APPLICATION OF THE CHARACTERISTIC ROOT INDEX MODEL TO THE ESTIMATION OF PHYSICO-CHEMICAL AND BIOLOGICAL PROPERTIES OF SELECTED ENDOCRINE DISRUPTING CHEMICALS

by

Mustafa Özkul

BS. in Environmental Engineering, İstanbul University, 2002

Submitted to the Institute of Environmental Sciences in partial fulfillment of the requirements for the degree of

Master of Science

in

Environmental Sciences

Boğaziçi University

2005

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis supervisor Assoc. Prof. Dr. Melek Türker Saçan for her great deal of interests, encouragement, guidance and support for completing my thesis.

I would like to thank my jury members Prof. Dr. Işıl Balcıoğlu and Assoc. Prof. Dr. Safiye Erdem for spending their valuable time for my thesis.

I wish to express my thanks to my friends Ayhan Yılmaz, İsmet Gül, Kamil Varınca, and Serkan Arslan.

I am thankful to my family for encouraging me to start and finish this study.

I owe a special thanks especially to my dear parents Sıdıka and Eyüp Özkul for their extreme patience, support and love.

ABSTRACT

Quantitative structure-property relationship (QSPR) models based on two different sets of parameters have been developed for water solubility (*S*), *n*-octanol/water partition coefficient (K_{OW}), and Henry's Law Constant (*H*) of selected endocrine disrupting chemicals (EDCs). Similarly, QSPR models have also been established for the biological properties including 50% effective inhibition concentration (48h-EC_{50}) to algae and bioconcentration factor (*BCF*) in fish for substituted benzenes having endocrine disrupting properties. The parameters used to develop these QSPR models were topology based Characteristic Root Index (*CRI*) and semi-empirical molecular descriptors, namely – energies of the highest occupied and the lowest unoccupied molecular orbital (E_{HOMO} and E_{LUMO}), and dipole moment (μ).

The best fit equation found by "forward multiple linear regression" showed that the topology based *CRI* was the most important parameter for the modeling of *S*, K_{OW} and *BCF*. For K_{OW} a two-parameter equation including the *CRI* and E_{HOMO} was obtained with a correlation coefficient of $r = 0.992$, whereas a three-parameter equation including the *CRI*, E_{LUMO} and μ was obtained for both solubility and Henry's Law Constant with a correlation coefficient of $r = 0.986$ and $r = 0.933$, respectively. E_{HOMO} and μ didn't appear in the same model because of the collinearity. The *CRI* and E_{LUMO} descriptors were used in regressions of 48h-EC_{50} with a correlation coefficient of $r = 0.926$ while only the *CRI* had appeared as a descriptor for the prediction of *BCF* with a correlation coefficient of $r = 0.850$. Since the number of *BCF* data is limited to 18 for the *CRI*-based QSPR model (model 6), reliability of this model was first checked by classical tests, and then it was compared with the other *CRI*-based model (model 7) reported in the literature. The results of modified jackknife tests indicated that the five models were statistically robust. Mean deviation of calculated values from experimental data amounted to 0.27, 0.17, 0.28, 0.19 and 0.29 log units for the *S*, K_{OW}, *H*, 48h-EC_{50}, and *BCF*, respectively. The developed models have been used to predict the *S*, K_{OW}, *H*, 48h-EC_{50}, and *BCF* of those compounds where there are no experimental measurements.

ÖZET

Seçilen hormon bozucu kimyasalların (EDCs) sudaki çözünürlüğü (*S*), *n*-oktanol/su katsayısı (K_{OW}) ve Henry Sabiti (*H*) için iki farklı parametre grubuna dayalı miktarsal yapı-özellik ilişkisi (QSPR) modelleri geliştirilmiştir. Benzer şekilde, hormon bozucu özelliğe sahip benzen türevleri için, alglerdeki %50 etkili inhibisyon konsantrasyonu (48h-EC_{50}) ve balıklardaki biyokonsantrasyon faktörü'nü (*BCF*) içeren biyolojik özelliklerin tahmin edilmesi için de miktarsal yapı-özellik ilişkisi modelleri geliştirilmiştir. Bu miktarsal yapı-özellik ilişkisi modellerini geliştirmek için kullanılan parametreler; topolojiye bağlı Karakteristik Kök İndeksi (*CRI*) ile en yüksek seviyedeki dolu moleküler orbital, en düşük seviyedeki boş moleküler orbital enerjileri (E_{HOMO} ve E_{LUMO}) ve dipol momenti (μ) içeren yarı-deneysel moleküler tanımlayıcılardır.

"İleri yönde çok değişkenli doğrusal regresyon" kullanılarak bulunan en iyi denklemler, topolojiye bağlı Karakteristik Kök İndeksi'nin *S*, K_{OW} ve *BCF* modellerinde en önemli parametre olduğunu ortaya koymuştur. *S* ve *H* için *CRI*, E_{LUMO} ve μ içeren üç parametreli ve sırasıyla $r = 0.986$ ve $r = 0.933$ korelasyon katsayılı modeller elde edilirken, K_{OW} için *CRI* ve E_{HOMO} içeren $r = 0.992$ korelasyon katsayılı iki parametreli bir model elde edilmiştir. E_{HOMO} ve μ ilişiklik yüzünden aynı modelde yer almamıştır. *BCF*'in tahmin edilmesinde yalnızca *CRI*'ın kullanıldığı modelin korelasyon katsayısı $r = 0.850$ iken, 48h-EC_{50} modelinin korelasyon katsayısı $r = 0.926$ olmuş ve modelde tanımlayıcı olarak *CRI* ve E_{LUMO} yer almıştır. BCF için elde edilen ve yalnızca *CRI*'ın kullanıldığı modelde (model 6), veri sayısı 18 ile sınırlı olduğu için modelin güvenilirliği önce klasik testlerle kontrol edilmiş, daha sonra literatürdeki yine *CRI*'ya bağlı başka bir model (model 7) ile karşılaştırılmıştır. Jackknife testlerinin sonuçları geliştirilen beş modelin de istatistiksel olarak güvenilir olduğunu göstermiştir. Hesaplanan değerlerin deneysel değerlerden ortalama sapmaları *S*, K_{OW}, *H*, 48h-EC_{50}, ve *BCF* için sırasıyla 0.27, 0.17, 0.28, 0.19 ve 0.29 logaritmik birim olarak bulunmuştur. Geliştirilen modeller, deneysel verileri olmayan ve veri setinde bulunmayan hormon bozucu kimyasalların *S*, K_{OW}, *H*, 48h-EC_{50} ve *BCF* değerlerinin tahmin edilmesinde kullanılmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation	Units used
BCF	Bioconcentration factor	
C	Solute concentration in water	(mol L^{-1})
E	Energy	(kcal mol^{-1})
EC_{50}	50% Effective inhibition concentration	(mol L^{-1})
E_{HOMO}	Energy of the highest occupied molecular orbital	(eV)
E_{LUMO}	Energy of the lowest unoccupied molecular orbital	(eV)
F	Fisher statistic	
G	Graph as a square matrix with the entries	
H	Henry's Law Constant	(atm m^3 mol^{-1})
h	Number of hydrogen atoms bound to the same atom	
K_{AW}	Air-water partition coefficient	
K_{OC}	Soil sorption coefficient	
K_{OW}	*n*-octanol/water partition coefficient	
LC_{50}	50% Lethal concentration	(mol L^{-1})
n	Number of observation	
OH	Hydroxyl groups	
P	Vapor pressure	(atm)
r	Coefficient of determination for multiple linear regressions	
R	Gas constant	(atm m^3 mol^{-1} K^{-1})
S	Water solubility	(mol L^{-1})
S.E.	Standard error of the estimate	
S^N_{av}	Superdelocalizability	
T	Temperature	(K^{-1})

UNIFAC	Group Contribution Method	
w_{ij}	Weighted distance traversed in moving from vertex i, to vertex j in *G*	
Z	Atomic Number	
Z^v	Number of valence electrons in an atom	
δ^v	Delta (valence) value	
μ	Dipole moment	(debye)
AAD	Average Absolute Deviation	
BBP	Butylbenzyl phthalate	
CAS	Chemical Abstracts Service	
CRI	Characteristic Root Index	
DCHP	Dicyclohexyl phthalate	
DEHP	Di (2-ethylhexyl) phthalate	
DEP	Diethyl phthalate	
DMP	Dimethyl phthalate	
DnBP	Di-*n*-butyl phthalate	
DnHP	Di-*n*-hexyl phthalate	
DnOP	Di-*n*-octyl phthalate	
EASI	Electrotopological Atomic State Indices	
EDCs	Endocrine Disrupting Chemicals	
GC-RIs	Gas Chromatographic Retention Indices	
HpCDD	Heptachlorodibenzo-*p*-Dioxin Derivatives	
HpCDF	Heptachlorodibenzo-*p*-Furan Derivatives	
HPLC	High Performance Liquid Chromatographic Retention Index	
HxCDD	Hexachlorodibenzo-*p*-Dioxin Derivatives	

HxCDF	Hexachlorodibenzo-*p*-Furan Derivatives
LR	Linear Regression
MCI	Molecular Connectivity Indices
MLR	Multiple Linear Regression
NLR	Nonlinear Regression
NR	Not Reported
O8CDD	Octachlorodibenzo-*p*-Dioxin
O8CDF	Octachlorodibenzo-*p*-Furan
PAEs	Phthalate Esters
PAHs	Polycyclic Aromatic Hydrocarbons
PCBs	Polychlorinated Biphenyls
PCDDs	Polychlorinated Dibenzo-*p*-Dioxins
PCDFs	Polychlorinated Dibenzo-*p*-Furans
PCDD/PCDFs	Polychlorinated Dibenzo-*p*-Dioxins and Polychlorinated Dibenzo-*p*-Furans
PCP	Pentachlorophenol
PeCDD	Pentachlorodibenzo-*p*-Dioxin Derivatives
PeCDF	Pentachlorodibenzo-*p*-Furan Derivatives
PI	Padmakar-Ivan Index
PLS	Partial Least-Squares Analysis
PVC	Polyvinyl Chloride
QSAR	Quantitative Structure-Activity Relationship
QSPR	Quantitative Structure-Property Relationship
QSTR	Quantitative Structure-Toxicity Relationship
RMS	Root-Mean Square Error
SAR	Structure-Activity Relationship

SOFA	Solubility Parameters for Fate Analysis
STORET	US Environmental Protection Agency STOrage and RETrieval Database
TCDD	Tetrachlorodibenzo-*p*-Dioxin Derivatives
TCDF	Tetrachlorodibenzo-*p*-Furan Derivatives
TSA	Total Surface Area
VIF	Variance Inflation Factor

1. INTRODUCTION

Endocrine disrupting chemicals (EDCs) have been of great attention to the scientific community due to their impact on public health and the environment. Endocrine disrupting chemicals comprise a diverse group of compounds of anthropogenic origin, including organochlorine pesticides, alkyl phenols, phthalate esters (PAEs), polychlorinated biphenyls (PCBs), polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzo-*p*-furans (PCDD/PCDFs) and polybrominated diphenyl ethers. EDCs are generally present in the environment at low concentrations but they are ubiquitous and persistent and, although environmental concentrations are low, they appear to exert a range of adverse effects on animals of many species, including humans. Their effects include disruption of reproductive function, of the immune system and they can also be carcinogenic (Rhind, 2002).

Phthalates have been in use for nearly 40 years, mainly in the manufacture of polyvinyl chloride (PVC) and to a lesser extent other resins, as well as in plasticisers for building materials and home furnishings, in food packaging and insect repellents. These esters have meanwhile been found in all types of environmental and many biological samples (Fromme et al., 2002). In Germany, 400,000 t phthalates (Di (2-ethylhexyl) phthalate (DEHP), 250,000 t; di-*n*-butyl phthalate (DnBP), 21,000 t; butylbenzyl phthalate (BBP), 9000 t) and worldwide, millions of tons were produced in 1994/1995 (Leisewitz, 1997). Phthalates can enter the environment from losses during the manufacturing processes and by leaching from final products, this is because they are not chemically bonded to the polymeric matrix (Fromme et al., 2002).

Widespread occurence of PAEs in the environment raise concerns about their toxicological effects on living organisms. The existing studies have indicated a variety of biological effects on humans and other organisms. DEHP is included in class B2 (probable human carcinogen) while BBP is in class C (possible human carcinogen). Other commercial phthalates (DnBP, diethyl phthalate (DEP) and dimethyl phthalate (DMP)) were included in class D (have not been classified as human carcinogens) (Mondragon et al., 2003).

Polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzo-*p*-furans which originate mainly from byproducts of the chlorine chemical industry, impurities in e.g., herbicides, municipal and industrial waste incineration, metallurgical processes, and bleaching processes in the pulp and paper industry, are typical persistent pollutants. PCDD/PCDFs are highly lipophilic and chemically stable, resulting in accumulation in organisms or organic phase of soils or sediments. Some of their isomers are strongly toxic, so they are considered as important priority contaminants that are distributed worldwide and with a great potential risk to human health (Chen et al., 2001).

Benzene derivatives have been used for many years in the chemical industry. Many of these chemicals were released into the environment and accumulated in nearly all environmental compartments, especially in aquatic ecosystems, so it is beneficial to make an extensive study of their potential hazards to aquatic organisms (Huang et al., 2003). Benzene is a by-product of various combustion processes, such as forest fires, the burning of wood, garbage, organic wastes, and cigarettes (Hattemer-Frey et al., 1990); it is also released into the air from crude oil seeps and volatilizes from plants (Brief et al., 1980).

The toxicity of these compounds can arise from a multitude of mechanisms of toxic action including a range of narcoses as well as reactive mechanisms in which the compounds are able to form covalent bonds with biological macromolecules (Netzeva et al., 2004). These compounds tend to magnify in the food chain and cause adverse effects including neuro, reproductional and immuno toxicities on human and wildlife situated on the top of the food chain (Dai et al., 1999).

The environmental fate of all these organic pollutants is governed by their physico-chemical and biological properties heavily depending on their chemical structures as stated by Ferreira (2001) and Dai et al. (1999). The physico-chemical and biological properties, such as *n*-octanol/water partition coefficient (K_{OW}), water solubility (S), Henry's Law Constant (H), bioconcentration factor (BCF), and toxicity; 50% effective inhibition concentration (EC_{50}) are of critical importance for evaluating phthalates, PCDD/PCDFs and benzene derivatives fate and potential exposure to environment and, consequently, for the whole process of environmental risk assessment (Dai et al., 1999).

The equilibrium distribution of an organic chemical between water and octanol (K_{OW}) is an important constant for predicting the tendency of a chemical to partition to water, animal lipids, sediment, and soil organic matter. The log K_{OW} value is defined as the ratio of the equilibrium concentration of a dissolved substance in a two-phase system consisting of two largely immiscible solvents, *n*-octanol and water (Staples et al., 1997). The *n*-octanol/water partition coefficient basically gives direct information on hydrophobicity of a compound and is frequently used to predict non-reactive toxicity, water solubility, bioconcentration factor and soil (sediment)/water partition coefficient (Yang et al., 2003). Compounds for which $K_{OW} > 1$ are lipophilic or hydrophobic and compounds for which $K_{OW} < 1$ are hydrophilic. By definition, K_{OW} is inversely proportional to aqueous solubility (Ferreira, 2001).

Water solubility is an extremely important property that influences the biodegradation and bioaccumulation potential of a chemical, as well as aquatic toxicity. Water solubility also is a significant factor controlling the environmental distribution of chemicals. Losses from wastewater treatment facilities, landfills and sludge-amended soil are a partial function of aqueous solubility (Staples et al., 1997). *S* affects processes such as evaporation, absorption, bioaccumulation and biodegradation/biodegradability. It provides an estimate of activity coefficient in the aqueous phase for a particular substance (Yang et al., 2003). Some studies showed that log *S* correlates with log K_{OW} significantly (Krop et al., 1997).

The Henry's Law Constant is essentially an air-water partition coefficient (K_{AW}) differing only by a scale factor when the temperature is kept constant since $K_{AW} = H/RT$, where R is the gas constant. K_{AW} is important for controlling the cycling of compounds across air-water interface and it can be determined by measurement of solute concentrations in both phases (Ferreira, 2001). Henry's Law Constant can also be expressed as the ratio of *P*/C where *P* is partial vapor pressure in gas or air and C is the solute concentration in water (Ferreira, 2001).

Biological properties of the selected endocrine disrupting chemicals which were used and predicted in this study are EC_{50} and bioconcentration factor. EC_{50} is the concentration of a toxicant that causes 50% of the maximum response of the test organisms (Ren et al.,

2003). In other words, EC_{50} is used to define the toxicity of the chemical, in terms of the concentration of the chemical that reduces the activity of the test bacteria by 50% (Chambers et al., 1996a). Quantitative structure-activity relationships (QSARs) and quantitative structure-toxicity relationships (QSTRs) have provided a valuable approach in research on the toxicity of organic chemicals in such studies (Huang et al., 2003).

Bioconcentration factor is one chemical property of interest in modeling the fate and persistence of chemicals such as PCDD/PCDFs, phthalates and benzene derivatives in the environment. Bioconcentration is the process of accumulation of chemicals by organisms through nondietary routes (Khadikar et al., 2003). In other words, bioconcentration is the process that causes an increased chemical concentration in an aquatic organism, compared to that in water, due to the uptake of chemicals by absorption from water only, which can occur via the respiratory surface and/or the skin. In addition, pollutants enter into biota along the food chain through dietary uptake, which is referred to as biomagnification. The sum of bioconcentration and biomagnification is referred to as bioaccumulation (Voutsas et al., 2002). In aquatic ecosystems, the bioconcentration factor of an organic chemical is defined as the ratio of its concentration in a target organism to that in water at a steady state. *BCF* gives a measure of partitioning of compounds between organisms and their surrounding environment (Khadikar et al., 2003).

Because experimental determination of the physico-chemical and biological properties of phthalate esters, PCDD/PCDFs and benzene derivatives are expensive, time consuming and experimental data are not available for all chemicals in use, many researchers tend to use estimation methods to supply the missing data. Quantitative structure-property relationship (QSPR) is a very practical approach for estimating the physico-chemical and biological properties since it only requires information of the chemical structure of a compound. As a result there has been a remarkable growth in interest in the area of quantitative structure-property relationships, which uses multivariate methods to model relevant properties as a function of molecular structure parameters (called descriptors) (Ferreira, 2001).

In a QSPR study, a mathematical model is developed which relates the structure of a set of compounds to a physical property such as aqueous solubility. The underlying

assumption in a QSPR study is that there is some sort of relationship between the physical property of interest and structural parameters (descriptors). These descriptors are numerical representations of structural features of molecules that attempt to encode important information that causes structurally different compounds to have different physical property values. The main types of molecular descriptors are constitutional, topological, electrostatic, geometrical and quantum chemical descriptors. The importance of highly specific QSPRs becomes obvious when the fate and transport of structurally similar compounds are considered (Delgado, 2002).

Various molecular structural descriptors, like constitutional descriptors, electrostatic descriptors, topological descriptors, geometrical descriptors, and quantum chemical descriptors have been used to develop QSPR models (Huuskonen et al., 1998; Beck et al., 2000; McClelland and Jurs, 2000; Chalk et al., 2001). The Characteristic Root Index (*CRI*) model which comprises all possible orders of path-type valence connectivity indices has been demonstrated correlating many physico-chemical and biological properties of organic compounds including solubility (Saçan and İnel, 1993), *n*-octanol/water partition coefficient (Saçan and İnel, 1995), soil sorption coefficient (K_{OC}) (Saçan and Balcıoğlu, 1996), vapor pressure (Saçan and Balcıoğlu, 1998), and bioconcentration factor (Saçan et al., 2003).

Saçan and İnel (1993) discussed the relationship between the *CRI* and aqueous solubilities including solid solubility and subcooled solubility of PCBs. They obtained high correlation coefficients (r = 0.980; r = 0.990) for solid solubilities and subcooled solubilities, respectively. Saçan and İnel (1995) developed the QSPR model by the use of the *CRI* for the prediction of *n*-octanol/water partition coefficient of PCBs, and the correlation coefficient of this model at 0.997 which is very significant demonstrating that the *CRI* itself has high predictive capacity to estimate K_{OW} of PCBs. Saçan et al. (2003) modeled the *CRI* together with four semi-empirical molecular descriptors and predicted the fish bioconcentration factor of 122 non-ionic organic compounds. They found that the topology based *CRI* was the most important parameter.

1.1. Purpose of the Study

Knowing the reasonable literature correlations between the *CRI* models and the physico-chemical and biological properties mentioned above, it seems logical to examine the relationship between topology based *CRI* model and physico-chemical and biological properties of selected endocrine disrupting chemicals including phthalate esters, polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzo-*p*-furans and substituted benzenes. As quantum chemical descriptors can be easily obtained by computation, they can clearly describe defined molecular properties and as they are not restricted to closely related compounds, the development of QSPR models in which quantum chemical descriptors are used, is of great importance (Chen et al., 2001). Therefore, this work investigates whether the *CRI* and the semi-empirical quantum chemical descriptors; the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), and dipole moment (μ) are useful parameters to predict the environmentally important physico-chemical and biological properties of chemicals having no structural similarity but a common property of disrupting endocrine systems. The physico-chemical properties which were used and predicted in this study include solubility, *n*-octanol/water partition coefficient, and Henry's Law Constant. The biological properties which were used and estimated in this study are 50% effective inhibition concentration and bioconcentration factor. Since there was not enough experimental literature data for the biological properties of the first group of endocrine disrupting chemicals including phthalate esters, polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzo-*p*-furans, we used another group of chemicals from the list of EDCs (substituted benzenes) for the prediction of the biological properties.

The objectives of this study are to develop QSPR models for the prediction of physico-chemical (K_{OW}, H, S) and biological (EC_{50}, BCF) properties, which are important in the partition of chemicals among air, water and living organisms by the application of the *CRI* model; and to increase the predictive power of the *CRI*-based models by including semi-empirical molecular descriptors (E_{HOMO} , E_{LUMO} , and μ); and finally to predict the unknown physico-chemical and biological parameters of organic compounds structurally similar to those compounds used in the data set.

2. THEORETICAL BACKGROUND

2.1. Studies on QSPR

Some previous studies showed that it was indeed feasible and successful to develop QSPR models for vapor pressure (Basak and Mills, 2001; Yaffe and Cohen, 2001), aqueous solubility (Ferreira, 2001; Delgado, 2002; Jorgensen and Duffy, 2002) and *n*-octanol/water partition coefficient (Ferreira, 2001).

Basak and Mills (2001) estimated vapor pressure of chemicals by using mathematical structural descriptors. Ferreira (2001) related solubility and *n*-octanol/water partition coefficient with a series of electronic, geometric and topological descriptors, and built several QSPR models. Yaffe and Cohen (2001) also predicted vapor pressure of hydrocarbons by their QSPR model. Delgado (2002) developed a quantitative structure-property relationship model to correlate the logarithm of solubility in water of 50 chlorinated hydrocarbons (including chlorinated benzenes, dibenzo-*p*-dioxins and biphenyls) from the molecular structure. Jorgensen and Duffy (2002) predicted drug solubility from the structure.

Basak et al. (1996) made a comparative study of topological and geometrical parameters in estimating normal boiling point and octanol/water partition coefficient. In another study, Basak et al. (1997) used topostructural, topochemical, and geometric parameters in their QSPR model to predict vapor pressure of chemicals.

Karickhoff (1981) used water solubility to estimate the organic carbon soil sorption coefficient for polycyclic aromatic hydrocarbons (PAHs). The same authors also correlated K_{OC} and K_{OW}. Govers et al. (1984) used topological index for K_{OC} prediction. Bahnick and Doucette (1988) used molecular connectivity indices (MCI) to estimate soil sorption coefficients for organic chemicals. Hawker (1989) estimated vapor pressures of polychlorinated biphenyls from molecular descriptors such as molecular volume or total surface area (TSA). Sabljic et al. (1995) also used topological index for K_{OC} prediction. Hong et al. (1996) predicted the soil sorption coefficient by reversed phase high

performance liquid chromatographic (HPLC) retention time. Another review on the K_{OC} prediction from topological index was given by Gawlik et al. (1997).

Some studies correlated K_{OW} to aqueous solubility (Mackay et al., 1980; Bruggeman et al., 1982; Miller et al., 1985). Mallhot and Peters (1988) studied the empirical relationships between *n*-octanol/water partition coefficient and nine physico-chemical properties. Warne et al. (1990) predicted aqueous solubility and *n*-octanol/water partition coefficient for lipophilic compounds using molecular descriptors and physico-chemical properties. Lohninger (1994) estimated soil sorption coefficients of pesticides from their chemical structure by combining both topological indices and structural fragments in which 26 structural fragments served as input variables to the model, and, Reddy and Locke (1994a) used the molecular modelling technique to establish relationships between the molecular properties of 45 substituted phenyl ureas and their octanol-water partition coefficient and soil sorption normalized to organic carbon.

Some works relate the Henry's Law Constant or air-water partition coefficient to structural molecular descriptors but usually focusing PCBs and pesticides. De Maagd et al. (1998) measured H besides aqueous solubilities and K_{OW}. Their results were related with molar volume. Bamford et al. (1999) also measured Henry's Law Constant at different temperatures, which relates log H with molar volume. Chen et al. (2002) developed a QSPR model for the prediction of octanol-air partition coefficients of polychlorinated biphenyls. Dearden and Schüürmann (2003) reviewed the QSPRs for predicting Henry's Law Constant from the molecular structure.

Quantitative structure-activity/-toxicity relationship models have emerged as useful tools to handle the data gap in toxicology and pharmacology (Hall et al., 1984; Basak and Grunwald, 1995; Gombar et al., 1995; Gute and Basak, 1997; Basak et al., 1999).

The majority of QSARs correlate toxicity using the octanol/water partition coefficient and/or various physico-chemical, topological, and quantum chemical parameters (Konemann, 1981; Nendza and Russom, 1991; Newsome et al., 1991; Schultz et al., 1991; Karabunarliev et al., 1996; Eldred et al., 1999).

Sixt et al. (1995) developed quantitative structure-toxicity relationships for 80 chlorinated compounds using quantum chemical descriptors. Kaiser and Niculescu (1999) developed a group contribution model to correlate the toxicity of 865 organic chemicals. Their model included groups to account for many different types of functional groups. Martin and Young (2001) have developed a group contribution method (UNIFAC) to correlate the acute toxicity to the fathead minnow for 397 organic chemicals. This linear model, which included four specific interaction terms, provided a rapid means of predicting the toxicity of a compound.

Numerous correlations have been developed relating *BCF* to K_{OW} (Devillers et al., 1996) as well as other descriptors, for non-ionic pollutants (Neely et al., 1974; Veith et al., 1979a; Kanazawa, 1981; Sabljic and Protic, 1982; Koch, 1983; Veith and Kosian, 1983; Govers et al., 1984; Isnard and Lambert, 1988; Schüürmann and Klein, 1988; Lu et al., 1999; Lu et al., 2000). Though good results were achieved in some studies, not all *BCF* always had a good correlation with K_{OW} with any kind of compound, and the application of these methods is limited by the availability of parameter data. Frequently, data on the parameters, when not available in the literature, have to be either experimentally determined or estimated by other methods. However; in this study, the estimations runs directly from structure to property by use of the *CRI*.

The correlation of *BCF* with K_{OW} is approached by Mackay (1982). Govers et al. (1984) modeled *BCF* and partition coefficients, using molecular weight and molecular connectivity indices. Sabljic (1987) also pioneered the study of the relationship between *BCF* and molecular connectivity indices. Lu et al. (1999) and Lu et al. (2000) used only connectivity indices, both connectivity indices and polarity factors to estimate *BCF* of non-ionic organic compounds. They found that molecular connectivity indices were not good descriptors for polar compounds. When polarity correction factors were introduced into the linear molecular connectivity model, the *BCF* estimation quality for polar compounds was much increased. Generally speaking, the bioconcentration process is controlled by polar and nonpolar interactions among chemicals, water and fish. Polarity of a chemical complicates this process. Khadikar et al. (2003) has made an attempt to estimate the accuracy, predictive power, and domain of application of the PI (Padmakar-Ivan) index for modeling *BCF* of polyhalogenated biphenyls.

