Iranian Journal of Analytical Chemistry
Volume 3, Issue 2, September 2016 (116-126)

مجله ایرانی شیمی تجزیه
دوره ۳، شماره ۲، مهر ماه ۱۳۹۵ (۱۲۶-۱۱۶)

مطالعه کمی جابجایی شیمیایی ^{13}C بتا نفتالن‌ها با استفاده از روش تصویر دو بعدی و محاسبه تئوری تابع چگالی

زهرا گرکانی نژاد*[۱]، مرضیه پشته شیرانی[۲]
۱. بخش شیمی، دانشکده علوم، دانشگاه شهید باهنر کرمان، کرمان، ایران
۲. بخش شیمی، دانشگاه صنعتی اصفهان، اصفهان، ایران
تاریخ دریافت: ۲۸ تیر ۱۳۹۵ تاریخ پذیرش: ۱۹ شهریور ۱۳۹۵

Quantitative Study of ^{13}C Chemical Shifts of β-Naphthalenes Using 2D Image Approach and Density Functional Theory Computation

Zahra Garkani-Nejad[1,*], Marziyeh Poshteh-Shirani[2]
1. Department of Chemistry, Faculty of Science, Shahid Bahonar University of Kerman, Kerman, Iran
2. Department of Chemistry, Isfahan University of Technology, Isfahan, Iran
Received: 18 July 2016 **Accepted:** 9 September 2016

چکیده
روش تصویر دو بعدی برای پیش بینی جابجایی شیمیایی ^{13}C-NMR مشتقات بتا نفتالن استفاده شده است. درمطالعه آنالیز چند متغیره تصویر- ارتباط کمی ساختار ویژگی (MIA-QSPR)، توصیف‌کننده‌ها نقاط دو بعدی ساختارهای شیمیایی می‌باشند. تغییر درساختار (استخلاف) موجب تغییر در ویژگی (جابجایی شیمیایی) می‌شود. در این مطالعه، جابجایی شیمیایی ^{13}C-NMR ۱۰ موقعیت کربن مربوط به ۲۴ بتا نفتالن تک استخلافی پیش‌بینی شده است. بر روی توصیف‌کننده‌های حاصل آنالیز اجزاء اصلی صورت گرفته و مهمترین اجزاء اصلی استخراج شده‌اند. سپس مدلسازی MIA-QSPR با روش‌های رگرسیون اجزاء اصلی (PCR) و اجزاء اصلی- شبکه عصبی مصنوعی (PC-ANN) انجام شده است. برای انتخاب مناسب‌ترین اجزاء اصلی به‌عنوان ورودی برای دو روش رگرسیون اجزاء اصلی و اجزاء اصلی- شبکه عصبی مصنوعی، ازروش ترتیب همبستگی استفاده شده است. همچنین جابجایی شیمیایی ^{13}C ترکیبات مورد مطالعه با استفاده از تئوری تابع چگالی و روش B3LYP/6-311++ G (GIAO) پیش‌بینی شده است. عملکردGIAO با دو روش PCR و PC-ANN مقایسه شده است. نتایج ارجحیت مدل PC-ANN را نسبت به دو روش GIAO و PCR نشان می‌دهد. نهایتا" جابجایی شیمیایی ترکیبات با استفاده از برنامه ChemDraw محاسبه شده است.

واژه‌های کلیدی
آنالیز چند متغیره تصویر؛ تئوری تابع چگالی؛ جابجایی شیمیایی ^{13}C؛ بتا نفتالن.

Abstract
A 2D image approach has been used to predict ^{13}C NMR chemical shifts of β-naphthalene derivatives. In multivariate image analysis-Quantitative structure property relationship (MIA-QSPR) study, descriptors correlating with dependent variable are pixels (binaries) of 2D chemical structures; Variant pixels in the structures (substitutes) account to explained variance in the property (chemical shifts). A case study is carried out in order to predict ^{13}C NMR chemical shifts of 10 carbon positions of 24 mono substituted β-naphthalenes. The resulted descriptors were subjected to principal component analysis (PCA) and the most significant principal components (PCs) were extracted. Then, MIA-QSPR modeling was done by means of principal component regression (PCR) and principal component –artificial neural network (PC-ANN) methods. A correlation ranking procedure is proposed here to select the most relevant set of PCs as inputs for PCR and PC-ANN modeling methods. Here, the ^{13}C chemical shifts of studied compounds were predicted using density functional theory (DFT) calculations, too. The widely applied method of gauge included atomic orbital (GIAO) B3LYP/6-311++ G have been used. The performance of the GIAO was also compared with PCR and PC-ANN models. Results showed the superiority of the PC-ANN over GIAO and PCR models. Finally, ^{13}C NMR chemical shifts of studied compounds were calculated using ChemDraw program.

Keywords
Multivariate Image Analysis; Density Functional Theory; ^{13}C Chemical Shift; β-Naphthalenes.

*Corresponding Author: z_garkani@uk.ac.ir

1. INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is undoubtedly one of the most important methods for elucidating complicated structures and processes, including structural configuration [1-2], reaction mechanisms, dynamic processes, chemical equilibrium and even three-dimensional structures of protein molecules in aqueous solution [3]. However, it is commonly not enough to obtain all structural parameters from experiments, due to the diverse natures of the structures and reactivities. Various different approaches have been developed and tested for chemical shift calculation, however, the most widely used technique is the Gauge Including Atomic Orbital (GIAO), calculation of NMR chemical shifts at the Density Functional Theory (DFT) B3LYP (Becke-3-Lee-Yang-Par) 6-311+G(2d,p) level which is suitable for organic molecules of medium size with a molecular weight usually less than 1000. Relatively accurate values of ^{1}H chemical shifts can be achieved using this technique, but predictions of ^{13}C chemical shifts are rather poor and do not always allow unambiguous solution of stereo chemical problems [4].

