QSAR and SAR Studies on the Reduction of Some Aromatic Nitro Compounds by Xanthine Oxidase

Mamta Thakur,[†] Abhilash Thakur,[‡] and Krishnan Balasubramanian*[,§]

Department of Chemistry, Softvision Institute of Biotechnology and Science, 13 Green Park Colony, Indore 452001, India, Department of Chemistry, Rishiraj Institute of Technology, Sanwer Road, Ravti Rang, Indore, India, Chemistry and Applied Material Science Directorate, Lawrence Livermore National Laboratory, University of California, Livermore, California 94550, Department of Mathematics and Computer Science, California State University East Bay, Hayward, California 94542, and Glenn T. Seaborg Center, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720

Received October 31, 2005

This work describes QSAR and SAR studies on the reduction of 27 aromatic nitro compounds by xanthine oxidase using both distance-based topological indices and quantum molecular descriptors along with indicator parameters. The application of a multiple linear regression analysis indicated that a combination of distance-based topological indices with the ad hoc molecular descriptors and the indicator parameters yielded a statistically significant model for the activity, log *K* (the reduction of aromatic nitro compounds by xanthine oxidase). The final selection of a potential aromatic nitro compound for the reduction by xanthine oxidase is made by quantum molecular modeling. We have found that, among the various parameters, the quantum Mulliken charge parameters on the fourth atom or para position relative to the nitro group correlated best with the activity.

INTRODUCTION

Substituted aromatic and polyaromatic compounds and, in particular, nitro aromatic compounds find extensive applications in many materials such as high-energy materials, medicinal applications, and so on, as seen from the review of Hansch.[1] It is, thus, important to understand the potential toxicity, electronic, and steric features. Thus, there is a compelling need to understand the mechanisms and correlation modes of potential toxicity and metabolic reduction pathways and reactivity with xanthine oxidase. Over the past several decades, the Hansch group has studied the effect and significance of the hydrophobic parameters as well as the steric parameters in the modeling of various biological activities of numerous compounds. In a review based on the role of hydrophobic parameters in QSAR studies, the Hansch group[1] has reported that the use of steric parameter σ yielded an excellent statistics on a set of 29 aromatic nitro compounds. Inspired by the pioneering work of Hansch,[1] and in the continuation of our earlier works,[2−9] we have revisited the work of Hansch[1] to see if we can further develop a significant QSAR model with an entirely different set of parameters. To achieve this objective, we have used a large set of topological indices, namely, the Wiener index (*W*),[10] the first-order connectivity index ($^1\chi$),[11] the Balaban index (*J*),[12] the Szeged index (Sz),[13] and molecular descriptors such as molar refractivity (MR), molar volume (MV), parachor (Pc), the refractive index (η), surface tension (ST), density (*D*), and quantum chemical parameters together with indicator parameters used for the structural and positional specifications.

To obtain a statistically significant model, we have used a number of methods including the maximum R^2 method, which was followed by stepwise regression analyses.[14−16] We note that the maximum R^2 method actually includes a combination of standard error, adjusted R^2 value, R_A^2, *R*, standard error of estimation, and the *F*-ratio value. The predictive ability of the model is discussed on the basis of the predictive correlation coefficients. To validate our model further, we have used quantum molecular modeling parameters, and on the basis of these parameters, we have analyzed the structural behavior of these molecules. For the molecular modeling, we have optimized the geometries of molecules using the molecular mechanics method by applying the MM+ force field method and ab initio density functional theory (DFT) and Møller−Plesset second-order perturbation (MP2) methods. Our motivation for the ab initio study is that the charge parameters and dipole moments strongly depend on the degree of sophistication, and such electronic parameters vary with the degree of the level of theory. The accuracy of a molecular mechanics or semiempirical quantum mechanical method depends on the database used to parametrize the method. This is true for the type of molecules and the physical and chemical data in the database. While these methods may provide satisfactory results for a set of properties of a class of molecules, these methods may not provide for electronically sensitive parameters such as dipole moments, electronic populations, and so forth. The ab initio method may overcome this problem or provide an alternative platform to validate the results. Thus, we need to gauge the accuracy of the techniques. The parent structure of the

* Corresponding author e-mail: balu@csueastbay.edu.
[†] Softvision Institute of Biotechnology and Science.
[‡] Rishiraj Institute of Technology.
[§] Lawrence Livermore National Laboratory, University of California, California State University East Bay, and University of California, Berkeley.

10.1021/ci050478s CCC: $33.50
Published on Web 01/23/2006

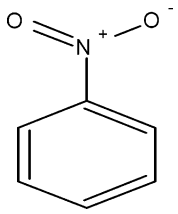

Figure 1. Parent structure for aromatic nitro compound used in the present study.

Table 1. Substituents of Aromatic Nitro Compounds, Their Activity, and Indicator Parameters Used in the Present Study[a]

compound number	substituents	log *K*	I_4	I_{OC}	I_{Cl}	I_A
1	4-NH_2	1.36	1	0	0	0
2	4-OMe	1.42	1	0	0	0
3	4-OH	1.46	1	0	0	0
4	4-SH	1.46	1	0	0	0
5	4-Me	1.46	1	0	0	0
6	H	1.68	0	0	0	0
7	4-C_6H_5	1.78	1	1	0	0
8	4-CH_2Cl	1.80	1	1	0	0
9	4-CH_2OH	1.89	1	1	0	0
10	4-Cl	2.20	1	0	1	0
11	4-Br	2.23	1	0	1	0
12	4-SO_2NH_2	2.25	1	0	0	1
13	4-CN	2.68	1	0	0	1
14	4-$CONH_2$	2.71	1	1	0	1
15	4-COOMe	2.72	1	1	0	1
16	4-NO_2	2.80	1	0	0	1
17	4-COMe	2.82	1	1	0	1
18	4-CHO	2.84	1	1	0	1
19	4-COC_6H_5	2.85	1	1	0	1
20	3-Cl	1.84	0	0	1	0
21	2-Cl	1.64	0	0	1	0
22	3-CHO	2.16	0	1	0	1
23	2-CHO	1.92	0	1	0	1
24	3-NO_2	2.29	0	0	0	1
25	2-NO_2	2.53	0	0	0	1
26	3-NH_2	1.80	0	0	0	0
27	2-NH_2	0.85	0	0	0	0

[a] Indicator parameters I_4, I_{Cl}, and I_A are defined as follows: $I_4 = 1$ if substitution at fourth/p position, 0 otherwise. $I_{Cl} = 1$ if substitution is chloride group, 0 otherwise. $I_A = 1$ if substitution is electron-attracting group, 0 otherwise.

aromatic nitro compounds used in the present study is shown in Figure 1, and various substituents are present in Table 1 along with their activities and the indicator parameters that we have used. Basak et al.[17] have reviewed the general topic of techniques for topological and other indices for such structure activity relations, and the readers are referred to this excellent review for further details on this topic.