2.2. Phthalate Esters (PAEs)

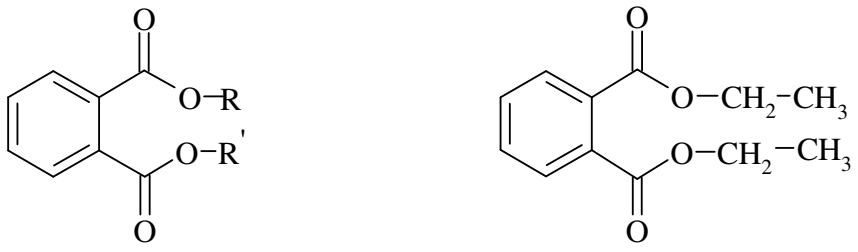
2.2.1. The Usage and Sources of Phthalates

Phthalic acid esters or phthalate esters are chemical compounds widely used as plasticizers giving plastics flexibility and durability. Plasticizers hold 65% of the 7.5 million ton world market for plastic additives. The majority of this, about 90% is used for PVC-based plastics. Industrial applications of PVC-based plastics include coatings, plumbing, and construction materials and in the manufacture of common plastic products such as vinyl upholstery, tablecloths, and shower curtains. PAEs are also present in plastic products for human use, e.g. teething rings, pacifiers, soft squeeze toys, plastic bottles, and enclosures for food containers and in medical products, e.g. flexible devices for administering parenteral solutions, and vinyl gloves. Formerly, they were used in an even wider range of applications (Mondragon et al., 2003).

Phthalate esters are ubiquitous compounds in the environment. Earlier studies reported the presence of phthalate esters in the air and precipitation at remote marine locations and surface waters and sediments. Phthalates are further found in waste such as source-sorted household solid waste. A recent survey reported detectable levels of phthalate esters in samples of foodstuffs, human mother's milk, dust, and textiles with DEHP and DnBP being the most abundant (Mondragon et al., 2003).

2.2.2. Chemical Structure of Phthalates

Phthalates are esters of 1,2-benzoldicarbonic acid (*ortho*-phthalic acid). The chemical structure of phthalates can be characterised by a planar aromatic ring with two slightly mobile side chains. For phthalates, in general, side chains are mainly alkyl groups with a secondary role played by allyl, benzyl, phenyl, cycloalkyl and alkoxy groups. With alkyl phthalates, a further distinction can be made between branched and unbranched side chains (http://www.ukmarinesac.org.uk/activities/water-quality/wq8_43.htm#a4, 2004). (a) General structure of phthalate esters and (b) structure of unbranched diethyl phthalate are shown in Figure 2.1.



(a) (b)

Figure 2.1. (a) General structure of phthalate esters and (b) diethyl phthalate

Phthalates are characterized by hydrocarbon functional groups that vary from 1 to 13 carbons. R and R´ in Fig. 2.1 (a) represent carbon chains that can range from 1 to 13 carbons.

2.2.3. Physico-Chemical and Biological Properties of Phthalates

Phthalates look like vegetable oil. They have little to no smell. Common commercial phthalate esters are liquids at ambient temperatures. Nearly all have melting points (or pour points) below -25 ºC although DMP and dicyclohexyl phthalate (DCHP) have melting points of 5.5 ^{o}C and 63 ^{o}C respectively. Phthalate esters have boiling points varying from about 230 ºC to 486 ºC (Staples et al., 1997).

Phthalate esters have generally low volatility (vapour pressures in the range 0.02 - 1.9 m Pa) which decrease with increasing length of the alcohol side-chain of the ester. Henry's Law Constants range between 10^{-8} and 10^{-5} atm m^3 mol^{-1}, indicating that phthalate esters will tend to volatilise from water, either not at all or very slowly. Since some phthalate esters have low water solubility and high octanol partition coefficient, they can become concentrated in suspended matter and sediment (http://www.ukmarinesac.org.uk/ activities/water-quality/wq8_43.htm#a4, 2004).

The most important aquatic degradation process for phthalate esters is biodegradation. Short side-chain phthalate esters, such as DMP, DEP and DnBP, are all likely to degrade quickly in aerobic surface waters. The longer chain PAEs, such as di-*n*-octyl phthalate

(DnOP), are likely to be more persistent, particularly as they will partition more strongly to sediments and so may be less available to microbial degradation. It is, therefore, probable that residues of DnOP in anaerobic sediments will tend to persist for long periods (http://www.ukmarinesac.org.uk/activities/water-quality/wq8_43.htm#a4, 2004).

2.2.4. Environmental Importance

2.2.4.1. Pathways of Phthalates to Humans. Once PAEs enter in the environment they partition between air, water, soil and sediments. However, given the low solubility and highly hydrophobic nature of these compounds, they will preferentially be sorbed to the organic fraction of soil or sediments, as well as to the organic matter suspended in water. PAEs are present in sewage through several non-point sources such as domestic and commercial discharges, street runoff, and aerial deposition and point sources such as industrial discharges (Mondragon et al., 2003).

The general population is exposed to phthalates primarily through the oral and dermal routes. Based on data for other phthalates, the most likely source of human exposure to di-*n*-hexyl phthalate (DnHP) is dietary intake. DnHP may be found in food as a result of environmental uptake during cultivation or as a result of migration from processing equipment or packaging materials (Kavlock et al., 2002).

Occupational exposure. Workers may be exposed to DnHP primarily through inhalation and dermal contact. Phthalates are manufactured within closed systems, but exposure to workers can occur during filtering or loading/unloading of tankcars. Higher exposures to phthalates can occur during the production of flexible PVC because the processes are open and run at higher temperatures (Kavlock et al., 2002).

Mouthing of toys is a potential source of oral phthalate exposure in children (Kavlock et al., 2002). Phthalates in children's toys can migrate from soft PVC products during child mouthing activities. Fish living in the aquatic environment polluted by phthalate containing wastewater is another pathway of phthalates to humans.

DEP is an important solvent and vehicle for fragrance and cosmetic ingredients. Thus, there is potential exposure to humans by the intentional application of such products to the skin (Api, 2001).

During sewage treatment, primary and ultimate degradation of PAEs occur during the aerobic phase. PAEs sorbed to the suspended organic matter will escape aerobic degradation and will be transferred to primary and secondary sludge streams. These streams are normally treated by anaerobic digestion of the sewage sludge. Members of the PAEs with lower molecular weight (i.e. DMP, DEP, DnBP and BBP) are easier to degrade than members with higher molecular weight (i.e. DnOP and DEHP) (Mondragon et al., 2003).

Sorbed PAEs that are not degraded during the anaerobic digestion of sewage sludge will accumulate in the sewage sludge solids (biosolids) at concentrations several orders of magnitude higher than in the influent wastewater (Mondragon et al., 2003). The United States of America National Sewage Sludge Survey reported detectable levels of DEHP in the range from 55.1 mg kg^{-1} to 163.3 mg kg^{-1} dry weight (USEPA, 1990).

Anaerobic digestion followed by land application is a common treatment and disposal practice for biosolids. Continuous practice of land application of biosolids may eventually lead to accumulation of the most persistent toxic organics in the soil, with a consequent threat to the ecosystem and to human beings in particular (Mondragon et al., 2003).

2.2.4.2. Adverse Health Effects of Phthalates. There is some validated scientific evidence to indicate that phthalate esters pose health problems for humans. Scientific research shows that rodents are uniquely sensitive to phthalate esters. As a result, some phthalates have shown the potential to induce liver or kidney tumors when fed to laboratory animals at high doses for extended periods of time. Adverse liver effects, however, are not seen in other mammals more closely related to humans (i.e., hamsters and monkeys) (Ganning et al., 1984). Based on the best evidence to date, we strongly believe phthalates may pose some hazard to children or adults when not properly used in vinyl products.

Phthalates have in recent years caused increasing concern due to reported weak carcinogenic and estrogenic effects of some of these compounds, thus possibly affecting the male reproductive health. In this connection the environmental behavior of phthalates calls for special attention (Thomsen et al., 1999). Based on the reproductive effects observed in high-dose studies of phthalate esters, this class of compounds has been included on lists of endocrine disrupters. The ability of a compound, at high dose levels, to cause reproductive effects does indicate that a compound will act as an endocrine disrupter. This assertion was primarily based upon work conducted in vitro, using receptor binding assays or reporter cell systems, but estrogenic activity was not a consistent finding (Moore, 2000).

The evidence is stronger that phthalates are a danger to young infants. In a 1988 German study, preterm infants receiving respiratory assistance utilizing PVC tubing accumulated significant amounts of DEHP in their lungs, developing unusual lung disorders resembling hyaline membrane disease. The authors make the point that the livers of very young infants do not metabolize DEHP as efficiently as those of adults, placing infants at higher risk. They further suggest that the damage to testicular and related structures seen in rats may be more relevant to preterm infants than to adult humans, again because of slower/less complete metabolisms (Jobling et al., 1995).

2.2.5. Recent Studies on Phthalates

Wolfe et al. (1980a) used linear free energy relationships to assess the fate and transport of phthalate esters in the aquatic environment. Rasmussen (1998) predicted aqueous solubilities and octanol-water coefficients of phthalates by the UNIFAC Group Contribution Method. Thomsen et al. (1999), described solubilities, octanol/water partitioning, and soil sorption coefficient of a series of phthalic acid esters through various structure-activity relationship concepts including molecular connectivity indices, electrotopological atomic state indices (EASI) and group-contribution. Cousins and Mackay (2000) described a quantitative structure-property relationship method for the correlation of physical-chemical properties and partition coefficients, namely the “three solubility” approach and applied to a group of 22 phthalate esters. They compiled the solubilities or “apparent-solubilities” of these substances in the liquid state and correlated

against Le Bas molar volume in the three primary media of air, water and octanol. Lu et al. (2000) developed a bioconcentration factor estimation model for non-ionic organic compounds including phthalates by the use of molecular connectivity indices and polarity correction factors.

2.3. Polychlorinated Dibenzo-*p*-Dioxins and Polychlorinated Dibenzo-*p*-Furans (PCDD/PCDFs)

2.3.1. The Usage and Sources of PCDD/PCDFs

Polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzo-*p*-furans are two groups of organic compounds which are ubiquitous in the environment at ultra-trace levels, but which have attracted considerable scientific and political concern because of their environmental persistence, tendency to bioaccumulate through the food chain and toxicity. In recent years, they have attracted particular interest because of their presence in sewage sludges and have been included on listings of the priority organic contaminants (http://www.fwr.org, 2005).

Since the 1930s, there has been a steady increase in environmental levels of dioxins coinciding with the large-scale production and use of chlorinated chemicals. Man-made sources of PCDD/PCDFs can be divided into a number of main categories (McKay, 2002). The USEPA (1994, 1998) have derived these emission estimates and identified the following major sources of dioxin releases in the USA.

2.3.1.1. Combustion Sources. PCDD/PCDFs are formed in most combustion systems. These can include waste incineration (such as municipal solid waste, sewage sludge, medical waste, and hazardous wastes), burning of various fuels, such as coal, wood, and petroleum products other high temperature sources such as cement kilns, and poorly controlled combustion sources such as building fires, and burning any chlorine compounds (McKay, 2002).

2.3.1.2. Metal Smelting and Refining Sources and Processing Sources. PCDD/PCDFs can be formed during various types of primary and secondary metals operations including iron ore sintering, steel production, and scrap metal recovery (McKay, 2002).

2.3.1.3. Chemical Manufacturing. PCDD/PCDFs can be formed as by-products from the manufacture of chlorine bleached wood pulp and chlorinated phenols, e.g. pentachlorophenol (PCP), PCBs, phenoxy herbicides (e.g. 2,4,5-trichloro-phenoxyacetic acid), and chlorinated aliphatic compounds (e.g. ethylene dichloride) (McKay, 2002).

2.3.1.4. Biological and Photochemical Processes. Recent studies have suggested that PCDD/PCDFs can be formed under certain environmental conditions, e.g. composting from the action of microorganisms on chlorinated phenolic compounds. Similarly, PCDD/PCDFs have been reported to be formed during photolysis of highly chlorinated phenols (McKay, 2002).

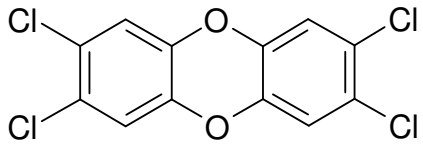
2.3.2. Chemical Structure of PCDD/PCDFs

Dioxins, as they are commonly called, are PCDD/PCDFs, which are compounds with similar chemical properties. Each compound comprises of two benzene rings interconnected by oxygen atoms. In the case of PCDDs, the benzene rings are joined by two oxygen bridges, and in the case of PCDFs, the benzene rings are connected by a carbon bond and an oxygen bridge. Figure 2.2 shows the generic structures of (a) PCDDs and (b) PCDFs, respectively (McKay, 2002).

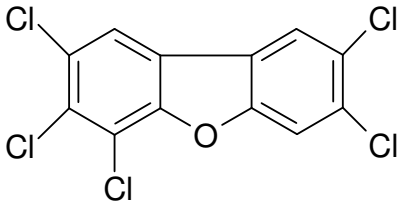
(a) (b)

Figure 2.2. Structures of (a) PCDDs and (b) PCDFs

PCDDs consist of 12 carbon atoms, forming two aromatic phenyl rings attached to one another through two oxygen bridges, and 8 atoms that can be either hydrogens or chlorines. Theoretically 75 various combinations of chlorine and hydrogen are possible, and the resulting dibenzo-*p*-dioxin derivatives are called *congeners.* Chlorine increases the stability of these compounds, and chlorines in positions 2,3,7, and 8 (lateral chlorines) are especially important, because they are essential for toxicity and also prevention of enzymatic destruction of PCDDs. Therefore the 7 *congeners* with 2,3,7,8-structure are toxicologically the most relevant. All additional chlorines to 2,3,7,8-structure decrease toxicity, but the spectrum of adverse effects remains similar. Tetra-, penta-, hexa-, and heptachloro-derivatives are often called TCDD, PeCDD, HxCDD, and HpCDD, respectively (http://www.ktl.fi/dioxin/ptoq.html, 2004). Figure 2.3 shows the structures of (a) 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and (b) 2,3,4,7,8-pentachlorodibenzo-*p*-furan.



(a)



(b)

Figure 2.3. Structures of (a) 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and (b) 2,3,4,7,8-pentachlorodibenzo-*p*-furan

PCDFs consist of 12 carbon atoms, forming two aromatic phenyl rings attached to one another through one carbon-carbon bond and one oxygen bridge, and 8 atoms which can be either hydrogens or chlorines. Theoretically 135 various combinations of chlorine and hydrogen are possible, and the resulting dibenzofuran derivatives are called *congeners*. Chlorine increases the stability of these compounds, and chlorines in positions 2,3,7, and 8 (lateral chlorines) are especially important, because they increase toxicity and also prevent enzymatic destruction of PCDFs. Therefore the 10 *congeners* with 2,3,7,8-Cl-structure are toxicologically the most relevant. Most additional chlorines to 2,3,7,8-structure decrease toxicity, but the spectrum of adverse effects remains similar. Tetra-, penta-, hexa-, and

heptachloro-derivatives are often called TCDF, PeCDF, HxCDF, HpCDF, respectively (http://www.ktl.fi/dioxin/ptoq.html, 2004).

2.3.3. Physico-Chemical and Biological Properties of PCDD/PCDFs

All PCDD/PCDFs are organic solids with high melting points and low vapour pressures. They are characterised by extremely low water solubilities, and have a tendency of being strongly adsorbed on surfaces of particulate matter. The water solubility of dioxin and furans decreases and the solubility in organic solvents and fats increases with increasing chlorine content (McKay, 2002). Their octanol/water distribution (indicating relative lipid to water solubility,) is in the order of one million to a hundred million (log K_{OW} 4.0 to 8.8) explaining the high tendency to move toward lipids from water (http://www.ktl.fi/dioxin/ptoq.html, 2004).

Many PCDD/PCDFs are extremely persistent in the environment. Increase in chlorination increases both stability and lipophilicity. Neither soil microbes nor animals are able to break down effectively those PCDDs with “lateral” chlorines, i.e. chlorines in positions 2,3,7, and 8. This causes especially slow elimination, and due to biomagnification those particular compounds are present in the organisms of higher trophic levels (such as birds and mammals). Therefore they concentrate along the food chain, and the species at the “top” of the food chain such as seals or eagles are in special danger (http://www.ktl.fi/dioxin/ptoq.html, 2004).

2.3.4. Environmental Importance

2.3.4.1. Pathways of PCDD/PCDFs to Humans. Atmospheric deposition of PCDD/PCDFs onto food crops and the food or soil consumed by domestic animals and fish are the primary pathway of PCDD/PCDFs to humans. In relation to biosolids exposure, the most significant health pathway is by direct ingestion. This can occur when animals graze surface fertilized fields or when particles of biosolids adhere to plants and are ingested. Exposure analysis indicates that the greatest animal intake occurred when liquid biosolids were applied to established crops, and animals had immediate access. Reduced animal ingestion occurs when access is delayed, or biosolids are incorporated into the soil. There

is no evidence of dioxin uptake by plants from typical biosolids applications, although some plant families absorb PCDD/PCDFs when grown on highly contaminated sites (http://www.ktl.fi/dioxin/ptoq.html, 2004).

USEPA estimates that most dioxin exposure occurs through diet, with over 95% coming through dietary intake of animal fats. Small amounts of exposure occur from breathing air containing trace amounts of dioxins on particles and in vapor form (USEPA, 1998).

Human exposure to PCDD/PCDFs is due almost solely to food consumption, mostly found in the fatty tissues of animals. PCDD/PCDFs exist in food (fatty foods: dairy products (butter, cheese, fatty milk), meat, egg, and fish) that we eat every day. This is another source of dioxins into wastewater treatment plants. Some subgroups within the society (e.g., nursing babies and people consuming plenty of fish) may be highly exposed to these compounds and are thus at greater risk (http://www.ktl.fi/dioxin/ptoq.html, 2004).

2.3.4.2. Adverse Health Effects of PCDD/PCDFs. Recent studies (Fossi et al., 1999; Younes, 1999) revealed that most PCDD/PCDFs are endocrine disrupting chemicals, i.e., they are exogenous substances that cause adverse health effects in an intact organism, or its progeny, consequent to changes in endocrine function. It is widely recognized that knowledge of the environmental fate of pollutants is a basic need for exposure assessment (Chen et al., 2001).

In humans, the highly chlorinated furans, dioxins and many organic compounds are stored in fatty tissues and are neither readily metabolised nor excreted. This persistence is a consequence of their structure: few of them contain hydrogen atoms on adjacent pairs of carbons at which hydroxyl groups (OH) can readily be added in the biochemical reactions necessary for their elimination. In contrast, those compounds with few chlorines always contain one or more such adjacent pairs of hydrogens and tend to be excreted rather than stored for a long time.

PCDD/PCDFs in the body are almost exclusively in fat because of their lipid solubility and poor water solubility. In some tissues dioxins may also be bound to specific

proteins. PCDD/PCDFs-elimination, process of discharging PCDD/PCDFs out of the body, is very slow in all mammals, because these compounds are lipophilic, and cannot be excreted in urine, and are also poorly degradable by the enzymatic machinery of the body (http://www.ktl.fi/dioxin/ptoq.html, 2004).

A number of other health effects have been linked to high exposure to dioxins, including mood alterations, reduced cognitive performance, diabetes, changes in white blood cells, dental defects, endometriosis, decreased male/female ratio of births, decreased testosterone and (in neonates) elevated thyroxin levels. As yet such effects have not been proven as caused by PCDD/PCDFs. The effect that has caused the greatest public concern is cancer. Another concern in the society are the possible developmental effects. There is recent data that dioxin exposure from breast milk is associated with abnormal development and mineralization of teeth (http://www.ktl.fi/dioxin/ptoq.html, 2004).

2.3.5. Recent Studies on PCDD/PCDFs

The K_{OW} of PCDD/PCDFs was estimated from water solubility (Chiou et al., 1977), high performance liquid chromatographic retention index (Webster et al., 1985), quantitative structure-activity relationships (Fiedler and Schramm, 1990) and structural group contributions (Lyman et al., 1990; Samiullah, 1990; Schwarzenbach et al., 1995). Voogt et al. (1990) applied quantitative structure-activity relationships for the *BCF* in fish of seven PCDDs. Govers and Krop (1998) have developed a solubility parameters for fate analysis (SOFA) method to predict the fate of a series of PCDD/PCDFs in the environment. Dai et al. (1999) described QSAR modeling techniques based on quantum chemical descriptors to predict K_{OW} and K_{OC} values of 70 polychlorinated organic compounds including PCDD/PCDFs. Lu et al. (2000) developed a bioconcentration factor estimation model for a wide range of non-ionic organic compounds including PCDD/PCDFs on the basis of molecular connectivity indices and polarity correction factors. Chen et al. (2001) obtained a significant quantitative structure-property relationship model for log K_{OW} of polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzo-*p*-furans by the use of partial least-squares (PLS) analysis, based on some fundamental quantum chemical descriptors computed by PM3 Hamiltonian. Wang and Wong (2002) estimated vapor pressure, water solubility, Henry's Law Constant and *n*-

octanol/ water partition coefficient of a series of PCDDs by using their GC-RIs (gas chromatographic retention indices). In another study of the same authors they estimated physico-chemical properties of PCDFs by using their GC-RIs.

2.4. Benzene Derivatives

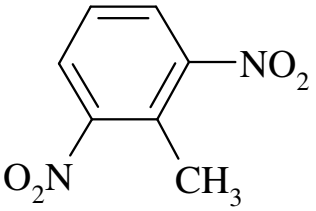
2.4.1. The Usage, Sources and Chemical Structures of Benzene Derivatives

Benzene is one of the world's major commodity chemicals. Its primary use (95% of production) is as an intermediate in the production of other chemicals, predominantly ethyl benzene (for styrofoam and other plastics), cumene (for various resins), and cyclohexane (for nylon and other synthetic fibers). Benzene is an important raw material for the manufacture of synthetic rubbers, gums, lubricants, dyes, pharmaceutical, and agricultural chemicals; it is also found in consumer products such as glues, paints, and marking pens (Mehlman, 1991).

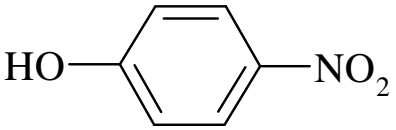
Benzene is a very stable (non-reactive) organic compound. Each carbon is joined to each adjacent carbon by a single sigma bond and a “partial” carbon-carbon double bond (a pi bond and a sigma bond). Benzene is a volatile, colorless, highly flammable liquid, most (98%) of which is commercially derived from petrochemical and petroleum refining industries (Hattemer-Frey et al., 1990).

Benzene derivatives are ubiquitous in the environment, having been measured in the air, water, and human biological samples (Wallace et al., 1987; Hattemer-Frey et al., 1990). The major environmental sources include automobile exhaust, automobile refueling, hazardous waste sites, underground storage tanks that leak, waste water from industries that use benzene, chemical spills, chemical manufacturing sites, and petrochemical and petroleum industries (Edgerton and Shah, 1992; Fishbein, 1992).

Chemical structures of two benzene derivatives: (a) 2,6-dinitrotoluene and (b) 4-nitrophenol were given in Figure 2.4.



(a)



(b)

Figure 2.4. Structures of (a) 2,6-dinitrotoluene and (b) 4-nitrophenol

2.4.2. Environmental Importance

2.4.2.1. Pathways of Benzene Derivatives to Humans. Occupational exposures predominantly account for human exposure to benzene derivatives. In 1987, approximately 238,000 workers employed by the rubber industry, oil refineries, chemical plants, the shoe manufacturing industry, gasoline storage facilities, and service stations were exposed to benzene derivatives (Mehlman, 1991).

As can be seen from the environmental sources given in section 2.4.1, inhalation accounts for up to 99% of the total daily intake of benzene derivatives (Hattemer-Frey et al., 1990; ATSDR, 1996). Anthropogenic emissions to the air are approximately 34,000 metric tons per year (USEPA, 1989), primarily because of industry-related releases to the environment. Smoking, however, is the largest anthropogenic source of benzene exposure for the general public.

Benzene derivatives are also found in other environmental media. They are released to water through discharge of industrial waste water, leachate from landfills, and gasoline leaks from underground storage tanks. According to the United States Environmental Protection Agency STOrage and RETrieval (STORET) database, benzene derivatives can be detected in surface water, unpolluted coastal water, polluted ocean water, and polluted coastal water. Benzene derivatives are found in soil following releases from industrial discharges, land disposal of benzene wastes, gasoline leaks from underground storage tanks, and underground injection of wastes containing benzene (ATSDR, 1996). Ingestion

of contaminated food items has been suggested as a potentially important pathway of human exposure to benzene derivatives (Gilbert et al., 1982).

2.4.2.2. Adverse Health Effects of Benzene Derivatives. Benzene is a known hematopoietic poison and bone marrow depressant (Zhu et al., 1995), having first been reported to produce anemia and leukopenia as long ago as 1897. In view of the reported literature, it is possible that chronic exposure to benzene derivatives in the environment can result in blood disorders. Some researchers have reported that a higher proportion of the toxic metabolites of benzene are produced at lower exposure concentrations (Sabourin et al., 1990). Sudden deaths from cardiovascular problems, predominantly ventricular fibrillation, have also been observed in laboratory animals following acute benzene poisoning (Leong, 1977). Several studies implicate the immune system as a target for benzene derivatives. Benzene is a known immunotoxicant (McMurry et al., 1991). Benzene derivatives does cause changes in circulating leukocytes, including lymphocytes.

Benzene derivatives have been shown to affect both the peripheral and central nervous systems. The signs and symptoms reported include vertigo, drowsiness, euphoria, headache, giddiness, narcosis, muscular incoordination, convulsions, paralysis, and unconsciousness (Davies and Levine, 1986). Chronic industrial exposure has also been reported to cause neurological abnormalities, such as global atrophy of the lower extremities and distal neuropathy of the upper extremities, as well as abnormal electroencephalograms (Kellerova, 1985).

According to the NIOSH (1988), acute exposure to relatively high levels of benzene in the air has been reported to cause respiratory irritation, pulmonary edema, and pneumonia. Chronic exposure to benzene in the air can cause labored breathing.

Benzene exposure, and its association with cancer, has been extensively covered in the literature. There is sufficient evidence to have declared benzene a carcinogen (Ning et al., 1991). Many authors (Bois et al., 1991; Eastmond, 1993) have demonstrated that the metabolites of benzene are primarily responsible for the carcinogenic action of benzene.

Exposure to benzene or mixtures containing benzene have also been implicated in causing such types of cancer in humans as bladder, stomach, prostate, and lung neoplasms (Parker et al., 1982; Bernardinelli et al., 1987), primarily observed in rubber workers. Stomach cancer was also reported for shoe workers (Fu et al., 1996).

2.4.3. Recent Studies on Benzene Derivatives

Hall et al. (1984) made structure-activity relationship studies on the toxicities of benzene derivatives. Hall et al. (1989) used an additivity model to correlate the acute toxicity of 105 substituted benzenes. Gao et al. (1992) developed a more general group contribution approach to correlate the toxicity of 130 mostly substituted benzene compounds. In this approach, chemical fragments, called groups, are used to describe a molecule. Karabunarliev et al. (1996) estimated the acute toxicity of substituted benzenes to the guppy (*Poecilia reticulata*) and fathead minnow (*Pimephales promelas*) by using quantum chemical descriptors. Gute and Basak (1997) predicted 50% lethal concentration (LC_{50}) of benzene derivatives using theoretical molecular descriptors. Basak et al. (2000) developed QSAR in constructing models for estimating physico-chemical, biomedicinal, and toxicological properties of interest. Lu et al. (2001) developed QSAR model by using E_{LUMO} and by using the hydrophobicity parameter (log K_{OW}) to predict EC_{50} of benzene derivatives and obtained a series of equations about the measured EC_{50} values of different subclasses of compounds.

3. MATERIALS AND METHODS

3.1. Data Set

The experimental data of K_{OW}, H, and S for 56 non-ionic organic compounds (PCDD/PCDFs and phthalates) were taken from (EPIWIN) software and extensive literature sources (Hollifield, 1979; Cowen and Baynes, 1980; Veith et al., 1980; Wolfe et al., 1980b; Sears and Darby, 1982; Leyder and Boulanger, 1983; Howard et al., 1985; Stephenson and Malanowski, 1987; Doucette and Andren, 1988; Shiu et al., 1988; Nielson and Bundgaard, 1989; Brooke et al., 1990; DeFoe et al., 1990; Friesen et al., 1993; Ellington and Floyd, 1996; Exxon, 1996; Govers and Krop, 1998; Ellington, 1999; Thomsen et al., 1999; Chen et al., 2001). These compounds included selected 18 polychlorinated dibenzo-*p*-dioxins, 14 polychlorinated dibenzo-*p*-furans, and 24 phthalate esters. Because of lack of the study concerning the biological properties of PCDD/PCDFs and phthalates in the literature, QSPR models were developed for EC_{50} and *BCF* of 36 substituted benzenes including nitrobenzenes, anilines and phenols. Experimental 48h-EC_{50} data of selected 36 substituted benzenes to algae (*Scenedesmus obliquus*) were taken from Lu et al. (2001). Experimental fish *BCF* values for selected 18 substituted benzenes were taken from Lu et al. (2000) and Devillers et al. (1996). All solubility data were converted to units of mol L^{-1} and the unit of log H is (atm m^3 mol^{-1}). The unit of 48h-EC_{50} is mol L^{-1}.