Chemometric tools, particularly principal component analysis (PCA), have been used to determine the ruling effects of substituents on chemical shifts [5]. However, few reports are dedicated to apply multivariate tools for quantitative NMR, determination of chemical shifts by using regression methods [6-8]. Quantitative structure-activity/property relationship (QSAR/QSPR) models have been introduced for calculating the physicochemical properties with various numerical descriptors of chemical structures. These relationships derive correlations between the similarities of individual compounds and their biological activity/chemical property [9-11].

A 2D image-based methodology recently developed (MIA-QSPR, multivariate image analysis applied to quantitative structure- property relationship) [12-13], which avoids conformational screening and 3D alignment steps, has also demonstrated to be nicely predictive with some operational advantages. Thus, this method can be used for predicting NMR spectra. In a previous work, we applied MIA-QSPR model for predicting ^{13}C NMR chemical shifts of α- mono substituted naphthalenes [14]. We proposed a correlation ranking- principal component - artificial neural network (CR-PC-ANN) method. CR-PC-ANN algorithm that selecting of PCs is based on correlation ranking for PC-ANN showed better results than the eigen-value ranking method. Goodarzi et al. have reported a quantitative structure–property relationship study on the ^{13}C chemical shifts of methoxyflavonol derivatives using MIA-QSPR method [15]. They revealed that the predictive ability of MIA descriptors is comparable or even superior to the Gauge Included Atomic Orbital (GIAO) procedure for ^{13}C chemical shift calculations. Geladi and Esbensen [16] have demonstrated that image analysis may provide useful information in chemistry, though the descriptors do not have a direct physicochemical meaning, since they are binaries. In QSPR, images (2D chemical structures) have shown to contain chemical information [17-18], allowing the correlation between chemical structures and properties.

In the present work, we use of 2D images, which are the proper structures of the compounds that can be drawn with aid of any appropriate program, as descriptors in QSPR. Then, multivariate image analysis-quantitative structure property relationship study (MIA-QSPR) is proposed to model and predict the ^{13}C chemical shifts of a series of β-naphthalene derivatives [19] using principal component regression (PCR) and principal component- artificial neural network (PC-ANN) modeling methods. Also, density functional theory (DFT) calculations have been used for calculating ^{13}C NMR chemical shift of compounds. Finally obtained results using different methods are compared.

2. EXPERIMENTAL

2.1 Data set

All data of the present investigation were obtained from literature [19]. This data set consists of 24 mono substituted β-naphthalenes. ^{13}C chemical shifts of studied compounds in ppm relative to TMS have been reported. The chemical structure of these compounds and their ^{13}C chemical shifts has been listed in Table 1. For PCR, data set was divided randomly into two groups of training set (16 compounds) and validation set (8 compounds) and for ANN modeling, data set was divided into three groups of training set (16 compounds), validation set (4 compounds) and test set (4 compounds).

2.2 MIA

MIA descriptors are binaries obtained from pixels of 2D chemical structures, which these pixels are correlated with dependent variables for making QSPR models. The 2D structures of each compound of Table 1 were systematically drawn in the ChemSketch module of ACDLabs program [20] and subsequently saved as bitmap files. The workspace size in each 2D image saved was 205×120 pixels and aligned by a common point

among them in a fixed window of 105×65 pixels, as illustrated in Fig. 1. The molecules used in the model should have some similarity in molecular structure so that one can proceed with calibration, in this case using a congeneric series. The 2D images were read and converted into double array by using Matlab [21]. Each image of dimension 205×120 pixels was unfolded to a 1×24600 row and then the 24 images were grouped to from 24×24600 matrix. Columns with zero variance were removed to minimize memory, reducing the size of matrices to 24×887.

2.3 PCR

In QSAR/QSPR studies, regression model of the form y = Xb + e may be used to describe a set of predictor variables (X) with a predicted variable (y) by means of regression vector (b). However, the colinearity, which often existed between independent variables, creates a severe problem in certain types of mathematical treatment such as matrix inversion [22]. Sometimes, the dimension of the input matrix is large, but the components of the matrices are highly correlated (redundant). It is useful in this situation to reduce the dimension of the input matrix. PCA is an effective procedure for reducing the dimensionality of large data sets. It permits identification of associations between variables, therefore reducing the dimensionality of the data set [23-25]. Different methods have been addressed to select the significant PCs for calibration purposes. In the most common one which is called correlation ranking, the factors are ranked by their correlation coefficient with the property to be correlated (a dependent variable) [26]. The factor with highest correlation coefficient is considered as the most significant one and, subsequently, the factors are introduced in to the calibration model until no further improvement of the calibration model is obtained.

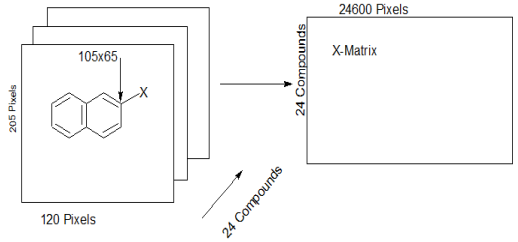
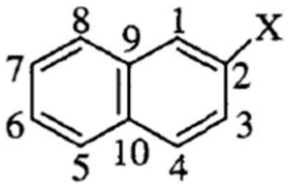


Fig. 1. 2D images and unfolding step of the 24 chemical structures to give the X-matrix.

Table 1. The structure of β-naphthalene derivatives and experimental ^{13}C chemical shifts for 10 carbon positions.

