Available online at www.sciencedirect.com
ScienceDirect
Journal of Environmental Sciences 19(2007) 800–804
JOURNAL OF ENVIRONMENTAL SCIENCES
ISSN 1001-0742
CN 11-2629/X
www.jesc.ac.cn

Quantitative structure-activity relationship study on the biodegradation of acid dyestuffs

LI Yin*, XI Dan-li

College of Environmental Science and Engineering, Dong Hua University, Shanghai 201620, China. E-mail: yinli@dhu.edu.cn

Received 18 September 2006; revised 21 November 2006; accepted 25 January 2007

Abstract

Quantitative structure-biodegradability relationships (QSBRs) were established to develop predictive models and mechanistic explanations for acid dyestuffs as well as biological activities. With a total of four descriptors, molecular weight (M_W), energies of the highest occupied molecular orbital (E_{HOMO}), the lowest unoccupied molecular orbital (E_{LUMO}), and the excited state (E_{ES}), calculated using quantum chemical semi-empirical methodology, a series of models were analyzed between the dye biodegradability and each descriptor. Results showed that E_{HOMO} and M_W were the dominant parameters controlling the biodegradability of acid dyes. A statistically robust QSBR model was developed for all studied dyes, with the combined application of E_{HOMO} and M_W. The calculated biodegradations fitted well with the experimental data monitored in a facultative-aerobic process, indicative of the reliable prediction and mechanistic character of the developed model.

Key words: quantitative structure-activity relationship (QSAR); acid dyestuff; biodegradability; decolorization

Introduction

Acid dyestuffs account for a large portion of the dyes used in textiles, leathers, papers, woods, and inks, with over 25000 t produced annually for these purposes in China (Yang and Yang, 2001). It has been recognized that brightly colored, water-soluble acid dyes are hardly biodegradable in the natural environment, resulting in problematic color pollution in water bodies (Venkataraman, 2000; Pagga and Brown, 1986; Brown and Laboureur, 1983). Therefore, the investigation of dye biodegradability seems to be essential, to prevent dye molecules from persisting in the aquatic environment, and will aid in efficiently decolorizing and degrading acid dye effluents by a biotreatment process.

With the development of computer science, computer-assisted modeling methods, relating chemical structure to qualitative biological activity and quantitative biological potency, have been applied to a diversity of problems. The basis for any quantitative structure activity relationship is that the biological activity of a new or untested chemical can be inferred from the molecular structure, or properties, of similar compounds whose activities have already been assessed (Luan *et al*., 2006; Schultz *et al*., 2003; Lu *et al*., 2001; Chen *et al*., 2000). In dye biodegradation, the outcome models are used to explain and represent the mechanism of dye molecule bioreduction, and most importantly, to predict the biodegradability associated with correlative dyes in the biological system.

The recent application, QSAR (quantitative structure-activity relationship) technique, to dye biodegradability is focusing on azo dyes (Mior *et al*., 2001; Oprea *et al*., 1997). A number of reports suggest that the biodegradabilities of azo dyes are governed by their properties of substituents in aromatic hydrocarbons, which in turn are determined by their chemical structures, such as, sites, numbers, and molecular weights of the substituents (Chen *et al*., 2003; Sun *et al*., 1999; Dai *et al*., 1998). There are several interrelationships between dye molecular structure and biodegradability or decolorization, however, models are still in short supply for those dyes that may possess congenic application function. This study is designed to develop predictive and mechanism-based QSBR (quantitative structure-biodegradability relationship) models for acid dye, in which the biodegradability of a dye is expressed as its decolorization efficiency, because the biodegradation effect predominates in the total decolorization (biodegradation and biosorption) after reaching biosorption equilibrium at a steady state operation.

Four descriptors reflecting the main characteristics of the dye molecule were employed in this study. Three quantum chemical descriptors, energies of the highest occupied molecular orbital (E_{HOMO}), the lowest unoccupied molecular orbital (E_{LUMO}), and the excited state (E_{ES}), were computed by PM3 Hamiltonian and the chemical descriptor, molecular weight (M_W), was also considered in the model's establishment. The relationship of dye biodegradability with each descriptor was investigated and

Project supported by the Natural Science Foundation of Shanghai, China (No. 06ZR14002). *Corresponding author. E-mail: yinli@dhu.edu.cn.

the impact of structure parameters on its biodegradability was discussed. Finally, the general QSBR modeling yield was tested and verified by the experimental decolorizations.

