J Mol Model (2008) 14:807–812
DOI 10.1007/s00894-008-0328-6

ORIGINAL PAPER

DFT and GA Studies on the QSAR of 2-aryl-5-nitro-1*H*-indole derivatives as NorA Efflux Pump Inhibitors

Yujie Dai · Xu Zhang · Xiuli Zhang · Huanjie Wang · Zhansheng Lu

Received: 18 February 2008 / Accepted: 30 May 2008 / Published online: 25 June 2008

Abstract The structures of 2-aryl-5-nitro-1*H*-indole derivatives were optimized with PM3 and DFT at b3lyp/6–31 g* level successively. Some structural and electric descriptors were obtained from the single point energy calculation and natural bond orbital analysis at the level of b3lyp/6–31 g*. As efflux pump inhibitors, a QSAR model was built with genetic algrithum (GA) and partial least square (PLS) analyses. The high R^2 and R^2_{CV} indicates the derived model has a good predictive power which can be used in prediction of activity for new 2-aryl-5-nitro-1*H*-indole derivatives. This model gives us a revelation that the activity of 2-aryl-5-nitro-1*H*-indole derivatives as efflux pump inhibitor can be improved by properly increasing the molecular volume and Mulliken atomic charge of C_3 (Q_{C3}) or lowering the dipole and Mulliken atomic charge of C_4 (Q_{C4}) in 2-aryl and it was found from this article that a QSAR relationship can be built for small samples with large descriptors by compressing the descriptors with GA and analyzing with PLS. With this model, a new compound, 2-(2-Azidomethyl-5-phenoxy-phenyl)-5-nitro-1*H*-indole was predicted to lower the MIC of berberine to 0.091 μg/mL for inhibiting K2361 of *S. aureus* with NorA efflux pump protein over expression. **Figure:** Basic structure of 2-aryl-5-nitro-1*H*-indoles

Keywords DFT · Efflux pump inhibitor · Genetic algrithum · QSAR · 2-aryl-5-nitro-1*H*-indole

Y. Dai (✉) · X. Zhang · Z. Lu
Tianjin Key Laboratory of Industrial Microbiology,
College of Biotechnology,
Tianjin University of Science and Technology,
Tianjin 300457, People's Republic of China
e-mail: yjdai@126.com

X. Zhang
College of Biochemistry, University of Missouri-Columbia,
Columbia, MO 65211, USA

H. Wang
Colledge of Chemistry and Chemical Engineering,
Shandong University,
Jinan 250100, People's Republic of China

Introduction

Efflux pump is one of the main mechanisms for the drug resistance of bacteria [1]. The over-expression of the efflux pump protein increases the pump-out effect and decreases the concentration of pharmaceuticals in bacterial cells, which leads to the decrease of the drug efficacy. Efflux pump inhibitor can inhibit the pump-out of the drugs from bacterial cells and increase the sensitivity of bacteria to drugs by combining with the efflux pump proteins. Since the discovery of reserpine, a kind of indole compound that can combine with NorA and other homologous and inhibit their pump-out effect to drugs [2], many investigations have been done on the inhibition effect of indole compounds on efflux pumps. Markham et al. [3] used the genetically engineered bacteria with NorA over expression and ethidium bromide (a kind of substrate of NorA) to build a screening model for efflux pump inhibitors. They found 370 active indole compounds from 9600 structurally diverse compounds and 7 of these active compounds were nitroindoles including the most potent nitroindole INF55. In the meanwhile, the quantitative structure-activity relationship (QSAR) studies were conducted. CoMFA (3D-QSAR) analysis showed that some related 2-aryl derivatives (for example, an o-anisole group) may be NorA pump inhibitors [4]. On this basis, some 2-aryl-5-nitro-1*H*-indole derivatives with a variety of functionalized groups in the aryl ring

were synthesized and the activity as inhibitors of the NorA efflux pump was evaluated by Samosorn et al [5]. These data can be used for establishing more accurate quantum structure activity relationships to predict more potent pump inhibitor and also give some indication for the QSAR study of other kinds of pharmaceuticals.

