


***Figure C2** Experimental octanol-water partitioning coefficients vs. UNIFAC derived values.*

As illustrated from the Figures, the average experimental aqueous solubilities are higher than UNIFAC predicted values at UNIFAC predicted solubilities below –6 in log units. Similar there is a tendency for the average experimental octanol-water partitioning coefficients to be lower than the UNIFAC predicted values for phthalates at logK_{ow} values above 6[v].

The observed non-linear increase in solubility at increasing numbers of carbon atoms in the alkyl chain has been explained by the folding of the hydrophobic aliphatic carbon chain, thereby reducing the solvent accessible surface area of the molecule. The effect of a folding of the alkyl chain is

however to small to account for the observed non-linearity[II], and the reason being inappropriate method for measuring the unimeric solubility of organic compounds forming micro-emulsions upon exceeding saturation of the aqueous bulk phase[I, II, v] (cf. Appendices D and E). Furthermore, the unimeric solubilities measured by the surface tension method are in agreement with the UNIFAC-predicted values at 25 ° C. At temperatures above or below 25 ° C the UNIFAC method seems to give wrong predictions.

Appendix D

Short description and comments on the OECD-guideline 105[155]

The Column elution method

A preliminary test consist of taking 0.1 g of the pure substance adding additional amounts of water to the sample device, shaking the mixture for 10 minutes. The procedure is continued until no undissolved parts of sample can be visualised, i.e., discriminating between visible heterogeneous and homogeneous mixtures.

In the column elution methods, an inert support material is coated with test substance by mixing support material and test substance dissolved in a volatile solvent, and evaporating the dispersion to dryness. After this the support material is soaked in water for two hours, or equilibrated by pouring it into the water-filled microcolumn and resting for two hours. After this the flow is started, and five bed volumes are discarded to remove impurities prior to recirculation of the system. The result of the solubility measurement is an apparent solubility determined by

$$C_{i,aq}^{app} = \frac{\sum_{i=1}^{n} C_{i,aq}}{n}, \quad RSD = \frac{s}{C_{i,aq}^{app}} \cdot 100 \leq 30\% \qquad \text{(D1)}$$

$C_{i,aq}^{app}$ is the so-called apparent solubility, $C_{i,aq}$ is repetitive measurements of the aqueous solubility, n is the number of measurements, s is the standard deviation on repetitive measurements and RSD is the relative standard deviation. The relative standard deviation on five repetitive measurements should be less than 30 %. It is suggested that a concentration versus time curve be used to show that equilibrium is reached.

The dissolution process from the solid inert phase to the aqueous phase is described through the equilibrium:

$$A_{adsorb} \rightleftharpoons A_{aq} \qquad \text{(D2)}$$

The test substance is partitioning from the solid to the aqueous bulk phase, and the solubility is defined at equilibrium. If the measured aqueous concentration increases after halving the flow rate, then the test should be continued at the low flow rate, i.e. a test for partitioning by limiting kinetics, and results according to Equation D1 obtained.

The result is highly dependent on the amounts of test substance used for coating the suggested 600 mg support material, as the equilibrium assumption do not relate to the true or unimeric solubility of the partitioning substance. There is one very important step in the guideline for the column elution method, and that is examination of the Tyndall effect. The Tyndall effect is measured to detect the presence of dispersed particles or emulsions within the bulk water phase. However, micro-emulsions are transparent and too small to be detected by spreading of light. Therefore the use of

the Tyndall effect to secure that the measured aqueous concentration corresponds to a true solution, is only useful when coarse dispersions are present within the aqueous bulk phase.

The guideline mentions problems in causes when test substance is deposit as oil, and states that these problems should be examined and reported. The formation of micro-emulsion, which form transparent systems with respect to the Tyndall effect[III,107,110], may therefore be an explanation for the missing quantitative prove of in the literature.

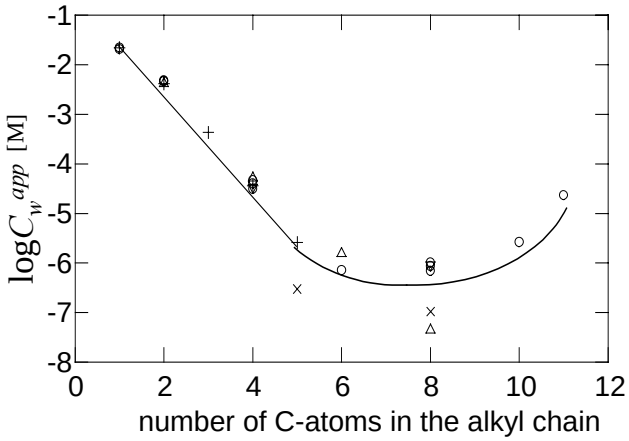


Figure D1 *Experimental solubilities measured by surface tension measurements (Δ)[II], method including turbidity measurements (∇), methods including a centrifugation step (o), solubility measured according to the OECD no.105 followed by extraction and quantification by GC (x), and solubility measured according to the OECD no.105 followed by quantification through measuring the UV absorbance (+). References cited in paper iii.*

The Figure clearly shows the inability of phase separation by use of centrifugation by the data points marked by a circle. The surface tension data as well as the data measured by the slow-stirring method[37,36,35] (not shown in Figure D1) is illustrates the need for methods which are able to discriminate between true solutions and mixtures defined as emulsions (or dispersions). As indicated by the surface tension measurements, as well as the UNIFAC-predicted solubilities, the unimeric, or true, solubility decreases lineary with an increase in the number of C-atoms in the alkyl chain of the phthalates esters.

Shake-Flask methods

Based on the preliminary measured solubility three vessels containing the test substance solution, at concentrations according to solubility range and analytical method, are prepared. The samples are stirred at 30 °C and after one, two, and three days, respectively, one samples at a time is equilibrated at 20±0.5 °C for 24 hours with occasionally shaking. The contents of the sample vessels are centrifuged at the test temperature, and the concentration of test substance in the clear aqueous phase is determined by a suitable analytical method. For the phthalates the micro-emulsion are not detected by UV-measurement as illustrated in Figure D1. Furthermore due to the similar densities of phthalates and water phase separation by centrifugation is virtually impossible. This is illustrated in Figure D2.

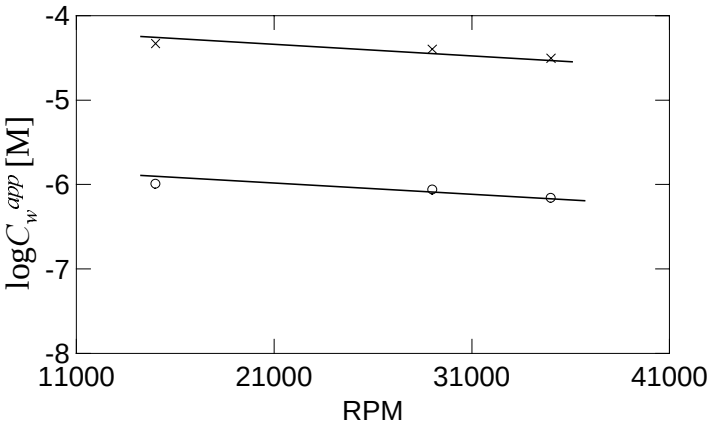


Figure D2 *The solubility of DnBP (x) and DEHP (o) as function of RPM used in the centrifugation step, respectively.*

Figure D1 and D2 shows the limitations in the OECD guideline for measuring the true solubilities of substances of colloidal nature above the unimeric solubility. Both of the methods show inability to discriminate between the borderline of a mixture and a true solution. As such, the result obtained from these methods relies on the starting concentration used in individual experimental systems, as well as on the colloidal nature of the mixture at concentrations that exceeds the unimeric solubility (cf. Appendix E).

Appendix E

Stability and process of the formation of emulsions

If sufficient mechanical energy is put into a two-phase system a microemulsion may be formed. Stability of the emulsion results if the reverse process, coalescence and phase separation is sufficiently slow[107,110]. The phthalates are surfactants, but they do not form low interfacial stable and ordered micellar structures within the bulk water phase. However they do form relatively stable microemulsions upon homogenisation.

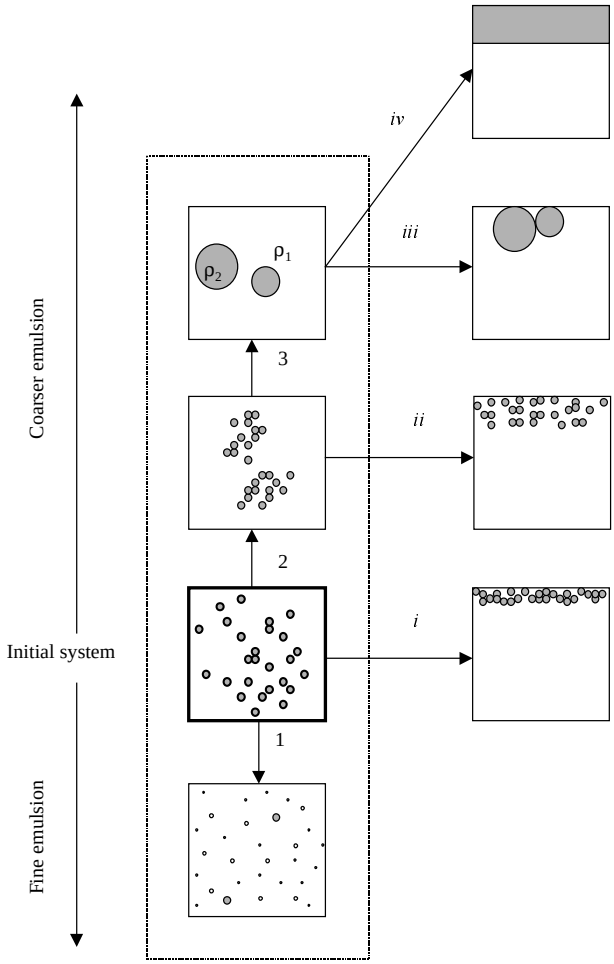


Figure E1 *Different stages and instabilities in emulsions. Reproduced from J.R. Hunter, 1995. An initial emulsion system is formed by thoroughly mixing of pure and liquid phthalate with water. The degree of disruption, illustrated by process 1) depends on the amount of mechanical energy put into the system. The process of aggregation, i.e. flocculation, is illustrated in process 2). Through flocculation small particles clump together, but do not fuse into a new particle. In process 3), coalescence, two or more particles fuse together to form a single larger particle. For the phthalates no macroscopic phase separation occurs, as only the stages included by the dashed box are relevant for the phthalates. In time scales of months, eventually macroscopic phase separation may occur.*

By mixing of phthalate and water an initial system such as illustrated in Figure E1 is formed. By mechanical mixing for 24 hours a disruption of the coarse emulsion system is processed and the

formation of finer dispersed microdroplets takes form[II]. The stability of the system is due to the similar densities of phthalates and water, i.e. the microdroplets of the dispersed phase are neutrally buoyant, F_b, and gravitional effects, F_g, are thereby insignificant. The netto force on a particle is described through

$$F_{net} = F_g - F_b = V_p(\rho_2 - \rho_1)g \qquad \text{(E1)}$$

where F_g is the gravitional force exerted on the particles, and F_b, is the buoyant force, i.e. force exerted by the continuous liquid on a particle, which equals the weight of the fluid displaced by the particle times acceleration of gravity, g. In the case of the phthalates $\rho_2 \approx \rho_1$, and thereby the net force on the microdroplets is insignificant, i.e. close to zero.

In Figure E1, the density ρ_2 of the particle is less than the density ρ_1 of the fluid, and the particles move upward (creaming). In the reverse case, $\rho_2 > \rho_1$, the particle will experience a net downward force and sedimentation of the particle will occur. The main difference between phthalates and other simple liquids are therefore microscopic phase separation instead of macroscopic phase separation.

By flocculation the particles remain their identity, but looses their kinetic independence, i.e. moves as a single unit. By this process there is only a maximal reduction in the total solvent assessable surface area. The process of coalescence, however, is driven by desired a reduction in the total surface area in contact with the medium.

The inherent properties of the clusters of molecular units, i.e. flocs and/or coarser microemulsion units are not described at a molecular level, and with respect to QSARs the behaviour of single molecules is not comparable to the behaviour microemulsions. The minimum requirement of QSARs, which are based on quantification of activities at a molecular level, is in contrast to the above a continuous and true solution.

Appendix F

Quality of solubility data

Phthalates and PCBs

Estimating the solubility of organic compounds in aqueous solutions at the point of saturation of the aqueous bulk phase should be an easy task compared to the spectrum of exposure endpoints as well as biomarker responses measured and modelled through the use of QSARs. However high variabilities in measured aqueous solubilities within all kinds of chemicals classes are a not unusual[34,38,iii]. Especially the conventional shake-flask methods seem to be associated with high variabilities (cf. Appendix D).

For the PCBs and the phthalates there is a relatively simple relation between variances on reported aqueous solubilities and the hydrophobicity of the solute as illustrated in Figure F1.

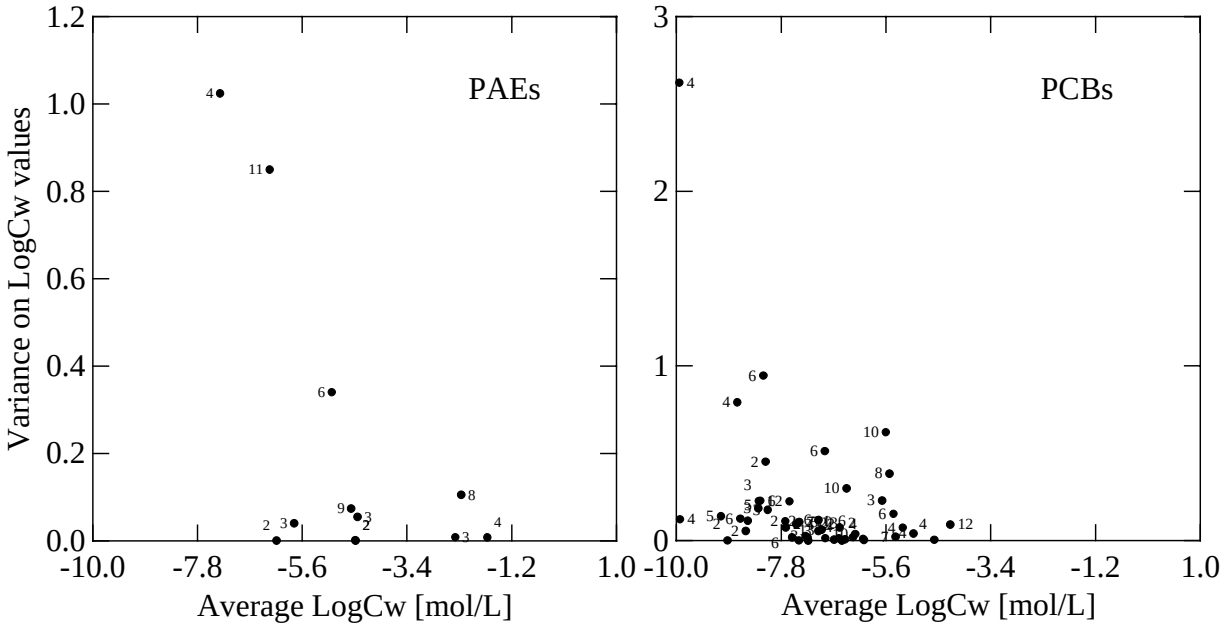


Figure F1 *Variances on reported solubility measurements for the a) phthalates and b)PCBs. Numbers at each point represent the number of experimental data used for calculating the variance between measurements.*

The variances illustrated in Figure F1, can be specified and analysed even further to show that the systematic increase in standard deviations between measurements are increasing significant, increasing in a systematic manner that is most pronounced for the shake-flask method. When dealing with reported experimental data, or data from databases data, ANOVAs can be used for evaluating data as well as the nature of variances. By testing if the variances between methods for each compound differs significantly and eliminating methods for which the variance is significantly higher than the remaining methods.

Through the elimination of single measurement outliers and methods outliers, variance homogeneity throughout the solubility range should prevail before using the data for model

calibration[x]. For the PCBs as well as the phthalates the relationship between molecular surface area and the solubility becomes linear after elimination of outliers and methods associated with high standard deviations on measurements (cf. Appendix C and D) [II, III]. For the phthalates the ANOVA and elimination of outlier data did not result in variance homogeneity between laboratories and/or methods and compounds[v]. This due to the sparse and low quality nature of data on the phthalates. The tendency for proportional systematic error in PAE solubility data results in a non-linear relationship between solubility and hydrophobicity of the phthalates.

For both classes of chemicals the solubility and partitioning seems to be driven mainly by hydrophobic effects[II] by these simple relations. The apparent solubility of the phthalates increases more than proportional with the molecular surface area of the solute[I,v]. The reason for this is to be found in the inherent properties of these compounds with respect to the formation of clusters within the aqueous bulk phase.

For both the phthalates (PAE) and the polychlorinated biphenyls (PCBs) a significant increase in variances on the measured solubilities is observed at increasing hydrophobicity of the molecular structure. An investigation of biases in variances between experimental methods behind the individual results, combined with experimental studies of the unimeric solubility of the phthalates, revealed that the a systematic increase in standard deviation that due to increasing systematic overestimating of the solubility by the shake-flask method. If enough data are available a between methods variance evaluation should be performed as described in Chapter 3[III].

PAHs/N,S,O-PAHs and Pesticides

As shown in Figure F2, the patterns in variances on the measured solubilities for pesticides[xvii] and the PAHs[xvii] are opposite to the PAEs[iii] and PCBs[III], i.e., is increasing with increasing solubility. Furthermore the variances on solubility measurements for the pesticides is significant larger than for the remaining classes of chemicals.

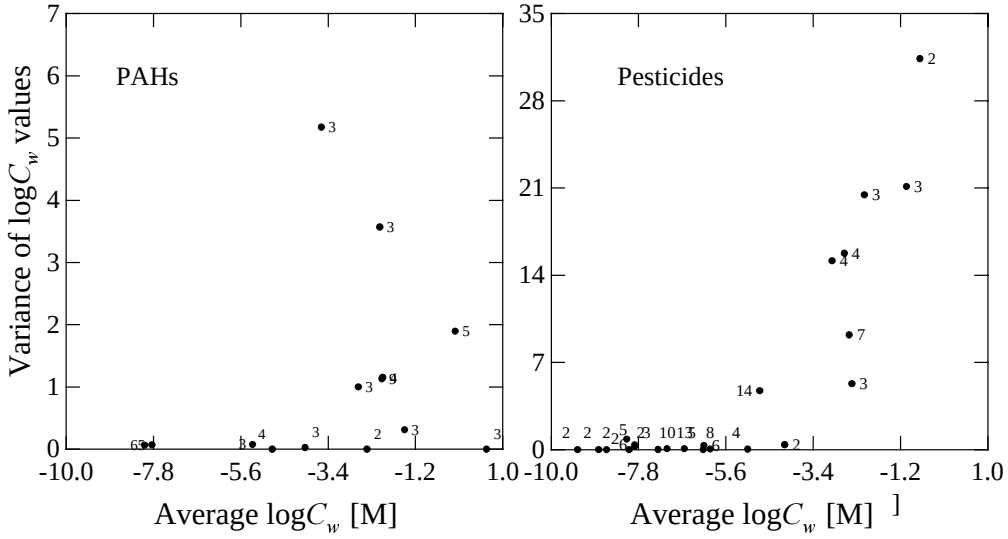


Figure F2 *Variances on reported solubility measurements for the PAHs/NSO-PAHs[xvii] and b) Pesticides[xvii]. Numbers at each point represent the number of experimental data used for calculating the variance between measurements.*

Most of the high molecular weight PAHs are solid molecular crystals, while the PAHs in high solubility range, with high variance on solubility measurements are solid leaflets or liquids. The insignificant variance on solubility measurements of the high molecular weight PAHs (Figure F2) are partly due to the high melting points, i.e. the formation of solids through strong van der Waals intermolecular forces by increasing number of conjungated planar ring structures. A more reasonable explanation may be the volatility of the low molecular weight PAHs.

For the pesticides the increasing variability in solubility measurements in the high solubility range, may be due to additional factors such as weak acid-base properties (pH-sensitivity) as well as strong solute-solute interactions. By screening the raw data including background data, i.e. pH and temperature, the significance of these parameters in explaining the variation in the individual solubilities was investigated and showed significant loadings of both pH and temperature[xvii].

Appendix G

Chemical property space of PAH and N,S,O-PAHs

The compounds included in the study of the sorption of PAH and N, S, O-PAHs to organic matter of different origin (cf. Chapter 5) are given in Table G1.

Table G1 *PAHs included in the case study in Chapter 5 on sorption of PAHs to organic originating from matrices denoted sediments and humic acids.*

name	id	cas.no.	type
Naphthalene	1	91-20-3	C
1-methylnaphthalene	2	90-12-0	C
2-methylnaphthalene	3	91-57-6	C
1-ethylnaphthalene	4	1127-76-0	C
1,5-dimethylnaphthalene	5	571-61-9	C
2,3-dimethylnaphthalene	6	581-40-8	C
2,6-dimethylnaphthalene	7	581-42-0	C
2,3,5-trimethylnaphthalene	8	2245-38-7	C
2,3,6-trimethylnaphthalene	9	829-26-5	C
trimethylnaphthalene	10	28652-77-9	C
1-naphthalenecarbonitrile	11	86-53-3	C
Acenaphthylene	12	208-96-8	C
Acenaphthene	13	83-32-9	C
Fluorene	14	86-73-7	C
1-methyl-9H-fluorene	15	1730-37-6	C
Benzo[a]fluorene A	16	238-84-6	C
Benzo[a]fluorene B	17	30777-18-5	C
Benzo[b]fluorene	18	243-17-4	C
Phenanthrene	19	85-01-8	C
Methylphenanthrene A	20	31711-53-2	C
2-methyl-phenanthrene	21	2531-84-2	C
3-methyl-phenanthrene	22	832-71-3	C
dimethyl phenanthrene	23	29062-98-4	C
9-cyanophenanthrene	24	2510-55-6	N
Anthracene	25	120-12-7	C
2-Methylanthracene	26	613-12-7	C
Benzo(a)anthracene	27	56-55-3	C
Naphthacene	28	92-24-0	C
Dibenz[a,h]anthracene	29	53-70-3	C
Fluoranthene	30	206-44-0	C
Benzo(b)fluoranthene	31	205-99-2	C
Benzo[j]fluoranthene	32	205-82-3	C
Benzo(k)fluoranthene	33	207-08-9	C
pyrene	34	129-00-0	C
1-methylpyrene	35	2381-21-7	C
benzo(a)pyrene	36	50-32-8	C
Benzo(e)pyrene	37	192-97-2	C
dibenzo(a,e)pyrene	38	192-65-4	C
dibenzo(a,i)pyrene	39	189-55-9	C

Indeno(1,2,3-cd)pyrene	40	193-39-5	C
Chrysene	41	218-01-9	C
Perylene	42	198-55-0	C
Benzo(ghi)perylene	43	191-24-2	C
Acridine	44	260-94-6	N
Benz[a]acridine	45	225-11-6	N
Benz[c]acridine	46	225-51-4	N
Dibenz[a,h]acridine	47	226-36-8	N
Quinolin	48	91-22-5	N
Isoquinolin	49	119-65-3	N
Methylquinoline	50	27601-00-9	N
4-methylquinoline	51	491-35-0	N
2-(1H)-quinoline	52	59-31-4	NOH
8-Quinolinol	53	148-24-3	NOH
Phenanthridine	54	229-87-8	N
Benzo[f]quinoline	55	85-02-9	N
Benzo[h]quinoline	56	230-27-3	N
Indole	57	120-72-9	N
Coronene	58	191-07-1	C
9(10H)-Acridone	59	578-95-0	NO
Carbazole	60	86-74-8	N
Naphthylamine	61	25168-10-9	N
Dibenzothiophene	62	132-65-0	S
Triphenylene	63	217-59-4	C
1-Indanone	64	83-33-0	O
2,3-Benzofuran	65	271-89-6	O
Dibenzofuran	66	132-64-9	O
Methylphenanthrene B	67	28652-81-5	C
6,7-dimethylquinoline	70	m	N
picene	71	213-46-7	C
quinolizine	72	m	N
5-Methylindole	73	614-96-0	N
1-methylindole	74	603-76-9	N
7-methylindole	75	933-67-5	N
4-methylindole	76	16096-32-5	N
6-methylindole	77	m	N
2-methylisoindole	78	33804-84-1	N
2-methylindole	79	95-20-5	N
3-methyl-1H-indole	80	83-34-1	N
6,7-Benzoquinoline	81	260-36-6	N
[1,1´-Biphenyl]-2-carbonitrile	82	24973-49-7	N
5,6-Benzoisoquinoline	83	229-67-4	N
m-cyanobiphenyl	84	m	N
benzoquinoline	85	39327-16-7	N
Phenyl benzonitrile	86	28804-96-8	N
1-Methylphenanthrene	87	832-69-9	C
Dibenz[a,c]anthracene	88	215-58-7	C
Dibenz[a,j]anthracene	89	224-41-9	C
2-ethylnaphthalene	91	939-27-5	C
Biphenyl	92	92-52-4	C
2,2´biquinoline	93	119-91-5	N
2,6-Dimethylquinoline	94	877-43-0	N
Quinaldine	95	91-63-4	N
6,8-Dimethylquinoline	96	2436-93-3	N
1-Acenaphenone	97	2235-15-6	O
13H-dibenzo[a,i]carbazole	98	239-64-5	N
Dibenz[a,j]acridine(1.2.7.8-dibenzacridine)	99	244-42-0	N
2,6-diphenylpyridine	102	3558-69-8	N

2,3-Benzofuran	105	271-89-6	O
4-hydroxyquinoline	106	3558-69-8	NOH
2-hydroxyquinoline	107	59-31-4	NOH
2-hydroxy-4-methylquinoline	108	607-66-9	NOH
1-Nitropyrene	110	5522-43-0	NOO
Benzothiophene	111	11095-43-5	S
6-Methylquinoline	113	91-62-3	N
Benzo[b]naphtho[2,3-d]furan	114	243-42-5	O
Benzo[b]naphtho[2,1-d]furan	115	239-30-5	O
2,3-5,6-dibenzoxalene	116	243-24-3	O
Benzo[b]naphtho[2,3-d]thiophen	117	243-46-9	S
Acenaphtho[1,2-b]pyridine	119	206-49-5	N
Indeno[1,2,3-ij]isoquinoline	120	206-56-4	N
Indeno[1,2,3-de]isoquinoline	121	7148-92-7	N
4-Azapyrene	122	m	N
11H-Indeno[1,2-b]quinoline	123	243-51-6	N
1-aminopyrene	124	1606-67-3	N
3-fluoranthenamine	125	2693-46-1	N
2-aminopyrene	126	1732-23-6	N
1,2-benzocarbazole	127	239-01-0	N
2,3-benzocarbazole	128	243-28-7	N
3,4-benzocarbazole	129	34777-33-8	N
4-phenanthreneacetonitril	130	74676-81-6	N
Diphenyl thiophene	131	1445-78-9	S
2,4-Diphenylthiophene	132	3328-86-7	S
4-methyl-dibenzothiophene	133	7372-88-5	S
Thioxanthene	134	261-31-4	S

The last column is denoted the type of hetero atom included in the ring structure or as substituent (amino-, cyano-, and hydroxy substituted PAHs are included)

Below an initial screening of the inherent structural and electronic properties of the compounds given in Table G1 is performed based on the non-empirical quantum-chemical descriptors.

The chemical property space of the PAHs/N, S, O-PAHs

The molecular structures was optimised by the unrestricted Hartree-Fock method[120], and both the alfa and beta orbital energies was included as the convergence criteria of 0.001 was larger than the differences between orbital energies calculated by the RHF-method[120].

First the non-significant, noisy descriptors are excluded. These are: The heat of formation, dH_f, the hydration energy, E_{hyd}, and the dipole moment as well as structures exhibiting high leverage in the model plane spanned by the principal components *PC1* and *PC2*. The original variable space is projected into a 3-dimensional space spanned by *PC1*, that explain 58 % of the variance in X, *PC2*, explaining 29 % and PC3 explaining 9 %.

Each original variable has a loading on each PC, which reflects 1) how much that variable contributes to the explained variance of the PC and 2) how well the variation of that variable is described by the PC over all data points. The loadings describe the data structure in terms of variable correlations. As seen from Figure G1.b the negative energy descriptors are inversely related to the polarisability, refractivity, volume and surface area descriptors, i.e. the total energy and the SCF atomic energy increases with increasing molecular size. The polarisability and refractivity are increasing with size as expected due the increase in the conjugated pi-electron

system. In the vertical direction of the loading plot, *PC2*, the energies of the highest and next highest molecular orbitals, E_{HOMO} and E_{NHOMO}, have highest loading while the electronegativity, Ena and Enb, and to a lesser extent the hardness of the molecules has negative loadings. The latter explained through the higher stability of the pi-system by the more electronegative and increased hardness of the heteroatoms O and N in respect to S.

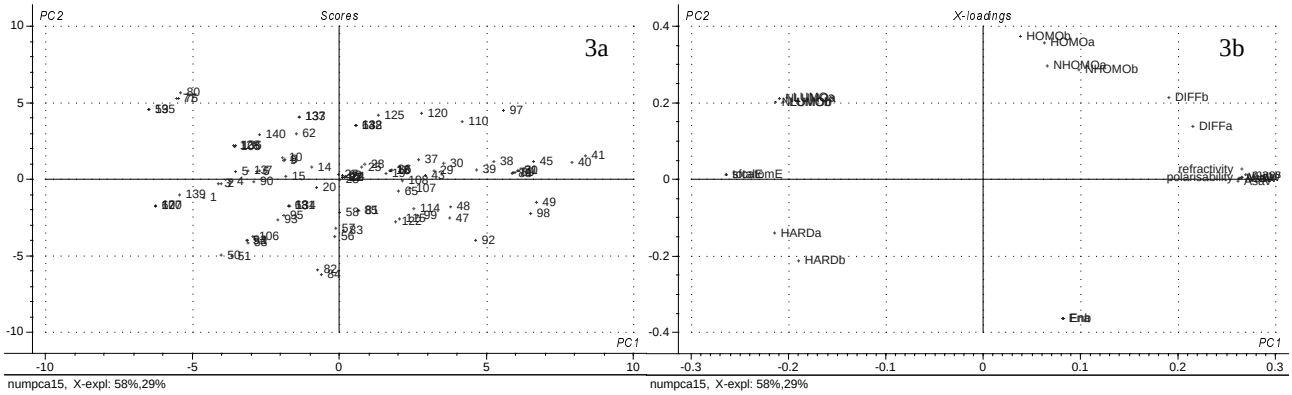


***Figure G1.a and G1.b**. Score and loading plots represented by, on the ordinates, the first principal components (PC1´s), and, on the abscissas, the second principal components (PC2´s). G1.a : the PAC scores, and G1.b: the loadings of original NEMQC-descriptors.*

The score plot (Figure G1.a) describes the data structure through the chemical patterns, and geometrical distances reflect similarities and dissimilarities between chemical structures. Graphically the loading plot can be used to explain and describe these similarities-dissimilarities, i.e. chemicals with high positive scores in *PC2* have highest E_{HOMO}, chemicals with high negative scores in *PC2* are the most electronegative molecules, e.g. quinoline and cyano substituted aromatic structures.

The structure of the score plot described by *PC1* and *PC2* is further shown, by assigned by the type of heteroatoms present within the molecular structure (Figure G1.c), and number aromatic rings (Figure G1.d), respectively.

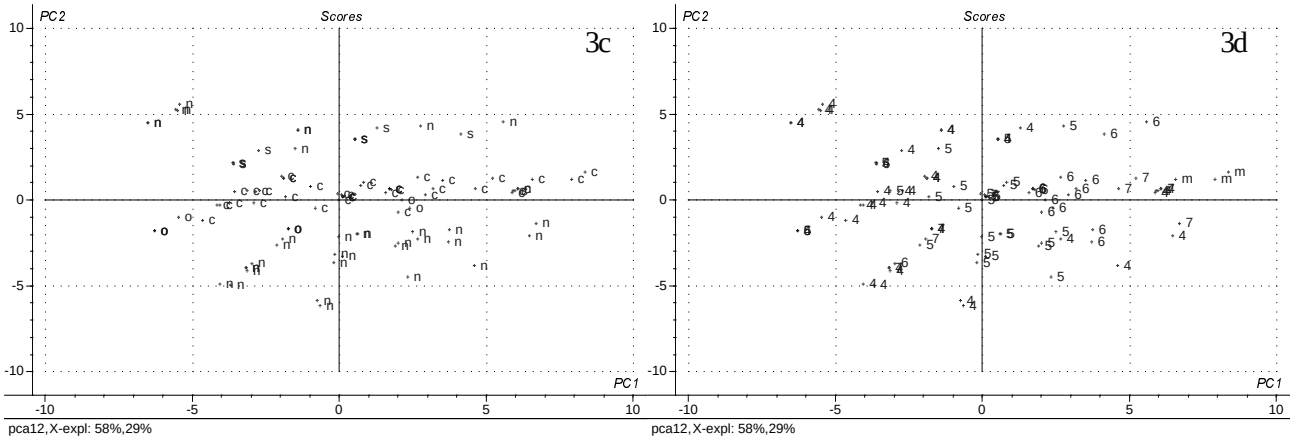


***Figure G1.c and G1.d**. Score plots represented by, on the ordinates, the first principal components (PC1´s), and, on the abscissas, the second principal components (PC2´s). The individual PAHs assigned by G1.c; analogue type and/or heteroatoms included in the molecular structure, and G1.d; the number of aromatic rings in the molecular structure.*

As seen in Figure G1.c, the hardness is decreasing from the third to the first quadrant, which explains a high density of O-PAHs in the third and S-PAHs in the first quadrant. The increasing size with increasing number of aromatic rings is seen in Figure G1.d. It should be emphasised, that the type, and not the number of heteroatoms are denoted in Figure G1.c. Likewise only the number of aromatic rings, not the size of the rings or the size of substituents are shown in Figure G1.d.

Therefore only the general trends in structure-chemical property relations can be extracted from Figures G1.c and G1.d.

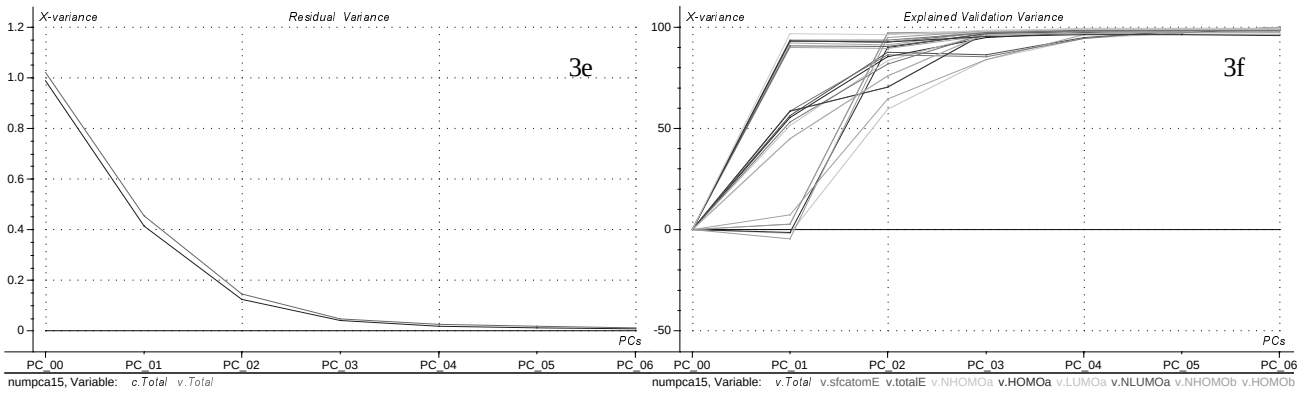


Figure G1.e and G1.f*. The total residual variance (G1.e) for the calibrated (blue line) and cross-validated (pink line) model as function of the number of PCs. The explained X-variance as function of the number of PCs (G1.f), showing the significance of each original variables in each PC.*

As seen from Figure G1.e, the optimum number of PCs is four after which there is no further decrease in the residual variance in the X-space. Figure G1.f illustrates how the inter-correlated variables with the highest explanatory capability in describing systematic variation in X is accentuated in each PC, whereas the orthogonal (non-correlated) variables are split into different PCs.

As illustrated from the above PCA no significant spitting of the molecular structures into groups of different type are observed. However by including specific toxicity descriptors in the initial PCA analysis, or by fitting to logEC_{50}, significant groupings are indicated[xvii]. Therefore, one should always analyse both the chemical property space based on solely molecular descriptors and PCA or PLS (if enough data are available) which will allow for a selection of the most significant descriptors with respect to the modelled endpoint, EC_{50}. If groupings are occurring upon including EC_{50} in the X-matrix, then there is a strong indication of different modes of action of the individual groups that should be analysed in separate models.

Experimental design[e.g. 116] is based on a selection of a few compounds to span the chemical domain of the model to be calibrated and validated. The including of specific endpoint toxicity data in the initial PCA are therefore crucial if test data for modelling, or designing experimental investigations, are to be selected. After the obtaining of a homogeneous spanning of the X-space including relevant descriptors, the most significant PCs, e.g. PC1 and PC2 may be used for experimental design. In the case of non-specific modes of action, the experimental design may be based on the most significant PC´s of the X-space spanned by inherent structural and electronic property descriptors[xvii].

Appendix H

Preprocessing of logK_{ow} data – PAH and N,S,O-PAHs

Missing data is an often-occurring limitation in SAR/QSAR studies. As logK_{ow} is often used in initial screening and in simple linear regression models it is of utmost importance that logK_{ow} data are evaluated and present for all other endpoint values to be investigated. Through a study in literature logK_{ow} values measured by use of different methods and in different laboratories were collected. These are given in Table H1. Column one contains the names of the individual PAHs, column two the id of the three-dimensional optimised structures, column three the cas-number, column four the indicated the types of hetero-atoms included in the molecular structure, column five the measured octanol-water partitioning coefficients, column six the standard deviation on the measured logK_{ow} when given in the original source, column seven the denotes the methods used for measuring logK_{ow}, and lastly pH, as this parameter was given in few cases.

Table H1 *Literature data on Octanol-water partition coefficients for 50 PAH and N, S, O-PAHs*

name	*id*	*cas.no.*	*type*	*logKow*	*std*	*method*	*pH*
1,5-dimethylnaphthalene	5	571-61-9	C	4.38	.	.	.
1,5-dimethylnaphthalene	5	571-61-9	C	4.38	.	shaking ang GC	.
10-azabenzo(a)pyrene	.	.	N	5.00	6.40E-01	ODS-65	.
10-azabenzo(a)pyrene	.	.	N	5.69	5.80E-01	Diol-35	.
13H-dibenzo[a,i]carbazole	98	239-64-5	N	6.40	.	.	.
1-ethylnaphthalene	4	1127-76-0	C	4.39	.	shaking ang GC	.
1-methylfluorene	15	1730-37-6	C	4.97	.	shaking ang GC	.
1-methylnaphthalene	2	90-12-0	C	3.86	.	.	.
1-methylnaphthalene	2	90-12-0	C	3.87	.	shaking ang GC	.
2,2´biquinoline	93	119-91-5	N	4.31	.	.	.
2,3-dimethylnaphthalene	6	581-40-8	C	4.38	.	.	.
2,3-dimethylnaphthalene	6	581-40-8	C	4.40	.	shaking ang GC	.
2.6-dimethylnaphthalene	7	581-42-0	C	4.31	.	shaking ang GC	.
2.6-dimethylnaphthalene	7	581-42-0	C	4.38	.	.	.
2-ethylnaphthalene	91	939-27-5	C	4.39	.	.	.
2-hydroxy-4-methylquinoline	108	607-66-9	NOH	1.70	1.20E-02	HPLC	6
2-hydroxyquinoline	107	59-31-4	NOH	1.30	2.72E-02	HPLC	6
2-Methylanthracene	26	613-12-7	C	5.15	.	.	.
2-methylnaphthalene	3	91-57-6	C	3.86	.	.	.
2-methylnaphthalene	3	91-57-6	C	3.86	.	shaking ang GC	.
4-azafluorene	.	.	N	2.59	6.50E-01	ODS-65	.
4-azafluorene	.	.	N	2.96	5.40E-01	Diol-35	.
4-hydroxyquinoline	106	3558-69-8	NOH	0.61	1.05E-02	HPLC	6
9(10H)-Acridone	59	578-95-0	NO	2.95	2.80E-02	HPLC	6
Acenaphthene	13	83-32-9	C	3.92	.	shaking ang GC	.
Acenaphthene	13	83-32-9	C	4.03	.	.	.
Acridine	44	260-94-6	N	3.18	6.40E-01	ODS-65	.
Acridine	44	260-94-6	N	3.27	5.30E-01	Diol-35	.
Acridine	44	260-94-6	N	3.23	4.83E-02	HPLC	6
Acridine	44	260-94-6	N	3.62	.	.	.
Anthracene	25	120-12-7	C	4.53	1.90E-01	ODS-65	.

Anthracene	25	120-12-7	C	4.54	.	shaking ang GC	.	
Anthracene	25	120-12-7	C	4.55	6.10E-01	Diol-35	.	
Anthracene	25	120-12-7	C	4.63	.	.	.	
Benz[a]acridine	45	225-11-6	N	4.05	6.30E-01	ODS-65	.	
Benz[a]acridine	45	225-11-6	N	4.48	5.30E-01	Diol-35	.	
Benzo(a)anthracene	27	56-55-3	C	5.50	6.40E-01	Diol-35	.	
Benzo(a)anthracene	27	56-55-3	C	5.54	1.90E-01	ODS-65	.	
benzo(a)pyrene	36	50-32-8	C	5.98	.	shaking ang GC	.	
benzo(a)pyrene	36	50-32-8	C	6.00	.	.	.	
benzo(a)pyrene	36	50-32-8	C	6.02	1.90E-01	ODS-65	.	
benzo(a)pyrene	36	50-32-8	C	6.14	7.10E-01	Diol-35	.	
benzo(a)pyrene	36	50-32-8	C	6.27	.	polydimethylsiloxane fibre and magnetic stirring of the sample and GC-MS	.	
benzo(a)pyrene	36	50-32-8	C	6.50	.	.	.	
Benzo(ghi)perylene	43	191-24-2	C	7.10	.	.	.	
Benzo[a]fluorene	16	238-84-6	C	5.75	.	.	.	
Benzo[b]fluorene	18	243-17-4	C	5.75	.	.	.	
Benzo[f]quinoline	55	85-02-9	N	3.37	5.30E-01	Diol-35	.	
Benzo[f]quinoline	55	85-02-9	N	3.46	6.40E-01	ODS-65	.	
Benzo[f]quinoline	55	85-02-9	N	3.51	5.30E-01	Diol-35	.	
Benzo[f]quinoline	55	85-02-9	N	3.69	6.30E-01	ODS-65	.	
Benzo[f]quinoline	55	85-02-9	N	3.57	6.53E-02	HPLC		6
Biphenyl	92	92-52-4	C	3.95	.	shaking ang GC	.	
Carbazole	60	86-74-8	N	3.22	5.30E-01	Diol-35	.	
Carbazole	60	86-74-8	N	3.47	6.30E-01	ODS-65	.	
Chrysene	41	218-01-9	C	5.91	.	.	.	
Dibenz[a,c]anthracene	88	215-58-7	C	6.40	1.90E-01	ODS-65	.	
Dibenz[a,c]anthracene	88	215-58-7	C	6.48	7.60E-01	Diol-35	.	
Dibenz[a,h]acridine	47	226-36-8	N	5.73	5.60E-01	Diol-35	.	
Dibenz[a,h]acridine	47	226-36-8	N	5.93	6.60E-01	ODS-65	.	
Dibenz[a,h]anthracene	29	53-70-3	C	6.54	1.90E-01	ODS-65	.	
Dibenz[a,h]anthracene	29	53-70-3	C	6.60	7.80E-01	Diol-35	.	
Dibenz[a,j]anthracene	89	224-41-9	C	4.99	6.40E-01	ODS-65	.	
Dibenz[a,j]anthracene	89	224-41-9	C	5.63	5.60E-01	Diol-35	.	
Dibenz[a,j]anthracene	89	224-41-9	C	6.44	7.50E-01	Diol-35	.	
Dibenz[a,j]anthracene	89	224-41-9	C	6.54	1.90E-01	ODS-65	.	
Dibenz[a.i]acridine	.	.	N	5.85	5.70E-01	Diol-35	.	
Dibenz[a.i]acridine	.	.	N	5.94	6.60E-01	ODS-65	.	
Dibenz[c,h]acridine	.	.	N	6.27	5.90E-01	Diol-35	.	
Dibenz[c,h]acridine	.	.	N	6.56	6.80E-01	ODS-65	.	
Dibenzofuran	66	132-64-9	O	4.06	6.30E-01	Diol-35	.	
Dibenzofuran	66	132-64-9	O	4.09	1.90E-01	ODS-65	.	
Dibenzothiophene	62	132-65-0	S	4.41	1.90E-01	ODS-65	.	
Dibenzothiophene	62	132-65-0	S	4.43	6.10E-01	Diol-35	.	
Fluoranthene	30	206-44-0	C	5.16	.	polydimethylsiloxane fibre and magnetic stirring of the sample and GC-MS	.	
Fluoranthene	30	206-44-0	C	5.22	.	.	.	
Fluorene	14	86-73-7	C	4.07	6.20E-01	Diol-35	.	
Fluorene	14	86-73-7	C	4.18	.	shaking ang GC	.	
Fluorene	14	86-73-7	C	4.32	1.90E-01	ODS-65	.	
Fluorene	14	86-73-7	C	4.47	.	.	.	
Indole	57	120-72-9	N	0.75	0.00E+00	HPLC	.	
Isoquinolin	49	119-65-3	N	2.21	6.60E-01	ODS-65	.	
Isoquinolin	49	119-65-3	N	2.26	5.60E-01	Diol-35	.	
Naphthalene	1	91-20-3	C	3.35	.	.	.	
Naphthalene	1	91-20-3	C	3.47	1.90E-01	ODS-65	.	
Naphthalene	1	91-20-3	C	3.58	6.60E-01	Diol-35	.	
Perylene	42	198-55-0	C	6.50	.	.	.	

