ERDC TR-09-3

Engineer Research and Development Center

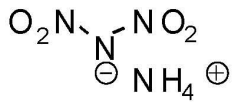
US Army Corps of Engineers®
Engineer Research and Development Center

Strategic Environmental Research and Development Program

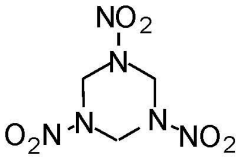
Sensitivity Analysis of QSAR Models for Assessing Novel Military Compounds

Jay L. Clausen, Erin Bennett, and Igor Linkov

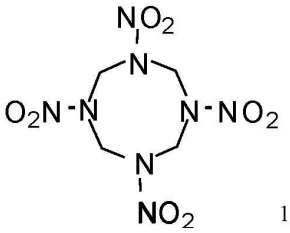
January 2009



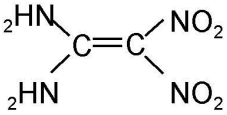
Ammonium dinitramide (ADN)



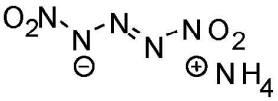
Cyclotrimethylenetrinitramine (RDX)



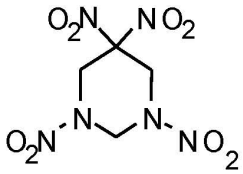
Cyclotetramethylenetetranitramine (HMX)



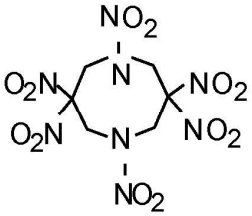
1,1-Diamino-2,2-dinitro-ethylene (FOX-7)



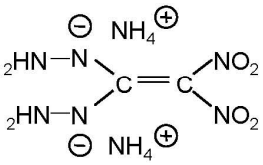
Ammonium di(nitramido) amine (ADNA)



Tetranitrohexahydropyrimidine (DNNC)



Hexanitrodiazacyclooctane (HCO)



Diammonium di(nitramido) dinitroethylene (ADNDNE)

Approved for public release; distribution is unlimited.

COVER: Compounds of interest and their analogs.

Strategic Environmental Research and Development Program

ERDC TR-09-3
January 2009

Sensitivity Analysis of QSAR Models for Assessing Novel Military Compound

Jay L. Clausen

Cold Regions Research and Engineering Laboratory
U.S. Army Engineer Research and Development Center
72 Lyme Road
Hanover, NH 03755-1290

Erin Bennett

Intertox Inc.
600 Stewart St. Suite 1101
Seattle, WA 98101

Igor Linkov

Environmental Laboratory
U.S. Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Approved for public release; distribution is unlimited.

Prepared for Strategic Environmental Research and Development Program (SERDP)
Arlington, VA 22203

Abstract: Reliable estimates of physical and biochemical properties of novel energetic compounds are essential before making the investment to synthesize, scale-up, and manufacture a new material for use in either military or civilian applications. Quantitative Structure-Activity Relationship (QSAR) software tools are available for predicting the physicochemical properties and environmental impacts of these emerging materials. The uncertainty and variability in melting point, solubility, half-lives, and related properties as a means of determining whether QSAR tools could provide meaningful results were evaluated. In particular, the octanol-water partition coefficient (K_{ow} or log P) was estimated for several proposed compounds. Log P was selected both because it typically can be measured with a high degree of certainty and because it correlates highly with water solubility and bioaccumulation. This study tested: 1) the variability in QSAR model predictions resulting from potential structural variants in emerging chemicals; and 2) the uncertainty from six different commercial K_{ow} calculators: KOWWIN, MarvinSketch, ACD/Labs, CLogP, SPARC, and ALOGPs. Analyses were performed on three military compounds [hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), butanetriol trinitrate (BTTN), and 2,4,6-trinitrotoluene (TNT)] and two pesticides [1-chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine (atrazine) and dichlorodiphenyldichloroethylene (DDE)]. Analyses of these compounds revealed that the uncertainty due to structural variations can be several orders of magnitude. Variability among the five software packages was as high as 10 orders of magnitude for emerging materials although lower for more well-studied chemicals such as DDE and atrazine. The magnitude of the uncertainty suggests use of existing QSAR models for emerging energetic materials is not appropriate.

DISCLAIMER: The contents of this report are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such commercial products. All product names and trademarks cited are the property of their respective owners. The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

DESTROY THIS REPORT WHEN NO LONGER NEEDED. DO NOT RETURN IT TO THE ORIGINATOR.

Contents

Figures and Tables iv

Preface vi

Nomenclature vii

1 Introduction 1

Estimation Program Interface (EPI) Suite model 2

Compounds previously evaluated 4

2 Objectives 6

3 Methods 7

Log P calculators 10

4 Results 11

Sensitivity analysis of EPA Suite 11

Comparison of log P calculators 12

Hexahydro-1,3,5-trinitro-1,3,5-triazine 13

2,4,6-Trinitrotoluene 15

Atrazine 17

Dichlorodiphenylethylene 19

Butanetriol trinitrate 21

5 Discussion 23

6 Conclusions 26

7 Recommendations 27

References 28

Appendix A 31

Report Documentation Page

Figures and Tables

Figures

Figure 1. Relationship between the polarity of a compound and its concentration in water and octanol. 4

Figure 2. Compounds of interest and their analogs in our previous work. 5

Figure 3. Compounds with the same molecular formula and similar structures to RDX based on a molecular formula search. 8

Figure 4. Compounds with similar structures to RDX based on a Tanimoto Algorithm search. 9

Figure 5. Comparison of experimental and predicted log P values. 13

Figure 6. Comparison of RDX-like log P predicted values for various calculators with RDX predictions. 14

Figure 7. Selection of RDX-like compounds using the Tanimoto Algorithm and simulated with log P calculators. 14

Figure 8. Comparison of TNT-like log P predicted values for various calculators with TNT predictions. 16

Figure 9. Chemical structure of 2,4,6-trinitrotoluene. 16

Figure 10. Comparison of atrazine-like log P predicted values for various calculators with atrazine predictions. 18

Figure 11. Chemical structure of 1-chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine (atrazine). 19

Figure 12. Comparison of DDE-135-like log P predicted values for various calculators with DDE predictions. 19

Figure 13. Chemical structure of dichlorodiphenylethylene. 20

Figure 14. Comparison of BTTN-like log P predicted values for various calculators with BTTN predictions. 21

Figure 15. Chemical structure of butanetriol trinitrate. 22

Tables

Table 1. Description of log P calculators used and their on-line availability. 10

Table 2: Compounds used for similarity analysis. 10

Table 3. Input parameters utilized by EPI Suite to make output predictions. 12

Table 4. Log P values for RDX and compounds with similar structures to RDX using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software. 15

Table 5. Log P values for TNT and compounds with similar structures to TNT using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software. 17

Table 6. Log P values for atrazine and compounds with similar structures to atrazine using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software. 18

Table 7. Log P values for DDE and compounds with similar structures to DDE using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software. 20

Table 8. Log P values for BTTN and compounds with similar structures to BTTN using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software. 21

Preface

This report was prepared by Jay L. Clausen, Biogeochemical Sciences Branch, U.S. Army Engineer Research and Development Center (ERDC), Cold Regions Research and Engineering Laboratory (CRREL), Hanover, NH; Erin Bennett, Intertox Inc., Salem, MA; and Igor Linkov, Environmental Risk Assessment Branch, ERDC, Environmental Laboratory (EL), Vicksburg, MS.

The Strategic Environmental Research and Development Program (SERDP) provided funding for this work. Dr. Charles M. Reynolds and Dr. David Ringleberg, CRREL, provided technical reviews.

This report was prepared under the general supervision of Terrence Sobecki, Chief, Biogeochemical Sciences Branch, CRREL; Dr. Justin B. Berman, Chief, Research and Engineering Division, CRREL; Dr. Lance D. Hansen, Deputy Director, CRREL; and Dr. Robert E. Davis, Director, CRREL.

The Commander and Executive Director of ERDC is COL Gary E. Johnston. The Director is Dr. James R. Houston.

Nomenclature

ADN	ammonium dinitramide
ADNA	ammonium di(nitramido)amine
ADNDNE	diammonium di(nitramido)dinitroethylene
AP	ammonium perchlorate
Atrazine	1-chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine
BP	boiling point
BTTN	butanetriol trinitrate
CAS	Chemical Abstract Service
DDE	dichlorodiphenyldichloroethylene
DNNC	1,3,5,5-tetranitrohexahydropyrimidine
EC_{50}	half-maximal effective concentration
EPI Suite	Estimation Program Interface Suite
FOX-7	1,1-diamino-2,2-dinitroethene
HCO	1,3,3,5,7,7-hexanitro-1,5-diazacyclooctane
HMX	octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine
K_{oc}	adsorption coefficient
LC_{50}	median lethal concentration
LOEC	lowest observed effect concentration
Log K_{ow}	octanol-water partition coefficient (log P)
Log P	octanol-water partition coefficient (K_{ow})
MP	melting point
QSAR	Quantum Structure Activity Relationship
RDX	hexahydro-1,3,5-trinitro-1,3,5-triazine
REACH	Registration, Evaluation, Authorization and Restriction of Chemicals
SERDP	Strategic Environmental Research and Development Program
SMILES	Simplified Molecular Input Line Entry System
TNT	2,4,6-trinitrotoluene
USEPA	U.S. Environmental Protection Agency
VP	vapor pressure

1 Introduction

The United States military is continuously performing research aimed at developing new energetic materials with greater safety, performance, and environmental acceptability. Reliable and early information regarding environmental toxicity, fate, and transport properties is essential before deciding to synthesize, scale-up, and manufacture a new explosive or propellant ingredient.

Environmental exposure assessment models are one set of tools for predicting a material's chemical/physical properties and environmental effects. For example, Quantitative Structure Activity Relationship (QSAR) screening level software predicts environmental effects using databases composed of properties of a wide range of compounds.

Log P, also known as the n-octanol/water partition coefficient (K_{ow}), is the ratio of the concentration of a chemical in n-octanol to that in water in a two-phase system at equilibrium. This chemical property has been shown to be one of the key parameters in QSAR studies, and is used to provide invaluable information in environmental (i.e., bioaccumulation, soil and sediment adsorption coefficients) and pharmacokinetic (i.e., adsorption from the gastrointestinal tract and biotransformation) assessments (Finizio and Sandroni 1997; Tetko et al. 2001; Benfenati et al. 2003; Linkov et al. 2005). Since log P is such a significant molecular descriptor, it is important that the measured or calculated value be as accurate as possible, especially if this value is used in conjunction with QSAR models. Typically, log P values are derived by analytical measurement or via computational methods. Several analytical methods, such as shake flask and chromatographic methods, are available for measuring a compound's log P but these methods can be time-consuming and unreliable, especially for compounds having high lipophilicity (Finizio and Sandroni 1997; Benfenati et al. 2003; Linkov et al. 2005; Gargadennec et al. 2005). The alternative to analytical measurement of log P is the use of calculators that predict log P based on a chemical's structural data. Computational methods may not be as accurate as measurement techniques, but may provide reasonable estimates, especially for assessments before chemical synthesis, formulation, and testing. Results from various log P calculators have been compared to measured values (training data sets) to assess their validity and accuracy

(Benfenati et al. 2003; Hilal et al. 2004; USEPA 2007a). Furthermore, various log P calculators have been compared to determine their validity and accuracy between programs relative to known measured values for specific sets of compounds (Mannhold and Petrauskas 2003; Zhou et al. 2005).