In a QSAR investigation, one of the key issues is to find suitable structural descriptors. Some structural and electric parameters from experiments such as the hydrophobicity parameter π [6] and Hammett constants σ [7] were used as descriptors in the early QSAR studies. However, these parameters were expensive to obtain for new compounds. In addition, some valuable descriptors can't be obtained from experiment for some compounds up to now. Therefore, the structural and electric parameters from theoretical calculation such as some energy of chemical bonds, the highest occupied molecular orbital energy (E_{HOMO}), the lowest unoccupied molecular orbital energy (E_{LUMO}) and the electric density of some atoms etc. offer a new approach to build QSAR model. Because of the huge number of these data required and the restriction of the calculation condition, these parameters were usually obtained from semi-empirical methods such as AM1 and PM3 [8–10] etc. In recent years, with the development of computer hardware and the improvement of quantum chemistry calculation methods, it is possible to obtain more accurate structural and electric parameters for QSAR processes. QSAR studies [11–13] have shown that the choice of a DFT method instead of the semi-empirical method of AM1 [14] or PM3 [15, 16] results in better correlation between calculated results and experimental data. Therefore, the DFT method is expected to give a more accurate QSAR model compared with the semi-empirical methods.

In this study, the 3D structures of ten 2-aryl-5-nitro-1*H*-indole derivatives were optimized with the semi-empirical PM3 method and density functional theory (DFT) at b3lyp/6–31 g* level successively. Some structural and electric parameters were subsequently obtained from single point energy and natural bond orbital analysis at b3lyp/6–31 g* level. On this basis, a QSAR model was built with a genetic algorithm method (GA) and partial least square analysis (PLS) for offering a theoretical reference to discover more potent NorA efflux pump inhibitors.

Theory and methods

Activities of 2-aryl-5-nitro-1*H*-indole derivatives

The activities of 2-aryl-5-nitro-1*H*-indole derivatives used in this study were taken from the work of Samosorn et al. [5] and are shown in Table 1. The MICs of berberine against NorA efflux pump protein over-expressed K2361 of *Staphylococcus aureus* strains with the existence of different 2-aryl-5-nitro-1*H*-indole derivatives were used to examine the inhibitory activity of these compounds. The lower the MIC value, the higher of the compound's inhibitory activity to NorA efflux pump is. In the process of the QSAR, the value of 50 for MIC was adopted when MIC>50 in the original literature of Ref. 4 (Table 1). The basic structures of these compounds are shown in Fig. 1 and the substituted groups are listed in Table 1.

Table 1 Structures and activities of ten 2-aryl-5-nitro-1*H*-indole derivatives used for QSAR analysis

Compound	R_1	R_2	MIC of $NorA^{++}$ K2361 plus berberine (30 μg/mL)
7a	COOH	H	50
8a	CH_2OH	H	12.5
9a	CH_2N_3	H	3
7b	COOH	OCH_3	50
8b	CH_2OH	OCH_3	6.25
9b	CH_2N_3	OCH_3	1.5
8c	CH_2OH	OCH_2Ph	0.8
10a	CH_2NH_2	H	12.5
10b	CH_2NH_2	OCH_3	12.5
INF55	H	H	3

Calculation and selection of structural and electric descriptors

The PM3 semi-empirical method was used for the initial structure optimization of the 2-aryl-5-nitro-1*H*-indole derivatives. On this basis, further structure optimizations were conducted with the density functional theory at the b3lyp/6–31 g* level in turn and the conformations with minimized energy were obtained. Their single point energy and natural bond orbital energy were calculated at b3lyp/6–31 g* level to get the molecular structural and electric parameters, the lowest unoccupied molecular orbital energy (E_{LUMO}), the highest occupied molecular orbital energy (E_{HOMO}), the polarizability (α), the nuclear-nuclear repulsion energy (E_{rep}), the molecular total energy (E_{tot}), the

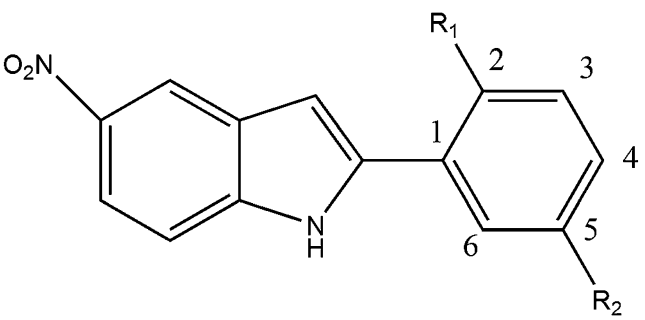

Fig. 1 Basic structure of 2-aryl-5-nitro-1*H*-indoles

dipole (μ), the molecular volume (V_{mol}), and the Mulliken atomic charges of the six carbons of the 2-aryl (Q_{C1}-Q_{C6}) obtained were employed as descriptors to construct the QSAR of the 2-aryl-5-nitro-1*H*-indole derivatives as NorA efflux inhibitors. All these structural and electric descriptors shown in Table 2 were calculated by Gaussian 03 package on Windows platform [17].

