

**Application of COSMO-SAC to Solid Solubility in Pure and Mixed Solvent
Mixtures for Organic Pharmacological Compounds**

Paul Eric Mullins

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Chemical Engineering

Dr. Y. A. Liu, Chair
Dr. Richey M. Davis
Dr. Amadeu K. Sum

24 January 2007
Blacksburg, Virginia

Keywords: COSMO, COSMO-SAC, Solubility, Pharmaceutical, Mixed Solvent

Application of COSMO-SAC to Solid Solubility in Pure and Mixed Solvent Mixtures for Organic Pharmacological Compounds

Paul Eric Mullins

Dr. Y. A. Liu, Chair

Abstract

In this work, we present two open literature databases, the VT-2005 Sigma Profile Database and the VT-2006 Solute Sigma Profile Database, that contain sigma profiles for 1,645 unique compounds. A sigma profile is a molecular-specific distribution of the surface-charge density, which enables the application of solvation-thermodynamic models to predict vapor-liquid and solid-liquid equilibria, and other properties. The VT-2005 Sigma Profile Database generally focuses on solvents and small molecules, while the VT-2006 Solute Sigma Profile Database primarily consists of larger, pharmaceutical-related solutes. We design both of these databases for use with the conductor-like screening model – segment activity coefficient (COSMO-SAC), a liquid-phase activity-coefficient model. The databases contain the necessary information to perform binary and multicomponent VLE and SLE predictions. We offer detailed tutorials and procedures for use with our programs so the reader may also use their own research on our research group website (www.design.che.vt.edu). We validate the VT-2005 Sigma Profile Database by pure component vapor pressure predictions and validate the VT-2006 Solute Sigma Profile Database by solid solubility predictions in pure solvents compared with literature data from multiple sources. Using both databases, we also explore the application of COSMO-SAC to solubility predictions in mixed solvents. This work also studies the effects of conformational isomerism on VLE and SLE property prediction. Finally, we compare COSMO-SAC solubility predictions to solubility predictions by the Non-Random Two-Liquid, Segment Activity Coefficient (NRTL-SAC) model. We find UNIFAC is a more accurate method for predicting VLE behavior than the COSMO-SAC model for many of the systems studied, and that COSMO-SAC predicts solute mole fraction in pure solvents with an average root-mean-squared error ($\log_{10}(x_{sol})$) of 0.74, excluding outliers, which is greater than the RMS error value of 0.43 using the NRTL-SAC model.

Acknowledgments

I would like to thank my advisor, Dr. Y. A. Liu, for his support and guidance during my tenure as a student at Virginia Tech. I would also like to thank the members of Dr. Liu's research group, especially Stephen Fast, Anthony Gaglione, Adel Ghaderi, Bruce Lucas, Richard Oldland, and Kevin Seavey, for being an excellent sounding board for ideas. My parents, Phillip and Joalyn Mullins, and my family deserve special thanks for pushing me to try and be my best in all endeavors in life. I would also like to thank my professional collaborators, Dr. Stanley I. Sandler and his research group at the University of Delaware and Dr. Chau-Chyun Chen, Senior Technology Fellow at Aspen Technology, Inc. Both of these individuals have provided sound advice and guidance backed by many years of experience, and I greatly value their opinions. I would also like to thank our industrial sponsors, Alliant Techsystems, Aspen Technology, Inc., Formosa Petrochemical Co., Honeywell International, Milliken & Co., and SINOPEC Corporation., for providing the funding for the research presented here.

Paul Eric Mullins

January 24, 2007

Blacksburg, Virginia

Table of Contents

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	ix
1 INTRODUCTION	1
1.1 BACKGROUND AND SIGNIFICANCE OF THIS RESEARCH.....	1
1.2 OBJECTIVES OF THIS RESEARCH.....	2
2 THEORY	4
2.1 OVERVIEW OF COSMO-SAC MODEL.....	4
2.1.1 <i>Derivation of Segment Activity-Coefficient Model</i>	5
2.1.2 <i>Sigma Profiles</i>	8
2.2 OVERVIEW OF DENSITY-FUNCTIONAL THEORY	11
2.3 SOLUBILITY THEORY.....	12
2.3.1 <i>Overview of Solubility</i>	12
2.3.2 <i>Derivation of Solubility Equation from Solid-Liquid Equilibrium</i>	12
2.4 ESTIMATION METHODS FOR MELTING TEMPERATURE AND HEATS OF FUSION	15
2.4.1 <i>Estimation of Normal Melting-Point Temperature</i>	15
2.4.2 <i>Estimation of Latent Heat of Fusion</i>	16
2.5 OVERVIEW OF NRTL-SAC THEORY	18
3 COMPUTATIONAL METHODS AND VALUES	20
3.1 PROCEDURE FOR GENERATING SIGMA PROFILES	20
3.1.1 <i>Ideal Gas-Phase Geometry Optimization</i>	21
3.1.2 <i>Condensed Phase Energy Calculation</i>	22
3.1.3 <i>COSMO Calculation – Surface Segment Averaging</i>	23
3.2 GEOMETRY OPTIMIZATION TOOLS	24
3.2.1 <i>Amber8</i>	24
3.2.2 <i>MS Forcite Plus Annealing Dynamics</i>	25
3.2.3 <i>Qualitative Comparison of Methods</i>	26
3.3 CONVERGENCE METHOD FOR SOLUBILITY PREDICTIONS.....	28
4 VT-2005 SIGMA PROFILE DATABASE	29
4.1 DESCRIPTION OF DATABASE.....	29
4.2 VALIDATION OF VT-2005 SIGMA PROFILE DATABASE.....	33
4.2.1 <i>Pure Component Vapor-Pressure Predictions</i>	34
4.2.2 <i>Activity-Coefficient Predictions</i>	35
4.3 DETAILS OF VT-2005 SIGMA PROFILE DATABASE RELEASE AND REFINEMENT.....	35
4.4 EFFECT OF CONFORMATIONAL ISOMERISM ON THERMODYNAMIC PROPERTY PREDICTION	36
4.4.1 <i>Small Molecule Example: 2-Methoxy-ethanol</i>	38
4.4.2 <i>Medium-Sized Molecule Example: Benzyl Benzoate</i>	47
4.5 CONFORMATIONAL IMPROVEMENTS OF SELECT RELEASED COMPOUNDS	52
4.5.1 <i>Polar Solvent: Ethanol</i>	53
4.5.2 <i>Nonpolar Solvent: Cyclohexane</i>	58
4.6 VLE PREDICTION FOR MULTICOMPONENT SYSTEMS	64
4.7 APPLICATION HEURISTICS FOR COSMO-BASED METHODS.....	65
4.7.1 <i>Applicability of General COSMO-Based Methods</i>	66

4.7.2	<i>Sensitivity Effects of Sigma Profiles</i>	66
4.7.3	<i>COSMO-Based Methods and Nitrogen-Containing Compounds</i>	66
4.7.4	<i>Polymer Applications</i>	67
4.7.5	<i>Model Accuracy Regarding Chloroform Mixtures with Ketones or Alcohols</i>	67
4.7.6	<i>Selection of Appropriate Model Parameters</i>	67
4.7.7	<i>Limitations of Validated Atomic Radii</i>	67
4.7.8	<i>Applicability of UNIFAC and COSMO-Based Methods</i>	68
4.7.9	<i>High Temperature and High Pressure Applications</i>	68
5	VT-2006 SOLUTE SIGMA PROFILE DATABASE	69
5.1	DESCRIPTION OF VT-2006 SOLUTE SIGMA PROFILE DATABASE	69
5.2	VALIDATION OF VT-2006 SOLUTE SIGMA PROFILE DATABASE	70
5.2.1	<i>Error Calculation Methodology and Quantification</i>	70
5.2.2	<i>Solid Solubility Predictions in Pure Solvents</i>	73
5.3	EFFECTS OF CONFORMATIONAL ISOMERISM ON SOLUBILITY PREDICTIONS	77
5.3.1	<i>Acetaminophen (VTSOL-082, CAS-RN: 103-90-2, C₈H₉NO₂)</i>	78
5.3.2	<i>Aspirin (VTSOL-088, VT-1422, CAS-RN: 50-78-2, C₉H₈O₄)</i>	80
5.3.3	<i>Ibuprofen (VTSOL-141, VT-1205, CAS-RN: 15687-27-1, C₁₃H₁₈O)</i>	84
5.3.4	<i>Lidocaine (VTSOL-148, CAS-RN: 137-58-6, C₁₄H₂₂N₂O)</i>	86
5.3.5	<i>Cholesterol (VTSOL-197, CAS-RN: 57-88-5, C₂₇H₄₆O)</i>	89
5.3.6	<i>Summary of Conformational Effects</i>	93
5.4	SOLID SOLUBILITY IN BINARY AND MIXED SOLVENTS.....	93
5.5	COMPARISON OF COSMO-SAC AND NRTL-SAC SOLUBILITY PREDICTIONS	95
5.6	APPLICATION GUIDELINES AND HEURISTICS OF COSMO-SAC SOLID SOLUBILITY PREDICTIONS	98
5.6.1	<i>Aqueous Solubility</i>	99
5.6.2	<i>n-Octanol Solubility</i>	100
5.6.3	<i>Solvent Conformational Effects on Solubility in Ethanol and Cyclohexane</i>	101
5.6.4	<i>Nitrogen-Containing Compound Solubility</i>	103
5.6.5	<i>Comparison of Exchange Energy Expressions of Lin and Sandler⁶ and Mathias et al¹³</i>	106
5.6.6	<i>Effect of Multicomponent Systems on Accuracy of Solubility Predictions</i>	106
5.6.7	<i>Overall Accuracy of COSMO-SAC as Solubility Predictor</i>	107
5.6.8	<i>Sensitivity to Melting-point Temperature and Latent Heat of Fusion</i>	107
5.6.9	<i>Effect of Conformation Pre-optimization with Amber 8 on Solubility Predictions</i>	108
6	RESOURCES	110
7	CONCLUSIONS	111
8	IMPROVEMENTS AND FUTURE WORK	112
9	NOMENCLATURE	113
10	REFERENCES	116
	APPENDICES	120
	APPENDIX A: SUPPLEMENTAL TABLES	121
	APPENDIX B: FORTRAN CODE AND SCRIPTS	157
	APPENDIX C: SAMPLE DATABASE ENTRY	180
VITA	196

List of Figures

FIGURE 2.1: CONCEPTUAL DIAGRAM OF IDEAL SOLVATION PROCESS WITH A COSMO-BASED MODEL.....	5
FIGURE 2.2: WATER (VT-1076) AND AMMONIA (VT-1070) SIGMA PROFILES AVAILABLE IN THE VT-2005 SIGMA PROFILE DATABASE. BOTH SIGMA PROFILES HAVE PEAKS ABOVE THE HYDROGEN BONDING CUTOFF SIGNIFYING THESE COMPOUNDS LIKELY FORM HYDROGEN BONDS.....	11
FIGURE 2.3: CARTOON REPRESENTATION OF THREE-STEP THERMODYNAMIC CYCLE TO CALCULATE FUGACITY RATIO OF SOLID PHASE FUGACITY F^S TO LIQUID-PHASE FUGACITY F^L	13
FIGURE 3.1: ENERGY MINIMIZATION OUTPUT FROM AMBER8 AND ANNEALING DYNAMICS OUTPUT FROM MS FORCITE PLUS FOR DI-N-DECYL-PHTHALATE (VT-0690). NEITHER CONFORMATION PRODUCED ACCEPTABLE RESULTS..	28
FIGURE 4.1: DISTRIBUTION OF PURE COMPONENT VAPOR-PRESSURE PREDICTIONS AT EACH COMPOUNDS RESPECTIVE NORMAL BOILING POINT TEMPERATURE WITH A REVISED COSMO-SAC-BP MODEL ⁵¹ ON A LOGARITHMIC SCALE.	35
FIGURE 4.2: TWO CONFORMATIONS OF 3-METHYL-2-BUTANOL (VT-0499), WHICH DIFFER ONLY IN THE RELATIVE BOND ANGLES OF THE HYDROXYL HYDROGEN ATOM.	37
FIGURE 4.3: TWO CONFORMATIONS AND SIGMA PROFILES OF 3-METHYL-2-BUTANOL (VT-0499) AFTER DMOL3 GEOMETRY OPTIMIZATION USING AMBER8 AND MS FORCITE AS INITIAL CONFORMATIONS.	38
FIGURE 4.4: LOW ENERGY CONFORMATIONS FOR 2-METHOXY-ETHANOL (VT-1397).	39
FIGURE 4.5: SIGMA PROFILES FOR VARIOUS CONFORMATIONS OF 2-METHOXY-ETHANOL (VT-1397). THE RELEASED CONFORMATION IS INCLUDED IN THE VT-2005 SIGMA PROFILE DATABASE.	39
FIGURE 4.6: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) CYCLOHEXANE AT 303.15 K. ⁵³	40
FIGURE 4.7: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) CYCLOHEXANE AT 313.15 K. ⁵³	41
FIGURE 4.8: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) CYCLOHEXANE AT 323.15 K. ⁵³	41
FIGURE 4.9: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) N-HEXANE AT 313.15 K. ⁵³	42
FIGURE 4.10: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) N-HEXANE AT 323.15 K. ⁵³	42
FIGURE 4.11: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) N-HEPTANE AT 323.15 K. ⁵³	43
FIGURE 4.12: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) METHANOL AT 298.15 K. ⁵³	43
FIGURE 4.13: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) METHYL ACETATE AT 298.15 K. ⁵⁴	44
FIGURE 4.14: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) DIISOPROPYL-ETHER AT 331.02 K. ⁵⁵	44
FIGURE 4.15: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) DIISOPROPYL-ETHER AT 341.01 K. ⁵⁵	45
FIGURE 4.16: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) TRICHLOROETHYLENE AT 341.01 K. ⁵⁵	45
FIGURE 4.17: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) 2-METHOXY-ETHANOL (2) TRICHLOROETHYLENE AT 357.98 K. ⁵⁵	46
FIGURE 4.18: FOUR CONFORMATIONS OF BENZYL BENZOATE AND THEIR RESPECTIVE SIGMA-PROFILE FOR COMPARISON.....	48
FIGURE 4.19: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) BENZENE (2) BENZYL BENZOATE AT 453.23 K. ⁵⁶ THE VAPOR PHASE IS A PREDICTION ONLY.	49
FIGURE 4.20: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) TOLUENE (2) BENZYL BENZOATE AT 453.25 K. ⁵⁶ THE VAPOR PHASE IS A PREDICTION ONLY.	50
FIGURE 4.21: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) BENZALDEHYDE (2) BENZYL BENZOATE AT 453.25 K. ⁵⁶ THE VAPOR PHASE IS A PREDICTION ONLY.	50
FIGURE 4.22: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) BENZYL ALCOHOL (2) BENZYL BENZOATE AT 453.23 K. ⁵⁶ THE VAPOR PHASE IS A PREDICTION ONLY.	51

FIGURE 4.23: PRESSURE-COMPOSITION DATA AND COSMO-SAC PREDICTIONS FOR (1) PHENOL (2) BENZYL BENZOATE AT 453.26 K. ⁵⁶ THE VAPOR PHASE IS A PREDICTION ONLY.	51
FIGURE 4.24: VARIATIONS IN THE RELEASED CONFORMATION OF ETHANOL (VT-0478) AND AN IMPROVED CONFORMATION AND THE EFFECTS ON THEIR RESPECTIVE SIGMA PROFILES.	53
FIGURE 4.25: PRESSURE-COMPOSITION CURVE OF (1) ETHANOL (2) WATER PREDICTED BY THE COSMO-SAC MODEL FOR THE RELEASED CONFORMATION AND CONFORMATION A AT 298.15 K, 323.15 K, AND 338.15 K IN COMPARISON TO LITERATURE DATA. ⁵⁷	54
FIGURE 4.26: PRESSURE-COMPOSITION CURVE FOR AN (1) ETHANOL (2) ACETONE MIXTURE AT 298.15 K AS PREDICTED BY THE COSMO-SAC MODEL FOR THE RELEASED CONFORMATION AND CONFORMATION A WITH A COMPARISON TO LITERATURE DATA. ⁵⁷ THE VAPOR PHASE IS A PREDICTION ONLY.	55
FIGURE 4.27: PRESSURE-COMPOSITION CURVE OF (1) METHANOL (2) ETHANOL PREDICTED BY THE COSMO-SAC MODEL FOR THE RELEASED CONFORMATION AND CONFORMATION A AT 298.15 K IN COMPARISON TO LITERATURE DATA. ⁵⁷	55
FIGURE 4.28: PRESSURE-COMPOSITION CURVE OF (1) ETHANOL (2) TETRACHLOROMETHANE PREDICTED BY THE COSMO-SAC MODEL FOR THE RELEASED CONFORMATION AND CONFORMATION A AT 298.15 K IN COMPARISON TO LITERATURE DATA. ⁵⁷	56
FIGURE 4.29: PRESSURE-COMPOSITION CURVE OF (1) ETHANOL (2) BENZENE PREDICTED BY THE COSMO-SAC MODEL FOR THE RELEASED CONFORMATION AND CONFORMATION A AT 298.15 K IN COMPARISON TO LITERATURE DATA. ⁵⁷	56
FIGURE 4.30: PRESSURE-COMPOSITION CURVE OF (1) ETHANOL (2) OCTANE PREDICTED BY THE COSMO-SAC MODEL FOR THE RELEASED CONFORMATION AND CONFORMATION A AT 318.15 K IN COMPARISON TO LITERATURE DATA. ⁵⁷	57
FIGURE 4.31: VARIATIONS IN THE RELEASED CONFORMATION OF “BOAT” CYCLOHEXANE (VT-0099) AND THE “CHAIR” CONFORMATION AND THE EFFECTS ON THEIR RESPECTIVE SIGMA PROFILES.	59
FIGURE 4.32: COSMO-SAC PREDICTED PRESSURE FOR (1) ETHANOL (2) CYCLOHEXANE AT 298.15 K. CYCLOHEXANE, CYCLOHEXANE A, ETHANOL, AND ETHANOL A REFER TO THE RELEASED CONFORMATION AND CONFORMATION A FOR THEIR RESPECTIVE SPECIES. ⁵⁷	60
FIGURE 4.33: COSMO-SAC PREDICTED PRESSURE TWO CYCLOHEXANE CONFORMATIONS FOR (1) BENZENE (2) CYCLOHEXANE AT 298.15 K. ⁵⁷	60
FIGURE 4.34: COSMO-SAC PREDICTED PRESSURE TWO CYCLOHEXANE CONFORMATIONS FOR (1) OCTANE (2) CYCLOHEXANE AT 298.15 K. ⁵⁷ THE VAPOR PHASE IS A PREDICTION ONLY.	61
FIGURE 4.35: COSMO-SAC PREDICTED PRESSURE TWO CYCLOHEXANE CONFORMATIONS FOR (1) PYRIDINE (2) CYCLOHEXANE AT 298.15 K. ⁵⁷ THE VAPOR PHASE IS A PREDICTION ONLY.	61
FIGURE 4.36: COSMO-SAC PREDICTED PRESSURE TWO CYCLOHEXANE CONFORMATIONS FOR (1) 1,4-DIOXANE (2) CYCLOHEXANE AT 298.15 K. ⁵⁷ THE VAPOR PHASE IS A PREDICTION ONLY.	62
FIGURE 4.37: COSMO-SAC PREDICTED PRESSURE TWO CYCLOHEXANE CONFORMATIONS FOR (1) 1,2-DICHLOROETHANE (2) CYCLOHEXANE AT 298.15 K. ⁵⁷ THE VAPOR PHASE IS A PREDICTION ONLY.	62
FIGURE 5.1: RMSE AND AA%E BEHAVIOR IN RELATION TO SOLUBILITY MODELING OVER-PREDICTION AND UNDER-PREDICTION ON A LOGARITHMIC BASE 10 SCALE.	72
FIGURE 5.2: PREDICTED BENZOIC ACID (VTSOL-044, VT-0610) SOLUBILITY IN 50 PURE SOLVENTS AT 298.15 K USING BOTH DEFINITIONS FOR THE EXCHANGE ENERGY COMPARED WITH EXPERIMENTAL VALUES ON A LOGARITHMIC BASE 10 SCALE. ²⁷	73
FIGURE 5.3: PREDICTED DIBENZOFURAN (VTSOL-120, VT-0770) SOLUBILITY IN FOUR PURE SOLVENTS AT VARIOUS TEMPERATURES USING BOTH EXCHANGE ENERGY EXPRESSIONS COMPARED WITH EXPERIMENTAL SOLUBILITY ON A LOGARITHMIC BASE 10 SCALE. ²⁷	75
FIGURE 5.4: THREE ACETAMINOPHEN CONFORMATIONS WITH VARIATIONS IN THE POSITION OF THE HYDROXYL GROUP HYDROGEN ATOM RELATIVE TO THE AROMATIC RING.	78
FIGURE 5.5: SIGMA PROFILES FOR THEIR RESPECTIVE ACETAMINOPHEN CONFORMATIONS, A, B, AND C.	79
FIGURE 5.6: PREDICTED ACETAMINOPHEN SOLUBILITY IN 26 PURE SOLVENTS COMPARED WITH EXPERIMENTAL VALUES ²⁷ AT VARIOUS TEMPERATURES FOR EACH CONFORMATION USING THE EXCHANGE ENERGY DEFINED BY LIN AND SANDLER ⁶ ON A LOGARITHMIC BASE 10 SCALE.	80
FIGURE 5.7: THREE OPTIMIZED ASPIRIN (VTSOL-088) CONFORMATIONS WITH VARIATIONS IN THE RELATIVE POSITIONS OF THE CARBOXYL GROUP IN PROXIMITY TO THE ESTER LINKAGE.	81
FIGURE 5.8: SIGMA PROFILES FOR THEIR RESPECTIVE ASPIRIN CONFORMATIONS, A, B, AND C.	81

FIGURE 5.9: PREDICTED ASPIRIN SOLUBILITY IN 15 PURE SOLVENTS COMPARED WITH EXPERIMENTAL VALUES ²⁷ AT VARIOUS TEMPERATURES FOR EACH CONFORMATION USING THE EXCHANGE ENERGY DEFINED BY LIN AND SANDLER ⁶ ON A LOGARITHMIC BASE 10 SCALE.	83
FIGURE 5.10: THREE OPTIMIZED IBUPROFEN CONFORMATIONS WITH VARIATIONS IN THE POSITION OF THE CARBOXYL GROUP IN RELATION TO THE ISOBUTYL FUNCTIONAL GROUP.	84
FIGURE 5.11: SIGMA PROFILES FOR THEIR RESPECTIVE IBUPROFEN CONFORMATIONS, A, B, AND C.	85
FIGURE 5.12: PREDICTED IBUPROFEN SOLUBILITY IN 25 PURE SOLVENTS ²⁷ FOR EACH CONFORMATION USING THE EXCHANGE ENERGY DEFINED BY LIN AND SANDLER ⁶ AT VARIOUS TEMPERATURES ON A LOGARITHMIC BASE 10 SCALE.	86
FIGURE 5.13: FOUR LIDOCAINE CONFORMATIONS AND THEIR VARIATIONS IN STRUCTURE CONCERNING THE RELATIVE POSITIONS OF THE AMIDE AND AMINE FUNCTIONAL GROUPS.	87
FIGURE 5.14: SIGMA PROFILES FOR THEIR RESPECTIVE LIDOCAINE CONFORMATIONS.	88
FIGURE 5.15: PREDICTED LIDOCAINE SOLUBILITY IN 20 PURE SOLVENTS USING THE EXCHANGE ENERGY EXPRESSION DEFINED BY LIN AND SANDLER ⁶ COMPARED TO THEIR EXPERIMENTAL VALUE AT 298.15 K ON A LOGARITHMIC BASE 10 SCALE. ²⁷	89
FIGURE 5.16: SIX CHOLESTEROL CONFORMATIONS WITH CIRCLED REGIONS TO EMPHASIZE THEIR RELATIVE DIFFERENCES. CONFORMATIONS A AND F HAVE SIMILAR ATOMIC POSITIONS FOR ALL ATOMS EXCEPT THE HYDROXYL HYDROGEN PLACEMENT.	90
FIGURE 5.17: SIGMA PROFILE FOR SIX CHOLESTEROL CONFORMATIONS.	91
FIGURE 5.18: PREDICTED CHOLESTEROL SOLUBILITY IN 60 PURE SOLVENTS FOR EACH CONFORMATION USING THE EXCHANGE ENERGY EQUATION DEFINED BY LIN AND SANDLER ⁶ AT VARIOUS TEMPERATURES COMPARED WITH THEIR EXPERIMENTAL VALUES ON A LOGARITHMIC BASE 10 SCALE. ²⁷	92
FIGURE 5.19: NRTL-SAC AND COSMO-SAC PREDICTED SOLUBILITIES FOR 17 SOLUTES AND 258 EXPERIMENTAL SOLUBILITY POINTS AT 298.15 K, EXCEPT ACETAMINOPHEN SOLUBILITY AT 303.15 K, ON A LOGARITHMIC BASE 10 SCALE. ²⁷	97
FIGURE 5.20: PREDICTED SOLUBILITY IN PURE WATER USING THE EXCHANGE ENERGY DEFINED BY LIN AND SANDLER ⁶ COMPARED WITH LITERATURE VALUES FOR 144 SOLUTES AND 438 SOLUBILITY POINTS ON A LOGARITHMIC BASE 10 SCALE. ²⁷	99
FIGURE 5.21: PREDICTED SOLUBILITY IN PURE N-OCTANOL USING THE EXCHANGE ENERGY DEFINED BY LIN AND SANDLER ⁶ COMPARED WITH LITERATURE VALUES FOR 52 SOLUTES AND 73 SOLUBILITY POINTS ON A LOGARITHMIC BASE 10 SCALE. ²⁷	101
FIGURE 5.22: PREDICTED SOLUBILITY IN PURE ETHANOL FOR 50 SOLUTES AT VARIOUS TEMPERATURES USING LIN AND SANDLER EXCHANGE ENERGY EXPRESSION COMPARED WITH THEIR EXPERIMENTAL VALUES ON A LOGARITHMIC BASE 10 SCALE. ²⁷	102
FIGURE 5.23: PREDICTED SOLUBILITY IN PURE CYCLOHEXANE FOR 39 SOLUTES AT VARIOUS TEMPERATURES USING LIN AND SANDLER EXCHANGE ENERGY EXPRESSION COMPARED WITH LITERATURE VALUES ON A LOGARITHMIC BASE 10 SCALE. ²⁷	102
FIGURE 5.24: PREDICTED NITROGEN-CONTAINING SOLUBILITY FOR BOTH EXCHANGE ENERGY EXPRESSIONS COMPARED WITH LITERATURE VALUES FOR 93 SOLUTES (664 SOLUBILITY POINTS) ON A LOGARITHMIC BASE 10 SCALE. ²⁷	104
FIGURE 5.25: GENERAL CHEMICAL STRUCTURES FOR THE FOLLOWING FUNCTIONAL GROUPS IN THIS ORDER: AMINE, AMIDE, NITRILE, NITRO, PYRIDINE, AMINOSULFONYL	105
FIGURE 5.26: RMSE SENSITIVITY OF PREDICTED SOLUBILITIES AS A FUNCTION OF NORMAL MELTING-POINT TEMPERATURE AND THE LATENT HEAT OF FUSION.	108

