A COMPUTATIONAL ROUTE TO HANSEN SOLUBILITY PARAMETERS OF NITRATED SPECIES THROUGH A QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR)

_____

A Thesis

Presented to

The Faculty of the Department of Chemistry

Sam Houston State University

_____

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

_____

by

Karl D. Kuklenz

May, 2009

A COMPUTATIONAL ROUTE TO HANSEN SOLUBILITY PARAMETERS OF NITRATED SPECIES THROUGH A QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR)

by

Karl D. Kuklenz

_____

APPROVED:

_____
Darren L. Williams
Thesis Director

_____
Benny E. Arney

_____
Rick C. White

Approved:

_____
Jaimie L. Hebert, Dean
College of Arts and Sciences

DEDICATION

I wish to dedicate this work to my family for more than twenty years of their love and support. I could not have been a chemist without them. I thank my mother for teaching me to be observant, observant of my environment and also of the thoughts and feelings of others. I thank my father for teaching me to foster an attention to detail and to always, always do it right the first time. I thank my sisters, Jill and Jackie, for their smile and song.

I wish a thank you to anyone has ever been a mentor or a teacher to me. Teaching is truly mankind's noblest profession.

ABSTRACT

Kuklenz, Karl Douglas, A computational route to Hansen solubility parameters of nitrated species through a quantitative structure activity relationship (QSAR). Master of Science (Chemistry), May, 2009, Sam Houston State University, Huntsville, Texas.

Purpose

The purpose of this research was to develop a universal method for the calculation of the Hansen solubility parameters (HSPs) of nitrated materials. These energetic materials are of use to the mining industry, the pyrotechnics industry, as well as the nation's military. The use of computational chemistry to approximate HSPs is a fast, cost-effective, and safe way to explore the solubility characteristics of these and other materials.

Methods

Electronic structure calculations of over 200 chemical species were performed using the Gaussian® 03W software package along with GaussView for construction and visualization of optimized geometries. Structure variables were extracted from the result files and catalogued with the Hansen solubility parameters of their respective species. A step-wise regression technique was employed to determine the coefficients of a quantitative structure activity relationship (QSAR) matrix. This matrix transforms the structure variables reported in a Gaussian® result file into the HSPs of the optimized species.

Findings

This method provides a fast and simple way to determine the HSPs of nitrated species. The method is universal in that it can be applied to any nitrated molecule that the Gaussian® program can optimize the geometry for. Larger and more diverse training

sets may allow the method to be utilized regardless of the species of interest. The method is also unambiguous in that it provides the same answer no matter how the molecule is put together. This is in contrast to functional group contribution methods which can yield varying results depending on how the molecule is assembled. This method allows for the qualitative analysis of the best solvent or blend of solvents for use with these nitrated species.

KEYWORDS: solubility parameters, solubility, computational chemistry, quantitative structure activity relationship, quantitative structure property relationship, QSAR, QSPR, energetic materials, step-wise regression.

Approved:

_____
Darren L. Williams
Thesis Director

ACKNOWLEDGEMENTS

I would especially like to thank Dr. Darren Williams for his constant guidance, support, and coffee. He made a difficult research topic easy to understand and execute.

Derek Blaylock and Katie Rothlisberger are acknowledged for their computational work. Thank you for lightening the load.

Thank you to Dr. Arney for his computational chemistry consultations.

TABLE OF CONTENTS

Page

LIST OF TABLES

TABLE **Page**

LIST OF FIGURES

FIGURE **Page**

CHAPTER I

INTRODUCTION

For many years scientists have used the "like-dissolves-like" criteria to select solvents used in their laboratories. This method is often based on trial and error and a subjective evaluation of a solvent's polarity. This method has had varying degrees of success. One of two outcomes is usually observed. If the solvent selected provided too much solubility, then the solute becomes trapped in the solvent. This is a problem if the solute is to be extracted from the solvent later. If the solvent selected does not provide enough solubility, then the mixture is left in a heterogeneous state. This requires the addition of more solvent to achieve complete solubility. Either problem is solved with a waste of materials, particularly the solvent. In recent years scientists have endeavored to remove the guesswork from solvent selection. They have turned to model-based approaches which are capable of quantifying the solubility or solvency of a solute or solvent, and the components of their models are termed solubility parameters.

History of Solubility Parameters

Solubility parameters have been around for a many years and are of use to a wide variety of industries. There are many models for solubility parameters, but the way in which all solubility parameters are used is this: materials with sufficiently similar solubility parameters will interact strongly with each other. Two liquids with similar solubility parameters will be miscible. A solvent and a solute with similar solubility parameters will likely result in a mixture that is homogeneous. A liquid on a smooth, solid surface with similar solubility parameters will spread evenly and with minimal contact angle. In this way solubility parameters can provide a systematic way of

selecting a single solvent or a blend of solvents for the dissolution of a given solute. In the paint and polymer industry this is applicable to solvent and pigment/binder formulations. Solubility parameters have also been useful to the cleaning industry where the cleaning solution is targeted to the particular soil and surface.

The term solubility parameter was first coined by Hildebrand and Scott (1,2). Hildebrand's parameter is defined as the square root of the cohesion energy per unit of volume:

$$\delta = \sqrt{E_c/V_m} \qquad (1)$$

Where E_c is the energy of vaporization and V_m is the molar volume. The energy of vaporization is proportional to the amount of cohesive energy of the material in the following way. The cohesive energy of a system is the sum of all intermolecular forces together. The energy required for any given molecule to escape the condensed phase and become a vapor would necessitate the breaking of all intermolecular associations.

Thermodynamics require that a spontaneous process will be characterized by a negative free-energy change. As such, the thermodynamics of mixing are governed by the free-energy equation of thermodynamics:

$$\Delta G_M = \Delta H_M - T\Delta S_M \qquad (2)$$

Where ΔG_M is the free-energy of mixing, ΔH_M is the enthalpy of mixing, T is the absolute temperature of the system, and ΔS_M is the entropy of mixing. Hildebrand and Scott proposed an equation for the determination of the enthalpy of mixing:

$$\Delta H_M = \varphi_1 \varphi_2 V_T (\delta_1 - \delta_2)^2 \qquad (3)$$

Where φ_1 and φ_2 are the volume fractions of mixture components 1 and 2, δ_1 and δ_2 are the Hildebrand solubility parameters of components 1 and 2, and V_T is the total

volume of the mixture. Equations 2 and 3 illustrate the fundamental usage of solubility parameters. As the Hildebrand parameters of components 1 and 2 approach the same value the enthalpy of mixing will approach zero independent of their respective volume fractions. When the solubility parameters of components 1 and 2 are sufficiently similar the enthalpy of mixing will reach a value where it can no longer dominate the entropy term of equation 2. This results in a negative free-energy of mixing and spontaneous mixing will occur.

One of the fundamental weaknesses of Hildebrand's work is that it is only applicable to "regular solutions." Regular solutions have no "orienting or chemical effects" and all molecules are oriented and distributed randomly throughout (3). Unfortunately most chemical mixtures have dipole or hydrogen-bond orientations. Multi-component solubility parameter systems offer some measure of improvement by accounting for arrangement and orientation effects.

Hansen's widely popular three-component model of solubility parameters treats intermolecular attraction forces independently by giving each attractive force its own solubility parameter. Hansen chose three parameters for his model and gave them the symbols δ_D, δ_P, and δ_H to account for dispersion (induced dipole-induced dipole) forces, polar (dipole-dipole) forces, and hydrogen-bonding forces respectively. They may be related to Hildebrand's "total solubility parameter" by the square root of the sum of the squares of the three Hansen parameters (4).

$$\delta_T = \sqrt{\delta_D^2 + \delta_P^2 + \delta_H^2} \qquad (4)$$

It now may be easy to see in Table 1 how Hildebrand's parameter predicts two substances such as ethylene carbonate and methanol ($\delta_T = 29.6$) to be miscible when they

are not. A glance at the Hansen solubility parameters (HSPs) of each species provides an answer. Although they possess the same total quantity of cohesive energy per unit volume, each species has that energy partitioned into different attractive forces. The fundamental usage of solubility parameters still applies, however, for two species to interact strongly they must now have three sufficiently similar parameters instead of one.

Table 1. A comparison of two immiscible liquids predicted to be miscible by Hildebrand's equation.

Species	δ_D / $MPa^{1/2}$	δ_P / $MPa^{1/2}$	δ_H / $MPa^{1/2}$	δ_T / $MPa^{1/2}$
ethylene carbonate	19.4	21.7	5.1	29.6
methanol	15.1	12.3	22.3	29.6

One may think of this another way. If a species' Hansen solubility parameters were plotted as a small sphere in a Cartesian coordinate space with its center defined as a point located by the three parameters, those species which share a significant volume within that coordinate space should be mutually soluble. Two species may reside at the same distance from the origin (represented by δ_T) yet occupy two separate volumes of space and therefore would not be soluble.

There are other multi-component solubility parameter models with varying strengths and weaknesses. All have achieved some measure of success in academia or the private sector. For example, Karger, Snyder, and Eon have developed a solubility parameter system for chromatographic solvent selection (5). Another system developed by Beerbower, Martin, and Wu provides for a more sophisticated treatment of Hansen's hydrogen-bonding parameter (6,7). Because of this increased ability to handle the complex nature of hydrogen-bonding interactions Beerbower, Martin, and Wu's four-

component system is rapidly gaining acceptance in many fields of study. However, with four components, one loses the ability to visualize the interaction in a three-dimensional Cartesian space.

Previous Methods of Calculating Solubility Parameters

One weakness of this model of solubility is the confusing and tedious methods used to calculate the parameters. These methods attempt to be universal but still do not apply to many cases. This is dependant upon how much information is known about the species of interest. For some species the methods yield wholly inaccurate results. It is this weakness that the work presented in this document aims to resolve.

Perhaps the most straightforward way in which to determine solubility parameters is to use solvents with well determined solubility parameters to experimentally determine the parameters of the unknown. This is akin to using radio tower triangulation to locate a signal between them. Picture the Cartesian coordinate system mentioned earlier. An experiment is performed where ten solvents with well known solubility parameters are used to solvate a species of unknown parameters. The solubility of the unknown species can then be ranked by percent mass composition. The user then determines an operational definition of "good" solubility. The solute is represented as a point in space with a radius of interaction which encompasses the "good" solvents and excludes the "bad" solvents (8). This produces the most reliable parameters as they are dependant upon empirical evidence alone. The obvious drawback to this method is the impracticality and cost of applying it to a large number of species. Solute dimerization can give unusual results on rare occasions. The other obvious drawback is the

fundamental criteria for the experiment: the unknown species must be soluble in something.

A cost-effective estimation of solubility parameters can be achieved via group contribution methods. This method operates on the principle that each functional group or part of a molecule accounts for some piece of the solubility parameters of the whole molecule. The molecule can then be built in pieces and the contributions of the individual pieces can then be summed to give the solubility parameters of the whole. This method was pioneered by van Krevelen and Hoftyzer(9) to build up the solubility parameters of polymers. It was later refined by Stefanis and Panayioutou(10).

The universality of the group contribution method is often a problem. Most functional groups are not isolated from others and as such do not exist in an independent fashion. On the contrary, groups are absolutely affected by their nearest neighbors either electrostatically or through the conjugative effects of molecular orbitals. This has an effect on the overall solubility parameters of the molecule. The other phenomenon which this method of calculation does not address is the orientation of groups. This is important to the polar parameter. If two moieties have separate dipole moments but are oriented in opposite directions then they will destructively interfere. Conversely, if they are oriented in the same direction they will constructively interfere. If the two moieties provide the only significant source of dipole-dipole interaction then this method will likely produce inaccurate results for the polar solubility parameter. The use of planes of symmetry to correct for these effects introduces more complexity for very little benefit.

Of the two problems, electronic and orientation, Stefanis and Panayioutou have successfully addressed the former. They have used conjugation theory to provide a more

sophisticated group contribution table with first order and second order groups. Unfortunately what results is a longer and more confusing process of group contribution, albeit more successful than that of van Krevelen and Hoftyzer for many molecules. The fundamental problem of group contribution methods is that by their very nature they cannot be entirely inclusive. There are an infinite number of ways in which functional groups can be assembled and this results in an infinite number of group contributions.

There are ways of calculating individual solubility parameters. When possible, the dispersion parameter is calculated from the critical temperature and molar volume of an alkane homomorph in a procedure devised by Blanks and Prausnitz (11). The homomorph structure is defined as the non-polar analog of similar size and shape as the species of interest. Unfortunately this method breaks down when it comes to ringed structures and halogenated species. Quite often the polar parameter is calculated from Hansen and Beerbower's simplification of the Böttcher equation (12):

$$\delta_P = 37.4\mu/V_M^{1/2} \qquad (5)$$

Where μ is the dipole moment in Debye and V_M is the molar volume in $cm^{3/}mol$. This requires the user to know or estimate these values with some degree of certainty. When this cannot be done, group contribution methods are usually employed.

When the dispersion and polar solubility parameters can be estimated with confidence the hydrogen-bonding parameter can be calculated from equation 4; this requires the use of energy of vaporization data to determine the total solubility parameter. When either δ_P or δ_D cannot be estimated to solve for δ_H, then the group contribution method outlined by Stefanis and Panayioutou appear to be sufficient to calculate the hydrogen-bonding parameter independently.

Current Project

The funding agency for this project, B&W Pantex in Amarillo, TX is interested in the solubility of nitrated aromatic and nitramine species such as 1,3,5-Trinitro-2-[2-(2,4,6-trinitrophenyl)ethenyl]benzene (HNS) and 1,3,5,7-tetranitro-1,3,5,7-tetrazocane (HMX), Figure 1. HNS is an example of a nitrated aromatic species and HMX is an example of a nitramine. These species exhibit poor solubility in common industrial solvents. The best solvents for many of these applications are N-methylpyrrolidone and N,N-dimethylformamide though they do not perform as well as desired. The Hansen solubility parameters of these explosives and other energetic materials will be explored to determine if these two solvents are the best solvents, or if blended solvents are better for dissolving energetic materials.

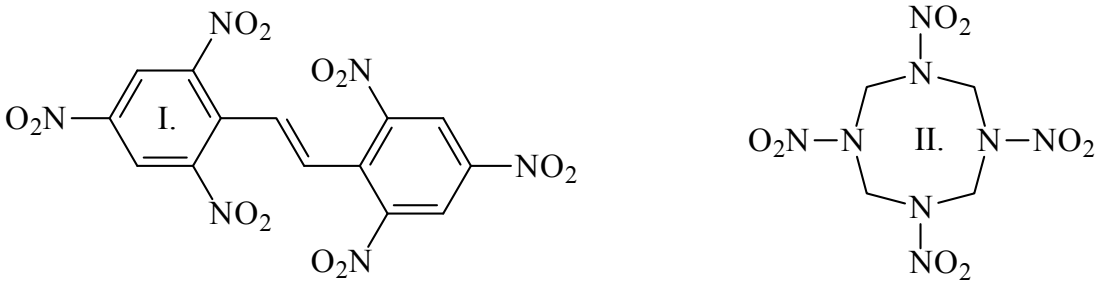

Figure 1. Chemical Structures of HNS (I) and HMX (II).

There are significant problems with the estimation of the HSPs of nitrated aromatics and nitramines. Primarily, they have poor solubility in most common industrial solvents. This precludes a wet chemistry solubility study. Secondly, there are no entries in group contribution methods for aromatic nitro-groups or nitramines. This precludes all group contribution methods for the estimation of HSPs. The Yamamoto

group contribution method present in HSPiP contains nitro groups, but gives unreasonable results, and is still under development.

One could approach this problem with a brute force technique and try various combinations of two, three, and four component solvent blends to determine which matches the HSPs of the unknown. The number of combinations of two, three, and four component blends of a possible ten solvents is 375. If each combination is tested at three different volume fractions, the number of experiments top one thousand. If the number of possible solvents is expanded by just one more to eleven, then the number of experiments would top two thousand. In this case, time and cost preclude the use of this method for the determination of HSPs.

Problems of this size and complexity often necessitate the use of computers. It is the aim of this research to simplify and accelerate the determination of HSPs through the use of electronic structure calculations. These calculations are cheap, relatively accurate, and if employed with modern parallel processing computers they can be fast. Structure variables will be extracted from calculation result files and used to develop a quantitative structure activity relationship (QSAR). Electronic structure calculations will provide the structure variables and a matrix of coefficients will then relate them to HSPs (the desired activity). The result will be an inexpensive and unambiguous method for calculating the HSPs of nitrated aromatics and nitramines.

Theory

As recent as the 1960s, chemists have begun to use computers to aid them in the modeling of molecular structure (13). There have been many developments since then and many programs now are capable of the accurate modeling of bond lengths, molecular

orbitals, and thermodynamic properties. These computational chemistry models allow scientists to predict the bulk properties of a species from the calculation of a single molecule.

There are significant advantages to this. One advantage is cost. If a scientist can research a species or a group of species without ever having to synthesize them, then the cost of conducting that research is tremendously reduced. Aside from the obvious cost of consumable materials required to synthesize a species there are many less tangible costs. For example, the user's time spent on synthesis, purification, characterization, and evaluation. Reduced instrument maintenance costs as well as less exposure to laboratory chemicals are also seen as benefits.

The least computationally expensive (fastest) model chemistry is molecular mechanics which uses pair potentials between all atoms in the system. These potentials have been parameterized to match the most common types of molecules. Unfortunately the highly nitrated species of interest are not well targeted by the molecular mechanics parameter set.

Electronic structure calculations seek to account for the electrons' contribution to the chemistry of a species. Because the exact location and velocity of an electron cannot be known, the numerical Schrödinger equation cannot be solved explicitly. The pieces which cannot be solved must be approximated by various levels of electron-correlation theory. Most structure calculations require a designation of level of theory and basis set of wavefunctions. Some popular levels of theory include HF (Hartree-Fock) and B3LYP (Becke type, 3-parameter exchange with Lee-Yang-Parr correlation functional) (14). The basis set defines the number and type of functions that will be used to approximate the

wavefunction of each electron in the species. A typical basis set is 6-31G(d). This nomenclature indicates that six functions will be used to approximate the wavefunction of the core electrons, three functions will be used for the valence electrons of heavy atoms (all atoms which are not hydrogen), and one function will be used for hydrogens. The "G" indicates that the type of functions to be used will be Gaussian functions; and the "(d)" indicates that d-orbital polarization functions will be added to heavy atoms.

Typically the input geometry is not the lowest energy conformation and so the molecule will require optimization. Any number of subsequent calculations can be performed on the resulting optimized geometry. The most common calculation is a "frequency calculation." Many of the electronic structure parameters are calculated in this step, the most prominent of which are the vibrational frequencies, and infrared(IR) and Raman spectral intensities. The publication syntax for the specification of a geometry optimization with Hartree-Fock theory and a 3-21G(d) basis set followed by a frequency calculation of higher level of theory and basis set would be: B3LYP/6-31G(d,p)//HF/3-21G(d). The B3LYP/6-31G(d) model chemistry was chosen for this project because of its high accuracy and its low computational cost. The published mean average deviation for this method on a standard set of molecules is 7.9 kcal/mol with a standard deviation of 9.5 kcal/mol (14).

The utility of electronic structure calculations to approximate bulk properties via a structure activity relationship has been demonstrated (15). This technique employs the use of structure variables to predict an unknown activity. Each variable is assigned a coefficient which indicates its weight or effect toward the total activity. For a quantitative structure activity relationship (QSAR) to be established the relationship

matrix must be trained to a set of species for which the activity and structure variables are well defined. This will yield a matrix which can be applied to the prediction of activity for species which are sufficiently similar to the training set. This relationship "training" can be accomplished by a step-wise regression technique.

The step-wise regression is a methodical way for including or excluding structure variables from the relationship matrix. The driving force for the inclusion or exclusion of variables is their correlation with activity. Correlation is quantified in the form of an R-squared value. The criteria for the decision to include or exclude a particular variable is the probability that random noise could produce the coefficient of that variable in the structure-activity transformation matrix. This criteria is quantified by a P-test which results in the appropriately named P-value. A variable is chosen and a regression is performed, producing a transformation coefficient. There will be some level of correlation with this first structure variable. Another variable is included and another regression is performed. The coefficients of the two variables and their P-values are compared to the exclusion criteria. If a third variable is inserted and causes the P-value of another variable coefficient already in the model to exceed the exclusion criteria and also causes an increase in the R-squared value, then the third variable is kept and the criteria-exceeding variable is ejected from the model. This process is repeated until all variables in the model meet the inclusion criteria.

In the case of this research each Hansen solubility parameter will have a relationship matrix and thus will require a step-wise regression. This means that for every model there will be three step-wise regressions on *n* number of variables which will yield three $1 \times n$ matrices ($[D]_n$, $[P]_n$, and $[H]_n$) or one $3 \times n$ matrix ($[D\ P\ H]_n$). These

matrices can be simplified by eliminating those variables which do not have a statistical contribution to the activity (coefficient = 0) to yield matrices with numbers of rows less than n.

CHAPTER II

EXPERIMENTAL

Species which were used for training QSAR matrices 1 and 2 were selected from a single reference (16). The species in this reference were originally the only source of solubility parameters the author had access to. As the author acquired references with more comprehensive tables of solubility parameters (4, 20) QSAR3 was able to more effectively target the nitrated functionality.

Computational Chemistry

All electronic structure calculations were made using the Gaussian® '03W (17) and GaussView computational package. Calculations were run on either one of two systems: 1) a Dell™ Optiplex745 desktop PC running Microsoft Windows XP™ with a dual-core central processing unit speed of 3.4 GHz and 2 Gb of random access memory, or 2) a Dell™ Precision690 running Linux with two quad-core processors at a speed of 2.66 GHz and 32 Gb of memory. A parallel processing version of the Gaussian software package was used on the eight core machine so that as many as eight calculations could be performed simultaneously. This was also helpful in assigning multiple processors to one job to increase the computational power for difficult or long calculations. A complete list of the computational times for all structures in this work are located in Appendix A, pg 44.

Each molecule's geometry optimization and subsequent frequency calculation was made using the B3LYP hybrid density functional (18, 19) and a 6-31G(d) basis set. The majority of the calculations were performed by the author however acknowledgement is given to Derek Blaylock and Katie Rothlisberger who did perform

some calculations under the direction of the author and Dr. Darren Williams. All calculations were optimized to a minimum with no observed negative frequencies.