Phenanthrene	19	85-01-8	C	4.48	1.90E-01	ODS-65	.	
Phenanthrene	19	85-01-8	C	4.54	6.10E-01	Diol-35	.	
Phenanthrene	19	85-01-8	C	4.57	.	shaking ang GC	.	
Phenanthrene	19	85-01-8	C	4.63	.	.	.	
Phenanthridine	54	229-87-8	N	3.17	6.40E-01	ODS-65	.	
Phenanthridine	54	229-87-8	N	3.44	5.30E-01	Diol-35	.	
Phenanthridine	54	229-87-8	N	3.43	6.00E-02	HPLC		6
pyrene	34	129-00-0	C	4.84	1.90E-01	ODS-65	.	
pyrene	34	129-00-0	C	5.14	6.20E-01	Diol-35	.	
pyrene	34	129-00-0	C	5.18	.	shaking ang GC	.	
pyrene	34	129-00-0	C	5.20	.	.	.	
pyrene	34	129-00-0	C	5.22	.	.	.	
Quinolin	48	91-22-5	N	2.17	6.60E-01	ODS-65	.	
Quinolin	48	91-22-5	N	2.23	5.60E-01	Diol-35	.	
Quinolin	48	91-22-5	N	2.02	4.12E-03	HPLC		6
Quinolin	48	91-22-5	N	2.04	.	HPLC	.	
Triphenylene	63	217-59-4	C	5.45	.	.	.	

Level 1. Initial screening of endpoint value significance

The data in Table H2 is sorted by increasing average logK_{ow}, on which a one-sided t-test[x] were performed to secure that increase in endpoint values with respect to the uncertainty was significant. The number of samples used for calculating the t-values are given in column five, the critical t-value, $t_{critical}$, are given in column ten. Degrees of freedom, *df* (N-1), are given in column nine.

Due to the low number of samples per compound in some cases, a simple value significance testing was performed as well. The simple value significance testing is based on the criteria

$$d\log K_{ow} = \left[\left(\log K_{ow}^{higher} - std_{higher}\right) - \left(\log K_{ow}^{lower} + std_{lower}\right)\right] > 0 \qquad \text{(H1)}$$

through a pairwise comparison of average logK_{ow} values including the standard deviation on substance specific average, i.e. the average value of the compound in row n minus the standard deviation should be higher than the average value of compound in row n-1 plus the standard deviation. This is the minimum requirement to data.

The hypothesis H_o for the t-test as well as the simple value significance test, is that the average logK_{ow} value of row n is significant higher than the average logK_{ow} value of row n-1 or not. The result of H_o is given in column eleven and thirteen, respectively. If the hypothesis is rejected the answer is no, if not the column assignment is yes.

Table H2 *Mean value difference testing by i) t-test (P=0.10) and ii) simple value significance testing.*

name	*id*	*casno*	*Average logK_{ow}*	*N*	*std*	*var*	*t*	*df*	*$t_{critical}$*	*H_0 i)*	*dlogK_{ow}*	*H_0 ii)*
Quinolin	48	91-22-5	2.12	4	0.10	0.01	2.12	5	2.02	no	0.05	yes
Isoquinolin	49	119-65-3	2.24	2	0.04	0.00	2.89	1	6.31	no	0.77	yes
4-azafluorene	.	.	2.78	2	0.26	0.07	2.61	3	2.35	no	0.49	yes
Acridine	44	260-94-6	3.33	4	0.20	0.04	0.12	5	2.02	no	-0.01	no
Carbazole	60	86-74-8	3.35	2	0.18	0.03	0.01	4	2.13	no	-0.02	no
Phenanthridine	54	229-87-8	3.35	3	0.15	0.02	1.08	5	2.02	no	0.08	yes
Naphthalene	1	91-20-3	3.47	3	0.12	0.01	0.62	7	1.89	no	0.06	yes
Benzo[f]quinoline	55	85-02-9	3.52	5	0.12	0.01	6.42	4	2.13	no	0.23	yes
1-methylnaphthalene	2	90-12-0	3.87	2	0.01	0.00	1.99	1	6.31	no	0.18	yes
Acenaphthene	13	83-32-9	3.98	2	0.08	0.01	1.75	1	6.31	no	0.04	yes
Dibenzofuran	66	132-64-9	4.08	2	0.02	0.00	2.10	3	2.35	no	0.34	yes

Fluorene	14	86-73-7	4.26	4	0.17	0.03	0.02	2	2.92	no	0.14	yes
Benz[a]acridine	45	225-11-6	4.27	2	0.30	0.09	0.37	1	6.31	no	-0.17	no
2,6-dimethylnaphthalene	7	581-42-0	4.35	2	0.05	0.00	1.24	1	6.31	no	0.01	yes
2,3-dimethylnaphthalene	6	581-40-8	4.39	2	0.01	0.00	2.12	4	2.13	no	0.03	yes
Dibenzothiophene	62	132-65-0	4.42	2	0.01	0.00	4.12	4	2.13	yes	0.18	yes
Phenanthrene	19	85-01-8	4.56	4	0.06	0.00	0.19	7	1.89	no	-0.01	no
Anthracene	25	120-12-7	4.56	4	0.05	0.00	7.49	5	2.02	yes	0.66	yes
pyrene	34	129-00-0	5.12	5	0.16	0.02	0.97	6	1.94	no	-0.04	no
Fluoranthene	30	206-44-0	5.19	2	0.04	0.00	0.45	1	6.31	no	0.60	yes
10-azabenzo(a)pyrene			5.35	2	0.49	0.24	0.51	1	6.31	no	-0.28	no
Benzo(a)anthracene	27	56-55-3	5.52	2	0.03	0.00	3.04	1	6.31	no	0.42	yes
Dibenz[a,h]acridine	47	226-36-8	5.83	2	0.14	0.02	0.59	2	2.92	no	-0.01	no
Dibenz[a.i]acridine	.	.	5.90	2	0.06	0.00	0.01	3	2.35	no	0.67	yes
Dibenz[a,j]anthracene	89	224-41-9	5.90	4	0.73	0.53	0.67	4	2.13	no	-0.28	no
benzo(a)pyrene	36	50-32-8	6.15	6	0.20	0.04	1.58	3	2.35	no	0.27	yes
Dibenz[c,h]acridine	.	.	6.42	2	0.21	0.04	0.17	1	6.31	no	-0.12	no
Dibenz[a,c]anthracene	88	215-58-7	6.44	2	0.06	0.00	2.60	4	2.13	no	0.12	yes
Dibenz[a,h]anthracene	29	53-70-3	6.57	2	0.04	0.00	.	.	.	.	.	.

Level 2. ANOVA, Variance significance testing by F-tests

The purpose of the level 2 tesing is to secure that the variance on individual average logK_{ow} is significantly lower than the mean variance between logK_{ow} values. The results of the the hypothesis, H_1, that the variance on the substance specific average logK_{ow} is significantly lower than the mean between variance is given in Table H2 in the last column.

Table H2 *Testing of hypotesis, H_1.*

name	*id*	*cas.no*	*average logKow*	*N*	*std*	*var*	*Tl**	*F_1*	*H_1*
1,5-dimethylnaphthalene	5	571-61-9	4.38	1	.	.	.		
10-azabenzo(a)pyrene	id	casno	5.345	2	0.487904	0.23805	10.04486	251.667	no
13H-dibenzo[a,i]carbazole	98	239-64-5	6.40037	1	.	.	.		
1-ethylnaphthalene	4	1127-76-0	4.39	1	.	.	.		
1-methylfluorene	15	1730-37-6	4.97	1	.	.	.		
1-methylnaphthalene	2	90-12-0	3.865	2	0.007071	5.00E-05	47823.6	251.667	yes
2,2´biquinoline	93	119-91-5	4.30535	1	.	.	.		
2,3-dimethylnaphthalene	6	581-40-8	4.39	2	0.014142	0.0002	11955.9	251.667	yes
2,6-dimethylnaphthalene	7	581-42-0	4.345	2	0.049498	0.00245	975.9918	251.667	yes
2-ethylnaphthalene	91	939-27-5	4.39	1	.	.	.		
2-hydroxy-4-methylquinoline	108	607-66-9	1.69984	1	.	.	.		
2-hydroxyquinoline	107	59-31-4	1.3032	1	.	.	.		
2-Methylanthracene	26	613-12-7	5.15	1	.	.	.		
2-methylnaphthalene	3	91-57-6	3.86	1	.	.	.		
4-azafluorene	.	.	2.775	2	0.26163	0.06845	34.93324	251.667	no
4-hydroxyquinoline	106	3558-69-8	0.612784	1	.	.	.		
9(10H)-Acridone	59	578-95-0	2.95424	1	.	.	.		
Acenaphthene	13	83-32-9	3.975	2	0.077782	0.00605	395.2364	251.667	yes
Acridine	44	260-94-6	3.32592	4	0.201609	0.040646	58.82883	8.583	yes
Anthracene	25	120-12-7	4.5625	4	0.045735	0.002092	1143.192	8.583	yes
Benz[a]acridine	45	225-11-6	4.265	2	0.304056	0.09245	25.86458	251.667	yes
Benzo(a)anthracene	27	56-55-3	5.52	2	0.028284	0.0008	2988.975	251.667	yes
benzo(a)pyrene	36	50-32-8	6.15167	6	0.202624	0.041057	58.24092	4.447	yes
Benzo(ghi)perylene	43	191-24-2	7.1	1	.	.	.		
Benzo[a]fluorene	16	238-84-6	5.75	1	.	.	.		
Benzo[b]fluorene	18	243-17-4	5.75	1	.	.	.		

Benzo[f]quinoline	55	85-02-9	3.51964	5	0.119815	0.014356	166.5666	5.702	yes
Biphenyl	92	92-52-4	3.95	1	.	.	.		
Carbazole	60	86-74-8	3.345	2	0.176777	0.03125	76.51776	251.667	no
Chrysene	41	218-01-9	5.91	1	.	.	.		
Dibenz[a,c]anthracene	88	215-58-7	6.44	2	0.056569	0.0032	747.2438	251.667	yes
Dibenz[a,h]acridine	47	226-36-8	5.83	2	0.141421	0.02	119.559	251.667	no
Dibenz[a,h]anthracene	29	53-70-3	6.57	2	0.042426	0.0018	1328.433	251.667	yes
Dibenz[a,j]anthracene	89	224-41-9	5.9	4	0.730799	0.534067	4.477303	8.583	no
Dibenz[a.i]acridine	.	.	5.895	2	0.06364	0.00405	590.4148	251.667	yes
Dibenz[c,h]acridine	.	.	6.415	2	0.205061	0.04205	56.86516	251.667	no
Dibenzofuran	66	132-64-9	4.075	2	0.021213	0.00045	5313.733	251.667	yes
Dibenzothiophene	62	132-65-0	4.42	2	0.014142	0.0002	11955.9	251.667	yes
Fluoranthene	30	206-44-0	5.19	2	0.042426	0.0018	1328.433	251.667	yes
Fluorene	14	86-73-7	4.26	4	0.173397	0.030067	79.52918	8.583	yes
Indole	57	120-72-9	0.749736	1	.	.	.		
Isoquinolin	49	119-65-3	2.235	2	0.035355	0.00125	1912.944	251.667	yes
Naphthalene	1	91-20-3	3.46667	3	0.115036	0.013233	180.6942	19.474	yes
Perylene	42	198-55-0	6.5	1	.	.	.		
Phenanthrene	19	85-01-8	4.555	4	0.06245	0.0039	613.1231	8.583	yes
Phenanthridine	54	229-87-8	3.34712	3	0.153452	0.023548	101.5467	19.474	yes
pyrene	34	129-00-0	5.116	5	0.157099	0.02468	96.88736	5.702	yes
Quinolin	48	91-22-5	2.1153	4	0.101117	0.010225	233.8654	8.583	yes
Triphenylene	63	217-59-4	5.45	1	.	.	.		

**The variance between average logK_{ow} values for the individual compounds is 2.39, the degree f freedom (N-1) 49. The degrees of freedom in the H_1 hypothesis is therefore (49, n-1) for the individuaal PAHs.*

The H_1 test is rejected for six of the PAHs. In all six cases the present data set, given in Table 1, only contains two logK_{ow} values. Therefore outlier tests (e.g. for small samples Dixon´s Q) are not possible[114], and the six PAHs are eliminated from the data set.

Level 3. Evaluation of between method variances by F-testing

Table H3 *Testing of hypothesis, H2, the method.*

name	method	average logK_{ow}	n	std	var	df(f_a,f_b)*	F2	T2
Quinolin	HPLC	2.03	2	1.33E-02	1.77E-04	3,1	57.74	215.7
Benzo[f]quinoline	Diol-35	3.44	2	9.90E-02	9.80E-03	4,1	1.46	224.6
Benzo[f]quinoline	ODS-65	3.58	2	1.63E-01	2.65E-02	4,1	0.56	224.6
Dibenz[a,j]anthracene	Diol-35	6.04	2	5.73E-01	3.28E-01	3,1	0.01	215.7
Dibenz[a,j]anthracene	ODS-65	5.77	2	1.10E+00	1.20E+00	3,1	0.01	215.7

**df:degree of freedom, f_a=n-1 for the individual PAHs given in Table 2, f_b=n-1 for the method specific logK_{ow} values.*

As shown in Table H3, only few data on different methods are available. Identification of low, method specific, uncertainty is not possible, as, e.g., in the case of the phthalates and the PCBs[III, iii].

Appendix I

Model predictions and influence of noise in data on QSARs for estimating sorption to NOM

In Figure I1 the loading weight of the standard deviation on logK_{oc} values is shown to be significant in *PC2*.

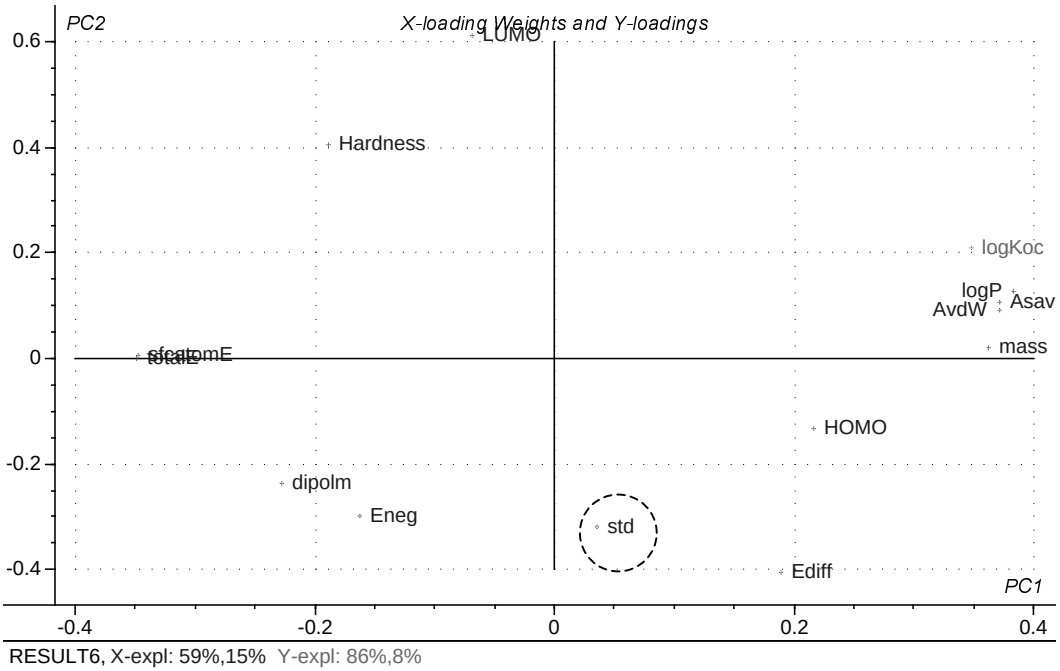


Figure I1 *LogK_{oc} loading and X-loading weights showing the correlation patterns, and significance, of the individual descriptors with respect to logK_{oc} and each other. As seen from PC1, logK_{ow} (denoted logP) is strongly correlated to logK_{oc}. In addition logP is strongly correlated to the mass and area descriptors. LogK_{oc} is inversely related to total energy as well as the self-consistent field energy descriptors. Only 8% of the variance in logK_{oc} in explained in PC2, by use of 15% explained X-variance. The explained X-variance decreases slightly, whereas the explained Y-variance increases slightly by including the standard deviation on average logK_{oc}. The effect is however insignificant comparred to models based on non-preprocessed data as described in Appendix H.*

The information content concerning the electronic descriptors should be evaluated with caution, as the uncertainty in logK_{oc} values is probably due to the heterogeneity of the matrix, i.e. variations the inherent properties of the sorbent. Therefore *PC2*, which is quantifying intrinsic molecular properties, can only be interpreted for uniform sorbent characteristics.

The PLS-QSAR estimated logK_{oc} values for the matrix type, HA, compared to the experimental data are shown in Table I1.

Table I1 *Model predicted and average measured logK_{oc} values for sorption to HA of different origin.*

name	id	Predicted log*Koc*	model std	average measured log*Koc*	cas.no.
2-hydroxyquinoline	107	2,88	0,3	2,70	59-31-4
Quinolin	48	2,90	0,2	2,89	91-22-5
Isoquinolin	49	2,61	0,2	3,09	119-65-3
4-Azapyrene	122	4,28	0,1	3,63	m
Naphthalene	1	3,76	0,2	3,74	91-20-3
Acridine	44	3,89	0,2	4,00	260-94-6
Phenanthridine	54	3,95	0,2	4,06	229-87-8
Benzo[f]quinoline	55	4,05	0,2	4,10	85-02-9
Dibenzofuran	66	3,99	0,1	4,15	132-64-9
pyrene	34	5,03	0,2	4,34	129-00-0
Anthracene	25	4,62	0,2	4,46	120-12-7
Phenanthrene	19	4,80	0,1	4,58	85-01-8
Dibenzothiophene	62	4,96	0,3	4,59	132-65-0
Phenanthrene	19	4,80	0,1	4,65	85-01-8
Fluorene	14	4,54	0,2	4,68	86-73-7
Carbazole	60	4,86	0,3	4,74	86-74-8
Benz[a]acridine	45	4,94	0,2	5,00	225-11-6
Benzo(a)anthracene	27	5,65	0,1	5,62	56-55-3
2-Methylanthracene	26	5,04	0,2	5,78	613-12-7
Dibenz[a,j]acridine	99	5,87	0,3	5,86	244-42-0
Triphenylene	63	5,66	0,2	5,89	217-59-4
Perylene	42	5,83	0,5	5,90	198-55-0
Dibenz[a,h]acridine	47	5,98	0,2	6,05	226-36-8
benzo(a)pyrene	36	5,90	0,1	6,27	50-32-8
Dibenz[a,h]anthracene	29	6,53	0,2	6,44	53-70-3
Dibenz[a,c]anthracene	88	6,46	0,2	6,54	215-58-7
Dibenz[a,j]anthracene	89	6,56	0,2	6,58	224-41-9

As seen from Table I1 the standard deviations on predicted values are insignificant compared to, the often, observed standard deviation of one to two log units by the use of the linear regression models. By comparing the predicted and measured logK_{oc} values the same insignificant differences compared to models based on non-pre-processed data are observed.

Appendix J

Linearity of log*BCF*-logK_{ow} relationships for the PCBs

Several studies have evaluated the bioconcentration of the hydrophobic, including the PCBs, to be parabolic or bilinear function of size or hydrophobicity descriptors. The bilinear or parabolic log*BCF*-logK_{ow} relation has been explained by size restrictions in relation to penetration of the membrane. However, simple linear regression models estimating log*BCF* as a function of logK_{ow} show no sign of steric hindrance due to restrictions caused by size or steric hindrance as illustrated in Figure J1.

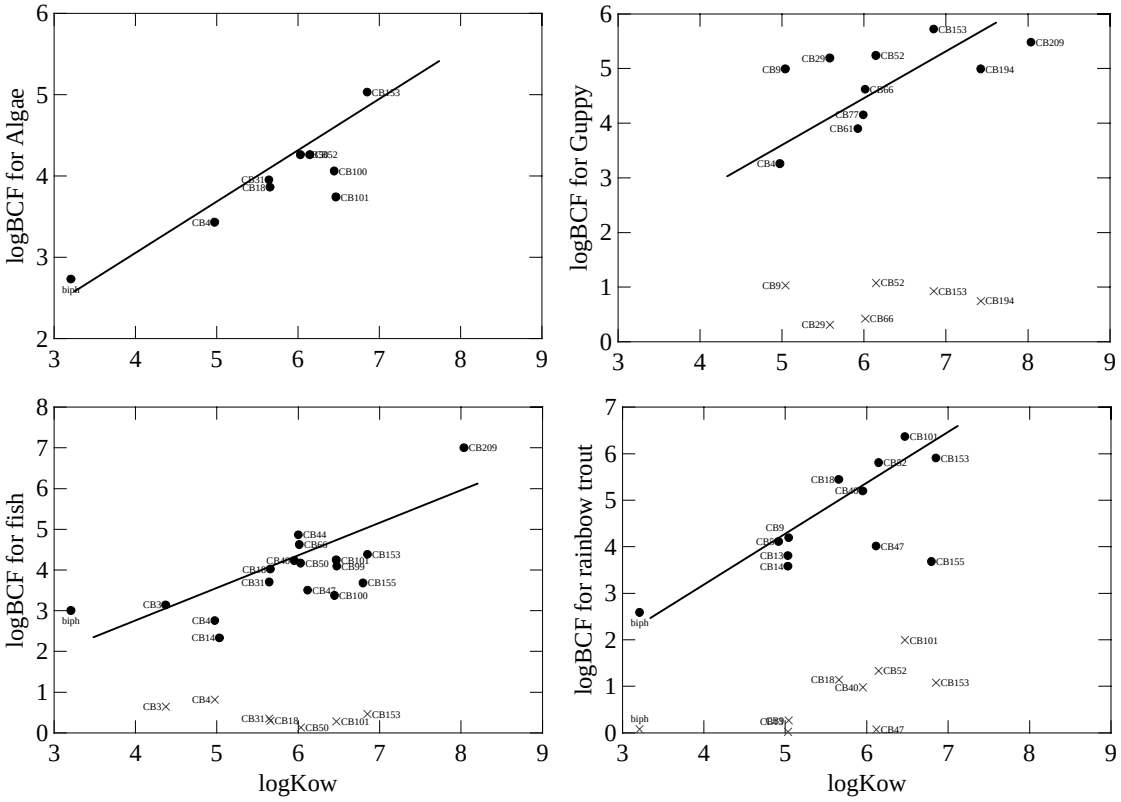


Figure J1 *LogBCF values as function of predicted logK_{ow}^{III} showing no significant sign of parabolic or bilinear relationship. The red crosses are the calculated standard deviation on logBCF values found in the literature.*

As illustrated by the red crosses, the standard deviation on the organism denoted fish, and the Guppy data are fairly homogeneous and below one log-unit. For the rainbow trout data a systematic increase in standard deviation with increasing hydrophobicity is observed. The log*BCF*-logK_{ow} for the Rainbow Trout data, denoting the CB-congeners by the number of ortho-substituted Cl-atoms, are shown in Figure J2.

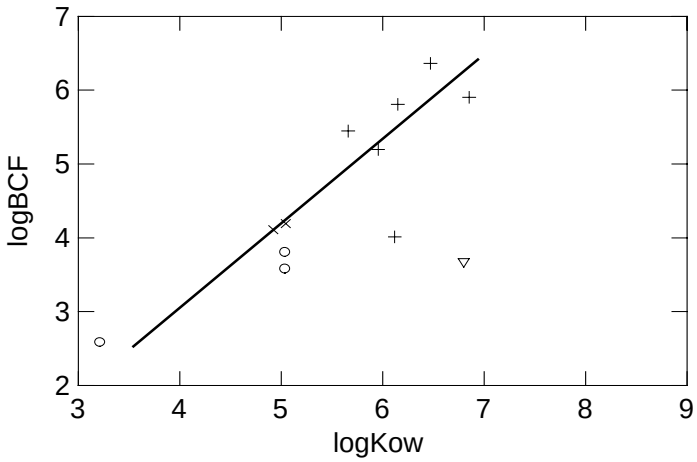


Figure J2 *In the figure of logBCF vs. logK_{ow}, the CB congeners are denoted by the number of ortho-substituted Cl-atoms, i.e. o: zero, +: one, x: two, and ∇: four ortho-substituted Cl-atoms. Clearly grouping are observed, but no information on the total number of Cl atoms or the angle are included, and therefore it is difficult to draw any conclusions concerning the observed grouping. However, two of the CB congeners are significant outliers, which indicates that the twist-angle between the two phenyl rings may be a significant factor with respect to steric hindrance by penetration of the membrane.*

As seen from figure J2 there is an indication that an increase in the number of ortho-substituted Cl-atoms results in decreased bioconcentration. However, only one of the di-substituted CB- congeners show decreased bioconcentration, and additional parameters may contribute in the explanation of the outliers shown in Figure J2.

With respect to the PCBs it has been shown that the degree of co-planarity, Angle, affect the measured fixed toxicity endpoints[III]. Furthermore, the degree of co-planarity show strong influence on the effective charge delocalization between the two aromatic rings, thereby affecting the dipole moment of the individual PCBs.

In Figure J3, the loading weights and log*BCF* loadings of a PLS-QSAR for estimating the bioconcentration shows that the angle may in fact contribute in explaining variations in BCF. However, the significance is strongly influenced by systematic variance inhomogeneities, as shown by including the standard deviation (stdspe22) on measurements obtained from various sources in the literature. The significance of the standard deviation is revealed from the PLS model given in Figure J3

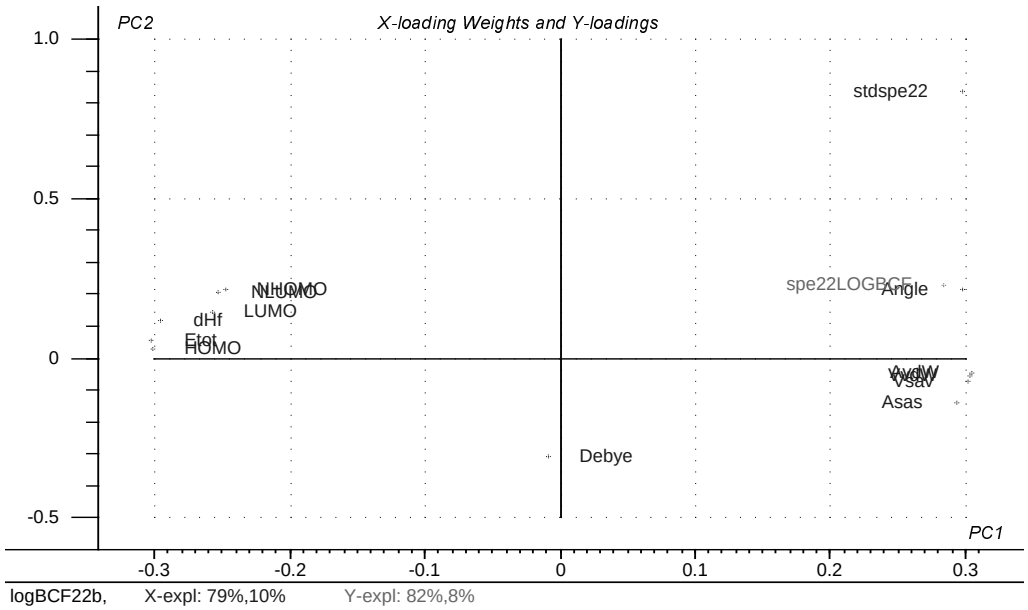


Figure J3 *Plot of X-loading weights in respect to the logBCF-loading (spe22LOGBCF), showing the significance of the Angle in explaining the variation in the bioconcentration factor. However, as observed from the high loading of the standard deviation, STDSPE22, a systematic increase in the uncertainty of experimental data is observed by increasing hydrophobicity (cf. Figure J1).*

Bioconcentration is a quantitative measure of the effective exposure concentration of pollutants towards aquatic organisms. For chemicals acting by non-specific mode of action the bioconcentration is expected to be linear related to the observed fixed toxicity endpoint. With respect to the PCBs it has been shown that the degree of co-planarity, Angle, affect the measured fixed toxicity endpoints. Furthermore, the degree of co-planarity show strong influence on the effective charge delocalization between the two aromatic rings, thereby affecting the dipole moment of the individual PCBs. From Figure J3 it is observed that the non-orthogonality of the biphenyl rings are not a hindrance for penetrating the membrane. The bioconcentration factor increases with size as well as hydrophobicity as illustrated by Figure J2 and J3. However, within group of equal number of Cl-substituent decrease in BCF may be observed at indicated by Figure J2[III].

Paper I

Solubility of Phthalates Revisited. Environmental Implications

Accepted for publication in *Handbook on QSARs for predicting Environmental Fate of Chemicals* (J.D. Walker, Ed.). Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA.

In press.

Solubility of Phthalates Revisited. Environmental Implications

Marianne Thomsen[1], Søren Hvidt[2] and Lars Carlsen[1]*

[1] National Environmental Research Institute,
Department of Environmental Chemistry, DK-4000 Roskilde, Denmark
[2] Department of Chemistry and Life Science,
Roskilde University, DK-4000 Roskilde, Denmark

Corresponding Author:
Lars Carlsen, National Environmental Research Institute,
Department of Environmental Chemistry, DK-4000 Roskilde, Denmark
Tel: + 45 46 30 13 50
Fax: +45 46 30 12 14
E-mail: lc@dmu.dk

ABSTRACT

SAR/QSAR studies based on various techniques (the UNIversal Functional group Activity Coefficients (UNIFAC), Molecular Connectivity Indices (MCI), Electrotopological Indices (EASI), and Molecular Modeling) revealed a pronounced discrepancy between experimentally determined solubilities and theoretically predicted values of phthalates . The latter are found to be several orders of magnitude lower than the experimentally determined values for phthalates with more than 6 carbon atoms in the alkyl ester groups, whereas an excellent agreement was found for the lower molecular weight phthalates**.** The surface tension of the phthalates, DnBP (di-n-butyl phthalate) and DEHP (di-(2-ethyl-hexyl) phthalate) in water, is measured as a function of concentration to investigate any possible surfactant nature of the phthalates and further to determine the unimer solubility of the phthalates in bulk water. The formation of micro-droplet dispersions is more pronounced for phthalates with more than 6 carbons in the ester chain. The environmental implications, such as the migration potential of the phthalates, are discussed.

1. INTRODUCTION

Phthalates have in recent years caused increasing concern due to reported weak carcinogenic (Toppari et al. 1995; Rhodes et al. 1995) and estrogenic effects (Toppari et al. 1995; Nordic Council of ministers 1996; Nielsen and Larsen 1996) of some of these compounds, thus, possibly affecting the male reproductive health. Phthalates are continuously used in very high amounts (Hoffmann 1996), and these compounds must be regarded as being omnipresent throughout the environment, typically in significant concentrations (Thomsen and Carlsen 1998). These compounds are present in significant concentrations in, e.g., sludge-amended soils (Vikelsøe et al. 1998), and are as such potentially leachable. Hence, the environmental behavior of phthalates calls for special attention. Further, recent studies have demonstrated that phthalates may be leached from plastic toys when sucked by children (Rastogi et al. 1997).

The solubility of the phthalates in aqueous solution is a crucial parameter in determining the environmental fate of these compounds. Thus, the leaching of phthalates as well as the subsequent migration in terrestrial environmental matrices will be closely linked to their solubility. This study includes 16 phthalic acid esters for which experimentally determined solubility data are available in the literature (Thomsen and Carlsen 1998). Structural characteristics as well as solubility data are summarized in Table 1.

Previous SAR/QSAR studies (Thomsen et al. 1998a) revealed a pronounced discrepancy between experimentally determined solubilities and theoretically predicted values. The latter are found to be several order of magnitudes lower

than the experimentally determined values for phthalates with more than 6 carbon atoms in the alkyl ester groups, whereas an excellent agreement was found for the lower molecular weight phthalates**.** This fact and the general expectation of an increased hydrophobic effect with an increase in the solute hydrophobic surface area suggest that the UNIFAC predicted values correspond to the solubilities of unimer phthalates (Thomsen et al. 1998a). This paper discusses the solubility behavior of phthalates based on the SAR/QSAR studies in combination with surface-tension measurements of aqueous solutions of the phthalates DnBP and DEHP.

2. METHODS

The UNIFAC method calculates the activity coefficients of solutes in phase equilibrium in water-air and water-octanol systems respectively (Hansen et al., 1991; Chen et al., 1993; Kan and Tomson, 1996; Rasmussen, 1998). The model is based on the assumption that a given compound is composed of a mixture of groups (sub-structural fractions). The activity coefficient of each component or molecule can be described through a combinatory part, which depends on the size and shape of the molecule, i.e., an entropy effect, and a residual part governed by the interactions between the groups, i.e., an energy related effect (Rasmussen 1998). The UNIFAC method is further based on the fundamental assumption applying to group contribution methods, i.e., additivity, indicating that the contribution from one structural group of the molecule is assumed to be independent of that of another group in the same molecule. This is valid only if the contribution of one group within the molecule is not affected by the nature of any of the other groups present in the same molecule.

The Molecular Connectivity Index concept is based on the δ and δ^v connectivity values reflecting the electronic nature of the individual atoms in their valence states. The δ value for a given atom is a simple count of non-hydrogenic sigma bonds to neighboring atoms. The molecular connectivity index is quantified by dissecting the molecule into selected bond contributions (Hall and Kier 1991; Kier and Hall 1976). Thus, the simple first-order MCI is given by

$$^1X = \Sigma_{i,j}(\delta_i \delta_j)^{-0.5} \qquad (1)$$

Through the corresponding first order valence-corrected connectivity index $^1X^v$ it is possible to differentiate between saturated and unsaturated carbon atoms, as well as take into account the lone pair electrons in possible heteroatoms present within the molecular structure (Kier and Hall 1981; 1991), whereas the simple indices encode structural features only.

Molecular Modeling was carried out applying the PCMODEL software implemented for Power Macintosh computers (Serena 1996). The force field used in PCMODEL is called MMX and is derived from MM2 force field (Allinger and Yuh 1981). The molecular structures are geometry-energy optimized.

This study utilizes the build-in option of PCMODEL to calculate the exposed surface area of the investigated molecules in their energy-optimized conformation (Serena 1996). Thus, a stochastic algorithm is used that 1) generates points on the surface randomly, 2) determines if a given point is exposed or not, and 3) counts the number of exposed and unexposed points. The number of calculated points and the number of times the calculation is repeated are used as input parameters (Serena 1996). Since the method is stochastic, the average surface area, which is the result of the repeated calculations, is reported**.** Based on the type of atoms in the molecule under investigation, the total surface area is subdivided into saturated surface area, unsaturated surface area, and polar surface area.

3. EXPERIMENTAL

Surface-tension measurements were performed on a Krüss K10 tensiometer (Hamburg FRG) using the ring method at 25 °C. Stock solutions of the phthalates, DnBP (37.05 mM) and DEHP (0.045 and 0.224 mM), was made in methanol from which standard concentration ranges of the phthalates in water were prepared. The standard solutions were prepared to include the typical values for the water solubility of DnBP and DEHP based on a literature study (Thomsen and Carlsen 1998) as well as the UNIFAC-predicted values (Rasmussen 1998).

4. RESULTS AND DISCUSSIONS

4.1 SAR/QSAR APPROACHES TO PHTHALATE SOLUBILITIES

In a previous study we have demonstrated that the experimentally derived solubilities of phthalates apparently are satisfactorily reproduced by the UNIFAC method for phthalates with ester alkyl chains containing up to ca. 6 carbon atoms (Thomsen et al. 1998a). However, it was obvious that for phthalates with a higher number of carbon atoms in the ester chains, significant deviations between the experimentally and theoretically derived values are noted, as shown in Figure 1. It should be noted that several studies indicate that the experimentally derived values for aqueous solubilities of the high molecular weight phthalates systematically have been overestimated (Staples et al. 1997; Leyder and Boulanger 1983; Pedersen and Larsen 1996; DeFoe et al. 1990).

In order to elucidate the existence of latent informational factors quantifying possible differences between the phthalates with short and long ester chains, respectively, principal component analysis (PCA) was carried out (Höskuldsson, 1996; Bro, 1996). After elimination of non significant descriptors within the original data matrix, specific molecular, structural, and electronic information given by the molecular descriptor variables ^{1}X, ^{2}X, $^{3}X_{p,}$ $^{1}X^{v}$, $^{2}X^{v}$, $^{3}X^{v}_{p}$, $d^{1}X$, $d^{2}X$, S_{ester} and S_{alkyl} (Hall and Kier 1991; 1995; Kier and Hall 1990; 1997; Kier et al. 1991) as well as PCMODEL-derived molecular surface areas and volumes are included in the final PCA. The 2 dimensional representation of the principal components $\mathbf{t}_1$ and $\mathbf{t}_2$ is illustrated in Figure 2.

Based on the score plot in Figure 2, it is clear that the phthalates included in the present study apparently constitute 3 separate groups, containing a) esters with aliphatic ester alkyl groups with up to 6 carbon atoms, b) esters with aliphatic ester alkyl groups with more than 6 carbon atoms and c) esters where one or both ester groups include cyclic or aromatic structures, whereas DAP (di-allyl phthalate) is identified as an outlier with respect to the 3 remaining groups possibly due to the 2 unsaturated allyl ester moieties**.** This grouping within a homologue class of compounds suggests that different mechanisms of dissolution may be involved for the different groups of phthalates. The variable loadings, as shown in Figure 3, demonstrate that the MCI indices and the molecular surface area and volume are highly intercorrelated.

The high degree of intercorrelation between the structural descriptors dominating $\mathbf{p}_1$, which accounts for 74.55 % of the total variation, are not surprising, since ^{1}X is a measure of the relative size of the molecules, whereas the higher order indices include the shape effects, e.g., degree of branching within the molecular structure. The molecular difference indices $d^{1}X$ and $d^{2}X$, which primarily describe the non-sigma electronic (Hall and Kier 1991), and to a lesser extent the ester group electrotopological index S_{ester} (Thomsen et al. 1998), dominate $\mathbf{p}_2$, which accounts for 19.13 % of the total variation. It is also $\mathbf{p}_2$, which describes the outlier status of the compounds DAP, BzBP ,DcHP and DBoEP compared to the di-alkyl esters of the phthalic acid (group a and b in Fig. 2). It has not been possible to assign separate descriptors to the splitting of the phthalates into group a and b, and as such it is not possible to elucidate the different mechanism describing the solubilization of the low and high molecular weight phthalates, respectively. All of the original variables, except $d^{1}X$ and $d^{2}X$, contribute to the separation of the phthalates into a low and high molecular weight group. The spread of the molecular structures in the horisontal direction in the score plot (Fig. 2) is caused by the first principal component, p1, of the loading plot (Fig 3).

Because of the low explained variance of $\mathbf{p}_2$ and the intercorrelation between original variables in $\mathbf{p}_1$, it appears reasonable to apply the simple first-order

MCI in an attempt to describe the unimer solubility of the phthalates, since hydrophobic effects are related to the size of the solute. In Fig. 4 the correlations between ^1X and surface area, respectively, and the experimentally and UNIFAC-derived solubilities, respectively, are displayed**.**

The observed linear correlation between the UNIFAC-derived solubilities and ^1X and the molecular surface area suggests that the solubilities of the unimeric phthalates are correctly reproduced by the UNIFAC calculations.

4.2 THE THRESHOLD OF FORMATION OF DISPERSED PHASES OF DnBP and DEHP IN BULK WATER

In order to explain the apparent discrepancies between the experimentally derived solubility data and the UNIFAC-calculated unimer solubilities, we must understand the actual molecular structures of the phthalates as well as the significance of the environment surrounding the molecule.

The results of the surface-tension measurements for DnBP and DEHP, respectively, can be seen in Figure 5.

The type of behavior indicated by the curves of surface-tension as function of the bulk water concentration is typical for amphiphatic species (Hiemenz and Rajagopalan 1997). The break in the curves represents the threshold of the formation of a dispersion of liquid phthalate in water, i.e., the borderline between a continuous phase of solubilized unimeric phthalate molecules and an emulsion. At the point of saturation no further shift in the equilibrium [$Phthalate_{bulk\ water} \rightleftarrows Phthalate_{air\text{-}water\ interface}$] occurs, which causes the surface-tension to approach a constant value. The surface-tension measurements show that the unimeric molar solubilities of DnBP and DEHP are -4.2 and -7.4 in log units, respectively. While DnBP lowers the surface-tension about 15 mN/m, DEHP reduces the surface-tension by only 5 mN/m. Neither of the phthalates can be characterized as very effective surfactants, but the surface-tension measurements are nonetheless a unique method for distinction between a true solution and a dispersion of clusters of molecules. The results of the surface-tension measurements is in perfect agreement with the UNIFAC-predicted solubilities which are -4.11 and -7.63 for DnBP and DEHP, respectively (Thomsen et al. 1998a) . The explanation for the apparent inconsistency in the experimental solubility data found in the literature (Thomsen and Carlsen, 1998) as well as the apparent discrepancy between experimentally and theoretically derived solubilities are probably to be found, partly, in the unawareness of the slow kinetics of macroscopic phase separation when using the conventional shake-flask method (Staples at al. 1997). More appropriate methods such as centrifugation and turbidity inflection results in less overestimated unimeric solubilities (DeFoe et al.

1990). Experimental procedures which include extraction of the aqueous phase is not appropriate since these methods make it impossible to elucidate the presence of a dispersed oily phase within the bulk water. The concentrations of DnBP and DEHP in natural surface waters exceed their unimeric solubilities in several cases (cf. Thomsen and Carlsen 1998 and references therein). For this reason it is important to be aware of the colloidal nature of the phthalates, and to consider ways to include the colloidal nature of organic compounds in fate studies and thus risk assessments.

4.3 ENVIRONMENTAL IMPLICATIONS

The hypothesis that phthalates with more than 6 carbon atoms in the ester chains in aqueous solution may be present in the form of micro-droplets obviously may have some pronounced effects with respect to the environmental fate of these species. Phthalates have been reported to be significantly retained in soil columns (cf. Thomsen and Carlsen 1998). The mean value of $\log K_{oc}$ found in the literature are 4.37±0.7 and 7.06±0.4 for DnBP and DEHP, respectively (Thomsen and Carlsen 1998). The variation in the K_{oc} values found in the literature may partly be caused by the presence of dispersions of micro-droplets in the bulk water phase. It seems highly unlikely that possible micro-droplets will be retained in soil columns to any significant extent. If a top soil is contaminated with phthalates, e.g. through sludge amendment, the solubilization of the phthalates as micro-droplets may significantly increase the potential risk of leaching, whereby the underlying aquifer may be subject to phthalate pollution. It should, however, in this connection be emphasized that as long as sludge amendment is carried out rationally, i.e. such as to avoid the any risk of supersaturating the soil water, the microbial degradation of phthalates most properly will remedy this problem (Rasmussen et al. 1998; Thomsen at al. 1998b; Vikelsøe et al. 1998).