List of Tables

TABLE 2.1: SUMMARY OF ERROR FOR MELTING-POINT TEMPERATURE AND ENTHALPY OF FUSION ESTIMATION METHODS	18
TABLE 3.1: COSMO KEYWORDS USED IN CALCULATING SURFACE SEGMENT CHARGES IN DMOL3 ^A	22
TABLE 3.2: ATOMIC RADII FOR CREATING THE COSMO MOLECULAR CAVITY ^{2,4,49}	23
TABLE 3.3: PARAMETER VALUES USED IN THE COSMO-SAC MODEL. ^{2,6}	23
TABLE 3.4: MS FORCITE PLUS ANNEAL DYNAMICS CALCULATION SETTINGS ⁴⁵	26
TABLE 3.5: COMPARISON OF METHODS, AMBER8 AND MS FORCITE PLUS, AS GEOMETRY OPTIMIZATION TOOLS BASED ON CONDENSED PHASE DMOL3 ENERGY VALUES E_H AND PURE COMPONENT VAPOR PRESSURE AS PREDICTED BY COSMO-SAC-BP ⁵¹ FOR 93 COMPOUNDS WHICH ORIGINALLY FAILED TO MEET ACCEPTABLE ERROR CRITERIA.	27
TABLE 4.1: SUMMARY OF THE VT-2005 SIGMA PROFILE DATABASE	29
TABLE 4.2: AA%E OF VAPOR MOLE FRACTION 2-METHOXY-ETHANOL, % M.F.	47
TABLE 4.3: PREDICTED AZEOTROPIC COMPOSITIONS FOR NON-POLAR SOLVENTS WITH 2-METHOXY-ETHANOL, M.F.	47
TABLE 4.4: AA%E OF SOLVENT VAPOR PRESSURE, KPA, FOR EACH BENZYL BENZOATE SYSTEM.	52
TABLE 4.5: AA%E OF SYSTEM PRESSURE, MM HG, FOR MULTIPLE SYSTEMS WITH VARIOUS SOLVENTS AT VARIOUS TEMPERATURES FOR BOTH CONFORMATIONS OF ETHANOL.	58
TABLE 4.6: AA%E OF SYSTEM PRESSURE, MM HG, FOR MULTIPLE SYSTEMS WITH VARIOUS SOLVENTS AT VARIOUS TEMPERATURES FOR BOTH CONFORMATIONS OF CYCLOHEXANE.	63
TABLE 4.7: AA%E OF PREDICTED PRESSURE AND VAPOR COMPOSITION COMPARED WITH LITERATURE VALUES. ⁵⁷ WE USE BOTH CONFORMATIONS FOR CYCLOHEXANE AND ETHANOL FROM SECTION 4.5 FOR APPLICABLE SYSTEMS.	65
TABLE 5.1: OVERALL ERROR FOR PREDICTED BENZOIC ACID SOLUBILITY IN 50 PURE SOLVENTS.	74
TABLE 5.2: ERROR SUMMARY, RMSE AND AA%E, OF PREDICTED SOLUBILITY IN PURE SOLVENTS IN COMPARISON TO EXPERIMENTAL VALUES FOR THE ENTIRE SAMPLE SET AND SELECT SOLUTES WHICH EXCLUDE OUTLIERS.	76
TABLE 5.3: NINETEEN SOLUTES WITH AN OVERALL AA%E EXCEEDING THE ERROR CUTOFF VALUE OF 5000%.	77
TABLE 5.4: ERROR SUMMARY, RMSE AND AA%E, FOR EACH ACETAMINOPHEN CONFORMATION IN 26 PURE SOLVENTS AT VARIOUS TEMPERATURES COMPARING BOTH EXCHANGE ENERGY EXPRESSIONS.	79
TABLE 5.5: ERROR SUMMARY, RMSE AND AA%E, FOR EACH ASPIRIN CONFORMATION IN 15 PURE SOLVENTS AT VARIOUS TEMPERATURES COMPARING BOTH EXCHANGE ENERGY EXPRESSIONS.	82
TABLE 5.6: ERROR SUMMARY FOR EACH ASPIRIN CONFORMATION IN 14 PURE SOLVENTS, EXCLUDING CYCLOHEXANE, COMPARING BOTH EXCHANGE ENERGY EXPRESSIONS TO EXPERIMENTAL VALUES. ²⁷	83
TABLE 5.7: ERROR SUMMARY FOR EACH IBUPROFEN CONFORMATION IN 20 PURE SOLVENTS, EXCLUDING ACETAPHENONE, 1,2-PROPANEDIOL, AND DIETHYL ETHER, COMPARING BOTH EXCHANGE ENERGY EXPRESSIONS TO EXPERIMENTAL VALUES. ²⁷	85
TABLE 5.8: ERROR SUMMARY FOR EACH LIDOCAINE CONFORMATION IN 20 PURE SOLVENTS COMPARING BOTH EXCHANGE ENERGY EXPRESSIONS TO EXPERIMENTAL VALUES. ²⁷	88
TABLE 5.9: ERROR SUMMARY FOR EACH CHOLESTEROL CONFORMATION IN 60 PURE SOLVENTS COMPARING BOTH EXCHANGE ENERGY EXPRESSIONS TO EXPERIMENTAL VALUES. ²⁷	92
TABLE 5.10: MIXED SOLVENT SOLUBILITY ERROR FOR ACETAMINOPHEN ⁷¹ AND NAPHTHALENE ⁷² USING BOTH EXCHANGE ENERGY EXPRESSIONS FROM EXPERIMENTAL SOLUTE MOLE FRACTION.	95
TABLE 5.11: COMPARISON ERROR SUMMARY FOR NRTL-SAC AND COSMO-SAC PREDICTED SOLUBILITIES FOR 17 SOLUTES AND 258 SOLUTE/SOLVENT PAIRINGS AT 298.15 K, EXCEPT ACETAMINOPHEN AT 303.15 K.	98
TABLE 5.12: SUMMARY AND COMPARISON OF SOLUBILITY PREDICTION ERROR IN N-OCTANOL USING BOTH EXCHANGE ENERGY EXPRESSIONS FOR 52 SOLUTES AT VARIOUS TEMPERATURES. ²⁷	100
TABLE 5.13: COSMO-SAC PURE-SOLVENT PREDICTED SOLUBILITY ERROR SUMMARY FOR NITROGEN-CONTAINING AND NITROGEN FREE SOLUTES. COMPARE WITH SAMPLE SET NO. 1 IN TABLE 5.2.	104
TABLE 5.14: ERROR SUMMARY OF THE PREDICTED SOLUBILITY FOR EACH NITROGEN-CONTAINING SOLUTE CATEGORIZED BY FUNCTIONAL GROUP. EACH ERROR MEASUREMENT WEIGHTS EACH SOLUTE/SOLVENT PAIR EQUALLY, AND EACH SOLUTE EQUALLY. COMPARE WITH THE SAMPLE SET NO. 1 ERROR VALUES FROM TABLE 5.2.	105
TABLE 5.15: RECOMMENDED SOLUTES AND SOLVENTS FOR USE WITH THE MATHIAS ET AL ¹³ EXCHANGE ENERGY EXPRESSION.	106
TABLE A.1: VT-2006 SOLUTE SIGMA PROFILE DATABASE COMPOUND AND PROPERTY FULL LISTING.	121

TABLE A.2: ERROR SUMMARY OF PREDICTED COSMO-SAC SOLUBILITY FROM EXPERIMENTAL MOLE FRACTION ²⁷ FOR EACH ACETAMINOPHEN CONFORMATION IN 26 PURE SOLVENTS AT VARIOUS TEMPERATURES COMPARING OF BOTH EXCHANGE ENERGY EXPRESSIONS (PART 1: RMSE, PART 2: AA%E).....	128
TABLE A.3: ERROR SUMMARY OF PREDICTED COSMO-SAC SOLUBILITY FROM EXPERIMENTAL MOLE FRACTION ²⁷ FOR EACH ASPIRIN CONFORMATION IN 26 PURE SOLVENTS AT VARIOUS TEMPERATURES COMPARING OF BOTH EXCHANGE ENERGY EXPRESSIONS (PART 1: RMSE, PART 2: AA%E).....	130
TABLE A.4: ERROR SUMMARY OF PREDICTED COSMO-SAC SOLUBILITY FROM EXPERIMENTAL MOLE FRACTION ²⁷ FOR EACH IBUPROFEN CONFORMATION IN 26 PURE SOLVENTS AT VARIOUS TEMPERATURES COMPARING OF BOTH EXCHANGE ENERGY EXPRESSIONS (PART 1: RMSE, PART 2: AA%E).....	132
TABLE A.5: ERROR SUMMARY OF PREDICTED COSMO-SAC SOLUBILITY FROM EXPERIMENTAL MOLE FRACTION ²⁷ FOR EACH LIDOCAINE CONFORMATION IN 26 PURE SOLVENTS AT VARIOUS TEMPERATURES COMPARING OF BOTH EXCHANGE ENERGY EXPRESSIONS (PART 1: RMSE, PART 2: AA%E).....	134
TABLE A.6: ERROR SUMMARY OF PREDICTED COSMO-SAC SOLUBILITY FROM EXPERIMENTAL MOLE FRACTION ²⁷ FOR EACH CHOLESTEROL CONFORMATION IN 26 PURE SOLVENTS AT VARIOUS TEMPERATURES COMPARING OF BOTH EXCHANGE ENERGY EXPRESSIONS (PART 1: RMSE, PART 2: AA%E).....	136
TABLE A.7: ERROR SUMMARY OF PREDICTED SOLUBILITY IN BINARY SOLVENTS FOR 37 SYSTEMS WITH CONSIDERATION FOR MULTIPLE CONFORMATIONS OF ACETAMINOPHEN, ETHANOL, AND CYCLOHEXANE.	140
TABLE A.8: COSMO-SAC PREDICTED AQUEOUS PURE-SOLVENT SOLUBILITY ERROR SUMMARY.....	143
TABLE A.9: COSMO-SAC PREDICTED N-OCTANOL PURE-SOLVENT SOLUBILITY ERROR SUMMARY FOR ALL SOLUTES.	148
TABLE A.10: COMPARISON OF COSMO-SAC PREDICTED SOLUTE MOLE FRACTION IN ETHANOL AS A PURE-SOLVENT USING THE SIGMA PROFILE FROM BOTH ETHANOL CONFORMATIONS USING THE BOTH EXCHANGE ENERGY EXPRESSIONS TO LITERATURE SOLUTE MOLE FRACTIONS.	150
TABLE A.11: COMPARISON OF COSMO-SAC PREDICTED SOLUTE MOLE FRACTION IN CYCLOHEXANE AS A PURE-SOLVENT USING THE SIGMA PROFILE FROM BOTH CYCLOHEXANE CONFORMATIONS USING THE EXCHANGE ENERGY EXPRESSION DEFINED BY LIN ⁶ TO LITERATURE SOLUTE MOLE FRACTIONS.....	152
TABLE A.12: COSMO-SAC PREDICTED NITROGEN-CONTAINING SOLUTE PURE-SOLVENT SOLUBILITY ERROR SUMMARY.....	154
TABLE C.1: EXAMPLE WATER (VT-1076) SIGMA PROFILE FROM THE VT-2005 SIGMA PROFILE DATABASE.	180
TABLE C.2: ACCELRY'S MS GEOMETRY OPTIMIZATION TASK OUTPUT (OUTMOL FILE) FOR WATER, VT-1076. GEOMETRY OPTIMIZATION TASK OUTPUT IS ITERATIVE, THEREFORE WE TRUNCATE DISPLAY ONLY THE LAST ITERATION.	182
TABLE C.3: ACCELRY'S MS ENERGY CALCULATION TASK OUTPUT (OUTMOL FILE) FOR WATER, VT-1076. THE ENERGY CALCULATION TASK OUTPUT CALCULATES ATOMIC COORDINATES, TOTAL ENERGY, AND MOLECULAR CAVITY DIMENSIONS. THE ENERGY CALCULATION TASK RUNS SIMULTANEOUSLY WITH THE COSMO CALCULATION FOR A CONDENSED PHASE MOLECULE.....	189

1 Introduction

Predictive thermodynamic models are in high demand in the current engineering practice. Their need is high enough that engineers are willing to accept inaccuracies for their promised benefits such as time and cost savings. Solvent selection for new or existing processes is just one of the many potential applications for predictive models. For example, in pharmaceutical drug development, synthesis, and delivery, we use solvents as a medium for the synthesis reaction or to separate and purify the desired components from unwanted byproducts. Even after the drug is synthesized, researchers spend considerable amounts of time finding suitable solvents and scale up from a lab bench-scale to high-volume manufacturing. Reducing both time and cost with a predictive property method could significantly enhance the success of developing and manufacturing a new drug. Conceptually, one could simulate the synthesis of a new compound and select a variety of solvents for different applications spending far less time and money using predictive thermodynamic models.

1.1 Background and Significance of This Research

Solvent selection is just one potential application for predictive models. There are many instances where researchers need phase-equilibrium data and the quickest way to obtain the necessary values is with a predictive thermodynamic model. Currently, group-contribution methods like UNIFAC, and activity-coefficient models like NRTL also perform these functions; however, these methods require regressed parameters from experimental data, and therefore have little or no applicability to compounds with new functional groups (in the case of UNIFAC) or new compounds (in the case of NRTL) without substantial experimentation. Solvation thermodynamics is another approach to characterize molecular interactions and account for liquid-phase nonidealities. Solvation thermodynamic models, theoretically based on computational quantum mechanics, allow researchers to predict physical properties without experimental data. The Conductor-like Screening Model – Realistic Solvation, (COSMO-RS)¹⁻⁴ and Conductor-like Screening Model – Segment Activity Coefficient (COSMO-SAC),^{5,6} are two *a priori* models, which predict intermolecular interactions based on molecular structure and a few adjustable parameters. COSMO-RS is the first extension of a dielectric continuum – solvation model to liquid – phase thermodynamics, and COSMO-SAC is a variation of COSMO-

RS. COSMO-based models predict liquid-phase activity coefficients, which we use in this work for phase-equilibrium calculations.

Both COSMO-based models require input in the form of a sigma profile, which is a molecular-specific distribution of the surface-charge density, in a manner similar to the way UNIFAC requires parameter databases with one exception. Sigma profiles for COSMO-based methods are molecule-specific, and UNIFAC binary interaction parameters are specific to each functional group. We generate molecular-specific sigma profiles from a single structure by performing quantum-mechanical calculations. These calculations represent the most time-consuming task involved with using COSMO-based methods with over 90% of the computational effort devoted to this step. The lack of a comprehensive, open-literature sigma-profile database also hinders the ability to apply or improve the COSMO approach.⁷ The groundwork for the first open-literature database of sigma profiles and the motivation for this work come from the graduate work of Richard J. Oldland.⁸

1.2 Objectives of This Research

We present two open literature databases, the VT-2005 Sigma Profile Database^{8,9} and the VT-2006 Solute Sigma Profile Database, that contain sigma profiles for 1,645 compounds. The databases include chemicals composed of the following ten atoms only: hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine. The VT-2005 Sigma Profile Database includes many common solvents and is the larger of the two databases. The VT-2006 Solute Sigma Profile Database includes 206 solutes and 32 solvents, and focuses primarily on larger pharmacological compounds and their derivatives. Both databases are available free of charge from our website (www.design.che.vt.edu). We continue to update the sigma profiles as new results become available. Our website also includes detailed procedures for generating additional sigma profiles for any compound, along with FORTRAN programs for the sigma-averaging algorithm and the COSMO-SAC model. We refer to specific compounds from both databases throughout this work with the following nomenclature: VT-(Index No.) refers to a compound from the VT-2005 Sigma Profile Database^{8,9} with a 4-digit index number, VTSOL-(Index No.) refers to a compound from the VT-2006 Solute Sigma Profile Database with a 3-digit index number.

Through our validation effort, we create a set of application guidelines for researchers to follow when using COSMO-based models, which covers both favorable and unfavorable systems for this approach. To accomplish our objectives, we validate each database based on comparisons to several physical properties and phase-equilibrium data from literature. We demonstrate the applicability of COSMO-based methods to predict pure component vapor pressure, binary and multicomponent vapor-liquid equilibrium systems, solid solubility in pure and mixed solvents. Our work focuses on solid solubility, not gas solubility, in liquids unless specifically stated. We also discuss several factors that affect the accuracy of COSMO-based methods, such as conformational isomerism, melting point temperature and latent heat of fusion sensitivity, parameter regression, and variations in model definitions. Ultimately, we hope to implement this model into the solvent selection process by demonstrating its abilities to reduce costs and design time. We also assist practitioners by providing guidelines for its use.

2 Theory

In this chapter, we give an overview of COSMO-based thermodynamic models including a brief derivation of COSMO-SAC (Segment Activity Coefficient) and assumptions inherent to this model. We introduce the concept of a sigma profile, which is a molecule-specific distribution of the surface-charge density. To calculate this distribution, we discretize the molecule's surface into multiple charged surface segments with a specific surface-charge density. We also briefly discuss density-functional theory (DFT), which we use to perform COSMO calculations, and the governing equations for solid-liquid equilibria. Finally, we discuss the estimation of normal melting point temperatures and latent heats of fusion from group and non-group-contribution methods, which users can use to predict solubility in the absence of available literature values. This work focuses on COSMO-SAC model applications, developed by Lin and Sandler,⁶ although COSMO-RS, a precursor to COSMO-SAC developed in 1995 by Klamt et al^{1-4,10} is also available for license with several software packages. (<http://www.cosmologic.de/>).

2.1 Overview of COSMO-SAC Model

The basic principle behind COSMO-based thermodynamic models is the “solvent-accessible surface” of a solute molecule.^{2,10} Conceptually, COSMO-based models create a cavity with the exact size of a molecule within a homogeneous medium, or solvent, of a dielectric constant ϵ then place the molecule inside the cavity. We illustrate the ideal solvation process with COSMO-based methods in Figure 2.1.

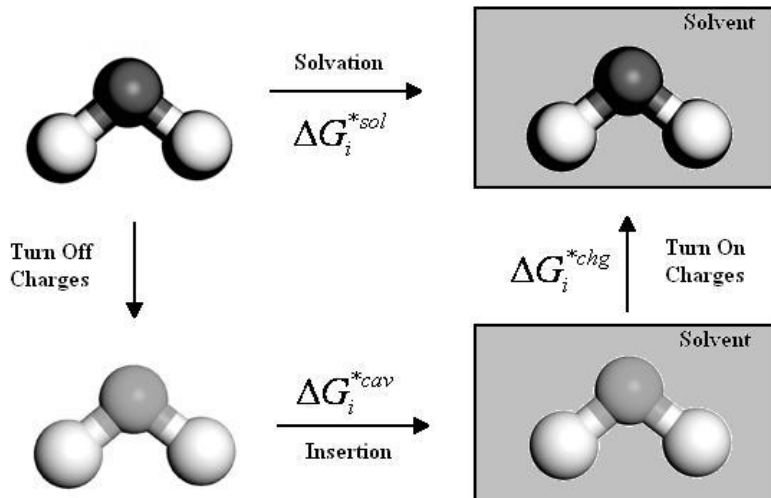


Figure 2.1: Conceptual diagram of ideal solvation process with a COSMO-based model.

In Figure 2.1, the solvation free energy ΔG_i^{*sol} represents the change in Gibbs free energy associated with moving a molecule i from a fixed position in an ideal gas to a fixed position in a solution S . The cavity-formation free energy ΔG_i^{*cav} represents the change in Gibbs free energy required to form a cavity within a solution S of the exact size of the molecule i . The charging free energy ΔG_i^{*chg} represents the Gibbs free energy required to remove the screening charges from the surface of the molecular cavity. We write an expression for the solvation free energy as the sum of the cavity-formation free energy and the charging free energy in equation (2.1).

$$\Delta G_i^{*sol} = \Delta G_i^{*cav} + \Delta G_i^{*chg} \quad (2.1)$$

2.1.1 Derivation of Segment Activity-Coefficient Model

In this section, we summarize the derivation of the COSMO-SAC model presented by Lin and Sandler^{5,6}. Lin and Sandler initially define the activity coefficient in terms of the difference in the charging Gibbs free energy as a pure species and as a pure species within solution; however Lin and Sandler^{11,12} later find that using the Staverman-Guggenheim combinatorial term improves the calculation of the cavity-formation free energy.

$$\ln \gamma_{i/S} = \frac{\Delta G_{i/S}^{*chg} - \Delta G_{i/i}^{*chg}}{RT} + \ln \gamma_{i/S}^{SG} \quad (2.2)$$

We summarize the definition for the Staverman-Guggenheim combinatorial term with the following equation. It accounts for size and shape effects of the solute and the solvent.

$$\ln \gamma_{i/S}^{SG} = \ln \frac{\phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\theta_i}{\phi_i} + l_i - \frac{\phi_i}{x_i} \sum_j x_j l_j \quad (2.3)$$

Here $\phi_i = x_i q_i / \sum_j x_j q_j$, $\theta_i = x_i r_i / \sum_j x_j r_j$, and $l_i = (z/2)(r_i - q_i) - (r_i - 1)$, where x_i is the mole fraction of component i , z is the coordinate number generally valued at 10. The normalized volume and surface area, r_i and q_i , respectively, are $r_i = V_i / r$ and $q_i = A_i / q$, where A_i and V_i are the cavity surface area and cavity volume, respectively, and r and q are COSMO-SAC model parameters. We assign values for these parameters and others in Table 3.3.

Because Klamt¹ treats the homogeneous medium as a perfect conductor, Lin and Sandler⁶ define the charging free energy as the sum of two terms, the ideal solvation energy, ΔG^{IS} , and the restoring free energy, ΔG^{*res} , which represents the free energy necessary to remove the screening charges from the cavity surface after placing the solute in the cavity. Since the ideal solvation energy for a solute in a solvent S or a pure liquid i are equivalent, the equation for the activity coefficient simplifies to the following equation.^{5,6}

$$\ln \gamma_{i/S} = \frac{\Delta G_{i/S}^{*res} - \Delta G_{i/i}^{*res}}{RT} + \ln \gamma_{i/S}^{SG} \quad (2.4)$$

Lin and Sandler⁶ define the restoring free energy as the sum of the products of the sigma profile and the natural log of the segment activity coefficients over all surface segments

$$\frac{\Delta G_{i/S}^{*res}}{RT} = \sum_{\sigma_m} \left[n_i(\sigma_m) \frac{\Delta G_{\sigma_m/S}^{*res}}{RT} \right] = n_i \sum_{\sigma_m} p_i(\sigma_m) \ln \Gamma_s(\sigma_m) \quad (2.5)$$

where $\Gamma_s(\sigma_m)$ is the activity coefficient for a segment m of charge density, σ . We calculate the segment activity coefficient for the segment in a solution $\Gamma_s(\sigma_m)$ and in a pure liquid $\Gamma_i(\sigma_m)$ as derived rigorously using statistical mechanics.⁶

$$\begin{aligned} \ln \Gamma_s(\sigma_m) &= -\ln \left\{ \sum_{\sigma_n} p_s(\sigma_n) \Gamma_s(\sigma_n) \exp \left[\frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\} \\ \ln \Gamma_i(\sigma_m) &= -\ln \left\{ \sum_{\sigma_n} p_i(\sigma_n) \Gamma_i(\sigma_n) \exp \left[\frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\} \end{aligned} \quad (2.6)$$

The exchange energy $\Delta W(\sigma_m, \sigma_n)$ is the energy required to obtain one pair of segments from a neutral pair of segments. It contains contributions from electrostatic interactions, or misfit

energy E_{mf} , hydrogen-bonding interactions E_{hb} , and nonelectrostatic interactions E_{ne} of the segment pairs.^{4,5} The first term of equation (2.7) describes the electrostatic contribution and the second term describes the hydrogen-bonding contribution to the exchange energy. The nonelectrostatic energy contribution is assumed constant and therefore drops out of the final equation (2.7).

$$\Delta W(\sigma_m, \sigma_n) = \left(\frac{\alpha'}{2} \right) (\sigma_m + \sigma_n)^2 + c_{hb} \max[0, \sigma_{acc} - \sigma_{hb}] \min[0, \sigma_{don} + \sigma_{hb}] \quad (2.7)$$

$$[\equiv] \text{ kcal} \cdot \text{mol}^{-1}$$

The electrostatic contribution or misfit energy is a correction for induced surface charges which screen any polarization effects by placing the molecule in an ideal conductor rather than next to another molecular cavity. Klamt and co-workers^{2,4} fit the misfit energy constant α' , the hydrogen bonding constant c_{hb} , and the hydrogen bonding sigma cutoff value σ_{hb} to experimental data. The misfit energy constant α' is the product of the polarizability factor f_{pol} and constant α . Generally, the polarizability factor is a function of the dielectric constant of the medium, but Klamt suggests setting this factor to a constant 0.64.

$$f_{pol} = \frac{\epsilon - 1}{\epsilon + 1/2} = 0.64 \quad (2.8)$$

$$\alpha = \frac{0.3a_{eff}^{3/2}}{\epsilon_0} [\equiv] \text{ kcal} \cdot \text{\AA}^4 \cdot \text{mol}^{-1} \cdot \text{e}^{-2} \quad (2.9)$$

$$\alpha' = f_{pol}\alpha [\equiv] \text{ kcal} \cdot \text{\AA}^4 \cdot \text{mol}^{-1} \cdot \text{e}^{-2} \quad (2.10)$$

Mathias et al¹³ suggest another model for the exchange energy, which defines a new contribution term to account for hydrogen bonding, ΔW_{hb} . We show the equations for this new definition of the exchange energy below.

$$\Delta W(\sigma_m, \sigma_n) = \frac{\alpha'}{2} (\sigma_m + \sigma_n)^2 + \Delta W_{hb}(\sigma_m, \sigma_n) [\equiv] \text{ kcal} \cdot \text{mol}^{-1} \quad (2.11)$$

$$\Delta W_{hb} = -c_{hb} \left\{ \max[0, |\sigma_m - \sigma_n| - \sigma_{hb}^n] \right\}^2 [\equiv] \text{ kcal} \cdot \text{mol}^{-1} \quad (2.12)$$

For our work concerning solubility, we use both expressions to calculate and compare the effect of each exchange energy definition to gain knowledge concerning acceptable cases for both.