Structure variables were then extracted from the result files. The result file was opened and checked for proper optimization by verifying the absence of negative harmonic frequencies. The dipole moment μ, electronic spatial extent e_{se}, Δ charge d, polarizability α, and molar volume V_m were extracted for use in the first structure activity relationship. The dipole moment is reported explicitly in the Gaussian® result file in units of Debye. The electronic spatial extent is also reported explicitly in the Gaussian® result file and has units of $Å^2$. The Δ charge value is calculated from the difference in Mulliken charge of the most negative heteroatom (not carbon or hydrogen) and the most positive hydrogen atom. For species in which there are no heteroatoms or in which the heteroatoms were all positive the Mulliken charge for the most negative heteroatom was assigned the value of zero. The charge of the most positive hydrogen atom was assigned a value of zero for species in which there were no hydrogen atoms present. The polarizability was calculated as the square root of the sum of the squares of the six exact polarizability tensors reported in the result file and has units of $Å^3$.

The molar volume is calculated in a rather roundabout way. The reason for this is that the electronic structure calculation does a poor job of estimating the molar volume. Hansen solubility parameters are typically reported along with a molar volume so one could look the experimental value up if one so chooses, however, it is the goal of this project to provide all the information necessary to calculate the HSPs of a nitrated species from a computational result file. To calculate the molar volume a regression was

performed on all of the training set species' experimental molar volumes with respect to the molecular formula which is an explicit output of the Gaussian® result file.

For the second and third structure activity relationships the energy of the highest occupied molecular orbital (HOMO) I, was extracted from the result files as well. This was done to approximate the first ionization energy of the species which was was reported to have some correlation with the dispersion properties (20).

Structure Activity Relationship Training

Once structure variables have been calculated the relationship matrix must be optimized or "trained" to the activities of the calculated species. These species are termed the "training set." The activities of interest are the three Hansen solubility parameters. To accomplish this, a step-wise regression technique was employed. The training set structure variables and activities were imported to Minitab® 15 (21) and the program's step-wise regression application was used. The step-wise regression of each HSP was performed individually so that for a given training set there were three regression matrices, one for each HSP. The criteria for inclusion and exclusion was set at a p-value of 0.2. This is analogous to a minimum %80 confidence that the coefficient is not due to random noise.

The size of the training set as well as the structure variables selected as potential predictors are what set the various QSAR models apart. The first model contains 183 chemical species comprising a wide variety of functionality (Appendix B, pg 54). Structure variable predictors for this model were: dipole moment μ, electronic spatial extent e_{se}, Δ charge d, polarizability α, and molar volume V_m. The matrix resulting from this first model is termed "QSAR1."

The second model which is termed "QSAR2" shares the same 183 chemical species as QSAR1 but with a larger structure variable set. In this model the set of structure variables is expanded by using mathematical transforms of the original five structure variables and the energy of the HOMO, *I*. The square, square-root, inverse, natural logarithm, and exponential of each variable was included in the model with the exception of electronic spatial extent. In addition to these simple mathematical transforms, some of the original structure variables were combined by multiplication/division. These combinations include a $\mu\,(V_m)^{-0.5}$ term, a $d\alpha$ term, an $I\alpha$ term, and an $(I\alpha)^{-1}$ term. These combination terms were included after studying equations put forth in Barton's Handbook of Solubility Parameters (20) as well as Hansen's own handbook (4).

The third model, "QSAR3" has a shortened list of chemical species in the training set. Unfortunately the accepted values for HSPs cannot be regarded as canon due to the many ways in which they can be calculated. The list of 183 chemical species was narrowed down to 39 species which were believed to be calculated from experiment based on their cost and ubiquity. Added to this list, however, are 15 more species which have similarity to the functionality depicted in Figure 1. This was done in an effort to further enhance the predictive power of the third model toward nitrated species. Because this new functionality was included, a new regression to calculate molar volume was also performed (Table 3, and Figure 3). A complete list of the training set species for QSAR3 can be found in Appendix C, page 62. In addition to the structure variables used in QSAR2, stoichiometry was also included. It was postulated that if activities such as

molar volume could be approximated using the molecular formula then they might also be useful for the approximation of HSPs.

Calculation of Previously Unknown HSPs

Finally, the models were put to use by calculating the HSPs for various energetic materials. Each species was calculated by Gaussian '03W at the same level of theory and basis set as the training set species. The appropriate structure variables were extracted from the result files and were operated on by the three QSAR matricies. Structure variables for the nitrated unknowns are tabulated in Appendix D, page65. The result is three models of HSPs for these energetic materials. The strengths and weaknesses of these models will be discussed later.

CHAPTER III

RESULTS

Molar Volumes

Molar volumes were calculated using the method described in Chapter 2. QSAR1 and QSAR2 were calculated using the molar volume regression shown in Table 2 and Figure 2. Figures 2 and 3 show an equation for the correlation trendline. This is not to indicate how the correlation applies to the variable "x" but rather to show the slope of the trendline. As the correlation improves the slope will approach a value of one.

Table 2. Regression coefficients of stoichiometry modeling molar volume (N=183)

Variable	Coefficient	Standard Error
Intercept	15.51199403	1.656059157
#C	9.023012684	0.32021995
#H	3.807253279	0.178497756
#N	1.36510039	1.139692896
#O	5.16678748	0.729221579
#S	15.97578012	3.555444864
#P	14.78730787	8.866345126
#F	14.34459346	1.588282536
#Cl	16.50292516	1.016481139
#Br	19.0710091	1.768516926

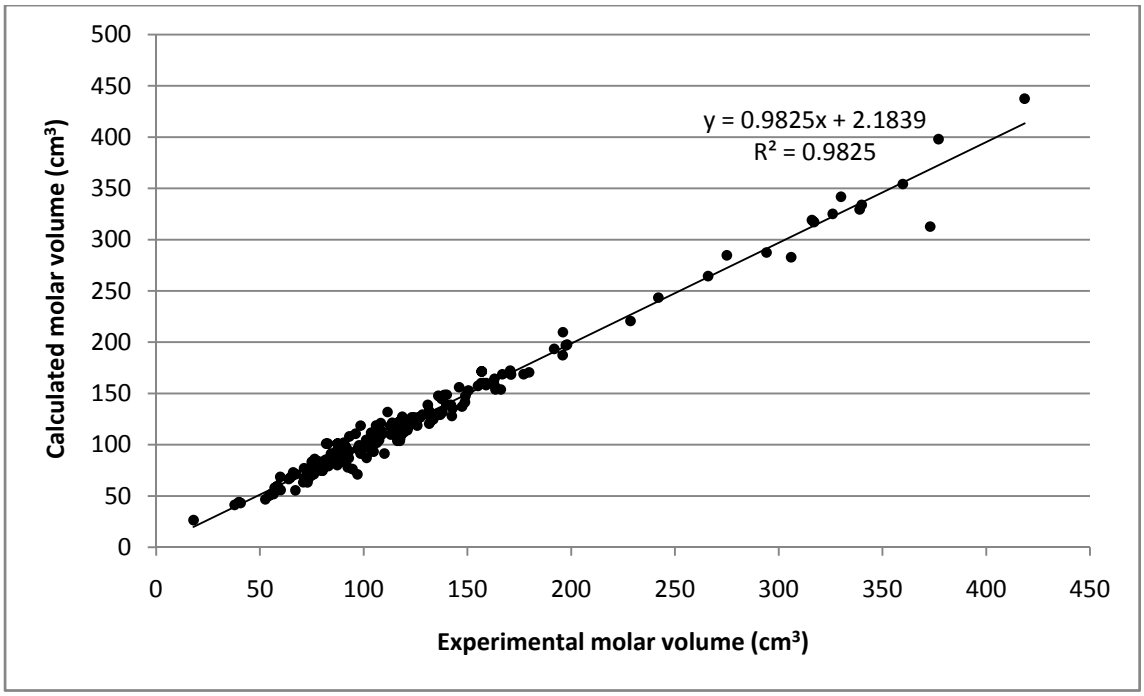

Figure 2. Correlation chart of regression calculated molar volume to experimental molar volume. (N=183)

The molar volumes for QSAR3 were calculated using a different regression because the number and identity of species in the training set changed. This training set includes some species which are also in the training set of QSAR1 and QSAR2 as well as several new nitrated species to enhance the predictive power of QSAR 3 toward the target nitrated species. Results of this regression are shown in Table 3 and Figure 3. Note that in this training set there were no phosphorous or fluorine containing species.

Table 3. Regression coefficients of stoichiometry modeling molar volume (N=54)

Variable	Coefficient	Standard Error
Intercept	9.84325433	4.236976949
#C	6.852084724	0.56839236
#H	4.910317572	0.569191206
#N	7.217614235	1.773869329
#O	6.484731548	0.926807038
#S	19.68305223	3.801321124
#P	0	0
#F	0	0
#Cl	20.2641589	1.721525705
#Br	27.96045898	5.510054368

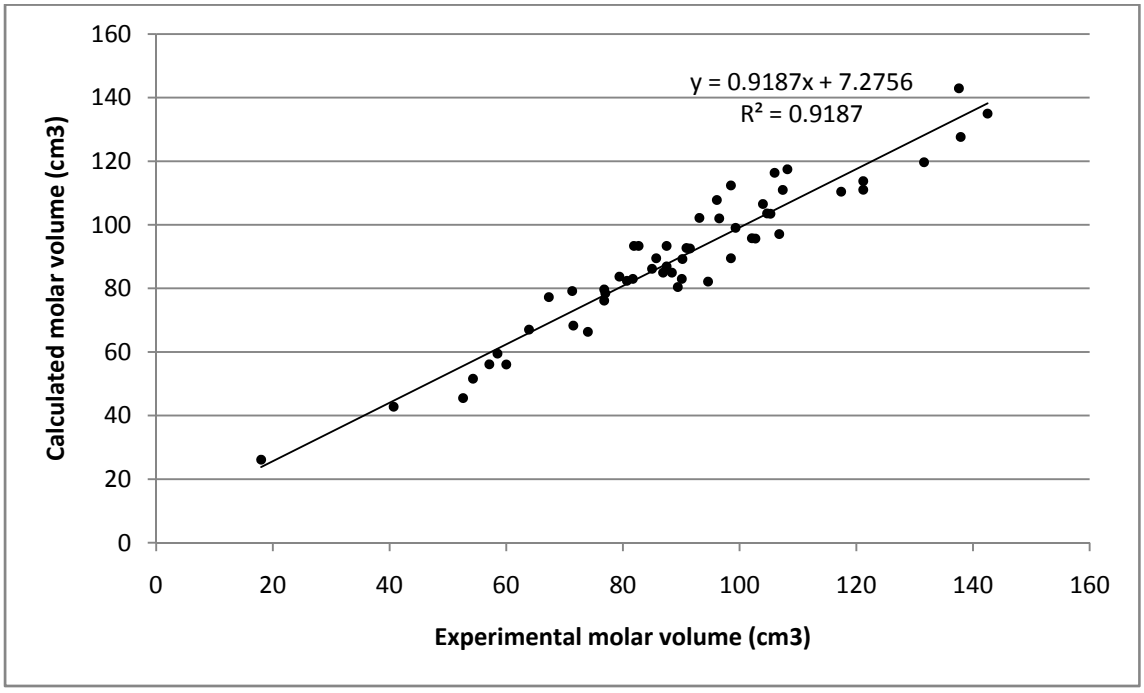

Figure 3. Correlation chart of regression calculated molar volume to experimental molar volume. (N=54)

QSAR Matrices

Once the QSAR matrices were generated they were then used to calculate the HSPs of the training set to look for predictive trends. Figures 4 through 6 depict the transformation matrices and how they are used to generate HSPs.

$$[1 \quad V_m \quad \mu \quad e_{se} \quad d \quad \alpha] \cdot \begin{bmatrix} 18.07 & 3.8558 & 6.148 \\ -0.073 & -0.0513 & -0.0677 \\ 0.111 & 2.51 & 0.0 \\ -.00006 & 0.00008 & 0.00013 \\ -0.68 & 1.86 & 11.85 \\ 0.0632 & 0.0233 & 0.0255 \end{bmatrix} = [\delta_D \quad \delta_P \quad \delta_H]$$

Figure 4. QSAR1 transformation matrix

$$\begin{bmatrix} 1 & V_m & \frac{\mu}{\sqrt{V_m}} & \mu^{0.5} & e^{\mu} & d & d^{0.5} & d^2 & \dots \\ \dots & e^d & d\alpha & \alpha^{0.5} & \alpha^2 & Ln(\alpha) & I\alpha & I\alpha^{-1} & I^2 \end{bmatrix} \cdot \begin{bmatrix} 12.39 & 0.2602 & -0.452 \\ -0.0712 & -0.034 & -0.0283 \\ 0 & 37.4 & 0 \\ 0 & -5 & 0 \\ 0.0172 & 0 & 0 \\ -13.2 & -6.5 & 0 \\ 3.7 & 12.7 & 18.9 \\ 0 & 0 & 29 \\ 5.3 & 0 & -18.69 \\ 0 & 0 & -0.0202 \\ 0.53 & 0 & 0 \\ 0 & 0 & .00003 \\ 0 & 0 & 3.83 \\ -0.0155 & -0.132 & 0 \\ 0 & -12.8 & -91 \\ -47.2 & 0 & 0 \end{bmatrix} = [\delta_D \quad \delta_P \quad \delta_H]$$

Figure 5. QSAR2 transformation matrix. Here the ellipses do not indicate a sparse matrix. They indicate the continuation of the 1 by 16 input matrix.

$$\begin{bmatrix} 1 & \alpha^{0.5} & \mu^2 & \#N & \#H & I^{-1} & \mu^{-1} & \dots \\ \dots & \frac{\mu}{\sqrt{V_m}} & \#O & \mu^{0.5} & d^{0.5} & \#Br & \#Cl & \alpha^{-1} & d \end{bmatrix} \cdot \begin{bmatrix} 8.995 & -1.304 & 0.5078 \\ 0.596 & 0 & 0 \\ 0.081 & 0 & 0 \\ -0.6 & 0 & 0 \\ -0.31 & 0 & 0 \\ -1.09 & -1.45 & 0 \\ 0.208 & 0 & 0 \\ 0 & 59.8 & -7 \\ 0 & 1.55 & 0 \\ 0 & -10.9 & 0 \\ 0 & 3.3 & 0 \\ 0 & 4.4 & 0 \\ 0 & 1.3 & 0 \\ 0 & 0 & 252 \\ 0 & 0 & 13.41 \end{bmatrix} = [\delta_D \quad \delta_P \quad \delta_H]$$

Figure 6. QSAR3 transformation matrix. Here the ellipses do not indicate a sparse matrix. They indicate the continuation of the 1 by 15 input matrix.

QSAR Correlation with Literature Values

Figures 7 through 15 illustrate the correlation of the various transformation matrices to the literature values used in each training set. The R-squared values shown are not a direct measure of the predictive power of the model. They only indicate the consistency of the model with the training set values. The equation of the trendline is included to show the slope of the trendline and not to indicate any correlation to some variable "x." Data used to generate Figures 7 through 15 are located in Appendices E and F on pages 67 and 74 respectively.

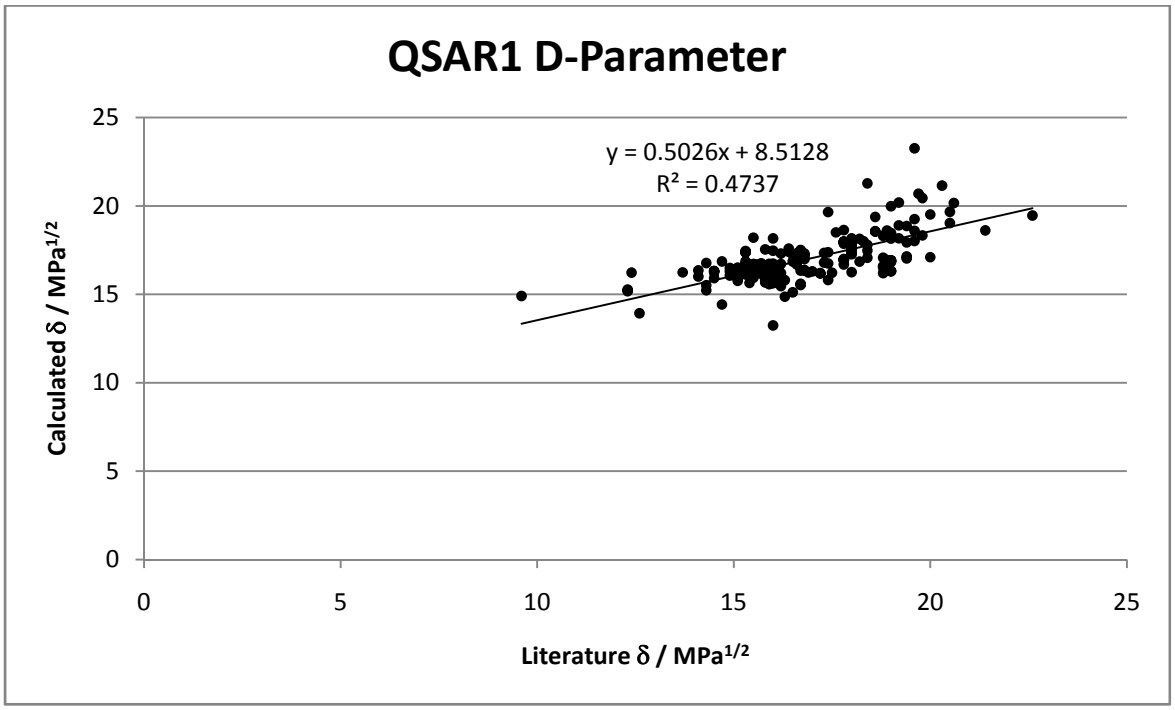

Figure 7. Correlation of QSAR1 predicted δ_D parameters of training set species to literature values.

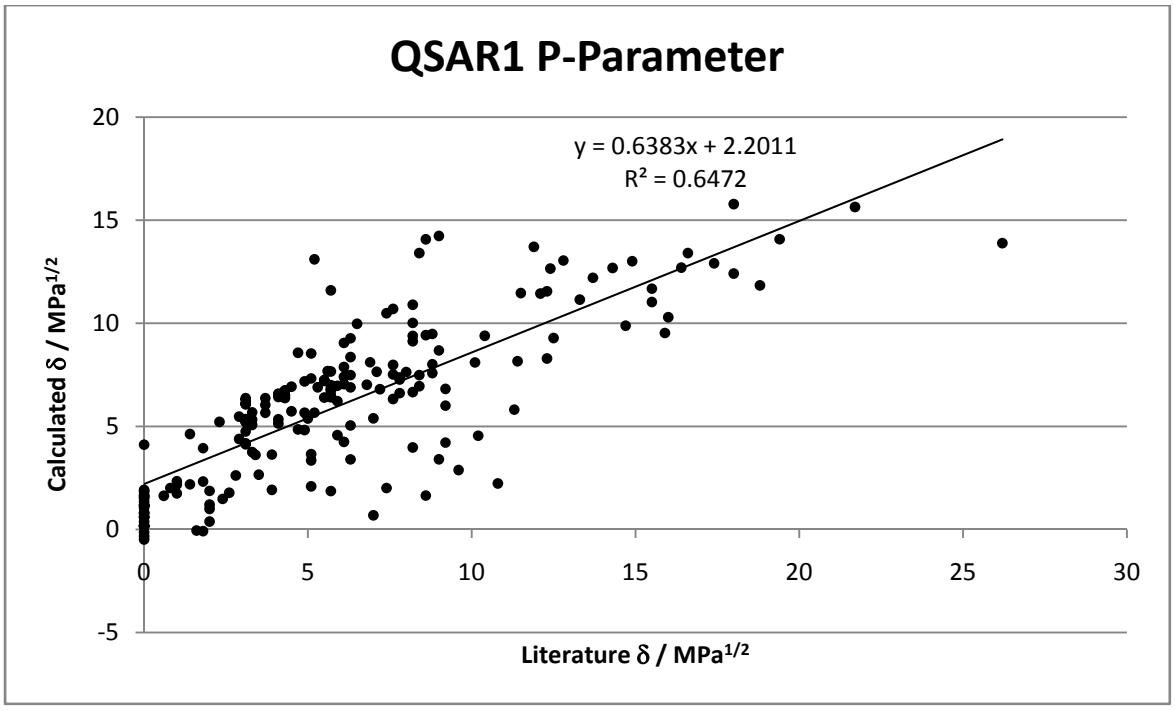

Figure 8. Correlation of QSAR1 predicted δ_P parameters of training set species to literature values.

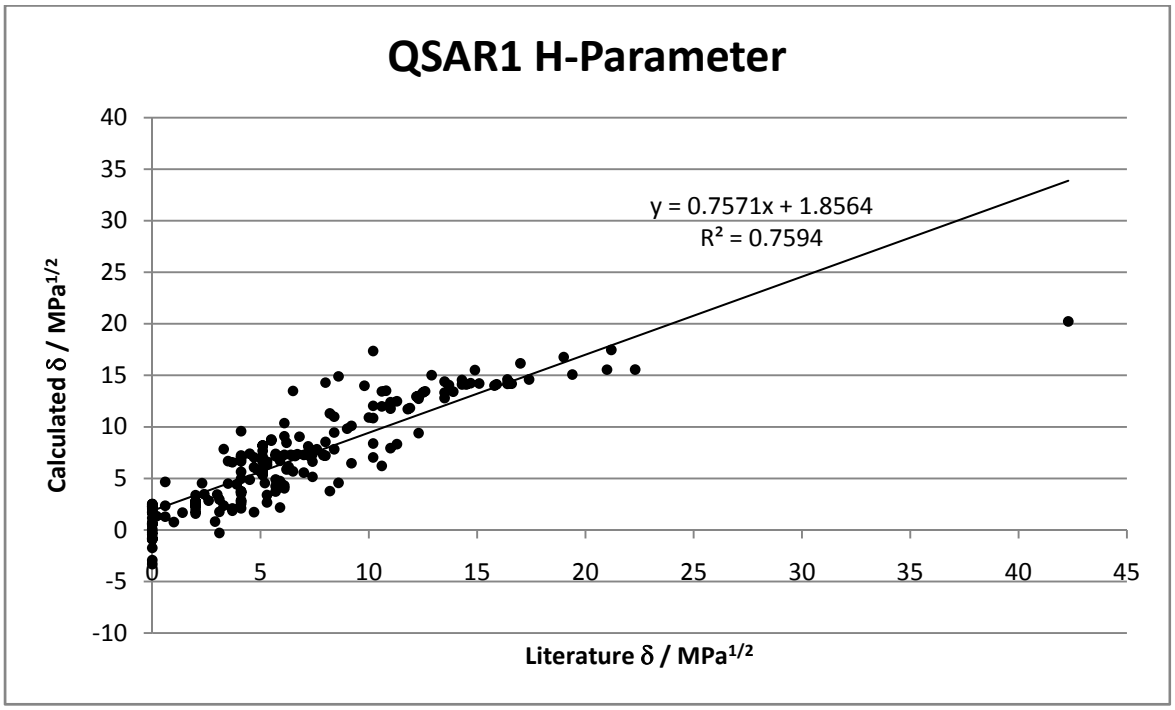

Figure 9. Correlation of QSAR1 predicted δ_H parameters of training set species to literature values.

