Bull Environ Contam Toxicol (2012) 89:950–954
DOI 10.1007/s00128-012-0819-0

QSAR of Acute Toxicity of Halogenated Phenol to Green Fluorescent Protein by Using Density Functional Theory

X. X. Liu · J. Y. Li · J. X. Yu · Sh. Q. Sun · Y. J. Wang · H. X. Liu

Received: 19 June 2012 / Accepted: 31 August 2012 / Published online: 16 September 2012

Abstract A novel approach was established to predict toxicity of environmental pollutants by using green fluorescent protein (GFP) as bio-marker. In the approach, recombinant *Escherichia coli* was constructed to express GFP. The toxicity values ($-\lg EC_{50}$) of 14 halogenated phenols to recombinant *E. coli* with GFP gene were measured. And optimized calculation was carried out at B3LYP/6–31G* level using density functional theory method. Based on the MTLSER model, the obtained parameters were taken as theoretical descriptors to establish the novel QSAR model for predicting $-\lg EC_{50}$ ($R^2 = 0.922$). The model includes two variables (standard entropy (S^{θ}) and the most negative atomic net charges of the molecule (q^-)). The results of cross–validation test ($q^2 = 0.868$) indicate the model of this study has optimum stability, which shows that it is feasible to predict to toxicity of chemistry utilizing recombinant *E. coli* with GFP gene.

Keywords Quantitative structure–activity relationship (QSAR) · Toxicity ($-\lg EC_{50}$) · Green fluorescent protein (GFP) · Halogenated phenol · Density functional theory (DFT)

Bacteria, algae, fish and other organisms have been used to evaluate the toxicity of environmental pollutants. Among the organism, the luminescent bacteria are the most favoured because of their convenience, sensitivity and efficiency in the evaluation. The luminescent bacteria fluoresce due to the reaction of fluorescence enzyme, and green fluorescent protein (GFP) can fluoresce itself. In recent years, GFP was widely used in microbial metabolic engineering, microbial ecology, microbial cell biology as molecular marker duo to its easy detection, high sensitivity, fluorescence stability, intoxicity and so on (Romero et al. 2008; Yang 2010). One of the many applications is to detect degradation of environmental pollutants using GFP expressed in appropriate bacteria (Yao et al. 2003; Jin and Zhang 2008; Tian and Wang 2005). It is well known that protein expression responses to cell growth and physiological state. Therefore, GFP expression correlates well to the cell growth. Thus GFP can be an appropriate organism to evaluate the toxicity of environmental pollutants. But the application of GFP in this aspect has not been reported in litenature to the best of our knowledge.

Several methods have been used to detect environmental pollutants, such as health risk evaluation method and parameter determination method (Tang et al. 2009), capillary gas chromatography method (Jin 2009). But novel substances increase on the daily base and their environmental hazards may not be known. Their screening for such hazards are time consuming and costly. Quantitative structure–activity/property relationship (QSAR/QSPR) method is an effective way to solve this problem (Yu and Hu 2005; Zhang and Zhang 2010; Li et al. 2010). Xie et al. (2008) studied the QSPR of solubility and *n*-Octanol/Water partition coefficients for substituted phenols, the results showed the stability and predictive power were more advantageous than those based on AM1 molecular orbital method and molecular connectivity method. Loader et al. (2006) researched QSAR of the cytotoxicity of ortho alkyl substituted 4-X-phenols based on theoretical bond lengths and electron densities.

X. X. Liu · J. Y. Li · J. X. Yu · Sh. Q. Sun · Y. J. Wang · H. X. Liu (✉)
College of Biological, Chemical Sciences and Engineering, Jiaxing University, 118 Jiahang Road, Jiaxing 314001, Zhejiang, People's Republic of China
e-mail: wulhx@126.com

In this study, the toxicities of 14 halogenated phenols to recombinant strain with GFP gene were determined. Using Gaussian 03 program (Frisch et al. 2003), the structural parameters and thermodynamic parameters of these compounds have been calculated by density functional theory (DFT) at the B3LYP/6-31G* level. The QSAR model was established, and then the toxicity of compounds was discussed on the basis of the molecular structure. The feasibility of using the constructed model to predict the toxicity of environmental pollutants was investigated.