TECHNIQUES

The quantum chemical computations were carried out using ab initio quantum chemical methods based on DFT. We have carried out DFT[18] and MP2 calculations. Since electronic properties such as Mulliken populations and electric dipole moments are sensitive to the level of quantum chemical approaches, we have used a variety of quantum chemical approaches in the present study, particularly to gauge the accuracies of different methods. The DFT approach utilized the standard B3LYP technique, which is comprised of Becke's three-parameter functional[20] with Vosko et al.'s local correlation part[21] and Lee et al.'s nonlocal part[22]. The geometry was optimized using the Berny algorithm with redundant internal coordinates.[23] We have also computed the vibrational frequencies of all structures to ensure that the optimized geometries were true minima. At the optimized equilibrium geometries, we have obtained the Mulliken charges for the various atoms in the molecule for each of the nitro aromatic compounds considered here. We have used the 6-311g(d,p) Gaussian basis sets for the quantum chemical computations. We have also obtained the optimized equilibrium geometries using the molecular mechanics method applying the MM+ force field. All of the quantum calculations were carried out using the Gaussian 03 package of programs.[24]

Table 2. Topological Indices Calculated for Aromatic Nitro Compounds Used in the Present Study[a]

compound number	*W*	$^1\chi$	*J*	Sz
1	120	4.698 38	2.2599	192
2	162	5.236 38	2.2427	252
3	120	4.698 38	2.2599	192
4	120	4.698 38	2.2599	192
5	120	4.698 38	2.2599	192
6	088	4.304 53	2.228 36	142
7	388	7.270 86	1.766 99	658
8	162	5.236 38	2.2427	252
9	162	5.236 38	2.2427	252
10	120	4.698 38	2.2599	192
11	120	4.698 38	2.2599	192
12	252	5.9097	2.392 39	378
13	162	5.236 38	2.2427	252
14	206	5.609 06	2.295 14	314
15	262	6.147 07	2.300 11	388
16	206	5.609 06	2.295 14	314
17	215	5.736 38	2.200 76	323
18	162	5.236 38	2.2427	252
19	606	8.254 02	1.606 58	930
20	117	4.698 38	2.319 85	186
21	114	4.715 21	2.396 04	180
22	156	5.236 38	2.330 31	240
23	150	5.253 22	2.442 97	228
24	197	5.609 06	2.402 39	296
25	188	5.6259	2.540 87	278
26	117	4.698 38	2.319 85	186
27	114	4.715 21	2.396 04	180

[a] *W* = Wiener index, $^1\chi$ = first-order connectivity index, *J* = Balaban index, Sz = Szeged index.

RESULT AND DISCUSSION

The set of 27 aromatic nitro compounds and their adopted activities, that is, the reduction of aromatic nitro compounds by xanthine oxidase, have been expressed as log *K* and are presented in Table 1. As may be seen from Table 1, a very low-level degeneracy is present in the activity log *K*. As a result of the occurrence of degeneracy in activity log *K*, it becomes essential to examine the degeneracy in the molecular descriptors also. A perusal of Table 2, which contains the topological indices calculated for aromatic nitro compounds, shows that high to low degeneracy is observed in the topological indices. Balaban et al.[25,26] have shown that these indices, despite their degeneracies, can be used successfully in developing statistically significant QSAR models.

Table 1 also shows that the $-NH_2$ substitution at the ortho position gives the lowest numerical value for the activity

Table 3. Physicochemical Properties of Aromatic Nitro Compounds Used in the Present Study[a]

compound number	MR	MV	Pc	η	ST	D
1	37.03	103.5	288.5	1.634	60.3	1.333
2	39.47	125.2	319.4	1.542	42.2	1.222
3	34.67	99.7	277.7	1.612	60.2	1.395
4	40.97	113.8	311.2	1.638	55.7	1.362
5	37.62	117.5	300.4	1.553	42.6	1.166
6	32.79	101.2	262.7	1.561	45.3	1.215
7	57.39	166.5	436.1	1.605	47	1.196
8	42.56	128.9	338	1.574	47.1	1.33
9	39.24	115	316.2	1.597	57	1.33
10	37.69	113.2	298.6	1.58	48.3	1.391
11	40.48	117.4	313.2	1.605	50.5	1.719
12	45.22	130.2	370	1.611	65.2	1.552
13	37.35	112.2	310	1.579	58.2	1.31
14	41.72	119.9	337.1	1.612	62.3	1.384
15	44.57	139.1	367.6	1.553	48.6	1.301
16	39.34	113.1	318.2	1.612	62.6	1.486
17	42.82	132.8	347.9	1.558	47.1	1.243
18	39.55	112.9	307.8	1.617	55.1	1.338
19	62.59	179.3	482.3	1.614	52.2	1.266
20	37.69	113.2	298.6	1.58	48.3	1.391
21	37.69	113.2	298.6	1.58	48.3	1.391
22	39.55	112.9	307.8	1.617	55.1	1.338
23	39.55	112.9	307.8	1.617	55.1	1.338
24	39.34	113.1	318.2	1.612	62.6	1.486
25	39.34	113.1	318.2	1.612	62.6	1.486
26	37.03	103.5	288.5	1.634	60.3	1.333
27	37.03	103.5	288.5	1.634	60.3	1.333

[a] MR = molar refractivity, MV = molar volume, Pc = parachor, η = index of refraction, ST = surface tension, and D = density.

log K, while $-COC_6H_5$ substitution at the para position gives the highest value for the same.

