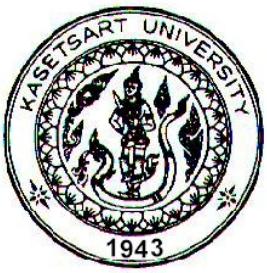


THESIS

COMPUTATIONAL TOXICITY IN SUBSTITUTED BENZENE DERIVATIVES FOR DYES PRODUCTION: A QSAR INVESTIGATION

WARAPORN JUNGTANASOMBUT

GRADUATE SCHOOL, KASETSART UNIVERSITY

2008

THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Chemistry)
DEGREE

Chemistry	Chemistry
FIELD	**DEPARTMENT**

TITLE: Computational Toxicity in Substituted Benzene Derivatives for Dyes Production: A QSAR Investigation

NAME: Miss Waraporn Jungtanasombut

THIS THESIS HAS BEEN ACCEPTED BY

Supa Hannongbua — **THESIS ADVISOR**
(Associate Professor Supa Hannongbua, Dr.rer.nat.)

Chak Sangma — **THESIS CO-ADVISOR**
(Mr. Chak Sangma, Ph.D.)

M. Thitinun — **THESIS CO-ADVISOR**
(Miss. Thitinun Monhaphol, Ph.D.)

Noojaree Prasitpan — **DEPARTMENT HEAD**
(Assistant Professor Noojaree Prasitpan, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON 16 May 2008

Gunjana Theeragool — **DEAN**
(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

COMPUTATIONAL TOXICITY IN SUBSTITUTED BENZENE DERIVATIVES FOR DYES PRODUCTION: A QSAR INVESTIGATION

WARAPORN JUNGTANASOMBUT

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Chemistry)
Graduate School, Kasetsart University
2008

Waraporn Jungtanasombut 2008: Computational Toxicity in Substituted Benzene Derivatives for Dyes Production: A QSAR Investigation. Master of Science (Chemistry), Major Field: Chemistry, Department of Chemistry. Thesis Advisor: Associate Professor Supa Hannongbua, Dr.rer.nat. 56 pages.

Benzene derivatives are widely used for dyes production. Some of these compounds show a significant part of the chemicals that are of environmental concern and also to human health. Therefore, quantitative structure-activity relationship (QSAR) method has been applied to investigate the relationships between the structure and toxicity of substituted benzene derivative against *Tetrahymena pyriformis*, protozoa. Initial three dimensional molecular structures of chemicals were searched from SciFinder Database. All geometries were minimized and calculated for atomic charge by using SYBYL 7.0. Then structural descriptors of compounds were calculated by using MOE program. The best log (IGC^{-1}_{50}) model (the negative logarithm of 50 percent growth inhibitory concentration against *T. pyriformis*), which is satisfactory in both statistical significance and predictive ability, was derived from multiple linear regression (MLR). This model consists of three physical properties descriptors as the following: molecular weight (Mw), sum of the atomic polarizabilities (apol), and octanol-water partition coefficient (log P). The equation is log (IGC^{-1}_{50}) = -1.308(±0974) + 0.016(±0.003) Mw -0.053(±0.046) apol + 0.178(±0.190) log P, yielding statistics: r^2 = 0.773, s = 0.277, and F = 37.439. From obtained QSAR model showed that molecular bulk, molecular polarizability, and hydrophobicity of compounds are directly relate to toxicity. The major parameter affect to toxicity is log P that indicated the partitioning of each toxicant into biophase and may have a role in the toxicity.

Waraporn Jungtanasombut — Student's signature

Supa Hannongbua — Thesis Advisor's signature

12 / May / 2008

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my advisor, Associate Professor Dr. Supa Hannongbua, for her excellent supervision, generous guidance, continuous support and encouragement throughout the course of my study. I am particulary grateful to Dr. Chak Sangma and Dr. Thitinun Monhapol for their extremely useful comments and discussion to make my thesis complete.

I owe a great debt of gratitude of Ms. Chonticha Suwattanasopon who calculate molecular descriptors by Molecular Operating Environment (MOE) program for me. This thesis would not have been completed without her.

Financial support from the Thailand Research Fund (TRF) and the Postgraduate program in education and research on Petroleum and Petrochemical Technology (ADB-MUA) were gratefully acknowledged.

Special thanks are due to University of Vienna and the High Performance Computing Center of National Electric and Computer Technology (NECTEC) in providing SYBYL and MOE program, respectively. The Laboratory for Computational and Applied Chemistry (LCAC) at Kasetsart University are also greatfully acknowledge for computational resource and software facilities.

I appreciate all of my colleagues at LCAC at Kasetsart University for their helpful assistance, friendship and great entertainment value during the past three years.

Finally, my utmost gratitude must be extended to my family for their encouragement, love, sincere care and support me throughout the entire study.

Waraporn Jungtanasombut

March, 2008

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

LIST OF ABBREVIATIONS

apol	=	Sum of the atomic polarizabilities
CAS	=	Chemical Abstracts Service Registry Number
EU	=	European Union
F	=	F-test value
LD_{50}	=	Median lethal dose
log (IGC^{-1}_{50})	=	Negative logarithm of 50% impairment growth concentration
log P	=	Log of the octanol-water partition coefficieeent
MLR	=	Multiple linear regression
Mw	=	Molecular weight
MR	=	Molecular refractivity
n	=	Number of cases/observations
QSAR	=	Quantitative structure-activity relationship
r^2	=	Squared correlation coefficient
REACH	=	Registration, Evaluation, and Authorisation of chemicals
s	=	Standard deviation
SSQ	=	Predictive error sum of square
SS_T	=	Sum of squares of deviation between the affinities of the fitting set and their mean affinity

COMPUTATIONAL TOXICITY IN SUBSTITUTED BENZENE DERIVATIVES FOR DYES PRODUCTION: A QSAR INVESTIGATION

INTRODUCTION

In December 2006 the European Parliament and the European Council have passed the new EU chemicals legislation (European Comission, 2006). Under the EU legislation for chemicals, called REACH (Registration, Evaluation and Authorisation of Chemicals), the declared aims are "to increase the protection of human health and environment from exposure to chemicals while at the same time to maintain and enhance the competitive and innovative capability of the EU chemical industry" (Women's Environmental Network, 2003). From this objective, the industry gives greater responsibility for managing the risks from chemicals and providing safety information on substance. The REACH entry into force estimated to be in the first half of 2007 (Lowell Center for Sustainable production, 2006). Therefore, Thai manufactures that want to export their products to EU should aware this legislation.

Synthetic organic colorants (e.g. azo dyes) are used universally in manufacturing processes. The synthetic dye industry was founded in 1857 by the young and latened English chemist William Henry Perkin in setting up a factory for manufacture of Mauviene from coal tar benzene (Abrahart, 1968). Several million different colored chemical compounds were synthesized. The chemical industry contained colors including red, blue, green, brown and black through the production of color different containing dyes (Sponza, 2006). Dyes are applied for coloration of various substrates (textile materials, leather, paper, plastics, food, drug, cosmetic *etc.*). A list of dyes by using classification comprises acetate rayon dyes, acid dyes, azoic dyes, basic dyes, direct dyes, mordant or chrome dyes, lake or pigment dyes, sulfur or sulfide dyes and vat dyes (World Bank Group, 1998). Approximately 10-15% of the total World production of dyes are released into the environment during synthesis and dyeing of substrate (Anliker, 1979). The main loss is to be found in residual liquors,

because of incomplete exhaustion of dye (Zollinger, 1987). Wastewaters produced by dye manufacturers typically comprise mixtures of the various dyes and their intermediate precursors (Plumb *et al.*, 2001). Even the very low concentrations (10-15 mg/L) of water soluble azo dyes can cause waste streams (Bae and Freeman, 2007). Therefore, the removal of dyes from effluent is one major environmental problem. Additional, dye manufacturing can produce air pollutants which are classified as volatile organic compounds (VOCs), nitrogen oxides (NO_x), hydrogen chloride (HCl) and sulfur oxides (SO_x) (World Bank Group, 1998).

Most of dye materials are irritants to the skin, eyes, and respiratory system and may be toxic by inhalation and ingestion (David, 2003). Moreover, some of dye types are known to be toxic or corrosive, some cause allergies (sensitizing), and some have long-term (chronic) health effects such as cancer e.g., Lissamine green (FD & C green dye no. 2) and Fast green FCF (FD &C green dye no. 3) were listed as an irritant with an oral LD_{50} 2 g/kg and more than 2 g/kg (possible carcinogen) in rats, respectively (Paterson *et al*., 2001). Especially, the chemicals such as substituted benzene derivatives that widely used for dye manufacturing processes, also show a significant part of the chemicals that impact to human. Additionally, many of these compounds were released into the environment compartments, especially in aquatic system. Therefore, dye manufacturers of Thailand get the effect from REACH if they don't have enough information about effect, uses and safety. This leads to testing requirements and classification. Toxic chemical requires an indicator organism for testing (Botsford *et al.*, n.d.). Some dyes, which damage living material (rat, mice, dogs, rabbits, fish, or *Daphnia*), are defined as being toxic but these tests are complicated and expensive. The development of alternative (non-animal) method for toxicity prediction is a widely explored research area. Moreover, alternative hazard testing methods and hazard assessment for human and environment are designed to potentially reduce time and monetary cost to the chemical industry needed for compliance under REACH. Therefore, the molecular modeling and statistic methods, used in QSAR (quantitative structure-activity relationship) field, constitute an important new tool for studying toxicity in chemical dyes. Because it is now being

increasingly viewed as one of the most cost effective alternatives to estimate ecological and health effects of chemicals.

QSAR, introduced by Hansch *et al.* (1964), is a computer-based mathematical model, which relates the biological activity of compounds to theoretically calculated or experimental descriptors of their chemical structures (Lessigiarska *et al*., 2006). It is sometimes called *in silico* models that have been nearly 40 year since the QSAR paradigm firstly found the way into the practice of agrochemistry, pharmaceutical chemistry, and toxicology (Hansch and Leo, 1979). In present, QSAR approach has increasing interest in assessing the toxicological effect of chemicals more than other methods, animal-base and in vitro methods, since it is easy to apply and efficient in terms of time and financial cost. Its staying power may be attributed to the strength of its initial postulate that activity was a function of structure as described by electronic attribution, hydrophobicity, and steric properties as well as the rapid and extensive development in methodologies and computational techniques (Selassie, 2003). QSAR is an estimation method which developed and used in order to predict properties of chemical substance. Several physico-chemical and fate parameters are used for establishing an exposure level based on modeling. Because of the lack of reliable experimental data, these parameters may be derived by QSAR method (European Commission, 1996). Moreover, it may provide an understanding of the effect of structure on activity.

In the present work, toxicity of substituted benzene derivatives for dyes production studied by using QSAR. Since these chemicals widely used in dye manufacturer and many of these chemicals were released into the environment, especially in aquatic system, therefore the goals of this study are following:

1. To set up the toxicity quantitative structure-activity relationship model of substituted benzene derivatives against *Tetrahymena pyriformis* (protozoa) for predicting the toxicity. Due to toxicity to aquatic organisms such as protozoa was used to indicate some of the effect on the environment (Rose and Hall, 2003).

2. To determine obtained quantitative structure-activity relationship model.

3. To reduce the use of animal testing, as required by the European Directive on the Protection of Laboratory Animals.

The obtained results are expected to help dye industry for coping and preparing with EU legislation. Additionaly, derived quantitative structure-activity relationship can be used to estimate toxicity values of other molecules outside the dataset.