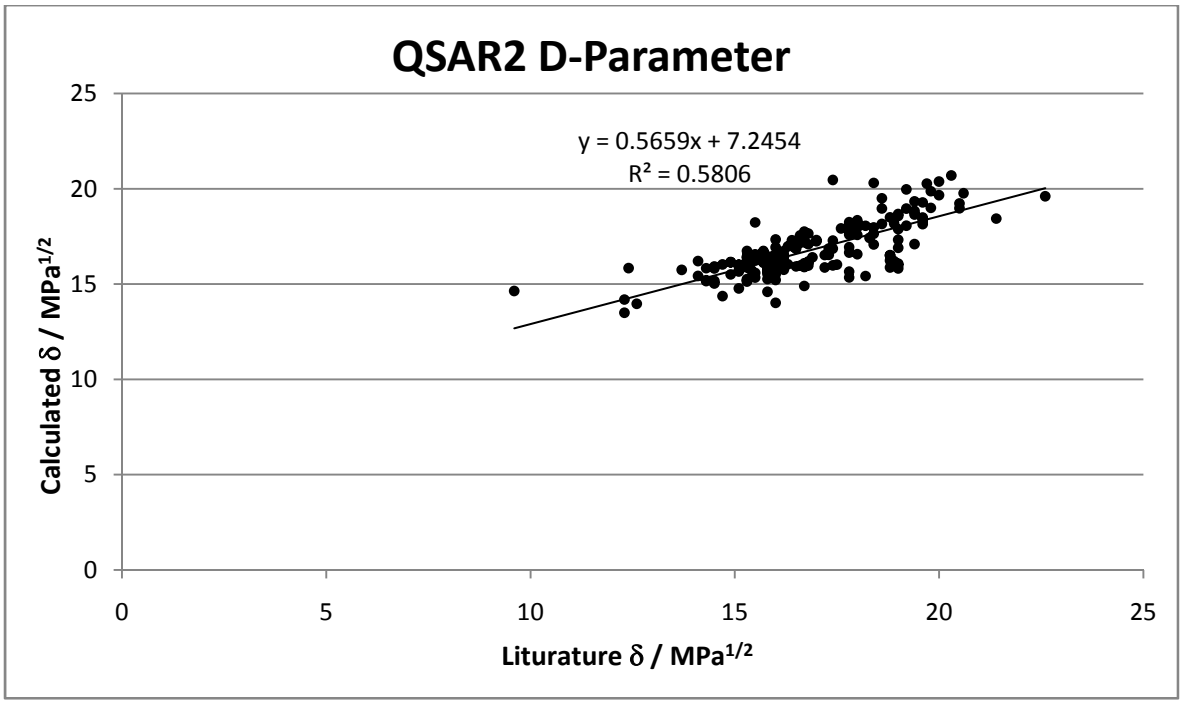

Figure 10. Correlation of QSAR2 predicted δ_D parameters of training set species to literature values.

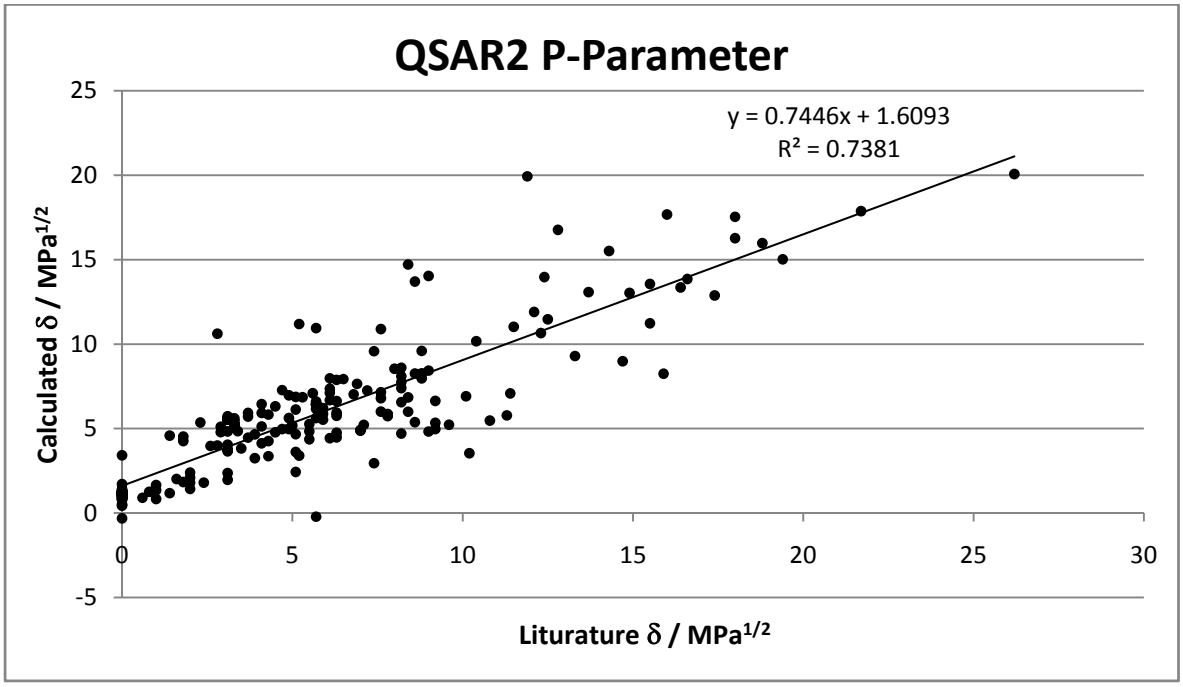

Figure 11. Correlation of QSAR2 predicted δ_P parameters of training set species to literature values.

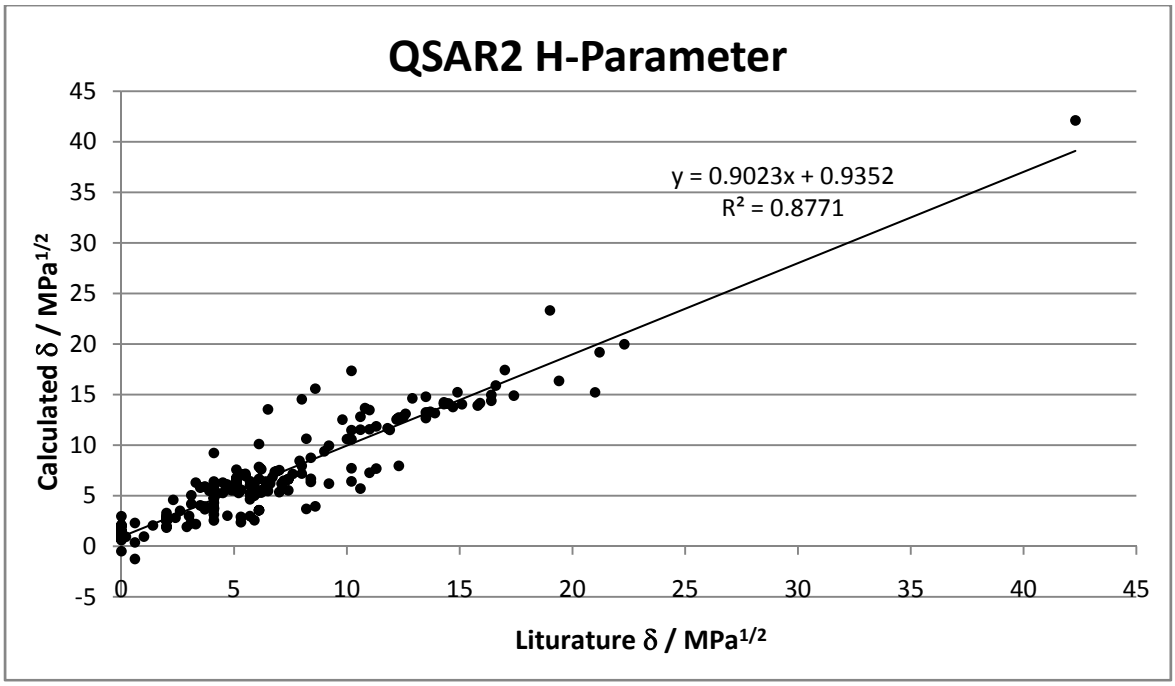

Figure 12. Correlation of QSAR2 predicted δ_H parameters of training set species to literature values.

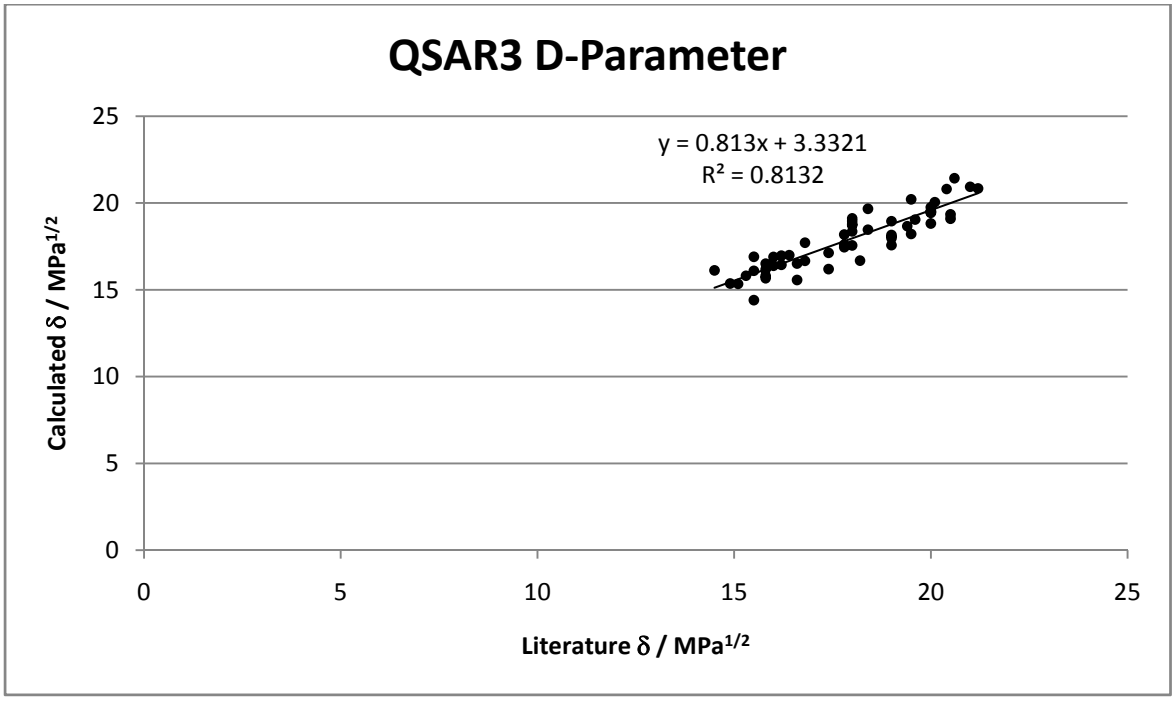

Figure 13. Correlation of QSAR3 predicted δ_D parameters of training set species to literature values.

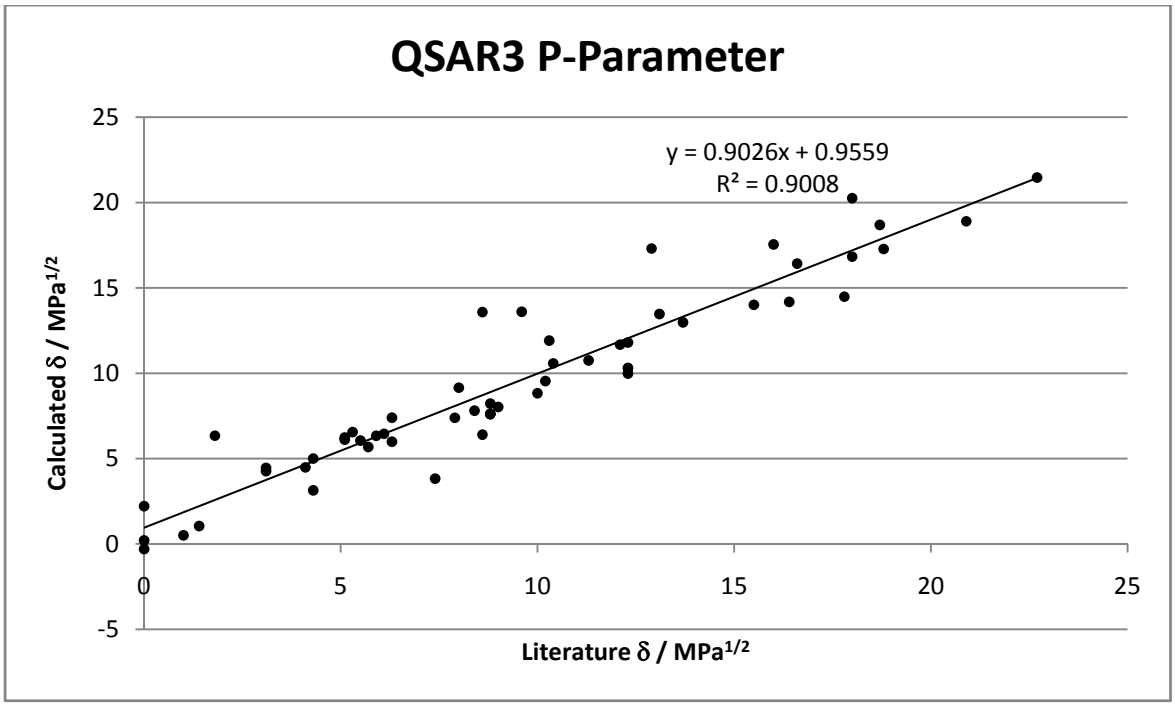

Figure 14. Correlation of QSAR3 predicted δ_P parameters of training set species to literature values.

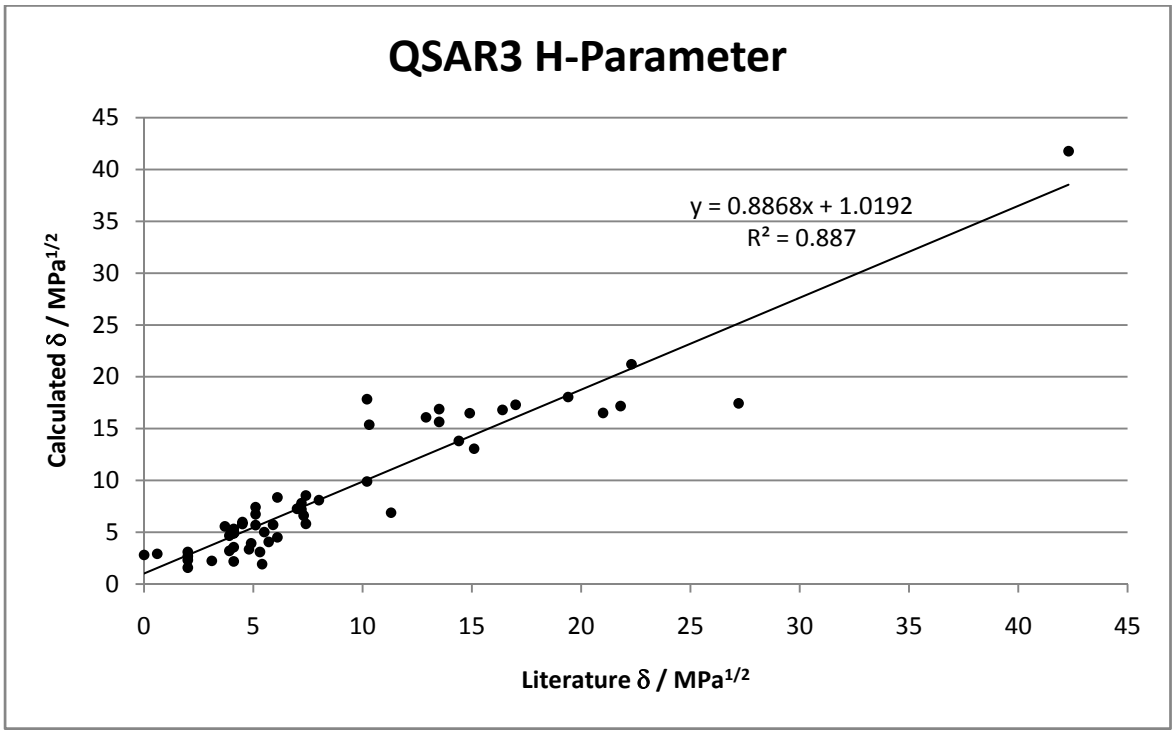

Figure 15. Correlation of QSAR3 predicted δ_H parameters of training set species to literature values.

Predicted HSPs of Nitrated Unknowns

Tables 4 through 6 display the predicted Hansen solubility parameters of the nitrated species of interest.

Table 4. HSPs of Explosive Unknowns Predicted by QSAR1.

Explosive Unknowns		Calculated HSPs		
Abbreviated Name	IUPAC Nomenclature	δ_D	δ_P	δ_H
HNS	(*E*)1,3,5-Trinitro-2-[2-(2,4,6-trinitrophenyl)ethenyl]benzene	30.2	7.3	8.7
HNAB	2,2',4,4',6,6'-hexanitroazobenzene	32.0	6.1	10.5
TACOT	tetranitrodibenzo-1,3a,4,6a-tetrazapentalene	27.7	16.3	8.1
TNC	1,3,6,8-tetranitrocarbazole	28.3	8.1	22.4
tetryl	2,4,6-trinitrophenylmethylnitramine	23.1	11.4	8.3
ethyltetryl	2,4,6-trinitrophenylethylnitramine	22.9	14.0	8.0
picric acid	2,4,6-trinitrophenol	21.6	9.2	16.6
picramide	1,3,5-trinitroaniline	22.7	13.7	21.5
TATB	1,3,5-triamino-2,4,6-trinitrobenzene	24.2	6.9	22.0
DATB	1,3-diamino-2,4,6-trinitrobenzene	23.9	12.8	21.8
TNB	1,3,5-trinitrobenzene	21.4	3.7	8.0
TNCB	2,4,6-trinitrochlorobenzene	21.6	3.8	7.6
TNR	3-hydroxyl-2,4,6-trinitrophenol	21.9	9.1	17.1
TNBA	2,4,6-trinitrobenzoic acid	21.4	7.1	14.3
TNA	2,4,6-trinitroanisole	21.7	9.4	9.4
TNX	3-methyl-2,4,6-trinitrotoluene	21.3	6.7	6.7
PETN	3-nitrooxy-2,2-bis(nitrooxymethyl)propyl nitrate	19.9	3.6	6.3
HMX	1,3,5,7-tetranitro-1,3,5,7-tetrazocane	21.7	12.6	8.2
RDX	1,3,5-trinitro-1,3,5-triazinane	20.4	10.6	8.8
phloroglucinol	benzene-1,3,5-triol	17.8	10.3	15.5
TNPG	2,4,6-trinitrobenzene-1,3,5-triol	22.3	11.3	17.1
TETNB	2,4,6-trinitro-1,3,5-triethoxybenzene	21.5	4.9	6.5
ethylpicrate	2,4,6-trinitroethoxybenzene	21.8	10.1	9.3

Table 5. HSPs of Explosive Unknowns Predicted by QSAR2.

Explosive Unknowns		Calculated HSPs		
Abbreviated Name	IUPAC Nomenclature	δ_D	δ_P	δ_H
HNS	(*E*)1,3,5-Trinitro-2-[2-(2,4,6-trinitrophenyl)ethenyl]benzene	30.2	13.8	9.0
HNAB	2,2',4,4',6,6'-hexanitroazobenzene	30.9	15.2	9.7
TACOT	tetranitrodibenzo-1,3a,4,6a-tetrazapentalene	29.3	14.7	7.5
TNC	1,3,6,8-tetranitrocarbazole	32.2	13.0	15.1
tetryl	2,4,6-trinitrophenylmethylnitramine	23.2	11.3	6.5
ethyltetryl	2,4,6-trinitrophenylethylnitramine	23.8	12.6	6.5
picric acid	2,4,6-trinitrophenol	22.3	10.0	13.4
picramide	1,3,5-trinitroaniline	26.4	11.6	17.3
TATB	1,3,5-triamino-2,4,6-trinitrobenzene	27.7	9.6	16.9
DATB	1,3-diamino-2,4,6-trinitrobenzene	27.2	10.8	17.1
TNB	1,3,5-trinitrobenzene	21.3	8.8	5.9
TNCB	2,4,6-trinitrochlorobenzene	21.7	8.5	5.9
TNR	3-hydroxyl-2,4,6-trinitrophenol	22.7	9.7	13.8
TNBA	2,4,6-trinitrobenzoic acid	21.6	9.4	11.2
TNA	2,4,6-trinitroanisole	21.5	10.3	7.1
TNX	3-methyl-2,4,6-trinitrotoluene	21.3	8.3	5.7
PETN	3-nitrooxy-2,2-bis(nitrooxymethyl)propyl nitrate	20.3	8.0	5.2
HMX	1,3,5,7-tetranitro-1,3,5,7-tetrazocane	22.1	11.0	6.2
RDX	1,3,5-trinitro-1,3,5-triazinane	20.0	9.7	6.5
phloroglucinol	benzene-1,3,5-triol	18.4	8.7	15.0
TNPG	2,4,6-trinitrobenzene-1,3,5-triol	23.1	10.7	13.7
TETNB	2,4,6-trinitro-1,3,5-triethoxybenzene	20.8	7.7	6.3
ethylpicrate	2,4,6-trinitroethoxybenzene	21.8	10.8	7.1

Table 6. HSPs of Explosive Unknowns Predicted by QSAR3.