3.2. Descriptors

3.2.1. The Characteristic Root Index (*CRI*)

The *CRI* is a hybrid of path-type valence connectivity index and a distance matrix (Saçan and İnel, 1993). It is the insertion of valence connectivity into a distance matrix. Graph, G, is defined as a square matrix with the entries, w_{ij}, representing the weighted distance traversed in moving from vertex i, to vertex j in G. Thus, the constructed matrix comprises all possible orders of path-type valence connectivity index for a molecule, except zero order (Saçan et al., 2003).

It is very difficult to show the stepwise calculation of the *CRI* for any congener. The smallest PCDD congener, dibenzo-*p*-dioxin, has a matrix dimension of 14×14. Therefore, stepwise calculation of the *CRI* is demonstrated for a simple molecule, *n*-propane.

Stepwise calculation of the CRI for n-propane

(a) Molecular Structure; CH_3-CH_2-CH_3

n-propane

(b) Hydrogen-Depleted structural graph, *G*, with vertex valencies.

G: •—•—•

1 2 1

(c) Arbitrarily numbering of each vertex which will form the rows and columns of the matrix.

G: •—•—•

1 2 3

(d) Computation of matrix entries, w_{ij}.

$w_{11} = 0$	$w_{12} = (2.1)^{-1/2}$	$w_{13} = (1.2.1)^{-1/2}$
$w_{21} = (1.2)^{-1/2}$	$w_{22} = 0$	$w_{23} = (1.2)^{-1/2}$
$w_{31} = (1.2.1)^{-1/2}$	$w_{32} = (2.1)^{-1/2}$	$w_{33} = 0$

(e) Constructed matrix, *C*(*G*), with the above entries.

$$C(G) = \begin{matrix} 0.0000 & 0.7071 & 0.7071 \\ 0.7071 & 0.0000 & 0.7071 \\ 0.7071 & 0.7071 & 0.0000 \end{matrix}$$

(f) Computation of the coefficients of the characteristic polynomial using Bocher's formula.

$$x^3 - 1.4999x - 0.7068$$

(g) Computed characteristic roots of the polynomial of *n-propane* using the SCIENTIFIC WORK PLACE 3.0.

$$x_1 = 1.4139$$
$$x_2 = -0.7105$$
$$x_3 = -0.7204$$

(h) The *CRI* for *n-propane* is 1.4139.

The starting point of the derivation of the *CRI* is to draw the structural graph, *G*, of a molecule (a).

The calculation of the *CRI* starts from the hydrogen suppressed skeleton of a molecule (b). First, each non-hydrogen atom is numbered arbitrarily (c) and assigned a delta (valence) value, which is calculated from their electronic configuration by eq. (3.1).

$$\delta^v = \frac{(Z^v - h)}{(Z - Z^v - 1)} \qquad (3.1)$$

Z^v is the number of valence electrons in an atom, Z is its atomic number and h is the number of hydrogen atoms bound to the same atom. These delta (valence) values used have been reported by Kier and Hall (Kier and Hall, 1976). The *CRI* is the sum of the positive characteristic roots obtained from the characteristic polynomial of the matrix with the entries calculated from the electronic input information (atomic δ^v values) by eq. (3.2).

$$w_{ij} = (\delta_i^v \, . \, \delta_j^v ..... \delta_n^v)^{-½} \qquad (3.2)$$

where *i*,*j*,.....,*n* correspond to the consecutive non-hydrogen atoms. The entries, w_{ij}, of the matrix are calculated by considering the shortest path to any other non-hydrogen atoms (d). In case of equal paths ($w_{ij} = w_{ji}$) clockwise direction was chosen. So, all possible orders of the connectivity index-except zero order for each chemical are included in the constructed square matrix ($m \times m$) (e). The main diagonal of the matrix contains only zeros, because there is no path bonding to the atom itself. Therefore, entries of this sort (w_{ii}, w_{jj}) are zero (Saçan and Balcıoğlu, 1996).

The following procedure after the construction of the matrix is to convert it to a polynomial form using Bocher's (Istefanopulos, 1987) formula, which states that the sum of the diagonal elements of a square matrix is equal to the sum of its characteristic values. The constructed matrix was loaded into Excel 2002 and the program written in TURBO PASCAL for personal computer uses of this Excel file and converts the stored matrix into the characteristic polynomial (f). Then the *CRI* was obtained by summing up the positive characteristic roots of the polynomial calculated with SCIENTIFIC WORKPLACE 3.0. (g).

3.2.2. Semi-Empirical Quantum Chemical Descriptors

In this study, the individual primary semi-empirical parameters influencing octanol/water partition coefficient, water solubility, Henry's Law Constant, 50% effective inhibition concentration and bioconcentration factor were inclusive of the energy of the highest occupied molecular orbital, the energy of the lowest unoccupied molecular orbital, and dipole moment. They were calculated by the quantum chemical PM3 method (Stewart, 1989) which utilizes semi-empirical parameters. Units of E_{HOMO}, E_{LUMO} energies and dipole moment are electron volts (eV) and debye, respectively. Geometries of 92 non-ionic organic compounds were fully optimized. For each flexible molecule, conformational analysis has been performed with the semi-empirical PM3 method including the effect of aqueous solvent. Aqueous solvation energies of all the conformers have been estimated using the SM54 model (Chambers et al., 1996b) and added to gas phase energies. For every compound molecular orbital energies and the dipole moments have been calculated for the conformer having the lowest total energy. SPARTAN PRO software was used for quantum chemical computations.

3.3. Model Development and Robustness Test

Multiple linear regressions were used to fit the physico-chemical and biological parameters of compounds in the data set. The coefficient of determination for multiple linear regressions (r) and the standard error of the estimate (*S.E.*), Fisher statistic (F), and t-values for individual variables were taken into consideration in testing the quality of the regression. The studentized residual was used as an additional statistic in testing the quality of the regression. Additionally, the predictive performance of the model was checked by the classical internal validation. A modified jackknife test (Dietrich et al., 1980) (leave one/three-out, leave less than 10% out) was performed for regression validation. To verify predictive stability of the model, a strongest cross-validation by the leave-subclass-out procedure was also performed. A multiple linear regression was performed after each deletion to calculate the jackknife coefficient of determination and then compared with the same parameter derived prior to deletion of a chemical group for purpose of robustness evaluation.

The best-fit multiple linear regression was calculated for each parameter in a stepwise manner including descriptors only if they provide a statistically significant improvement in the model. The best-fit model was further tested by the exchange of descriptors, i.e. descriptors previously excluded were exchanged for those found to give a significant improvement, to ensure that the model was stable.

Intercorrelation between variables was adhered to two conditions. All of the multiple correlation coefficients are zero and all of the variance inflation factors (VIFs) are unity. The VIF is defined as $1/(1-r^2)$, where r is the correlation coefficient of one independent variable against others. Large VIF values imply a strong correlation. With the best developed equations, three physico-chemical properties of chemicals outside the sample set were predicted and compared with the reported literature data. All of the statistical analyses were performed using STATISTICA 6.0. Software.

Futhermore, to compare the predictive performance of the models in this study with the literature models, RMS (i.e., root-mean square error) values (eq. 3.3) of five models was calculated (Devillers et al., 1996). RMS value for solubility was calculated according

to the below equation. RMS values of the other properties were calculated similarly. For example; for the calculation of the RMS value of Henry's Law Constant, log *S* term was replaced by log *H* in the following equation.

$$\text{RMS} = ((\sum(\log S_{obs} - \log S_{calc})^2) / \text{number of observation})^{1/2} \qquad (3.3)$$

4. RESULTS AND DISCUSSION

For 92 compounds, the *CRI* and the semi-empirical quantum chemical descriptors- E_{HOMO}, E_{LUMO} and μ- were calculated and listed in Table 4.1 together with CAS registry numbers. A number of physico-chemical and biological properties are of environmental importance, as the chapter 2 given above illustrate. The results of the developed models for the prediction of *S*, K_{OW}, *H*, 48h-EC_{50}, and *BCF* were given and discussed in the following sections.

Table 4.1. Calculated values for the descriptors of selected 92 endocrine disrupting chemicals

No	Name	CAS Number	*CRI*	E_{HOMO} (eV)	E_{LUMO} (eV)	Dipole (μ) debye
1	dibenzo-*p*-dioxin	262-12-4	2.6611	-8.65	-0.18	0.00
2	1-monochlorodibenzo-*p*-dioxin	39227-53-7	3.0935	-8.75	-0.30	0.74
3	2-monochlorodibenzo-*p*-dioxin	39227-54-8	3.1402	-8.70	-0.37	1.05
4	2,3-dichlorodibenzo-*p*-dioxin	29446-15-9	3.5619	-8.72	-0.53	1.57
5	2,7-dichlorodibenzo-*p*-dioxin	33857-26-0	3.5815	-8.76	-0.52	0.00
6	2,8-dichlorodibenzo-*p*-dioxin	38964-22-6	3.5723	-8.76	-0.53	0.87
7	1,2,4-trichlorodibenzo-*p*-dioxin	39227-58-2	4.0048	-8.86	-0.60	1.17
8	1,2,3,4-tetrachlorodibenzo-*p*-dioxin	30746-58-8	4.4838	-8.85	-0.71	1.59
9	1,2,3,7-tetrachlorodibenzo-*p*-dioxin	67028-18-6	4.5787	-8.84	-0.75	0.64
10	1,3,6,8-tetrachlorodibenzo-*p*-dioxin	33423-92-6	4.6206	-8.93	-0.72	0.00
11	2,3,7,8-tetrachlorodibenzo-*p*-dioxin	1746-01-6	4.5386	-8.80	-0.78	0.00
12	1,2,3,4,7-pentachlorodibenzo-*p*-dioxin	39227-61-7	4.9909	-8.89	-0.82	0.77
13	1,2,3,7,8-pentachlorodibenzo-*p*-dioxin	40321-76-4	4.9948	-8.85	-0.85	0.43
14	1,2,3,4,7,8-hexachlorodibenzo-*p*-dioxin	39227-28-6	5.4201	-8.90	-0.92	0.06
15	1,2,3,6,7,8-hexachlorodibenzo-*p*-dioxin	57653-85-7	5.4276	-8.91	-0.92	0.00
16	1,2,3,7,8,9-hexachlorodibenzo-*p*-dioxin	19408-74-3	5.4321	-8.90	-0.92	0.80
17	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-dioxin	35822-46-9	6.0133	-8.96	-0.98	0.38
18	octachlorodibenzo-*p*-dioxin	3268-87-9	6.4683	-9.01	-1.04	0.00
19	dibenzo-*p*-furan	132-64-9	2.4600	-9.02	-0.48	0.83
20	2,8-dichlorodibenzo-*p*-furan	5409-83-6	3.5013	-9.10	-0.79	0.47
21	1,2,7,8-tetrachlorodibenzo-*p*-furan	58802-20-3	4.4592	-9.05	-1.04	1.02
22	2,3,7,8-tetrachlorodibenzo-*p*-furan	51207-31-9	4.4703	-9.03	-1.07	0.12
23	2,3,4,7,8-pentachlorodibenzo-*p*-furan	57117-31-4	4.9503	-9.07	-1.17	0.62
24	1,2,3,8,9-pentachlorodibenzo-*p*-furan	83704-54-5	4.9245	-9.10	-1.14	1.20
25	1,2,3,4,7,8-hexachlorodibenzo-*p*-furan	70648-26-9	5.3758	-9.02	-1.27	0.16
26	1,2,3,4,8,9-hexachlorodibenzo-*p*-furan	92341-07-6	5.3550	-9.05	-1.24	0.92
27	1,2,3,6,7,8-hexachlorodibenzo-*p*-furan	57117-44-9	5.3925	-9.10	-1.26	0.35

Table 4.1. Continued

No	Name	CAS Number	*CRI*	E_{HOMO} (eV)	E_{LUMO} (eV)	Dipole (μ) debye
28	1,2,3,7,8,9-hexachlorodibenzo-*p*-furan	72918-21-9	5.3805	-9.06	-1.26	0.70
29	2,3,4,6,7,8-hexachlorodibenzo-*p*-furan	60851-34-5	5.4052	-9.12	-1.25	1.04
30	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-furan	67562-39-4	5.8205	-9.07	-1.35	0.62
31	1,2,3,4,7,8,9-heptachlorodibenzo-*p*-furan	55673-89-7	5.8021	-9.05	-1.35	0.26
32	octachlorodibenzo-*p*-furan	39001-02-0	6.3087	-9.09	-1.43	0.22
33	dimethyl phthalate	131-11-3	2.4899	-10.71	-0.84	7.76
34	diethyl phthalate	84-66-2	3.5713	-10.66	-0.81	5.98
35	diallyl phthalate	131-17-9	3.4875	-10.61	-0.86	7.81
36	dipropyl phthalate	131-16-8	4.3391	-10.68	-0.82	7.82
37	di-*n*-butyl phthalate	84-74-2	5.0230	-10.65	-0.80	8.09
38	diisobutyl phthalate	84-69-5	5.0586	-10.68	-0.81	7.69
39	butylbenzyl phthalate	85-68-7	4.9061	-9.86	-0.83	7.78
40	dihexyl phthalate	84-75-3	6.3282	-10.47	-0.68	5.87
41	di-*n*-octyl phthalate	117-84-0	7.8247	-10.32	-0.60	3.24
42	butyl 2-ethylhexyl phthalate	85-69-8	6.4121	-10.66	-0.84	7.74
43	decylhexyl phthalate	25724-58-7	7.6932	-10.61	-0.84	8.15
44	di (2-ethylhexyl) phthalate	117-81-7	7.8023	-10.50	-0.69	5.31
45	diisooctyl phthalate	27554-26-3	7.8173	-10.66	-0.83	7.91
46	diisononyl phthalate	28553-12-0	8.4663	-10.48	-0.64	5.49
47	diisodecyl phthalate	26761-40-0	11.2205	-10.32	-0.45	3.27
48	di-C9-11-branched alkyl phthalate	68515-49-1	10.3898	-10.32	-0.55	3.09
49	diundecyl phthalate	3648-20-2	12.3086	-10.32	-0.43	1.34
50	diheptyl phthalate	68515-44-6	6.9082	-10.31	-0.43	1.31
51	dinonyl phthalate	68515-45-7	8.4678	-10.49	-0.63	5.33
52	heptylnonyl phthalate	111381-89-6	7.6197	-10.67	-0.84	7.92
53	heptylundecyl phthalate	111381-90-9	8.2577	-10.32	-0.50	3.23
54	nonylundecyl phthalate	111381-91-0	11.0900	-10.31	-0.43	1.10
55	ditridecyl phthalate	119-06-2	15.7062	-10.32	-0.43	1.40
56	di-C11-14-branched alkyl phthalate	68515-47-9	13.7780	-10.32	-0.60	3.18
57	4-nitrotoluene	99-99-0	1.8382	-10.47	-1.11	5.74
58	1,3-dinitrobenzene	99-65-0	1.7718	-11.47	-1.96	4.85
59	1,4-dinitrobenzene	100-25-4	1.6960	-11.30	-2.25	0.00
60	2,4-dinitrotoluene	121-14-2	2.0865	-11.19	-1.83	5.39
61	2,6-dinitrotoluene	606-20-2	2.0482	-11.08	-1.65	3.65
62	nitrobenzene	98-95-3	1.4648	-10.60	-1.13	5.25
63	2-nitroanisole	91-23-6	1.9195	-9.98	-0.74	5.88
64	3-nitroanisole	555-03-3	1.9069	-9.90	-1.14	6.27
65	4-nitroanisole	100-17-4	1.8535	-10.07	-1.03	6.03
66	3-bromonitrobenzene	585-79-5	2.5154	-10.52	-1.35	4.77
67	4-bromonitrobenzene	586-78-7	2.4516	-10.70	-1.39	4.26
68	2,4-dinitroaniline	97-02-9	1.8738	-9.87	-1.63	6.62
69	2-methylaniline	95-53-4	1.8126	-8.54	0.41	1.32
70	2,5-dichloroaniline	95-82-9	2.4309	-8.77	-0.17	1.38

Table 4.1. Continued

No	Name	CAS Number	*CRI*	E_{HOMO} (eV)	E_{LUMO} (eV)	Dipole (μ) debye
71	3-bromoaniline	591-19-5	2.4224	-8.83	-0.01	1.99
72	2,4,6-tribromoaniline	147-82-0	4.3782	-9.00	-0.71	1.59
73	4-nitrophenol	100-02-7	1.5750	-10.17	-1.08	5.57
74	2,4-dinitrophenol	51-28-5	1.8087	-10.95	-1.75	6.11
75	2-chloronitrobenzene	88-73-3	2.0236	-9.99	-1.19	5.27
76	3-chloronitrobenzene	121-73-3	1.9745	-10.06	-1.30	4.83
77	4-chloronitrobenzene	100-00-5	1.9661	-10.22	-1.36	4.59
78	3,4-dichloronitrobenzene	99-54-7	2.4639	-9.82	-1.32	4.86
79	2,5-dichloronitrobenzene	89-61-2	2.5248	-9.78	-1.36	4.85
80	2-nitroaniline	88-74-4	1.6568	-9.16	-1.00	4.88
81	3-nitroaniline	99-09-2	1.6506	-9.29	-1.06	5.70
82	4-nitroaniline	100-01-6	1.6215	-9.42	-1.01	6.64
83	aniline	62-53-3	1.4048	-8.61	0.42	1.30
84	2-chloroaniline	95-51-2	1.9269	-8.66	0.13	1.47
85	3-chloroaniline	108-42-9	1.9378	-8.76	0.12	1.79
86	2,4-dichloroaniline	554-00-7	2.3276	-8.65	-0.12	1.85
87	2-nitrophenol	88-75-5	1.5831	-9.90	-1.22	4.18
88	3-nitrophenol	554-84-7	1.5792	-9.96	-1.18	6.12
89	phenol	108-95-2	1.3596	-9.17	0.29	1.14
90	2,4-dichlorophenol	120-83-2	2.3675	-9.09	-0.24	0.52
91	2,4,6-trichlorophenol	88-06-2	2.8488	-9.12	-0.44	1.00
92	pentachlorophenol	87-86-5	3.6987	-9.14	-0.79	1.10

4.1. Solubility (*S*)

The forward stepwise multiple linear regression (MLR) using the topology based and semi-empirical quantum chemical descriptors resulted in a four-parameter model (model 1) for the estimation of solubility.

$$\log S = -0.546\ (\pm 0.129)\ CRI - 1.648\ (\pm 0.906)\ E_{HOMO} + 3.322\ (\pm 0.733)\ E_{LUMO} - 0.017\ (\pm 0.212)\ \mu - 16.713\ (\pm 7.570) \quad \text{Model (1)}$$

$$n = 40; \quad r = 0.807; \quad S.E. = 1.356; \quad F_{(4,35)} = 16.30$$

However, the high correlation coefficient ($r = 0.879$) between E_{HOMO} and μ, and VIF for the initial model (model 1) which was given in Table 4.2 implies that these variables

are collinear. The correlation of relevant pairs of explanatory variables is needed to be tested when three or more variables are involved in multicollinearities. The ideal situation in regression is that all of the multiple correlation coefficients are zero, and all of the variance inflation factors are unity. Generally, a value of 1.0 indicates no correlation, a value greater than 2.0 is usually considered problematic. Values over 10.0 indicate an unstable regression that must be reexamined. Multiple linear regression analysis with log *S* as the dependent variable, and one topological indices and the two quantum chemical descriptors as the independent variables, for the PCDD/PCDFs and phthalates having no outlier and no collinearity resulted in another QSPR model (model 2). The stepwise multiple linear regression excluded E_{HOMO} from model (1). We did examine both E_{HOMO} and μ as independent variables in regressions of log *S*, and μ significantly improved the correlation. The *t*-test method was used to test the correlation of each independent variable. The numbers in the parentheses are the 95% confidence intervals of that coefficient.

Table 4.2. Correlation coefficient matrix for independent variables and the variance inflation factors (VIFs) for model (1) and model (2)

Parameter	**Correlation Matrix**				**VIF**	
	CRI	E_{HOMO}	E_{LUMO}	μ	Model (1)	Model (2)
CRI	1.000				2.573	1.308
E_{HOMO}	-0.434	1.000			11.194	
E_{LUMO}	0.064	-0.063	1.000		1.008	1.208
μ	0.071	-0.879	0.059	1.000	9.145	1.121

$$\log S = -1.028\ (\pm 0.042)\ CRI + 0.810\ (\pm 0.230)\ E_{LUMO} + 0.443\ (\pm 0.020)\ \mu - 2.050\ (\pm 0.219) \qquad \text{Model (2)}$$

$$n = 35; \quad r = 0.986; \quad S.E. = 0.347; \quad F_{(3,31)} = 362.44$$

t-values for partial correlation coefficients in model (2) were listed in Table 4.3.

Table 4.3. *t*-values for partial correlation coefficients in model (2)

	***t*(31)**	***p*-level**
Intercept	9.353	1.55×10^{-10}
CRI	24.462	7.87×10^{-22}
E_{LUMO} (eV)	-3.529	1.33×10^{-3}
μ (debye)	-22.645	7.62×10^{-21}

The values of *S* at 298 K (25 ºC) ranged from 1.45×10^{-2} mol L^{-1} for dimethyl phthalate to 1.41×10^{-10} mol L^{-1} for octachlorodibenzo-*p*-dioxin. The *CRI* was the most important factor for the prediction of log *S*. Model (2) implies that for the studied EDCs with high molecular size or the *CRI*, and low E_{LUMO} and μ values, have low water solubilities. This is reasonable since the molecular properties that affect aqueous solubility are mostly the size, shape, and polarity of the molecule. These features of a compound can be represented numerically in many ways such as molecular weight, surface area, polarity, steric effects and the ability of the molecule to participate in hydrogen bonding. The *CRI* which comprises all possible orders of path-type molecular connectivity indices encodes global molecular properties such as size, volume and surface area emphasizing the strong dependence of log *S* of the studied compounds on the compound size. Local structural properties and possibly long-range interactions described by path-type molecular indices may also be encoded in the *CRI*. Since the size and shape of the molecule are very important factors in a compound's aqueous solubility, it is not surprising that the *CRI* would be an important factor in determining aqueous solubility. On the other hand, E_{LUMO} is related to the hydrogen-bonding term. It reflects the reactivity of the whole molecule and depends on the number and type of hetero atoms as well as on the type of the polarized group. Therefore, energy levels are normally important factors determining the thermodynamics of all kinds of specific interactions; e.g. interaction between chemicals and water. In addition, since the polarity of a molecule is also a very important factor in a compound's aqueous solubility, it is not surprising that the dipole moment would be an important factor in determining aqueous solubility.

The relative errors in model (2) for the individual chemicals studied were examined using the studentized residuals compared with the calculated log *S* values. The studentized

residuals are symmetrically distributed around zero with no specific trend for model (2) as shown in Figure 4.1. All compounds were estimated within a two standard deviation range.

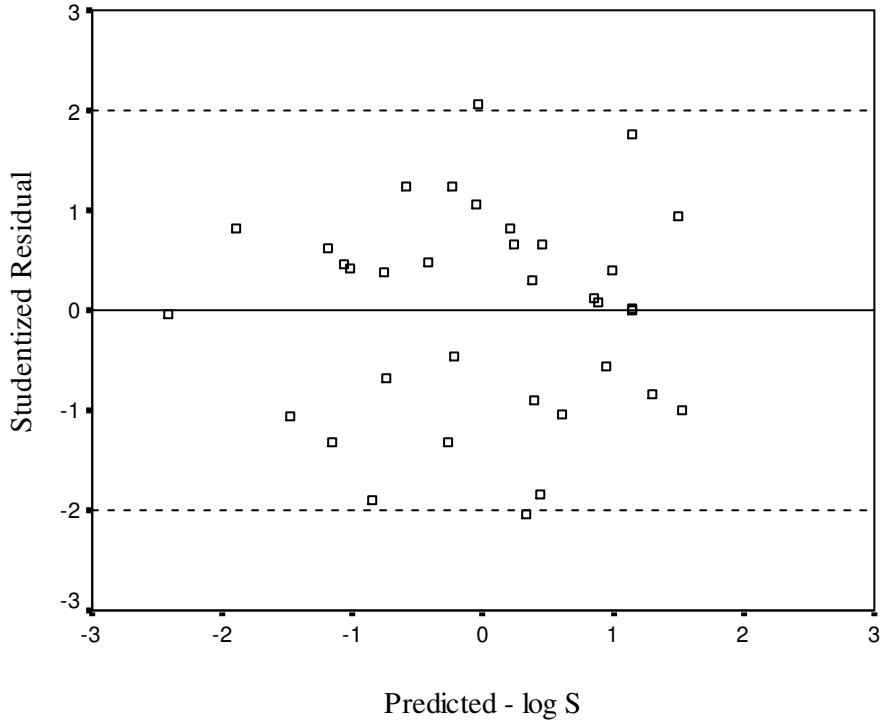


Figure 4.1. Plot of studentized residual versus calculated - log *S* from model (2)

For model (2), plot of observed and predicted values is shown by Figure 4.2. The observed values, fitted values obtained from the developed model and the residuals for the studied chemicals are listed in Table 4.4 together with the reported literature ranges.

Considering the relatively small number of collected experimental log *S* values in this study, the database was not divided into modeling and testing sets for purposes of model validation. Instead, an alternative approach using a modified jackknife test was applied. We performed a jackknife test with removal of less than 10% randomly selected compounds in each run, the regression then was rerun for all the other observed values. The overall results of the deletion study (leave one/three-out method) for model (2) are summarized in Table 4.5.

Table 4.4. Experimental, predicted and reported literature log *S* (at 25 °C)

No	Name	CAS Number	- log *S* (mol L^{-1}) Observed	- log *S* (mol L^{-1}) Pred. by Model (2)	Residual	- log *S* (mol L^{-1}) Literature Range	
						Minimum	Maximum
1	dibenzo-*p*-dioxin	262-12-4	4.36 [1]	4.93	-0.57	4.36 [1]	5.34 [14]
2	1-monochlorodibenzo-*p*-dioxin	39227-53-7	4.92 [1]	5.14	-0.22	4.92 [1]	5.96 [2]
3	2-monochlorodibenzo-*p*-dioxin	39227-54-8	5.24 [1]	5.11	0.13	5.24 [1]	6.18 [2]
4	2,3-dichlorodibenzo-*p*-dioxin	29446-15-9	5.86 [1]	5.45	0.41	5.86 [1]	7.23 [14]
5	2,7-dichlorodibenzo-*p*-dioxin	33857-26-0	6.00 [1]	6.15	-0.15	5.93 [1]	7.83 [2]
6	2,8-dichlorodibenzo-*p*-dioxin	38964-22-6	5.93 [1]	5.76	0.17	5.93 [1]	7.18 [14]
7	1,2,4-trichlorodibenzo-*p*-dioxin	39227-58-2	6.55 [1]	6.14	0.41	6.55 [1]	7.83 [2]
8	1,2,3,4-tetrachlorodibenzo-*p*-dioxin	30746-58-8	7.20 [1]	6.53	0.67	7.20 [1]	8.77 [14]
9	1,2,3,7-tetrachlorodibenzo-*p*-dioxin	67028-18-6	7.30 [1]	7.08	0.22	7.30 [1]	8.89 [15]
10	1,3,6,8-tetrachlorodibenzo-*p*-dioxin	33423-92-6	7.08 [1]	7.39	-0.31	6.91 [1]	9.00 [15]
11	2,3,7,8-tetrachlorodibenzo-*p*-dioxin	1746-01-6	7.45 [1]	7.35	0.10	7.45 [1]	11.61 [16]
12	1,2,3,4,7-pentachlorodibenzo-*p*-dioxin	39227-61-7	7.73 [1]	7.51	0.22	7.73 [1]	9.48 [15]
14	1,2,3,4,7,8-hexachlorodibenzo-*p*-dioxin	39227-28-6	8.37 [1]	8.34	0.03	8.37 [1]	10.95 [15]
17	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-dioxin	35822-46-9	8.86 [1]	8.86	0.00	8.86 [1]	10.41 [2]
18	octachlorodibenzo-*p*-dioxin	3268-87-9	9.85 [1]	9.54	0.31	9.60 [1]	12.79 [15]
19	dibenzo-*p*-furan	132-64-9	4.74 [2]	4.60	0.14	3.43 [1]	5.06 [2]
20	2,8-dichlorodibenzo-*p*-furan	5409-83-6	5.64 [1]	6.08	-0.44	5.05 [1]	7.21 [17]
22	2,3,7,8-tetrachlorodibenzo-*p*-furan	51207-31-9	6.87 [1]	7.46	-0.59	6.87 [1]	8.12 [2]
23	2,3,4,7,8-pentachlorodibenzo-*p*-furan	57117-31-4	7.47 [1]	7.81	-0.34	7.47 [1]	8.59 [2]
25	1,2,3,4,7,8-hexachlorodibenzo-*p*-furan	70648-26-9	8.67 [1]	8.54	0.13	8.15 [1]	9.80 [2]
27	1,2,3,6,7,8-hexachlorodibenzo-*p*-furan	57117-44-9	8.28 [1]	8.46	-0.18	8.22 [1]	9.80 [2]
30	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-furan	67562-39-4	9.40 [1]	8.85	0.55	8.76 [1]	11.48 [2]
32	octachlorodibenzo-*p*-furan	39001-02-0	9.28 [1]	9.60	-0.32	9.28 [1]	11.58 [17]
33	dimethyl phthalate	131-11-3	1.84 [3]	1.85	-0.01	1.42 [2]	1.98 [2]
35	diallyl phthalate	131-17-9	3.13 [4]	2.87	0.26	3.13 [4]	3.76 [18]
36	dipropyl phthalate	131-16-8	3.37 [4]	3.71	-0.34	3.37 [4]	3.82 [18]
37	di-*n*-butyl phthalate	84-74-2	4.48 [5]	4.28	0.21	4.33 [9]	5.27 [18]
38	diisobutyl phthalate	84-69-5	4.65 [6]	4.50	0.15	4.14 [12]	4.74 [2]
39	butylbenzyl phthalate	85-68-7	3.89 [7]	4.32	-0.43	3.89 [7]	5.67 [18]
40	dihexyl phthalate	84-75-3	6.86 [8]	6.50	0.36	6.14 [10]	7.46 [2]
41	di-*n*-octyl phthalate	117-84-0	8.88 [8]	9.14	-0.26	5.12 [9]	8.99 [19]
43	decylhexyl phthalate	25724-58-7	7.29 [5]	7.03	0.26	7.29 [5]	8.97 [2]
44	di (2-ethylhexyl) phthalate	117-81-7	8.31 [9]	8.28	0.04	5.99 [13]	8.81 [20]
45	diisooctyl phthalate	27554-26-3	6.64 [10]	7.25	-0.61	6.64 [10]	9.21 [2]
46	diisononyl phthalate	28553-12-0	8.85 [11]	8.84	0.01	6.32 [10]	10.26 [18]

[1]. Govers and Krop, 1998; [2]. EPIWIN; [3]. Nielson and Bundgaard, 1989; [4]. Leyder and Boulanger, 1983; [5]. DeFoe et al., 1990; [6]. Hollifield, 1979; [7]. Veith et al., 1980; [8]. Ellington, 1999; [9]. Wolfe et al., 1980b; [10]. Howard et al., 1985; [11]. Exxon, 1996; [12]. Gledhill et al., 1980; [13]. Russell and McDuffie, 1986; [14]. Shiu et al., 1988; [15]. Shiu and Ma, 2000; [16]. Adams and Blaine, 1986; [17]. Doucette and Andren, 1988; [18]. Meylan and Howard, 1995; [19]. Letinksi et al., 1999; [20]. Staples et al., 1997.