The correlation among the descriptors such as topological indices (Table 2), physicochemical properties (Table 3), indicator parameters (Table 1), and activities (Table 4) shows that, except for the Balaban index, all other topological indices are highly mutually correlated, while this is not so with the other physicochemical parameters used. Furthermore, the data presented in Table 4 (correlation matrix) express that none of the topological or molecular descriptors correlates well with the activity (log K); only the indicator parameter I_A accounts for the presence of an electrophilic group at the benzene moiety, it and shows a good correlation of the activity log K with an R value of 0.8099. A uniparametric model of indicator parameter I_A is given below:

$$\log K = 0.8895(\pm 0.1289) I_A + 1.656 \quad (1)$$

$$n = 27,\ \mathrm{Se} = 0.3357,\ R = 0.8099,\ R^2 = 0.6559,\ F = 47.646,\ Q = 2.4343$$

Equation 1 signifies the presence of an electrophilic group on the benzene moiety, but this model alone is unable to provide sufficient structural information required for the modeling activity log K.

Initial biparametric regression analyses indicate that, out of the large set of molecular descriptors used, the Balaban index (J) and the indictor parameter I_A play a significant role in modeling the activity log K with the following statistics:

$$\log K = -0.6618(\pm 0.337) J + 0.9072(\pm 0.1224) I_A + 3.1456 \quad (2)$$

$$n = 27,\ \mathrm{Se} = 0.3152,\ R = 0.8388,\ R^2 = 0.7035,\ R_A^{\,2} = 0.6788,\ F = 28.473,\ Q = 2.66$$

As seen in the above model, eq 2 demonstrates that the increase in the size of the molecule along with branching reduces the numerical value of the activity log K, while the presence of an electrophilic group on the benzene moiety favors the activity of xanthine oxidase on aromatic nitro compounds. We also employed other topological indices such as the Wiener index, the Randic first-order connectivity index, and the Szeged index, but out of all such indices, the Balaban J index is found to correlate best.

We obtain improved models with better correlation statistics when three descriptors are used. The triparametric

Table 4. Correlation Matrix

	log K	W	R	J	Sz	MR	MV	Pc
log K	1.000 00							
W	0.469 96	1.000 00						
R	0.516 66	0.980 09	1.000 00					
J	-0.158 14	−0.776 79	−0.702 75	1.000 00				
Sz	0.432 58	0.996 78	0.977 44	−0.809 47	1.000 00			
MR	0.353 29	0.948 88	0.939 85	−0.799 74	0.960 36	1.000 00		
MV	0.367 46	0.921 50	0.918 40	−0.796 84	0.930 97	0.968 88	1.000 00	
Pc	0.431 64	0.962 38	0.964 30	−0.756 09	0.965 96	0.986 37	0.977 64	1.000 00
η	−0.080 63	0.061 25	0.047 23	0.046 55	0.067 59	0.082 41	−0.165 29	−0.004 78
ST	0.179 34 -	0.011 36	0.020 30	0.345 18	−0.036 34	−0.124 36	−0.310 83	−0.105 21
D	0.226 90	−0.173 04	−0.162 44	0.420 28	−0.195 69	−0.159 52	−0.257 70	−0.148 86
I_4	0.254 80	0.299 30	0.288 12	−0.453 25	0.306 10	0.344 94	0.392 63	0.380 27
I_{OC}	0.415 50	0.483 06	0.547 80	−0.392 73	0.480 29	0.528 95	0.538 70	0.530 60
I_{Cl}	−0.057 94	−0.257 65	−0.310 35	0.114 19	−0.246 18	−0.161 78	−0.123 20	−0.192 43
I_A	0.809 86	0.418 75	0.496 45	0.073 50	0.369 98	0.268 15	0.240 09	0.348 51

η	ST	D	I_4	I_{OC}	I_{Cl}	I_A
η	1.000 00					
ST	0.765 89	1.000 00				
D	0.396 16	0.519 42	1.000 00			
I_4	−0.189 87	−0.129 95	−0.067 93	1.000 00		
I_{OC}	−0.041 81	−0.160 59	−0.336 82	0.216 93	1.000 00	
I_{Cl}	−0.177 72	−0.321 97	0.421 63	−0.147 44	−0.319 85	1.000 00
I_A	0.107 35	0.415 72	0.159 46	0.000 00	0.394 45	−0.373 00	1.000 00

model containing three descriptors J, I_A, and I_{Cl} is shown below:

$$\log K = -0.8028(\pm 0.2907)J + 1.045(\pm 0.1131)I_A + 0.5028(\pm 0.1587)I_{Cl} + 3.3282 \quad (3)$$

$$n = 27,\ \mathrm{Se} = 0.2687,\ R = 0.8909,\ R^2 = 0.7935,\ R_A{}^2 = 0.7666,\ F = 29.467,\ Q = 3.315$$

As stated earlier, the negative coefficient of J in eq 3 again indicates that the bigger size and high branching play a negative role in the exhibition of activity. Also, the positive coefficient of the indicator parameter I_A expresses the dominant characteristics of the electrophilic group on the benzene moiety for the modeling of activity log K. Equation 3 also exhibits a positive role of indicator parameter I_{Cl} in the modeling; that is, the presence of −Cl substituents at the benzene moiety enhances the reduction of aromatic nitro compound by xanthine oxidase.

The stepwise regression has indicated that an additional parameter, the Szeged index (Sz), can be added to further improve the statistics:

$$\log K = -1.6636(\pm 0.6559)J - 0.0011(\pm 7.85 \times 10^{-4})\mathrm{Sz} + 0.5095(\pm 0.1551)I_{Cl} + 1.207(\pm 0.1567)I_A + 5.5234 \quad (4)$$

$$n = 27,\ \mathrm{Se} = 0.2624,\ R = 0.9009,\ R^2 = 0.8117,\ R_A{}^2 = 0.7774,\ F = 23.707,\ Q = 3.34$$

The physical significance of J, I_A, and I_{Cl} is the same as that discussed for eq 3. The negative correlation coefficient of Sz in eq 4 confirms the negative role of size in the activity. This further proves the negative role of the size factor over log K.