Explosive Unknowns		Calculated HSPs		
Abbreviated Name	IUPAC Nomenclature	δ_D	δ_P	δ_H
HNS	(*E*)1,3,5-Trinitro-2-[2-(2,4,6-trinitrophenyl)ethenyl]benzene	20.7	17.1	5.6
HNAB	2,2',4,4',6,6'-hexanitroazobenzene	23.0	22.0	6.5
TACOT	tetranitrodibenzo-1,3a,4,6a-tetrazapentalene	20.3	13.4	3.4
TNC	1,3,6,8-tetranitrocarbazole	20.9	14.8	21.0
tetryl	2,4,6-trinitrophenylmethylnitramine	18.5	13.2	5.6
ethyltetryl	2,4,6-trinitrophenylethylnitramine	18.9	14.6	5.4
picric acid	2,4,6-trinitrophenol	18.9	12.5	15.4
picramide	1,3,5-trinitroaniline	19.2	13.8	19.4
TATB	1,3,5-triamino-2,4,6-trinitrobenzene	18.5	13.4	21.1
DATB	1,3-diamino-2,4,6-trinitrobenzene	18.8	12.6	19.7
TNB	1,3,5-trinitrobenzene	19.1	10.8	6.7
TNCB	2,4,6-trinitrochlorobenzene	19.5	11.4	6.5
TNR	3-hydroxyl-2,4,6-trinitrophenol	19.3	14.0	15.9
TNBA	2,4,6-trinitrobenzoic acid	19.0	13.0	13.7
TNA	2,4,6-trinitroanisole	18.8	11.7	7.6
TNX	3-methyl-2,4,6-trinitrotoluene	18.1	8.7	5.7
PETN	3-nitrooxy-2,2-bis(nitrooxymethyl)propyl nitrate	16.8	17.8	6.4
HMX	1,3,5,7-tetranitro-1,3,5,7-tetrazocane	15.3	13.4	5.6
RDX	1,3,5-trinitro-1,3,5-triazinane	15.5	11.1	6.5
phloroglucinol	benzene-1,3,5-triol	19.7	11.9	15.5
TNPG	2,4,6-trinitrobenzene-1,3,5-triol	19.9	16.6	15.4
TETNB	2,4,6-trinitro-1,3,5-triethoxybenzene	17.8	12.9	8.2
ethylpicrate	2,4,6-trinitroethoxybenzene	18.7	11.6	7.8

CHAPTER IV

DISCUSSION

QSAR1

As one might expect, the dipole moment is significant for the determination of the polar solubility parameter. The Δ charge variable was also significant for this parameter. The Δ charge variable appears to also be significant for determining the hydrogen-bonding parameter, see Figure 4.

Correlation of the predicted dispersion parameters to the dispersion parameters in the training set appears to be exceptionally poor (R^2 = 0.4737), however this is an effect of the predicted dispersion parameter's variability over a relatively small range. Nearly all tabulated dispersion parameters exist somewhere between 15 and 30 $MPa^{1/2}$. Most of the dispersion parameters in the training sets of this work exist between 15 and 20 $MPa^{1/2}$. The variability of the predicted values when compared to the training set values over this small range exhibit the small R-squared value seen in Figure 7. In actuality it appears as though most of the dispersion values between 15 and 20 $MPa^{1/2}$ are predicted to within 2 $MPa^{1/2}$. This is good considering that it has been shown that the very literature values the predictions are based on and compared to are possibly erroneous (REF). The only major outliers appear to be species with literature values of approximately 10 and 13 $MPa^{1/2}$ which are predicted to be around 15 $MPa^{1/2}$.

This model's prediction of the polar solubility parameter is also poor. In this case the variability is over a wider range so the previous explanation of poor correlation does not apply. Here the prediction is actually the problem. As evidenced in Figure 8, there are few species which are correctly predicted by the model (R^2=0.6472). Many are

predicted between 3 and 8 $MPa^{1/2}$ off of their target value. There appears to be no trend which can be noted in this figure. The model does not predict one sub-set of chemical species better than any other, nor does it err in any consistent fashion. The reason for this is likely due to the amount of variability in the dipole moment when calculated with the B3LYP/6-31g(d) model chemistry. A species can be optimized three times at this model chemistry to the same geometry and the resulting dipole moment calculation can vary by as much as a Debye or more. A variation of 1 Debye in the QSAR1 model will result in a 2.51 $MPa^{1/2}$ change in δ_P. For this reason, any calculation which is based on these dipole moments can be compromised. The solution for this is to calculate all species at a more computationally expensive model chemistry which will remove some of this variability. Because of the high computational cost, this change will likely have to be addressed in a later work.

Hydrogen-bonding parameter predictions from QSAR1 show considerable promise (R^2=0.7594). Many of the species with literature values between 0 and 20 $MPa^{1/2}$ are modeled to within 3 $MPa^{1/2}$. The correlation drops off above 20 $MPa^{1/2}$; the notable outlier is water (d_H = 42.3) which is predicted at 20 $MPa^{1/2}$. If water is removed from the model and the correlation is quantified the R-squared value is greater than 0.8. As with the polar parameter predictions from this model there is no discernable trend which can be noted about the predictions. No chemical group is predicted any better or worse than the others. The only trend appears to be that the hydrogen-bonding parameter becomes harder to predict the larger it becomes. This is problematic for the prediction of unknowns in that there can be no confidence in values predicted at 20 $MPa^{1/2}$ or higher. This means that this model with likely have trouble predicting the nitramine species

discussed in Chapter 1. Based on their structure, these species are suspected of having large hydrogen-bonding contributions to their total cohesive energy.

If these observations are applied to the predictions of the explosive unknowns then the reliability of these predictions (Table 4) can be evaluated. First, the dispersion values are predicted largely between 20 and 30 $MPa^{1/2}$. This is outside of the normal range of dispersion parameters; however, the size of these species is not to be underestimated. It is likely that the presence of two, three, or sometimes four nitro groups in these species serves to swell the electron cloud and increase the size of the molecule and also the value of the dispersion parameter. This might also help explain why these molecules exhibit such poor solubility in many common solvents which have dispersion parameters between 15 and 20 $MPa^{1/2}$. It may be wise to discount the polar parameter predictions from this model because of the poor correlation which was achieved in that model. The hydrogen-bonding parameter predictions should prove reliable up to 20 $MPa^{1/2}$ as noted earlier. Unfortunately, this caveat discredits the values for nitramine species such as tetranitrocarbazole, picramide, TATB, and DATB.

QSAR2

As with the first model, the polar parameter depends heavily on dipole moment and Δ charge structure variables. The hydrogen-bonding parameter in this model depends largely upon Δ charge (*d*), polarizability (α), and HOMO energy (*I*) terms as was expected. It is interesting to note that the coefficient which is calculated for the $\mu/V_m^{1/2}$ term (37.4) is the same as the coefficient in Hansen and Beerbower's simplified Böttcher equation. This tells more about the way in which most of the δ_P parameters were calculated in the first place and not necessarily a validation of the model. What this also

means is that Gaussian® calculated dipole moments may be applied to the simplified Böttcher equation to calculate δ_P.

Correlation of the dispersion parameter predictions from this model (Figure 10) looks very similar to that of the first model. The correlation is slightly improved (R^2= 0.5806) however this nominally poor correlation is subject to the same explanation as the other dispersion parameter predictions.

The polar parameter is predicted better by this model. The overall correlation is better (R^2 = 0.7381). More of the species are predicted accurately by the expanded variable set this model uses, however there still are some major outliers. Some species are predicted as much as 10 $MPa^{1/2}$ off of their literature values. In these extreme cases it is likely that the literature value was not determined in the same way as the rest of values.

The hydrogen-bonding parameter is also predicted better by this model. A ten percent increase in correlation with the training set values (R^2 = 0.8771) is accompanied by the ability to predict even the highest hydrogen-bonding parameter (water, δ_H = 42.3 $MPa^{1/2}$) with a degree of confidence. One thing that can be noted from Figure 12 is that a large number of the species which are calculated to be higher than the literature values are primary amines. In fact, nearly all of these species are calculated as having a hydrogen-bonding parameter 6-8 $MPa^{1/2}$ larger than the literature value. The reason for this is unknown. It is possible that the ability of primary amine to form two hydrogen-bonds is somehow involved in this error. If more primary amines are investigated with this model it may be possible to institute some subtractive term for a primary amine which can be applied after the transformation matrix to correct for this effect. In practice

however this may prove difficult as there is no easy way to determine from a Gaussian® result file if the geometry present is a primary amine or not.

As with the first model, the observations about the correlations may help evaluate the prediction of the explosive unknowns. The predictions of these species from the QSAR2 model are shown in Table 5. Dispersion parameters are very similar to those calculated by QSAR1. They range from the low twenties to the low thirties and are consistent with the approximate size of those species. The polar parameters calculated by QSAR2 can now be accepted a little more than in QSAR1. It should be noted that the polar parameters are the least reliable numbers compared to the dispersion and hydrogen-bonding parameters of this model. This is owing to the variability of the dipole moment calculations and also the varied methods with which the training set polar parameters were calculated. The hydrogen-bonding parameter predictions fall within the normal range of values, however, species with large expected hydrogen-bonding parameters (δ_H>20 $MPa^{1/2}$) such as DATB and picramide were not predicted to be so. They were indeed larger than their companion explosive unknowns; however, this fact coupled with the primary amine discrepancy leaves some doubt with these numbers.

QSAR3

In general, the correlation of this model's prediction of the training set values appears to be increased. This is due to the smaller training set and so any comparison of correlation with QSAR1 or QSAR2 would be erroneous. There are some important features to notice in this model that may serve to interpret the predictions of the explosive unknowns later on.

The dispersion parameter predictions of the training set species in this model are very accurate. The overall correlation is great for the variability over the small range of dispersion parameters ($R^2 = 0.8132$). A look at Figure 13 shows that most species are predicted to within 1 $MPa^{1/2}$ of the literature values. The deviation that is present likely results from the variability in the computational route to the structure variables. A higher basis set such as 6-311G(d,p) could possibly tighten these values up; however, they are accurate enough for use with other solubility parameters in their current state.

Polar parameter predictions of training set species are accurate as well. Figure 14 shows a good overall correlation ($R^2 = 0.9008$). There are a few species which are predicted to be 4 -5 $MPa^{1/2}$ from their target values, most notably on the high side. The majority of species are predicted within 3 $MPa^{1/2}$ and 44.4% are calculated to within 1 $MPa^{1/2}$. The major outliers are not exclusively of one chemical group or another. Two are oxygen containing ringed systems (1,4-dioxacyclohexane and 4-methyl-1,3-dioxolan-2-one). The rest are species which contain nitro-groups (nitrobenzene, 4-nitrotoluene, and 3,5-dinitrophenol). This is not to say that all species with nitro-groups were predicted poorly. In fact, several species of this category were very accurately predicted. 3-nitroaniline, 2,4-dinitrotoluene, and the nitropropanes were some of the best.

QSAR3 consistently predicts hydrogen-bonding parameters to within 5 $MPa^{1/2}$ or less. This is not ideal; however the high correlation which is achieved suggests that the model may be on the right track. One thing that may tighten these values up may be achieved with an increase the computational effort. Figure 6 shows that the model for the hydrogen-bonding parameter in QSAR3 is extremely simple. One variable is the $\mu/V_m^{1/2}$ term. An increase in the computational effort as discussed earlier may determine this

$\mu/V_m^{1/2}$ term more accurately and lead to a better prediction of the hydrogen-bonding parameter.

Dispersion parameter predictions of the explosive unknowns are much lower than the other models. It has been noted that this model seems to predict dispersion parameters as much as 5 $MPa^{1/2}$ below the actual value of the parameter. This suggests that the real dispersion parameters of these species may lie between 20 and 26 $MPa^{1/2}$ instead of between 15 and 21 $MPa^{1/2}$ as shown in Table 6. Predictions of the polar parameters for the explosive unknowns must be taken with a grain of salt. QSAR3 predicts polar parameters between 8 and 22 $MPa^{1/2}$. According to the training set correlation these predictions should be lower than actual; however these predictions are higher than the previous two models. The hydrogen-bonding parameters may be off by as much as 5 $MPa^{1/2}$.

CHAPTER V

CONCLUSION

In conclusion, the author has demonstrated a fast, unambiguous tool for the determination of the solubility parameters of nitrated materials. There is good evidence to suggest that the explosive unknowns presented in this work have dispersion parameters in the high teens to around 20 $MPa^{1/2}$. Polar parameters were determined but users should be cautious of values which were calculated by QSAR models 1 and 2 because many values were calculated to be as much as 5 $MPa^{1/2}$ off of their target values. However, the polar parameters predicted by QSAR3 appear to be very reliable. The most interesting findings in this work may be the hydrogen-bonding parameters. The precision with which the hydrogen-bonding parameter can be calculated by QSARs 2 and 3 suggests that this work is well on its way to the reliable determination of hydrogen-bonding parameters. Recent work in the area of solute precipitation has indicated that the hydrogen-bonding character of a solvent may influence the crystal morphology of the precipitate (22). This is exciting for explosives manufacturers because certain crystal morphologies have been linked to desirable detonation properties.

This information may be enough for the estimation of where to begin probing blended solvents for the solubility of nitrated species. Table 7 shows recently obtained experimental data from the Pantex facility (23). R_a is a value that represents the "distance" between two solubility parameters in the (D,P,H) Cartesian space. Clearly a small R_a would indicate a prediction is close to accuracy. This column shows that QSAR3 predicts the HSPs of HMX, RDX, PETN, and HNS better than either group contribution method.

Table 7. A comparison of group contribution methods and QSAR3 HSP predictions of HMX, RDX, PETN, and HNS with experimentally determined HSPs

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$^{1/2}$
HMX	exp	0.10	17.7	11.6	13.7	12.1	1.00	
	GC[a]		22.5	22.9	6.9			16.4
	GC[b]		23.5	45.2	7.3			36.1
	QSAR3		15.3	13.4	5.6			8.6
RDX	exp	1.40	17.3	12.4	9.1	8.4	1.00	
	GC[a]		18.5	18.8	5.7			7.7
	GC[b]		22.0	35.8	7.5			25.3
	QSAR3		15.5	11.1	6.5			3.4
PETN	exp	1.50	16.7	12.0	8.4	7.8	1.00	
	GC[a]		21.4	21.2	9.5			13.1
	GC[b]		18.8	50.4	3.0			39.0
	QSAR3		16.8	17.8	6.4			6.1
HNS	exp	0.15	18.9	13.9	6.1	6.0	0.98	
	GC[a]		21.0	13.3	8.6			4.9
	GC[b]		28.0	33.0	1.9			26.7
	QSAR3		20.7	17.1	5.6			3.7

[a]VanKrevelen method, [b]Stefanis method

Another exciting aspect of this work is that the methodology can be applied to predict the solubility parameters of any chemical group. Indeed, it may be possible to expand this work to predict the Hansen solubility parameters of species that are important to other disciplines such as ink pigments, polymer additives, surfactants, etc.

LITERATURE CITED

(1) Hildebrand, J.; Scott, R.L. *The Solubility of Nonelectrolytes,* 3rd ed.; Reinhold: New York, 1950, 1 ff.

(2) Hildebrand, J.; Scott, R.L. *Regular Solutions*; Prentice-Hall: Englewood Cliffs, NJ, 1962, 1 ff.

(3) Hildebrand, J.H. The term 'Regular Solution.' *Nature*. **1951**, 168, 868.

(4) Hansen, C.M. *Hansen Solubility Parameters: A Users Handbook*, 2nd ed.; CRC Press: Boca Raton, FL, 2007, 45-46.

(5) Karger, B.L.; Snyder, L.R.; Eon, C. An expanded solubility parameter treatment for classification and use of chromatographic solvents and adsorbents. Parameters for dispersion, dipole, and hydrogen bonding interactions. *J. Chromatogr*. **1976**, 125, 71.

(6) Beerbower, A., Wu, P.L., and Martin, A., Expanded solubility parameter approach I. Naphthalene and benzoic acid in individual solvents, *J. Pharm. Sci.* **1984**, 73, 179.

(7) Martin, A.; Wu, P.L.; Beerbower, A. Expanded solubility parameter approach II. *p*-hydroxybenzoic acid and *p*-hydroxybenzoate in individual solvents. *J. Pharm. Sci.* **1984**, 73, 188.

(8) Hansen, C.M.; Skaarup, K. The three dimensional solubility parameter – key to paint component affinities III. *J. Paint Technol*. **1967,** 39, 511-514.

(9) van Krevelen, D.W.; Hoftyzer, P.J. *Properties of Polymers: Their Estimation and Correlation with Chemical Structure*, 2nd ed.; Elsevier: Amsterdam, 1976, pp 189-225.

(10) Stefanis, E.; Panayioutou, C. Prediction of Hansen Solubility Parameters with a New Group-Contribution Method. *Int. J. Thermophys*. **2008,** 29, 568-585.

(11) Blanks, R.F.; Prausnitz, J.M. Thermodynamics of polymer solubility in polar and nonpolar systems. *Ind. Eng. Chem. Fundam.* **1964,** 3, 1-8.

(12) Hansen, C.M.; Beerbower, A. Solubility parameters, in *Kirk-Othmer Encyclopedia of Chemical Technology*, Suppl. Vol., 2nd ed.; Standen, A., Ed.; Interscience: New York, 1971; 889-910.

(13) Barnett, M.P. Mechanized Molecular Calculations – The POLYATOM system. *Rev. Mod. Phys*. **1963**, 35, 571-572.

(14) Foresman, J.; Frisch, A. *Exploring Chemistry with Electronic Stucture Methods*, 2nd ed.; Gaussian, Inc.: Pittsburgh, 1996, pp 157-159.

(15) Zeng, X.; Wang, Z.; Ge, Z.; Liu, H. Quantitative structure–property relationships for predicting subcooled liquid vapor pressure (PL) of 209 polychlorinated diphenyl ethers (PCDEs) by DFT and the position of Cl substitution (PCS) methods. *Atmos. Environ.* **2007**, 41, 3590-3603.

(16) Mollet, H.; Grubenmann, A. *Formulation Technology*; Wiley-VCH: Weinheim, 2001, pp 268-272.

(17) 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.

(18) Becke, A. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, 98, 5648-5652.

(19) Lee, C.; Wang, W.; Parr, R.G. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, 37, 785-789.

(20) Barton, A.F. *Handbook of Solubility Parameters and Other Cohesion Parameters*, 2nd ed.; CRC Press: Boca Raton, FL, 1991, 57-58.

(21) Minitab® 15.1.0.0; Minitab, Inc.: State College, PA, 2006.

(22) Williams, D.L. *Measurement of Constants of Crystalline Explosives;* Final Report to B&W Pantex LLC. On Grant Contract 60785; Sam Houston State University: Huntsville, 2008.

(23) Duncan, A., Unpublished compilation of HE Solubility Data, B&W Pantex LLC, Amarillo, TX 2008.

