QSPR Studies on Impact Sensitivities of High Energy Density Molecules

Chan Kyung Kim,[*] Soo Gyeong Cho,[†,*] Jun Li, Chang Kon Kim, and Hai Whang Lee

High Energy Material Research Center and Department of Chemistry, Inha University, Incheon 402-751, Korea
[*]E-mail: kckyung@inha.ac.kr
[†]Agency for Defense Development, P.O. Box 35-42, Yuseong, Daejeon 305-600, Korea. [*]E-mail: soocho@add.re.kr

Impact sensitivity, one of the most important screening factors for novel high energy density materials (HEDMs), was predicted by use of quantitative structure-property relationship (QSPR) based on the electrostatic potential (ESP) values calculated on the van der Waals molecular surface (MSEP). Among various 3D descriptors derived from MSEP, we utilized total and positive variance of MSEP, and devised a new QSPR equation by combining three other parameters. We employed 37 HEDMs bearing a benzene scaffold and nitro substituents, which were also utilized by Rice and Hare. All the molecular structures were optimized at the B3LYP/6-31G(d) level of theory and confirmed as minima by the frequency calculations. Our new QSPR equation provided a good result to predict the impact sensitivities of the molecules in the training set including zwitterionic molecules.

Key Words : High energy density molecule (HEDM), Impact sensitivity, MSEP approach, Quantitative structure-property relationship

Introduction

Prediction of physicochemical properties of organic compounds is an essential step in designing new materials with improved property and performance. QSPR studies[1] have been known to be one of the most efficient approaches because one can predict various molecular properties before synthesis. With the help of QSPR methodology, one can save lots of time and efforts without performing trial-and-error based synthetic works. One way to estimate such properties was to use the General Interaction Properties Function (GIPF) developed by Politzer and coworkers.[2] In GIPF approach, several molecular descriptors, such as molecular weight, surface area, volume and other statistical parameters derived from the electrostatic potentials (ESPs) on the surface of molecules were evaluated using the high level *ab initio* program packages and these descriptors were subject to a linear or non-linear multiple regression with properties. This approach, denoted as Density approach, has been successfully applied to predict various physicochemical properties - heats of fusion,[3] boiling point,[2] heats of sublimation,[4] liquid density,[3] and solid density.[3]

Prediction of the physicochemical parameters of high energy density molecules (HEDMs) before synthesis is even more desirable in military science because a great deal of synthetic efforts should be given to devise new synthetic routes for novel HEDMs, and to obtain a relatively large amount of HEDMs to perform qualification tests for new HEDMs. In addition, there exist potential unwanted hazards in synthesis laboratories. The GIPF approach developed by Politzer *et al*. was applied to predict several molecular properties inherent to HEDMs, such as impact sensitivity (in $h_{50\%}$).[5] Recently, we simplified Politzer's approach by using

the van der Waals (vdW) surface of molecules, denoted hereafter as MSEP approach.[6] In our approach, molecular surface of a molecule was constructed from the vdW radii[7] of all constituent atoms and the ESP values were calculated on this surface. Our approach was applied successfully to estimate the solid densities of HEDMs.[8] In Density model, a relatively large density box was selected to accommodate all the atoms in a molecule and the molecular surface was constructed from the 0.001 electron/bohr^3 isosurface of electron density.[2,9] In MSEP approach, however, such a procedure was not necessary. If one has the 3D coordinates of a molecule determined experimentally or quantum-chemically, vdW surface can be generated automatically from the radii of all atoms in a molecule. Regardless of the surface models, these two methods gave almost similar trends on various physicochemical properties of organic molecules because all the independent variables in the QSPR equations from MSEP approach are the same (labelled as surface-independent, SI, variables) or linearly dependent (labelled as surface-dependent, SD, variables) to the respective independent variables from Density approach.[6]

Recently, Rice and Hare extended Politzer's scheme in predicting impact sensitivities of HEDMs with five different QSPR equations.[10] Four equations were employed parameters related to the surface ESPs. In this work, we wish to validate the performance of our MSEP approach in the cases of nonlinear models, particularly in predicting impact sensitivity. Moreover, we want to develop new linear QSPR equations including MSEP parameters.