Estimation Program Interface (EPI) Suite model

A QSAR tool readily and publicly available is the U.S. Environmental Protection Agency's (USEPA) Windows-based software program Estimation Program Interface (EPI) Suite (USEPA 2007a). EPI Suite comprises individual chemical/physical estimating modules; each is designed to estimate a specific physical or chemical property of a given structure. The model comprises the following modules:

- KOWWIN: Estimates the log octanol-water partition coefficient, log K_{ow}, of chemicals using an atom/fragment contribution method;
- AOPWIN: Estimates the gas-phase reaction rate for the reaction between the most prevalent atmospheric oxidant, hydroxyl radicals, and a chemical. Gas-phase ozone radical reaction rates are also estimated for olefins and acetylenes. In addition, AOPWIN informs the user if nitrate radical reaction will be important. Atmospheric half-lives for each chemical are automatically calculated using assumed average hydroxyl radical and ozone concentrations;
- HENRYWIN: Calculates the Henry's Law constant (air/water partition coefficient) using both the group contribution and the bond contribution methods;
- MPBPWIN: Melting point, boiling point, and vapor pressure of organic chemicals are estimated using a combination of techniques;
- BIOWIN: Estimates aerobic biodegradability of organic chemicals using six different models; two of these are the original Biodegradation Probability Program (BPP);
- PCKOCWIN: The ability of a chemical to sorb to soil and sediment, the soil adsorption coefficient (K_{oc}), is estimated by this module. EPI's K_{oc} estimations are based on the Sabljic molecular connectivity method with improved correction factors;
- WSKOWWIN: Estimates an octanol-water partition coefficient using the algorithms in the KOWWIN program and estimates a chemical's water solubility from this value. This method uses correction factors to modify the water solubility estimate based on regression against log K_{ow};

- HYDROWIN: Acid- and base-catalyzed hydrolysis constants for specific organic classes are estimated by HYDROWIN. A chemical's hydrolytic half-life under typical environmental conditions is also determined. Neutral hydrolysis rates are currently not estimated;
- BCFWIN: This module calculates the bioconcentration factor and its logarithm from the log K_{ow}. The methodology is analogous to that for WSKOWWIN. Both are based on log K_{ow} and correction factors;
- WVOLWIN: Estimates the rate of volatilization of a chemical from rivers and lakes; calculates the half-life for these two processes from their rates. The model makes certain default assumptions such as water body depth, wind velocity, etc;
- STPWIN: Using several outputs from EPIWIN, this module predicts the removal of a chemical in a sewage treatment plant; values are given for the total removal and three contributing processes (biodegradation, sorption to sludge, and stripping to air) for a standard system and set of operating conditions; and
- LEV3EPI: This level III fugacity model predicts partitioning of chemicals between air, soil, sediment, and water under steady-state conditions for a default model "environment." Various defaults can be changed by the user.

EPI Suite runs from a single input [i.e., the chemical structure in Simplified Molecular Input Line Entry System (SMILES) notation]. The notation can be created and pasted into the input screen or obtained from a linked file of CAS numbers. The EPI Suite interface screen permits empirically derived physical data to be entered and has the capability of predicting a number of environmental properties including octanol-water partition coefficient (log K_{ow}, hereafter referred to as log P), gas-phase reaction rate, atmospheric half-lives, Henry's Law constant, melting point (MP), boiling point (BP), vapor pressure (VP), aerobic biodegradability, soil adsorption coefficient (K_{oc}), water solubility, hydrolytic half-life, bioconcentration factor, volatilization rate, removal efficiency in a sewage treatment plant, and aquatic toxicity values for selected species such as green algae, selected freshwater crustaceans (daphnids) such as *Daphnia* spp., or fish, including various exposure durations and endpoints.

This study placed particular emphasis on log P because of its importance for estimating fate and transport in the environment. Log P also has pharmokinetic importance because it is useful for estimating the distribution of drugs within the body. Log P is a measure of the partitioning of a

chemical between water (polar) and octanol (nonpolar), thereby describing the chemical's hydrophobic (lipophilic) nature (Figure 1). A higher log P indicates a more lipophilic (hydrophobic) nature and thus a greater tendency to partition into nonwater matrices. Octanol/water partition coefficients are generally presented using a log scale because the values range over several orders of magnitude.

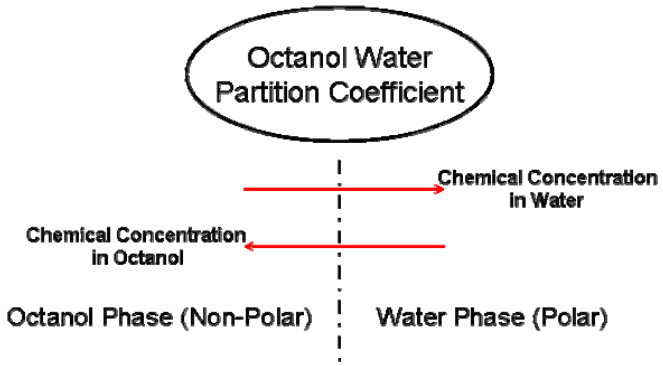


Figure 1. Relationship between the polarity of a compound and its concentration in water and octanol.

Compounds previously evaluated

One of the goals of Strategic Environmental Research and Development Program (SERDP) project PP-1403, "Synthesis, Evaluation, and Formulation Studies on New Oxidizers as Alternatives to Ammonium Perchlorate (AP) in Department of Defense (DoD) Missile Propulsion Applications," was to evaluate four possible replacement compounds: ammonium di(nitramido)amine (ADNA); 1,3,5,5-tetranitrohexahydropyrimidine (DNNC); 1,3,3,5,7,7-hexanitro-1,5-diazacyclooctane (HCO); and diammonium di(nitramido)dinitroethylene (ADNDNE). The focus of the evaluation was to determine whether the replacement compounds' environmental fate and transport and toxicity properties were more or less favorable than those for AP (Clausen et al. 2007). An important issue was the uncertainty of the model-predicted values for the four perchlorate-replacement compounds. Because these materials had not been synthesized, physical property data were not available for comparison. Consequently, existing compounds for which data were available that had structures similar to the proposed replacement compounds were evaluated as a means of assessing the accuracy of the predicted values. The existing energetic compounds evaluated previously (Clausen et al. 2007) with EPI Suite were: ammonium dinitramide (ADN), hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (HMX), 1,1-diamino-2,2-dinitroethene (FOX-7), and AP. In this earlier work, ADN

was used as an analog for ADNA; RDX as an analog for DNNC; HMX as an analog for HCO; and FOX-7 as an analog for ADNDNE (Figure 2). AP was also evaluated for a baseline comparison. The results from the Clausen et al. (2007) study suggested a large degree of uncertainty of the EPI Suite model output values.

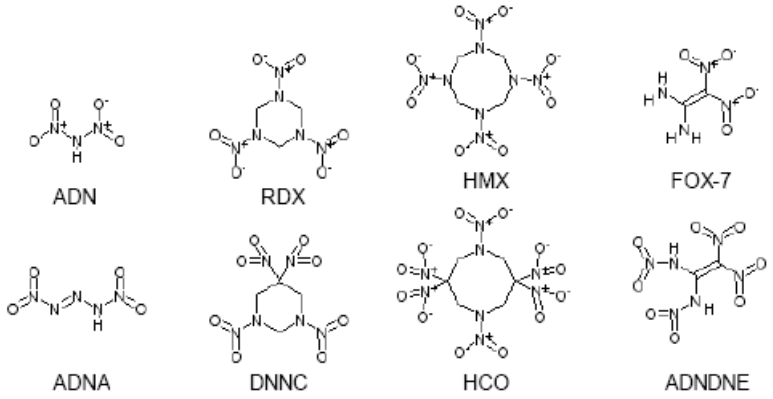


Figure 2. Compounds of interest and their analogs in our previous work.

2 Objectives

The previous study (Clausen et al. 2007) called into question the utility of using existing QSAR models such as EPI Suite to predict the physical and biochemical properties of novel energetic compounds. SERDP funded this study to verify the usefulness and suitability of EPI Suite, a readily available and commonly used set of QSAR models, as a predictive assessment tool to determine the environmental properties (log P, solubility, half-lives, etc.) for energetic materials. The objective was to conduct a sensitivity analysis of systematic changes in model input terms and their impact on model predictions. EPI Suite predictions are largely determined by a compound's log P or its chemical structure (molecular geometry and electronic structure). Therefore, this effort focused on these same two characteristics using the EPI Suite v3.2 QSAR model (USEPA 2007a).

The aim of this work is to develop a simple method for testing multiple log P calculators against a database of propellant-like or explosive-like compounds with limited physical chemical data to assess each calculator's applicability and potential range in predictions. As shown above, it is extremely important that a QSAR user has confidence in the predicted log P value before using it as input to QSAR models that predict the compound's environmental fate. If users have information regarding the reliability of the predicted values, this allows them to further assess the uncertainty associated with their QSAR model outputs. This is especially important for new explosive or propellant ingredients for use in military munitions. Furthermore, with the recent European Union implementation of a new regulation on Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), the use of QSARs to predict various physical-chemical properties of chemicals will become more widespread indicating the need for novel methods for decreasing the level of uncertainty in QSAR predictions.

3 Methods

EPI Suite runs from a single input: the chemical structure in SMILES notation. SMILES notation represents molecular structure by a linear string of symbols. The notation can be created via chemical drawing software and then pasted into the input screen, or obtained from a linked file of CAS (Chemical Abstract
System) numbers. The EPI Suite interface screen also permits entry of empirically derived physical data from the literature. Initially, a traditional sensitivity analysis of EPI Suite was attempted where one variable was selected and systematically varied for ADN, RDX, HMX, and FOX-7. The effort showed which variables affected which output parameters; however, a sensitivity analysis based on chemical structure within EPI Suite is not possible. Because of this limitation, predictions from several models for compounds with known physical properties were also compared. Customarily, measured log Ps are compared to values produced by the calculation programs. This comparison method works well when measurements are available.

Initially, compounds with the same molecular formulas and similar structures to our target compounds were identified. Compounds examined were: two explosives [RDX, 2,4,6-trinitrotoluene (TNT)], one propellant [butanetriol trinitrate (BTTN)], one pesticide [1-chloro-3-ethylamino-5-isopropylamino-2,4,6-triazeine (atrazine)], and one pesticide metabolite [dichlorodiphenyldichloroethylene (DDE)]. Both atrazine and DDE are essentially “controls” because they are both well-studied and commonly occur in the environment.

A similarity search based on molecular formula was performed using SciFinder of Chemical Abstract Service in Columbus, OH (SciFinder 2007). SciFinder is an internet-based tool for searching CAS databases. SciFinder identified compounds with structures similar to the target compound. The molecular formula similarity searches for RDX, TNT, atrazine, BTTN, and DDE required drawing chemical structures for each of the compounds using MarvinSketch v4.1.8 (ChemAxon 2007). Figure 3 shows the chemical structures identified by molecular formula to be similar to RDX. Outputs were limited to a maximum of 50, when