No.	Substituent X	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
1	H	128	125.9	125.9	128	128	125.9	125.9	128	133.6	133.6
2	CH_3	126.7	135.2	127.9	127.2	127.5	124.8	125.7	127.4	133.5	131.6
3	$C(CH_3)_3$	124.7	148.4	122.9	127.6	128	125.2	125.7	127.4	134	132.3
4	CH2Br	126.3	134.9	127.8	127.6	127.5	126.6	126.2	128.6	133	132.9
5	CH_2OH	125.3	138.2	121.5	127.9	127.6	125.7	126	127.8	133.3	132.8
6	CF_3	126	127.2	121.7	129.1	128.1	128.3	127.4	129.1	132.5	134.9
7	F	111	160.8	116.3	130.4	128	125.2	127	127.4	134.3	130.6
8	Cl	126.6	131.6	126.7	129.5	127.8	126.1	127	126.9	134.3	131.7
9	Br	129.8	119.7	129	129.5	127.7	126.1	126.7	126.9	131.6	134.3
10	I	137.2	91.8	134.9	130.3	128.5	127.2	127.4	127.4	135.7	132.8
11	OH	109.4	153.2	117.6	129.8	127.7	123.5	126.4	126.3	134.5	128.9
12	OCH_3	105.8	157.7	118.8	129.5	127.7	123.7	126.4	126.8	134.6	129.3
13	$OCOCH_3$	118.5	148.4	121.1	129.3	127.6	125.6	126.5	127.6	133.7	131.4
14	NH_2	107.4	142.6	117	127.8	126.4	121.1	126.8	124.5	133.5	126.6
15	$N(CH_3)_2$	106.9	149.5	117.1	129.2	128	122.5	126.6	126.9	136	127.7
16	NH_3^+	122.1	125.6	119.4	131.2	128.2	128.2	127.9	128.2	133.7	133.3
17	NO_2	124.6	145.9	119.2	129.7	128.1	129.9	128.1	130.1	132.5	136
18	CN	133.8	109.2	126	129	127.8	128.9	127.5	128.2	132	134.3
19	CHO	134.2	133.8	122.3	128.8	127.7	128.8	126.8	129.8	136	132.2
20	$COCH_3$	129.9	134.2	123.7	128.2	127.6	128.2	126.6	129.4	135.4	132.3
21	COOH	130.7	128.3	125.3	128.2	127.7	128.3	126.8	129.3	132.3	135.1
22	$COOCH_3$	131	127.7	125.4	128.2	127.9	128.3	126.8	129.4	132.6	135.5
23	COCl	130.5	135	125.2	128.2	127.6	128.1	126.7	129.2	133.6	132.2
24	$Si(CH_3)_3$	133.8	137.8	129.8	127	128.1	126.2	125.7	128.1	133.1	133.8



In the present work, First PCA was carried out on data matrix using Minitab program [27]. After achieving PCs, PCR analysis including correlation ranking based-PCR (CR-PCR) was employed. In the CR- PCR procedure, the scores of PCs were entered to the PCR model, consecutively, based on decreasing their correlation with the ^{13}C chemical shifts. $R^2 \geq 0.75$ was used to select the optimum number of PCs in the PCR models. For regression analysis, data set was separated into two groups: training set including 16 compounds and validation set including 8 compounds. Training set was used for the construction of the PCR models and then the generated models were applied to the validation set. Obtained models were summarized in Table 2 for CR-PCR method. In all 10 CR-PCR equations, the factor with highest correlation coefficient with the ^{13}C chemical shifts was considered as the most significant one and, subsequently, the factors were introduced into the calibration model until $R^2 \geq 0.75$ is achieved. PCs with higher correlation have greater information about the variation in the ^{13}C chemical shifts. Calculated ^{13}C chemical shifts using CR-PCR equations were shown in Table 3.

2.4 ANN

Artificial neural network (ANN) is a computer based system derived from the simplified concept of the brain in which a number of nodes, called processing elements or neurons, are interconnected in a netlike structure. The ANN characteristics have been found to be nonlinear making them suitable for data processing in which the relationship between cause and results cannot be linearly defined. Three components constitute an ANN: the processing elements, the topology of connection between the nodes, and the learning rules. An ANN is arranged into discrete layers, consisting of input, hidden and output, each of which includes one or more individual nodes or processing elements. The number of input variables necessary for predicting the desired output variable determines the number of input nodes. The optimum number of hidden nodes and hidden layers is dependent on the complexity of the modeling problem [28].

Table 2. CR-PCR models for C1-C10 positions of β-naphthalene derivatives.

Position	Model
1	124.175(±1.053) -0.356 PC_2(±0.114) +0.667 PC_4(±0.130) +0.376 PC_5(±0.155) +0.336 PC_6(±0.168) +0.484 PC_{12}(±0.220) -0.673 PC_{14} (±0.253) -0.510 PC_{19}(±0.324)
2	135.108(±1.719)-0.561 PC_3(±0.194) -0.827 PC_5(±0.253) -0.896 PC_7(±0.278) +0.998 PC_8(±0.289)-0.710 PC_{10}(±0.312) +1.132 PC_{14}(±0.413) -0.661 PC_{17}(±0.473)+0.975 PC_{19}(±0.528)+0.913 PC_{21}(±0.577)
3	123.437(±0.572) +0.142 PC_3(±0.064) +0.283 PC_4(±0.071) +0.209 PC_6(±0.091)+0.175 PC_7(±0.092)-0.247 PC_8(±0.096) +0.272 PC_{12}(±0.119) +0.218 PC_{18}(±0.165) -0.417 PC_{19}(0.175) -0.162 PC_{20}(±0.183) -0.315 PC_{21}(±0.192) -0.200 PC_{23}(±0.229)
4	128.800(±0.147) +0.061 PC_7(±0.024)-0.057 PC_9(±0.025) -0.073 PC_{11}(±0.027) -0.044 PC_{12}(±0.031)-0.087 PC_{14}(±0.035) +0.059 PC_{15} (±0.037) +0.089 PC_{16}(±0.037) +0.045 PC_{17}(±0.040) -0.035 PC_{18}(±0.042) +0.071 PC_{20}(±0.047)
5	127.783(±0.041) +0.015 PC_4(±0.005) -0.025 PC_{10}(±0.007) -0.027 PC_{11}(±0.008) -0.224 PC_{14}(±0.010) +0.028P C_{16}(±0.011) -0.031 PC_{19} (±0.013) +0.047 PC_{20}(±0.013) -0.033 PC_{21}(±0.014)
6	126.350(±0.281) +0.120 PC_4(±0.035) +0.086 PC_5(±0.041) +0.062 PC_7(±0.045) +0.080 PC_9(±0.049) -0.074 PC_{11}(±0.053) +0.089 PC_{12} (±0.059) -0.256 PC_{14}(±0.068) +0.156 PC_{16}(±0.072) -0.068 PC_{17}(±.077) +0.095 PC_{20}(±0.090) -0.067 PC_{21}(±0.094)
7	126.692(±0.098) +0.041 PC_7(±0.016) -0.021 PC_{11}(±0.018) +0.034 PC_{13}(±0.023) -0.081 PC_{14}(±0.024) +0.049 PC_{16}(±0.025) +0.027 PC_{20} (±0.031) -0.005 PC_{21}(±0.033)
8	127.946(±0.155) -0.058 PC_3(±0.017) +0.083 PC_4(±0.019) +0.061 PC_9(±0.027) -0.042 PC_{11}(±0.029) +0.039 PC_{12}(±0.032) -0.118 PC_{14} (±0.037) +0.056 PC_{16}(±0.040) +0.060 PC_{20}(±0.050) +0.041 PC_{23}(±0.062)
9	133.721(±0.149) -0.058 PC_9(±0.026) -0.059 PC_{12}(±0.031) -0.061 PC_{16}(±0.035) +0.141 PC_{15}(±0.038) -0.072 PC_{17}(±0.041) -0.123 PC_{18} (±0.043) -0.067 PC_{19}(±0.046) -0.170 PC_{22}(±0.055)
10	132.337(±0.324) -0.059 PC_2(±0.035) +0.139 PC_4(±0.040) +0.099 PC_5(±0.048) -0.088 PC_8(±0.055) -0.024 PC_{10}(±0.059) -0.071 PC_{11}(±0.061) +0.103 PC_{12}(±0.068) -0.170 PC_{14}(±0.078) -0.137 PC_{15}(±0.082) +0.170 PC_{16}(±0.083) +0.161 PC_{20}(±0.104)