Standardization of the structural and electric descriptors

In order to avoid the effect of dimensions on the significance of the descriptors, all the descriptors used in QASR were standardized as follows:

$$x_s = (x - \bar{x})/SD \quad (1)$$

where x_s is the standardized value of a descriptor; $\bar{x}$ is the mean value of the descriptor and *SD* is the standard deviation of the descriptor.

Genetic algorithm feature selection and descriptor reduction

Under the condition of a small quantity of samples with large parameters, it is difficult to choose the adequate descriptors for predictor training in QSAR studies because there are no absolute rules that govern this choice. Recently, evolutionary algorithms and specifically genetic algorithms have been used for variable selection problems [18, 19]. Although there are 13 parameters that were obtained from the quantum chemistry calculations, only a subset of them is statistically significant in terms of correlation with the compounds' activity. In this sense, a genetic algorithm-based multi-linear regression analysis (GA-MRA) was carried out for building optimum QASR models.

The GA was generated from the apocalypse of the evolution rules, which reveals survival of the fittest [20]. This algorithm is inspired by the concept of evolution via natural selection and is based on the idea of evolving solutions to problems in a way analogous to the way organisms evolve. This method creates a series of potential solutions to a problem (the population of organisms) and then this set is iteratively modified and tested until a near-optimal solution is found. As in our study, the genetic algorithm module in materials studio 4.0 package [21] was used to reduce the number of the structural and electric parameters and to get the optimized set of descriptors. Because the sample number is small, the obtained optimized set was further analyzed by the partial least square analysis to improve the predictive power of the model.

A leave-one-out (LOO) cross-validation was used for testing the model's predictive power. In the cross-validation process, a descriptor subset of nine samples is used for training the GA model and that of the remaining one is then predicted. This process is repeated until each sample has

Table 2 Data of structural and electric descriptors of 2-aryl-5-nitro-1*H*-indole derivatives from the DFT calculation at b3lyp/6–31 g*

Compound	E_{HOMO} (ev)	E_{LUMO} (ev)	α(a.u.)	E_{rep} (Hartrees)	E_{tot} (Hartrees)	μ (Debye)	V_{mol} (cm^3/mol)	Q_{C1}	Q_{C2}	Q_{C3}	Q_{C4}	Q_{C5}	Q_{C6}
7a	−5.972	−2.098	122.21	1517.56	−987.97	9.62	172.41	0.094	−0.023	−0.143	−0.080	−0.083	−0.119
8a	−6.068	−2.014	184.77	1412.24	−913.93	8.39	166.98	0.058	0.062	−0.133	−0.077	−0.087	−0.127
9a	−5.878	−1.959	178.17	1611.43	−1002.29	8.71	200.90	0.029	0.095	−0.118	−0.075	−0.087	−0.114
7b	−5.932	−2.15	198.30	1770.87	−1102.50	11.5	202.62	0.093	−0.021	−0.155	−0.128	0.362	−0.162
8b	−5.878	−2.092	196.48	1659.47	−1028.45	9.91	208.23	0.059	0.060	−0.145	−0.127	0.358	−0.169
9b	−6.15	−2.068	190.26	1872.01	−1116.82	10.16	234.15	0.032	0.091	−0.131	−0.123	0.358	−0.157
8c	−5.66	−1.769	241.07	2321.14	−1259.51	10.12	265.47	0.060	0.060	−0.149	−0.110	0.357	−0.175
10a	−5.605	−1.714	181.87	1417.18	−894.08	10.24	191.80	0.072	0.089	−0.050	0.0145	0.003	−0.013
10b	−5.905	−1.986	194.89	1662.32	−1008.60	11.60	224.62	0.071	0.090	−0.146	−0.127	0.351	−0.153
INF55	−5.85	−2.068	164.57	1150.64	−799.40	7.80	176.57	0.093	−0.125	−0.088	−0.080	−0.092	−0.109
8d	−5.943	−2.155	254.12	2560.32	−1347.88	8.65	280.78	0.054	0.107	−0.146	−0.107	0.359	−0.172