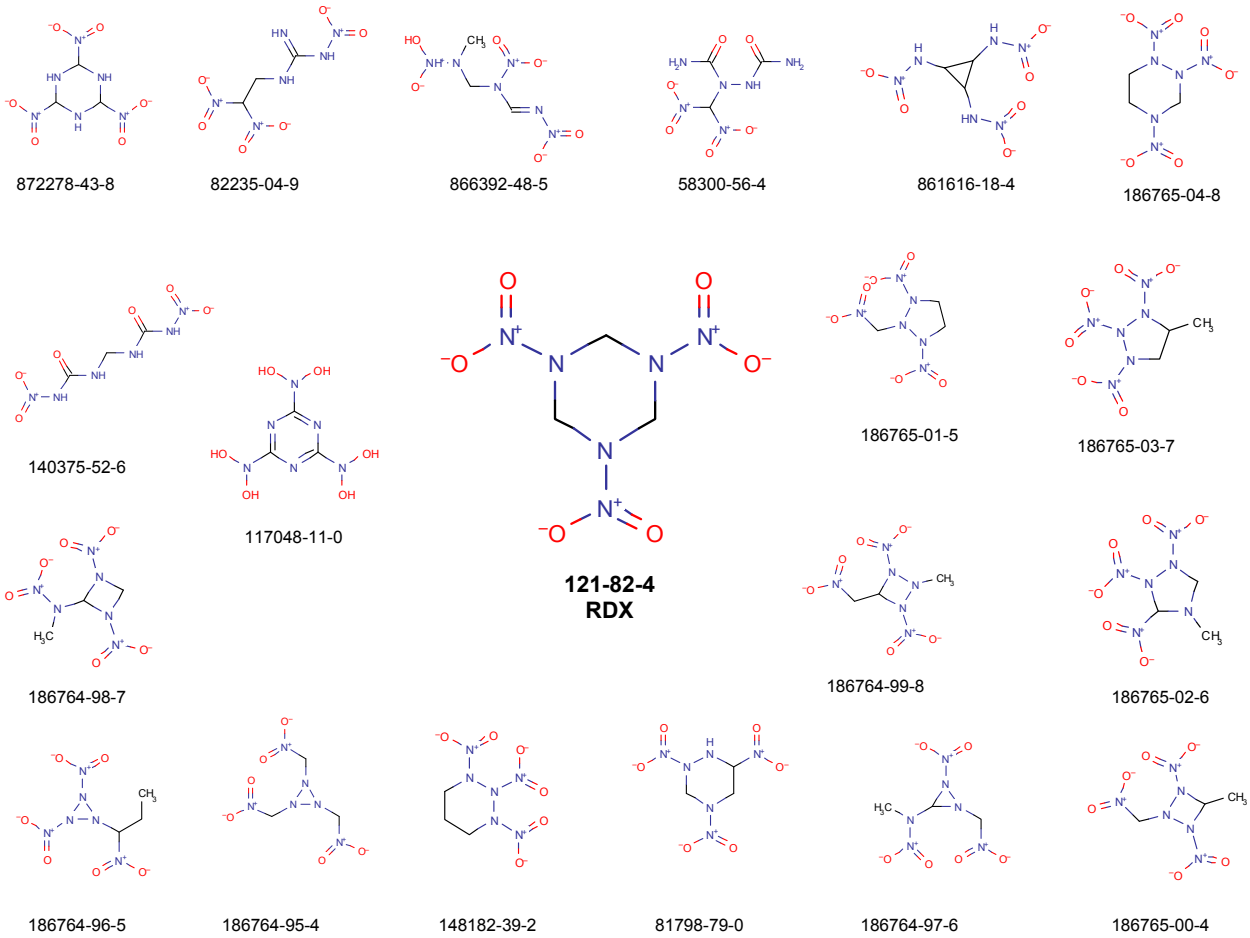


Figure 3. Compounds with the same molecular formula and similar structures to RDX based on a molecular formula search; CAS numbers included.

available, although none of the searches identified more than 19 similar compounds (Appendix A). None of the other compounds comparable to RDX shown in Figure 3 had experimental data available. Because of the lack of available data, it was decided to evaluate TNT and BTTN as part of this study.

A second similarity search was performed for RDX only. This search, based on the Tanimoto Algorithm (Willett et al. 1998), was also performed using SciFinder. The Tanimoto approach is based on the premise that "given an adequate set of descriptors, it is desirable to have a measure of similarity or alternatively a distance measure under which chemically equivalent molecules have a high level of similarity or small distance, and nonequivalent compounds have a low level of similarity or large distance. The most common measure of similarity amongst sets of molecular descriptors is the so-called "Tanimoto coefficient" (Karakoc et al. 2006). It

is possible with the Tanimoto Algorithm to assess not only chemical bond similarity between compounds but also their respective distance from each other within a trained data set. Performing structural similarity searches with the Tanimoto Algorithm is a standard approach used in drug discovery (Karakoc et al. 2006). In this case, RDX was drawn using the structure drawing software provided within the "Chemical Structure" window in SciFinder. Next, a similarity search using the Tanimoto Algorithm was performed to identify compounds with varying degrees of similarity. Compounds with greater than 87% similarity (within three standard deviations), as calculated with the algorithm, were selected for the log P comparisons. Compounds identified with the Tanimoto Algorithm are presented in Figure 4.

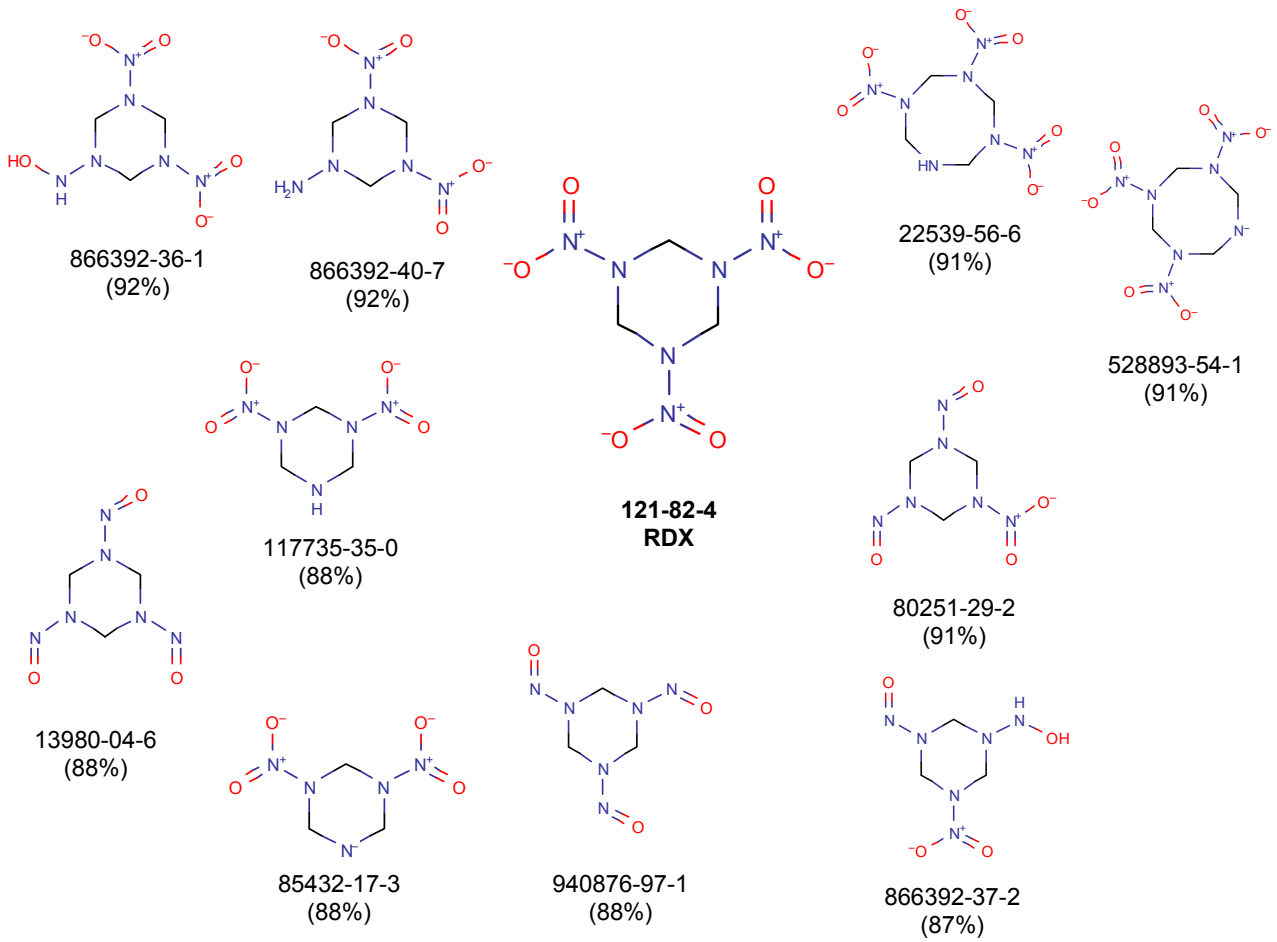


Figure 4. Compounds with similar structures to RDX based on a Tanimoto Algorithm search; CAS numbers and percent similarity to RDX are included.

Log P calculators

Six freeware programs (Table 1) were used to calculate log P values for RDX, TNT, BTTN, atrazine, and DDE (Table 2) and for the corresponding compounds identified from the SciFinder similarity search (Appendix A). The six programs were: KOWWIN v1.67, a subprogram within EPI Suites (USEPA 2007a); MarvinSketch v4.1.8 (ChemAxon 2007); ClogP v4.9 (Daylight Systems 2007); ACD/Chemsketch 10.2 (ACD/Labs 2006); SPARC v4.0 (SPARC 2007); and ALOGPs v2.1 (VCCLAB 2007). Log P calculations were performed either by entering an individual compound's SMILES string (ClogP, SPARC, KOWWIN) or its chemical structure (MarvinSketch, ACD/Labs). Calculated values were sorted, averaged, and plotted for each of the five compounds for comparison.

Table 1. Description of log P calculators used and their on-line availability.

Calculator	ACD/log P v10.2	ALOGPs v2.1	ClogP v4.9	KOWWIN v1.67	MarvinSketch/ log P v4.1.8	SPARC v4.0
Method	Fragment-based	E-state indexes	Fragment-based	Fragment-based	Atom contribution	Fragment-based
Training Set	> 10,000	> 10,000	> 10,000	2464	n/a	n/a
Validation Set	> 13,000	n/a	n/a	> 10,000	120 atoms from 893 compounds	n/a
Availability	Download	Online	Online*	Download	Online	Online
Data Entry	Structure	SMILES	SMILES	SMILES	Structure	SMILES
Reference	ACD/Labs 2006	Tetko et al. 2001	Daylight Systems 2007	USEPA 2007b	Viswanadhan et al. 1989	SPARC 2007
*As of June 2007, this program was no longer available free online; n/a, not available.						

Table 2. Compounds used for similarity analysis.

Chemical Name	IUPAC	RN (CAS)	Molecular Formula
RDX	1,3,5-Trinitroperhydro-1,3,5-triazine	121-82-4	$C_3H_6N_6O_6$
TNT	2-Methyl-1,3,5-trinitrobenzene	118-96-7	$C_7H_5N_3O_6$
BTTN	1,2,4- Butanetriol trinitrate	6659-60-5	$C_4H_7N_3O_9$
Atrazine	1-Chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine	001912-24-9	$C_8H_{14}ClN_5$
DDE	1,1-Bis-(4-chlorophenyl)-2,2-dichloroethene	72-55-9	$C_{14}H_8Cl_4$

4 Results

Sensitivity analysis of EPA Suite

The sensitivity analysis consisted of systematically entering measurements from the literature into EPA Suite to determine the relative impact on the predictions. Changing the MP had no effect on boiling point, log P, K_{oc}, Henry's Law constant, or half-lives for air, water, sediment, and soil for any of the compounds evaluated. Adjustments in MP also had no impact on predicted solubility for AP, ADN, and FOX-7 or on vapor pressure for HCO. The addition of MP as an input term changed the vapor pressure for RDX, but adding log P had no effect. Adding log P to MP did change the Daphnid LC_{50} and LOEC Daphnid EC_{50} RDX values. The addition of MP and log P had no effect on the boiling point and half-lives and only changed the K_{oc} and LOEC Daphnid EC_{50} values for HCO, Henry's Law constant for AP, and Daphnid LC_{50} and LOEC Daphnid EC_{50} values for HMX. Adding MP and log P changed the predicted solubility of RDX but not that of ADN, AP, HMX, or FOX-7 (Clausen et al. 2007).

These previous results suggest the EPI Suite subprograms rely primarily on chemical structure and to a lesser degree on log P, MP, and solubility to predict chemical and physical properties (Table 3). For example, changes to MP affect only three submodules in EPI Suite: ECOSAR, which predicts aquatic toxicity; MPMPWIN, which predicts the boiling point; and WSKOWWIN, which estimates the water solubility. However, there is no simple means within EPI Suite to conduct a sensitivity analysis on chemical structure.

In summary, this limited sensitivity analysis found the EPI Suite output terms such as boiling point, log P, K_{oc}, Henry's Law constant, and half-lives to be generally insensitive to changes in MP and log P input values. The results further indicated that increasing the MP above the SMILES-predicted MP value resulted in a decrease in solubility and vapor pressure while increasing the LOEC Daphnid EC_{50} values. Entering a lower, experimentally derived MP than the SMILES-predicted value had the opposite effect. Adding a higher K_{ow} value than predicted by SMILES alone resulted in a decrease in predicted solubility, Daphnid LC_{50}, and LOEC Daphnid EC_{50} values.

