

# Application of Quantum Chemical Approximations to Environmental Problems: Prediction of Water Solubility for Nitro Compounds

YANA A. KHOLOD

*Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Jackson State University, Jackson, Mississippi, 39217*

EUGENE N. MURATOV

*Laboratory of Theoretical Chemistry, Department of Molecular Structure, A.V. Bogatsky Physical-Chemical Institute, National Academy of Sciences of Ukraine, Odessa, Ukraine, and Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Jackson State University, Jackson, Mississippi, 39217*

LEONID G. GORB

*SpecPro, Inc., Huntsville, Alabama 35805*

FRANCES C. HILL

*U.S. Army ERDC, Vicksburg, Mississippi, 39180*

ANATOLY G. ARTEMENKO

*Laboratory of Theoretical Chemistry, Department of Molecular Structure, A.V. Bogatsky Physical-Chemical Institute, National Academy of Sciences of Ukraine, Odessa, Ukraine*

VICTOR E. KUZ'MIN

*Laboratory of Theoretical Chemistry, Department of Molecular Structure, A.V. Bogatsky Physical-Chemical Institute, National Academy of Sciences of Ukraine, Odessa, Ukraine*

MOHAMMAD QASIM

*U.S. Army ERDC, Vicksburg, Mississippi, 39180*

JERZY LESZCZYNSKI\*

*Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Jackson State University, Jackson, Mississippi, 39217, and U.S. Army ERDC, Vicksburg, Mississippi 39180*

*Received August 23, 2009. Revised manuscript received October 8, 2009. Accepted October 13, 2009.*

Water solubility values for 27 nitro compounds with experimentally measured values were computed using the conductor-like screening model for real solvent (COSMO-RS) based on the density functional theory and COSMO technique. We have found that the accuracy of the COSMO-RS approach for prediction of water solubility of liquid nitro compounds is impressively high (the errors are lower than 0.1 LU). However, for some solid nitro compounds, especially nitramines, there is sufficient disagreement between calculated and experimental values. In order to increase the accuracy of predictions the quantitative structure–property relationship (QSPR) part of the COSMO-RS approach has been modified. The solubility values calculated

by the modified COSMO-RS method have shown much better agreement with the experimental values (the mean absolute errors are lower than 0.5 LU). Furthermore, this technique has been used for prediction of water solubility for an expanded set of 23 nitro compounds including nitroaromatic, nitramines, nitroanisoles, nitrogen rich compounds, and some their nitroso and amino derivatives with unknown experimental values. The solubility values predicted using the proposed computational technique could be useful for the determination of the environmental fate of military and industrial wastes and the development of remediation strategies for contaminated soils and waters. This predictive capability is especially important for unstable compounds and for compounds that have yet to be synthesized.

## 1. Introduction

Nitroaromatics, nitramines, nitroanisoles, nitrogen rich compounds, and their nitroso and amino derivatives are of primary use as explosives and propellants in the military and in industry (1, 2). Waste from nitro compounds are easily disseminated during manufacturing, storage, transportation, and utilization of munitions, leading to a potential hazard for humans and the environment (3). A number of studies have shown that nitro compounds, as well as their metabolites of environmental transformation, byproduct of synthesis, or incomplete combustion are harmful for the biosphere due to their toxicity (3–6). Therefore, remediation of these contaminants is a point of prominent concern for environmental science. The main methods for the remediation of nitro compounds are biological (7–10), physical (11–15), and chemical (14, 16–21) treatment of contaminated soils and waters. The environmental fate of chemicals, as well as the efficiency of in-solution treatment is determined largely by their solubility in water. In particular, high water solubility leads to expeditious distribution in water, i.e., chemicals that rapidly and completely dissolve in water are transported along with the general water flow. Low water solubility is generally associated with high affinity for adsorption to solids in water (soil particles or sediment). In addition, high water solubility is associated with absorption of pollutants by living organisms. Hence, solubility of hazardous chemicals is one of the most fundamental physical properties of ecological importance.

Unfortunately, due to the highly exothermic nature of explosives, experimental data on many of their properties are quite limited. Therefore, several robust partly- or fully theoretical models have been developed for efficient estimation of solubility. The general solubility equation proposed by Jain and Yalkowsky in 2001 (22) estimates water solubility of solid organic nonelectrolytes based on two parameters: melting point (experimental or theoretically predicted) and octanol–water partition coefficient. Group contribution methods are based on the numbers and types of functional groups in a molecule of interest (23). Quantitative structure–property relationship (QSPR) methods were developed in order to associate the structure of molecules represented by a set of descriptors with their physical or chemical properties (24).

The conductor-like screening model for real solvents (COSMO-RS) recently developed by Klamt and Eckert (25, 26) is one of the most promising methods for the prediction of physicochemical properties, in particular water solubility, for virtually any compound of interest. This model is based on high level quantum chemical calculations combined with

\* Corresponding author phone : (601) 979-3723; fax: (601) 979-7823; e-mail: jerzy@ccmsi.us.

# Report Documentation Page

Form Approved  
OMB No. 0704-0188

Public reporting burden for the 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 the 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 Washington Headquarters Services, Directorate for Information Operations and Reports, 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 a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE <b>2009</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2009 to 00-00-2009</b>	
4. TITLE AND SUBTITLE <b>Application of Quantum Chemical Approximations to Environmental Problems: Prediction of Water Solubility for Nitro Compounds</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Jackson State University, Jackson, MS, 39217</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
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 <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

the conductor-like screening model (COSMO) (27) and an efficient statistical thermodynamics of interfacial surfaces. In the study by Klamt et al. (28) the COSMO-RS model has shown impressively high accuracy for prediction of the solubility values of liquid compounds (0.3 logarithmic units (LU)). In the same study this method was adapted for solid at given temperature solutes and tested on drugs and pesticides. The intrinsic error of ca. 0.5 LU obtained for solid solutes is comparable with accuracy of experimental results (28).

The COSMO-RS and group contribution methods have been used by Toghiani and co-workers in 2008 (30) to calculate the water solubility of nine high energetic materials at different temperatures. The authors have shown that both methods give quite accurate results when compared with available experimental data. In 2007 the COSMO-RS method was applied by our group (31) for the prediction of the major ecologically important physicochemical properties, in particular, water solubility of 10 nitro compounds: trinitrotoluene (TNT) and nitro and amino derivatives of toluene. It has been shown that the COSMO-RS method give an intrinsic error of 0.52 LU for prediction for water solubility for those compounds, a value that is close to the COSMO-RS limitations for solids estimated by Klamt et al. (28). However, as it will be shown below, when we expanded the set of target compounds to include nitramines, nitroanisoles, nitrogen rich compounds, and their nitroso and amino derivatives, the accuracy of predictions decreased sufficiently. That fact has motivated us to (i) find the reasons for worse predictions of water solubility in case of expanding the set of nitro compounds; (ii) improve the solubility predictions for nitro compounds within the COSMO-RS method; (iii) predict water solubility with reasonable accuracy for those nitro compounds that lack experimental data.

