Development of quantitative structure–property relationships for predictive modeling and design of energetic materials

Jason A. Morrill [a],*, Edward F.C. Byrd [b]

[a] Department of Chemistry, William Jewell College, 500 College Hill, Liberty, MO 64068, United States
[b] United States Army Research Laboratories, AMSRD-ARL-WM-BD, Aberdeen Proving Ground, MD 21005-5069, United States

ABSTRACT

A quantitative structure–property relationship (QSPR) based on the AM1 semiempirical quantum mechanical method was derived using the program, CODESSA, to describe published drop height impact sensitivities for 227 nitroorganic compounds. An eight-descriptor correlation equation having $R^2 = 0.8141$ was obtained through a robust least median squares regression. The resulting model is the most comprehensive and systematic quantum mechanically derived QSPR for energetic materials of those that have been published. The predictive capability of the model is also presented and discussed.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Background

The impact sensitivity of an energetic material is a measure of the tendency of the material to undergo an explosive detonation when experiencing an impact, such as being struck by a hammer [1]. Impact sensitivities are experimentally determined by the drop weight impact test, which entails dropping a weight having a typical mass of 2.5 kg from a predetermined height onto a milligram quantity sample placed on a striker plate. Following each test, evidence of reaction or non-reaction is recorded. The height at which 50% of the drop weight impact test trials result in reaction is recorded and termed the $h_{50\%}$ value for the energetic material. Materials having smaller $h_{50\%}$ values are considered more sensitive to impact because less kinetic energy results in reaction in the drop weight impact test. The handling of such materials is extremely dangerous and presents a hazard to the scientist undertaking their development. For the scientist developing notional energetic materials it would therefore be useful to have a tool with which to screen the impact sensitivity of a candidate energetic material [2]. Since the development of notional materials requires extensive performance evaluation such a tool would also

allow the scientist to avoid the costly development and inherent danger of those candidate materials, which are inevitably later found to be unsuitable.

The development of predictive tools for impact sensitivity of energetic materials has been an ongoing area of research over several years. Some of the earliest published attempts related the constitutional makeup of explosive compounds to their performance properties [3]. Other attempts have sought correlations between impact sensitivities and such properties as electronegativity [4], estimated reaction heats and molecular composition [5–7].

More recently, efforts have focused on the relationships between impact sensitivity and molecular quantities derived from the use of electronic structure methods, including structural modeling [8–11] and reaction modeling [12–14].

More relevant to the study reported here, Rice and Hare mapped the electrostatic surface potentials (ESPs) onto the total electronic density of a series of C, H, N, O energetic materials at the B3LYP/6-31G* level of theory [2]. The ESPs were analyzed in order to identify features of molecular electronic distribution that related to the degree of impact sensitivity of the materials investigated. In their study, Rice and Hare pointed out that the build-up of positive charge over covalent bonds within the molecular framework of energetic materials was related to the degree of impact sensitivity.

Klapotke, et al., reported on the synthesis and drop weight impact testing of tetrazole azide and also applied Rice's ESP

---

* Corresponding author.
  E-mail address: morrillj@william.jewell.edu (J.A. Morrill).

mapping approach to understand the impact sensitivity of the compound. In order to suggest the likely mechanism of the detonation reaction, Klapotke, et al., also performed reaction modeling for reactions corresponding to three proposed initial dissociation pathways corresponding to explosive detonation at the B3LYP/6-31G** level of theory [15]. In concurrence with the findings of Rice, he found that the lack of electron density over the tetrazole ring and C–N bonds was related to the degree of impact sensitivity when tetrazole azide was compared with other tetrazoles. In particular, he found the less evenly the electron density is distributed throughout the framework of a molecule, the greater its sensitivity to impact.

Edwards et al., compared the correlation of impact sensitivities to various quantum mechanically derived quantities, such as the energies of the highest energy occupied molecular orbital (HOMO), lowest energy unoccupied molecular orbital (LUMO) and heats of detonation corresponding to various detonation reaction mechanisms [16]. Although, coefficients of multiple determinations were not reported for all correlations examined, Edwards found moderate degrees of correlation for the orbital and reaction energies investigated. Interestingly, he found higher degrees of correlation with quantities that describe the constitutional makeup of energetic materials (number of NO_2 groups) than with orbital energies.

Zhang et al., computed the total charge on nitro groups Q_{NO_2} as the sum of individual atomic charges of each of the atoms in the group:

$$Q_{NO_2} = Q_N + Q_{O_1} + Q_{O_2}$$

The nitro group charges were computed from Mulliken population analysis using density functional theory (B3LYP) and the DNP basis set. The nitro group charges were then investigated with respect to their degree of correlation with impact sensitivity [17]. Zhang found that when the most negatively charged nitro group of a compound has a net negative charge, $Q_{NO_2} < 0.23e$ that impact sensitivity was less than 40 cm.