Table 3. Calculated values of ^{13}C chemical shifts using CR-PCR method for all 10 carbon positions – Training and Validation sets.

Substituent X	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
Training										
H	123.684	134.591	124.278	128.500	128.071	126.397	126.160	127.513	134.081	133.656
CH_3	121.595	134.702	128.701	127.793	127.618	124.626	126.010	127.546	133.660	131.081
$C(CH_3)_3$	132.990	150.155	120.905	127.649	127.598	126.839	125.699	127.972	133.174	132.919
CH_2Br	125.560	142.204	125.828	127.705	127.408	125.015	125.635	127.718	133.161	131.466
CH_2OH	122.433	141.350	119.766	128.505	127.527	124.274	126.110	127.327	133.498	132.969
Cl	127.170	127.211	128.152	129.315	127.685	126.820	127.209	127.993	134.341	131.901
Br	130.536	124.427	126.721	129.154	127.620	126.821	127.229	126.878	132.513	135.094
OCH_3	113.291	145.881	120.480	128.974	127.581	124.412	126.329	127.592	134.093	129.856
$OCOCH_3$	123.724	146.459	122.354	129.307	127.693	125.843	126.502	127.755	134.575	132.194
NH_2	103.210	133.616	117.154	127.341	126.582	121.012	125.420	124.774	133.725	125.844
NH_3^+	117.042	120.678	120.592	131.404	128.109	128.557	128.138	128.453	133.940	131.624
NO_2	121.867	139.155	120.290	129.280	128.19	127.851	127.656	129.404	132.645	134.866
CN	126.603	104.104	126.566	128.706	127.605	129.448	127.042	128.989	131.843	134.948
$COCH_3$	130.119	136.255	120.625	128.743	127.718	126.115	126.484	128.634	135.136	133.051
$COOCH_3$	131.127	131.072	127.638	128.423	128.085	128.261	126.592	129.157	132.673	133.004
$Si(CH_3)_3$	132.750	140.994	132.205	127.872	128.127	126.100	126.482	128.639	133.178	132.900
Validation										
CF_3	126.686	122.087	122.872	128.770	128.156	127.961	127.048	128.381	133.461	134.963
F	112.538	163.405	119.353	129.912	128.068	125.259	126.921	128.120	133.888	131.423
I	134.206	100.322	130.549	130.702	128.191	127.177	126.883	126.885	136.115	133.557
OH	114.147	140.509	117.801	128.471	127.812	124.200	126.728	125.718	133.368	129.242
$N(CH_3)_2$	113.552	148.248	118.548	128.323	127.889	124.328	126.194	127.582	134.787	130.271
CHO	135.776	141.205	120.079	129.244	127.700	128.313	127.153	129.803	135.519	133.209
COOH	133.997	139.947	123.545	128.859	127.898	128.022	126.999	129.249	131.728	133.157
COCl	125.589	132.722	127.495	128.245	127.868	128.749	126.576	128.615	134.197	132.903

The PC-ANN, which combines the PCA with ANN and models the non-linear relationships between the PCs and dependent variable, was proposed by Gemperline to improve the training speed and decrease the overall calibration error [29]. At the present work, we used the PCs which were selected by CR-PCR method as input variables of ANN. An artificial neural network with back-propagation algorithm was constructed. Our network had an input layer, a hidden layer and an output layer. The input vectors were the set of PCs which were selected by correlation ranking procedure. The number of nodes in the input layer depended on the number of PCs in the PCR equations. The number of nodes in the hidden layer was optimized through learning procedure. The training, validation and test datasets including 16, 4 and 4 compounds, respectively, were used to optimize the network performance. Obtained results using CR-PC-ANN method were shown in Table 4. For comparison, R^2 and standard error (SE) of different models for training, validation and test sets were summarized in Table 5.

2.5 DFT

Density functional theory (DFT) calculations have been used extensively for calculating a wide variety of molecular properties such as equilibrium structure, charge distribution, FTIR and NMR spectra, and provided reliable results which are in agreement with experimental data [30]. DFT calculations were carried out using Gaussian 03 [31]. NMR chemical shifts were computed at the B3LYP/6-311++G (2d, p) level using the GIAO method [32], but for compound no. 10, (B3LYP/LanL2DZ) level was used. The chemical shift values are given relative to TMS calculated values at the same level of theory. For compound no. 10, TMS calculated values with (B3LYP/LanL2DZ) level were used. Obtained results using DFT method were shown in Table 6.