Table 3. Input parameters utilized by EPI Suite to make output predictions.

Model	SMILES	M.W.	log K_{ow}	WS	VP	HLC	MP	BP	Temp for VP	Function/Calculation
AEROWIN					√					Atmospheric particulate adsorption
AOPWIN	√									Atmospheric oxidation
BIOWIN	√									Biodegradation probability
BCFWIN	√		√							BCF
ECOSAR			√	√			√			SARS
HenryWIN	√			√	√					Henry's Law constant
HYDRWIN	√									Aqueous hydrolysis rate
KoaWIN	√		√			√				Octanol-air partition coefficient (K_{oa})
KowWIN	√									Log octanol-water partition coefficient
MPBPWIN	√						√	√	√	BP, MP, and VP
PCKOCWIN	√									Soil adsorption coefficient
WSKOWWIN	√		√				√			Water solubility 1
WATERNT	√									Water solubility 2
WVOLNET		√		√	√	√				Volatilization half-lives (aquatic)
Level III Fugacity	√		√		√	√	√	√		Mass balance (steady-state)

Comparison of log P calculators

Historically, the accuracy of QSAR programs, such as KOWWIN within EPI Suite, has been evaluated by comparing empirically derived data with the model predictions (Figure 5; USEPA 2007b). However, the utility of such analysis is questionable because the assessment is based on evaluation of the same data (trained data set) used to develop the software. Such an evaluation does little to assess the accuracy of model predictions for compounds not part of or similar to those used to develop the model. In this instance, the desire was to examine how well these programs predict the properties of compounds not part of a trained data set (RDX, HMX, TNT).

In the following discussion, atrazine and DDE are both known to be present in the trained data set. Although RDX, BTTN, and TNT are likely to be part of the trained data, they only have a limited number of compounds of similar structure for comparison. Our assessment employed two different approaches to find compounds similar in structure to our target compounds (see section 3). Next, log P results of the similar compounds were compared to the target compound. Each of the target compounds selected had existing measured log P values.

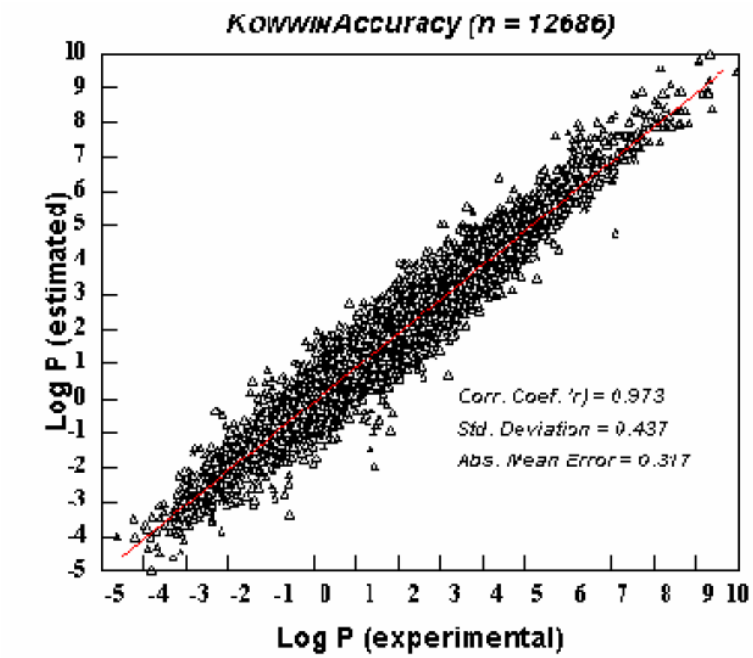


Figure 5. Comparison of experimental and predicted log P values (from USEPA 2007b).

Hexahydro-1,3,5-trinitro-1,3,5-triazine

For the first evaluation, RDX (Table 2) and compounds with a similar structure were evaluated with the various log P calculators (Figures 6 and 7; Appendix A). RDX has a chemical formula of $C_3H_6N_6O_6$ and is a heterocyclic nitramine also known as 1,3,5-trinitroperhydro-1,3,5-triazine. It is a six member heteroaliphatic ring with three nitro groups attached to nitrogen in the ring.

The log P of RDX is reported to be 0.87 (Sangster 1993) or 0.86 (Groom et al. 2001; Tomkins 2000). Compound 20 in Figure 6 is RDX and the log P calculators yielded predicted values ranging from -2.19 with ACD/Labs to 1.13 with Marvin (Table 4). ClogP yielded a value of 0.81, which was closest to the measured value. Log P values for those compounds similar in structure to RDX based on molecular formula yielded values ranging from -4.07 to 4.03 (Figure 6). The mean log P of the 19 RDX-like compounds (0.66) using Marvin came closest to predicting the actual value (Table 4). KowWin, which is a part of EPI Suite, yielded the poorest prediction (-2.36). The other log P calculators (SPARC, ACD/Labs, and ClogP) consistently yielded values less than the measured RDX values. Calculations for Compound 2 had the smallest variance and yielded a

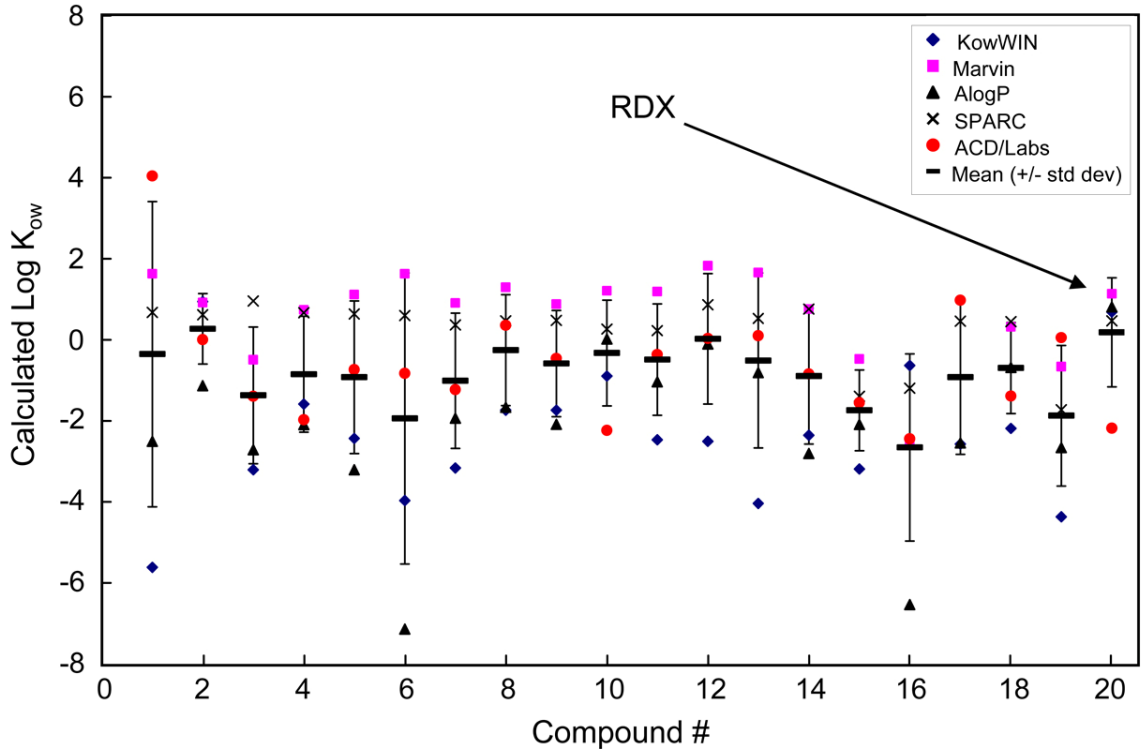


Figure 6. Comparison of RDX-like log P predicted values for various calculators with RDX predictions.

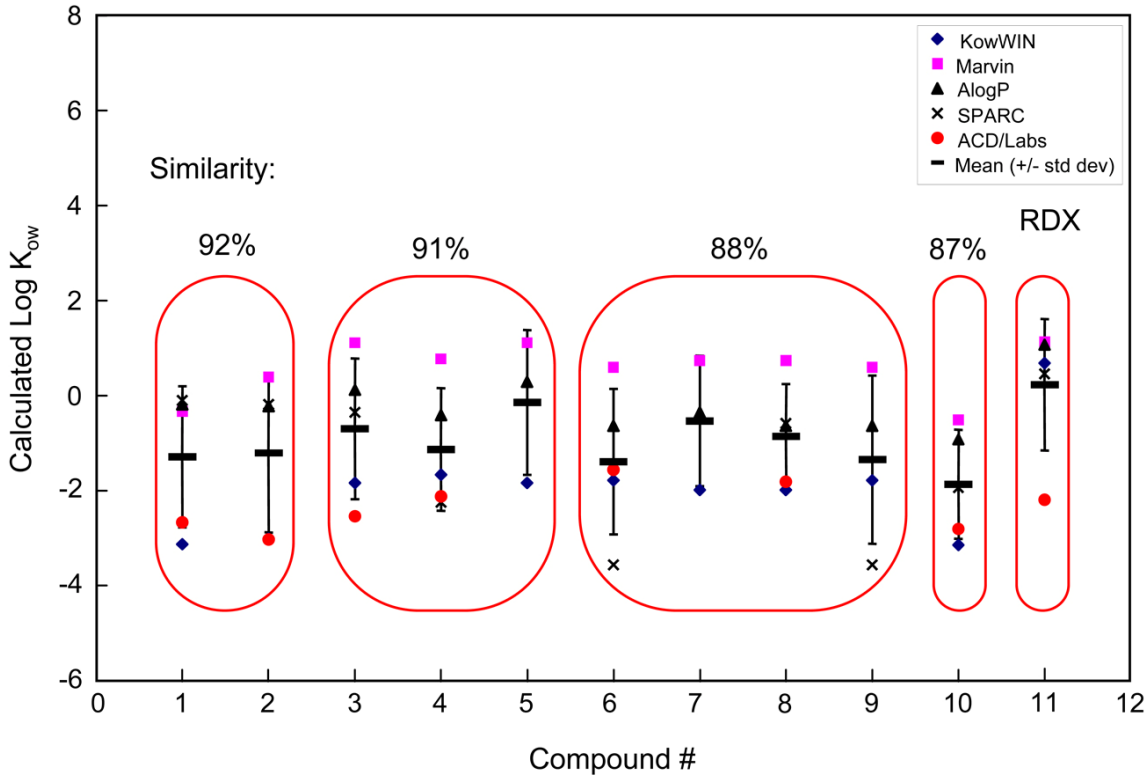


Figure 7. Selection of RDX-like compounds using the Tanimoto Algorithm and simulated with log P calculators (percent similarity is a calculation performed with the Tanimoto Algorithm).

Table 4. Log P values for RDX and compounds with similar structures to RDX using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software.