Cheng et al., investigated the relationship between impact sensitivity and bond dissociation energies computed at the B3LYP/6-31G* level of theory [18]. Cheng modeled bond dissociation reactions corresponding to breaking the C–NO_2 bond for all NO_2 groups of a small group of nitroorganic compounds. He found that the computed bond dissociation energies, when divided by the total molecular energy, were roughly positively correlated with impact sensitivity.

Cho et al., used a neural network approach to identify descriptors of molecular structure that were a priori correlated to published impact sensitivity [19] ($\log_{10} h_{50\%}$) for a series of 234 energetic materials [20]. In their study there were 39 descriptors that were constitutional and quantum mechanical in nature. The quantum mechanical descriptors were computed using the AM1 semiempirical method. Subsets of the descriptors (treated as regression variables) were correlated into regression equations using neural network architectures with either one or two hidden layers in the architecture. The authors reported obtaining several correlations generally having between 11 and 20 descriptors and R^2 generally between 0.6 and 0.8. One regression equation was composed of 39 molecular descriptors and having R^2 value of 0.848.

Cho made several comparisons to a similar study that was conducted by Legendre et al. Legendre also used a neural network approach to find correlations between subsets of 39 constitutional and AM1 quantum mechanical descriptors and impact sensitivities ($\log_{10} h_{50\%}$) [21] for 204 energetic materials, coming from the same molecular database as Cho [19]. Legendre obtained several regression equations composed of 11 or 13 descriptors and having cross-validated (leave-one-out) R^2 values that varied between 0.799 and 0.882. One of the best regressions was obtained using an average prediction from a 13-descriptor regression equation from partial least squares analysis and an 11-descriptor regression equation from a neural network analysis.

1.2. Objective and general approach

The work of Cho and Legendre assumes that from relatively few descriptors of molecular structure one can develop predictive tools for impact sensitivity. Their work also assumes that those descriptors are among relatively small groups of constitutional or quantum mechanical descriptors. Their assumptions are not entirely unreasonable given the good fits to impact sensitivity observed. Their success is likely due, in a gross sense, to the nature of what is known about the relationship between molecular structure and impact sensitivity (i.e. increasing the number of nitro groups in a molecule generally tends to increase its sensitivity to impact). This investigation makes no such assumptions about the relationships between molecular structure and impact sensitivity. Rather, in the work presented here a large number of descriptors belonging to several different classes (constitutional, topological, geometric, quantum mechanical, and thermodynamic) were computed for every molecule analyzed and were chosen on the basis of their quality of fit and statistical significance. In this report, the results for the prediction of the impact sensitivities of 227 nitroorganic energetic materials using QSPRs derived from a large set of molecular descriptors using the computer program CODESSA [23] is reported. The relationships between the molecular descriptors appearing in the QSPRs and impact sensitivity are also discussed.

2. Technical discussion

2.1. Molecular modeling

All of the energetic materials investigated were modeled using the AM1 semiempirical quantum mechanical method [22] as implemented in the program AMPAC With Graphical User Interface [23]. Several of the descriptors computed within CODESSA have values that depend explicitly on the three-dimensional structure of the molecule being investigated. Therefore, we assumed that the most representative conformer of each of the materials investigated corresponded to the conformer having the lowest obtainable computed enthalpy of formation (ΔH_f) for that material. To obtain the lowest ΔH_f a conformational analysis of each molecule was conducted. In the conformational analysis each rotable bond in the molecule was rotated either 180° or 360° (in 10° increments), depending on the geometry of the molecule around the bond at which the group was being rotated. The lowest enthalpy conformer was then fully optimized and characterized, being found to give zero negative eigen-values when subjected to a frequency analysis. The lowest enthalpy conformers for each of the rotable bonds were then compared and conformer having the lowest ΔH_f was retained. The conformational analysis was conducted in this way in order to approximately obtain the global minimum enthalpy conformer for each material investigated. The assumption utilized here is that the global minimum energy conformer is the most representative conformation among those accessible to the materials under investigation, although there may be other conformations present in the condensed crystalline environment as a result of entropy and packing effects.

2.2. Computation of descriptors and QSPR derivation

Generation of descriptors within CODESSA is accomplished in one of two ways. First, several descriptors are read directly into the

program from the output files created using AMPAC for each molecule in a training set of molecular structures. Second, the remaining descriptors are computed once the AMPAC output file is opened by CODESSA. The largest number of molecular descriptors that was generated for any one material in this study was 530. Since some molecular descriptors are atom-type dependent, and thus are not necessarily defined for all molecules investigated, those descriptors not having defined values for all molecules were eliminated to give a working set of descriptors for consideration in the derivation of multivariate linear correlation equations. CODESSA has a number of algorithms for deriving correlation equations from a working set of descriptors through simple least squares (LS) regression. In this investigation the two algorithms that were chosen were the best multilinear regression (BMLR) and the Heuristic regression algorithms.