Table 4. Calculated values of ^{13}C chemical shifts using CR-PC-ANN method for all 10 carbon positions – Training, Validation and Test sets.

SubstituentX	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
TrainingH	125.082	125.594	125.131	128.178	128.103	125.901	125.695	127.839	133.428	134.223
CH_3	128.292	135.236	127.4	127.246	127.411	124.639	125.606	127.558	133.549	131.653
$C(CH_3)_3$	121.817	148.175	122.207	127.768	127.932	124.384	125.759	127.875	133.909	132.219
CH_2Br	122.959	134.689	126.751	127.633	127.575	126.408	126.443	128.638	132.931	133.019
CH_2OH	125.545	138.099	121.093	127.918	127.516	125.719	126.340	127.800	133.239	132.982
Cl	128.953	127.233	126.827	129.505	127.813	125.971	127.117	127.135	134.226	130.991
Br	129.040	127.068	127.838	129.489	127.669	125.532	126.889	126.692	131.972	135.341
OCH_3	107.950	157.418	118.963	129.489	127.684	123.706	126.544	126.724	134.669	129.781
$OCOCH_3$	120.299	157.700	120.543	129.374	127.737	125.550	126.538	128.000	133.800	132.197
NH_2	107.350	142.502	117.067	127.874	126.347	121.201	125.688	124.291	133.522	126.693
NH_3^+	124.309	124.303	119.160	131.063	128.109	128.102	127.885	128.055	133.806	132.938
NO_2	125.454	153.793	119.004	129.752	128.183	129.785	128.069	130.043	132.384	135.346
CN	131.245	109.206	125.437	129.267	127.731	128.767	127.614	128.548	132.030	134.009
$COCH_3$	128.601	133.395	123.219	128.232	127.569	127.896	126.722	129.059	134.972	132.782
$COOCH_3$	133.331	127.175	124.628	128.355	127.869	128.160	126.870	129.035	132.793	134.994
$Si(CH_3)_3$	132.762	138.047	127.452	127.296	128.009	125.985	126.303	128.316	133.039	133.121
Validation										
CF_3	126.868	132.350	121.604	129.090	128.077	128.088	127.369	129.084	132.372	134.886
I	132.719	88.9821	134.614	130.279	128.498	127.162	127.395	127.353	135.700	133.504
$N(CH_3)_2$	106.944	149.500	116.744	129.150	128.013	122.540	126.629	126.889	135.931	127.479
COCl	133.593	130.725	124.980	128.167	127.601	128.018	126.700	129.026	133.582	132.913
Test										
F	109.994	160.800	115.901	130.354	128.000	125.097	126.976	127.497	134.306	130.393
OH	113.520	152.269	117.530	129.784	127.676	123.498	126.399	126.323	134.470	128.866
CHO	135.926	129.562	122.116	128.803	127.707	128.573	126.795	129.793	135.904	131.646
COOH	133.869	127.383	124.838	128.184	127.700	128.259	126.795	129.326	132.307	134.425

Table 5. The statistical parameters for CR-PCR, CR-PC-ANN, DFT/GIAO and ChemDraw program.

Method	Set	C-1		C-2		C-3		C-4		C-5		C-6		C-7		C-8		C-9		C-10	
		R^2	SE	R^2	SE	R^2	SE	R^2	SE	R^2	SE	R^2	SE	R^2	SE	R^2	SE	R^2	SE	R^2	SE
PCR	Training	0.707	4.379	0.768	5.722	0.817	1.857	0.855	0.385	0.859	0.152	0.750	1.054	0.599	0.490	0.783	0.528	0.790	0.405	0.806	1.026
	Valid	0.924	2.958	0.860	7.551	0.862	1.825	0.448	0.687	0.704	0.655	0.943	0.481	0.245	0.290	0.817	0.603	0.714	0.791	0.808	0.891
PC-ANN	Training	0.951	1.933	0.924	3.792	0.985	0.451	0.993	0.092	0.969	0.078	0.989	0.235	0.944	0.188	0.960	0.263	0.992	0.085	0.944	0.542
	Valid	0.940	3.713	0.967	5.740	0.999	0.135	0.999	0.017	0.998	0.018	0.999	0.077	0.998	0.018	0.997	0.077	0.986	0.217	0.981	0.545
	Test	0.973	2.704	0.991	1.875	0.998	0.210	0.999	0.020	0.992	0.016	0.999	0.093	0.999	0.008	0.999	0.047	0.993	0.19	0.998	0.109
DFT/GIAO	All	0.668	5.192	0.517	9.282	0.561	4.455	0.494	1.440	0.012	0.734	0.550	1.874	0.369	1.419	0.474	1.397	0.008	1.490	0.330	2.295
Chemdraw	All	0.976	1.500	0.995	1.135	0.847	1.907	0.477	0.778	0.985	0.048	0.927	0.610	0.505	0.534	0.960	0.266	0.121	1.367	0.920	0.674

Table 6. Calculated values of ^{13}C chemical shifts using DFT/GIAO method.