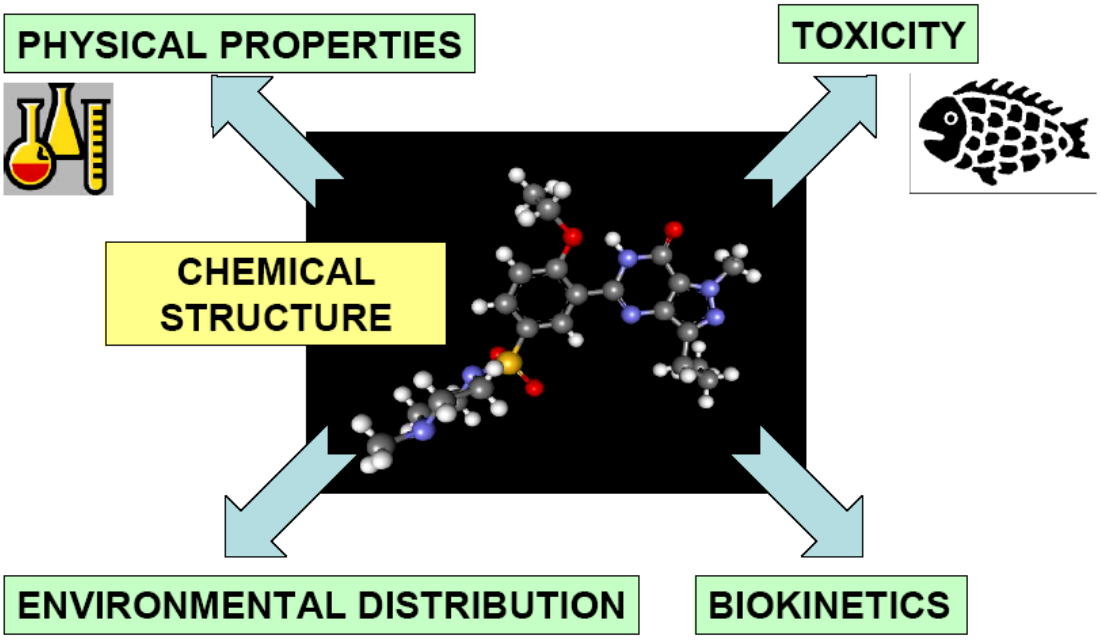


Figure 1 Information from chemical structure which will be used to construct QSAR model, related to toxicity and distribution of chemical in environment.

Source: Worth (2006)

LITERATURE REVIEW

Crum-Brown and Fraser (1868) proposed an equation that is considered to be the first general formulation of quantitative structure-activity relationship (QSAR) as such $\Phi = f(C)$ (Φ is the physiological activity and f(C) is the fuction of the chemical structure C). This equation linking change in biological activity to change in chemical structure that show the physiological activity must be a function of chemical structure.

Hansch *et al.* (1962) studied the structure-activity relationships of plant growth regulators which depended on Hammett constant and hydrophobicity. Their equation is $\Pi_x = \log P_x - \log P_H$ (P_x and P_H represent the partition coefficients between octanol and water of a derivative and the parent molecule, respectively). The parameter Π is the relative hydrophobicity of a substituent, was defined in a manner analogous to the definition of sigma (σ).

Hansch *et al.* (1964) used important variables, the substituent constant σ and the substituent constant Π, for studying QSAR. They investigated the effect of substituents on the biological activity of benzoic acid on mosquito larvae, phenols on gram-positives by using regression analysis. From this analysis they found that the constants related with Π provided insight into the nature of the cellular material through which a molecule must make its way before reaching the site of action. The results indicated that when strong electronic interaction is absent, Π is constant and additive in character. This should make the construction of isolipotropic molecules relatively easy.

Hansch and Steward (1964) utilized the substituent-constant analysis for correlating the relative activity of a series of 22 penicillin derivatives by mean of σ and Π parameters. σ (Hammet constant) is a measure of electronic effect of one or more substituents on the phenoxy ring and Π is constant that was defined as Hansch and Muir (1962). From this analysis it clearly shown that the lipophilic character modification of penicillins by changing side-chain substituents is primary effect to

biological activity. Additionally, electronic and negative steric effects of substituents on the phenoxy ring appear to be of minor importance.

Iwasa *et al.* (1965) determined the partition constant (Π) of aliphatic functions from the partition coefficients of a variety of compounds of the type $C_6H_5(CH_2)_nX$. The biological activity (log (1/C) values for the isonarcotic concentration of tadpoles by alcohols, esters, ketones and ether) and the additive nature of log P chemical structure were fitted to the equation log (1/C) = k log P + c by the method of least squares.They suggested that the log P or Π would appear to be the most useful parameter through variation of the solvents, one should be able to delvelope a model system more closely approximating the biophases. Moreover, They have shown the relation of Π to ΔR_M (a chromatograpiccally determined substituent constant) that is a very good linear correlation.

Hansch *et al.* (1965) separated electronic and adsorption effects on the rates of enzymatics reactions by using electronic, steric, “hydrophobic bonding” substituents and regression analysis. They concluded that regression analysis and substituent constants, especially Π and σ, provide a new approach for the difficult task of mapping the site of action on enzymes.

Fujita and Hansch (1967) applied structure-activity relationship for the correlation of biological activity and chemical structure of dissociable compounds under conditions of physiological pH using the Hammett constant (σ) and the hydrophobicity constant (Π). They also applied the correlation of effect of dissociation for analyzing the bacteriosatic activity and protein binding of the sulfonamide drugs. They summarized that the hydrophobicity of drugs play a definite role on the activity and the optimal hydrophobic character for activity can deduced from the relationships. In addition, the most favorable dissociation constant for the maximum activity and the optimal hydrophobicity for a series of sulfonamide can suggested for designing of new sulfonamide drugs.

Hansch *et al.* (1967) examined a variety of different hypnotics by fitting the experimental results to equation log (1/C) = - k $(\log P)^2$ + k' log P + k'', C represents the moles of drug per kilogram of test animal producing "hypnosis" and k, k', and k'' are constants obtained from the method of least squares, and compared the log P_0 values (come from setting the derivative d log (1/C) / d log P equal to zero, then solving the resulting equation for log P yielgs) for the different sets. The substituent effects of a single series by using Π values for substituents and log P for barbituric acid were correlated. They utilized 5,5-substituted barbiturate function (Figures 2) that has Π equal to -1.35, taking advantage of the additive-constitutive character of Π and log P, for studying this work. It shows that whenever aromatic rings are with polar function in a side chain, log P is lower than one would expect from the simple additivity principle. The obtained results indicated that the hypnotic activity of barbiturates groups depend almost entirely on their relative lipophilic character as defined by their octanol-water partition coefficients. From constant log P_0, it also shown that the other set of hypnotics structurally unrelated to the barbiturate, but the metabolism rate of barbiturates is linearly related to their partition coefficients.

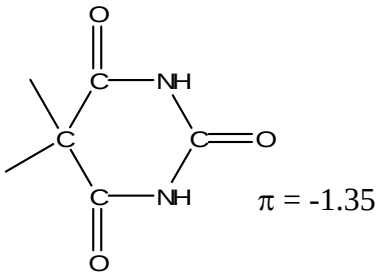


Figure 2 Structure of 5,5-substituted barbiturate function.

Leo *et al.* (1969) compared octanol-water partition coefficients with other physical constants (polarizability (α), molar attraction constant (F), parachor (P_r), adjusted parachor (P_r^*), and molecular weight) for studying the structure-activity relationships, which one could correlate the different in biological activity of the member of a set of congeners. They correlated the reactivity of a large group of the miscellaneous molecules in four different biological systems (tadpole narcosis, frog

muscle narcosis, complex with bovine serum albumin (BSA), and chick embryo hatching) with six physical constants and found that three of the systems show a linear dependence upon these parameters and the fourth requires the addition of a squared term. Additionally, the log P (octanol-water) can correlate a greater percentage of the biological activity than the other parameters.

Venger *et al.* (1979) tested the mutagenicity of 1-(X-phenyl)-3,3-dialkyltriazenes in the Ames test using *Salmonella typhimurium* TA92. Then they constructed QSAR model which can be expressed as following: log 1/C = 1.090 log P – 1.63 σ^+ + 5.58. C is the molar concentration of triazene producing 30 mutations/ 10^8 bacterial above background. From this equation shows that increasing lipophilicity and increased electron release via through resonance increases mutagenicity. In addition, the QSAR for mutagenicity was also compared with QSAR for antileukemia action and toxicity (LD_{50}) in mice. From these they summarized that both gross toxicity and antitumor activity in mice show the same dependence on the electronic effects of substituents, but they did not clear why mutagenicity is so much more sensitivity than antitumor activity to the electronic effect of substituents.

Debnath *et al.* (1991) made a survey of literature to find the mutagenicity data of nitroaromatic and heteroaromatic compounds for setting up the QSAR model. From the derived QSAR model of 188 congeners they found that the hydrophobicity which can modeled by octanol/water partition coefficients and the energies of the lowest unoccupied molecular orbitals which calculated by using the AM1 method are the main determinants of mutagenicity of these compounds. Additionally, its also shown that chemicals composing three or more fused ring much greater mutagenic potency than compound with one or two fused rings.

Debnath and Hansch (1993) analyzed the mutagenic activityof a series of 15 methanesulfonate esters activity on *Salmonella typhimurium* TA100 by using QSAR approach. They have converted Eder and Kutt (1989) values, reported mutagenic activity in term of rev/μmol, to rev/nmol for easier comparison with their other studies. The derived QSAR model can be ccorrelated as the following equation:

log TA100 = 1.10 log P + 0.73 log MMI - 2.53. P is the octanol/water partition coefficient that was calculated using the CLOGP program version 3.54. MMI is the relative rate of reaction of the sulfonate esters with N-methyl-2-mercaptionmidazole that appears to take care of the stereo-electronic interactions of the DNA action on the esters. However, they conclude that the most important parameter for mutagenicity is log P, not MMI. Becauese of these esters do not need microsomal activation and are mutagenic per se, the hydrophobicity (log P) must be associated with the penatration of the esters to the sites of interaction within the cell and reaction with DNA.

Oprea *et al.* (1997) analyzed the nature of dye-fiber interaction for 27 disperse azo dyes using several quantitative structure-activity relationship methods by correlating variations in the chemical structure with the affinity to cellulose fiber ($-\Delta\mu^\circ$). They found the obtained classical QSAR suggested that sterio, not hydrophobic effects, were important for enhancing $-\Delta\mu^\circ$. For Comparative Molecular Field Analysis (COMFA) results, a three-dimentional QSAR (3D-QSAR) method, implied that the pharmacophore theory of dye –fiber interaction holded true.

Giorgi *et al.* (1997) compared the fastness performance of the series of acid azo dyes, designed by a chemometric approach (Figure 3), on different fibres (silk, wool and nylon) of analogous structure. Fastness data were modelle as a function of the structure by partial least square (PLS) and the model used to identify the best dye for each fibre. They conclude that the fastness performance of the series on silk and on nylon appears to be similar but on wool is neatly inferior. Additionally, the best dye for each fibre was produced from different amine components. From derived QSAR models, gamma acid shown to be the optimum agent for the three fibres whereas the validity of the application of Experiment Design and QSAR modeling by PLS to find out the best dyes in a series for each of the fibres examined.

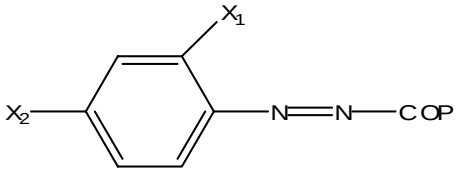


Figure 3 The general formular in a series of azo dyes.

He *et al.* (2000) evaluated the photodynamic effects of hypocrellins and their derivatives to human oral cavity epithethelial carcinoma KB cell line and surveyed the structure-activity relationships, which facilitate to select the among hypocrellin dyes suitable for photodynamic therapy (PDT) treatment. From obtained results they conclude that the photophysics and phototoxicity of these photosensitizers depended on the chemicals structure. The amphiphilicity, has been evaluated by measuring their partition coefficient between n-octanol and phosphate-bufferered saline (PBS) buffer (pH 7.4), as well as the singlet oxygen generating quantum yield of the hypocrellin dyes affected their photodynamic activity. The hydrophobic and amphiphilic dye with higher singlet oxygen-generating quantum yield exhibited high photodynamic activity, while the most hydrophilic dyes exhibited the lowest phototoxic activity. Cysteamine mono- and di-substituted hypocrellin B (Figure 4a and 4b) and cysteine mono-substituted hyprocrellin B (figure 4c) appropriate hydrophobic and amphiphilic property and high photodynamic activity to KB cells, might prove to be potential phototherapeutic agents for PDT.