Initially, in the BMLR analysis those descriptors having values that did not vary for the molecules in the molecular training set were further eliminated from the working set of descriptors. This resulted in 412 descriptors being considered for regression analysis. Following this elimination, single-descriptor correlation equations were computed for all descriptors. Also, the inter-correlation between all descriptor pairs was computed and those descriptor pairs having a low degree of intercorrelation ($R_{i,j}^2 < 0.1$) were regressed on the experimental values ($\log_{10} h_{50\%}$). The 400 two-descriptor correlation equations having the highest coefficients of multiple determination were then considered for a forward regression derivation. Provided that additional descriptors were not significantly intercorrelated ($R_{\text{collinear}}^2 < 0.6$) with any descriptor already in the top 400 two-descriptor equation, descriptors are added in the best multilinear algorithm. The addition of descriptors proceeds while they improve the Fisher criterion (at the given probability level) and the coefficient of multiple determination over the previous top 400 correlation equations until a threshold improvement in the coefficient of multiple determination for the equation is no longer satisfied ($\Delta R_{\text{min}}^2 > 0.02$).

In CODESSA the heuristic regression algorithm also begins by eliminating descriptors when they are not defined or do not vary across the training set of molecules. Descriptors are also eliminated from consideration if they fail to meet a number of user-definable criteria, including: a Fisher test value of less than 1.0, insignificant correlation of a single descriptor ($R_{\text{min}}^2 > 0.01$), low significance according to the student's t-test for a descriptor ($t < 0.1$), significant intercorrelation with another descriptor ($R_{i,j}^2 > 0.8$) when the other descriptor has a higher single-descriptor coefficient of determination. This resulted in 347 descriptors being considered for further regression analysis. The Heuristic algorithm also operates via a forward regression derivation, but it differs from the best multilinear algorithm in that descriptors are added to generate two-descriptor correlation equations so as to improve the Fisher test value until addition of a descriptor fails to improve the correlation beyond one-third of the maximum Fisher value for all of the equations in the working set. The algorithm then proceeds to add descriptors to each equation until the Fisher value can no longer be improved, or until the maximum number of descriptors allowed by the user have been added. CODESSA then provides the user-defined number of correlation equations, in order, according to their highest coefficients of multiple determination.

2.3. Initial screening of energetic materials for further analysis

The property chosen for the derivation of the correlations was obtained from a published database (SSK) of drop height impact sensitivity ($\log_{10} h_{50\%}$) containing 258 entries [19]. From the SSK database, only those compounds for which there is no ambiguity regarding the molecular structure were retained. Thus, experimental data and the corresponding molecular structures were eliminated from modeling if they met one of the following criteria: they are salts, there are multiple database entries with the same compound name having differing experimental values reported, or the name given is ambiguously incorrect. This screening led to the inclusion of 227 compounds being included in the first training set of molecular structures, resulting in the most comprehensive and systematic published quantum mechanically derived quantitative structure–property relationship for energetic materials.

3. Results

A comparison was made of the single best eight-descriptor correlation equations obtained using the BMLR and Heuristic

Table 1
Correlation equation parameters and associated significance values for the single best eight-parameter correlation equation from the CODESSA best multilinear regression (BMLR) and Heuristic regression algorithms

Algorithm	Descriptor	Coefficient	Student's t[a]
BMLR ($R^2 = 0.7512$, $F = 82.3$,[b] $R_{CV}^2 = 0.7260$[c])	Avg. bond order of a C atom	1.21	7.332
	Tot. molecular 1-center e-n attraction/# of atoms	4.44×10^{-3}	4.942
	Relative number of double bonds	−2.74	−5.788
	Avg. bond order of a N atom	0.909	6.556
	Max. bond order of a O atom	0.798	6.261
	Max. π–π bond order	−0.252	−2.542
	Max. σ–π bond order	−9.51	−4.409
	Avg. 1-electron reactivity index for a C atom	50.9	3.433
	Intercept	0.752	2.241
Heuristic ($R^2 = 0.7442$, $F = 79.3$,[b] $R_{CV}^2 = 0.7174$[c])	Relative number of double bonds	−3.22	−7.602
	Avg. bond order of a O atom	0.284	1.035
	Max. σ–π bond order	−11.0	−5.156
	Max. bond order of a O atom	0.583	5.235
	Avg. bond order of a C atom	1.19	5.916
	No. of occupied electronic levels/# of atoms	−0.693	−4.817
	Avg. bond order of a N atom	0.754	3.428
	Avg. 1-electron reactivity index for a C atom [31]	49.9	3.312
	Intercept	1.25	3.652

