A QSAR MODEL FOR PREDICTING SOLVENTS AND SOLVENT BLENDS FOR ENERGETIC MATERIALS

Darren L. Williams, Karl D. Kuklenz, *Chemistry Department, Sam Houston State University, 1003 Bowers Blvd, Huntsville, TX 77341-2117, Email: williams@shsu.edu*

Abstract

Researchers in the paint and polymer industry have shown that the Hansen solubility parameters (HSP) are useful for predicting suitable solvents for the filled-polymer formulation process. To apply this work to the high explosive formulation process, the HSPs of the various energetic materials must be determined or predicted.

A quantitative structure activity relationship (QSAR) was developed that is based upon the output of a density functional theory optimization and frequency calculation (B3LYP/6-31G(d)//B3LYP/6-31G(d)) using the Gaussian 03 computational package. Structural parameters were extracted from the Gaussian output files of each molecular species. These consisted of the geometric mean of the exact polarizability tensors (α, Å^3), the dipole moment (μ, Debye) the highest occupied molecular orbital energy (*HOMO*, Hartree), the number of each type of atom, and the delta charge (Δq) – defined as the difference between the most negative heteroatom and the most positive hydrogen in the molecule. The value of $\Delta q = 0$ was given to hydrocarbons by fiat. A stepwise linear regression was used to determine the correlation of these inputs and mathematical transformations of these inputs to the HSPs for a training set of 54 solvents and nitrated compounds. The resulting QSAR matrix was then applied to 23 energetic materials and precursors yielding the HSPs (δ_D, δ_P, δ_H) in MPa$^{1/2}$.

The HSPs were also determined for HMX, RDX, PETN, and HNS using experimental solubility data and the group additivity methods of Van Krevelen and Stefanis. The QSAR model outperformed the group additivity methods in matching the experimentally determined HSPs using the Hansen distance parameter (R_a) as the figure of merit.

En route to the QSAR model, a very simple model of molar volume was developed wherein the molar volume is computed directly from the molecular formula $C_aH_bN_cO_dS_eP_fF_gCl_hBr_i$ via the following equation: $V_m = 12.53 + 8.77a + 3.96b + 4.87c + 6.12d + 17.22e + 19.45f + 9.70g + 18.66h + 20.74i$. The correlation of this equation with the literature values of 183 molecules was 99.67% with an $R^2 = 0.9847$ over a range of 400 cm^3/mol.

1 Introduction

Solubility of explosives and polymer binders is a major factor in the formulation, recrystallization, and demilitarization processes. Solvents, co-solvents, and non-solvents are used to form molding powders, to change particle size distributions, and to separate HE from binder.

Since these activities are not new endeavors, there are well-established processes and process fluids. However, the area of solubility is active for many reasons. There is a constant push to reduce or eliminate the use of toxic solvents with more benign alternatives. There is also a mandate to eliminate ozone-depleting solvents. And lastly, there are efforts to remove solvents from the process stream that leave behind corrosion-promoting chlorides. [1]

The paint and polymer industries have led the way in solvent substitution and solvent blend prediction activities. In particular, the Hansen solubility parameters (HSPs) have proven themselves and are described in detail in the literature [2 – 6]. The HSPs may be used to qualitatively rank solvents and blends in terms of their interaction with a given solute. Solutes and solvents that have similar HSPs are predicted to mix spontaneously. Conversely, if the HSPs are drastically different between solute and solvent, then there will be little interaction and limited mixing.

Several group additivity methods have been developed to build up the Hansen solubility parameters from the various constituent chemical groups present in a solute. In an effort to produce a universal method, the developers of group additivity methods must include every possible chemical group that represents all bonding types. In an effort to increase the accuracy of these methods, the developers have introduced secondary structure arrangements. The drive towards universality has led to multiple ways to build a given solute, and non-obvious priority is given to certain bonding types and secondary structures.