## 2. Methods

Quantum chemical calculations were performed for all molecules in this study. We employed two levels of theory proposed by Klamt and Eckert (26): DFT/COSMO/SVP (abbreviated as SVP level further on) and DFT/COSMO/TZVP (abbreviated as TZVP level). For the SVP level the geometries of all compounds were optimized by the semiempirical AM1 (33) method combined with the COSMO (27) model to count the solvent effects using the MOPAC2009 (34) program. Based on the optimized geometries the COSMO polarization charge densities ( $\sigma$ ) of the molecular surfaces were computed at the DFT level of theory by application of the BVP86 functional based on Perdew's 1986 functional (35) combined with Ahlrichs' contracted basis set SVP (36) using the Gaussian03 program suite (37). For the TZVP level, the geometry optimizations were carried out by the BVP86 functional combined with Ahlrichs' contracted Gaussian basis set of triple- $\zeta$  valence quality (TZVP) (38) using the Gaussian03 program. The solvent effect was taken into account by a conductor solvent model (CPCM) (39) based on the COSMO model (27) using atomic radii and other parameters as suggested by Klamt (27). Then, the single-point calculation at the same level of theory was performed for each molecule to generate the polarization charge density ( $\sigma$ ) for the COSMOtherm program (32).

As the next step, the polarization charge density for each molecule was computed at both SVP and TZVP levels of theory, then converted into a distribution function ( $\sigma$ -profile) and treated statistically using the COSMO-RS model in the COSMOtherm program to get the values of standard chemical potentials ( $\mu^0$ ). The obtaining of standard chemical potentials of the compound in pure state and in dissolved state is possible due to computation of charge distribution on the surface of a specific cavity that surrounds a molecule of solute or water (25–27). Then, estimating the interactions of such

molecular cavities (solute with other solute molecules for pure state, or solute with water molecules for solution state) as described by Klamt (25, 26) one can calculate interaction chemical potentials. For the calculation of the chemical potential of solute in solution state the condition of infinite solution is applied to ensure absence of solute–solute interactions. Thus, the resulting chemical potentials are standard chemical potentials. The solubility values were computed according to the equations proposed by Eckert and Klamt (32) in the COSMO-RS theory:

(a) in the case of liquid solute:

$$\log(X_s) = [\mu_{\text{self}}^0 - \mu_{\text{solv}}^0] / RT \ln(10) \quad (1)$$

where  $X_s$  is solubility in mole fractions,  $\mu_{\text{self}}^0$  and  $\mu_{\text{solv}}^0$  are the standard chemical potentials of a solute in its liquid state and solute in solution, accordingly,  $R$  is the universal gas constant, and  $T$  is temperature;

(b) in the case of solid solute:

$$\log(X_s) = [\mu_{\text{self}}^0 - \mu_{\text{solv}}^0 - \Delta G'_{\text{fus}}] / RT \ln(10), \text{ where } \Delta G'_{\text{fus}} = \begin{cases} 0, \Delta G_{\text{fus}} \leq 0 \\ \Delta G_{\text{fus}}, \Delta G_{\text{fus}} > 0 \end{cases} \quad (2)$$

where  $\Delta G_{\text{fus}}$  is Gibbs free energy of fusion. According to definition  $\Delta G_{\text{fus}}$  should have all positive values for solid at given temperature substances. Since for liquids  $\Delta G_{\text{fus}}$  is negative by definition, eq 2 reduces to eq 1 for liquid solute. The COSMO-RS approach uses a QSPR technique described in detail by Klamt et al. (28) to estimate  $\Delta G_{\text{fus}}$  values for solids.

For alternative  $\Delta G_{\text{fus}}$  estimation, the simplex representation of molecular structure (SiRMS) QSPR approach developed by Kuz'min and co-workers (40) was used in the present study. The principal physical basis of SiRMS, in application to  $\Delta G_{\text{fus}}$  based on hypothesis that  $\Delta G_{\text{fus}}$  mainly determined not by the whole molecule, but by its structural parts, e.g., substituents, "pharmacophores", etc (40). Bounded and unbounded simplex descriptors of fixed composition and topology (2D level) were generated. Simplexes were differentiated by atom nature (element and type), partial charge, lipophilicity, refraction, and its ability to be donor or acceptor in H-bonding (see the Supporting Information (SI) for more information). Statistical models have been obtained by the partial least-squares or projections on latent structures (PLS) method (41). Usage of SiRMS in combination with PLS allows determining such fragments (simplexes) and their combinations which are important to  $\Delta G_{\text{fus}}$  variation. Each molecule is represented by unified set of simplexes and can be explicitly retrieved from this set.

## 3. Results and Discussion

**3.1. Calculation of Water Solubility with the Original COSMO-RS Approximation.** Initially, available experimental data on water solubility of nitro compounds was tabulated from the literature (20, 42). These compounds, their CAS registry numbers and corresponding solubility values are shown in Table 1. Full names of all the compounds used in this study along with structural formulas are given in SI Table S1).

Some of those chemicals have been previously studied using the COSMO-RS approach at the TZVP level in our recent work (31). For all remaining species we have computed the solubility values at the corresponding temperatures using the original COSMO-RS procedure at both SVP and TZVP levels as described in the Methods section.

