ELSEVIER

Journal of Molecular Structure: THEOCHEM 764 (2006) 141–148

THEO CHEM

www.elsevier.com/locate/theochem

A comparison of semiempirical and first principle methods for establishing toxicological QSARs of nitroaromatics

Xiu-Fen Yan, He-Ming Xiao *, Xue-Dong Gong, Xue-Hai Ju

Department of Chemistry, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China

Received 9 July 2005; received in revised form 22 February 2006; accepted 23 February 2006
Available online 19 April 2006

Abstract

Semiempirical MO AM1 and PM3, ab initio MO HF and density functional theory (DFT) B3LYP methods were used to calculate the electronic and structural properties of 28 nitroaromatics. QSARs were established based on these properties and the toxicity of nitroaromatics to the fathead minnow. The results showed that the models established based on the first principle methods (HF and DFT-B3LYP) are better than that based on semiempirical methods (AM1 and PM3). HF model is a little better than DFT-B3LYP model on correlation and significance. But the B3LYP model gives more reasonable interpretation of nitroaromatics toxic mechanism. Based on the model, the toxic mechanism of the nitroaromatics was discussed.

Keywords: Semiempirical MO method; First principle method; Nitroaromatics; QSAR; Toxicity

1. Introduction

Nitroaromatics are widely used either as materials or as intermediates in explosives, dyestuffs, pesticides and organic synthesis. They occur as industrial wastes and direct pollutants in the environment, and are relatively soluble in water and detectable in rivers, ponds and soil. Nitrobenzenes are representative of electrophilic toxicants. At present, nitroaromatic compounds are of acute concern because of their varied toxic effects, such as narcosis, mutagenicity and carcinogenicity [1–4]. Furthermore, some of them can be degraded to more toxic molecules. Thus, it is necessary to study the toxicities of nitroaromatics.

Quantitative structure–activity relationship (QSAR) is a mathematical model describing the relationship between toxic potency and one or more descriptors of the chemical. The descriptors include physicochemical properties, various stereo-electronic characteristics, topological indices, and presence/absence of functional groups [5]. With the development of the computer technology and quantum chemistry (QC), the QC treatment of electronic and structural parameters is potentially more powerful than the other approaches because it can produce the parameters accurately and comparatively more easily, and allows greater flexibility in the construction of the datum sets. In the earlier QSAR studies [6–8], many electronic and structural parameters were obtained from MO calculations. However, almost all MO calculations in the earlier QSAR work were made with the semiempirical methodologies, such as AM1, PM3 and et al. Only recently, B3LYP method was used to study the QSAR of chemicals toxicity [9,10].

In this paper, four different QC methods were used to calculate the descriptors of all object chemicals. QSARs were established based on the QC descriptors and the toxicity of nitroaromatics to the fathead minnow. The purpose of this study was to select a good QC method for getting better descriptors and establishing a better QSAR. Finally the mechanism of nitroaromatics' toxicity was discussed based on the QSAR.

2. Calculation methods

Four methods were employed to calculate the QC descriptors. These were the semiempirical AM1 and PM3 methods; ab initio HF/6-31G* method and DFT-B3LYP/6-311G** method as provided by the Gaussian 98 package [11]. At first, the geometries of the compounds were established by Chemoffice 7.0. Then the geometries of all compounds were optimized at the four different levels of theory followed by frequency calculations, which showed that all the optimized structures were energy minima on the potential energy surface.

* Corresponding author. Tel./fax: +86 25 84313919.
E-mail address: xiao@mail.njust.edu.cn (H.-M. Xiao).