Computational chemistry programs are now available to almost all chemical researchers, and these programs have delivered the universal ability to model solutes that the group additivity methods have sought to achieve. This presentation outlines a new approach wherein a quantitative structure activity relationship (QSAR) is developed that converts the structural parameters from a computational chemistry output file into the HSPs for the solute.

2 Experimental

The experimental solubility data for HMX, HNS, PETN, and RDX was obtained from Pantex [8] and the US Army [9]. The data was analyzed using the Hansen solubility sphere optimization method [4] using Microsoft Excel as described in an earlier paper [10]. The operational definition of solvent was varied for each optimization in order to find the solubility sphere that included all of the solvents and contained the fewest non-solvents. The cutoff values for the definition of solvent versus non-solvent for each HE are given in Table 1.

The group additivity methods of Van Krevelen [5] and Stefanis [6] were used to estimate the Hansen solubility parameters of HMX, HNS, PETN, and RDX. Hoy's method [2] was not used because of the need for critical temperatures and boiling points – both of which are unavailable for explosives.

2.1 QSAR Method Training

The Gaussian 03 (G03) computational chemistry package [7] was used to optimize the geometry and compute the vibrational frequencies of 54 solvents and nitrated compounds from [4] with the B3LYP/6-31G(d) density functional model chemistry. Although more accurate model chemistries were available, this one was chosen so that the resulting QSAR model would be accessible to the large number of researchers who are operating Gaussian on desktop PCs.

Structural parameters were extracted from the Gaussian output files of each molecular species. These consisted of the geometric mean of the exact polarizability tensors (α, $Å^3$), the dipole moment (μ, Debye) the highest occupied molecular orbital energy (*HOMO*, Hartree), the number of each type of atom, and the delta charge (Δq) – defined as the difference between the most negative heteroatom and the most positive hydrogen in the molecule. The value of $\Delta q = 0$ was given to hydrocarbons by fiat. The molar volume was calculated based upon a linear regression of the molecular formulae for the 54 chemical species.

Mathematical transformations of each of the structural parameters (i.e. the square, the inverse, etc.) were included in the statistical model. Many were guided by Barton's chapter on molecular interactions [2]. For example, dipole-dipole interactions involve the square of the dipole moment. Dipole-induced dipole interactions depend upon the polarizability times the square of the dipole moment. Induced dipole-induced dipole interactions are proportional to the product of the polarizability and the first ionization potential (I_1). Our model uses the HOMO energy in lieu

of the first ionization potential because explicit calculation of I_1 would require a time-consuming set of calculations on the cations of all species of interest. The evaluation of the HOMO approximation of I_1 was performed on the reported ionization potentials of H – Ar [11]. A linear relationship *(I_1 = 0.711(HOMO)-54.68)* was found for the B3LYP/6-31G(d) single point energy HOMO of the atom versus the literature value of I_1 of the atom with an R^2 = 0.9849.

The resulting twenty-seven structural parameters from G03 were used as inputs to a stepwise regression analysis. The dependence of the literature δ_D, δ_P, and δ_H upon the twenty-seven input parameters was determined using the Minitab 15 statistical software package. The structural parameters were inserted or removed based upon their p-value. A p-value greater than 0.2 was needed for a parameter to be retained, and the overall correlation constant was used to drive the regression optimization. Only fourteen structural parameters survived the regression analysis.

The QSAR matrix shown in Eq. (1) is defined as those coefficients (C_{uv}) that transform the structural parameters (S_u) into the activity terms δ_D, δ_P, and δ_H. The zeroth structural parameter allows for the calculation of a constant term in the coefficient matrix (C_{0v}).

$$\begin{bmatrix} 1 & S_1 & S_2 & \cdots & S_u \end{bmatrix} \begin{bmatrix} C_{0,\delta_D} & C_{0,\delta_P} & C_{0,\delta_H} \\ C_{1,\delta_D} & C_{1,\delta_P} & C_{1,\delta_H} \\ \vdots & \vdots & \vdots \\ C_{u,\delta_D} & C_{u,\delta_P} & C_{u,\delta_H} \end{bmatrix} = \begin{bmatrix} \delta_D & \delta_P & \delta_H \end{bmatrix} \quad (1)$$