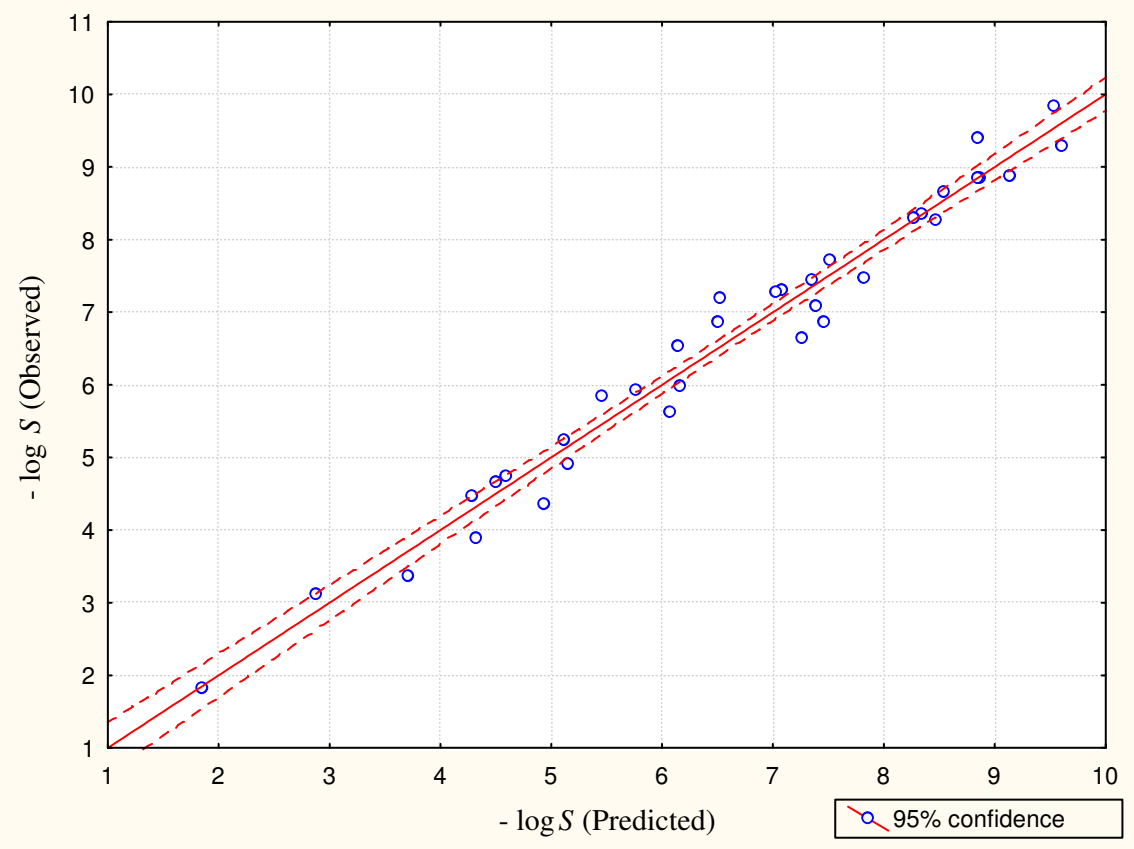


Figure 4.2. Plot of observed and predicted - log *S* values from model (2)

Table 4.5. Summary of results of random deletion test for log *S*

No. of Cases Deleted (<10% of *n*)	**No. of Regression Runs**	**av *r***	**av *S.E.***	**av *F***
1	35	0.986	0.347	352.38
2	70	0.986	0.348	339.75
3	70	0.986	0.348	330.45

The modified jackknife test validated that the developed model (2) was statistically robust. The average *r* values do not have any unduly high variation suggesting that the data set is fairly consistent and the models are not biased by any particular data point. Based on model (2), log *S* for the other 21 chemicals not present in the data set were predicted and listed in Table 4.6.

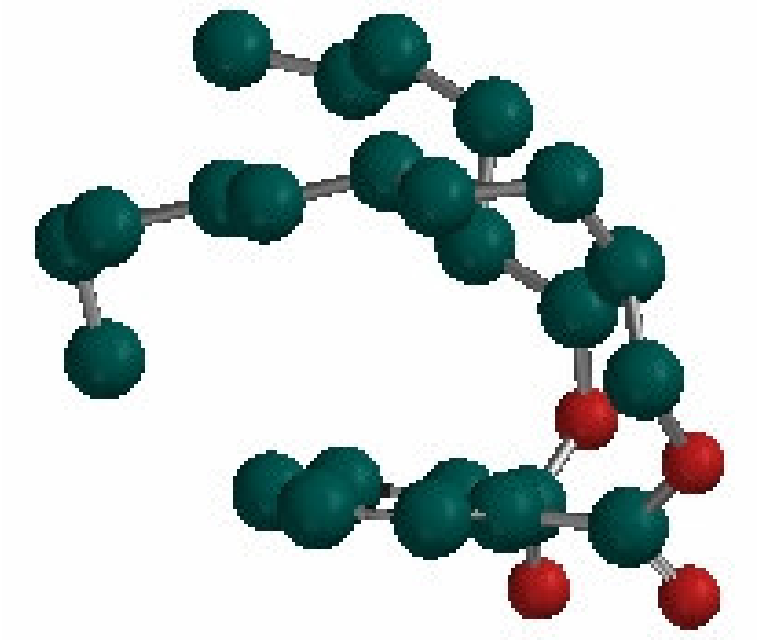
Table 4.6. The reported literature and predicted log *S* from model (2)

No	Name	CAS Number	- log *S* (mol L^{-1}) Pred. by Model (2)	- log *S*, (mol L^{-1}) Literature Range	
				Minimum	Maximum
13	1,2,3,7,8-pentachlorodibenzo-*p*-dioxin	40321-76-4	7.69	8.11 [1]	8.62 [2]
15	1,2,3,6,7,8-hexachlorodibenzo-*p*-dioxin	57653-85-7	8.37	8.65 [1]	10.17 [2]
16	1,2,3,7,8,9-hexachlorodibenzo-*p*-dioxin	19408-74-3	8.02	9.31 [2]	10.17 [2]
21	1,2,7,8-tetrachlorodibenzo-*p*-furan	58802-20-3	7.03	6.79 [1]	7.86 [2]
24	1,2,3,8,9-pentachlorodibenzo-*p*-furan	83704-54-5	7.50	7.69 [2]	8.14 [2]
26	1,2,3,4,8,9-hexachlorodibenzo-*p*-furan	92341-07-6	8.15	8.38 [2]	9.52 [2]
28	1,2,3,7,8,9-hexachlorodibenzo-*p*-furan	72918-21-9	8.29	8.38 [2]	9.52 [2]
29	2,3,4,6,7,8-hexachlorodibenzo-*p*-furan	60851-34-5	8.16	8.38 [1]	9.80 [2]
31	1,2,3,4,7,8,9-heptachlorodibenzo-*p*-furan	55673-89-7	8.99	9.07 [2]	11.48 [2]
34	diethyl phthalate	84-66-2	3.73	2.31 [20]	2.93 [18]
42	butyl 2-ethylhexyl phthalate	85-69-8	5.89	5.94 [21]	7.22 [18]
47	diisodecyl phthalate	26761-40-0	12.50	5.58 [10]	11.31 [18]
48	di-C9-11-branched alkyl phthalate	68515-49-1	11.81	10.30 [2]	11.24 [2]
49	diundecyl phthalate	3648-20-2	14.46	5.64 [10]	12.45 [22]
50	diheptyl phthalate	68515-44-6	8.92	7.41 [2]	8.17 [2]
51	dinonyl phthalate	68515-45-7	8.90	10.01 [2]	10.38 [2]
52	heptylnonyl phthalate	111381-89-6	7.05	8.49 [2]	9.21 [2]
53	heptylundecyl phthalate	111381-90-9	9.51	9.56 [2]	10.26 [2]
54	nonylundecyl phthalate	111381-91-0	13.31	10.41 [2]	11.24 [2]
55	ditridecyl phthalate	119-06-2	17.92	6.19 [6]	14.56 [2]
56	di-C11-14-branched alkyl phthalate	68515-47-9	15.29	12.00 [2]	13.32 [2]

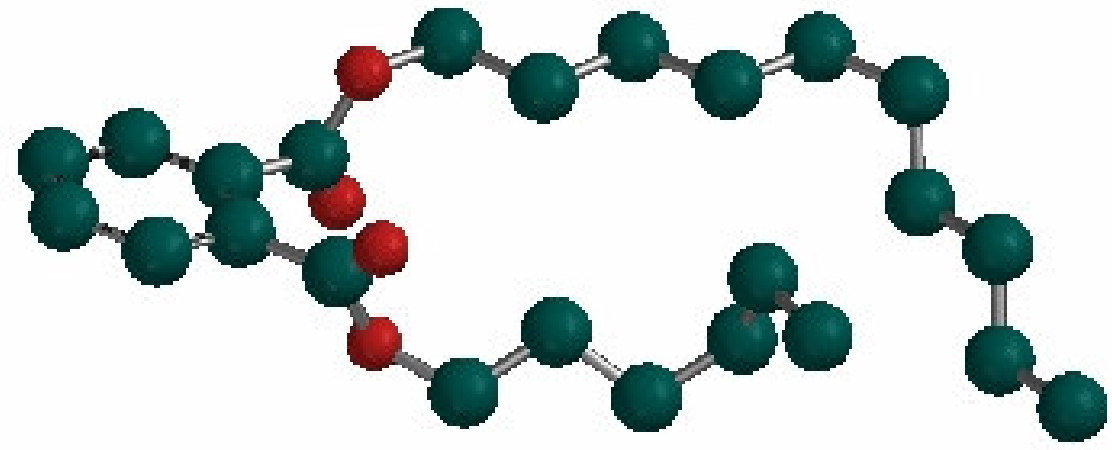
[1]. Govers and Krop, 1998; [2]. EPIWIN; [6]. Hollifield, 1979; [10]. Howard et al., 1985; [18]. Meylan and Howard, 1995; [20]. Staples et al., 1997; [21]. Cousins and Mackay, 2000; [22]. Long, 1995.

The present solubility values were in relatively good agreement with the literature data. The expected decreasing trend in water solubility with increasing alkyl chain length is not apparent for these compounds (Staples et al., 1997). Many of the measured water solubilities for high molecular weight phthalate esters reported in the literature are erroneously too high. Ellington (1999) has suggested that phthalate esters with long alkyl side chains, may rotate and fold to assume conformations of lower energy that more closely resemble a branched-alkyl chain, as opposed to a straight chain. We observed folding and formation of a completely closed loop for phthalates in either their first or their higher lowest energy conformers. For example, decylhexyl phthalate (Figure 4.3) and dinonyl phthalate (Figure 4.4) have folding in their first lowest energy conformers. This folding was represented by the E_{LUMO} and μ values in model (2).

However, some molecules have folding and/or rotation in their higher lowest energy conformer. For example, di-C11-14-branched alkyl phthalate (Figure 4.5) has folding and rotation in the third lowest energy conformer (Figure 4.5). However, the energy of the third stable conformer is only 0.61 kcal mol^{-1} higher than the stable conformer. Therefore, it is expected that it also contributes to the actual shape of di-C11-14-branched alkyl phthalate, as well as to the most stable conformer. Unexpected high solubility of some high molecular weight phthalates can be attributed to the oily nature of the phthalates which leads to the formation of microdroplets in the water phase as stated by Thomsen et al. (1999) or the molecules having folding and/or rotation in their higher energy conformer may exist together with the first lowest energy conformers to some extent in water phase and lead to an increase in solubility.

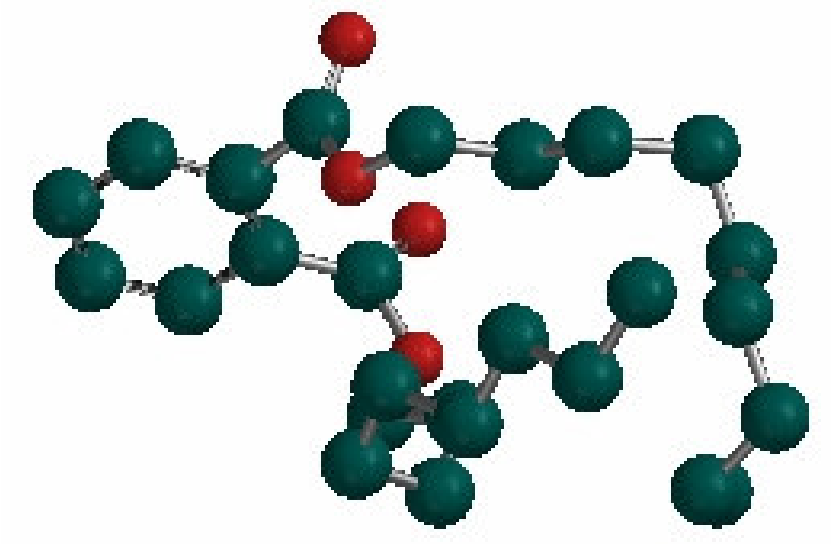


(a)

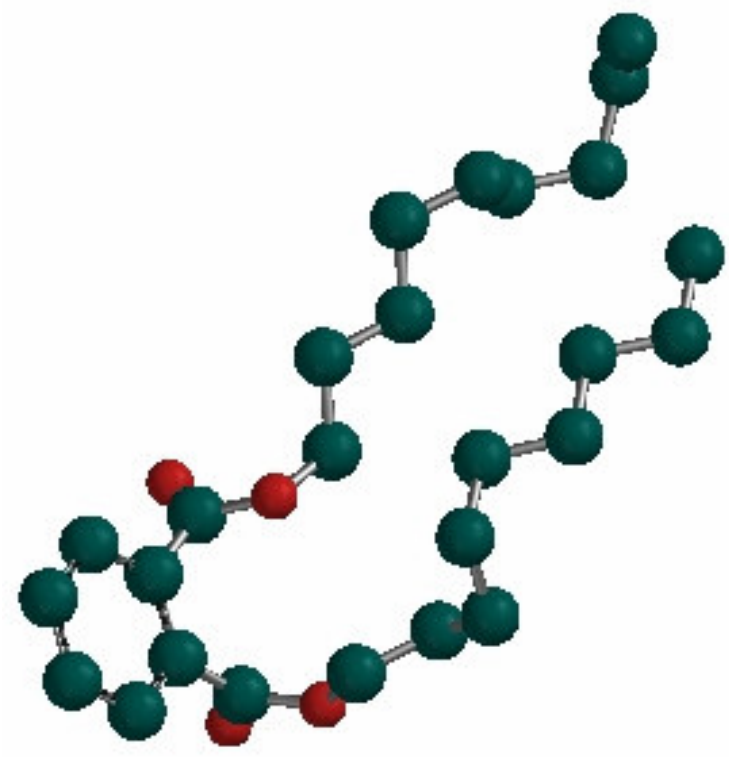


(b)

Figure 4.3. Hydrogen-depleted structure of decylhexyl phthalate conformers: (a) the most stable conformer having side-chain folding (E = -227.396 kcal mol^{-1}), (b) the second stable conformer without folding (E = -222.104 kcal mol^{-1}), (Red balls indicate oxygen atoms)

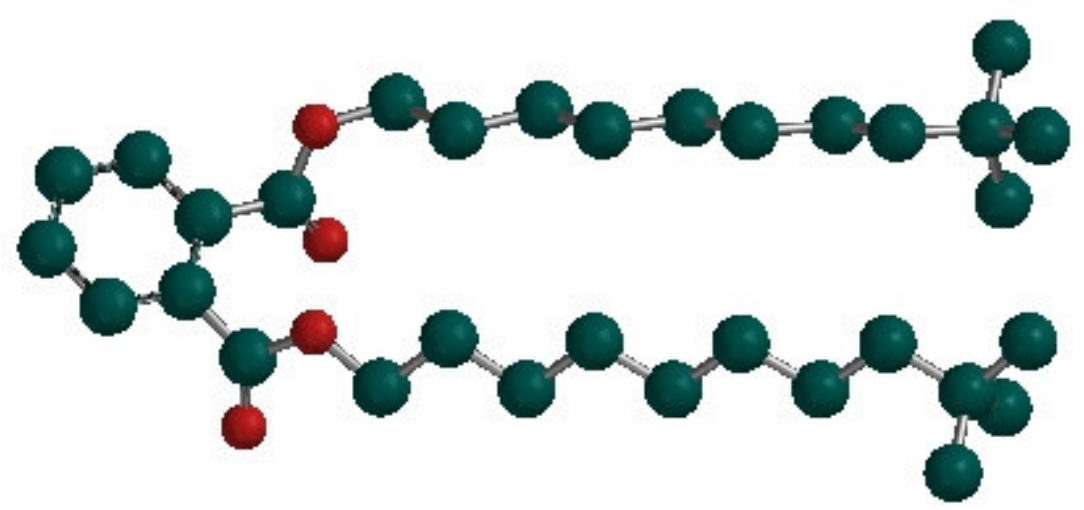


(a)

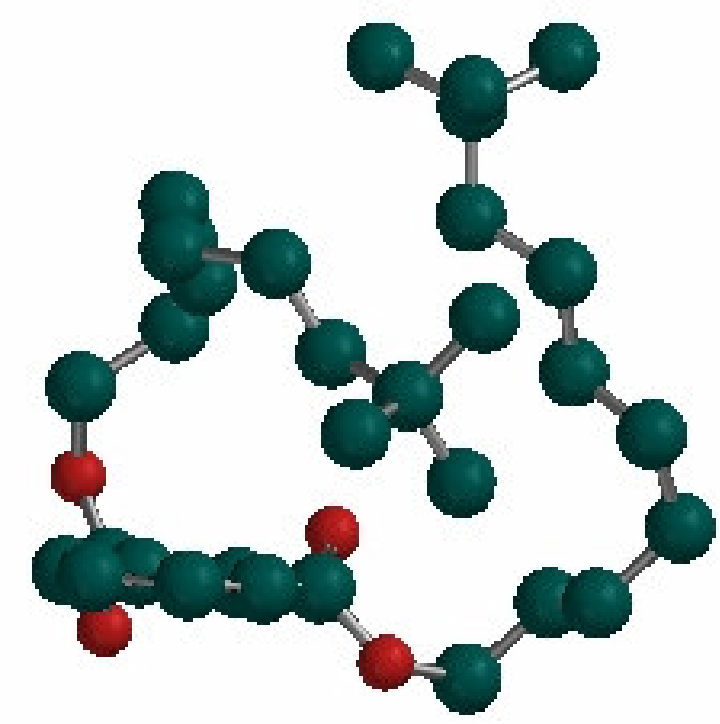


(b)

Figure 4.4. Hydrogen-depleted structure of dinonyl phthalate conformers: (a) the most stable conformer having side-chain folding (E = -235.753 kcal mol^{-1}), (b) moderately stable conformer without folding (E = -231.702 kcal mol^{-1}), (Red balls indicate oxygen atoms)



(a)



(b)

Figure 4.5. Hydrogen-depleted structure of di-C11-14-branched alkyl phthalate conformers: (a) the most stable conformer (E = -270.782 kcal mol^{-1}), (b) the third stable conformer having side-chain folding (E = -270.118 kcal mol^{-1}), (Red balls indicate oxygen atoms)

The average absolute deviation (AAD) of 0.27 log unit for *S* is reasonable for the predictive stability of model (2).

For comparison of the predictive power of model (2) with the literature models, RMS value of model (2) was calculated and provided in Table 4.7. Of the results reported in Table 4.7, SOFA and GC-RI methods for PCDDs in the case of *S* have lower RMS values than the corresponding models developed by our group. However, it must be noted that the total number of compounds in their data set was about 57% less than the data set used by our group.

Table 4.7. RMS deviations of log *S* calculations from different methods

	Model (2)	**SOFA [1]**		**GC-RI [24,25]**		**UNIFAC [23]**
log *S*	PCDD/PCDFs/PAEs	PCDDs	PCDFs	PCDDs	PCDFs	PAEs
RMS	0.327	0.226	0.448	0.139	0.325	1.286
Data points	35	15	6	15	6	15

[1]. Govers and Krop, 1998; [23]. Thomsen et al., 1999; [24]. Wang and Wong, 2002; [25]. Wang and Wong, 2003.

4.2. *n*-Octanol/Water Partition Coefficient (K_{OW})

The following QSPR equation without the outliers was obtained from multiple regression analysis for the *n*-octanol/water partition coefficient of the studied EDCs. The numbers in the parentheses are the 95% confidence intervals of that coefficient.

$$\log K_{OW} = 1.179\ (\pm 0.029)\ CRI + 1.472\ (\pm 0.053)\ E_{HOMO} + 14.271\ (\pm 0.479) \qquad \text{Model (3)}$$

$$n = 34; \quad r = 0.992; \quad S.E. = 0.229; \quad F_{(2,31)} = 994.02$$

Table 4.8. *t*-values for partial correlation coefficients in model (3)

	***t*(31)**	***p*-level**
Intercept	29.767	2.28×10^{-24}
CRI	41.333	1.11×10^{-28}
E_{HOMO} (eV)	27.786	1.79×10^{-23}

As shown in Table 4.8, *t*-values for partial correlation coefficients in model (3) were 41.333 and 27.786 for the *CRI* and E_{HOMO}, respectively.

For model (3), plot of observed and predicted values is shown by Figure 4.6. The observed values, fitted values obtained from the developed model and the residuals for the studied chemicals are listed in Table 4.9 together with the reported literature ranges.

Table 4.9. Experimental, predicted (model 3) and reported literature *n*-octanol/water partition coefficient (at 25 ºC)

No	Name	CAS Number	log K_{OW} Observed	log K_{OW} Pred. by Model (3)	Residual	log K_{OW} Literature Range	
						Minimum	Maximum
1	dibenzo-*p*-dioxin	262-12-4	4.30 [14]	4.69	-0.39	4.15 [29]	4.66 [26]
2	1-monochlorodibenzo-*p*-dioxin	39227-53-7	4.75 [26]	5.04	-0.29	4.75 [26]	5.19 [26]
3	2-monochlorodibenzo-*p*-dioxin	39227-54-8	5.00 [26]	5.16	-0.16	4.87 [26]	5.10 [1]
4	2,3-dichlorodibenzo-*p*-dioxin	29446-15-9	5.60 [26]	5.63	-0.03	5.60 [26]	5.77 [1]
5	2,7-dichlorodibenzo-*p*-dioxin	33857-26-0	5.75 [14]	5.60	0.15	5.63 [2]	5.75 [14]
6	2,8-dichlorodibenzo-*p*-dioxin	38964-22-6	5.60 [26]	5.60	0.00	5.60 [26]	5.68 [1]
7	1,2,4-trichlorodibenzo-*p*-dioxin	39227-58-2	6.35 [26]	5.95	0.40	6.27 [26]	6.45 [1]
8	1,2,3,4-tetrachlorodibenzo-*p*-dioxin	30746-58-8	6.60 [14]	6.53	0.07	6.44 [26]	7.18 [31]
9	1,2,3,7-tetrachlorodibenzo-*p*-dioxin	67028-18-6	6.91 [1]	6.66	0.25	6.81 [26]	7.23 [31]
11	2,3,7,8-tetrachlorodibenzo-*p*-dioxin	1746-01-6	6.80 [14]	6.67	0.13	6.66 [26]	7.44 [31]
12	1,2,3,4,7-pentachlorodibenzo-*p*-dioxin	39227-61-7	7.40 [14]	7.07	0.33	7.01 [26]	7.56 [2]
14	1,2,3,4,7,8-hexachlorodibenzo-*p*-dioxin	39227-28-6	7.79 [1]	7.56	0.23	7.63 [26]	8.21 [2]
17	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-dioxin	35822-46-9	8.20 [14]	8.17	0.03	8.00 [26]	8.85 [2]
18	octachlorodibenzo-*p*-dioxin	3268-87-9	8.60 [1]	8.64	-0.04	8.20 [14]	9.50 [2]
19	dibenzo-*p*-furan	132-64-9	3.92 [1]	3.90	0.02	3.68 [1]	4.31 [26]
20	2,8-dichlorodibenzo-*p*-furan	5409-83-6	5.30 [1]	5.01	0.29	5.00 [2]	5.67 [31]
21	1,2,7,8-tetrachlorodibenzo-*p*-furan	58802-20-3	6.23 [2]	6.21	0.02	5.96 [26]	6.41 [1]
22	2,3,7,8-tetrachlorodibenzo-*p*-furan	51207-31-9	6.10 [26]	6.25	-0.15	6.10 [26]	6.72 [31]
23	2,3,4,7,8-pentachlorodibenzo-*p*-furan	57117-31-4	6.79 [2]	6.76	0.03	6.50 [26]	7.11 [1]
24	1,2,3,8,9-pentachlorodibenzo-*p*-furan	83704-54-5	6.26 [2]	6.69	-0.43	6.26 [2]	7.22 [1]
25	1,2,3,4,7,8-hexachlorodibenzo-*p*-furan	70648-26-9	7.00 [26]	7.33	-0.33	7.00 [26]	7.92 [2]
30	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-furan	67562-39-4	7.92 [2]	7.78	0.14	7.40 [26]	8.23 [2]
31	1,2,3,4,7,8,9-heptachlorodibenzo-*p*-furan	55673-89-7	7.92 [2]	7.80	0.12	7.62 [26]	8.23 [1]
32	octachlorodibenzo-*p*-furan	39001-02-0	8.00 [26]	8.33	-0.33	7.99 [26]	8.87 [2]
33	dimethyl phthalate	131-11-3	1.47 [10]	1.44	0.03	1.47 [10]	1.90 [32]
34	diethyl phthalate	84-66-2	2.76 [23]	2.79	-0.03	2.21 [30]	3.27 [4]
35	diallyl phthalate	131-17-9	2.98 [23]	2.77	0.21	2.98 [23]	3.61 [23]
36	dipropyl phthalate	131-16-8	3.64 [23]	3.67	-0.03	3.27 [2]	5.62 [27]
37	di-*n*-butyl phthalate	84-74-2	4.50 [27]	4.52	-0.02	3.74 [10]	5.15 [33]

Table 4.9. Continued

No	Name	CAS Number	log K_{OW} Observed	log K_{OW} Pred. by Model (3)	Residual	log K_{OW} Literature Range	
						Minimum	Maximum
38	diisobutyl phthalate	84-69-5	4.48 [23]	4.52	-0.04	4.11 [2]	4.48 [23]
40	dihexyl phthalate	84-75-3	5.93 [10]	6.32	-0.39	5.65 [10]	6.82 [27]
41	di-*n*-octyl phthalate	117-84-0	8.10 [27]	8.31	-0.21	5.22 [4]	8.54 [18]
43	decylhexyl phthalate	25724-58-7	8.10 [27]	7.72	0.38	8.10 [27]	8.54 [2]
44	di (2-ethylhexyl) phthalate	117-81-7	8.05 [28]	8.02	0.03	4.20 [28]	8.39 [18]

[1]. Govers and Krop, 1998; [2]. EPIWIN; [4]. Leyder and Boulanger, 1983; [10]. Howard et al., 1985; [14]. Shiu et al., 1988; [18]. Meylan and Howard, 1995; [23]. Thomsen et al., 1999; [26]. Chen et al., 2001; [27]. Ellington and Floyd, 1996; [28]. Brooke et al., 1990; [29]. Delgado, 2002; [30]. McDuffie, 1981; [31]. Dai et al., 1999; [32]. Renberg et al., 1985; [33]. Veith et al., 1979b.