There are two reasons why we did calculations using both basis sets. First, to compare the accuracy of our results and those reported by Klamt et al. (28) (where only accuracy of SVP level is estimated). Second, since the BVP86/SVP calculations based on COSMO/AM1 geometry optimization

**TABLE 1. Comparison of the Water Solubilities ( $S_w$ ) Computed Using the COSMO-RS Approach and Available Experimentally Measured Water Solubility Values**

compound	CAS	COSMO-RS $S_w$ , log(mg/L) <sup>a</sup>		errors		exp. $S_w$ , log(mg/L) <sup>b</sup>	temp, °C	
		SVP	TZVP	SVP	TZVP			
Solids								
1	TNB	99-35-4	1.79	2.12	0.65	0.32	2.44	15
2	1,2-DNB	528-29-0	2.33	2.83	0.21	0.71	2.12	25
3	1,3-DNB	99-65-0	2.76	2.88	0.03	0.15	2.73	25
4	1,4-DNB	100-25-4	2.47	2.86	0.63	1.02	1.84	25
5	1,4-DNsB	105-12-4	3.40	3.57	0.06	0.11	3.46	25
6	NsB	586-96-9	3.45	3.74	0.12	0.41	3.33	25
7	2,4,6-TNT	118-96-7	1.72	2.05	0.24	0.01	2.06	23
8	2,3,4-TNT	602-29-9	2.07	2.06	0.05	0.06	2.12	25
9	2,4,5-TNT	610-25-3	1.71	2.04	0.48	0.15	2.19	25
10	2,3-DNT	602-01-7	2.57	2.38	0.23	0.04	2.34	25
11	2,4-DNT	121-14-2	2.50	2.55	0.07	0.12	2.43	22
12	2,5-DNT	619-15-8	2.44	2.54	0.1	0.2	2.34	25
13	2,6-DNT	606-20-2	2.62	2.65	0.32	0.35	2.30	25
14	3,4-DNT	610-39-9	2.33	2.42	0.33	0.42	2.00	25
15	4-NT	99-99-0	2.98	2.94	0.33	0.29	2.65	30
16	2-Am-46-DNT	35572-78-2	3.53	3.80	0.44	0.71	3.09	25
17	4-Am-26-DNT	19406-51-0	3.79	4.03	0.7	0.94	3.09	25
18	2,4,6-TNAN	606-35-9	1.73	1.71	0.57	0.59	2.30	15
19	2,4-DNAN	119-27-7	2.79	2.61	0.6	0.42	2.19	15
20	3,5-DNAN	5327-44-6	2.01	2.22	0.47	0.26	2.48	25
21	RDX	121-82-4	3.54	3.74	1.79	1.99	1.75	25
22	HMX	2691-41-0	2.42	3.72	1.77	3.07	0.65	25
23	CL-20	135285-90-4	1.80	2.99	1.24	2.43	0.56	25
24	tetryl	479-45-8	1.84	1.84	0.03	0.03	1.87	25
	MAE <sup>c</sup>		0.51	0.62				
Liquids								
25	NB	98-95-3	3.28	3.21	0.04	0.11	3.32	25
26	2-NT	88-72-2	3.05	2.95	0.24	0.14	2.81	30
27	3-NT	99-08-1	2.87	2.83	0.17	0.13	2.70	30
	MAE <sup>c</sup>		0.15	0.13				

<sup>a</sup> All COSMO-RS calculations were performed at the temperature of experimental measurements. <sup>b</sup> References for experimental solubility values: CL-20 (26), all other compounds (53). <sup>c</sup> Mean absolute error.

are much faster, than BVP86/TZVP calculations including COSMO/BVP86/TZVP geometry optimization, the SVP level is more attractive when it is a need to treat a large set of chemicals.

The theoretically obtained values of solubility were then compared with the collected experimental ones. The results are displayed in Table 1. One can see that both the SVP level and the TZVP level predicted solubility values for liquid compounds (NB, 2-NT and 3-NT) are in very close agreement with the experimental values (ca. 0.1 LU error). This is in agreement with the reported accuracy for the COSMO-RS method for liquids (0.3 LU) (28). However, the accuracy decreases sufficiently in the case of solid solutes. Although for some solids the predicted water solubility values are in good agreement with the experimental values, the errors for prediction of solubilities for some others, in particular, RDX, HMX, and CL-20, are quite large: 1.6–2.5 LU. As a result, the mean absolute error (MAE) for the SVP level is equal to 0.51 LU, and it is even worse for TZVP level: 0.62 LU. The errors obtained for the SVP level is beyond the limitations of the intrinsic error of the COSMO-RS method for solids, reported as 0.5 LU (28). A higher degree of accuracy was expected for predictions at the TZVP level since it is based on the higher and more acute level of theory.

In order to determine the possible reasons for such contradictions we have analyzed the contributions into eqs 1 and 2 in detail. All the terms of eq 1 for the liquid compounds at experimental measurement temperature and the terms of eq 2 for solids are given in Table 2. Since the most compounds of interest are solids under given temperatures we have paid special attention to the values of  $\Delta G_{fus}$  which are obtained statistically in the COSMO-RS theory.

Analysis of the data presented in Table 2 suggests that the difference of  $\mu_{self}^0 - \mu_{solv}^0$  is negative. This corroborates the meaning of these terms according to eqs 1 and 2, since COSMO-RS is oriented to the prediction of solubility of rather low soluble solutes (28). The difference  $\mu_{self}^0 - \mu_{solv}^0$  is negative for low-soluble compounds, when energy of the compound in pure state is lower, than in solution. The only valuable parameter is the difference between  $\mu_{self}^0$  and  $\mu_{solv}^0$ , thus we put this difference value in Table 2 instead of individual values of standard chemical potentials.

$\Delta G_{fus}$  is not applicable for liquid compounds, therefore this value is not estimated by the COSMO-RS procedure for liquid solute, and solubility values are computed according to eq 1. However, in the case of solid compounds, the COSMO-RS estimated values of  $\Delta G_{fus}$  are surprisingly negative. This finding contradicts the meaning of the  $\Delta G_{fus}$  term, which has a positive value for solids, interpreted as the energy necessary to break intermolecular bonds in a crystal lattice. A selective check of the value of this parameter for organic nonelectrolyte compounds of different classes (in particular those studied in the work (28) and implemented in the COSMOtherm database) reveals that  $\Delta G_{fus}$  values are always positive.

Therefore, we have concluded that the QSPR equation used in the original COSMO-RS method to predict  $\Delta G_{fus}$  is inadequate for this class of nitro compounds. We can also assume that the large error for nitramines is because those compounds should have higher  $\Delta G_{fus}$  than nitroaromatic compounds, and neglecting this value causes inaccuracy. In order to increase the accuracy of the predictions of the solubility values of nitro compounds and to expand the usage of the COSMO-RS method for military compounds we have