The B3LYP/6-31G(d) model chemistry was used to calculate the geometry and vibrational frequencies of the 23 molecules shown in Fig. 1: diaminotrinitrobenzene (DATB), ethylpicrate, ethyltetryl, cyclotetramethylenetetranitramine (HMX), hexanitroazobenzene (HNAB), hexanitrostilbene (HNS), pentaerythritol tetranitrate (PETN), cyclotrimethylenetrinitramine (RDX), phloroglucinol, picric acid, picramide, tetranitrodibenzotetrazapentalene (TACOT), triaminotrinitrobenzene (TATB), triethyltrinitrobenzene (TETNB), tetryl, trinitroanisole (TNA), trinitrobenzene (TNB), trinitrobenzoic acid (TNBA), tetranitrocarbazole (TNC), trinitrochlorobenzene (TNCB), trinitrophloroglucinol (TNPG), trinitroresorcinol (TNR), and trinitroxylene (TNX). These were chosen to cover a variety of species relevant to the explosive formulation industry. The QSAR matrix from the 54-molecule training set was used to convert the Gaussian 03-derived structural parameters of these molecules into HSPs.

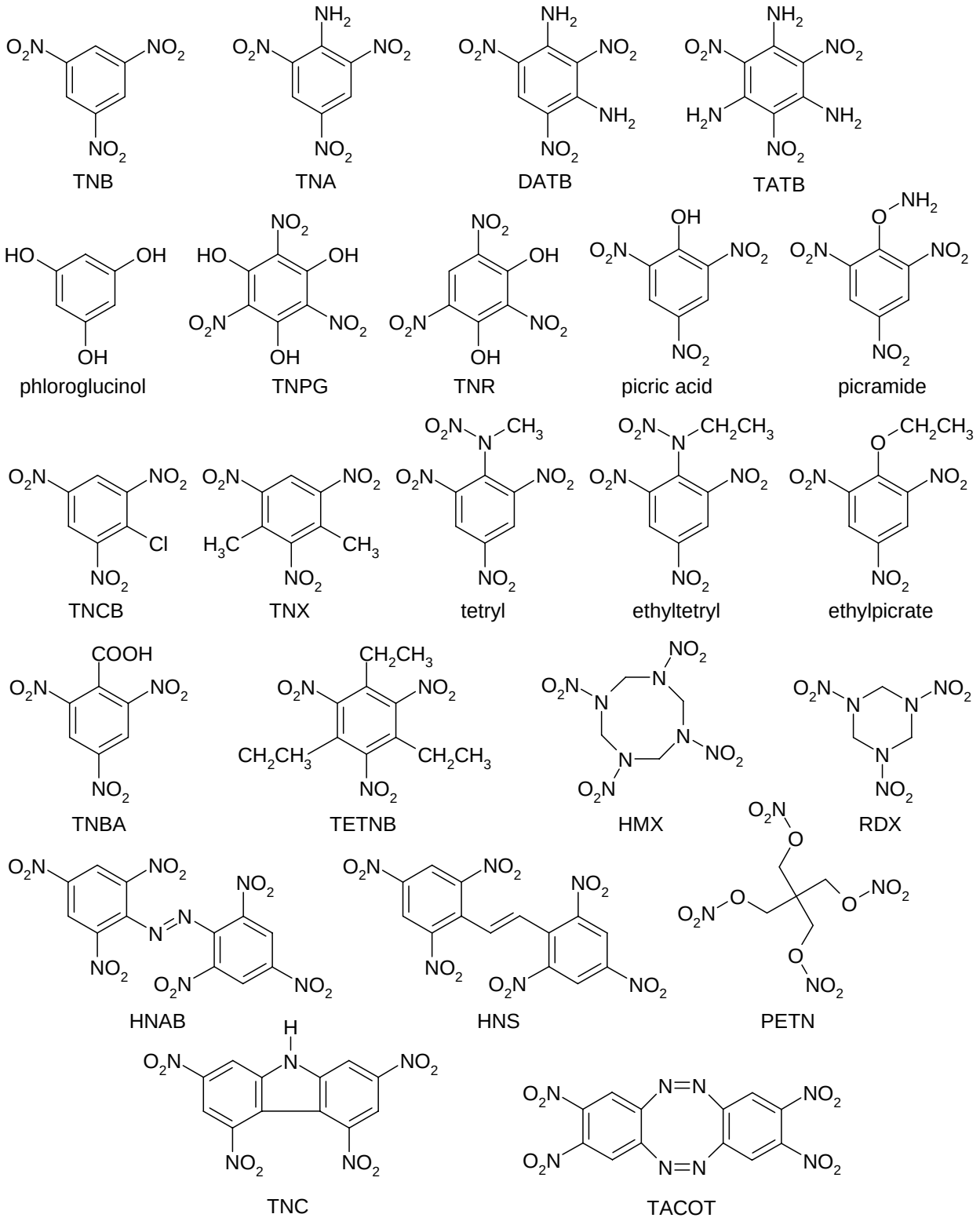

Figure 1. The molecules studied in this work.

3 Results and Discussion

The coefficients of the molar volume (V_m) regression analysis of the 183 solvents in [3] are shown in Eq. (2). This relationship allows one to calculate the molar volume and the density of a species at 25 °C directly from the molecular formula ($C_aH_bN_cO_dS_eP_fF_gCl_hBr_i$).