Compound	KowWIN	Marvin	ClogP	SPARC	ACD/Labs	Mean	Std Dev	Error
1	-5.62	1.62	-2.51	0.67	4.035	-0.36	3.76	1.68
2	0.933	0.91	-1.14	0.61	-0.003	0.26	0.87	0.39
3	-3.21	-0.50	-2.72	0.95	-1.40	-1.38	1.68	0.75
4	-1.59	0.73	-2.10	0.66	-1.98	-0.86	1.43	0.64
5	-2.44	1.11	-3.21	0.63	-0.74	-0.93	1.88	0.84
6	-3.97	1.62	-7.14	0.59	-0.83	-1.95	3.59	1.60
7	-3.17	0.90	-1.94	0.36	-1.23	-1.02	1.67	0.75
8	-1.74	1.29	-1.67	0.46	0.35	-0.26	1.37	0.61
9	-1.74	0.87	-2.09	0.47	-0.47	-0.59	1.31	0.58
10	-0.9	1.20	0.014	0.26	-2.24	-0.33	1.30	0.58
11	-2.47	1.18	-1.04	0.22	-0.37	-0.50	1.37	0.61
12	-2.51	1.82	-0.11	0.86	0.03	0.02	1.61	0.72
13	-4.04	1.65	-0.81	0.52	0.10	-0.52	2.16	0.97
14	-2.36	0.75	-2.81	0.75	-0.85	-0.90	1.68	0.75
15	-3.19	-0.48	-2.09	-1.40	-1.56	-1.74	1.00	0.45
16	-0.64	-2.48	-6.54	-1.20	-2.45	-2.66	2.31	1.03
17	-2.58	na	-2.55	0.45	0.97	-0.93	1.90	0.95
18	-2.19	0.31	-0.69	0.44	-1.40	-0.70	1.12	0.50
19	-4.37	-0.67	-2.66	-4.19	0.04	-2.37	2.01	0.90
20 (RDX)	0.68	1.13	0.81	0.46	-2.19	0.18	1.34	0.60
Mean	-2.52	0.66	-2.31	0.11	-0.53			
Actual	0.86/0.87	0.86/0.87	0.86/0.87	0.86/0.87	0.86/0.87			

na, not analyzed; Std Dev, standard deviation.

mean (0.26) closest to the value. As shown in Figure 7, Compound 2 or CAS 866392-40-7 had a 92% similarity with RDX based on the Tanimoto Algorithm. All of the calculators consistently underestimated the log P of RDX and appear to have underestimated the log P values for those compounds with similar structures to RDX when based on molecular formula.

2,4,6-Trinitrotoluene

TNT (Table 2) and compounds with a similar structure were evaluated next (Figure 8; Table 5). TNT is a poly-nitroaromatic consisting of symmetrical ring planar structure with a chemical formula of $C_7H_5N_3O_6$ (Figure 9). The six-carbon ring is stereochemically different than RDX, having three nitrogen groups attached to the electron-deficient aromatic ring with

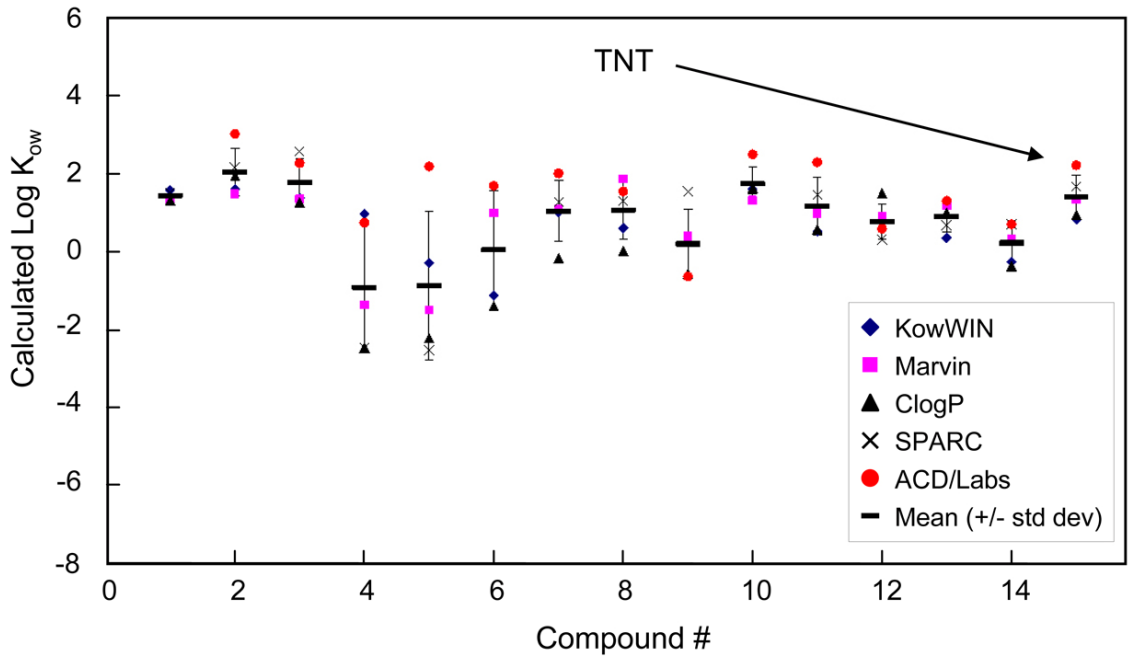


Figure 8. Comparison of TNT-like log P predicted values for various calculators with TNT predictions.

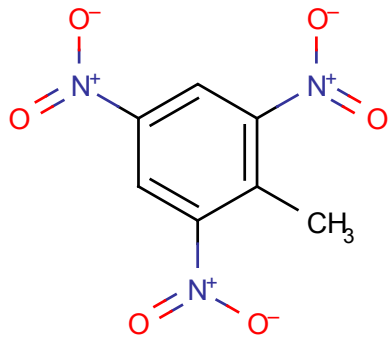


Figure 9. Chemical structure of 2,4,6-trinitrotoluene.

methyl group attached. The log P of TNT is reported to range from 1.6 to 1.99 (Hansch et al. 1995; Tomkins 2000; Groom et al. 2001; Sung et al. 2002). KowWIN, ClogP, and ACD/Labs yielded estimates similar to measured values (Table 5); both the Marvin and SPARC calculators produced over-estimates. The ACD/Labs average of 1.63 was closest to measured values. KowWIN tended to underestimate log P based on the average of the 19 compounds. Compounds 2, 3, and 10 yielded values closest to the measured value and showed the least difference between calculators (Figure 8).

Table 5. Log P values for TNT and compounds with similar structures to TNT using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software.

Compound	KowWIN	Marvin	ClogP	SPARC	ACD/Labs	Mean	Std Dev	Error
1	1.58	1.34	1.31	1.51	na	1.43	0.13	0.06
2	1.60	1.48	1.94	2.16	3.02	2.04	0.61	0.27
3	1.38	1.34	1.25	2.57	2.27	1.76	0.61	0.27
4	0.96	-1.38	-2.49	-2.47	0.74	-0.93	1.68	0.75
5	-0.29	-1.50	-2.23	-2.54	2.18	-0.88	1.92	0.86
6	-1.12	0.99	-1.41	na	1.68	0.04	1.53	0.77
7	1.00	1.09	-0.18	1.27	2.00	1.04	0.79	0.35
8	0.60	1.87	0.001	1.29	1.54	1.06	0.75	0.34
9	0.25	0.39	-0.58	1.54	-0.64	0.19	0.89	0.40
10*	1.60	1.31	1.61	1.64	2.49	1.73	0.45	0.20
11	0.51	0.97	0.57	1.45	2.29	1.16	0.74	0.33
12	0.58	0.89	1.49	0.31	0.58	0.77	0.45	0.20
13	0.34	1.18	0.99	0.67	1.30	0.89	0.39	0.17
14	-0.26	0.32	-0.38	0.70	0.69	0.21	0.51	0.23
15	0.82	1.34	0.94	1.66	2.21	1.39	0.57	0.25
16	0.47	0.32	0.52	0.12	1.05	0.50	0.35	0.16
17	1.99	2.37	1.79	2.63	1.65	2.09	0.41	0.18
18	1.99	2.37	1.87	2.80	2.12	2.23	0.37	0.16
19	1.99	2.37	1.79	2.91	2.12	2.24	0.43	0.19
20 (TNT)	1.99	2.37	1.71	2.34	1.68	2.02	0.33	0.15
Mean	0.84	1.00	0.46	1.12	1.63			
Actual	1.6-1.99	1.6-1.99	1.6-1.99	1.6-1.99	1.6-1.99			

*Compound 10 is a nonionic species; na, not analyzed; Std Dev, standard deviation.

Atrazine

Atrazine (Table 2) and compounds with similar structures were evaluated next (Figure 10; Table 6). Atrazine is an aromatic compound with the chemical formula $C_8H_{14}ClN_5$ (Figure 11). The log P of atrazine has been measured as 2.61 (Sangster 1993). Using the chemical formula for atrazine, 14 compounds with similar structures were identified (Figure 10; Appendix A). The various log P calculators yielded values ranging from -1.0 to 3.96 (Table 6). Results were similar for the atrazine-like compounds as well as consistent with the measured value (Figure 10). Compounds 1–7, 11, and 12, in particular, yielded log P values close to the 2.61 for atrazine. Results for compounds 8–10, 13, and 14 exhibited greater variance. Assessing the individual log P calculators, four out of five yielded predicted values for atrazine within 10% of the measured value (Table 6).

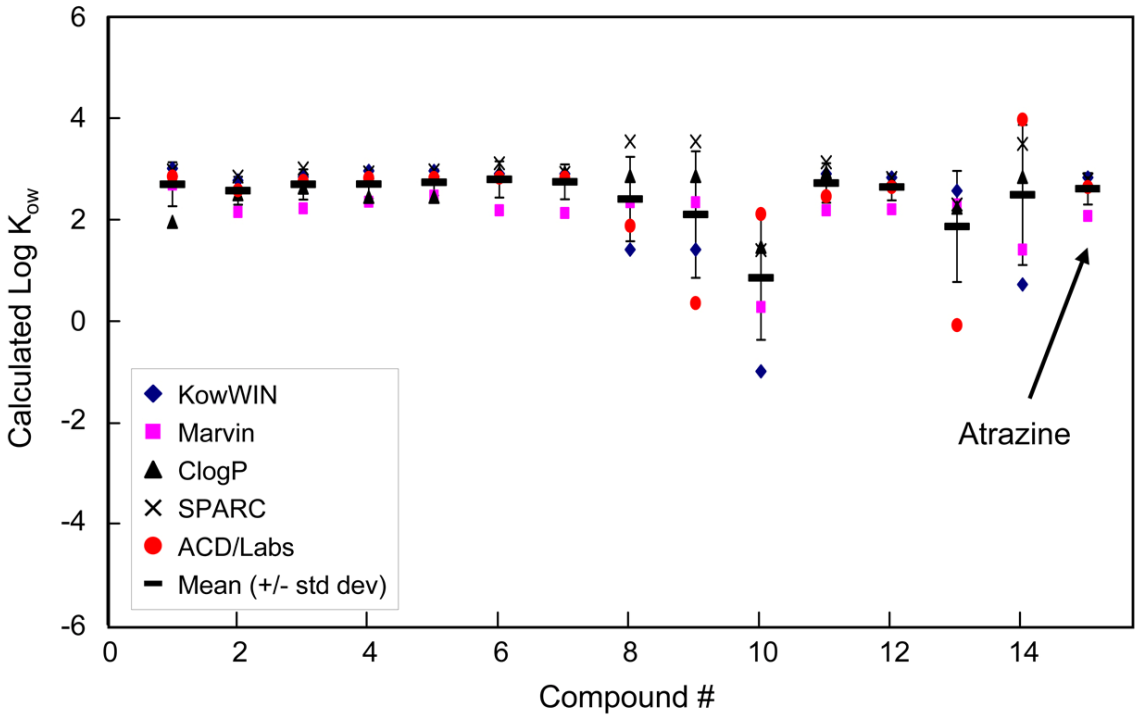


Figure 10. Comparison of atrazine-like log P predicted values for various calculators with atrazine predictions.

Table 6. Log P values for atrazine and compounds with similar structures to atrazine using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software.

Compound	KowWIN	Marvin	ClogP	SPARC	ACD/Labs	Mean	Std Dev	Error
1	3.00	2.68	1.94	2.96	2.84	2.69	0.43	0.19
2	2.76	2.14	2.49	2.84	2.57	2.56	0.27	0.12
3	2.83	2.21	2.62	3.00	2.75	2.68	0.30	0.13
4	2.95	2.34	2.44	2.92	2.81	2.69	0.28	0.13
5	2.95	2.46	2.44	2.96	2.81	2.72	0.26	0.12
6	2.89	2.17	2.92	3.10	2.82	2.78	0.36	0.16
7	2.89	2.12	2.92	2.91	2.82	2.73	0.34	0.15
8	1.40	2.33	2.85	3.53	1.87	2.39	0.83	0.37
9	1.40	2.33	2.85	3.53	0.34	2.09	1.25	0.56
10	-1.00	0.27	1.45	1.39	2.10	0.84	1.22	0.55
11	2.89	2.17	2.92	3.12	2.45	2.71	0.39	0.17
12	2.82	2.19	2.70	2.81	2.63	2.63	0.26	0.12
13	2.56	2.29	2.23	2.29	-0.09	1.85	1.09	0.49
14	0.71	1.40	2.83	3.48	3.96	2.48	1.38	0.62
15 (atrazine)	2.82	2.06	2.70	2.78	2.63	2.60	0.31	0.14
Mean	2.22	2.08	2.54	2.92	2.33			
Actual	2.61	2.61	2.61	2.61	2.61			
Std Dev, standard deviation.								