5. CONCLUSIONS

Within a homologue class of compounds such as the phthalic acid esters, different mechanisms of dissolution may be involved for the different groups of compounds shown by the PCA. It has not been possible to assign separate descriptors to the grouping of the di-alkyl phthalates. Thus, it is not possible to elucidate the different mechanisms describing the solubilization of the low and high molecular-weight phthalates, respectively.

The surface-tension measurements support the presence of 2 distinct phases as the concentration in water exceeds the unimeric solubility.

It seems that the formation of a dispersed phase is more pronounced for the high molecular weight phthalates. A full understanding of this process of

dissolution requires knowledge about the energetics of the process and thereby the surrounding media to be included in the modeling.

The presence of dispersions in the form of micro-droplets obviously may have pronounced effects with respect to the environmental fate of the phthalates. The migration potential of the phthalates can be expected to increase significantly due to the presence of phthalate micro-droplets in the bulk water.

REFERENCES

Allinger NL, Yuh YH (1981): MM2: Molecular Mechanics II, QCPE 13, 395.

Bro R (1996): Håndbog i multivariate kalibrering. Kemometri. 1. ed., The Royal Veterinary and Agricultural University, Copenhagen, Denmark.

Chen F, Holten-Andersen J, Tyle H (1993): New Developments of the UNIFAC Model for Environmental Application. *Chemosphere*; 26, 1325-1354.

DeFoe DL, Holcombe GW, Hammermeister DE (1990): Solubility and Toxicity of Eight Phthalate esters to four Aquatic Organisms, *Environ.Tox.Chem.*9, 623-636.

Hall LH, Kier LB (1991): The Molecular Connectivity Chi Indexes in Structure-Property Modeling. In: Reviews in Computational Chemistry. Vol. 2. (Eds: Lipkowitz,Kenny B; Boyd,Donald B) VCH publishers, New York.

Hall LH, Kier LB (1995): Electropological state indices for atom types: A novel combination of electronic, topological, and valens state information. *J.Chem.Inf. Comput.Sci.* 35, 1039-1045.

Hansen HK, Rasmussen P, Fredenslund AA, Schiller, M, Gmehling J (1991) Vapor-Liquid Equilibria by UNIFAC Group Contribution. 5. Revision and Extension *Ind. Eng. Chem. Res.*, 30, pp. 2352-2355.

Hiemenz PC, Rajagopalan R (1997): Principles of Colloid and Surface Chemistry, 3rd Ed., Marcel Dekker, New York.

Hoffmann L (1996) Massestrømsanalyse for Phthalater. Miljøprojekt No. 320, Danish Environmental Protection Agency, Copenhagen, Denmark.

Höskuldsson A (1996): Prediction methods in Science and Tecnology. Vol. 1 Basic Theory, Thor Publishing, Denmark.

Kan AT, Tomson MB (1996): UNIFAC prediction of aqueous and nonaqueous solubilities of chemicals with environmental interest. *Environ. Sci. Technol.* 30, 1369-1376.

Kier LB, Hall LH (1976): Molecular Connectivity in Chemistry and Drug Research. Vol. 14. Academic Press, NY.

Kier LB, Hall LH (1981): Derivation and significance of valence molecular connectivity. *J.Pharm.Sci.* 70, 583-589.

Kier LB, Hall LH (1990): The molecular connectivity of non-sigma electrons. *Rep.Mol. Theory* 1, 121-125.

Kier LB, Hall LH (1991): A Differential molecular connectivity index. *QSAR* 10, 134-140.

Kier LB, Hall LH, Frazer JW (1991): An index of electropological state for atoms in molecules. *J.Math.Chem.* 7, 229-241.

Kier LB, Hall LH (1997): The E-state as an Extended Free Valence. *J.Chem.Inf. Comput.Sci.* 37, 548-552.

Leyder F, Boulanger P (1983): Ultraviolet Absorption, Aqueous Solubility, and Octanol-Water Partition for Several Phthalates, *Bull. Environ.Contam. Toxicol.* 30, 152-157.

Nielsen E, Larsen PB (1996): Toxicological evaluation and limit values for DEHP and phthalates, other than DEHP. Environmental Review No. 6, Danish Environmental Protection Agency, Copenhagen, Denmark.

Nordic Council of Ministers (1996): Chemicals with Estrogen-like Effects, TemaNord, report no. 580.

Pedersen F, Larsen J (1996): Review of environmental fate and effects of di(2-ethylhexyl)phthalate. Working report. No. 54, Danish Environmental Protection Agency, Copenhagen, Denmark.

Rasmussen AG (1998): Prediction of Aqueous Solubilities and Octanol-Water Coefficients of Phthtalates by the UNIFAC Group Contribution Method. NERI Research Notes No. 69, National Environmental Research Institute, Roskilde, Denmark.

Rasmussen AG, Sørensen PB, Thomsen M, Vikelsøe J, Carlsen L (1998): Model Analysis for the Fate of Antropogenic Substances in Differenly Dressed Soils. Contribution to the Eight Annual Meeting of SETAC-Europe, Bordeaux, Apr. 1998.

Rastogi SC, Vikelsøe J, Jensen GH, Johansen E, Carlsen L (1997): Migration of phthalates from teetheers.NERI Research Notes No. 64, National Environmental Research Institute, Roskilde, Denmark.

Rhodes JE, Adams WJ, Biddinger GR, Robillard KA, Gorsuch JW (1995): Chronic toxicity of 14 phthalate esters to Daphnia Magna and Rainbow Trout (oncorhyncnus mykiss). *Environ.Toxicol.Chem.* 14, 1967-1976.

Serena (1996): PCMODEL v. 6.0. Molecular Modelling Software. Serena Software, Box 3076, Bloomington, IN 47402-3076.

Staples CA, Peterson DR, Parkerton TF, Adams WJ (1997): The Environmental Fate of the Phthalates Esters: A Litterature Review, *Chemosphere*, 35, 667-749.

Thomsen M, Carlsen L (1998): Phthalater i Miljøet. Opløselighed, Sorption og Transport. NERI technical report (in press), National Environmental Research Institute, Roskilde, Denmark.

Thomsen M, Rasmussen AG, Carlsen L (1998a): SAR/QSAR Approaches to Solubility, Partitioning and Sorption of Phthalates. Submitted for publication.

Thomsen M, Vikerlsøe J, Johansen E, Carlsen L (1998b): The Fate of Xenobiotic Compounds in Soil. Contribution to the Eight Annual Meeting of SETAC-Europe, Bordeaux, April 1998.

Toppari J, Larsen JC, Christiansen P, Giwercman A, Grandjean P, Gulillette LJ Jr., Jégou B, Jensen TK, Jounannet P, Keiding N, Leffers H, McLachlan JA, Meyer O, Müller J, Meyts ER, Scheike T, Sharpe R, Sumpter J, Skakkebæk NE (1995): Male Reproductive Health and Environmental Chemicals with Estrogenic Effects. Miljøprojekt No. 290, Danish Environmental Protection Agency, Copenhagen, Denmark.

Vikelsøe J, Thomsen M, Johansen E, Carlsen L (1998): The Occurrence of Phthalates and Nonylphenoles in Soil. A Field Study of Different Soil Profiles. NERI technical report (in press), National Environmental Research Institute, Roskilde, Denmark.

Indeks

Figure Captions

Figure 1. Experimental solubilities of selected phthalates versus the UNIFAC predicted solubility values. The diagonal line represents the potential perfect agreement between the UNIFAC calculated and experimental solubilities.

Figure 2. Two-dimensional representation of the principal component scores, $\mathbf{t}_1$ and $\mathbf{t}_2$, based on specific molecular structural and electronic information for the selected phthalates.

Figure 3. Two-dimensional representation of the first two loading vectors, $\mathbf{p}_1$ and $\mathbf{p}_2$,.

Figure 4. Experimental and UNIFAC derived solubilities of selected phthalates described through the simple first-order molecular connectivity index ^1X in (a) and the calculated exposed molecular surface area in (b). (Open circles are experimental data. Filled circles are UNIFAC calculated values)

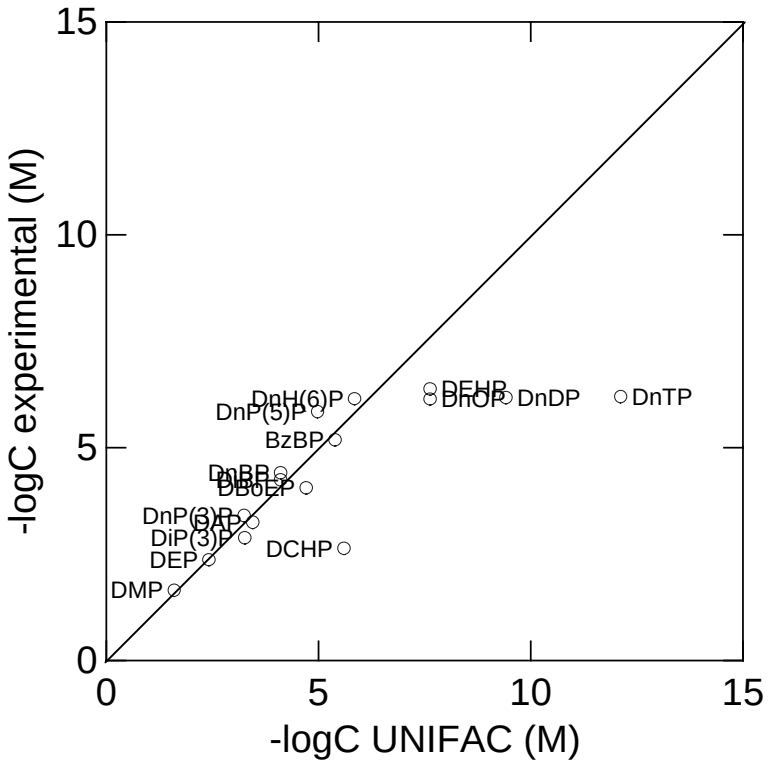
Figure 5. Measured surface tensions for DnBP (a) and DEHP (b) as function of the bulk water concentration.

Table 1. Structural characteristics and solubility data for the 16 phthalic acid esters included in this study.

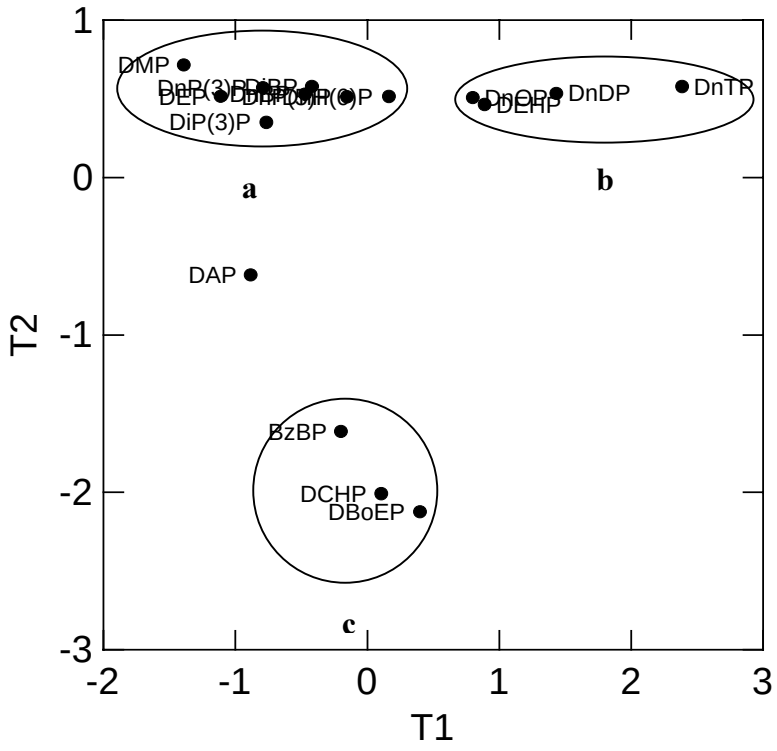
Cas-nr.	acronym*	-R_1	-R_2	$\log C_w^{sat}$	$\log C_w^{sat,UNIFAC}$
131-11-3	DMP	$-CH_3$	$-CH_3$	1.646	1.60
84-66-2	DEP	$-CH_2CH_3$	$-CH_2CH_3$	2.364	2.42
131-17-9	DAP	$-CH_2CH=CH_2$	$-CH_2CH=CH_2$	3.242	3.45
605-45-8	DiP(3)P	$-C(CH_3)_2$	$-C(CH_3)_2$	2.877	3.26
131-16-8	DnP(3)P	$-CH_2CH_2CH_3$	$-CH_2CH_2CH_3$	3.401	3.25
84-69-5	DiBP	$-CH_2CH(CH_3)_2$	$-CH_2CH(CH_3)_2$	4.238	4.11
84-74-2	DnBP	$-CH_2(CH_2)_2CH_3$	$-CH_2(CH_2)_2CH_3$	4.402	4.11
131-18-0	DnP(5)P	$-CH_2(CH_2)_3CH_3$	$-CH_2(CH_2)_3CH_3$	5.839	4.98
84-75-3	DnH(6)P	$-CH_2(CH_2)_4CH_3$	$-CH_2(CH_2)_4CH_3$	6.144	5.85
85-68-7	BzBP	$-CH_2(CH_2)_2CH_3$	$-CH_2$-phenyl	5.180	5.39
117-84-0	DnOP	$-CH_2(CH_2)_6CH_3$	$-CH_2(CH_2)_6CH_3$	6.137	7.63
117-81-7	DEHP	$-CH_2CH(CH_2CH_3)(CH_2)_3CH_3$	$-CH_2CH(CH_2CH_3)(CH_2)_3CH_3$	6.374	7.63
84-61-7	DCHP	-cyclohexyl	-cyclohexyl	2.630	5.60
84-77-5	DnDP	$-CH_2(CH_2)_8CH_3$	$-CH_2(CH_2)_8CH_3$	6.173	9.42
119-06-2	DnTP	$-CH_2(CH_2)_{11}CH_3$	$-CH_2(CH_2)_{11}CH_3$	6.194	12.12
117-83-9	DBoEP	$-(CH_2)_2$-O-$(CH_2)_3CH_3$	$-(CH_2)_2$-O-$(CH_2)_3CH_3$	4.049	4.71

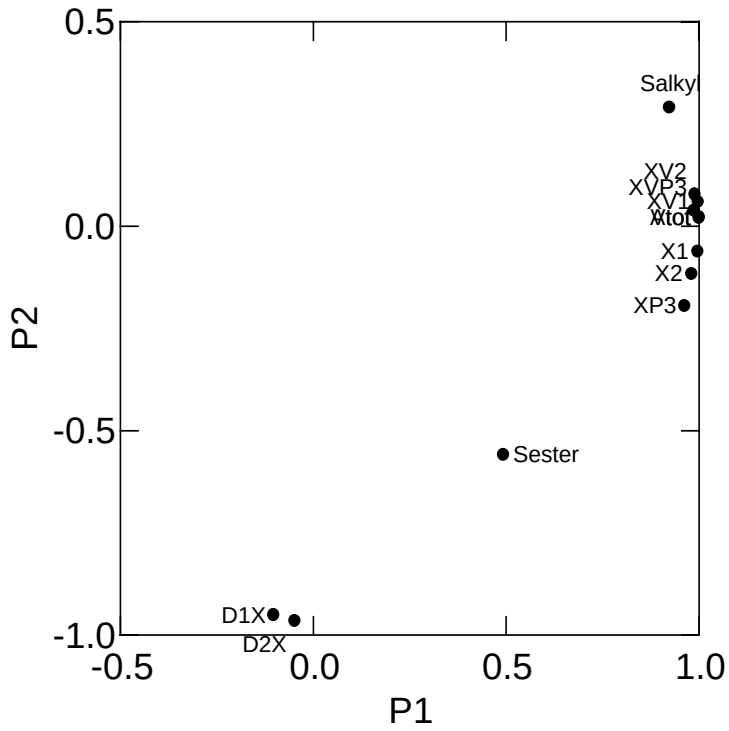
*DMP (di-methyl phthalate), DEP (di-ethyl phthalate), DAP (di-allyl phthalate), DiP(3)P (di-iso-propyl phthalate), DnP(3)P (di-n-propyl phthalate), DiBP (di-iso-butyl phthalate), DnBP (di-n-butyl phthalate), DnP(5)P (di-n-pentyl phthalate), DnH(6)P (di-n-hexyl phthalate), BzBP (benzyl butyl phthalate), DBzP (di-benzyl phthalate), DnOP (di-n-octyl phthalate), DEHP (di(2-ethyl-hexyl) phthalate), DCHP (di-cyclohexyl phthalate), DnDP (di-n-decyl phthalate), DnTP (di-n-tridecyl phthalate), DBoEP (di-2-butoxyethyl phthalate)

Paper I. *Solubility of Phthalates Revisited. Environmental Implications, Figure 1*

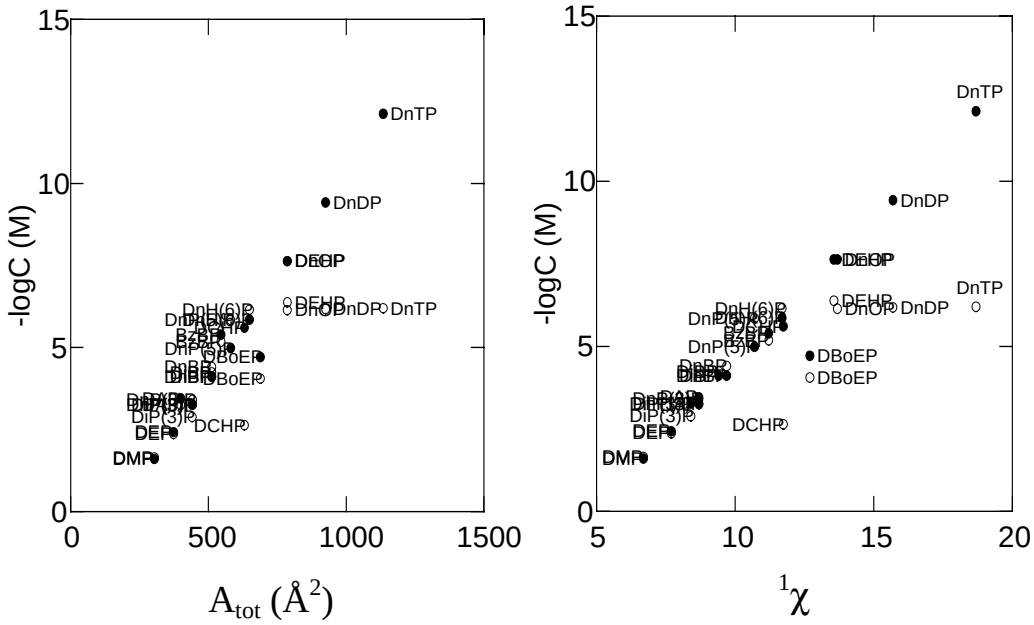


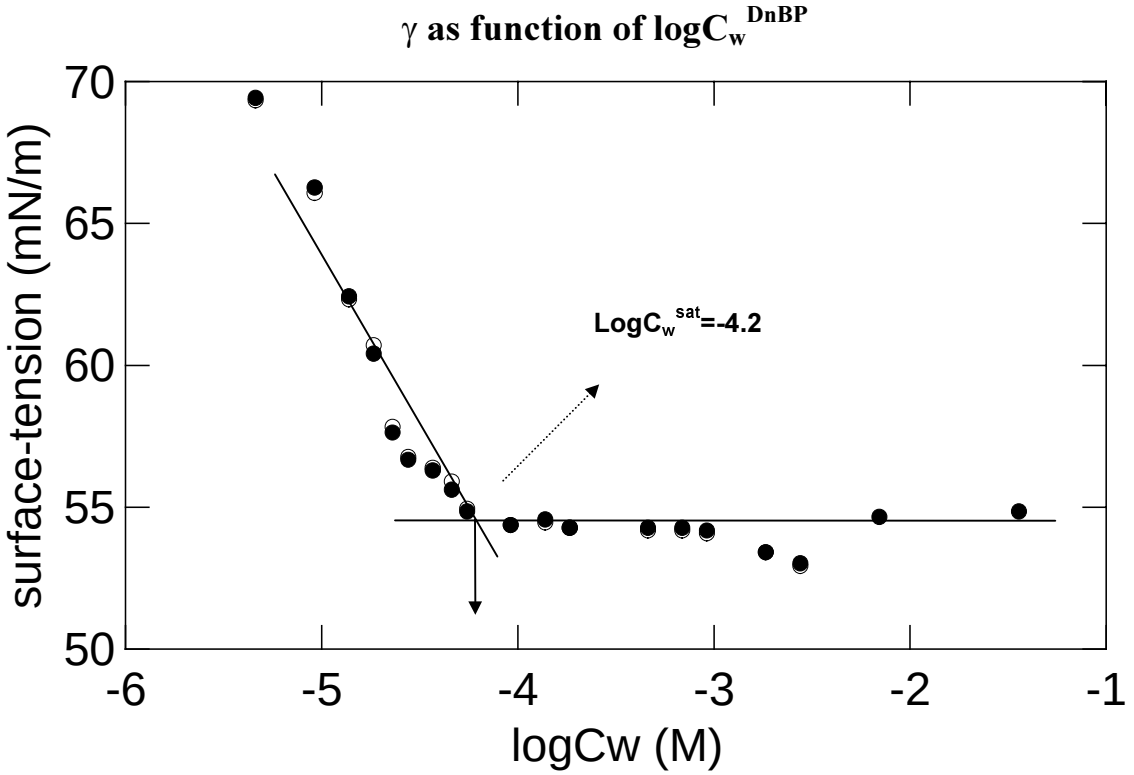
Paper I. Solubility of Phthalates Revisited. Environmental Implications, Figure 2

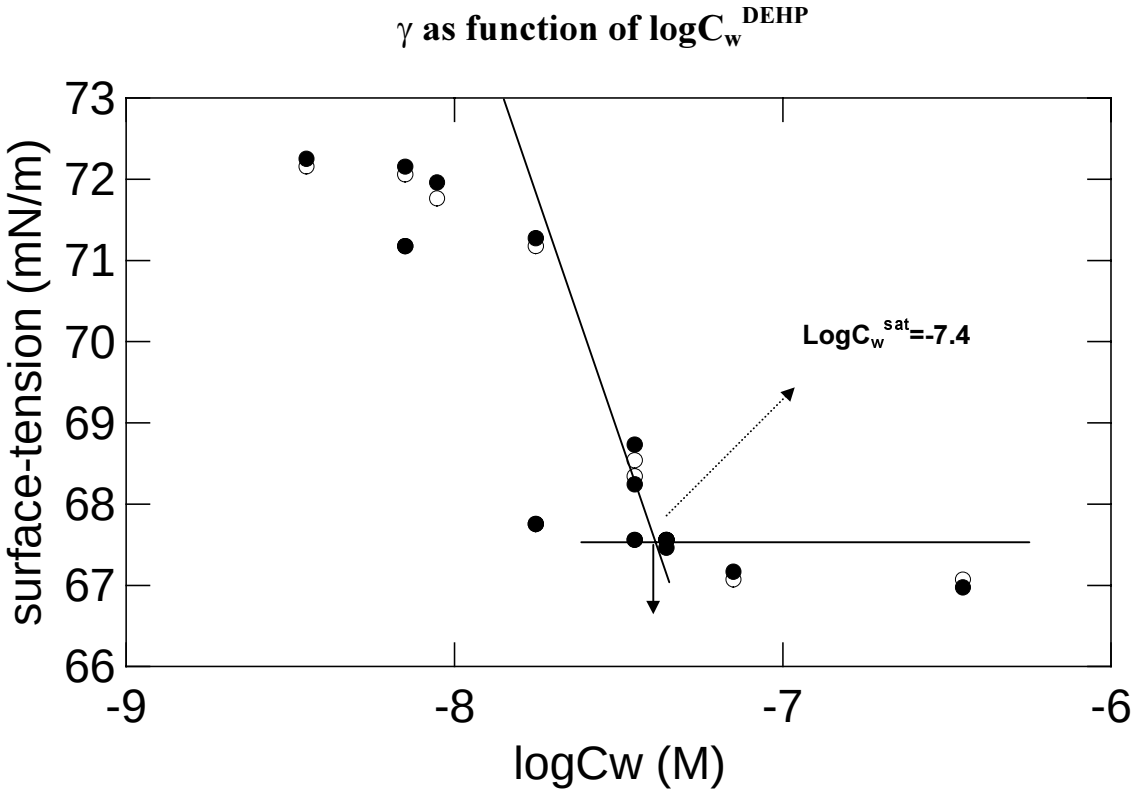



0.5
0.0
-0.5
-1.0
P2
-0.5
0.0
0.5
1.0
P1
Salky
XV2
XVP3
X1
X2
XP3
Sester
D1X
D2X

Paper I. Solubility of Phthalates Revisited. Environmental Implications, Figure 4




γ as function of logC_w^DnBP
surface-tension (mN/m)
logCw (M)
LogC_w^sat=-4.2
70
65
60
55
50
-6
-5
-4
-3
-2
-1


γ as function of logC_w^DEHP
surface-tension (mN/m)
logCw (M)
LogC_w^sat=-7.4
73
72
71
70
69
68
67
66
-9
-8
-7
-6

Paper II

Solubilities and surface activities of phthalates investigated by surface tension measurements,

Environmental Toxicology and Chemistry 20, pp. 127-132.

SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 20, No. 1, pp. 127–132, 2001

Printed in the USA
0730-7268/01 $9.00 + .00

SOLUBILITIES AND SURFACE ACTIVITIES OF PHTHALATES INVESTIGATED BY SURFACE TENSION MEASUREMENTS

MARIANNE THOMSEN,†‡ LARS CARLSEN,† and SØREN HVIDT*‡
†Department of Environmental Chemistry, National Environmental Research Institute, DK-4000 Roskilde, Denmark
‡Department of Chemistry, Roskilde University, DK-4000 Roskilde, Denmark

(*Received* 17 *November* 1999; *Accepted* 18 *May* 2000)

Abstract—Aqueous solutions of DEP (di-ethyl), DnBP (di-*n*-butyl), DnH(6)P (di-*n*-hexyl), and DEHP (di-[2-ethyl-hexyl]) phthalates have been investigated by use of surface tension measurements at temperatures between 10 and 35°C. A tensiometric approach allows for the determination of unimeric solubilities and ΔG°, which is the standard Gibbs free energy change, for the dissolution of phthalates in water. The unimeric solubility of the phthalates increase with decreasing temperature. The ΔG° shows a linear increase with increasing phthalate alkyl chain length. The contribution of enthalpy (ΔH°) and entropy (ΔS°) to ΔG° were calculated from the temperature-dependent solubilities. The contributions of both ΔH° and ΔS° are negative and increase in magnitude with increasing alkyl chain length, suggesting hydrophobic interactions between phthalates and water. The ability of different phthalates to lower the surface tension decreases with increasing alkyl chain length, whereas the relative affinity for adsorption in the air–water interface increases drastically for long-chain phthalates. Despite the low surface activity of phthalates compared with that of common surfactants, they show significant affinity for adsorption in air–water interfaces of natural surface waters. This property, combined with their low solubilities, may affect the fate of these compounds within the natural environment, because they form emulsions above unimeric saturation in aqueous media.

Keywords—Aqueous solubility Dissolution thermodynamics Phthalates Sorption Surface tension

INTRODUCTION

Phthalates are used as plasticizers in a variety of applications, and they are among the most abundant manmade chemicals in the environment. Thousands of tons of plastics are disposed of annually in landfills, thus possibly enabling phthalate esters to migrate into soil and groundwater [1]. The compounds have the ability to bioconcentrate in animals [2,3]. As such, the physicochemical properties of the phthalates determining, for example, migration potential and bioconcentration are crucial when considering the fate of these chemicals in the environment, and among these properties, solubility is of major importance. The unimeric solubility (C_w^{sat}), above which the water phase has the physicochemical characteristics of a dispersion, is crucial in the field of environmental chemistry when considering fate modeling as well as risk assessment.

Conventional shake-flask methods [4,5] have been used to determine the solubilities and partitioning coefficients [5,6] of phthalates. However, results show significant variation, which we have attributed to phthalate and water densities being very similar [7,8]. An aqueous bulk phase that is supersaturated with phthalates may reach local thermodynamic equilibrium between dissolved and dispersed phthalates fairly rapidly. However, macroscopic phase separation of a supersaturated bulk phase may be a slow process, because the densities of phthalates and water are very similar [9,10]. Some authors have noted difficulties during experimental studies of phthalate behavior in aqueous media because of the small difference between the densities of water and phthalates [4,6,9,10]. Thus, Howard et al. [4] mentioned difficulties in obtaining samples of the bulk without contamination from undissolved, light-density phthalates floating on the surface. This problem is reflected through a measured solubility of DEHP (di-[2-ethyl-hexyl]; 0.34 mg/L ± 0.04 mg/L), which is 20-fold greater than the result of the present study. Solubility data for phthalates in the literature [11] show an increasing variance between results obtained using different methods with increasing size of the alkyl chain length of the phthalates [7,8,11]. We have compared our results with those of others who have been aware of the colloidal nature of the phthalates. Centrifugation has been used to minimize the amount of emulsified phthalate within the bulk water phase, and a turbidity inflection method has been used to detect aggregates or microemulsions [4,6,9,12]. The latter makes it possible to differentiate between chemicals in solution and dispersed chemicals in the form of colloids or emulsions, which scatter light. However, for true solutions of phthalates with an alkyl chain length greater than C-6, a surprising increase in solubility has been noted [8]. This may partly be caused by alkyl folding of the long-chain phthalates [13,14]. For highly overestimated solubilities, a more adequate explanation is the aggregation of phthalate molecules into clusters consisting of few or many molecules, with the latter having a more pronounced effect in reducing the solvent-accessible surface area [15].

The lowest solubilities for the long-chain phthalates have been obtained by Ellington [13] and Letinksi et al. [14] by use of the slow-stirring method that was developed to avoid the occurrence of associating phthalates in bulk water. The method has been used for measuring octanol/water partitioning coefficients [10,13] and water solubilities [9,13,14] of the long-chain phthalates. The measured solubility of DEHP is 0.0019 mg/L [14], which is approximately one order of magnitude less than the solubility obtained in the present study. Several authors have suggested the colloidal nature of phthalates and their surface activities [16,17] as a possible explanation for

* To whom correspondence may be addressed (hvidt@ruc.dk).

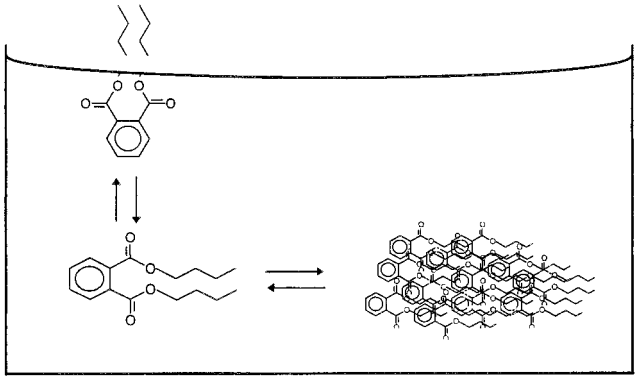


Fig. 1. Model representation of phthalates (di-*n*-butyl phthalate [DnBP]) in equilibria in bulk and surface aqueous solution. K_1 is the equilibrium partitioning coefficient of unimeric phthalates between surface and bulk, whereas K_2 is the equilibrium partitioning coefficient between aggregated liquid phthalate and bulk water phase.

the high variations in both the measured solubilities and octanol/water partitioning coefficients when using different techniques. This study uses a tensiometric approach to measure unimeric solubilities of phthalates as a function of temperature. Through the tensiometric approach, we can measure the bulk activity of the unimeric solutes based on the measured activity of the solute molecules in the surface. When the bulk water phase becomes saturated, only a minimal further change occurs in the activity at the surface. Above the unimeric saturation, a dispersion of aggregates or a formation of microemulsion occurs within the bulk water phase. The objective of this study was to investigate the surface activities of the phthalates and to test the capability of this method for estimating the unimeric solubilities of the high-molecular-weight phthalates. The limit of detection of this method depends on a combination of the numerical solubility of the solute and the effect they produce on the surface tension.

MATERIALS AND METHODS

Theory

Phthalates in bulk water and at the air–water interface. The partition of phthalates between a liquid phthalate phase and the bulk water phase system is described by the equilibrium ($P^* \rightleftharpoons P_{(aq)}$) and the partition of phthalates between the bulk water phase and the surface ($P_{(aq)} \rightleftharpoons P_{surface}$). This system is illustrated in Figure 1. Because of similar densities of the solute and the solvent, an emulsion is formed when the unimeric solubility is exceeded. At concentration levels greater than the unimeric solubility, the aggregates are distributed throughout the bulk water phase, and equilibrium distributions are obtained through partitioning of unimeric phthalates dissolved in bulk to the surface and the microemulsion, respectively, according to Figure 1. Our model is analogous to micellar systems; however, in contrast to micelles, the emulsions are not expected to consist of well-defined, ordered structures. This results from the less pronounced amphiphatic nature of the phthalates.

Relation between the measured surface tension and the bulk water concentration. In dilute solutions, in which the concentration of phthalates within the bulk water phase is far less than saturation, we only need to consider the equilibrium ($P_{(aq)} \rightleftharpoons P_{surface}$). At each experimental concentration, the partitioning of phthalate between aqueous bulk and air–water interface at equilibrium is determined by $\mu_p^s = \mu_p^w$ (i.e., the chemical potentials of phthalates at the surface and in the bulk water phase are equal). For adsorption from dilute solutions, the activity coefficient of phthalate in the aqueous bulk phase is approximately equal to unity when the standard chemical potential (μ_p^o) is defined as the infinite dilution standard chemical potential. So defined, the activity can be replaced by the molar concentration (C), and the change in surface tension (γ) is determined by the phthalate concentration in the bulk water phase through

$$\Gamma_p^w = \frac{n_p^s}{A_s} = -\frac{1}{\ln 10 \cdot RT}\frac{\partial \gamma}{\partial \log C} \quad (1)$$

where Γ_p^w is the surface excess concentration of phthalates given as mol solutes at the air–water interface (n_p^s) per surface area (A_s). The R and T are the gas constant and the absolute temperature, respectively. From the slope of the surface tension (γ) versus the bulk concentration in water (log C), we can determine the surface excess concentration of phthalates [17–19]. The latter can be expressed as the area occupied per molecule at the surface by dividing the inverse of Γ_p^w by Avogadro's number (N_A). The unimeric solubilities are determined as the concentration with a significant change of slope ($\partial\gamma/\partial$ log C). Such determinations were made at 10, 25, and 35°C.

The surface pressure (π) of a solution is defined as the surface tension of pure water (γ_o) minus the surface tension of the solution. The surface pressure of saturated aqueous phthalate solutions (π^{sat}) was calculated at each experimental temperature.

Standard free energy of adsorption at the air–water interface. In the low surface pressure area of the curves, we also evaluated the standard free energy of surface adsorption ($\Delta_a\mu^o$), in which the standard states are hypothetically infinite dilution states using molar fractions as concentrations [19]:

$$\Delta_a\mu^o = -RT \ln\left(\frac{\pi}{a}\right)_0 = -RT \ln\left(-\frac{\partial x}{\partial \gamma}\right)^{-1} \quad (2)$$

where

$$\left(\frac{\pi}{a}\right) = -\left(\frac{\partial \gamma}{\partial a}\right)_{a=0}$$

is the initial slope of the π versus a plot, where a is the activity that can be replaced by the mol fraction (x) for dilute solutions [18,19]. The surface pressure data were analyzed using the method described by Aveyard and Chapman [19].

Experiments

The DEP (di-ethyl; 99%; Aldrich Chemical, Milwaukee, WI, USA), DnBP (di-*n*-butyl; >98%; Fluka Chemical, Milwaukee, WI, USA), DnH(6)P (di-*n*-hexyl; >98%; Chem-Service, West Chester, PA, USA), and DEHP (> 98%; Fluka Chemical) were used as received. A stock solution of each phthalate in methanol (MeOH) was prepared. The concentrations of DEP, DnBP, and DnH(6)P were 40,000, 10,000, and 250 mg/L, respectively. For DEHP, two stock solutions of 100 and 10 mg/L in MeOH were prepared. For each of the phthalates, a series of dilutions in Millipore (Bedford, MA, USA) water was prepared, keeping the fraction of methanol less than 5% (w/w) [20]. Because of the very similar densities of the phthalates and water, no phase separation occurs on exceeding the unimeric solubility. Instead, a microscopic phase separation occurs, and a dispersion of oily aggregates of phthalates forms throughout the bulk water phase. Samples were mixed

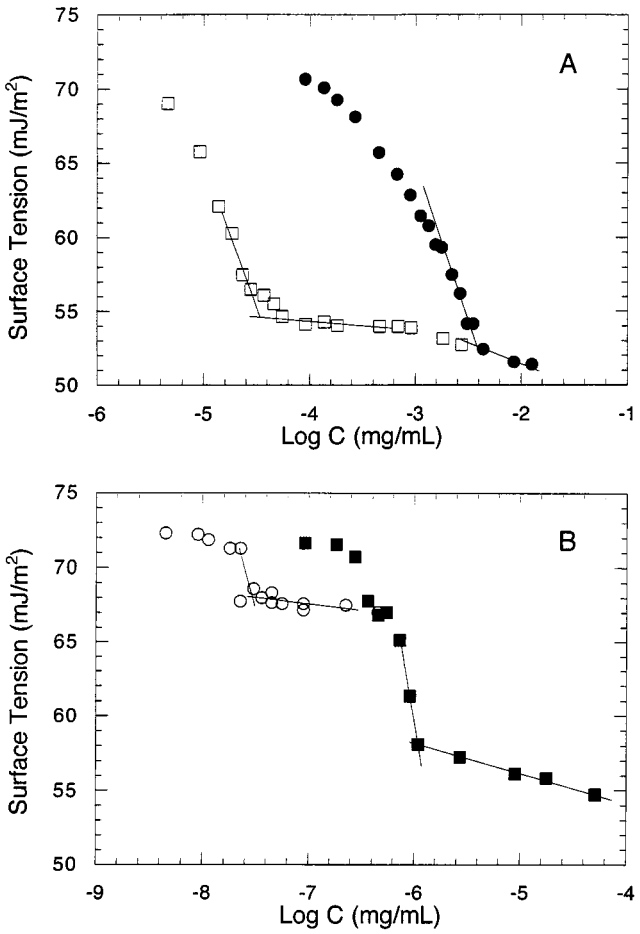


Fig. 2. Surface tension of phthalates as function of concentration in bulk water. **A.** DEP (●) and DnBP (□) at 25°C. **B.** DnH(6)P (■) at 25°C and DEHP (○) at 22°C. DEP = di-ethyl, DnBP = di-*n*-butyl, DnH(6)P = di-*n*-hexyl, DEHP = di-2-ethyl-hexyl.

in a thermostated, shaking water bath for 24 h at 10, 25, and 35°C, respectively, to increase the rate of formation of a homogeneous solution or mixture of phthalates within the bulk water phase. After this, the samples were left at rest in the thermostat for another 24 h, allowing the equilibria, as described in Figure 1, to be established. Solutions were then individually transferred to the thermostated sample container within the K10 Krüss Tensiometer (Hamburg, Germany). The surface tension was measured with the Du Nouy ring method [18].

Before use, all glassware was carefully washed in 10% (v/v) ethanolic HCl, which was followed by heat treatment at 110°C for several hours. The glassware was kept packed in aluminum foil until use to avoid contamination.

RESULTS

Figure 2 shows the surface tension measurements as a function of bulk concentrations in water for DEP, DnBP, and DnH(6)P at 25°C and DEHP at 22°C. A critical aggregation concentration can be determined when the surface tension exhibits two distinct linear concentration regions [18]. The unimeric solubilities (log C_w^{sat}) of the phthalates are determined as the intercept concentrations between the steepest, first part of the curves and the flat, later part of the surface tension curves, as illustrated in Figure 2. The solubilities so determined are summarized in Table 1. The shapes of the curves look similar to curves for typical detergents [18], in which the concentration at the break equals the critical micelle concentration. However, for phthalates, it represents the critical aggregation concentration above which a dispersion of aggregates or clusters is formed, as illustrated in Figure 1. Figure 2 does not show any undershoot of the surface tension just before the solubilities are reached, which would have indicated the presence of more hydrophobic or surface active components than the main component [18]. In the present experiments, the concentration of MeOH in the phthalate solutions was kept less than 5% (w/w) and, in most cases, much lower. At these concentrations, MeOH was not expected to affect the solubility of phthalates [20]. The influence of the MeOH in the phthalate solutions on the measured surface tension was estimated by comparing the experimental measured graphs with the MeOH-corrected graphs, assuming additive contributions of MeOH and phthalates. The corrected graphs were obtained by simply adding the surface tension lowering of MeOH at relevant bulk water concentrations to the measured surface tensions of phthalate solutions. The changes in the slope and intercept of the curves because of MeOH were approximately 1.4 and 0.4%, respectively. Thus, the uncertainty introduced by MeOH is negligible compared with that of each measurement (Table 1), and as such, we did not correct for MeOH in the further interpretation of the experimental data.

Table 1. Unimeric solubilities and standard Gibbs free energy of dissolution ($\Delta G°$) of phthalates at different temperatures[a]

Compound[b]	T (°C)	C_w^{sat} (mg/L)	$\Delta G°$ (kJ/mol)	$T\Delta S°$ (kJ/mol)	$\Delta H°$ (kJ/mol)
DEP	10	1113 (±14)[c]	21.93		
DEP	25	938	23.51	−35.2	−11.6
DEP	35	741 (±7)	24.91		
DnBP	10	13.3 (±1.4)	32.68		
DnBP	25	14.6 (±0.8)	34.39	−57.0	−21.5
DnBP	35	5.5 (±0.8)	38.03		
DnH(6)P	10	0.94	39.55		
DnH(6)P	25	0.52 (±0.03)	43.11	−69.4	−26.4
DnH(6)P	35	0.38 (±0.06)	45.36		
DEHP	22	0.017 (±0.003)	51.39	ND[d]	ND

[a] Standard enthalpy and entropy contributions at 25°C.
[b] DEP = di-ethyl, DnBP = di-*n*-butyl, DnH(6)P = di-*n*-hexyl, DEHP = di-2-ethyl-hexyl.
[c] Values in parentheses represent standard deviations.
[d] ND = not determined.

At the low concentration side of the break in the curves, the form of the Gibbs equation given in Equation 1 can be used for analysis of the surface tension data. From the slope of the linear range of concentrations in Figure 2, the surface excess concentrations of the respective phthalates at C_w^{sat} were calculated by use of Equation 1. In Table 2, surface excess concentrations and surface areas occupied per molecule at the air–water interface at the point of saturation of the aqueous phase are given.