Calculations

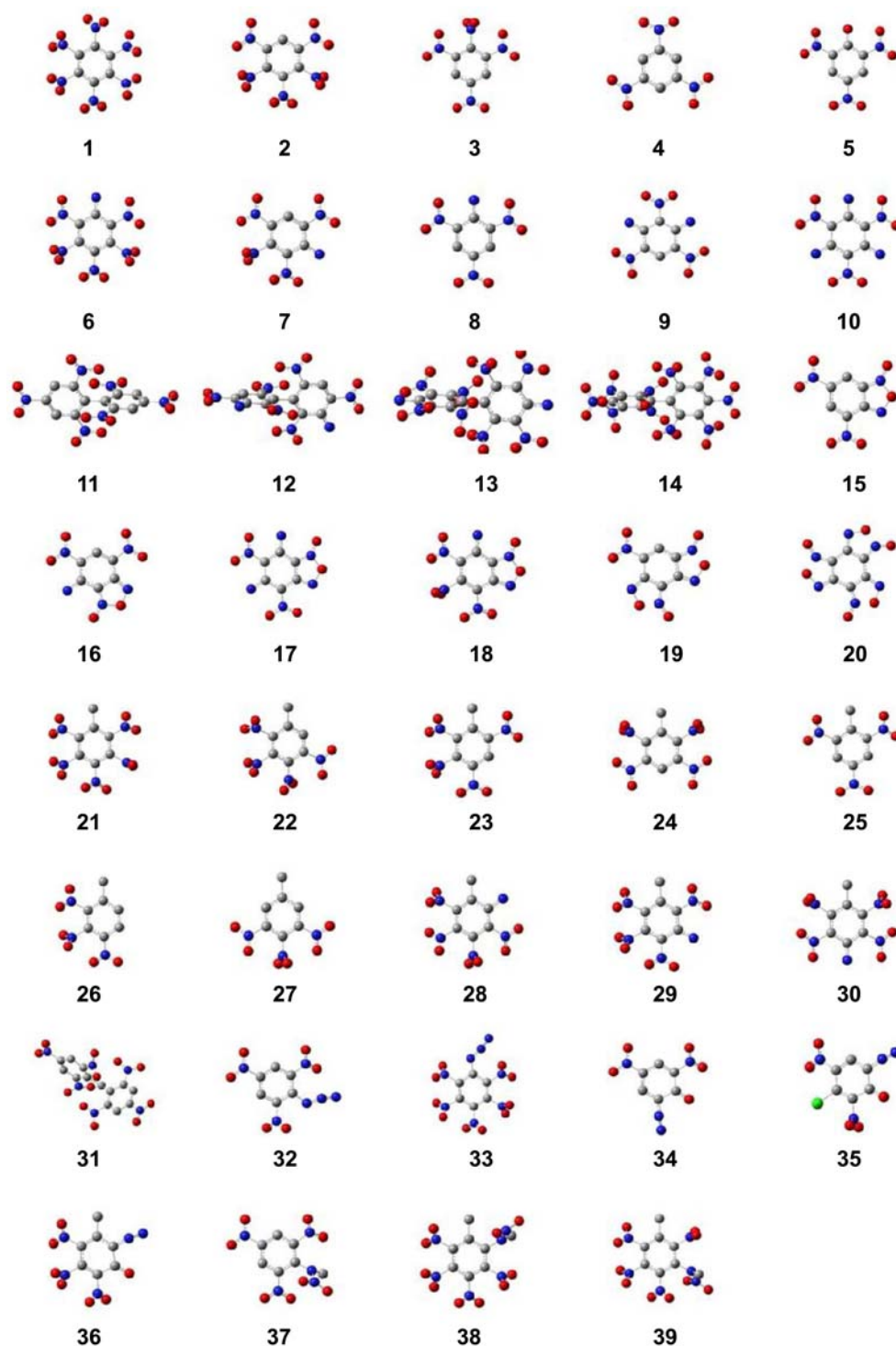The data set for prediction of impact sensitivities of CHNO

Figure 1. Optimized 3D structures of training set molecules.