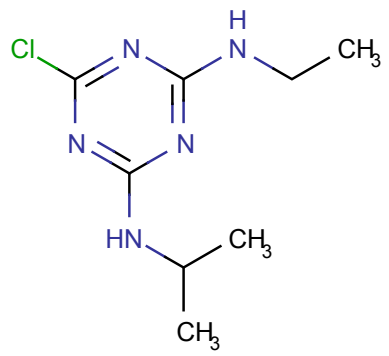


Figure 11. Chemical structure of 1-chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine (atrazine).

Dichlorodiphenylethylene

Log P results for DDE or 1,1-bis-(4-chlorophenyl)-2,2-dichlorethene (Table 2) and compounds with a similar structure are shown in Figure 12 and Table 7. DDE has a chemical formula of $C_{14}H_8Cl_{14}$ and is an organochlorine compound with two aromatic rings (Figure 13). The log P of DDE is reported to be 6.51 (Sangster 1993). Fifteen compounds with similar structures were identified. The various log P calculators yielded values ranging from 4.85 to 7.69. The SPARC and ACD/Labs results were closest to the measured value (Table 7). Out of the five log P calculators, the SPARC program result for the 14 compounds with similar structures

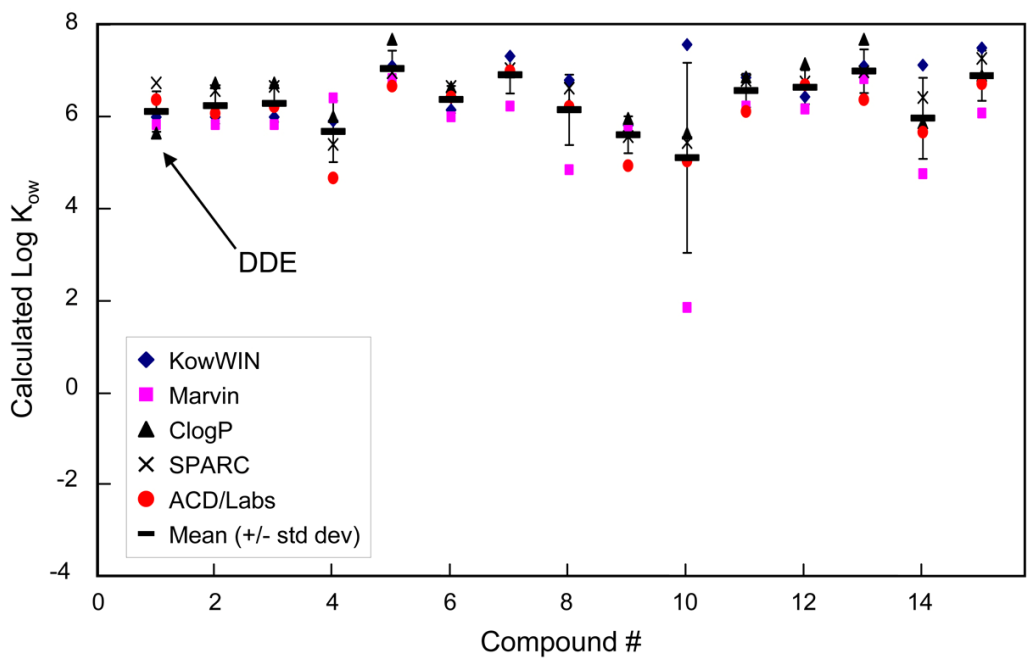


Figure 12. Comparison of DDE-135-like log P predicted values for various calculators with DDE predictions.

Table 7. Log P values for DDE and compounds with similar structures to DDE using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software.

Compound	KowWIN	Marvin	ClogP	SPARC	ACD/Labs	Mean	Std Dev	Error
1 (DDE)	6.00	5.83	5.65	6.74	6.37	6.12	0.44	0.20
2	6.00	5.83	6.74	6.57	6.08	6.24	0.39	0.17
3	6.00	5.83	6.74	6.66	6.22	6.29	0.40	0.18
4	5.91	6.41	6.01	5.40	4.67	5.68	0.67	0.30
5	7.10	6.83	7.69	6.96	6.66	7.05	0.39	0.18
6	6.15	6.00	6.65	6.66	6.43	6.38	0.30	0.13
7	7.32	6.23	6.98	7.05	7.00	6.91	0.41	0.18
8	6.80	4.85	6.27	6.62	6.22	6.15	0.77	0.34
9	5.82	5.74	5.97	5.57	4.94	5.61	0.40	0.18
10	7.57	1.85	5.65	5.44	5.04	5.11	2.07	0.92
11	6.85	6.23	6.87	6.79	6.12	6.57	0.37	0.16
12	6.43	6.17	7.16	6.76	6.69	6.64	0.37	0.17
13	7.10	6.83	7.69	6.98	6.37	6.99	0.48	0.21
14	7.13	4.76	5.87	6.42	5.67	5.97	0.88	0.39
15	7.50	6.08	6.90	7.27	6.72	6.89	0.55	0.24
Mean	6.69	5.69	6.65	6.51	6.06			
Actual	6.51	6.51	6.51	6.51	6.51			
Std Dev, standard deviation.								

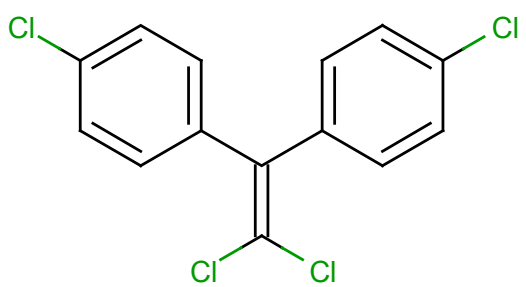


Figure 13. Chemical structure of dichlorodiphenylethylene.

yielded a mean value identical to the measured DDE value of 6.51 (Table 7). ClogP and KowWIN were closest for DDE itself, with greater discrepancies observed with Marvin and ACL/Labs. Overall, as shown in Figure 12, the majority of compounds identified as structurally similar yielded log P values close to that of DDE. The largest difference among the five calculators was in the predictions for Compound 10.

Butanetriol trinitrate

The final test evaluated BTTN or butane-1,2,4-triol (Table 2) and compounds with a similar structure (Figure 14; Table 8). BTTN has a chemical formula of $C_4H_7N_3O_9$ consisting of three nitrate groups attached to oxygen attached to butane (Figure 15). BTTN's structure is similar to nitroglycerin, a propellant, which has three nitrogen groups attached to a three-carbon structure. A search of the literature revealed no measured log P values for BTTN. The log P calculators yielded values ranging from 1.54 (Marvin) to 3.00 (ACD/Labs) with a mean value of 2.15 and standard deviation of 0.55 (Table 8). The mean log P for the three compounds identified to have a similar structure to BTTN ranged from 1.14 (ClogP) to 2.84 (ACD/Labs).

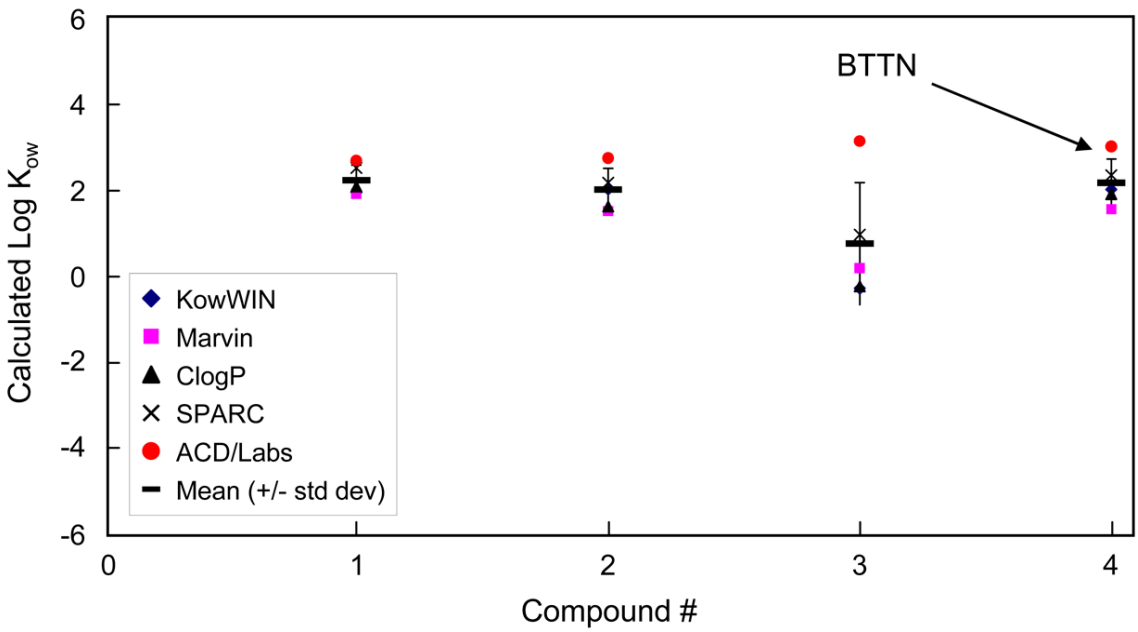


Figure 14. Comparison of BTTN-like log P predicted values for various calculators with BTTN predictions.

Table 8. Log P values for BTTN and compounds with similar structures to BTTN using KowWIN, Marvin, ClogP, SPARC, and ACD/Labs software.

Compound	KowWIN	Marvin	ClogP	SPARC	ACD/Labs	Mean	Std Dev	Error
1	1.93	1.90	2.07	2.50	2.67	2.21	0.35	0.16
2	2.00	1.50	1.60	2.16	2.73	2.00	0.49	0.22
3	-0.29	0.18	-0.24	0.95	3.12	0.74	1.42	0.63
4 (BTTN)	2.00	1.54	1.90	2.33	3.00	2.15	0.55	0.25
Mean	1.21	1.19	1.14	1.87	2.84			
Std Dev, standard deviation.								

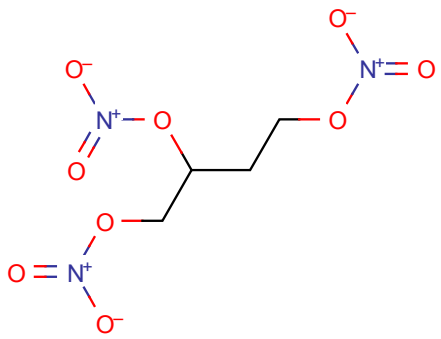


Figure 15. Chemical structure of butanetriol trinitrate.

5 Discussion

Current methods for testing the effectiveness of log P calculators involve comparing a chemical's measured log P with calculated values. This is a valid method when measured values are readily available. When dealing with compounds that have little to no physicochemical data and/or have not been synthesized, the current approach is not practical. For example, there are many military compounds, such as explosives and propellants, which have minimal chemical data, especially those in early stages of development (Clausen et al. 2007). Upfront reliable data for chemical properties, such as log P, would aid in estimating a chemical's environmental fate and toxicology when applied to QSAR modeling. Granted, QSAR models, such as EPI Suite, are available to provide screening-level predictions to help estimate chemical properties, but having some confidence in log P values before using this program would help reduce uncertainty in outputs that rely on this parameter.