Materials and Methods

Escherichia coli BL21 (DE3), DH5α, plasmid pUC19-GFP were from our laboratory, plasmid pET28a as gift was from the Zhejiang Medical University, Xu Jian associate Professor.

Taq DNA polymerase, dNTP, T4DNA ligase, DNA marker and Isopropyl beta-D-thiogalactopyranoside (IPTG) were purchased from Shanghai Sangon company. *Hind* III and *Eco*RI were purchased from Dalian TaKaRa company. Oligonucleotide primers were synthesized by Shanghai Sangon company. Halogen compounds were purchased from Sigma–Aldrich (Shanghai) Trading Co Ltd, and other biochemical reagents of AR grade were provided by local supplier.

The plasmid pUC19-GFP and pET28a were simultaneously digested with *Eco*RI and *Hind*III, and then ligated with T4 ligase before being transferred into competent cell *E. coli* BL21 (DE3). The process of recombinant plasmid construction was described in Fig. 1.

The positive clones were chosen by PCR. The forward primer and reverse primer are5′-GGAATTCCATATGGTG AGCAAGGGCGAGG-3′ and 5′-CCCAAGCTTCT-TGTA CAG-CTCGTCCATGC-3′, respectively.

The recombinant strain BL21(DE3) pET28a-GFP was cultured in LB medium (peptone 10 g L^{-1}, yeast extract 5 g L^{-1}, NaCl 10 g L^{-1}) for 2 h, and then the culture temperature was adjusted to 28°C, while 0.2 mM IPTG was added. Under the conditions, the recombinant strain was cultured for 2 h before half inhibitory concentration was measured.

The half inhibitory concentrations of 14 halogenated phenols were measured using ModulusTM microplate photometer. And 5 gradients were set with even logarithm interval, each concentration was parallel measured 6 times.

Using Gaussian 03 program, the selected compounds have been calculated by DFT at the B3LYP/6-31G* level. The term, Opt Freq, refers to the optimization of the molecular structure (Opt) followed by frequencies calculations (Freq) performed at the stationary point on the potential energy surface. The properties of parameters were obtained directly from DFT calculations (Gaussian's output file).

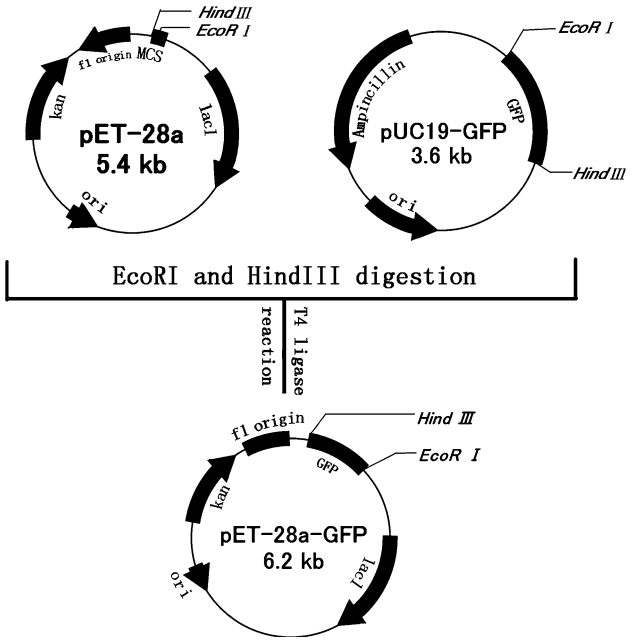

Fig. 1 Construction of the recombinant plasmid pET28a-GFP

Zhou et al. (2005), Zhai et al. (2006) and Wang et al. (2008) have calculated some POPs (PCBs, PCNs and PBDEs) by DFT and the relevant QSAR models are better than those from semi-empirical methods. Especially, the correlation increases obviously once thermodynamic parameters are included in the QSAR model (Chen et al. 2007; Liu et al. 2007) and the model such as Eq. (1):

$$XYZ = XYZ_0 + aX + bY + cZ + \cdots \quad (1)$$

where XYZ is the property of solute and solvent interaction, such as the distribution coefficient, retention capacity, toxicity (EC_{50} and LC_{50}) and so on, which are often expressed as the value of the logarithm or negative logarithm, XYZ_0 is the constant term, a, b, c the coefficient. The coefficients were determined by stepwise regression analysis.