[a] All are significant at the $p < 0.02$ level with 203 degrees of freedom.
[b] Fisher statistic value.
[c] Leave-one-out cross-validated coefficient of multiple determination.

regression algorithms (see Table 1). Interestingly, both algorithms nearly converge on the same result for this training set of compounds with nearly equal coefficients of multiple determination and six out of eight descriptors in common. Since these correlation equations are very similar the one having the higher coefficient of multiple determination (since this equation should have a somewhat greater predictive capability) was used in subsequent analyses. In an effort to improve the predictive capability resulting from this analysis a correlation equation containing 11 descriptors was derived using the Heuristic algorithm. The resulting correlation equation is given in Table 2. Comparison of the 8 ($R^2 = 0.7512$) and 11 descriptor ($R^2 = 0.7568$) models indicates little improvement in the predictive capability as evidenced by the small difference in the respective coefficients of multiple determination. In addition, both the 8 descriptor and the 11 descriptor correlation equations have 7 out of 8 descriptors in common and 2 of the descriptors were found not to have coefficients that were better than zero at the $p < 0.2$ level. For these reasons the model was not considered for further analysis.

It has been shown that ordinary LS regression is highly sensitive to the presence of outliers in either experimental data or the descriptors that are fit to the data [24]. A solution to the problem of outliers is the use of robust regression. In our robust regression analysis least median of squares regression (LMS) was performed using Stata 10 [25]. In the LMS regression the original 8 descriptors were used. The residuals from the LS regression ($R^2 = 0.7512$) and the LMS regression were then standardized by dividing the raw residual by the scale estimate corresponding to each of the respective fits. The scale estimate is given by $s = \sqrt{s^2}$, where

$$s^2 = \frac{1}{n-p} \sum_{i=1}^{n} r_i^2$$

for n observations and $p = 2$ degrees of freedom, and where r_i is the residual for the ith experimental quantity. The scaled residuals (r_s) were then plotted vs. the calculated values in Fig. 1. Those scaled residuals having quantities such that $|r_s| \geq 2.5$ were then identified as outliers [24]. As shown in Fig. 1 six data points were identified as outliers in the LS regression and seven data points were identified as outliers in the LMS regression. The seven outliers identified in the LMS regression were then removed from further consideration.

Following the removal of the outliers, a group of materials was left out of LS regression analysis and treated as a test set for external validation. In order to choose the test set of molecules
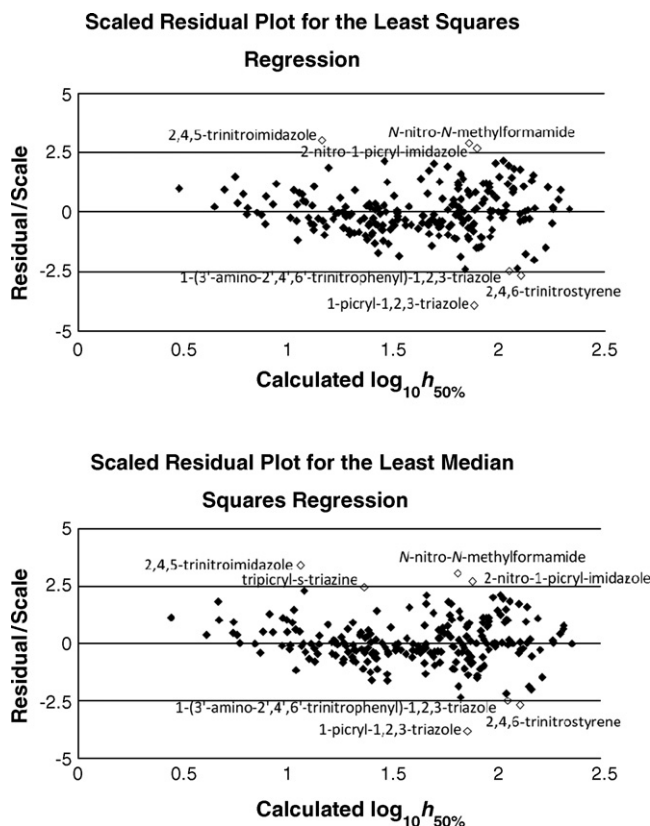

Fig. 1. Scaled residual plots for the LS and LMS regressions.

from the training set of molecules (less seven outliers) a histogram analysis was performed. The number of subclasses of experimental data points in the training set was determined using Sturges' rule, $k = 1 + \log_2 N$, where N is the number of compounds in the training set, 220. This resulted in eight subclasses of experimental data points and one compound was chosen from each subclass. The range of data values was divided into eight equal sub-ranges, with the midpoint of each sub-range being chosen as the target experimental data point for selection. The compound corresponding to the data point that was closest to the midpoint of each of the eight sub-ranges was chosen as the compound to remove for the leave-group-out cross-validation. Following removal of the eight test set compounds the BMLR regression algorithm was then used on the 212 compounds in the