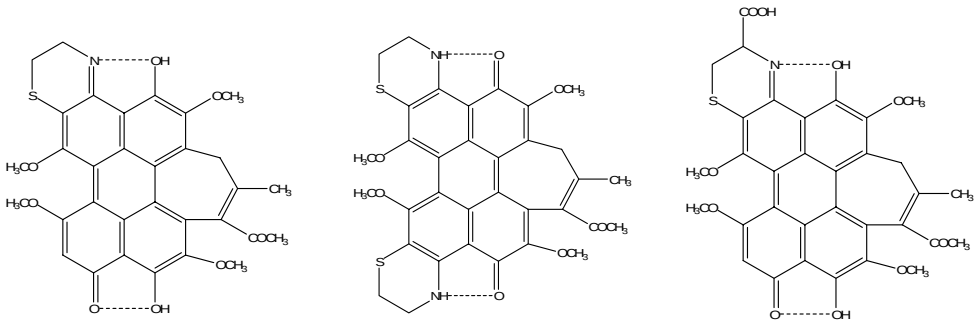


Figure 4 Chemical structures of (a) cysteamine mono-substituted hypocrellins B
(b) cysteamine di-substituted hypocrellins B
(c) cysteine mono-substituted hyprocrellin B

Timofei *et al.* (2000) reviewed the QSAR for dye affinity for celluloses fibres. They studied series of anthraquinones vat dyes, mono and biazo and disperse dyes by several variants of classical QSAR and 3DQSAR and of other modern methods. They conclude that QSAR studied concerning the effect of structural features of dye molecules upon adsorption on cellulose fibres. For attractive dye-cellulose interaction, electrostatic and steric interactions are more important for binding than hydrophobic interactions whereas anionic mono azo dyes an increase in lipophilicity contributes to dye-cellulose inding. In the case of disperse dyes, steric interactions are also important for this binding.

Greaves *et al.* (2001) developed much more refined bioelimination-molecular structure relationship of anionic, water-soluble dyes. They found the larger ratio of molecule weight to number of sulphonate group the greater the bioelimination. Because of an increase in the number of aromatic primirary amines, the number of unsulphonated naphthalene nuclei, the size/charge ratio and the length of the dyes, the bioelimination increase. Besides, they also found the decreased number of aliphatic alcohol group and the increased bioelimination.

Garg *et al.* (2002) derived and developed quantitative structure-activity/property-activity relationships (QSAR/QPARs) of 43 aminoazobenzene derivatives (Figure 5), 4-aminoazobenzene (AAB), N-methyl-4-aminoazobenzene (MAB) and n,n-dimethyl-4-aminoazobenzene (DAB) derivatives, by correlating between the observed mutagenic activity of these chemicals in the *S.typhimurium* TA98 bacterial strain with S9 activation (TA98+S9) to various molecular descriptors. They have been shown that models based on multilinear regression techniques and artificial neural networks are capable of accounting for more than 80% of the variation in TA98+S9. From the best multilinear regression (BMLR) models, they found that the hydrophobicity descriptor log P does not appear in any of these BMLR correlation equations. Science these aminoazobenzene compounds can have rather similar values of log P but quite different mutagen activities. Moreover, all of the BMLR models suggest that the relative mutagenicity of these aminoazobenzene derivatives increase as the average electrophilic reactivity index, $\overline{\text{ERI}}$, for all the

nitrogen in the compound increase. For artificial neural network (ANN) models they summarized that only three descriptors were needed to adequately train these ANN models because of the limited data available. However, they found that these several 3-descriptor ANN models can account for over 90% of the variation in mutagenicity of the compounds.

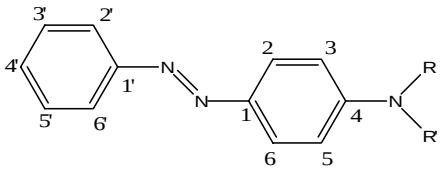
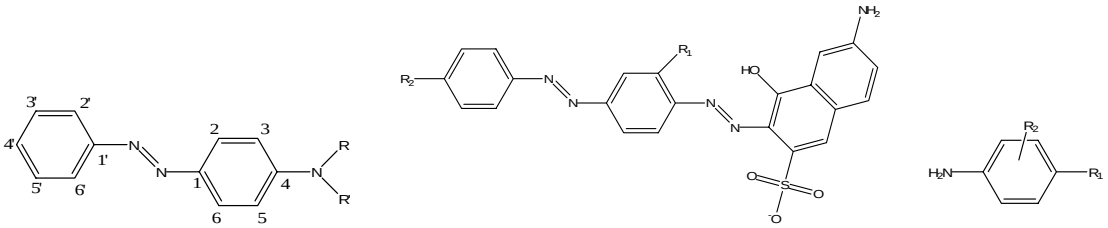


Figure 5 The general structure of aminoazobenzene derivatives.

Sztandera *et al.* (2003) constructed and modified QSAR/QPARs for the observed mutagenicity in *Salamonella typhimurium* TA98 bacterial tester strain in the present of an induced rat-liver S9 (TA98+S9) mix of an enhanced collection of aminoazo dyes and their reductive-cleavage metabolites. The quantitative mutagenicity data of 74 compounds (Figures 6), aminoazobenzene derivatives, disazo derivatives and reductive cleavage products, they had used come from a variety of laboratories which the ranges of the mutagenic activity of these compounds about 10^{-3} to 10^{2} rev/nmol. They concluded that multilinear regression techniques using 8 descriptors were shown to account for about 73% of the variation in the relative mutagenic activity of aminoazo dyes (62 aminoazo derivatives and 12 of their reductive cleavage products). For aminoazobenzene derivatives and disazo derivatives, the main descriptor is either the minimum or maximum values of the electrophilic reactivity index for an N atom, EIR_N, while the reductive cleavage compounds do not have azo linkages, the main descriptor is average of this index. The large values of this index occur at azo nitrogen atoms and smaller values occur at amino nitrogen atoms. The mutagenicity increases as the maximum values of EIR_N increase for 8 descriptors equation. Additionally, they also used artificial neural networks (ANNs) for the relative mutagenic activity in TA98+S9 of these 74 compounds. They summarized that the 8-descriptors ANN can account for about 95% of the reported

variation. The total dipole moment and various polarizabilities play important roles in the neural networks.



6a 6b 6c

Figure 6 The general structures of (a) aminoazobenzene derivatives
(b) disazo derivatives
(c) reductive cleavage products

Schüürmann and Funar-Timofei (2003) analyzed QSAR model of 30 direct azo dyes about the affinity of the dye molecules for the cellulose fiber, based on dye-fiber interactions or dye-dye interactions or both. These studies employed Comparative Molecular Field Analysis (CoMFA) for analysizing the intermolecular interactios in term of site-specific steric and electrostatic fields of the substrate. Additionally, their investigation also included the solution-phase molecular descriptors. They found the dyefiber affinity increases with increasing electron donor capacity that corresponds to an increasing hydrogen bond acceptor strength of azo dyes. With the azo dyes, likely hydrogen bonding acceptor site is the azo nitrogen, interacting with OH group of the cellulose polysaccharide.

Estrada *et al.* (2003) used topological substructural molecular desciptors (TOPS-MODE) to derived models for understanding the molecular structural contribution to skin sensitization. They conclude that the structural contributions to skin sensitization for various classes of chemical are presented on the basis of bond contributions. Moreover, the derived models able to identify potential structure alerts forchemicals requiring metabolic activation.

Søsted *et al.* (2004) published paper in the topic of ranking of hair dye substance according to predicted sensitization potency: quantitative structure-activity relationships. This work proposed to identify all hair dye substances registered in Europe and provided their tonnage data. They substance according to their estimated potency which being to supplement the current diagnostic work-up for hair dye allergy with new potential allergens. From QSAR model (Estrada *et al.*, 2003) they found that 75%, 22% and 3% of hair dye substances were predicted to be strong moderate, weak and extremely weak or non-sensititizing, respectively.

Wang *et al.* (2005) applied three-dimensional quantitative structure activity relationships (3D-QSAR) model, utilizing comparative molecular field analysis (CoMFA) and comparative similarity indices analysis (CoMSIA), for predicting the toxicity of 46 polybrominated diphenyl ethers (PBDEs, Figure 7). They built 3D-QSAR of 18 PBDEs congeners, based on calculated indices and reported experimental toxicology index (aryl hydrocarbon receptor relative bunding affinities, RBA), then determined the factors that required for RBA of these PBDEs. The satisfactory results of 3D-QSAR models were obtained with cross-validation Q^2 and R^2 values equal to 0.58 and 0.995 by CoMFA model and 0.68 and 0.982 by CoMSIA model. From the steric and electrostatic contour maps of CoMFA and CoMSIA models they found that both 3D-QSAR models have similary contribution. Additionally, they also concluded that the nonplanar conformations result in the lowest energy level for PBDEs and the main factor reflecting the RBA of PBDEs was electrostatic index.

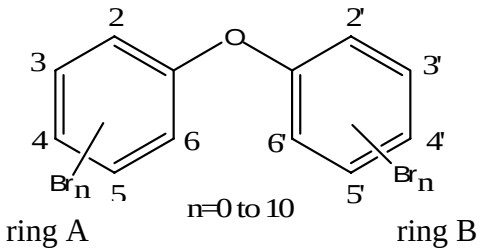


Figure 7 Chemical structure of the PBDEs molecule.

Costescu and Diudea (2006) studied the quantitative structure-toxicity relationship (QSTR) in the field of aquatic toxicology, fish and protozoa toxicity of various aromatic compounds. They devided these compounds by functional group in to four sets: phenol substituteds, benzene substitutes, nitrobenzene substitutes, and aniline substitutes. The model of four sets derived from correlation analysis and multiple linear regression (MLR). From obtained toxicity models of four sets against *Poecila reticulate* and *Tetrahymena pyriformis*, they concluded that topological indices base on Cluj matrices show a good predictive ability of the aquatic toxicity against *Poecila reticulate* and *Tetrahymena pyriformis*. These indices account for molecular bulk, halogen, nitro and amino substitution in benzene ring.

Furnar-Timofei *et al.* (2006) applied traditional and rational QSAR/QSPR modeling techniques for finding quantitative structure-retention relationship (QSRR) of disazo and triasazo 4,4'-diaminobenzaniline-based direct dye molecules. They conclude that dye lipophilicity can be described by hydrophobic and polarity dye structure parameters. Additionally, comparative molecular field analysis (CoMFA) and comparative molecular similarity index approach (CoMSIA) also indicated importance of steric and electrostatic interactions to chromatographic mobility.

Lessigiarska *et al.* (2006) investigated the *in vivo* human and rodent toxicity including a combination of toxicity endpoint and structural descriptors of chemicals (inorganics, simple organic chemicals, alkaloids and drugs) by using the quantitative structure-activity-activity relationship (QSAAR) technique. The obtained results can be concluded that the human peak blood/serum LC_{50} concentration strongly relate to human liver cell toxicity, the in vivo oral human lethal dose most closely relate to the in vivo rodent LD_{50} values. Moreover, they also developed quantitative structure-activity relationships (QSARs) for both *in vivo* and *in vitro* endpoint. The derived QSAR models used structural descriptors for setting up, accounting for molecular hydrophobicity, size and shape, and electronic properties, have the potential in the development of non-animal methods for predicting human and mammalian toxicity.

Kumar *et al.* (2006) applied the quantitative structure-activity relationship approach for understanding the affinity and selectivity of triaryl imidazole derivatives to wards glucagon receptor. The obtained QSAR models of glucagon receptor inhibition by triaryl imidazoles were constructed from QuaSAR descriptors of molecular operating environment (MOE). From the generated QSAR models they found that the hydrophobicity, molecular shaoe and geometry predominant to glucagon receptor binding affinity of the triaryl imidazoles. Therefore, the relevance of shape has specific steric interactions between the molecule and the receptor.

Papa *et al.* (2007) developed quantitative structure-activity relationship models of the large and heterogeneous data set of 640 organic chemicals, such as dyes, for predicting the logarithm of fish bioconcentration factor (logBCF) that is steady-state ratio of the chemical concentration in exposed organism to the concentration of dissolved chemical in the aqutic environment. From derived QSAR models, they found polarizability, hydrogen bonding and chemical dimension are important descriptors for predicting logBCF.