Finally, we combine equations (2.4) through (2.12) to calculate the activity coefficient using the following expression.

$$\ln \gamma_{i/s} = n_i \sum_{\sigma_m} p_i(\sigma_m) [\ln \Gamma_s(\sigma_m) - \ln \Gamma_i(\sigma_m)] + \ln \gamma_{i/s}^{SG} \quad (2.13)$$

Because of the slightly different definition of our sigma profiles for pure components, see Section 2.1.2, we modify equation (2.6) derived by Lin and Sandler⁶ by substituting the area-weighted sigma profile $p'_i(\sigma)$ in place of the standard sigma profile $p_i(\sigma)$, which is analogous to the equation used by Klamt.⁴ Each term in equation (2.14) has the same definition as in equation (2.6). We also redefine equation (2.13) to incorporate the change in definition of the sigma profile.

$$\ln \Gamma_i(\sigma_m) = -\ln \left\{ \sum_i \frac{p'_i(\sigma_n)}{A_i} \Gamma_i(\sigma_n) \exp \left[\frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\} \quad (2.14)$$

$$\ln \gamma_{i/s} = \frac{1}{a_{eff}} \sum_{\sigma_m} p'_i(\sigma_m) [\ln(\Gamma_s(\sigma_m)) - \ln(\Gamma_i(\sigma_m))] + \ln \gamma_{i/s}^{SG} \quad (2.15)$$

We refer to equation (2.15) as the COSMO-SAC model from this point forward.

2.1.2 Sigma Profiles

A sigma profile is a probability distribution of the surface-charge density of a molecule or a mixture. COSMO-based models construct the molecular shaped cavity within the perfect conductor¹ according to a specific set of rules and atom-specific dimensions. Then, the molecule's dipole and higher moments draw charges from the surrounding medium to the surface of the cavity to screen, or cancel, the electric field both inside the conductor and tangential to the surface, allowing the molecule to move freely within the system without altering the system's overall energy. We calculate the induced charges on the solute surface in discretized space from Poisson's equation and the zero total potential boundary condition.

$$\Phi_{tot} = \Phi_i + \Phi(q^*) = \Phi_i + Aq^* = 0 \quad (2.16)$$

In equation (2.16), Φ_{tot} is the total potential on the cavity surface, Φ_i is the potential due to the charge distribution of the solute molecule i , $\Phi(q^*)$ is the potential as a function of the ideal screening charge q^* . We set $\Phi(q^*)$ equal to the product of the ideal screening charge q^* and Coulomb Interaction Matrix A , which describes potential interactions between surface-charges and is a function of the cavity geometry.⁴ The surface-charge distribution in a finite

dielectric solvent is well-approximated by a simple scaling of the surface-charge density in a conductor σ^* .

We average these segment surface-charge densities σ^* , from COSMO calculation output, and obtain a new surface-charge density, $\sigma = q_{avg} / a_{eff}$, where q_{avg} is the average screening charge for a given segment. The effective area of a standard surface segment a_{eff} represents the contact area between different segments, e.g., a theoretical bonding site. Klamt sets this adjustable parameter, a_{eff} , to 7.1 \AA^2 . Klamt¹ then defines the sigma profile $p_i(\sigma)$ for a molecule i as the probability of finding a segment with a surface-charge density with the following equations.

$$p_i(\sigma) = n_i(\sigma) / n_i = A_i(\sigma) / A_i \quad (2.17)$$

$$n_i = \sum_{\sigma} n_i(\sigma) = A_i / a_{eff} \quad (2.18)$$

$$A_i = \sum_{\sigma} A_i(\sigma) [\equiv] \text{ \AA}^2 \quad (2.19)$$

Here, $n_i(\sigma)$ is the number of segments with a surface-charge density σ , n_i is the total number of surface segments around the molecular cavity, A_i is the surface area of the molecular cavity, and $A_i(\sigma)$ is the total surface area of all of the segments with a particular charge density σ . $A_i(\sigma)$ and $n_i(\sigma)$ are proportional by a_{eff} , $A_i(\sigma) = a_{eff} n_i(\sigma)$ as defined by Lin and Sandler.⁶ We use area-weighted sigma profiles $p'_i(\sigma)$ in both databases. We show the equation below.

$$p'_i(\sigma) = p_i(\sigma) A_i = A_i(\sigma) [\equiv] \text{ \AA}^2 \quad (2.20)$$

The sigma profile of a mixture is also a weighted average of the pure component sigma profiles. In principle, the mixture sigma profile $p_s(\sigma)$ is not limited to a specific number of components.

$$p_s(\sigma) = \frac{\sum_i x_i n_i p_i(\sigma)}{\sum_i x_i n_i} = \frac{\sum_i x_i A_i p_i(\sigma)}{\sum_i x_i A_i} = \frac{\sum_i x_i p'_i(\sigma)}{\sum_i x_i A_i} \quad (2.21)$$

Lin and Sandler⁶ also define the averaging algorithm for the segment surface-charge densities σ^* to calculate the new surface-charge densities σ .

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{eff}^2}{r_n^2 + r_{eff}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{eff}^2}\right)}{\sum_n \frac{r_n^2 r_{eff}^2}{r_n^2 + r_{eff}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{eff}^2}\right)} [\equiv] \text{e}/\text{\AA}^2 \quad (2.22)$$

In equation (2.22), σ_m is the average surface-charge density on segment m , the summation is over n segments from the COSMO output, and r_n is the radius of the actual surface segment, which have an assumed circular geometry. The effective radius, $r_{eff} = \sqrt{a_{eff}/\pi}$, is an adjustable parameter in this model, and d_{mn} is the distance between the two segments m and n .^{2,6} The paired segments m and n have segment charge densities σ_m and σ_n , respectively. We use an averaging radius³, $r_{av} = 0.81764 \text{ \AA}$, for the sigma-averaging algorithm in place of the effective segment radius r_{eff} . This corresponds to the average segment surface area of $a_{av} = \pi r_{av}^2 = 2.100265 \text{ \AA}^2$. We use the averaging algorithm in equation (2.23) to calculate the average surface-charge density σ_m .

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{av}^2}\right)}{\sum_n \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp\left(-\frac{d_{mn}^2}{r_n^2 + r_{av}^2}\right)} [\equiv] \text{e}/\text{\AA}^2 \quad (2.23)$$

Our averaging algorithm is identical to those determined by Lin and Sandler⁶ and Klamt and co-workers,^{1,2} except that we use a different value for r_{av} . Klamt and co-workers report using averaging radii ranging between 0.5 and 1.0 \AA stating that “the best value for the averaging radius r_{av} turns out to be 0.5 \AA . This is less than the initially assumed value of about 1 \AA .”² The deliverables for our two databases include calculation results from the density-functional theory (DFT) calculations, thus enabling future work in optimization of r_{av} . We use equation (2.23) when generating all sigma profiles for the VT-2005 Sigma Profile Database. Each sigma profile contains 50 segments, ranging from $-0.025 \text{ e}/\text{\AA}^2$ to $0.025 \text{ e}/\text{\AA}^2$ with a step size of $0.001 \text{ e}/\text{\AA}^2$. Given the sigma profiles and the COSMO-based models, we compute various physical properties, such as partition coefficients, infinite-dilution activity coefficients, and phase-

equilibrium behavior, etc.^{1,14,15} Figure 2.2 shows examples of area-weighted sigma profiles from the VT-2005 Sigma Profile Database.

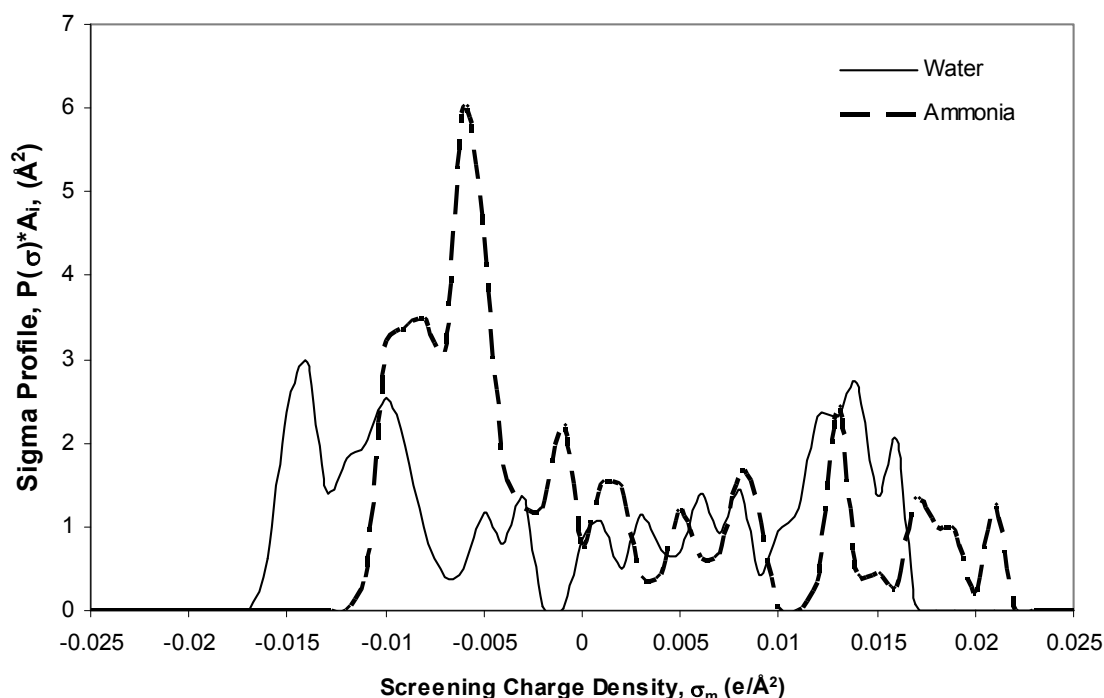


Figure 2.2: Water (VT-1076) and Ammonia (VT-1070) sigma profiles available in the VT-2005 Sigma Profile Database. Both sigma profiles have peaks above the hydrogen bonding cutoff signifying these compounds likely form hydrogen bonds.

Wang et al.¹⁶ discuss an alternative method for generating sigma profiles by essentially dividing the sigma profile into two parts, a hydrogen-bonding sigma profile and a non-bonding sigma profile. They show improvement in some cases where hydrogen bonding is prevalent; however, all sigma profiles discussed in this work have not undergone such treatment.

2.2 Overview of Density-Functional Theory

We calculate each molecule-specific sigma profile from the output of a series of calculations using the DMol3 module, which incorporates both density-functional theory (DFT),¹⁷⁻²¹ available in Accelrys' Materials Studio v4.0 (MS), and the COSMO-based models.^{1,2,4-6,10} Many researchers contribute to the derivation of density-functional theory (DFT). DFT approximates ground-state energies by calculating the Hamiltonian operator as

determined by the electron density. Researchers arrive at this conclusion after applying several simplifying assumptions to the Schrödinger equation to avoid calculating a solution in its full form. The full solution to the Schrödinger equation is very computationally intensive and oftentimes not practical, thus the need for an approximation. Some of the approximations include the local-density approximation (LDA) and the generalized gradient approximation (GGA). We use the generalized gradient approximation for our DFT calculations in MS, which we detail in Section 3.1.1. For further information on DFT and its supporting principles, assumptions, and functionals see Thijssen,²² Levine,²³ and Harrison.²⁴

2.3 Solubility Theory

In this section, we discuss the background and underlying assumptions concerning solid-liquid mixtures. We also summarize the derivation of ideal solubility of a solid in solution.

2.3.1 Overview of Solubility

Before we discuss the solubility theory, we must first define the term and its nomenclature. Ben-Naim²⁵ defines "... the solvation process of a molecule s in a fluid l as the process of transferring the molecule s from a fixed position in an ideal gas phase g into a fixed position in the fluid or liquid phase l . The process is carried out at a constant temperature T and pressure P . Also, the composition of the system is unchanged." The molecule s serves as our solute and the fluid l as our solvent in the conventional sense. From this definition, we also take the idea that solubility should focus on interactions between the solvated molecule and its surroundings. COSMO-based thermodynamic methods incorporate these molecular interactions by calculating "the chemical potential of any species in any mixture,"⁶ and therefore are potential tools for modeling solubility for a wide range of solutes.

2.3.2 Derivation of Solubility Equation from Solid-Liquid Equilibrium

Now we outline the derivation of the equation describing ideal solubility based on Prausnitz et al.²⁶ Solubility is largely dependant on the interactions of the solute molecules with its surrounding environment. For a solute molecule to dissolve in a liquid there must be a favorable change in Gibbs energy. When the solute and solvent equilibrate, the fugacity of the

pure solute equals the fugacity of the solute in solution. The equation describing this equilibrium is as follows.

$$f_{(\text{pure solute})} = f_{(\text{solute in liquid solution})} \quad (2.24)$$

The fugacity of the pure solute is the product of its composition x_{sol} , liquid-phase activity coefficient γ_{sol} and standard state fugacity f^0 . The following equation shows this relationship where γ_{sol} and x_{sol} are the activity coefficient and mole fraction of the solute.

$$f^S = \gamma_{sol} x_{sol} f^0 \quad (2.25)$$

Here the reference state is the compound as a pure subcooled liquid at the same temperature as the pure solid f^0 . We calculate this fugacity ratio by breaking the transformation of a solute molecule to a liquid molecule down into a three-step thermodynamic cycle. Figure 2.3 illustrates the thermodynamic steps taken in the solvation process.

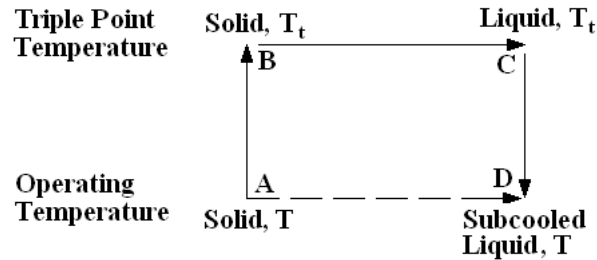


Figure 2.3: Cartoon representation of three-step thermodynamic cycle to calculate fugacity ratio of solid phase fugacity f^S to liquid-phase fugacity f^L

The initial state, A, is the solute as a solid at the desired operating temperature. The first step requires the solute be raised to a solid at the triple point temperature, or the temperature and pressure at which all three phases, solid, liquid, and gas, co-exist at equilibrium, which is state B. The second step is then the transformation of the solute from a solid at the triple point temperature to a liquid at the triple point temperature, state C. The final step cools the solute to a subcooled liquid, state D. The change in Gibbs free energy from state A to D depends on the ratio of fugacity of the solute as a pure subcooled liquid f^L and the fugacity of the solute as a pure solid f^S . We show the enthalpic and entropic contributions to the Gibbs free energy below.

$$\Delta G = RT \ln \frac{f^L}{f^S} \quad [\equiv] \text{ kJ} \cdot \text{mol}^{-1} \quad (2.26)$$

$$\Delta H = \Delta H_{fus}^{T_t} + \int_{T_t}^T \Delta C_p dT \quad [\equiv] \text{ kJ} \cdot \text{mol}^{-1} \quad (2.27)$$

$$\Delta S = \Delta S_{fus}^{T_t} + \int_{T_t}^T \frac{\Delta C_p}{T} dT \quad [\equiv] \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} \quad (2.28)$$

Combining equations (2.26) – (2.28) gives us the following relationship for the fugacity ratio assuming that the heat capacity remains constant over the specified temperature range.

$$\ln \frac{f^L}{f^S} = \frac{\Delta H_{fus}^{T_t}}{RT_t} \left(\frac{T_t}{T} - 1 \right) - \frac{\Delta C_p}{R} \left(\frac{T_t}{T} - 1 \right) + \frac{\Delta C_p}{R} \ln \frac{T_t}{T} \quad (2.29)$$

By assuming that the effect of heat capacity is negligible in comparison to the enthalpic energy contribution, substitution of equation (2.25) into equation (2.29) gives:

$$\ln x_{sol} = \frac{\Delta H_{fus}^{T_m}}{RT_m} \left(1 - \frac{T_m}{T} \right) - \ln \gamma_{sol} \quad (2.30)$$

We also substitute the normal melting temperature for the triple point temperature which does increase error, but only slightly because the difference in enthalpy at the normal melting temperature and the triple point temperature is generally small. We refer to equation (2.30) as the solubility equation, which remains valid assuming we satisfy the condition that the operating temperature T is less than the normal melting point temperature T_m . We rearrange equation (2.30) and set the constant on the right-hand side of the equation equal to $\ln K_{sp}$, or the solubility product constant. This is another common representation of the solubility equation. This form of the solubility equation is useful, especially in industrial applications, for regression of physical constants, such as the latent heat of fusion and the melting temperature, from experimental solubility and an activity coefficient model. Using experimental solubility data and an activity coefficient model to estimate K_{sp} is an alternative to using differential scanning calorimetry to

determine physical properties. We show the relationship of K_{sp} to the solute composition and the solute activity coefficient in equation (2.31).

$$\ln x_{sol} + \ln \gamma_{sol} = \ln K_{sp} = \frac{\Delta H_{fus}^{T_m}}{RT_m} \left(1 - \frac{T_m}{T} \right) \quad (2.31)$$

For all solubility predictions in our comparison of COSMO-SAC and NRTL-SAC in Section 5.5, we use a constant value of K_{sp} from equation (2.31) from literature.²⁷

2.4 Estimation Methods for Melting Temperature and Heats of Fusion

In order to predict solubility for a new compound, we need estimates for both the normal melting point temperature and the latent heat of fusion. However, these methods are still inadequate for the pharmaceutical industry. Due to multiple crystalline structures of a given solute, multiple melting point temperatures or latent heats of fusion may exist with similar values. Generally in the pharmaceutical industry, researchers determine heats of fusion from solubility experiments at multiple temperatures, which depending on the breadth of the experiment can be very costly. For simple solutes, these estimation methods may still be satisfactory for quick solubility calculations. We briefly summarize several methods for estimating both of these physical properties, including both group-contribution and non-group-contribution methods, and their relative accuracies. Later, in Section 5.6, we address the sensitivity of COSMO-SAC solubility predictions to deviations in normal melting temperature and the latent heat of fusion. We also recommend a resource for tabulated group-contribution parameters by Poling et al.²⁸

2.4.1 Estimation of Normal Melting-Point Temperature

Here, we summarize three methods for estimating normal melting-point temperatures, two group-contribution methods and one method that combines group-contribution terms and non-group-contribution terms. Joback^{29,30} presented the first method for estimating T_m . We write the Joback group-contribution method in equation (2.32). Poling et al.²⁸ tabulate the group-contribution parameters, $N_k(tfpk)$.

$$T_m = 122 + \sum_k N_k(tfpk) [\equiv] \text{K} \quad (2.32)$$

Constantinou and Gani^{31,32} improve the Joback method by incorporating second-order group-contribution parameters, which are also tabulated in Poling et al.²⁸

$$T_m = 102.425 \ln \left[\sum_k N_k(tfp1k) + W \sum_j M_j(tfp2j) \right] [\equiv] \text{K} \quad (2.33)$$

In equation (2.33), $N_k(tfpk)$ and $M_j(tfp2j)$ are first- and second-order group-contribution parameters, respectively, and we set W to zero for compounds with only first order contributions and set W to one when including both first and second order contributions.

Finally, we summarize the method presented by Yalkowsky,³³ Krzyzaniak et al,³⁴ and Zhao et al,³⁵ which incorporates both group-contribution terms and molecular descriptor contributions. Zhao et al tabulate the molecular descriptors and group-contribution parameters.³⁵

$$T_m = \frac{\Delta H_{fus}^{T_m}}{\Delta S_{fus}^{T_m}} = \frac{\sum_i n_i^t m_i}{56.5 - 19.2 \log_{10} \sigma + 9.2\tau} [\equiv] \text{K} \quad (2.34)$$

In equation (2.34), n_i and m_i are group-contribution parameters tabulated by Zhao et al.³⁵ Dannenfelser et al³⁶ define the molecular rotational symmetry number σ as the number of conformations identical to the reference conformation when rotating a rigid molecule 360° on two axes, φ and z , in spherical coordinates. Dannenfelser et al³⁶ calculate the number of torsional angles τ using equation (2.35).

$$\tau = \text{SP3} + 0.5\text{SP2} + 0.5\text{RING} - 1 \quad (2.35)$$

Excluding end-groups and rings, SP3 represents the number of sp3 hybridized atoms, and SP2 represents the number of sp2 hybridized atoms. RING represents the number of monocyclic fused ring systems. Yalkowsky³³ and Dannenfelser et al³⁶ provides more detailed definitions and discussion of the equation variables.

2.4.2 Estimation of Latent Heat of Fusion

We briefly summarize the methods presented by Dannenfelser and Yalkowsky,^{33,36} Joback,^{29,30} and Chickos^{37,38} to estimate entropies and enthalpies of fusion. The Dannenfelser – Yalkowsky method is a non-group-contribution method which estimates the entropy of fusion at

the normal melting temperature. The total entropy of a solid – liquid phase change includes contributions due to positional ΔS_m^{pos} , rotational ΔS_m^{rot} , and conformational changes ΔS_m^{conf} . Dannenfelser and Yalkowsky define the total entropy in equation (2.36).

$$\Delta S_m^{tot} = \Delta S_m^{pos} + \Delta S_m^{rot} + \Delta S_m^{conf} \quad (2.36)$$

Dannenfelser and Yalkowsky arrive at the equation (2.37) for the total entropy change.

$$\Delta S_m^{tot} = 50 - R \ln \sigma + R \ln \phi \quad [\equiv] \text{ J/K} \cdot \text{ mol} \quad (2.37)$$

In equation (2.37), the molecular symmetry number σ has the same definition as defined in Section 2.4.1, and the molecular flexibility number ϕ is a function of number of effective torsional angles τ , which we define in Section 2.4.1. Equation (2.38) represents the relationship between molecular flexibility number and the number of torsional angles.

$$\phi = 2.85^\tau \quad (2.38)$$

The Joback method for estimating enthalpy of fusion is a group-contribution method. Equation (2.39) defines the enthalpy of fusion with the necessary group-contribution parameters tabulated by Poling et al.²⁸

$$\Delta H_{fus}^{T_m} = -0.88 + \sum_k N_k (hmk) \times 0.004184 \quad [\equiv] \text{ J/mol} \quad (2.39)$$

Finally, Chickos³⁷ developed a group-contribution method for estimating heats of fusion, which they divide into four categories: (1) acyclic and aromatic hydrocarbons, (2) acyclic and aromatic hydrocarbon derivatives, (3) non aromatic cyclic and polycyclic hydrocarbons, and (4) cyclic and polycyclic hydrocarbon derivatives. Chickos et al³⁷ tabulate all necessary equations and parameters for each of the four groups listed above. Table 2.1 summarizes the error associated with the three estimation methods for normal melting point temperatures and the three estimation methods for enthalpies of fusion at the normal melting point.

Table 2.1: Summary of error for melting-point temperature and enthalpy of fusion estimation methods

Normal Melting Point Temperature Estimation Methods			
	Joback ^a	Constantiou / Gani ^a	Yalkowsky / Krzyzaniak / Zhao ^a
Abs Avg Error	28.8 K	25.8 K	26.70 K
Abs % Error	14.4%	13.2%	15.1%
Test Set Size	307	273	146
Error < 5%	80	80	35
Error > 10%	154	116	80

Enthalpy of Fusion Estimation Methods			
	Joback ^b	Dannenfesler / Yalkowsky ^a	Chickos / Acree / Liebman ^c
Abs Avg Error	2030 J/mol	1700 J/mol	3520 J/mol
% Error	38.70%	18%	17%
Test Set Size	155	43	1858

^a Values from Poling et al²⁸

^b Values from Joback²⁹

^c Values from Chickos et al³⁷

2.5 Overview of NRTL-SAC Theory

The Non-Random Two-Liquid – Segment Activity Coefficient (NRTL-SAC) model developed by Chen and Song³⁹ is a variation of the original NRTL model which incorporates the segment-based method similar to the polymer NRTL model.⁴⁰ We compare solubility predictions in pure solvents for COSMO-SAC to NRTL-SAC for a comparison to a regressed parameter model, and here we review the key concepts and equations. Chen and Song³⁹ develop NRTL-SAC from the polymer NRTL model with the specific purpose of predicting solubility of organic non-electrolyte solvents, and have since expanded this model to include electrolyte solubility modeling⁴¹ and mixed-solvent solubility modeling.⁴² They write the activity coefficient as the sum of combinatorial and residual terms.

$$\ln \gamma_I = \ln \gamma_I^C + \ln \gamma_I^R \quad (2.40)$$

The authors use the Flory-Huggins term as the combinatorial term $\ln \gamma_I^C$ and the segment-based activity coefficient $\ln \gamma_I^R$ is the residual term.

$$\ln \gamma_I^R = \ln \gamma_I^{lc} = \sum_m r_{m,I} \left[\ln \Gamma_m^{lc} - \ln \Gamma_m^{lc,I} \right] \quad (2.41)$$

This model uses a conceptual segment contribution where each molecule is potentially divided into as many as four conceptual segments, hydrophobicity, denoted by X , hydrophilicity Z , and two polar segments (positive Y^+ and negative Y^-). Chen and Song³⁹ regress the values for X , Y^- , Y^+ , and Z from experimental solubility data. These conceptual segments are in equation (2.43) and represented by $r_{j,I}$ where j is the conceptual segment of component I . If a compound does not contain a certain segment, then we exclude its segment parameter from the summation. We show the essential model equations below.

$$\ln \Gamma_m^{lc} = \frac{\sum_j x_j G_{jm} \tau_{jm}}{\sum_k x_k G_{km}} + \sum_{m'} \frac{x_{m'} G_{mm'}}{\sum_k x_k G_{km'}} \left(\tau_{mm'} - \frac{\sum_j x_j G_{jm'} \tau_{jm'}}{\sum_k x_k G_{km'}} \right) \quad (2.42)$$

$$\ln \Gamma_m^{lc,I} = \frac{\sum_j x_{j,I} G_{jm} \tau_{jm}}{\sum_k x_{k,I} G_{km}} + \sum_{m'} \frac{x_{m',I} G_{mm'}}{\sum_k x_{k,I} G_{km'}} \left(\tau_{mm'} - \frac{\sum_j x_{j,I} G_{jm'} \tau_{jm'}}{\sum_k x_{k,I} G_{km'}} \right)$$

$$x_j = \frac{\sum_I x_{j,I} r_{j,I}}{\sum_I \sum_i x_{i,I} r_{i,I}} \quad x_{j,I} = \frac{r_{j,I}}{\sum_i r_{i,I}} \quad (2.43)$$

In equations (2.42) and (2.43), $\ln \Gamma_m^{lc,I}$ and $\ln \Gamma_m^{lc}$ are the segment activity coefficients of segment m in component I only and segment m , respectively. G and τ are local binary parameters regressed from VLE and LLE data for specific interactions between conceptual segments, setting the parameter α constant. Chen and Song³⁹ summarize values for G , τ , and α . The terms, x_j and x_I , are the segment-based mole fraction of the segment j and mole fraction of component I , respectively. Chen and Song define the segment-based parameters $r_{j,I}$ as the number of segment species j in component I .