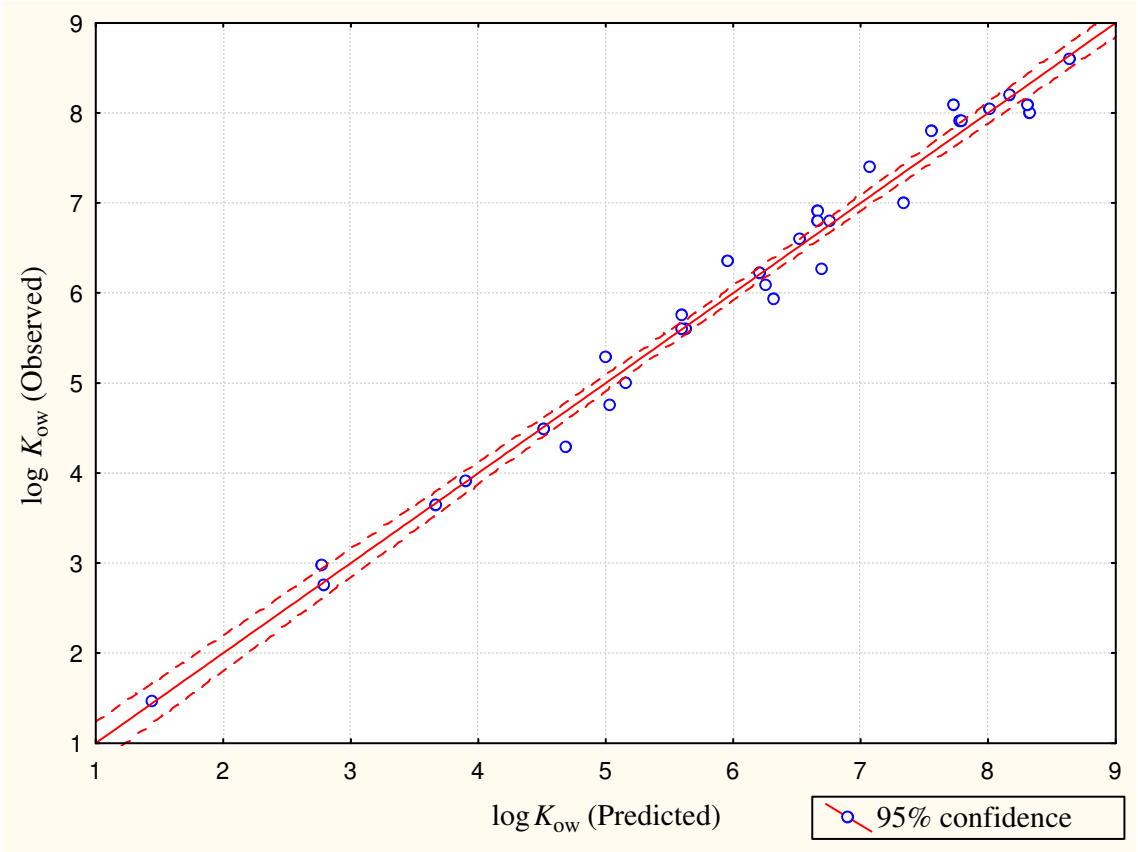


Figure 4.6. Plot of observed and predicted log K_{OW} values from model (3)

Phthalate esters exhibits an eight order of magnitude increase in octanol-water partition coefficients as alkyl chain length increases from 1 to 13 carbons. The values of log K_{OW} at 298 K ranged from 1.47 for dimethyl phthalate to 8.60 for octachlorodibenzo-*p*-

dioxin. The stepwise multiple linear regression excluded E_{LUMO} and μ from model (3). E_{LUMO} and μ did not improve the correlation coefficient of model (3). There was no collinearity between E_{LUMO} and μ. Similar to model (1), VIF values (Table 4.10) between E_{HOMO} and μ shows that these variables are collinear. All compounds were estimated within a two standard deviation range for model (3).

Table 4.10. Correlation coefficient matrix for independent variables and the variance inflation factors (VIFs) for the model not shown here and model (3)

Parameter	**Correlation Matrix**				**VIF**	
	CRI	E_{HOMO}	E_{LUMO}	μ	Model Not Shown Here	Model (3)
CRI	1.000				1.542	1.088
E_{HOMO}	-0.285	1.000			10.527	1.088
E_{LUMO}	-0.463	0.050	1.000		1.326	
μ	0.125	-0.935	0.073	1.000	9.956	

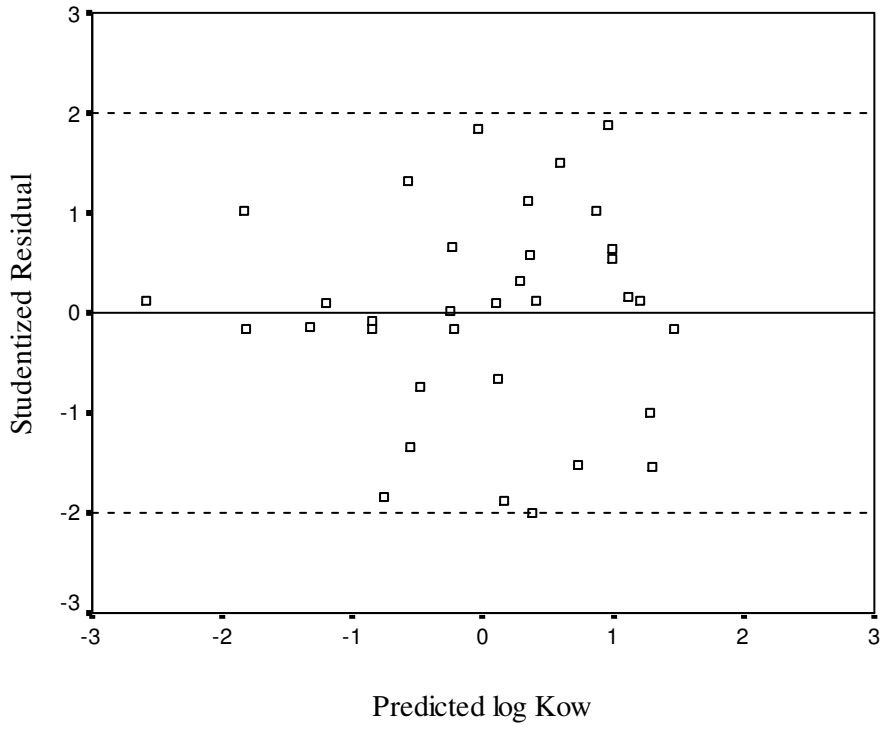


Figure 4.7. Plot of studentized residual versus calculated log K_{OW} from model (3)

The relative errors in model (3) for the individual chemicals studied were examined using the studentized residuals compared with the calculated log K_{OW} values. The studentized residuals are symmetrically distributed around zero with no specific trend for model (3) as shown in Figure 4.7.

Based on model (3), log K_{OW} for the other 22 chemicals not present in the data set were predicted, as shown in Table 4.11. The predicted results agree well with the reported literature range in case of log K_{OW}.

Table 4.11. The predicted and reported literature data of log K_{OW}

No	Name	Cas Number	log *K*ow Pred. by Model (3)	log *K*ow Literature Range	
				Minimum	Maximum
10	1,3,6,8-tetrachlorodibenzo-*p*-dioxin	33423-92-6	6.57	6.64 [1]	7.37 [31]
13	1,2,3,7,8-pentachlorodibenzo-*p*-dioxin	40321-76-4	7.13	6.64 [2]	7.56 [2]
15	1,2,3,6,7,8-hexachlorodibenzo-*p*-dioxin	57653-85-7	7.55	7.64 [26]	8.21 [2]
16	1,2,3,7,8,9-hexachlorodibenzo-*p*-dioxin	19408-74-3	7.57	7.61 [26]	8.21 [2]
26	1,2,3,4,8,9-hexachlorodibenzo-*p*-furan	92341-07-6	7.27	6.96 [26]	7.72 [1]
27	1,2,3,6,7,8-hexachlorodibenzo-*p*-furan	57117-44-9	7.24	7.20 [26]	7.92 [2]
28	1,2,3,7,8,9-hexachlorodibenzo-*p*-furan	72918-21-9	7.28	7.13 [26]	7.76 [1]
29	2,3,4,6,7,8-hexachlorodibenzo-*p*-furan	60851-34-5	7.22	6.94 [26]	7.92 [2]
39	butylbenzyl phthalate	85-68-7	5.54	3.57 [10]	5.33 [23]
42	butyl 2-ethylhexyl phthalate	85-69-8	6.14	5.64 [21]	6.50 [18]
45	diisooctyl phthalate	27554-26-3	7.80	7.73 [21]	8.39 [18]
46	diisononyl phthalate	28553-12-0	8.83	8.60 [21]	9.40 [18]
47	diisodecyl phthalate	26761-40-0	12.31	9.46 [21]	10.36 [2]
48	di-C9-11-branched alkyl phthalate	68515-49-1	11.33	10.28 [2]	10.28 [2]
49	diundecyl phthalate	3648-20-2	13.59	10.33 [21]	11.50 [18]
50	diheptyl phthalate	68515-44-6	7.24	7.41 [2]	7.41 [2]
51	dinonyl phthalate	68515-45-7	8.81	9.52 [2]	9.52 [2]
52	heptylnonyl phthalate	111381-89-6	7.55	8.39 [2]	8.39 [2]
53	heptylundecyl phthalate	111381-90-9	8.82	9.37 [2]	9.37 [2]
54	nonylundecyl phthalate	111381-91-0	12.17	10.28 [2]	10.28 [2]
55	ditridecyl phthalate	119-06-2	17.60	12.06 [21]	13.45 [2]
56	di-C11-14-branched alkyl phthalate	68515-47-9	15.32	12.25 [2]	12.25 [2]

[1]. Govers and Krop, 1998; [2]. EPIWIN; [10]. Howard et al., 1985; [18]. Meylan and Howard, 1995; [21]. Cousins and Mackay, 2000; [23]. Thomsen et al., 1999; [26]. Chen et al., 2001; [31]. Dai et al., 1999.

We performed a jackknife test with removal of less than 10% randomly selected compounds in each run, the regression then was rerun for all the other observed values.

The overall results of the deletion study (leave one/three-out method) for model (3) are summarized in Table 4.12.

Table 4.12. Summary of results of random deletion test for log K_{OW}

No. of Cases Deleted (<10% of *n*)	**No. of Regression Runs**	**av *r***	**av *S.E.***	**av *F***
1	34	0.992	0.229	965.70
2	70	0.992	0.229	937.63
3	70	0.992	0.229	906.84

The modified jackknife test validated that the developed model (3) was statistically robust. The average *r* values do not have any unduly high variation suggesting that the data set is fairly consistent and the models are not biased by any particular data point.

Multiple linear regression analysis with log K_{OW} as the dependent variable, and the two descriptors as the independent variables, for the studied compounds resulted in QSPR model (3), with a high correlation coefficient (r = 0.992). Based on the descriptors, it can be concluded that log K_{OW} of EDCs is mainly governed by the size of the molecule since the *CRI* reflects structural characteristics such as molecular size, molecular surface area and includes information concerning the complexity and branching of the molecular structure. The bigger the molecules, the greater the *CRI* values and thus the greater the log K_{OW} values. The electron-transfer interactions represented by E_{HOMO} have the second significance in the model. The E_{HOMO} describes the covalent contribution to hydrogen bonding donor. E_{HOMO} is also related to ionization potential. Increasing the E_{HOMO} values of the molecules also lead to increasing log K_{OW} values. The higher the E_{HOMO} values, the greater the tendency of EDCs to donate electrons in intermolecular interactions, the greater the intermolecular interactions between octanol molecules and EDCs, the greater the K_{OW} values. Similarly, in a study reported by Reddy and Locke (1994b) E_{HOMO} was found to be significant rather than E_{LUMO} in describing the log K_{OW} of 90 herbicides. It is not surprising that E_{LUMO} is excluded from the model, because it is widely used in combination with hydrophobicity for the determination of LC_{50}/EC_{50} values (Lu et al., 2001; Huang et al., 2003; Lu et al., 2003). E_{LUMO} does not seem particularly related to the tendency to dissolve in octanol phase.

The estimated log K_{OW} of the PCDDs at low degree of chlorination were reported to be in good agreement with the actual ones, while significant deviations between the estimated and experimental data prevailed for the PCDD/PCDFs at high degrees of chlorination (Kuramochi et al., 2002) with the AAD of 0.81. We did not observe over and/or under estimation of log K_{OW} systematically. On the basis of the good agreement between our predicted data and the experimental data and having the AAD of 0.17 which is smaller than that of the UNIFAC derived model (Kuramochi et al., 2002) by a factor of 4.76, we believe that the present K_{OW} data obtained using the *CRI*-E_{HOMO} based model are reasonable.

The average absolute deviation of 0.17 for K_{OW} shows that model (3) can be used for the prediction of K_{OW} values of EDCs not present in the data set.

RMS value of model (3) was calculated and provided in Table 4.13. Of the results reported in Table 4.13, GC-RI method for PCDDs and PCDFs in the case of K_{OW} have lower RMS values than the corresponding models developed by our group. However, it must be noted that the total number of compounds in their data set was about 65% for PCDDs and 82% for PCDFs respectively less than the data set used by our group.

Table 4.13. RMS deviations of log K_{OW} calculations from different methods

	Model (3)	**SOFA [1]**		**GC-RI [24,25]**		**UNIFAC [23]**
log K_{OW}	PCDD/PCDFs/PAEs	PCDDs	PCDFs	PCDDs	PCDFs	PAEs
RMS	0.218	0.232	0.526	0.148	0.047	0.574
Data points	34	12	6	12	6	13

[1]. Govers and Krop, 1998; [23]. Thomsen et al., 1999; [24]. Wang and Wong, 2002; [25]. Wang and Wong, 2003.

4.3. Henry's Law Constant (*H*)

The Henry's Law Constant plays an important role to quantify local equilibria at the interface of water and air. However, often experimental values of the Henry's Law Constant are missing and estimations are necessary. The Henry's Law Constant can be estimated from vapour pressure and water solubility (Lyman, 1985; Prausnitz et al., 1986).

This method will give reliable results, if accurate data on both quantities are available (Lyman, 1985).

The Henry's Law Constant for 10 chemicals was calculated from the ratio of experimental literature vapor pressure and water solubility data and listed in Table 4.14.

Table 4.14. Calculated Henry's Law Constant for 10 chemicals from the direct ratio of reported experimental literature vapour pressure and aqueous solubility

Name	Water Solubility S, (mol m^{-3})	Vapour Pressure P, (atm)	- log H (atm m^3 $mole^{-1}$)
dibenzo-*p*-dioxin	4.89×10^{-3} [17]	5.08×10^{-6} [1]	2.983
1-monochlorodibenzo-*p*-dioxin	1.91×10^{-3} [14]	7.51×10^{-7} [1]	3.406
2,3-dichlorodibenzo-*p*-dioxin	1.38×10^{-3} [1]	9.24×10^{-8} [1]	4.174
dimethyl phthalate	1.45×10^{1} [3]	7.11×10^{-6} [34]	6.310
diallyl phthalate	7.39×10^{-1} [4]	2.11×10^{-7} [35]	6.544
di-*n*-butyl phthalate	3.29×10^{-2} [5]	1.22×10^{-8} [35]	6.431
butylbenzyl phthalate	2.24×10^{-3} [6]	1.14×10^{-9} [35]	6.293
di (2-ethylhexyl) phthalate	6.91×10^{-4} [5]	5.00×10^{-10} [36]	6.140
diisodecyl phthalate	2.66×10^{-3} [10]	6.95×10^{-10} [2]	6.583
diundecyl phthalate	2.32×10^{-3} [10]	6.97×10^{-10} [35]	6.523

[1]. Govers and Krop, 1998; [2]. EPIWIN; [3]. Nielson and Bundgaard, 1989; [4]. Leyder and Boulanger, 1983; [5]. DeFoe et al., 1990; [6]. Hollifield, 1979; [10]. Howard et al., 1985; [14]. Shiu et al., 1988; [17]. Doucette and Andren, 1988; [34]. Cowen and Baynes, 1980; [35]. Sears and Darby, 1982; [36]. Stephenson and Malanowski, 1987.

The following QSPR equation without the outliers was obtained from multiple regression analysis for the Henry's Law Constant of the studied EDCs. The numbers in the parentheses are the 95% confidence intervals of that coefficient.

$$\log H = -0.219\ (\pm 0.029)\ CRI + 1.133\ (\pm 0.230)\ E_{LUMO} - 0.238\ (\pm 0.023)\ \mu - 2.645\ (\pm 0.240) \qquad \text{Model (4)}$$

$$n = 33; \quad r = 0.933; \quad S.E. = 0.374; \quad F_{(3,29)} = 64.69$$

t-values for partial correlation coefficients in model (4) are 7.534, -4.922, and 10.323 for the *CRI*, E_{LUMO}, and μ, respectively and are given in Table 4.15.

Table 4.15. *t*-values for partial correlation coefficients in model (4)

	***t*(29)**	***p*-level**
Intercept	11.038	6.70×10^{-12}
CRI	7.534	2.64×10^{-8}
E_{LUMO} (eV)	-4.922	3.15×10^{-5}
μ (debye)	10.323	3.19×10^{-11}

For model (4), plot of observed and predicted values is shown by Figure 4.8. The observed values, fitted values obtained from the developed model and the residuals for the studied chemicals are listed in Table 4.16 together with the reported literature ranges.

Table 4.16. Experimental, predicted and reported literature log *H* data (at 25 ºC)

No	Name	CAS Number	- log *H* (atm m^3 $mole^{-1}$) Observed	- log *H* (atm m^3 $mole^{-1}$) Pred. by Model (4)	Residual	- log *H* (atm m^3 $mole^{-1}$) Literature Range	
						Minimum	Maximum
1	dibenzo-*p*-dioxin	262-12-4	2.983	3.430	-0.447	3.936 [1]	4.932 [2]
2	1-monochlorodibenzo-*p*-dioxin	39227-53-7	3.406	3.836	-0.430	4.204 [1]	5.062 [2]
3	2-monochlorodibenzo-*p*-dioxin	39227-54-8	3.903 [1]	3.999	-0.096	3.840 [15]	5.062 [2]
4	2,3-dichlorodibenzo-*p*-dioxin	29446-15-9	4.174	4.402	-0.228	4.184 [15]	5.192 [2]
5	2,7-dichlorodibenzo-*p*-dioxin	33857-26-0	4.094 [1]	4.021	0.073	4.094 [1]	5.192 [2]
6	2,8-dichlorodibenzo-*p*-dioxin	38964-22-6	4.674 [1]	4.231	0.443	4.312 [2]	5.192 [2]
7	1,2,4-trichlorodibenzo-*p*-dioxin	39227-58-2	4.425 [1]	4.484	-0.059	4.382 [2]	5.322 [2]
8	1,2,3,4-tetrachlorodibenzo-*p*-dioxin	30746-58-8	4.699 [2]	4.808	-0.109	4.430 [15]	5.452 [2]
9	1,2,3,7-tetrachlorodibenzo-*p*-dioxin	67028-18-6	5.119 [2]	4.644	0.475	4.451 [2]	5.245 [1]
10	1,3,6,8-tetrachlorodibenzo-*p*-dioxin	33423-92-6	5.154 [1]	4.475	0.679	3.167 [2]	5.452 [2]
11	2,3,7,8-tetrachlorodibenzo-*p*-dioxin	1746-01-6	4.301 [2]	4.523	-0.222	4.301 [2]	5.452 [2]
12	1,2,3,4,7-pentachlorodibenzo-*p*-dioxin	39227-61-7	5.588 [2]	4.853	0.735	4.521 [2]	5.654 [1]
14	1,2,3,4,7,8-hexachlorodibenzo-*p*-dioxin	39227-28-6	5.204 [1]	4.891	0.313	4.592 [2]	5.712 [2]
17	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-dioxin	35822-46-9	5.365 [1]	5.164	0.201	3.757 [2]	5.845 [2]
18	octachlorodibenzo-*p*-dioxin	3268-87-9	5.171 [2]	5.237	-0.066	4.730 [2]	5.975 [2]
19	dibenzo-*p*-furan	132-64-9	3.863 [1]	3.922	-0.059	3.672 [2]	4.291 [2]
20	2,8-dichlorodibenzo-*p*-furan	5409-83-6	4.210 [1]	4.413	-0.203	4.200 [15]	4.551 [2]
22	2,3,7,8-tetrachlorodibenzo-*p*-furan	51207-31-9	4.775 [37]	4.865	-0.090	4.574 [1]	4.940 [1]
23	2,3,4,7,8-pentachlorodibenzo-*p*-furan	57117-31-4	5.245 [1]	5.201	0.044	4.594 [1]	5.245 [1]
25	1,2,3,4,7,8-hexachlorodibenzo-*p*-furan	70648-26-9	4.585 [1]	5.303	-0.718	4.411 [2]	5.167 [2]
27	1,2,3,6,7,8-hexachlorodibenzo-*p*-furan	57117-44-9	4.963 [1]	5.341	-0.378	4.411 [2]	5.167 [2]
32	octachlorodibenzo-*p*-furan	39001-02-0	5.724 [2]	5.699	0.025	4.883 [1]	5.724 [2]

Table 4.16. Continued

No	Name	CAS Number	- log *H* (atm m^3 $mole^{-1}$) Observed	- log *H* (atm m^3 $mole^{-1}$) Pred. by Model (4)	Residual	- log *H* (atm m^3 $mole^{-1}$) Literature Range	
						Minimum	Maximum
33	dimethyl phthalate	131-11-3	6.310	5.986	0.324	6.650 [2]	7.212 [2]
34	diethyl phthalate	84-66-2	6.215 [2]	5.766	0.449	6.215 [2]	6.951 [2]
35	diallyl phthalate	131-17-9	6.544	6.239	0.305	6.373 [21]	6.932 [2]
36	dipropyl phthalate	131-16-8	6.395 [2]	6.383	0.012	6.158 [2]	6.652 [2]
37	di-*n*-butyl phthalate	84-74-2	6.431	6.574	-0.143	5.742 [2]	6.352 [2]
39	butylbenzyl phthalate	85-68-7	6.293	6.509	-0.216	5.693 [21]	8.672 [2]
41	di-*n*-octyl phthalate	117-84-0	5.590 [2]	5.809	-0.219	3.987 [20]	5.590 [2]
44	di (2-ethylhexyl) phthalate	117-81-7	6.140	6.398	-0.258	4.408 [21]	6.569 [2]
46	diisononyl phthalate	28553-12-0	5.827 [2]	6.530	-0.703	4.038 [21]	5.827 [2]
47	diisodecyl phthalate	26761-40-0	6.583	6.390	0.193	3.670 [21]	5.955 [2]
49	diundecyl phthalate	3648-20-2	6.523	6.148	0.375	3.301 [21]	4.252 [2]

[1]. Govers and Krop, 1998; [2]. EPIWIN; [15]. Shiu and Ma, 2000; [20]. Staples et al., 1997; [21]. Cousins and Mackay, 2000; [37]. Friesen et al., 1993.

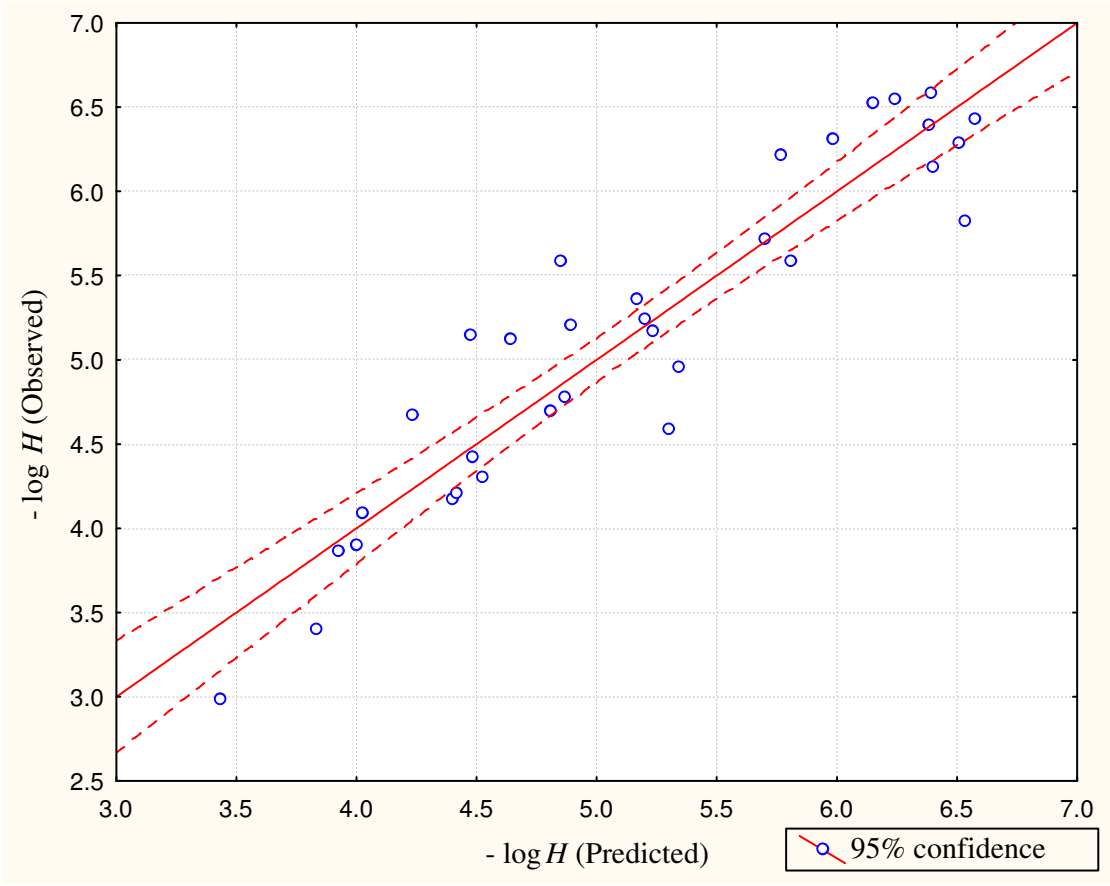


Figure 4.8. Plot of observed and predicted - log *H* values from model (4)

The values of H at 298 K ranged from 2.61×10^{-7} atm m^3 $mole^{-1}$ for diisodecyl phthalate to 1.04×10^{-3} atm m^3 $mole^{-1}$ for dibenzo-*p*-dioxin. The high correlation coefficient (r = 0.895) between E_{HOMO} and μ (Table 4.17) implies that these variables are collinear. The stepwise multiple linear regression excluded E_{HOMO} from model (4). We again examined both E_{HOMO} and μ as independent variables in regressions of log H, but μ significantly improved the correlation as similar to model (2). All compounds were estimated within a two standard deviation range for model (4).