1 Materials and methods

1.1 Studied dyes

All studied dyes were provided by Anhui Fengyang Dyestuffs & Chemicals Co. Ltd., China. Twenty acid dyes were selected on the basis of availability and structural non-congener, to allow interpretation of differences in efficiencies of decolorization. The inclusion of such structural diversity presented a significant challenge for QSAR/QSBR development. The names and Color Index number of the dyes are listed in Table 1.

For comparison of the calculated and experimental results, each studied dye effluent was treated in a facultative-aerobic process. The operation of this biological treatment system and determination of the biodegradation degree and decolorization efficiency were described in the early report (Li and Xi, 2004).

1.2 Quantum chemical descriptors and statistical analysis

The energies of frontier orbitals, E_{HOMO} and E_{LUMO}, which directed the nucleophilic and electrophilic reactivity of a compound respectively, were calculated using the PM3 Hamiltonian contained in the quantum chemical computation software HYPER (version 5.0) (Stewart, 1989). This program allowed one to draw any structure of interest, to create the input file for the calculations, and to perform the actual calculations of interest via simple pull-down menu options. The method of PM3 was selected because it was a recently developed semi-empirical molecular orbital algorithm and the computational time was much shorter than needed by the ab initio methods. All statistical analyses were carried out by ORIGIN software (version 6.0). The regression analysis was used to establish the model. Model quality was characterized by the number of observations (n), the correlation coefficient (R), the standard deviation (SD), the Fisher criterion (F), and the significance level (P).

1.3 Computer calculation method

The method for performing quantum mechanical calculations includes the following steps: (1) draw the molecular carbon-frame structure. The initial geometry can generally be designed by applying chemically reasonable atomic distances, angles, and dihedral angles. With a molecular editor the molecule can be built interactively including the chemically reasonable bonds and angles; (2) set up the molecular dynamics and optimize the structure parameters. Once the molecular geometry has been defined, considerations involved in selecting the quantum mechanical method, checking numbers of unpaired electrons present in the molecule, and understanding the effect of the environment, are required; (3) select a calculation model. The semi-empirical method, PM3, or AM1, is selected. The method of PM3 can automatically optimize the bond length, angle, and twist angle, and yield a lot of information on the structure; (4) define and select the output parameter. Quantum chemical calculations produce a large amount of output data. Therefore it is important to decide which output is relevant for the chemical, biochemical, or biological mechanism underlying the phenomenon under investigation. In this study, E_{HOMO}, E_{LUMO}, and E_{ES} can be expected to be of importance for dye biodegradability in the relevant process; (5) run the calculation job. After the method and all generated output have been defined, the calculation can begin. The geometry optimization is the most time-consuming part of many quantum mechanical calculations. The extent to which the geometry is converging to an energy minimum is listed in the output file and indicates when the calculation is about to be completed.

Table 1 Names, Color Index number, molecular weights (M_W), calculated E_{HOMO}, E_{LUMO}, and E_{ES} of the acid dyes used in this study and their biodegradabilities