3 Computational Methods and Values

We use the following procedure to calculate the sigma profile for each compound in the VT-2005 Sigma Profile Database and the VT-2006 Solute Sigma Profile Database. This procedure includes three essential steps and one optional step. The optional step is using a pre-optimization tool such as Amber8^{43,44} or Accelrys MS Forcite Plus.⁴⁵ These tools provide an initial guess for the optimum low energy geometry of a molecule, and we discuss their use in Section 3.2. For the first essential step, we calculate the optimum low energy geometry of an individual molecule in the ideal gas phase based on the Hamiltonian energy using density-functional theory (DFT). Second, we calculate the charge and position of each segment on the surface of the geometrically optimized molecule in the condensed phase using both DFT and COSMO calculations; we assume that the low energy optimal geometry does not change from the ideal gas phase to the condensed phase. We believe that this is a weak assumption for some molecules, and we present supporting evidence in later in this chapter and in Sections 4.4, 4.5, and 5.3. However, presently, we do not have a better alternative to select an appropriate condensed phase conformation without exhaustive computational effort. Finally, we average the charged surface segments using equation (2.23) in a program originally coded by Oldland⁸ to generate the sigma profile, or screening charge density distribution over the entire surface of the molecule.

3.1 Procedure for Generating Sigma Profiles

The first step of this procedure involves drawing or importing the molecular structure from another source into MS. In the VT-2006 Solute Sigma Profile Database, we import pre-optimized structures from Amber8.^{43,44} In the VT-2005 Sigma Profile Database, we manually draw structures for the majority of the compounds and import pre-optimized structures from Amber8 and MS Forcite Plus for molecules with pure component vapor-pressure predictions originally outside the acceptable release criterion, $10.0 \leq \ln P_{vap} \leq 13.0$, \ln Pa units. We download most structures from the National Library of Medicine, (<http://chem.sis.nlm.nih.gov/chemidplus>), and use them as the initial inputs for the pre-optimization tools, Amber8 and MS Forcite Plus. We also use the NIST online database (<http://webbook.nist.gov/chemistry/>) and the American Chemical Society's Scifinder™

(<http://www.cas.org/SCIFINDER/SCHOLAR/index.html>) as alternative sources and checks for correct molecular structures.

The final “optimized” structure is a function of initial structure, for example, a poorly drawn molecule may result in a structure in a local energy minimum instead of a global energy minimum. We draw each molecule, generally starting with the carbon backbone and major functional groups. Next, we draw the hydrogen atoms to complete the structure using an “Add Hydrogen” tool in MS. This tool calculates and adds the appropriate number of hydrogen atoms to balance the remaining unbonded valence electrons. Finally, we use the “Clean” tool in MS to create a rough estimate of an optimal molecular conformation. The “Clean” tool rearranges the atoms and corrects for improper bond angles and bond lengths with fairly good accuracy. Because the “Clean” tool is only a rough estimate and possibly creates local low energy conformations, we draw each molecule in several random conformations following these steps, and then relax the geometry to determine if the conformation is indeed a low energy conformation.

The issue of conformational isomerism arises when a molecule’s structure is flexible and could exist in multiple stable forms. From our experience, we find that diols and other hydroxyl-containing compounds are likely to find a local low energy conformation if special care is not taken. Other molecules with multiple functional groups or long carbon chains are also susceptible to conformational isomerism. We discuss conformational isomerism and its effects on COSMO-SAC activity coefficient predictions in Sections 4.4, 4.5, and 5.3.

3.1.1 Ideal Gas-Phase Geometry Optimization

Now that we have an initial structure, we optimize the molecular geometry in the ideal gas phase using a “Geometry Optimization” task in the DMol3 module in MS with the following calculation settings. We use the DNP v4.0.0 basis set, which Accelrys’ MS documentation recommends for COSMO applications.⁴⁵ DNP refers to Double Numerical basis with Polarization functions, i.e., functions with angular momentum one higher than that of the highest occupied orbital in free atom. According to Koch and Holthausen,¹⁷ the DNP basis set is generally very reliable. We use the GGA/VWN-BP functional setting with a real space cutoff of 5.5 Å for all DMol3 calculations. Here, GGA stands for the generalized gradient approximation, and VWN-BP represents the Becke-Perdew version of the Volsko-Wilk-Nusair functional.^{17,46-48}

We optimize the geometry under “fine” tolerances, with 1.0E-6 Hartree energy unit or E_h , for convergence of the self-consistent field (SCF) equations, and 0.002 $E_h/\text{Å}$ for the convergence of the geometry-optimization calculations.

3.1.2 Condensed Phase Energy Calculation

For the second step, we calculate the surface screening charges surrounding the molecule in the condensed phase by performing an “Energy Calculation” task in the DMol3 module with the addition of several keywords, which calls an integrated COSMO calculation scheme in MS. These keywords specify settings for the COSMO solvation calculation and include optimized atomic radii for ten elements. We add the keywords to the .INPUT file in MS prior to running the “Energy Calculation” task. Table 3.1 summarizes the keywords and settings we use for the COSMO calculation as part of the “Energy Calculation” task.

Table 3.1: COSMO Keywords Used in Calculating Surface Segment Charges in DMol3^a

Keyword name	Default Value	Description
Cosmo	on	Turns on COSMO solvation procedure
Cosmo_Grid_Size	1082	Tells DMol3 how many basic grid points per atom to consider
Cosmo_Segments	92	Specifies the maximum number of segments on each atomic surface
Cosmo_Solvent_Radius	1.300000	Solvent probe radius
Cosmo_A-Matrix_Cutoff	7.000000	Determines the accuracy of the electrostatic interactions on the COSMO surface
Cosmo_Radius_Incr	0.000000	Specifies the increment to the atomic radii used in the construction of the COSMO cavity
Cosmo_RadCorr_Incr	0.150000	Used to construct the outer cavity for the outlying charge correction
Cosmo_A-Constraint	1.882190	Used to approximate the non-electrostatic contribution to the solvation energy within the COSMO model
Cosmo_B-Constraint	0.010140	Used to approximate the non-electrostatic contribution to the solvation energy within the COSMO model

^a Detailed parameter descriptions are available in the Accelrys MS software documentation.⁴⁵

We include the atomic radii for the following ten elements in the COSMO calculations keywords: hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorous, chlorine, bromine, and iodine. Klamt et al^{2,4} optimize these radii and suggest using 117% of the van der Waals bonding radius when approximating atomic radii for other elements. Both VT databases include compounds containing only the ten elements listed above; however, Klamt⁴⁹ recently published an optimized radius for silicon in addition to the ten elements listed in Table 3.2.

Table 3.2: Atomic Radii for Creating the COSMO Molecular Cavity^{2,4,49}

Element	Cavity radius [\AA]	Element	Cavity radius [\AA]
H	1.30	S	2.16
C	2.00	P	2.12
N	1.83	Cl	2.05
O	1.72	Br	2.16
F	1.72	I	2.32
Si	2.48		

3.1.3 COSMO Calculation – Surface Segment Averaging

For the third and final step, we average the surface segment charges generated from the COSMO calculation using equation (2.30) to yield the sigma profile. Table 3.3 summarizes the parameters values we use for the averaging calculation. We provide the source code, the executable FORTRAN program, and graphical procedures detailing the proper input formats required on our research group website (www.design.che.vt.edu).^{8,9}

Table 3.3: Parameter Values Used in the COSMO-SAC model.^{2,6}

Symbol	Units	Value	Description
r_{av}	\AA	0.81764	sigma averaging radius
a_{av}	\AA^2	2.100265	sigma averaging area
a_{eff}	\AA^2	7.5	effective surface segment surface area
c_{hb}	$\text{\AA}^4 \cdot \text{kcal}/(\text{e}^2 \cdot \text{mol})$	85580.0	hydrogen-bonding constant
σ_{hb}	$\text{e}/\text{\AA}^2$	0.0084	sigma cutoff for hydrogen-bonding
α'	$\text{\AA}^4 \cdot \text{kcal}/(\text{e}^2 \cdot \text{mol})$	16466.72	misfit energy constant
z	dimensionless	10	coordination number
q	\AA^2	79.53	standard area parameter
r	\AA^3	66.69	standard volume parameter

Computation times vary depending on the size and initial geometry of the molecule. On average, small molecules, such as methane and ethane, require less than five minutes of total computational time on a 3.6 GHz Pentium IV-equipped PC. For very large molecules, such as cholesterol, calculation takes up to 24 hours on the same machine. With the exception of our

case studies of conformational variations, both VT databases represent a single conformation for each compound; however, we acknowledge that large molecules can exist in alternate low energy conformations due to rotational freedoms.

3.2 *Geometry Optimization Tools*

In some cases, the above mentioned procedure fails to produce a conformation which meets the validation criteria set in Section 4.2. At this point, we seek further insurance that the DFT and COSMO calculations generate global low energy conformations by improving our initial guess for molecular structure. We use two supplemental software packages: (1) Amber8, developed by Pearlman et al⁴⁴ in 1995, which simulates energy minimization as well as several other tasks; and (2) MS Forcite Plus, a simulation module for annealing dynamics, in conjunction with our DMol3 calculations. We hope that in using an annealing simulator, the molecular conformation overcomes any energy barriers during the gradual heating and cooling and ultimately results in a global low energy conformation. We discuss these tools and the relevant parameters in the following sections.

3.2.1 Amber8

We summarize the calculation settings we use for energy minimization tasks in this section. We obtain the relevant information from the Amber8 User Manual⁴³ and the description of methods in Amber8 by Pearlman et al.⁴⁴ We perform a minimization task using the modules Antechamber, LEaP, and SANDER in Amber8. LEaP stands for Link, Edit, and Parm, which are all modules from previous versions of Amber, and SANDER stands for Simulated Annealing and NMR-Derived Energy Restraints. We use the Antechamber module to prepare our input files in Brookhaven PDB (Protein Data Bank) file format for use with LEaP. Antechamber generates molecular topology and is designed for use with GAFF, or General Amber Force Field. GAFF is best suited for organic molecules and contains 33 basic atom types and 22 special atom types covering the following atoms, carbon, nitrogen, oxygen, sulfur, phosphorous, hydrogen, fluorine, chlorine, bromine, and iodine. Therefore, this force field is well suited for our purposes. LEaP, which has both graphically interfaced and non-graphical versions, is a preparatory program similar to Antechamber. After we treat our input file with the Antechamber and LEaP module, we begin our energy minimization with SANDER, which uses possible

combinations of molecular dynamics, NMR structural refinement, and integrated Newtonian equations of motion for its calculations. Our minimization calculation uses only the integrated Newtonian equations of motion, not molecular dynamics or NMR structural refinement. We include our calculation scripts⁵⁰ in Appendix B. The final result from using Amber8 as a geometry optimization tool is a structure with a minimized molecular mechanics energy, which serves as an input structure for our sigma profile generation procedure. All force field, energy equations, and other relevant information contained within Amber8 are available in the program documentation⁴³ and in the summary by Pearlman et al.⁴⁴

3.2.2 MS Forcite Plus Annealing Dynamics

Accelrys' MS Forcite Plus is a molecular mechanics energy-optimization module which offers several tasks, including energy calculation, energy minimization or geometry optimization, molecular dynamics calculations, quench dynamics calculations, and anneal dynamics calculations in combination for use with the COMPASS, UFF, and Dreiding force fields. COMPASS is an ab initio force field which stands for Condensed phase Optimized Molecular Potentials for Atomistic Simulation Studies, and it is a high quality force field best suited for organics. UFF, or the Universal Force Field, covers all atoms of the periodic table but only moderately accurate for predicting conformational energies of organic molecules. The Dreiding force field uses parameters determined from hybridization rules and is also only moderately accurate for predicting conformational geometries and energies.⁴⁵ We generate multiple conformations for a single molecule using the "Anneal" task under "Ultra Fine" tolerances for all subtasks using the COMPASS force field. We summarize the relevant settings of our "Anneal" task in with default settings duly noted. MS Online Help⁴⁵ lists more detailed descriptions for the parameters in Table 3.4

Table 3.4: MS Forcite Plus Anneal Dynamics Calculation Settings⁴⁵

Parameter	Value	Default
COMPASS Force field Settings		
Charges	Force field assigned	yes
Quality	Ultra-fine	
Summation Method:		
Electrostatic	Atom-based	yes
van der Waals	Atom-based	yes
Forcite Anneal Dynamics Settings		
Anneal Cycles	5 - 20	5
Initial Temperature, K	300 - 400	300
Mid-Cycle Temp, K	> 500	500
Heating Ramps/Cycle	5 – 40, Variable	5
Dynamic Steps/Ramp	100 – 200, Variable	100
Total Steps	80000 – 200000, Variable	5000
Ensemble	NVE	yes
Time Step	1 f.s.	yes
Forcite Geometry Optimization After Each Anneal Cycle		
Algorithm	Smart	yes
Quality	Ultra-fine	
Max Iterations	500	

In Table 3.4, the NVE ensemble represents the microcanonical ensemble, which holds volume and energy constant. The NVE ensemble is adiabatic and conserves energy by allowing temperature and pressure to vary. The time step is in units of femtoseconds (10^{-15} seconds). MS Forcite Plus outputs the lowest energy conformation of each anneal cycle, and we use the lowest energy conformation from all anneal cycles as our initial guess for an optimized molecular structure. We find that the resulting low energy conformations are not necessarily unique for each anneal cycle, and several low energy conformations are identical, indicating that we have a higher probability of a global low energy conformation.

3.2.3 Qualitative Comparison of Methods

We compare low energy, condensed phase conformations from the DMol3 energy calculations using both Amber8 and MS Forcite Plus optimized output structures as our initial structure for 101 compounds based on predicted pure component vapor pressure. All 101 compounds initially failed to satisfy the validation criteria for release in the VT-2005 Sigma

Profile Database, see Section 4.2. Of the 101 compounds tests, Amber8 fails to generate results for six compounds and MS Forcite Plus fails on three occasions, narrowing our study to 93 total cases. We release 83 of these 93 cases because the improved structures produce improvements in vapor pressure predictions meeting our validation criteria. Of the released cases, 41 cases gives similar conformations using Amber8 and MS Forcite Plus output as an initial conformation. Overall, when using Amber8 for an initial conformation, 49 of the 83 cases generate better predictions than using MS Forcite Plus output, although some only marginally. Table 3.5 shows the breakdown for all 93 cases.

Table 3.5: Comparison of methods, Amber8 and MS Forcite Plus, as geometry optimization tools based on condensed phase DMol3 energy values E_h and pure component vapor pressure as predicted by COSMO-SAC-BP⁵¹ for 93 compounds which originally failed to meet acceptable error criteria.

Comparison Cases for MS Forcite and Amber 8		Best Prediction based on $\ln P^{\text{vap}}$ of released cases		
Total cases	93	MS Forcite	34/83	
Released cases	83	Amber 8	49/83	
	Cases	Cases not lowest energy ^a	Difference $\ln P^{\text{vap}}$ (pred.)	Difference Condensed phase Energy ^b , E_h
Similar Conformation		41/93		
Best Prediction: MS Forcite	19/41	11/19	5.23E-02	-2.00E-05
Best Prediction: Amber 8	16/41	7/16	9.48E-02	3.31E-04
Not Released	6/41	1/6	3.90E-02	7.73E-05
Different Conformation		52/93		
Best Prediction: MS Forcite	15/52	8/15	4.48E-01	-6.31E-04
Best Prediction: Amber 8	33/52	12/33	9.33E-01	1.60E-03
Not Released	4/52	3/4	1.45E+00	1.35E-03

^a In these cases, the lowest energy conformation did not provide the best prediction of vapor pressure

^b We calculate these values by subtracting converged DMol3 condensed phase energies (Energy Difference = Energy(MS Forcite) - Energy(Amber 8))

We also observe that for larger molecules like those with longer carbon backbones, MS Forcite Plus appears to generate “folded” structures more often than “unfolded” structures. The “folded” nature could be likened to protein denaturing, which occurs naturally. Figure 3.1 depicts the “folding” and “unfolding” of the alkyl chains with di-n-decyl-phthalate (VT-0690) generated with Amber8 and MS Forcite Plus.

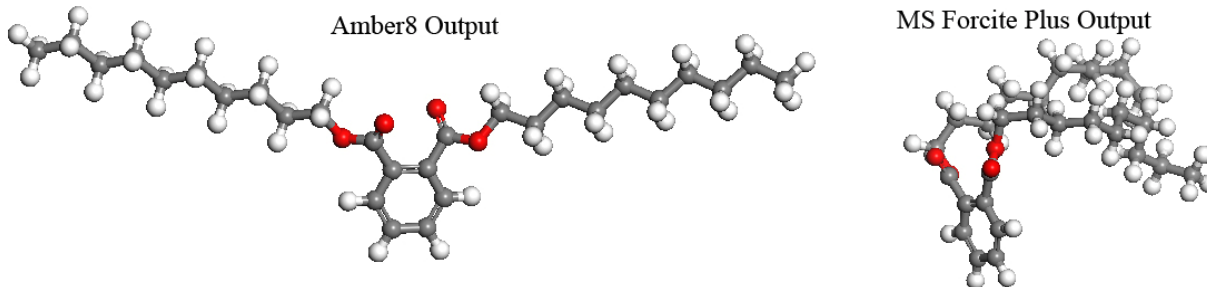


Figure 3.1: Energy minimization output from Amber8 and Annealing Dynamics output from MS Forcite Plus for di-n-decyl-phthalate (VT-0690). Neither conformation produced acceptable results.

However, the molecules we study are generally much smaller than proteins, so we question the possibility of a folded structure in some cases. It has yet to be explored extensively, but these folded structures may or may not improve property predictions for large molecules. For our purposes, we use Amber8 optimized conformations as input for the DMol3 calculation for all remaining unsatisfactory compounds in the VT-2005 Sigma Profile Database and for all compounds in the VT-2006 Solute Sigma Profile Database.

3.3 Convergence Method for Solubility Predictions

To insure convergence of the solute mole fraction, we damp the solution to the solubility equation (2.30) by a damping factor $\omega = 1/3$, which increases total computation time but is more stable than Newton's method. We also normalize all mole fractions prior to the next iteration step to prevent calculating mole fractions outside of the possible physical range. When normalizing mole fractions for a system with more than two solvents, we keep the mole fraction ratio of those two solvents constant throughout each iteration for a given temperature. Equation (3.1) shows the convergence scheme we use for generating the initial guess for the solute mole fraction for the $i+1$ iteration based on a damping factor ω , the value for the solute mole fraction from the previous iteration I , and the solution to the solubility equation (2.30).

$$x_{sol}^{i+1} = (\omega) x_{sol}^{Eqn(2.30)} + (1 - \omega) x_{sol}^i \quad (3.1)$$

This scheme requires the user to supply an initial guess for the solute mole fraction, and we use the literature value for validation purposes; however, if a literature value is not available, we recommend using the ideal solubility as an initial guess.

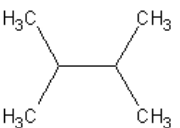
4 VT-2005 Sigma Profile Database

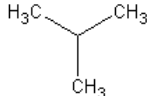

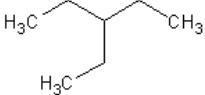

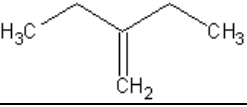
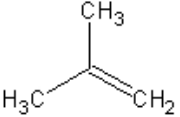
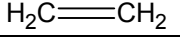
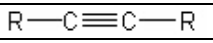
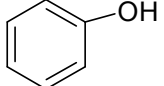
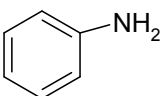
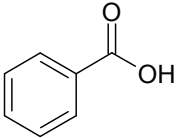
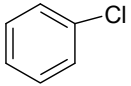
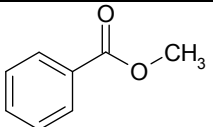

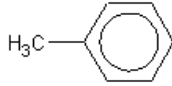
In this chapter, we discuss the VT-2005 Sigma Profile Database including its contents, the validation criterion, the effect of conformational variations on thermodynamic property prediction, improvements made since its initial release in *Industrial & Engineering Chemical Research* in 2006, and several application guidelines.

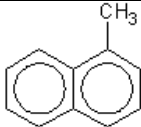
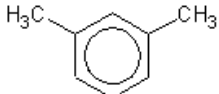
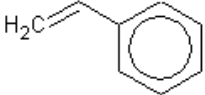
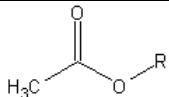
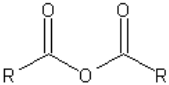
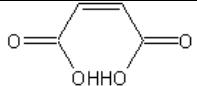
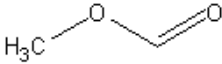
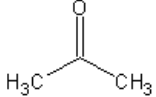

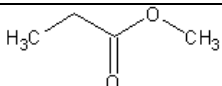
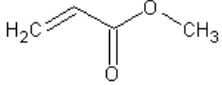
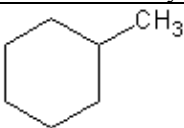
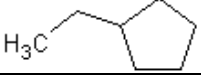
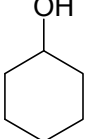

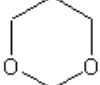
4.1 Description of Database

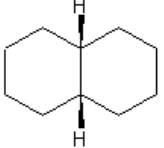
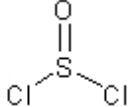
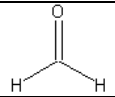
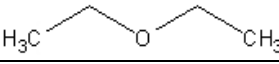

We identify each compound by a unique VT-2005 index number, its CAS-RN (registry number), chemical formula, and name in the VT-2005 index, which we tabulate in a spreadsheet file. We recommend searching the database by CAS-RN. The chemical family classification is another searchable category; however, many molecules fit into more than one group classification. Some of the functional groups represented in the VT-2005 Sigma Profile Database are acetates, alcohols, aldehydes, alkanes, alkenes, alkynes, anhydrides, aromatics, carboxylic acids, cyclic compounds, cyanates, elements, epoxides, esters, ethers, formates, halogenated compounds, inorganic acids and bases, ketones, mercaptans, nitriles, nitro compounds, peroxides, sulfides, thiophenes, etc. Table 4.1 summarizes the chemical family classification and their respective number of compounds listed in the database. There are a total of 1432 compounds listed in this database, each having one sigma profile unless otherwise stated. Conformationally flexible molecules can have multiple sigma profiles, and we present several examples of the effect of conformational variations on sigma profiles later in this chapter. The large majority of molecules in this database are small to medium-sized molecules (containing ten carbons or less) and liquids at room temperature.