Table 4.17. Correlation coefficient matrix for independent variables and the variance inflation factors (VIFs) for the model not shown here and model (4)

Parameter	**Correlation Matrix**				**VIF**	
	CRI	E_{HOMO}	E_{LUMO}	μ	Model Not Shown Here	Model (4)
CRI	1.000				2.089	1.003
E_{HOMO}	-0.455	1.000			10.277	
E_{LUMO}	-0.059	-0.009	1.000		1.007	1.002
μ	0.158	-0.895	0.042	1.000	8.368	1.002

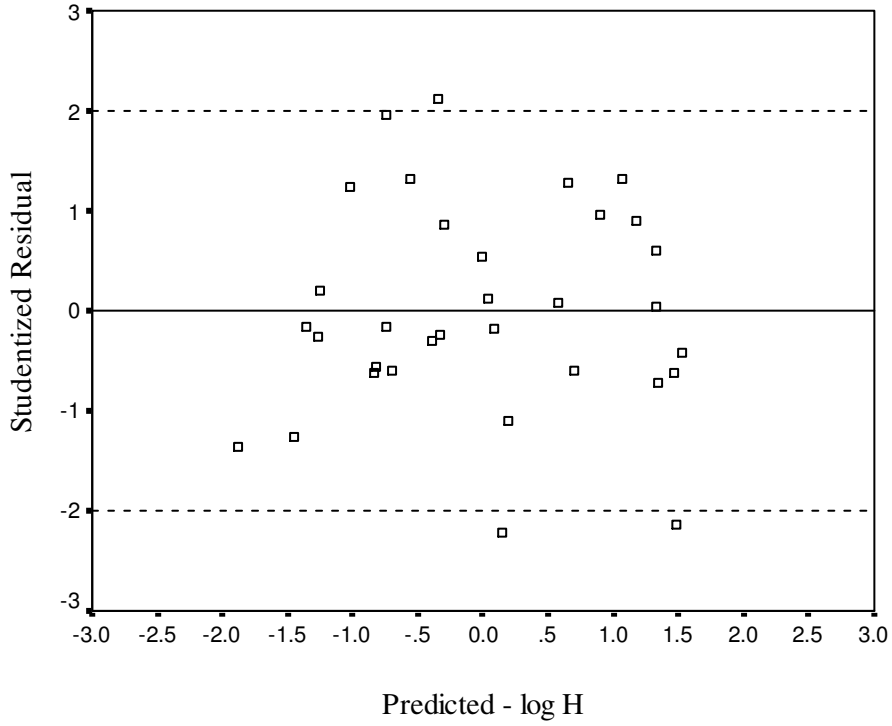


Figure 4.9. Plot of studentized residual versus calculated - log H from model (4)

The relative errors in model (4) for the individual chemicals studied were examined using the studentized residuals compared with the calculated log *H* values. As can be seen in Figure 4.9, the studentized residuals are symmetrically distributed around zero with no specific trend for model (4). The three chemicals with the highest residuals are 1,2,3,4,7-pentachlorodibenzo-*p*-dioxin, 1,2,3,4,7,8-hexachlorodibenzo-*p*-furan, and diisononyl phthalate. One possible explanation for this result could be the potential inaccuracies in the measured data itself, since the maximum difference of the multiple measured *H* values for some of the compounds used in this study reported as 1.99 log units. However, the highest residual in the *H* data set is 0.735 for 1,2,3,4,7-pentachlorodibenzo-*p*-dioxin. And also 100% of plots show good agreement within ±1.0 log unit and the response plot (Figure 4.8) confirms the model quality.

Based on model (4), log *H* for the other 23 chemicals not present in the data set were predicted and listed in Table 4.18.

Table 4.18. The predicted and reported log *H* data

No	Name	CAS Number	- log *H* (atm m^3 $mole^{-1}$) Pred. by Model (4)	- log *H* (atm m^3 $mole^{-1}$) Literature Range	
				Minimum	Maximum
13	1,2,3,7,8-pentachlorodibenzo-*p*-dioxin	40321-76-4	4.809	4.521 [2]	5.583 [2]
15	1,2,3,6,7,8-hexachlorodibenzo-*p*-dioxin	57653-85-7	4.876	4.592 [2]	5.712 [2]
16	1,2,3,7,8,9-hexachlorodibenzo-*p*-dioxin	19408-74-3	5.063	4.592 [2]	5.712 [2]
21	1,2,7,8-tetrachlorodibenzo-*p*-furan	58802-20-3	5.044	4.684 [1]	4.812 [2]
24	1,2,3,8,9-pentachlorodibenzo-*p*-furan	83704-54-5	5.298	4.943 [2]	5.044 [1]
26	1,2,3,4,8,9-hexachlorodibenzo-*p*-furan	92341-07-6	5.436	5.064 [1]	5.072 [2]
28	1,2,3,7,8,9-hexachlorodibenzo-*p*-furan	72918-21-9	5.422	5.024 [1]	5.072 [2]
29	2,3,4,6,7,8-hexachlorodibenzo-*p*-furan	60851-34-5	5.497	4.411 [2]	5.167 [2]
30	1,2,3,4,6,7,8-heptachlorodibenzo-*p*-furan	67562-39-4	5.598	4.465 [1]	5.202 [2]
31	1,2,3,4,7,8,9-heptachlorodibenzo-*p*-furan	55673-89-7	5.508	5.004 [1]	5.202 [2]
38	diisobutyl phthalate	84-69-5	6.501	5.881 [21]	6.738 [20]
40	dihexyl phthalate	84-75-3	6.198	4.357 [20]	5.752 [2]
42	butyl 2-ethylhexyl phthalate	85-69-8	6.843	5.336 [21]	6.398 [20]
43	decylhexyl phthalate	25724-58-7	7.221	4.928 [2]	5.590 [2]
45	diisooctyl phthalate	27554-26-3	7.180	4.408 [21]	4.991 [2]
48	di-C9-11-branched alkyl phthalate	68515-49-1	6.279	4.391 [2]	4.435 [2]
50	diheptyl phthalate	68515-44-6	4.957	5.174 [2]	5.292 [2]

Table 4.18. Continued

No	Name	CAS Number	- log *H* (atm m^3 $mole^{-1}$) Pred. by Model (4)	- log *H* (atm m^3 $mole^{-1}$) Literature Range	
				Minimum	Maximum
51	dinonyl phthalate	68515-45-7	6.482	4.682 [2]	4.851 [2]
52	heptylnonyl phthalate	111381-89-6	7.150	4.928 [2]	4.991 [2]
53	heptylundecyl phthalate	111381-90-9	5.789	4.682 [2]	4.693 [2]
54	nonylundecyl phthalate	111381-91-0	5.823	4.312 [2]	4.435 [2]
55	ditridecyl phthalate	119-06-2	6.905	2.565 [21]	3.697 [2]
56	di-C11-14-branched alkyl phthalate	68515-47-9	7.099	3.790 [2]	3.943 [2]

[1]. Govers and Krop, 1998; [2]. EPIWIN; [20]. Staples et al., 1997; [21]. Cousins and Mackay, 2000.

We performed a jackknife test with removal of less than 10% randomly selected compounds in each run, the regression then was rerun for all the other observed values. The overall results of the deletion study (leave one/three-out method) for model (4) are summarized in Table 4.19.

Table 4.19. Summary of results of random deletion test for log *H*

No. of Cases Deleted (<10% of *n*)	No. of Regression Runs	av *r*	av *S.E.*	av *F*
1	33	0.933	0.374	62.88
2	70	0.933	0.373	61.21
3	70	0.933	0.373	59.72

The modified jackknife test validated that the developed model (4) was statistically robust. The average *r* values do not have any unduly high variation suggesting that the data set is fairly consistent and the models are not biased by any particular data point.

Although the aqueous solubility of studied chemicals significantly decreased depending on the chlorine substitution which had increased and the carbon number of the side chain that increased the experimental data of *H* were scattered. Despite the highly linear correlations, estimated and observed values of log *H* deviated more strongly than those of log *S* and log K_{OW} from equality (Figure 4.8) and the correlation between log *H*

and the selected variables was the lowest among the three physico-chemical parameters studied. Because of that, *H* is affected by both pressure and solubility of the chemicals. *H* increased with *P* and decreased with *S*, the correlation patterns between *H* and *P* and *H* and *S* are different. Another problem concerning this parameter is almost certainly the sources of the data set.

The overall agreement between the observed and the calculated values was reasonable (AAD for log *H* was 0.28 log unit). Henry's Law Constant depends more on the electronic factors than on the *CRI*. The *t*-values for partial correlation coefficients in *H* model are 7.534, -4.922, and 10.323 for the *CRI*, E_{LUMO}, and μ, respectively. This indicated that μ is the most significant factor for *H*. Dipole moment is related to the polarity i.e., solubility of organic molecules. In general, as the dipole moment increases, the aqueous solubility also increases and *H* decreases. The positive coefficient on E_{LUMO} reflects the fact that E_{LUMO} values are negative (i.e., electron attracting ability increases interaction with the solvent) as also stated by Dearden et al. (2000). log *H* of the studied compounds are based on molecular bulkiness and molecular ability of accepting electrons. The E_{LUMO} also accounts for possible hydrogen bonding. Polar interactions have significant role in solubility of a molecule as stated above. Appearance of electronic descriptor; dipole moment of a molecule in model (4) reveals the role of electronic interactions on the solubility of molecule in water. Model (4) was used to predict the Henry's Law Constant of the 23 compounds at 25 ºC and listed together with the reported predicted values in Table 4.18. The predicted literature values of 23 compounds not present in the data set agree with the reported literature range except large molecular mass, less soluble phthalate esters. For solubilities, measurement problems are encountered below 1.0 mg L^{-1} and measurements below 1.0 µg L^{-1} are often unreliable. For vapor pressure, measurement problems are encountered below 1.0×10^{-4} Pa and measurement below 1.0×10^{-6} Pa are often unreliable (Cousins and Mackay, 2000). Thus, experimental measurements are believed to be reliable for lower molecular weight phthalates.

For octachlorodibenzo-*p*-dioxin/octachlorodibenzo-*p*-furan (O8CDD/O8CDF) the reported deviations were 1.55 and 1.90 log unit in UNIFAC derived *H* in which the data set included only PCDD/PCDFs (Kuramochi et al., 2002). However, our model has not got the disadvantage that the UNIFAC model has for O8CDD/O8CDF because the deviations for

the same compounds are 0.066 and 0.025 log unit, respectively. The log *H* value of O8CDD was not included in the study of Wang and Wong (2002) to build the mathematical relationship between the log *H* and GC-RIs of PCDDs because it's log *H* value reported by Govers and Krop (1998) was found significantly different from the reported data of Mackay et al. (1992).

RMS value of model (4) was calculated and provided in Table 4.20. Of the results reported in Table 4.20, SOFA method for PCDDs and PCDFs in the case of *H* have higher RMS values than the corresponding models developed by our group. However, GC-RI methods for PCDDs and PCDFs have lower RMS values than our model. In addition, it must be noted that the total number of compounds in their data set was about 58% for PCDDs and 82% for PCDFs respectively less than the data set used by our group.

Table 4.20. RMS deviations of log *H* calculations from different methods

	Model (4)	**SOFA [1]**		**GC-RI [24,25]**	
log *H*	PCDD/PCDFs/PAEs	PCDDs	PCDFs	PCDDs	PCDFs
RMS	0.350	0.366	0.364	0.299	0.276
Data points	33	14	6	14	6

[1]. Govers and Krop, 1998; [24]. Wang and Wong, 2002; [25]. Wang and Wong, 2003.

4.4. 48h-EC_{50} of Substituted Benzenes to Algae (*Scenedesmus obliquus*)

We have developed the following QSPR model (5) without the outliers from multiple regression analysis with the *CRI* and E_{LUMO} descriptors for the 48h-EC_{50} of the studied EDCs. The numbers in the parentheses are the 95% confidence intervals of that coefficient.

$$\log 1/EC_{50} = 0.494\ (\pm 0.072)\ CRI - 0.798\ (\pm 0.063)\ E_{LUMO} + 1.985\ (\pm 0.169) \qquad \text{Model (5)}$$

$$n = 36; \quad r = 0.926; \quad S.E. = 0.258; \quad F_{(2,33)} = 99.47$$

t-values for partial correlation coefficients in model (5) are 6.871 and -12.717 for the *CRI* and E_{LUMO}, respectively as given in Table 4.21.

Table 4.21. *t*-values for partial correlation coefficients in model (5)

	***t*(33)**	***p*-level**
Intercept	11.746	2.49×10^{-13}
CRI	6.871	7.63×10^{-8}
E_{LUMO} (eV)	-12.717	2.87×10^{-14}

For model (5), plot of observed and predicted values is shown by Figure 4.10. The observed values, fitted values obtained from the developed model and the residuals for the studied chemicals are listed in Table 4.22.

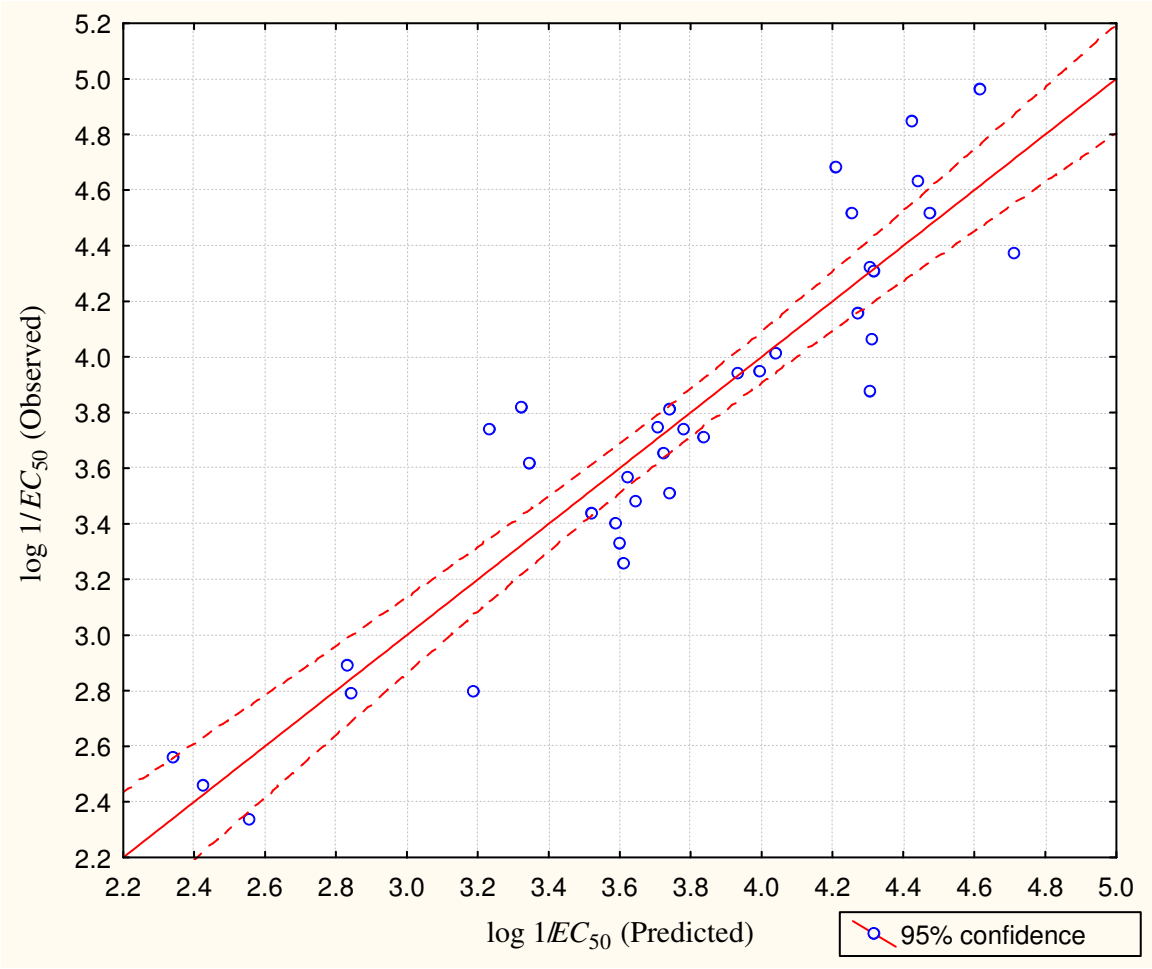


Figure 4.10. Plot of observed and predicted log $1/EC_{50}$ values from model (5)

Table 4.22. Experimental and predicted (model 5) toxicity values of algae

No	Name	Cas Number	log 1/EC_{50} (mol L^{-1}) Observed*	log 1/EC_{50} (mol L^{-1}) Pred. by Model (5)	Residual
57	4-nitrotoluene	99-99-0	3.74	3.78	-0.04
58	1,3-dinitrobenzene	99-65-0	4.85	4.42	0.43
59	1,4-dinitrobenzene	100-25-4	4.96	4.62	0.34
60	2,4-dinitrotoluene	121-14-2	4.52	4.48	0.04
61	2,6-dinitrotoluene	606-20-2	4.06	4.31	-0.25
62	nitrobenzene	98-95-3	3.26	3.61	-0.35
63	2-nitroanisole	91-23-6	3.44	3.52	-0.08
64	3-nitroanisole	555-03-3	3.71	3.84	-0.13
65	4-nitroanisole	100-17-4	3.65	3.72	-0.07
66	3-bromonitrobenzene	585-79-5	4.32	4.30	0.02
67	4-bromonitrobenzene	586-78-7	3.88	4.30	-0.42
68	2,4-dinitroaniline	97-02-9	4.68	4.21	0.47
69	2-methylaniline	95-53-4	2.34	2.55	-0.21
70	2,5-dichloroaniline	95-82-9	3.82	3.32	0.50
71	3-bromoaniline	591-19-5	2.80	3.19	-0.39
72	2,4,6-tribromoaniline	147-82-0	4.37	4.71	-0.34
73	4-nitrophenol	100-02-7	3.57	3.62	-0.05
74	2,4-dinitrophenol	51-28-5	4.16	4.27	-0.11
75	2-chloronitrobenzene	88-73-3	3.94	3.93	0.01
76	3-chloronitrobenzene	121-73-3	3.95	4.00	-0.05
77	4-chloronitrobenzene	100-00-5	4.01	4.04	-0.03
78	3,4-dichloronitrobenzene	99-54-7	4.52	4.25	0.27
79	2,5-dichloronitrobenzene	89-61-2	4.31	4.32	-0.01
80	2-nitroaniline	88-74-4	3.33	3.60	-0.27
81	3-nitroaniline	99-09-2	3.48	3.65	-0.17
82	4-nitroaniline	100-01-6	3.40	3.59	-0.19
83	aniline	62-53-3	2.56	2.34	0.22
84	2-chloroaniline	95-51-2	2.89	2.83	0.06
85	3-chloroaniline	108-42-9	2.79	2.85	-0.06
86	2,4-dichloroaniline	554-00-7	3.74	3.23	0.51
87	2-nitrophenol	88-75-5	3.51	3.74	-0.23
88	3-nitrophenol	554-84-7	3.75	3.71	0.04
89	phenol	108-95-2	2.46	2.43	0.03
90	2,4-dichlorophenol	120-83-2	3.62	3.35	0.27
91	2,4,6-trichlorophenol	88-06-2	3.81	3.74	0.07
92	pentachlorophenol	87-86-5	4.63	4.44	0.19

* Data were taken from Lu et al. (2001).

The values of log 1/EC_{50} at 20±1 ºC ranged from 2.34 mol L^{-1} for 2-methylaniline to 4.96 mol L^{-1} for 1,4-dinitrobenzene. The stepwise multiple linear regression excluded E_{HOMO} and μ from the model (5). VIF values are given in Table 4.23.

Table 4.23. Correlation coefficient matrix for independent variables and the variance inflation factors (VIFs) for the model not shown here and model (5)

Parameter	**Correlation Matrix**				**VIF**	
	CRI	E_{HOMO}	E_{LUMO}	μ	Model Not Shown Here	Model (5)
CRI	1.000				1.434	1.004
E_{HOMO}	0.223	1.000			6.407	
E_{LUMO}	0.068	0.902	1.000		7.602	1.004
μ	-0.349	-0.589	-0.649	1.000	2.119	

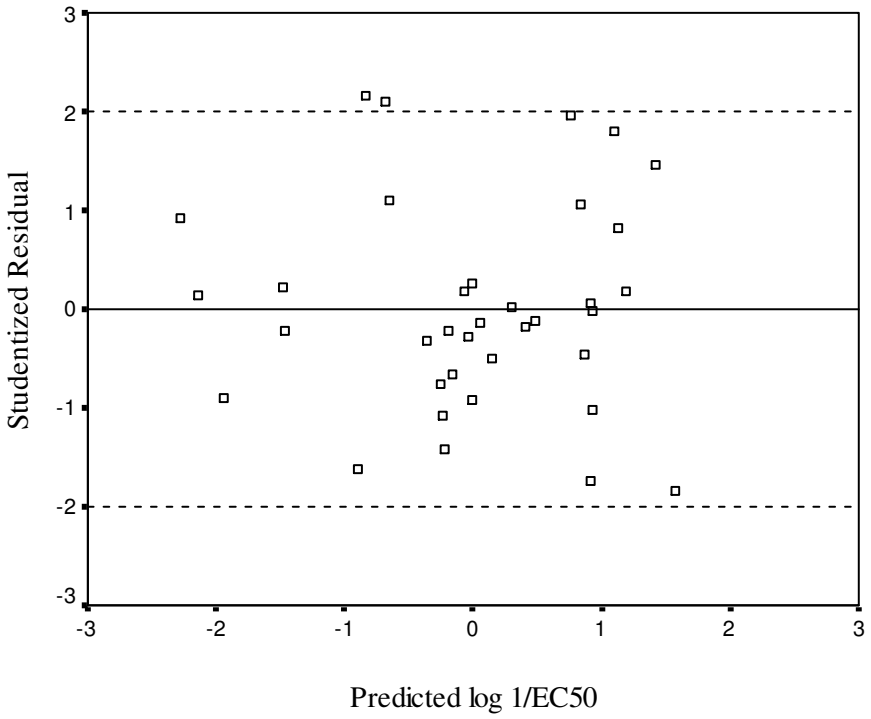


Figure 4.11. Plot of studentized residual versus calculated log $1/EC_{50}$ from model (5)

The relative errors in model (5) for the individual chemicals studied were examined using the studentized residuals compared with the calculated log $1/EC_{50}$ values. The studentized residuals are symmetrically distributed around zero with no specific trend for model (5) as shown in Figure 4.11. All compounds were estimated within a two standard deviation range for model (5). The two chemicals with the highest residuals are 2,5-dichloroaniline and 2,4-dichloroaniline. These two compounds have residuals of 0.50 and

0.51, respectively. These values for residuals are reasonable because 100% of plots show good agreement within ±1.0 log unit and the response plot (Figure 4.10) confirms the model quality. It shows a good alignment of the studied organic compounds along the optimal line.

We performed a jackknife test with removal of less than 10% randomly selected compounds in each run, the regression then was rerun for all the other observed values. The overall results of the deletion study (leave one/three-out method) for model (5) are summarized in Table 4.24.

Table 4.24. Summary of results of random deletion test for log $1/EC_{50}$

No. of Cases Deleted (<10% of *n*)	**No. of Regression Runs**	**av *r***	**av *S.E.***	**av *F***
1	36	0.926	0.258	96.92
2	70	0.926	0.257	94.18
3	70	0.927	0.257	92.42

The modified jackknife test validated that the developed model (5) was statistically robust. The average *r* values do not have any unduly high variation suggesting that the data set is fairly consistent and the models are not biased by any particular data point.

Veith and Mekenyan (1993) established a set of QSARs for aromatic chemicals by using average superdelocalizability (S^N_{av}), E_{LUMO} and log K_{OW} to explain the variation in the acute toxicity of substituted benzenes, phenols, and anilines to fish. The square of correlation coefficient r^2 is identical up to 0.81 by using either S^N_{av} and log K_{OW} or E_{LUMO} and log K_{OW}. Both S^N_{av} and E_{LUMO} show the tendency of chemicals to undergo orbital-controlled reactions, and r^2 is 0.94 between them. The QSAR for acute toxicity using these molecular descriptors defined a toxicity plane, which included several modes of toxic action. Type (I) narcotics are chemicals located in the region of low reactivity where toxicity varies with hydrophobicity alone. Type (II) narcotics are more toxic than Type (I) narcotics at similar values of log K_{OW}, and the increase can be explained by stronger electronic interactions with cellular soft nucleophiles. Dearden et al. (1995) analyzed the data on acute toxicity of nitrobenzenes to *Tetrahymena pyriformis* and concluded that the

toxicity of 2- and 3- substituted nitrobenzenes is probably controlled largely by hydrophobic and electronic factors. In this study, the *CRI* represents hydrophobicity since the bigger the molecules, the greater the *CRI* and thus the greater the hydrophobicity. Hydrophobicity is dependent on the *CRI* as stated in section 4.2. The *CRI* descriptor is used in place of log K_{OW} because the *CRI* represents many molecular properties besides hydrophobicity while log K_{OW} only gives us information about hydrophobicity. In addition, E_{LUMO} characterizes the toxicity of substituted benzenes and is generally used related to the toxicity parameter. On the basis of the *t*-values for the *CRI* and E_{LUMO}, it can be concluded that E_{LUMO} controls the toxicity mostly. Yuan et al. (1997) investigated the toxicity of nitrobenzenes to river bacteria and *Photobacterium phosphorcum* and obtained QSARs with E_{LUMO} and log K_{OW} and concluded that the toxicity of substituted nitrobenzenes is controlled mainly by electronic factors.

The average residual of substituted benzenes in the data set of Lu et al. (2001) and our data set for log $1/EC_{50}$ are 0.27 and 0.19 log unit, respectively. RMS values are 0.325 and 0.247 respectively. Both AAD and RMS values of the study of Lu et al. (2001) and our study show that our model is superior to the model of Lu et al. (2001) in respect to the predictive capability for log $1/EC_{50}$.

The higher the *CRI*, the stronger the hydrophobicity, the easier the compound is bioconcentrated in an organism; E_{LUMO} is an electrophilicity parameter, and it appears as directly proportional to the electronic affinity of the compound. The lower the E_{LUMO} values, the stronger the electrophilicity (Wang and Li, 1981). The developed QSPR model (model 5) implies that the toxicity of substituted benzenes to the algae is related chiefly to their ability to penetrate to the cell by its cell membrane and to the electronic interactions of the chemicals with the active site of action through a variety of electronic processes as stated by Lu et al. (2001).

4.5. Bioconcentration Factors (*BCF*) of Substituted Benzenes in Fish

The experimental *BCF* values of only 18 chemicals present in the second data set were available in the literature. Therefore we first carried out the forward stepwise multiple linear regression with these chemicals, using the *CRI* and the three quantum chemical descriptors. However, the stepwise multiple linear regression excluded E_{HOMO}, E_{LUMO} and μ from the regression equation. High VIF values (Table 4.25) between *CRI*-E_{HOMO}-E_{LUMO}-μ, *CRI*-E_{HOMO}-E_{LUMO}, E_{HOMO}-E_{LUMO}-μ, E_{HOMO}-E_{LUMO}, E_{LUMO}-μ, *CRI*-E_{LUMO}-μ, *CRI*-E_{HOMO}-μ, and E_{HOMO}-μ all indicate that these four variables are collinear. Therefore, their simultaneous presence in the regression equation violates the basic rule of the MLR method. Because of this multicollinearity between descriptors, we reexamined the regression by using *CRI*-E_{HOMO}, *CRI*-E_{LUMO}, and *CRI*-μ as independent variables, separately. In all three cases, E_{HOMO}, E_{LUMO}, and μ were excluded from the regression equation in each run since neither of them significantly improved the correlation.

Table 4.25. The variance inflation factors (VIFs) of models for *BCF* in fish not shown here and in model (6)

Descriptors Used	Meaningful Descriptors	*r*	*S.E.*	VIF for Descriptors, respectively
CRI-E_{HOMO}-E_{LUMO}-μ	*CRI*-E_{HOMO}	0.963	0.207	4.130 , 6.251 , 20.707 , 10.826
CRI-E_{HOMO}-E_{LUMO}	*CRI*-E_{HOMO}-E_{LUMO}	0.953	0.225	1.153 , 4.750 , 4.820
CRI-E_{HOMO}	*CRI*-E_{HOMO}	0.926	0.272	1.003 , 1.003
CRI	*CRI*	0.850	0.367	1.000
E_{HOMO}-E_{LUMO}-μ	E_{LUMO}-μ	0.896	0.330	4.193 , 6.247 , 3.022
E_{HOMO}-E_{LUMO}		0.337	0.677	4.193 , 4.193
E_{HOMO}		0.322	0.659	1.000
E_{LUMO}-μ	E_{LUMO}-μ	0.894	0.322	3.021 , 3.021
E_{LUMO}		0.330	0.657	1.000
μ		0.208	0.681	1.000
CRI-E_{LUMO}-μ	*CRI*-E_{LUMO}-μ	0.929	0.275	2.770 , 7.363 , 8.226
CRI-E_{HOMO}-μ	*CRI*-E_{HOMO}-μ	0.963	0.200	1.246 , 2.223 , 2.520
CRI-E_{LUMO}	*CRI*	0.878	0.344	1.017 , 1.017
CRI-μ	*CRI*	0.855	0.373	1.137 , 1.137
E_{HOMO}-μ	E_{HOMO}-μ	0.701	0.513	2.028 , 2.028

The *CRI* itself was employed in the modeling of experimental log *BCF* values ranging from 0.41 for aniline to 2.99 for pentachlorophenol and resulted in the following QSPR

model (6) having a correlation coefficient of 0.850 and standard error of the estimate of 0.366. For model (6), the observed values, fitted values, and the residuals for 18 chemicals are listed in Table 4.27 and are plotted in Figure 4.12.

log *BCF* = 0.983 (±0.152) *CRI* - 0.412 (±0.324) Model (6)

$n = 18$; $r = 0.850$; $S.E. = 0.366$; $F_{(1,16)} = 41.73$

t-value for partial correlation coefficient in model (6) is 6.460 for the *CRI*.