When the Szeged index is replaced by an indicator parameter I_4 (the effect due to substitution at the 4 position) in eq 4, the statistics are significantly improved. Thus, the resulting tetraparametric model containing J, I_A, I_{Cl}, and I_4 gives the following equation:

$$\log K = -0.4919(\pm 0.2951)J + 1.0461(\pm 0.103)I_A + 0.538(\pm 0.1454)I_{Cl} + 0.2694(\pm 0.1128)I_4 + 2.4406 \quad (5)$$

$$n = 27,\ \mathrm{Se} = 0.2448,\ R = 0.9144,\ R^2 = 0.8361,\ R_A{}^2 = 0.8063,\ F = 28.05,\ Q = 3.735$$

The significance attached to parameters J, I_A, and I_{Cl} is the same as that discussed above. Added indicator parameter I_4 has a positive coefficient. This parameter accounts for the presence of a substituent at the fourth position (para position). Thus, its coefficient indicates that the substitution at the para position is favorable for the exhibition of the activity log K.

From the comparison of the statistics from eqs 4 and 5, we have found that the indicator parameter I_4 has a higher magnitude than Sz, and as a result of the replacement of Sz by I_4, there is a great improvement in the value of R; also, the predictive power/quality factor Q is increased for model 5. This shows the dominating role of the substitution at the fourth position over the size of the molecule. It is interesting to note that Dearden et al.[29] have also found that the fourth position is important in predicting the toxicity of nitrobenzenes.

Table 5. Experimental and Calculated log K Values Using the Models Expressed by Eqs 5 and 6

compound number	log K (obsd)	log K (calcd)[a]	residual	log K (calcd)[b]	residual
1	1.36	1.60	−0.24	1.56	−0.20
2	1.42	1.61	−0.19	1.57	−0.15
3	1.46	1.60	−0.14	1.56	−0.10
4	1.46	1.60	−0.14	1.56	−0.10
5	1.46	1.60	−0.14	1.56	−0.10
6	1.68	1.34	0.34	1.23	0.45
7	1.78	1.84	−0.06	1.79	−0.01
8	1.80	1.61	0.19	1.57	0.23
9	1.89	1.61	0.28	1.57	0.32
10	2.20	2.14	0.06	2.17	0.03
11	2.23	2.14	0.09	2.17	0.06
12	2.25	2.58	−0.33	2.61	−0.36
13	2.68	2.65	0.03	2.68	0.00
14	2.71	2.63	0.08	2.65	0.06
15	2.72	2.62	0.10	2.65	0.07
16	2.80	2.63	0.17	2.65	0.15
17	2.82	2.67	0.15	2.70	0.12
18	2.84	2.65	0.19	2.68	0.16
19	2.85	2.97	−0.12	2.97	−0.12
20	1.84	1.84	0.00	1.80	0.04
21	1.64	1.80	−0.16	1.76	−0.12
22	2.16	2.34	−0.18	2.29	−0.13
23	1.92	2.29	−0.37	2.24	−0.32
24	2.29	2.30	−0.01	2.25	0.04
25	2.53	2.24	0.29	2.19	0.34
26	1.80	1.30	0.50	1.19[c]	0.61
27	0.85	1.26	−0.41	1.15	−0.30

[a] Calculated from eq 5. [b] Calculated from eq 6. [c] Data points are not included in calculation from eq 6.

We have observed that the current model containing 27 compounds has one outlier (compound 26), and it was deleted from the regression procedure. When we deleted the outlier, we observed a large increase in the correlation statistics, giving the following equation:

$$\log K = -0.467(\pm 0.2648)J + 1.1054(\pm 0.0953)I_A + 0.6099(\pm 0.1334)I_{Cl} + 0.3489(\pm 0.106)I_4 + 2.2709 \quad (6)$$

$$n = 26,\ \mathrm{Se} = 0.2195,\ R = 0.9344,\ R^2 = 0.8751,\ R_A{}^2 = 0.8491,\ F = 36.313,\ Q = 4.26$$

The physical significance of the model is same as that discussed above for eq 5. Equation 6 shows that the dummy parameters I_A, I_{Cl}, and I_4 played the dominating role over the distance-based topological indices and ad hoc molecular descriptors. The indicator parameter I_A demonstrates higher magnitude in all the models, indicating the role of the electron-attracting group on the benzene moiety of the aromatic compound.

We also compare our model with the earlier proposed model of Hansch and co-workers.[1] The present correlation statistics are improved over the model proposed by Hansch and co-workers,[1] and our model has more parameters. Moreover, the present model is significant ($R = 0.9344$), with an entirely different set of structural features.

To confirms our findings, we have compared the calculated log K values from eqs 5 and 6 with the observed values of log K. Such a comparison is shown in Table 5 and demonstrated in Figures 2 and 3.

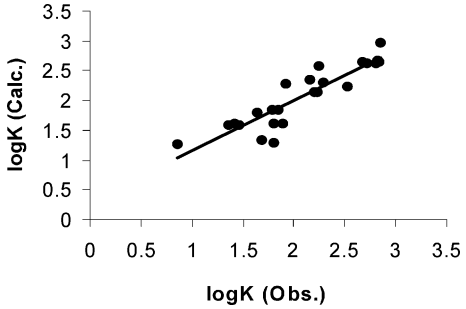

Figure 2. Graph obtained between observed and calculated log *K* values from eq 5.

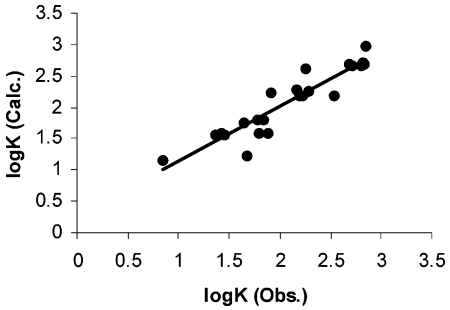

Figure 3. Graph obtained between observed and calculated log *K* values from eq 6.