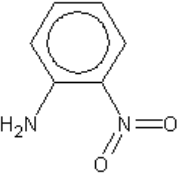
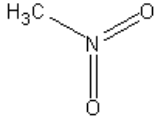
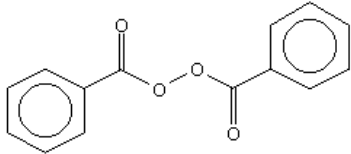
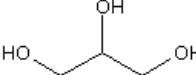
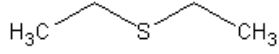
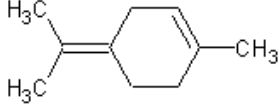
Table 4.1: Summary of the VT-2005 Sigma Profile Database

Chemical Family	Characteristic Structure	Number of Compounds in VT-2005	Example Compounds
Alkanes			
Dimethyl-Alkanes		21	2,3-dimethyl-butane

Methyl-Alkanes		17	Isobutane
n-Alkanes		27	Propane
Other Alkanes		27	3-Ethylpentane
Alkenes			
Dialkenes		25	1,4-pentadiene
Ethyl/higher-Alkenes		11	2-ethyl-1-butene (C ₆ H ₁₂)
Methyl-Alkenes		22	Isobutylene
n-Alkenes		41	Ethylene, 1-Butene, cis-2-Butene
Alkynes			
n-Alkynes		18	1-Butyne (C ₄ H ₆)
Aromatic structures			
Aromatic Alcohols		31	Phenol
Aromatic Amines		37	Pyridine and Aniline
Aromatic Carboxylic Acids		9	Benzoic Acid (C ₇ H ₆ O ₂)
Aromatic Chlorides		15	Benzyl Chloride (C ₇ H ₇ Cl)
Aromatic Esters		19	Benzyl Acetate (C ₉ H ₁₀ O ₂)
Diphenol/Polyaromatics		17	Biphenyl (C ₁₂ H ₁₀)
n-Alkyl Benzenes		19	Toluene

Naphthalenes		15	1-Methylnaphthalene (C ₁₁ H ₁₀)
Other Alkyl Benzenes		44	m-Xylene (C ₈ H ₁₀)
Other Monoaromatics		15	Styrene
Carboxylic compounds			
Acetates		22	Methyl Acetate, Vinyl Acetate, Allyl Acetate
Anhydrides		8	Acetic Anhydride
Dicarboxylic Acids		11	Maleic Acid (C ₄ H ₄ O ₄)
Formates		13	Ethyl Formate (C ₃ H ₆ O ₂)
Ketones		33	Acetone
n-Aliphatic Acids		20	Formic Acid, Acetic Acid, Propionic Acid
Propionates and Butyrates		13	Methyl Propionate (C ₄ H ₈ O ₂)
Unsaturated Aliphatic Esters		23	Methyl Acrylate (C ₄ H ₆ O ₂)
Cyclic compounds			
Alkylcyclohexanes		19	Methylcyclohexane
Alkylcyclo-pentanes		22	Ethylcyclopentane
Cycloaliphatic Alcohols		10	Cyclohexanol
Cycloalkanes/ alkenes		15	Cyclooctene (C ₈ H ₁₄) or Cyclobutane (C ₄ H ₈)
Epoxides		14	1,3-Dioxane (C ₄ H ₈ O ₂), Furan, Ethylene Oxide

Multiring-cycloalkanes		3	Cis-Decalin (C ₁₀ H ₁₈)
Other Hydrocarbon Rings	No Common Structure	16	Indene (C ₁₀ H ₁₆)
Halogenated compounds			
F, Cl, Br, I Compounds	H ₃ C—X	94	Methyl Bromide, Ethyl Chloride, Difluoromethane
Multihalogenated Alkanes	X—CH ₂ —X	37	Chlorofluormethane
Inorganic compounds			
Inorganic Acids/Bases	No Common Structure	9	Sulfuric Acid, Nitric Acid, Ammonia
Inorganic Gases	No Common Structure	8	Ozone (O ₃)
Inorganic Halides		5	Thionyl Chloride (SOCl ₂)
Other Inorganics	No Common Structure	3	Water
Polyfunctional compounds			
Polyfunctional Acids	No Common Structure	16	Glycolic Acid (C ₂ H ₄ O ₃)
Polyfunctional Amides/Amines	No Common Structure	26	Formamide
Polyfunctional C,H,O,halides	No Common Structure	36	Chloroacetic Acid
Polyfunctional-C,H,N,halide,(O)	No Common Structure	12	O-Chloroaniline
Polyfunctional C,H,O,N	No Common Structure	27	Niacin
Polyfunctional C,H,O,S	No Common Structure	13	Sulfolane (C ₄ H ₈ O ₂ S)
Polyfunctional Esters	No Common Structure	21	Ethyl Lactate (C ₅ H ₁₀ O ₃)
Polyfunctional Nitriles	No Common Structure	7	Aminocapronitrile (C ₆ H ₁₂ N ₂)
Other Polyfunctional C,H,O	No Common Structure	35	Furfural
Other Polyfunctional Organics	No Common Structure	4	Malathion (C ₁₀ H ₁₉ O ₆ PS ₂)
Other structures			
Aldehydes		31	Formaldehyde
Aliphatic Ethers		33	Diethyl Ether
Elements	No Common Structure	5	Hydrogen, Bromine, Iodine
Isocyanates/Diisocyanates	R—N=C=O	6	n-butyl-diisocyanate (C ₅ H ₉ NO)
Mercaptans	HS—CH ₃	22	Methyl Mercaptan (CH ₄ S)
n-Alcohols	R—OH	20	Methanol, Ethanol, 1-Octanol
n-Aliphatic Primary Amines		13	Ethyl Amine

Nitriles		28	Acetonitrile (C ₂ H ₃ N)
Nitroamines		4	o-Nitroaniline (C ₆ H ₆ N ₂ O ₂)
C, H, NO ₂ Compounds		20	Nitromethane (CH ₃ NO ₂)
Sulfates, Nitrates, Phosphates, Carbonates	No Common Structure	11	Ethylene Carbonate (C ₃ H ₄ O ₃)
Peroxides		10	Benzoyl Peroxide (C ₁₄ H ₁₀ O ₄)
Polyols		33	Glycerol
Sulfides/ Thiophenes		21	Diethyl Sulfide (C ₄ H ₁₀ S)
Terpenes		6	Terpinolene (C ₁₀ H ₁₆)
Other Aliphatic Acids	No Common Structure	19	Isobutyric Acid (C ₄ H ₈ O ₂)
Other Aliphatic Alcohols	No Common Structure	26	Isobutanol
Other Aliphatic Amines	No Common Structure	16	Dimethylamine
Other Amines/Imines	No Common Structure	36	Pyrazine (C ₄ H ₄ N ₂)
Other Ethers/Diethers	No Common Structure	17	Anethole (C ₁₀ H ₁₂ O)
Other Saturated Aliphatic Esters	No Common Structure	19	Caprolactone (C ₆ H ₁₀ O ₂)
Miscellaneous Other	No Common Structure	3	Ethanesulfonyl-Chloride (C ₂ H ₅ ClO ₂ S)
Condensed Rings	No Common Structure	8	Anthracene (C ₁₄ H ₁₀), Pyrene (C ₁₆ H ₁₀)

4.2 Validation of VT-2005 Sigma Profile Database

In this section, we discuss the validation criterion for releasing each database entry and present several examples of validated predictions of activity coefficients and pure component vapor pressure using the COSMO-SAC model. For the purposes of evaluating the error

associated with the various predicted properties, we use absolute average relative percent error (AA%E) to quantify the prediction error.

$$AA\%E = \frac{1}{n} \sum_i^n \left| \frac{f_i^{exp} - f_i^{pred}}{f_i^{exp}} \right| \times 100\% \quad (4.1)$$

In each equation, f^{exp} and f^{pred} represent the experimental variable and the predicted variable, respectively. These are generic variables, so we specify which variable we refer to in the text, but generally we calculate the prediction error in terms of pressure and composition. We discuss the differences in procedure and generated sigma profiles generated by Lin and Sandler and the sigma profiles in our databases in Appendix A of Mullins et al.^{8,9}

4.2.1 Pure Component Vapor-Pressure Predictions

Each compound in the VT-2005 Sigma Profile Database has a predicted pure component vapor pressure calculated at its respective normal boiling temperature with the VT-2005 Sigma Profile using a revised COSMO-SAC-BP model developed by Wang et al.⁵¹ Here, we summarize the predicted values for each compound using the natural logarithm of the vapor pressure. Theoretically, each compound's vapor pressure should have a pure component vapor pressure of 101325 Pa at its boiling point (on a logarithmic scale, $\ln(101325) = 11.5261$). Figure 4.1 summarizes the distribution of vapor-pressure predictions for every compound in the database. The average predicted $\ln(P_{vap})$ is 11.5405 with a standard deviation of 0.4435.

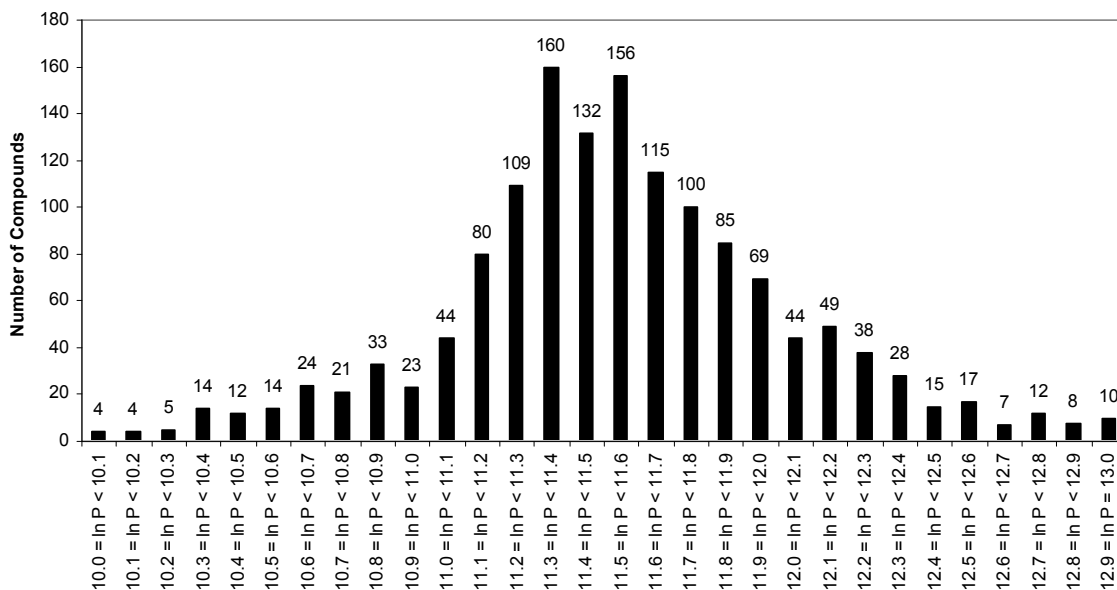


Figure 4.1: Distribution of pure component vapor-pressure predictions at each compounds respective normal boiling point temperature with a revised COSMO-SAC-BP model⁵¹ on a logarithmic scale.

We use these pure component vapor-pressure predictions as the sole criterion for releasing the sigma profile for compound. Our validation criterion requires that the logarithm of the vapor pressure must be in the range of 10.0 – 13.0 ($10.0 \leq \ln P_{\text{vap}} \leq 13.0$), ln Pa units.

4.2.2 Activity-Coefficient Predictions

We validated our COSMO-SAC predictions against work by Lin and Sandler⁶ and COSMO-RS predictions by Eckert¹⁴ and Giles.⁵² Oldland⁸ performed this work originally and published it in Section 4.2.2 of Mullins et al.^{8,9}

4.3 Details of VT-2005 Sigma Profile Database Release and Refinement

We release the VT-2005 Sigma Profile Database in *Industrial Chemical and Engineering Research* in June 2006.⁹ We initially consider 1515 compounds, which we later narrowed to 1432 compounds for the publication version. We finalize the number of compounds at 1432 after two rounds of pure component vapor-pressure calculations. The first round includes vapor-pressure calculations using a revised COSMO-SAC-BP model⁵¹ for all 1515 compounds. After the first round, 1266 compounds satisfy the release validation criterion.

The second round includes the remaining 249 compounds, which do not meet the validation criterion. Checks of these compounds find several problems, including improper calculation settings in the DMol3 calculation, incorrectly drawn structures regarding proper atom connectivity or chemical formula, or an unconverged solution to a DMol3 calculation. Prior to performing the second round of vapor-pressure calculations, we refine each conformation using the pre-optimization tools, Amber8 and MS Forcite Plus, and the procedure laid out in Chapter 3. Of the second round compounds, 166 compounds meet the validation criterion after conformation refinement with Amber8 and MS Forcite Plus, which results in a total of 1432 compounds. Despite correcting the problems with convergence, calculation settings, and incorrect structures, 83 compounds still fail to satisfy the validation criterion. We may release the remaining 83 compounds, provided we can make further improvements to where the vapor-pressure predictions fall within the acceptable range.

4.4 Effect of Conformational Isomerism on Thermodynamic Property

Prediction

We describe the possible effects and potential causes of conformational isomerism in structurally flexible molecules and the resulting variations in COSMO-SAC activity coefficient prediction. The COSMO-SAC model makes two crucial assumptions. The first assumption is that the optimized geometry from the DMol3 calculation in the vapor phase is identical to the optimal geometry in the condensed phase. Factors such as solvent polarity, molecule size, and solvent-solute interactions could affect a solute molecule's structural conformation. The second assumption requires that the molecule is in lowest energy conformation, but several low energy structural conformations may exist because of the freedom in choosing dihedral angles. Each conformation results in a slightly different sigma profile and therefore may effect property predictions.

We cannot guarantee with complete certainty that our geometry-optimization DMol3 calculations yield the lowest energy state possible, simply because energy optimizations can result in local minima instead of global minima. We present evidence that these assumptions may not hold for certain cases in Table 3.5 and in Section 5.3. We compare two low energy conformations of several molecules, using Amber8 and MS Forcite Plus output for initial

conformations in the DMol3 calculations, and find in some cases that the lower-energy conformation predicts the less accurate pure component vapor-pressure. This situation occurs both when the final conformations are similar and different. We define similar conformations as structures with similar relative atomic spatial positions and bond angles, and different conformation as any isomer of that conformation. Figure 4.2 and Figure 4.3 shows an example of the different conformations of 3-methyl-2-butanol (VT-0499), which results from using the two pre-optimization tool outputs as initial conformations, and their respective sigma profiles. In nature, the different bond angles of the hydroxyl hydrogen would represent rotational freedom; however, because we calculate sigma profiles from rigid structures, these conformations are different.

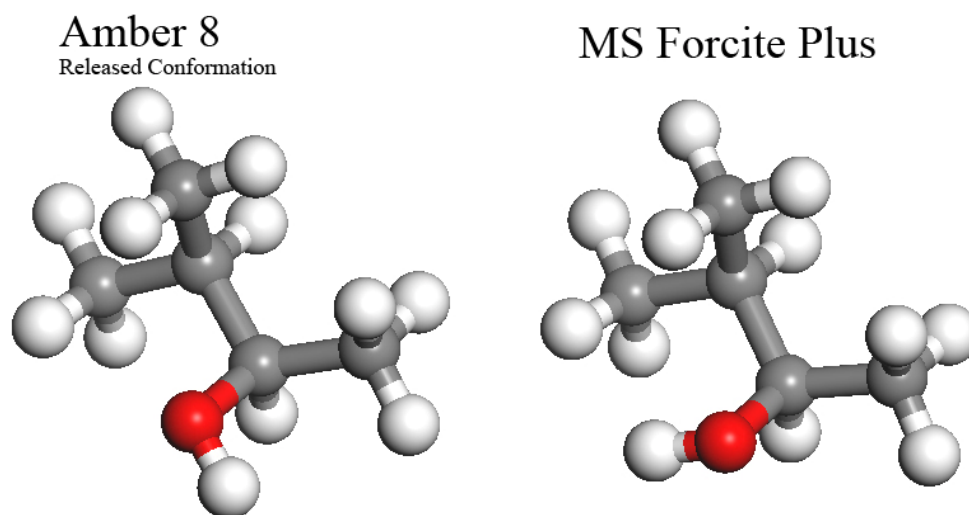


Figure 4.2: Two conformations of 3-methyl-2-butanol (VT-0499), which differ only in the relative bond angles of the hydroxyl hydrogen atom.

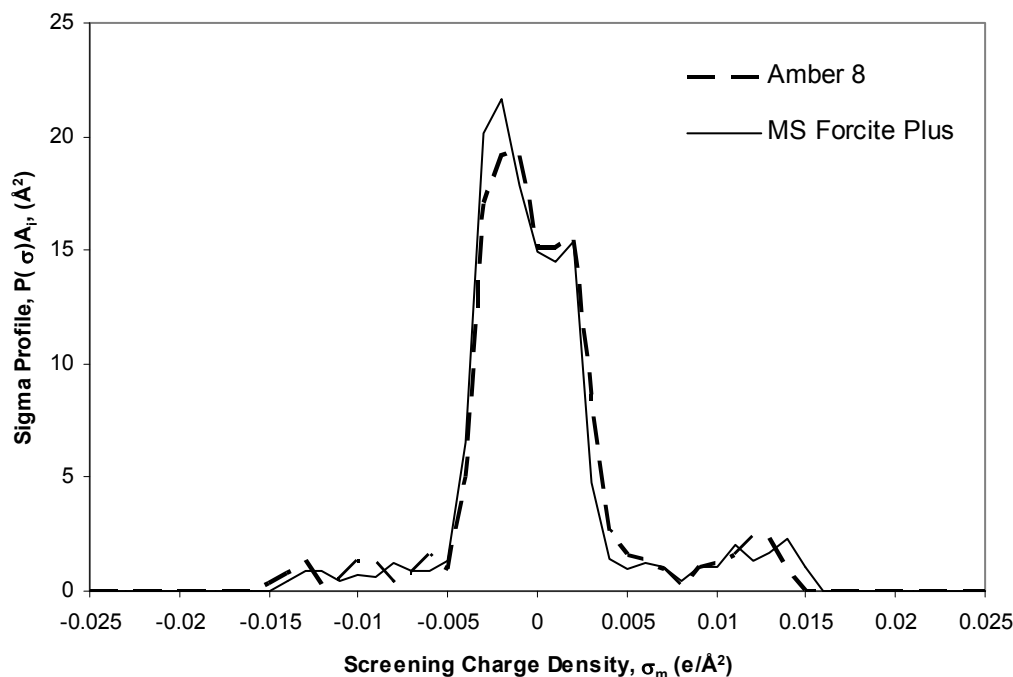


Figure 4.3: Two conformations and sigma profiles of 3-methyl-2-butanol (VT-0499) after DMol3 geometry optimization using Amber8 and MS Forcite as initial conformations.

We present two cases for comparison, a small and a medium-sized molecule. We use 2-methoxy-ethanol (VT-1397) for the small molecule and benzyl benzoate (VT-0676) for the medium-sized molecule. We compare sigma profiles and pressure-composition predictions to published data for several solvents for each case.

4.4.1 Small Molecule Example: 2-Methoxy-ethanol

We use the COSMO-SAC model to predict activity coefficients for binary systems of *n*-hexane (VT-0009), *n*-heptane (VT-0014), cyclohexane (VT-0099), methanol (VT-0477), methyl acetate (VT-0638), diisopropyl ether (VT-0713), and trichloroethylene (VT-0802) with 2-methoxy-ethanol (VT-1397) at various temperatures. We generate *P*-*x*-*y* predictions using the modified Raoult's Law and compare these predictions to literature data. We use pure component vapor-pressure data or predictions from the Antoine equation if data are unavailable. Figure 4.4 shows the released conformation from the VT-2005 Sigma Profile Database and three alternative conformations (labeled A – C) for comparison. We use Amber8 output to generate conformation

A and manually drew conformations B and C. We optimize each conformation following the procedure outlined in Chapter 3.

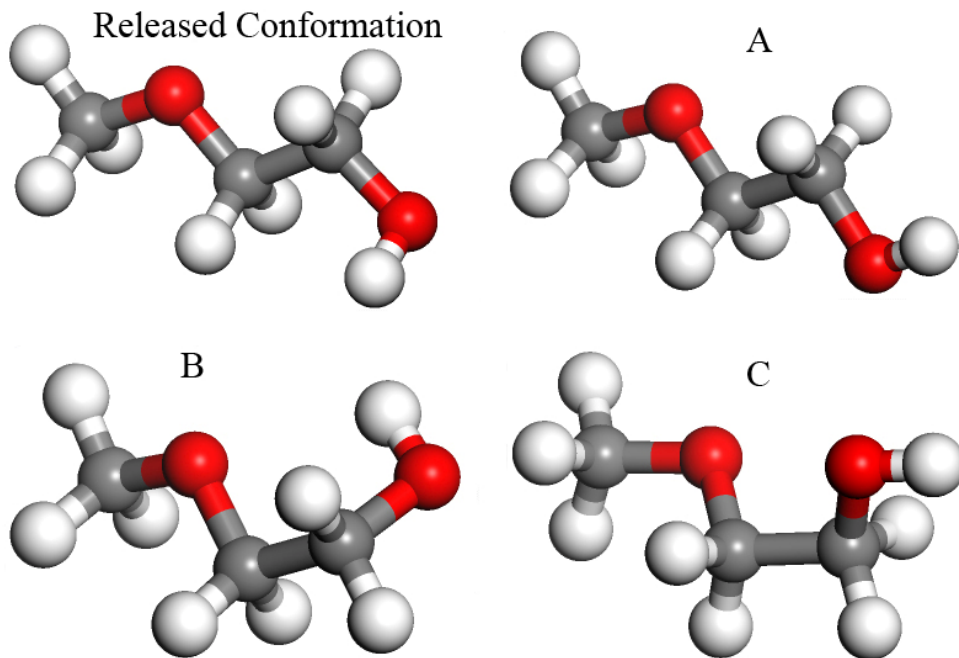


Figure 4.4: Low energy conformations for 2-methoxy-ethanol (VT-1397).

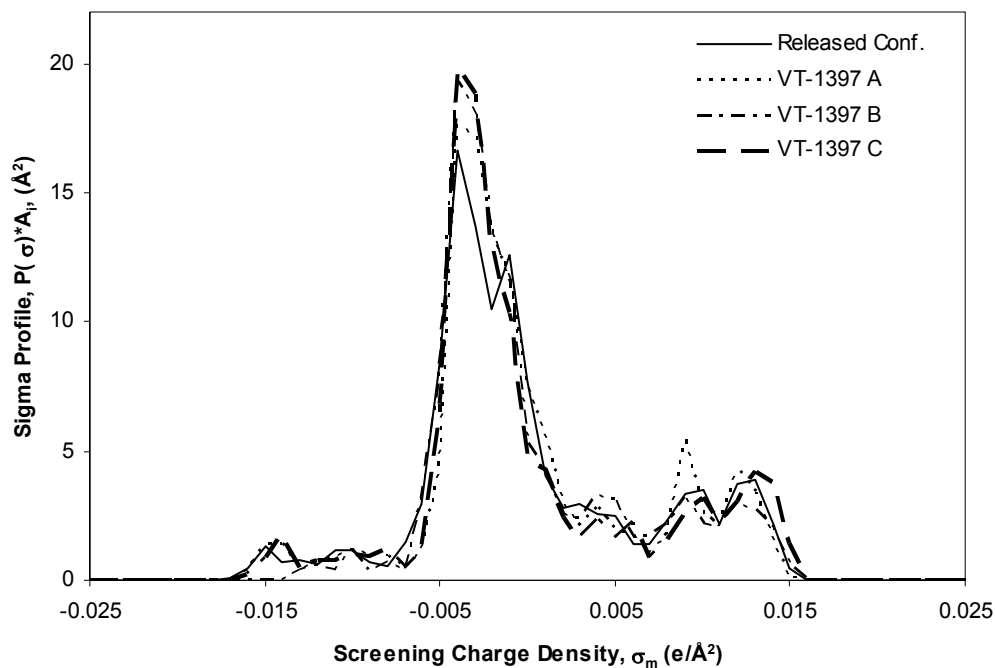


Figure 4.5: Sigma Profiles for various conformations of 2-methoxy-ethanol (VT-1397). The released conformation is included in the VT-2005 Sigma Profile Database.

Figure 4.5 illustrates that each of these structures produces a slightly different sigma profile. Conformation B contains an intramolecular hydrogen bond, whereas the other structures would be more likely to form intermolecular hydrogen bonds in solution. There are no open literature resources available concerning what percentage of each structure actually exists in solution. We examine pressure-composition data for systems with cyclohexane. Figure 4.6 through Figure 4.8 compare the COSMO-SAC predictions with experimental data⁵³ at temperatures of 303.15, 313.15, and 323.15K, respectively. Figure 4.9 through Figure 4.11 compare our predictions for binary systems with other nonpolar solvents, n-hexane and n-heptane with experimental data.⁵³ Figure 4.12 through Figure 4.17 compare COSMO-SAC predictions with polar solvents (i.e. methanol, methyl acetate, diisopropyl ether, and trichloroethylene) to published data.^{54,55}

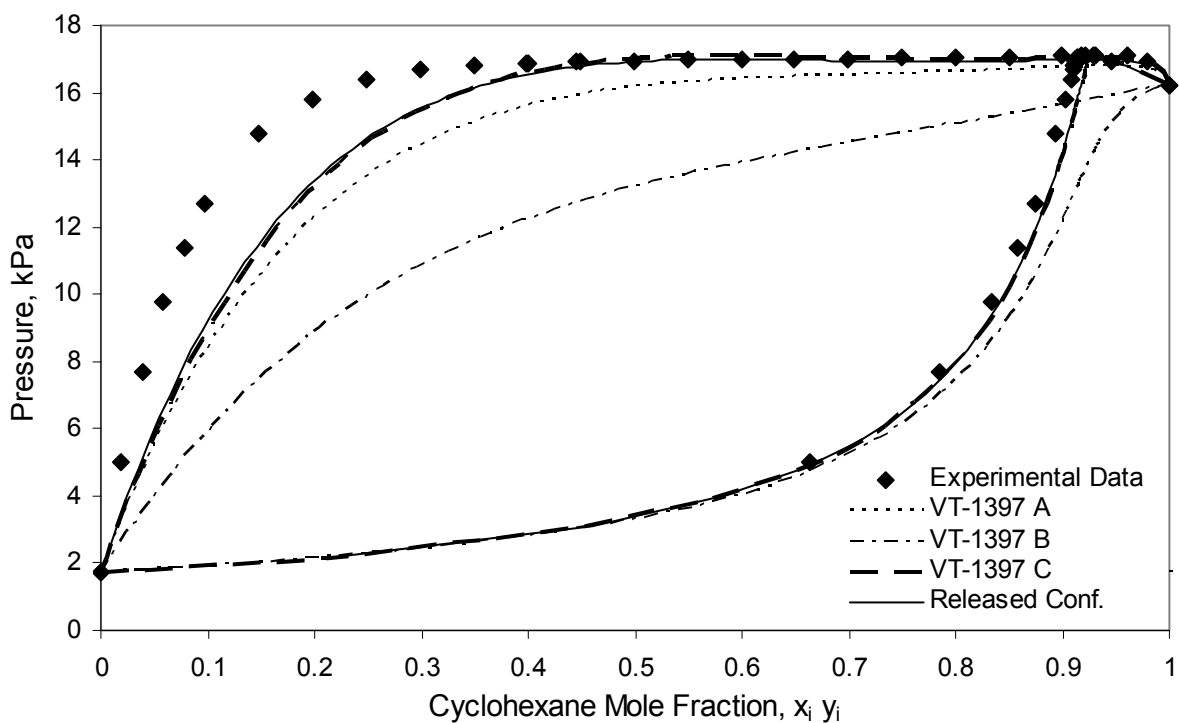


Figure 4.6: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) cyclohexane at 303.15 K.⁵³