explosives were taken from the earlier work.[10] Initial structures of 39 data set molecules were retrieved from the Cambridge Structural Database.[11] If no exact match was found, an analogous structure was modified to obtain a reasonable starting geometry. All the structures considered in this work were optimized using the B3LYP/6-31G(d) level of theory using Gaussian 98[12] and confirmed as minima by frequency calculations. In order to remove small negative frequency observed in each case of multi-substituted nitro compounds, GaussView program[13] was used to generate a new structure by adjusting the atomic positions according to the normal modes of vibration and the structure was re-optimized. This process was repeated until the structure showed no imaginary frequency. Final optimized structures are shown in Figure 1. The procedure for MSEP approach was given in detail earlier.[6] Multiple non-linear regressions

were performed using Origin 6.0,[14] and the final coefficients were obtained until the coefficients showed no further changes. The QSPR equation developed in this work was also applied to the molecules of the test set in order to compare the performance with previously published results.

Results and Discussion

Rice and Hare proposed five different models (eqs. 1-5) to predict impact sensitivities of HEDMs.[10] Among 39 molecules shown in Table 1, only 34 training set molecules were selected for QSPR studies because **14** had no experimental $h_{50\%}$, and **13** had wide varying experimental values and **34-36** were zwitterionic species. Among the five models proposed, we reexamined three models, Models 2, 3, and 5, using our MSEP approach because Models 1 and 4 (using the descriptor derived from the positive charge buildup over C-NO$_2$ bonds, $\overline{V}_{Mid}$, and the heat of detonation, Q, respectively) were independent of the definitions of molecular surfaces. Especially, we included **34-36** in the training set because the zwitterions could be important species in developing candidates for HEDMs.

Table 1. Experimental and calculated impact sensitivities for the training set molecules

No.	Chemical name	Exp. $h_{50\%}$[a]	Calc. $h_{50\%}$[a]	
			Model 5	Model 7
1	hexanitrobenzene	11	1	20
2	pentanitrobenzene	11	2	17
3	1,2,3,5-tetranitrobenzene	28	7	22
4	1,3,5-trinitrobenzene	71	52	66
5	2,4,6-trinitrophenol	64	77	63
6	pentanitroaniline	22	4	22
7	2,3,4,6-tetranitroaniline	47	19	41
8	2,4,6-trinitroaniline	141	141	121
9	1,3-diamino-2,4,6-trinitrobenzene	320	301	256
10	1,3,5-triamino-2,4,6-trinitrobenzene	490	478	600
11	2,2,4,4,6,6-hexanitrobiphenyl	70	26	55
12	3,3-diamino-2,2,4,4,6,6-hexanitrobiphenyl	67	72	67
13[b,c]	4,4-diamino-2,2,3,3,5,5,6,6-octanitrobiphenyl	20-95	-	-
14[b,c]	2,2,3,3,4,4,5,5,6,6-decanitrobiphenyl	-	-	-
15	4,6-dinitrobenzofuroxan	76	14	53
16	7-amino-4,6-dinitrobenzofuroxan	100	45	69
17	5,7-diamino-4,6-dinitrobenzofuroxan	120	102	155
18	7-amino-4,5,6-trinitrobenzofuroxan	56	7	26
19	8-amino-7-nitrobenzobisfuroxan	56	15	49
20	benzotrifuroxan	53	1	56
21	pentanitrotoluene	18	4	13
22	2,3,4,5-tetranitrotoluene	15	19	24
23	2,3,4,6-tetranitrotoluene	19	20	38
24	2,3,5,6-tetranitrotoluene	25	20	34
25	2,4,6-trinitrotoluene	98	143	72
26	2,3,4-trinitrotoluene	56	119	51
27	3,4,5-trinitrotoluene	107	140	54
28	2-amino-3,4,5,6-tetranitrotoluene	36	30	31
29	3-amino-2,4,5,6-tetranitrotoluene	37	47	53
30	4-amino-2,3,5,6-tetranitrotoluene	47	38	80
31	2,2,4,4,6,6-hexanitrodiphenylmethane	39	41	38
32	2-azido-1,3,5-trinitrobenzene	19	17	22
33	azidopentanitrobenzene	17	1	9
34[b]	2-diazo-4,6-dinitrophenol	9	-	21
35[b]	5-chloro-2-diazo-4,6-dinitrophenol	8	-	8
36[b]	3-methyl-2-diazo-4,5,6-trinitrophenol	8	-	9
37	*N*-methyl-*N*,2,4,6-tetranitroaniline	25	34	30
38	*N*-methyl-2-amino-*N*,3,4,5,6-pentanitrotoluene	21	14	16
39	*N*-methyl-3-amino-*N*,2,4,5,6-pentanitrotoluene	18	20	18
	rms deviation (cm)		28.1	26.8

[a] in cm. [b] Not included in earlier work. [c] Not included in this study.