Name (Color Index number)	M_W	E_{HOMO} (eV)	E_{LUMO} (eV)	E_{ES} (eV)	Biodegradability (%)
Acid Red 1 (C. I. 18050)	509	–8.817	–4.698	–4.268	95.0
Acid Brown 2 (C. I. 17605)	515	–8.678	–1.919	–2.834	76.0
Acid Green 20 (C. I. 20495)	586	–7.602	–2.119	–2.567	49.4
Acid Orange 156 (C. I. 26501)	462	–8.403	–4.654	–2.161	94.3
Acid Red 151 (C. I. 26900)	454	–8.330	–2.186	–2.249	60.2
Acid Blue 113 (C. I. 26360)	681	–7.839	–2.508	–2.470	92.6
Acid Blue 25 (C. I. 62055)	416	–8.717	–1.972	–2.243	96.1
Acid Blue 40 (C. I. 62125)	473	–8.679	–4.457	–3.410	79.8
Acid Blue 129 (C. I. 62058)	458	–7.858	–1.898	–2.256	84.0
Acid Blue 324 (NA) [a]	473	–8.460	–2.172	–3.298	93.6
Acid Violet 17 (C. I. 42650)	761	–4.481	–0.413	–2.899	21.1
Acid Blue 7 (C. I. 42080)	690	–5.038	–0.830	–10.447	21.3
Acid Red 52 (C. I. 45100)	596	–4.763	–0.921	–2.540	19.7
Acid Red 87 (C. I. 45380)	692	–8.631	–2.444	–2.993	54.6
Acid Black 2 (C. I. 50420)	637	–4.423	–0.987	–1.914	66.3
Acid Orange 3 (C. I. 10385)	452	–9.348	–2.104	–4.803	89.5
Acid Violet 68 (NA) [a]	811	–8.466	–2.415	–2.631	42.9
Acid Violet 90 (C. I. 18762)	963	–8.685	–3.649	97.817	38.5
Mordant Red 7 (C. I. 18760)	446	–8.320	–1.686	–2.810	94.6
Acid Black 52 (C. I. 15711)	461	–8.545	–2.355	–3.132	94.1

[a] NA presents that the Color Index number is not available.

The calculation results of quantum chemical descriptors are also listed in Table 1.

2 Results and discussion

The statistical models developed, primarily employed simple linear regression analysis to investigate the effect of each chemical or molecular descriptor of the selected dye on its biodegradability (B). On the basis of the experimental data, a plot of dye biodegradations versus their molecular weights developed model 1.

Model 1

$B = 143.54(\pm 18.44) - 0.13(\pm 0.03)\ M_W$

$n = 20$, $R^2 = 0.75$, SD = 20.19, $P = 0.000$,

$F\ (31.14) > F_{0.01(1,\ 18)}\ (8.28)$

Where, B is the biodegradability; M_W is the molecular weight; n is the number of observations; R^2 is the correlation coefficient; P is the significance level; F is the Fisher criterion.

The result of simple regress analysis showed that model 1 was highly significant at 99% confidence intervals (M_W was in the range of 416 to 963), indicating that the molecular weight was closely correlated with the property of dye biodegradation and had an important effect on dye decolorization. This result was consistent with the studies of Moir *et al*. (2001), and Sun *et al*. (1999). Generally, in dyes with high molecular weights it was rather difficult to penetrate the cell membrane of a microorganism. Early qualitative study also found that the higher the molecular weights, the worse the biodegradabilities for dyes with similar structures in the same biotreatment process.

The energy of the highest occupied molecular orbital, E_{HOMO}, was widely used in developing the QSAR model, for studying the different mechanisms of chemicals, such as, toxicity, biotransformation, and bioreduction (Hu *et al*., 2005; Soffers *et al*., 2001; Wang *et al*., 2001). The quantum chemical descriptor E_{HOMO} usually directed the nuclephilic reactivity of a compound. In this study, E_{HOMO} was calculated and the relationship with biodegradability in all the 20 studied dyes with E_{HOMO} was established.

Model 2

$B = -19.47(\pm 4.13) - 11.38(\pm 3.07)\ E_{HOMO}$

$n = 20$, $R^2 = 0.81$, SD = 21.44, $P = 0.001$,

$F\ (13.73) > F_{0.01(1,18)}\ (8.28)$

Obvious correlation between decolorizations of the selected dyes and E_{HOMO} (E_{HOMO} was between -9.346 eV and -4.423 eV) implied that the biodegradabilities of these dyes were strongly involved in their molecular reactivities of donating electrons. This finding would illustrate the reason why acid dyes normally reacted in acid conditions.

The use of frontier orbital characteristics to describe relative chemical reactivities of a series of compounds, is based on the so-called frontier orbital theory. Although E_{LUMO} is an electronphilicity parameter and it appears as directly proportional to the electronic affinity of a compound, to explore the biodegradation mechanism of the tested dyes, the E_{LUMO}-dependent relationship has been investigated, and thus Model 3 was obtained.

