

UDC 541.6:548.737

## PREDICTION OF MONOMER REACTIVITY PARAMETERS USING QUANTUM CHEMICAL DESCRIPTORS

X.L. Yu<sup>1,2</sup>, Z.D. Tan<sup>1</sup>, X.Y. Wang<sup>2</sup>

<sup>1</sup>College of Chemistry and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, P. R. China, e-mail: yxliang5602@sina.com.cn

<sup>2</sup>Key Laboratory of Environmentally Friendly Chemistry and Applications of Ministry of Education, College of Chemistry, Xiangtan University, Xiangtan, P. R. China

Received December, 14, 2010

To construct artificial neural network (ANN) models for the prediction of reactivity parameters ( $u$ ,  $v$ ), density functional theory (DFT) calculations are carried out for 55 vinyl monomers, at the B3LYP level of theory with a 6-31G( $d$ ) basis set. After the generation of quantum chemical descriptors, the stepwise multiple linear regression (MLR) analysis and the ANN method are used to develop quantitative structure-property relationship (QSPR) models of parameters  $u$  and  $v$ . The ANN models produced test set root-mean-square (rms) errors of 0.35 for the parameter  $u$  and 0.34 for the parameter  $v$ . Research results indicate that the QSPR models based on DFT calculations and ANN techniques are accurate and possess the ability to generalize.

**Keywords:** artificial neural network, density functional theory, monomer, radical copolymerization, QSPR, quantum chemical descriptors.

## INTRODUCTION

For binary copolymerization with monomers  $M_1$  and  $M_2$ , the copolymer equation can be expressed as [ 1 ]

$$R_p = R_m(r_{12}R_m + 1) / (r_{21} + R_m), \quad (1)$$

where  $R_m$  is equal to  $[M_1]/[M_2]$  in the monomer mixture and  $R_p$  is equal to  $[M_1]/[M_2]$  in the copolymer formed,  $r_{12}$  ( $= k_{11}/k_{12}$ ) and  $r_{21}$  ( $= k_{21}/k_{22}$ ) are the monomer reactivity ratios. Therefore it would be extremely useful to predict the values of  $r_{12}$  and  $r_{21}$ , to control copolymer composition in the copolymer [ 1 ]. The reactivity ratios would be obtained from the  $Q - e$  scheme [ 1—5 ]

$$r_{12} = \frac{Q_1}{Q_2} \exp[-e_1(e_1 - e_2)], \quad (2)$$

where  $Q_1$  and  $Q_2$  are measures of the resonance stabilization of  $M_1$  and  $M_2$  respectively and  $e_1$  and  $e_2$  are measures of their respective polarities. Although the  $Q - e$  scheme is now widely used, it has serious limitations. Recently, the revised patterns of the reactivity scheme have greatly improved both its accessibility and accuracy, which can be expressed as [ 2—5 ]

$$\log r_{12} = \log r_{1s} - u_2\pi_1 - v_2, \quad (3)$$

where  $S$  denotes the monomer styrene;  $u_2$  and  $v_2$  represent the polarity and the intrinsic reactivity of the monomer  $M_2$  respectively. The counterpart of  $e_1$  is the polarity parameter  $\pi_1$  [ 2 ].

Parameters  $u$  and  $v$  can be determined approximately from experimental data. The development of reliable quantitative structure-property relationship (QSPR) models to predict the basic parameters  $u$  and  $v$  is of real interest, particularly for new monomers for which experimental investigation would