Model 1: $h_{50\%} = a_1 + a_2 \exp(-a_3 \overline{V}_{Mid}) + a_4 \overline{V}_{Mid}$ (1)

Model 2: $h_{50\%} = a_1 + a_2 \exp[-(a_3|\overline{V}_S^+ - |\overline{V}_S^-||)]$ (2)

Model 3: $h_{50\%} = a_1 + a_2 \exp(a_3 \nu)$ (3)

Model 4: $h_{50\%} = a_1 + a_2 \exp(-a_3[Q - a_4])$ (4)

Model 5: $h_{50\%} = a_1 \exp(a_2 \nu - a_3[Q - a_4])$ (5)

Here, $h_{50\%}$, a typical parameter for impact sensitivity, is the height (in cm) from which 50% of the drops result in detonation of the sample by dropping 2.5 kg weight of a drop hammer. $\overline{V}_S^+$ and $\overline{V}_S^-$ are the averages of the positive and negative ESPs on molecular surface, respectively. ν is the balance parameter and is described as the degree of balance between positive and negative potentials on the isosurface. Q is the heat of detonation for CHNO explosives and defined as the heat of the reaction for the following reaction, eq. (6).

$$C_a H_b N_c O_d \rightarrow \frac{1}{2}cN_2 + \frac{1}{2}bH_2O + \left(\frac{1}{2}d - \frac{1}{4}b\right)CO_2 + \left(a - \frac{1}{2}d + \frac{1}{4}b\right)C \quad (6)$$
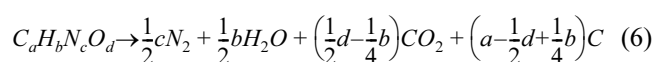
In this work, $\overline{V}_S^+$, $\overline{V}_S^-$, and ν were recalculated using MSEP approach and the results for the non-linear regressions are summarized in Table 2 along with the results reported by Density approach.

Since the ESP values obtained from Density and MSEP approaches were different from each other, all the independent variables derived from these ESP values should be method dependent. Even though the coefficients in equations (2), (3) and (5) had marked differences, it is meaningless to compare them directly because the corresponding independent variables obtained using Density and MSEP approaches have different averages and standard deviations. One way to circumvent this problem is to introduce Z-scores, as suggested by a reviewer. In Z-scores, the independent (x) and dependent (y) variables are standardized by using eq (7).

$$Z_x = \frac{x - \overline{x}}{\sigma_x}, \quad Z_y = \frac{y - \overline{y}}{\sigma_y} \quad (7)$$

In eq. (7), $\overline{x}$ and $\overline{y}$ are average values of x and y, respectively and σ_x and σ_y are their standard deviations. Introducing eq. (7) to eqs. (2), (3) and (5) can give the follow eqs. (8), (9), and (10), respectively:

Model 2: $Z_{h_{50\%}} = \beta_1 + \beta_2 \exp[-(\beta_3 Z_{|\overline{V}_S^+ - |\overline{V}_S^-||})]$ (8)

Model 3: $Z_{h_{50\%}} = \beta_1 + \beta_2 \exp(\beta_3 Z_\nu)$ (9)

Model 5: $Z'_{h_{50\%}} = \beta_1 \exp(\beta_3 Z_\nu - \beta_3 Z_Q)$ (10)

where, β_1, β_2 and are β_3 the coefficients and $Z'_{h_{50\%}} = Z_{h_{50\%}} + \frac{\overline{h}_{50\%}}{\sigma_{h_{50\%}}}$.