$$V_m = 12.53 + 8.77a + 3.96b + 4.87c + 6.12d + 17.22e + 19.45f + 9.70g + 18.66h + 20.74i \quad (2)$$

The comparison of the V_m model to the literature values of the molar volume in [3] is shown in Fig. 2. The V_m model of Eq. 2 has a 0.9967 slope over a range of 400 cm^3/mol with an R^2 = 0.9847. Readers who are interested in a broadly-applicable V_m model are encouraged to use Eq. 2.

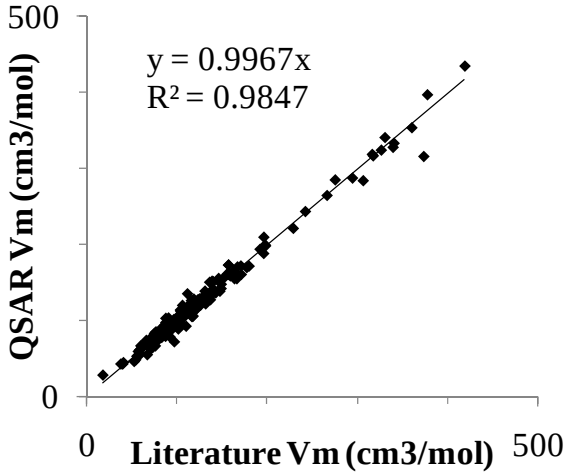

Figure 2. The correlation of the QSAR molar volume calculation to the literature values of 183 solvents in [3].

This model was applied to the 54 solvents and nitrated species in the QSAR training set used in this study. The correlation was not as good (R^2 = 0.8447) because very few of the nitrated molecules in the QSAR training set were included in the regression that yielded Eq. 2. Therefore, the molar volume regression was repeated on the QSAR training set. The results of this regression are given in Eq. 3.

$$V_m = 9.84 + 6.85a + 4.91b + 7.22c + 6.48d + 19.68e + 20.26h + 27.96i \quad (3)$$

There are no coefficients for P or F because these atoms are not represented in the 54-molecule QSAR training set. This particular regression has a 0.9950 slope with an R^2 = 0.9120. This model produced the molar volumes used in the subsequent stepwise regression.