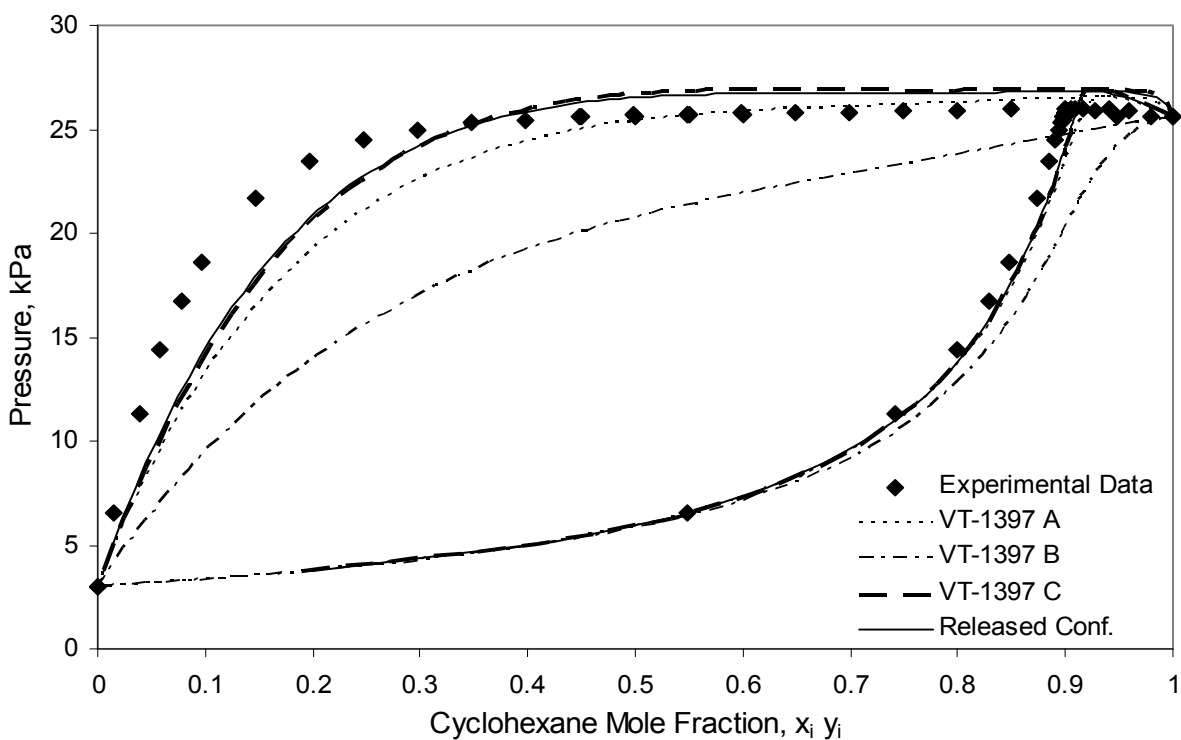


Figure 4.7: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) cyclohexane at 313.15 K.⁵³

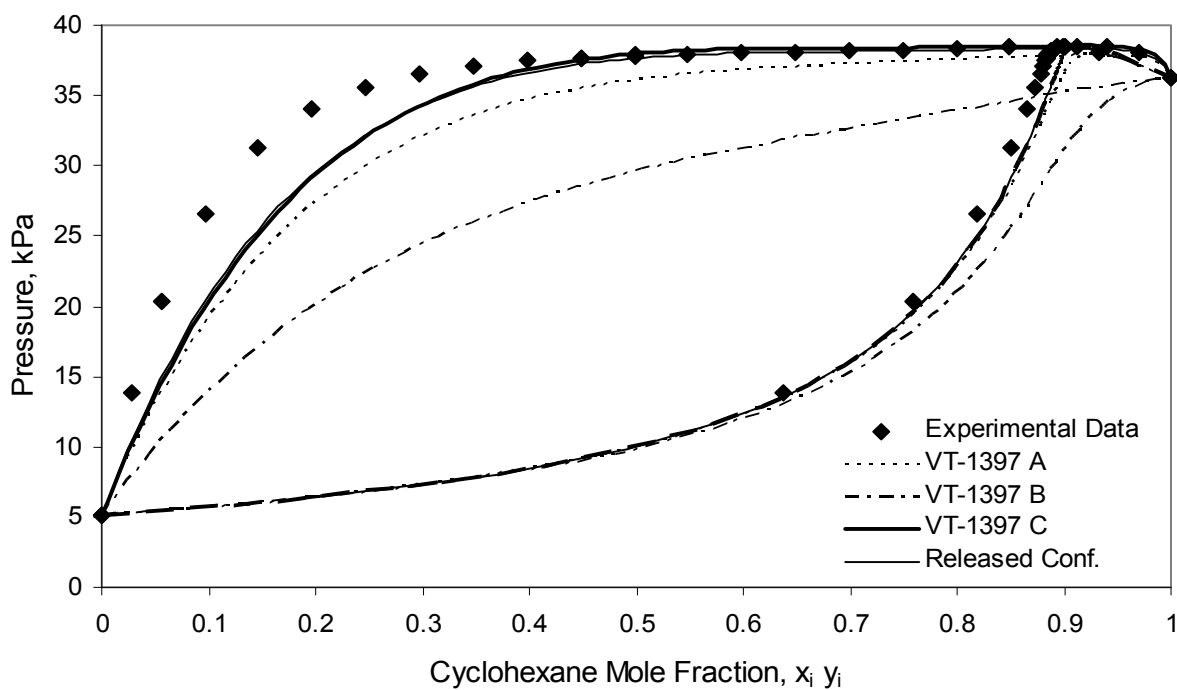


Figure 4.8: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) cyclohexane at 323.15 K.⁵³

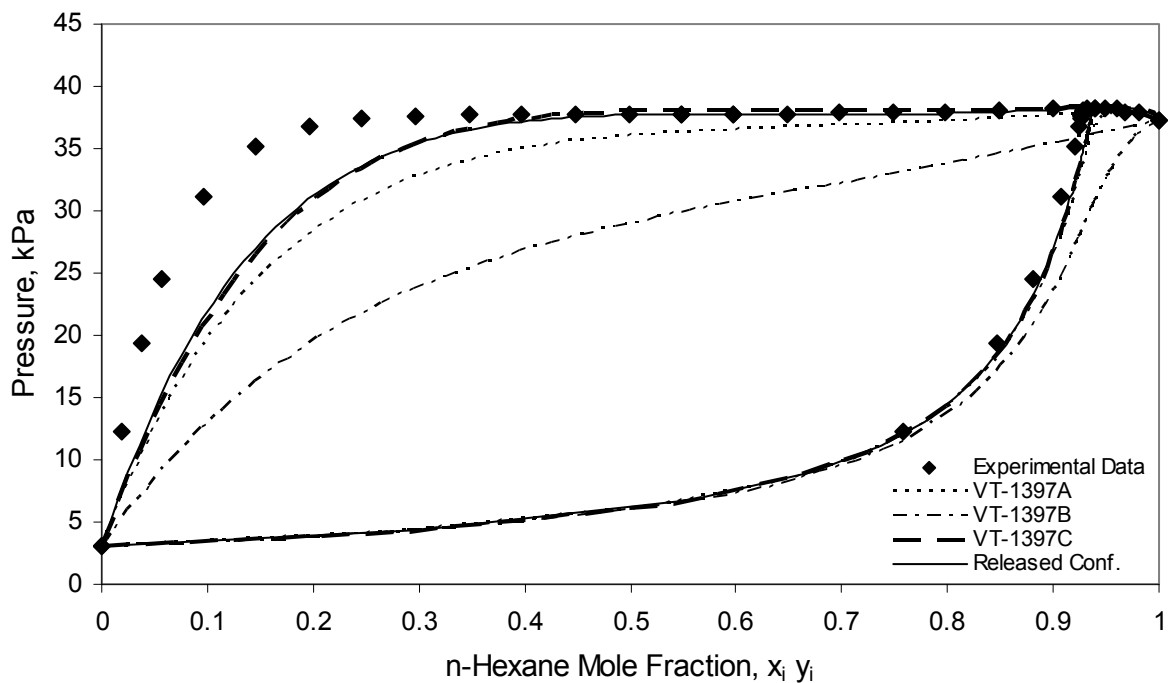


Figure 4.9: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) n-hexane at 313.15 K.⁵³

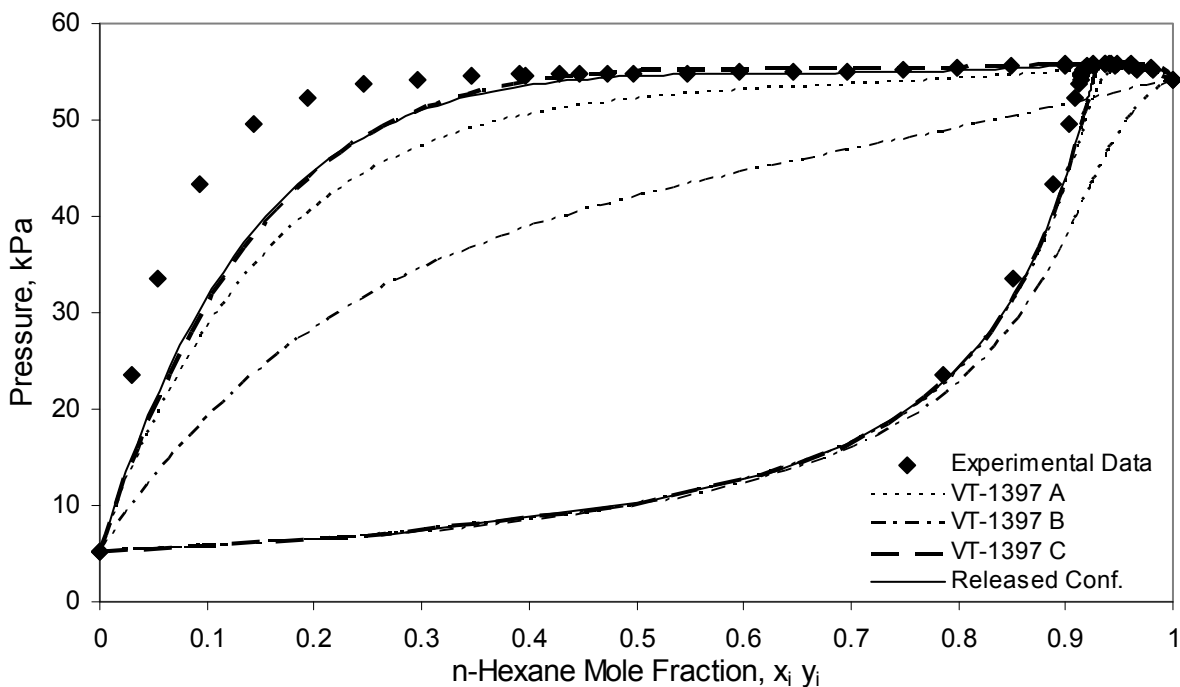


Figure 4.10: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) n-hexane at 323.15 K.⁵³

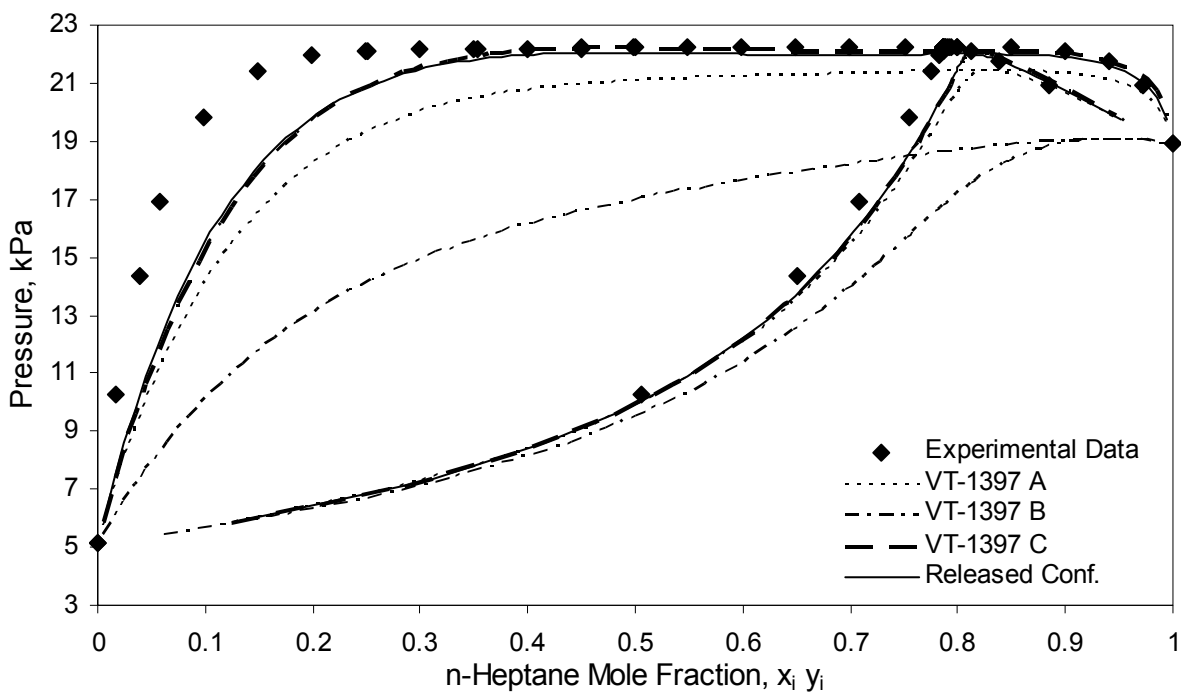


Figure 4.11: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) n-heptane at 323.15 K.⁵³

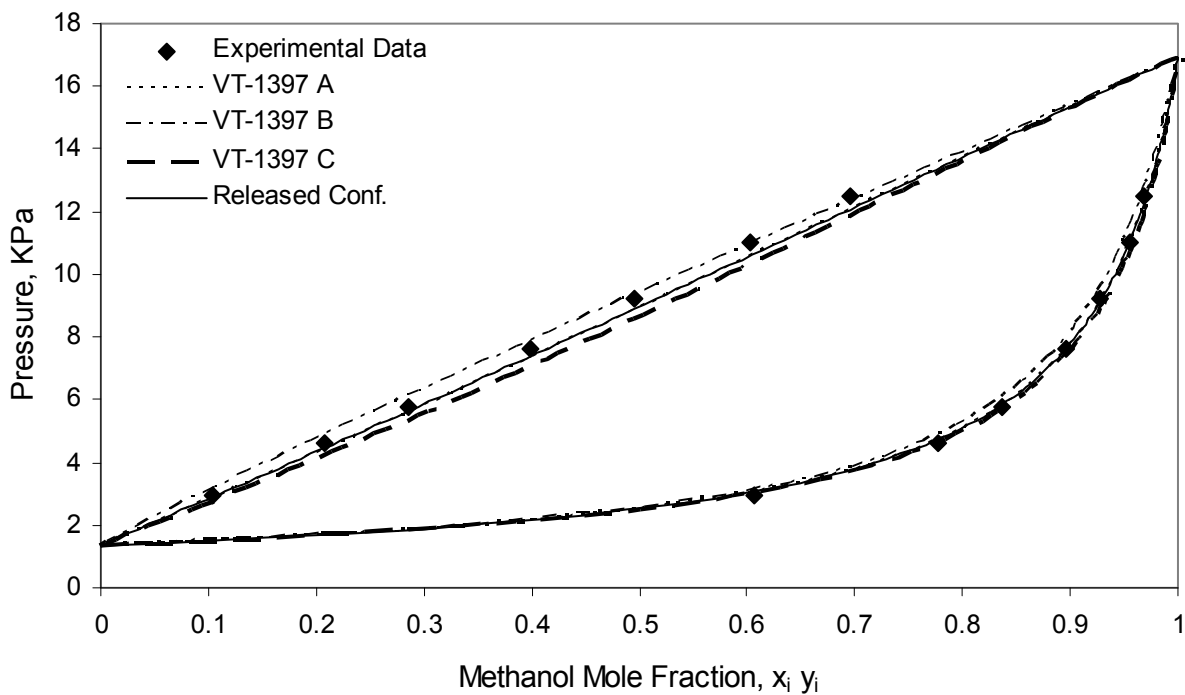


Figure 4.12: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) methanol at 298.15 K.⁵³

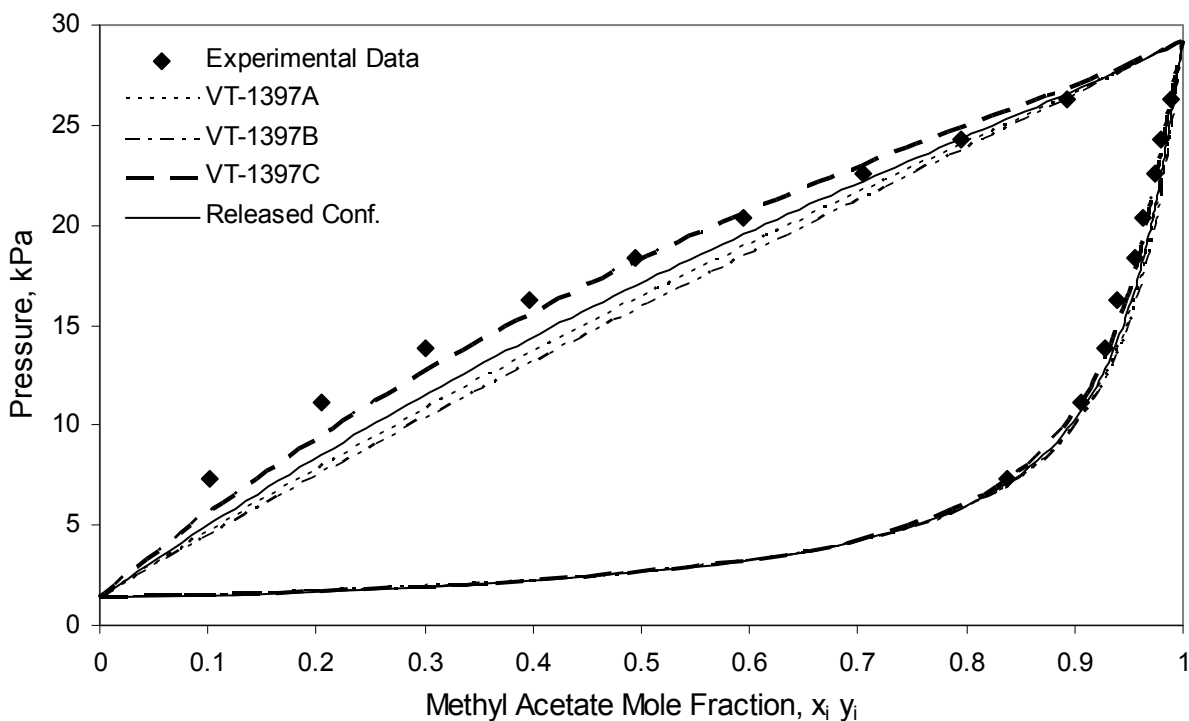


Figure 4.13: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) methyl acetate at 298.15 K.⁵⁴

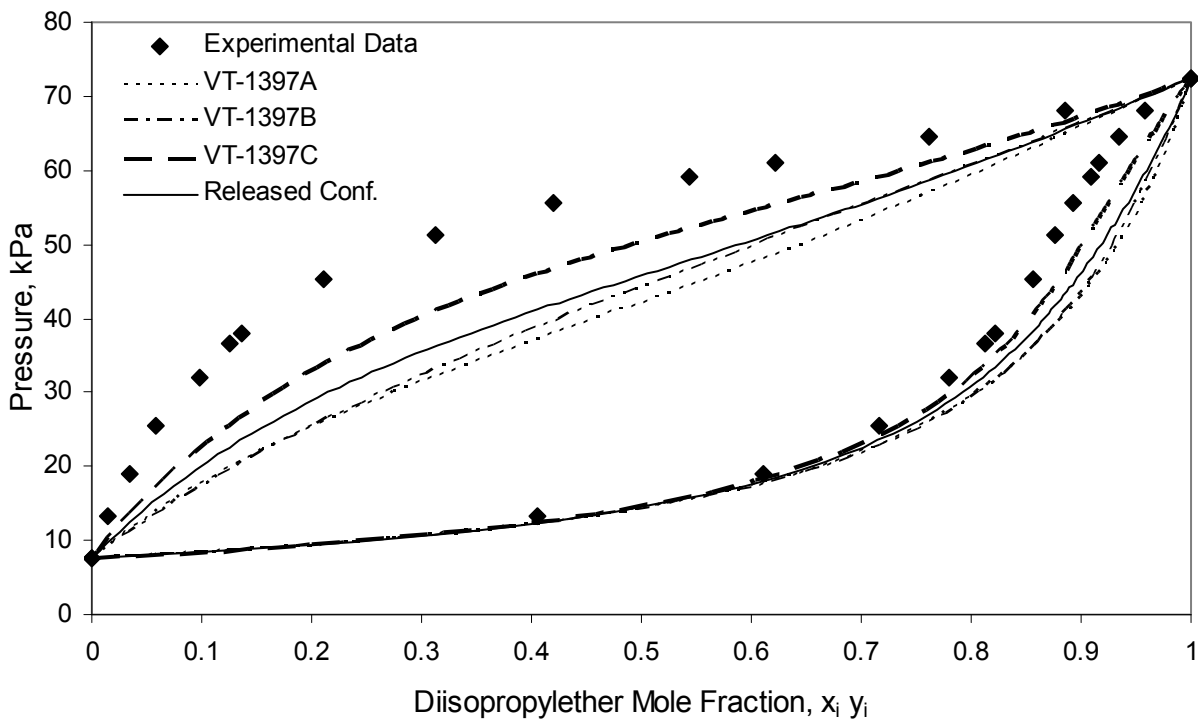


Figure 4.14: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) diisopropyl-ether at 331.02 K.⁵⁵

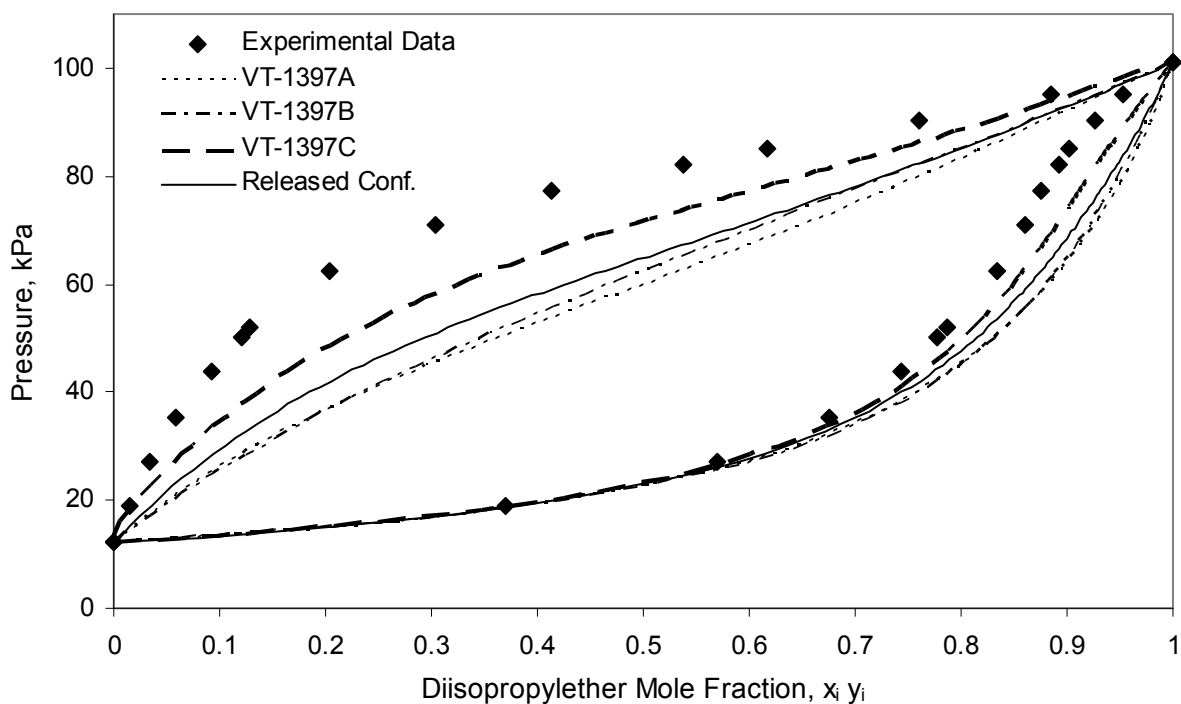


Figure 4.15: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) diisopropyl-ether at 341.01 K.⁵⁵

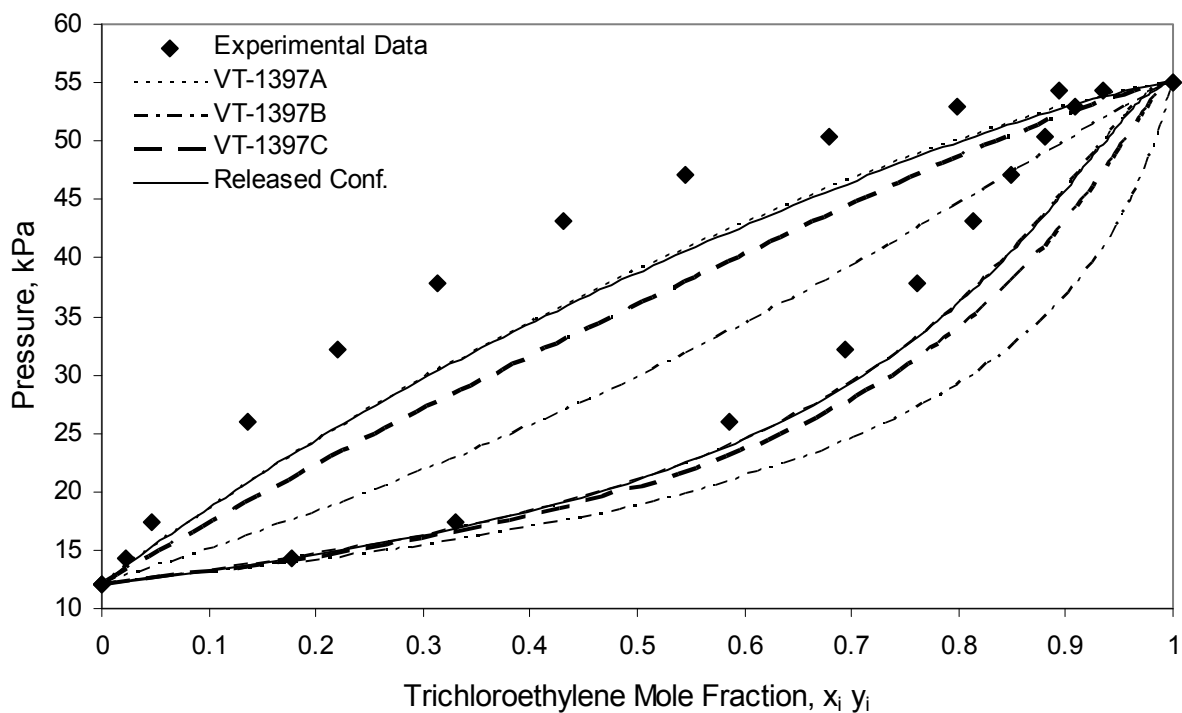


Figure 4.16: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) trichloroethylene at 341.01 K.⁵⁵

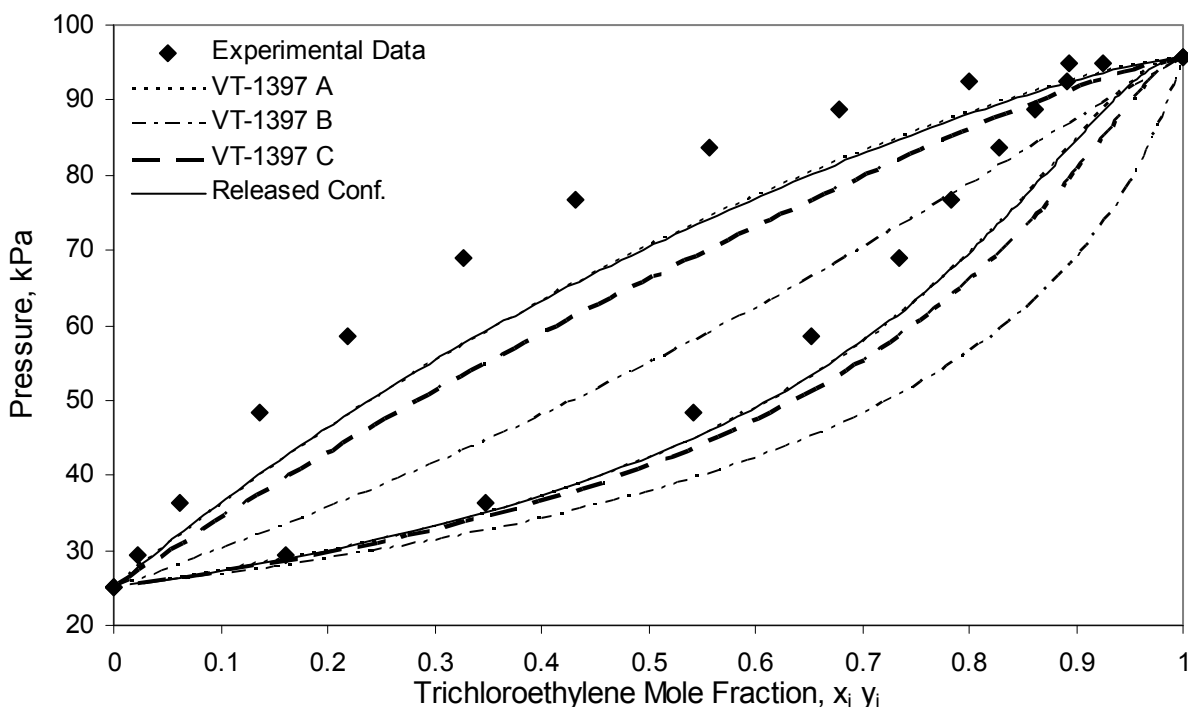


Figure 4.17: Pressure-composition data and COSMO-SAC predictions for (1) 2-methoxy-ethanol (2) trichloroethylene at 357.98 K.⁵⁵

The COSMO-SAC predictions generally improve as the temperature increases for each conformation, based on absolute average relative percent error (AA%E) calculations using equation (4.1) for the vapor mole fraction (m.f.) of 2-methoxy-ethanol, with the exception of diisopropyl ether. Table 4.2 summarizes AA%E values for each system and conformation of 2-methoxy-ethanol. The released conformation from the published database produces the best overall predictions for all of the solvents, with an AA%E of 5.66%. Conformation C also adequately predicts pressure-composition data, with a similar AA%E of 5.81%. The difference in error between the released conformation and conformation C is essentially insignificant, despite the apparent differences in their conformations. The COSMO-SAC predictions are the most accurate for the (1) 2-methoxy-ethanol (2) methanol system, which seems consistent with Lin and Sandler's observation that the COSMO-SAC model accuracy decreases for highly hydrophobic systems.⁶

Each of the six nonpolar solvents forms an azeotrope with 2-methoxy-ethanol. Table 4.3 compares the predicted azeotropic composition with the experimental value.⁵³ We note that conformation B does not predict an azeotrope in five of the six systems and predicts poorly in the

case of *n*-heptane. We interpolate the azeotropic composition from experimental data. In most cases, conformation C generates the most accurate azeotropic composition.