**TABLE 2. Difference of the Standard Chemical Potentials and Free Energies of Fusion ( $\Delta G_{fus}$ ) Computed Using the COSMO-RS Approach and the Water Solubility ( $S_w$ ) Values Calculated Using These Terms According to Eq 1 for Liquid and Eq 2 for Solid Solute**

compound	SVP			TZVP			
	$\mu_{self} - \mu_{solvr}$ kcal/mol	$\Delta G_{fus}$ kcal/mol	$S_w \log(X)$	$\mu_{self} - \mu_{solvr}$ kcal/mol	$\Delta G_{fus}$ kcal/mol	$S_w \log(X)$	
Solids							
1	TNB	-6.96	-1.42	-5.28	-6.54	-1.84	-4.95
2	1,2-DNB	-6.34	-0.75	-4.64	-5.65	-1.08	-4.14
3	1,3-DNB	-5.75	-0.58	-4.21	-5.58	-1.01	-4.09
4	1,4-DNB	-6.14	-0.68	-4.50	-5.6	-1.02	-4.10
5	1,4-DNsB	-4.75	-0.09	-3.48	-4.52	-0.33	-3.31
6	NsB	-4.28	0.25	-3.33	-3.96	0.18	-3.03
7	2,4,6-TNT	-7.29	-1.45	-5.38	-6.84	-1.95	-5.05
8	2,3,4-TNT	-6.87	-1.30	-5.03	-6.88	-1.77	-5.04
9	2,4,5-TNT	-7.35	-1.45	-5.39	-6.9	-1.94	-5.06
10	2,3-DNT	-6.06	-0.58	-4.44	-6.31	-1.09	-4.62
11	2,4-DNT	-6.08	-0.53	-4.50	-6.02	-1.03	-4.46
12	2,5-DNT	-6.23	-0.61	-4.56	-6.09	-1.08	-4.46
13	2,6-DNT	-5.99	-0.56	-4.39	-5.94	-1.08	-4.36
14	3,4-DNT	-6.38	-0.68	-4.68	-6.26	-1.15	-4.59
15	4-NT	-5.41	-0.10	-3.90	-5.47	-0.35	-3.94
16	2-Am-46-DNT	-4.79	-0.17	-3.51	-4.42	-0.47	-3.24
17	4-Am-26-DNT	-4.43	-0.04	-3.25	-4.1	-0.40	-3.01
18	2,4,6-TNAN	-7.37	-1.53	-5.40	-7.14	-1.90	-5.42
19	2,4-DNAN	-5.8	-0.55	-4.25	-5.84	-1.13	-4.43
20	3,5-DNAN	-6.86	-0.88	-5.03	-6.58	-1.30	-4.82
21	RDX	-4.84	-0.96	-3.55	-4.57	-1.64	-3.35
22	HMX	-6.55	-1.37	-4.80	-4.77	-1.94	-3.49
23	CL-20	-7.61	-2.60	-5.58	-6	-2.77	-4.39
24	Tetryl	-8.2	-2.25	-6.01	-7.31	-2.70	-5.36
Liquids							
25	NB	-4.86		-3.56	-4.95		-3.63
26	2-NT	-5.32		-3.83	-5.46		-3.93
27	3-NT	-5.56		-4.01	-5.62		-4.05

modified the QSPR procedure of  $\Delta G_{fus}$  estimation as described in the next section.

**3.2. New QSPR Equation for Prediction of  $\Delta G_{fus}$  for COSMO-RS.** QSPR procedures require that a training set be constructed for the property of interest. In the present study, values for the training set for the free energy of fusion ( $\Delta G_{fus}$ ) were estimated according to eq 3 proposed in the work (28) for free energy of fusion estimation:

$$\Delta G_{fus}^* = RT \ln(10) [\log(x_{COSMO-RS}) - \log(x_{exp})] \quad (3)$$

where  $x_{EXP}$  is the experimental solubility values in mole fraction at 25 °C;  $x_{COSMO-RS}$  is COSMO-RS solubility values in mole fraction at 25 °C calculated according to eq 1; and  $T$  is the temperature considered equal to 25 °C.

The data set  $\{x_{exp}\}$  of experimentally measured water solubility values was composed of 150 common organic and drug compounds which are solid at 25 °C. This data set was initially proposed by Duffy and Jorgensen (43) and then used by Klamt et al. (28) for obtaining a regression equation for  $\Delta G_{fus}$  estimation in the COSMO-RS approach for solid solutes. The  $\sigma$ -profiles of these compounds implemented in the COSMOtherm database were used for computation of the solubility values in COSMO-RS at both the SVP and TZVP levels. Since all the compounds of the  $\{x_{exp}\}$  set are solids at 25 °C, all of the corresponding  $\Delta G_{fus}^*$  values are constrained to be positive. Thus, all the compounds of the training set with negative  $\Delta G_{fus}^*$  values were excluded. Then a data set  $\{x_{exp}\}$  of 125 remaining compounds and seven additionally included nitro compounds (2-NT, 3-NT, 4-NT, 2,4-DNT, 2,6-DNT, RDX, and 5-nitro-1,10-phenanthroline) was used to create the training work set (shown in SI Table S2) for  $\Delta G_{fus}$  estimation.

Starting from this step the QSPR protocol used in (28) was changed. Thus, instead of using the fitting equation with

predetermined descriptors, the SiRMS QSAR approach (40) was used for prediction of  $\Delta G_{fus}$  values (see ref 40 and SI Table S3 for more details regarding the developed models). Basic statistical parameters which characterize the quality of 2D PLS models for the SVP and TZVP levels are presented in Table 3.

As follows from the analysis of the data collected in Table 3 we have succeeded in obtaining QSPR equations of near the same quality for both SVP and TZVP levels that are both robust and well-constrained.

**3.3. Verification of the Developed QSPR Model.** In order to validate the predictive ability of the proposed models we have applied these models to predict the  $\Delta G_{fus}$  and solubility values of 53 drugs and pesticides collected from the literature (28, 44, 45) plus five nitro compounds: TNB, TNT, 4-Am-2,6-DNT, CL-20, and HMX. Our external validation set consisting of 58 compounds shown in SI Table S4. The standard chemical potentials,  $\mu_{self}^0$  and  $\mu_{solvr}^0$  were computed with the COSMO-RS approach. The modified solubility values for this set of compounds were calculated according to eq

**TABLE 3. Basic Statistical Parameters of Selected SiRMS QSPR Models<sup>a</sup>**

	$R^2$	$Q^2$	$S(ws)$	$S(cv)$	A	D	M
SVP	0.86	0.78	0.43	0.55	2	28	132
TZVP	0.87	0.82	0.52	0.62	2	46	132

<sup>a</sup> Note:  $R^2$  is determination coefficient,  $Q^2$  is the cross validation determination coefficient,  $S(ws)$  is the standard error of prediction for the work set,  $S(cv)$  is the standard error of prediction for the work set in cross validation terms, A is the number of PLS latent variables, D is the number of descriptors, and M is the number of molecules in the training set.