The fourteen structural parameters that survived the stepwise regression and the coefficient matrix are given in Table 1. The overall matrix uses fourteen structural terms, but in reality the individual δ_D, δ_P, and δ_H terms required 7, 8, and 4 coefficients, respectively. This is a satisfactory over-determination using 54 equations to find 7, 8, and 4 unknowns. The correlation coefficients (R^2) of the training set are 0.7710, 0.8924, and 0.8784 for δ_D, δ_P, and δ_H, respectively (Fig. 3). The Van Krevelen group additivity-derived δ_D and δ_P values did not correlate as well

with the literature values (R^2 = 0.0343 and 0.3998, respectively). The Van Krevelen method correlated better with the literature for the δ_H parameter with an R^2 = 0.9315. However, of the 54 molecules in the training set, the QSAR method has a smaller Hansen distance (R_a) to the literature values than the Van Krevelen method for 42 of them. (Fig. 4) Also, the Van Krevelen method is incapable of calculating the HSPs for any species with a sulfur such as DMSO or CS_2.

Table 2 shows the Hansen solubility parameters for HMX, RDX, PETN, and HNS. The HSPs were determined five different ways: 1) the Hansen solubility sphere analysis of experimental solubility data, 2) the centroid of the HSPs of the solvents exhibiting solubility greater than the cutoff values, 3) the Van Krevelen [5] group additivity method, 4) the Stefanis [6] group additivity method, and 5) the QSAR method. The use of the centroid was included because the Hansen solubility spheres for HMX and RDX were not completely bound by non-solvents, and the optimization tended to walk the center of the interaction sphere toward unreasonable extremes in δ_P and δ_H.

Table 1. The structural parameters and QSAR matrix resulting from the stepwise regression. The structural parameters *c*, *b*, *d*, *h*, and *i* are defined by the molecular formula ($C_aH_bN_cO_dS_eP_fF_gCl_hBr_i$).

S_u	$C_u\delta_D$	$C_u\delta_P$	$C_u\delta_H$
1	8.995	-1.304	0.5078
$\alpha^{1/2}$	0.596	0	0
μ^2	0.081	0	0
c	-0.6	0	0
b	-0.31	0	0
$HOMO^{-1}$	-1.09	-1.45	0
μ^{-1}	0.208	0	0
$(\mu / V_m)^{1/2}$	0	59.8	-7
d	0	1.55	0
$\mu^{-1/2}$	0	-10.9	0
$\Delta q^{1/2}$	0	3.3	0
i	0	4.4	0
h	0	1.3	0
α^{-1}	0	0	252
Δq	0	0	13.41

The Hansen distances (R_a) [4] between the two experimental data HSPs and the three computational models were calculated to evaluate the performance of the three HSP models. The QSAR method was closer to the experimentally-determined HSPs for all four explosives. The Stefanis method [6] did not perform well primarily due to its over-estimation of the contribution of nitro groups to δ_P (33 to 50 $MPa^{1/2}$). The results of the QSAR determination of the HSPs of the 23 energetic materials in this study are presented in Table 3.

Table 2. The Hansen solubility parameters determined experimentally (exp), by the average of the solvents defined as good (centroid), by group additivity (GA), and by the QSAR model.

HE	Method	Cutoff g/100mL	δ_D /$MPa^{1/2}$	δ_P /$MPa^{1/2}$	δ_H /$MPa^{1/2}$	R_0 / $MPa^{1/2}$	FIT	R_a Exp/ $MPa^{1/2}$	R_a Centroid/ $MPa^{1/2}$
HMX	exp	0.10	17.7	11.6	13.7	12.1	1.00		
HMX	centroid	0.10	16.7	9.2	6.6				
HMX	GA[a]		22.5	22.9	6.9			16.4	18.0
HMX	GA[b]		23.5	45.2	7.3			36.1	38.5
HMX	QSAR		15.3	13.4	5.6			9.5	5.1
RDX	exp	1.40	17.3	12.4	9.1	8.4	1.00		
RDX	centroid	1.40	15.9	11.0	8.8				
RDX	GA[a]		18.5	18.8	5.7			7.7	9.9
RDX	GA[b]		22.0	35.8	7.5			25.3	27.7
RDX	QSAR		15.5	11.1	6.5			4.7	2.5
PETN	exp	1.50	16.7	12.0	8.4	7.8	1.00		
PETN	centroid	1.50	16.3	9.7	6.9				
PETN	GA[a]		21.4	21.2	9.5			13.1	15.5
PETN	GA[b]		18.8	50.4	3.0			39.0	41.2
PETN	QSAR		16.8	17.8	6.4			6.2	8.2
HNS	exp	0.15	18.9	13.9	6.1	6.0	0.98		
HNS	centroid	0.15	18.0	14.0	8.7				
HNS	GA[a]		21.0	13.3	8.6			4.9	6.0
HNS	GA[b]		28.0	33.0	1.9			26.7	28.4
HNS	QSAR		20.7	17.1	5.6			4.8	6.9