The solubility of phthalates depends on temperature, as shown for DEP in Figure 3, which illustrates that the solubility determined by the concentration break in the curve decreases with increasing temperature. Decreasing solubilities with increasing temperature are common for hydrophobic organic compounds at environmentally relevant temperatures [15]. The measured unimeric solubilities at the different temperatures are summarized in Table 1. For the phthalates DEP, DnBP, and DnH(6)P, these data show an overall decreasing solubility with

Table 2. Surface excess concentration, area occupied per molecule at the surface, and surface pressure at the unimeric saturation of the aqueous phase

Compound[a]	T (°C)	log C_w^{sat} (M)	$\Gamma_p^w \cdot 10^6$ (mol/m²)	$1/(\Gamma_p^w \cdot N_A)$ (Å²/molecule)	π^{sat} (mJ/m²)	log (π^{sat}/x_{sat})[b] (mJ/m²)
DEP	10	−2.30	6.3	26.5	NC[c]	NC
DEP	25	−2.37	6.9	24.0	19.6	5.4
DEP	35	−2.48	8.0	20.7	NC	NC
DnBP	10	−4.32	5.7	29.4	NC	NC
DnBP	25	−4.28	5.2	32.3	17.8	7.3
DnBP	35	−4.70	13.0	15.3	NC	NC
DnH(6)P	10	−5.55	7.5	22.2	NC	NC
DnH(6)P	25	−5.81	11.0	15.9	16.6	8.8
DnH(6)P	35	−5.95	12.8	13.7	NC	NC
DEHP	22	−7.35	3.3	54.3	4.5	9.8

[a] DEP = di-ethyl, DnBP = di-*n*-butyl, DnH(6)P = di-*n*-hexyl, DEHP = di-2-ethyl-hexyl.
[b] x_{sat} = The unimeric solubility given in molar fractions.
[c] NC = not calculated.

increasing temperature. Furthermore, the change in the standard Gibbs free energy ($\Delta G°$) for the dissolution of the phthalates from pure liquid to the aqueous phase is calculated from the molar fraction of phthalates in the saturated solution (x_w^{sat}) by $\Delta G° = -RT \ln x_w^{sat}$, assuming an activity coefficient that is equal to unity. The temperature dependence of the solubility and, thus, of $\Delta G°$ allows the determination of $\Delta S°$. The change in entropy on dissolution is evaluated through the partial derivative of $\Delta G°$ with respect to temperature. Knowing $\Delta G°$ at the three experimental temperatures and $\Delta S°$, the standard enthalpy ($\Delta H°$) is easily obtained as $\Delta H° = \Delta G° + T\Delta S°$. Values of $T\Delta S°$ at 25°C and of $\Delta H°$ are summarized in Table 1. As shown, the $\Delta G°$ values are positive, and the magnitudes increase with increasing temperature. The entropy and enthalpy contributions to $\Delta G°$ are large and negative for all the phthalates. A one-tailed F-test ($p = 0.05$) demonstrated that the observed variation in solubility with temperature was significant compared with the variation on repeated measurements at each experimental temperature.

DISCUSSION

Surface activity and adsorption of phthalates

The phthalates studied reduced the measured air–solution surface tension from the value of 71.9 μJ/m² at 25°C for water

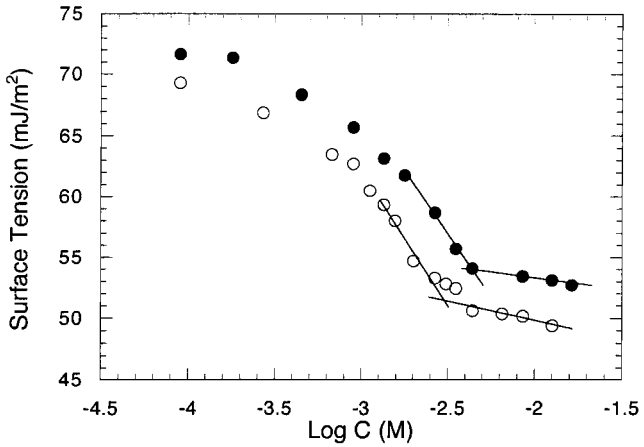


Fig. 3. The surface tension of di-ethyl phthalate (DEP) as function of concentration in bulk water at 10°C (●) and at 35°C (○).

(γ_o). The difference between the surface tension of water and a saturated phthalate solution (π^{sat}) decreases with increasing alkyl chain length, as seen in Table 2. Changes in surface excess concentration, and in area occupied per molecule in the surface, at C_w^{sat} are observed with temperature when looking separately at the data for DEP, DnBP, and DnH(6)P, as shown in Table 2. The surface excess concentration at C_w^{sat} increases with decreasing aqueous solubility. The increase in surface excess concentration corresponds to closer packing of solutes present in the surface (i.e., the surface area occupied per molecule in the surface decreases) [18].

When comparing the data between different phthalates, the surface pressure, as given at 25°C for DEP, DnBP, and DnH(6)P and at 22°C for DEHP, decreases with increasing alkyl chain length. This reflects a decrease in surface excess concentration at C_w^{sat} for the high-molecular-weight phthalates. The insignificant lowering of the surface tension by DEHP might partly result from the branching in the alkyl chain, which is not present in the lower-molecular-weight phthalates included in the study. According to Traube's rule [15], the ratio of the concentration at the surface layer to that in the bulk medium, expressed as π^{sat}/x_w^{sat}, increases approximately threefold for each CH_2 group added to the alkyl chain of a homologous series. In going from DEP to DnH(6)P, an increase of eight CH_2 groups occurs in total. From DnH(6)P to DEHP, an additional increase of two CH_2 and two CH_3 groups occurs. Neglecting the possible effect of the branching of DEHP, we would expect the ratio of the surface to the bulk concentration of the phthalates to increase by a factor of 3^{12} when going from DEP to DEHP. The surface to bulk concentration ratio increases by a log factor of approximately 4.5, as shown in Table 2, which is less than the expected value of 5.7 according to Traube's rule [15]. This might very well result from the branching of DEHP, because the effect of alkyl folding for the solubility of high-molecular-weight phthalates has been estimated to involve a significant increase in unimeric solubility [14]. The increase in the surface to bulk ratio has been observed for other homologous series with different polar groups [15], and these results suggest that migration to the surface results from a low affinity between hydrocarbons and water (i.e., hydrophobic effects). Furthermore, if the attraction between the hydrocarbon chains in the bulk phase was important for migration to the surface, the process would show cooperativity at low solute concentrations, and the bulk and surface concentrations would not be linearly related [15].

Despite the absolute lower surface coverage, expressed as the surface area occupied per molecule at the surface, of DEHP compared with DEP, the relative affinity for adsorption at the surface increases significantly with increasing alkyl chain length. The numbers in Table 2 refer to a saturated bulk phase. At a concentration of 0.017 mg/L (i.e., the solubility of DEHP), none of the lower-molecular-weight phthalates would be present at the surface to any significant degree (compare with the concentration ranges in Fig. 2). The increase in surface to bulk concentration ratios is further supported by the calculated standard free energy of adsorption at infinite dilution, as shown in Table 3.

The average $\Delta_a\mu°$ for the adsorption of one methylene group from dilute aqueous solution is −2.43 kJ/mol according to Aveyard et al. [19]. However, they found a slightly smaller value for diesters, which was explained by a modification of the hydration energy close to the polar ester group. In Table 3, we compare the increment in the calculated standard free

Table 3. Measured and calculated standard free energies of phthalate absorption at the air–water interface[a]

Compound	DEP	DnBP	DnH(6)P	DEHP
$\Delta_a\mu^\circ$ (kJ/mol) (experimental)	−34.1	−44.2	−53.2	−60.8
$\Delta_a\mu^\circ$ (kJ/mol) (predicted)[b]	−34.1	−43.8	−53.9	−63.0

[a] DEP = di-ethyl, DnBP = di-*n*-butyl, DnH(6)P = di-*n*-hexyl, DEHP = di-2-ethyl-hexyl.
[b] Predicted assuming −2.43 kJ/mol per CH_2 group [19].

energy of adsorption with the predicted value of −2.43 kJ/mol for each methylene group compared with DEP as a starting point. The increment in $\Delta_a\mu^\circ$ corresponds closely with an increase of four methylene groups when going from DEP to DnH(6)P. This increment has been argued to reflect a flat orientation of the alkyl chains, parallel to the surface at infinite dilution [19]. However, the low area occupied per molecule at the surface at C_w^{sat}, as shown in Table 2, compared with a calculated [21] surface area of approximately 374 Å^2 for DEP to 787 Å^2 for DEHP, supports an orthogonal-to-the-surface orientation for all the phthalates at bulk saturation, as illustrated in Figure 1. For DEHP, the experimentally determined increase in $\Delta_a\mu^\circ$ is slightly lower, suggesting that branching of the alkyl chain is reducing the solvent-accessible surface area.

Standard free energy of dissolution

The calculated ΔG°, ΔH°, and $T\Delta S^\circ$ values (expressed as kJ/mol) at 25°C for DEP, DnBP, and DnH(6)P have been linearly related to the number of C atoms in the alkyl chains (n_C) of the phthalates:

$$\Delta G^\circ = 2.4n_C + 14.3 \qquad \Delta H^\circ = -1.9n_C - 5.0$$

$$T\Delta S^\circ = -4.3n_C - 19.4$$

If the model of dissolution of unimeric phthalate molecules from the oily phthalate phase ($P^* \rightleftharpoons P_{(aq)}$) is correct, we would expect the slope of the $\Delta G^\circ - n_C$ relationship to correspond with the Gibbs free energy change when a methylene group is transferred from the organic phthalate phase into water. The intercept would correspond with the ΔG° for transferring the phthalic acid part of the solute from the organic phase into the water phase. For aliphatic hydrocarbons, Tanford [15] gives a value of 3.4 kJ/mol for the dissolution of one methylene group from bulk liquid hydrocarbon into bulk water, which should be compared with the value of 2.4 kJ/mol found in the present study. Overall, the solubility is favored by the negative enthalpic contribution but is controlled by the increasing and dominating negative entropy contribution because of hydrophobic effects [15].

Comparison with other methods for estimating unimeric solubilities of phthalates

The solubility ($\log C_w^{sat}$) decreases with increasing alkyl chain length, and the measured solubilities are in agreement with results of earlier quantitative structure-activity relationship studies on the phthalates [7,8]. The UNIFAC-calculated solubility of DEHP at 25°C is 0.0092 mg/L, which is in good agreement with the measured [7] solubility of 0.017 mg/L in this study.

The surface tension method shows a linear decrease in solubility up to C_8 phthalates. Hence, this method appears to be superior to conventional shake-flask methods followed by centrifugation [4,9,12]. The latter obviously obscures measurement of the true solubility because of the inability to differentiate between true solutions and dispersions. The limitations of the shake-flask method followed by centrifugation lie in the very similar densities of phthalates and water, which make a macroscopic phase separation virtually impossible. Thus, the results obtained with this method are a function of the effectiveness of the centrifugation (i.e., the time range and revolutions per minute) [4,9,12].

The limitations of the turbidity approach [6] lie in its dependence on cooperativity of the cluster formation within the bulk phase. Thus, the high solubility of DEHP found by use of the turbidity method could indicate low cooperativity of the aggregation process of DEHP within the bulk phase and, thereby, a reduced ability to scatter light at low aggregation numbers.

The limitations of the surface tension method lie in the low surface activity of the high-molecular-weight phthalates. The resolution or accuracy of our tension measurement is approximately 0.1 to 0.2 mJ/m^2, and the temperature precision is ±1°C. The resolution of this type of method can be increased to 1 μJ/m^2 and the temperature constancy to ±0.1°C. An improvement in the sensitivity and accuracy of the method might make it possible to measure the solubilities of phthalates with more than six carbons in the alkyl chain.

Even the high relative standard deviation of the measured solubility for DEHP does not explain the difference between the measured solubilities for DEHP in this study, which are in agreement with the UNIFAC-calculated solubilities [7], and the results obtained by Letinksi et al. [14]. The reason for the lower solubility found by Letinksi et al. may partly be the experimental procedure of the slow-stirring methods. The long-chain phthalates in particular have high affinities for adsorption at interfaces of laboratory equipment, and increasing the preprocessing of the sample may reduce the recovery significantly [22–24]. Furthermore, because the affinity for sorption at interfaces varies significantly between different phthalates and as function of bulk concentration, the quantified recoveries obtained by use of internal standard may be inaccurate [13].

Impacts on the environment

That phthalates with more than six carbon atoms in their ester chains apparently exhibit much lower unimeric solubilities in pure water than previously argued may have some pronounced effects regarding the environmental fate of these compounds in natural waters. At environmental concentration levels [11,25], the adsorption of DEHP at air–water interfaces can be expected to be most significant when compared with that of the lower-molecular-weight phthalates. Because surface excess concentration increases with increasing temperature, in contrast to the corresponding water solubility, we can expect to observe a similar, seasonal variation in natural surface water. An increased presence of phthalates in surface waters [25], partially through adsorption in the air–water interface, may increase the bioconcentration in microbes feeding in surface waters [24,26].

The inverse relationship between measured bioconcentration factors and octanol/water partitioning for the high-molecular-weight phthalates might very well result from exposure concentration levels greater than the unimeric solubility [3]. The 24-h exposure data on DEHP in the work of Mayer [2] confirm this. Mayer measured an increase in bioconcentration

with increasing exposure concentration from 0.002 to 0.0081 mg/L, with the latter value being in excellent agreement with the unimeric saturation concentration found in the present study. When the exposure concentration is further increased, from 0.014 to 0.062 mg/L, the bioconcentration factor decreases [2].

Because of the very similar densities of phthalates and water, transport from natural surface waters most likely occurs through adsorption to dissolved organic matter, followed by sedimentation, or through uptake by microbes feeding on surface waters, followed by turnover sedimentation of nondegraded phthalates. Transport through continental surface waters might differ from that expected when assessing the risk of this class of compound based on solubilities in the literature, because the nature of an emulsion differs significantly from the properties of a true solution. An aqueous emulsion of phthalates can be expected to behave similarly to dispersed organic matter, because both are aqueous dispersions of colloidal size. In the terrestrial environment, phthalates in the form of microemulsions may, in respect to the earlier comparison, cause an increased potential risk for leaching of these substances to underlying aquifers [27], because microemulsions are unlikely to be retained in soil to any significant extent compared with phthalates in their unimeric form. Furthermore, even at low concentrations, surfactants seem to alter the soil physics, chemistry, and biology in various ways (e.g., the wetability and stability of soil particles) [28].

The solubility of the phthalates decreases significantly with increasing temperature in the range from 10 to 35°C. Even though the surface pressure of the phthalates decreases significantly with an increase in the alkyl chain length, the relative affinity for adsorption at the air–water interface increases drastically for the high-molecular-weight phthalates. Compared with common surfactants (e.g., sodium dodecyl sulfate), linear alkylbenzenesulfonates, and nonylphenol ethoxylates, the surface activities of the phthalates are less pronounced. Despite this, they show high affinity for adsorption at air–water interfaces, with the standard free energy for adsorption increasing with an increasing number of methylene groups in the alkyl chain. The results of the increment in the standard free energy of adsorption per methylene group, the increase in surface to bulk ratio at C_w^{sat}, and the $\Delta G^\circ - n_C$ relationship support the idea that migration to the surface as well as the process of dissolution are caused solely by hydrophobic effects.

Acknowledgement—Financial support from the Danish Research Academy to M. Thomsen and from the Danish Natural Science Foundation to S. Hvidt is greatly appreciated. We thank two reviewers for valuable comments and suggestions.

REFERENCES

1. Giam CS, Atlas E, Powers MA Jr, Leonard JE. 1984. *The Handbook of Environmental Chemistry.* Springer-Verlag, Heidelberg, Germany.
2. Mayer FL. 1976. Residue dynamics of di-2-ethylhexyl phthalate in fathead minnows (*Pimephales promelas*). *J Fish Res Board Can* 33:2610–2613.
3. Parkerton TF, Konkel WJ. 2000. Application of quantitative structure activity relationships for assessing the aquatic toxicity of phthalate esters. *Ecotoxicol Environ Saf* 45:61–78.
4. Howard PH, Banerjee S, Robillard KH. 1985. Measurement of water solubilities, octanol/water partition coefficients and vapor pressures of commercial phthalate esters. *Environ Toxicol Chem* 4:653–661.
5. Staples CA, Peterson DR, Parkerton TF, Adams WJ. 1997. The environmental fate of phthalate esters: A literature review. *Chemosphere* 35:667–749.
6. Leyder F, Boulanger P. 1983. Ultraviolet absorption, aqueous solubility, and octanol-water partition for several phthalates. *Bull Environ Contam Toxicol* 30:152–157.
7. Thomsen M, Rasmussen AG, Carlsen L. 1998. SAR/qsQSAR approaches to solubility, partitioning and sorption of phthalates. *Chemosphere* 38:2613–2624.
8. Thomsen M, Hvidt S, Carlsen L. 2000. Solubility of phthalates revisited. Environmental implications. In Walker JD, ed, *Handbook on QSARS for Predicting Environmental Fate of Chemicals.* Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA.
9. DeFoe DL, Holcombe GW, Hammermeister DE. 1990. Solubility and toxicity of eight phthalate esters to four aquatic organisms. *Environ Toxicol Chem* 9:623–636.
10. De Bruijn J, Busser F, Seinen W, Hermens J. 1989. Determination of octanol/water partition coefficients for hydrophobic organic chemicals with the ''slow-stirring'' method. *Environ Toxicol Chem* 8:499–512.
11. Thomsen M, Carlsen L. 1998. Phthalater i Miljøet. Opløselighed, Sorption og Transport. Technical Report 249. National Environmental Research Institute, Roskilde, Denmark.
12. Wolfe NL, Syeen WC, Burns LA. 1980. Phthalate ester hydrolysis: Linear free energy relationships. *Chemosphere* 9:430–408.
13. Ellington JJ. 1999. Octanol/water partition coefficients and water solubilities of phthalate esters. *J Chem Eng Data* 44:1414–1418.
14. Letinksi DJ, Connelly MJ, Parkerton TF. 1999. Slow-stir water solubility measurements for phthalate ester plasticizers. *Abstracts,* Ninth Annual Meeting of SETAC Europe, Leipzig, Germany, May 25–29, pp 159–160.
15. Tanford C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes.* John Wiley & Sons, New York, NY, USA.
16. Cini R, Desideri P, Lepri L. 1994. Transport of organic compounds across the air/sea interface of artificial and natural marine aerosols. *Anal Chim Acta* 291:329–340.
17. Aveyard R, Binks PB, Fletcher PDI, Kingston PA. 1994. Surface chemistry and microemulsion formation in systems containing dialkylphthalate esters as oils. *J Chem Soc Faraday Trans* 90: 2743–2751.
18. Hiemenz PC, Rajagopalan R. 1997. *Principles of Colloid and Surface Chemistry.* Marcel Dekker, New York, NY, USA.
19. Aveyard R, Chapman J. 1974. Adsorption of polar organic molecules at oil-water and air-water interfaces. *Can J Chem* 53:916–925.
20. Schwarzenbach RP, Gschwend PM, Imboden DM. 1993. *Environmental Organic Chemistry.* Wiley-Interscience, New York, NY, USA.
21. HyperCube. 1997. HyperChem™, Release 5.1 Pro for Windows. Gainesville, FL, USA.
22. Williams MD, Adams WJ, Parkerton TF, Biddinger GR, Robillard KA. 1995. Sediment sorption coefficient measurements for four phthalate esters: Experimental results and model theory. *Environ Toxicol Chem* 14:1477–1486.
23. Vikelsoe J, Thomsen M, Johansen E, Carlsen L. 1999. Phthalate and nonylphenols in soil. Field study of soil depth profiles. Technical Report 268. National Environmental Research Institute, Roskilde, Denmark.
24. Södergren A. 1982. Significance of interfaces in the distribution and metabolism of di-2-ethylhexyl phthalate in an aquatic laboratory model ecosystem. *Environ Pollut* 27:263–274.
25. Fatoki OS, Ogunfowokan AO. 1993. Determination of phthalate ester plasticizers in the aquatic environment of southwestern Nigeria. *Environ Int* 19:619–623.
26. Boese B. 1984. Uptake efficiency of the gills of English sole (*Parophrys vetulus*) for four phthalate esters. *Can J Fish Aquat Sci* 41:1713–1718.
27. Muszkat L. 1989. Large scale contamination of deep groundwaters by organic pollutants. In Longevialle P, ed, *Advances in Mass Spectrometry,* Vol 11B. Heyden & Sons, London, UK, pp 1628–1629.
28. Kuhnt G. 1993. Behavior and fate of surfactants in soil. *Environ Toxicol Chem* 12:1813–1820.

Paper III

Evaluation of Empirical contra non-empirical Descriptors

Accepted for publication in *SAR and QSAR in Environmental Chemistry*
In press.

EVALUATION OF EMPIRICAL VERSUS NON-EMPIRICAL DESCRIPTORS

Marianne Thomsen* and Lars Carlsen[1]

National Environmental Research Institute

Dept. of Environmental Chemistry

DK-4000 Roskilde, Denmark

Running Title: Evaluation of Descriptors

*author to whom correspondence should be addressed; e-mail mth@dmu.dk

[1] Present address: Awareness Center, Hyldeholm 4, Veddelev, DK-4000 Roskilde, Denmark

ABSTRACT

In this case study, based on the notorious PCBs, the performance of selected empirical (EM) versus non-empirical and quantum chemical (NEM-QC) descriptors in multivariate QSARs have been evaluated. The informational content of the EM descriptors has been evaluated with respect to the physical understandable NEM-QC descriptors. Models for estimating key parameters for risk assessment have been developed, based on two-dimensional EM and three-dimensional NEM-QC descriptors, respectively. In spite of the simplicity of the two-dimensional descriptors, no evidence of lower predictive ability of the EM descriptors compared to the NEM-QC descriptors was observed. Homogeneity of variance within and between experimental methods, species and PCB congeners has been analysed, disclosing the importance for the need for handling uncertainty aspects of the results obtained by different laboratories and methods. The latter appears crucial for the model developments when using data from more than one source.

Key words: QSAR, solubility, partitioning, Chromatographic retention, CYP1A induction

INTRODUCTION

The empirical descriptors developed by Kier and Hall [1-10] have been used in a variety of QSAR-studies for estimating endpoints determining the fate and effects of chemical substances within the environment [1-4,7,10-19]. Due to the simplicity of the empirical descriptors their performance in QSAR models have been questioned, especially due to their apparent low capacity in quantifying changes in the electronic nature of molecules [20]. Three-dimensional descriptors, derived from semi-empirical or ab initio geometry optimised molecular conformations, have found pronounced use especially for estimating endpoints involving interaction with bioreceptors [20]. This approach allows for a search for micromechanisms of interactions of the active molecule with the bioreceptor, if the structural characteristics of the bioreceptor are known. Often, however, the nature of the bioreceptor is unknown, and therefore the greatest task within multivariate QSARs is to quantify every possible variation in the structural and electronic properties of the active molecules. The three-dimensional descriptors exhibit a higher degree of specificity toward the spatial arrangements of atoms. In contrast to the empirical descriptors, they correspond directly to real molecular properties, which makes it easier to gain insight and elucidate possible cause-related interactions in respect to an observed endpoint. The informational content of the two-dimensional empirical descriptors, in contrast, is difficult to understand, as they do not have an intuitive physical meaning. The present case study, based on the notorious PCBs, focuses on development of models for estimating endpoints of varying complexity. Endpoints included in this study are bulk properties, such as solubility ($\log C_w^{sat}$) and octanol-water partitioning ($\log K_{ow}$), surface adsorption elucidated through chromatographic retention volumes (V_R) [21], and enzymatic induction properties through EC_{50} values on the induction of CYP1A activity [22]. The main objective of the present study is to evaluate the informational content and model performance, based on 1) empirical (EM) and 2) non-empirical and quantum-chemical (NEM-QC) descriptors, respectively.

DESCRIPTORS

The three-dimensional conformations of the 209 PCB congeners have been geometry optimised by the AM1 semi-empirical, unrestricted Hartree-Fock method [23]. A number of the two-dimensional empirical descriptors are calculated from input files incoding molecular configurations [24].

Empirical Descriptors

The empirical descriptors used comprise: simple molecular connectivity indices (MCIs), ${}^{p}\chi_0$-${}^{p}\chi_{10}$, ${}^{c}\chi_3$, ${}^{pc}\chi_4$, quantifying variations in structural characteristics [2], valence MCIs,${}^{p}\chi^{v}_0$-${}^{p}\chi^{v}_{10}$,${}^{c}\chi^{v}_3$,${}^{pc}\chi^{v}_5$ [8,10], the sum of electrotopological state indices for aromatic carbons and chlorine atoms, S_C and S_{Cl}, quantifying information on structural and electronic properties of the molecules [6], the $d\chi_n = {}^{p}\chi_n - {}^{p}\chi^{v}_n$ for n=0 to 10 as descriptors for strictly electronic properties [5-6,25] and the number of ortho-substituted Cl atoms, as a simple descriptor for the degree of coplanarity in the CB-congeners, respectively. The empirical descriptors are calculated using the programme Molconn-Z [24].

Non-Empirical and Quantum Chemical Descriptors

The quantum chemical descriptors used comprise energies of frontier orbitals, E_{HOMO} and E_{LUMO}, the second lowest and second highest MO energies, E_{NHOMO} and $E_{NLUMO,}$ the Debye dipole moment, the twist-angle between the two phenyl rings, Angle, heat of formation, dH_f, solvent accessible area and volume, A_{sas} and V_{sav} and van der Waals molecular surface area and volume, A_{vdW} and V_{vdW}, respectively. The quantum chemical descriptors are obtained from optimisations run by the commercial software HyperChem [23]. The solvent accessible area and volume, and van der Waals molecular surface area and volume, were calculated by ChemPlus included in HyperChem [23].

INFORMATIONAL CONTENTS OF THE EM AND QC-NEM X-MATRICES

The capability of the EM and the NEM-QM descriptors in quantifying variations in structural and electronic properties of the 209 PCB congeners are evaluated through principal component analysis (PCA). The main objective is to investigate the informational content of the empirical descriptors in reference to the physically interpretable NEM-QC descriptors. Results are shown in Figure 1.

As illustrated in Figure 1, there are several similarities between the PC-analysis based on the EM and NEM-QC descriptors, respectively. The projection of the descriptors onto the 2-dimensional space spanned by PC1 and PC2 captures 75 % and 7% of the original variance of the EM descriptors, and 77 % and 9 % of the original variance of the NEM-QC descriptors. Both the EM and the NEM-QC descriptors contain class variables grouping the PCBs into clusters of different degree of Cl-substitution. This can be seen from the score plots of PCBs described by the EM descriptors (Figure1, A1), and NEM-QC descriptors (Figure1, B1). Class variables are descriptors which groups the chemical structures into clusters, and is normally used for classification of chemicals into group-memberships before further analysis. However, the objective of this study is to model fate and effect related parameters for of all 209 PCBs in the same model, and for this reason it is of outmost importance to obtain an optimal homogeneous span of the PC-space. The latter is obtained by identifying and excluding class variables from the X-matrix.

PCA based on the NEM-QC Descriptors

For the NEM-QC descriptors the class variables are the molecular orbital energy descriptors (E_{HOMO}, $E_{LUMO,}$ E_{NHOMO} and E_{NLUMO}) which show insignificant within-group variance compared to between-group variance. By eliminating these descriptors, the PC-space gets more homogeneously

spanned with less grouping as shown in Figure 2.

A homogeneous span of the PC-plane is obtained by excluding the most significant class variables. The van der Waals surface and volume descriptors are excluded to reduce groupings according to increasing number of Cl-substituents. The twist-angle, which is restricting the overlap between the pi-orbital clouds of the two aromatic ring systems, adopt values of close to 40° for non-ortho substituted PCB, close to 60° for mono-ortho substituted PCB and close to 90° for di, tri and tetra-ortho substituted PCB. There is no significant correlation between the Angle, and the remaining descriptors, i.e. the dipole moment, dH_f, E_{tot}, A_{SAS} and V_{SAV}, and accordingly this descriptor reduces score-groupings in the plane spanned by PC1 and PC2.

The explained variance in horizontal direction, PC1, is 72 %, while the explained variance of PC2 is 17 %. The significance of the dipole moment is reflected in the score plot (upper plot, Figure 2). All the symmetrically substituted chlorinated biphenyl (CB) congeners have high negative scores in PC2, i.e. CB15, CB54, CB80, CB153, CB155, CB169, and CB209. The dipole moment increases with increasing PC2 score value, and in the top of the score plot CB125, with the highest dipole moment is placed. From left to right in horizontal direction the surface are and volume is increasing with increasing number of Cl-substituents. Last the twist-angle is increasing from the fourth to the second quadrant in overlapping areas parallel to the full-drawn bold line shown in the score plot (upper plot, Figure 2).

PCA based on the EM Descriptors

First step was to exclude all “degree of Cl-substitution” class variable from the data set. These are S_{Cl}, S_C, ${}^p\chi_0$, ${}^p\chi^v_0$, $d\chi_0$, ${}^p\chi_1$, ${}^p\chi^v_1$, ${}^p\chi_2$, and ${}^p\chi^v_2$. Second step was to exclude variables with insignificant within group variance ($\sigma^2 < 0.6$), and, thirdly, we excluded variables with insignificant variance when grouped by “number of ortho-Cl substituents”. By this approach, analogue to the optimisation

process of the physical interpretable NEM-QC descriptors, the result of the PC analysis developed as shown in Figure 3.

The explained variance in horizontal direction, PC1, is 74 %, whereas the explained variance of PC2, on the abscissa, is 11 %. Thus, a total explained variance of 85 % compared to 89 % for the NEM-QM descriptors in Figure 2. The twist-angle, which in the EM space, is described by the number of ortho Cl-substituents, $orthoCl_{tot}$, is increasing when going from the second to the fourth quadrant. The empirical descriptor $d\chi_1$ seems to be reflecting the MO-energy levels to some extent, the Pearson correlations varying between 0.63 and 0.92. The $d\chi_{2,}$ with highest loading in PC2, does to a minor extent reflect the variation in dipole moment, however low inverse correlations between ÷0.3 and ÷0.4 exists. The groupings in scores have been minimised, allowing for PLS regression without splitting into submodels according to group-memberships.

ANALYSIS OF VARIANCE

Both in traditional regression analysis and in multivariate projection methods the quality of experimental data used in model calibration are very important. Results on different endpoints vary between laboratories, as well as between methods used for obtaining the result. Most of the results included in the $logC_w^{sat}$ data set are measured by are shake-flask methods with UV and GC/ECD detection, generator column coupled with HPLC-UV detection, and slow-stirring methods [26]. Our data includes results from different methods and different laboratories, and therefore the standard deviations are increased compared to, e.g., standard deviation levels on physical-chemical properties of the PCBs obtained by Hansen et al. [27]. However, we need to be able to deal with the differences in results obtained from different sources, by eliminating outlier values and ensuring that the variation on the mean value is acceptable, i.e. low and homogeneous. For all of the 47 PCBs included in the solubility data set, the mean solubility and confidence interval (CI) of the

mean value at 95 % confidence level was calculated. Values exceeding the upper or lower CI were excluded from the data set. The mean variance on the solubility data, which include the variance between methods, was then calculated as

$$\sigma_{mean-within-PCB} = \frac{\sum_{h=1}^{47} \sigma_{within-PCB}}{47} \qquad \text{Eqn.1}$$

where h is the number of PCBs, and $\sigma_{within\text{-}PCB}$ is the variance of results for the individual PCBs calculated as:

$$\sigma_{within-PCB} = \frac{\left(x_{method} - \bar{x}\right)^2}{n-1} \qquad \text{Eqn.2}$$

where n is the number of measurements per PCB, $\bar{x}$ is the mean solubility for the individual PCBs, and x_{method} is the results obtained for the individual methods. The total number of measurements equals 243.

The mean variance given in Eqn.1 should be significantly lower than the variance between the mean values of each PCB, calculated as:

$$\sigma_{between-PCB} = \frac{\sum_{h=1}^{47} \left(\bar{x} - \bar{\bar{x}}\right)^2}{h-1} \qquad \text{Eqn.3}$$

where $\bar{\bar{x}}$ is the grand mean of all results. The degree of freedom, df, calculated F-value and critical F-value at 95 % confidence level, is given in Table I below.

Table I shows, that the variation in solubilities between PCBs is significantly higher than the variance between mean values. The latter variance, mainly caused by variations between methods and laboratories, is verified as insignificant compared to the variation in solubilities between the individual PCBs according to Table I.

In addition to the above variance analysis, if the variance within each PCB (cf. Eqn.2) exceeds the mean within variance, $\sigma_{within\text{-}PCB}$ is submitted to an F-test against the mean within variance (cf. Eqn.1) to secure a variance homogeneity of the input data. The results are given in Table II, and in causes where the F-value exceeds the critical F-value the PCB are left out of the data set.

The degrees of freedom for the mean within-PCB variance 242, and for each PCB n-1, where n is given in Table II. The calculated F-value exceeds the critical value in the case of CB3, CB194 and CB209, which are excluded from the data set.

The same kind of variance analysis was performed for the $\log K_{ow}$-values to insure the quality of data before further analysis.

In the following data analysis, data was mean centred and variable weighted by the individual standard deviation, to obtain unit variance of the individual variables before Projection into Latent Structures (PLS) regression.

RESULTS AND DISCUSSION

Solubility and Partitioning

If the interpreted informational content of the EM descriptors versus the NEM-QC descriptors are correct we would expect the variability in solubility between the PCBs to be described, most significant by the size ($^{p}\chi_3$, $^{p}\chi_7$, $^{p}\chi_9$, $^{p}\chi^{v}_3$, $^{p}\chi^{v}_7$, $^{p}\chi^{v}_9$, $^{c}\chi_3$) and shape ($^{pc}\chi_4$, $^{c}\chi^{v}_3$, $^{pc}\chi^{v}_4$) descriptors, and

secondary by the electronic property descriptors ($d\chi^v_1$, $d\chi^v_2$, $d\chi^v_3$, $d\chi^v_7$, $d\chi^v_9$). The results of the PLS regressions, based on the EM and NEM-QC descriptors, using the variance minimised mean $\log C_w^{sat}$ data are shown in Figure 4.

Both models are one-component models; i.e. PC2 shown in Figure 4 does not contribute to the estimated variance in solubility. The EM and NEM-QC models using 77 % and 79 %, respectively, of the variance in X to explain 93 and 90 % of the variation in the solubility, respectively. The informational contents of the EM contra NEM-QC descriptors can be evaluated qualitatively by comparing the correlation patterns between descriptors and the solubility in plot A2 and B2, Figure 4. As can be seen from A2 and B2 the descriptor $d\chi_1$, is related to the solubility in similar manner to the heat of formation and the total energy of the molecules. This supports the results of the PCA analysis, that the informational content of $d\chi_1$, which is independent of the size of the molecules, is quantifying information concerning the electronic properties of the molecules. The descriptor $d\chi_2$ shows some tendency to decrease with increasing dipole moment, however the Pearson correlation coefficient is insignificant. Lastly the simple and valence chi indices incode information concerning the structural properties of the molecules, and in analogy to the surface accessible area and volume, these descriptors have negative loading weights in PC1, i.e. are inversely related to the solubility. Dipole moment descriptors are insignificant in explaining the variability in solubilities in respect to the size and total energy descriptors for both models.

The score and loading plots of the models for estimating $\log K_{ow}$ are shown in Figure 5.

The correlation patterns between descriptors and octanol-water partitioning, $\log K_{ow}$, are reversed compared to the solubility models as expected. The partitioning increases with increasing size, and decreases with increasing energy of the frontier orbitals, the latter being expressed through $d\chi_1$ in the empirical model.

The insignificance of $d\chi_2$ in reflecting the dipole is revealed by, e.g. the high score of CB116 in Figure 5, A1. In general the significance of the dipole moment is unexpected low. The explanation is to be found in the degree of coplanarity, Angle, strongly influencing the effective charge delocalization between the two aromatic rings, i.e. the dipole moment is describing variations within classes which is suppressed in the models including all 209 PCBs.

Chromatographic Retention

The modelling of retention volumes by adsorption of selected non-ortho substituted CB-congeners resulted in the lowest calibration (r^2) and validation (q^2) correlation coefficients, and highest, RMSEP, root mean square error of predictions. The influence of the dipole moment and Angle is increased compared to the models of bulk properties. V_R increases with increasing polarity and decreases with increasing twist-angle between the biphenyl rings.

The EM model is a three component model (cf. Table III) using a total of 89 % X variance in explaining 90 % of the variation in retention volumes. The explained X and Y variance in the NEM-QC model was improved by including the descriptor Angle, the number of principal components included increasing from one to three. The X variance increased from 76 % to 100 % and the explained variance in V_r increased from 59 % to 89 %. Furthermore, the RMSEP was decreased by including two more PC´s (cf. Table III). The increase in principal components indicates an increase in complexity of the endpoint compared to the bulk properties $\log C_w^{sat}$ and $\log K_{ow}$. Increased model performance by including the twist-angle descriptor is not surprising since the interaction with the solid phase is strongly influence by the planarity of the CB-congeners.

As shown in plots B1 and B2, Figure 6, the retention by the adsorbent is strongly related to the twist angle between the two phenyls. The explained variance by PC2 is significant in both of the models, i.e., 46 % Y variance in the EM model and 23 % Y variance in the NEM-QC model. The inverse

relation between dipole moment and $d\chi_1$ is strong and verified by the score plot, i.e. the symmetrically substituted CB-congeners having high negative score values in PC2, while the CB-congeners with highest dipole moment have high positive score values in PC2. The higher significance of PC2 in the EM model relative to the NEM-QC model is to be found in the ability of the MCI descriptors to incode information concerning substitution patterns of the PCBs.

CYP1A Induction

The toxicity of the PCB congeners has been shown to depend on the substitution patterns, the more toxic PCB congeners being the non-ortho substituted coplanar CB congeners [22]. The X-space turned out to be best explained by including all of the original variables in both models. This is due to the small and equal number of CB-congeners from each class, which leaves the X-space homogeneously spanned as shown in Figure 7, A1 and B1. The data set consists of three tetra-, three penta-, two hexa- and one hepta-substituted PCBs, and the results of the PLS regressions are shown in Figure 7 below.

The X-loadings in PC1 and PC2 (61 % and 30 %, respectively) for the EM model, are used to explain a total of 93 % variance in EC50. For the NEM-QC model 79 % and 14 % are used to explain a total of 94 % variance in EC50. As seen from Figure 7 the twist-angle is the most significant descriptor as expected, and again the structural MCI descriptors show high specificity in describing differences in substitution patterns.

SUMMARY AND CONCLUSIONS

All of the multivariate data analysis in this study was performed by use of Unscrambler and Systat [28,29]. A summary of the performance of the individual models is given in Table III.

The study has revealed the significance and informational contents of the empirical descriptors. Despite the simplicity of the empirical descriptors, they do seem to incode information's concerning the electronic properties of the PCBs. Furthermore the ability to quantify substitution patterns seems to be significantly contribution in explaining endpoint which include complex interactions such as surface adsorption and enzyme induction processes. From models included in this study (cf. Table III) no indication of decreased model performance due to the simplicity of the empirical descriptors is observed.

References

[1] Hall, L.H., Kier, L.B., and Murray, W.J. (1975). Molecular connectivity II: Relationship to water solubility and boiling point. *J. Pharm. Sci.* **64**, 1974-1977.

[2] Kier, L.B., and Hall, L.H. (1976). *Molecular connectivity in chemistry and drug research.* Academic Press, New York, vol. 14, p. 257.

[3] Murray, W.J., Hall, L.H., and Kier, L.B. (1979). Molecular connectivity III: Relationship to partition coefficients. *J. Pharm. Sci.* **64**, 1978-1981.

[4] Hall, L.H., and Kier, L.B. (1984). Molecular connectivity of phenols and their toxicity to fish. *Bull. Environm. Contam. Toxicol.* **32**, 354-362.

[5] Kier, L.B., and Hall, L.H. (1990). The molecular connectivity of non-sigma electrons. *Rep. Mol. Theory* **1**, 121-125.

[6] Kier, L.B., and Hall, L.H. (1991). A differential molecular connectivity index. *Quant. Struct.-Act. Relat.* **10**, 134-140.

[7] Hall, L.H., and Stewart, D. (1994). Response surface analysis of bioconcentration by chlorinated organics using molecular connectivity. *SAR QSAR Environ. Res.*, **2**, 181-191.

[8] Kier, L.B., and Hall, L.H. (1999). The E-state in database analysis: the PCBs as an example. *Il Farm.* **54**, 346-353.

[9] Gough, J.D., and Hall, L.H. (1999). Modeling the toxicity of amide herbicides using the electrotopological State. *Environ. Toxicol. Chem.* **18**, 1069-1075.

[10] Kier, L.B., and Hall, L.H. (1999) *Molecular structure description. The Electrotopological state.* Academic Press, San Diego, p. 245.

[11] Koch, R. (1983). Molecular connectivity index for assessing ecotoxicological behaviour of organic compounds. *Toxicol. Environ. Chem.* **6**, 87-96.

[12] Bahnick, D.A., and Doucette, W.J. (1988). Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals. *Chemosphere* **17**, 1703-1715.

[13] Nirmalakhandan, N.N., Speece, R.E. (1989). Prediction of aqueous solubility of organic chemicals based on molecular structure. 2. Application to PNAs, PCBs, PCDDs, etc. *Environ. Sci. Technol.* **23**, 708-713.

[14] Hong, H., Wang, L., and Han, S. (1996). Prediction adsorption coefficients (K_{oc}) for aromatic compounds by HPLC retention factors (k'). *Chemosphere* **32**, 343-351.

[15] Pogliani, L. (1996). Modelling with special descriptors derived from a medium-sized set of connectivity indices. *J. Phys. Chem.* **100**, 18065-18077.

[16] Pogliani, L. (1996). Modeling purines and pyrimidines with the linear combination of connectivity indices-molecular connectivity "LCCI-MC Method. *J. Chem. Inf. Comput. Sci.* **36**, 1082-1091.

[17] Dowdy, D.L., and McKone, E.T. (1997). Predicting plant uptake of organic chemicals from soil or air using octanol/water and octanol/air partition ratios and molecular connectivity index. *Environ. Toxicol. Chem.* **16**, 2448-2456.

[18] Landrum, P.F., Fisher, S.W., Heang, H., and Hickey, J. (1999). Hazard evaluation of ten organophosphorus insecticides against the midge, Chironomus Riparius via QSAR. *SAR QSAR Environn. Res.* **10**, 423-450.

[19] Thomsen, M., Hvidt, S., and Carlsen, L. (2000). Solubility of phthalates revisited. Environmental implications. In, *Handbook on QSARs for predicting Environmental Fate of Chemicals* (J.D.Walker, Ed.). Society of Environmental Toxicology and Chemistry, Pensacola, FL, in press.

[20] Bersuker, I.B., and Dimoglo, A.S. (1991). The electron-topological approach to the QSAR problem. In, *Reviews of computational chemistry*, vol. 2, (D. Boyd and K. Lipkowitz, Eds.). VCH Publishers, New York, pp. 423-460.

[21] Storr-Hansen, E., Cleemann, M., Cederberg, T., and Jansson, B. (1992). Selective retention of non-ortho substituted coplanar chlorinated biphenyl congeners on adsorbents for column chromatography, *Chemosphere* **24**, 323-333.

[22] Van der Burght, A.S.A.M., Tysklind, M., Andersson, P.L., Horbach, G.J., and van der Berg, M. (2000). Structure dependent induction of CYP1A by polyhlorinated biphenyls in hepatocytes of male castrated pigs. *Chemosphere* **41**, 1697-1708.

[23] Hypercube Inc. (1995). *Hyperchem 4.5 for SGI*, Gainsville, FL, U.S.A.

[24] Hall Associates Consulting (1999). *Molconn-Z 3.50,* Quincy, MA, U.S.A.

[25] Thomsen, M., Rasmussen, A.G., and Carlsen, L. (1999). SAR/QSAR approaches to solubility, partitioning and sorption of phthalates. *Chemosphere* **38**, 2613-2624

[26] Mackay, D., Shiu, W. Y., and Ma, K. C. (1992). *Illustrated Handbook of physical-chemical properties and environmental fate for organic chemicals.* Vol.1, Monoaromatic Hydrocarbons, Chlorobenzenes, and PCBs. Lewis Publishers, London, p. 697.

[27] Hansen, B.G., Paya-Perez, A.B., Rahman, M., and Larsen, B.R. (1999). QSAR's for K_{ow} and K_{oc} of PCB congeners: A critical examination of data, assumptions and statistical approaches, *Chemosphere* **39**, 2209-2228.

[28] CAMO ASA (1998). *The Unscrambler 7.01,* Oslo, Norway.

[29] SPSS Inc. (1996). *Systat 6.0 for Windows,* Chicago, USA.

Table I Variances within and between individual PCB solubility data.

Type	σ^2 *(type)*	*df*	*(df1,df2)*	*F-ratio*	*F-critical*
mean variance within each PCB, Eqn.1	0.2	242	(46, 242)	8.0	1.6
variance between mean PCB´s, Eqn.3	1.7	46			

df:degrees of freedom

Table II Variances in solubility data found in the literature.