In regression analysis, all independent variables were verified using t test in a significant level $\alpha = 0.05$. The Variation inflation factor (VIF) was used to check the multicollinearity between independent variables. Here $VIF = 1/(1 - r^2)$ in which r is the correlation coefficient of multiple regressions between one variable and the others in the equation. If $VIF = 1.0$, no intercorrelation exists for each variable; if VIF ranges from 1.0 to 5.0, the related equation is acceptable; and if VIF is larger than 10.0, the regression equation is unstable and recheck is necessary. The quality of the model was evaluated in terms of the leave-one-out (LOO) cross-validation correlation

coefficient q^2, the squared conventional correlation coefficient R^2, and the standard error *SE*.

Results and Discussion

The 1.0 % agarose electrophoresis of PCR products was identified after amplification (see Fig. 2). The results showed the size of DNA band amplified matchs to theoretical speculation value of GFP fragment (850 bp), which indicated that the recombinant plasmid pET28a-GFP was constructed successfully.

The partial calculated structural and thermodynamic parameters are listed in Table 1. To obtain correlation equation of $-\lg EC_{50}$, structural parameters and thermodynamic parameters of halogenated phenols were taken as theoretical descriptors. By employing SPSS 12.0 for Windows, forward stepwise regression analysis was carried out on observed values of $-\lg EC_{50}$ to GFP, and the results are presented in Eq. (2).

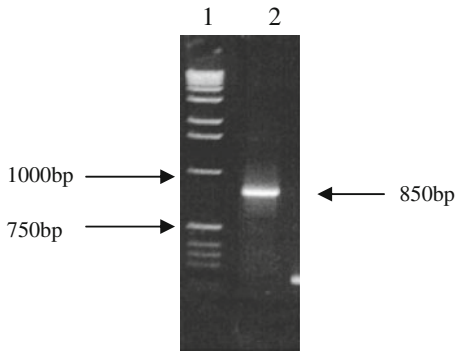

Fig. 2 Agarose electrophoresis of PCR products

$$-\lg EC_{50} = 17.662 + 10.578S^{\theta} \times 10^{-3} + 29.975q^{-} \quad (2)$$

$$n = 14,\ R^2 = 0.922(R = 0.960),\ SE = 0.126,$$

$$q^2 = 0.868,\ RMS_{LOO} = 0.149$$

where n, R^2, *SE*, q^2 and RMS_{LOO} are the number of regression equations, correlation coefficient, standard deviation, cross-validation correlation coefficient and RMS error, respectively. As Eq. (2) gives high correlation coefficients and small standard deviations, the correlation of Eq. (2) is excellent. For comparison, the relevance of the model was hardly affected by adding other parameters or increasing the number of parameters.

Equation (2) shows: (a) $-\lg EC_{50}$ increases with the most negative atomic charge (q^-), this is because the charge of the molecular impact property of fluorescent protein. (b) $-\lg EC_{50}$ is directly proportional to the entropy (S^{θ}) due to the degree of disorder caused by S^{θ} expression, and the larger the degree of disorder, the stronger the toxicity.

It can be seen from the Table 2, the *VIF* for Eq. (2) are both 4.119, and all the absolute *t* values are far larger than the standard one ($t_{\alpha/2} = 2.160$), suggesting Eq. (2) has obvious statistic significance. The *VIF* is less than 5.0, which shows that the correlation between the variables from Eq. (2) is very low. The standard regression coefficients (*SR*) and *t* values are also listed in Table 2. As shown in table 2, the value of $t_{\alpha/2}$ is 2.160 with confidence level of 95 %. The results shown in Table 2 that the absolute value of *SR* and *t*-scores for S^{θ} are greater than

Table 1 Measured and predicted $-\lg EC_{50}$ of halogenated phenols and corresponding structural and thermodynamic parameters by B3LYP/6-31G*