Table 1

*Reactivity parameters  $u$  and  $v$  for 55 monomers*

No.	Monomer	exp. $u$	exp. $v$	calc. $u$ (MLR)	calc. $u$ (ANN)	calc. $v$ (MLR)	calc. $v$ (ANN)
1	2	3	4	5	6	7	8
Training set							
1	Vinyl ethyl sulfide	0.66	-0.77	0.16	0.36	-1.17	-0.82
2	Vinyl hendecanoate	-0.34	-1.33	-0.10	0.18	-1.07	-1.28
3	Vinyl phenyl sulfide	0.20	-0.58	-0.32	-0.12	-0.69	-0.41
4	Vinyl stearate	-1.19	-1.15	-0.89	-1.24	-1.02	-1.17
5	Vinyl 2-chloroethyl ether	1.18	-2.06	1.15	0.75	-1.85	-2.04
6	Styrene, p-methyl-	-0.20	0.08	-0.61	-0.51	-0.38	-0.11
7	Acrylate, a-chloro-, methyl	-1.55	0.64	-1.18	-1.19	0.28	0.38
8	Acrylate, methyl	-2.34	0.16	-2.09	-2.22	0.24	0.31
9	Acrylate, octadecyl	-3.01	0.24	-3.29	-2.82	0.59	0.30
10	Acrylamide	-1.82	-0.07	-1.74	-1.95	-0.27	-0.03
11	Methacrylonitrile	-2.08	0.44	-1.44	-1.39	0.45	0.36
12	Methacrylate, benzyl	-1.4	0.26	-1.48	-1.50	0.35	0.25
13	Methacrylate, butyl	-1.49	0.26	-1.42	-1.37	0.37	0.22
14	Methacrylate, phenyl	-2.22	0.62	-1.64	-1.76	0.49	0.41
15	Methacrylate, 2-hydroxyethyl	-1.45	0.34	-1.46	-1.44	0.23	0.26
16	Vinyl acetate	-0.44	-1.56	0.06	0.07	-1.26	-1.58
17	Methacrylate, 2-chloroethyl	-1.39	0.42	-1.67	-1.82	0.38	0.37
18	Acrylate, butyl	-2.22	0.12	-2.13	-2.24	0.38	0.33
19	Acrylate, 2-chloroethyl	-2.38	0.25	-2.35	-2.35	0.44	0.45
20	Acrylate, 2-nitrobutyl	-2.39	0.44	-2.62	-2.43	0.71	0.54
21	Naphthalene, 1-vinyl-	-0.83	0.16	-1.30	-1.20	0.23	0.29
22	Pyridine, 2-vinyl-	-0.98	0.32	-1.42	-1.34	0.18	0.26
23	Pyridine, 4-vinyl-	-0.94	0.30	-1.51	-1.51	0.43	0.39
24	Styrene, 2,5-dichloro-	-1.98	0.67	-1.58	-1.65	0.28	0.39
25	Vinyltrimethylsilane	-1.41	0.21	-1.14	-1.31	0.07	0.33
Validation set							
26	Acrolein	-2.75	0.59	-2.62	-2.38	1.01	0.53
27	Acrolein, methyl	-1.73	0.77	-2.39	-2.31	0.61	0.46
28	Isopropenyl methyl ketone	-1.91	0.42	-1.79	-2.00	0.29	0.38
29	Hexatriene, tetrachloro-	-2.39	0.86	-2.24	-2.39	1.06	0.54
30	Styrene, p-chloromethyl-	-0.62	-1.15	-1.21	-1.20	-1.12	-0.92
31	Acrylate, benzyl	-2.68	0.28	-2.23	-2.32	0.44	0.35
32	Acrylate, ethyl	-1.99	0.08	-2.08	-2.21	0.29	0.31
33	Acrylonitrile	-2.6	0.42	-2.19	-2.29	0.54	0.47
34	Methacrylate, 2-bromoethyl	-1.97	0.47	-1.65	-1.82	0.37	0.37
35	Methacrylate, glycidyl	-1.32	0.32	-1.63	-1.80	0.38	0.30
36	Methacrylate, isobutyl	-1.43	0.27	-1.44	-1.43	0.33	0.23
37	Methacrylamide, N-phenyl-	-1.26	-0.08	-1.43	-1.41	0.23	0.26
38	Styrene, p-l-(2-hydroxypropyl)-	-0.6	0.16	-0.57	-0.45	-0.5	-0.22
39	Pyridine, 2-methyl-5-vinyl-	-1.04	0.28	-0.93	-0.86	-0.2	0.11
40	Styrene, a-methyl	-0.04	-0.03	-0.36	-0.34	-0.54	-0.30
41	Vinylidene chloride	-0.34	-1.38	0.49	0.06	-0.93	-1.79