**TABLE 4. Statistical Parameters for the External Validation Set Prediction by the Considered SiRMS QSPR Models**

	$R^2_{\text{test}}$	$Q^2_{\text{test}}$	$S(\text{ts})$
COSMO-RS/SVP, original	0.49	0.67	1.05
COSMO-RS/SVP, modified	0.46	0.51	1.29
COSMO-RS/TZVP, original	0.44	0.62	1.13
COSMO-RS/TZVP, modified	0.66	0.71	0.99

Note:  $R^2_{\text{test}}$  is the squared correlation coefficient for external test set;  $Q^2_{\text{test}}$  is the determination coefficient for external test set.  $S(\text{ts})$  is standard error of a prediction for external test set.

2. The original COSMO-RS method for solids, with the QSPR procedure implemented with the COSMOtherm program was also applied to compute solubility values. The results of solubility predictions by both of these models were compared with the experimental values (SI Table S4). The statistical parameters of the developed models are presented in Table 4.

The analysis of Table 4 shows that the modified COSMO-RS/TZVP model can provide quite accurate predictions. The modified COSMO-RS/SVP model does not fully satisfy the predictivity requirements described by Golbraikh and Tropsha (46), so predictions by this model are expected to be less reliable.

At the next step, to check how our modified model works for nitro compounds, we have estimated the Gibbs free energies of fusion for nitro compounds listed in Table 1. To obtain a correct statistical picture of the influence of our modification on the solubility prediction results, 4-NT, 2,4-DNT, 2,6-DNT, and RDX were excluded from consideration at this point, since, as mentioned above these compounds were the constituents of the training set for the QSPR model development. Since our QSPR model is developed to estimate the  $\Delta G_{\text{fus}}$  values at 25 °C, adjustments had to be done for TNB, 2,4,6-TNT, 2,4,6-TNAN, and 2,4-DNAN to estimate their  $\Delta G_{\text{fus}}$  at the temperatures of experimental measurements carried out at temperatures other than 25 °C.

As shown in the work (29) and then applied in ref 30 the solubility can be estimated as

$$\ln(\gamma_s x_s) = \frac{\Delta H_{\text{fus}}}{RT} \left(1 - \frac{T}{T_{\text{melt}}}\right) \quad (4)$$

where  $\gamma_s$  is the activity coefficient,  $\Delta H_{\text{fus}}$  is the enthalpy change for melting the solute,  $T_{\text{melt}}$  is the melting point temperature. In the case of ideal solubility  $\gamma_s = 1$ , and the solubility is determined only by the intermolecular forces in solid solute crystal that are characterized by the free energy of fusion. Thus, the eq 4 can be rewritten for  $\Delta G_{\text{fus}}$  (47)

$$\Delta G_{\text{fus}}(T) = -\Delta H_{\text{fus}} \left(1 - \frac{T}{T_{\text{melt}}}\right) \quad (5)$$

Using the eq 5 the  $\Delta G_{\text{fus}}$  at 25 °C (298K) obtained using the proposed above QSPR model can be adjusted for temperature  $T$  if the melting point temperature  $T_{\text{melt}}$  is known:

$$\Delta G_{\text{fus}}^T = \Delta G_{\text{fus}}^{298} \left(\frac{T_{\text{melt}} - T}{T_{\text{melt}} - 298}\right) \quad (6)$$

In the current work, the  $\Delta G_{\text{fus}}$  values for TNB, 2,4,6-TNT, 2,4,6-TNAN, and 2,4-DNAN were adjusted depending on the temperature of the experiment according to eq 6. Melting point temperatures were taken from the PhysProp database (42).

We believe that neglecting of nonlinear behavior of  $\Delta G_{\text{fus}}$  dependence on temperature in this case does not generate

significant errors, since we do the adjustments in the small region of temperatures ( $\leq 10\text{K}$ ).

Based on the  $\Delta G_{\text{fus}}$  values and standard chemical potentials calculated with new QSPR model combined with the COSMO-RS approach, we have calculated the new water solubility values for solids, according to eq 2. In addition, we have predicted the solubility values of these chemicals using the U.S. Environmental Protection Agency recommended EPI Suite and SPARC predictive tools. All the results are collected in Table 5. As seen from Table 5, all the  $\Delta G_{\text{fus}}$  values, estimated by our QSPR model are positive, that is in agreement with the physical meaning of this term.

For comparison of accuracy, we have calculated the mean absolute errors for the predictions made for the 21 nitro compounds listed in Table 5 for all models used. One may see that the accuracy of the COSMO-RS technique with the modified QSPR procedure is higher than the accuracy of solubility predictions by the EPI and SPARC predictive tools. Similar to the original COSMO-RS model, in EPI and SPARC the highest errors are observed for nitramines. The errors of the solubility values computed by the modified COSMO-RS approach for both the SVP and TZVP levels are now within the limits of the COSMO-RS method for solid solute. The errors for all compounds under study are quite low. The mean absolute error for the solubility values calculated with the modified procedure for  $\Delta G_{\text{fus}}$  decreased by 0.13 LU when compared to the original COSMO-RS method at the SVP level. In the case of the TZVP level the predictions for all compounds have improved, and the MAE decreased by 0.15 LU if compared with the original model.

The comparison of the results obtained for the current set of compounds at both levels shows that SVP method gives quite accurate results, as well as TZVP. That allows using of SVP level for calculation of water solubilities of large sets of compounds. However, as statistical results in Table 4 show, the reliability of SVP model is lower, that TZVP (it means that in some cases SVP can result in higher errors), so if computational resources allow, we recommend using TZVP level, as more reliable.

**3.4. Application of the Modified COSMO-RS Approach for Water Solubility Predictions for Nitro Compounds with Unknown Experimental Values.** At the final step of the present work we have demonstrated the predictive role of the method discussed to compute the solubility values at room temperature for some more nitro-compounds of interest for those experimental values have been never measured, and some of those compounds have never even been synthesized. Since the SVP and TZVP levels give almost equally accurate results as shown in Table 5, we have used both levels of theory for the prediction of water solubility values for nitro compounds with unknown experimental values at 25 °C, collected in Table 6.

Table 6 clearly demonstrates that all predicted  $\Delta G_{\text{fus}}$  values are positive that means this term is counted during the calculation of solubility values by eq 2 for solid solute. Thus, based on the results for the nitro-compounds with known experimental values, we can expect these solubility values are trustworthy and their accuracy is close to experimental. The analysis of Table 6 demonstrates that all compounds under study except four last nitrogen rich compounds are low soluble. In contrast, the nitrogen rich compounds ANTA, FOX-7, and HBT have shown the extremely high water solubility values predicted by all applied models. Although there is no available experimental data, we can assume that these compounds are highly soluble, since several different models give similar results on those compounds.