Li and Xi (2007) established and developed quantitative structure-biodegradability relationships (QSBRs) of acid dyestyffs, used in textiles, leathers, paper, wood and inks, for studying the predictive and mechanism-based relationships on the biodegradation that seems to be essential and to prevent dye molecules from persisting in the aquatic environment. They used chemical and molecular descriptors, molecular weight (M_w), energies of the highest occupied molecular orbital (E_{HOMO}), the lowest unoccupied molecular orbital (E_{LUMO}), and excited state (E_{ES}), to seted up QSAR models. The QSAR models were obtained through each descriptor for investigating the controlling mechanism of biodegradation. From the obtained results, they conclude that the dominant parameters for controlling the biodegradability of acid dyes were E_{HOMO} and M_w.

Lin (n.d.) published papers for explaining MOE (Molecular Operating Environment) system for QSAR and QuaSAR-descriptor. TheQSAR/QSPR functions in MOE are provided in a QuaSAR package. The QuaSAR system uses SVL (Scientific Vecter Language) program modules as the source for all molecular descriptors. The calculation of QuaSAR-descriptor can proceeds as follow. Given a molecular database with a molecule field, a set of numerical properties will be calculated for each molecule and stored in database. Every descriptor is given a unique name, or code, which identifies the descriptor. After descriptors are calculated, they are sent to the suite of MOE QuaSAR modle-building functions for model fitting and evaluation.

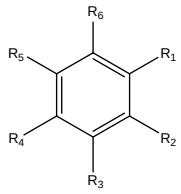
Porcelli *et al.* (2008) used the DEMETRA, Development of Environmental Modules for Evaluation of Toxicity of pesticide Residues in Agriculter, acute toxicity model toward the water flea (*Daphnia magna*) as a case study to outline a varidation method compatible with regulatory use. The DEMETRA model for *Daphnia* is based on a training set of 220 compounds. This model was built through a hybrid model approach. The final model is composed of three individual models joined in a mathematical function that leads them toward a single predictive value. They verified the reliability, predictive power, uncertainly and applicability of the DEMETRA acute toxicity model by using an external test set of pesticides. Moreover, they also compared the prediction of this external set using the DEMETRA model with the results using EOSAR and TOPKAT as benchmarks. They found that the DEMETRA gave good statistical predictions and the maximum error of the outliner was lower than that the other two models.

METHODS OF CALCULATIONS

Chemical Structures and Toxicity Data

In order to investigate the relationships between structure and toxicity of substituted benzene derivatives to *Tetrahymena pyriformis*, 50 compounds were selected. The chemical structures of substituted benzene derivatives and their toxicity values used in this study are shown in Table 1. The aquatic toxicity values of *Tetrahymena pyriformis* (protozoa) obtained IGC_{50} values, the 50% impairment growth concentration, were derived from Schultz, 1997. These toxicity values were converted into negative logarithm, log (IGC^{-1}_{50}). Where IGC_{50} is impairment growth concentration for 50% of control populations. In other word, it is the effective concentration that reduces population density to 50 percent of controls. In this work the log (IGC^{-1}_{50}) was used as a dependent variable in QSAR study. The whole toxicity data of substituted benzene derivatives against *Tetrahymena pyriformis*, based on the range of these toxicity data, was divided into two subsets as following: the training set 37 compounds (Tables 1) and the test set 13 compounds (Table 2), respectively. The training set of substituted benzene derivatives was used to set up QSAR model on the aquatic toxicity and to examine for predictive ability of model. In addition, the selected compounds from various structures of substituted benzene derivatives (test set) were utilized to test the reliability, predictive power and uncertainly of the model.

Table 1 Structure of substituted benzene derivatives and their observed toxicity values (log (IGC^{-1}_{50})), used in the training set.



No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Observed log (IGC^{-1}_{50})
1	98-85-1	CH_2CH_3 | OH	H	H	H	H	H	-0.66
2	87-62-7	NH_2	CH_3	H	H	H	CH_3	-0.43
3	108-95-2	OH	H	H	H	H	H	-0.35
4	95-48-7	OH	CH_3	H	H	H	H	-0.29
5	95-68-1	NH_2	CH_3	H	CH_3	H	H	-0.29
6	108-44-1	NH_2	H	CH_3	H	H	H	-0.28
7	100-52-7	CH=O	H	H	H	H	H	-0.20
8	95-51-2	NH_2	Cl	H	H	H	H	-0.17
9	95-53-4	NH_2	CH_3	H	H	H	H	-0.16
10	106-44-5	OH	H	H	CH_3	H	H	-0.16
11	108-90-7	Cl	H	H	H	H	H	-0.13
12	91-23-6	OCH_3	NO_2	H	H	H	H	-0.07
13	88-05-1	NH_2	CH_3	H	CH_3	H	CH_3	-0.05
14	99-09-2	NH_2	H	NO_2	H	H	H	0.03
15	100-44-7	CH_2Cl	H	H	H	H	H	0.06
16	103-69-5	$NHCH_2CH_3$	H	H	H	H	H	0.07
17	88-74-4	NH_2	NO_2	H	H	H	H	0.08
18	98-95-3	NO_2	H	H	H	H	H	0.14
19	123-07-9	OH	H	H	H	H	H	0.21
20	108-88-3	CH_3	H	H	H	H	H	0.25
21	88-72-2	CH_3	NO_2	H	H	H	H	0.26
22	121-89-1	$COCH_3$	H	NO_2	H	H	H	0.32
23	95-74-9	NH_2	H	Cl	CH_3	H	H	0.39
24	95-79-4	NH_2	CH_3	H	H	Cl	H	0.5
25	106-48-9	OH	H	H	Cl	H	H	0.54
26	95-82-9	NH_2	Cl	H	H	Cl	H	0.58
27	99-99-0	CH_3	H	H	NO_2	H	H	0.65
28	88-73-3	Cl	NO_2	H	H	H	H	0.68

Table 1 (Continued)

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Observed log (IGC^{-1}_{50})
29	97-02-9	NH_2	NO_2	H	NO_2	H	H	0.72
30	121-73-3	Cl	H	NO_2	H	H	H	0.73
31	121-87-9	NH_2	Cl	H	NO_2	H	H	0.75
32	99-65-0	NO_2	H	NO_2	H	H	H	0.76
33	108-70-3	Cl	H	Cl	H	Cl	H	0.87
34	121-14-2	CH_3	NO_2	H	NO_2	H	H	0.87
35	51-28-5	OH	NO_2	H	NO_2	H	H	1.06
36	99-54-7	NO_2	H	Cl	Cl	H	H	1.16
37	97-00-7	Cl	NO_2	H	NO_2	H	H	2.16

*CAS: Chemical Abstracts Service Registry Number

Table 2 Structure of substituted benzene derivatives, used in the test set and their observed toxicity values (log (IGC^{-1}_{50})).

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Observed log (IGC^{-1}_{50})
1	108-46-3	OH	H	OH	H	H	H	-0.65
2	95-78-3	NH_2	CH_3	H	H	CH_3	H	-0.33
3	62-53-3	NH_2	H	H	H	H	H	-0.23
4	71-43-2	H	H	H	H	H	H	-0.12
5	106-49-0	CH_3	H	H	NH_2	H	H	-0.05
6	106-47-8	NH_2	H	H	Cl	H	H	0.05
7	108-42-9	NH_2	H	Cl	H	H	H	0.22
8	87-60-5	NH_2	CH_3	Cl	H	H	H	0.38
9	95-50-1	Cl	Cl	H	H	H	H	0.53
10	626-43-7	NH_2	H	Cl	H	Cl	H	0.71
11	573-56-8	OH	NO_2	H	H	H	NO_2	0.83
12	329-71-5	OH	NO_2	H	H	NO_2	H	1.40
13	100-25-4	NO_2	H	H	NO_2	H	H	1.30

*CAS: Chemical Abstracts Service Registry Number

Quantitative Structure-Activity Relationships Analysis

One of the current interests in medicinal chemistry, environmental sciences and toxicology is the ranking of chemical substances with respect to their potential hazardous effects on humans, wild life and aqutic flora. The European legislation REACH specifies the use of QSAR models to predict properties of industrial chemicals. The toxicological-based quantitative structure-activity relationships (QSARs) have increasing interest in the using as non-animal methods to provide data for priority setting, risk assessment and chemical classification and labeling.

QSAR has great advantages over both experimental techniques and other computational methods. Since QSAR is a purely computational method that does not require the use of expensive equipment or hazardous chemicals and being computationally inexpensive. The fundamental hypothesis of the QSAR methodology is that the structure of a molecule must contain the feauters responsible for its physical, chemical, and biological properties and on the possibility of representing a molecule by numerical descriptors.

In this study, QSAR is developed using a variety of parameters as descriptors of structural properties of molecules. These physical properties descriptors, which capture electronic, lipophilic and steric characteristics of a molecule, are determined by computational methods. Activities used in QSAR are toxicity values.

Calculations of Molecular Descriptors

In this work, initial 3D molecular structures of chemicals were searched from SciFinder database. All geometries were minimized and calculated for atomic charge by using SYBYL 7.0 package and save as mol2 files. Then, physical properties descriptors of compounds were calculated by using MOE (Molecular Operating Environment) program. MOE is a software system designed by the Chemical Computing Group (http://www.chemcomp.com) that can read most common file formats (sdf, SMILE, pdb, mol2) as well as the internal .moe file type. In addition, the calculations are empirical, and so, generally, are fast. This program can calculate physical properties descriptors such as following:

1. Lipophilicity (Log P: the logarithm of the octanol-water partition coefficient)

One of the most important physicochemical properties much interest in QSAR studies is lipophilicity (or hydrophobicity). Because it directly relates to solubility in aqueous phase, to membrane permeation (an important factor contributing to the toxicity of chemicals), and to its contribution to ligand binding at the receptor site. Therefore, it is also a measure of a compound's tendency to bioaccumulate in the environment and has even been used as an estimate of a compound's toxicity.

Lipophilicity is defined by the partition of a compound between an aqueous and a nonaqueous phase. The most widely used lipophilic descriptor is the octanol-water partition coefficient (Log P) proposed by Hansch. P is a quotient between solubility in octanol and water. It is defined by following equation:

$$P = \frac{[\text{compound}]_{\text{octanol}}}{[\text{compound}]_{\text{water}}} \quad (1)$$

Hydrophobic molecules will prefer to dissolve in the octanol layer, while hydrophilic molecules will prefer to dissolve in the aqueous layer. Thus, hydrophobic compounds will have a high P value, whereas hydrophilic compounds will have a low P values

Log P represents the over all hydrophobicity of a molecule, which can be precisely measured. On the other hand, it can be easily calculated by contribution that includes the sum of the hydrophobic contributions of the "parent" molecule and its substituent. This contribution is known as the substituent hydrophobicity constant (Π).

Partition coefficients are measured for a standard compound with and without a substituent (X). The hydrophobicity constant (Π_x) for the substituent (X) is then obtained using the following:

$$\Pi_x = \log P_{R\text{-}X} - \log P_{R\text{-}H} \quad (2)$$

where $P_{R\text{-}H}$ is the coefficient for the parent molecule, and $P_{R\text{-}X}$ is the partition coefficient for the derivative molecule.

A positive value of Π indicates that the substituent is more hydrophobic than hydrogen. On the other hand, a negative value indicates that the substituent is less hydrophobic. These Π values are characteristic for the substituents and can be used to calculate how the partition coefficient of compound would be affected by adding these substituents.

2. Molecular Refractivity (MR)

The measure of the steric factor is provided by a parameter known as molar refractivity (MR). This parameter is a measure of the volume occupied by an atom or group of atoms. The molar refractivity is a constitutive-additive property (like Log P parameter) that is calculated by Lorenz-Lorentz formular:

$$MR = \frac{n^2 - 1}{n^2 + 2} \times \frac{Mw}{\rho} \quad (3)$$

when n is the refraction index, Mw is the molecular weight, and ρ is the density. The $(n^2 - 1)/(n^2 + 2)$ term provides a correction factor by defining how easily the substituent can be poralized, whereas the Mw/ρ term defines a volume. Molar refractivity is related to the lipophilicity, volume, and steric of the molecules. Moreover, it has been correlated with the London dispersive force that acts in the drug-receptor interaction. Due to its Mw/ρ component, it is indeed related to volume and size of a substituent. MR value depends on the polar part of molecule. Because of the refractive index-related correction term in MR accounts for the polarizability and size and polarity of a certain group. The MR value will be large when the polar part of molecule is large.