The assessment of BTTN is analogous to what would occurs while evaluating a novel energetic compound because no measured log P values are available in the literature. Predictions for BTTN and its analogs averaged 1.65. However, the result for a given BTTN analog across all five calculators or an average of all analogs within an individual calculator resulted in differences of up to 1.5 log units. This discrepancy associated with BTTN displays the deficiencies of the models when dealing with compounds without known log P values. This amount of variation will be especially problematic for compounds not yet synthesized and thus no material is available for measuring log P.

Results for atrazine and DDE demonstrate that selecting the mean log P value of all calculators and analogs yields a value within 0.5 log units. In the case of atrazine, the measured log P value is 2.61 and the average for all calculators and analogs is 2.42. Similarly, the results for DDE indicate a measured value of 6.51 with a predicted value of 6.32 for all analogs and calculators. These data suggest that unknown compounds similar in structure to the trained data set in the QSAR database yield generally reliable predictions. The atrazine and DDE data demonstrate that assessing a compound with a single log P calculator might be acceptable. However, only by using a variety of calculators would an individual have an idea of

the variation in results. Reliance on a single log P calculator could result in significant error. Therefore, the use of several log P calculators is recommended if QSAR modeling is being considered for an unknown compound.

Problems arise, however, when modeling the more oxygenated compounds and compounds with more complex functional groups residing outside the training sets used for the log P calculators, as shown in the results for RDX and TNT. This problem has been similarly noted by Mullee (2007). Mullee (2007) measured the log P of 298 compounds and found KOWWIN predicted values within ±1 log unit of the measured value, but log P predictions for oxygen-rich substances were consistently underestimated.

Log P calculations for the RDX-like validation data set produced a broad range of log P predictions per calculator program for each compound. Of the 20 compounds in the validation set, the predicted range of values all had standard deviations greater than 1 log unit. Predicted log P value per compound ranged from a separation of 2 log units up to 8 log units.

Three of the five log P calculations for TNT were within ±0.5 log units of the measured value with one result (1.68; ACD/Labs) within the range of measured values (1.6–1.99). The other four log P estimates range from 0.5 to 1.5 log units from the measured value. Comparing the TNT and RDX results shows it is not possible to select one program as better than the others. Consequently, if dealing with a new energetic compound, it is not possible to know whether the predicted log P values are reasonable. Because EPI Suite and other QSAR models utilize structural relationships and log P for many of their computations, the problems identified for log P must apply to many of the other subprograms within EPI Suite (e.g. BCFWIN, ECOSAR, WSKOWIN, and Level III Fugacity). For this reason caution must be exercised when considering the use of QSAR models for the energetic compounds.

These results suggest that when faced with the need to predict properties of an unknown oxygenated energetic compound, a reasonable approach involves: 1) identifying compounds with structure similar to the target compound based on molecular formula, Tanimoto Algorithm, or some other method; 2) using several log P calculators to determine the log P of the target compound and the similarly structured compounds; and 3) analyzing the target compound and those closest in structural similarity with EPI Suite. If the log P calculators yield values within one log unit of each

other, the predicted values for the other properties such as solubility, bioconcentration factors, etc. are probably within an order of magnitude of the measured value based on the compounds tested. However, if the log P calculators yield values that vary by more than an order of magnitude, then results of any QSAR simulation such as with EPI Suite are qualitative. Over-reliance on a single log P calculator may result in an incorrect conclusion regarding the log P of an unknown compound. Therefore, it is recommended that if QSAR modeling is being considered for an unknown compound that several log P calculators be used and that the range of values be assessed.

Finally, these data suggest that other approaches be considered. For example, recent efforts have focused on modifications to the QSAR algorithms (Qasim 2007) or on the development of quantum models for environmental predictions of novel energetic compounds (Hurley 2007). Modification of CosmoTherm, an existing commercial software program (CosmoLogic Inc.), to determine the properties of log P, K_{oc}, solubility, and VP holds great promise. Recent work at the Army Research Laboratory comparing model-predicted values for RDX and 33 nitro-compounds yielded results within 0.5 log units of empirical measurements (Hurley 2007). One of the advantages of the quantum approach is the ability to utilize tuning parameters, i.e. additions to the trained data, as well as the ability to evaluate ionic compounds. QSAR models such as EPI Suite cannot handle ionic compounds such as perchlorate. Consequently, traditional QSAR models did not accurately predict the fate-and-transport behavior of perchlorate.

6 Conclusions

This work suggests the following conclusions:

- Log P calculations for compounds similar in structure to those used in the training data sets are reasonably accurate.
- Oxygenated compounds, such as those used by the United States military, are not well-predicted by the various calculators and there is no obvious approach for determining which of the calculators is more accurate.
- The uncertainty in predicting log P for oxygenated compounds is believed to be a consequence of the lack of data for compounds with similar structures.
- A variety of log P calculators should be compared when assessing a proposed compound. Only if the results are within an order of magnitude or less should the data be considered more than a guide.
- The log K_{ow} or log P calculators performed well for those compounds, similar in structure to the compounds in their respective training sets. However, the calculators yielded significantly different results for the more oxygenated compounds such as the military unique compounds. The high uncertainty of predicted values for the military compounds is due to the calculators lacking enough compounds with similar chemical structures for comparison. Consequently, existing QSAR models, including EPI Suite, are not appropriate for assessing energetic compounds.

It is established that EPI Suite and similar QSAR programs are only used as a tool for estimating environmental fate-and-transport values, but the work herein further illustrates the importance of reducing the uncertainty of predicted values before using them for this type of environmental model. This is especially important with the recent implementation of REACH in Europe. Because of the abundance of chemicals to be registered under REACH—both new chemicals and those in current use—there has been a shift from using expensive conventional animal testing to less expensive computational models and/or read-across approaches for assessing aquatic toxicity (Pavan et al. 2006; Lilienblum et al. 2008). It seems likely that a similar shift away from physical testing to computational models will occur in the United States over the next decade. Since log P is directly related or utilized in these prediction models, more emphasis on reducing the uncertainty of predicted log P values is extremely important.

7 Recommendations

- Develop an internal protocol for the use of EPI Suite by addressing the importance of using multiple log P calculators to improve the predicted outputs from QSAR simulations.
- Conduct further research to assess if one or more of the calculators used in this study (and/or additional calculators) are more applicable for highly oxygenated compounds with limited physicochemical properties. This can be assessed by reviewing individual training sets to determine whether these training sets contain this class of compound. Furthermore, the need exists for a detailed description of the methodologies each calculator program uses to derive log P values. For example, both ClogP and KOWWIN use a substructure approach but their methodologies differ by use of a constructionist and reductionist approach, respectively (Mannhold and Petrauskas 2003).
- Use predicted data calculated by each log P calculator for each of the five selected compounds and input the high, low, and average values into EPI Suite to evaluate BCFWIN (BCF), ECOSAR (aquatic toxicity estimations), KoaWIN (log K_{oa}), and WSKOWWIN (water solubility) outputs. These data ranges could be used to calculate or estimate the uncertainty for each of the predicted values. Data produced for this exercise could be used as an example data set for the internal protocol proposed above.
- Support efforts into the development of quantum models for predicting environmental properties of novel energetic compounds.

References

ACD/Labs. 2006. Chemsketch 10.2. Toronto, Ontario, Canada: Advanced Chemistry Development. http://www.acdlabs.com

Benfenati, E., G. Gini, N. Piclin, A. Roncaglioni, and M. R. Varì. 2003. Predicting log P of pesticides using different software. *Chemosphere*. 53:1155–1164.

ChemAxon. 2007. MarvinSketch 4.1.8. ChemAxon Kft. Budapest, Hungary.

Clausen, J. L., R. Cramer, S. Clough, P. Gwinn, and M. Gray. 2007. *Environmental screening assessment of perchlorate replacements*. ERDC/CRREL TR 07-12. Hanover, NH: Cold Regions Research and Engineering Laboratory, Environment Research and Development Center.

Daylight Systems. 2007. *ClogP Daylight Version 4.9*. Aliso Viejo, CA: Daylight Chemical Information Systems Inc.

Finizio, A., and V. D. Sandroni. 1997. Determination of n-octanol/water partition coefficient (Kow) of pesticide critical review and comparison of methods. *Chemopshere*. 34(1):131–161.

Gargadennec, S., G. Burgot, J.-L. Burgot, R. Mannhold, and R. F. Rekker. 2005. log P estimation of 1,2-dithiole-3-thiones and 1,2-dithiole-3-ones: a comparison of experimental and calculative approaches. *Pharma. Res.* 22(1):875–881.

Groom, C. A., S. Beaudet, A. Halasz, L. Paquet, and J. Hawari. 2001. Detection of the cyclic nitramine explosives hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) and octahydro1,3,5,7-tetranitro-1,3,5,7-tetrazine (HMX) and their degradation products in soil environments. *Journal of Chromatography Abstracts*. 909:53–60.

Hansch. C., A. Leo, and D. Hoekman. 1995. *Exploring QSAR. hydrophobic, electronic, and steric constants. ACS professional reference book*. Washington, DC: American Chemical Society.

Hilal, S. H., S. W. Karickhoff, and L. A. Carreira. 2004. Prediction of the solubility, activity coefficient and liquid/liquid partition coefficients of organic compounds. *QSAR Comb. Sci.* 23:709–720.

Hurley, M. 2007. Intelligent design of energetic materials. In *SERDP/ESTCP Partners in Environmental Technology Technical Symposium and Workshop, 4–6 December, Washington, DC.*

Karakoc, E., A. Cherkasov, and S. C. Sahinalp. 2006. Distance based algorithms for small biomolecule classification and structural similarity search. *Bioinformatics*. 15:243–251.

Lilienblum, W., W. Dekant, H. Foth, T. Gebel, J. G. Hengstler, R. Kahl, P. J. Kramer, H. Schweinfurth, and K. M. Wollin. 2008. Alternative methods to safety studies in experimental animals: role in risk assessment of chemicals under the new European Chemicals Legislation (REACH). *Arch. Toxicol.* 82(4):211–236.

Linkov, I., M. R. Ames, E. A. C. Crouch, and F. K. Satterstrom. 2005. Uncertainty in octanol-water partition coefficient: implications for risk assessment and remedial costs. *Environ. Sci. Technol.* 39(18):6917–6922.

Mannhold, R., and A. Petrauskas. 2003. Substructure versus whole-molecule approaches for calculating Log P. *QSAR Comb. Sci.* 22:466–745.

Mullee, D. M. 2007. Comparison of predicted vs actual partition coefficient results. In *SETAC 2007 Conference, 11–17 November, Milwaukee, WI.*

Qasim, M. 2007. New QSAR equations for accurate prediction of solubility for military compounds. In *SERDP/ESTCP Partners in Environmental Technology Technical Symposium and Workshop, December 4–6, 2007, Washington, DC.*

Pavan, M., T. I. Netzeva, and A. P. Worth. 2006. Validation of QSAR model for toxicity. *SAR Envrion. Res.* 17(2):147–171.

Sangster, J. 1993. *LOGKOW Databank. A databank of evaluated octanol-water partition coefficients (Log P) on microcomputer diskette*. Montreal, Quebec, Canada: Sangster Research Laboratories.

SciFinder. 2007. *Chemical Abstracts Service*. Columbus, OH: American Chemical Society. http://www.cas.org/products/scifindr/index.html

SPARC. 2007. *SPARC Performs Automated Reasoning in Chemistry (SPARC) v4.0*. Athens, GA: University of Georgia. http://ibmlc2.chem.uga.edu/sparc/

Sung, K., C. L. Munster, R. Rhykerd, M. C. Drew, and M. Y. Corapcioglu. 2002. The use of box lysimeters with freshly contaminated soils to study the phyto-remediation of recalcitrant organic contaminants. *Environmental Science and Technology*. 36:2249–2255.

Tetko, I. V., V. Y. Tanchuk, and A. E. P. Villa. 2001. Prediction of n-octanol partition coefficients from PHYSPROP database using artificial neural networks and e-state indices. *J. Chem. Inf. Comput. Sci.* 41:1407–1421.

Tomkins, B. A. 2000. Explosives analysis in the environment. In *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers, 2402–2441. Hoboken, NJ: John Wiley & Sons.