Table 6. Molecular Modeling Parameter (Total Energy) Calculated for Aromatic Nitro Compounds (MM+)

compound number	T_{energy}[a]	compound number	T_{energy}[a]
1	5.91	15	17.88
2	5.31	16	4.82
3	1.35	17	8.12
4	1.11	18	7.15
5	0.88	19	32.26
6	1.06	20	1.44
7	0.41	21	4.49
8	2.89	22	7.20
9	1.82	23	15.99
10	1.41	24	4.92
11	1.55	25	16.00
12	−3.29	26	5.93
13	1.50	27	7.43
14	0.098		

[a] T_{energy} = total energy.

We have also obtained quantum chemically derived parameters since some of the properties depend strongly on electronic features such as electrophilic regions of the nitro compounds. To carry out quantum computations, we have first carried out the molecular geometry optimizations[17] to find out the structural behavior of these aromatic compounds as a function of attached groups and their positions. The corresponding molecular modeling parameters are shown in Table 6. From the perusal of Table 1, we categorized these compounds into two classes, one comprised of those having electron-repelling substituents and the other having the electron-attracting substituents.

As can be seen from Table 6, we can observe that, as the substituents are closer to the NO_2 group in the case of electron-repelling substituents, the total energy decreases; that is, there is an increase in stability of the compounds from the para position to the ortho position. In the case of electron-attracting substituents, the total energy follows the same pattern as that of the electron-repelling substituents.

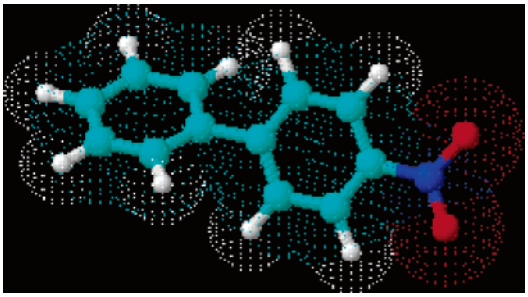

Figure 4. Optimized structure for compound number 7.

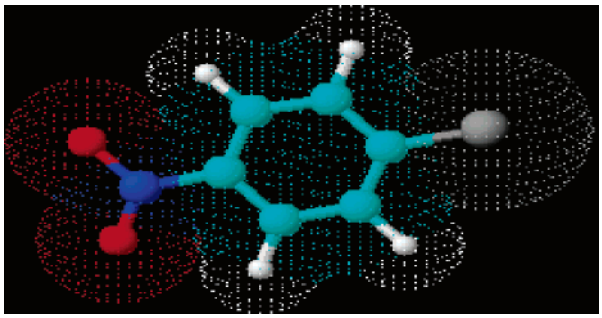

Figure 5. Optimized structure for compound number 10.

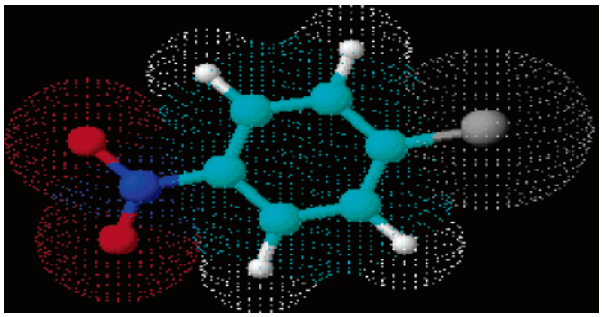

Figure 6. Optimized structure for compound number 11.

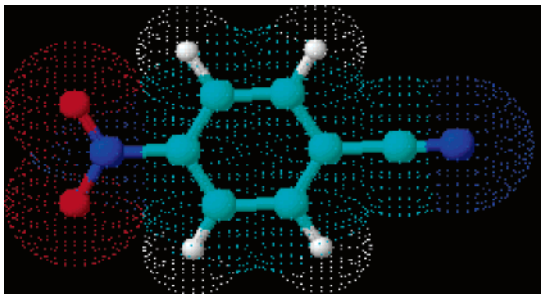

Figure 7. Optimized structure for compound number 13.

On the basis of the above study and the magnitude of the residues, we have selected eight compounds, namely, compounds 7, 10, 11, 13, 14, 15, 20, and 24, to correlate their modeling parameters with the activities. We have done it to find out which aromatic nitro compound has the highest correlation and predictive potential for the same category. The optimized molecular structures of these compounds are presented in Figures 4−11, respectively, while Figure 12 contains the Mulliken population charges obtained using the DFT/B3LYP method. The correlations of the computed parameters with log *K* are shown in Table 8 and, in the form of graphs, presented in Figures 13 and 14.

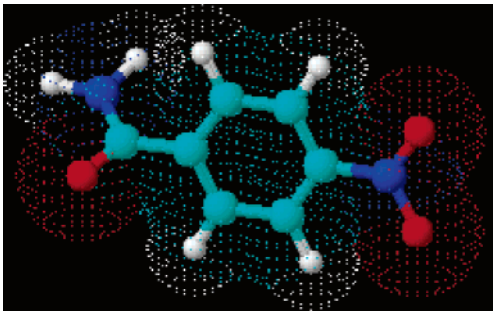

Figure 8. Optimized structure for compound number 14.

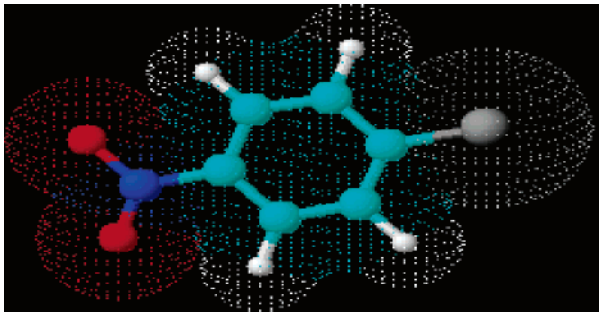

Figure 9. Optimized structure for compound number 15.