Table 2
Eleven-descriptor correlation equation obtained using the Heuristic algorithm

Algorithm	Descriptor	Coefficient	Student's t [a]
Heuristic ($R^2 = 0.7568$, $F = 77.7$,[b] $R_{CV}^2 = 0.7206$[c])	Relative number of double bonds	−2.91	−5.554
	Avg. bond order of a O atom	0.306	1.047[d]
	Final heat of formation/# of atoms	-3.18×10^{-3}	−0.409[d]
	Avg. bond order of a C atom	0.871	3.928
	Max. σ–π bond order	−8.23	−3.575
	Max. bond order of a O atom	0.588	4.464
	Avg. 1-electron reactivity index for a C atom	50.1	3.352
	Tot. molecular 1-center e-n attraction/# of atoms	4.9×10^{-3}	4.416
	Avg. bond order of a N atom	0.968	3.737
	Max. e–n attraction for a N–O bond	-3.16×10^{-2}	−3.002
	Min. valency of a C atom	−1.43	−2.468
	Intercept	23.3	3.627

[a] Unless otherwise indicated significant at the $p < 0.01$ level.
[b] Fisher statistic value, significant at the $p < 0.01$ level.
[c] Leave-one-out cross-validated coefficient of multiple determination.
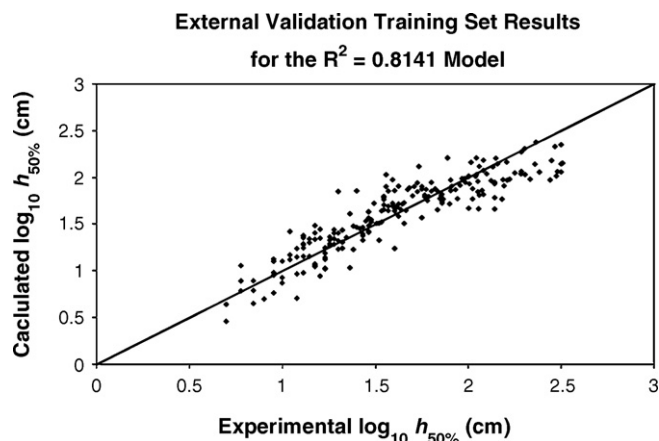[d] Not significant at the $p < 0.2$ level.

External Validation Training Set Results

for the R^2 = 0.8141 Model

Fig. 2. Comparison of calculated and experimental values resulting from the training set for external validation.

Residual Plot for the R^2 = 0.8141 Model

Fig. 3. Residual plot for the R^2 = 0.8141 model.

External Validation Test Set Results

for the R^2 = 0.8141 Model

Fig. 4. Experimental and calculated values for the test set for external validation of the R^2 = 0.8141 model.

training set and predictions were made on the compounds removed for the test set. The resulting correlation equation is given in Table 3. The resulting calculated and experimental values for the test set are compared in Fig. 2 [26]. The corresponding residuals are plotted in Fig. 3. From Fig. 3 it is apparent that the residuals are not randomly distributed. The curvature in Fig. 2 for compounds which exhibit the lowest impact sensitivity ($\log_{10} h_{50\%} > 2.0$) indicates that for these compounds the impact sensitivity is generally under predicted. It is possible that a non-linear fit may improve the model's performance for highly insensitive compounds. The lack of performance in the model could also be attributed to a greater amount of experimental error for these compounds.