a Van Krevelen method, b Stefanis method

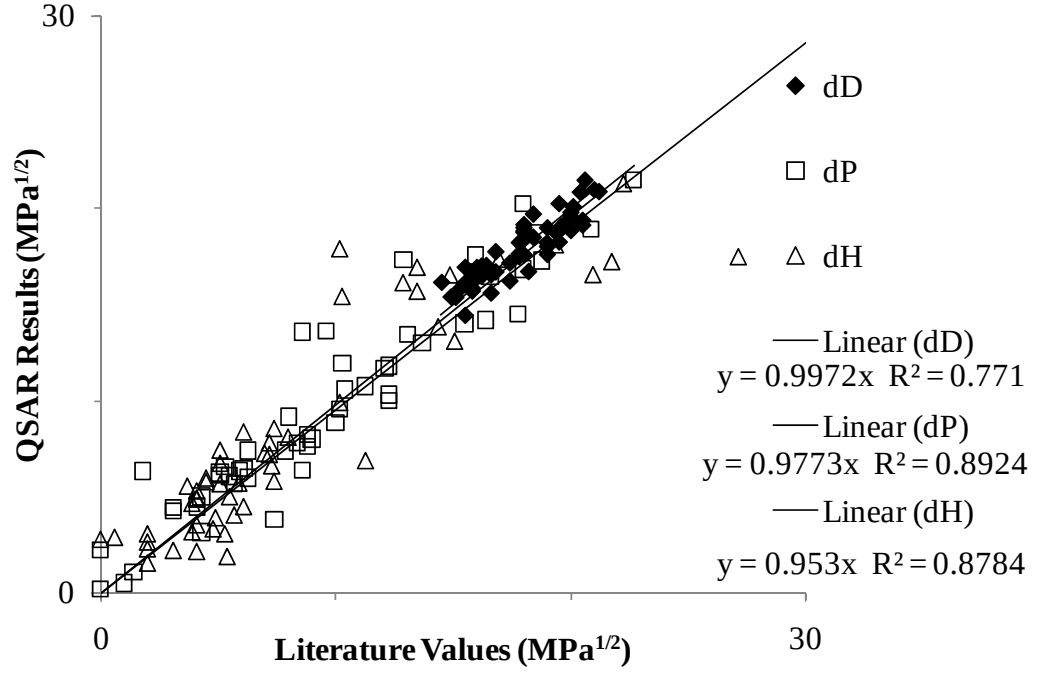

Figure 3. The correlation of the QSAR results to the literature values of δ_D, δ_P, and δ_H for 54 solvents and nitrated compounds in [4].

4 Conclusions

A QSAR model was trained against 54 molecules and was used to predict the HSPs of 23 energetic materials (Table 3).

Table 3. The QSAR-determined HSPs for energetic materials of interest to the explosive formulation community.