Model 3

$B = 43.24(\pm 12.34) - 10.75(\pm 4.74)\ E_{LUMO}$

$n = 20$, $R^2 = 0.47$, SD = 25.10, $P = 0.036$

$|R| < \left|R_{0.01(18)}\right| = 0.56$

$F\ (5.15) < F_{0.01(1,18)}\ (8.28)$

Because the absolute value of the correlation coefficient and F value was less than 0.56 and 8.28 respectively, a weak correlation was found from the quantum chemical descriptor E_{LUMO} (E_{LUMO} fluctuated from -4.698 eV to -0.413 eV). This finding suggested that the acid dye biodegradability depended more on E_{HOMO} than on E_{LUMO}, which was not in good agreement with the results of Dai *et al*. (1998) and Sun *et al*. (1999), who studied QSBRs on azo dyes. The lack of statistical significance correlation observed for B and E_{LUMO} was most likely a result of the fact that complicated mechanisms existed in the studied biochemical process, especially for enzymatic reduction by microbes, in the facultative bio-film reactor. Biosorption might have, to a great extent, synchronized with biodegradation in the microbial zoogloea during the dye decolorization. Walker and Weatherley (2000), also found that biosorption accounted for 13%–19% of decolorization in the aerobic biodegradation of acid blue 277:1. Furthermore, an attempt was made to check the influence of ΔE ($E_{LUMO} - E_{HOMO}$) on the dye biodegradability, however, no statistical linear relationship could be found between them.

Although the excited probability between the singlet state and the triplet state in a dye molecule is rather low, the energy of the triplet state is a significant parameter of photochemical reaction for color change. Therefore, E_{ES} was still taken into consideration in this study. Regression analysis on decolorization against the energy of the molecular excited state was performed and model 4 was obtained.

Model 4

$B = 68.68(\pm 6.22) - 0.28(\pm 0.02)\ E_{ES}$

$n = 20$, $R^2 = 0.48$, SD = 27.72, $P = 0.336$

$|R| < \left|R_{0.01(18)}\right| = 0.56$

$F\ (1.02) < F_{0.01(1,18)}\ (8.28)$

Similar to the effect of E_{LUMO} on the dye biodegradability, model 4 explained the E_{ES}-dependent correlation without statistical meaning (E_{ES} was in the range of –1.914 eV to 97.817 eV).

In view of the complex effects in the biological treatment, the description of quantitative structure-biodegradability relationships often does not follow a one-descriptor mechanistic approach. A multi-parameter approach is widely accepted by researchers, which is of the idea that a single descriptor will not contain enough information to capture the dominant features of a given biological phenomenon and that the characterization of

Table 2 Correlation matrix of independent variables (calculated from 20 cases)

	M_W	E_{HOMO}	E_{LUMO}	ΔE (E_{LUMO}-E_{HOMO})	E_{ES}
M_W	1				
E_{HOMO}	0.334	1			
E_{LUMO}	0.133	0.670	1		
ΔE (E_{LUMO}-E_{HOMO})	–0.311	–0.657	0.119	1	
E_{ES}	0.596	–0.161	–0.268	–0.057	1

Table 3 Comparison of observed biodegradation and their predicated values from model 5

Name (Color Index number)	E_{HOMO} (eV)	M_W	Predicted value (%)	Observed value (%)	Difference (observed value–predicted value) (%)
Acid Orange 7 (C. I. 15510)	–8.577	350	98.7	96.4	–2.3
Acid Black 1 (C. I. 20470)	–7.420	616	62.6	59.8	–2.8
Acid Yellow 219 (NA) [a]	–8.437	448	87.8	86.9	–0.9
Acid Blue 145 (NA)	–7.577	430	84.0	85.1	1.1
Acid Green 27 (C. I. 61580)	–8.030	706	58.7	57.3	–1.4

[a] NA presents that the Color Index number is not available.

chemicals should be multivariate (Leonard and Roy, 2006; Mon *et al*., 2006; Verma and Hansch, 2005). On the basis of previous analyses, it can be seen that the studied independent variables such as E_{HOMO}, E_{LUMO}, E_{ES}, ΔE, and M_W are strongly inter-correlated, probably owing to a similar dependence on the molecular wave function (Table 2), and these one-parameter models are far from complete. However, the main trend is apparent: E_{HOMO} and M_W have dominating effects on dye biodegradation. Thus, two descriptors including E_{HOMO} and M_W have been applied, to determine a general and robust QSBR model. Multiple linear regression technique and variance analysis have been used in model 5.