C o n t i n u e d T a b l e 1

1	2	3	4	5	6	7	8
		Test set					
42	Butadiene	-0.30	0.41	-0.92	-0.58	-0.21	-0.04
43	Butadiene-1-carboxylic acid	-2.00	0.94	-2.01	-2.23	0.29	0.48
44	Butadiene-1-carboxylate, ethyl	-2.57	0.92	-1.84	-2.05	0.37	0.49
45	Oxazoline, 2-isopropenyl-	-0.95	0.19	-0.92	-0.62	-0.42	-0.15
46	Silane, 3-methacryloxypropyl,trimethoxy-	-1.06	0.37	-1.23	-1.05	-0.4	-0.11
47	Styrene,P-acetoxy-	-0.44	0.07	-0.8	-0.77	-0.5	-0.23
48	Methacrylate, methyl	-1.18	0.23	-1.45	-1.42	0.25	0.26
49	Acrylate,3,4-epoxyhexahydrobenzyl	-1.95	0.57	-2.12	-2.25	0.54	0.35
50	Benzothiazole, vinylmercapto-	-0.38	-0.38	-0.73	-0.83	-0.44	-0.18
51	Vinyl ethyl ether	1.11	-2.00	1.66	0.75	-2.24	-2.01
52	Vinyl isothiocyanate	-1.56	0.18	-0.66	-0.97	-0.25	0.03
53	Vinyl methyl ketone	-2.46	0.54	-2.24	-2.30	0.54	0.47
54	Vinyl isocyanate	0.29	-0.91	-0.15	-0.21	-1.08	-1.35
55	Isoprene	-0.32	0.34	-0.82	-0.42	-0.37	-0.16

be expensive. The QSPR approach can conserve resources and accelerate the process of development of new molecules [ 6—11 ]. This paper is the first to produce robust back propagation artificial neural network (BPANN) models for parameters  $u$  and  $v$ . The quantum chemical descriptors used were selected by the multiple linear regression (MLR) analysis.

#### MATERIALS AND METHODS

Table 1 shows the data of parameters  $u$  and  $v$  of 55 vinyl monomers [ 5 ]. Each data set of parameters  $u$  and  $v$  was randomly divided into three groups, including training, validation, and test sets, which consist of 25, 16, and 14 monomers respectively. In ANN modeling, the training set was used to fit the parameters, the validation set was used to optimize the parameters and prevent overtraining, and the test set was used to assess the performance (i.e., generalization).

All monomers were fully optimized and the single point energy calculations were carried out with a DFT method in the Gaussian 03 program at the B3LYP level of theory with the 6-31G( $d$ ) basis set, to obtain some quantum chemical descriptors. A total of 11 descriptors were calculated, which are Mulliken and atomic polar tensor (APT) [ 12 ] charges of C<sup>1</sup> and C<sup>2</sup> ( $q_{MC^1}$ ,  $q_{MC^2}$ ,  $q_{AC^1}$ , and  $q_{AC^2}$ ), the mean positive net atomic charge in a molecule ( $q_m^+$ ), the mean negative net atomic charge in a molecule ( $q_m^-$ ), the total dipole moment ( $\mu$ ), the molecular traceless quadrupole moment ( $\Theta$ ), the molecular mean hexadecapole moment ( $\Phi$ ) [ 13 ], the energies of the highest occupied molecular orbital ( $E_{HOMO}$ ) and the lowest unoccupied molecular orbital ( $E_{LUMO}$ ) [ 14, 15 ]. In addition, the descriptor  $q_a^+$  was defined, which is equal to  $q_m^+$  for vinyl monomers and 0 for acrylate (and methacrylate) monomers respectively.

#### RESULTS AND DISCUSSION

Correlation between the parameters  $u$  and  $v$  in the training sets and 12 descriptors by the stepwise MLR analysis resulted in the following best regression equations:

$$u = -1.284 + 6.720 \cdot 10^{-5} \Phi + 18.261 E_{LUMO} - 7.413 q_{AC^1}, \quad (4)$$

$$R = 0.933, \quad R^2 = 0.871, \quad s = 0.392, \quad F = 47.370, \quad n = 25,$$

$$v = 0.453 - 18.740 E_{LUMO} - 4.526 q_m^+ + 4.223 q_a^+, \quad (5)$$

$$R = 0.947, \quad R^2 = 0.897, \quad s = 0.254, \quad F = 60.738, \quad n = 25,$$