APPENDICES

Appendix A: IUPAC Nomenclature and CPU Times for Gaussian® Calculations

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
(chloromethyl)benzene	C_7H_7Cl	33 minutes 18.0 seconds	1 hours 13 minutes 3.0 seconds
(*E*)1,3,5-Trinitro-2-[2-(2,4,6-trinitrophenyl)ethenyl]benzene	$C_{14}H_6N_6O_{12}$	8 hours 21 minutes 15.5 seconds	9 hours 24 minutes 9.2 seconds
(*Z*)-octadec-9-en-1-ol	$C_{18}H_{36}O$	3 hours 33 minutes 30.5 seconds	18 hours 5 minutes 12.1 seconds
(*Z*)-octadec-9-enoic acid	$C_{18}H_{34}O_2$	4 hours 14 minutes 37.4 seconds	19 hours 45 minutes 40.6 seconds
1-(phenoxy)-3-[3-(phenoxy)phenoxy]benzene	$C_{24}H_{18}O_3$	4 hours 36 minutes 46.5 seconds	11 hours 45 minutes 10.4 seconds
1,1,1-trichloroethane	$C_2H_3Cl_3$	1 minutes 27.5 seconds	13 minutes 58.4 seconds
1,1,2,2-tetrabromoethane	$C_2H_2Br_4$	7 minutes 49.3 seconds	30 minutes 43.2 seconds
1,1,2,2-tetrachloroethane	$C_2H_2Cl_4$	3 minutes 7.6 seconds	13 minutes 38.4 seconds
1,1,2-trichlorotrifluoroethane	$C_2Cl_3F_3$	10 minutes 47.0 seconds	33 minutes 60.0 seconds
1,1,3,3-tetramethylurea	$C_5H_{12}N_2O$	1 hours 6 minutes 26.2 seconds	1 hours 38 minutes 48.4 seconds
1,1-dichloroethane	$C_2H_4Cl_2$	50.0 seconds	6 minutes 10.5 seconds
1,1-dichloroethene	$C_2H_2Cl_2$	40.4 seconds	4 minutes 5.5 seconds
1,1-dimethylhydrazine	$C_2H_8N_2$	2 minutes 38.6 seconds	15 minutes 0.4 seconds
1,1-thiobisethane	$C_4H_{10}S$	3 minutes 56.1 seconds	29 minutes 2.1 seconds
1,2,3,4-tetrahydronaphthalene	$C_{10}H_{12}$	14 minutes 36.3 seconds	1 hours 44 minutes 47.0 seconds
1,2-diaminoethane	$C_2H_8N_2$	5 minutes 16.4 seconds	14 minutes 48.0 seconds
1,2-dibromoethane	$C_2H_4Br_2$	1 minutes 27.2 seconds	7 minutes 57.6 seconds
1,2-dichlorobenzene	$C_6H_4Cl_2$	5 minutes 7.0 seconds	52 minutes 2.2 seconds
1,2-dichloroethane	$C_2H_4Cl_2$	1 minutes 16.0 seconds	6 minutes 1.0 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
1,2-dichlorotetrafluoroethane	$C_2Cl_2F_4$	3 minutes 1.1 seconds	20 minutes 18.2 seconds
1,2-dimethylbenzene	C_8H_{10}	15 minutes 37.7 seconds	1 hours 29 minutes 10.6 seconds
1,2-dinitrobenzene	$C_6H_4N_2O_4$	4 hours 6 minutes 12.8 seconds	2 hours 8 minutes 31.8 seconds
1,3,5,7-tetranitro-1,3,5,7-tetrazocane	$C_4H_8N_8O_8$	5 hours 16 minutes 25.6 seconds	6 hours 9 minutes 14.5 seconds
1,3,5-triamino-2,4,6-trinitrobenzene	$C_6H_6N_6O_6$	44 minutes 39.0 seconds	23 hours 1 minutes 9.0 seconds
1,3,5-trimethylbenzene	C_9H_{12}	5 minutes 11.9 seconds	42 minutes 51.9 seconds
1,3,5-trinitro-1,3,5-triazinane	$C_3H_6N_6O_6$	6 hours 44 minutes 13.4 seconds	4 hours 42 minutes 59.9 seconds
1,3,5-trinitroaniline	$C_6H_4N_4O_6$	9 hours 52 minutes 10.9 seconds	4 hours 30 minutes 8.5 seconds
1,3,5-trinitrobenzene	$C_6H_3N_3O_6$	3 hours 5 minutes 59.7 seconds	3 hours 23 minutes 27.5 seconds
1,3,5-trinitrotoluene	$C_7H_5N_3O_6$	5 hours 1 minutes 10.6 seconds	5 hours 15 minutes 53.6 seconds
1,3,6,8-tetranitrocarbazole	$C_{12}H_5N_5O_8$	1 days 8 hours 18 minutes 11.4 seconds	21 hours 24 minutes 18.5 seconds
1,3-diamino-2,4,6-trinitrobenzene	$C_6H_5N_5O_6$	15 hours 18 minutes 9.9 seconds	6 hours 25 minutes 21.6 seconds
1,3-dimethylbutan-1-ol	$C_6H_{14}O$	20 minutes 53.3 seconds	1 hours 19 minutes 41.4 seconds
1,3-dinitrooxypropan-2-yl nitrate	$C_3H_5N_3O_9$	3 hours 43 minutes 11.3 seconds	3 hours 9 minutes 13.1 seconds
1,3-dioxolan-2-one	$C_3H_4O_3$	18 minutes 56.0 seconds	21 minutes 45.0 seconds
1,4-dioxacyclohexane	$C_4H_8O_2$	6 minutes 10.5 seconds	37 minutes 10.2 seconds
1-bromoethane	C_2H_5Br	1 minutes 38.2 seconds	7 minutes 6.6 seconds
1-bromonaphthalene	$C_{10}H_7Br$	16 minutes 16.3 seconds	2 hours 32 minutes 35.7 seconds
1-bromopropane	C_3H_7Br	5 minutes 10.0 seconds	22 minutes 28.0 seconds
1-chlorobutane	C_4H_9Cl	5 minutes 1.2 seconds	28 minutes 12.4 seconds
1-chloropropane	C_3H_7Cl	2 minutes 26.9 seconds	12 minutes 58.5 seconds
1-decanol	$C_{10}H_{22}O$	21 minutes 22.1 seconds	4 hours 16 minutes 21.3 seconds
1-methyl-2-pyrrolidone	C_5H_9NO	2 hours 10 minutes 5.0 seconds	1 hours 4 minutes 33.0 seconds
1-methylnaphthalene	$C_{11}H_{10}$	18 minutes 24.7 seconds	2 hours 53 minutes 3.6 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
1-nitropropane	$C_3H_7NO_2$	10 minutes 16.0 seconds	33 minutes 37.0 seconds
1-octanol	$C_8H_{18}O$	13 minutes 49.0 seconds	2 hours 11 minutes 52.3 seconds
1-pentanol	$C_5H_{12}O$	6 minutes 12.7 seconds	48 minutes 49.0 seconds
1-propanol	C_3H_8O	2 minutes 26.0 seconds	15 minutes 37.0 seconds
1-tridecanol	$C_{13}H_{28}O$	35 minutes 45.4 seconds	7 hours 23 minutes 57.8 seconds
1-trifluoromethyl-1,2,2,3,3,4,4,5,5,6,6-unedecafluorocyclohexane	C_7F_{14}	2 hours 21 minutes 12.9 seconds	10 hours 32 minutes 59.6 seconds
2-(2-butoxyethoxy)ethanol	$C_8H_{18}O_3$	15 minutes 15.3 seconds	2 hours 55 minutes 30.4 seconds
2-(2-ethoxyethoxy)ethanol	$C_6H_{14}O_3$	11 minutes 9.4 seconds	1 hours 42 minutes 11.9 seconds
2-(2-Methoxyethoxy)ethanol	$C_5H_{12}O_3$	7 minutes 38.8 seconds	1 hours 12 minutes 0.3 seconds
2,2',4,4',6,6'-hexanitroazobenzene	$C_{12}H_4N_8O_{12}$	7 hours 39 minutes 29.5 seconds	1 days 8 hours 55 minutes 27.8 seconds
2,2,4-trimethylpentane	C_8H_{18}	1 hours 22 minutes 24.1 seconds	2 hours 33 minutes 48.8 seconds
2,4,6-trinitro-1,3,5-triethoxybenzene	$C_{12}H_{15}N_3O_9$	18 hours 54 minutes 3.4 seconds	1 days 9 hours 32 minutes 55.4 seconds
2,4,6-trinitroanisole	$C_7H_5N_3O_7$	4 hours 6 minutes 48.4 seconds	6 hours 39 minutes 52.0 seconds
2,4,6-trinitrobenzene-1,3,5-triol	$C_6H_3N_3O_9$	5 hours 52 minutes 27.7 seconds	7 hours 4 minutes 8.2 seconds
2,4,6-trinitrobenzoic acid	$C_7H_3N_3O_8$	11 hours 10 minutes 32.0 seconds	6 hours 49 minutes 44.2 seconds
2,4,6-trinitrochlorobenzene	$C_6H_2ClN_3O_6$	2 hours 30 minutes 2.7 seconds	4 hours 13 minutes 1.1 seconds
2,4,6-trinitroethoxybenzene	$C_8H_7N_3O_7$	5 hours 4 minutes 40.6 seconds	8 hours 17 minutes 11.2 seconds
2,4,6-trinitrophenol	$C_6H_3N_3O_7$	11 minutes 55.6 seconds	4 hours 26 minutes 28.4 seconds
2,4,6-trinitrophenylethylnitramine	$C_8H_7N_5O_8$	8 hours 23 minutes 32.8 seconds	15 hours 55 minutes 44.5 seconds
2,4,6-trinitrophenylmethylnitramine	$C_7H_5N_5O_8$	6 hours 14 minutes 15.3 seconds	12 hours 18 minutes 37.6 seconds
2,4-dinitrotoluene	$C_7H_6N_2O_4$	3 hours 49 minutes 46.4 seconds	2 hours 47 minutes 16.4 seconds
2,6-dimethylheptan-4-one	$C_9H_{18}O$	2 hours 8 minutes 57.0 seconds	6 hours 28 minutes 34.0 seconds
2-aminoethanol	C_2H_7NO	2 minutes 49.0 seconds	13 minutes 41.0 seconds
2-butanol	$C_4H_{10}O$	6 minutes 59.7 seconds	29 minutes 4.4 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
2-butoxyethanol	$C_6H_{14}O_2$	9 minutes 27.6 seconds	1 hours 26 minutes 52.3 seconds
2-ethoxyethanol	$C_4H_{10}O_2$	4 minutes 59.9 seconds	39 minutes 21.8 seconds
2-ethoxyethyl acetate	$C_6H_{12}O_3$	19 minutes 1.8 seconds	1 hours 32 minutes 1.3 seconds
2-ethylbutan-1-ol	$C_6H_{14}O$	53 minutes 3.3 seconds	1 hours 22 minutes 25.4 seconds
2-ethylhexan-1-ol	$C_8H_{18}O$	1 hours 37 minutes 21.3 seconds	2 hours 36 minutes 24.8 seconds
2-furanmethanol	$C_5H_6O_2$	16 minutes 25.9 seconds	38 minutes 40.6 seconds
2-methoxyethanol	$C_3H_8O_2$	8 minutes 51.0 seconds	25 minutes 33.0 seconds
2-methyl-1-propanol	$C_4H_{10}O$	6 minutes 43.6 seconds	32 minutes 7.6 seconds
2-methylbutane	C_5H_{12}	7 minutes 4.7 seconds	40 minutes 29.3 seconds
2-methylpropyl 2-methylpropanote	$C_8H_{16}O_2$	1 hours 45 minutes 26.5 seconds	2 hours 58 minutes 19.7 seconds
2-methylpropyl ethanoate	$C_6H_{12}O_2$	1 hours 35 minutes 4.0 seconds	2 hours 46 minutes 35.0 seconds
2-nitropropane	$C_3H_7NO_2$	9 minutes 7.0 seconds	33 minutes 12.5 seconds
2-octanol	$C_8H_{18}O$	29 minutes 52.9 seconds	2 hours 24 minutes 20.6 seconds
2-propanol	C_3H_8O	3 minutes 12.0 seconds	16 minutes 27.0 seconds
2-pyrrolidone	C_4H_7NO	19 minutes 27.0 seconds	37 minutes 12.0 seconds
3,5,5-trimethylcyclohex-2-en-1-one	$C_9H_{14}O$	43 minutes 53.3 seconds	3 hours 2 minutes 46.3 seconds
3,5-dinitrophenol	$C_6H_4N_2O_5$	5 hours 30 minutes 46.1 seconds	2 hours 17 minutes 21.2 seconds
3,6,9-trioxa-(18*Z*)-heptacosan-1ol	$C_{24}H_{48}O_4$	7 hours 44 minutes 0.9 seconds	2 days 9 hours 37 minutes 33.9 seconds
3-chloropropan-1-ol	C_3H_7ClO	4 minutes 0.7 seconds	21 minutes 34.2 seconds
3-hydroxyl-2,4,6-trinitrophenol	$C_6H_3N_3O_8$	3 hours 25 minutes 11.9 seconds	5 hours 52 minutes 33.7 seconds
3-methyl-2,4,6-trinitrotoluene	$C_8H_7N_3O_6$	5 hours 5 minutes 28.1 seconds	7 hours 40 minutes 40.5 seconds
3-methylbutyl ethanoate	$C_7H_{14}O_2$	33 minutes 49.8 seconds	1 hours 59 minutes 18.8 seconds
3-methylphenol	C_7H_8O	24 minutes 28.7 seconds	1 hours 6 minutes 45.1 seconds
3-nitroaniline	$C_6H_6N_2O_2$	3 hours 23 minutes 27.7 seconds	1 hours 30 minutes 47.2 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
3-nitrooxy-2,2-bis(nitrooxymethyl)propyl nitrate	$C_5H_8N_4O_{12}$	2 minutes 26.7 seconds*	1 minutes 45.3 seconds*
4-hydroxy-4-methylpentan-2-one	$C_6H_{12}O_2$	57 minutes 52.0 seconds	1 hours 35 minutes 29.3 seconds
4-methyl-1,3-dioxolan-2-one	$C_4H_6O_3$	1 hours 8 minutes 9.0 seconds	45 minutes 2.0 seconds
4-methylpent-3-en-2-one	$C_6H_{10}O$	25 minutes 50.8 seconds	57 minutes 11.1 seconds
4-methylpentan-2-one	$C_6H_{12}O$	36 minutes 47.2 seconds	1 hours 8 minutes 45.0 seconds
4-nitrochlorobenzene	$C_6H_4ClNO_2$	32 minutes 24.7 seconds	1 hours 17 minutes 8.9 seconds
4-nitrophenol	$C_6H_5NO_3$	2 hours 15 minutes 42.7 seconds	1 hours 19 minutes 43.1 seconds
4-nitrotoluene	$C_7H_7NO_2$	2 hours 27 minutes 10.5 seconds	1 hours 30 minutes 37.3 seconds
5-methylhexan-2-one	$C_7H_{14}O$	2 hours 5 minutes 33.4 seconds	1 hours 36 minutes 4.5 seconds
acetaldehyde	C_2H_4O	12.3 seconds	3 minutes 52.4 seconds
acetic acid	$C_2H_4O_2$	2 minutes 47.3 seconds	6 minutes 37.0 seconds
acetonitrile	C_2H_3N	26.0 seconds	2 minutes 44.0 seconds
acetophenone	C_8H_8O	32 minutes 12.0 seconds	1 hours 36 minutes 44.0 seconds
aniline	C_6H_7N	1 hours 23 minutes 47.8 seconds	45 minutes 46.3 seconds
benzaldehyde	C_7H_6O	21 minutes 10.0 seconds	1 hours 1 minutes 27.0 seconds
benzene	C_6H_6	3 minutes 37.0 seconds	1 hours 2 minutes 3.0 seconds
benzene-1,2-diol	$C_6H_6O_2$	18 minutes 47.4 seconds	54 minutes 41.3 seconds
benzene-1,3,5-triol	$C_6H_6O_3$	25 minutes 31.5 seconds	1 hours 7 minutes 36.6 seconds
benzene-1,3-diol	$C_6H_6O_3$	23 minutes 3.2 seconds	56 minutes 37.6 seconds
benzene-1,4-diol	$C_6H_6O_4$	16 minutes 27.9 seconds	50 minutes 46.1 seconds
benzoic acid	$C_7H_6O_2$	1 hours 17 minutes 4.0 seconds	1 hours 20 minutes 2.0 seconds
benzonitrile	C_7H_5N	8 minutes 7.0 seconds	58 minutes 36.0 seconds
benzyl n-butyl phthalate	$C_{19}H_{20}O_4$	9 minutes 40.3 seconds	8 hours 39 minutes 8.7 seconds
biphenyl	$C_{12}H_{10}$	22 minutes 19.9 seconds	3 hours 6 minutes 4.0 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
bis(2-aminoethyl)amine	$C_4H_{13}N_3$	12 minutes 9.0 seconds	1 hours 35 minutes 5.0 seconds
bis(2-chloroethyl) ether	$C_4H_8Cl_2O$	11 minutes 3.0 seconds	52 minutes 27.0 seconds
bis(2-chloroisopropyl) ether	$C_6H_{12}Cl_2O$	1 hours 1 minutes 29.6 seconds	2 hours 8 minutes 52.1 seconds
bis(2-methoxyethyl) ether	$C_6H_{14}O_3$	11 minutes 15.4 seconds	1 hours 39 minutes 17.7 seconds
bromobenzene	C_6H_5Br	7 minutes 24.3 seconds	48 minutes 50.0 seconds
bromochloromethane	CH_2BrCl	51.0 seconds	4 minutes 7.0 seconds
bromotrifluoromethane	$CBrF_3$	1 minutes 1.6 seconds	7 minutes 29.6 seconds
butan-1-ol	$C_4H_{10}O$	3 minutes 0.8 seconds	26 minutes 48.1 seconds
butanal	C_4H_8O	9 minutes 4.0 seconds	26 minutes 5.0 seconds
butane	C_4H_{10}	8 minutes 11.3 seconds	17 minutes 26.5 seconds
butanenitrile	C_4H_7N	4 minutes 55.0 seconds	39 minutes 15.0 seconds
butano-4-lactone	$C_4H_6O_2$	9 minutes 26.9 seconds	28 minutes 22.9 seconds
butanoic acid	$C_4H_8O_2$	26 minutes 13.0 seconds	1 hours 3 minutes 5.0 seconds
butanone	C_4H_8O	21 minutes 14.0 seconds	45 minutes 44.0 seconds
butyl (2R)-hydroxypropanoate	$C_7H_{14}O_3$	38 minutes 14.3 seconds	2 hours 23 minutes 37.4 seconds
butyl ethanoate	$C_6H_{12}O_2$	15 minutes 32.7 seconds	1 hours 16 minutes 17.9 seconds
butylamine	$C_4H_{11}N$	16 minutes 27.0 seconds	1 hours 0 minutes 33.0 seconds
carbon disulfide	CS_2	10.0 seconds	55.0 seconds
chlorobenzene	C_6H_5Cl	6 minutes 23.0 seconds	43 minutes 53.0 seconds
chlorocyclohexane	$C_6H_{11}Cl$	17 minutes 59.3 seconds	1 hours 14 minutes 5.5 seconds
chlorodifluoromethane	$CHClF_2$	45.0 seconds	3 minutes 42.0 seconds
chloromethane	CH_4Cl	23.0 seconds	1 minutes 38.0 seconds
chloromethyloxirane	C_3H_5ClO	3 minutes 15.2 seconds	17 minutes 32.0 seconds
cis-bicyclo[4.4.0]decane	$C_{10}H_{18}$	35 minutes 16.4 seconds	4 hours 33 minutes 28.1 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
cyclohexanamine	$C_6H_{13}N$	9 minutes 58.0 seconds	1 hours 26 minutes 46.0 seconds
cyclohexane	C_6H_{12}	9 minutes 55.3 seconds	1 hours 0 minutes 9.5 seconds
cyclohexanol	$C_6H_{12}O$	15 minutes 54.6 seconds	1 hours 22 minutes 39.7 seconds
cyclohexanone	$C_6H_{10}O$	18 minutes 49.9 seconds	1 hours 7 minutes 1.5 seconds
decane	$C_{10}H_{22}$	7 minutes 14.3 seconds	46 minutes 53.5 seconds
dibutyl phthalate	$C_{16}H_{22}O_4$	7 hours 14 minutes 15.6 seconds	22 hours 15 minutes 8.6 seconds
dibutyl sebacate	$C_{18}H_{34}O_4$	4 hours 37 minutes 14.4 seconds	1 days 1 hours 20 minutes 49.7 seconds
dichlorodifluoromethane	CCl_2F_2	41.0 seconds	4 minutes 59.8 seconds
dichlorofluoromethane	$CHCl_2F$	39.7 seconds	3 minutes 30.7 seconds
dichloromethane	CH_2Cl_2	27.0 seconds	2 minutes 11.0 seconds
diethyl carbonate	$C_5H_{10}O_3$	7 minutes 1.2 seconds	47 minutes 41.8 seconds
diethyl phthalate	$C_{12}H_{14}O_4$	3 hours 9 minutes 48.2 seconds	8 hours 45 minutes 8.0 seconds
diethyl sulfate	$C_4H_{10}O_4S$	1 hours 17 minutes 6.0 seconds	1 hours 48 minutes 51.0 seconds
diethylamine	$C_4H_{11}N$	4 minutes 57.4 seconds	31 minutes 37.0 seconds
dimethyl phthalate	$C_{10}H_{10}O_4$	1 hours 34 minutes 14.4 seconds	2 hours 12 minutes 29.0 seconds
dimethyl sulfone	$C_2H_6O_2S$	20 minutes 25.0 seconds	23 minutes 42.0 seconds
dimethyl sulfoxide	C_2H_6OS	9 minutes 56.0 seconds	13 minutes 5.0 seconds
dioctyl phthalate	$C_{24}H_{38}O_4$	7 hours 0 minutes 1.8 seconds	14 hours 48 minutes 36.7 seconds
dodecane	$C_{12}H_{26}$	24 minutes 34.5 seconds	5 hours 0 minutes 23.8 seconds
ethanol	C_2H_6O	1 minutes 19.0 seconds	7 minutes 14.0 seconds
ethoxyethane	$C_4H_{10}O$	3 minutes 20.0 seconds	28 minutes 59.8 seconds
ethyl 2-hydroxypropanoate	$C_5H_{10}O_3$	43 minutes 58.9 seconds	1 hours 6 minutes 52.5 seconds
ethyl 3-phenyl-2-propenonate	$C_{11}H_{12}O_2$	1 hours 6 minutes 42.1 seconds	3 hours 34 minutes 45.1 seconds
ethyl ethanoate	$C_4H_8O_2$	7 minutes 9.0 seconds	32 minutes 7.1 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
ethyl methanoate	$C_3H_6O_2$	24 minutes 55.0 seconds	19 minutes 38.0 seconds
ethylbenzene	C_8H_{10}	36 minutes 26.8 seconds	1 hours 21 minutes 50.0 seconds
furan	C_4H_4O	3 minutes 42.7 seconds	14 minutes 8.0 seconds
furan-2-carbaldehyde	$C_5H_4O_2$	10 minutes 48.0 seconds	34 minutes 48.0 seconds
heptane	C_7H_{16}	8 minutes 50.3 seconds	1 hours 27 minutes 2.0 seconds
hexadecane	$C_{16}H_{34}$	37 minutes 55.5 seconds	11 hours 12 minutes 19.4 seconds
hexane	C_6H_{14}	5 minutes 35.7 seconds	55 minutes 56.6 seconds
icosane	$C_{20}H_{42}$	1 hours 31 minutes 37.9 seconds	20 hours 35 minutes 58.3 seconds
isopropyl hexadecanoate	$C_{19}H_{38}O_2$	3 hours 7 minutes 15.9 seconds	23 hours 34 minutes 31.2 seconds
methanamide	CH_3NO	48.0 seconds	3 minutes 44.0 seconds
methanoic acid	CH_2O_2	1 minutes 24.0 seconds	4 minutes 41.0 seconds
methanol	CH_4O	41.0 seconds	2 minutes 24.0 seconds
methoxybenzene	C_7H_8O	22 minutes 22.9 seconds	1 hours 3 minutes 37.2 seconds
methyl (*Z*)-octadec-9-eneoate	$C_{19}H_{36}O_2$	2 hours 25 minutes 22.9 seconds	7 hours 33 minutes 40.4 seconds
methyl acetate	$C_3H_6O_2$	20 minutes 23.0 seconds	20 minutes 29.0 seconds
methyl benzoate	$C_8H_8O_2$	32 minutes 6.2 seconds	1 hours 34 minutes 41.1 seconds
methylcyclohexane	C_7H_{14}	9 minutes 32.7 seconds	1 hours 37 minutes 12.0 seconds
morpholine	C_4H_9NO	3 minutes 48.0 seconds	32 minutes 23.7 seconds
N,N-dimethylacetamide	C_4H_9NO	1 hours 19 minutes 5.0 seconds	1 hours 17 minutes 12.0 seconds
N,N-dimethylmethanamide	C_3H_7NO	23 minutes 50.0 seconds	24 minutes 6.0 seconds
naphthalene	$C_{10}H_8$	12 minutes 20.1 seconds	1 hours 54 minutes 5.1 seconds
nitrobenzene	$C_6H_5NO_2$	19 minutes 25.3 seconds	1 hours 5 minutes 7.9 seconds
nitroethane	$C_2H_5NO_2$	12 minutes 12.0 seconds	15 minutes 38.7 seconds
nitromethane	CH_3NO_2	2 minutes 34.0 seconds	6 minutes 47.0 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
nonane	C_9H_{20}	15 minutes 26.5 seconds	2 hours 33 minutes 35.7 seconds
N-propylpropanamine	$C_6H_{15}N$	9 minutes 49.7 seconds	1 hours 16 minutes 29.8 seconds
octadecanoic acid	$C_{18}H_{36}O_2$	1 hours 55 minutes 15.0 seconds	5 hours 54 minutes 19.1 seconds
octane	C_8H_{18}	13 minutes 24.1 seconds	1 hours 49 minutes 23.9 seconds
octanoic acid	$C_8H_{16}O_2$	22 minutes 16.4 seconds	2 hours 14 minutes 18.3 seconds
oxacyclopentane	C_4H_8O	18 minutes 46.6 seconds	25 minutes 52.6 seconds
oxidane (water)	H_2O	13.0 seconds	29.0 seconds
pentan-3-one	$C_5H_{10}O$	1 hours 40 minutes 22.0 seconds	1 hours 20 minutes 11.0 seconds
pentane	C_5H_{12}	4 minutes 12.0 seconds	36 minutes 28.9 seconds
phenol	C_6H_6O	8 minutes 36.2 seconds	39 minutes 41.0 seconds
phenyl ethanoate	$C_8H_8O_2$	43 minutes 35.3 seconds	1 hours 40 minutes 33.5 seconds
phenylethene	C_8H_8	1 hours 23 minutes 30.6 seconds	1 hours 8 minutes 19.9 seconds
phenylmethanol	C_7H_8O	23 minutes 42.1 seconds	1 hours 3 minutes 42.4 seconds
phenylmethoxymethylbenzene	$C_{14}H_{14}O$	2 hours 49 minutes 2.5 seconds	6 hours 24 minutes 36.6 seconds
p-nonylphenoxyethanol	$C_{17}H_{28}O_2$	4 hours 12 minutes 51.4 seconds	5 hours 54 minutes 45.4 seconds
propanenitrile	C_3H_5N	1 minutes 27.0 seconds	9 minutes 37.0 seconds
propanone	C_3H_6O	8 minutes 12.0 seconds	19 minutes 29.0 seconds
propenonitrile	C_3H_3N	1 minutes 8.0 seconds	5 minutes 13.0 seconds
propylamine	C_3H_9N	10 minutes 19.0 seconds	32 minutes 1.0 seconds
pyridine	C_5H_5N	2 minutes 20.1 seconds	22 minutes 49.7 seconds
quinoline	C_9H_7N	14 minutes 25.0 seconds	2 hours 0 minutes 16.0 seconds
tetrachloroethene	C_2Cl_4	1 minutes 3.8 seconds	6 minutes 56.2 seconds
tetrachloromethane	CCl_4	38.0 seconds	5 minutes 24.0 seconds
tetranitrodibenzo-1,3a,4,6a-tetrazapentalene	$C_{12}H_4N_8O_8$	13 hours 53 minutes 46.5 seconds	1 days 10 hours 22 minutes 52.3 seconds