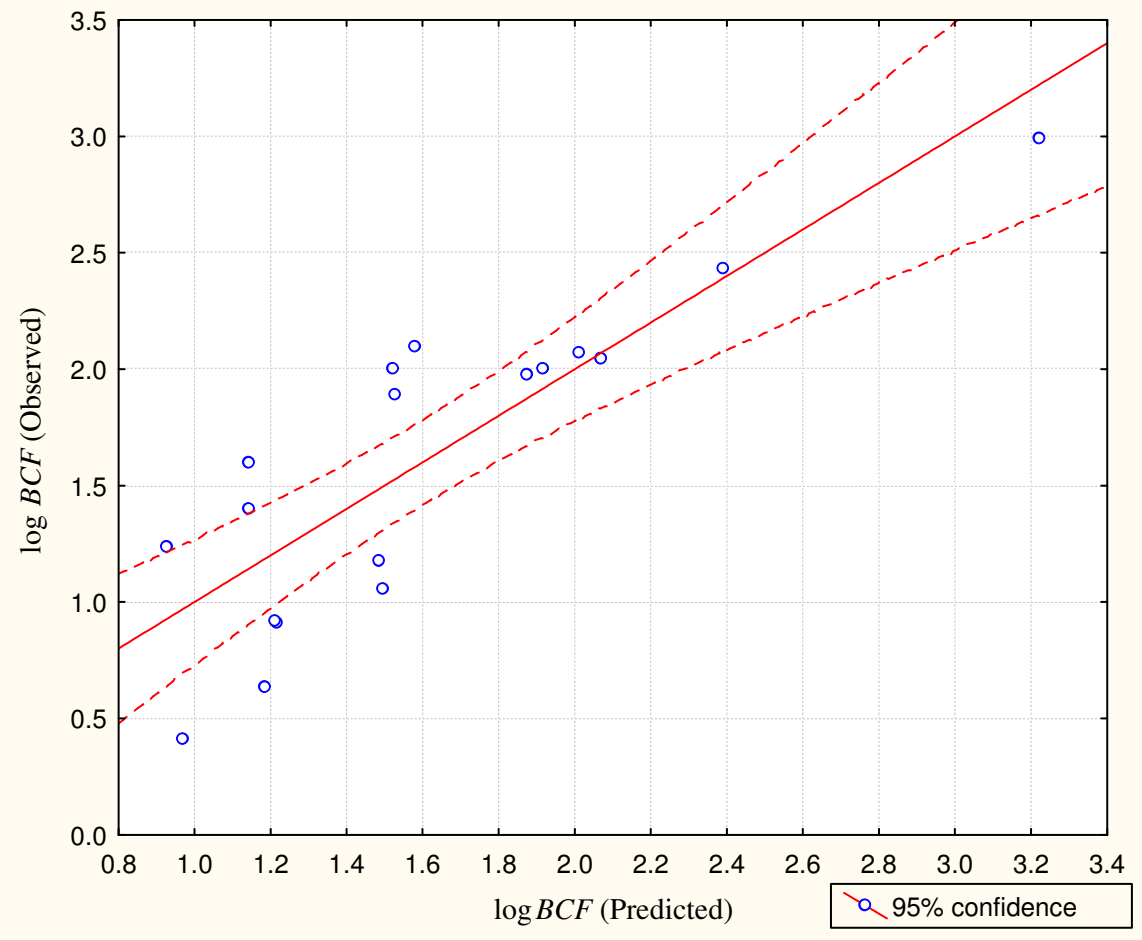


Figure 4.12. Plot of observed and predicted log *BCF* values from model (6)

The relative errors in model (6) for the individual chemicals studied were examined using the studentized residuals compared with the calculated log *BCF* values (Figure 4.13).

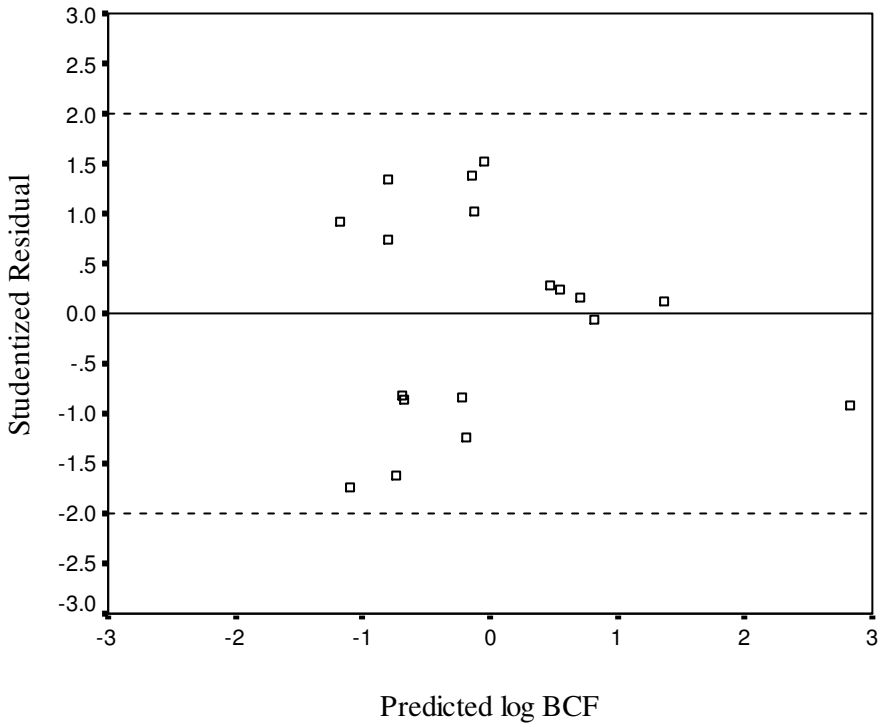


Figure 4.13. Plot of studentized residual versus calculated log *BCF* from model (6)

Although all compounds were estimated within a two standard deviation range (Figure 4.13), the *t*-value for the intercept is not significant ($p>0.05$) which is not shown here. Therefore, we looked for another QSPR model to predict the *BCF* values of the selected benzene derivatives. The predictive capability of the *CRI* was shown previously by Saçan et al. (2003) in which they developed a QSPR model by the application of the *CRI* and quantum chemical descriptor E_{HOMO} to predict the fish bioconcentration factor of 122 non-ionic organic compounds. Their data set included halogenated organic compounds such as polychlorinated biphenyls, polybrominated biphenyls, chlorinated aliphatic hydrocarbons, polychlorinated benzenes, polybrominated benzenes, polychlorinated anilines, polychlorinated nitro benzenes and phenols, and alkyl benzenes and phenols. Most of these compounds have endocrine disrupting properties similar to the those used in this study such as substituted benzenes, phthalates, polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzo-*p*-furans. They found that the *CRI* was the most important parameter for *BCF* prediction and the addition of quantum chemical descriptors made only a slight improvement in the predictive capability of their QSPR model. Therefore, it is of interest to use their model in the prediction of the *BCF* values of benzene derivatives.

$\log BCF = 4.623 + 1.045\ CRI + 0.546\ E_{HOMO}$ Model (7) (Saçan et al. (2003))

$$n = 122;\ r = 0.921;\ S.E. = 0.599;\ F_{(2,\ 119)} = 332.50$$

Considering model (6) and model (7), the higher the *CRI* was, the higher the *BCF* was. The Characteristic Root Index reflected the complexity and branching of the molecular structure. The *CRI* which comprises all possible orders of path-type molecular connectivity indices encodes global molecular properties such as size, volume, and surface area emphasizing the strong dependence of *BCF* of non-ionic organic compounds on the compound size. Local structural properties and possibly long-range interactions described by path-type molecular indices may also be encoded in the *CRI* as stated in previous sections.

Some differences were apparent for the two QSPRs. The AAD and RMS values of substituted benzenes in model (6) and model (7) developed by Saçan et al. (2003) are 0.29, 0.49, and 0.346, 0.592, log unit respectively. Although the AAD and RMS values of model (6) are less than that of model (7), the latter model has the advantage of containing more compounds in the data set. On the other hand, use of fewer descriptors has important advantages when constructing regression equations (Randic, 2001). In this respect, the *CRI* based model (model 6) is superior to the models reported in the literature in terms of number of parameters. Of the results reported in Table 4.26, all models except model 2 and 4 have higher errors than the *CRI* based model; however, it must be noted that the total number of descriptors in their data set was higher than the data set used by our group. Model numbers 1, 2, and 4 have lower errors than the *CRI*-E_{HOMO} based model while the total number of compounds in their data set was 34%, 58%, and 78% less than the data set used by Saçan et al. (2003), respectively. Since generally a model having compounds less than 30 in its data set is not considered to be very reliable statistically, both model (6) and (7) were used to estimate the *BCF* values of the rest of the benzene derivatives and the predicted values are presented in Table 4.27 together with the reported literature log *BCF* values based on log K_{OW}.

Table 4.26. QSPR bioconcentration factor prediction models

Model No	Model Type	No. of Parameters	*n*	r^2	*S.E.*	*F*	Investigators
1	MLR	5	9/44[b]	0.792	0.432	24	Fatemi et al. (2003)
2	MLR	4	27	0.931	0.137	103	Wei et al. (2001)
3	NLR[a]	13	239	0.810	0.615	NR[c]	Lu et al. (2000)
4	MLR	5	80	0.907	0.364	NR	Lu et al. (1999)
5	MLR	2	122	0.851	0.599	332.50	Saçan et al. (2003)
6	LR[d]	1	18	0.723	0.366	41.73	current study

[a] NLR = nonlinear regression. [b] Notation indicates: $n_{\text{prediction set}}/n_{\text{training set}}$. [c] NR = not reported. [d] LR = linear regression

The physical significance of the developed model (6) and model (7) could be explainable. To be absorbed in a biological system, a chemical must penetrate a sequence of hydrophobic and hydrophilic barriers, and therefore the bioconcentration process is controlled by polar and nonpolar interactions among water and fish (Lu et al., 2000).

All of the estimated log *BCF* values of aniline compounds except 2,4-dichloroaniline were found to be higher than the experimental values (Table 4.27). For 2,4-dichloroaniline, observed and predicted log *BCF* values are 1.98 and 1.88, respectively. Anilines either may not share a common mode of action which is based on their reactivity and/or polar interactions, which are utilized to provide an attractive force between anilines and water, rather than fish tissue. In addition, a significant body of evidence now supports the role of the metabolism in producing *BCF* values that are lower than expected (Meylan et al., 1999). Examples include chlorinated anilines (De Wolf et al., 1992) and certain organophosphates (De Bruijn et al., 1993). Thus, these compounds cannot reach their maximum in bioaccumulation. On the other hand, the predicted *BCF* values of all halogenated nitro benzenes were found to be lower than the experimental ones except 2,5-dichloronitrobenzene. Only for 2,5-dichloronitrobenzene, the predicted *BCF* value is higher than the experimental one. It is likely that part of the estimation error might be caused by the observed data themselves. Similar to the results obtained in this study, Saçan et al. (2003) found that most of the estimated log *BCF* values of aniline compounds are higher than the experimental values while the predicted *BCF* values of halogenated nitro benzenes are lower than the experimental ones.

The selection of test species also induces a source of error in the estimation of *BCF*s. In aquatic toxicology, it is generally accepted that lipid content of an animal is an important determinant of bioconcentration (Barron, 1990). The *BCF* data was not corrected for fish lipid content because these data are not available in most publications. Normalizing data to lipid content can eliminate some interspecies variation in *BCF* values. Error in the prediction model may also be generated from other sources. Uncertainty associated with a given measured *BCF* may arise from exposure concentration, test conditions, duration of the experiment, and the determination of the concentration in water and fish. It is also likely that the one-descriptor or two descriptors selected for the modeling may still not be sufficient for a full explanation of all of the structural features of these chemicals. Nevertheless, these are the common problems of all the other estimation methods based on topological and other descriptors which uses experimental values of *BCF*.

Table 4.27. *BCF* values of substituted benzenes in fish taken from EPIWIN software and predicted by model (6) and (7)

No	Name	CAS Number	log *BCF* Observed[§]	log *BCF* Pred. by Model (6)	Residual from Model (6)	log *BCF* Pred. by Model (7)	Residual from Model (7)	log *BCF* (EPIWIN) Pred. from log K_{OW}
57	4-nitrotoluene	99-99-0		1.39		0.83		1.13
58	1,3-dinitrobenzene	99-65-0		1.33		0.21		0.45
59	1,4-dinitrobenzene	100-25-4		1.26		0.23		0.42
60	2,4-dinitrotoluene	121-14-2		1.64		0.69		0.83
61	2,6-dinitrotoluene	606-20-2		1.60		0.71		0.92
62	nitrobenzene	98-95-3		1.03		0.37		0.73
63	2-nitroanisole	91-23-6		1.47		1.18		0.63
64	3-nitroanisole	555-03-3		1.46		1.21		0.96
65	4-nitroanisole	100-17-4		1.41		1.06		0.86
66	3-bromonitrobenzene	585-79-5		2.06		1.51		1.33
67	4-bromonitrobenzene	586-78-7		2.00		1.34		1.26
68	2,4-dinitroaniline	97-02-9		1.43		1.19		0.71
69	2-methylaniline	95-53-4		1.37		1.85		0.32
70	2,5-dichloroaniline	95-82-9		1.98		2.37		1.42
71	3-bromoaniline	591-19-5		1.97		2.33		0.92
72	2,4,6-tribromoaniline	147-82-0		3.89		4.28		2.18
73	4-nitrophenol	100-02-7		1.14		0.72		0.77
74	2,4-dinitrophenol	51-28-5		1.37		0.53		0.59
75	2-chloronitrobenzene	88-73-3	2.10	1.58	0.52	1.28	0.82	1.03
76	3-chloronitrobenzene	121-73-3	1.89	1.53	0.36	1.19	0.70	1.19
77	4-chloronitrobenzene	100-00-5	2.00	1.52	0.48	1.10	0.90	1.14

Table 4.27. Continued

No	Name	CAS Number	log *BCF* Observed[§]	log *BCF* Pred. by Model (6)	Residual from Model (6)	log *BCF* Pred. by Model (7)	Residual from Model (7)	log *BCF* (EPIWIN) Pred. from log K_{OW}
78	3,4-dichloronitrobenzene	99-54-7	2.07	2.01	0.06	1.84	0.23	1.70
79	2,5-dichloronitrobenzene	89-61-2	2.05	2.07	-0.02	1.92	0.13	1.68
80	2-nitroaniline	88-74-4	0.91	1.22	-0.31	1.35	-0.44	0.73
81	3-nitroaniline	99-09-2	0.92	1.21	-0.29	1.28	-0.36	0.36
82	4-nitroaniline	100-01-6	0.64	1.18	-0.54	1.17	-0.53	0.37
83	aniline	62-53-3	0.41	0.97	-0.56	1.39	-0.98	0.50
84	2-chloroaniline	95-51-2	1.18	1.48	-0.30	1.91	-0.73	0.76
85	3-chloroaniline	108-42-9	1.06	1.49	-0.43	1.87	-0.81	0.75
86	2,4-dichloroaniline	554-00-7	1.98	1.88	0.10	2.33	-0.35	1.44
87	2-nitrophenol	88-75-5	1.60	1.14	0.46	0.87	0.73	0.68
88	3-nitrophenol	554-84-7	1.40	1.14	0.26	0.84	0.56	0.84
89	phenol	108-95-2	1.24	0.92	0.32	1.04	0.20	0.42
90	2,4-dichlorophenol	120-83-2	2.00	1.92	0.08	2.13	-0.13	1.26
91	2,4,6-trichlorophenol	88-06-2	2.43	2.39	0.04	2.62	-0.19	1.74
92	pentachlorophenol	87-86-5	2.99	3.22	-0.23	3.50	-0.51	2.84

§ Data were taken from Lu et al. (2000) and Devillers et al. (1996).

As opposite to the model (6) and (7), all of the EPIWIN estimated log *BCF* values of aniline compounds are lower than the experimental values. On the other hand, for all halogenated nitro benzenes the predicted *BCF* values from model (6), model (7) and EPIWIN were found to be lower than the experimental ones. For nitrotoluene, nitrobenzene, nitroanisole, bromonitrobenzene, nitrophenol, and chloronitrobenzene groups, the predicted log *BCF* values by model (7) are lower than the those predicted by model (6). However, for nitroaniline, chloroaniline, chlorophenol, and bromoaniline groups, the predicted log *BCF* values by model (7) are higher than those predicted by model (6). For some compounds, the predicted values by model (6) are much closer to the EPIWIN prediction, whereas the predicted log *BCF* data by model (7) of some other compounds are much closer to the those values predicted by EPIWIN software. As a result, model (6) and model (7) can be used for *BCF* prediction but we preferred model (7) developed by Saçan et al. (2003) to predict the *BCF* of organic compounds since their data set is more diverse and has much more compounds than our data set, which might be considered a more reliable QSPR prediction model.

5. CONCLUSIONS

The physico-chemical and biological properties of PCDD/PCDFs, phthalate esters and benzene derivatives are a direct function of their size, topology and electronic distribution. The use of multivariate regressions has great importance in correlating physical-chemical and biological properties with molecular descriptors. Molecular descriptors used in this study included topology based characteristic root index (*CRI*) and three semi-empirical molecular descriptors, namely – energies of the highest occupied and the lowest unoccupied molecular orbital (E_{HOMO} and E_{LUMO}), and dipole moment (μ).

S, K_{OW}, *H*, 48h-EC_{50} and *BCF* of chemicals having endocrine disrupting properties in common can be calculated with a good precision from log linear equations including the *CRI* and the semi-empirical quantum chemical descriptors. All the models obtained showed to have reasonably high correlation coefficients and in average small leave-one/three out cross-validated residuals. Of the five properties, log K_{OW} has the highest correlation coefficient, which is directly related to the reliability of the experimental data.

The best fit equation found by "forward multiple linear regression" indicated that the topology based *CRI* was the most important parameter for the modeling of solubility, *n*-octanol/water partition coefficient and bioconcentration factor. For the modeling of *n*-octanol/water partition coefficient a two-parameter equation included the *CRI* and E_{HOMO} with a correlation coefficient of $r = 0.992$, while for the modeling of solubility and Henry's Law Constant a three-parameter equation included the *CRI*, E_{LUMO} and μ with a correlation coefficient of $r = 0.986$ and $r = 0.933$, respectively. E_{HOMO} and μ didn't appear in the same model because of the collinearity. The *CRI* and E_{LUMO} descriptors were used for modeling of 48h-EC_{50} with a correlation coefficient of $r = 0.926$ while only the *CRI* descriptor was used for the prediction of *BCF* with a correlation coefficient of $r = 0.850$. The results of modified jackknife tests indicated that the five models were statistically robust.

Analyzing the descriptors we see that the *CRI* is one of the fundamental variables. This indicates that the parameters that represent the size of the molecule, branching and global molecular properties, such as size, volume and surface area are important in the

prediction of pyhsico-chemical and biological properties. The average absolute deviations are 0.27, 0.17, 0.28, 0.19, and 0.29 log units for *S*, K_{OW}, *H*, 48h-EC_{50}, and *BCF* respectively, indicating that the developed methods could represent the *S*, K_{OW}, *H*, 48h-EC_{50}, and *BCF* of the selected endocrine disrupting chemicals. Depending on the predictive power of the *CRI* based models which were built in this study, physico-chemical and biological properties of some other chemicals - not being used in the data set and having environmental importance - can be predicted through these models where there are no experimental measurements. These compounds might include polychlorinated biphenyls, polychlorinated benzenes, etc., which have similar structural characteristics with the modeled compounds. The physico-chemical and biological properties estimated in this study determine the distribution of PCDD/PCDFs and phthalate esters in the environment, therefore they can be used in modeling the environmental fate of these compounds.

The results suggest that a small number of chemically meaningful descriptors will provide the most predictive QSPR for the three physico-chemical and two biological properties. It should be stated that this group of chemicals is not a congeneric series in the usual sense in which that term used in QSAR/QSPRs. The descriptors used in the QSPR models in this study can be calculated easily and rapidly and are error free.

On the other hand, our study highlights the urgent need for more reliable measurements of water solubility, Henry's Law Constant and *n*-octanol/water partition coefficient for phthalate esters, and *BCF* and EC_{50} values of endocrine disrupting compounds. Despite discrepancies for phthalates with longer alkyl chains, it appears reasonable to calculate solubility and Henry's Law Constant for compounds where no experimental data have been reported to obtain a rough estimate of these parameters. Only with better experimental data it will be posssible to derive and evaluate more reliable predictive methods for these important fate-controlling parameters of EDCs.

REFERENCES

Adams, W.J., Blaine, K.M., 1986. A water solubility determination of 2,3,7,8,TCDD. Chemosphere, 15, 1397-1400.

Api, A.M., 2001. Toxicological profile of diethyl phthalate: A vehicle for fragnance and cosmetic ingredients. Food and Chemical Toxicology, 39, 97-108.

ATSDR, 1996. Agency for Toxic Substances and Disease Registry, Toxicological Profile for Benzene (Update). U.S. Department of Health and Human Services, Public Health Service. Atlanta, GA.

Bahnick, D.A., Doucette, W.J., 1988. Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals. Chemosphere, 17, 1703-1715.

Bamford, H.A., Poster, D.L., Baker, J.E., 1999. Temperature dependence of Henry's Law Constants of thirteen polycyclic aromatic hydrocarbons between 4˚C and 31˚C. Environmental Toxicology and Chemistry, 18, 1905-1912.

Barron, M.G., 1990. Bioconcentration. Environmental Science and Technology, 24, 1612-1618.

Basak, S.C., Grunwald, G.D., 1995. Predicting mutagenicity of chemicals using topological and quantum chemical parameters: A similarity based study. Chemosphere, 31, 2529-2546.

Basak, S.C., Gute, B.D., Grunwald, G.D., 1996. A comparative study of topological and geometrical parameters in estimating normal boiling point and octanol/water partition coefficient. Journal of Chemical Information and Computer Sciences, 36, 1054-1060.

Basak, S.C., Gute, B.D., Grunwald, G.D., 1997. Use of topostructural, topochemical, and geometric parameters in the prediction of vapor pressure: A hierarchical QSAR approach. Journal of Chemical Information and Computer Sciences, 37, 651-655.

Basak, S.C., Gute, B.D., Grunwald, G.D., 1999. Assessment of the mutagenicity of chemicals from theoretical structural parameters: A hierarchical approach. SAR and QSAR in Environmental Research, 10, 117-129.

Basak, S.C., Grunwald, G.D., Gute, B.D., Balasubramanian, K., Opitz, D., 2000. Use of statistical and neural net approaches in predicting toxicity of chemicals. Journal of Chemical Information and Computer Sciences, 40, 885-890.

Basak, S.C., Mills, D., 2001. Quantitative structure-property relationships (QSPRs) for the estimation of vapor pressure: A hierarchical approach using mathematical structural descriptors. Journal of Chemical Information and Computer Sciences, 41, 692-701.

Beck, B., Breindl, A., Clark, T., 2000. QM/NN QSPR models with error estimation: vapor pressure and logP. Journal of Chemical Information and Computer Sciences, 40, 1046-1051.

Bernardinelli, L., Demarco, R., Tinelli, C., 1987. Cancer mortality in an Italian rubber factory. British Journal of Industrial Medicine, 44, 187-191.

Bois, F.Y., Smith, M.T., Spear, R.C., 1991. Mechanisms of benzene carcinogenesis: application of a physiological model of benzene pharmacokinetics and metabolism. Toxicology Letters, 56, 283-298.

Brief, R.S., Lynch, J., Bernath, T., Scala, R.A., 1980. Benzene in the workplace. American Industrial Hygiene Association Journal, 41, 616-623.

Brooke, D., Nielsen, I., Bruijn, 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.

Bruggeman, W.A., Van der Steen, J., Hutzinger, O., 1982. Relationship with hydrophobicity as measured by aqueous solubility and octanol-water partition coefficient. Journal of Chromatography, 238, 335-346.

Chalk, A.J., Beck, B., Clark, T., 2001. A temperature-dependent quantum mechanical/ neural net model for vapor pressure. Journal of Chemical Information and Computer Sciences, 41, 1053-1059.

Chambers, P.L, Chambers, C.M., Kennedy, A.C., 1996a. Validation with in vitro test systems of the quantitative structure activity relationships (QSAR) used for the prediction of toxicology of new fire retardants. Polymer Degradation and Stability, 54, 323-328.

Chambers, C.C., Hawkins, G.D., Cramer, C.J., Truhlar, D.G., 1996b. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. Journal of Physical Chemistry, 100, 16385.

Chen, J., Quan, X., Yazhi, Z., Yan, Y., Yang, F., 2001. Quantitative structure-property relationship studies on n-octanol/water partitioning coefficients of PCDD/Fs. Chemosphere, 44, 1369-1374.

Chen, J.W., Xue, X.Y., Schramm, K.W., Quan, X., Yang, F.L., Kettrup, A., 2002. Quantitative structure-property relationships for octanol-air partition coefficients of polychlorinated biphenyls. Chemosphere, 48, 535-544.

Chiou, C.T., Freed, V.H., Schmedding, D.W., Kohnert, R.L., 1977. Partition coefficients and bioaccumulation of selected organic chemicals. Environmental Science and Technology, 11, 475-478.

Cousins, I., Mackay, D., 2000. Correlating the physical-chemical properties of phthalate esters using the “three solubility” approach. Chemosphere, 41, 1389-1399.

Cowen, W.F., Baynes, R.K., 1980. Estimated application of gas chromatographic headspace analysis to priority pollutants. Journal of Environmental Science and Health, A15(5), 413-427.

Dai, J., Sun, C., Han, S., Wang, L., 1999. QSAR for polychlorinated organic compounds (PCOCs). I. Prediction of partition properties for PCOCs using quantum chemical parameters. Bulletin of Environmental Contamination and Toxicology, 62, 530-538.

Davies, J.E., Levine, R.S., 1986. Human health effects of benzene. In: Benzene in Florida Groundwater: An Assessment of the Significance to Human Health. Florida Petroleum Council, Tallahassee, FL., 16-35.

Dearden, J.C., Cronin, M.T.D., Schultz, T.M., 1995. QSAR study of the toxicity of nitrobenzenes to *Tetrehymena pyrifirmis*. Quantitative Structure-Activity Relationships, 14, 427-432.

Dearden, J.C., Ahmed, S.A., Cronin, M.T.D., Sharra, J.A., 2000. QSPR prediction of Henry's Law Constant: Improved correlation with new parameters. In Gundertofte, K., Jørgensen, F.S., (Eds.), Molecular Modeling and Prediction of Bioactivity, Plenum, 273-274, New York, NY, USA.

Dearden, J.C., Schüürmann, G., 2003. Quantitative structure-property relationships for predicting Henry's Law Constant from molecular structure. Environmental Toxicology and Chemistry, 22, 1755-1770.

De Bruijn, J., Seinen, W., Hermens, J., 1993. Biotransformation of organic compounds by Rainbow Trout (Oncorhynchus mykiss) liver in relation to bioconcentration. Environmental Toxicology and Chemistry, 12, 1041-1050.

DeFoe, D.L., Holcombe, G.W., Hammermeister, D.E., 1990. Solubility and toxicity of eight phthalate esters to four aquatic organisms. Environmental Toxicology and Chemistry, 9, 623-636.

Delgado, E.J., 2002. Predicting aqueous solubility of chlorinated hydrocarbons from molecular structure. Fluid Phase Equilibria, 199, 101-107.

De Maagd, P.G.J., TenHulscher, D.T.E.M., VandenHeuvel, H., Opperhuizen, A., Sijm, D.T.H.M., 1998. Physicochemical properties of polycyclic aromatic hydrocarbons: aqueous solubilities, *n*-octanol/water partition coefficients, and Henry's Law Constants. Environmental Toxicology and Chemistry, 17, 251-257.

Devillers, J., Bintein, S., Domine, D., 1996. Comparison of BCF models based on log P. Chemosphere, 33, 1047-1065.

De Wolf, W., Seinen, W., Opperhuizen, A., Hermens, J.L.M., 1992. The influence of biotransformation on the relationship between bioconcentration factors and octanol-water partition coefficients. Environmental Science and Technology, 26, 1197-1201.

Dietrich, D.W., Dreyer, N.D., Hansch, C.J., 1980. Confidence interval estimators for parameters associated with quantitative structure-activity relationships. Journal of Medicinal Chemistry, 23, 1201-1205.