3. Molecular Polarizability (apol: Sum of the atomic polarizabilities)

Polarizability of an atom or molecule is the relative tendency of a charge distribution, like the electron cloud, to be distorted from its normal shape by an external electric field, which may be caused by the presence of a nearby ion or dipole. It is widely used to describe the inductive and dispersive interaction of a molecule or molecular system. In addition, polarizability values have been shown to be related to hydrophobicity and thus to other biological activities. It is one of the descriptors that are extensively used in QSAR study.

Molecular polarizability can be expressed as sum of atomic polarizabilities, plus correlations depending on the types of bonds present. Moreover, it can also be expressed approximately as sums of bond polarizabilities. On the other hand, the polarizability of a molecule can be obtained by summing up the contributions of a variety of atoms and/or fuctional groups in the molecule.

Highly polarizable molecules can be expected to have strong attractions with other molecules. The molecular polarizabity (α) is also related to molecular refractivity (MR) by the Lorentz-Lorenz equation 4 where N_0 is the Avogadro constant.

$$MR = \frac{n^2 - 1}{n^2 + 2} \times \frac{Mw}{\rho} = \frac{4}{3}\pi N_0 \alpha \quad (4)$$

Statistical Analysis for QSAR Analysis

Toxicity QSAR uses log-based toxicological data and molecular descriptor data for linking together by a statistical method which the most common use the multiple linear regression (MLR) method. In addition, MLR was first used to set up the QSAR equations and it is still the most widely used statistical tool in QSAR today. Equation 5 describes a linear model containing X_k variable, which is simplify for considering the multiple linear regression.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \quad (5)$$

where Y is dependent variable (*e.g.* biological data or experimental data) that represents one of the characteristics of the phenomenon studied. β_0 is the regression constant obtained from the fit. β_1 to β_k are coefficients to be estimated. X_1 to X_k are independent variables or predictor or explanatory variables (*e.g.* physicochemical parameters).

Certain assumptions are made with regrade to this procedure:

1. The independent variables (X), including the physical parameters, are measured without error. Unfortunately, this is not always the case, although the error in these variables is small compared to that in the dependent variable (Y).

2. For any given value of X, the Y values are independent and follow a normal distribution. The error possesses a normal distribution with a mean of zero.

3. The expected mean value for the variable Y, for all values of X, lies on a straight line.

4. The variance around the regression line is constant. The “best” straight line for model is drawn through the data points.

The objective was a model certaining as few parameters and variables as possible, but still describing the toxicity correctly within the limits set by physiological variation of chemical data. Using substituent benzene derivatives, multiple linear regression models were develop based on regression algorithms in the SPSS for windows package. The quality of the model was considered as statistically satistically on the basis of multiple correlation coefficients (r), standard deviation (s), and F-test value (F) which define as

$$r^2 = 1 - \frac{SSQ}{SS_T} \tag{6}$$

$$s = \sqrt{\frac{SSQ}{n-k-1}} \tag{7}$$

$$F = \frac{r^2 \cdot (n-k-1)}{k(1-r^2)} \tag{8}$$

where *k* and *n* are number of variables and observations, respectively. SSQ is the prediction error sum of squares or unexplained variance:

$$SSQ = \sum (Y_{obs} - Y_{calc})^2 \quad (9)$$

and SS_T is the sum of squares of deviation between the affinities of the fitting set and their mean affinity:

$$SS_T = \sum (Y_{obs} - Y_{mean})^2 \quad (10)$$
$$= \sum y^2 - (\sum y)^2/n$$

where Y_{obs}, Y_{mean} and Y_{calc} are observed, mean and calculated values of the target property, respectively.

The over all step of QSAR analysis in this study are summarized in Figure 8.

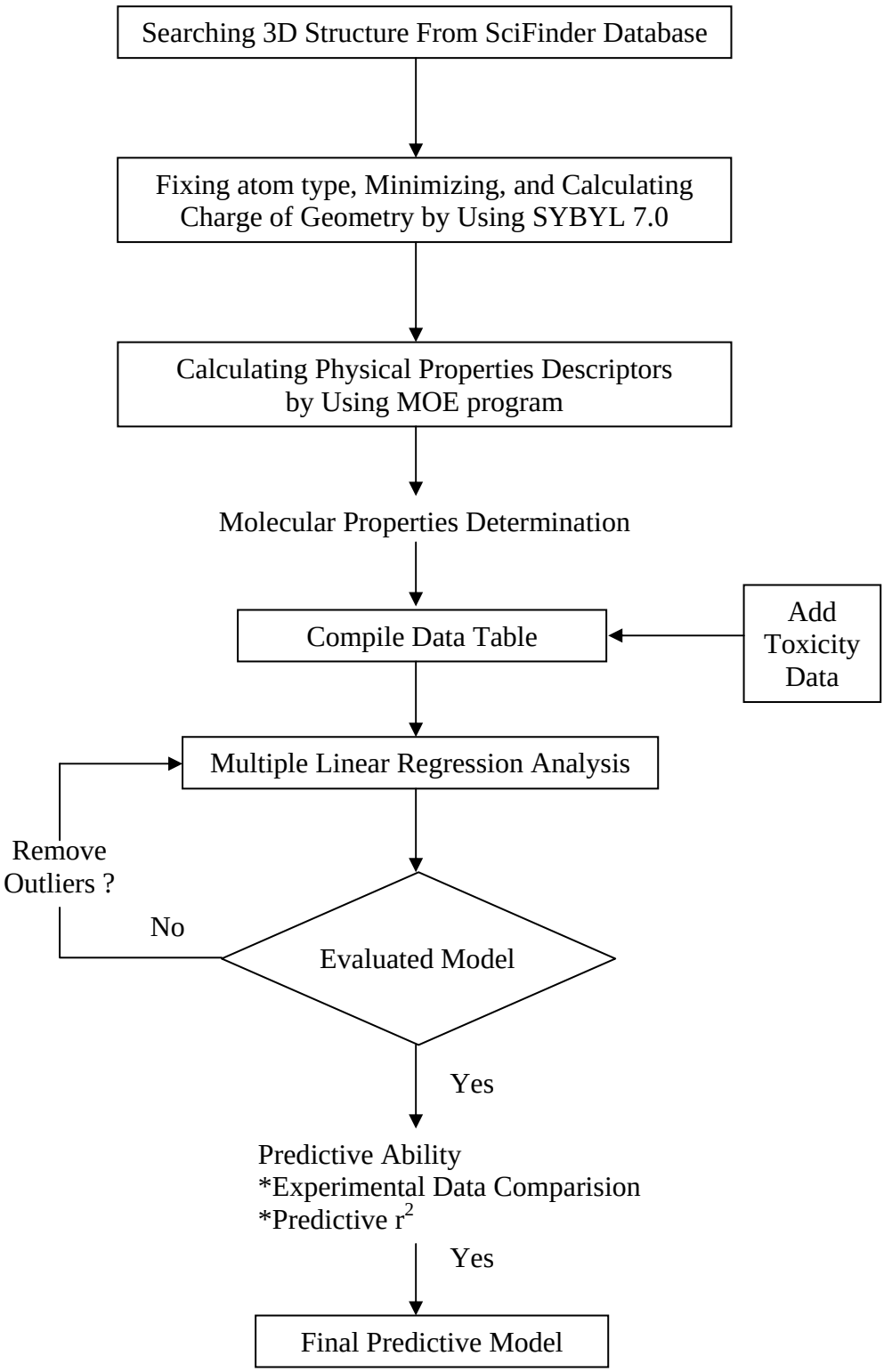


Figure 8 Flow chart of QSAR approach in this study.

RESULTS AND DISCUSSION

Aquatic Toxicity QSAR Model of Substituted Benzene Derivatives Against *Tetrahymena pyriformis*

The 37 observed benzene derivatives toxicity values to *Tetrahymena pyriformis,* are shown in Table 1 together with their chemical structures. These structures were used to set up QSAR model and defined as training set. The QSAR model was constructed by using multiple linear regression analysis. The data analysis was performed by the SPSS programme which process for building up a model through stepwise addition of descriptors. In a stepwise manner, F statistic values are used in automated algorithms to derive a regression model (strating with the best single variable and adding further significant variable; forward selection). This series of steps has effect for finding the equations and the models based on combinations of descriptors with high coefficient of determination r^2 were generated as an output.

All possible combinations of molecular descriptors were considered to construct the QSAR model. These molecular descriptors used in this study are molecular weight (Mw), sum of the atomic polarizabilities (apol) and partition coefficient (log P). The quality of model is shown by r^2, s, and F. From correlation equation, n is the number of chemicals employed to set up model, r^2 is the squared correlation coefficient, S is the standard deviation, and F indicates overall F statistics for the addition of each successive molecular descriptor and values in parentheses are the 95% confidence limit for each coefficient. The calculated molecular descriptors are shown in Table 3.

Table 3 Molecular descriptors (Mw, apol, and log P) of the training set obtained from MOE calculations.

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Mw	apol	log P
1	98-85-1	CH_2CH_3 OH	H	H	H	H	H	122.17	21.55	1.85
2	87-62-7	NH_2	CH_3	H	H	H	CH_3	121.18	22.51	1.83
3	108-95-2	OH	H	H	H	H	H	94.11	15.36	1.60
4	95-48-7	OH	CH_3	H	H	H	H	108.14	18.46	1.89
5	95-68-1	NH_2	CH_3	H	CH_3	H	H	121.18	22.51	1.86
6	108-44-1	NH_2	H	CH_3	H	H	H	107.16	19.42	1.57
7	100-52-7	CH=O	H	H	H	H	H	106.12	17.12	1.83
8	95-51-2	NH_2	Cl	H	H	H	H	127.57	17.84	1.82
9	95-53-4	NH_2	CH_3	H	H	H	H	107.16	19.42	1.53
10	106-44-5	OH	H	H	CH_3	H	H	108.14	18.46	1.90
11	108-90-7	Cl	H	H	H	H	H	112.56	16.07	2.50
12	91-23-6	OCH_3	NO_2	H	H	H	H	153.14	20.49	1.80
13	88-05-1	NH_2	CH_3	H	CH_3	H	CH_3	135.21	25.61	2.20
14	99-09-2	NH_2	H	NO_2	H	H	H	138.13	18.36	1.21
15	100-44-7	CH_2Cl	H	H	H	H	H	126.59	19.17	2.63
16	103-69-5	$NHCH_2CH_3$	H	H	H	H	H	121.18	22.51	1.89
17	88-74-4	NH_2	NO_2	H	H	H	H	138.13	18.36	1.17
18	98-95-3	NO_2	H	H	H	H	H	123.11	16.60	1.84

Table 3 (Continued)

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Mw	apol	log P
19	123-07-9	OH	H	H	CH_2CH_3	H	H	122.17	21.55	2.37
20	108-88-3	CH_3	H	H	H	H	H	92.14	17.65	2.37
21	88-72-2	CH_3	NO_2	H	H	H	H	137.12	19.69	2.14
22	121-89-1	$COCH_3$	H	NO_2	H	H	H	165.15	22.25	1.73
23	95-74-9	NH_2	H	Cl	CH_3	H	H	141.60	20.93	2.16
24	95-79-4	NH_2	CH_3	H	H	Cl	H	141.60	20.93	2.16
25	106-48-9	OH	H	H	Cl	H	H	128.56	16.88	2.19
26	95-82-9	NH_2	Cl	H	H	Cl	H	162.02	19.35	2.45
27	99-99-0	CH_3	H	H	NO_2	H	H	137.14	19.69	2.14
28	88-73-3	Cl	NO_2	H	H	H	H	157.56	18.11	2.43
29	97-02-9	NH_2	NO_2	H	NO_2	H	H	183.12	20.40	1.14
30	121-73-3	Cl	H	NO_2	H	H	H	157.56	18.11	2.47
31	121-87-9	NH_2	Cl	H	NO_2	H	H	172.57	19.88	1.80
32	99-65-0	NO_2	H	NO_2	H	H	H	168.11	18.64	1.81
33	108-70-3	Cl	H	Cl	H	Cl	H	181.45	19.10	3.79
34	121-14-2	CH_3	NO_2	H	NO_2	H	H	182.14	21.73	2.11
35	51-28-5	OH	NO_2	H	NO_2	H	H	184.11	19.44	1.50
36	99-54-7	NO_2	H	Cl	Cl	H	H	192.00	19.62	3.06
37	97-00-7	Cl	NO_2	H	NO_2	H	H	202.55	20.15	2.40