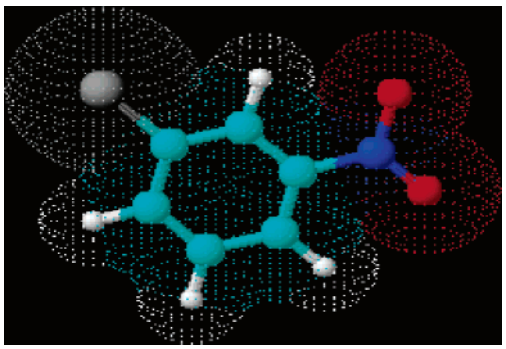

Figure 10. Optimized structure for compound number 20.

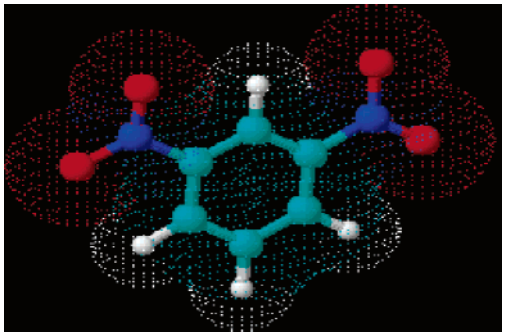

Figure 11. Optimized structure for compound number 24.

As can be seen from Table 8, the Mulliken population for the fourth C atom (M4; calculated by applying the B3LYP method) adds to the correlation value, while the correlation shown by HL (HOMO−LUMO) is closer to the line of significance. The model obtained from M4 is shown below:

$$\log K = -1.7380(\pm 1.2138)\mathrm{M4} + 2.1142 \qquad (7)$$

$$n = 8,\ \mathrm{Se} = 0.3499,\ R = -0.5047,\ R^2 = 0.2547,\ F = 2.050,\ Q = 1.44$$

Using this model, we have calculated and compared the log *K* values and observed that the residue is the least for the

Table 7. Modeling Parameters Obtained from B3LYP Force Field

compound number	HL	DpM(B3LYP)	Mulliken population (At.4)
7	0.151 13	5.6093	−0.048 359
10	0.1806	2.8616	−0.226 022
11	0.174 31	3.0606	−0.179 88
13	0.182 47	0.0698	−0.033 078
14	0.171 49	4.8503	−0.187 643
15	0.180 94	4.0085	−0.237 05
20	0.175 53	3.8349	0.046 543
24	0.194 87	4.2096	−0.018 552

Table 8. Comparison of Various Results from Modeling Parameters[a]

Sr. number	variable	R^2	R	Se	F ratio	Q
1	HL	0.2186	0.4675	0.3583	1.678	1.30
2	DpM	0.1510	−0.3886	0.3734	1.067	1.04
3	M4	0.2547	−0.5047	0.3499	2.050	1.44
4	HL					
	M4	0.4696	0.6853	0.3234	2.213	2.12
5	DpM					
	M4	0.4324	0.6576	0.3345	1.904	1.97

[a] HL = HOMO−LUMO, DpM2 = dipole moment from B3LYP force field, M4 = Mulliken population on the fourth C atom.

4-$CONH_2$ aromatic nitro compounds, as mentioned earlier (Table 6), the same compound having the maximum energy. Considering that eq 7 is for a single parameter correlation and $R > 0.5$, this parameter has the greatest potential for modeling the reduction activity of xanthine oxidases on aromatic nitro compounds. Equation 7 also indicates that, in these compounds (having minimum residue), an increase in the Mulliken population at the fourth C (M4) increases the activity.

In the case of bivariate combinations, a significant statistical correlation is exhibited by a combination of DpM (dipole moment calculated from the B3LYP method) and M4, but a better correlation is obtained by the combination of HL and M4. As can be seen from Table 8, the electron population at the fourth C atom (M4) plays the predominant role in the modeling of log *K*, while in combinations, the magnitude of HL is higher than the magnitude of the electron population at the fourth C atom (M4) and DpM (dipole moment calculated from B3LYP method). The best biparametric correlation is obtained by the combination of M4 and HL, and the equation obtained from this combination is as below:

$$\log K = -1.7254(\pm 1.1218)\mathrm{M4} + 13.9781(\pm 9.8213)\mathrm{HL} - 0.3504 \qquad (8)$$

$$n = 8,\ \mathrm{Se} = 0.3234,\ R = 0.6853,\ R^2 = 0.4696,\ F = 2.213,\ Q = 2.12$$

According to the rule of thumb, eq 8 should be considered as a marginal one. Equation 8 shows the dominating role of HL over M4 in the case of combinations for modeling of the activity log *K* and modeling the compound with the highest correlation and predictive potential for the same category. Equation 8 also exhibits an enhancement in the value of log *K* with an increase in the value of M4 and confirms the result of eq 7, but neither equation suffices by itself as the best QSAR model. Thus, a combination of a set of parameters serves as a better QSAR predictor than any