The test set of compounds that were removed and their respective experimental and calculated impact sensitivity values are compared in Fig. 4. As indicated by Fig. 4 the regression equation appearing in Table 3 predicts the trend in drop height impact sensitivities for the test set of materials reasonably well, with the exception of two compounds. For 2,4,6-trinitrobenzyl alcohol the drop height impact sensitivity value is over-predicted and for 4,4,8,8-tetranitro-1,11-dinitrato-6-nitrazaundecane the drop height impact sensitivity value is under predicted. The residuals and relative errors for the test set are given in their experimentally determined units (cm) in Table 4. It is worthy to note that for compounds exhibiting the highest impact sensitivity ($h_{50\%} < 50$ cm) that the mean unsigned error is 3.4 cm. This suggests that the predictions of the impact sensitivities of the most sensitive and dangerous materials are the most accurate for the obtained model.
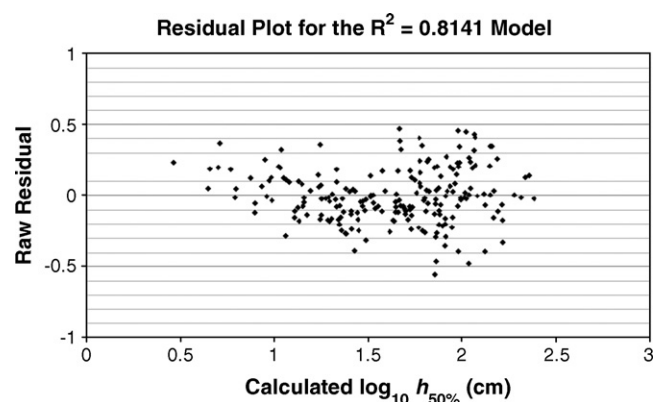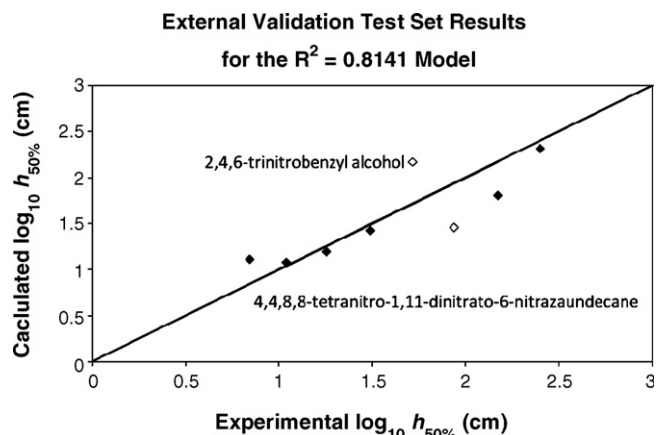
The exact value of many of the descriptors in the correlation equation used to predict this trend explicitly depends on the conformation of the molecule from which they are calculated. Because the adopted conformation for each molecule may not accurately reflect the condensed-phase conformation of the molecules investigated, conformers corresponding to rotation of the bond between the aromatic ring and the alcoholic carbon of 2,4,6-trinitrobenzyl alcohol were built and descriptors were computed for each of these conformers. The descriptors for each of these conformers were then used to calculate a value for the $\log_{10} h_{50\%}$ value for each conformer as a function of the dihedral angle for the rotated bond. The results for the eight-descriptor

Table 3

Eight-descriptor correlation equation obtained using the BMLR algorithm having removed seven outliers and eight materials for external validation

Algorithm	Descriptor	Coefficient	Student's t[a]
BMLR (R^2 = 0.8141, F = 111.2,[b] $R^2_{CV} = 0.7951$[c])	Min. internuclear repulsion for a N–O bond	1.046×10^{-2}	2.090
	Relative number of double bonds	−3.238	−7.496
	Avg. bond order of a C atom	1.015	6.990
	Max. bond order of a O atom	0.565	5.951
	Max. σ–π bond order	−10.694	−5.346
	Avg. bond order of a N atom	0.890	6.726
	Tot. molecular 1-center e-n attraction/# of atoms	4.023×10^{-3}	4.776
	Avg. 1-electron reactivity index for a C atom	62.00	4.730
	Intercept	−1.262	−1.046[d]

[a] Unless otherwise indicated, all are significant at the $p < 0.05$ level.

[b] Fisher statistic value, significant at the $p < 0.001$ level with 8 and 203 degrees of freedom.

[c] Leave-one-out cross-validated coefficient of multiple determination.

[d] Not significant at the $p < 0.2$ level.

Table 4
Calculated and experimental values with errors and relative errors (%) with average values for the test set of compounds predicted using the $R^2 = 0.8141$ correlation equation

Structure	Calc. $h_{50\%}$ (cm)	Exp. $h_{50\%}$ (cm)	Error (cm)	Relative error (%)
1,1,1,6,6,6-Hexanitro-3-hexyne	13.0	7.0	6.0	85.6
1,1,1,4,6,6,8,11,11,11-Decanitro-4,8-diazaundecane	12.1	11.0	1.1	9.6
2,2,2-Trinitroethyl-2,4,6,6-tetranitro-2,4-diazaheptaneoate	15.7	18.0	−2.3	12.7
1-(2,2,2-Trinitroethyl)-2,4-dinitrobenzene	26.6	31.0	−4.4	14.3
2,4,6-Trinitrobenzyl alcohol	148.3	52.0	96.3	185.2
4,4,8,8-Tetranitro-1,11-dinitrato-6-nitrazaundecane	28.7	87.0	−58.3	67.1
4-Nitro-1-picryl-3-picrylaminopyrazole	63.7	149.0	−85.3	57.3
1,3-Dimethoxy-2,4,6-trinitrobenzene	204.1	251.0	−46.9	18.7
Average			−11.7	56.3
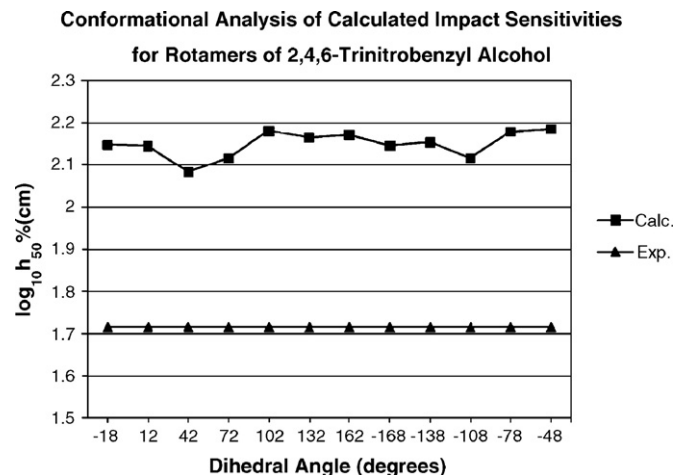
Fig. 5. Comparison of calculated and experimental impact sensitivities for rotamers of 2,4,6-trinitrobenzyl alcohol.