The Z-score formulas are very similar to the original ones because of exponential nature of QSPR equations. Application of Z-scores to the linear models, however, usually removes the intercepts from the formulas. In the case of Model 5, three coefficients, β_1, β_2 and β_3 are enough to describe the formula, which suggests that eq. (5) can be written as $h_{50\%} = A_1 \exp(a_2 \nu - a_3 Q)$, where $A_1 = a_1 \exp(a_3 a_4)$ = 29,824. Results of regressions using eqs. (8)-(10) are summarized in Table 3. Table 3 shows that the coefficients of the regression are quite similar for eqs. (8) and (10) but show marked difference in eq. (9). However, the correlation coefficients (r in the last column) were almost same for both methods. These are not unexpected because SD variables ($\overline{V}_S^+$, $\overline{V}_S^-$, and ν) have linear correlations between two approaches and SI variable (Q) can give similar coefficients, as mentioned above.[6] Therefore, this result again confirms that MSEP approach can give similar trends even in the cases of non-linear regressions when compared to Density approach.

As shown in eqs. (1)-(5), the correlations of impact sensitivity with charge distribution in the molecular surface followed exponential equations. In most QSPR studies, however, linear equations have been widely used for simplicity and ease of interpretation. In this work, we tried several linear equations to find a good relationship between $h_{50\%}$ and

Table 3. Results of Z-scores regressions on the training set Molecules using MSEP[a] and Density[b] approaches

Eq.	Method	β_1	β_2	β_3	r[a]
8	MSEP	-0.5518	0.2593	1.4555	0.94
	Density	-0.6533	0.3951	0.9326	0.94
9	MSEP	-0.4380	0.1311	1.5951	0.83
	Density	-0.3893	0.02069	3.2769	0.80
10	MSEP	0.2856	0.4380	1.0335	0.96
	Density	0.2436	0.4421	1.2230	0.95

[a]regression coefficient.

Table 2. Results of regressions on the training set molecules using MSEP[a] and Density[b] approaches

Eq.	Method	a_1	a_2	a_3	a_4	a_5	a_6	r[c]
2	MSEP	33.087	360890	0.5602	-	-	-	0.94
	Density	9.1949	803.4464	0.3663	-	-	-	0.94
3	MSEP	19.3724	0.9447	24.0947	-	-	-	0.83
	Density	29.3248	0.001386	48.8381	-	-	-	0.80
5	MSEP	2.1714	8.1508	5.7395	1.5073	-	-	0.96
	Density	1.3410	8.1389	6.7922	1.4737	-	-	0.95
10	MSEP	15.13	-0.5113	0.7838	-0.06246	0.2714	-0.2687	0.97

[a]this work. [b]reference 10. [c]regression coefficient.
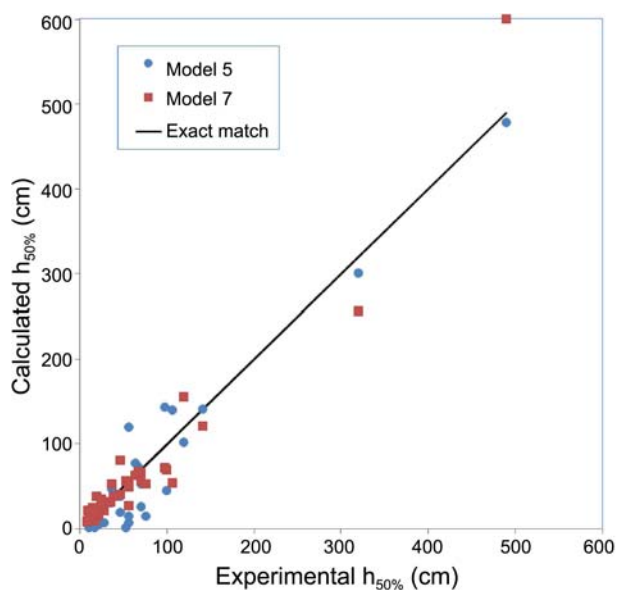
Figure 2. A plot of calculated $h_{50\%}$ *vs.* experimental $h_{50\%}$ derived from Model 7. Results from Model 5 were also shown for comparison.