been predicted once. The stability of the correlations was tested against the cross-validated coefficient, R^2_{cv}, which indicates the stability of an obtained model by focusing on its sensitivity to the elimination of any single data point. R^2_{cv} can be calculated from the following equation:

$$R^2_{cv} = \frac{\sum_{k=1}^{n} (y_k - \hat{y}_k)^2}{\sum_{k=1}^{n} \left[y_k - \left(\sum_{k=1}^{n} y_k \Big/ n \right) \right]} \quad (2)$$

where y_k is the desired output, and $\hat{y}_k$ is the actual output of the model and n is the number of samples in the analyzed set.

The multi-linear equation between -lgMIC and the structural and electric descriptors (Eq. 3) was taken as the object function and the square of the correlation coefficient R^2 was used as individual fitness or cost function for GA optimization process.

$$-\lg \mathrm{MIC} = k_i \sum_{i=1}^{m} x_i + b \quad (3)$$

In Equation 3, m is the number of selected descriptors used to build the QASR multi-linear relation. x_i is the selected descriptor for the and QASR. k_i is the coefficient of the corresponding descriptor and b is the constant.

The first step for the GA is to select the number of descriptors (m) in the multi-linear relationship between -lgMIC and the selected descriptors, and then a population of N individuals is created. Each individual was generated from the encode of the same number (m) of descriptors. The descriptors are randomly chosen from a data matrix in a way to make sure that (a) no two individuals have exactly the same set of descriptors and (b) all descriptors in a given individual must be different from each other. The fitness of each individual in this generation is evaluated by R^2. A fraction of the individuals with top scores of fitness was selected as parents to reproduce the next generation.

For the next step, a fraction of children of the next generation is produced by cross-over (cross-over children) and the others by mutation (mutation children) from the parents. Through the sexual and asexual reproductions the new offspring generated contains characteristics from one or both of its parents. In a sexual reproduction, two individuals are chosen probabilistically according to their scaled fitness scores and serve as parents. Then in a cross-over operation, two parent individuals provide a random selection of their partial descriptor set to exchange and a child is produced by combining the two parent of "genetic code". On the other hand, the rest of the individuals in the new generation are generated by asexual reproduction from a random mutation in one of genes of the randomly selected parents; i.e., one descriptor is replaced by another. Elitism was also included to protect the fittest individual in any given generation from cross-over or mutation during reproduction. Thus, cross-over and mutation processes are repeated until all of the N parents in the population are replaced by their children. The fitness score of each individual of this new generation is again evaluated, and the reproductive cycle is continued until all the possible descriptor combinations were tested. Some initial parameters for GA optimization are listed in Table 3.

Results and discussion

Construction of QSAR model and the variables reduction

The structural and electric parameters of 2-aryl-5-nitro-1*H*-indole derivatives are listed in Table 2. Because the main difference of these compounds lies in the substituted groups in 2-aryl, which can alter the Mulliken atomic charge distribution of 2-aryl, the Mulliken atomic charges of the six carbons of the 2-aryl, Q_{C1}, Q_{C2}, Q_{C3}, Q_{C4}, Q_{C5}, Q_{C6} and some other descriptors usually considered in QSAR such as E_{LUMO}, E_{HOMO}, α, E_{rep}, E_{tot}, μ and V_{mol} coming from the quantum chemistry calculation were used as candidate descriptors. -lgMIC of the substituted 2-aryl-5-nitro-1*H*-indole was taken as the dependent variable and the partial parameters m selected from the 13 descriptors were taken as the independent variables to build the QSAR model. In order to obtain the optimum data set for constructing the QASR model and omit the insignificant descriptors, m variables were selected as data set for each time to build QASR. There are C^m_{13} kinds of variable's combinations. Then the optimum combinations containing the same numbers of descriptors were screened out with GA method. For each m of 2–7, two combinations of the descriptors with the top square correlation coefficients R^2 are listed in Table 4.