model ($R^2 = 0.8141$) are shown in Fig. 5. The calculated values in Fig. 5 vary by approximately 0.1 log units over the range of dihedral angles analyzed, suggesting that the descriptor values are somewhat insensitive to these conformations of 2,4,6-trinitrobenzyl alcohol.

For the eight-descriptor correlation equation from the BMLR algorithm ($R^2 = 0.8141$) the standard errors and the variance inflation factors (VIFs) are given in Table 5. The VIF values serve as a measure of the extent of collinearity between the descriptors in the correlation equation and are calculated by VIF = $1/(1 - R^2)$ where R^2 corresponds to the coefficient of multiple determination resulting from regressing each descriptor on the remaining descriptors in the model. Generally VIF values less than 5.0 are considered acceptable [27,28]. The values given in Table 5 suggest that the descriptors in this model are approximately orthogonal and valid.

Table 5
Standard errors and variance inflation factors (VIF) for each descriptor appearing in the $R^2 = 0.8141$ correlation equation

Descriptor	S.E.	VIF
Min. internuclear repulsion for a N–O bond	5.00×10^{-3}	1.61
Relative number of double bonds	0.432	3.93
Avg. bond order of a C atom	0.145	1.90
Max. bond order of a O atom	9.50×10^{-2}	1.33
Max. σ–π bond order	2.00	1.45
Avg. bond order of a N atom	0.132	1.47
Tot. molecular 1-center e-n attraction/# of atoms	8.42×10^{-4}	3.25
Avg. 1-electron reactivity index for a C atom	13.1	1.04
Intercept	1.21	

4. Conclusions

The eight-descriptor multivariate linear regression equation ($R^2 = 0.8141$) obtained in this study is capable of distinguishing the relative drop height impact sensitivities of a structurally diverse set of nitroorganic compounds. The descriptors appearing in the regression model were obtained from a large initial set and chosen on the basis of their goodness of fit and statistical significance. The regression model predicts impact sensitivities most accurately for compounds having the highest sensitivity to impact (i.e. those compounds having small values for their drop height impact sensitivity).

The error in the prediction of impact sensitivities may be a result of a number of different factors. The drop height impact sensitivity test is relatively easy to implement, but often provides experimental results with a low degree of reproducibility. Consequently, a regression model that is built on experimental data having a large amount of indeterminate variance will exhibit greater error. Additionally, the large variation in structural types has probably diminished the predictive capability of this model since it is a well-known phenomenon in QSAR studies that sets of molecules possessing very similar structures tend to give the best correlations. In biological QSAR models compounds that have very different activities in a QSAR model can be the result of the compound acting by a very different mechanism that the other molecules in the training set [29]. A similar phenomenon is likely occurring within the present data set, that is, it is likely that different types of energetic materials are exhibiting their drop height impact sensitivity by different detonation mechanisms. Additionally, the method of modeling the energetic materials as isolated adiabatic molecules is an approximation and does not take into account the effects due to the crystalline environment, which are known to be important and vary from material to material. The error in any model can also be attributed to limitations in the theory used to derive the model. The descriptor space from which the descriptors are selected may be too limited in this case, or there may be discontinuities within the descriptor space [30]. The AM1 semiempirical quantum mechanical model was used for molecular modeling and the computation of descriptors used in this regression model. Although the AM1 method is computationally efficient and capable of reproducing experimental trends accurately, other theoretical methods may be employed in an attempt to improve the accuracy with which drop height impact sensitivities can be predicted.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2008.06.003.