Number	Molecule	Exp.	Eq. (1)		Eq. (2)		q^- (e)	S^{θ} (J mol^{-1} K^{-1})
			Cal.	Diff.	Cal.	Diff.		
1	*p*-chlorophenol	2.146	2.128	−0.018	2.132	−0.014	−0.6390	342.339
2	3-chlorophenol	2.230	2.170	−0.060	2.171	−0.059	−0.6376	342.084
3	2,3-Dichlorophenol	2.614	2.641	0.027	2.629	0.015	−0.6311	368.443
4*	2,4-Dichlorophenol	2.768	2.620	−0.148	2.611	−0.157	−0.6325	370.267
5	2,5-Dichlorophenol	2.743	2.684	−0.059	2.670	−0.073	−0.6303	370.033
6	3,4-Dichlorophenol	2.712	2.526	−0.185	2.525	−0.187	−0.6359	371.004
7	3,5-Dichlorophenol	2.700	2.630	−0.070	2.622	−0.078	−0.6327	371.962
8*	*m*-Fluorophenol	1.905	1.949	0.043	1.956	0.051	−0.6407	330.201
9	*o*-Fluorophenol	1.754	1.890	0.136	1.902	0.147	−0.6423	329.252
10	2,4-Difluorophenol	2.149	2.052	−0.097	2.068	−0.081	−0.6437	348.280
11	2-Chloro-4-fluorophenol	2.133	2.346	0.213	2.348	0.215	−0.6376	358.933
12*	3-Chloro-4-fluorophenol	2.232	2.276	0.044	2.285	0.053	−0.6405	360.460
13	2,3,5,6-Tetrafluorophenol	3.045	3.041	−0.004	3.015	−0.030	−0.6245	387.447
14	Pentafluorophenol	3.070	3.236	0.166	3.214	0.144	−0.6256	408.927

The compounds with asterisk aren't in training set for the validation of Eq. (1)

Table 2 Correlation coefficients and the variance inflation factors (*VIF*) and standard regression coefficients (*SR*) and *t* scores ($\alpha = 0.05$) for Eq. (1)

Variable	r^2	*VIF*	*SR*	$t(t_{\alpha/2} = 2.160)$, $\alpha = 0.05$
q^-	0.870	4.119	0.433	2.530
S^θ	0.870	4.119	0.559	3.265

those of q^- indicates that the impact on $-\lg EC_{50}$ for S^θ is greater than that of q^-.

In order to identify "outliers" and examine the correlation of the model, the cross-validation technique 'LOO' (Liu et al. 2003) was used. In this approach, one sample was removed each time and the remaining samples were used as training set. The correlation coefficient q^2 Eq. (2) was 0.868 and mean square error (RMS_{LOO}) was 0.149. High correlation coefficient and lower RMS_{LOO} suggests that the model is of good stability and predictability.

In addition, we also verified the model with group validation. Fourteen halogenated phenols were divided into two groups: the first three compounds in every four in-order are included in the first group (training sets); and the left ones are all in second group (test sets). Using the same regression method as mentioned above, the validation models of the 11 compounds (training sets) were created Eq. (3).

$$-\lg EC_{50} = 16.192 + 10.638S^\theta \times 10^{-3} + 27.700q^- \quad (3)$$

$$\mathrm{n} = 11,\ R^2 = 0.918(R = 0.958),\ SE = 0.136,$$

$$q^2 = 0.787,\ RMS_{LOO} = 0.190$$

The correlation coefficient ($R^2 = 0.918$), cross-validation correlation coefficient ($q^2 = 0.787$) of Eq. (3) are close to the corresponding value of Eq. (2) ($R^2 = 0.922$, $q^2 = 0.868$). Furthermore, we predicted the $-\lg EC_{50}$ to GFP by Eqs. (2) and (3), respectively, and the predicted values and errors are listed in Table 1. The maximum deviation of $-\lg EC_{50}$ values with Eq. (3) is -0.215 (No. 11 2-chloro-4-fluorophenol), which are close to the corresponding values (0.213) from Eq. (2). The results show that Eq. (2) is reliable and possesses good predictive ability.

Jiang et al. (2011) and You et al. (2009) also studied the toxicity of substituted phenol. Although the indication organisms are different, the orders of the toxicity are inconsistent, in which the toxicity of flurophenol is larger than that of chlorophenol. Furthermore, the parameters of models in our study are less than those in their models.

Acknowledgments We gratefully acknowledge Prof. Liu Xiao Ming, for revising the manuscript. This work was supported by the National Natural Science Foundation of China (No. 31100053) and the general science and technology project of Jiaxing (No. 2011AY1034).