Table 4.2: AA%E of vapor mole fraction 2-methoxy-ethanol, % m.f..

Solvent	Temp, K	VT-1397 A	VT-1397 B	VT-1397 C	Released Conf.
Cyclohexane	303.15	3.89	7.71	3.35	3.03
	313.15	4.04	7.75	3.58	3.26
	323.15	3.44	7.13	2.86	2.63
n-Hexane	313.15	3.36	6.71	2.88	2.62
	323.15	2.42	5.31	1.97	1.81
n-Heptane	323.15	6.01	12.78	5.11	4.61
Methanol	298.15	1.67	0.39	2.44	1.55
Methyl acetate	298.15	2.29	2.63	1.30	1.85
Diisopropyl ether	331.02	18.51	19.29	10.48	14.15
	341.01	19.71	20.83	11.37	15.22
Trichloroethylene	341.01	8.96	20.21	12.52	8.79
	357.98	8.63	19.90	11.87	8.40

Table 4.3: Predicted azeotropic compositions for non-polar solvents with 2-methoxy-ethanol, m.f.

Solvent	Temp, K	VT-1397 A	VT-1397 B	VT-1397 C	Released Conf.	Experimental
Cyclohexane	303.15	0.9440	N/A	0.9307	0.9316	0.9204
	313.15	0.9393	N/A	0.9251	0.9264	0.9112
	323.15	0.9273	N/A	0.9123	0.9136	0.9001
n-Hexane	313.15	0.9645	N/A	0.9519	0.9538	0.9408
	323.15	0.9587	N/A	0.9447	0.9470	0.9349
n-Heptane	323.15	0.8298	0.9443	0.8124	0.8116	0.7939

4.4.2 Medium-Sized Molecule Example: Benzyl Benzoate

Next, we examine conformational effects due to varying structures with a medium-sized molecule from the VT-2005 Sigma Profile Database. Medium-sized molecules generally contain more single bonds and allow more rotations, and, thus, more conformations. We identify four low energy states for benzyl benzoate (VT-0676) and predict pressure-composition behavior using these conformations. Figure 4.18 shows the released conformation from the VT-2005 Sigma Profile Database and three other low energy conformations and their respective sigma profiles. As with the previous example, conformations A, B, and C are varied from the released

conformation and undergo a geometry optimization, following the procedure in Chapter 3. The benzyl benzoate sigma profiles exhibit larger variations, based on standard deviation and variance calculations, when compared to 2-methoxy-ethanol sigma profiles. This suggests that conformational differences in larger molecules create larger variations in sigma profiles than smaller molecules.

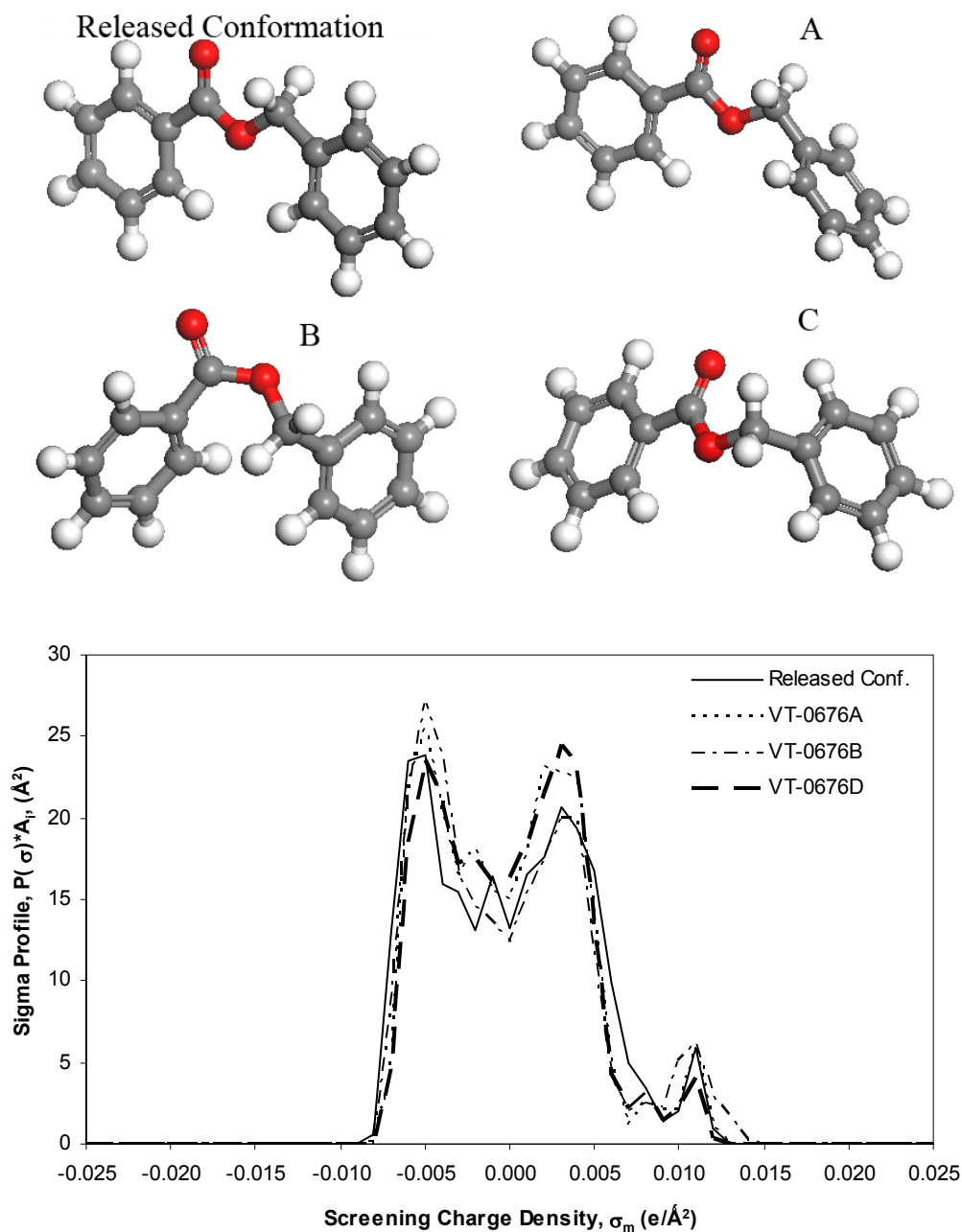


Figure 4.18: Four conformations of benzyl benzoate and their respective sigma-profile for comparison.

We compare predictions for binary systems of benzyl benzoate and five solvents: benzene (VT-0242), toluene (VT-0243), benzaldehyde (VT-0432), benzyl alcohol (VT-0541), and phenol (VT-0542). Figure 4.19 through Figure 4.23 show COSMO-SAC predictions for those systems in comparison with experimental data.

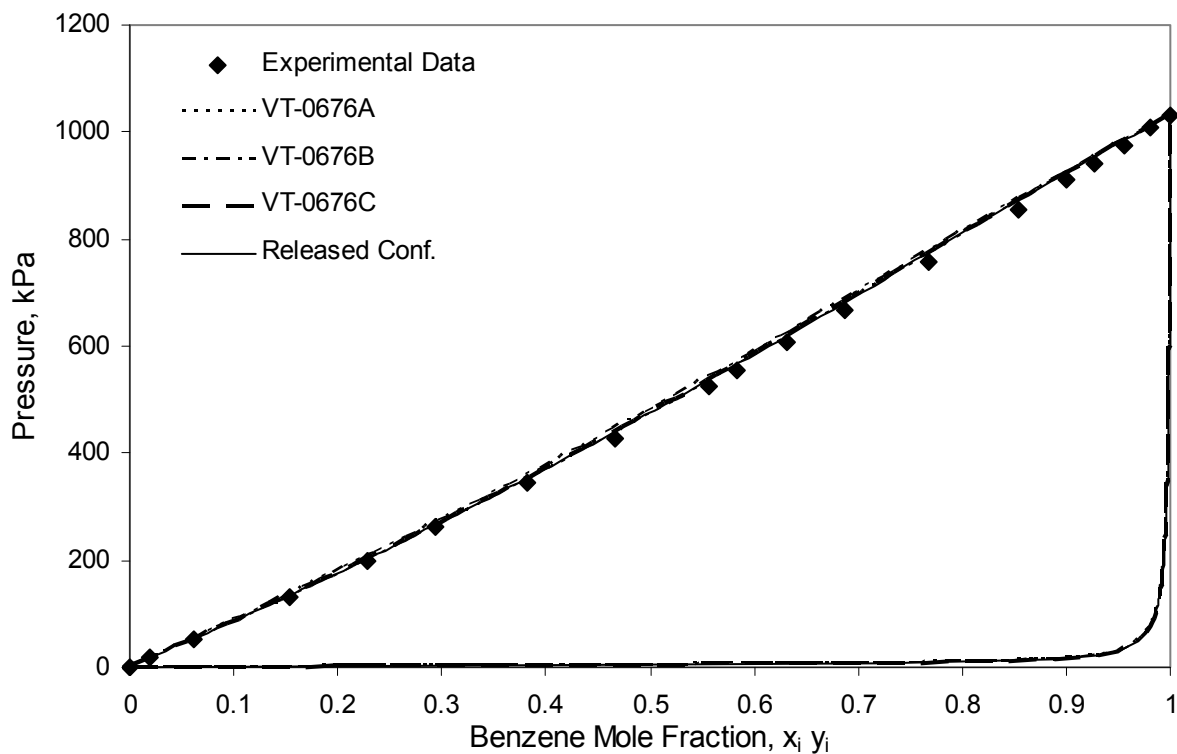


Figure 4.19: Pressure-composition data and COSMO-SAC predictions for (1) benzene (2) benzyl benzoate at 453.23 K.⁵⁶ The vapor phase is a prediction only.

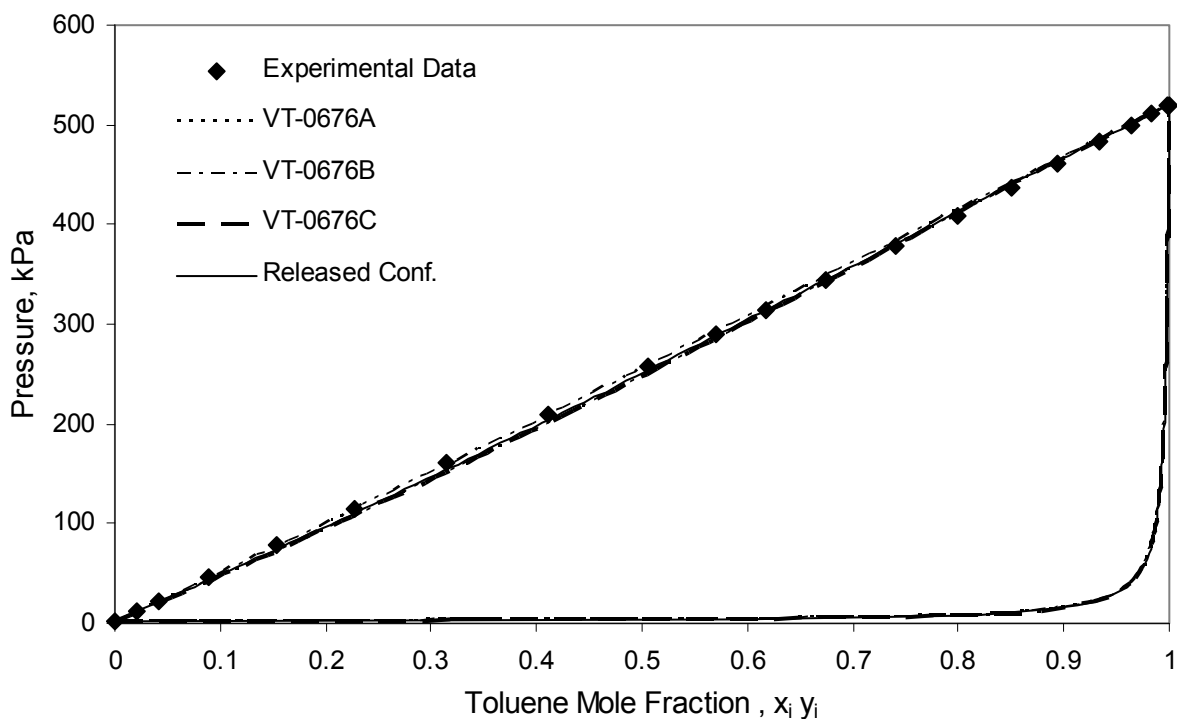


Figure 4.20: Pressure-composition data and COSMO-SAC predictions for (1) toluene (2) benzyl benzoate at 453.25 K.⁵⁶ The vapor phase is a prediction only.

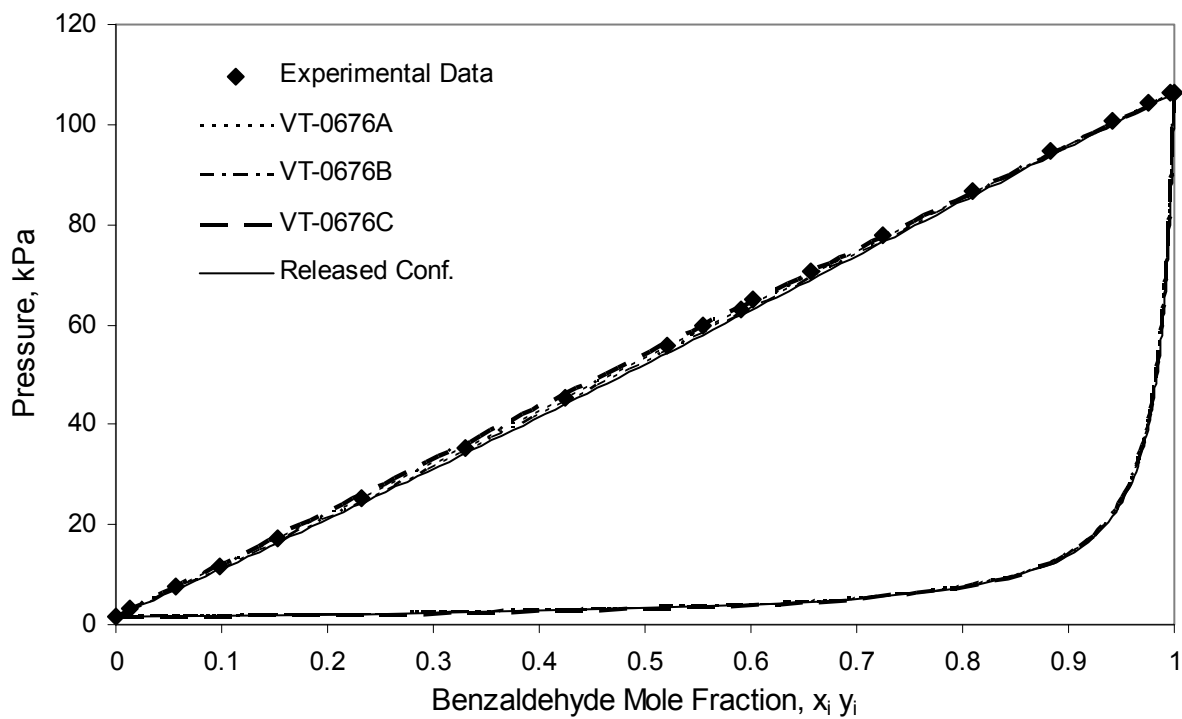


Figure 4.21: Pressure-composition data and COSMO-SAC predictions for (1) benzaldehyde (2) benzyl benzoate at 453.25 K.⁵⁶ The vapor phase is a prediction only.

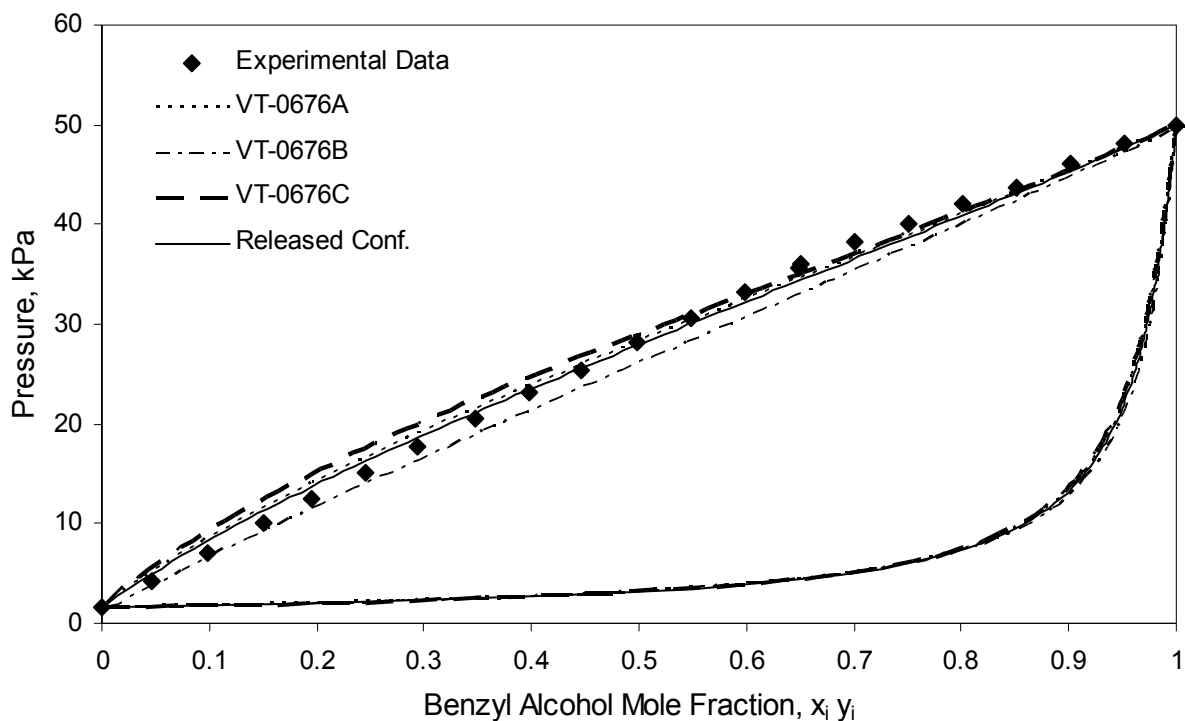


Figure 4.22: Pressure-composition data and COSMO-SAC predictions for (1) benzyl alcohol (2) benzyl benzoate at 453.23 K.⁵⁶ The vapor phase is a prediction only.

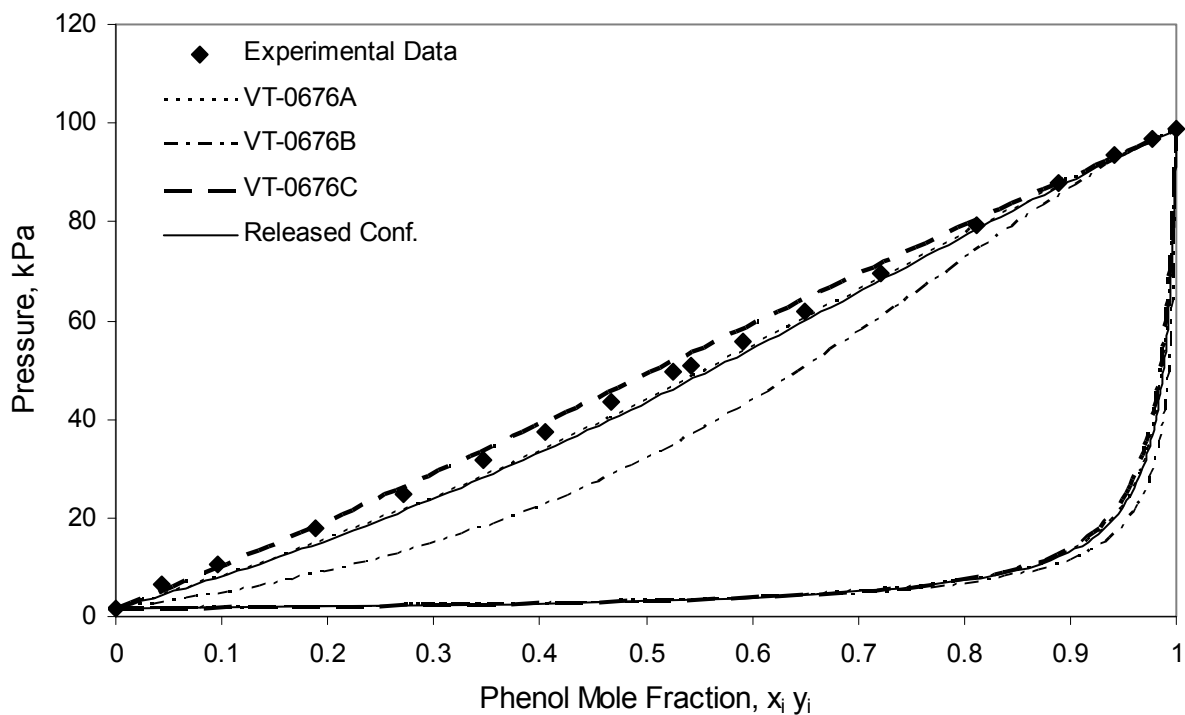


Figure 4.23: Pressure-composition data and COSMO-SAC predictions for (1) phenol (2) benzyl benzoate at 453.26 K.⁵⁶ The vapor phase is a prediction only.

All of these binary systems exhibit ideal or near-ideal behavior. With a few exceptions, the COSMO-SAC model predictions match ideal or near-ideal behavior, to varying degrees of accuracy. Table 4.4 summarizes the absolute average relative percent error (AA%E) calculation using equation (4.1) for the solvent vapor pressure, kPa, for each conformation of benzyl benzoate with the five solvents studied. There is a small difference in AA%E values concerning each prediction for conformations A and C, and the released conformation. Conformation B is only marginally worse, with the exception of the prediction for the (1) phenol (2) benzyl benzoate system, which is significantly different from the published values and the other predictions. We observe the largest deviation of the COSMO-SAC model predictions from experimental data for the (1) phenol (2) benzyl benzoate system for all conformations. Overall, conformation C predicts the pressure-composition data with the most accuracy.

Table 4.4: AA%E of solvent vapor pressure, kPa, for each benzyl benzoate system.