References

[1] A. Macek, Chem. Rev. 62 (1) (1962) 41–63.
[2] B.M. Rice, J.J. Hare, J. Phys. Chem. A 106 (9) (2002) 1770–1783.
[3] W.C. Lothrop, G.R. Handrick, Chem. Rev. 44 (3) (1949) 419–445.
[4] J. Mullay, Propell. Explos. Pyrotech. 12 (1987) 60–63.
[5] M. Vaullerin, A. Espagnacq, L. Morin-Allory, Propell. Explos. Pyrotech. 23 (1998) 237–239.
[6] M.H. Keshavarz, H.R. Pouretedal, J. Hazard. Mater. A 124 (2005) 27–33.
[7] M.H. Keshavarz, H.R. Pouretedal, A. Semnani, J. Hazard. Mater. 141 (2007) 803–807.
[8] C.B. Storm, R.R. Ryan, J.P. Ritchie, J.H. Hall, S.M. Bachrach, J. Phys. Chem. 93 (2) (1989) 1000–1007.
[9] J. Sharma, B.C. Beard, M. Chaykovsky, J. Phys. Chem. 95 (3) (1991) 1209–1213.
[10] Y. Kohno, K. Maekawa, T. Tsuchioka, T. Hasizume, A. Imamura, Chem. Phys. Lett. 214 (6) (1993) 603–608.
[11] X.J. Xu, W.H. Zhu, H.M. Zhao, J. Phys. Chem. B 111 (8) (2007) 2090–2097.
[12] F. Jianfen, X. Heming, Theochemistry 365 (1996) 225–229.
[13] C. Mereditch, T.P. Russell, R.C. Mowrey, J.R. McDonald, J. Phys. Chem. A 102 (2) (1998) 471–477.
[14] X.J. Xu, H.M. Xiao, X.D. Gong, X.H. Ju, Z.X. Chen, J. Phys. Chem. A 109 (49) (2005) 11268–11274.
[15] A. Hammerl, T.M. Klapotke, H. Noth, M. Warchold, Propell. Explos. Pyrotech. 28 (4) (2003) 165–173.
[16] J. Edwards, C. Eybl, B. Johnson, Int. J. Quant. Chem. 100 (5) (2004) 713–719.
[17] C. Zhang, Y. Shu, Y. Huang, X. Zhao, H. Dong, J. Phys. Chem. B. 109 (18) (2005) 8978–8982.
[18] X.S. Song, X.L. Cheng, X.D. Yang, B. He, Propell. Explos. Pyrotech. 31 (4) (2006) 306–310.
[19] C.B. Storm, J.R. Stine, J.F. Kramer, Sensitivity relationships in energetic materials, in: S.N. Bulusu (Ed.), Chemistry and Physics of Energetic Materials, Kluwer Academic, Netherlands, 1990.
[20] S.G. Cho, K.T. No, E.M. Goh, J.K. Kim, J.H. Shin, Y.D. Joo, S. Seong, Bull. Korean Chem. Soc. 26 (3) (2005) 399–408.
[21] H. Nefati, J.M. Cense, J.J. Legendre, J. Chem. Inf. Comput. Sci. 36 (4) (1996) 804–810.
[22] M.J.S. Dewar, E.G. Zoebisch, E.F. Healey, J.J.P. Stewart, AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (13) (1985) 3902–3909.
[23] Semichem Inc., Shawnee, KS.
[24] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, Hoboken, NJ, 2003.
[25] StataCorp., Stata Statistical Software: Release 10, StataCorp. LP, College Station, TX, 2007.
[26] The calculated and experimental values for all compounds in the resulting correlation equation ($R^2 = 0.8141$) are available as supplementary material.
[27] J.D. Jobson, Applied Multivariate Data Analysis, vol. 1, Springer-Verlag, New York, 1991.
[28] D.A. Belsey, E. Kuh, R.E. Welsche, Regression Diagnostics, Wiley, New York, 1980.
[29] R.P. Verma, C. Hansch, Bioorg. Med. Chem. 13 (2005) 4597–4621.
[30] G.M. Maggoria, J. Chem. Inf. Comput. Sci. 46 (4) (2006) 1535.
[31] The average one-electron reactivity index ($\bar{R}_C$) for a carbon atom is computed by, $\bar{R}_C = \sum_{i \in C} \sum_{j \in C} c_{i\text{LUMO}} c_{j\text{HOMO}} / (\varepsilon_{\text{LUMO}} - \varepsilon_{\text{HOMO}})$ where $c_{i\text{LUMO}}$ and $c_{j\text{HOMO}}$ are the atomic orbital coefficients to the LUMO and the HOMO, respectively, $\varepsilon_{\text{LUMO}}$ and $\varepsilon_{\text{HOMO}}$ are the respective energies of these orbitals, and the summation performed over all of the carbon atoms in the molecule. For a discussion of this descriptor, see K. Fukui, Theory of Orientation and Stereoselection, Springer-Verlag, Berlin, 1975.