0166-1280/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2006.02.018

Table 1
QC descriptors obtained by b3lyp/6-311g** method of 28 nitroaromatics

No.	Compounds	E_{HOMO} (a.u.)	E_{LUMO} (a.u.)	ΔE (a.u.)	Q_C (e.u.)	Q_N (e.u.)	Q_{NO_2} (e.u.)	μ (Debye)
1	Nitrobenzene	−0.2875	−0.0967	0.1908	0.1113	0.1769	−0.3575	4.541
2	1,2-Dinitrobenzene	−0.2995	−0.1172	0.1823	0.1495	0.1733	−0.3142	6.666
3	1,3-Dinitrobenzene	−0.3168	−0.1218	0.1950	0.1006	0.1808	−0.3290	4.209
4	1,4-Dinitrobenzene	−0.3145	−0.1344	0.1801	0.1297	0.1786	−0.3304	0.000
5	1,3,5-Trinitrobenzene	−0.3366	−0.1414	0.1952	0.0941	0.1831	−0.3048	0.001
6	2-Nitrotoluene	−0.2752	−0.0914	0.1838	0.1092	0.1550	−0.3802	4.263
7	3-Nitrotoluene	−0.2745	−0.0936	0.1809	0.1201	0.1750	−0.3627	4.791
8	4-Nitrotoluene	−0.2783	−0.0922	0.1861	0.1136	0.1731	−0.3687	5.200
9	2,3-Dinitrotoluene	−0.2910	−0.1110	0.1800	0.1900	0.1605	−0.3404	6.673
10	2,4-Dinitrotoluene	−0.3049	−0.1155	0.1894	0.1004	0.1759	−0.3406	4.854
11	2,5-Dinitrotoluene	−0.3012	−0.1285	0.1727	0.1348	0.1755	−0.3371	0.909
12	2,6-Dinitrotoluene	−0.2977	−0.1112	0.1865	0.0741	0.1472	−0.3513	2.938
13	3,4-Dinitrotoluene	−0.2914	−0.1127	0.1787	0.1563	0.1730	−0.3159	7.321
14	3,5-Dinitrotoluene	−0.3017	−0.1185	0.1832	0.1108	0.1782	−0.3353	4.875
15	2,4,6-Trinitrotoluene	−0.3178	−0.1341	0.1837	0.0880	0.1764	−0.3188	1.519
16	2,3,6-Trinitrotoluene	−0.3103	−0.1419	0.1684	0.2198	0.1593	−0.2988	2.843
17	2-Nitroaniline	−0.2303	−0.0874	0.1429	0.2067	0.1781	−0.4331	4.726
18	2-Methyl-3-nitroaniline	−0.2276	−0.0822	0.1454	0.0899	0.1454	−0.3805	5.518
19	2-Methyl-4-nitroaniline	−0.2334	−0.0769	0.1565	0.1389	0.1665	−0.4047	7.294
20	2-Methyl-5-nitroaniline	−0.2303	−0.0859	0.1444	0.1057	0.1713	−0.3756	6.063
21	2-Methyl-6-nitroaniline	−0.2271	−0.0855	0.1416	0.2014	0.1756	−0.4065	5.042
22	4-Methyl-3-nitroaniiline	−0.2259	−0.0856	0.1403	0.1006	0.1553	−0.3874	5.254
23	2,4-Dinitroaniline	−0.2603	−0.1099	0.1504	0.2332	0.1857	−0.3930	6.687
24	2-Methyl-4,6-dinitroaniline	−0.2555	−0.1070	0.1485	0.2525	0.1839	−0.3985	7.125
25	2-Methyl-3,6-dinitroaniline	−0.2497	−0.1178	0.1319	0.1945	0.1770	−0.4091	3.013
26	3-Methyl-2,6-dinitroaniline	−0.2522	−0.1143	0.1379	0.2454	0.1799	−0.4091	2.678
27	3-Methyl-2,4-dinitroaniline	−0.2531	−0.1062	0.1469	0.2250	0.1442	−0.3910	5.256
28	4-Methyl-3,5-dinitroaniline	−0.2477	−0.1049	0.1428	0.0960	0.1461	−0.3581	4.991

Referring to the previous studies [12,13], six molecular descriptors were selected in this study, which include the energy of the lowest unoccupied molecular orbital (E_{LUMO}), the energy of the highest occupied molecular orbital (E_{HOMO}), the frontier orbital energy gap ($\Delta E = E_{LUMO} - E_{HOMO}$), the maximum net atomic charge at the aromatic carbon conjoining the nitro group (Q_C), the net charge of the nitro nitrogen (Q_N), the charge of nitro group (Q_{NO_2}) and the dipole moment (μ). The acute toxicities (96 h-LC_{50}, mol/L) of these chemicals to fathead minnow were obtained from Hall et al. [14], which were measured using a standard flow-through method with laboratory reared fathead minnows. Only the QC descriptors obtained by B3LYP method are listed in Table 1. The experimental toxicity values of the nitroaromatic chemicals are listed in Table 2. Descriptors derived from the other three methods are listed in Appendix A.

Statistical analyses were conducted using the SPSS program package [15]. Quantitative structure–activity relationships (QSARs) were developed using stepwise linear regression analysis at the confidence level of 95%. The models were assessed with the R^2_{adj} (adjustment of the R^2), the SE (root of the mean square error) value, *t*-test and its *P* value (the probability of the regression coefficients), the *F* value (Fisher statistic) and its *P* value (the probability of the *F*). The correlation between the variables in the model was estimated by the pairwise correlation matrix for all variables and the variance inflation factor (VIF). Correlation coefficients above 0.8 are considered excessive and one of the pair of variables must be dropped from the model, above 0.7 is conservative. If VIF > 10, a high correlation existed between different variables, and the regression model was not a stable one. If VIF < 5, the regression model was acceptable. Finally, the analysis of residuals was done. The number of the compounds (*n*) was also noted.

3. Results

Using the toxicity value as the dependent variable and the QC descriptors as independent variables, the QSARs were established by multiple linear stepwise regression analysis, and are shown as the following equations:

$$\text{AM1:} \quad -\lg LC_{50} = 2.447 - 32.877E_{LUMO}$$
$$n = 28,\ R^2_{adj} = 0.678,\ SE = 0.462,\ F = 57.733,\ P = 0.000 \tag{1}$$

$$\text{PM3:} \quad -\lg LC_{50} = 2.368 - 33.458E_{LUMO}$$
$$n = 28,\ R^2_{adj} = 0.649,\ SE = 0.482,\ F = 50.826,\ P = 0.000 \tag{2}$$