various descriptors. In these trials, the maximum number of descriptors was set to five. The best correlation coefficient in the linear model was 0.90, better than that of Model 3, but worse than that of Models 2 and 5 and the rms deviation of the results from experiment was 38.5 cm. This model, denoted as Model 6 (not shown), was unacceptable in two reasons: (1) for 2 compounds among the 37 of the training set molecules, the calculated $h_{50\%}$ is within 10% of the experimental data. (2) the predicted $h_{50\%}$ values were negative in the cases of **2** (−26 cm), **3** (−6 cm), and **34** (−63 cm). Especially, such negative values were unavoidable in the linear model because we did not impose any condition for the predicted values. In the models developed by Rice and Hare,[10] only positive values were possible because of using exponential functions. To avoid such an unrealistic situation, we tried some linear correlations between "ln $h_{50\%}$" and various molecular descriptors. The best model using five descriptors had the form

$$\text{Model 7: ln } h_{50\%} = a_1 + a_2(\text{H}) + a_3(HBD) + a_4(PSA) + a_5(\sigma) + a_6(\sigma_+^2) \tag{11}$$

where *H*, *HBD*, and *PSA* are the number of hydrogen atoms,[15] the number of hydrogen bond donor,[16] and the polar surface of the molecule,[17] respectively, and σ is the sum of MSEP values and σ_+^2 is the variance of positive MSEPs defined in eq. (12).[2,9] The numerical values for these descriptors are summarized in Table S1. The regression coefficient for this fit was 0.97 and the rms deviation of the predicted values from the experimental values is 26.8 cm, which was better than that from Model 5 (28.1 cm). Comparison of the results with experiments is given in Table 1 and depicted in Figure 2. The largest positive and negative deviations were found for **9** (64 cm) and **10** (−110 cm), respectively.

$$\sigma_+^2 = \frac{1}{m} \sum_{i=1}^{m} [V^+(r_i) - \overline{V}_S^+]^2 \tag{12}$$

To see how this model worked for the test set molecules selected by Rice and Hare, the impact sensitivities of these molecules (**T1-T15**) were predicted using eq. (11). The results are summarized in Table 4 and Figure 3. The regression coefficients for Models 5 and 7 were moderate and comparable (0.74 *vs.* 0.72). However, both models failed to give reasonable regression lines: the slopes for Models 5 and 7 were larger (3.19) and smaller (0.14) than the unit slope of exact match, respectively. Such poor results were easily understandable because the test set molecules were structurally very different from the molecules in the training set or the models developed using the GIPF parameters may be

Table 4. Experimental and calculated $h_{50\%}$ values (cm) for the test set molecules

| No.	Chemical Name	Exp. $h_{50\%}$	Calc. $h_{50\%}$	
			Model 5	Model 7
T1	tetranitrate pentaerythritol	13, 16, 12	16	4
T2	2,4,6,8,10,12-hexanitrohexaazaiso-wurtzitane (ε-polymorph)	12, 16, 17, 21	3	3
T3	2,4,6,8,10,12-hexanitrohexaaza-isowurtzitane (β-polymorph)	14	3	3
T4	hexahydro-1,3,5-trinitro-1,3,5-s-triazine	28, 26, 24	22	10
T5	1,3,5,7-tetranitro-1,3,5,7-tetraaza cyclooctane	32, 29, 26	22	2
T6	*N,N*′-dinitro-1,2-ethanediamine	34	153	24
T7	2,4,6-trinitroresorcinol	43	106	46
T8	2,2′,4,4′,6,6′-hexanitrostilbene	54	47	12
T9	1,4-dinitroimidazole	55	38	36
T10	2,4,5-trinitroimidazole	68	4	10
T11	2,4-dinitroimidazole	105	41	18
T12	1,1-diamino-2,2-dinitro-ethylene	126	133	30
T13	2-methoxy-1,3,5-trinitrobenzene	192	128	51
T14	3-nitro-1,2,4-triazole-5-one	291	296	41
T15	nitroguanidine	> 320	1800	37
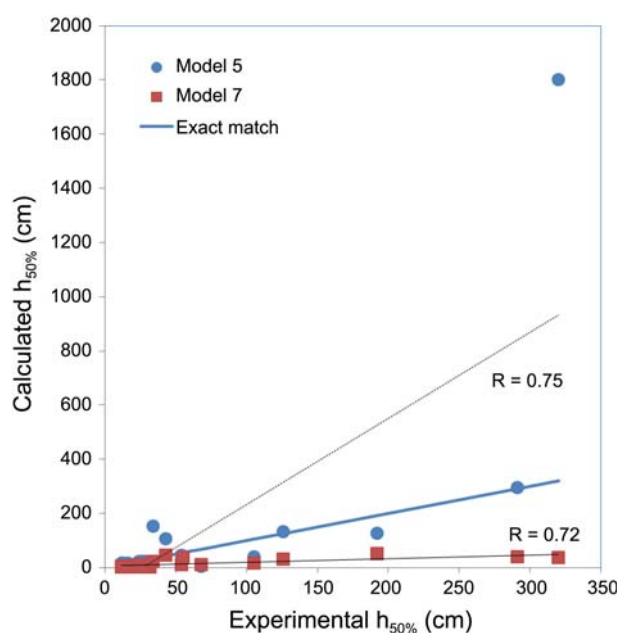
Figure 3. A plot of calculated $h_{50\%}$ *vs.* experimental $h_{50\%}$ derived from Models 5 and 7 for the test set molecules. Regression lines for Models 5 (···) and 7 (—) were shown with the corresponding R values.