*CAS: Chemical Abstracts Service Registry Number

For this work, the log (IGC^{-1}_{50}) models (the negative logarithm of 50% growth inhibitory concentration to *Tetrahymena pyriformis*) of a training set of 37substituted benzene derivatives were obtained as shown in equation 11 and 12, respectively.

$$\log (IGC^{-1}_{50}) = -1.958(\pm 0.508) + 0.016(\pm 0.004)Mw \quad (11)$$

$(n = 37, r^2 = 0.706, s = 0.306, F = 84.088)$

$$\log (IGC^{-1}_{50}) = -0.999(\pm 0.951) + 0.017(\pm 0.003)Mw - 0.056(\pm 0.048)apol \quad (12)$$

$(n = 37, r^2 = 0.748, s = 0.288, F = 50.414)$

With regard to the relative importance of various statistical criteria, the aquatic toxicity equations (log (IGC^{-1}_{50})) derived from stepwise addition of descriptors are revealed that the quality of QSAR model in equation 12 is higher than QSAR model in equation 11 and the addition of sum of the atomic polarizabilities (apol) can improve the quality of the fit. The predictive ability of both models are satisfied since the squared correlation coefficients (r^2) are higher than 0.6 (Afantitis *et al.*, 2006). However, the squared correlation coefficients (r^2) of the use of one-term and two-term models are low.

In previous works, Schultz *et al.* (2003) analysed the toxicity of aromatic compounds on *Tetrahymena pyriformis* by using QSAR approach. They summarized that the regression-based QSAR of aromatic compounds (benzenes) were developed based on the hydrophobicity (log K_{ow}) and electrophilicity. The acute lethal toxicity QSAR of benzene derivatives to *Rana japonica* tadpoles was determined by Huang *et al.* (2003). Their model was contained the variables reflecting hydrophobicity (log Kow), electronic property (E_{lumo}), and molecular size, respectively. From these lead to the way for developing the QSAR model in this work by adding partition coefficient (log P), van der Waals volume (vdw_vol) and energy of the lowest unoocupied molecular orbital (E_{LUMO}). Due to log P indicates the hydrophobicity of a compound which directly relates to the partitioning of the compound into the membrane of

organism. In addition, it is an important factor contributing to the toxicity of compounds (Donlon *et al.*, 1995). For van der Waals volume (vdw_vol) and energy of the lowest unoccupied molecular orbital (E_{LUMO}), they relate to toxicity of benzene derivatives to *Rana japonica* tadpoles as previous work studied by Huang *et al.* (2003). Van der Waals volume is parameter accounting for the size or bluk of a molecule or a substituent (Khadikar *et al.*, 2002). E_{LUMO} is an electrophilicity parameter that characterizes the susceptibility of chemicals towards nucleophilic attack (Schultz and Deweese, 1999).

The log (IGC^{-1}_{50}) in equation 12 can be improven by adding hydrophobicity parameter (log P) as shown in equation 13. An addition of hydrophobicity term to equation 12 as presented in equation 13, produced better statistical results. Therefore, the final model is satisfied by r^2 equals to 0.773 (77%). Additionally, the mutual correlations between the independent variables (Mw, apol, and log P) are not hight as shown in Table 4. The use of more than three independent variables (molecular descriptors) can not provided significant log (IGC^{-1}_{50}) model. Since r^2 of four-term model has a little different from r^2 of three-term model. Additional, the mutual correlations between the independent variables (Mw, apol, log P, vdw_vol, and E_{LUMO}) of the log (IGC^{-1}_{50}) model in equation 14 are very hight as shown in Table 5. Therefore, the suitable QSAR model in this study is the model in equation 13. This model based on three molecular descriptors (logP, apol, and Mw) although molecular descriptors from MOE program have more than 30 types as shown in Appendix II. The reason for this is the mutual correlations between molecular descriptors that are not hight (Table 4). From equation 13, the lipophilicity (log P) seems to be the major parameter to the toxicity of substituted benzene derivatives. The positive coefficients of logarithm of octanol-water partition coefficient (log P) and weight indicate that the aquatic toxicity to *Tetrahymena pyriformis* increase with rise in hydrophobicity and weight of molecular chemical, while the negative coefficient of sum of the atomic polarizabilities (apol) implies the decrease of toxicity with increasing molecular polarizability.

$$\log (IGC^{-1}_{50}) = -1.308(\pm 0.974) + 0.016(\pm 0.003)Mw - 0.053(\pm 0.046)apol + 0.178(\pm 0.190)\log P \quad (13)$$

($n = 37$, $r^2 = 0.773$, $s = 0.277$, $F = 37.439$)

$$\log (IGC^{-1}_{50}) = -1.341(\pm 1.007) + 0.016(\pm 0.004)Mw - 0.061(\pm 0.065)apol + 0.188(\pm 0.201)\log P + 0.001(\pm 0.007)vdw\_vol \quad (14)$$

($n = 37$, $r^2 = 0.774$, $s = 0.281$, $F = 27.356$)

$$\log (IGC^{-1}_{50}) = -1.467(\pm 0.983) + 0.013(\pm 0.005)Mw - 0.164(\pm 0.133)apol + 0.160(\pm 0.198)\log P + 0.016(\pm 0.018)vdw\_vol + 0.055(\pm 0.061)E_{LUMO} \quad (15)$$

($n = 37$, $r^2 = 0.796$, $s = 0.271$, $F = 24.140$)

Table 4 The mutual correlations between independent variables (Mw, apol, and log P).

correlations	log P	apol	Mw
log P	1.000	0.074	-0.258
apol		1.000	-0.257
Mw			1.000

Table 5 The mutual correlations between independent variables (Mw, apol, log P, vdw_vol, and E_{LUMO}).

correlations	log P	apol	Mw	vdw_vol	E_{LUMO}
log P	1.000	0.070	-0.166	-0.036	-0.162
apol		1.000	0.642	-0.938	-0.877
Mw			1.000	-0.750	-0.649
vdw_vol				1.000	0.919
E_{LUMO}					1.000

Prediction of Model

The best log (IGC^{-1}_{50}) model, equation 13, was then used to predict the aquatic toxicity of the compounds in training set and test set (37 and 13 substituted benzene derivatives, respectively). Molecular descriptors of 13 compounds in the test set for utilizing to predict the toxicity values are listed in the Table 6. The difference between predicted and observed toxicities, is call residual. The protozoa toxicity to *T. pyriformis* are listed in Table 7 for training set and Table 8 for test set, respectively. The plot of observed toxicity versus predicted toxicity for training and test set are shown in Figure 9 and 10, respectively.

Based on the residual value of training set and test set, the results indicated that model 13 can be useful to predict the toxicity of substituent bezene derivatives. Due to the both data sets do not have residual higher than 1.00 and only three compounds exceed 0.50. The largest residual, 0.87, is recorded for compound 37 in the training set; the second is 0.60 for compound 20 in the training set; and the third is 0.59 for compound 13 in the test set. Compoud 37 shows the highest residual, since it has a high molecular polarizability value (20.1484) and the range of toxicity value rather far from other compounds. Therefore, considering the plot of predicted versus observed values as shown in Figure 9, this compound seems to be an outlier. The other reason for compound 37 in the training set and 13 in the test set having large residual are substituent as nitro group with to aromatic ring high molecular polarizability. For compound 20 in the training, this compound has a big residual. The reason for this is due to a methyl group attached to the benzene ring that shows a low molecular weight. In addition, these three compounds are different structure from most of other compounds in the series.

Table 6 Molecular weight (Mw), sum of atomic polarizabilities (apol), and lipophilicity (log P) values of the test set.

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Mw	apol	log P
1	108-46-3	OH	H	OH	H	H	H	110.11	16.16	1.33
2	95-78-3	NH_2	CH_3	H	H	CH_3	H	121.18	22.51	1.86
3	62-53-3	NH_2	H	H	H	H	H	93.13	16.33	1.23
4	71-43-2	H	H	H	H	H	H	78.11	14.56	1.91
5	106-49-0	CH_3	H	H	NH_2	H	H	107.16	19.42	1.53
6	106-47-8	NH_2	H	H	Cl	H	H	127.57	17.84	1.83
7	108-42-9	NH_2	H	Cl	H	H	H	127.57	17.84	1.86
8	87-60-5	NH_2	CH_3	Cl	H	H	H	141.60	20.93	2.12
9	95-50-1	Cl	Cl	H	H	H	H	147.00	17.59	3.09
10	626-43-7	NH_2	H	Cl	H	Cl	H	162.02	19.35	2.53
11	573-56-8	OH	NO_2	H	H	H	NO_2	184.11	19.44	1.46
12	329-71-5	OH	NO_2	H	H	NO_2	H	184.11	19.44	1.50
13	100-25-4	NO_2	H	H	NO_2	H	H	168.11	18.64	1.78

*CAS: Chemical Abstracts Service Registry Number

Table 7 Observed, predicted, and residual aquatic toxicity values of benzene derivatives againt *T. pyriformis*, used in the training set.

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Observed log (IGC^{-1}_{50})	Predicted log (IGC^{-1}_{50})	residual
1	98-85-1	CH_2CH_3 OH	H	H	H	H	H	-0.66	-0.17	-0.49
2	87-62-7	NH_2	CH_3	H	H	H	CH_3	-0.43	-0.24	-0.19
3	108-95-2	OH	H	H	H	H	H	-0.35	-0.33	-0.02
4	95-48-7	OH	CH_3	H	H	H	H	-0.29	-0.22	-0.07
5	95-68-1	NH_2	CH_3	H	CH_3	H	H	-0.29	-0.23	-0.06
6	108-44-1	NH_2	H	CH_3	H	H	H	-0.28	-0.34	0.06
7	100-52-7	CH=O	H	H	H	H	H	-0.20	-0.19	-0.01
8	95-51-2	NH_2	Cl	H	H	H	H	-0.17	0.11	-0.28
9	95-53-4	NH_2	CH_3	H	H	H	H	-0.16	-0.35	0.19
10	106-44-5	OH	H	H	CH_3	H	H	-0.16	-0.22	0.06
11	108-90-7	Cl	H	H	H	H	H	-0.13	0.09	-0.22
12	91-23-6	OCH_3	NO_2	H	H	H	H	-0.07	0.38	-0.45
13	88-05-1	NH_2	CH_3	H	CH_3	H	CH_3	-0.05	-0.11	0.06
14	99-09-2	NH_2	H	NO_2	H	H	H	0.03	0.14	-0.11
15	100-44-7	CH_2Cl	H	H	H	H	H	0.06	0.17	-0.11
16	103-69-5	$NHCH_2CH_3$	H	H	H	H	H	0.07	-0.23	0.30
17	88-74-4	NH_2	NO_2	H	H	H	H	0.08	0.14	-0.06
18	98-95-3	NO_2	H	H	H	H	H	0.14	0.11	0.03

Table 7 (Continued)

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Observed log (IGC^{-1}_{50})	Predicted log (IGC^{-1}_{50})	residual
19	123-07-9	OH	H	H	CH_2CH_3	H	H	0.21	-0.07	0.28
20	108-88-3	CH_3	H	H	H	H	H	0.25	-0.35	0.60
21	88-72-2	CH_3	NO_2	H	H	H	H	0.26	0.22	0.04
22	121-89-1	$COCH_3$	H	NO_2	H	H	H	0.32	0.46	-0.14
23	95-74-9	NH_2	H	Cl	CH_3	H	H	0.39	0.23	0.16
24	95-79-4	NH_2	CH_3	H	H	Cl	H	0.5	0.23	0.27
25	106-48-9	OH	H	H	Cl	H	H	0.54	0.24	0.30
26	95-82-9	NH_2	Cl	H	H	Cl	H	0.58	0.70	-0.12
27	99-99-0	CH_3	H	H	NO_2	H	H	0.65	0.22	0.43
28	88-73-3	Cl	NO_2	H	H	H	H	0.68	0.69	-0.01
29	97-02-9	NH_2	NO_2	H	NO_2	H	H	0.72	0.74	-0.02
30	121-73-3	Cl	H	NO_2	H	H	H	0.73	0.69	0.04
31	121-87-9	NH_2	Cl	H	NO_2	H	H	0.75	0.72	0.03
32	99-65-0	NO_2	H	NO_2	H	H	H	0.76	0.72	0.04
33	108-70-3	Cl	H	Cl	H	Cl	H	0.87	1.26	-0.39
34	121-14-2	CH_3	NO_2	H	NO_2	H	H	0.87	0.83	0.04
35	51-28-5	OH	NO_2	H	NO_2	H	H	1.06	0.88	0.18
36	99-54-7	NO_2	H	Cl	Cl	H	H	1.16	1.27	-0.11
37	97-00-7	Cl	NO_2	H	NO_2	H	H	2.16	1.29	0.87

*CAS: Chemical Abstracts Service Registry Number

Table 8 Observed, predicted, and residual aquatic toxicity values of benzene derivatives againt *T. pyriformis*, used in the test set.