With regard to the change of water solubility depending on polarity of the compound, Table 6 shows a clear trend of

**TABLE 5. Free Energies of Fusion ( $\Delta G_{fus}$ ) Estimated on the Base of SiRSM Equations and Corresponding Water Solubility Values ( $S_w$ ) Computed at SVP and TZVP Levels along with Available EPI and SPARC Data**

compound	SVP <sup>a</sup>			TZVP <sup>a</sup>			$S_w$ , log(mg/L)				
	$\Delta G_{fus}$ , kcal/mol	$S_w$ , log(mg/L)	error	$\Delta G_{fus}$ , kcal/mol	$S_w$ , log(mg/L)	error	EPI <sup>b</sup>		SPARC <sup>a</sup>	exp <sup>c</sup>	
							frag	$K_{ow}$			
1 TNB	0.35	1.71	0.73	0.66	1.80	0.64	2.38	3.18	2.96	2.44	
2 1,2-DNB	0.32	2.10	0.02	0.69	2.32	0.2	2.67	2.96	3.17	2.12	
3 1,3-DNB	0.19	2.61	0.12	0.55	2.47	0.26	2.67	3.14	3.35	2.73	
4 1,4-DNB	0.19	2.33	0.49	0.45	2.54	0.7	2.67	3.16	2.93	1.84	
5 1,4-DNsB	0.38	3.12	0.34	0.16	3.45	0.01	2.04	3.46	6.13	3.46	
6 NsB	0.48	3.28	0.05	0.41	3.57	0.24	2.60	3.33	4.47	3.33	
7 2,4,6-TNT	0.37	1.49	0.57	0.69	1.53	0.53	1.88	2.74	2.32	2.06	
8 2,3,4-TNT	0.36	1.81	0.31	0.64	1.59	0.53	1.88	2.41	2.08	2.12	
9 2,4,5-TNT	0.36	1.45	0.74	0.97	1.33	0.86	1.88	2.41	2.44	2.19	
10 2,3-DNT	0.36	2.30	0.04	0.76	1.82	0.52	2.18	2.48	2.69	2.34	
11 2,5-DNT	0.23	2.27	0.07	0.52	2.16	0.18	2.18	2.48	3.04	2.34	
12 3,4-DNT	0.36	2.07	0.07	0.76	1.86	0.14	2.18	2.56	2.96	2.00	
13 2-Am-4,6-DNT	0.72	3.00	0.09	0.82	3.20	0.11	2.50	3.09	2.06	3.09	
14 4-Am-2,6-DNT	0.72	3.26	0.17	0.82	3.43	0.34	2.50	3.09	2.27	3.09	
15 2,4,6-TNAN	0.49	1.37	0.93	1.01	1.15	1.15	2.11	2.72	2.22	2.30	
16 2,4-DNAN	0.31	2.56	0.37	0.90	2.10	0.09	2.41	2.80	2.43	2.19	
17 3,5-DNAN	0.27	1.81	0.67	0.13	2.13	0.35	2.41	2.80	2.33	2.48	
18 HMX	2.76	0.39	0.26	3.03	1.50	0.85	6.00	3.97	0.67	0.65	
19 CL-20	2.21	0.18	0.38	3.77	0.23	0.33	6.00	2.95	-5.73	0.56	
20 Tetryl	0.56	0.78	1.09	1.34	0.86	1.01	2.48	2.76	1.06	1.87	
MAE (original) <sup>d</sup>		0.50			0.60		0.90	0.66	0.95		
MAE (modified) <sup>40</sup>		0.37			0.45						

<sup>a</sup> All COSMO-RS and SPARC calculations are at the temperature of experimental measurements indicated in Table 2. <sup>b</sup> All EPI calculations are done at 25, because it is the only available option, frag -  $S_w$  values calculated from fragments,  $K_{ow}$  -  $S_w$  values calculated from octanol/water partition coefficients ( $K_{ow}$ s are theoretically estimated in EPI as well). <sup>c</sup> The experimental values here duplicate those in Table 2 and are displayed just for reader's convenience. <sup>d</sup> Mean absolute errors for original and modified COSMO-RS models. MAE for the original COSMO-RS at both SVP and TZVP levels were recalculated at this step for the set of 20 compounds collected in the current table in purpose of correct comparison of predictive ability of original and modified COSMO-RS models.

**TABLE 6. Modified COSMO-RS, EPI, and SPARC Predictions of Water Solubility Values for Nitro and Nitroso Compounds with Unknown Experimental Values at Temperature 25 °C**

compound <sup>a</sup>	CAS number	SVP		TZVP		EPI		
	CAS number	$\Delta G_{fus}$ , kcal/mol	$S_w$ , log(mg/L)	$\Delta G_{fus}$ , kcal/mol	$S_w$ , log(mg/L)	frag <sup>b</sup>	$K_{ow}$ <sup>b</sup>	SPARC
1 TNsB		0.60	2.69	0.16	3.12	1.46	3.44	>6.00
2 1,3-DNsB		0.38	3.25	0.26	3.55	2.04	3.46	6.00
3 2,3,6-TNT	18292-97-2	0.36	1.42	0.97	1.37	1.88	2.41	2.10
4 3-Am-2,6-DNT		0.82	3.60	0.85	3.80	2.50	2.62	2.62
5 2,4-Am-6-NT	6629-29-4	1.08	3.87	0.97	4.53	3.10	4.33	2.76
6 2,6-Am-4-NT	59229-75-3	1.08	3.87	0.97	4.56	3.10	4.33	2.73
7 2,4,6-TAmT	88-02-8	1.44	4.51	1.12	5.22	3.99	5.58	>6.00
8 2-NHOH-4,6-DNT	59283-76-0	0.54	3.67	0.72	3.82	2.88	3.04	3.19
9 4-NHOH-2,6-DNT	59283-75-9	0.54	4.40	0.72	4.33	2.88	3.04	3.17
10 2,3DNAN		0.27	2.02	0.65	1.81	2.41	2.80	2.06
11 2,5DNAN		0.27	2.15	0.52	2.09	2.41	2.80	2.69
12 3,4DNAN		0.27	2.08	0.15	2.33	2.41	2.80	2.35
13 1,3-DN-5-NsTz		2.10	2.27	2.25	2.31	6.00	5.33	1.74
14 1-N-3,5-DNsTz		2.20	2.40	2.44	2.43	6.00	6.00	3.20
15 1,3,5-TNsTz		0.96	3.63	1.08	3.69	6.00	6.00	>6.00
16 1-HxAm-3,5-DNTz		2.19	4.23	2.20	5.42	6.00	6.00	5.00
17 1-HxAm-3-Ns-5-NTz		2.28	4.17	2.39	4.37	6.00	6.00	4.50
18 1-HxAm-3,5-DNsTz		1.04	5.33	1.03	5.48	6.00	6.00	5.87
19 CL-14	117907-74-1	1.72	3.30	1.70	3.30	6.00	3.50	N/A
20 Pentryl	4481-55-4	1.48	-1.65	2.98	-1.73	1.54	1.96	0.14
21 ANTA		1.44	>6.00	1.59	>6.00	6.00	5.13	4.54
22 FOX-7	145250-81-3	1.13	>6.00	0.47	>6.00	6.00	6.00	5.70
23 HBT		2.50	>6.00	1.60	>6.00	6.00	6.00	>6.00