In our study, R^2, R^2_{CV} and F (a=0.05) were used to evaluate the regression model. It can be seen from the F value in Table 4 that all the descriptor sets except for that of two descriptors of V_{mol} and Q_{C3} were in the confidence interval of 95%. With the increase of the descriptor number (m) of the data sets, R^2 value increases, however, R^2_{CV} increases first, but decreases subsequently. When m comes to 5, the highest R^2_{CV} value was obtained with a good squared correlation coefficient of 0.9982. This indicates that further increasing of the linear correlation with the descriptor numbers was acquired at the expense of

Table 3 Initial parameters for GA optimization

Population	Maximum generations	Scoring function	Mutation probability
50	500	*R*-squared	0.100

Table 4 Optimum descriptor combinations with different *m*

m	Best variable set							R^2	R^2_{CV}	F	F>$F_{0.05}$
2	μ	V_{mol}						0.9327	0.8062	48.5	Yes
2	V_{mol}	Q_{C5}						0.6184	0.2891	5.67	No
3	E_{rep}	μ	V_{mol}					0.9785	0.9152	90.9	Yes
3	E_{tot}	V_{mol}	Q_{C1}					0.9745	0.9032	76.4	Yes
4	E_{LUMO}	α	μ	V_{mol}				0.9885	0.9002	107.7	Yes
4	E_{rep}	μ	V_{mol}	Q_{C1}				0.9885	0.9502	107.6	Yes
5	μ	V_{mol}	Q_{C4}	Q_{C3}	Q_{C2}			0.9982	0.9941	441.8	Yes
5	μ	V_{mol}	Q_{C1}	Q_{C4}	Q_{C3}			0.9965	0.9653	228.7	Yes
6	E_{HOMO}	E_{LUMO}	E_{rep}	E_{tot}	Q_{C5}	Q_{C2}		0.9997	0.9939	1441	Yes
6	E_{HOMO}	E_{LUMO}	E_{rep}	E_{tot}	Q_{C1}	Q_{C5}		0.9994	0.9443	1299	Yes
7	E_{HOMO}	E_{rep}	E_{tot}	μ	Q_{C1}	Q_{C4}	Q_{C3}	0.9999	0.9932	6921	Yes
7	E_{rep}	E_{tot}	μ	Q_{C1}	Q_{C6}	Q_{C5}	Q_{C4}	0.9999	0.9594	5619	Yes

increasing the risk of over fitting and decreasing the predictive ability of the model.

From the analysis above, the model containing five descriptors of μ, V_{mol}, Q_{C2}, Q_{C3} and Q_{C4} with the highest R^2_{CV} of 0.9941 and a good R^2 of 0.9982 can be considered as a better model for QSAR of 2-aryl-5-nitro-1*H*-indole derivatives.

Partial least square (PLS) analysis

The descriptor set obtained from GA with the highest R^2_{CV} was reanalyzed with partial least square analysis (PLS), and the LOO cross validation method was used to evaluate the predictive ability. The QSAR model with 5 independent variables obtained from PLS analysis is shown as follows:

$$-\lg MIC = -0.397 \times \mu + 0.019 \times V_{mol} + 1.231 \times Q_{C2} + 9.976 \times Q_{C3} - 6.177 \times Q_{C4} - 0.198$$
$$R^2 = 0.998,\ R^2_{CV} = 0.996 \quad (4)$$

The high values of R^2 and R^2_{CV} showed this model had a good predicability to the activity of the unknown samples. The predicted value, the experimental value and the residue are listed in Table 5. It shows that the predicted value and the experimental value correlate well.

This model showed that the inhibitory activity of the 2-aryl-5-nitro-1*H*-indole derivatives increases with the decrease of the dipole (μ) and the increase of the molecular volume. The Mulliken atomic charges of C_3 and C_4 have more significant effects than that of other carbons in 2-aryl. Increasing Q_{C3} and decreasing Q_{C4} can improve the object compounds inhibitory activity. For compound 8c, its highest molecular volume and the comparatively low Mulliken atomic charge of C_4 counteract its low value of Q_{C3}, which make it have the highest activity.