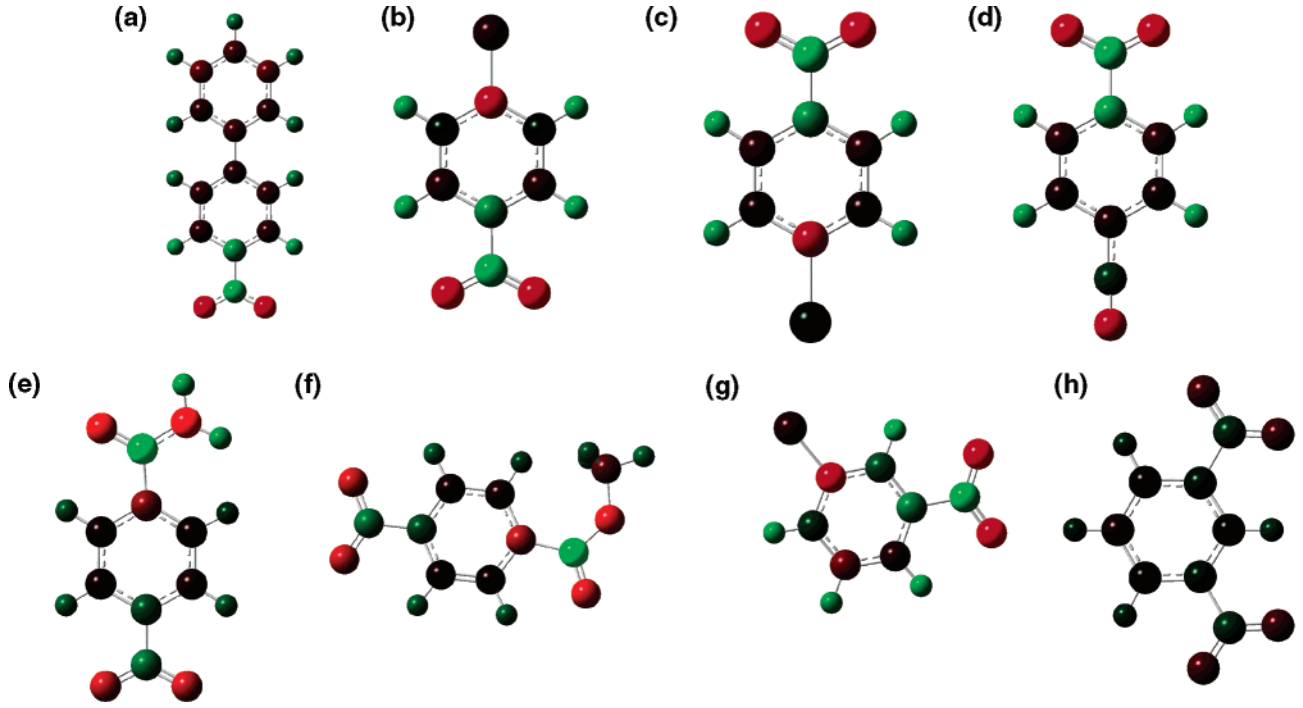

Figure 12. Mulliken charges of all molecules considered here. Red circles mean negative charges, green mean positive charges, and black mean nearly neutral charges. (a) A × 4, (b) A × 5, (c) A × 6, (d) A × 7, (e) A × 8, (f) A × 9, (g) A × 10, and (h) A × 11 Mulliken charges.

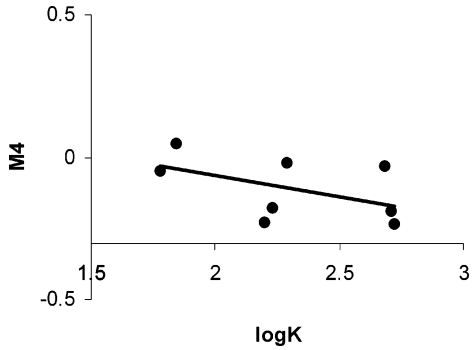

Figure 13. Graph obtained between log *K* and the modeling parameter Mulliken charge.

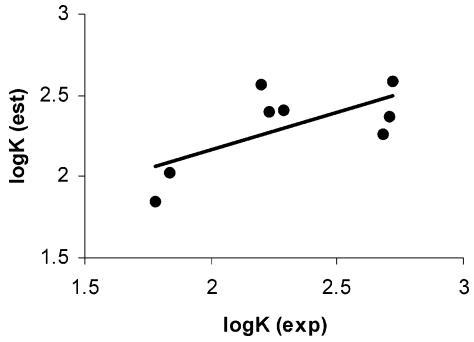

Figure 14. Graph obtained between log *K*(exp) and log *K*(est) from eq 8.

single parameter by itself. Considering that this was obtained just with two parameters, again, an *R* value of 0.6853 is a statistically significant bivariate model.

CONCLUSION

From the results and the discussion above, we conclude that the distance-based topological indices can be used successfully for modeling the reduction activities of aromatic nitro compounds by xanthine oxidase and that, for the present set of aromatic nitro compounds, the Balaban index (*J*) is found to be the prominent one. The results also indicate that a combination of topological indices and molecular (3D) modeling can be used for understanding the structural behavior and selecting the compound with potential activity.

EXPERIMENTAL SECTION

Biological Activity. The biological activities for the set of 27 compounds used in the present study, expressed as log *K*, were taken from the literature.[1]

Topological Indices. All the topological indices were calculated from the hydrogen-suppressed molecular graphs. Since the calculations of these topological indices are well-documented in the literature, it is not necessary to duplicate the same here.

Wiener Index (W).[10] The Wiener index $W = W(G)$ of cyclic graph G is defined as half the sum of the elements of the distance matrix.

$$W = W(G) = \frac{1}{2}\sum_{i=1}\sum_{j=1}(D)_{ij}$$

where $(D)_{ij}$ is the ijth element of the distance matrix, which denotes the shortest graph—the theoretical distance between sites i and j of G.

Connectivity Index($^1\chi$).[11] The connectivity index $^1\chi = {}^1\chi(G)$ of G is defined by Randic as

$$^1\chi = {}^1\chi(G) = \sum_{i,j}[d(i)\,d(j)]^{-0.5}$$

Balaban Index (J).[12] The Balaban index $J = J(G)$ of G is defined as

$$J = M/\mu + 1 \sum_{\text{bonds}} (d_i d_j)^{-0.5}$$

where M is the number of bonds in G, μ is the cyclomatic number of G, and d_i ($i = 1, 2, 3, ..., N$; N is the number of vertices in G) is the distance sum.

The cyclomatic number $\mu = \mu(G)$ of a cyclic graph G is equal to the minimum number of edges necessary to be erased from G in order to transform it into the related acyclic graph. In the case of a monocyclic graph, $\mu = 1$; otherwise, it is calculated by means of the following expression:

$$M = M - N + 1$$

Szeged Index (Sz).[13] The Szeged index, $Sz = Sz(G)$, is calculated according to the following expression:

$$Sz = Sz(G) = \sum_{\text{edges}} n_u n_v$$

where n_u is the number of vertices lying closer to one end of the edge $e = uv$; the meaning of n_v is analogous. Edges equidistant from both ends of an edge $e = uv$ are not taken into account.