limited in their predictive capability as found earlier.[10] In conventional QSPR studies, the test set molecules are selected to be structurally similar to the training set molecules. In other words, it will be necessary to study a general QSPR relationship using a wider range of molecular types. This work is currently in progress in our lab.

Conclusion

Impact sensitivities of HEDMs were evaluated using the MSEP approach and the results were compared to those predicted from Density approach. For the first time, we confirmed that the MSEP results could be equally applicable to the exponential QSPR equations derived from Density approach by giving comparable regression coefficients. We developed a new linear model (Model 7), where "ln $h_{50\%}$" was correlated with three MSEP and two topological parameters. Model 7 was quite comparable to the best model (Model 5) developed by Rice and Hare. It proved that impact sensitivities of aromatic CHNO HEDMs were able to be predicted with the MSEP approach without performing complex computation of isosurface envelop of a certain electron density. In addition, our new model appeared to predict the impact sensitivities of the zwitterionic species successfully.

References

1. Jurs, P. C. In *the Encyclopedia of Computational Chemistry*; Vol. 4, Schleyer, P. v. R., Ed.; John Wiley & Sons: 1998; p 2320.
2. Politzer, P.; Murray, J. S. In *the Quantitative Treatments of Solute/Solvent Interactions*; Elsevier Amsterdam: 1994; p 243.
3. Murray, J. S.; Brinck, T.; Politzer, P. *Chem. Phys.* **1996**, *204*, 289.
4. Politzer, P.; Murray, J. S.; Grice, M. E.; Desalvo, M.; Miller, E. *Mol. Phys.* **1997**, *91*, 923.
5. Murray, J. S.; Lane, P.; Politzer, P. *Mol. Phys.* **1998**, *93*, 187.
6. Kim, C. K.; Lee, K. A.; Hyun K. H.; Park, H. J.; Kwack, I. Y.; Kim, C. K.; Lee, H. W.; Lee, B.-S. *J. Comput. Chem.* **2004**, *25*, 2073.
7. (a) Pauling, L. In *the Nature of Chemical Bond and the Structure of Molecules and Crystals*; Cornell Univ. Press Ithaca, 1960. (b) Valvani, S. C.; Yalkowsky, S. H.; Amidon, G. L. *J. Phys. Chem.* **1976**, *80*, 829.
8. Kim, C. K.; Cho, S. G.; Kim, C. K.; Park, H.-Y.; Zhang, H.; Lee, H. W. *J. Comput. Chem.* **2008**, *29*, 1818.
9. Murray, J. S.; Brinck, T.; Lane, P.; Paulsen, K.; Politzer, P. *J. Mol. Struct. (THEOCHEM)* **1994**, *93*, 187.
10. Rice, B. M.; Hare, J. J. *J. Phys. Chem. A* **2002**, *106*, 1770.
11. Cambridge Structural Database, ver. 5.24, 2005.
12. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, Revision A.6. Gaussian, Inc., Pittsburgh PA, 1998.
13. GaussView 3.0, Gaussian Inc., Carnegie Office Park - Building 6, Pittsburgh, PA, 15106 USA.
14. Microcal Origin ver. 6.0, Microcal Software, Inc., One Roundhouse Plaza, Northampton, MA 01060 USA.
15. Ouvrard, C.; Mitchell, J. B. O. *Acta Cryst.* **2003**, *B59*, 676.
16. Charlton, M. H.; Docherty, R.; Hutchings, M. G. *J. Chem. Soc. Perkin Trans 2* **1995**, 2023.
17. Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714.