Table 2

*Characteristics of descriptors appearing in the models*

Model	Descriptors	Coefficients	Std. Error	Sig.-test	<i>t</i> -test	VIF
<i>u</i>	Constant	-1.284	0.279	0.000	-4.602	/
	$\Phi$	$6.836 \cdot 10^{-5}$	0.000	0.000	4.146	1.010
	$E_{\text{LUMO}}$	18.261	4.826	0.001	3.784	1.793
	$q_{\text{AC}}^{\text{I}}$	-7.413	1.311	0.000	-5.656	1.787
<i>v</i>	Constant	0.453	0.414	0.287	1.093	/
	$E_{\text{LUMO}}$	-18.740	4.241	0.000	-4.418	2.959
	$q_m^+$	-4.526	1.491	0.006	-3.036	1.442
	$q_a^+$	4.223	1.049	0.001	4.024	3.689

where  $R$  is the correlation coefficient,  $s$  is the standard error,  $F$  is Fischer ratio, and  $n$  is the number of samples. The best subset of descriptors in the model of  $u$  comprises  $\Phi$ ,  $E_{\text{LUMO}}$ , and  $q_{\text{AC}}^{\text{I}}$ ; while the best subset of descriptors in the model of  $v$  consists of  $E_{\text{LUMO}}$ ,  $q_m^+$ , and  $q_a^+$ . The characteristics of the descriptors in each model are shown in Table 2, which suggests that all these descriptors are significant descriptors from the Sig-value test. All the VIF values are less than 10, which shows that multicollinearities do not exist among the descriptors in each model.

The two optimal MLR models were assessed with respective validation and test sets. The calculated values of the parameters  $u$  and  $v$  are listed in Table 1 and plotted in Fig. 1. For the parameter  $u$ , the root-mean-square (rms) errors of training, validation, and test sets are 0.36 ( $R = 0.933$ ), 0.37 ( $R = 0.897$ ), and 0.46 ( $R = 0.897$ ) respectively; for the parameter  $v$  the rms errors of three sets are 0.23 ( $R = 0.947$ ), 0.31 ( $R = 0.863$ ), and 0.48 ( $R = 0.926$ ) respectively.

Each best subset of descriptors selected from the stepwise MLR method was then fed to respective ANNs of the parameters  $u$  and  $v$  as input vectors. The optimal conditions of neural networks were obtained by adjusting various parameters by trial-and-error. The architectures of the final optimum neural networks are 3-3-1 for  $u$  and 3-5-1 for  $v$ , with the sigmoid parameter  $Q$  being 0.9, the learning rate  $\eta$  being 0.1, the momentum parameter  $\alpha_0$  being 0.6, the permission network error  $E$  being 0.00001, and the maximum number of epochs being 5000. After that, each of the optimum ANN models was assessed by the respective test set.

The results calculated from the ANN models of the parameters  $u$  and  $v$  are listed in Table 1 and depicted in Fig. 2. The calculated values for each monomer were obtained by averaging the output

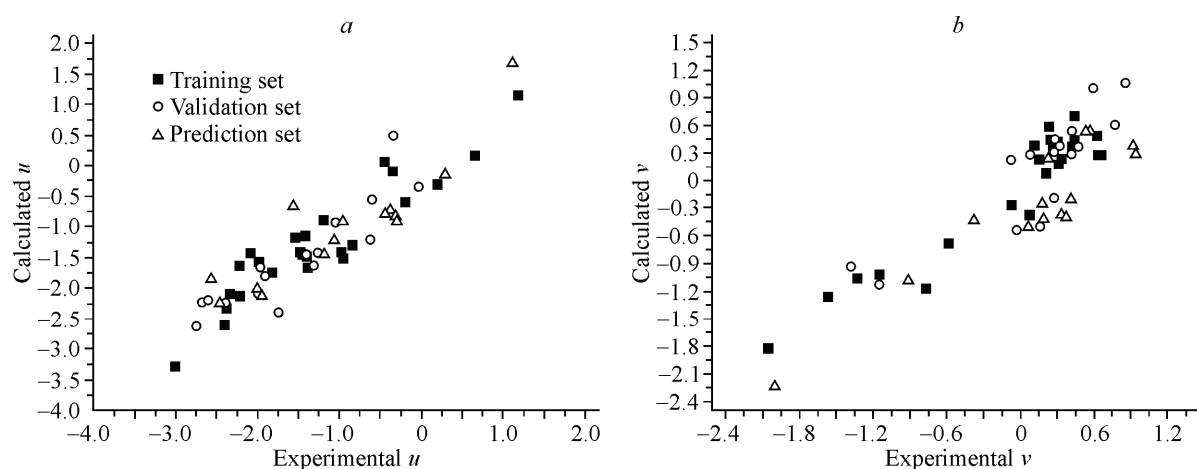


Fig. 1. Plot of the experimental versus calculated values (a) for the MLR model of  $u$ ; (b) for the MLR model of  $v$