PCB	$logC_w^{sat}$	σ^2	*F-value*	$F_{critical}$	*n*
biphenyl	0.940	0.09	0.43		12
CB1	0.693	0.00	0.02		4
CB2	0.258	0.04	0.19		4
CB3	-0.324	0.62	2.91	2.54	10
CB4	-0.174	0.38	1.80	2.77	8
CB5	-0.046	0.02	0.10		7
CB8	-0.330	0.23	1.07	4.76	3
CB9	-0.090	0.15	0.72		6
CB10	0.103	0.07	0.34		4
CB15	-1.234	0.01	0.06		11
CB18	-1.013	0.30	1.40	2.54	10
CB26	-0.665	0.01	0.04		2
CB28	-0.831	0.04	0.17		10
CB29	-0.877	0.02	0.08		6
CB30	-0.648	0.00	0.02		4
CB31	-1.043	0.01	0.03		3
CB40	-1.547	0.06	0.26		3
CB44	-1.103	0.07	0.35		4
CB47	-1.220	0.01	0.02		2
CB52	-1.482	0.06	0.29		13
CB53	-1.413	0.51	2.40	3.15	6
CB60	-1.406	0.01	0.06		5
CB61	-1.776	0.02	0.08		10
CB66	-1.550	0.12	0.55		6
CB77	-2.707	0.94	4.43	3.15	6
CB80	-3.075	0.05	0.26		2
CB86	-1.903	0.11	0.50		7
CB87	-1.964	0.09	0.43		5
CB88	-1.913	0.00	0.00		4
CB95	-1.773	0.03	0.12		3
CB99	-2.195	0.11	0.52		2
CB100	-2.055	0.02	0.08		2
CB101	-2.111	0.22	1.05	2.32	12
CB116	-2.187	0.07	0.35		6
CB128	-3.095	0.12	0.58		6
CB129	-2.720	0.18	0.87		3
CB134	-2.565	0.45	2.12	6.82	2
CB136	-2.523	0.17	0.82		3
CB153	-2.677	0.23	1.07		6
CB155	-2.938	0.11	0.53		5
CB171	-2.670	0.23	1.06	4.76	3
CB185	-3.325	0.00	0.00		2
CB194	-3.085	0.79	3.71	3.93	4
CB202	-3.426	0.14	0.65		5
CB206	-4.255	0.12	0.57		4
CB208	-4.490	0.13	0.59		2
CB209	-4.235	2.62	12.30	2.82	4

Table III Model performance parameters.

		NEM-QC models				*EM models*			
Endpoints	*n*	*n PC´s*	q^2	r^r	*RMSEP*	*n PC´s*	q^2	r^r	*RMSEP*
$logC_w^{sat}$	42	1	0.95	0.97	0.311	1	0.96	0.96	0.364
$logK_{ow}$	44	1	0.93	0.94	0.405	3	0.95	0.96	0.274
V_R	16	1	0.79	0.67	11.523	3	0.87	0.95	9.143
V_R[a]	16	3	0.83	0.94	8.683				
EC50EROD[b]	19	3	0.88	0.99	1.821	2	0.85	0.96	1.824
EC50MROD[b]	20	2	0.98	0.99	0.524	2	0.97	0.99	0.946

[a]The descriptor Angle included
[b]Class specific descriptors included

FIGURE CAPTIONS

Figure 1 Score and loading plots represented by, on the ordinates, the first principal components (PC1´s), and, on the abscissa´s, the second principal components (PC2´s). A1 : the PCB scores, and A2: the loadings of original EM-descriptors. B1: the PCB scores and B2: the loadings of original NEM-QC descriptors. The score plots, A1 and B1, where the individual CB congeners are denoted by the number of ortho-substituted Cl-atoms on the two phenyl-rings, shows that variation in the degree of ortho-substitutions vary, to some extent, independent of the total number of Cl-atoms within the biphenyls as quantified by PC2. The within group variation in the number of ortho-substituted Cl-atoms is quantified in PC2 by the total number of ortho-substituted Cl-atoms, ortoCltot, in the empirical model (A2). In the NEM-QC model the change in dipole moment is the most significant descriptor in PC2, i.e. the total substitution pattern of the biphenyl is more pronounced in describing the variations in electronic properties than the twist-angle in PC2.

Figure 2 Score plot of the 209 PCB´s, and loading plot of the NEM-QC descriptors excluding the most significant class variables.

Figure 3 Score of the 209 PCB´s and loading plot of EM descriptors excluding class variables.

Figure 4 Bi-plots showing the score and X-loading weights in the EM-model (A1) and the NEM-QC-mode (B1), respectively. Plots of X-loading weights with respect to the Y-loading for the EM descriptors and NEM-QC descriptors as shown in A2 and B2, respectively.

Figure 5 Bi-plots showing the score and X-loading weights in the EM-model (A1) and the NEM-QC-mode (B1), respectively. Plots of X-loading weights with respect to the Y-loading for the EM descriptors and NEM-QC descriptors as shown in A2 and B2, respectively.

Figure 6 Bi-plots showing the score and X-loading weights in the EM-model (A1) and the NEM-QC-mode (B1), respectively. Plots of X-loading weights with respect to the Y-loading for the EM descriptors and NEM-QC descriptors as shown in A2 and B2, respectively.

Figure 7 Bi-plot showing the score and X-loading weights in the EM-model (A1) and the NEM-QC-mode (B1), respectively. Plot of X-loading weights in respect to the Y-loading for the EM descriptors (A2), and N *(A2), and NEM-QC descriptors (B2).*

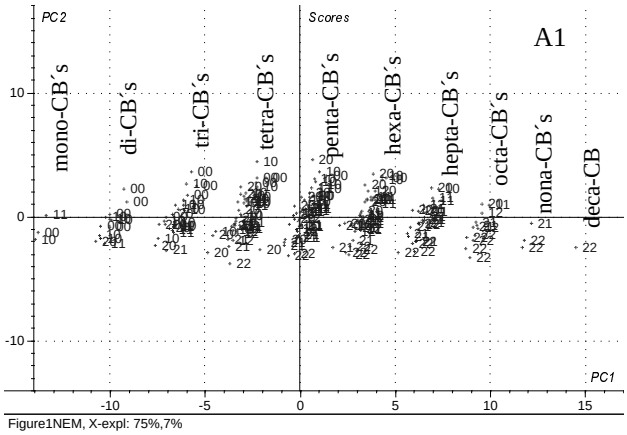


Figure1NEM, X-expl: 75%,7%

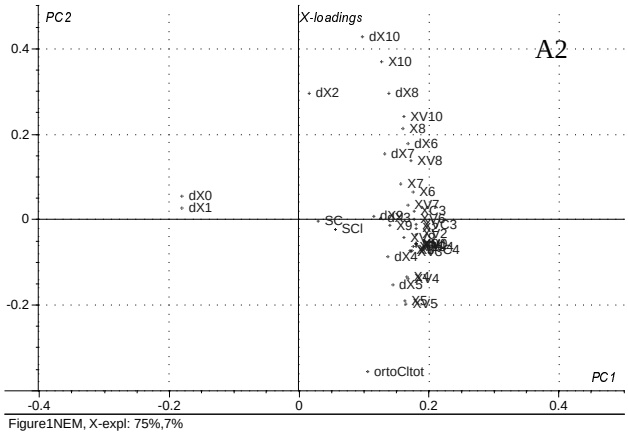


Figure1NEM, X-expl: 75%,7%

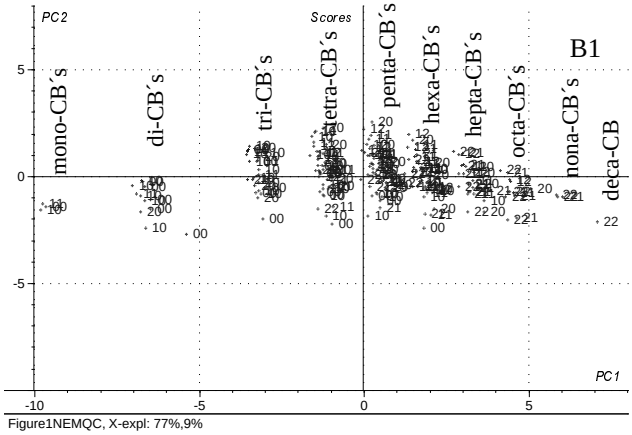


Figure1NEMQC, X-expl: 77%,9%

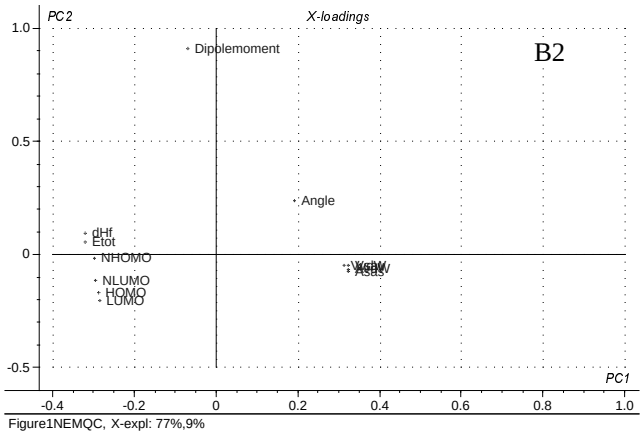
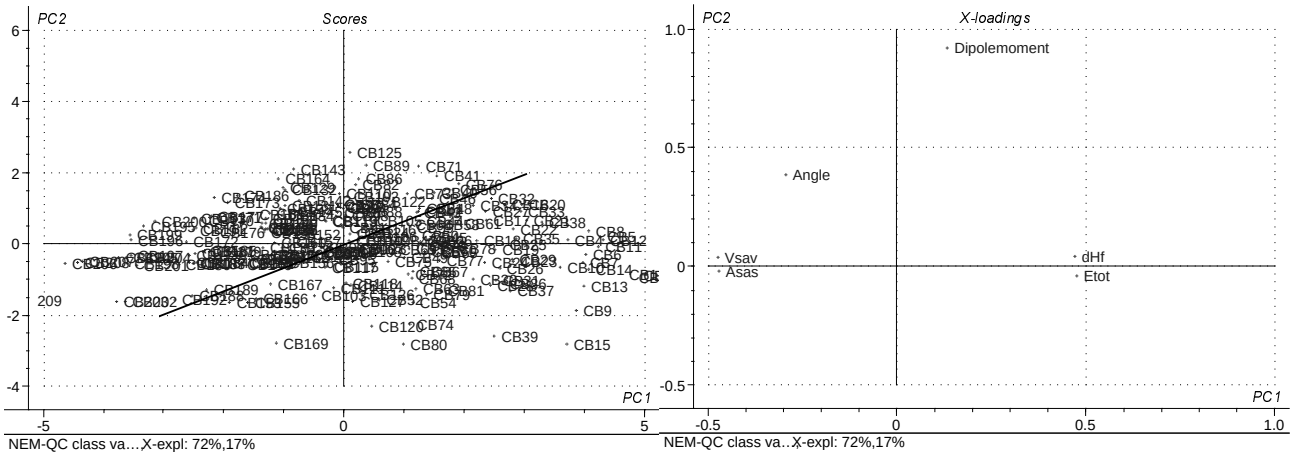
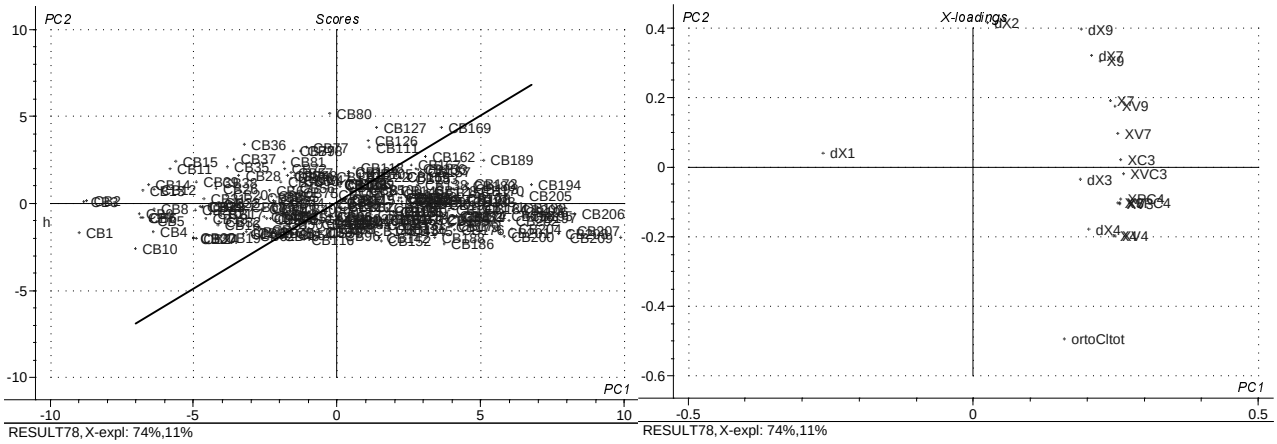


Figure1NEMQC, X-expl: 77%,9%

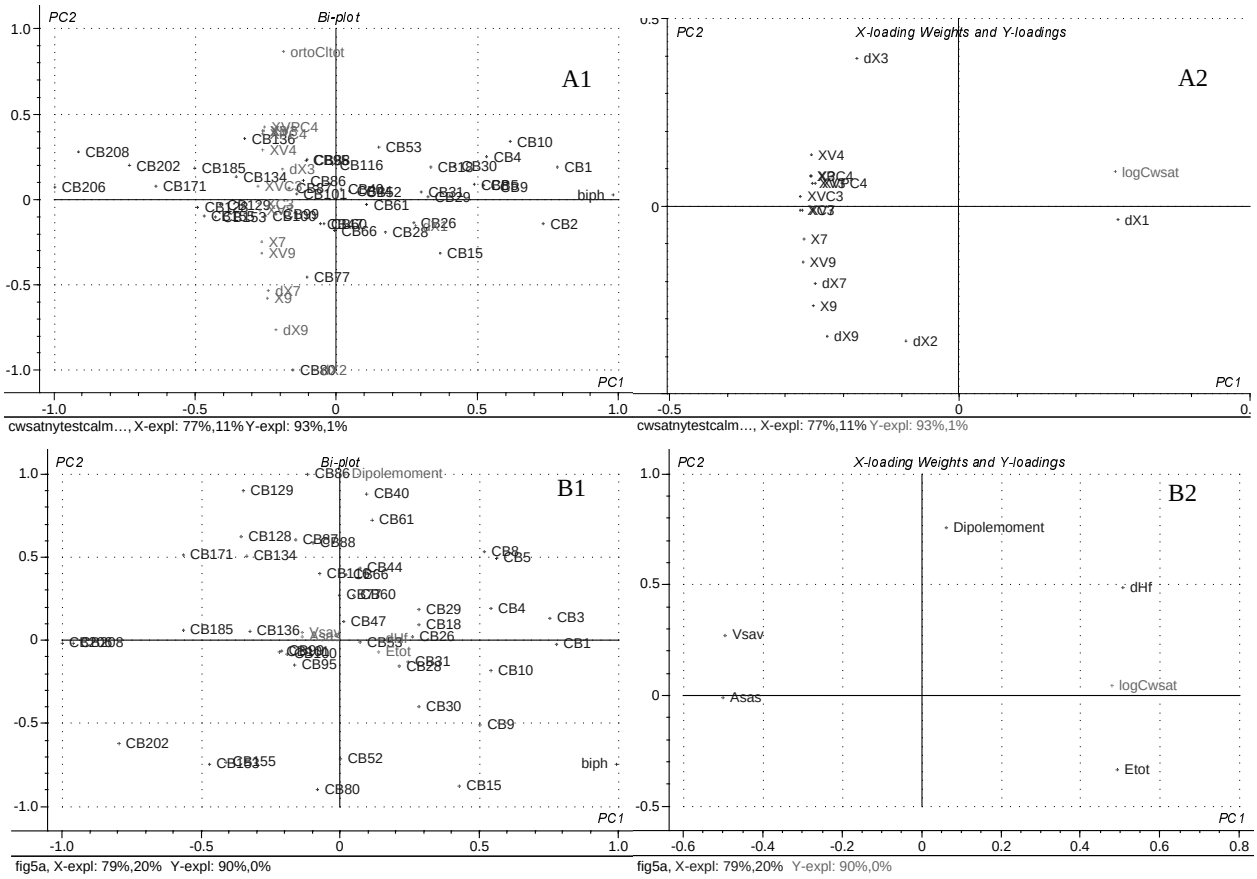
Paper III. EVALUATION OF EMPIRICAL VERSUS NON-EMPIRICAL DESCRIPTORS,FIGURE 2



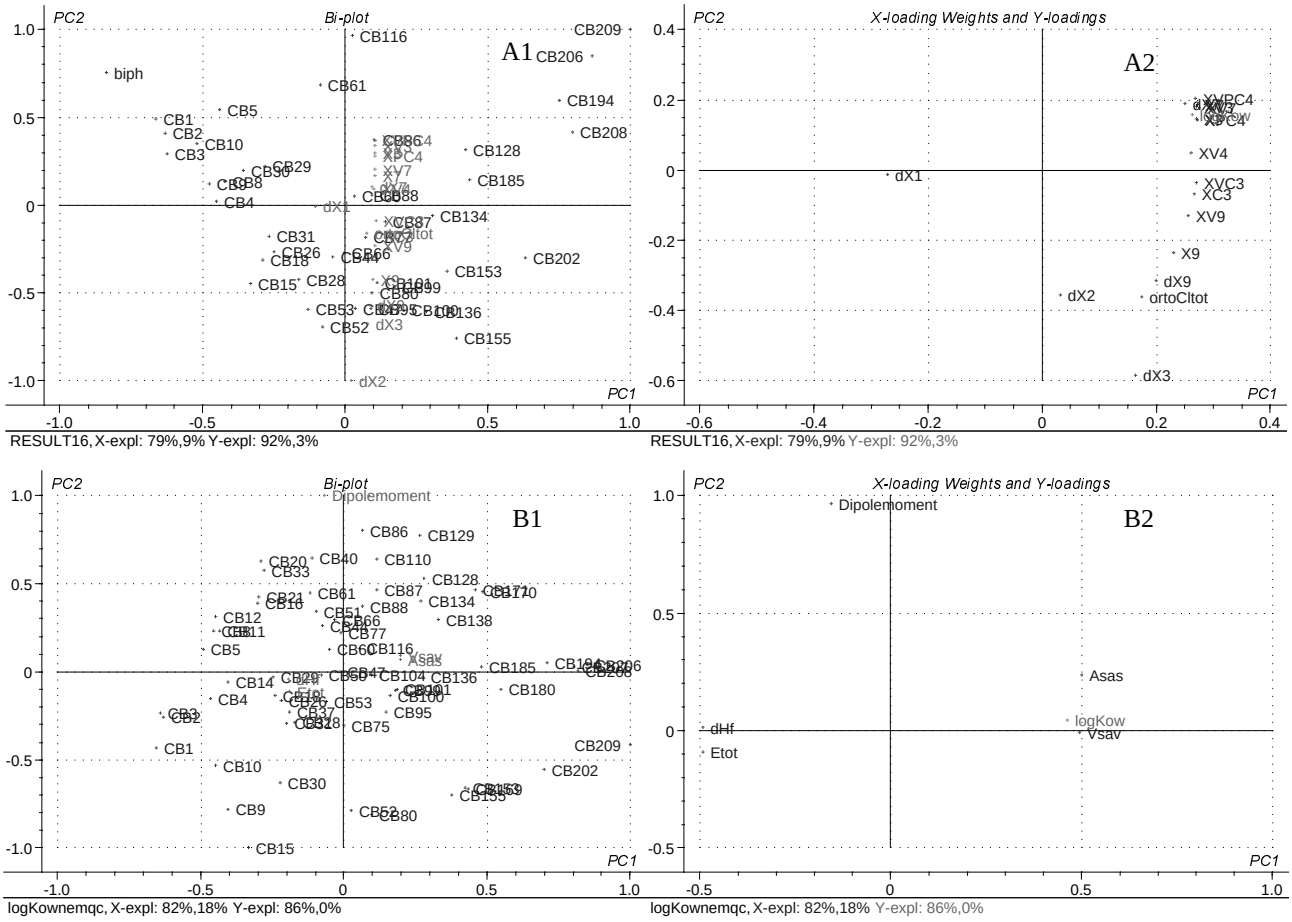
Paper III. EVALUATION OF EMPIRICAL VERSUS NON-EMPIRICAL DESCRIPTORS,FIGURE 3

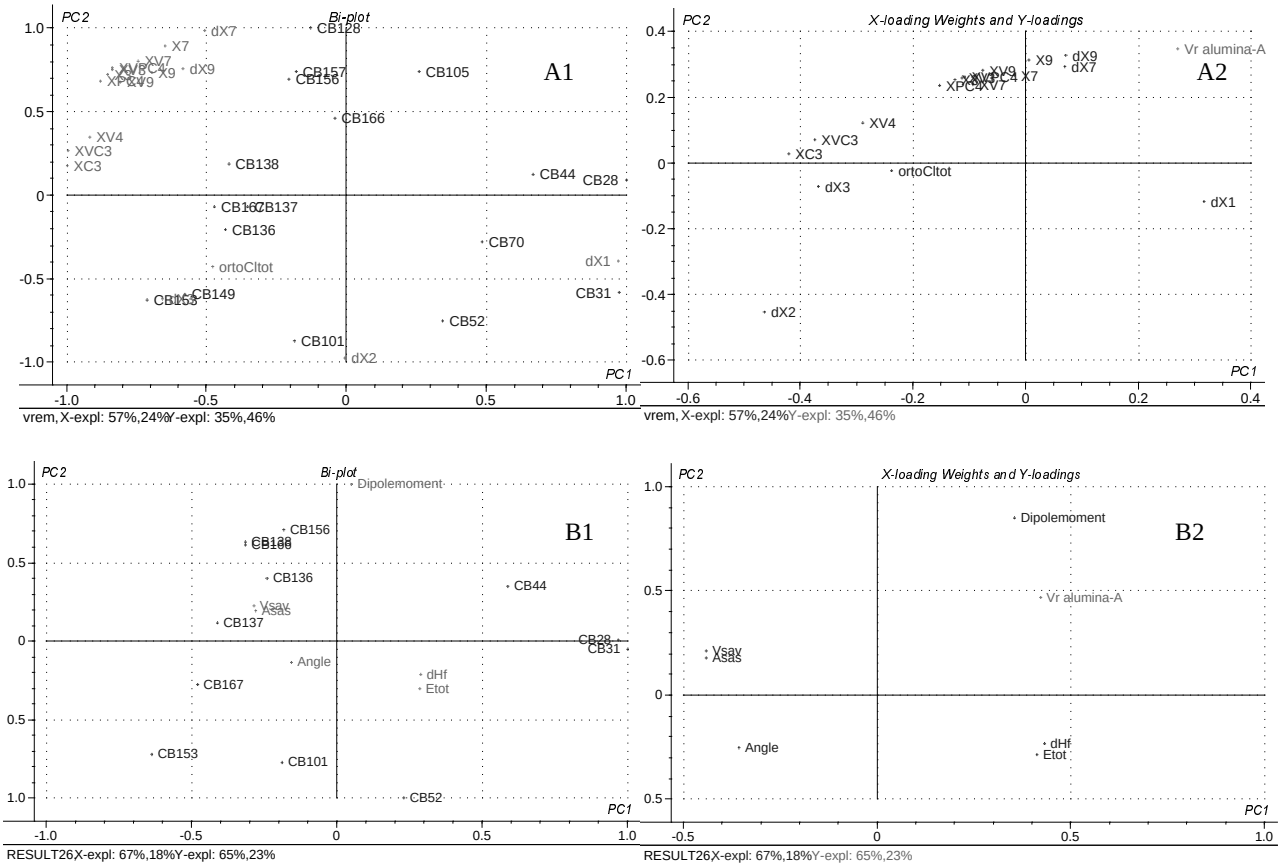


Paper III. EVALUATION OF EMPIRICAL VERSUS NON-EMPIRICAL DESCRIPTORS,FIGURE 4

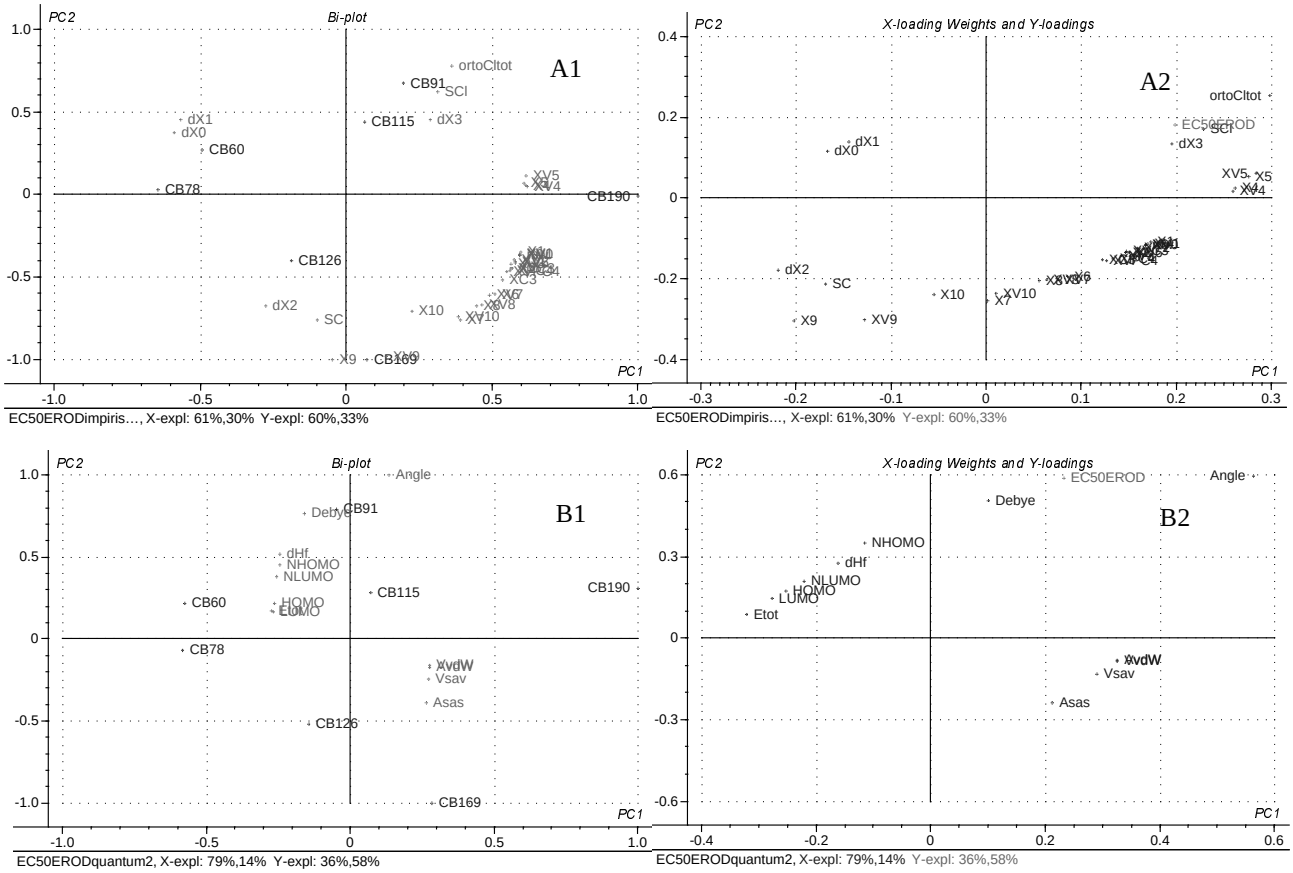


Paper III. EVALUATION OF EMPIRICAL VERSUS NON-EMPIRICAL DESCRIPTORS,FIGURE 5




PC2
Bi-plot
dX7
CB128
X7
XV7
XPC4
X9
dX9
CB157
CB156
CB105
A1
CB166
XV4
XVC3
XC3
CB138
CB44
CB28
CB167
CB137
CB136
CB70
dX1
ortoCltot
CB149
CB153
CB31
CB52
CB101
dX2
PC1
vrem, X-expl: 57%,24%Y-expl: 35%,46%
PC2
X-loading Weights and Y-loadings
Vr alumina-A
X9
dX9
dX7
XV9
XPC4
XV7
A2
XV4
XVC3
XC3
ortoCltot
dX3
dX1
dX2
PC1
vrem, X-expl: 57%,24%Y-expl: 35%,46%
PC2
Bi-plot
Dipolemoment
CB156
CB136
CB44
Vsav
Asas
CB137
B1
CB28
CB31
Angle
dHf
Etot
CB167
CB153
CB101
CB52
PC1
RESULT26X-expl: 67%,18%Y-expl: 65%,23%
PC2
X-loading Weights and Y-loadings
Dipolemoment
Vr alumina-A
B2
Vsav
Asas
Angle
dHf
Etot
PC1
RESULT26X-expl: 67%,18%Y-expl: 65%,23%

Paper III. EVALUATION OF EMPIRICAL VERSUS NON-EMPIRICAL DESCRIPTORS,FIGURE 7



EC50ERODimpiris..., X-expl: 61%,30% Y-expl: 60%,33%

EC50ERODimpiris..., X-expl: 61%,30% Y-expl: 60%,33%

EC50ERODquantum2, X-expl: 79%,14% Y-expl: 36%,58%

EC50ERODquantum2, X-expl: 79%,14% Y-expl: 36%,58%

Paper IV

Characterisation of humic materials of different origin:
A multivariate approach for quantifying the latent properties of dissolved organic matter.

Submitted to *Chemosphere*

Characterisation of humic materials of different origin: A multivariate approach for quantifying the latent properties of dissolved organic matter

Marianne Thomsen[a,*], Pia Lassen[a], Shima Dobel[b], Poul Erik Hansen[b], Lars Carlsen[d], and Betty Bügel Mogensen[a]

[a]*National Environmental Research Institute, Department of Environmental Chemistry, DK-4000 Roskilde, Denmark*

[b]*Danish Environmental Protection Agency, Office of Chemicals, Strandgade 29, DK-1401 Copenhagen-K, Denmark*

[c]*Department of Life Science and Chemistry, Roskilde University, DK-4000 Roskilde, Denmark*

[d]*Department of Environment, Technology and Social Studies, Roskilde University, DK-4000 Roskilde, Denmark*

*Corresponding author. Tel.: +45 +46 30 13 58; fax: +45 +46 30 11 14.

E-mail address: mth@dmu.dk (M. Thomsen)

Abstract

The inherent chemical properties of eight different dissolved organic matters (DOMs) originating from soil, surface- and groundwater have been analysed. The samples consist of isolated fulvic acids (FA), humic acids (HA), and humic substances (HS), i.e. natural mixtures containing a humic and a fulvic fraction. The humic substances have been characterised by elemental analysis (EA), size exclusion chromatography (SEC), E2/E3 and E4/E6 UV absorption ratios, and liquid-state ^{13}C-NMR spectroscopy. The information contents of the different analytical methods have been investigated by pattern recognition, i.e. cluster analysis (CA) and principal component analysis (PCA). A comparative study of the information contents of DOM descriptors derived from different analytical methods is presented. Through extraction of information content of the individual analytical methods the inherent properties of DOM are quantified. Pattern recognition revealed significant quantitative differences in the inherent properties of DOM of different origin and type. PCA based on the NMR descriptors showed highest explained variance. However, all models showed low robustness due to the limited number of samples. The supervised pattern recognition indicates a classification of DOMs into groups of similar properties by an increase in the number of samples.

Keywords: Dissolved Organic Matter (DOM); UV-VIS; liquid-state ^{13}C-NMR spectroscopy (NMR); Size Exclusion Chromatography (SEC); Elemental Analysis (EA); Principal Component Analysis (PCA), Cluster Analysis (CA).

1. Introduction

Humic materials of different origin constitute a heterogeneous group of macromolecules with varying sorption capacity towards pollutants. The dissolved fraction of humic materials, i.e. dissolved organic matter (DOM), therefore affects the fate, and hence the exposure and potential effects, of pollutants to a varying degree. DOM is the major organic constituent in natural waterbodies (Holmström et al., 2000; Ran et al., 2000) and its presence in aqueous bulk phases of environmental compartments allows for third-phase effects (Kalbitz et al., 2000; Periago et al., 2000; Shen, 1999) or colloid-mediated transport of pollutants to occur. Impacts of the presence of DOM may be increased apparent solubility of pollutants of low polarity and thereby increased transport through the aquatic compartment. Another example is vertical colloid-mediated transport by binding to mobile fraction of organic matter, DOM, within the terrestrial compartment (Kaiser et al., 2000).

The equilibrium partitioning of hydrophobic organic pollutants between natural organic matter and an aqueous bulk phase have traditionally been estimated by the use of simple linear $\log K_{oc}$-$\log K_{ow}$ regression models using the octanolic phase as a model phase for natural organic matter (EC, 1996). The chemical domain of $\log K_{oc}$-$\log K_{ow}$ regression models is limited to simple hydrophobic pollutants for which the partitioning is driven mainly by the hydrophobic effect. However, for more complex organic pollutants, where specific interactions between the pollutant and the sorbent phase are pronounced, the electronic and structural characteristics of the sorbent, e.g. DOM, affects the equilibrium partitioning coefficients. In the latter case a quantification of the inherent properties of organic sorbent phase is needed to gain knowledge about organic sorbent properties parameters influencing the sorption capacity towards different pollutants.

Several studies focussing on the characterisation of structural properties of natural organic matter exist (Artinger et al., 1999; Ran et al., 2000; Zsolnay et al., 1999). However, a simple

quantitative description of the nature and diversity of DOM is missing. Furthermore, there is a need for an evaluation and harmonisation of the large number of characterisation methodologies in order to obtain simple and consistent methods for classification (Artinger et al., 1999; Davis et al., 1999) and quantification of the inherent properties of humic materials.

The objective of the present study is, by use of classical characterisation methodologies (Artinger et al., 1999; Hautala et al., 2000; Schulten, 1996; Steelink, 1985), to extract and evaluate relevant information for characterising the similarities and dissimilarities in the inherent properties of DOMs of different origin. The explanatory capability of descriptors, based on NMR, UV, SEC and EA data, for quantifying inherent properties DOMs are evaluated. Focus will be on the mobile, non-fixed, organic matter, i.e. DOM originating from soil pore water, as well as organic matter extracted from surface and groundwater. In spite of the relatively low number of samples, and varying uncertainties in experimental data, the present study presents an approach of how to explore and quantify the inherent properties of DOM.

2. Materials and methods

2.1 Humic substances

The origin of the individual humic substances included in the present study is given in Table 1.

2.2 Methods

2.2.1 Liquid-state ^{13}C-NMR

Liquid-state ^{13}C-NMR spectra were recorded in 0.5 M NaOD (50 mg in 0.5 ml) on a Bruker 250 MHz instrument. Spectral width set to 17 kHz, and pulse width 3.5 µs (45°). No NOE build up during the waiting period. Typically 70000 FIDs were collected. A line broadening of 80 Hz was applied. The spectra were referenced to external TMS.

The spectra were divided into the following regions, here given as averages of all spectra: 3(± 6)-52(±9) ppm, un-substituted aliphatic C (*AlkHC*); 55(±10)-91(±7) ppm, N-alkyl and methoxy C including a major fraction of carbohydrates (*AlkO*); 94(±8)-146(±8) ppm, un-substituted aromatic C (*ArHC*); 145(±5)-165(±6) ppm, aromatic O-substituted C (*ArO*); 165(± 6)-188(±2) ppm, carboxylic and ester C (*COO*); 190(±3)-214(±4), C atoms of quinonic and ketonic groups (*CO*). Variation in chemical shift regions is partly due to the complexity and heterogeneous nature of DOM, and partly because of the presence of paramagnetic inorganic species, which may affect both the width and shape of each peak (cf. section 3.1). The regions were integrated and the relative carbon content of each area was calculated. Peak widths were measured. ^{13}C-NMR was not performed on water pond HS due to limited amount of sample, nor on Aldrich HA (Na^+) due to the presence of inorganic impurities consisting mainly of Fe^{3+}, which prevents observation.

2.2.2 Size exclusion chromatography (SEC)

SEC was performed on solutions of the humic substances with a concentration 300 mg/l at pH 8.3 using a Sephadex® G-50 Medium gel (Code No. 17-0043-01 Pharmacia Biotech AB). The diameter of the column used was 13 mm, the height of gel 12 cm, and the flow rate was approximately 15 ml/h. The SEC elution profiles were monitored by UV absorbance at 400 nm.

2.2.3 UV-VIS spectroscopy and elemental analysis (EA)

UV-VIS spectroscopy of the solutions of humic substances was carried out using a Cary 50 UV-Visible Spectrophotometer. Concentrations of each humic substance solution were determined at 400 nm. UV-VIS spectra of the humic substances were recorded from 700 to 200 nm. The E2/E3 ratio (the absorbance at 250 nm divided by the absorbance at 365 nm) and the E4/E6 ratio (the absorbance at 465 nm divided by the absorbance at 665 nm) were calculated (Malcolm, 1989; Chen et al., 1977; Lassen et al., 1994). The absorptivity at 272 nm is used as descriptor quantifying the aromaticity of the samples.

EA of the humic substances was done on an EA 1110 CHNS analyser, using 5 mg of dry sample per measurement.

2.2.4 Cluster Analysis (CA)

CA was performed separately on SEC, UV and NMR data and compared to CA based on all data. In the CA based on all data, the descriptor variables were block-standardised by range prior to CA, to avoid any effects of scale of units on the distance measurements. Similarities-dissimilarities were quantified through Euclidean distance measurements, the distance between two objects (humic materials), *i* and *j,* is given as (Everitt, 1993):

$$d_e = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \qquad (1)$$

where d_e denotes the Euclidean distance, x_{ik} and x_{jk} are the values of variable k for object i and j, respectively, and m is the number of variables. The dendograms shown in Figure 3 are based on the method of complete linkage (Everitt, 1993). The most similar objects, i_1 and i_2, are united in one cluster i_{12} and the distance, d, of this cluster to all the remaining objects (humic samples) j is calculated as:

$$d(i_{12}, j) = 0.5d_{i1,j} + 0.5d_{i2,j} + 0.5\left|d_{i1,j} - d_{i2,j}\right| \qquad (2)$$

In the complete linkage method the distance between clusters are determined by the greatest distance between any two objects in different clusters.

2.2.5 Principal Component Analysis (PCA)

The main goal of PCA is a quantification of the significance of variables that explain the observed groupings and patterns of the inherent properties of the individual DOMs. Through a linear combination of the original property variables (measured characterisation properties) in the data matrix **X**, the property space is reduced and explained by a set of principal components (PCs). The PCA in matrix form is a least-square model and is expressed by

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{F} + \mathbf{E} \qquad (3)$$

X is the original data matrix, **F** the values of the object in the projection space, and **A** is the loadings of the original variables in the in hyperspace projected by the principal components (Höskuldsson, 1993) and **E** contains the residuals.

The PCs account for the maximum explainable variance of all original property parameters in a descending order, and are non-correlated.

$$PC_j = \alpha_{j1}x_1 + \alpha_{j2}x_2 + ... + \alpha_{jn}x_n \qquad (4)$$

The loadings, α, of each original characterisation variable (x_1 to x_n) in *PC* number *j*, reflects the importance of variable 1 to *n* in describing the patterns in scores in the direction of PC_J.

3. Results

3.1 Liquid-state ^{13}C-NMR Spectra

In addition to the peak area the peak width is included since, e.g., the width of the chemical shift range of O-substituted aromatic carbons depends on the substitution patterns of the aromatic system. Hence, the width of chemical shift zone in the NMR-spectra may be just as important as the intensity when quantifying the chemically and structurally characteristics of humic substances. Furthermore, integrated peak areas divided by the corresponding peak widths are included as a descriptor quantifying the shape of the individual peaks. Descriptors based on liquid-state ^{13}C-NMR measurements are given in Table 2.

The ^{13}C NMR spectra of the humic acids (Purified Aldrich HA, Kranichsee HA and Gohy-573-HA-(H^+)II), the fulvic acids (DE72 and FA-surface), and humic substances (Gohy-573-HS-(H^+)II) are shown in Figure 1.

The un-substituted aliphatic region (0-50 ppm), denoted *AlkHC*, is more pronounced for FA-surface compared to the remaining humic materials. Methoxy, carbohydrates mainly and N-alkyl groups in the region of 50-90 ppm, are denoted by *AlkO*. As seen from the spectra (Figure 1), Kranichsee HA has higher content of O- and/or N-substituted aliphatic carbons than the fulvic acids DE72 and FA-surface. This is atypical in respect to the general expectation of higher polarity of the fulvic acids compared to humic acids.

The aromatic region, often assigned by the region 110-160 ppm consists of a down-field signal region assigned to O-substituted aromatic substructures (*ArO)* (145-165 ppm), and lesser shielded aromatic C-nuclei region (*ArHC*). The content of un-substituted or C-substituted aromatics is highest for Gohy-573-HA-(H^+)II and Purified Aldrich HA. FA-surface is highly aliphatic and has virtually no detectable content of O-substituted aromatics. The fulvic acid fraction of Gohy-573-HS-(H^+)II, contains a significant amount of O-

substituted aromatics observed by comparing with a content of only 8 % in the isolated HA-fraction, i.e. Gohy-573-HA-(H^+)II.

Typically the most obvious difference between fulvic and humic acids is the large fraction of carbohydrates (*AlkO*) of the former. This is in agreement with a high content of carbohydrates in the fulvic acids DE72 and FA-surface compared to the humic acids Gohy-573-HA-(H^+)II and Purified Aldrich HA. Kranichsee HA originates from interstitial soil water of a raised bog, characterised by a slow humification process, which may explain the high content of carbohydrate or O-substituted carbons.

The content of carboxyl carbons, which include carboxylate ions, is highest for the fulvic acids, whereas the variability and contents of carboxy groups of esters and amides are low for all humic materials.

3.2 Size Exclusion Chromatography (SEC)

SEC was used to separate the single samples of humic substances by size. The Sephadex G50 column material has a fractionation range of 1,500-30,000 Da for globular proteins and 500-10,000 Da for dextrans. Since humic materials are more spherical than dextrans, the fractionation range for humic materials is possibly between dextrans and globular proteins with a higher similarity to dextrans. No external calibration standards for fulvic acids or humic acids were used, and the SEC results given in Figure 2 are therefore qualitative. However, relative high and low average molecular weight (MW) fractions of each sample were quantified by normalising to the total peak area (data given in Carlsen et al., 2000).

For humic substances and humic acids two well-separated elution peaks are observed. The low average MW fractions are retained on the column and separated according to size distribution. The high average MW fractions, with a molecular weight above 10,000 Da, are eluted with the eluent front. The SECs of the fulvic acids are different. FA-surface only

displays the low MW fraction peak, whereas a minor shoulder in the high MW elution range is observed for DE72. The missing resolution of the two average MW fractions of DE72 may be caused by a high degree of intermolecular hydrogen bonding network between humic monomeric units (Conte and Piccolo, 1999), caused by the significant higher content of carboxylic groups in DE72 compared to the other HSs.

The shape and size of the high MW elution peaks of Gohy-573-HS-(H^+)II versus Gohy-573-HA-(H^+)II and Aldrich HA (Na^+) versus Purified Aldrich HA, reveal that the nature of the high average MW fraction is significantly influenced both by fractionation and purification of the humic samples. The approximate 50 % decrease in intensity of the high average MW fraction in the Purified Aldrich HA compared to the non-purified Aldrich HA (Na^+), indicates a significant influence of inter- and intra-molecular metal ion-DOM bindings on the structural characteristics of the Aldrich HA (Na^+) (Robertson and Leckie, 1999). Aggregation of humic materials is promoted by the presence of positive ions in the solution, and the effect is amplified in cases where the salts contain divalent metal ions (Ragle et al., 1997; Kim et al., 1990). The ash content of Aldrich HA (Na^+) is close to 10 %, with major constituent being divalent ions such as Mg^{2+}, Fe^{2+}, Ca^{2+}, and Si^{2+} (Kim et al., 1990). Therefore, an increased amount of aggregated structures in Aldrich HA (Na^+), compared to Purified Aldrich HA, appears to be a very likely explanation for an increased intensity of the high average MW peak of Aldrich HA (Na^+). The same pattern, although less pronounced, is observed for Gohy-573-HA-(H^+)II compared to Gohy-573-HS-(H^+)II.