IUPAC nomenclature	Molecular Formula	CPU TIMES	
		OPT	FREQ
toluene	C_7H_8	4 minutes 57.6 seconds	49 minutes 23.1 seconds
trans-bicyclo[4.4.0]decane	$C_{10}H_{18}$	31 minutes 42.0 seconds	4 hours 24 minutes 0.9 seconds
tribromomethane	$CHBr_4$	1 minutes 9.0 seconds	14 minutes 49.0 seconds
trichloroethene	C_2HCl_3	1 minutes 38.2 seconds	8 minutes 31.0 seconds
trichlorofluoromethane	CCl_3F	54.5 seconds	6 minutes 49.9 seconds
trichloromethane	$CHCl_3$	33.0 seconds	4 minutes 3.0 seconds
triethyl phosphate	$C_6H_{15}O_4P$	2 hours 1 minutes 28.0 seconds	3 hours 38 minutes 44.2 seconds
trimethyl phosphate	$C_3H_9O_4P$	50 minutes 47.0 seconds	1 hours 23 minutes 40.0 seconds
trinitromethane	CHN_3O_6	1 hours 21 minutes 2.3 seconds	58 minutes 22.7 seconds

Appendix B: Training Set and Structure Variables for Species Used in QSAR1 and QSAR2

It should be noted that the molar volume, V_m, in this table is not the literature molar volume but rather a calculated molar volume from the regression in Table 2. The HSPs shown here are the literature values obtained from reference 4.

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
dimethyl sulfoxide	18.4	16.4	10.2	77.54	3.9039	0.67692	-0.2224	74.49	370.1573
methanamide	17.2	26.2	19	42.49	3.8323	1.08579	-0.25475	24.19	146.0894
acetonitrile	15.3	18	6.1	46.34	3.6498	0.42166	-0.3254	42.05	161.6442
1-methyl-2-pyrrolidone	18	12.3	7.2	101.42	3.7702	0.51374	-0.23342	103.99	728.9479
acetophenone	19.6	8.6	3.7	123.32	3.0018	0.40187	-0.24707	150.97	1237.2127
propanone	15.5	10.4	7	70.59	2.8181	0.36802	-0.24419	59.37	293.9463
oxidane (water)	15.5	16	42.3	28.29	2.069	1.32999	-0.2915	9.60	19.2
chloromethane	15.3	6.1	3.9	52.46	1.949	0.07602	-0.296	36.23	132.456
chlorobenzene	19	4.3	2	105.19	1.9341	0.03369	-0.24641	125.51	902.6719
trichloromethane	17.8	3.1	5.7	77.85	1.4013	0.08146	-0.31604	70.95	640.633
tribromomethane	21.4	4.1	6.1	85.56	1.1504	0.08599	-0.28099	107.73	1444.191
1,1,2,2-tetrabromoethane	22.6	5.1	8.2	117.46	0.5458	0.09293	-0.26844	160.51	2977.2654
2,2,4-trimethylpentane	14.1	0	0	156.23	0.4661	0.02262	-0.29376	148.47	1308.9261
methylcyclohexane	16	0	1	131.97	0.2641	0.02137	-0.28818	125.13	863.6494
2-methylbutane	13.7	0	0	106.31	0.1932	0.02200	-0.30895	94.29	566.9052
heptane	15.3	0	0	139.59	0.3282	0.02191	-0.29942	134.31	1711.535
nonane	15.7	0	0	172.86	0.4117	0.02190	-0.29265	173.69	3382.7661
benzene	18.4	0	2	92.49	0.1008	0.01652	-0.24629	102.94	458.2225
butane	14.1	0	0	89.68	0.19161	0.02196	-0.31762	76.70	427.376
1,3-dioxolan-2-one	19.4	21.7	5.1	73.31	5.2751	0.40194	-0.29404	65.56	460.0791

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
dimethyl sulfone	19	19.4	12.3	82.71	4.6138	0.57782	-0.28943	75.81	578.567
nitromethane	15.8	18.8	5.1	47.66	3.4781	0.34652	-0.29485	44.51	229.3976
4-methyl-1,3-dioxolan-2-one	20	18	4.1	89.95	5.4801	0.41627	-0.29061	84.07	704.7619
2-pyrrolidone	19.4	17.4	11.3	84.79	3.9274	0.83285	-0.23716	84.00	523.3845
propenonitrile	16	12.8	6.8	55.37	3.8834	0.42204	-0.28922	63.42	279.9072
butano-4-lactone	19	16.6	7.4	84.78	4.5015	0.41496	-0.26518	76.91	508.4396
trimethyl phosphate	16.7	15.9	10.2	124.59	3.3849	0.54809	-0.28222	105.71	1114.4989
2-aminoethanol	17	15.5	21.2	66.74	2.782	1.20894	-0.23398	57.56	361.9713
nitroethane	16	15.5	4.5	64.29	3.5934	0.33403	-0.29217	62.35	408.5659
furan-2-carbaldehyde	18.6	14.9	5.1	86.19	4.1695	0.34118	-0.24961	104.11	682.1196
diethyl sulfate	15.7	14.7	7.1	126.32	3.3649	0.53108	-0.29567	126.40	1622.4417
propanenitrile	15.3	14.3	5.5	62.98	3.901	0.44767	-0.32166	60.09	501.1152
N,N-dimethylmethanamide	17.4	13.7	11.3	75.76	3.832	0.45114	-0.24218	75.06	443.7825
bis(2-aminoethyl)amine	16.7	13.3	14.3	105.19	3.1558	1.04491	-0.21416	115.93	1548.8059
butanenitrile	15.3	12.4	5.1	79.62	4.0439	0.44996	-0.31886	79.41	587.8475
methanol	15.1	12.3	22.3	44.93	1.6952	0.99137	-0.26454	27.31	84.0942
2-nitropropane	16.2	12.1	4.1	80.93	3.6826	0.32963	-0.28935	79.06	563.2793
methanoic acid	14.3	11.9	16.6	42.48	3.8701	0.85316	-0.28742	31.00	134.5035
N,N-dimethylacetamide	16.8	11.5	10.2	92.40	3.6641	0.51174	-0.2338	92.49	624.6825
triethyl phosphate	16.7	11.4	9.2	162.21	3.0161	0.57835	-0.27294	163.14	2272.9746
benzyl n-butyl phthalate	19	11.3	3.1	283.76	2.438	0.41915	-0.24442	370.07	12438.84
dimethyl phthalate	18.6	10.8	4.9	164.48	0.4036	0.44132	-0.26017	204.50	2675.9393
chloromethyloxirane	18.9	7.6	6.6	83.29	3.4054	0.39048	-0.27989	76.98	615.8262
p-nonylphenoxyethanol	16.7	10.2	8.4	285.84	1.6643	0.99161	-0.20583	344.54	16330.2258
diethyl phthalate	17.6	9.6	4.5	197.76	0.9481	0.46125	-0.25758	242.08	3681.8645
2-(2-ethoxyethoxy)ethanol	16.1	9.2	12.2	138.45	1.5532	1.03811	-0.25044	137.23	2860.1985
2-ethoxyethanol	16.2	9.2	14.3	100.01	0.545	1.03825	-0.25097	90.68	967.167

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
2-methoxyethanol	16.2	9.2	16.4	83.37	1.4663	0.99445	-0.24792	71.56	544.3025
bis(2-chloroethyl) ether	18.8	9	5.7	120.23	0.7171	0.48550	-0.27964	120.79	2492.2329
benzonitrile	17.4	9	3.3	99.07	4.5624	0.40567	-0.26682	136.69	918.4615
butanone	16	9	5.1	87.23	2.6959	0.37638	-0.24346	77.27	492.1672
pyridine	19	8.8	5.9	81.03	2.1944	0.30260	-0.25258	94.79	434.0723
1,2-diaminoethane	16.6	8.8	17	66.75	2.202	1.08711	-0.21719	63.51	332.6017
ethanol	15.8	8.8	19.4	61.57	1.5675	1.00501	-0.26155	45.92	193.5658
nitrobenzene	20	8.6	4.1	100.39	4.5827	0.33397	-0.279	135.48	1117.4888
dibutyl phthalate	17.8	8.6	4.1	264.31	0.9529	0.44888	-0.25855	324.15	7137.8673
benzene-1,3-diol	18	8.4	21	102.83	1.3612	1.11131	-0.2123	120.84	912.9846
bis(2-chloroisopropyl) ether	19	8.2	5.1	153.51	1.2998	0.48666	-0.26999	155.17	2705.7551
ethyl 3-phenyl-2-propenonate	18.4	8.2	4.1	170.79	2.7836	0.46830	-0.23434	251.92	3905.4452
1,1,3,3-tetramethylurea	16.7	8.2	11	114.21	3.2205	0.52622	-0.22601	123.52	1073.4613
1,1-dichloroethane	16.5	7.8	3	81.79	2.2841	0.07753	-0.30404	72.67	531.7321
3,5,5-trimethylcyclohex-2-en-1-one	16.6	8.2	7.4	155.19	4.072	0.42893	-0.23097	166.04	1534.6009
4-hydroxy-4-methylpentan-2-one	15.8	8.2	10.8	125.67	2.8396	1.07093	-0.23606	119.26	1087.3152
acetaldehyde	14.7	12.5	7.9	53.95	2.6425	0.30970	-0.25523	42.06	167.6061
acetic acid	14.5	8	13.5	59.12	1.5812	0.93049	-0.27543	46.86	253.6915
2-(2-Methoxyethoxy)ethanol	16.2	7.8	12.6	121.81	1.6649	1.03810	-0.25178	117.32	2057.7486
1-chloropropane	16	7.8	2	85.73	2.3735	0.08826	-0.29042	75.15	550.5629
2-furanmethanol	17.4	7.6	15.1	93.80	1.7212	1.00213	-0.22533	95.91	743.6226
ethyl 2-hydroxypropanoate	16	7.6	12.5	114.20	1.5138	1.00959	-0.26651	109.89	1226.4035
pentan-3-one	15.8	7.6	4.7	103.87	2.5777	0.37323	-0.24248	95.59	795.0458
benzaldehyde	19.4	7.4	5.3	106.68	3.3048	0.32526	-0.25518	134.40	935.8264
1,2-dichloroethane	19	7.4	4.1	81.79	0.1412	0.08824	-0.30678	76.17	727.61
(chloromethyl)benzene	18.8	7.1	2.6	121.83	2.5685	0.09727	-0.25308	142.25	1298.774
ethyl methanoate	15.5	8.4	8.4	75.76	4.2487	0.38306	-0.27789	86.81	510.8863

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
methyl acetate	15.5	7.2	7.6	75.76	1.7847	0.42947	-0.26949	65.42	429.5196
quinoline	19.8	5.6	5.7	124.73	2.0254	0.40808	-0.23118	183.70	1250.647
benzoic acid	18.2	6.9	9.8	111.85	1.9228	0.99022	-0.26037	138.69	1178.2062
dioctyl phthalate	16.6	7	3.1	397.41	1.3629	0.43678	-0.25404	469.21	25664.5047
2-(2-butoxyethoxy)ethanol	16	7	10.6	171.73	1.5497	1.03815	-0.24994	176.45	5157.8992
1,2-dibromoethane	19.2	3.5	8.6	86.93	0.1908	0.12196	-0.28216	104.41	1538.7058
1,1-dichloroethene	16.4	5.2	2.4	74.18	1.5158	0.03457	-0.26627	73.36	487.9885
1-propanol	16	6.8	17.4	78.21	1.4974	1.01769	-0.26149	64.18	383.7464
tetrachloroethene	18.3	5.7	0	99.57	0.1335	0.00000	-0.26166	114.90	1295.0247
butyl (2R)-hydroxypropanoate	15.8	6.5	10.2	147.47	3.2554	0.99160	-0.26375	148.44	2676.08
1,2-dichlorobenzene	19.2	6.3	3.3	117.88	2.7892	0.02653	-0.2516	145.96	1298.2544
phenylmethanol	18.4	6.3	13.7	114.30	1.771	1.02738	-0.23467	131.47	993.644
dichloromethane	18.2	6.3	6.1	65.16	1.9757	0.07448	-0.30952	53.71	379.8284
cyclohexanone	17.8	6.3	5.1	112.89	3.1674	0.37200	-0.23433	107.68	750.0492
chlorodifluoromethane	12.3	6.3	5.7	73.53	1.479	0.17318	-0.33357	38.70	327.3132
4-methylpent-3-en-2-one	16.4	6.1	6.1	112.89	2.8532	0.46562	-0.2351	123.65	961.7145
2-propanol	15.8	6.1	16.4	78.21	1.6521	1.01999	-0.26038	63.92	319.1177
bis(2-methoxyethyl) ether	15.7	6.1	6.5	138.45	1.2988	0.42377	-0.25108	138.15	2856.6365
4-methylpentan-2-one	15.3	6.1	4.1	120.50	2.6897	0.39080	-0.24353	113.66	1004.7263
phenol	18	5.9	14.9	97.66	1.3592	1.10033	-0.21895	112.15	668.2477
1,1-dimethylhydrazine	15.3	5.9	11	66.75	0.4663	0.76471	-0.18899	65.66	305.9881
3-chloropropan-1-ol	17.5	5.7	14.7	90.90	3.4058	1.01784	-0.2755	81.48	868.457
bromochloromethane	17.3	5.7	3.5	67.72	1.7051	0.10176	-0.28962	65.29	534.6537
oxacyclopentane	16.8	5.7	8	87.23	1.8	0.42105	-0.23738	75.71	383.6015
butan-1-ol	16	5.7	15.8	94.84	1.5369	1.01956	-0.26112	83.03	670.7109
5-methylhexan-2-one	16	5.7	4.1	137.14	2.7532	0.38105	-0.24265	133.21	1555.7554
2-butanol	15.8	5.7	14.5	94.84	1.6465	1.03182	-0.26083	82.33	523.2595

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
2-methyl-1-propanol	15.1	5.7	15.9	94.84	1.5815	1.03451	-0.26193	81.91	517.6949
bromobenzene	20.5	5.5	4.1	107.76	1.8295	0.08339	-0.24192	138.59	1232.1253
chlorocyclohexane	17.3	5.5	2	128.03	2.7097	0.08783	-0.28319	124.91	1071.859
1-chlorobutane	16.2	5.5	2	102.37	2.458	0.08803	-0.28891	94.35	926.8829
ethyl ethanoate	15.8	5.3	7.2	92.40	1.9702	0.43149	-0.26663	84.83	721.9179
butanal	15.6	10.1	6.2	87.23	2.5092	0.30160	-0.25243	78.07	555.9574
aniline	19.4	5.1	10.2	97.67	1.7965	1.23558	-0.19822	121.52	686.297
1,1,2,2-tetrachloroethane	18.8	5.1	5.3	107.18	0.3791	0.06919	-0.30651	109.35	1346.8991
3-methylphenol	18	5.1	12.9	114.30	1.6351	1.10451	-0.21553	133.02	975.2173
2-butoxyethanol	16	5.1	12.3	133.29	0.6047	1.03831	-0.25035	129.25	2176.4616
morpholine	18.8	4.9	9.2	92.40	1.2747	0.66534	-0.20987	88.87	521.3311
propylamine	16.9	4.9	8.6	78.21	1.4973	1.03018	-0.22869	70.19	406.7962
2-octanol	16.1	4.9	11	161.39	1.3567	1.02319	-0.2604	159.54	2815.0888
2-ethoxyethyl acetate	15.9	4.7	10.6	130.84	2.927	0.44416	-0.25979	131.78	2332.475
butylamine	16.2	4.5	8	94.85	1.5503	1.03062	-0.2285	89.07	699.8184
1-pentanol	15.9	5.9	13.9	111.48	1.4852	1.01935	-0.26096	102.12	1083.4185
dibutyl sebacate	16.7	4.5	4.1	328.04	2.8204	0.42978	-0.26062	365.74	28826.8796
1,1,1-trichloroethane	16.8	4.3	2	94.49	2.2001	0.04038	-0.30879	89.39	788.7227
2-ethylbutan-1-ol	15.8	4.3	13.5	128.12	1.7786	1.02603	-0.25908	118.93	1046.2036
methoxybenzene	17.8	4.1	6.7	114.30	1.3189	0.45598	-0.21515	133.72	968.5391
cyclohexanol	17.4	4.1	13.5	120.50	1.7069	1.04430	-0.25673	112.40	812.2139
butanoic acid	14.9	4.1	10.6	92.40	1.4073	0.95585	-0.27428	83.49	765.1659
methyl (*Z*)-octadec-9-eneoate	14.5	3.9	3.7	334.34	1.3201	0.42230	-0.23297	380.90	28033.2811
isopropyl hexadecanoate	14.3	3.9	3.7	341.96	2.0062	0.43627	-0.26455	381.77	32237.0589
phenylmethoxymethylbenzene	19.6	3.4	5.2	200.30	1.1557	0.40164	-0.2369	257.15	4978.6927
2,6-dimethylheptan-4-one	16	3.7	4.1	170.42	2.2737	0.38363	-0.23651	169.49	2235.4482
butyl ethanoate	15.8	3.7	6.3	125.67	2.0483	0.43158	-0.26574	123.46	1743.5659

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
2-methylpropyl ethanoate	15.1	3.7	6.3	125.67	1.9516	0.42326	-0.26628	121.74	1393.7146
1-octanol	16	5	11.9	161.39	1.5551	1.01945	-0.26079	160.78	3293.7601
octadecanoic acid	16.3	3.3	5.5	325.32	1.5659	0.95712	-0.27332	361.67	30623.4809
2-ethylhexan-1-ol	15.9	3.3	11.8	161.39	1.7378	1.03216	-0.262	156.91	2104.4596
1,3-dimethylbutan-1-ol	15.4	3.3	12.3	128.12	1.3061	1.01922	-0.26027	119.26	1059.2858
octanoic acid	15.1	3.3	8.2	158.95	1.4119	0.95716	-0.27355	161.21	3652.5321
1-bromonaphthalene	20.3	3.1	4.1	151.46	1.7671	0.08192	-0.21831	223.76	2372.4633
1-(phenoxy)-3-[3-(phenoxy)phenoxy]benzene	19.6	3.1	5.1	316.10	2.1784	0.55931	-0.21212	464.85	16220.7806
trichloroethene	18	3.1	5.3	86.87	0.9553	0.04879	-0.2613	95.26	931.9411
cyclohexanamine	17.2	3.1	6.5	120.51	1.5626	1.04310	-0.22877	118.95	833.2777
1,1-thiobisethane	16.8	3.1	2	105.65	1.7067	0.02931	-0.2132	108.00	815.0563
diethyl carbonate	15.1	6.3	3.5	114.20	0.7399	0.44051	-0.27952	112.00	1475.1898
dichlorofluoromethane	15.8	3.1	5.7	75.69	1.4058	0.20306	-0.32221	55.53	475.9884
1-bromoethane	16.5	8.4	2.3	71.67	2.1801	0.12406	-0.26808	67.70	381.4604
3-methylbutyl ethanoate	15.3	3.1	7	142.31	2.1002	0.43681	-0.26587	141.35	2088.6388
1-tridecanol	16.2	3.1	9	244.58	1.5649	1.01939	-0.26073	260.55	11647.8941
(*Z*)-octadec-9-enoic acid	16	2.8	6.2	317.71	1.1631	0.96787	-0.23358	359.93	24466.5116
3,6,9-trioxa-(18*Z*)-heptacosan-1ol	16	3.1	8.4	435.48	1.5969	1.05058	-0.23338	501.45	69935.0318
2-methylpropyl 2-methylpropanote	15.1	2.9	5.9	158.95	2.0351	0.42379	-0.26296	158.74	2275.4177
ethoxyethane	14.5	2.9	5.1	94.84	1.0476	0.40007	-0.2481	84.78	632.4849
1-decanol	16	4.7	10	194.67	1.5757	1.01943	-0.26075	200.50	5817.3318
(*Z*)-octadec-9-en-1-ol	14.3	2.6	8	320.15	0.9059	1.00871	-0.23275	360.03	22639.5358
bromotrifluoromethane	9.6	2.4	0	86.64	0.302	0.04973	-0.31233	50.51	566.3108
diethylamine	14.9	2.3	6.1	94.85	1.1037	0.69586	-0.21477	90.72	671.9246
trichlorofluoromethane	15.3	2	0	88.39	0.0292	0.03571	-0.32523	72.38	731.3448
1,2,3,4-tetrahydronaphthalene	19.6	2	2.9	151.43	0.0189	0.02160	-0.22574	175.67	1442.0464