Substituent X	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
H	133.173	131.836	131.836	133.174	133.173	131.836	131.835	133.175	141.377	141.377
CH_3	133.902	142.833	133.978	133.791	134.130	130.301	132.036	133.367	140.191	138.056
$C(CH_3)_3$	130.028	155.898	131.419	133.932	133.574	130.808	131.692	134.564	140.024	137.840
CH_2Br	129.507	140.812	130.665	133.924	134.198	131.873	132.513	132.541	139.207	139.573
CH_2OH	128.412	145.052	128.771	133.617	132.982	131.864	130.774	134.307	139.702	138.920
CF_3	134.486	136.495	127.434	134.143	134.246	133.756	132.925	135.793	139.111	140.614
F	117.488	171.313	121.927	136.279	134.576	130.870	133.121	133.383	141.917	137.798
Cl	132.764	148.491	133.036	135.099	133.790	132.109	133.156	133.547	142.018	139.260
Br	136.325	146.165	135.837	135.035	134.209	132.267	132.966	133.521	140.976	138.332
I	140.509	146.082	138.644	133.255	132.532	131.206	131.652	131.629	136.387	133.859
OH	115.562	162.126	120.135	135.443	134.039	129.019	132.900	132.612	141.939	134.943
OCH_3	118.860	165.283	114.507	135.229	133.979	128.926	132.612	132.293	141.667	135.013
$OCOCH_3$	123.894	157.817	128.374	134.500	134.200	131.445	132.423	133.911	140.623	137.667
NH_2	136.544	149.898	135.253	133.729	133.740	131.293	131.968	134.067	140.580	138.324
$N(CH_3)_2$	111.131	154.293	119.007	135.115	133.944	127.162	132.329	132.446	141.790	132.925
$NH3^+$	125.952	125.970	117.404	142.975	136.452	140.686	140.323	134.630	138.366	140.833
NO_2	124.979	159.361	124.325	135.982	134.599	133.419	133.941	135.116	139.130	138.477
CN	141.386	116.468	134.084	135.020	134.625	134.919	133.379	135.566	139.543	141.540
CHO	135.622	139.342	134.645	134.456	133.999	134.795	132.632	137.934	139.345	143.233
$COCH_3$	136.958	139.106	131.191	133.703	133.789	134.189	132.407	137.975	138.677	141.531
COOH	141.168	131.806	132.308	133.501	133.977	134.364	132.350	137.168	138.594	141.769
$COOCH_3$	140.286	133.836	131.926	133.150	133.903	133.920	132.234	137.080	138.666	141.366
COCl	140.611	134.996	135.002	134.175	134.153	136.124	133.194	137.441	137.949	142.585
$Si(CH_3)_3$	139.810	144.815	135.842	133.207	134.737	131.809	131.588	133.702	139.703	140.294

3. RESULT AND DISCUSSION

Table 1 lists the names of the compounds used in this study and their corresponding experimental ^{13}C chemical shift values. In this list, the experimental ^{13}C chemical shift values for 10 carbon positions have been accessed. In order to find a correlation between MIA descriptors and these spectroscopic data, after eliminating the descriptors with zero variance, 887 MIA descriptors were remained. Then, PCA was applied on the descriptors data matrix. Twenty-three PCs were generated which were considered as the input variables of PCR and PC-ANN models. For each carbon position, separate PCR models based on correlation ranking were obtained. Obtained models were shown in Table 2. Calculated values of ^{13}C NMR chemical shifts using these CR-PCR equations were indicated in Table 3 for training and validation sets, respectively. The statistical parameters of these models were summarized in Table 5. To increase the predictive ability of the obtained models, a nonlinear modeling method was employed. Typically, superior models can be found using ANNs because they implement non-linear relationships and because they have more adjustable parameters than the linear models. Therefore, we suggested the use of ANN as the non-linear model. The order of PCs based on their decreasing correlation was shown in equations of Table 2. Thus, these subsets of PCs were used as input of ANN models. The calculated values of ^{13}C chemical shifts using ANN models were represented in Table 4 for training, validation and test sets, respectively. R^2 and standard error (SE) values using two different methods (CR-PCR and CR-PC-ANN) were summarized in Table 5. As can be seen from this table, CR-PC-ANN model shows more predictive ability than the PCR models. This indicates that there are nonlinear relationship between PCs and ^{13}C chemical shifts. Plots of experimental ^{13}C chemical shifts versus calculated values using CR-PC-ANN method for all 10 carbon positions are shown in Fig. 2 (a–j), respectively. As it is observed, obtained models by the CR-PC-ANN method indicate high qualities. This means that there are non-linear relationships between the proposed MIA descriptors and the ^{13}C chemical shifts of the β-naphthalene derivatives. In order to compare the predictive ability of the MIA-QSPR model with a standard procedure for ^{13}C chemical shift calculations, the GIAO method [32] was utilized to compute the spectroscopic data for the studied compounds. Obtained results using GIAO method were shown in Table 6. The results of this table show that MIA-QSPR predictions are significantly better than the GIAO method. Also, ^{13}C chemical shifts of the studied compounds were calculated using ChemDraw program [33].