Qualitative evaluation of the average MW derived from the SECs are as expected: Fulvic acids < humic substances < humic acids. As FA-surface has no high average MW elution peak, this fulvic acid must be considered as having the lowest average MW compared to the other humic samples. The missing high MW peak for FA-surface is supported by the NMR results, as these show high aliphaticity, high acidity and insignificant amounts of aromatic

sub-units, e.g. leading to increased flexibility and thereby curling-up of the macromolecular colloids (Tombácz, 1999).

3.3 UV-VIS and EA data

The EA data are based on conventional C, H, N- analysis (the content of oxygen is found as the difference from 100 % even in cases in which ash is present) (Steelink, 1985). Due to significant ash content of some samples EA data was not included in the pattern recognition analysis (cf. section 4). Elemental analysis and the UV-spectroscopic data are given in Table 3.

Generally, the E_4/E_6 ratio is expected to decrease with increasing MW and content of condensed aromatic rings (Malcolm, 1989). In addition the ratio is expected to increase with an increase in oxygen contents (Chen et al., 1977). The E_4/E_6 ratios for FA-surface, Gohy-573-HS-(H^+)II, Purified Aldrich HA, and Aldrich HA (Na^+), are decreasing with increasing molecular weight of the high average MW fraction, whereas the opposite is seen for the low average MW fraction (Carlsen et al., 2000). The E_4/E_6 measurements indicate also that the Water pond HS sample should have the lowest MW range, which is in disagreement with the SEC data, but high E_4/E_6 ratio and H/C ratio are in agreement with a low aromaticity of the Water pond HS. However, this could not be checked by NMR data, due to lack of sample material to record the ^{13}C-NMR spectrum of the Water pond HS.

3.4 CA

The ambiguity of the quantified characterisation data is reflected in Figure 3, showing the results of analysing the similarities-dissimilarities between the humic materials based on SEC, NMR, UV data separately and all data, respectively.

3.4.1 CA based on NMR data

The dendogram based on NMR data (cf. Figure 3a) unites the two humic acids Purified Aldrich HA and Gohy-573-HA-(H^+)II by the lowest distance. In spite of the same origin Gohy-573-HA-(H^+)II is classified more similar to Aldrich HA (Na^+) than to Gohy-573-HS-(H^+)II. This indicates the significant change in inherent properties by the presence of a polar fulvic acid fraction. DE72 and Kranichsee HA have the same degree of similarity as Purified Aldrich HA and Gohy-573-HA-(H^+)II, whereas the distance between the two clusters are significant. These four humic substances are linked together through Gohy-573-HS-(H^+)II, by a small distance, indicating intermediate characteristics of Gohy-573-HS-(H^+)II in respect to the four sub-clustered humic substances. FA-surface is dissimilar to all humic substances, most pronounced compared to Gohy-573-HS-(H^+)II. This is due to the low degree of aromaticity and very high aliphaticity of FA-surface (cf. Table 2).

3.4.2 CA based on SEC data

The dendogram based on SEC data (cf. Figure 3b) links Gohy-573-HS-(H^+)II and Purified Aldrich HA by shortest distance, i.e. indicating highest degree of similarity between these in respect to the remaining sub-cluster humic materials. The relatively high linking distance of FA-surface and Water pond HS indicates significant differences in inherent properties, even though they are in the same sub-cluster. The low average MW of DE72 and Kranichsee, are identified by the high linking distance to the remaining clustered humic materials.

3.4.3 CA based on UV data

The UV data (cf. Figure 3c) unites Water pond HS and FA-surface, Kranichsee and DE72, respectively, in the same clusters. These cluster-groupings are similar to the results of the SEC data. However, the degree of similarity differs markedly in the two dendograms.

Aldrich HA (Na^+) is clustered together with Aldrich HA and Gohy-573-HA-(H^+)II, in respect to the content of chromophores, most similar to the latter. Similar Gohy-573-HS-(H^+)II show properties most similar to DE72 and Kranichsee. This indicates that the fulvic acid fraction of Gohy-573-HS-(H^+)II is significant, i.e. changes the properties significantly with respect the isolated HA fraction (i.e. Gohy-573-HA-(H^+)II). This change in cluster patterns in respect to the dendogram based on NMR data could be due to Aldrich HA (Na^+), on which no ^{13}C-NMR spectra could be obtained.

3.4.4 CA based on all data

In spite of missing data, as well as data of varying quality, the dendogram based on all data (cf. Figure 3d) seems reasonable, and most significantly supported by the NMR data.

Furthermore, it should be noted that none of the characterisation data results in groupings of the two conventionally defined (cf. Table 1) fulvic acids, DE72 and FA-surface, in the same cluster. As unsupervised pattern recognition, e.g. CA, does not reveal information concerning the causes to observed similarities-dissimilarities, the cluster patterns can only be compared to the conventional definition of humic acids, fulvic acids and humic substances, as defined in Table 1. The dendograms in Figure 3, however, indicates that the inherent properties of DOM may be more varied, than grouping according to origin and fractionation procedure. Classification by the conventional method could therefore be inadequate. The dendograms based on the UV and SEC data indicate a significant similarity of Water pond HS and FA-surface as seen from Figures 3b, 3c and 3d, which indicates that the inherent properties of Water pond HS is closest to average fulvic acids properties if such an average exists.

The similarity between Kranichsee HA and DE72 in all dendograms is attributed to the high content of carbohydrates in Kranichsee HA, which indicates that the definition of “type”, i.e. humic acid, fulvic acid and humic substance (cf. Table 1), according to extraction procedures is insufficient. This reflects the inadequacy in the general statements of differences in

properties of humic and fulvic acids, i.e. by high/low size, polarity, aromaticity, aliphaticity. These general qualitative classification indices may not be valid as aquatic fulvic acids originating from surface runoff may be a class in between FA of aquatic origin and HA.

3.5 PCA

Figure 4 illustrates the results of a PCA model including all types of NMR variables (cf. Table 2). The figure is a bi-plot of the two most significant variables of the projection space, i.e. principal component number two, PC_2, as function of principal component number one, PC_1. The bi-plot includes the scores of the individual humic materials and the loadings of each original variable in PC_1 and PC_2.

The model explains 76 % of the variance in the original X space, and illustrates the usefulness of PCAs for getting a simple summary of the correlation patterns of the descriptors, as well as the overall span in characteristics of the humic materials.

In the horizontal direction, i.e. PC_1, the aromaticity is decreasing from left to right, which is seen from the large and positive loadings of the descriptors $ArHC_i$, $ArHC_w$, $ArHC_{i/w}$, $ArHC_i/AlHC_i$ *and* $ArHC_i/AlkO_i$. In opposite direction an increase in the aliphatic content is observed, which is seen from the large and negative loading values of the descriptors $AlkO_i$, $AlkO_{i/w}$ and $AlkHC_{i/w}$.

The vertical direction, i.e. PC_2, reflects decreasing contents of ketonic and O-substituted aromatic groups downward explained by the large and positive loadings of the descriptors ArO_i, $ArO_{i/w}$, ArO_w, $CO_{i/w}$ and CO_i. In addition the shape and width of the peak quantifying the content of ester groups seems significant in explaining the variation in PC_2 score values. The score values of the individual humic materials indicates four grouping, i.e. “conventional” a high aliphatic high polar group (FA-surface and DE72), a high aromaticity

and low polarity group (Purified Aldrich HA and Gohy-573-HA-(H^+)II), a high aromatic and more specific high content of O-substituted aromatics (Gohy-573-HS-(H^+)II). In addition, the score of Kranichsee HA, indicates a group of properties, intermediate to that of fulvic and humic acids exists, i.e. significant degree of aromaticity and high polarity, distinct from Gohy-573-HS-(H^+)II humic substance properties. Model performance parameters for the PCA corresponding to Figure 4 is given in Table 4, model 2.

In Table 4 the explained variance of the individual principal components for four different models are given in the order of descending total explained variance. Model 1 is based on NMR integrated area (NMR_i), model 2 on the NMR data given in Table 2, model 3 on NMR and UV (*NMR/UV*) data and model 4 on NMR and SEC (*NMR/SEC*) data.

In general, the robustness of the PCAs in Table 4 are low due to limited number of humic samples, the groupings, and high influence of e.g. FA-surface, and inevitable the presence of variables of skewed distribution. The PCA model 1, based on the peak area integrated NMR data is insignificant by cross-validation. This is due to the significant skewness in the aliphaticity descriptor based on the current data set. By eliminating the $AlkHC_i$ descriptor and furthermore expanding the explanatory X-space by the width and area-to-width descriptors (cf. Table 2), the robustness of model 2 is increased noticeable. The explained variance in PC_1 is slightly increased by including the absorptivity in model 3, whereas the SEC data decreases the explained X-variance by calibration and validation due to the skewness of these variables in analogy to model 1. On the present basis the NMR-derived descriptors including the molar absorptivity seem to be the most well performing parameters for describing the patterns in properties of the individual humic materials. More samples are needed to reveal the significance of groupings, or increase homogeneity of spanning the X-space. A significant variation in the inherent properties between humic substances of different origin, as well as different size fractions has been quantified.

4. Discussion

According to previous investigations, the absorptivity of humic materials increases with increasing molecular weight, % C, degree of condensation, and ratio of aromatic C to aliphatic C, whereas the absorptivity of fulvic acids from various sources are fairly similar (Malcolm, 1989; Korshin et al., 1999). The abscissa of the SEC spectra in Figure 2, reflects the relative absorbance as a function of time, and as no external standard was applied in the study, it is not possible to quantify absolute size ranges of the two peak eluents. Furthermore, a calculation of the absolute concentration would require known absorptivities of the two size fractions, as the capacity for light absorbance generally differs for humic samples of different origin according to the nature of chromophores present within the different size fractions.

Generally, a high E4/E6 ratio indicate low average MW and particle size, whereas low E2/E3 indicate high degree of aromaticity (Hautala et al., 2000). However, the absorptivities and absorbances were measured on the non-fractionated samples, and an interpretation of the SECs is possible only by assuming equal distribution of content and type of functional groups between the two MW fractions. In this respect, the SEC data is virtually of no use without correcting for differences in absorptivities, and concentration determination by obtained calibration curves of the individual size fractions.

In spite of the above considerations, the correlation between the total area normalised low average MW peak and the E_4/E_6, E_2/E_3 ratio, respectively, is –0.94 and –0.67. This is in agreement with the general observance of increasing E_4/E_6 with decreasing MW and particle size. Furthermore, the degree of aromaticity quantified by the E_2/E_3 ratio increases with decreasing average MW. The relative area of the high MW peak eluent is, oppositely, positive correlated with the E_4/E_6 and the E_2/E_3 ratio.

E4/E6 have also been shown to increase with an increase in the oxygen content (Chen et al, 1977; Hautala et al., 2000). This aspect is again tested by simple correlation analysis, which

shows that negative correlation between the E_4/E_6 ratio and the descriptors $ArHC_i$, $ArHC_{iw}$, $ArHC_w$, respectively, are observed, whereas a positive correlation between the E_4/E_6 ratio and the descriptors COO_{iw} and COO_i are observed.

This study shows problem with the quality of data, i.e. the influence of impurities on the EA data, and design of experiments, i.e. SEC-fractionation contra hole sample UV-measurements. The ash content is the most critical parameter, i.e. obscuring the elemental composition, as well as data derived from size exclusion chromatograms. With respect to the usual range for the elemental composition of humic materials, strong indication of underestimated contents of C, H, and N, most significant for % C, is observed for Gohy-573-HS-(H^+)II and Aldrich HA (Na^+) (Steelink, 1985). As the purity of the humic material is crucial for the elemental analysis, these data was left out of the pattern recognition analysis, due to a significant influence from impurities.

5. Concluding remarks

In spite of inadequate spanning and homogeneity of the X-space, principal component analysis based on liquid-state ^{13}C-NMR and to a lesser extent UV-VIS spectroscopy showed highest specificity and capability of characterising the inherent properties in DOM. Inclusion of the width and area-to width descriptors increased the explained variance in PC_1. Pattern recognition in the inherent properties of DOM, indicates a continuous spectrum of properties ranging from high to low aromaticity and degree of aliphatic content, and at the same time a wide range of polarities, which is more specific quantified by content of O-substituted aromatics, carbohydrates, and ester groups. The PCA analysis indicates that the inherent properties of humic materials are more continuously distributed than a classification as fulvic acids, humic acids or humic substances.

A further investigation of the heterogeneity in the structural and compositional characteristics of humic materials is needed, as this is the basis for explaining the varying sorption capacities crucial for the fate (exposure) and effects (bioavailability) of pollutants by the presence of DOM.

Acknowledgements

Financial support from the Danish Research Academy to M. Thomsen is greatly appreciated. The authors are grateful to Drs. V. Moulin, CEA, France, G. Buchau, Forschungzentrum Karlsruhe, Germany, Karl-Heinz Heise, Forschungszentrum Rossendorf, Germany, J. Higgo, British Geological Survey, UK and A. Maes, Katholic University of Leuven, Belgium for providing samples of humic materials for the present investigation.

References

Artinger, R., Buckau, G., Kim, J.I., and Geyer, S., 1999. Characterization of groundwater humic and fulvic acids of different origin by GPC with UV/Vis and fluorescence detection. Fresenius J. Anal. Chem. 364, 737-745.

Carlsen, L., Thomsen, M., Dobel, S., Lassen, P., Mogensen, B.B. and Hansen, P.E., 2000. The interaction between esfenvalerate and humic substances of different origin. In: Ghabbour, E.A. and Davies, G. (Eds.). Humic substances. Versatile components of plants, soil and water. Royal Society of Chemistry, Cambridge, Special Publication No. 259, pp 177-189.

Chen, Y., Senesi, N. and Schnitzer, M., 1977. Information Provided on Humic Substances by E4/E6 Ratios, Soil Sci. Soc. Am. J., Vol. 41, 352-358.

Conte, P., and Piccolo, A., 1999. Conformational Arrangement of Dissolved Humic Substances. Influence of Solution Composition on Association of Humic Molecules. Environ. Sci. Technol. 33, 1682-1690.

Davis, W.M., Erickson, C.L., Johnston, C.T., Delfino, J.J., and Porter, J.E., 1999. Quantitative fourier transform infrared spectroscopic investigation of humic substance functional group composition. Chemosphere 38, 2913-2928.

EC European Commission (1996) Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New and Notified Substances and Commission Directive (EC) No. 1488/94 on Risk Assessment for Existing Substances. 1996. Luxembourg: Office for Official Publications of the European Communities. Part III. Use of QSAR, Use Categories, Risk Assessment Format, CR-48-96-003-En-C, ISBN 92-827-8013-9.

Everitt, B.S., 1993. Cluster Analysis. Third ed. Edward Arnold. A division of Holder & Stoughton, London, p. 170.

Haitzer, M., Höss, S., Traunspurger, W., and Steinberg, C., 1999. Relationship between concentration of dissolved organic matter (DOM) and the effect of DOM on the bioconcentration of benzo(a)pyrene. Aquatic toxicology 45, 147-158.

Hautala, K., Peuravuori, J., and Pihlaja, K., 2000. Measurement of aquatic humus content by spectroscopic analysis. Wat. Res. 34, 246-258.

Holmström, H., Ljungberg, J., and Öhlander, B., 2000. The characterization of the suspended and dissolved phases in the water cover of the flooded mine tailings at Stekenjokk, nothern Sweden. Sci. Total Environ. 247, 15-31.

Höskuldsson, A., 1996. Prediction Methods in Science and Technology. Thor Publishing, Denmark, p. 405.

Kaiser, K., Haumaier,L., and Zech, W., 2000. The sorption of organic matter in soils as affected by the nature of soil carbon. Soil Sci. 165, 305-313.

Kalbitz, K., Solinger, S., Park, J.-H., Michalzik, B., and Matzner, E., 2000. Controls on the dynamics of dissolved organic matter in soil: A review. Soil Sci. 165, 277-304.

Kim, J.I., Buckau, G., Li, G.H., Duschner, H. and Psarros, N., 1990. Characterization of humic and fulvic acids from Gorleben groundwater. Fresenius J. Anal. Chem. 338, 245-252.

Korshin, G.V., Croué, J.-P., Li, C.-W. and Benjamin, M.M., 1999. In: Ghabbour, E.A. and Davies, G. (Eds.). Understanding Humic Substances. Advanced methods, properties and applications. The Royal Society of Chemistry, Cambridge, pp.147-157.

Lassen, P., Carlsen, L., Warwick, P., Randall, A. and Zhao, R., 1994. Radioactive labelling and characterization of humic materials, Environ.Int. 20, 127-134.

Malcolm, R.L., 1990. The uniqueness of humic substances in each of soil, stream and marine environments. Anal. Chim. Acta 232, 19-30.

Malcolm, R.L., 1989. Applications of solid-state 13-C-NMR spectroscopy to geochemical studies of humic substancs. In: Hayes, M.H.B., MacCarthy, P., Malcolm, R.L. and Swift, R.S.

(Eds.). Humic Substances II. Search of structure. John Wiley and Sons, Chichester, pp. 339-372.

Periago, E.L., Delgado, A.N., and Diaz-Fierros, F., 2000. Groundwater contamination due to cattle slurry: Modelling infiltration on the basis of soil column experiments. Wat. Res. 34, 1017-1029.

Ragle, C.S., Engebretson, R.R., and von Wandruszka, R., 1997. The Sequestration of Hydrophobic Micropollutants by Dissolved Humic Acids. Soil Sci. 162, 106-114.

Ran, Y., Fu, J.M., Sheng, G.Y., Beckett, R., and Hart, B.T., 2000. Fractionation and composition of colloidal and suspended particulate materials in rivers. Chemosphere 41, 33-43.

Schulten, H.-R., 1996. A new approach to the structural analysis of humic substances in water and soils. In: Gaffney, J.S., Marley, N.A. and Clark, S.B. (Eds.). Humic and fulvic acids. Isolation, structure, and environmental role. American Chemical Society, Washington DC, vol. 651, pp. 42-56.

Robertson, A.P., and Leckie, J.O., 1999. Acid/Base, copper binding, and Cu2+/H+ exchange properties of a soil humic acid, an experimental and modeling study. Environ. Sci. Technol. 33, 786-795.

Shen, Y.-H., 1999. Sorption of humic acid to soil: The role of soil mineral composition. Chemosphere 38, 2489-2499.

Steelink, C., 1985. Implications of elemental characteristics of humic substances. In: Aiken, G.R., McKnight, D.M. and Wershaw, R.L. (Eds.). Humic substances in soil, sediment, and water. Geochemistry, isolation, and characterization. John Wiley and Sons, New York, pp. 457-476.

Tombácz, E., 1999. Colloidal Properties of Humic Acids and Spontaneous Changes of Their Colloidal State under Variable Solution Conditions. Soil Sci. 164, 814-824.

Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., and Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soil subjected to drying. Chemosphere 38, 45-50.

Figure captions

Figure 1 Liquid-state ^{13}C-NMR of six humic substances.

Figure 2 Size exclusion chromatograms of the humic substances, expressed as the relative absorbances at 285 nm as function of retention time.

Figure 3 Results of hierarchical cluster analysis, showing dendograms of range-normalised data a) based on the NMR data given in Table 2, b) based on SEC data, c) based on UV data, and d) all data, and by use of data derived from the different characterisation methods (bottom 4), respectively.

Figure 4 Bi-plot of loadings and scores, showing PC_1 on the ordinate and PC_2 on the abscissa. PC_1 explains 57 %, and PC2 29 % of the variation in X.

Table 1 Types, names and origin of the eight humic substances included in the present study.

Type	***Names***	***Origin***
Fulvic acids	DE72	Fulvic acid fraction, Dervent Reservoir, Derbyshire, U.K.
	FA-surface	Fulvic acid fraction of surface water, Soulaines, France
Humic acids	Aldrich HA (Na^+)	Commercial
	Purified Aldrich HA	
	Kranichsee HA	From pore water in raised bog, Kleiner Kranichsee, Germany
	Gohy-573-HA-(H^+)II	Isolated humic acid fraction of dissolved organic matter in groundwater, Gorleben, Germany
Humic substances	Gohy-573-HS(H^+)II	Dissolved organic matter in groundwater, Gorleben, Germany
	Water pond HS	Dissolved organic matter from surface water, National Environmental Research Institute, Roskilde, Denmark

The size distribution and chemical composition of humic materials depends on the type and origin as shown in Table 1, and are classified, by "type", as fulvic acids (FA), humic acids (HA) and humic substances (HS). This way of classifying humic materials is traditionally based on specific steps in the experimental fractionation and isolation procedures (e.g. Malcolm, 1990).

Table 2 The percentage of the different fractions of chemical building blocks of the humic substances as determined by ^{13}C-NMR-spectroscopy.

HS, Peak area[a]:	CO_i	COO_i	ArO_i	$ArHC_i$	$AlkO_i$	$AlkHC_i$
DE72	3.01	31.47	7.07	24.22	13.04	21.18
FA-surface	5.00	22.46	0.00	7.08	13.94	51.52
Gohy-573-HA-(H^+)II	1.21	8.90	8.01	52.68	1.95	27.25
Gohy-573-HS-(H^+)II	6.42	12.80	18.58	39.47	6.75	15.99
Kranichsee HA	1.76	18.31	9.74	35.85	18.09	15.18
Purified Aldrich HA	3.78	16.10	5.42	50.55	4.55	19.58
HS, Peak width[b]:	CO_w	COO_w	ArO_w	$ArHC_w$	$AlkO_w$	$AlkHC_w$
DE72	29.97	26.89	23.01	49.96	44.91	40.02
FA-surface	31.95	29.94	0.00	48.00	33.96	58.99
Gohy-573-HA-(H^+)II	24.95	24.99	19.96	55.94	23.98	70.04
Gohy-573-HS-(H^+)II	20.97	15.67	26.27	56.98	29.94	42.01
Kranichsee HA	21.97	21.94	22.94	52.96	41.97	39.96
Purified Aldrich HA	22.01	27.97	14.94	53.03	39.96	61.97
HS, Peak Area/Peak width[c]:	$CO_{i/w}$	$COO_{i/w}$	$ArO_{i/w}$	$ArHC_{i/w}$	$AlkO_{i/w}$	$AlkHC_{i/w}$
DE72	0.10	1.17	0.31	0.48	0.29	0.53
FA-surface	0.16	0.75	0.00	0.15	0.41	0.87
Gohy-573-HA-(H^+)II	0.05	0.36	0.40	0.94	0.08	0.39
Gohy-573-HS-(H^+)II	0.31	0.82	0.71	0.69	0.23	0.38
Kranichsee HA	0.08	0.83	0.42	0.68	0.43	0.38
Purified Aldrich HA	0.17	0.58	0.36	0.95	0.11	0.32

The descriptors represents quinolic and ketonic carbon atoms (*CO*), carboxylic and ester carbon atoms (*COO*), O-substituted aromatics (*ArO*), un-substituted or C-substituted aromatics (*ArHC*), carbohydrate or O-substituted aliphatic carbons (*AlkO*) and un-substituted aliphatic carbons (*AlkHC*), respectively.

[a] The subscript *i* denotes descriptors quantified by integrated peak area.

[b] The subscript *w* denotes descriptors quantified by peak width.

[c] The subscript *i/w* denotes descriptors quantified by peak area divided by peak width.

Table 3 Elemental composition and spectroscopic data.[a]

Humic substances	*Elemental analysis*					*UV-VIS data*			
	%C	%H	%N	%O	H/C	(N+O)/ C	E_4/E_6	E_2/E_3	ε 272 *
Gohy-573-HS-(H^+)II	34.03	5.05	1.14	59.48	1.78	1.35	8.27	3.30	0.016
Gohy-573-HA-(H^+)II	57.32	4.76	1.77	36.15	1.00	0.5	6.20	2.89	0.034
Aldrich HA (Na^+)	38.37	4.68	0.57	56.38	1.46	1.12	5.56	2.67	0.029
Purified Aldrich HA	53.27	4.88	0.93	40.92	1.10	0.59	7.43	2.48	0.039
Kranichsee HA	49.34	4.07	1.60	44.99	0.99	0.71	8.82	3.10	0.024
DE72	49.06	4.18	0.60	46.16	1.02	0.72	11.80	3.00	0.022
FA-surface	48.76	5.03	1.14	45.07	1.24	0.71	10.97	7.92	0.007
Water pond HS	28.29	4.83	1.81	65.07	2.05	1.78	21.32	7.04	0.004

• L $(mg\ cm)^{-1}$

[a] Data taken from Ref. 59.

Table 4 Explained X-variance of two-component PCAs, and total explained variance, based on NMR, UV and SEC derived descriptors

Model no., descriptors	***Explained variance of PC_1, PC_2 and in total***					
	by calibration			*by cross- validation*		
	PC_1	PC_2	$\Sigma varPC_{cal}$	PC_1	PC_2	$\Sigma varPC_{val}$
1, NMR_i[a]	54	23	77	3	12	15
2, NMR[b]	47	29	76	18	10	28
3, NMR/UV[c]	50	26	76	21	8	29
4, *NMR/SEC*	44	27	71	14	1	15

[a] based on peak area integrated NMR data (cf. Table 2), including a non-normal distributed variables $AlkHC_i$,

[b] based on all NMR data given in Table 2, excluding non-normal distributed variables $AlkHC_i$ and CO_w

[c] non-normal distributed variables, E2/E3 and E4/E6 eliminated, i.e. only one additional descriptor, the absorptivity, are included.

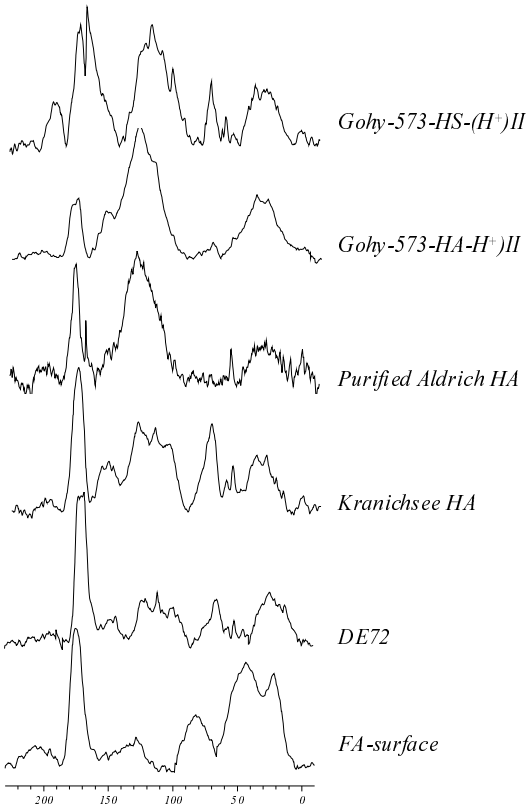

Gohy-573-HS-(H+)II
Gohy-573-HA-H+)II
Purified Aldrich HA
Kranichsee HA
DE72
FA-surface
200
150
100
50
0

Figure 1

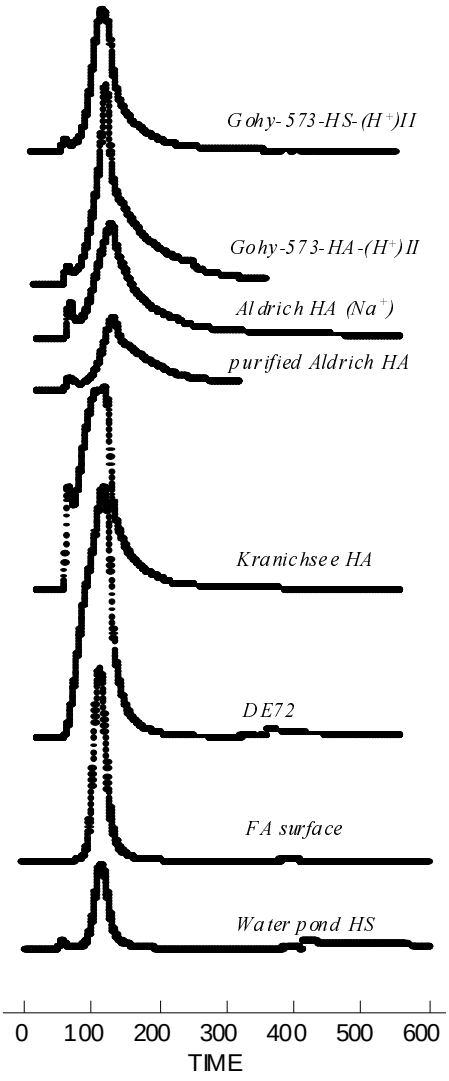

Gohy-573-HS-(H+)II
Gohy-573-HA-(H+)II
Aldrich HA (Na+)
purified Aldrich HA
Kranichsee HA
DE72
FA surface
Water pond HS
0
100
200
300
400
500
600
TIME

Figure 2

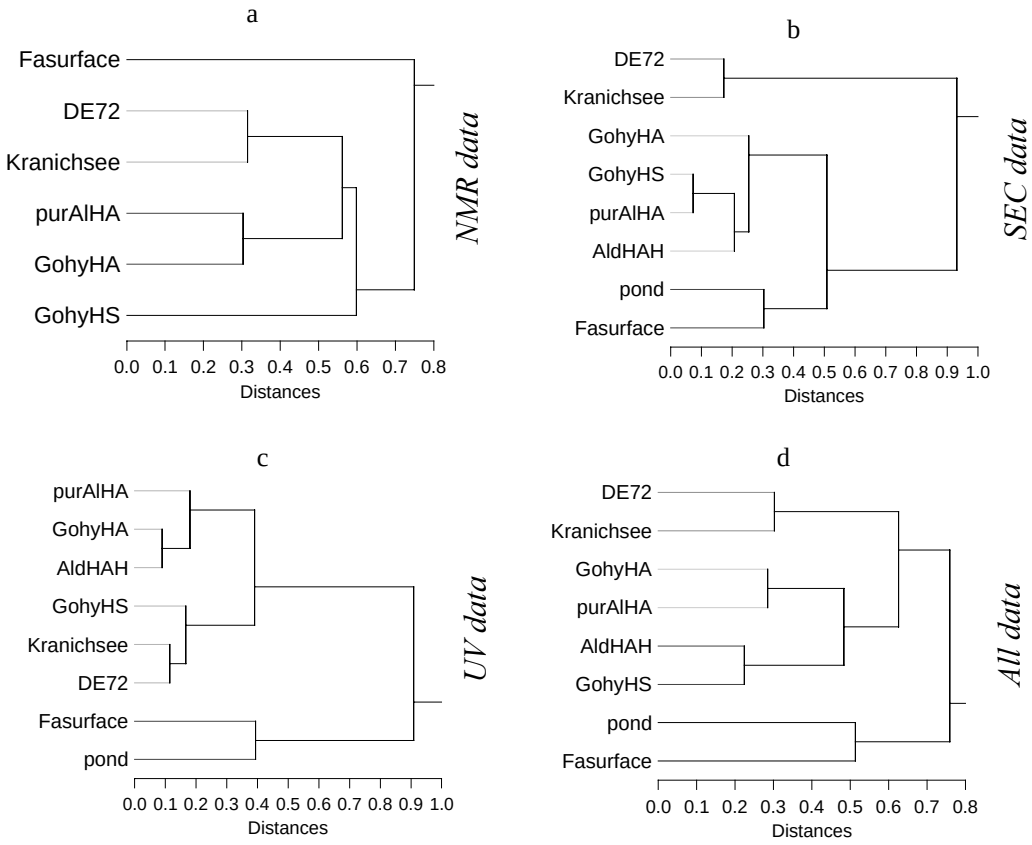

a
Fasurface
DE72
Kranichsee
purAlHA
GohyHA
GohyHS
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Distances
NMR data
b
DE72
Kranichsee
GohyHA
GohyHS
purAlHA
AldHAH
pond
Fasurface
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Distances
SEC data
c
purAlHA
GohyHA
AldHAH
GohyHS
Kranichsee
DE72
Fasurface
pond
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Distances
UV data
d
DE72
Kranichsee
GohyHA
purAlHA
AldHAH
GohyHS
pond
Fasurface
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Distances
All data

Figure 3

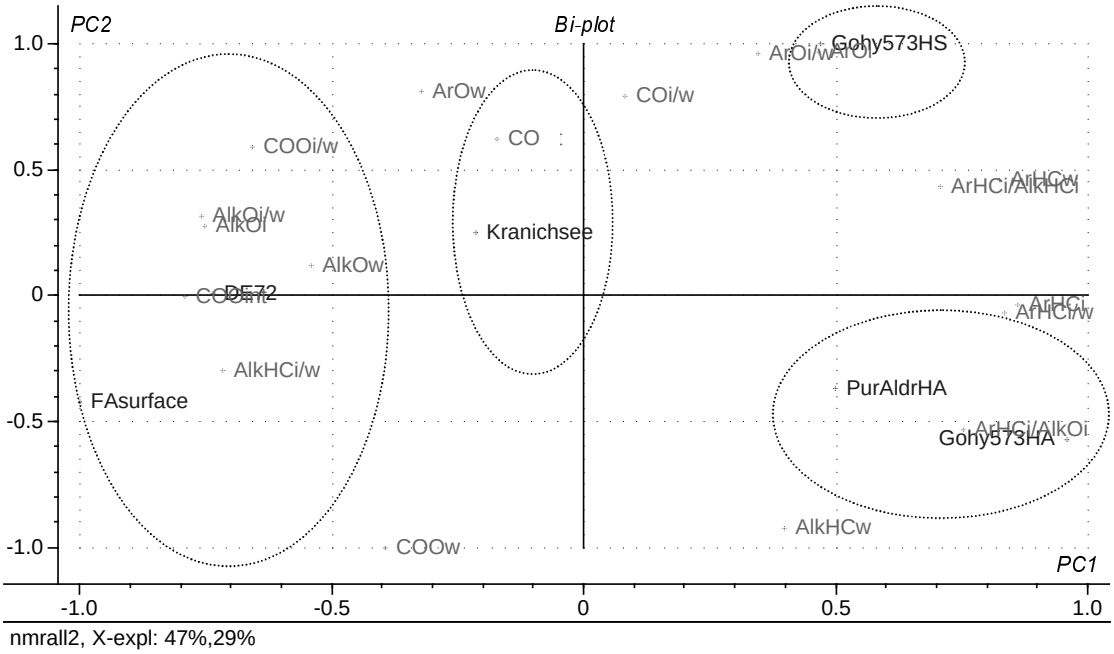

PC2
Bi-plot
PC1
1.0
0.5
0
-0.5
-1.0
-1.0
-0.5
0
0.5
1.0
Gohy573HS
ArOi/w
ArOi
ArOw
COi/w
CO
COOi/w
ArHCi/AlkHCi
ArHCw
AlkOi/w
AlkOi
Kranichsee
AlkOw
DE72
ArHCi
ArHCi/w
AlkHCi/w
FAsurface
PurAldrHA
ArHCi/AlkOi
Gohy573HA
AlkHCw
COOw
nmrall2, X-expl: 47%,29%

Figure 4

Paper V

Reverse QSAR for modelling the sorption of esfenvalerate to dissolved organic matter (DOM). A multivariate approach

Submitted to *Chemosphere*

Reverse QSAR for modelling the sorption of esfenvalerate to dissolved organic matter (DOM). A multivariate approach

Marianne Thomsen[a,*], Shima Dobel[b], Pia Lassen[a], Lars Carlsen[c], Betty Bügel Mogensen[a] and Poul Erik Hansen[d]

[a]*National Environmental Research Institute, Department of Environmental Chemistry, DK-4000 Roskilde, Denmark*

[b]*Danish Environmental Protection Agency, Office of Chemicals, Strandgade 29, DK-1401 Copenhagen-K, Denmark*

[c]*Department of Environment, Technology and Social Studies, Roskilde University, DK-4000 Roskilde, Denmark*

[d]*Department of Life Science and Chemistry, Roskilde University, DK-4000 Roskilde, Denmark*

*Corresponding author. Tel.: +45 +46 30 13 58; fax: +45 +46 30 11 14.

E-mail address: mth@dmu.dk (M. Thomsen)

Abstract

The sorption of the pyrethroid, esfenvalerate, to the dissolved fraction of eight different natural humic compounds has been investigated and modelled at DOM concentration levels where equilibrium partitioning of esfenvalerate between DOM and the aqueous bulk phase prevails. The inherent characteristics of the eight different humic materials have been used as explanatory variables for modelling the equilibrium partitioning of esfenvalerate between bulk water and DOM of different origin. Through this reverse QSAR approach based on by PLS-R (Projection-into-Latent-Structure Regression) inherent sorbent properties determining for the sorption affinity of esfenvalerate to DOM were analysed. Significant variations in equilibrium partitioning coefficients, K_{DOM}, to DOM of different origin was found at DOM concentrations of 75 and 100 ppm, respectively. The latter is a strong indication of variations in sorption mechanisms to DOM of varying inherent properties. Groupings in the principal property space quantifying DOMs may indicate that separate models are needed for quantifying the equilibrium partitioning to different classes of DOM. DOMs included in this study origins from ground water, soil pore water, and DOM of aqueous origin.

Keywords: Dissolved Organic Matter (DOM); Esfenvalerate, Reverse QSAR; equilibrium partitioning coefficient; KDOM

1. Introduction

Environmental risk assessment of pollutants have traditionally been based solely on the inherent properties of the individual chemical compounds, i.e. physicochemical properties such as solubility, octanol-water partitioning, sorption and degradation rate in different media. These physicochemical properties are used as input parameters in models for calculating the fate, e.g. transport and distribution, of chemicals in environmental compartments air, water and soil (EC, 1996). However the effect of binding of pollutants to a mobile fraction of organic matter is not included in the current assessment of the fate and effects of pollutants within the environments (EC, 1996).

The natural occurrences of DOM in surface waters, soil and sediment pore water have been found in the range of 2-50, 10-1000 and above 100 mg C/L, respectively (Cao et al., 1999; Caron and Suffet, 1989; Kukkonen and Oikari, 1991). The impact of dissolved organic matter, and thereby the possibility of a third-phase effect (Lee and Kuo, 1999), may be a significant factor in relation to the fate and effects of environmental pollutants. Sorption to dissolved organic matter (DOM) may significantly affect the apparent solubility and migration potential of especially hydrophobic substances (Fauser and Thomsen, 2002). Furthermore, the presence of dissolved organic matter reduce the bioavailability, and thereby toxicity, of pollutants in most cases (Haitzer et al., 1998; Kukkonen and Oikari, 1991; Steinberg et al., 2000).

Experimental parameters, such as concentration of DOM, pH and ionic strength (Ghosh and Schnitzer, 1980), contribute to the generally high uncertainty or variability in measured equilibrium partitioning to DOM (Cousins and Makay, 2000; Staples et al., 1997). Still partitioning coefficients are often considered unique values independent of varying inherent properties and concentration of the organic sorbent. Thus the sorption of PAHs to DOM increases with an increase in aromaticity of the DOM due to van der Waals interactions between the polycyclic aromatic hydrocarbons (PAHs) and aromatic substructures of DOM

(Lassen and Carlsen, 1999; Nielsen et al., 1997; Perminova et al., 1999). However, for structurally more complicated molecules such as the pesticides, including the pyrethroids, no such clear relationship can be quantified (Oesterreich et al., 1999; Piccolo et al., 1996). The apparent lack of simple relationships describing the organic matter partitioning coefficients for the pyrethroids may be explained through the structural complexity of the this group of compounds.

The main goal of this study is to analyse the explanatory significance of DOM descriptors for estimating the sorption of esfenvalerate to different types of DOM.

Through the use of reverse QSARs, i.e. using the inherent properties of the humic substances for quantifying the $\log K_{DOM}$ of esfenvalerate, the predictability of PLS-regressions based on different type of DOM descriptors is evaluated.

2. Descriptors quantifying the inherent properties of DOM

Descriptors included are derived from liquid-state ^{13}C-NMR and UV-VIS spectroscopy, respectively (data given in Thomsen et al., 2002). These descriptors, quantifies the amount and distribution of the different functional groups within the DOMs. The aliphaticity and aromaticity descriptors contain secondary information concerning the rigidity/flexibility of the DOM macromolecules.

The NMR data comprises eighteen descriptors quantifying the percentage distribution of sub-structural groups within the humic macromolecular structure. The descriptors are integrated peak areas quantifying the content of C atoms of quinonic and ketonic groups CO_i, , carboxylic and ester C, COO_i, aromatic O-substituted C, ArO_i, un-substituted aromatic C, $ArHC_i$, N-alkyl and methoxy C including a major fraction of carbohydrates, $AlkO_i$ and un-substituted aliphatic C, $AlkHC_i$. The same type of descriptors was calculated based on peak widths, the subscript being replaced by w, and for peak areas divided by widths, the subscript

replaced by *i/w*. All descriptors calculated according to the chemical shift ranges (ppm) of the functional subgroups within the molecular structure of DOM.

The UV-VIS descriptors quantify information concerning the degree of aromaticity, average molecular weight (MW), size and oxygen content. The E_4/E_6 ratio is found to increase with decreasing average MW and increasing oxygen content, the E_2/E_3 ratio increases with a decrease in aromaticity, and finally the absorptivity quantifies the aromatic $\pi \rightarrow \pi^*$ transitions at 272 nm.

The selection of descriptors, included in the modelling of the sorption of esfenvalerate to DOM of different origin, is based on a preceding evaluation of the explanatory significance of different spectroscopic methods and elemental analysis data (Thomsen et al., 2002).

3. Equilibrium partitioning of esfenvalerat to DOM - K_{DOM}

The nominal concentration of the solute, esfenvalerate, was 2.57 µg/L in all measurements, and the aqueous media of the experimental systems, were kept at constant pH, ionic strength and temperature (Carlsen et al., 2000). As such, the only varying parameter is the change in DOM concentration. The equilibrium constant for the process of complex formation between esfenvalerate and DOM in an aqueous bulk phase were determined DOM concentrations of 10, 20, 30, 40, 50, 75 and 100 mg/L. The so-called DOM-normalised partitioning coefficient calculated as

$$K_{DOM} = \frac{C_{esfen}^{DOM}}{C_{esfen}^{aq} C_{DOM}} \qquad (1)$$

where C_{esfen}^{DOM} and C_{esfen}^{aq} are the concentration of esfenvalerate bound to DOM and the concentration of freely dissolved esfenvalerate, respectively. The equilibrium constant, as expressed in equation 1, refers to the process

$$esfenvalerate + DOM \rightleftharpoons [\ esfenvalerate\text{-}DOM] \qquad (2)$$

A system described as in equation 1 and 2 is defined as being a mixture of dilute pollutant, e.g. esfenvalerate, and DOM in an aqueous bulk phase, and as seen from Table 1, the equilibrium process, in equation 2, is displaced towards the right at increasing DOM concentrations. An equilibrium partitioning coefficients, independent of the DOM concentration is expected, only when a two-phase system prevails, i.e. the activity of DOM phase equals one. This seems to be the case at concentration levels of DOM above 60-90 mg/L (cf. Figure 1).

The equilibrium concentration of esfenvalerate in bulk water and bound to DOM by complex formation, as described by equation 2, was determined by an inverse column elution method (Kukkonen et al, 1990). The method is based on separating the freely dissolved fraction of esfenvalerate from the DOM-complexed fraction by eluting the aqueous equilibrium mixtures through an inert support column. Freely dissolved hydrophobic compounds will be retained on the column, while the DOM molecules, and thereby the DOM-bound fraction of esfenvalerate, will show lower retention times (Landrum et al., 1984). The esfenvalerate concentration, in the continuous bulk water phase and bound to DOM, is determined by scinitilisation using ^{14}C-labbeled esfenvalerate. Average K_{DOM} values based on triple to quintuple repeated sample measurements is given in Table 1.

The equilibrium partitioning coefficients as function of DOM concentration, for Water pond HS and Gohy-573-HS-(H^+)II, are given in Figure 1.

As seen from Figure 1, as well as given in Table 1 and 2, the dependence of K_{DOM} on C_{DOM} is most significant at low concentrations of DOM. At high DOM concentrations, above approximately 90 ppm for Water pond HS and 60 ppm for Gohy-573-HS-(H^+)II, K_{DOM} is approaching constant values, i.e. the partitioning coefficient becomes independent of the concentration of DOM (Hiemenz and Rajagopalan, 1997; Schwarzenbach et al., 1993). The standard deviations on reproduced determinations of K_{DOM}, given as error bars in Figure 1, are observed to decrease significantly at increasing DOM concentration. This pattern is similar for all humic samples, as given in Table 2.