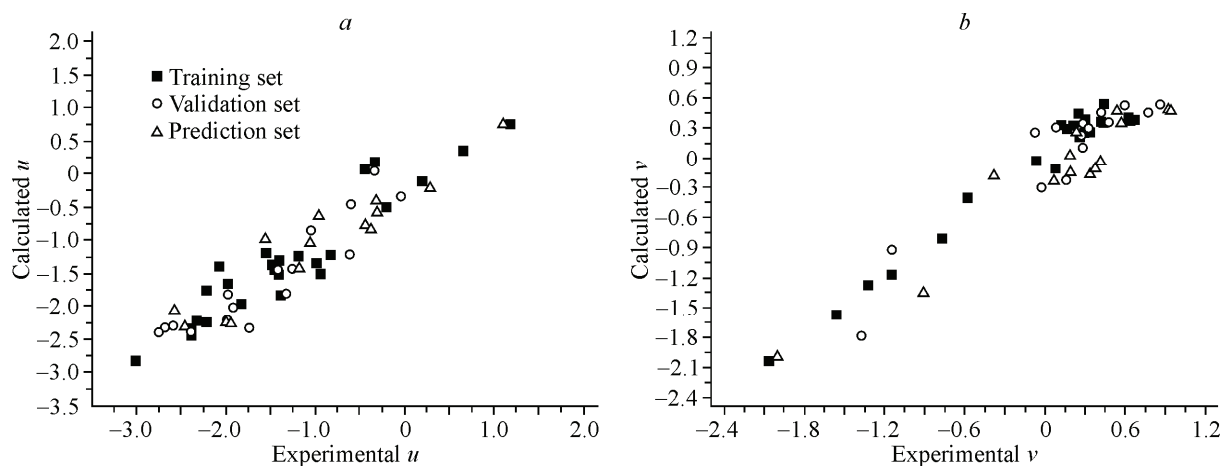


Fig. 2. Plot of the experimental versus calculated value (a) for the ANN model of  $u$ ; (b) for the ANN model of  $v$

values over 10 individual network trainings [18]. The rms errors of training, validation, and test sets for the parameter  $u$  are 0.34 ( $R = 0.943$ ), 0.32 ( $R = 0.922$ ), and 0.35 ( $R = 0.948$ ) respectively; the rms errors of three sets for the parameter  $v$  are 0.13 ( $R = 0.983$ ), 0.23 ( $R = 0.932$ ), and 0.34 ( $R = 0.957$ ) respectively. The results calculated from the ANN models are better than those from the MLR models, which denotes that the correlations between the parameters  $u$  and  $v$  and the descriptors outlined above are nonlinear rather than linear.

The parameter  $u$  represents the polarity of a monomer. According to the  $t$ -test (Table 2), the most significant descriptor appearing in Eq. 4 is APT charges of  $C^1$ :  $q_{AC^1}$ . Indeed, it has been proven that local electron densities or charges are important in many chemical reactions and the physicochemical properties of compounds and can reflect the polarity of monomers [6]. Monomers, such as vinyl ethyl ether (No.51) with an ether group, have small polarity, while ATP charges on  $C^1$  ( $q_{AC^1}$ ) of these monomers are obviously less than the others. This phenomenon denotes that  $u$  is negative correlation with  $q_{AC^1}$ . Reversely, these monomers have larger  $\Phi$  values. Thus,  $u$  is positive correlation with  $\Phi$ .

According to the frontier molecular orbital (FMO) theory of chemical reactivity [14, 15], the descriptors  $E_{HOMO}$ ,  $E_{LUMO}$  and the energy gap ( $\Delta E_g$ ) between LUMO and HOMO play major roles in governing many chemical reactions and are correlated with the stability of the molecule [20]. The descriptor  $E_{LUMO}$  is directly related to the electron affinity and characterizes the susceptibility of the molecule towards attack by nucleophiles. In fact, a decrease of the energy gap  $\Delta E_g$  usually leads to easier polarization of the molecule [6]. Therefore the parameter  $u$  is correlated with  $E_{LUMO}$ .