Doucette, W.J., Andren, A.W., 1988. Aqueous solubility of selected biphenyl, furan, and dioxin congeners. Chemosphere**,** 17, 243-252.

Eastmond, D.A., 1993. Induction of micronuclei and aneuploidy by the quinone-forming agents benzene and o-phenylphenol. Toxicology Letters, 67, 1-3, 105-118.

Edgerton, S.A., Shah, J.J., 1992. Assessing total exposures to gasoline vapor using the source exposure model. Journal of Exposure Analysis and Environmental Epidemiology, 2, 1, 109-115.

Eldred, D.V., Weikel, C.L., Jurs, P.C., Kaiser, K.L.E., 1999. Prediction of fathead minnow acute toxicity of organic compounds from molecular structure. Chemical Research in Toxicology, 12, 670-678.

Ellington, J.J., Floyd, T.L., 1996. Octanol/water partition coefficients for eight phthalate esters. United States Environmental Protection Agency: Environmental Research Brief, Report no. EPA/600/S-96/006.

Ellington, J.J., 1999. Octanol/water partition coefficients and water solubilities of phthalate esters. Journal of Chemical and Engineering Data, 44, 1414-1418.

EPA (U.S. Environmental Protection Agency) and SRC (Environmental Science Center), EPIWIN Version 3.11., 2000.

Exxon Biomedical Sciences Inc., 1996. Water Solubility Study No. 199638. East Millstone, NJ, USA.

Fatemi, M.H., Jalali-Heravi, M., Konuze, E., 2003. Prediction of bioconcentration factor using genetic algorithm and artifical neural network. Analytica Chimica Acta, 486, 101-108.

Ferreira, M.M.C., 2001. Polycyclic aromatic hydrocarbons: a QSPR study. Chemosphere, 44, 125-146.

Fiedler, H., Schramm, K.W., 1990. QSAR generated octanol-water partition coefficients of selected mixed halogenated dibenzodioxins and dibenzofurans. Chemosphere, 20, 1597-1602.

Fishbein, L., 1992. Exposure from occupational versus other sources. Scandinavian Journal of Work, Environment and Health, 8(Suppl 1), 5-16.

Fossi, M.C., Casini, S., Marsili, L., 1999. Nondestructive biomarkers of exposure to endocrine disrupting chemicals in endangered species of wildlife. Chemosphere, 39, 1273-1285.

Friesen, K.J., Fairchild, W.L., Loewen, M.D., Lawrence, S.G., Holoka, M.H., Muir, D.C.G., 1993. Evidence for particle-mediated transport of 2,3,7,8-tetrachlorodibenzofuran during gas sparging of natural water. Environmental Toxicology and Chemistry, 12, 2037-2044.

Fromme, H., Küchler, T., Otto, T., Pilz, K., Müller, J., Wenzel, A., 2002. Occurrence of phthalates and bisphenol A and F in the environment. Water Research, 36, 1429-1438.

Fu, H., Demers, P.A., Costantini, A.S., Winter, P., Colin, D., Kogevinas, M., Boffetta, P., 1996. Cancer mortality among shoe manufacturing workers: an analysis of two cohorts. Occupational and Environmental Medicine, 53, 6, 394-398.

Ganning, A.E., Brunk, U., Dallner, J., 1984. Phthalate esters and their effect on the liver. Hepatology, 4, 3, 541-547.

Gao, C., Govind, R., Tabak, H.H., 1992. Application of the group contribution method for predicting the toxicity of organic chemicals. Environmental Toxicology and Chemistry, 11, 631-636.

Gawlik, B.M., Sotiriou, N., Feicht, E.A., Schulte-Hostede, S., Kettrup, A., 1997. Alternatives for the determination of the soil adsorption coefficient, K_{OC}, of non-ionic organic compounds - A review. Chemosphere, 34, 2525-2551.

Gilbert, D., Byrne, M., Harris, J., Steber, W., Woodruff, C., 1982. An exposure and risk assessment for benzene, final draft report, United States Environmental Protection Agency, Office of Water and Waste Management, Washington, DC.

Gledhill, W.E., Kaley, R.G., Adams, W.J., Hicks, O., Michael, P.R., Saeger, V.W., 1980. An environmental safety assessment of butyl benzyl phthalate. Environmental Science and Technology, 14, 301-305.

Gombar, V.K., Enslein, K., Blake, B.W., 1995. Assessment of developmental toxicity potential of chemicals by quantitative structure-toxicity relationship models. Chemosphere, 31, 2499-2510.

Govers, H., Ruepert, C., Aiking, H., 1984. Quantitative structure-activity relationships for polycyclic aromatic hydrocarbons: correlation between molecular connectivity, physico-chemical properties, bioconcentration and toxicity in *Daphinia pulex*. Chemosphere, 13, 227-236.

Govers, H.A.J., Krop, H.B., 1998. Partition constants of polychlorinated dibenzofurans and dibenzo-*p*-dioxins. Chemosphere, 37, 2139-2152.

Gute, B.D., Basak, S.C., 1997. Predicting acute toxicity (LC_{50}) of benzene derivatives using theoretical molecular descriptors: A hierarchical QSAR approach. SAR and QSAR in Environmental Research, 7, 117-131.

Hall, L., Kier, L., Phipps, G., 1984. Structure-activity relationship studies on the toxicities of benzene derivatives: I. An additivity model. Environmental Toxicology and Chemistry, 3, 355-365.

Hall, L.H., Maynard, E.L., Kier, L.B., 1989. Structure-activity relationship studies on the toxicity of benzene derivatives: III. Predictions and extensions to new substituents. Environmental Toxicology and Chemistry, 8, 431-436.

Hattemer-Frey, H.A., Travis, C.C., Land, M.L., 1990. Benzene: Environmental partitioning and human exposure. Environmental Research, 53, 221-232.

Hawker, D.W., 1989. Vapor pressures and Henry's Law Constants of polychlorinated biphenyls. Environmental Science and Technology, 23, 1250-1253.

Hollifield, H.C., 1979. Rapid nephelometric estimate of water solubility of highly insoluble organic chemicals of environmental interest. Bulletin of Environmental Contamination and Toxicology, 23, 579-586.

Hong, H., Wang, L.S., Han, S.K., Zou, G.W., 1996. Prediction adsorption coefficients (K_{OC}) for aromatic compounds by HPLC retention factors (k^1). Chemosphere, 32, 343-351.

Howard, P.H., Banerjee, S., Robillard, K.H., 1985. Measurement of water solubilities, octanol/water partition coefficients and vapor pressures of commercial phthalate esters. Environmental Toxicology and Chemistry, 4, 653-661.

Huang, H., Wang, X., Ou, W., Zhao, J., Shao, Y., Wang, L., 2003. Acute toxicity of benzene derivatives to the tadpoles (Rana japonica) and QSAR analyses. Chemosphere, 53, 963-970.

Huuskonen, J., Salo, M., Taskinen, J., 1998. Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. Journal of Chemical Information and Computer Sciences, 38, 450-456.

Istefanopulos, Y., 1987. State Variables and Linear Control Systems, 61-63, Boğaziçi University Publications, İstanbul.

Isnard, P., Lambert, S., 1988. Estimating bioconcentration factors from octanol-water partition coefficient and aqueous solubility. Chemosphere, 17, 21-34.

Jobling, S., Reynolds, T., White, R., Parker, M.G., Sumpter, J.P., 1995. A variety of environmentally persistent chemicals, including some phthalate plasticizers, are weakly estrogenic. Environmental Health Perspect, 103, 582-587.

Jorgensen, W.L., Duffy, E.M., 2002. Prediction of drug solubility from structure. Advanced Drug Delivery Reviews, 54, 355-366.

Kaiser, K.L.E., Niculescu, S.P., 1999. Using probabilistic neural networks to model the toxicity of chemicals to the fathead minnow (*Pimephales promelas*): A study based on 965 compounds. Chemosphere, 38, 3237-3245.

Kanazawa, J., 1981. Measurement of the bioconcentration factors of pesticides by freshwater fish and their correlation with physicochemical properties or acute toxicities. Pesticide Science, 12, 417-424.

Karabunarliev, S., Mekenyan, O.G., Karcher, W., Russom, C.L., Bradbury, S.P., 1996. Quantum-chemical descriptors for estimating the acute toxicity of substituted benzenes to the guppy (*Poecilia reticulata*) and fathead minnow (*Pimephales promelas*). Quantitative Structure-Activity Relationships, 15, 311-320.

Karickhoff, S.W., 1981. Semiempirical estimation of sorption of hydrophobic pollutants on natural sediments and soils. Chemosphere, 10, 833-846.

Kavlock, R., Boekelheide, K., Chapin, R., Cunningham, M., Faustman, E., Foster, P., Golub, M., Henderson, R., Hinberg, I., Little, R., Seed, J., Shea, K., Tabacova, S., Tyl, R., Williams, P., Zacharewski, T., 2002. NTP Center for the evaluation of risks to human reproduction: phthalates expert panel report on the reproductive and developmental toxicity of di-n-hexyl phthalate. Reproductive Toxicology, 16, 709-719.

Kellerova, V., 1985. Electroencephalographic findings in workers exposed to benzene. Journal of Hygiene, Epidemiology, Microbiology and Immunology, 29, 4, 337-346.

Khadikar, P.V., Singh, S., Mandloi, D., Joshi, S., Bajac, A.V., 2003. QSAR study on bioconcentration factor (BCF) of polyhalogented biphenyls using the PI Index. Bioorganic and Medicinal Chemistry, 11, 5045-5050.

Kier, L.B., Hall, L.H., 1976. Molecular connectivity in chemistry and drug design, 16-195, Academic Press.

Koch, R., 1983. Molecular Connectivity Index for assessing ecotoxicological behaviour of organic compounds. Toxicological and Environmental Chemistry, 6, 87-96.

Konemann, H., 1981. Quantitative structure-activity relationships in fish toxicity studies. Toxicology, 19, 209-221.

Krop, H.B., Van Velzen, M.J.M., Parsons, J.R., Govers, H.A.J., 1997. Determination of the adsorption-coefficient of organic substances on sewage sludges. Chemosphere, 34, 107-119.

Kuramochi, H., Ohsako, M., Maeda, K., Sakai, S., 2002. Prediction of physico-chemical properties for PCDD/Fs using the UNIFAC model with an alternative approximation for group assignment. Chemosphere, 49, 135-142.

Leisewitz, A., 1997. Stoffströme wichtiger hormonell wirkender substanzen (Behavior of important hormonal active substances). Project No. 10601076. German Federal Environmental Agency, Berlin.

Leong, B.K.J., 1977. Experimental benzene intoxication. Journal of Toxicology and Environmental Health. Supplement, 2, 45-61.

Letinksi, D.J., Connelly, M.J., Parkerton, T.F., 1999. Slow-stir water solubility measurements for phthalate ester plasticizers. Paper presented at SETAC Europe, Leipzig, Germany, May 25-29.

Lewis, S., Howe, A., Comber, S., Reynolds, P., Mascarenhas, R., Sutton, A., Rogers, H., 1998. Proposed environmental quality standards for phthalates in water. Final Report to the Department of the Environment, Transport and the Regions, DoE 3929, 3.

Leyder, F., Boulanger, P., 1983. Ultraviolet absorption, aqueous solubility and octanol-water partition coefficients for several phthalates. Bulletin of Environmental Contamination and Toxicology, 30, 152-157.

Lohninger, H., 1994. Estimation of soil partition coefficients of pesticides from their chemical structure. Chemosphere, 9, 1611-1626.

Long, J.M., 1995. SPARC estimations of K_{OW}s and solubilities for a series of phthalate esters. U.S. Environmental Protection Agency, NERL, Athens, Georgia.

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

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.

Lu, G.H., Yuan, X., Zhao, Y.H., 2001. QSAR study on the toxicity of substituted benzenes to the algae (*Scenedesmus obliquus*). Chemosphere, 44, 437-440.

Lu, G.H., Yuan, X., Wang, C., 2003. Quantitative structure-toxicity relationships for substituted aromatic compounds to *Vibrio fischeri*. Bulletin of Environmental Contamination and Toxicology, 70, 832-838.

Lyman, W.J., 1985. Estimation of Physical Properties. In Environmental Exposure from Chemicals, W.B. Neely and G.E. Blau (Eds.), CRC Press: Boca Raton FL, Vol 1, Chapter 2,13-47.

Lyman, W.J., Reehl, W.F., Rosenblatt, D.H., 1990. Handbook of Chemical Property Estimation Methods-Environmental Behavior of Organic Compounds. American Chemical Society, Washington, DC.

Mackay, D., Bobra, A., Shiu, W.Y., Yalkowsky, S.H., 1980. Relationships between aqueous solubility and octanol/water partition coefficients. Chemosphere, 9, 701-711.

Mackay, D., 1982. Correlation of bioconcentration factors. Environmental Science and Technology, 16, 274-278.

Mackay, D., Shiu, W.Y., Ma, K.C., 1992. Illustrated handbook of physical-chemical properties and environmental fate of organic chemicals. Vol.II: polynuclear aromatic hydrocarbons polychlorinated dioxins and dibenzofurans. Boca Raton: Lewis Publishers.

Mallhot, H., Peters, R.H., 1988. Empirical relationships between the 1-octanol water partition coefficient and nine physicochemical properties. Environmental Science and Technology, 22, 1479-1488.

Martin, T.M., Young, D.M., 2001. Prediction of the acute toxicity (96-h LC_{50}) of organic compounds to the fathead minnow (*Pimephales promelas*) using a group contribution method. Chemical Research in Toxicology, 14, 1378-1385.

McClelland, H.E., Jurs, P.C., 2000. Quantitative structure-property relationships for the prediction of vapor pressures of organic compounds from molecular structures. Journal of Chemical Information and Computer Sciences, 40, 967-975.

McDuffie, B., 1981. Estimation of octanol-water partition coefficients for organic pollutants using reverse-phase HPLC. Chemosphere, 10, 73-83.

McKay, G., 2002. Dioxin characterisation, formation and minimisation during municipal solid waste (MSW) incineration:review. Chemical Engineering Journal, 86, 343-368.

McMurry, S.T., Lochmiller, R.L., Vestey, M.R., Qualls, C.W., Elangbam, C.S., 1991. Acute effects of benzene and cyclophosphamide exposure on cellular and humoral immunity of cotton rats, *Sigmodon hispidus*. Bulletin of Environmental Contamination and Toxicology, 46, 937-945.

Mehlman, M.A., 1991. Dangerous and cancer-causing properties of products and chemicals in the oil refining and petrochemical industry: Part I carcinogenicity of motor fuels: Gasoline. Toxicology and Industrial Health, 7, 143-152.

Meylan, W.M., Howard, P.H., 1995. User's Guide for EPIWIN. Syracuse Research Corporation, Syracuse, NY,USA.

Meylan, W.M., Howard, H.P., Boethling, R.S., Aronson, D., Printup, H., Gouchie, S., 1999. Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. Environmental Toxicology and Chemistry, 18, 664-672.

Miller, M.M., Wasik, S.P., Huang, G.L., Shiu, W.Y., Mackay, D., 1985. Relationship between octanol/water partition coefficient and aqueous solubility. Environmental Science and Technology, 19, 6, 522-529.

Mondragon, F.A., Iranpour, R., Ahring, B.K., 2003. Toxicity of di-(2-ethylhexyl) phthalate on the anaerobic digestion of wastewater sludge. Water Research, 37, 1260-1269.

Moore, N.P., 2000. The oestrogenic potential of the phthalate esters. Reproductive Toxicology, 14, 183-192.

Neely, B.W., Branson, D.R., Blau, G.E., 1974. Partition coefficient to measure bioconcentration potential of organic chemicals in fish. Environmental Science and Technology, 8, 1113-1115.

Nendza, M., Russom, C.L., 1991. QSAR modeling of the ERL-D fathead minnow acute toxicity database. Xenobiotica, 21, 147-170.

Netzeva, T.I., Dearden, J.C., Edwards, R., Worgan, A.D.P., Cronin, M.T.D., 2004. QSAR analysis of the toxicity of aromatic compounds to *Chlorella vulgaris* in a novel short-term assay. Journal of Chemical Information and Computer Sciences, 44, 258-265.

Newsome, L.D., Johnson, D.E., Lipnick, R.L., Broderius, S.J., Russom, C.L., 1991. A QSAR study of the toxicity of amines to the fathead minnow. The Science of the Total Environment, 109/110, 537-551.

Nielson, N.M., Bundgaard, H., 1989. Evaluation of glycolamide esters and various other esters of aspirin as true aspirin prodrugs. Journal of Medicinal Chemistry, 32, 727-734.

Ning, H., Kado, N.Y., Kuzmicky, P.A., Hsieh, D.P., 1991. Benzene-induced micronuclei formation in mouse fetal liver blood, peripheral blood, and maternal bone marrow cells. Environmental and Molecular Mutagenesis, 18, 1:1-5.

NIOSH, National Institute for Occupational Safety and Health, 1988. Occupational Safety and Health Guideline for Benzene: Potential Human Carcinogen, U.S. Department of Health and Human Services, Public Health Service, Cincinnati, OH. Pub. No. 89-104, Supp. II.

Parker, H.G., Veys, C.A., Waterhouse, J.A.H., Peters, A., 1982. Cancer mortality in the British rubber industry. British Journal of Industrial Medicine, 39, 209-220.

Prausnitz, J.M., Lichtenthaler, R.N., De Azevedo, E.G., 1986. Molecular thermodynamics of fluid phase equilibria. Prentice Hall Inc., Englewood Cliffs.

Randic, M., 2001. The connectivity index 25 years after. Journal of Molecular Graphics and Modelling, 20, 19-35.

Rasmussen, A.G., 1998. Prediction of aqueous solubilities and octanol-water coefficients of phthalates by the UNIFAC group contribution method. NERI Research Notes No. 69, National Environmental Research Institute.

Reddy, K.N., Locke, M.A., 1994a. Relationships between molecular properties and log P and soil sorption (K_{OC}) of substituted phenylureas: QSAR Models. Chemosphere, 28, 1929-1941.

Reddy, K.N., Locke, M.A., 1994b. Molecular properties as descriptors of octanol-water partition coefficients of herbicides. Water, Air, and Soil Pollution, 86, 389-405.

Ren, S., Frymier, P.D., Schultz, T.W., 2003. An exploratory study of the use of multivariate techniques to determine mechanisms of toxic action. Ecotoxicology and Environmental Safety, 55, 86-97.

Renberg, L.O., Sundström, S.G., Rosen-Olofsson, A.C., 1985. The determination of partition coefficients of organic compounds in technical products and wastewaters for the estimation of their bioaccumulation potential using reversed phase thin layer chromatography. Toxicological and Environmental Chemistry, 10, 333-349.

Rhind, S.M., 2002. Endocrine disrupting compounds and farm animals: their properties, actions and routes of exposure. Domestic Animal Endocrinology, 23, 179-187.

Russell, D.J., McDuffie, B., 1986. Chemodynamic properties of phthalate esters: partitioning and soil migration. Chemosphere, 15, 1003-1021.

Sabljic, A., Protic, M., 1982. Molecular connectivity: A novel method for prediction of hazardous chemicals. Chemico-Biological Interactions, 42, 301-310.

Sabljic, A., 1987. The prediction of fish bioconcentration factors of organic pollutants from the molecular connectivity model. Zeits Chrift fur die Gesamte Hygiene und ihre Grenzgebiete, 33, 493-496.

Sabljic, A., Gutsen, H., Verhaar, H., Hermens, J., 1995. QSAR modeling of soil sorption-improvements and systematics of log K_{OC} vs log K_{OW} correlations. Chemosphere, 31, 4489-4515.

Sabourin, P.J., Sun, J.D., MacGregor, J.T., Wehr, C.M., Birnbaum, L.S., Lucier, G., Henderson, R.F., 1990. Effect of repeated benzene inhalation exposures on benzene metabolism, binding to hemoglobin, and induction of micronuclei. Toxicology and Applied Pharmacology, 103, 3, 452-462.

Saçan, M.T., İnel, Y., 1993. Prediction of aqueous solubility of PCBs related to molecular structure. Turkish Journal of Chemistry, 17, 188-195.

Saçan, M.T., İnel, Y., 1995. Application of the characteristic root index model to the estimation of octanol/water partition coefficients: polychlorinated biphenyls. Chemosphere, 30, 39-50.

Saçan, M.T., Balcıoğlu, I.A., 1996. Prediction of the soil sorption coefficient of organic pollutants by the characteristic root index model. Chemosphere, 32, 10, 1993-2001.

Saçan, M.T., Balcıoğlu, I.A., 1998. Estimation of liquid vapor pressures for low volatility environmental chemicals. Chemosphere, 36, 451-460.

Saçan, M.T., Erdem, S.S., Özpınar, G.A., Balcıoğlu, I.A., 2003. QSPR study on the bioconcentration factors of nonionic organic compounds in fish by characteristic root index and semiempirical molecular descriptors. Journal of Chemical Information and Computer Sciences, 44, 985-992.

Samiullah, Y., 1990. Prediction of the environmental fate of chemicals. Elsevier Science Publishers Ltd, London.

Schultz, T.W., Wilke, T.S., Bryant, S.E., Hosein, L.M., 1991. QSARS for selected aliphatic and aromatic amines. The Science of the Total Environment, 109/110, 581-587.

Schüürmann, G., Klein, W., 1988. Advances in bioconcentration prediction. Chemosphere, 17, 1551-1574.

Schwarzenbach, R.P., Gschwend, P.M., Imboden, D.M., 1995. Environmental organic chemistry- Illustrative examples, problems, and case studies. John Wiley & Sons, Inc., New York.

Sears, J.K., Darby, J.R. (Ed.), 1982. The Technology of Plasticizers, 906-907, Wiley Publishers, New York.

Shiu, W.Y., Doucette, W.J., Gobas, F.A.P.C., Andren, A.W., Mackay, D., 1988. Physical-chemical properties of chlorinated dibenzo-p-dioxins. Environmental Science and Technology, 22, 651-657.

Shiu, W.Y., Ma, K.C., 2000. Temperature dependence of physical-chemical properties of selected chemicals of environmental interest. 2. Chlorobenzenes, polychlorinated biphenyls, polychlorinated dibenzo-p-dioxins and dibenzofurans. Journal of Physical and Chemical Reference Data, 29, 387-462.

Sixt, S., Altschuh, J., Brüggemann, R., 1995. Quantitative structure-toxicity relationships for 80 chlorinated compounds using quantum chemical descriptors. Chemosphere, 30, 2397-2414.

Spartan 04 Windows, Wavefunction, Inc. 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612 USA.

Staples, C.A., Peterson, D.R., Parkerton, T.F., Adams, W.J., 1997. The environmental fate of phthalate esters: a literature review. Chemosphere, 35, 4, 667-749.

Stephenson, R.M., Malanowski, S., 1987. Handbook of the thermodynamics of organic compounds. Elsevier Science Publishing Co., Inc., New York.

Stewart, J.J.P., 1989. Optimization of parameters for semiempirical methods. Journal of Computational Chemistry, 10, 209-220.

Thomsen, M., Rasmussen, A.G., Carlsen, L., 1999. SAR/QSAR approaches to solubility, partitioning and sorption of phthalates. Chemosphere, 38, 11, 2613-2624.

USEPA., United States Environmental Protection Agency, 1989. National Emissions Standards for Hazardous Air Pollutants: Benzene Emissions from Maleic Anhydride Plants, Ethylbenzene Styrene Plants, Benzene Storage Vessels, Benzene Equipment Leaks, and Coke-Byproduct Recovery Plants, Fed. Reg. 54:38044-38139.

USEPA., United States Environmental Protection Agency, 1990. National Sewage Sludge Survey. Fed Reg, 55(218) : 47210 – 83.

USEPA., United States Environmental Protection Agency, 1994. Combustion Emissions Technical Resource Document (CETRED), Report No. EPA 530-R-94-014, Washington, DC.

USEPA., United States Environmental Protection Agency, 1998. The Inventory of Sources of Dioxin in the United States, Review Draft, EPA/600/P, 98002Aa (NTIS PB98-137037), National Centre for Environmental Assessment, Washington, DC, April.

Veith, G.D., DeFoe, D.L., Bergstedt, B.V., 1979a. Measuring and estimating the bioconcentration factor of chemicals in fish. Journal of the Fisheries Research Board of Canada, 36, 1040-1048.

Veith, G.D., Austin, N.M., Morris, R.T., 1979b. A rapid method for estimating Log P for organic chemicals. Water Research, 13, 43-47.

Veith, G.D., Macek, K.J., Petrocelli, S.R., Carroll, J., 1980. An evaluation of using partition coefficients and water solubility to estimate bioconcentration factors for organic chemicals in fish. In Eaton J.G., Parrish P.R., Hendricks A.C. (Eds), Aquatic Toxicology ASTM STP 707, American Society for Testing and Materials, 116-129.

Veith, G.D., Kosian, P., 1983. Estimating bioconcentration potential from octanol/water partition coefficients. In Mackay, D. (Ed), Physical Behaviour of PCBs in the Great Lakes, Ann Arbor Science Pub., Ann Arbor, MI, USA, 269-282.

Veith, G.D., Mekenyan, O.C., 1993. A QSAR approach for estimating the aquatic toxicity of soft electronphiles. Quantitative Structure-Activity Relationships, 12, 349-356.

Voogt, P.de., Muir, D.C.G., Webster, G.R.B., Govers, H., 1990. Quantitative structure-activity relationships for the bioconcentration in fish of seven polychlorinated dibenzodioxins. Chemosphere, 21, 12, 1385-1396.

Voutsas, E., Magoulas, K., Tassios, D., 2002. Prediction of the bioaccumulation of persistent organic pollutants in aquatic food webs. Chemosphere, 48, 645-651.

Wallace, L.A., Pellizzari, E.D., Hartwell, T.D., Perritt, R., Ziegenfus, R., 1987. Exposure to benzene and other volatile compounds from active and passive smoking. Archives of Environmental Health, 42, 272-279.

Wang, Y.H., Wong, P.K., 2002. Mathematical relationships between vapor pressure, water solubility, Henry's Law Constant, *n*-octanol/water partition coefficent and gas chromatographic retention index of polychlorinated-dibenzo-dioxins. Water Research, 36, 350-355.

Wang, Y.H., Wong, P.K., 2003. Correlation relationships between physico-chemical properties and gas chromatographic retention index of polychlorinated-dibenfurans. Chemosphere, 50, 499-505.

Wang, Z.Z., Li, X.D., 1981. Semi-experimental molecular orbital theory and practice. Science, Moscow.

Warne, M.St.J., Connell, D.W., Hawker, D.W., Schürmann, G., 1990. Prediction of aqueous solubility and the octanol-water partition coefficient for lipophilic compounds using molecular descriptors and physicochemical properties. Chemosphere, 21, 7, 877-888.

Webster, G.R.B., Friesen, K.J., Sarna, L.P., Muir, D.C.G., 1985. Environmental fate modelling of chlorodioxins: Determination of physical constants. Chemosphere, 14, 609-622.

Wei, D., Zhang, A., Wu, C., Han, S., Wang, L., 2001. Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds (PCOCs). Chemosphere, 44, 1421-1428.

Wolfe, N.L., Burns, L.A., Steen, W.C., 1980a. Use of linear free energy relationships and evaluative model to assess the fate and transport of phthalate esters in the aquatic environment. Chemosphere, 9, 393-402.

Wolfe, N.L., Steen, W.C., Burns, L.A., 1980b. Phthalate ester hydrolysis: linear free energy relationships. Chemosphere, 9, 403-408.

Yaffe, D., Cohen, Y., 2001. Neural networks based temperature-dependent quantitative structure property relations (QSPR) for predicting vapor pressure of hydrocarbons. Journal of Chemical Information and Computer Sciences, 41, 463-477.

Yang, P., Chen, J., Chen, S., Yuan, X., Schramm, K.W., Kettrup, A., 2003. QSPR models for physicochemical properties of polychlorinated diphenyl ethers. The Science of the Total Environment, 305, 65-76.

Younes, M., 1999. Specific issues in health risk assessment of endocrine disrupting chemicals and international activities. Chemosphere, 39, 8, 1253-1257.

Yuan, X., Lu, G., Lang, P., 1997. QSAR study of the toxicity of nitrobenzenes to river bacteria and *Photobacterium phosphoreum*. Bulletin of Environmental Contamination and Toxicology, 58, 123-127.

Zhu, H., Li, Y.B., Trush, M.A., 1995. Differences in xenobiotic detoxifying activities between bone marrow stromal cells from mice and rats-implications for benzene-induced hematotoxicity. Journal of Toxicology and Environmental Health, 46, 2, 183-201.

http://www.fwr.org, 2005. Foundation for Water Research Home Page.

http://www.ktl.fi/dioxin/ptoq.html, 2004. KTL, National Public Health Institute Home Page, Synopsis on dioxins and PCBs.

http://www.ukmarinesac.org.uk/activities/water-quality/wq8_43.htm#a4, 2004. U.K. Marine Special Areas of Conservation Home Page.