Model 5

$$B = 63.18(\pm 8.04) - 8.23(\pm 2.44)E_{HOMO} - 0.10(\pm 0.02)M_W$$

$$n = 20,\ R^2 = 0.80,\ \mathrm{SD} = 16.10,\ P < 0.0001$$

$$F\ (29.64) > F_{0.01(2,17)}\ (6.11)$$

The great significance of the multiple-variable-containing model justified by F-value was obtained in model 5, with the majority of variances (99%) being explained, and the significance of all variables in the model. Parameters reflecting the electronic properties and penetration were selected to construct the QSBR model 5, which covered the major factors controlling the biodegradation mechanism. No overlapped information was included in the model.

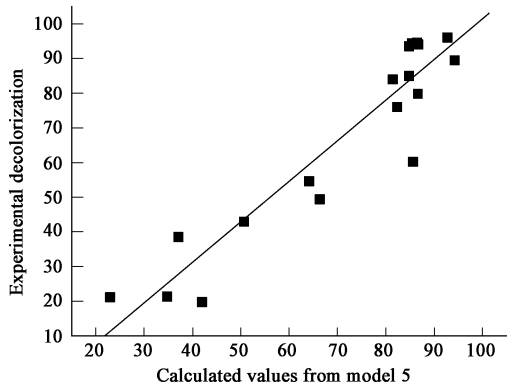

Fig. 1 Correction plot of experimental decolorization efficiencies versus calculated values by model 5.

For the sake of verifying the accuracy of the combined QSBR model, comparison of calculated values with observed color removal efficiencies was made. It could be found from Fig.1 that a good correlation existed, except for very few outliers, because of a slight biosorption in the process of decolorization and fluctuation of environmental factors during the experiment. In addition, another five acid dyes were induced in the facultative-aerobic process to test the practicality of the obtained model (Table 3). Calculated data from model 5 fitted well with the experimental values, demonstrating that the developed QSBR model was able to serve as a general predictor for unknown acid dyes. It would be significant for theoretical guidance, for dye wastewater biotreatment and dye product development for environmental protection.

3 Conclusions

QSAR approaches have been used extensively to study many types of biological activities, including chemical toxicity of a variety of different classes of compounds (Juranić *et al*., 2006; He and Jurs, 2005; Netzeva and Schultz, 2005). The use of these approaches in dealing with problems of acid dye biodegradability, however, has been rather limited. In view of the widespread use of acid dyes in various consumer products, developing practical, reliable QSAR/QSBRs to screen new and existing acid dyes, as well as their metabolites, for potential biodegradation problems is of vital importance.

In this study, a QSAR method was employed to study the predictive and mechanism-based relationships on the biodegradability of acid dyes. Chemical and molecular descriptors, M_W, E_{HOMO}, E_{LUMO}, and E_{ES}, were computed and QSBRs were obtained through each descriptor to investigate the controlling mechanism of biodegradation. Results showed that E_{HOMO} and M_W were the dominant parameters governing biodegradability of acid dyes. A general and robust QSBR model, B=63.18(±8.04)–

8.23(±2.44) E_{HOMO}–0.10(±0.02) M_W, was developed by 20 acid dyes with the combined application of variables reflecting nuclephilic reactivity and molecular property.

Tendency in the calculated values from the general model with molecular descriptors versus their observed biodegradations in a facultative-aerobic process presented quite a linear behavior, indicating that the developed model 5 could be explained mechanistically and applied practically for reliable predictions.

References

Brown D, Laboureur P, 1983. The degradation of dyestuffs: Part I. Primary biodegradation under anaerobic conditions[J]. Chemosphere, 12(3): 397–404.

Chen C, Pu L, Shi H, 2003. Relationship between dye structure and decolorization efficiency of iron chip technique[J]. China Environ Sci, 23(4): 376–379.

Chen J, Peijnenburg W J G M, Quan X *et al*., 2000. Quantitative structure-property relationships for direct photolysis quantum yields of selected polycyclic aromatic hydrocarbons[J]. Sci Total Environ, 246(1): 11–20.