<sup>a</sup> Full names and structural formulas are given in SI Table S1. <sup>b</sup> Frag- $S_w$  values calculated from fragments,  $K_{ow}$ - $S_w$  values calculated from octanol/water partition coefficients ( $K_{ow}$ s are theoretically estimated in EPI as well).

an increase of water solubility with substitution of nitro groups in with nitroso and, especially, amino groups in the derivatives.

**3.5. Environmental Implications.** The environmental fate of chemicals, as well as the efficiency of in-solution treatment is determined largely by their solubility in water.

Hence, solubility of hazardous chemicals is one of the most fundamental physical properties of ecological importance (48). Unfortunately, due to the highly exothermic nature of explosives, experimental data on many of their properties are quite limited.

The proposed in the current study model is based on the quantum chemical methods, COSMO-RS, and SiRMS QSPR techniques. This powerful theoretical tool serves for prediction of water solubility of nitro compounds and their nitroso and amino derivatives. Furthermore, this computational approach can be extended for treatment of a wide range of organic contaminants. Thus, this tool can be useful for the determination of the environmental fate and development of degradation and remediation pathways of existing and possible environmental contaminants, hazardous chemicals, drugs, etc. It is especially useful for treatment of hazardous, toxic or unstable compounds which are hard to be threaded experimentally.

It should be noted that the current model limited to study of rather low soluble compounds, and can result in large errors in the case of highly soluble solutes. In addition, for moderately to highly soluble compounds the errors of 0.5 order of magnitude give large numerical errors if compared to low-soluble compounds.

## Acknowledgments

The support by the Basic Research Project BT25-08-41 grant from the U.S. Army Engineer Research and Development Center (ERDC) is acknowledged. Permission was granted by the Chief of Engineers to publish this information. 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. We are thankful to Dr. Andreas Klamt and Dr. Frank Eckert from COSMOlogic GmbH & Co KG for providing the information for the training set compounds and enlightening discussions of the COSMO-RS theory.

## Supporting Information Available

Table S1 contains the names, CAS registry numbers and structural formulas of the substances studied in the current work. Table S2 presents the parameters for training set for SiRMS model are collected. Table S3 displays the information on the descriptors for the SiRMS QSPR model. Table S4 shows the parameters obtained during validation of the developed QSPR model. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Literature Cited