The standard deviation on the DOM-normalised partitioning coefficients for esfenvalerate, is similar to the standard deviations found for benzo[α] pyrene, anthracene, biphenyl, *p,p'*-DDT, 2,5,2´,5'-tetrachlorobiphenyl and 2,4,5,2',4',5'-hexachlorobiphenyl based on the same reversed-phase method (Landrum et al.,1984).

Analysis of homogeneity of variances (Funk et al., 1995) of measured DOM-complexed, C_{esfen}^{DOM}, and freely dissolved concentrations of Esfenvalerate, C_{esfen}^{aq}, (cf. Eq.1) as function of the DOM concentrations was performed to evaluate the relative contributions to the standard deviation in K_{DOM} values. The results showed that variance inhomogeneity is highest for measuring of the dissolved esfenvalerate concentrations in respect to the variance in measurements on the DOM-complexed fraction (data not given).

Significance testing of the changes in the DOM-normalised partitioning coefficients in the concentration range of 10 to 100 mg DOM per Litre is relevant in relation to the need for correction factors when quantifying the partitioning to DOM. This due to natural variations in DOM concentration levels in different natural compartment systems. If the dependence of K_{DOM} on the concentration DOM concentration is not significant, then there is no need to investigate this aspect any further (cf. Figure 1 and Table 2).

Another aspect of partitioning to DOM is the effect of variations in the inherent properties of DOM. Modelling the complex-formation between esfenvalerate and DOM at the different concentration levels, require that significant differences in K_{DOM} at the individual DOM concentration levels exist.

In summery it is of utmost importance that the variation between object, i.e. in this case K_{DOM} values for the partitioning of esfenvalerate to DOM of different origin, is significant. Therefore, before calibration of any SAR/QSAR model, the test data should be validated by pre-processing of endpoint data (Cousins and Mackay, 2000; Thomsen, 2001), e.g. as described below.

3.1 Significance testing of K_{DOM} data

Pre-processing of data is an overlooked step in the majority of SAR/QSAR investigations for estimating endpoints for use in environmental risk assessment at any level. Simple endpoints such as the aqueous solubility and octanol-water partition coefficients shown significant variations according to experimental standard methods used for the measuring the endpoint.

Prior to the development of QSARs, validation of the quality of endpoint data, as well as significance testing, in respect to the variation in endpoint-data, is performed (Thomsen, 2001). Furthermore, the homogeneity in spanning of the descriptor space used for quantifying the endpoint is investigated (Thomsen and Carlsen, 2001).

As partition coefficients to natural organic matter in general display high variabilities (Cousins and Mackay, 2000; Staples et al., 1997), the variances in K_{DOM} for each DOM at each concentration level, σ_{within}, is compared to the variance between DOMs at each DOM concentration level, $\sigma_{between,1}$. The minimum requirement to data is that the variance of each average K_{DOM} value is significantly lower than the variance in K_{DOM} between DOMs at each

concentration level. A one-side F-test (Miller and Miller, 1988) for this hypothesis is defined as

$$H_1(\sigma^2_{between,1} > \sigma^2_{within}): \quad T_1 > F_{1\,(1-0.05)} \qquad (3)$$

If H_1 is fulfilled at 95% confidence level, then the variance in K_{DOM} values between DOMs is significant higher than the variance in K_{DOM} for the individual DOMs at individual DOM concentration levels. Thus, the data quality accepted for further analysis. PLS-regression models may be developed for each concentration level, where H_1 is fulfilled.

A second hypothesis, H_2, was performed to validate the concentration dependence of K_{DOM}. Hypothesis, H_2, is a tests to evaluate if the variance in K_{DOM} between DOM concentration levels for each specific DOM ($\sigma^2_{between,2}$) is significantly higher than the variance on K_{DOM} for each DOM at each concentration level, σ_{within}, i.e.

$$H_2(\sigma^2_{between,2} > \sigma^2_{within}): \quad T_2 > F_{2\,(1-0.05)} \qquad (4)$$

The rejection or acceptance of the hypothesis H_1 and H_2 are summaries in Table 3.

In general the variance of K_{DOM} values between the different DOMs, at constant concentration level of DOM, are in the same order of magnitude as found for, e.g., the low molecular weight phthalates, and lower than variances of K_{DOM} values for the high molecular weight phthalates (Staples et al, 1997). Furthermore the variances within each K_{DOM} measurement are much lower than found the pesticides between different unknown sorbents (NERI, 2001). It is observed that the frequency of rejected hypothesis, H_1 and H_2 increases at decreasing DOM concentration level. The results of the ANOVA indicates that variations in K_{DOM} values may not only be due to the heterogeneity and complexity of natural organic matter, but also to

uncertainties in the experimental methods used for measuring the partition to of organic pollutants to natural organic matter. This due to the systematic increase in standard deviation for all DOMs, as well as high contribution from uncertainties from the measuring the concentration of freely dissolved esfenvalerate.

The variation in K_{DOM} between DOMs of different origin at DOM concentration of 10, 20, 30, 40 and 50 mg/L is insignificant, as tested by H_1 (cf. Table 3). Therefore, it is only reasonable to model the sorption of esfenvalerate to DOMs of different origin at the two highest DOM concentration levels, i.e., 75 and 100 mg/L, respectively.

4. QSARs based on partial least-squares regression

The QSAR paradigm is based on the assumption the variation in the activity of chemical compounds can be modelled through a quantification of the molecular inherent structural and electronic properties. In this study however the approach is reversed and the change in the partitioning of single compound, esfenvalerate, to DOM is modelled through a quantitative description of variations in the inherent properties natural dissolved organic matter. The objective is to study the influence of changes in microenvironment surrounding a single pollutant molecule on the measured endpoint, i.e. K_{DOM}.

The absolute value of the DOM-normalised equilibrium coefficient (cf. Eq.´s 1 and 2) as described in the preceding sections depend on the DOM concentration as well as the origin of DOM. This is a major problem for most environmental and ecotoxicological QSARs as they do not include any effects from varying environmental conditions.

The results of the four best performing PLS-models at DOM concentration levels of 75 and 100 mg/L are given in Table 4.

The correlation coefficient, R^2, expresses the fraction of the variance in $\log K_{DOM}$ that is explained by the models, and Q^2, the cross-validated correlation coefficient, expresses the fraction of predicted variance according to the leave-one-out method (Höskuldsson, 1996, CAMO ASA, 1998). The standard deviations on endpoint values, i.e. the K_{DOM} measurements, show significant influence on the model performance. The robustness of the models is reflected in the differences between Q^2 and R^2, and as seen from Table 4, the robustness decreases significantly by decreasing DOM concentration. The same trend is observed for the root mean square error of calibration (RMSEC) versus the root mean square error of predictions (RMSEP) by the leave-on-out cross-validation method as seen from Table 4. The RMSEP is a measure of the average differences between predicted and measured K_{DOM} values. In this case the RMSEP is obtained through a full cross validation with six different DOMs, and six different models based on five possible subsets of five DOMs, i.e. leaving one sample out at a time. For each of the left-out DOM samples predictions are made, and the RMSEP is calculated through the expression

$$RMSEP = \sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{I}} \qquad (5)$$

I is the number of calibration samples, i is the left out sample, $\hat{y}_i$ is the predicted K_{DOM} by the model calibrated on the I samples, and y_i is the K_{DOM} values estimated by the reference models based on the six DOMs.

A bi-plot of second, *PC2* versus the first principal component, *PC1*, showing the loading weights of the NMR descriptors and the K_{DOM} loading are given in Figure 2.

The most significant principal component, *PC1*, explains 78 % of the variation in K_{DOM}. From Figure 2, it is showed that the partitioning to DOM increases with increasing un-substituted or

C-substituted aromaticity, quantified by the descriptors $ArHC_I$ and $ArHC_i/AlkHC_i$. The logK_{DOM} value is inversely related to the aliphatic carbon shape descriptors, *AlkHCi/w*, and carbohydrate, $AlkO_{i/w}$, descriptors. Furthermore a significant inverse relation to the width descriptors, ArO_w and $ArO_{i/w}$ is observed.

In *PC2* explaining 16 % of the variation in K_{DOM}, the descriptors COO_i, $AlkO_w$ and COO_w have high positive loading weights, whereas the descriptors ArO_I, $ArO_{i/w}$ and $ArHC_w$ have high negative loading weights. The partitioning to DOM is inversely related to the O-substituted aromaticity shape and width descriptors in the third quadrant. The NMR shape and width descriptors increase the model performance significantly in addition to models based on solely the integrated area descriptors. The effect is mainly on the robustness of the model. In the above models the width of carbohydrate peaks has high loading weight in *PC2*, which contribute significant to the homogeneous spanning of the DOM-property space in the PLS regression models when compared to a preceding study of the inherent properties of DOM. This is illustrated below in Figure 3.

By use of the reverse QSAR concept it is possible to quantify varying sorbent sorption capacities. The standard errors of predicted K_{DOM} values are significantly lower than the standard deviation given in Table 2 in all of the tested models. The sorption generally increases with increasing aromaticity, but still the aliphaticity and carbohydrate descriptors is just as significant descriptors in the investigated models for quantifying K_{DOM} of esfenvalerate. By fitting the X-matrix to logK_{DOM} the weighting of the original descriptor variables changes compared to the PCA model based exclusively on the DOM property descriptors as investigated in a preceding study (Thomsen et al., 2001). The shape and size descriptors generally show high significance in explaining the variation in logK_{DOM}.

5. Discussion and conclusions

There are different theories concerning the dependence of K_{DOM} on the DOM concentration. DOM may consist of different types of sorption sites, and in this case the solute will show highest affinity towards the most energetically favourable sorption sites, secondly the next highest etc., which could explain a change in K_{DOM} by a change in the DOM concentration (Schwarzenbach, 1993). Another explanation that may contribute to explaining the concentration dependence of K_{DOM} is colloidal nature of DOMs, and the degree of inter- and intra-molecular associations within and between the organic macromolecules (Wershaw, 1999). At low concentration of DOM in the aqueous bulk, the system may be described as an aqueous true solution of humic monomeric macromolecules and esfenvalerate, respectively, as described by equation 1 and 2. However, at increasing concentration the size of the organic macromolecular colloids increases, and the system changes from being a solution to being a two-phase systems. In this case, under the right conditions of dilute solutions of esfenvalerate, the partition to DOM is expected to be independent of the DOM concentration (Hiemenz and Rajagopalan, 1997; Schwarzenbach et al., 1993).

Clearly the uncertainty level of K_{DOM} measurements is expected to be significant higher than compared to the uncertainty on octanol-water partitioning measurements. This simply due to the complexity of DOMs, caused by the heterogeneity in structural characteristics such as shape and size as function of concentration level. Significant variations in the partitioning of esfenvalerate to DOM have been quantified solely based on a quantification of sub-structural functional group within the different humic materials. For an in depth analysis of the effects of the inherent properties of DOM of aqueous origin, the inclusion of more fulvic type DOMs is needed.

The molecular structure of esfenvalerate includes an ester group, a cyano-group, a biphenylether, a chlorophenyl and an alkyl group. As observed for other pesticides

(Oesterreich et al., 1999; Piccolo et al, 1996), this may increase the complexity, and the number of possible mechanisms, of sorbate-sorbent interactions. For this reason the patterns in loadings weights of the inherent DOM property descriptors may very well vary for different e.g. pesticides. This aspect, as well as the variance inhomogeneity in K_{DOM} values for DOMs between concentration levels as well as inherent properties, suggests that classification of humic materials into similar inherent properties is required. The latter at least if the variation in sorption affinities to DOM of different classes of environmental pollutants is to be quantified by conventional QSARs. Separate QSAR models for the different classes of humic substances will probably increase the robustness and predictability of QSAR models for estimating sorption to DOM provided that the measured equilibrium partitioning values are independent of the DOM concentration.

The present study has showed significant variation in sorption affinities of esfenvalerate to dissolved humic substances of different origin. The significance of this variation needs to be further investigated for other pesticides with respect to generally high variabilities in measured partitioning coefficient to sorbent of varying composition. With respect to the bioavailability and mobility of environmental pollutants, more focus on the presence of a non-fixed mobile organic matter or third phase effects from a dispersed colloidal phase. The presence of DOM has shown impacts on the potential risks of environmental pollutants to a degree that is yet only sparsely elucidated.

References

CAMO ASA, 1998. The Unscrambler 7.01, Oslo, Norway.

Cao, J., Tao, S. and Li, B.G., 1999. Leaching kinetics of water soluble organic carbon (WSOC) from upland soil, Chemosphere 39, 1771-1780.

Carlsen, L., Thomsen, M., Dobel, S., Lassen, P., Mogensen, B.B. and Hansen, P.E., 2000. The interaction between esfenvalerate and humic substances of different origin. In: Ghabbour, E.A. and Davies, G. (Eds.). Humic substances. Versatile components of plants, soil and water. Royal Society of Chemistry, Cambridge, Special Publication No. 259, pp. 177-189.

Caron, G. and Suffet, I.H., 1989. Binding of nonpolar pollutants to dissolved organic carbon. Environmental fate modelling. In: Suffet, I.H. and MacCarthy, P. (Eds.). Aquatic Humic Substances. Influence on Fate and Treatment of Pollutants. Advances in Chemistry Series, No. 219, pp. 117-130.

Cousins, I., Mackay, D., 2000. Correlating the physical-chemical properties of phthalate esters using the "three solubility" approach. Chemosphere 41, 1389-1399.

EC European Commission, 1996. Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New and Notified Substances and Commission Directive (EC) No. 1488/94 on Risk Assessment for Existing Substances. 1996. Luxembourg: Office for Official Publications of the European Communities. Part III. Use of QSAR, Use Categories, Risk Assessment Format, CR-48-96-003-En-C, ISBN 92-827-8013-9.

Funk,W., Dammann, V., Donnevert, G., 1995. Quality assurrance in analytical chemistry. VCH Verlagsgesellschaft, Weinheim.

Ghosh, K., Schnitzer, M., 1980. Macromolecular structures of humic substances. Soil Sci. 129, 266-276.

Haitzer, M., Höss, S., Traunspurger, W., and Steinberg, C., 1998. Effects of dissolved organic matter (DOM) on the bioconcentration of organic chemicals in aquatic organisms - A review. Chemosphere 37, 1335-1362.

Hiemenz, P.C., Rajagopalan, R., 1997. Principles of Colloid and Surface Chemistry. Third Edition, Revised and Expanded. Marcel Dekker, Inc.

Kukkonen, J., McCarthy, J.F., and Oikari, A., 1990. Effects of XAD-8 fractions of dissolved organic carbon on the sorption and bioavailability of organic micropollutants. Arch. Environ. Contam. Toxicol. 19, 551-557.

Kukkonen, J., and Oikari, A., 1991. Bioavalability of Organic Pollutants in Boreal Waters with Varying Levels of Dissolved Organic Material. Wat. Res. 25, 455-463.

Landrum, P.L., Nihart, S.R., Eadie,B.J., and Gardner,W.S., 1984. Reverse-phase separation method for determining pollutant binding to Aldrich humic acid and dissolved organic carbon of natural waters, Environ. Sci. Technol. 18, 187-192.

Lassen, P. and Carlsen, L., 1999. The effect of humic acids and the water solubility and water-organic carbon partitioning of fluorene and its NSO-heteroanalogues: Carbazole, dibenzofurane and dibenzothiophene, Chemosphere 38, 2959-2968

Lee, C.-L., and Kuo, L.-J., 1999. Quantification of the dissolved organic matter effect on the sorption of hydrophobic organic pollutant: Application of an overall mechanistic sorption model, Chemosphere 38, 807-821.

Miller, J. C., Miller, J. N., 1988. Statistics for analytical chemistry. 2nd ed. Ellis Horwood, Chichester.

Nielsen, T., Siigur, K., Helweg, C., Jørgensen, O., Hansen, P.E., and Kirso, U., 1997. Sorption of polycyclic aromatic compounds to humic acid as studied by high-performance liquid chromatography. Environ. Sci. Technol. 31, 1102-1108.

NERI National Environmental Research Institute, 2001. The Pesticide and Environmental Database (PATE).

Oesterreich, T., Klaus, U., Volk, M., Neidhart, B., and Spiteller, M., 1999. Environmental fate of amitrole: Influence of dissolved organic matter. Chemosphere 38, 379-392.

Perminova, I.V., Grechishcheva, N.Y., and Petrosyan, V.S., 1999. Relationships between structure and binding affinity of humic substances for polycyclic aromatic hydrocarbons: Relevance of molecular descriptors. Environ. Sci. Technol. 33, 3781-3787.

Steinberg, C.E.W., Haitzer, M., Brüggemann, R., Perminova, I.V., Yashchenko, N.Y., and Petrosyan, S., 2000. Towards a Quantitative Structure Activity Relationship (QSAR) of dissolved humic substances as detoxifying agents in freshwaters. Internat. Rev. Hydrobiol. 85, 253-266.

Piccolo, A., Conte, P., and Saccomandi, F., 1996. Interactions of different classes of herbicides with humic substances from european soils under monocultures and in a climate gradient. In: Del Re, A.A.M., Capri, E., Evans, S.P. and Trevisan, M. (Eds.). X Symposium Pesticide Chemistry - Basic Processes, pp. 239-244.

Schwarzenbach, R. P., Gschwend, P. M., Imboden, D. M., 1993. Environmental organic chemistry. Wiley-Interscience, New York.

Staples, C.A., Peterson, D.R., Parkerton, T.F., and Adams, W.J., 1997. The environmental fate of phthalate esters: A litterature review. Chemosphere 35, 667-749.

Thomsen, M., 2001. QSARs in Environmental Risk Asessment. Interpretation and validation of SAR/QSAR based on multivariate data analysis, PhD-dissertation, Department of Environmental Chemistry, National Environmental Research Institute and Department of Life Science and Chemistry, Roskilde University, DK-4000, Denmark.

Thomsen, M., Lassen, P., Dobel, S., Hansen, P.E., Carlsen, L., and Mogensen, B.B., 2002. Characterisation of humic substances of different origin: A multivariate approach for quantifying the latent properties of dissolved organic matter. Chemosphere x, x-x.

Thomsen, M., and Carlsen, L., 2001. Evaluation of Empirical contra non-empirical descriptors. SAR QSAR Environ. Chem., in press.

Wershaw, R.L., 1999. Molecular Aggregation of Humic Substances. Soil Sci. 164, 803-813.

FIGURE CAPTIONS

Figure 1 Average DOM-normalised partitioning coefficients, K_{DOM}, as function of the concentration of dissolved organic matter, DOM, for Water pond HS (left) and Gohy-573-HS-(H^+)II (right). A decrease in the standard deviations on repeated K_{DOM} measurements, are illustrated by error bars on each average value.

Figure 2 Bi-plots of second versus first principal component, *PC2* versus *PC1*, showing the loading weights, i.e. correlation patterns and importance, of individual original NMR-descriptor with respect to the $K_{DOM.}$

Figure 3 The picture to the left illustrates the calibrated model-predicted versus measured K_{DOM} values. The illustration to the right shows a homogeneous spanned X-space consisting of FA-surface, DE72, Gohy-573-HS-(H^+)II, Kranichsee HA, Purified Aldrich HA and Gohy-573-HA-(H^+)II.

Table 1 Average DOM normalised partitioning, K_{DOM} values for esfenvalerate[a] as function of DOM-concentrations.

DOM [mg/L]	*Aldrich HA (Na+)*	*DE72*	*FA surface*	*Gohy-573-HA-(H+)II*	*Gohy-573-HS-(H+)II*	*Kranichsee HA*	*Purified Aldrich HA (Na+)*	*Water pond HS*
10	23828.5	5316.1	10089.0	7612.2	5701.7	20329.0	28484.0	13282.7
20	18166.5	4588.8	6092.9	5282.3	3124.9	7923.9	16878.0	9213.0
30	11494.2	3338.4	2356.2	3998.7	2214.8	4485.8	13700.8	8014.0
40	9512.0	2294.3	2041.9	3023.8	1558.9	3312.1	9976.4	4942.1
50	8222.0	1902.9	1896.8	3187.1	1090.6	2356.5	8981.0	4227.0
75	6513.8	1513.0	820.0	1857.0	792.3	1776.2	7716.2	2311.3
100	5150.3	1393.1	377.4	2051.8	584.0	1251.2	6232.7	1669.4

[a]The nominal concentration of esfenvalerate was 2.57 µg/L in all partition experiments.

Table 2 Standard deviations on the reproducibility of K_{DOM} measurements.

[a]*Standard deviations on K_{DOM} measurements*							
HS concentration [mg/L]	*10*	*20*	*30*	*40*	*50*	*75*	*100*
Humic substances							
Aldrich HA (Na+)	4186.1	2070.3	2066.3	1144.6	1676.5	533.6	651.8
DE72	2476.2	2514.0	1881.4	976.5	1003.1	406.9	470.1
FA surface	6027.1	622.6	307.1	835.3	984.0	284.8	54.2
Gohy-573-HA-(H+)II	2509.1	1441.1	1164.3	346.8	1079.1	412.9	617.2
Gohy-573-HS-(H+)II	1378.2	747.1	643.7	176.5	173.9	98.9	95.4
Kranichsee HA	3506.0	5202.2	2037.8	1355.0	993.2	75.3	108.9
Purified Aldrich HA (Na+)	936.8	2280.3	1382.6	751.5	1003.8	561.3	121.3
Water pond HS	850.7	2225.9	1284.7	1081.8	1322.4	515.5	498.8

[a] based on triple to quintuple measurements

Table 3 Results of hypothesis testing based on analysis of variance on individual K_{DOM} values, as well as within and between DOM concentration levels.

HS concentration [mg/L]:		10	20	30	40	50	75	100
Humic substances	*F1*[a]			$T1 = \sigma^2_{between,1} > \sigma^2_{within}$				
Aldrich HA (Na^+)	19.353	4.5	7.4	4.5	8.5	3.2	24.9	11.0
DE72	19.353	12.7	5.1	5.4	11.7	9.0	42.8	21.1
FA surface	19.353	2.1	82.4	203.3	16.1	9.3	87.5	1583.8
Gohy-573-HA-(H^+)II	6.094	12.4	15.4	14.1	93.1	7.8	41.6	12.2
Gohy-573-HS-(H^+)II	19.353	41.1	57.2	46.3	359.5	299.2	725.0	511.2
Kranichsee HA	19.353	6.4	1.2	4.6	6.1	9.2	1251.7	392.7
Purified Aldrich HA	19.353	89.0	6.1	10.0	19.8	9.0	22.5	316.6
Water pond HS	19.353	107.9	6.4	11.6	9.6	5.2	26.7	18.7
				Rejection (N) or acceptance (Y) of H_1				
Aldrich HA (Na^+)		N	N	N	N	N	Y	N
DE72		N	N	N	N	N	Y	Y
FA surface		N	Y	Y	N	N	Y	Y
Gohy-573-HA-(H^+)II		Y	Y	N	Y	Y	Y	Y
Gohy-573-HS-(H^+)II		N	Y	Y	Y	Y	Y	Y
Kranichsee HA		N	N	N	N	N	Y	Y
Purified Aldrich HA		Y	N	N	Y	N	Y	Y
Water pond HS		Y	N	N	N	N	Y	N
Humic substances	*F2*[b]			$T2 = \sigma^2_{between,2} > \sigma^2_{within}$				
Aldrich HA (Na^+)	19.296	2.6	10.7	10.8	35.1	16.3	161.3	108.1
DE72	19.296	0.4	0.4	0.7	2.5	2.4	14.5	10.9
FA surface	19.296	0.3	31.4	129.0	17.4	12.6	150.0	4136.3
Gohy-573-HA-(H^+)II	6.256	0.7	2.0	3.0	34.1	3.5	24.1	10.8
Gohy-573-HS-(H^+)II	19.296	1.7	5.8	7.8	103.4	106.6	329.5	353.8
Kranichsee HA	19.296	4.4	2.0	13.1	29.6	55.2	8006.1	3825.1

Purified Aldrich HA	19.296	61.6	10.4	28.3	95.7	53.7	171.6	3676.2
Water pond HS	19.296	23.9	3.5	10.5	14.8	9.9	65.0	69.4

	Rejection (N) or acceptance (Y) of H_2						
Aldrich HA (Na^+)	N	N	N	Y	N	Y	Y
DE72	N	N	N	N	N	N	N
FA surface	N	Y	Y	Y	N	Y	Y
Gohy-573-HA-(H^+)II	N	N	N	Y	N	Y	Y
Gohy-573-HS-(H^+)II	N	N	N	Y	Y	Y	Y
Kranichsee HA	N	N	N	Y	Y	Y	Y
Purified Aldrich HA	Y	N	Y	Y	Y	Y	Y
Water pond HS	Y	N	N	N	N	Y	Y

[a] The degrees of freedom for the variance between DOMs in the H_1-test is seven (N-1), as there are eight different DOMs at each concentration level.

[b] The degrees of freedom for the variance for each DOM between DOM concentration levels in H_2-test is five (N-1), as there are six concentration levels of DOM

[a,b] The degrees of freedom for the variance within each K_{DOM} measurement is from two and four, as the K_{DOM} measurements was reproduced three to five times.

Table 4 Model performance parameters of reverse QSAR models for estimating the sorption of esfenvalerate to DOM of different origin

Descriptors	*Endpoint*	*N*	R^2	Q^2	*RMSEC*	*RMSEP*	*ExpXcal*	*ExpXval*	*ExpYcal*	*ExpYval*
NMR	LogK_{DOM}(100)	3	0.99	0.65	0.036	0.325	82	27	97	61
NMR/UV	LogK_{DOM}(100)	3	0.99	0.60	0.031	0.340	74	19	98	56
NMR	LogK_{DOM}(75)	2	0.94	0.33	0.083	0.385	28	39	20	67
NMR/UV	LogK_{DOM}(75)	2	0.93	0.28	0.088	0.386	63	2	86	31

N is the number of principal components included in the model, R^2 the correlation coefficient, Q^2 the cross-validated correlation coefficient. *RMSEC* is the root mean square of error of the model, RMSEP the root mean square error of predictions. *ExpXcal* is the total explained X-variance used for explaining the variation in logK_{DOM} in the calibrated model, *ExpXval* the total explained X-variance used for explaining the variation in logK_{DOM} in the cross-validated model. *ExpYcal* is the total explained variance in logK_{DOM} , *ExpYval* the total explained variance in logK_{DOM} by cross-validation .

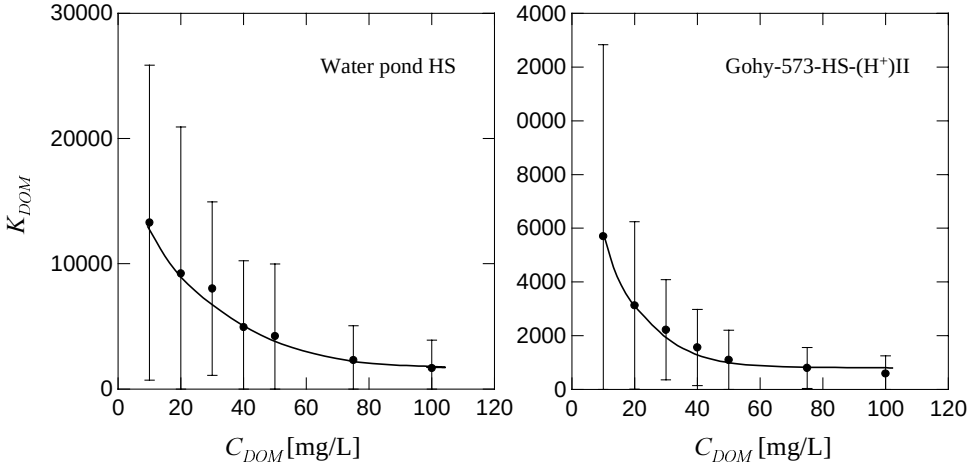


Figure 1

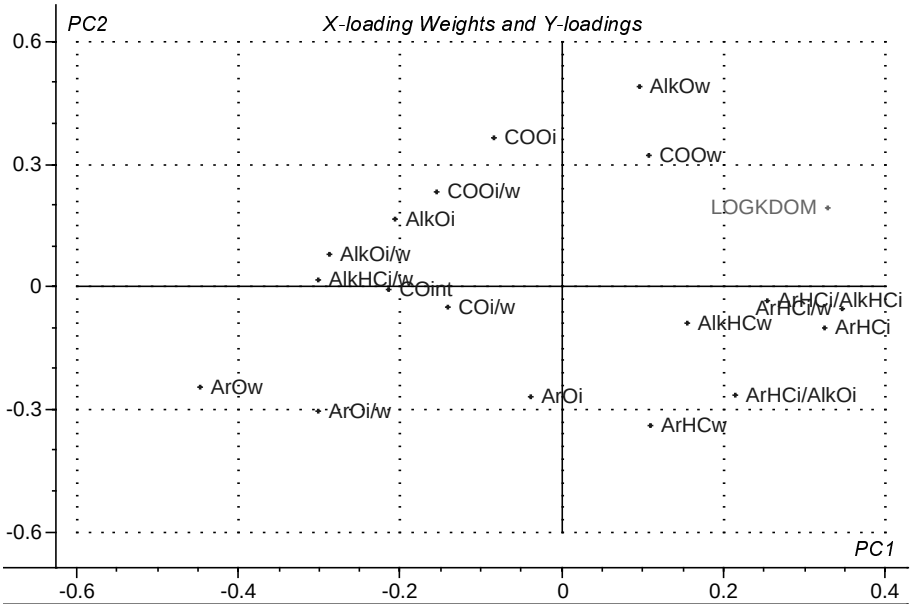

X-loading Weights and Y-loadings
PC2
PC1
0.6
0.3
0
-0.3
-0.6
-0.6
-0.4
-0.2
0
0.2
0.4
AlkOw
COOi
COOw
COOi/w
AlkOi
LOGKDOM
AlkOi/w
AlkHCi/w
COint
COi/w
ArHCi/AlkHCi
ArHCi/w
AlkHCw
ArHCi
ArOw
ArOi
ArHCi/AlkOi
ArOi/w
ArHCw

Figure 2

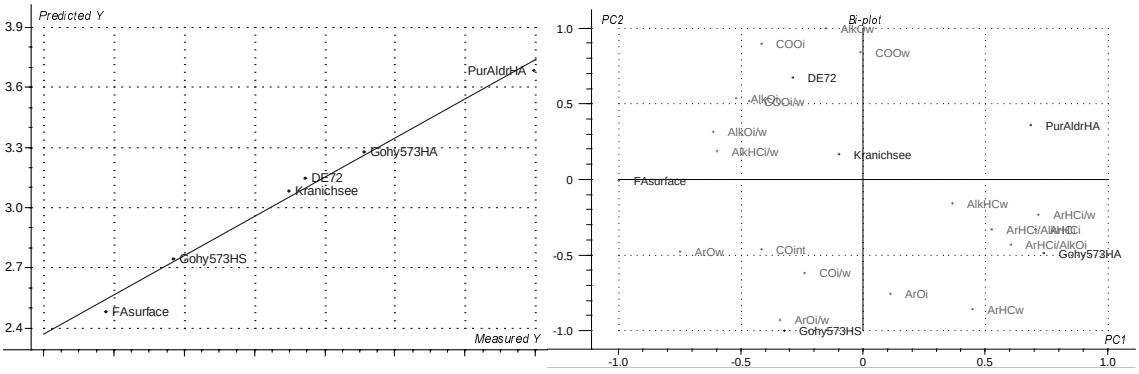

Predicted Y
Measured Y
3.9
3.6
3.3
3.0
2.7
2.4
PurAldrHA
Gohy573HA
DE72
Kranichsee
Gohy573HS
FAsurface
Bi-plot
PC2
PC1
1.0
0.5
0
-0.5
-1.0
COOi
COOw
DE72
AlkOi/w
AlkHCi/w
Kranichsee
PurAldrHA
FAsurface
AlkHCw
ArHCi/w
ArOw
COint
COi/w
ArOi
ArHCw
ArOi/w
Gohy573HS
Gohy573HA

Figure 3

Paper VI

Sensitivity analysis of calculated exposure concentrations and dissipation of DEHP in a topsoil compartment – The influence of the third phase effect and Dissolved Organic Matter (DOM)

Submitted to *The Science of the Total Environment*

Sensitivity analysis of calculated exposure concentrations and dissipation of DEHP in a topsoil compartment

– The influence of the third phase effect and Dissolved Organic Matter (DOM)

Patrik Fauser* and Marianne Thomsen

National Environmental Research Institute, Department of Environmental Chemistry, Frederiksborgvej 399, postbox 358, DK- 4000 Roskilde, Denmark

* Corresponding author, tel.: +45 46 30 12 36, fax: +45 46 30 11 14, e-mail: paf@dmu.dk

Abstract

The fate and risk assessment of hydrophobic substances in the terrestrial environment can be associated with large errors. These can be attributed to the partitioning and process coefficients derived in experimental studies and to the model set-up that is designed to calculate the exposure concentrations. In many cases the concentration of xenobiotics are low in the environment, which gives the aqueous phase the characteristics of a true solution, which are in accordance with the thermodynamic description of dilute solutions. Under these circumstances the conventional equilibrium coefficients, such as K_d, Henry's Law constant H, and the bioconcentration factor, BCF, are independent of the activity coefficient of the partitioning compound in the respective phases. However, for hydrophobic substances, these coefficients are often measured in laboratory experiments where the nominal concentration levels are above the substance saturation point within the bulk water phase. In the case of the phthalates, the hydrophobic effect induces the formation of microdroplets (third phase) in the bulk water phase, by which the system is characterised as a heterogeneous mixture. Consequently the linearity between dissolved and sorbed concentration is no longer true. Furthermore, in the terrestrial and aquatic environment the presence of natural Dissolved Organic Matter (DOM) will have an influence on the fate and effects of hydrophobic substances. Hydrophobic compounds show large affinity for sorption to DOM, and contrary to Fixed Organic Matter (FOM) DOM is mobile and can be transported through the soil pores with the advective flow. It is therefore crucial that dispersed or emulsified phases within the continuous aqueous phase, e.g. DOM and microemulsions of phthalates, are distinguished from true solutions in the experimental

measurements of partitioning coefficients, e.g. in order not to underestimate the mobility of sorbed substance. These aspects are treated in this study, where the exposure concentration, vertical transport and microbial degradation of Di-(2-ethylhexyl)-phthalate (DEHP) is modelled in an organic rich topsoil compartment, using experimental partitioning coefficients and degradation rates from the literature. Two model set-ups are derived for the topsoil compartment, i.e., 1) A system with dilute solution of substance, and 2) A system with the presence of a third phase of microdroplets. In both models the presence of DOM is incorporated. The first model shows that the error in the calculated exposure concentration by using partitioning coefficients derived under unfavourable experimental conditions compared to realistic conditions amounts to 1400 %. A comparison between the two models shows, that when emulsion formation is not incorporated in the model, the calculated flux will be overestimated with a factor of 60.

Keywords: Third phase, microdroplets, compartment modelling, dissolved organic matter, DOM, DEHP, topsoil, degradation, sorption, evaporation, bioaccumulation

Introduction

The phthalates were introduced in the 1920's as softeners in plastic materials and are among the most important chemicals in various industrial products. The predominant use is as additives in polyvinylchlorid (PVC) where the presence of phthalates gives rise to products that are soft and workable (e.g. Poppe, 1986). They are not chemically bound in the plastic matrix which enables them to migrate to the surface of the material where they can be transported to the surrounding environment, e.g. air, water, soil etc (e.g. Smistad and Waaler, 1989; Furtmann, 1996). Di-(2ethylhexyl)-phthalate (DEHP) is one of the most commonly used phthalates and due to its suspected hormone-disrupting effects (Nielsen et al., 1996), there is a great need to investigate the fate and effects of this substance in the environment.

Inherent in most mathematical compartment models is the assumption of equilibrium distribution, i.e. a constant ratio between the concentrations in organic matter, i.e. soil and organic tissue, and air and the exposure concentration. The equilibrium state being defined at the point of equal chemical

potentials of the specific substance in the respective phases of the compartment system. The exposure concentration is the concentration of substance in its unimeric form in the bulk water. However, hydrophobic substances, such as the phthalates have low water solubilities and when the concentration exceeds the unimeric saturation point, the water phase has the physicochemical characteristics of an emulsion (Hiemenz and Rajagopalan, 1997; Thomsen et al., 2001a and 2001b). The bulk phase now comprises the unimeric molecules as well as the microdroplets distributed homogeneously in the water phase, on account of a density similar to that of water. In relation to risk assessment the exposure concentration must still be related to the dissolved unimeric molecules only, as several studies have shown that only the dissolved fraction of phthalates in bulk water are bioavailable and biodegradable. However, if present in real systems, the third phase effect needs to be included in order to quantify the fate processes, e.g. mobility, correctly.

If the concentration levels in the experiments for measuring the partitioning coefficients, such as the suspended matter-water partitioning coefficient (K_d), Henry's Law Constant (H) and the bio-concentration factor (BCF), is not well below the bulk water saturation point, then these parameters can be significantly underestimated (Thomsen et al., 2001a and 2001c). This creates a great challenge with respect to the experimental determination of the process parameters that are used in the models. A first step must be to know, or measure the solubility, C_w^{sat}, of the respective substances. The experimental set-up and analytical methods must be very precise and sensitive due to the low concentration levels, which must be held below saturation if realistic data on bioconcentration and dose-response measurements are to be obtained.

Conventional methods, e.g. shake-flask methods, have been used to determine the solubilities and partitioning coefficients of phthalates (Staples et al., 1997), but the results showed significant variations on account of the similar densities of phthalates and water and the presence of micro-droplets (Thomsen et al., 2001a). This is due to the inability to discriminate between a true solution

and a mixture. However a tensiometric approach has been used for determining the bulk activity of the unimeric solutes through measurements of the activity of the solute molecules at the surface (Thomsen et al., 2001a and b). As such the true, i.e. unimeric, solubility of e.g. DEHP has been found to be $C_w^{sat}(DEHP) = 1.7 \cdot 10^{-2}$ mg pr. litre, which is a factor of approximately 20 lower compared to previous investigations (Thomsen et al., 2001a), but that are in accordance with the UNIFAC estimates (Thomsen et al., 1998).

Hydrophobic substances show large affinity of sorption to Dissolved Organic Matter (DOM). DOM originates from decomposition of plant and animal residues and consists of organic macromolecules resembling the fixed organic matter (FOM) in soils and sediments. The molecular weights range from 600 to 500,000 Da and the chemical composition depends on the type and the origin. DOM can be present in soil and sediment pore water in concentrations of 10 – 1000 mg C pr. litre (Caron and Suffet, 1989). At low concentrations the DOM molecules can be described as subunits that are dissolved in the bulk phase. However, in the environment the DOM concentration in most cases, exceeds the solubility point, where the macromolecules aggregate to form a dispersion (Thomsen et al., 2001c). Under such conditions the soil compartment consists of a fixed phase (FOM) and a dispersion (DOM), that both have sorption potential towards hydrophobic substances. In the terrestrial environment sorption is significant, and accordingly the substance bioavailability, degradability and transport, in relation to surface runoff and advective flow, is affected (Landrum et al., 1987; Matthiessen, 1994; Kukkonen and Oikari, 1991; Tanaka et al., 1997; Cao et al., 1999). Proper investigations of the influence of the presence of DOM in relation to the fate and effects of xenobiotics has only been sparsely elucidated in the literature to date (Schlautman and Morgan, 1993; Lassen et al., 1997; Schulze at al., 1999; Tao and Lin, 2000). For investigating the effect of DOM, the methods used in experiments for measuring partition coefficients must be able to separate and detect components of a true solution versus a dispersion, e.g. the DOM macromolecules, substances sorbed to DOM as well as freely dissolved substances (Eadie et al.,

1990). So far the majority of experimental investigations does not discriminate between substances sorbed to mobile and fixed organic matter, but only quantifies the substance affinity for sorption to organic matter.

In this work emphasis is on the influence of the characteristics of the aqueous phase, of experimentally measured “apparent” partitioning coefficients (Schwarzenbach and Gschwend, 1993; Mackay, 1991). A sensitivity analysis is performed, where the measured coefficients are applied in a fate model of DEHP in a topsoil compartment. This comprises sorption to soil, biota, distribution to the soil pore water, diffusive and advective vertical flow as well as bio-degradation. Exposure concentration profiles and the vertical flux of substance from the topsoil compartment are used as quality parameters and the errors in using process parameters from inappropriate laboratory conditions are quantified.

Theory

In fate assessment models the environment is divided into air, water and soil compartments. Each compartment comprises air, liquid and solid phases. The fate, i.e. the exposure concentrations, of substances is found from measurements and/or from calculations of the transport, removal and distribution processes that occur in and between phases.

In this work a simple mass balance is set up that describes the vertical concentration profile and flux of DEHP in a typical topsoil compartment. In Fig. 1 a three-dimensional infinitesimal topsoil unit is shown. The soil unit is repeated infinitely throughout the topsoil compartment and due to symmetry, the net horizontal substance transport will be zero. Accordingly only substance flux in the vertical direction, due to advection and the vertical concentration gradient, will be considered.

DEHP is hydrophobic, implying that apart from being in the aqueous phase as freely dissolved molecules, $\underline{C}_w$, it can occur in the following phases:

- Microemulsion phase, $\underline{C}_{emul}$.
- Sorbed to fixed organic material, $\underline{C}_{FOM}$.
- Sorbed to dissolved organic material, $\underline{C}_{DOM}$.
- Vapour in soil air, $\underline{C}_{air}$.
- Sorbed to lipids in organisms, $\underline{C}_{org}$.

where $\underline{C}$ denotes the concentration of DEHP in mass per volume. The phases are considered to be homogeneously mixed throughout the topsoil compartment.

In spite of the high hydrophobicity of DEHP, the concentration in the environment is in most cases below saturation, which will rule out the presence of emulsions in the aqueous phase. Because of dilution, the activity coefficient of the aqueous DEHP will be unity and the activity will be equal to the concentration. This very important assumption of dilute solution is treated in depth in (Thomsen et al., 2001a and 2001c), and the consequences with respect to this study will be considered below.

Most distribution processes in the terrestrial environment are rapid compared to degradation and transport processes. Studies on sorption kinetics have shown an initial phase, lasting few minutes to hours, where up to 50 % of the dissolved substance is rapidly sorbed (Brusseau and Rao, 1989) and a remaining sorption in the following days to months. Biodegradation half-lives for xenobiotics, and the hydraulic retention time due to diffusive and advective transport in the terrestrial compartment, varies from months to years. Therefore it is reasonable to assume equilibrium conditions for the

distribution processes. At equilibrium, the chemical potentials of the substance in the aqueous and solid/air/lipid phases are equal, (e.g. Winn, 1995) and the following distribution equilibrium expressions are used

Sorption: $$K_{DOM} = \frac{C_{DOM}}{C_w},\ K_{FOM} = \frac{C_{FOM}}{C_w} \quad \left(\frac{litre\ water}{kg\ dry\ matter\ soil} \right) \tag{1}$$

Evaporation: $$H = \frac{C_{air}}{C_w} \quad \left(\frac{Pa \cdot m^3}{mol}\ or\ \frac{litre\ water}{litre\ air} \right) \tag{2}$$

Bio-accumulation: $$BCF = \frac{C_{org}}{C_w} \quad \left(\frac{litre\ water}{mg\ lipid\ in\ organism} \right) \tag{3}$$

Where $\underline{K_{DOM}}$ is the partition coefficient between water and dry DOM, $\underline{K_{FOM}}$ is the partition coefficient between water and dry FOM, $\underline{H}$ is Henry's Law Constant and $\underline{BCF}$ is the bio-concentration factor.

In the topsoil unit in Fig. 1, the total substance concentration in mass per total volume is thus

$$C_{total} = C_w \cdot \left(\theta_w + K_{DOM} \cdot M_{D.DOM} + K_{FOM} \cdot M_{D.FOM} + H \cdot \theta_{air} + BCF \cdot M_{lipid\ org} \right) = C_w \cdot R \tag{4}$$

where $\underline{C_{total}}$ is mass total DEHP per total volume, $\underline{C_w}$ is mass dissolved DEHP per water volume, $\underline{\theta_w}$ is water volume per total volume, $\underline{M_{D.DOM}}$ is mass dry DOM per total volume, $\underline{M_{D.FOM}}$ is mass dry FOM per total volume, $\underline{\theta_{air}}$ is air volume per total volume and $\underline{M_{lipid\ org}}$ is mass lipid in organisms per total volume. $\underline{R}$ is the retention factor in water volume per total volume (e.g. Fauser et al., 2001).