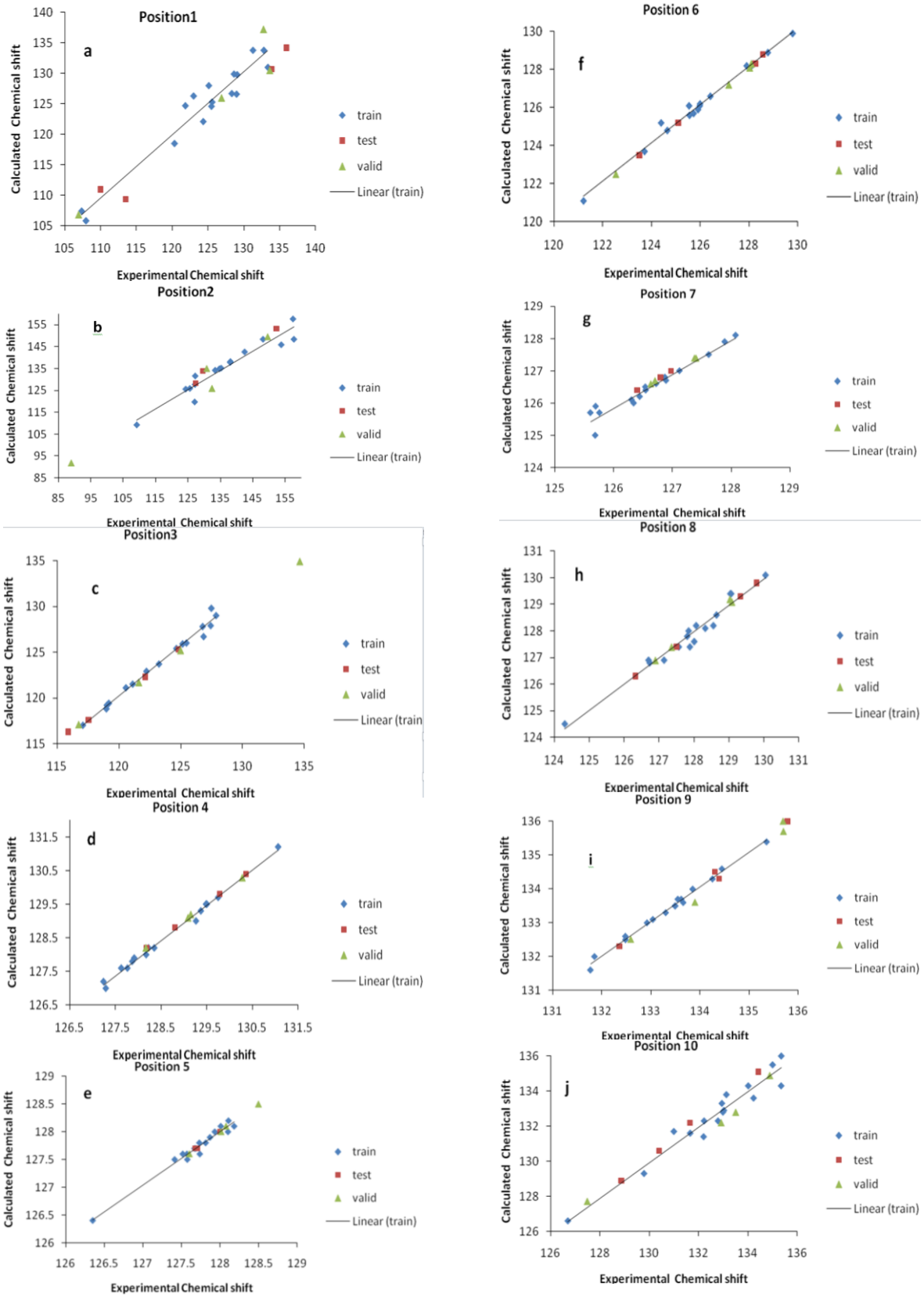


Fig. 2. (a-j): Plot of experimental ^{13}C chemical shifts of β-naphthalene derivatives against the calculated values using CR-PC-ANN model for C1-C10 positions, respectively.

Table 7. Calculated values of ^{13}C chemical shifts using ChemDraw program.

Substituent X	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
H	128	125.9	125.9	128	128	125.9	125.9	128	133.6	133.6
CH_3	127	135.3	128	127.3	127.6	125.1	126	127.5	133.7	131.8
$C(CH_3)_3$	122.9	148.4	124.7	127.7	128.1	125.5	126	127.5	134.2	132.5
CH_2Br	127	135.2	128	127.3	127.6	125.1	126	127.5	133.7	131.8
CH_2OH	127	135.2	128	127.3	127.6	125.1	126	127.5	133.7	131.8
CF_3	126.3	127.2	121.8	129.2	128.2	128.6	127.7	129.2	132.7	135.1
F	111	161	116.1	130.5	128.1	125.5	127.3	127.5	134.5	130.8
Cl	126.7	131.8	126.8	129.6	127.9	126.4	127.3	127	134.5	131.9
Br	130.1	119.9	129.4	129.6	127.8	126.4	127	127	133.6	131.6
I	137.2	91.8	134.9	130.4	128.6	127.5	127.7	127.5	135.9	133
OH	109.5	155.8	117.6	129.9	127.8	123.8	126.7	126.4	134.7	129.1
OCH_3	105.9	157.2	118.8	129.6	127.8	124	126.7	126.9	129.4	129.5
$OCOCH_3$	118.6	148.6	121.1	129.4	127.7	125.9	126.8	127.7	133.9	131.6
NH2	108.4	142.7	118.1	129	126.5	121.4	125.3	124.6	133.7	126.8
$N(CH_3)_2$	105.6	146.3	116.1	130.4	128.1	122.8	126.9	127	136.2	127.9
NH_3^+	128.1	126	126	128.1	128.1	126.2	126.2	128.1	133.8	132.8
NO_2	123.6	145.2	118.9	127.9	128.2	130.2	128.4	130.2	131.7	136.2
CN	134	109.1	126	129.1	127.9	129.2	127.8	128.3	132.2	134.5
CHO	134.2	133.3	122.8	128.9	127.8	129.1	127.1	129.9	133.1	132.4
$COCH_3$	129.9	134.2	124.2	128.3	127.7	128.5	126.9	129.5	132.5	132.5
COOH	130.4	127.1	125.8	128.3	127.8	128.6	127.1	129.4	132.5	135.3
$COOCH_3$	130.7	128	125.9	128.3	128	128.6	127.1	129.5	132.8	135.7
COCl	128.6	134.2	124.2	128.3	127.7	128.5	126.9	129.5	135.6	132.5
$Si(CH_3)_3$	133.8	137.8	129.8	127.1	128.2	126.5	126	128.2	133.3	134

Obtained values using ChemDraw program were shown in Table 7. For comparison, statistical parameters of these values were indicated in Table 5. As can be seen from this table, calculated ^{13}C chemical shifts using CR-PC-ANN models are more accurate than the calculated values by ChemDraw program.

4. CONCLUSION

The main aim of the present work was to investigate relationship between 2D images and ^{13}C chemical shifts. Obtained results indicated that though MIA descriptors do not have a direct physicochemical meaning, but may provide useful information and are capable to predict the ^{13}C chemical shifts of studied compounds.The MIA-QSPR results presented the advantage of being based on a calibration model, and thus the model may be validated and predictions performed more reliably.

REFERENCES

[1] S.Witkowski, D. Maciejewska and I. Wawer, ^{13}C NMR studies of conformational dynamics in 2,2,5,7,8-pentamethylchroman-6-ol derivatives in solution and the solid state, *J. Chem. Soc. Perkin Trans*. 2 (2000) 1471–1476.

[2] H. Neuvonen and K. Neuvonen, Correlation analysis of carbonyl carbon ^{13C} NMR chemical shifts, IR absorption frequencies and rate coefficients of nucleophilic acyl substitutions. A novel explanation for the substituent dependence of reactivity, *J. Chem. Soc. Perkin Trans.* 2 (1999) 1497–1502.