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
naphthalene	19.2	2	5.9	136.20	0.1569	0.01678	-0.21269	192.01	1291.6749
dichlorodifluoromethane	12.3	2	0	86.23	0.0558	0.03989	-0.3337	56.16	565.771
1,4-dioxacyclohexane	19	1.8	7.4	92.40	0.2076	0.38459	-0.23558	82.93	500.4187
furan	17.8	1.8	5.3	72.00	0.6375	0.30119	-0.22446	68.61	290.1752
1,2-dichlorotetrafluoroethane	12.6	1.8	0	123.94	0.1579	0.05151	-0.34256	79.01	1126.1404
1,1,2-trichlorotrifluoroethane	14.7	1.6	0	126.10	0.1102	0.05083	-0.33527	89.53	1299.3857
toluene	18	1.4	2	109.13	0.3678	0.03020	-0.23531	124.24	719.366
N-propylpropanamine	15.3	1.4	4.1	128.12	1.1378	0.72676	-0.21415	129.16	1625.3211
biphenyl	19.7	1	2	161.86	0.2435	0.01853	-0.22221	230.47	2312.2183
phenylethene	18.6	1	4.1	118.15	0.2317	0.02084	-0.22167	157.89	991.3716
1,2-dimethylbenzene	17.8	1	3.1	125.77	0.5785	0.02742	-0.22949	144.31	954.6567
1-methylnaphthalene	20.6	0.8	4.7	152.84	0.3578	0.02887	-0.20851	210.91	1613.8337
ethylbenzene	17.8	0.6	1.4	125.77	0.3087	0.02389	-0.23582	142.75	1043.6853
carbon disulfide	20.5	0	0.6	56.49	0.4104	0.00000	-0.27779	90.13	347.3342
cis-bicyclo[4.4.0]decane	18.8	0	0	174.27	0.4323	0.01743	-0.27573	172.53	1518.8489
1,3,5-trimethylbenzene	18	0	0.6	142.41	0.3562	0.02950	-0.22718	167.05	1372.873
trans-bicyclo[4.4.0]decane	18	0	0	174.27	0.4164	0.01920	-0.2742	173.57	1627.2903
tetrachloromethane	17.8	0	0.6	90.55	0.2421	0.00000	-0.32352	87.29	902.2646
cyclohexane	16.8	0	0.2	115.34	0.2712	0.01821	-0.29179	106.26	595.2914
icosane	16.5	0	0	355.88	0.8702	0.02189	-0.27977	394.46	33190.4532
hexadecane	16.3	0	0	289.33	0.7024	0.02189	-0.28223	313.79	17325.9855
dodecane	16	0	0	222.78	0.5344	0.02189	-0.28661	233.46	7582.0399
decane	15.7	0	0	189.50	0.4502	0.02190	-0.29023	192.60	4533.6833
octane	15.5	0	0	156.23	0.3659	0.02190	-0.29566	153.95	2451.2248
hexane	14.9	0	0	122.95	0.2814	0.02192	-0.3044	114.90	1145.3461
pentane	14.5	0	0	106.31	0.245	0.02195	-0.31087	95.63	722.8208

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
1-trifluoromethyl-1,2,2,3,3,4,4,5,5,6,6-unedecafluorocyclohexane	12.4	0	0	131.97	0.4479	0.08976	-0.35625	127.27	4052.2529
methyl benzoate	18.9	8.2	4.7	128.49	1.8755	0.44843	-0.25706	159.95	1623.5705
phenyl ethanoate	19.8	5.2	6.4	128.49	4.4985	0.48931	-0.25093	151.31	1451.8616

Appendix C: Training Set and Structure Variables for Species Used in QSAR3

It should be noted that the molar volume, V_m, in this table is not the literature molar volume but rather a calculated molar volume from the regression in Table 3. The HSPs shown here are the literature values obtained from reference 4.

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	*#C*	*#H*	*#N*	*#O*	*#S*	*#Cl*	*#Br*	V_m / cm^3 mol^{-1}	μ / Debye	*d* / charge	*I* / hartrees	α / Å^3
dimethyl sulfoxide	18.4	16.4	10.2	2	6	0	1	1	0	0	77.5441	3.9039	0.67692	-0.2224	74.4905
acetonitrile	15.3	18	6.1	2	3	1	0	0	0	0	46.3449	3.6498	0.42166	-0.3254	42.047
1-methyl-2-pyrrolidone	18	12.3	7.2	5	9	1	1	0	0	0	101.424	3.7702	0.51374	-0.2334	103.989
acetophenone	19.6	8.6	3.7	8	8	0	1	0	0	0	123.321	3.0018	0.40187	-0.2471	150.972
propanone	15.5	10.4	7	3	6	0	1	0	0	0	70.5913	2.8181	0.36802	-0.2442	59.3683
oxidane	15.5	16	42.3	0	2	0	1	0	0	0	28.2933	2.069	1.32999	-0.2915	9.59574
chlorobenzene	19	4.3	2	6	5	0	0	0	1	0	105.189	1.9341	0.03369	-0.2464	125.512
trichloromethane	17.8	3.1	5.7	1	1	0	0	0	3	0	77.851	1.4013	0.08146	-0.316	70.9478
benzene	18.4	0	2	6	6	0	0	0	0	0	92.4936	0.1008	0.01652	-0.2463	102.937
nitromethane	15.8	18.8	5.1	1	3	1	2	0	0	0	47.6554	3.4781	0.34652	-0.2949	44.507
4-methyl-1,3-dioxolan-2-one	20	18	4.1	4	6	0	3	0	0	0	89.9479	5.4801	0.41627	-0.2906	84.0749
butano-4-lactone	19	16.6	7.4	4	6	0	2	0	0	0	84.7811	4.5015	0.41496	-0.2652	76.911
N,N-dimethylmethanamide	17.4	13.7	11.3	3	7	1	1	0	0	0	75.7637	3.832	0.45114	-0.2422	75.0584
methanol	15.1	12.3	22.3	1	4	0	1	0	0	0	44.9308	1.6952	0.99137	-0.2645	27.3115
butanone	16	9	5.1	4	8	0	1	0	0	0	87.2289	2.6959	0.37638	-0.2435	77.2734
1,2-diaminoethane	16.6	8.8	17	2	8	2	0	0	0	0	66.7462	2.202	1.08711	-0.2172	63.5081
ethanol	15.8	8.8	19.4	2	6	0	1	0	0	0	61.5683	1.5675	1.00501	-0.2616	45.9189

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	*#C*	*#H*	*#N*	*#O*	*#S*	*#Cl*	*#Br*	V_m / cm^3 mol^{-1}	μ / Debye	*d* / charge	*I* / hartrees	α / Å^3
nitrobenzene	20	8.6	4.1	6	5	1	2	0	0	0	100.385	4.5827	0.33397	-0.279	135.481
acetic acid	14.5	8	13.5	2	4	0	2	0	0	0	59.1206	1.5812	0.93049	-0.2754	46.8561
1,2-dichloroethane	19	7.4	4.1	2	4	0	0	0	2	0	81.7929	0.1412	0.08824	-0.3068	76.1709
dichloromethane	18.2	6.3	6.1	1	2	0	0	0	2	0	65.1554	1.9757	0.07448	-0.3095	53.707
cyclohexanone	17.8	6.3	5.1	6	10	0	1	0	0	0	112.889	3.1674	0.372	-0.2343	107.685
2-propanol	15.8	6.1	16.4	3	8	0	1	0	0	0	78.2058	1.6521	1.01999	-0.2604	63.9234
phenol	18	5.9	14.9	6	6	0	1	0	0	0	97.6604	1.3592	1.10033	-0.219	112.152
oxacyclopentane	16.8	5.7	8	4	8	0	1	0	0	0	87.2289	1.8	0.42105	-0.2374	75.7101
bromobenzene	20.5	5.5	4.1	6	5	0	0	0	0	1	107.757	1.8295	0.08339	-0.2419	138.591
ethyl ethanoate	15.8	5.3	7.2	4	8	0	2	0	0	0	92.3956	1.9702	0.43149	-0.2666	84.8265
aniline	19.4	5.1	10.2	6	7	1	0	0	0	0	97.6659	1.7965	1.23558	-0.1982	121.523
3-methylphenol	18	5.1	12.9	7	8	0	1	0	0	0	114.298	1.6351	1.10451	-0.2155	133.018
1,1,1-trichloroethane	16.8	4.3	2	2	3	0	0	0	3	0	94.4886	2.2001	0.04038	-0.3088	89.3883
cyclohexanol	17.4	4.1	13.5	6	12	0	1	0	0	0	120.504	1.7069	1.0443	-0.2567	112.395
trichloroethene	18	3.1	5.3	2	1	0	0	0	3	0	86.874	0.9553	0.04879	-0.2613	95.2609
1,4-dioxacyclohexane	19	1.8	7.4	4	8	0	2	0	0	0	92.3956	0.2076	0.38459	-0.2356	82.9323
toluene	18	1.4	2	7	8	0	0	0	0	0	109.131	0.3678	0.0302	-0.2353	124.239
1,2-dimethylbenzene	17.8	1	3.1	8	10	0	0	0	0	0	125.769	0.5785	0.02742	-0.2295	144.311
carbon disulfide	20.5	0	0.6	1	0	0	0	2	0	0	56.4866	0.4104	9.6E-09	-0.2778	90.1255
hexane	14.9	0	0	6	14	0	0	0	0	0	122.952	0.2814	0.02192	-0.3044	114.895
1,3,5-trinitrotoluene	19.5	10	4.5	7	5	3	6	0	0	0	132.805	1.5037	0.3726	-0.3108	215.289
1,3-dinitrooxypropan-2-yl nitrate	16.2	17.8	5.9	3	5	3	9	0	0	0	112.214	2.3959	0.37794	-0.3334	158.571
trinitromethane	15.5	10.3	7.3	1	1	3	6	0	0	0	63.4383	1.996	0.36188	-0.3482	90.1046

IUPAC Nomenclature	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	*#C*	*#H*	*#N*	*#O*	*#S*	*#Cl*	*#Br*	V_m / cm^3 mol^{-1}	μ / Debye	*d* / charge	*I* / hartrees	α / Å^3
nitroethane	16	15.5	4.5	2	5	1	2	0	0	0	64.293	3.5934	0.33403	-0.2922	62.3524
1-nitropropane	16.6	12.3	5.5	3	7	1	2	0	0	0	80.9305	3.7349	0.31526	-0.2906	80.7281
2-nitropropane	16.2	12.1	4.1	3	7	1	2	0	0	0	80.9305	3.6825	0.32963	-0.2894	79.0555
4-nitrochlorobenzene	20	8.8	3.9	6	4	1	2	0	1	0	113.081	2.9345	0.3421	-0.278	164.702
4-nitrophenol	20.4	20.9	15.1	6	5	1	3	0	0	0	105.552	5.3422	1.08795	-0.2543	151.241
3,5-dinitrophenol	19.5	12.9	14.4	6	4	2	5	0	0	0	113.443	4.6426	1.10934	-0.2761	176.726
3-nitroaniline	21.2	18.7	10.3	6	6	2	2	0	0	0	105.557	5.6821	1.27339	-0.2256	155.089
1,2-dinitrobenzene	20.6	22.7	5.4	6	4	2	4	0	0	0	108.276	6.6758	0.31796	-0.2916	158.338
benzene-1,2-diol	20	11.3	21.8	6	6	0	2	0	0	0	102.827	2.5192	1.22295	-0.2066	120.123
benzene-1,3-diol	18	8.4	21	6	6	0	2	0	0	0	102.827	1.3612	1.11131	-0.2123	120.836
benzene-1,4-diol	21	10.2	27.2	6	6	0	2	0	0	0	102.827	0.1189	1.1142	-0.1989	121.603
4-nitrotoluene	20.1	9.6	3.9	7	7	1	2	0	0	0	117.023	5.2119	0.34092	-0.2706	160.958
2,4-dinitrotoluene	20	13.1	4.9	7	6	2	4	0	0	0	124.914	4.8537	0.3807	-0.2982	190.095
1-bromopropane	16.4	7.9	4.8	3	7	0	0	0	0	1	88.3028	2.2861	0.12227	-0.2672	87.7963

Appendix D: Structure Variables of Nitrated Unknowns

IUPAC Nomenclature	#C	#H	#N	#O	#Cl	V_m / cm^3 mol^{-1}	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
(*E*)1,3,5-Trinitro-2-[2-(2,4,6-trinitrophenyl)ethenyl]benzene	14	6	6	12	0	241.21	256.36	0.993	0.37668	-0.29459	487.886	15558.21
2,2',4,4',6,6'-hexanitroazobenzene	12	4	8	12	0	219.03	247.27	0.0719	0.41205	-0.2775	490.432	13476.36
tetranitrodibenzo-1,3a,4,6a-tetrazapentalene	12	4	8	8	0	196.59	221.33	4.794	0.33753	-0.28821	385.796	11142.67
1,3,6,8-tetranitrocarbazole	12	5	5	8	0	195.27	204.59	0.4336	1.49789	-0.28605	412.669	9604.23
2,4,6-trinitrophenylmethylnitramine	7	5	5	8	0	150.22	170.33	3.2055	0.43193	-0.30255	256.365	5034.97
2,4,6-trinitrophenylethylnitramine	8	7	5	8	0	166.79	187.00	4.409	0.46574	-0.2967	271.525	5381.35
2,4,6-trinitrophenol	6	3	3	7	0	124.64	132.73	1.7891	1.09962	-0.30272	212.508	3579.40
1,3,5-trinitroaniline	6	4	4	6	0	124.51	138.38	3.1233	1.46649	-0.2776	231.200	3619.48
1,3,5-triamino-2,4,6-trinitrobenzene	6	6	6	6	0	135.47	162.63	0.215	1.47604	-0.2659	273.549	4270.98
1,3-diamino-2,4,6-trinitrobenzene	6	5	5	6	0	129.99	150.50	2.6233	1.46860	-0.26463	257.277	3928.46
1,3,5-trinitrobenzene	6	3	3	6	0	119.03	126.25	0.1772	0.37561	-0.32819	196.557	3332.51
2,4,6-trinitrochlorobenzene	6	2	3	6	1	132.99	141.60	0.3013	0.37191	-0.3168	216.437	3882.47
3-hydroxyl-2,4,6-trinitrophenol	6	3	3	8	0	130.25	139.22	1.7182	1.14062	-0.29271	224.076	3822.26
2,4,6-trinitrobenzoic acid	7	3	3	8	0	139.26	146.07	1.2009	0.95232	-0.30926	226.267	4246.54
2,4,6-trinitroanisole	7	5	3	7	0	141.21	149.41	2.4685	0.55041	-0.30402	225.920	3983.32
3-methyl-2,4,6-trinitrotoluene	8	7	3	6	0	152.17	159.59	1.6875	0.37232	-0.29999	231.501	3934.40
3-nitrooxy-2,2-bis(nitrooxymethyl)propyl nitrate	5	8	4	12	0	164.28	190.07	0.6192	0.38313	-0.33269	229.001	6979.37
1,3,5,7-tetranitro-1,3,5,7-tetrazocane	4	8	8	8	0	139.63	186.15	3.8035	0.43886	-0.30898	221.069	4748.94
1,3,5-trinitro-1,3,5-triazinane	3	6	6	6	0	108.44	142.08	2.9441	0.46139	-0.30355	163.871	2762.93
benzene-1,3,5-triol	6	6	0	3	0	108.44	99.87	2.7166	1.11712	-0.21261	129.094	1186.07

IUPAC Nomenclature	#C	#H	#N	#O	#Cl	V_m / cm^3 mol^{-1}	V_m / cm^3 mol^{-1}	μ / Debye	d / charge	I / hartrees	α / Å^3	e_{se} / Å^2
2,4,6-trinitrobenzene-1,3,5-triol	6	3	3	9	0	135.86	145.70	2.5859	1.14472	-0.28219	235.368	4110.05
2,4,6-trinitro-1,3,5-triethoxybenzene	12	15	3	9	0	235.28	245.74	1.4262	0.56259	-0.26987	337.188	7830.15
2,4,6-trinitroethoxybenzene	8	7	3	7	0	157.78	166.08	2.8343	0.58435	-0.30711	247.746	4650.44

Appendix E: Predictions of training set HSPs by QSAR models 1 and 2 (data for Figures 7-12)

IUPAC Nomenclature	Literature HSPs			QSAR1 Predicted HSPs			QSAR2 Predicted HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
dimethyl sulfoxide	18.4	16.4	10.2	17.1	12.7	10.9	17.1	13.4	10.6
methanamide	17.2	26.2	19	16.2	13.9	16.8	15.9	20.1	23.3
acetonitrile	15.3	18	6.1	17.5	12.4	9.1	15.3	17.5	7.8
1-methyl-2-pyrrolidone	18	12.3	7.2	17.3	11.6	8.1	17.3	10.6	7.4
acetophenone	19.6	8.6	3.7	18.6	9.4	6.6	18.4	8.3	5.9
propanone	15.5	10.4	7	16.7	9.4	7.3	16.2	10.2	7.5
oxidane	15.5	16	42.3	15.9	10.3	20.2	15.3	17.7	42.1
chloromethane	15.3	6.1	3.9	16.7	7.1	4.4	15.2	7.1	5.5
chlorobenzene	19	4.3	2	18.5	6.4	2.7	18.6	3.4	2.6
trichloromethane	17.8	3.1	5.7	16.9	5.2	3.7	15.7	4.0	3.0
tribromomethane	21.4	4.1	6.1	18.6	5.1	4.3	18.4	4.1	3.6
1,1,2,2-tetrabromoethane	22.6	5.1	8.2	19.5	3.4	3.8	19.6	3.6	3.7
2,2,4-trimethylpentane	14.1	0	0	16.0	0.6	-0.2	16.2	0.8	0.7
methylcyclohexane	16	0	1	16.3	0.8	0.8	16.4	0.9	1.0
2-methylbutane	13.7	0	0	16.2	1.2	1.7	15.7	1.2	1.0
heptane	15.3	0	0	16.3	0.8	0.6	16.4	1.1	0.9
nonane	15.7	0	0	16.3	0.4	-0.4	16.7	1.1	0.8
benzene	18.4	0	2	17.8	1.8	2.8	18.0	1.3	2.0
butane	14.1	0	0	16.4	1.6	2.3	15.4	1.3	1.2
1,3-dioxolan-2-one	19.4	21.7	5.1	17.1	15.6	7.7	18.7	17.9	6.5
dimethyl sulfone	19	19.4	12.3	16.9	14.1	9.4	16.9	15.0	8.0
nitromethane	15.8	18.8	5.1	17.5	11.8	8.2	16.1	16.0	7.6
4-methyl-1,3-dioxolan-2-one	20	18	4.1	17.1	15.8	7.2	19.7	16.3	6.1
2-pyrrolidone	19.4	17.4	11.3	17.0	12.9	12.5	17.1	12.9	11.9
propenonitrile	16	12.8	6.8	18.2	13.0	9.1	17.3	16.8	7.4
butano-4-lactone	19	16.6	7.4	16.9	13.4	7.4	17.3	13.9	6.6
trimethyl phosphate	16.7	15.9	10.2	15.6	9.5	7.0	14.9	8.2	6.4