The parameter  $v$  describes the intrinsic reactivity of a monomer. In Eq. 5, the parameter  $v$  increases with decreasing  $E_{LUMO}$ . As stated above, the descriptor  $\Delta E_g$  is correlated with the stability of the molecule, i.e. a monomer with a small  $E_{LUMO}$  value is prone to forming a free radical and has a larger parameter  $v$  value. Acrylates and methacrylates with an unsaturated bond in side groups easily form free radicals and possess larger  $v$  values. Therefore, the descriptor  $q_a^+$  is positive correlation with  $v$ . In addition,  $v$  is negative correlation with  $q_m^+$ . The reason may be that a larger  $q_m^+$  value implies low stability for free radicals and possesses a small  $v$  value. Thus,  $v$  decreases with increasing  $q_m^+$ .

## CONCLUSIONS

ANN models have been obtained to predict the reactivity parameters ( $u$ ,  $v$ ) of the revised patterns scheme, which are useful in the prediction of reactivity ratios for vinyl monomers in radical polymerizations. The results indicate that the atom charges ( $q_{AC^1}$ ,  $q_m^+$  and  $q_a^+$ ), the energy of FMOs ( $E_{LUMO}$ ), and the molecular mean hexadecapole moment ( $\Phi$ ) are most variables to correlate with the parameters  $u$  and  $v$ .

This work was supported by the National Natural Science Foundation of China (No. 20972045), and the Open Project Program of Key Laboratory of Environmentally Friendly Chemistry and Applications of Ministry of Education, China (No. 10HJYH06).

#### REFERENCES

1. Jenkins A.D., Jenkins J. // *Macromol. Symp.* – 2001. – **174**. – P. 187.
2. Jenkins A.D. // *J. Polym. Sci. Part A: Polym. Chem.* – 1999. – **37**. – P. 113.
3. Jenkins A.D. // *Macromol. Rapid Commun.* – 1996. – **17**. – P. 275.
4. Jenkins A.D., Jenkins J. // *Polym. Int.* – 1997. – **44**. – P. 391.
5. Brandrup J., Immergut E.H., Grulke E.A. *Polymer Handbook*, 4th edn. – New York: Wiley-Interscience, 1999.
6. Karelson M., Lobanov V.S., Katritzky A.R. // *Chem. Rev.* – 1996. – **96**. – P. 1027.
7. Yu X.L., Yi B., Wang X.Y. // *Eur. Poly. J.* – 2008. – **44**. – P. 3997.
8. Yu X.L., Yi B., Liu F. *et al.* // *React. Funct. Poly.* – 2008. – **68**. – P. 1557.
9. Toropov V.O., Kudryshkin N.L., Voropaeva I.N. *et al.* // *J. Struct. Chem.* – 2004. – **45**. – P. 945.
10. Yu X.L., Yu W.H., Wang X.Y. // *J. Struct. Chem.* – 2009. – **50**. – P. 859.
11. Zhang Y.X. // *Chemomet. Intell. Lab. Syst.* – 2009. – **98**. – P. 1557.
12. Cioslowski J. // *J. Amer. Chem. Soc.* – 1989. – **111**. – P. 33.
13. Yu X.L., Yi B., Wang X.Y. *et al.* // *Chem. Phys.* – 2007. – **332**. – P. 115.
14. Fukui K. *Theory of Orientation and Stereoselection*. – New York: Springer, 1975. – P. 34 – 39.
15. Franke R. *Theoretical Drug Design Methods*. – Amsterdam: Elsevier, 1984. – P. 115 – 123.
16. Yu X.L., Yi B., Wang X.Y. // *J. Comput. Chem.* – 2007. – **28**. – P. 2336.
17. Liu P.X., Long W. // *Int. J. Mol. Sci.* – 2009. – **10**. – P. 1978.
18. Mattioni B.E., Jues P.C. // *J. Chem. Inf. Sci.* – 2002. – **42**. – P. 232.
19. Zhan C.G., Dixon D.A. // *J. Phys. Chem. A.* – 2002. – **106**. – P. 10311.
20. Zhou Z., Parr R.G. // *J. Amer. Chem. Soc.* – 1990. – **112**. – P. 5720.