USEPA. 2007a. *Estimation program interface suite (EPI Suite) version 3.2*. Washington, DC: U.S. Environmental Protection Agency. http://www.epa.gov/oppt/exposure/pubs/episuite.htm.

USEPA. 2007b. *EPI Suite review panel report*. Washington, DC: Science Advisory Board, U.S. Environmental Protection Agency. http://www.epa.gov/sab/panels/epi_suite_review_panel.html

VCCLAB. 2007. Virtual Computational Chemistry Laboratory. http://www.vcclab.org

Viswanadhan, V. N., A. K. Ghose, G. R. Revankar, and R. K. Robins. 1989. Atomic physiochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* 29(3):163–172.

Willett, P., J. M. Barnard, and G. M. Downs. 1998. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38(6):983–996.

Zhou W., Z. Zhai, Z. Wang, and L. Wang. 2005. Estimation of n-octanol/water partition coefficients (K_{ow}) of all PCB congeners by density functional theory. *THEOCHEM* 755:137–145.

Appendix A

Table A1. Compounds identified by SciFinder Search molecular formula similarity searches as having structures similar to the target compounds.

Compound	RN (CAS)	SMILES Notation
Compounds with structures similar to RDX ($C_3 H_6 N_6 O_6$)		
1	872278-43-8	O=N(=O)C1NC(NC(N1)N(=O)=O)N(=O)=O
2	866392-48-5	CN(CN(\C=N\N(=O)=O)N(=O)=O)N(=O)=O
3	861616-18-4	O=N(=O)NC1C(NN(=O)=O)C1NN(=O)=O
4	186765-04-8	O=N(=O)N1CCN(N(C1)N(=O)=O)N(=O)=O
5	186765-03-7	CC1CN(N(N1N(=O)=O)N(=O)=O)N(=O)=O
6	186765-02-6	CN1CN(N(C1N(=O)=O)N(=O)=O)N(=O)=O
7	186765-01-5	O=N(=O)CN1N(CCN1N(=O)=O)N(=O)=O
8	186765-00-4	CC1N(N(CN(=O)=O)N1N(=O)=O)N(=O)=O
9	186764-99-8	CN1N(C(CN(=O)=O)N1N(=O)=O)N(=O)=O
10	186764-98-7	CN(C1N(CN1N(=O)=O)N(=O)=O)N(=O)=O
11	186764-97-6	CN(C1N(CN(=O)=O)N1N(=O)=O)N(=O)=O
12	186764-96-5	CCC(N1N(N1N(=O)=O)N(=O)=O)N(=O)=O
13	186764-95-4	O=N(=O)CN1N(CN(=O)=O)N1CN(=O)=O
14	148182-39-2	O=N(=O)N1CCCN(N1N(=O)=O)N(=O)=O
15	140375-52-6	O=C(NCNC(=O)NN(=O)=O)NN(=O)=O
16	117048-11-0	ON(O)C1=NC(=NC(=N1)N(O)O)N(O)O
17	82235-04-9	N=C(NCC(N(=O)=O)N(=O)=O)NN(=O)=O
18	81798-79-0	O=N(=O)C1CN(CN(N1)N(=O)=O)N(=O)=O
19	58300-56-4	NC(=O)NN(C(N(=O)=O)N(=O)=O)C(N)=O
20 (RDX)	121-82-4	O=N(=O)N1CN(CN(C1)N(=O)=O)N(=O)=O
Compounds with structures similar to atrazine ($C_8 H_{14} Cl N_5$)		
1	914099-23-3	CCN(C)C1=NC(Cl)=NC(=N1)N(C)C
2	40605-32-1	CC(C)CCNC1=NC(N)=NC(Cl)=N1
3	857010-11-8	CCCCCNC1=NC(N)=NC(Cl)=N1
4	310901-77-0	CCNC1=NC(=NC(Cl)=N1)N(C)CC
5	99767-64-3	CCCNC1=NC(=NC(Cl)=N1)N(C)C
6	99767-63-2	CCCCNC1=NC(Cl)=NC(NC)=N1
7	90952-64-0	CCCNC1=NC(Cl)=NC(NCC)=N1
8	90952-63-9	CCN(CC)C1=NC(NN)=NC(Cl)=C1
9	90952-62-8	CCN(CC)C1=NC(Cl)=CC(NN)=N1

Table A1 (cont.). Compounds identified by SciFinder Search molecular formula similarity searches as having structures similar to the target compounds.

Compound	RN (CAS)	SMILES Notation
(cont.). Compounds with structures similar to atrazine ($C_8 H_{14} Cl N_5$)		
10	88442-83-5	CN(C)CCNC1=NC=NC(Cl)=C1N
11	34333-27-2	CCCCNC1=NC(NC)=NC(Cl)=N1
12	33692-99-8	CCC(C)NC1=NC(NC)=NC(Cl)=N1
13	30355-62-5	CCNC1=NC(NCC)=NC(CCl)=N1
14	6270-25-3	CCCCNC1=C(N)C(Cl)=NC(N)=N1
15 (Atrazine)	001912-24-9	CCNC1=NC(NC(C)C)=NC(Cl)=N1
Compounds with structures similar to TNT ($C_7 H_5 N_3 O_6$)		
1	875839-44-4	NC1=CC(=C(C=C1C(O)=O)N(=O)=O)N(=O)=O
2	861792-36-1	COC1=CC(N=O)=C(C=C1N(=O)=O)N(=O)=O
3	861601-01-6	NC1=C(C=C(C(=C1)C(O)=O)N(=O)=O)N(=O)=O
4	861558-95-4	OCC1=C(C=C(C(=O)\C1=N\O)N(=O)=O)N=O
5	861386-90-5	OCC1=C(N=O)C(=O)C(=C\C1=N\O)N(=O)=O
6	859763-55-6	ON(=O)=C(C1=CC(=CC=C1)N(=O)=O)N(=O)=O
7	855197-77-2	OC(=O)C1=C(C=C(NN(=O)=O)C=C1)N(=O)=O
8	693288-34-5	OC(=O)NC1=C(C=C(C=C1)N(=O)=O)N(=O)=O
9	680208-24-6	ONC(=O)C1=C(C=CC(=C1)N(=O)=O)N(=O)=O
10	359901-22-7	O\N=C\C1=CC(=CC(=C1O)N(=O)=O)N(=O)=O
11	312504-84-0	NC(=O)C1=C(O)C=C(C=C1N(=O)=O)N(=O)=O
12	193354-03-9	OCC1=C(C=C(C=C1N=O)N(=O)=O)N(=O)=O
13	181934-63-4	NC(=O)OC1=CC(=CC(=C1)N(=O)=O)N(=O)=O
14	171179-75-2	NC(=O)C1=CC(O)=C(C=C1N(=O)=O)N(=O)=O
15	140380-55-8	NC1=CC(=CC(=C1C(O)=O)N(=O)=O)N(=O)=O
16	135252-14-1	NC(=O)C1=CC(=C(O)C(=C1)N(=O)=O)N(=O)=O
17	609-74-5	CC1=CC(=CC(=C1N(=O)=O)N(=O)=O)N(=O)=O
18	603-15-6	CC1=CC(=C(C(=C1)N(=O)=O)N(=O)=O)N(=O)=O
19	602-29-9	CC1=CC=C(C(=C1N(=O)=O)N(=O)=O)N(=O)=O
20 (TNT)	118-96-7	CC1=C(C=C(C=C1N(=O)=O)N(=O)=O)N(=O)=O
Compounds with structures similar to BTTN ($C_4 H_7 N_3 O_9$)		
1	84002-64-2	CC(ON(=O)=O)C(CON(=O)=O)ON(=O)=O
2	82975-75-5	O=N(=O)OCC(CON(=O)=O)CON(=O)=O
3	13051-22-4	OCC(CON(=O)=O)(CON(=O)=O)N(=O)=O
4 (BTTN)	6659-60-5	O=N(=O)OCCC(CON(=O)=O)ON(=O)=O

Table A1 (cont.). Compounds identified by SciFinder Search molecular formula similarity searches as having structures similar to the target compounds.

Compound	RN (CAS)	SMILES Notation
Compounds with structures similar to DDE ($C_{14}H_8Cl_4$)		
1 (DDE)	72-55-9	Cl\C(Cl)=C(/C1=CC=C(Cl)C=C1)C2=CC=C(Cl)C=C2
2	3328-98-1	Cl\C(Cl)=C(\C1=C(Cl)C=CC=C1)C2=C(Cl)C=CC=C2
3	3424-82-6	Cl\C(Cl)=C(/C1=CC=C(Cl)C=C1)C2=C(Cl)C=CC=C2
4	3622-56-8	ClC1(Cl)C2=CC=CC=C2C(Cl)(Cl)C3=C1C=CC=C3
5	6271-50-7	ClC1=CC(Cl)=C(C=CC2=C(Cl)C=C(Cl)C=C2)C=C1
6	18264-62-5	Cl\C=C(\C1=CC=C(Cl)C=C1)C2=C(Cl)C=C(Cl)C=C2
7	23812-84-2	ClC1=C(Cl)C2=C(C(Cl)=C1Cl)C3=CC=CC=C3CC2
8	29801-72-7	Cl\C(=C(\Cl)C1=C(Cl)C=CC=C1)C2=CC=C(Cl)C=C2
9	36804-46-3	ClC(Cl)C1=CC=CC2=C1C(Cl)(Cl)C3=CC=CC=C23
10	52923-07-6	ClC1=C(Cl)C(Cl)=C(Cl)\C1=C/C=C2\C=CC=CC=C2
11	61601-17-0	ClC1=C(Cl)C2=C(CC3=CC=CC=C3C2)C(Cl)=C1Cl
12	62681-19-0	ClC1=CC(Cl)=C(C=C1)C(=C)C2=C(Cl)C=C(Cl)C=C2
13	67242-09-5	ClC1=CC=CC(Cl)=C1\C=C\C2=C(Cl)C=CC=C2Cl
14	74326-84-4	ClC1=C2C3C(C4=C3CC=CC4)C2=C(Cl)C(Cl)=C1Cl
15	74326-85-5	ClC1=C(Cl)C2=CC3=C(CC=CC3)C=C2C(Cl)=C1Cl

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*
January 2009

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Strategic Environmental Research and Development Program
Sensitivity Analysis of QSAR Models for Assessing Novel Military Compounds

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jay L. Clausen, Erin Bennett, and Igor Linkov

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Research and Development Center
Cold Regions Research and Engineering Laboratory
Environmental Laboratory
72 Lyme Road
Hanover, NH 03755-1290

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC TR-09-3

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Reliable estimates of physical and biochemical properties of novel energetic compounds are essential before making the investment to synthesize, scale-up, and manufacture a new material for use in either military or civilian applications. Quantitative Structure-Activity Relationship (QSAR) software tools are available for predicting the physicochemical properties and environmental impacts of these emerging materials. The uncertainty and variability in melting point, solubility, half-lives, and related properties as a means of determining whether QSAR tools could provide meaningful results were evaluated. In particular, the octanol-water partition coefficient (K_{ow} or log P) was estimated for several proposed compounds. Log P was selected both because it typically can be measured with a high degree of certainty and because it correlates highly with water solubility and bioaccumulation. This study tested: 1) the variability in QSAR model predictions resulting from potential structural variants in emerging chemicals; and 2) the uncertainty from six different commercial K_{ow} calculators: KOWWIN, MarvinSketch, ACD/Labs, CLogP, SPARC, and ALOGPs. Analyses were performed on three military compounds [hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), butanetriol trinitrate (BTTN), and 2,4,6-trinitrotoluene (TNT)] and two pesticides [1-chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine (atrazine) and dichlorodiphenyldichloroethylene (DDE)]. Analyses of these compounds revealed that the uncertainty due to structural variations can be several orders of magnitude. Variability among the five software packages was as high as 10 orders of magnitude for emerging materials although lower for more well-studied chemicals such as DDE and atrazine. The magnitude of the uncertainty suggests use of existing QSAR models for emerging energetic materials is not appropriate.

15. SUBJECT TERMS
sensitivity analysis, QSAR, novel military compounds

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	U	33	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18