No	CAS*	R_1	R_2	R_3	R_4	R_5	R_6	Observed log (IGC^{-1}_{50})	Predicted log (IGC^{-1}_{50})	residual
1	108-46-3	OH	H	OH	H	H	H	-0.65	-0.17	-0.48
2	95-78-3	NH_2	CH_3	H	H	CH_3	H	-0.33	-0.23	-0.10
3	62-53-3	NH_2	H	H	H	H	H	-0.23	-0.46	0.23
4	71-43-2	H	H	H	H	H	H	-0.12	-0.49	0.37
5	106-49-0	CH_3	H	H	NH_2	H	H	-0.05	-0.35	0.30
6	106-47-8	NH_2	H	H	Cl	H	H	0.05	0.11	-0.06
7	108-42-9	NH_2	H	Cl	H	H	H	0.22	0.12	0.10
8	87-60-5	NH_2	CH_3	Cl	H	H	H	0.38	0.23	0.15
9	95-50-1	Cl	Cl	H	H	H	H	0.53	0.66	-0.13
10	626-43-7	NH_2	H	Cl	H	Cl	H	0.71	0.71	0.00
11	573-56-8	OH	NO_2	H	H	H	NO_2	0.83	0.87	-0.04
12	329-71-5	OH	NO_2	H	H	NO_2	H	1.40	0.88	0.16
13	100-25-4	NO_2	H	H	NO_2	H	H	1.30	0.71	0.59

*CAS: Chemical Abstracts Service Registry Number

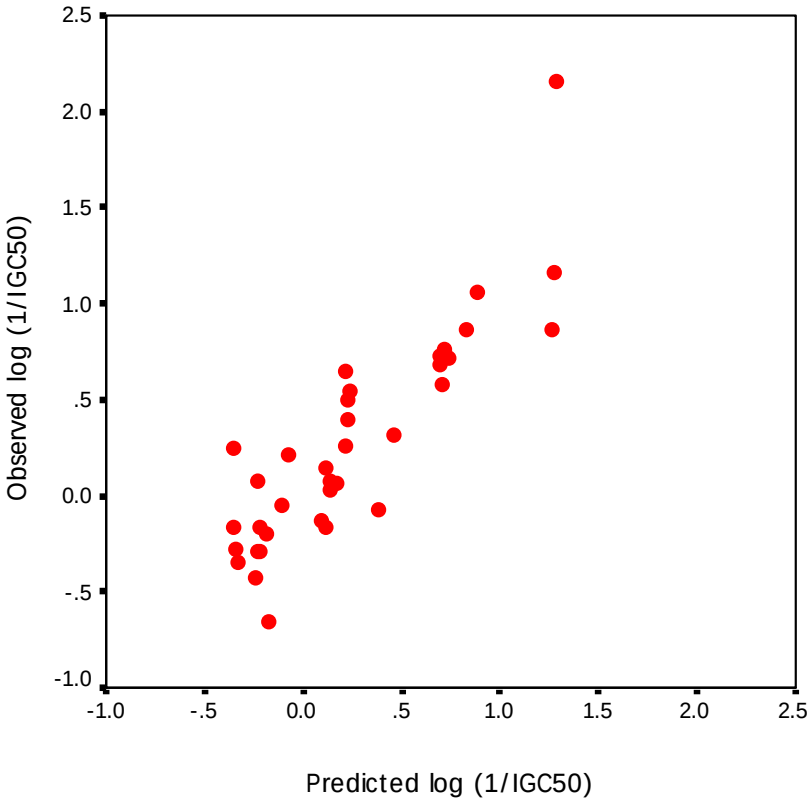


Figure 9 Plot of observed log (IGC^{-1}_{50}) values versus predicted log (IGC^{-1}_{50}) for 37 benzene derivatives: training set.

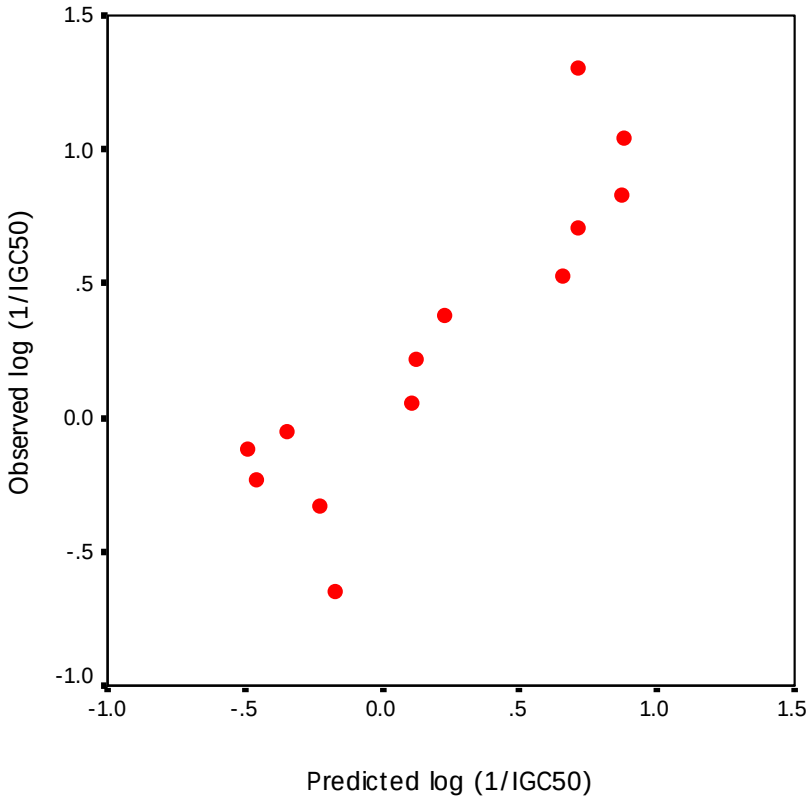


Figure 10 Plot of observed log (IGC^{-1}_{50}) values versus predicted log (IGC^{-1}_{50}) for 13 benzene derivatives: test set.

CONCLUSION

In this study, aquatic toxicity of substituted benzene derivatives to *Tetrahymena pyriformis* was investigated by using QSAR method. The study could be summarized as the following:

(1) The multiple linear regression process for setting up a model through stepwise addition of descriptors is a power tool for building QSAR model. Due to it can help to select significant independent variables (molecular descriptors) that correlate to dependent variables (toxicity values) by appropriate statistical procedure.

(2) The suitable QSAR model based on lipophilicity (log P), sum of the atomic polarizabilities (apol), and molecular weight (Mw). The formulation of this model is log (IGC^{-1}_{50}) = –1.308(±0.974) + 0.016(±0.003)Mw – 0.053(±0.046)apol + 0.178(±0.190)log P, yielding statistics: r^2 = 0.773, s = 0.277, and F = 37.439. The major parameter affect to toxicity of benzene derivatives is log P that represents the partitioning of each toxicant into biophase and may have a role in the toxicity. Additionally, the mutual correlations between three independent variables are not hight.

(3) The QSAR model was successfully applied to predict the toxicity of substituted benzene derivatives against *Tetrahymena pyriformis*. Since the predicted values are very close to the experimental ones that can see from the residuals lower than 1.00.

LITERATURE CITED

Abrahart, E.N. 1968. **Dye and their Intermediates**. Pergamon Press Ltd., Headington Hill Hall, Oxford.

Afantitis, A., G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos and O. Igglessi-Markopoulou. 2006. Prediction of intrinsic viscosity in polymer-solvent combinations using a QSPR model. **Polymer**. 47:3240-3248.

Anliker, R. 1977. Color chemistry and the environment. **Ecotoxicology and Environment Safety**. 1: 211-237.

Bae, J.-S. and H.S. Freeman. 2007. Aquatic toxicity evaluation of copper-complexed direct dyes to the *Daphnia magna*. **Dyes and Pigments**. 73: 126-132.

Botsford, J.L., T. Hillaker, B. Robertson, M. Gonzales, M. Benavidez, B. Jones, R. Baker, W. Steen, F. Pacheco, V. Homer, O. Lucero, M. Matthews and V. Koehler. n.d. **A Simple, Rapid, Inexpensive Assay for Toxic Chemicals Using a Bacterial Indicator**. Department of Biology, New Mexico State University.

Costescu, A. and M.V. Diudea. 2006. QSTR study on aquatic toxicity against *Poecilia Reticulate* and *Tetrahymena pyriformis* using topological indices. **Internet Electronic Journal of Molecular Design**. 5: 116-134.

David, A. 2003. **Dyes and Dyeing**. Chymist.com. Available Source: http://www.chymist.com/Dyes. Pdf, Febuary 2, 2007.

Debnath, A.K., R.L. Lopez de Compadre, G. Debnath, A.J. Shusterman and C. Hansch. 1991. Structure-Activity Relationship of mutagenic aromatic and hetroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity. **Journal of Medicinal Chemistry**. 34: 786-797.

Debnath, A.K. and C. Hansch. 1993. The importance of hydrophobicity in the mutagenicity of methanesulfonic acid esters with *Salmonella typhimurium* TA100. **Chemical Research in Toxicology**. 6: 310-312.

Estrada, E., G. Patlewicz, M. Chamberlain, D. Basketter and S. Larbey. 3003. Computer-aided knowledge generation for understanding skin sensitization mechanisms: the TOPS-MODE approach. **Chemical Research in Toxicology**. 16: 1226-1235.

European Comission. 1996. **Use of (quantitative) structure activity relationships ((Q)SAR)**. Toxicology and Chemical Substances European Chemicals Bureau. Available Source: http://ecb.jrc.it/documents/TECHNICAL_ GUIDANCE_DOCUMENT/EDITION_1/tgdpart3.pdf, February 2, 2007.

_____. 2006. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18, December 2006. **Official Journal of the European Union**. L 136: 3-280.

Fujita, T. and C. Hansch. 1967. Analysis of the structure-activity relationship of the sulfonamide drugs using substituent constants. **Journal of Medicinal Chemistry**. 10: 991-1000.

Funar-Timofei, S., M.F.W. Fabain, M.G. Simu and T. Suzuki. 2006. Quantitative structure-retention relationships (QSRR) for chromatographic of disazo and trisazo 4.4'-diaminobenzeneanilide-based dyes. **Croatica Chemica Acta**. 79: 227-236.

Garg, A., K.L. Bhat and C.W. Bock. 2002. Mutagenicity of aminoazobenzene dyes and related structure: a QSAR/QSPR investigation. **Dyes and Pigments**.55: 35-52.

Giorgi, M.R., R. Carpignano and G. Crisponi. 1997. Structure optimization in a series of acid dyes for wool and nylon. **Dyes and Pigments**. 34: 1-12.