References

Chen SD, Liu HX, Wang ZY (2007) Study of quantitative structure–retention relationship (QSRR) of gas chromatography for polychlorinated dibenzothiophenes on non-polar columns. QSAR Comb Sci 26(8):889–896

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery, Jr JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (2003) Gaussian 03, revision A. 1. Gaussian Inc., Pittsburgh, PA

Jiang ST, Nie CM, Lin YW, Shi CY (2011) Quantitative structure-activity analyses of anilines and phenols toxicity to *Daphnia magna Straus*. Comput Appl Chem 28(6):797–800

Jin L (2009) Capillary column gas chromatography of organic toxicants in the air. Chin J Health Lab Technol 11:2545–2546

Jin K, Zhang YJ (2008) GFP as a vital marker for investigation of pathogenic development of *Beauveria bassiana*. Mycosystema 27(3):377–384

Li YC, Zhang SL, Bai FM, Zhao GQ, Chen FL (2010) QSAR analysis of toxicity of chloro-benzene and alkylphenolic compounds to zebrafish. South China Fish Sci 6:19–23

Liu SS, Liu HL, Yin CS, Wang LS (2003) VSMP: a novel variable selection and modeling method based on the prediction. Chem Inf Comput Sci 43:964–968

Liu HX, Wang ZY, Zhai ZC, Liu HY, Wang LS (2007) Studies of *n*-Octanol/water partition coefficients (lg*K*ow) for organophosphate compounds by density functional theory. Chin J Struct Chem 26(3):367–373

Loader RJ, Singh N, O'Malley PJ, Popelier PLA (2006) The cytotoxicity of ortho alkyl substituted 4-X-phenols: a QSAR based on theoretical bond lengths and electron densities. Bioorgan Med Chem Let 16(5):1249–1254

Romero A, Figueras A, Thoulouze MI, Bremont M, Novoa B (2008) Recombinant infectious hematopoietic necrosis viruses induce protection for rainbow trout Oncorhynchus mykiss. Dis Aquat Org 80:123–135

Tang ZW, Cheng JL, Zhang HY, Fan XY (2009) Health risk assessment of organic pollutants in water of Wuhan reach of Yangtze River. J Hydraul Eng 9:1064–1069

Tian T, Wang Q (2005) The application of green fluorescent protein (GFP) as molecular marker in microbiology. J Microbiol 25:68–73

Wang ZY, Zeng XL, Zhai ZC (2008) Prediction of supercooled liquid vapor pressures and n-octanol/air partition coefficients for polybrominated diphenyl ethers by means of molecular descriptors from DFT method. Sci Total Environ 389:296–305

Xie YJ, Liu H, Liu HX, Zhai ZC, Wang ZY (2008) Determination of solubilities and n-octanol/Water partition coefficients and QSPR study for substituted phenols. Bull Environ Contam Toxicol 80:319–323

Yang XJ (2010) The application of green fluorescent protein in tumor research. J Tongji Univ (Med Sci) 31(2):107–110

Yao ZS, Chen ZY, Chen ZY, Zheng XB, Zhang J, Huang DF (2003) Genetically marking of natural biocontrol bacterium *Bacillus subtilis* strains with green fluorescent protein gene. Chin J Biotechnol 9(19):551–555

You XJ, Liu H, Yang GY, Wang ZY (2009) Determination and QSAR study on the toxicity of substituted phenol against Qinghaiensis sp (Q67). Chin J Struct Chem 28(10):1311–1316

Yu RL, Hu GR (2005) Toxicity of phenols to photobacterium at different pH and QSAR. Environ Sci Technol 28(4):20–22

Zhai ZC, Wang ZY, Chen SD (2006) Quantitative structure-retention relationship for gaschromatography of polychlorinated naphthalenes by abinitio quantummechanical calculations and a Cl substitution position method. QSAR Comb Sci 25(1):7–14

Zhang WY, Zhang YT (2010) The relationship between model predictive ability and validated method in QSAR/QSPR study. Comput Appl Chem 2:202–205

Zhou W, Zhai ZC, Wang ZY, Wang LS (2005) Estimation of n-octanol/water partition coefficients (Kow) of all PCB congeners by density functional theory. J. Mol Struct (Theochem) 755(1–3):137–145