Abbreviated Name	δ_D /MPa$^{1/2}$	δ_P /MPa$^{1/2}$	δ_H /MPa$^{1/2}$
DATB	18.8	13.6	19.6
ethylpicrate	18.7	11.9	7.8
ethyltetryl	18.9	15.8	5.3
HMX	15.3	16.0	5.3
HNAB	23.0	22.0	6.5
HNS	20.7	17.2	5.6
PETN	16.8	18.0	6.4
phloroglucinol	19.7	11.3	15.6
picramide	19.2	14.7	19.3
picric acid	18.9	12.8	15.3
RDX	15.5	13.2	6.3
TACOT	20.3	14.6	3.3
TATB	18.5	13.5	21.1
TETNB	17.8	13.0	8.1
tetryl	18.5	14.2	5.5
TNA	18.8	12.1	7.6
TNB	19.1	10.8	6.7
TNBA	19.0	13.1	13.7
TNC	20.9	14.9	21.0
TNCB	19.5	11.5	6.5
TNPG	19.9	17.1	15.4
TNR	19.3	14.3	15.9
TNX	18.1	8.9	5.6

The Hansen solubility parameters were determined for HMX, RDX, PETN, and HNS using solubility data. The QSAR method matched these experimental results more closely than the currently-popular group additivity methods.

The strengths of the QSAR method:

1. The computational chemistry approach is as universal as the basis set of the model chemistry, and the 6-31G(d) basis set covers the elements H through Kr. See [7] for a description of other basis sets.
2. The QSAR approach produces the HSPs in a less ambiguous manner than the group additivity methods since the QSAR approach avoids priority ranking of chemical groups or secondary structure categories.
3. The computational modeling, QSAR training, and subsequent transformation is amenable to a standard procedure approach. Once the user becomes familiar with the simple steps needed to calculate a geometry optimization and frequency calculation, the HSP determination via QSAR is a simple matrix transformation of output file parameters.
4. The QSAR method is easily adapted to target a particular class of compounds by training it against that class of compounds. The only drawback to this targeted approach is the desire to create a large training set to improve the accuracy of the model. If there is limited HSP data for a particular class of compounds, then the training set and the accuracy will be limited.