Greaves, A.J., J.H. Churchley, M.G. Hutchings, D.A.S. Phillips and J.A. Taylor. 2001. A chemometric approach to understanding the bioelimination of anionic, water-soluble dyes by a biomass using empirical and semi-empirical molecular descriptors. **Water Research**. 35: 1225-1239.

Hansch, C., E.W. Deutsch and R.N. Smith. 1965. The use of substituent constants and regression analysis in the study of enzymatic reaction mechanisms. **Journal of the American Chemical Society**. 87: 2738-3742.

_____ and A. Leo. 1979. **Substituent constants for correlation analysis in chemistry and biology**. John Wileyand Sons, Inc., New York.

_____ and A.R. Steward. 1964. The use of substituent constants in the analysis of the structure-activity relationship in penicillin derivatives. **Journal of Medicinal Chemistry**. 7: 691-694.

_____, _____, S.M. Anderson and D. Bentley. 1967. The parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. **Journal of Medicinal Chemical**. 11: 1-11.

He, Y.-Y., H.-Y. Liu, J.-Y. An, R. Han and L.-J. Jiang. 2000. Photodynamic action of hypocrellin Dyes: structure-activity relationships. **Dyes and Pigments**. 44: 63-67.

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

Iwasa, J., T. Fujita and C. Hansch. 1965. Substituent constants for aliphatic functions obtained from partition coefficients. **Journal of Medicinal Chemistry**. 8: 150-153.

Khadikar, P.V., S. Karmarkar, S. Singh and A. Shrivastava. 2002. Use of the PI index predicting toxicity of nitrobenzene derivatives. **Bioorganic and Medicinal Chemistry**. 10: 3163-3170.

Kumar, P.M., C. Karthikeyan, N.S. Moorthy and P. Trivedi. 2006. Quantitative Structure-Activity Relationships of selective antagonists of glucagon receptor using QuaSAR descriptors. **Chemical and Pharmaceutical Bulletin**. 54: 1586-1591.

Leo, A., C. Hansch and C. Church. 1969. Comparision of parameters currently used in the study of structure-activity relationship. **Journal of Medicinal Chemistry**. 12: 766-771.

Li, Y. and D.L. Xi. 2007. Quantitative structure-activity relationship study on the biodegradation of acid dye stuffs. **Journal of Environmental Sciences**. 19: 800-804.

Lin, A. n.d. QuaSAR: The MOE system for QSAR. **Journal of Chemical Computing Group**.

_____ n.d. QuaSAR-descriptor. **Journal of Chemical Computing Group**.

Lowell Center for Sustainable Production. 2006. **A brief introduction to the European Commission's regulatory proposal on Registration, Authorisation and Evaluation of Chemicals (REACH)**. Chemicals Policy Initiative. Avialable Source: http://www.chemicalspolicy.org/downloads/briefintroductionto.pdf, January 17, 2008.

Lessigiarska, I., P.A. Worth, T.I. Netzeva, J.C. Dearden and T.D. Cronin. 2006. Quantitative structure-activity-activity and quantitative structure-activity investigations of human and rodent toxicity. **Chemosphere**. 65: 1878-1887.

Opera, T.I., L. Kurunczi and S. Timofei. 1997. QSAR studies of disperse azo dyes. Towards the negation of the pharmacophore theory of dye-fiber interaction. **Dyes and Pigments**. 33: 41-64.

Papa, E., J.C. Dearden and P. Gramatica. 2007.Linear QSAR regression models for the prediction of bioconcentration factors by physiochemical properties and structural theoretical molecular descriptors. **Chemosphere**. 67: 351-358.

Paterson, D.J., L.A. Hurst and W.D. Comper. 2001. Variable toxicity of dyes used in micropuncture studies. **Nephrology**. 6:221-222.

Plum, J.J., J. Bell and D.C. Stuckey. 2001. Microbial population associated with treatment of an industrial dye effluent in an anaerobic baffled reactor. **Applied and Environmental Microbiology**. 67: 3226-3235.

Porcelli, C., E. Boriani, A. Roncaglioni, A. Chana and E. Benfenati. 2008. Regulatory perspectives in the use and validation of QSAR. A case study: DEMETRA model for *Daphnia* toxicity. **Environment Science and Technology**. 42: 491-496.

Rose, K. and L.H. Hall. 2003. E-state modeling of fish toxicity independent of 3D structure information. **SAR and QSAR in Environmental Research**. 14: 113-129.

Schultz, T.W. 1997. TETRATOX: Tetrahymena pyriformis population growth impairment endpoint-a surrogate for fish lethality. **Toxicology Methods**. 7: 289-309.

Schultz, T.W., A.D. Deweese. 1999. Structure-Toxicity relationships for seleted lactone to *Tetrahymena pyriformis*. **Bulletin of Environmental Contamination and Toxicology**. 62: 463-468.

_____, T.I. Netzeva and M.T.D. Cronin. 2003. Selection of data sets foe QSARs: analyses of *Tetrahymena* toxicity from aromatic compounds. **SAR and QSAR in Environmental Research**. 14: 59-81.

Schüürmann, G. and S. Funar-Timofei. 2003. Multilinear regression and comparative molecular field analysis (CoMFA) of azo dye-fiber affinities. 2. Inclusion of solution-phase molecular orbital descriptors. **Journal of Chemical Information and Computer Sciences**. 43: 1502-1512.

Selassie, C.D. 2003. **History of quantitative structure-activity relationships**. John Wiley and Sons, Inc., New York.

Søsted, H., D.A. Basketier, J.D. Johansen and G.Y. Patlewicz. Ranking of hair dye substances according to predicted senstitization potency: quantitative structure-activity relationships. **Contact Dermatitis**. 51: 241-254.

Sponza, D.T. 2006. Toxicity studies in a chemical dye production industry in turkey. **Journal of Hazardous Materials**. A138: 438-447.

Sztandera, L., A. Grag, S. Hayik, K.L. Bhat and C.W. Bock. 2003. Mutagenicity of Aminoazo dyes and their reductive-cleavage metabolites: a QSAR/QSPR investigation. **Dyes and Pigments**. 59: 117-133.

Timofei, S., W. Schmidt, L. Kurunczi and Z. Simon. 2000. A review of QSAR for dye affinity for cellulose fibres. **Dyes and Pigments**. 47: 5-16.

Venger, B.H., C. Hansch, J. Hatheway and U. Amrein. 1997. Ames test of 1-(X-Phenyl)-3,3-dialkyltriazenes. A quantitative structure-activity study. **Journal of Medicinal Chemistry**. 22: 473476.

Wang, Y., H. Liu, C. Zhao, H. Liu, Z. Cai and G. Jiang. 2005. Quantitative structure-activity relationship models for prediction of toxicity of polybrominated diphenyl ether congeners. **Environmental Science and Technology**. 39: 4961-4966.

Women's Environmental Network. 2003. **Chemicals under the sportlight: act now to protect The "reach" of European reforms**. Educating, empowering and men who care about the environment. Campaigning on environment and health issues from a female perspective. Available Source: http://www.wen.org.uk/Health/ Reports/cosmetics_spread2.pdf, January 28, 2007.

World Bank Group. 1998. **Dye manufacturing**. International Finance Corporation. Available Source: http://www.ifc.org/Ifcext/enviro.nsf/AttachmentsByTitle/ p_ppah/$FILE/PPAH.pdf, Febuary 14, 2007.

Zollinger, Heinrich. 1987. **Color chemistry: syntheses, properties and applications of organic dyes and pigment**. Weinheim: VCH.

APPENDICES

APPENDIX A

Appendix Table A1 The Physical Properties descriptors from MOE Program.

MOE-Code	Description
AM1_dipole	The dipole mement calculated using the AM1 Hamiltonian.
AM1_E	The total energy (kcal/mol) calculated using the AM1 Hamiltonian.
AM1_Eele	The electronic energy (kcal/mol) calculated using AM1 Hamiltonian.
AM1_HF	The heat of formation (kcal/mol) calculated using AM1 Hamiltonian.
AM1_IP	The ionization potential (kcal/mol) calculated using AM1 Hamiltonian.
AM1_LUMO	The energy (eV) of the Lowest Unoccupied Molecular Orbital calculated using the MOPAC AM1 Hamiltonian.
AM1_HOMO	The energy (eV) of the Highest Occupied Molecular Orbital calculated using the MOPAC AM1 Hamiltonian.
MNDO_dipole	The dipole mement calculated using the MNDO Hamiltonian.
MNDO_E	The total energy (kcal/mol) calculated using the MNDO Hamiltonian.
MNDO_Eele	The electronic energy (kcal/mol) calculated using MNDO Hamiltonian.
MNDO_HF	The heat of formation (kcal/mol) calculated using MNDO Hamiltonian.
MNDO_IP	The ionization potential (kcal/mol) calculated using MNDO Hamiltonian.
MNDO_LUMO	The energy (eV) of the Lowest Unoccupied Molecular Orbital calculated using the MNDO Hamiltonian.
MNDO_HOMO	The energy (eV) of the Highest Occupied Molecular Orbital calculated using the MNDO Hamiltonian.
PM3_dipole	The dipole mement calculated using the PM3 Hamiltonian.
PM3_E	The total energy (kcal/mol) calculated using the PM3 Hamiltonian.

Appendix Table A1 (Continued)

MOE-Code	Description
PM3_Eele	The electronic energy (kcal/mol) calculated using PM3 Hamiltonian.
PM3_HF	The heat of formation (kcal/mol) calculated using PM3 Hamiltonian.
PM3_IP	The ionization potential (kcal/mol) calculated using PM3 Hamiltonian.
PM3_LUMO	The energy (eV) of the Lowest Unoccupied Molecular Orbital calculated using the PM3 Hamiltonian.
PM3_HOMO	The energy (eV) of the Highest Occupied Molecular Orbital calculated using the PM3 Hamiltonian.
apol	Sum of the atomic polarizabilities (including implicit hydrogens).
bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens).
FCharge	Total charge of the molecule (sum of formal charges).
Reactive	Indicator of the presence of reactive groups. A non-zero value indicate that the molecule cantains a reactive group.
mr	Molecular refractivity (including implicit hydrogens).
SMR	Molecular refractivity (including implicit hydrogens).
Weight	Molecular weight (including implicit hydrogens).
logP(o/w)	Log of the octanol/water partition coefficient (including implicit hydrogens).
SlogP	Log of the octanol/water partition coefficient (including implicit hydrogens).
TPSA	Polar surface area calculated using group contributions to approximate the polar surface area from connection table information only.
Vdw_are	Area of Van der Waals surface calculated using a connection table approximation.
vdw_vol	Van der Waals volume calculated using a connection table approximation.
density	Molecular mass density: Weight divided by vdw_vol.

APPENDIX B

Oral Presentation

1. Waraporn Jungtanasombut, Patchareenart Suparpakorn, and Supa Hannongbua. **Quantitative Structure-Property Relationship Modeling of Polycyclic Aromatic Hydrocarbons.** The 11th Annual National Symposium on Computational Science and Engineering (ANSCSE 2007), Faculty of Science, Prince of Songkla University, Phuket Campus, Thailand, 28-30 March 2007.

Proceeding

1. Waraporn Jungtanasombut, Pornthip Poonsri, and Supa Hannongbua. **QSAR Study on Toxicity of Benzene Derivatives Against *Poecilia reticulate* and *Tetrahymena pyriformis* Using Molecular Descriptors.** Pure & Applied Chemistry Conference (PACCON 2008), Sofitel Central Hotel, Bangkok, Thailand, January 30- February 1, 2008.

Waraporn Jungtanasombut

Laboratory for Computational and Applied Chemistry
Department of Chemistry, Faculty of Science, Kasetsart University
50 Pahonyothin Road, Lat Yao Subdistrict, Jatujak District, Bangkok, THAILAND 10900.

PERSONAL DETAILS

Date of Birth:	25th May 1983
Place of Birth:	Ratchaburi, Thailand
Nationality:	Thai
Marital Status:	Single
Age:	25
Phone Number:	085-0314322
E-mail address:	g4864153@ku.ac.th, wj48641534@yahoo.com

EDUCATION

Kasetsart University, Bangkok, Thailand

- Ms in chemistry Chemistry 2005-Present

Maejo University, Thailand

- B.Sc. in Chemistry 2002-2005