IUPAC Nomenclature	Literature HSPs			QSAR1 Predicted HSPs			QSAR2 Predicted HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
2-aminoethanol	17	15.5	21.2	16.3	11.0	17.5	17.3	11.2	19.2
nitroethane	16	15.5	4.5	17.5	11.7	7.4	16.5	13.6	6.3
furan-2-carbaldehyde	18.6	14.9	5.1	18.5	13.0	7.1	19.0	13.0	6.1
diethyl sulfate	15.7	14.7	7.1	16.8	9.9	7.3	16.1	9.0	6.2
propanenitrile	15.3	14.3	5.5	17.4	12.7	8.8	15.9	15.5	7.0
N,N-dimethylmethanamide	17.4	13.7	11.3	17.4	12.2	8.3	17.3	13.1	7.7
bis(2-aminoethyl)amine	16.7	13.3	14.3	17.3	11.1	14.6	17.8	9.3	14.2
butanenitrile	15.3	12.4	5.1	17.4	12.7	8.2	16.4	14.0	6.4
methanol	15.1	12.3	22.3	16.0	8.3	15.6	14.8	10.7	20.0
2-nitropropane	16.2	12.1	4.1	17.3	11.4	6.7	16.7	11.9	5.6
methanoic acid	14.3	11.9	16.6	16.8	13.7	14.2	15.2	19.9	15.9
N,N-dimethylacetamide	16.8	11.5	10.2	17.2	11.5	8.4	17.1	11.0	7.7
triethyl phosphate	16.7	11.4	9.2	16.3	8.2	6.5	16.1	7.1	6.2
benzyl n-butyl phthalate	19	11.3	3.1	20.0	5.8	3.0	18.7	5.8	5.0
dimethyl phthalate	18.6	10.8	4.9	18.6	2.2	5.8	18.2	5.5	5.5
chloromethyloxirane	18.9	7.6	6.6	16.9	10.7	7.2	16.2	10.9	6.2
p-nonylphenoxyethanol	16.7	10.2	8.4	17.5	4.5	9.5	15.9	3.5	8.8
diethyl phthalate	17.6	9.6	4.5	18.5	2.9	4.9	17.9	5.2	5.3
2-(2-ethoxyethoxy)ethanol	16.1	9.2	12.2	15.9	6.0	12.9	16.2	5.3	12.5
2-ethoxyethanol	16.2	9.2	14.3	15.8	4.2	14.1	15.9	5.0	14.1
2-methoxyethanol	16.2	9.2	16.4	16.0	6.8	14.2	15.8	6.6	14.4
bis(2-chloroethyl) ether	18.8	9	5.7	16.5	3.4	7.2	15.9	4.8	6.0
benzonitrile	17.4	9	3.3	19.7	14.2	7.9	20.5	14.0	6.3
butanone	16	9	5.1	16.6	8.7	6.7	16.3	8.4	6.6
pyridine	19	8.8	5.9	18.2	8.0	6.7	17.9	8.0	6.0
1,2-diaminoethane	16.6	8.8	17	16.7	9.5	16.2	17.2	9.6	17.4
ethanol	15.8	8.8	19.4	16.0	7.6	15.1	15.3	8.3	16.4
nitrobenzene	20	8.6	4.1	19.5	14.1	6.9	20.4	13.7	5.6
dibutyl phthalate	17.8	8.6	4.1	18.6	1.6	2.8	17.8	5.4	4.7
benzene-1,3-diol	18	8.4	21	17.5	7.0	15.6	18.1	6.0	15.2

IUPAC Nomenclature	Literature HSPs			QSAR1 Predicted HSPs			QSAR2 Predicted HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
bis(2-chloroisopropyl) ether	19	8.2	5.1	16.3	4.0	5.8	15.8	4.7	5.5
ethyl 3-phenyl-2-propenonate	18.4	8.2	4.1	21.3	9.1	7.1	20.3	7.7	6.4
1,1,3,3-tetramethylurea	16.7	8.2	11	17.5	10.0	7.9	17.2	8.6	7.3
1,1-dichloroethane	16.5	7.8	3	16.9	7.3	3.5	15.9	5.7	3.0
3,5,5-trimethylcyclohex-2-en-1-one	16.6	8.2	7.4	17.3	10.9	5.2	17.6	8.1	5.5
4-hydroxy-4-methylpentan-2-one	15.8	8.2	10.8	16.0	9.4	13.5	16.4	7.4	13.7
acetaldehyde	14.7	12.5	7.9	16.9	9.3	7.3	16.0	11.5	8.4
acetic acid	14.5	8	13.5	16.2	7.6	14.4	15.0	8.5	14.8
2-(2-Methoxyethoxy)ethanol	16.2	7.8	12.6	15.9	6.6	13.5	16.1	5.8	13.1
1-chloropropane	16	7.8	2	16.7	7.4	3.4	16.2	5.9	3.3
2-furanmethanol	17.4	7.6	15.1	16.7	7.5	14.2	16.9	6.8	14.0
ethyl 2-hydroxypropanoate	16	7.6	12.5	16.1	6.3	13.3	16.0	6.0	12.8
pentan-3-one	15.8	7.6	4.7	16.5	8.0	6.1	16.3	7.1	6.0
benzaldehyde	19.4	7.4	5.3	18.9	10.5	6.3	18.8	9.6	5.6
1,2-dichloroethane	19	7.4	4.1	16.8	2.0	3.7	16.0	3.0	3.1
(chloromethyl)benzene	18.8	7.1	2.6	18.3	7.7	2.8	18.5	5.2	3.5
ethyl methanoate	15.5	8.4	8.4	18.2	13.4	7.8	18.2	14.7	6.4
methyl acetate	15.5	7.2	7.6	16.6	6.8	7.8	15.6	7.3	7.2
quinoline	19.8	5.6	5.7	20.4	7.7	7.4	19.9	7.1	6.4
benzoic acid	18.2	6.9	9.8	18.1	8.1	14.0	18.1	7.6	12.5
dioctyl phthalate	16.6	7	3.1	17.0	0.7	-0.3	16.0	4.9	4.2
2-(2-butoxyethoxy)ethanol	16	7	10.6	15.8	5.4	12.0	16.2	4.9	11.5
1,2-dibromoethane	19.2	3.5	8.6	18.2	2.7	4.6	18.1	3.8	3.9
1,1-dichloroethene	16.4	5.2	2.4	17.4	5.7	3.5	17.0	3.4	2.8
1-propanol	16	6.8	17.4	15.9	7.0	14.6	15.5	7.0	14.9
tetrachloroethene	18.3	5.7	0	18.0	1.9	2.5	17.4	-0.2	-0.5
butyl (2R)-hydroxypropanoate	15.8	6.5	10.2	16.2	10.0	12.0	16.4	7.9	11.5
1,2-dichlorobenzene	19.2	6.3	3.3	18.9	8.4	2.4	19.0	4.5	2.2
phenylmethanol	18.4	6.3	13.7	17.5	7.5	14.1	17.6	6.6	13.3
dichloromethane	18.2	6.3	6.1	16.9	6.9	4.0	15.4	5.9	3.6
cyclohexanone	17.8	6.3	5.1	16.7	9.3	5.8	16.7	7.9	5.9

IUPAC Nomenclature	Literature HSPs			QSAR1 Predicted HSPs			QSAR2 Predicted HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
chlorodifluoromethane	12.3	6.3	5.7	15.2	5.0	4.3	13.5	5.8	5.2
4-methylpent-3-en-2-one	16.4	6.1	6.1	17.6	9.1	7.3	17.3	8.0	6.7
2-propanol	15.8	6.1	16.4	15.9	7.4	14.6	15.5	7.4	15.0
bis(2-methoxyethyl) ether	15.7	6.1	6.5	16.4	4.2	5.7	16.1	4.4	5.5
4-methylpentan-2-one	15.3	6.1	4.1	16.4	7.9	5.6	16.2	6.7	5.7
phenol	18	5.9	14.9	17.4	7.0	15.5	18.0	6.2	15.2
1,1-dimethylhydrazine	15.3	5.9	11	16.9	4.6	12.4	16.7	5.5	13.5
3-chloropropan-1-ol	17.5	5.7	14.7	16.2	11.6	14.2	16.0	10.9	13.8
bromochloromethane	17.3	5.7	3.5	17.3	6.4	4.5	16.5	5.6	4.0
oxacyclopentane	16.8	5.7	8	16.4	6.5	7.2	16.0	6.4	7.2
butan-1-ol	16	5.7	15.8	15.8	6.7	14.0	15.7	6.4	13.9
5-methylhexan-2-one	16	5.7	4.1	16.4	7.7	5.0	16.3	6.2	5.3
2-butanol	15.8	5.7	14.5	15.8	7.0	14.1	15.7	6.6	14.1
2-methyl-1-propanol	15.1	5.7	15.9	15.8	6.8	14.1	15.7	6.5	14.2
bromobenzene	20.5	5.5	4.1	19.0	6.4	3.5	19.2	4.4	3.8
chlorocyclohexane	17.3	5.5	2	16.8	7.2	1.8	16.9	4.8	2.6
1-chlorobutane	16.2	5.5	2	16.7	7.2	2.8	16.4	5.3	2.9
ethyl ethanoate	15.8	5.3	7.2	16.6	6.9	7.3	15.9	6.9	6.5
butanal	15.6	10.1	6.2	16.7	8.1	5.9	16.5	6.9	5.3
aniline	19.4	5.1	10.2	17.9	8.5	17.4	19.3	6.9	17.4
1,1,2,2-tetrachloroethane	18.8	5.1	5.3	17.1	2.1	2.7	16.5	2.4	2.4
3-methylphenol	18	5.1	12.9	17.5	7.3	15.0	18.0	6.1	14.6
2-butoxyethanol	16	5.1	12.3	15.7	3.7	13.0	16.1	4.7	12.7
morpholine	18.8	4.9	9.2	16.6	5.7	10.1	16.2	5.6	9.9
propylamine	16.9	4.9	8.6	16.2	7.2	14.9	16.4	7.0	15.6
2-octanol	16.1	4.9	11	15.7	4.8	11.8	15.9	5.0	11.6
2-ethoxyethyl acetate	15.9	4.7	10.6	16.7	8.6	6.2	16.4	7.3	5.7
butylamine	16.2	4.5	8	16.2	6.9	14.3	16.5	6.3	14.5
1-pentanol	15.9	5.9	13.9	15.8	6.2	13.4	15.8	5.9	13.2
dibutyl sebacate	16.7	4.5	4.1	15.5	5.7	2.1	15.9	4.8	3.7

IUPAC Nomenclature	Literature HSPs			QSAR1 Predicted HSPs			QSAR2 Predicted HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
1,1,1-trichloroethane	16.8	4.3	2	17.0	6.8	2.6	16.1	4.3	1.8
2-ethylbutan-1-ol	15.8	4.3	13.5	15.7	6.5	12.8	15.8	5.8	12.7
methoxybenzene	17.8	4.1	6.7	18.0	5.3	7.3	17.6	5.1	6.9
cyclohexanol	17.4	4.1	13.5	15.8	6.6	13.3	16.0	5.9	13.2
butanoic acid	14.9	4.1	10.6	16.1	6.4	13.4	15.5	6.4	12.8
methyl (*Z*)-octadec-9-eneoate	14.5	3.9	3.7	15.9	1.9	1.9	15.2	3.3	3.9
isopropyl hexadecanoate	14.3	3.9	3.7	15.2	3.6	2.1	15.8	4.7	3.7
phenylmethoxymethylbenzene	19.6	3.4	5.2	19.3	3.6	4.6	18.5	4.8	5.3
2,6-dimethylheptan-4-one	16	3.7	4.1	16.2	5.7	3.8	16.0	4.5	4.8
butyl ethanoate	15.8	3.7	6.3	16.5	6.4	6.1	16.1	5.9	5.6
2-methylpropyl ethanoate	15.1	3.7	6.3	16.4	6.0	5.9	16.0	5.7	5.6
1-octanol	16	5	11.9	15.7	5.4	11.8	16.0	5.2	11.5
octadecanoic acid	16.3	3.3	5.5	14.9	3.8	8.7	16.0	5.6	7.2
2-ethylhexan-1-ol	15.9	3.3	11.8	15.6	5.7	11.7	15.8	5.3	11.7
1,3-dimethylbutan-1-ol	15.4	3.3	12.3	15.6	5.3	12.7	15.7	5.2	12.6
octanoic acid	15.1	3.3	8.2	15.9	5.1	11.3	16.0	5.4	10.6
1-bromonaphthalene	20.3	3.1	4.1	21.2	6.1	2.9	20.7	3.7	4.3
1-(phenoxy)-3-[3-(phenoxy)phenoxy]benzene	19.6	3.1	5.1	23.3	6.3	5.3	19.3	5.7	6.8
trichloroethene	18	3.1	5.3	17.8	4.2	3.4	17.6	2.4	2.9
cyclohexanamine	17.2	3.1	6.5	16.2	6.4	13.5	16.5	5.5	13.5
1,1-thiobisethane	16.8	3.1	2	17.3	5.4	2.2	17.7	2.0	2.8
diethyl carbonate	15.1	6.3	3.5	16.5	3.4	6.7	15.9	4.8	5.8
dichlorofluoromethane	15.8	3.1	5.7	16.0	5.2	4.9	14.6	5.5	4.6
1-bromoethane	16.5	8.4	2.3	17.3	7.5	4.5	16.9	6.8	4.6
3-methylbutyl ethanoate	15.3	3.1	7	16.4	6.1	5.6	16.1	5.6	5.4
1-tridecanol	16.2	3.1	9	15.5	4.1	9.8	16.0	4.8	9.4
(*Z*)-octadec-9-enoic acid	16	2.8	6.2	15.6	2.6	8.5	15.2	4.0	7.6
3,6,9-trioxa-(18*Z*)-heptacosan-1ol	16	3.1	8.4	13.2	4.8	11.0	14.0	3.8	6.7
2-methylpropyl 2-methylpropanote	15.1	2.9	5.9	16.3	5.5	4.8	16.0	5.1	5.0
ethoxyethane	14.5	2.9	5.1	16.3	4.4	6.7	15.9	4.8	6.5

IUPAC Nomenclature	Literature HSPs			QSAR1 Predicted HSPs			QSAR2 Predicted HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
1-decanol	16	4.7	10	15.7	4.9	10.9	16.0	5.0	10.6
(*Z*)-octadec-9-en-1-ol	14.3	2.6	8	15.5	1.8	8.6	15.2	4.0	8.0
bromotrifluoromethane	9.6	2.4	0	14.9	1.5	2.2	14.6	1.8	3.0
diethylamine	14.9	2.3	6.1	16.5	5.2	10.4	16.2	5.4	10.1
trichlorofluoromethane	15.3	2	0	16.1	1.2	2.5	15.1	2.4	1.7
1,2,3,4-tetrahydronaphthalene	19.6	2	2.9	18.0	0.4	0.8	18.1	1.8	1.9
naphthalene	19.2	2	5.9	20.2	1.9	2.2	20.0	1.4	2.6
dichlorodifluoromethane	12.3	2	0	15.3	1.0	2.3	14.2	2.1	2.0
1,4-dioxacyclohexane	19	1.8	7.4	16.3	2.3	6.6	16.1	4.3	6.6
furan	17.8	1.8	5.3	17.0	3.9	6.6	16.9	4.5	7.1
1,2-dichlorotetrafluoroethane	12.6	1.8	0	13.9	-0.1	0.5	14.0	1.8	1.4
1,1,2-trichlorotrifluoroethane	14.7	1.6	0	14.4	-0.1	0.7	14.4	2.0	1.4
toluene	18	1.4	2	17.9	2.2	2.4	18.2	1.2	2.5
N-propylpropanamine	15.3	1.4	4.1	16.4	4.6	9.6	16.0	4.6	9.2
biphenyl	19.7	1	2	20.7	1.8	1.6	20.3	1.7	2.7
phenylethene	18.6	1	4.1	19.4	2.2	2.6	19.5	1.4	2.6
1,2-dimethylbenzene	17.8	1	3.1	18.0	2.3	1.8	18.3	0.8	2.3
1-methylnaphthalene	20.6	0.8	4.7	20.2	2.0	1.7	19.8	1.3	3.0
ethylbenzene	17.8	0.6	1.4	17.9	1.6	1.7	18.1	0.9	2.1
carbon disulfide	20.5	0	0.6	19.7	4.1	4.7	19.0	1.0	0.4
cis-bicyclo[4.4.0]decane	18.8	0	0	16.2	0.2	-0.8	16.5	0.4	0.6
1,3,5-trimethylbenzene	18	0	0.6	18.2	1.5	1.3	18.4	0.9	2.3
trans-bicyclo[4.4.0]decane	18	0	0	16.3	0.2	-0.8	16.6	0.5	0.7
tetrachloromethane	17.8	0	0.6	16.9	1.9	2.4	15.3	-0.3	-1.3
cyclohexane	16.8	0	0.2	16.3	1.2	1.3	16.2	0.8	0.9
icosane	16.5	0	0	15.1	-0.3	-3.3	16.9	1.7	1.5
hexadecane	16.3	0	0	15.8	-0.5	-2.9	17.0	1.4	1.0
dodecane	16	0	0	16.2	-0.1	-1.7	16.9	1.2	0.8
decane	15.7	0	0	16.2	0.2	-0.9	16.7	1.1	0.7
octane	15.5	0	0	16.3	0.6	0.1	16.6	1.1	0.8
hexane	14.9	0	0	16.3	1.1	1.2	16.1	1.1	0.9

IUPAC Nomenclature	Literature HSPs			QSAR1 Predicted HSPs			QSAR2 Predicted HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
pentane	14.5	0	0	16.3	1.3	1.7	15.8	1.2	1.0
1-trifluoromethyl-1,2,2,3,3,4,4,5,5,6,6-unedecafluorocyclohexane	12.4	0	0	16.2	1.7	2.0	15.8	3.4	2.1
methyl benzoate	18.9	8.2	4.7	18.6	6.7	7.1	18.2	6.6	6.1
phenyl ethanoate	19.8	5.2	6.4	18.3	13.1	7.3	19.0	11.2	6.4

Appendix F: Predictions of training set HSPs by QSAR3 (data for Figures 13-15)

IUPAC Nomenclature	Literature HSPs			QSAR3 HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
dimethyl sulfoxide	18.4	16.4	10.2	18.5	14.2	9.9
acetonitrile	15.3	18	6.1	15.8	16.8	8.4
1-methyl-2-pyrrolidone	18	12.3	7.2	17.6	10.0	7.2
acetophenone	19.6	8.6	3.7	19.0	6.4	5.6
propanone	15.5	10.4	7	16.9	10.6	7.3
oxidane	15.5	16	42.3	14.4	17.5	41.8
chlorobenzene	19	4.3	2	19.0	3.1	1.6
trichloromethane	17.8	3.1	5.7	17.5	4.5	4.1
benzene	18.4	0	2	19.7	2.2	3.1
nitromethane	15.8	18.8	5.1	16.2	17.3	7.4
4-methyl-1,3-dioxolan-2-one	20	18	4.1	18.8	20.3	5.0
butano-4-lactone	19	16.6	7.4	18.2	16.4	5.8
N,N-dimethylmethanamide	17.4	13.7	11.3	17.1	13.0	6.9
methanol	15.1	12.3	22.3	15.3	10.3	21.2
butanone	16	9	5.1	16.9	8.0	6.7
1,2-diaminoethane	16.6	8.8	17	15.6	7.6	17.3
ethanol	15.8	8.8	19.4	15.7	7.6	18.1
nitrobenzene	20	8.6	4.1	19.4	13.6	3.6
acetic acid	14.5	8	13.5	16.1	9.2	16.9
1,2-dichloroethane	19	7.4	4.1	18.0	3.8	4.9
dichloromethane	18.2	6.3	6.1	16.7	6.0	4.5
cyclohexanone	17.8	6.3	5.1	17.6	7.4	5.7

IUPAC Nomenclature	Literature HSPs			QSAR3 HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
2-propanol	15.8	6.1	16.4	15.8	6.5	16.8
phenol	18	5.9	14.9	18.7	6.3	16.5
oxacyclopentane	16.8	5.7	8	16.7	5.7	8.1
bromobenzene	20.5	5.5	4.1	19.4	6.1	2.2
ethyl ethanoate	15.8	5.3	7.2	16.5	6.6	7.8
aniline	19.4	5.1	10.2	18.7	6.2	17.8
3-methylphenol	18	5.1	12.9	18.8	6.1	16.1
1,1,1-trichloroethane	16.8	4.3	2	17.7	5.0	2.3
cyclohexanol	17.4	4.1	13.5	16.2	4.5	15.6
trichloroethene	18	3.1	5.3	19.0	4.3	3.1
1,4-dioxacyclohexane	19	1.8	7.4	17.6	6.3	8.6
toluene	18	1.4	2	18.4	1.1	2.7
1,2-dimethylbenzene	17.8	1	3.1	18.2	0.5	2.2
carbon disulfide	20.5	0	0.6	19.1	0.2	2.9
hexane	14.9	0	0	15.4	-0.3	2.8
1,3,5-trinitrotoluene	19.5	10	4.5	18.2	8.8	5.8
1,3-dinitrooxypropan-2-yl nitrate	16.2	17.8	5.9	17.0	14.5	5.7
trinitromethane	15.5	10.3	7.3	16.1	11.9	6.6
nitroethane	16	15.5	4.5	16.4	14.0	6.0
1-nitropropane	16.6	12.3	5.5	16.5	11.8	5.0
2-nitropropane	16.2	12.1	4.1	16.4	11.7	5.3
4-nitrochlorobenzene	20	8.8	3.9	19.5	8.2	4.7
4-nitrophenol	20.4	20.9	15.1	20.8	18.9	13.1
3,5-dinitrophenol	19.5	12.9	14.4	20.2	17.3	13.8

IUPAC Nomenclature	Literature HSPs			QSAR3 HSPs		
	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$	δ_D / MPa$^{1/2}$	δ_P / MPa$^{1/2}$	δ_H / MPa$^{1/2}$
3-nitroaniline	21.2	18.7	10.3	20.8	18.7	15.4
1,2-dinitrobenzene	20.6	22.7	5.4	21.4	21.5	1.9
benzene-1,2-diol	20	11.3	21.8	19.5	10.8	17.2
benzene-1,3-diol	18	8.4	21	19.1	7.8	16.5
benzene-1,4-diol	21	10.2	27.2	20.9	9.5	17.4
4-nitrotoluene	20.1	9.6	3.9	20.1	13.6	3.2
2,4-dinitrotoluene	20	13.1	4.9	19.8	13.5	3.9
1-bromopropane	16.4	7.9	4.8	17.0	7.4	3.4

VITA

Karl D. Kuklenz was born in Edina, Minnesota on August 31, 1983. He attended public school in Brooklyn Park, Minnesota until age fourteen. Karl graduated from The Colony High School in The Colony, Texas in the spring of 2002 and began his undergraduate studies at Sam Houston State University later that fall. In 2007, he graduated with a bachelor of science degree in forensic chemistry. Karl liked chemistry so much that he stayed on the Sam Houston State campus to earn his master of science degree in chemistry in the spring of 2009.

Mr. Kuklenz's many accomplishments include presentations at the American Chemical Society's south-west regional meetings in 2007 and 2008 as well as a presentation at the Texas Academy of Science meeting in 2008. At the time of this writing he has publications in *Industrial and Engineering Chemistry* as well as the *Journal of the Texas Academy of Science*. Mr. Kuklenz is proficient in several areas of chemistry having spent lots of time teaching physical chemistry laboratories as well as many undergraduate laboratories. He especially enjoyed expanding his knowledge of organic chemistry with help from the many accomplished professors and students of that field at SHSU.