The general non steady-state mass balance for the total substance in the topsoil unit is given by:

$$\frac{\partial C_w \cdot R}{\partial t} = \frac{N}{\Delta z} - \frac{\left(N + \frac{\partial N}{\partial z} \cdot \Delta z\right)}{\Delta z} - \sum k_1 \cdot C_w \cdot \theta_w = -\frac{\partial N}{\partial z} - \sum k_1 \cdot C_w \cdot \theta_w \quad (5)$$

where N is the vertical flux, in mass total substance per surface area per time and Σk_1 is the sum of biotic and abiotic pseudo 1st order process rates in second^{-1}. In this study only bio-degradation is included.

Dissolved concentration is **below** saturation

Due to molecular diffusion and advection N can be expressed as follows:

$$N = - D_{diff} \cdot \frac{\partial C_w \cdot \theta_w}{\partial z} + \frac{q \cdot C_w \cdot (\theta_w + K_{DOM} \cdot M_{D.DOM})}{\beta_w} \quad (6)$$

where D_{diff} is the molecular diffusion coefficient in m^2 per second and q is the vertical flow in dm water per second. The mean soil pore diameter is assumed to be larger than the characteristic diameter of the substance emulsions and dissolved DOM molecules, which enable them to be transported by advection. β_w is the fraction of the total unit depth that consists of water. It is equal to θ_w but has the units m water pr. m total.

Insertion of Equation 6 in Equation 5 gives:

$$\frac{\partial C_w \cdot R}{\partial t} = D_{diff} \cdot \frac{\partial^2 C_w \cdot \theta_w}{\partial z^2} - \frac{q \cdot (\theta_w + K_{DOM} \cdot M_{D.DOM})}{\beta_w} \cdot \frac{\partial C_w}{\partial z} - k_1 \cdot C_w \cdot \theta_w \quad (7)$$

In this study the fluctuations originating from changes in precipitation intensity, q, or inlet substrate concentration are not of interest and will be set to constants. Focus will be on the influence of the estimated distribution and transport coefficients on the vertical distribution profile and flux. Under such circumstances, the situation will be steady-state and Equation 7 can be simplified to the following homogeneous 2nd order differential equation:

$$0 = D_{diff} \cdot \frac{\partial^2 C_w \cdot \theta_w}{\partial z^2} - \frac{q \cdot (\theta_w + K_{DOM} \cdot M_{D.DOM})}{\beta_w} \cdot \frac{\partial C_w}{\partial z} - k_1 \cdot C_w \cdot \theta_w \quad \Leftrightarrow$$

$$0 = \frac{\partial^2 C_w}{\partial z^2} - \frac{q \cdot (\theta_w + K_{DOM} \cdot M_{D.DOM})}{D_{diff} \cdot \theta_w \cdot \beta_w} \cdot \frac{\partial C_w}{\partial z} - \frac{k_1 \cdot C_w}{D_{diff}} \quad (8)$$

With the boundary conditions: $C_w \rightarrow 0$ for $z \rightarrow \infty$ (assuming same concentration profile below topsoil compartment).

$C_w = C_{w,surface}$ for $z = 0$

the solution becomes (Spiegel, 1968):

$$C_w(z) = C_{w,surface} \cdot \exp\left(\left(\frac{A}{2} - \sqrt{\left(\frac{A}{2}\right)^2 + \frac{k_1}{D_{diff}}}\right) \cdot z\right) \quad (9)$$

where $A = \frac{q \cdot (\theta_w + K_{DOM} \cdot M_{D.DOM})}{D_{diff} \cdot \theta_w \cdot \beta_w}$ (10)

The vertical substance flux at steady-state, at any depth, z, in the topsoil compartment, can be found by differentiating Equation 9 and insertion in Equation 6:

$$N(z) = - D_{diff} \cdot \theta_w \cdot A \cdot C_{w,surface} \cdot \exp\left(\left(\frac{A}{2} - \sqrt{\left(\frac{A}{2}\right)^2 + \frac{k_1}{D_{diff}}}\right) \cdot z\right) \cdot \left(\left(\frac{1}{2} - \sqrt{\left(\frac{1}{2}\right)^2 + \frac{k_1}{D_{diff} \cdot A^2}}\right) - 1\right) \quad (11)$$

Dissolved concentration is **above** saturation

When environmental concentrations exceed saturation an approach to calculate the concentration profile and the flux is to differentiate the apparent dissolved phase into a

1) Unimeric phase, that participates in partitioning processes, degradation, diffusion and advection.
2) Emulsion that only participates in advection.

The unimeric concentration in the water bulk phase is assumed to be constantly equal to $C_{w,sat}$ throughout the topsoil compartment. The difference between the nominal apparent dissolved

substance concentration, C_w^{app}, and the saturation concentration will be the emulsion phase: $C_w^{app} = C_{w,sat} + C_{emul}$.

Equilibrium partitioning of the substance between soil pore water and DOM is assumed, and at constant DOM concentrations the influx and outflux of sorbed substance, within a topsoil unit, are therefore equal. The advective flow of DOM will thus not contribute to changes in the concentration gradient within the topsoil compartment. The steady-state mass balance, analogous to Equation 8, becomes

$$0 = 0 - \frac{q}{\beta_w} \cdot \frac{\partial C_w^{app}}{\partial z} - k_1 \cdot C_{w,sat} \qquad (12)$$

The diffusion term will be zero, because the unimeric concentration will be $C_{w,sat}$ throughout the topsoil compartment, i.e. the term $\frac{\partial C_{w,sat}}{\partial z}$ is zero.

With the boundary condition

$C_w^{app} = C_{w,surface}^{app}$ for $z = 0$, the solution becomes

$$C_w^{app}(z) = C_{w,surface}^{app} - \frac{k_1 \cdot C_{w,sat} \cdot \beta_w}{q} \cdot z \qquad (13)$$

The vertical flux expression will again only include advection because the unimeric concentration gradient is zero.

$$N(z) = q \cdot \left(\frac{C_w^{app} \cdot \theta_w}{\beta_w} - \frac{k_1 \cdot \beta_w \cdot \theta_w \cdot C_{w,sat} \cdot z}{q} + \frac{C_{w,sat} \cdot K_{DOM} \cdot M_{D,DOM}}{\beta_w} \right) \quad (14)$$

Results and discussion

Experimental and model parameters

The derived equations will be used to investigate the influence of the third phase effect, e.g. the presence of microemulsions of pure phthalate and the presence of DOM dispersed in the aqueous phase, on the fate of DEHP in the topsoil compartment. In Table 1 the used experimental data from the literature is stated. Saturation is not exceeded, where either the substance concentration is stated to be below saturation or the measured equilibrium partitioning coefficients are the highest found. Data for saturated conditions are identified where the substance concentration is stated to be above saturation or the measured equilibrium partitioning coefficients are the lowest found.

Partitioning coefficients and degradation rates are defined for the following experimental conditions:

Case A: Partitioning coefficients and degradation rate are found for measured systems, where the aqueous concentration of DEHP is below saturation. DOM is separated and sorption specifically to DOM has been measured. The experimental conditions describe a three-phase system comprising 1) Unimeric substance molecules dissolved in the bulk phase 2) Dispersion of DOM molecules in the bulk water phase and 3) FOM phase.

Case B: Partitioning coefficients and degradation rate are found for measured systems, where the aqueous concentration of DEHP is below saturation. Same data as in case A. DOM is not separated from FOM and accordingly the substance sorbed to DOM is incorporated as sorbed to FOM. The sorption coefficient K_{FOM} is theoretically adjusted as stated in Table 1. The experimental conditions describe a three-phase system comprising 1) Unimeric substance molecules dissolved in the bulk water phase 2) Dispersion of immobile DOM molecules in the bulk water phase and 3) FOM phase.

Case C: Partitioning coefficients and degradation rate are found for measured systems, where the aqueous concentration of DEHP is above saturation. DOM is separated and sorption specifically to DOM has been measured. The sorption coefficient K_{DOM} is theoretically adjusted as stated in Table 1. The experimental conditions describe a four-phase system comprising 1) Unimeric substance molecules dissolved in the bulk phase 2) Dispersion of DOM molecules in the bulk water phase 3) Emulsion of substance in the bulk water phase and 4) FOM phase.

Case D: Partitioning coefficients and degradation rate are found for measured systems, where the aqueous concentration of DEHP is above saturation. Same data as in case C. DOM is not separated from FOM and accordingly the substance sorbed to DOM is incorporated as sorbed to FOM. The sorption coefficient K_{FOM} is theoretically adjusted as stated in Table 1. The experimental conditions describe a four-phase system comprising 1) Unimeric substance molecules dissolved in the bulk water phase 2) Emulsion of substance in the bulk water phase 3) Dispersion of immobile DOM molecules in the bulk water phase and 4) FOM phase.

Case E: Partitioning coefficients and degradation rate are found for measured systems, where the aqueous concentration of DEHP is above saturation. Same data as in case C. DOM is separated from FOM and accordingly the substance sorbed to DOM is incorporated in the apparent dissolved phase. The sorption coefficient K_{FOM} is theoretically adjusted as stated in Table 1. The experimental conditions describe a four-phase system comprising 1) Unimeric substance molecules dissolved in the bulk water phase 2) Emulsion of substance in the bulk water phase 3) Dispersion of DOM molecules in the bulk water phase and 4) FOM phase.

The following data for an organic rich topsoil is used:

Density of soil:	$M_{D.DOM}$ =0.1 g dry matter per litre total.
	$M_{D.FOM}$ = 1.4 g dry matter per litre total.
Water content (field capacity):	θ_w = 0.3 litre water per litre total.
	β_w = 0.3 m per m.
Air content:	θ_{air} = 0.15 litre air per litre total.
Lipid content in organisms:	$M_{lipid\ org}$ = $1 \cdot 10^{-4}$ kg lipid per litre total.
Molecular diffusion of DEHP:	D_{diff} = $2 \cdot 10^{-10}$ m^2 per second.
Vertical water flow:	q = 200 mm per year = $6.34 \cdot 10^{-9}$ m per second.
Depth of topsoil compartment:	$h_{topsoil}$ = 0.2 m.

Environmental dissolved concentration is **below** saturation (Equations 9 and 11)

The coefficient, A, defined in Equation 10 is 172.8 m^{-1} in cases A and C, where DOM follows the advective flow and 105.7 m^{-1} in cases B and D, where DOM is considered to be fixed and immobile, and 105.7 m^{-1} in case E, where DOM is defined as dissolved.

In Table 2 the model set-up is used to calculate the occurrence of DEHP in the different phases, the microbial degradation and the vertical flux caused by molecular diffusion and advection, respectively, for the five cases. In Fig. 2 the vertical profiles (concentration in soil divided by the surface concentration) are shown. In each case the experimentally measured parameters stated in Table 1 are used.

The error that is being done by using partitioning coefficients that are determined under laboratory conditions that are not in accordance with the actual environmental conditions can be found as follows.

The coefficients found in case A, where the dissolved phase only consists of dissolved unimeric substance molecules and DOM, will be the most realistic to use in this situation. If the total concentration at the surface is 50 ng DEHP per. litre, the vertical flux from the topsoil layer and the degradation rate pr. surface area, respectively, can be calculated from Table 2, case A

Flux (diffusion and advection): N(20 cm) = $11 \cdot 50 \cdot 10^{-3} = 0.5$ mg DEHP $\cdot (m^2 \cdot year)^{-1}$

Degradation rate: $k_1 \cdot \overline{C}_w \cdot \theta_w \cdot h_{topsoil} = 28 \cdot 50 \cdot 10^{-3} = 1.4$ mg DEHP $\cdot (m^2 \cdot year)^{-1}$

The exposure concentration profile and DEHP distribution between the phases are shown in Fig. 3.

The error that is being done by using coefficients measured under unfavourable conditions in the model can be seen in Table 3.

In case A, the experiments are performed with substance solutions that are below the aqueous saturation point, which is also the case in most terrestrial compartments. Furthermore case A distinguishes between dissolved and fixed organic matter in the soil compartment. Overall the use of the partitioning coefficients found in case A, are considered to be the most appropriate in fate modelling of hydrophobic substances in soil. From Table 2 (and Fig. 3) it is seen, that the substance sorbed to FOM makes up 94 % of the total substance in the soil, about 2 % is sorbed to DOM, and about 4 % is present as unimeric molecules in the pore water. The fractions in soil air and in organisms are practically negligible. However, in risk studies the effects of the accumulated substance on organisms must be dealt with in detail. The removal rate from the topsoil compartment by bio-degradation is a factor of 3 higher than the removal by vertical transport to the underlying soil. The transport is dominated by advection due to the presence of DOM that can be transported in the soil pore water.

When partitioning coefficients from case B are used, DOM is not considered as dissolved in the bulk water, but fixed and thus not mobile. The fraction of substance sorbed to FOM increases slightly to about 97 %. The change in chemical potentials of the substance is assumed negligible and the substance associated with the remaining media are identical to case A. Due to the reduced mobility of the substance the exposure concentrations decreases with 30 % compared to case A, diffusion decreases in the same order of magnitude, whereas advection decreases 70 %. These underestimations are critical in relation to risk assessment where worst case scenarios are preferred due to the principle of safety. It must therefore be emphasised that the presence of DOM must be taken into account when the fate of, especially hydrophobic, substances is assessed in the terrestrial environment. The decreased bio-degradation compensates the underestimated exposure

concentration to a certain extent, but still the resulting concentration profile, cf. Fig. 3, and vertical flux are underestimated.

In experimental case C the substance concentration in the bulk water phase are above saturation and the presence of DOM is respected. The system thus consists of three phases, a dissolved phase of unimeric substance and DOM molecules, an emulsion of microdroplets in the bulk water and FOM. The crucial point in relation to the realistic coefficient values in case A, is the significant decrease in partition coefficient values by a factor of 10 to 1000, cf. Table 1. Only 65 % of the total substance is sorbed to FOM, whereas about 30 % is now in the bulk water as apparent dissolved substance. This implies that the exposure concentrations, advective as well as diffusive flows are greatly overestimated with 1400 %, 235 % and 350 %, respectively. The calculated degradation rate is increased with 150 %. In relation to risk assessment the overall calculations are on the safe side.

In experimental case D the substance concentration in the bulk water is above saturation, but DOM is assumed fixed. Analogous to the calculations based on case C the coefficients are greatly underestimated leading to high exposure concentrations. The diffusive transport is a factor of 5 higher than the realistic case A, but the advective transport has decreased compared to case C, due to the immobility of DOM. The calculations are with the exception of the overestimated degradation rate, on the safe side.

Finally, case E aggregates the unimeric molecules, microdroplets and DOM in an apparent dissolved phase. The experimental coefficients are similar to those in case D, and the errors compared to case A are approximately the same.

When environmental fate assessment modelling is performed for hydrophobic substances in soil it can be concluded that it is necessary to use partitioning coefficients measured under dilute solution

conditions. In addition, it is important to account for the presence of DOM. Neglecting the mobility of DOM will underestimate the exposure concentration with approximately 20 %. When partitioning coefficients are measured under conditions where the bulk water concentration of the substance is above saturation, the exposure concentrations can be overestimated with over 1000 %!

Environmental dissolved concentration is **above** saturation (Equations 13 and 14)

The concentration of DEHP in natural surface waters exceed their unimeric solubilities in several cases (cf. Thomsen et al., 1998 and references therein). For this reason it is important to be aware of the colloidal nature of the phthalates, and to consider ways to include this in fate studies and thus risk assessments. In this situation the experimentally derived coefficients in case A will again be the most realistic. In the environment the soil will be a three-phase system comprising 1) Unimeric substance molecules and DOM molecules dissolved in the bulk water phase 2) Emulsion of substance in the bulk water phase and 3) FOM phase, corresponding to the experimental case C. However, in case C the partitioning coefficients are the ratio between the substance sorbed to FOM and the unimeric plus the emulsion. This is not correct as the emulsion does not participate in partitioning processes and degradation.

The correct way to calculate the fate of a substances that occur in concentrations above their saturation point is to use experimental coefficients determined as in case A, and assume saturation concentrations in the bulk phase. The difference between the nominal apparent dissolved substance concentration and saturation will be microdroplets. It is important to separate these two phases since the microdroplets do not sorb, bio-concentrate, degrade or diffuse. They will only participate in advective transport along with the DOM molecules.

At saturation in the water phase the total concentration in the topsoil unit is

$$C_{total} = C_{w,sat} \cdot R_A = 17 \text{ g pr. litre water} \cdot 8.69 = 150 \text{ g pr. litre total}$$

When this concentration is exceeded an emulsion will be formed in the water phase. The following surface concentrations are used as an example

$\underline{C_{total}}$ = 170 µg pr. litre total

$\underline{C_{w,sat}}$ = 17 µg pr. litre water

$\underline{C_{emul}}$ = $(170 - 17 \cdot 8.69) \cdot 0.3^{-1} \approx 70$ µg pr. litre water

$$C_{w,surface}^{app} = 17 + 70 \approx 90 \text{ g pr. litre water}$$

The vertical flux from the topsoil layer calculated from Equation 14 and the degradation rate pr. surface area, respectively, are stated below. The coefficients from experimental case A, Table 1, are used.

Flux (diffusion and advection): $\underline{N}(20 \text{ cm}) = 15 \text{ mg DEHP} \cdot (\text{m}^2 \cdot \text{year})^{-1}$

Degradation rate: $k_1 \cdot \overline{C}_w \cdot \theta_w \cdot h_{topsoil} = 7.2 \text{ mg DEHP} \cdot (\text{m}^2 \cdot \text{year})^{-1}$

Equations 13 gives the vertical apparent dissolved, $\underline{C_w}^{app}$, concentration profile in Fig. 4 together with distribution between phases, derived from Table 1, case A.

Equations 9 and 11 are derived under conditions assuming that the concentrations are below saturation. If they are used for a total concentration of $\underline{C_{total}}$ = 170 µg pr. litre total, the following values are found, cf. Table 1.

Flux (diffusion and advection): $N(20\ cm) = 11 \cdot 90 = 990$ mg DEHP $\cdot$ $(m^2 \cdot year)^{-1}$

Degradation rate: $k_1 \cdot \overline{C}_w \cdot \theta_w \cdot h_{topsoil} = 28 \cdot 90 = 2500$ mg DEHP $\cdot$ $(m^2 \cdot year)^{-1}$

Mean exposure concentration: $C_w(mean) = 53$ µg pr. litre

The exposure concentrations calculated with Equation 9, designed for unsaturated conditions, are close to the concentrations calculated with Equation 13 that respects microdroplets. However the calculated vertical transport and degradation rate are greatly overestimated. This is because the entire apparent dissolved phase participates in sorption to DOM and degradation, where it more accurately only should be the unimeric molecules.

The examples show that in addition to using high quality experimental data, it is important to use a model set-up that respects the actual environmental conditions regarding to the presence of a third phase and DOM.

Conclusion

The influence of the assumption of infinite dilute solution on the fate calculations of DEHP

The concentration of hydrophobic xenobiotics in the environment is typically low. However, in the majority of experimental studies that are performed to measure the distribution properties of hydrophobic substances, the nominal concentration is high, due to the desired reliability and reproducibility of the data. The dissolved concentration thus exceeds aqueous saturation and the introduction of a third phase, which does not participate in the distribution processes, can result in crucially underestimated partitioning coefficients. When these coefficients are employed in fate

models the quantification of the transport processes may be overestimated with up to 500 %, and the exposure concentrations with up to 1400 %. This is due to the greatly overestimated concentration of apparent dissolved substance on account of the underestimated partitioning coefficients. The degradation is overestimated with up to 150 %. The overall result will be associated with large error but on the safe side, in relation to risk assessment.

The influence of the presence of DOM on the fate of DEHP

The presence of natural Dissolved Organic Matter (DOM) in soil systems gives rise to an apparent dissolved phase of sorbed substances, that can follow the advective flow, but that will not be bio-degraded or available to organisms. If this effect is to be included in fate models the experimentally determined partitioning coefficients must have been measured under laboratory conditions where DOM is separated from the Fixed Organic Matter (FOM) phase of the soil. If DOM is considered immobile in the laboratory measurements the partitioning coefficients will lead to underestimations of the calculated dissolved concentrations, vertical flux and degradation of approximately 30, 70 and 20 %. Contrary to the errors in neglecting the third phase, the omission of DOM will lead to results that are critical in relation to risk assessment.

The influence of using the correct model set-up with respect to the presence of the third phase

In any case, the correct partitioning coefficients and degradation rates to be used in fate modelling are derived in experimental studies taking the third phase and DOM into account. When this criterion is fulfilled, it is furthermore necessary to use the correct model set-up depending on the environmental concentration of the substance in question. If the environmental concentration is below saturation the diffusion and degradation are calculated from the concentration of dissolved substance and advection is related to dissolved and substance sorbed to DOM. If the environmental

concentration is above saturation, the dissolved unimeric concentration is equal to the saturation concentration in the topsoil compartment, and accordingly the gradient is zero and no diffusive transport takes place. If the DOM concentration is constant throughout the topsoil compartment the concentration of sorbed substance is also constant and DOM will thus not contribute to the vertical concentration gradient. The net flux can be calculated from advection of unimeric molecules, microdroplets and substance sorbed to DOM. Degradation is related to the unimeric substance only.

The error from using a wrong model set-up, when the environmental concentrations exceed saturation, is e.g. an overestimation of the total vertical flux by a factor of 60 and an overestimation of the degradation rate by a factor of 350. The calculated exposure (dissolved unimeric substance) concentrations are, however, only overestimated by a factor of 3.

References

Bellobono IR, Marcandalli B, Selli, E, Polissi, A. A model study for realease of plasticizers from polymer films through vapor phase. J Appl Pol Sci 1984; 29: 3185-3195.

Brusseau ML, Rao PSC. The influence of sorbat-organic matter interactions on sorption nonequilibrium. Chemosphere 1989; 18: 1691-1706.

Cao J, Tao S, Li, BG. Leaching kinetics of water soluble organic carbon (WSOC) from upland soil, Chemosphere 1999; 39: 1771-1780.

Caron G, Suffet IH. Binding of nonpolar pollutants to dissolved organic carbon. Environmental fate modelling. In: Suffet IH, MacCarthy P, editors. Aquatic Humic Substances. Influence on Fate and Treatment of Pollutants. Advances in Chemistry Series, 1989, No. 219, pp. 117-130.

Fauser P, Sorensen PB, Carlsen L, Vikelsøe J. Phthalates and Nonylphenols and LAS in Roskilde Wastewater Treatment Plant. Fate Modelling Based on Measured Concentrations in Wastewater and

Sludge. National Environmental Research Institute, Roskilde. 110 pp. NERI Technical Report. No 354, 2001.

Frissell WJ. Volatility of vinyl plasticizers. Industr Eng Chem 1956; 48: 1096-1099.

Eadie BJ, Morehead NR, Landrum P. Three-phase partitioning of hydrophobic organic compounds in great lakes waters. Chemosphere 1990; 20: 161-178.

Furtmann K. Phthalates in the Aquatic Environment. European Council for Plasticisers & Intermediates 1996.

Gauthier TD, Seitz WR, Grant CL. Effects of structural and compositional variations of dissolved humic materials on pyrene K_{OC} values. Environ Sci Technol 1987; 21: 243-248.

Graham PR. Phthalate Ester Plasticizers-Why and How They Are Used. Environ. Health Perspect 1973; 3: 3-12.

Haitzer M, Höss S, Traunspurger W, Steinberg C. Effects of dissolved organic matter (DOM) on the bioconcentration of organic chemicals in aquatic organisms - A review. Chemosphere 1998; 37: 1335-1362.

Haitzer M, Höss S, Traunspurger W, Steinberg C. Relationship between concentration of dissolved organic matter (DOM) and the effect of DOM on the bioconcentration of benzo(a)pyrene. Aquatic Toxicol 1999; 45: 147-158.

Hiemenz PC, Rajagopalan R. Principles of Colloid and Surface Chemistry. Third Edition, Revised and Expanded. Marcel Dekker, Inc., 1997, 650 pp.

Kukkonen J, Oikari A. Bioavailability of organic pollutants in boreal waters with varying levels of dissolved organic material, Wat Res 1991; 25**:** 455-463.

Kukkonen J, Pellinen J. Binding of organic xenobiotics to dissolved organic macromolecules: Comparison of analytical methods, Sci Tot Environ 1994; 152: 19-29.

Landrum PF, Nihart SR, Eadie BJ, Herche LR. Reduction in bioavailability of organic contaminants to the amphipod *Pontoporeia hoyi* by dissolved organic matter of sediment interstitial water, Environ. Toxicol Chem 1987; 6: 11-20.

Lassen P, Poulsen ME, Stuer-Lauridsen F, Carlsen L. Leaching of selected PAH's nd hetero-analogues from an organic matrix into synthetic ground water. Influence of dissolved humic material. Chemosphere 1997; 34: 335-344.

Mackay D. Multimedia Environmental Models – The Fugacity Approach. Lewis Publishers, 1991, 257 pp.

Matthiessen A. Untersuchungen über den Einfluss von Huminstoffen auf den Transport polycyclischer aromatischer Kohlenwasserstoffe. Vom Wasser 1994; 82: 137-144.

Mato Y, Isobe T, Takada H, Kanehiro H, Ohtake C, Kaminuma T. Plastic Resin Pellets as a tranport medium for toxic chemicals in the marine environment. Environ Sci Technol 2001; 35: 318-324.

Muszkat L. Raucher D, Magaritz M, Ronen D, Amiel AJ. Unsaturated zone and groundwater contamination by organic pollutants in a seawage-effluent-irrigated site. G Water 1993; 31: 556-565.

Nielsen E, Larsen PB. Toxicological evaluation and limit values for DEHP and phthalates, other than DEHP. The Insitute of Toxicology, , NERI technical report, no. 6, National Food Agency, Denmark , 1996, 94 pp.

Perminova IV, Grechishcheva NY, Petrosyan VS. Relationships between structure and binding affinity of humic substances for polycyclic aromatic hydrocarbons: Relevance of molecular descriptors, Environ Sci Technol 1999; 33: 3781-3787.

Poppe AC. Migrationsgeschwindigkeit von phthalatweichmachern in weich-PVC bei raumtemperatur. Kunstst 1986; 76: 583-585.

Quackenbos HM Jr. Plasticizers in vinyl chloride resins. Migration of plasticizer. Industr Eng Chem 1954; 46: 1335-1345.

Schlautman MA, Morgan JJ. Effects of aqueous chemistry on the binding of polycyclic aromatic hydrocarbons by dissolved humic materials. Environ Sci Technol 1993; 27: 961-969.

Schulze M, Wilkes H, Vereecken H. Direct determination of hydrophobic organic compounds in aqueous solution in the presence of dissolved organic carbon by high-performance liquid chromatography. Chemosphere 1999; 39: 2365-2374.

Schwarzenbach RP, Gschwend PM, Imboden DM. Environmental organic chemistry. Wiley-Interscience, New York, 1993, 681 pp.

Seip HM, Alstad J, Carlberg GE, Martinsen K, Skaane R. Measurement of mobility of organic compounds in soil. Sci Tot Environ 1986; 50: 87-101.

Smistad G, Waaler T. Migration of plastic additives from soft polyvinyl chloride bags into normal saline and glucose infusions. Acta Pharm Nord 1989; 1: 287-290.

Staples CA, Peterson DR, Parkerton TF, Adams WJ. The environmental fate of phthalate esters: a litterature review. Chemosphere 1997; 35: 667-749.

Tanaka S, Oba K, Fukushima M, Nakayasu K, Hasebe K. Water solubility enhancement of pyrene in the presence of humic substances, Anal Chim Acta 1997; 337**:** 351-357.

Tao S, Lin B. Water soluble organic carbon and its measurement in soil and sediment. Wat. Res. 2000; 34: 1751-1755.

Thomsen M., Rasmussen AG, Carlsen L. SAR/QSAR approaches to solubility, partitioning and sorption of phthalates, Chemosphere 1998; 38: 2613-2624.

Thomsen M, Carlsen L, Hvidt S. Solubilities and surface activities of phthalates investigated by surface tension measurements, Environm Toxicol Chem 2001a; 20: 127-132.

Thomsen M, Hvidt S, Carlsen L. Solubility of phthalates revisited. Environmental implications. In, Handbook on QSARs for predicting Environmental Fate of Chemicals (J.D. Walker, Ed.). Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA, 2001b, in press.

Thomsen M, Lassen P, Dobel S, Mogensen BB, Carlsen L, Hansen PE. Inverse QSAR for modelling the sorption of esfenvalerate to dissolved organic matter (DOM). A multivariate approach. Submitted to SAR QSAR Environ Chem 2001c.

Winn JS. Physical Chemistry. Harper Collins College Publishers, 1995, 1159 pp.

Table 1 Experimental data for distribution coefficients and degradation rates. The data are shown to be dependent on the experimental condition, i.e. the fraction of substance that is measured as the "dissolved" phase. H is estimated from mean measured vapour pressures of pure substances and experimental C_w values. Units of the distribution coefficients are stated in Equations 1 to 3. The true aqueous saturation concentration of DEHP is $C_{w.sat} = 1.7 \cdot 10^{-2}$ mg per litre (Thomsen et al., 2001a).

Experimental system conditions defining "dissolved" phase.	***Sorption***		***Evaporation*** $H = \frac{C_{air}}{C_w}$	***Bio-accumulation*** $BCF = \frac{C_{org}}{C_w}$	***Bio-degradation*** k_1 ***(hour^{-1})***
	Dissolved: $K_{DOM} = \frac{C_{DOM}}{C_w}$	***Fixed:*** $K_{FOM} = \frac{C_{FOM}}{C_w}$			
A: Below sat. $\gamma = 1$ Only free molecules	*1905*	*5860*	*$1.7 \cdot 10^{-6}$*	*79*	*$8.01 \cdot 10^{-4}$*
B: Below sat. $\gamma = 1$ Only free molecules	*DOM is defined as FOM*	*5600[1)]*			
C: Above sat. $\gamma \neq 1$ Free molecules and emulsion	*150[2)]*	*452*	*$2.7 \cdot 10^{-9}$*	*251*	*$1.45 \cdot 10^{-4}$*
D: Above sat. $\gamma \neq 1$ Free molecules and emulsion	*DOM is defined as FOM*	*430[3)]*			
E: Above sat. $\gamma \neq 1$ Free molecules, emulsion and sorbed to DOM	*Substance adsorbed to DOM is defined as dissolved*	*445[4)]*	*$2.3 \cdot 10^{-9}$ [5)]*	*247[6)]*	

1) $\frac{K^A_{DOM} \cdot M_{D.DOM} + K^A_{FOM} \cdot M_{D.FOM}}{(M_{D.DOM} + M_{D.FOM})}$

2) $\frac{K^C_{FOM}}{K^A_{FOM}} \cdot K^A_{DOM}$

3) $\frac{K^C_{DOM} \cdot M_{D.DOM} + K^C_{FOM} \cdot M_{D.FOM}}{(M_{D.DOM} + M_{D.FOM})}$

4) $\frac{K^C_{FOM}}{(1 + K^C_{DOM} \cdot M_{D.DOM})}$

5) $\frac{H^C}{(1 + K^C_{DOM} \cdot M_{D.DOM})}$

6) $\frac{BCF^C}{(1 + K^C_{DOM} \cdot M_{D.DOM})}$

Table 2 Calculated retention factors (Equation 4), and percentages of substance in aqueous phase as free molecules, on DOM, FOM, in soil air, and in lipid phase of organism compared to the total substance concentration. The calculated mean bio-degradation in the topsoil compartment and vertical substance flux (Equation 11) are divided with the total surface concentration. The activity of the substance corresponds to the defined state in the respective cases and the calculations are based on the distribution coefficients and degradation rates in Table 1.

Distribution coefficients and degradation rates used in calculations	***Retention factor*** $R = \frac{C_{total}}{C}$	***Aqueous free molecules*** $\frac{\theta_w}{R}$	***Sorption***		***Evaporation*** $\frac{H \cdot \theta_{air}}{R}$	***Bio-accumulation*** $\frac{BCF \cdot M_{lipid\,org}}{R}$	***Bio-degradation*** $\frac{k_1 \cdot \overline{C}_w \cdot \theta_w \cdot h_{topsoil}}{C_{w,surface}}$ ***(m · year^{-1})***	***Vertical flux*** ***Equation 11 / $C_{w,surface}$ (m · year^{-1})***	
			Dissolved: $\frac{K_{DOM} \cdot M_{D,DOM}}{R}$	***Fixed:*** $\frac{K_{FOM} \cdot M_{D,FOM}}{R}$				***Diffusion***	***Advection***
A	*8.69*	*3.5 %*	*2.2 %*	*94.3 %*	*~0*	*0.1 %*	*28*	*0.4*	*11*
B	*8.69*	*3.5 %*	*DOM is defined as FOM*	*96.5 %*	*~0*	*0.1 %*	*22*	*0.3*	*3.4*
C	*0.97*	*30.9 %*	*1.5 %*	*65.1 %*	*~0*	*2.6 %*	*70*	*1.8*	*270*
D	*0.97*	*30.9 %*	*DOM is defined as FOM*	*66.6 %*	*~0*	*2.6 %*	*64*	*2.5*	*140*
E	*0.96*	*32.7 %*	*DOM is defined as dissolved*	*64.7 %*	*~0*	*2.6 %*	*64*	*2.5*	*140*

Table 3 Errors in model calculations of vertical flux (Equation 11), bio-degradation and exposure concentrations (Equation 9). The environmental concentrations are below saturation and the coefficients measured in experimental case A are the most appropriate to use in the model. – denotes underestimation, + denotes overestimation and the percentages indicate the deviation from the calculations where case A coefficients are used. The bold figures are the critical cases in environmental fate and risk assessment, and they are commented further in the text.

Distribution coefficients and degradation rates used in calculations	***Vertical flux***		***Bio-degradation***	***Exposure concentration, C_w(mean)***
	Diffusion	***Advection***		
A	*Realistic*	*Realistic*	*Realistic*	*Realistic*
B	***- 25 %***	***- 70 %***	*- 20 %*	***- 30 %***
C	*+ 350 %*	*+ 235 %*	***+ 150 %***	*+ 1400 %*
D	*+ 525 %*	*+ 120 %*	***+ 130 %***	*+ 1300 %*
E	*+ 525 %*	*+ 120 %*	***+ 130 %***	*+ 1300 %*

Figure Legends:

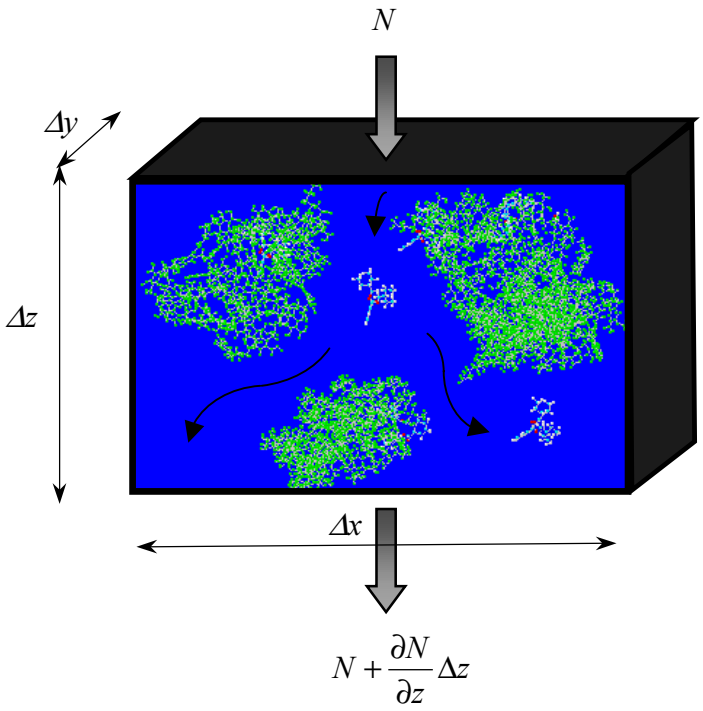
Fig. 1 Infinitesimal topsoil unit that is representative of the topsoil compartment. N denotes the flux of DEHP in mass per surface area per time. Due to symmetry net transport will only occur in the vertical direction. The vertical water flow, $\underline{q}$, is in dm water per second.

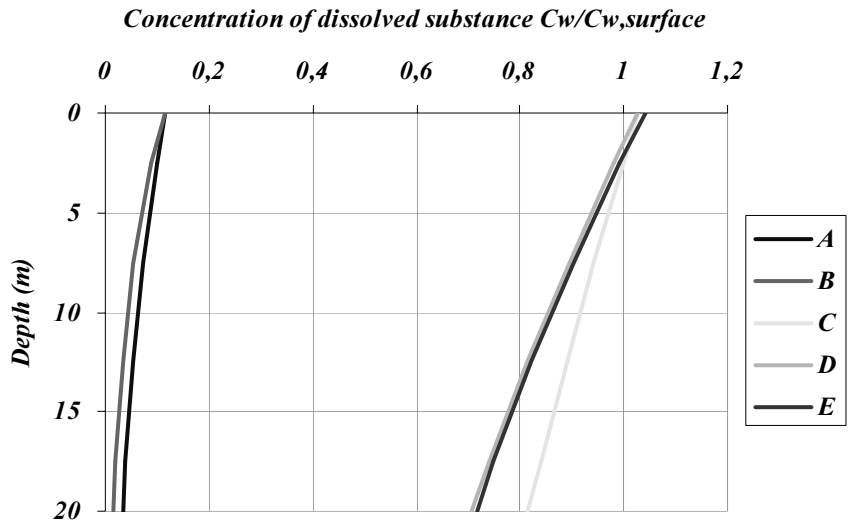
Fig. 2 Calculated vertical exposure (dissolved substance) concentration profiles from Equation 9. The calculations are based on the distribution coefficients and degradation rates in Table 1. In cases A and B the dissolved substance is the free unimeric substance molecules. In case C and D it is the apparent dissolved substance, $\underline{C_w}^{\underline{app}}$, comprising unimeric and emulsion of substance. In case E it is the apparent dissolved substance comprising unimeric and emulsion of substance and substance sorbed to DOM dispersion.

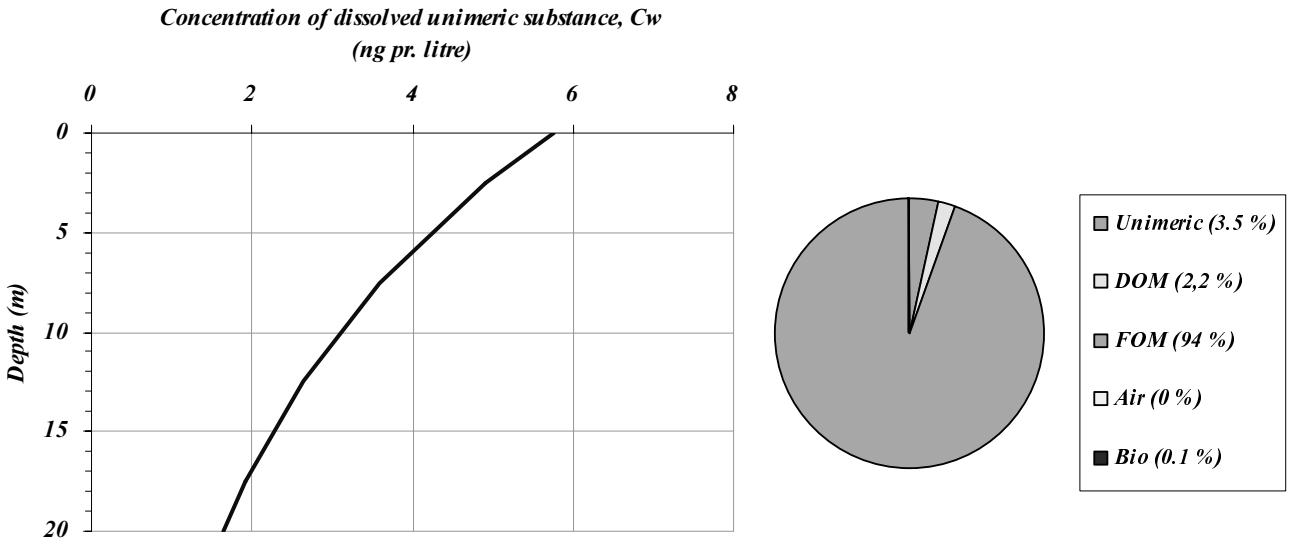
Fig. 3 Calculated exposure concentration profile in topsoil (Equation 9) and distribution of DEHP between different phases in soil, where coefficients found in case A are used. The distribution will be constant throughout the soil depth due to the homogeneous structure of the soil. The total concentration at the surface is 50 ng DEHP pr. litre, corresponding to a unimeric concentration of 6 ng DEHP pr. litre, which is below the saturation point.

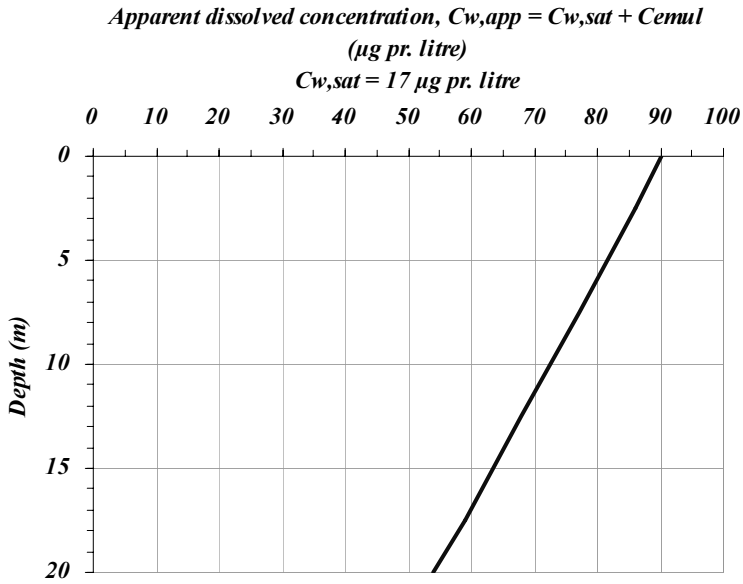
Fig. 4 Calculated apparent dissolved, $\underline{C_w}^{\underline{app}}$, concentration profile in topsoil (Equation 13) and distribution of DEHP between different phases at soil surface, where coefficients found in case A are used. The unimeric concentration will be $\underline{C_{w,sat}}$ and the difference between the apparent and the saturation concentration will consist of microdroplets. These will only participate in advective transport according to Equations 12 to 14. The distribution between phases will be constant throughout the soil depth due to the homogeneous structure of the soil. The total concentration at

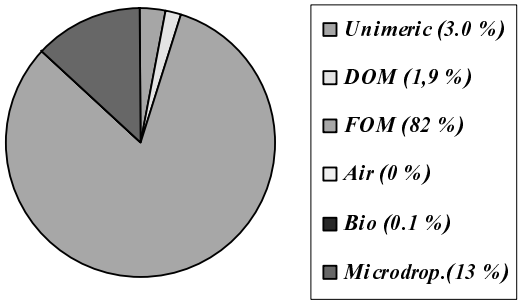
the surface is 170 ng DEHP pr. litre, corresponding to an apparent dissolved concentration of 90 ng DEHP pr. litre, which is above the saturation point.


N
Δy
Δz
Δx
$N + \frac{\partial N}{\partial z}\Delta z$


Concentration of dissolved substance Cw/Cw,surface
0
0,2
0,4
0,6
0,8
1
1,2
Depth (m)
0
5
10
15
20
A
B
C
D
E


Concentration of dissolved unimeric substance, Cw
(ng pr. litre)
0
2
4
6
8
0
5
10
15
20
Depth (m)
Unimeric (3.5 %)
DOM (2,2 %)
FOM (94 %)
Air (0 %)
Bio (0.1 %)


Apparent dissolved concentration, Cw,app = Cw,sat + Cemul
(µg pr. litre)
Cw,sat = 17 µg pr. litre
0
10
20
30
40
50
60
70
80
90
100
Depth (m)
0
5
10
15
20


Unimeric (3.0 %)
DOM (1,9 %)
FOM (82 %)
Air (0 %)
Bio (0.1 %)
Microdrop.(13 %)