From the analysis above, this model gives us a revelation that the activity of 2-aryl-5-nitro-1*H*-indole derivatives as efflux pump inhibitor can be improved by properly increasing the molecular volume and Q_{C3} or lowering the dipole and Q_{C4}. With this model, the electric and structural descriptors of a new compound, 2-(2-Azidomethyl-5-phenoxy-phenyl)-5-nitro-1*H*-indole (8d, R_1=-CH_2N_3, R_2=-OCH_2Ph) was calculated with the same quantum chemistry calculation method (seen in Table 2) and it was predicted from Eq. 4 to lower the MIC of berberine to 0.091 μg/mL for inhibiting K2361 of *S. aureus* with NorA efflux pump protein over expression.

Conclusions

According to the activity of 2-aryl-5-nitro-1*H*-indole derivatives as efflux pump inhibitor and the structural and electric descriptors obtained quantum chemistry calculation, a QSAR model was built with GA and partial least square analyses. The high R^2 and R^2_{CV} indicates the derived model has a good predictive power which can be used in

Table 5 Predicted and experimental data of -lgMIC

Compound	-lgMIC		
	Experiment	Predicted	Residual
7a	−1.699	−1.689	−0.01
8a	−1.097	−1.162	0.065
9a	−0.477	−0.452	−0.025
7b	−1.699	−1.661	−0.038
8b	−0.799	−0.764	−0.035
9b	−0.176	−0.230	0.054
8c	0.097	0.165	−0.068
10a	−1.097	−1.129	0.032
10b	−1.097	−1.086	−0.011
INF55	−0.477	−0.476	−0.001

prediction of activity for new 2-aryl-5-nitro-1*H*-indole derivatives. This model gives us an apocalypse that the activity of 2-aryl-5-nitro-1*H*-indole derivatives as efflux pump inhibitor can be improved by properly increasing the molecular volume and Q_{C3} or lowering the dipole and Q_{C4}, and it was predicted that a new compound, 2-(2-Azidomethyl-5-phenoxy-phenyl)-5-nitro-1*H*-indole may lower the MIC of berberine to 0.091 μg/mL for inhibiting K2361 of *S. aureus* with NorA efflux pump protein over expression.

Acknowledgements The calculation work was carried out at Shandong University and the financial supported by the Start-up Foundation for Scientific Research of Tianjin University of Science and Technology (Grant No. 0200036), which are all acknowledged.

References

1. Keith P (2001) J Curr Opin Microbiol 4:500–508
2. Neyfakh AA, Bidnenko VE, Chen LB (1991) J Proc Natl Acad Sci USA 88:4781–4785
3. Markham PN, Westhaus E, Klyachko K, Johnson ME, Neyfakh AA (1999) Antimicrob Agents Chemother 43:2404–2408
4. Markham PN, Mulhearn DC, Neyfakh AA, Crich D, Jaber M-R, Johnson ME (2000) WO Patent 0032196
5. Samosorn S, Bremner JB, Ball A, Lewis K (2006) J Bioorg Med Chem 14:857–865
6. Hansch C, Maloney PP, Fujita T, Muir RM (1962) Nature 194:178–180
7. Fujita T, Iwasa J, Hansch C (1964) J Am Chem Soc 86:5175–5180
8. Clare BW, Supuran CT (1998) J Mol Structure (Theochem) 428:109–121
9. Clare BW, Supuran CT (1999) Eur J Med Chem 34:463–474
10. Clare BW, Supuran CT (1999) Eur J Med Chem 34:41–50
11. Trohalaki S, Gifford E, Pachter R (2000) Comp Chem 24: 421–427
12. Zhan L, Wan J, Yang G (2004) Bioorg Med Chem 12:6183–6191
13. Singh PP, Srivastava HK, Pasha FA (2004) Bioorg Med Chem 12:171–177
14. Dewar M, Zoebisch EG, Healy EF, Stewart JJP (1993) J Am Chem Soc 115:5348–5356
15. Stewart JJP (1989) J Comp Chem 10:209–220
16. Stewart JJP (1989) J Comp Chem 10:221–264
17. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, T omasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03. Gaussian Inc, Wallingford CT
18. Caballero J, Fernández L, Garriga M, Abreu JI, Collina S, Fernández M (2007) J Mol Graph Mod 26:166–178
19. So S, Karplus M (1996) J Med Chem 39:1521–1530
20. Holland JH (1992) Adaptation in natural and artificial systems, 2nd edn. MIT, Cambridge, pp 15–18
21. Materials studio modeling. Accelrys Software Inc, San Diego, CA