[3] K. Wu¨thrich, The way of NMR structures of proteins, Nat. Struct. Biol. 8 (2001) 923–925.

[4] A.E. Aliev, D. Courtier-Murias and S. Zhou, Scaling factors for carbon NMR chemical shifts obtained from DFT B3LYP calculations, *J. Mol. Struct.* THEOCHEM 893 (2009) 1–5.

[5] E.L. Canto, L. Tasic, R.E. Bruns and R. Rittner, Principal component analysis in studies of substituent-induced chemical shifts of 1,4-disubstituted benzenes, *Magn. Reson. Chem*. 39 (2001) 316–322.

[6] M. Holík, Principal component regression with chemical-shift incerements .1. P-disubstituted benzenes and 2- naphtyl derivatives, *Chem. Commun*. 61 (1996) 713–725.

[7] M. Jalali-Heravi, S. Masoum and P. Shahbazikhah, Simulation of 13C nuclear magnetic resonance spectra of lignin compounds using principal component analysis and artificial neural networks, *J. Magn. Reson.* 171 (2004) 176–185.

[8] R. Kiralj and M.M.C. Ferreira, Simple Quantitative Structure−Property Relationship (QSPR) Modeling of 17O Carbonyl Chemical Shifts in Substituted Benzaldehydes Compared to DFT and Empirical Approaches, *J. Phys. Chem. A* 112 (2008) 6134–6149.

[9] R.A.A. Santos, J.B. Ghasemi, C.A. Braz, R. Safavi-Sohi and E.G. Barbosa, Mixed 2D–3D-LQTA-QSAR study of a series of Plasmodium

falciparum dUTPase inhibitors, *Med. Chem. Res*. 24 (2015) 1098-1111.
[10] M. Goodarzi, T. Chen and M.P. Freitas, QSPR predictions of heat of fusion of organic compounds using Bayesian regularized artificial neural networks, *Chemom. Intell. Lab. Sys.* 104 (2010) 260-264.
[11] P.R. Duchowicz, M.V. Mirifico, M.F. Rozas, J.A. Caram, F.M. Fernandez and E.A. Castro, Quantitative Structure-Spectral Property relationships for functional groups of novel 1,2,5-thiadiazole compounds, *Chemom. Intell. Lab. Sys*. 105 (2011) 27-37.
[12] C. Duchesne, J.J. Liu and J.F. MacGregor, Multivariate image analysis in the process industries: A review, *Chemom. Intell. Lab. Sys.* 117 (2012) 116-128.
[13] J.M. Prats-Montalban, A. de Juan and A. Ferrer, Multivariate image analysis: a review with applications, *Chemom. Intell. Lab. Sys*. 107 (2011) 1-23.
[14] Z. Garakani-Nejad and M. Poshteh-Shirani, Application of multivariate image analysis in QSPR study of ^{13}C chemical shifts of naphthalene derivatives: A comparative study, *Talanta* 83 (2010) 225-232.
[15] M. Goodarzi, M.P. Freitas and T.C. Ramalho, Prediction of 13C chemichal shifts in methoxyflavonol derivatives using MIA-QSPR, *Spectrochim. Acta Part A* 74 (2009) 563–568.
[16] P. Geladi and K. Esbensen, Can image analysis provide information useful in. chemistry?, *J. Chemometr*. 3 (1989) 419-429.
[17] Z. Garakani-Nejad and M. Poshteh-Shirani, Modeling of ^{13}C NMR chemical shifts of benzene derivatives using the RC–PC–ANN method: A comparative study of original molecular descriptors and multivariate image analysis descriptors, *Can. J. Chem*. 89 (2011) 598-607.
[18] Z. Garakani-Nejad and M. Ahmadvand, Comparative QSRR modeling of nitrobenzene derivatives based on original molecular descriptors and multivariate image analysis descriptors, *Chromatographia* 73 (2011) 733-742.
[19] E. Pretsch, P. Buhlmann and C. Affolter, *Structure determination of organic compounds*, Tables of spectral data, Ch. 4, p.p. 101.
[20] ACD/ChemSketch version 11.02, Advanced Chemistry Development, Inc., Toronto, Ont., Canada (2008).
[21] MATLAB Version 7.1 Mathworks Inc. (2005) www.Mathworks.com.
[22] D.C. Montgomery and E.A. Peck, *Introduction to Linear Regression Analysis*, Wiley, New York (1982).
[23] L.I. Smith, *A tutorial on Principal Component Analysis* (2002).
[24] C.B.Y. Cordella, R. Leardi and D.N. Rutledge, Three-way principal component analysis applied to noodles sensory data analysis, *Chemom. Intell. Lab. Sys*. 106 (2011) 125-130.
[25] I. Abdel-Qader, S. Pashaie-Rad and S. Yehia, PCA-Based algorithm for unsupervised bridge crack detection, *Adv. Eng. Soft*. 37 (2006) 771-778.
[26] J. Sun, A correlation principal component regression analysis of NIR data, *J. Chemometr*. 9 (1995) 21-29.
[27] Minitab version 15.1.0.0, (2006) www.Minitab.com.
[28] N. Fjodorova, M. Vracko, A. Jezierska and M. Novic, Counter propagation artificial neural network categorical models for prediction of carcinogenicity for non-congeneric chemicals, *SAR and QSAR in Environ. Res.* 21 (2010) 57-75.
[29] P.J. Gemperline, J.R. Long and V.G. Gregoriou, Nonlinear multivariate calibration using principal components regression and artificial neural networks, *Anal. Chem.* 63 (1991) 2313-2323.
[30] W. Koch and M.C. Holthausen, *A Chemist's Guide to Density Functional Theory*, second ed., Wiley-VCH Verlag GmbH, Weinheim, (2001).
[31] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M.Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C.

Gonzalez, J.A. Pople, Gaussian03, revision B.03, Gaussian, Inc., Pittsburgh, PA (2003).
[32] R. Ditchfield, Molecular Orbital Theory of Magnetic Shielding and Magnetic Susceptibility, *J. Chem. Phys*. 56 (1972) 5688-5691.
[33] ChemDraw program, www.cambridgesoft .com.