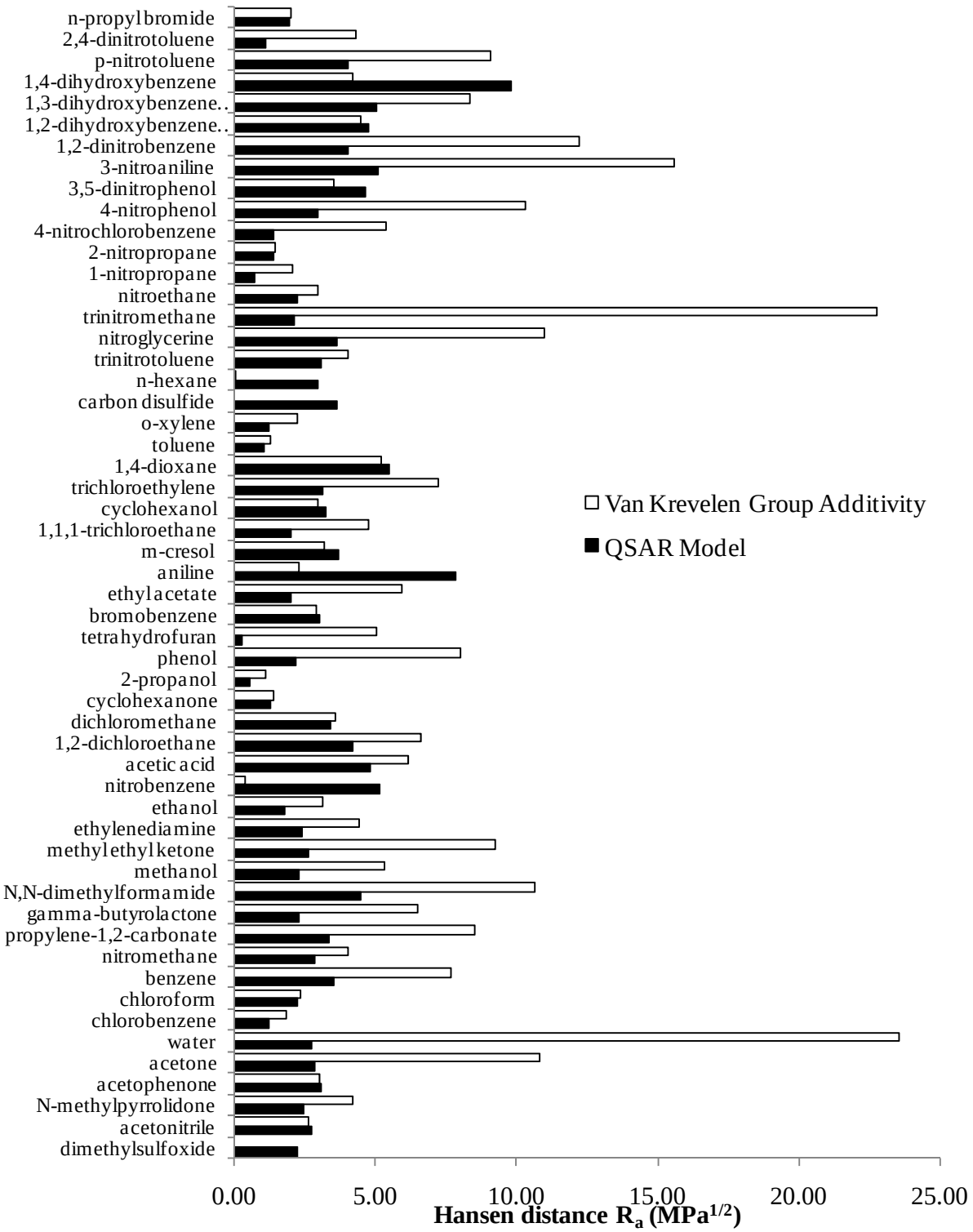

Figure 4. The Hansen distance (R_a) from the computational method to the literature values for the 54-molecule training set.

5 References

[1] U.S. Department of Energy, *Environmental Protection Program*, Order DOE O 450.1A, Attachment 1, Section 3, **2008**

[2] A. F. M. Barton, *CRC Handbook of Solubility Parameters and Other Cohesion Parameters, 2nd Ed.*, CRC Press, New York, NY, **1991**.

[3] H. Møllet, and A. Grubenmann, *Formulation Technology: Emulsions, Suspensions, Solid Forms* Tr. H. R. Payne, Wiley-VCH Weinheim (Fed. Rep. Germany), **2001**.

[4] C. M. Hansen, *Hansen Solubility Parameters: A User's Handbook 2nd Ed.*, CRC Press LLC, Boca Raton, FL, **2006**.

[5] D. W. Van Krevelen, *Properties of Polymers*, Elsevier Science Publishing Company Inc, New York **1990**, pp. 189 – 225.

[6] E. Stefanis, C. Panayiotou, Prediction of Hansen Solubility Parameters with a New Group-Contribution Method, *Int J Thermophys*, **2008**, *29*, 568 – 585.

[7] Gaussian 03, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, **2004**.

[8] A. Duncan, Unpublished Compilation of Solubility Data, B&W Pantex LLC, Amarillo, TX **2008**.

[9] US Department of the Army Technical Manual, Military Explosives, TM 9-1300-214, **1990**.

[10] D. L. Williams and K. D. Kuklenz, A Determination of the Hansen Solubility Parameters of Hexanitrostilbene (HNS), *Propellants, Explos., Pyrotech.* in press.

[11] D. R. Lide, *CRC Handbook of Chemistry and Physics, 76th Ed.*, CRC Press, New York, NY, **1995**.

Acknowledgments

This work was funded by the Sam Houston State University Faculty Research Grant program and the Plant Directed Research and Development program at B&W Pantex, LLC, under DOE prime contract DE-AC04-00AL66620. Derek Blaylock, Steven Doss, John Mayor, and Katie Rothlisberger are acknowledged for their assistance in the data analysis. Pantex scientists Arnie Duncan, Tim Quinlin, Monty Cates and technician Monte Moore provided the solubility data and input on the experimental methods.