- (1) Patai, S. *The Chemistry of Amino, Nitroso, and Nitro Compounds and Their Derivatives*; John Wiley & Sons Inc.: New York, 1982.
- (2) Feuer, H.; Nielsen, A. T., Eds. *Nitro Compounds: Recent Advances in Synthesis and Chemistry*; VCH Publishing: New York, 1990.
- (3) Neilson, A. H.; Allard, A.-S.; *Environmental Degradation and Transformation of Organic Chemicals*; CRC Press, Boca Raton, FL, 2008.
- (4) Talmage, S. S.; Oprea, D. M.; Maxwell, C. J.; Welsh, C. J. E.; Cretella, F. M.; Reno, P. H.; Daniel, F. B. Nitroaromatic munition compounds: environmental effects and screening values. *Rev. Environ. Contam. Toxicol.* **1999**, *161*, 1–156.
- (5) Rickert, D. E. *Toxicity of Nitroaromatic Compounds*; Hemisphere Publishing Corp.: Bristol, PA, 1984.
- (6) Robidoux, P. Y.; Svendsen, C.; Caumartin, J.; Hawari, J.; Ampleman, G.; Thiboutot, S.; Weeks, J. M.; Sunahara, G. I. Chronic toxicity of energetic compounds in soil determined using the earthworm (*Eisenia andrei*) reproduction test. *Environ. Toxicol. Chem.* **2000**, *19* (7), 1764–1773.
- (7) Lewandowski, G. A.; DeFilippi, L. J. *Biological Treatment of Hazardous Wastes*; John Wiley & Sons Inc.: New York, 1998.
- (8) Hawari, J.; Beaudet, S.; Halasz, A.; Thiboutot, S.; Ampleman, G. Microbial degradation of explosives: Biotransformation versus mineralization. *App. Microbiol. Biotechnol.* **2000**, *54* (5), 605–618.
- (9) Gorontzy, T.; Küver, J.; Blotevogel, K. H. Microbial transformation of nitroaromatic compounds under anaerobic conditions. *J. Gen. Microbiol.* **1993**, *139* (6), 1331–1336.
- (10) Gorontzy, T.; Drzyzga, O.; Kahl, M. W.; Bruns-Nagel, D.; Breitung, J.; von Loew, E.; Blotevogel, K. H. Microbial degradation of explosives and related compounds. *Crit. Rev. Microbiol.* **1994**, *20* (4), 265–284.
- (11) Kholod, Y.; Kosenkov, D.; Okovytyy, S.; Gorb, L.; Qasim, M.; Leszczynski, J. CL-20 photodecomposition: Ab initio foundations for identification of products. *Spectrochim. Acta, Part A* **2008**, *71* (1), 230–237.
- (12) Hawari, J.; Halasz, A.; Groom, C.; Deschamps, S.; Paquet, L.; Beaulieu, C.; Corriveau, A. Photodegradation of RDX in aqueous solution: A mechanistic probe for biodegradation with *Rhodococcus* sp. *Environ. Sci. Technol.* **2002**, *36* (23), 5117–5123.
- (13) Hawari, J.; Deschamps, S.; Beaulieu, C.; Paquet, L.; Halasz, A. Photodegradation of CL-20: Insights into the mechanisms of initial reactions and environmental fate. *Water Res.* **2004**, *38* (19), 4055–4064.
- (14) Qasim, M.; Fredrickson, H.; Honea, P.; Furey, J.; Leszczynski, J.; Okovytyy, S.; Szecsody, J.; Kholod, Y. Prediction of CL-20 chemical degradation pathways, theoretical and experimental evidence for dependence on competing modes of reaction. *SAR QSAR Environ. Res.* **2005**, *16* (5), 495–515.
- (15) Okovytyy, S.; Kholod, Y.; Qasim, M.; Fredrickson, H.; Leszczynski, J. The mechanism of unimolecular decomposition of 2,4,6,8,10,12-hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane. A computational DFT study. *J. Phys. Chem. A* **2005**, *109* (12), 2964–2970.
- (16) Liou, M.-J.; Lu, M.-C.; Chen, J.-N. Oxidation of explosives by Fenton and photo-Fenton processes. *Water Res.* **2003**, *37* (13), 3172–3179.
- (17) Kemper, J. M.; Ammar, E.; Mitch, W. A. Abiotic degradation of hexahydro-1,3,5-trinitro-1,3,5-triazine in the presence of hydrogen sulfide and black carbon. *Environ. Sci. Technol.* **2008**, *42* (6), 2118–2123.
- (18) Balakrishnan, V. K.; Halasz, A.; Hawari, J. Alkaline hydrolysis of the cyclic nitramine explosives RDX, HMX, and CL-20: New insights into degradation pathways obtained by the observation of novel intermediates. *Environ. Sci. Technol.* **2003**, *37* (9), 1838–1843.
- (19) Thorn, K. A.; Thorne, P. G.; Cox, L. G. Alkaline hydrolysis/polymerization of 2,4,6-trinitrotoluene: Characterization of products by <sup>13</sup>C and <sup>15</sup>N NMR. *Environ. Sci. Technol.* **2004**, *38* (7), 2224–2231.
- (20) Karakaya, P.; Sidhoum, M.; Christodoulatos, C.; Nicolich, S.; Balas, W. Aqueous solubility and alkaline hydrolysis of the novel high explosive hexanitrohexaazaisowurtzitane (CL-20). *J. Hazard. Mater.* **2005**, *120* (1–3), 183–191.
- (21) Kholod, Y.; Okovytyy, S.; Kuramshina, G.; Qasim, M.; Gorb, L.; Furey, J.; Honea, P.; Fredrickson, H.; Leszczynski, J. Are 1,5- and 1,7-dihydroimidazo(4,5-b:4,5'-e)pyrazine the main products of 2,4,6,8,10,12-hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane (CL-20) alkaline hydrolysis? A DFT study of vibrational spectra. *J. Mol. Struct.* **2006**, *794* (1–3), 288–302.
- (22) Jain, N.; Yalkowsky, S. H. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J. Pharm. Sci.* **2001**, *90* (2), 234–252.
- (23) Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of aqueous solubility of organic molecules by the group contribution approach: application to the study of biodegradation. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 474–482.
- (24) Tropsha, A.; Gramatica, P.; Gombar, V. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.
- (25) Klamt, A. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- (26) Klamt, A.; Eckert, F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.* **2000**, *172*, 43–72.
- (27) Klamt, A.; Schuurmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799–805.
- (28) Klamt, A.; Eckert, F.; Hornig, M.; Beck, M. E.; Brger, T. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. *J. Comput. Chem.* **2002**, *23* (2), 275–281.
- (29) Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P.; *The Properties of Gases and Liquids*, 5th ed.; McGraw-Hill: New York, 2001.



- (30) Toghiani, R. K.; Toghiani, H.; Maloney, S. W.; Boddu, V. M. Prediction of physicochemical properties of energetic materials. *Fluid Phase Equilib* **2008**, *264*, 86–92.
- (31) Qasim, M.; Kholod, Y.; Gorb, L.; Magers, D.; Honea, P.; Leszczynski, J. Application of quantum-chemical approximations to environmental problems: Prediction of physical and chemical properties of TNT and related species. *Chemosphere* **2007**, *69* (7), 1144–1150.
- (32) Eckert F.; Klamt A. *COSMOtherm*, Ver. C2.1 Rel. 01.05; COSMOlogicLeverkusen, Germany, 2002.
- (33) Dewar, M.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
- (34) Stewart, J. J. P. *MOPAC 2000*; Fujitsu Limited: Tokyo, Japan, 2009.
- (35) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- (36) Schaefer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03, Revision B.01*, Gaussian, Inc.: Wallingford CT, 2004.
- (38) Schafer, A.; Huber, C.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple- $\zeta$  valence quality for atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829–5835.
- (39) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- (40) Kuz'min, V. E.; Artemenko, A. G.; Muratov, E. N. Hierarchical QSAR technology on the base of simplex representation of molecular structure. *J. Comp. Aid. Mol. Des.* **2008**, *22*, 403–421.
- (41) Ranner, S.; Lindgren, F.; Geladi, P.; Wold, S. A PLS kernel algorithm for data sets with many variables and fewer objects. *J. Chemometr.* **1994**, *8*, 111–125.
- (42) PhysProp Database, SRCTec; <http://www.srcinc.com/what-we-do/databaseforms.aspx?id=386>.
- (43) Duffy, E. M.; Jorgensen, W. L. Prediction of properties from simulations: Free energies of solvation in hexadecane, octanol, and water. *J. Am. Chem. Soc.* **2000**, *122*, 2878–2888.
- (44) Wassvik, C. M.; Holmen, A. G.; Bergstrom, C. A. S.; Zamora, I.; Artursson, P. Contribution of solid-state properties to the aqueous solubility of drugs. *Eur. J. Pharm. Sci.* **2006**, *29*, 294–305.
- (45) McFarland, J. W.; Avdeef, A.; Berger, C. M.; Raevsky, O. A. *Poster presented at QSAR2000, Dusseldorf*, Germany, 2000.
- (46) Golbraikh, A.; Tropsha, A. Beware of  $q^2$ . *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.
- (47) Oleszek-Kudlak, S.; Grabda, M.; Shibata, E.; Eckert, F.; Nakamura, T. Application of the COSMO-RS method for prediction of the aqueous solubility of chlorobenzenes depending on temperature and salinity. *Environ. Toxicol. Chem.* **2005**, *24* (6), 1368–1375.
- (48) Lynch, J. C.; Myers, K. F.; Brannon, J. M.; Delfino, J. J. Effects of pH and temperature on the aqueous solubility and dissolution rate of 2,4,6-trinitrotoluene (TNT), hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), and octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (HMX). *J. Chem. Eng. Data* **2001**, *46*, 1549–1555.

ES902566B