Physicochemical Parameter. In the present study, molar refractivity (MR), molar volume (MV), parachor (Pc), the index of refraction (η), surface tension (ST), density (D), and polarizability (Pol), shown in Table 3, are tested and calculated using the computer software acdlabs (Chem Skech5.0).[27]

Regression Analysis. Regression analyses were made using the maximum R^2 method,[14−16] adopting a stepwise regression.

Molecular Modeling. Molecular optimization and calculation of the modeling parameters were made by using the software HyperChem7 (demo version).[28]

ACKNOWLEDGMENT

The work at Lawrence Livermore National Laboratory was carried out under the auspices of the U.S. Department of Energy under contract number W-7405-Eng-48, while the work at California State University was funded by the National Science Foundation, under Grant No. CHE-0540251.

REFERENCES AND NOTES

(1) Hansch, C.; Kurup, A.; Garg, R.; Gao, H. Chem-Bioinformatics and QSAR: A Review of QSAR Lacking Positive Hydrophobic Terms. *Chem. Rev.* **2001**, *101*, 619−671.
(2) Thakur, M.; Thakur, A.; Khadikar, P. V.; Supuran, C. T. QSAR Study On pK_a vis-à-vis Physiological Activity of Sulfonamides. *Bioorg. Med. Chem. Lett.* **2005**, *15* (1), 203−209.
(3) Thakur, A.; Thakur, M.; Khadikar, P. V.; Supurane, C. T.; Sudele, P. QSAR study on Benzene Sulphonamide Carbonic Anhydrase Inhibitors: Topological Approach Using Balaban Index. *Bioorg. Med. Chem.* **2004**, *12* (4), 789−793.
(4) Thakur, M., Thakur, A.; Khadikar, P. V. QSAR Studies on Psychotomimetic Phenylalkylamines. *Bioorg. Med. Chem.* **2004**, *12* (4), 825−831.
(5) Thakur, A., Thakur, M., Vishwakarma, S. QSAR Study of Flavonoid Derivative as p56lck Tyrosin Kinase Inhibitors. *Bioorg. Med. Chem.* **2004**, *12* (5), 1209−1214.
(6) Thakur, M.; Thakur, A.; Sudele, P. Comparative QSAR and QPAR Study of Benzodiazepines. *Indian J. Chem., Sect. B* **2004**, *43B*, 976−982.
(7) Thakur, M.; Agrawal, A.; Thakur, A.; Khadikar, P. V. QSAR Study on Phenolic Activity: Need of Positive Hydrophobic Term (logP) in QSAR. *Bioorg. Med. Chem.* **2004**, *12* (9), 2287−2293
(8) Thakur, A.; Thakur, M.; Kakani, N.; Joshi, A.; Gupta, A. Application of Topological and Physicochemical Descriptors: QSAR Study of Phenylamino-Acridine Derivatives. *ARKIVOC* **2004**, *10* (1), 36−43.
(9) Thakur, A.; Thakur, M.; Khadikar, P. V. Topological Modeling of Benzodiazepines Receptor Binding. *Bioorg. Med. Chem.* **2003**, *11* (23), 5203−5207.
(10) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−21.
(11) Randic, M. On Characterization of Molecular Branches. *J. Am. Chem. Soc*. **1975**, *97*, 6609−6615.
(12) Balaban, A. T. Highly Discriminating Distance Based Topological Indices.*Chem. Phys. Lett.* **1982**, *89*, 399−404.
(13) Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. Szeged Index of Hexagonal Chains. *MATCH* **2001**, *43*, 7−15.
(14) Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley VCH: New York, 2000.
(15) Dunn, W. J., Wold, S. In *Chemometric Method in Molecular Design*; Van de Waterbeemd, H., Ed.; VCH: Weinheim, Germany, 1995; pp 179−193.
(16) Devillers, J., Karcher, W. Applied Multivariate Analysis in SAR and Environmental Statistics. In *Chemical and Environmental Sciences*; Devillers, J., Karcher, W., Eds.; Kluwer Academic Publishers: Dordercht, The Netherlands, 1991; pp 153−205.
(17) Basak, S. C.; Frane, C. M.; Rosen M. E.; Magnuson, V. R. Molecular topology and acute toxicity: A QSAR study of monoketones. *Med. Sci. Res.* **1987**, *15*, 605−609.
(18) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, New York, 1989.
(19) Moller, C.; Plesset, M. General Properties of the Characteristic Matrix in the Theory of Elementary Particles. Parts I & II. *Phys. Rev.* **1943**, *46*, 618−628.
(20) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(21) Vosko, S. H.; Wilk, L.; Nusiar, M. Accurate Spin-Dependent Electron Liquid Correlation energies for local Spin Density Calculation: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200−1210.
(22) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti conelation energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(23) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. Using redundant internal coordinates to optimize geometries and transition states. *J. Comput. Chem.* **1996**, *17*, 49−56.
(24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. *Gaussian 98*, revision A7; Gaussian Inc.: Pittsburgh, PA, 1998.
(25) Balaban, A. T. Using Real Numbers as Vertex Invariant for Third Generation Topological Indices. *J. Chem. Inf. Comput. Sci*. **1992**, *32*, 23−28.
(26) Balaban, A. T.; Balaban, T. S. New Vertex Invariant and Topological Indices of Chemical Graphs Based on Information on Distance. *Math. Chem.* **1991**, *8*, 383−397.
(27) Chemsketh, S. O. *ACD-Lab*; Advanced Chemistry Development, Inc.: Toronto, Canada; software for calculating the referred physicochemical parameters. www.acdlabs.com.
(28) HyperChem 7 demo version from www.hyperchem.com.
(29) Dearden, J. C.; Cronin, M. T. D.; Schultz, T. M.; Lin, D. T. QSAR study of the toxicity of nitrobenzenes to tetrahymena pyriformis. *Quant. Struct.-Act. Relat.* **1995**, *14*, 427−432.

CI050478S