Dai S, Song W, Zhuang Y *et al*., 1998. The study of quantitative strucuture-biodegradability relationships for azo dyes[J]. Environ Chem, 17(2): 115–119.

He L, Jurs P C, 2005. Assessing the reliability of a QSAR model's predictions[J]. J Mol Graph Model, 23(6): 503–523.

Hu J, Eriksson L, Bergman Å *et al*., 2005. Molecular orbital studies on brominated diphenyle ethers. Part II. Reactivity and quantitative structure-activity (property) relationships[J]. Chemosphere, 59(7): 1043–1057.

Juranić I O, Drakulić B J, Petrović S D *et al*., 2006. A QSAR study of acute toxicity of N-substituted fluoroacetamides to rats[J]. Chemosphere, 62(4): 641–649.

Leonard J H, Roy K, 2006. QSAR by LFER model of HIV protease inhibitor mannitol derivatives using FA-MLR, PCRA, and PLS techniques[J]. Bioorgan Med Chem, 14(4): 1039–1046.

Li Y, Xi D L, 2004. Decolorization and biodegradation of dye wastewaters by a facultative- aerobic process[J]. Environ Sci Pollut Res, 11(6): 372–377.

Lu G H, Yuan X, Zhao Y H, 2001. QSAR study on the toxicity of substituted benzenes to the algae (*Scenedensmus obliquus*)[J]. Chemosphere, 44(3): 437–440.

Luan F, Ma W, Zhang X *et al*., 2006. Quantitative structure-activity relationship models for prediction of sensory irritants (log RD_{50}) for volatile organic chemicals[J]. Chemosphere, 63(7): 1142–1153.

Moir D, Masson S, Chu I, 2001. Structure-activity relationship study on the bioreduction of azo dyes by *Clostridium paraputerificum*[J]. Environ Toxicol Chem, 20(3): 479–484.

Mon J, Flury M, Harsh J B, 2006. A quantitative structure-activity relationships (QSAR) analysis of triarylemethane dye tracers[J]. J Hydrol, 316(1-4): 84–97.

Netzeva T I, Schultz T W, 2005. QSARs for the aquatic toxicity of aromatic aldehydes from *Tetrahymena* data[J]. Chemosphere, 61(11): 1632–1643.

Oprea T I, Kurunczi L, Timofei S, 1997. QSAR studies of disperse azo dyes: Towards the negation of the pharmacophore theory of dye-fiber interaction?[J]. Dyes and Pigments, 33(1): 41–64.

Pagga U, Brown D, 1986. The degradation of dyestuffs: Part II Behaviour of dyestuffs in aerobic biodegradation tests[J]. Chemosphere, 15(4): 479–491.

Schultz T W, Cronin M T D, Walker J D *et al*., 2003. Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective[J]. J Mol Struct (Theochem), 622(1/2): 1–22.

Soffers A E M F, Boersma M G, Vaes W H J *et al*., 2001. Computer-modeling-based QSARs for analyzing experimental data on biotransformation and toxicity[J]. Toxicol In Vitro, 15(4/5): 539–551.

Stewart J J P, 1989. Optimization of parameters for semiempirical methods: I. Method[J]. J Comp Chem, 10(2): 209–220.

Sun H, Wang G, Cong L *et al*., 1999. Biodegradation of azo dyes by algae and QSBR study[J]. China Environ Sci, 19(4): 289–292.

Venkataraman N, 2000. Colourful water is not colourful[J]. Asian Water, June: 33–35.

Verma R P, Hansch C, 2005. An approach toward the problem of outliers in QSAR[J]. Bioorgan Med Chem, 13(15): 4597–4621.

Walker G M, Weatherley L R, 2000. Biodegradation and biosorption of acid anthraquinon dye[J]. Environ Pollut, 108(2): 219–223.

Wang X, Dong Y, Wang L *et al*., 2001. Acute toxicity of substituted phenols to Rana japonica tadpoles and mechanism-based quantitative structure-activity relationship (QSAR) study[J]. Chemosphere, 44(3): 447–455.

Yang W, Yang X W, 2001. Domestic development of acid dyes [J]. Shanghai Dyestuffs, 29(1): 7–17, 24.