Solvent	Temperature, K	VT-0676A	VT-0676B	VT-0676C	Released Conf.
Benzene	453.23	1.44	2.80	1.99	1.81
Toluene	453.25	3.33	1.21	3.19	2.29
Phenol	453.26	8.59	28.20	4.69	9.08
Benzaldehyde	453.25	1.71	2.91	1.30	3.42
Benzyl Alcohol	453.23	6.07	4.63	9.02	5.33
Average Error, kPa		4.229	7.950	4.041	4.387

4.5 Conformational Improvements of Select Released Compounds

We present alternate conformations for compounds which satisfy the validation criterion based on pure component vapor pressure in hopes of further improvement of property predictions. This section covers the effects of conformational variations on vapor-liquid equilibrium behavior of binary systems. In Chapter 5, we discuss the effects of solvent conformational variations on solubility predictions for the solvents below. We compare each conformation's relative atomic positions, sigma profiles, and COSMO-SAC pressure-composition predictions to their respective literature values.

4.5.1 Polar Solvent: Ethanol

One common solvent in the VT-2005 Sigma Profile Database is ethanol. The released conformation generates a sigma profile which predicts an $\ln(P_{vap})$ of 11.9971 with the revised COSMO-SAC-BP model.⁵¹ We know from prior experience that alcohols and diols may form unlikely bond angles regarding the hydroxyl group placement. After visual inspection of the released optimized structure, we improved the conformation using Amber8 as a pre-optimization tool following the procedure laid out in Chapter 3. Figure 4.24 illustrates the differences in the released and improved conformations together with their sigma profiles.

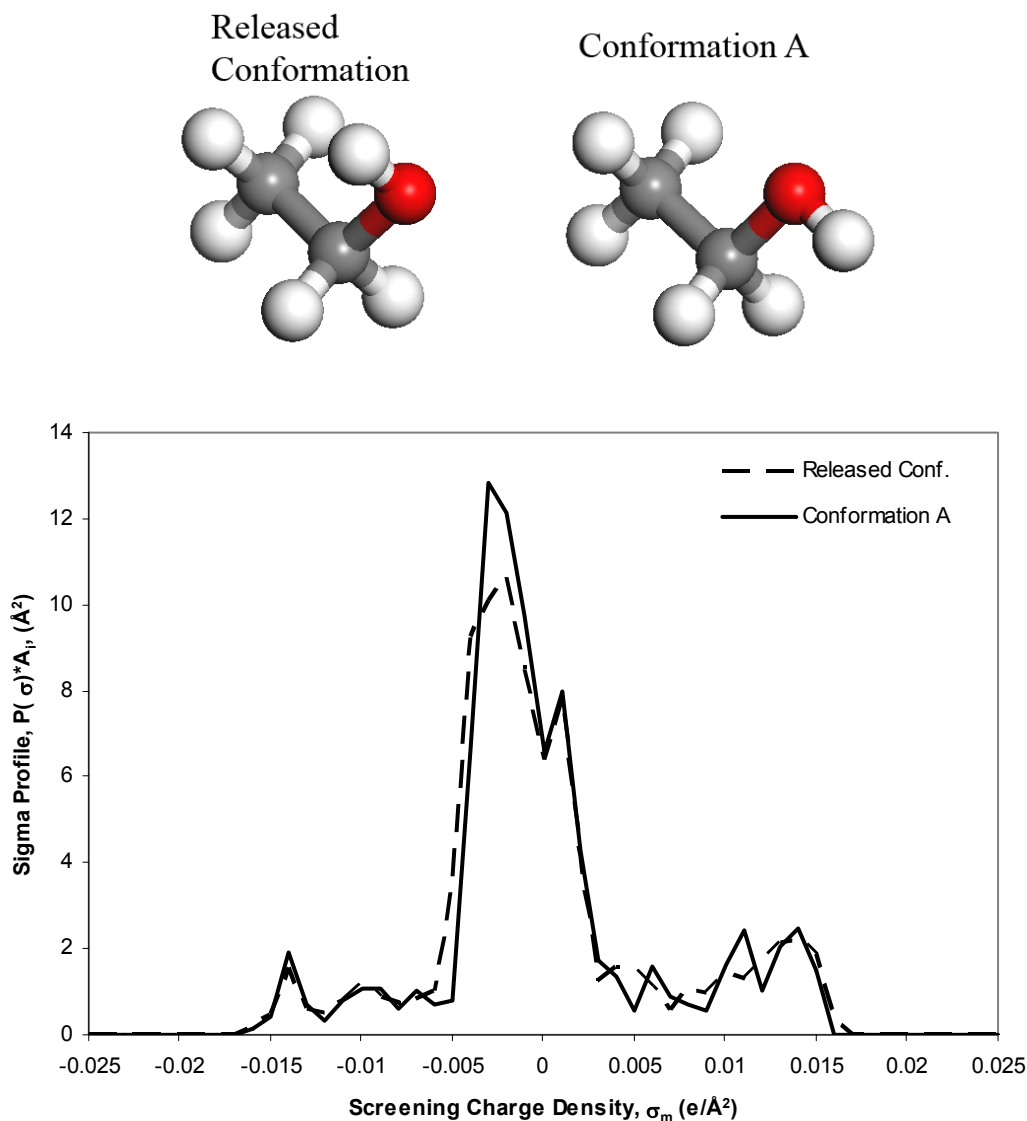


Figure 4.24: Variations in the released conformation of Ethanol (VT-0478) and an improved conformation and the effects on their respective sigma profiles.

The difference in the two conformations stems from a 120° rotation of the C – O bond but results in a 28% relative difference in peak height calculated with equation (4.1) over the relevant range of screening charge density. The released conformation has a lower condensed phase energy than conformation A, from the DMol3 calculation, (Conformation A: $-155.0867 E_h$, Released Conformation: $-155.0905 E_h$); however, pressure-composition predictions do improve with the new conformation for approximately half of the systems studied, including all of the aqueous systems. We use experimentally determined pure component vapor-pressures from literature data as reference points for use with the predicted COSMO-SAC activity coefficients or Antoine equation predictions when experimental data are not available. We show pressure-composition data and COSMO-SAC predictions for several example systems of ethanol with water, acetone, tetrachloromethane, benzene, and octane in Figure 4.25 through Figure 4.30. We use the aqueous ethanol system to demonstrate COSMO-SAC prediction accuracy dependence on temperature.

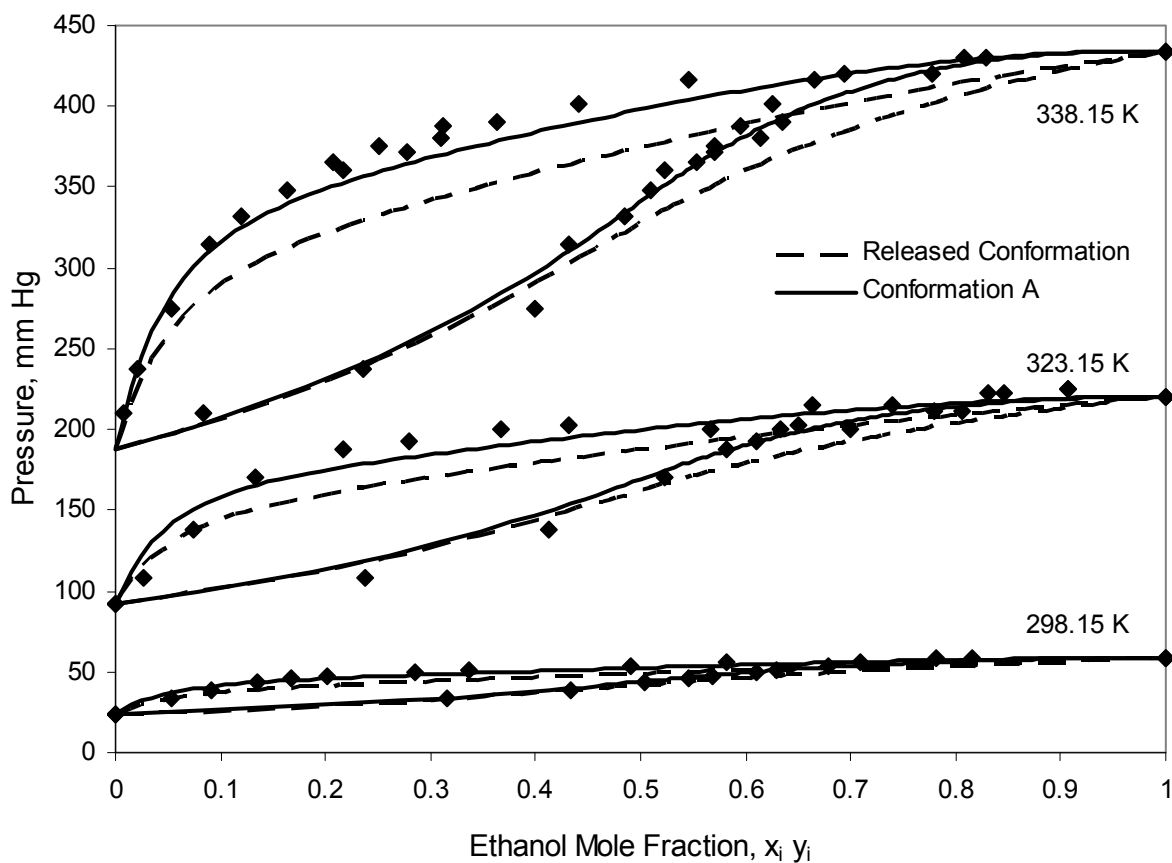


Figure 4.25: Pressure-composition curve of (1) ethanol (2) water predicted by the COSMO-SAC model for the released conformation and conformation A at 298.15 K, 323.15 K, and 338.15 K in comparison to literature data.⁵⁷

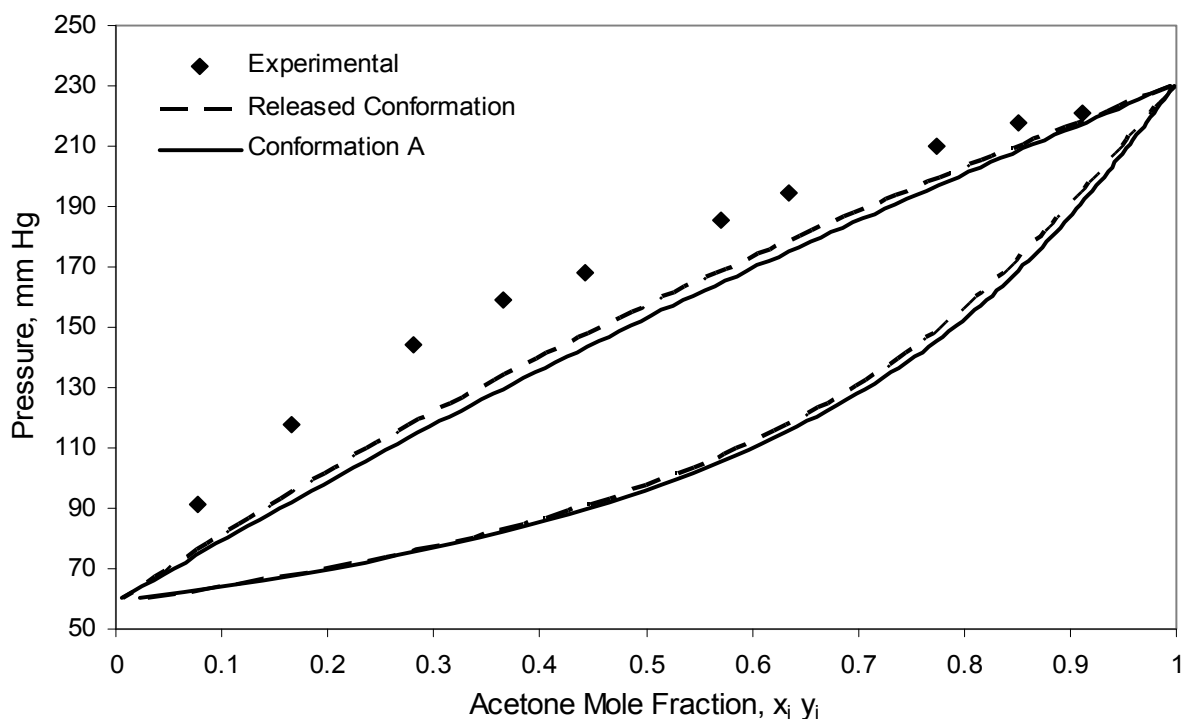


Figure 4.26: Pressure-composition curve for an (1) ethanol (2) acetone mixture at 298.15 K as predicted by the COSMO-SAC model for the released conformation and conformation A with a comparison to literature data.⁵⁷ The vapor phase is a prediction only.

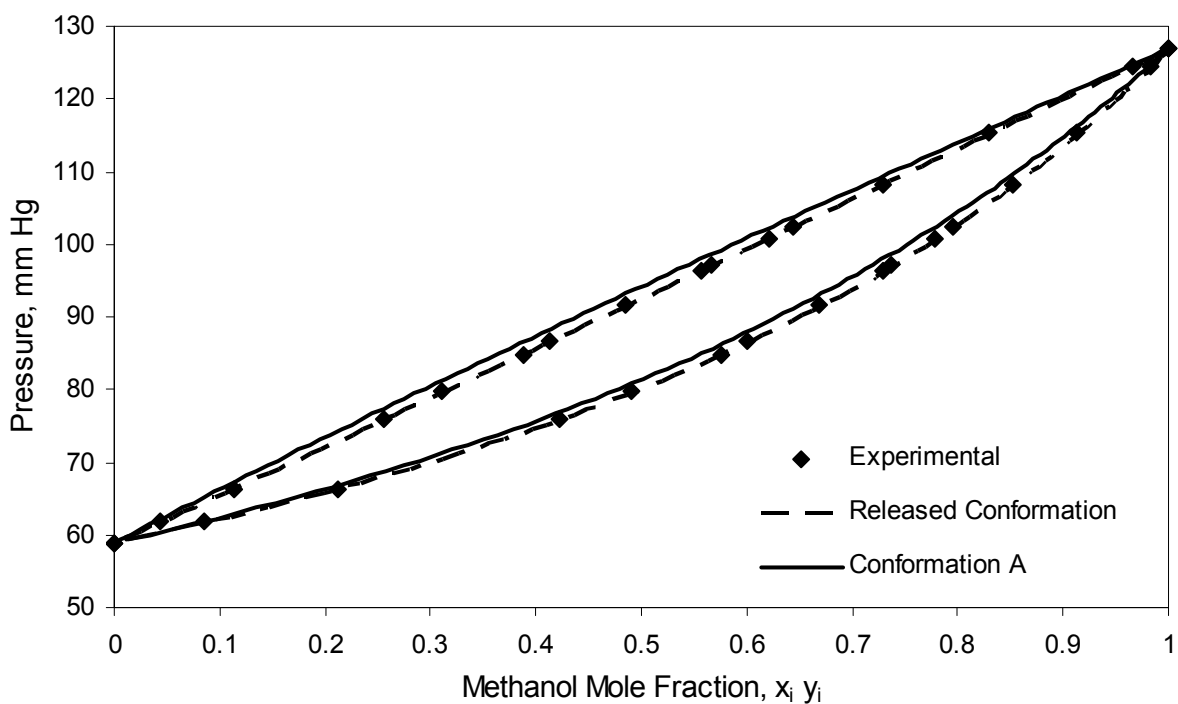


Figure 4.27: Pressure-composition curve of (1) methanol (2) ethanol predicted by the COSMO-SAC model for the released conformation and conformation A at 298.15 K in comparison to literature data.⁵⁷

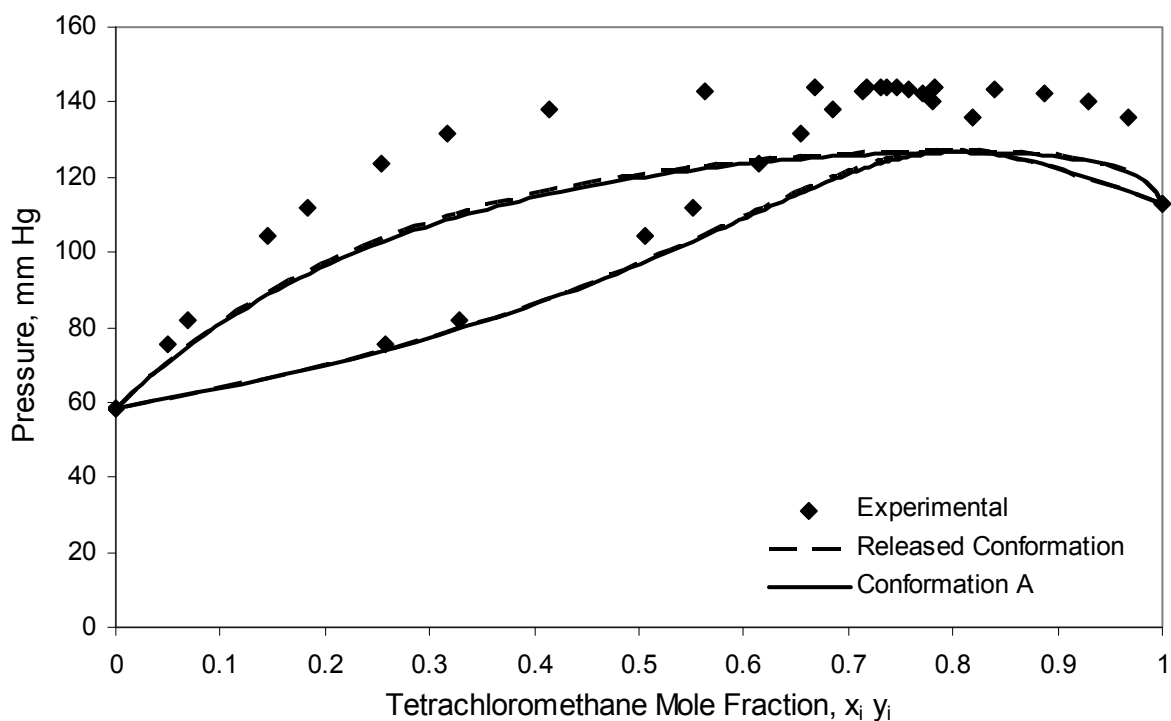


Figure 4.28: Pressure-composition curve of (1) ethanol (2) tetrachloromethane predicted by the COSMO-SAC model for the released conformation and conformation A at 298.15 K in comparison to literature data.⁵⁷

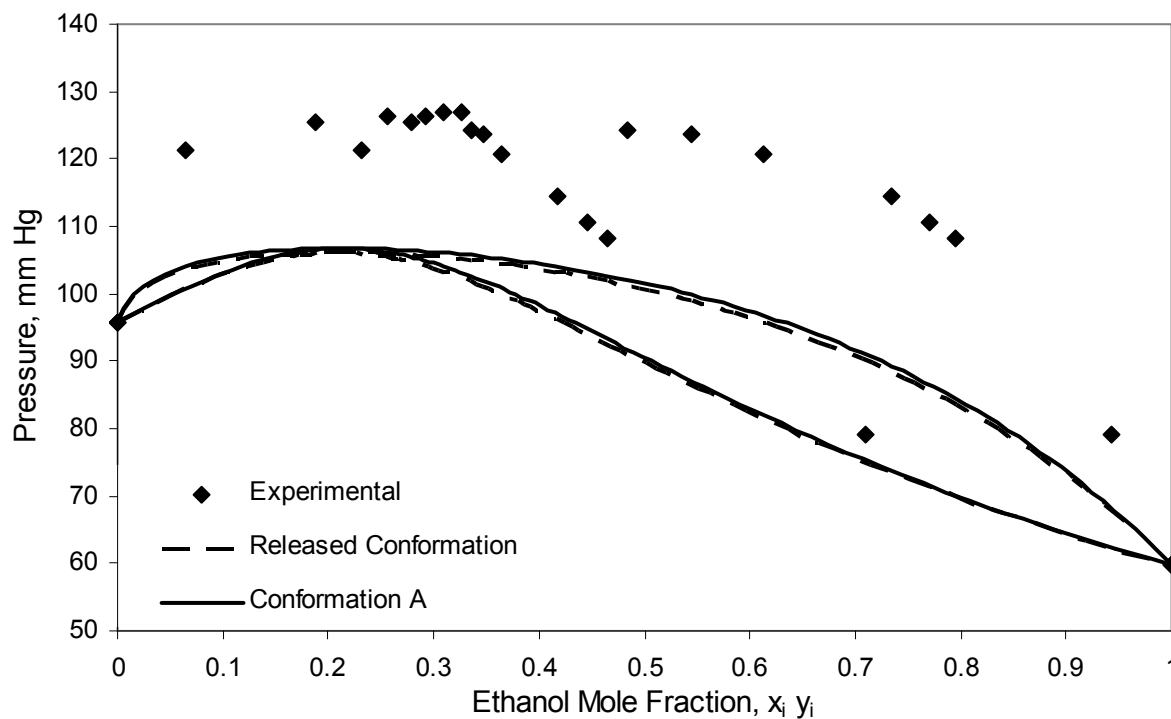


Figure 4.29: Pressure-composition curve of (1) ethanol (2) benzene predicted by the COSMO-SAC model for the released conformation and conformation A at 298.15 K in comparison to literature data.⁵⁷

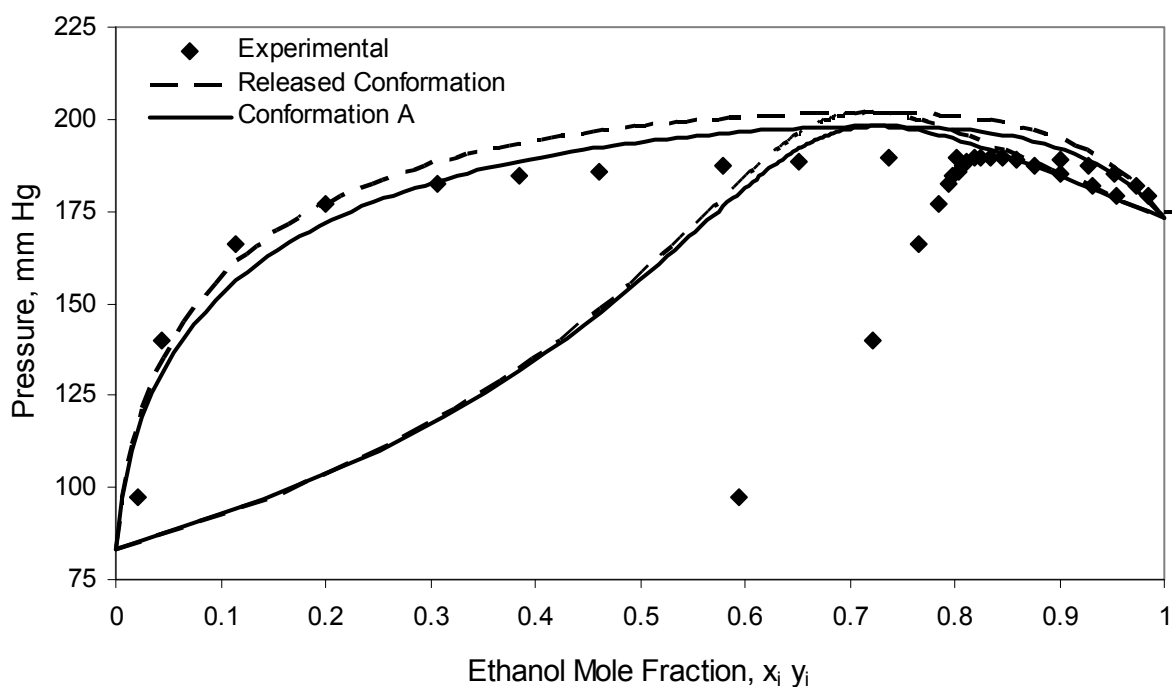


Figure 4.30: Pressure-composition curve of (1) ethanol (2) octane predicted by the COSMO-SAC model for the released conformation and conformation A at 318.15 K in comparison to literature data.⁵⁷

The COSMO-SAC model predicts azeotropes for systems with ethanol and benzene, octane, tetrachloromethane, and water. However, it only predicts azeotropes for water / ethanol systems at the higher temperatures studied. The predicted azeotropic compositions for these systems are less accurate than those predicted in Section 4.4.1. Here, the predicted azeotropic compositions vary by 30, 22, and 12 % mole fraction ethanol for the benzene, tetrachloromethane, and octane systems respectively. We also note that the predicted azeotropic compositions for each conformation are only slightly different from each other. Table 4.5 summarizes AA%E in mm Hg for the two conformations in comparison to literature data. In most cases where we consider multiple temperatures, we see that predictions improve with increasing temperature. We believe this case behaves in this manner because the resulting pressure is high enough that we cannot model the nonidealities properly with a modified Raoult's Law, approximately 6 atm.⁵⁷ The aqueous ethanol systems exhibit the most improvement with using conformation A, whereas predictions for the other systems are relatively similar between the two conformations.

Table 4.5: AA%E of system pressure, mm Hg, for multiple systems with various solvents at various temperatures for both conformations of ethanol.

Solvent	T, K	Released Conformation	Conformation A
		AA%E, mmHg	AA%E, mmHg
Acetone	288.15	16.06	18.09
Acetone	298.15	10.62	12.70
Acetone	308.15	8.12	10.17
Acetone	318.15	6.87	8.86
Acetone	328.15	4.81	6.77
Acetone	372.7	3.42	5.11
Benzene	298.15	18.62	17.96
Benzene	313.15	14.01	13.54
Benzene	318.15	15.73	15.21
Methanol	298.15	0.05	1.35
Methanol	313.15	0.04	1.19
Methanol	373.15	1.21	0.62
Methanol	393.15	0.35	0.80
Octane	313.15	14.26	16.34
Octane	318.15	4.68	3.83
Octane	338.15	6.83	6.72
Tetrachloromethane	298.15	12.43	13.00
Water	298.15	8.74	4.02
Water	323.15	7.45	4.72
Water	338.15	8.42	2.31
Water	343.15	8.11	2.69
Water	348.15	7.67	2.05

4.5.2 Nonpolar Solvent: Cyclohexane

For this example, we study the differences in property predictions between the “boat” and “chair” conformations of cyclohexane (VT-0099) by comparing their conformations, sigma profiles, condensed phase energy, and predicted VLE. The released cyclohexane conformation is in the “boat” configuration but predicts a pure component vapor-pressure prediction with a revised COSMO-SAC-BP model⁵¹ within the validation and release criterion specified in Section 4.2.1 with $\ln P_{\text{vap}}$ equal to 12.0364. We know that the “chair” conformation is the lower energy state, and our condensed phase energy calculations reflect that fact (Released Conformation “Boat”: $-235.9094 E_h$, Conformation A “Chair”: $-235.9136 E_h$). Figure 4.31 illustrates the difference between both conformations and their sigma profiles. We calculate the difference in

the sigma profiles as 23 % using equation (4.1) over the relevant range; however, the difference in the main peak is only 3.7 %.

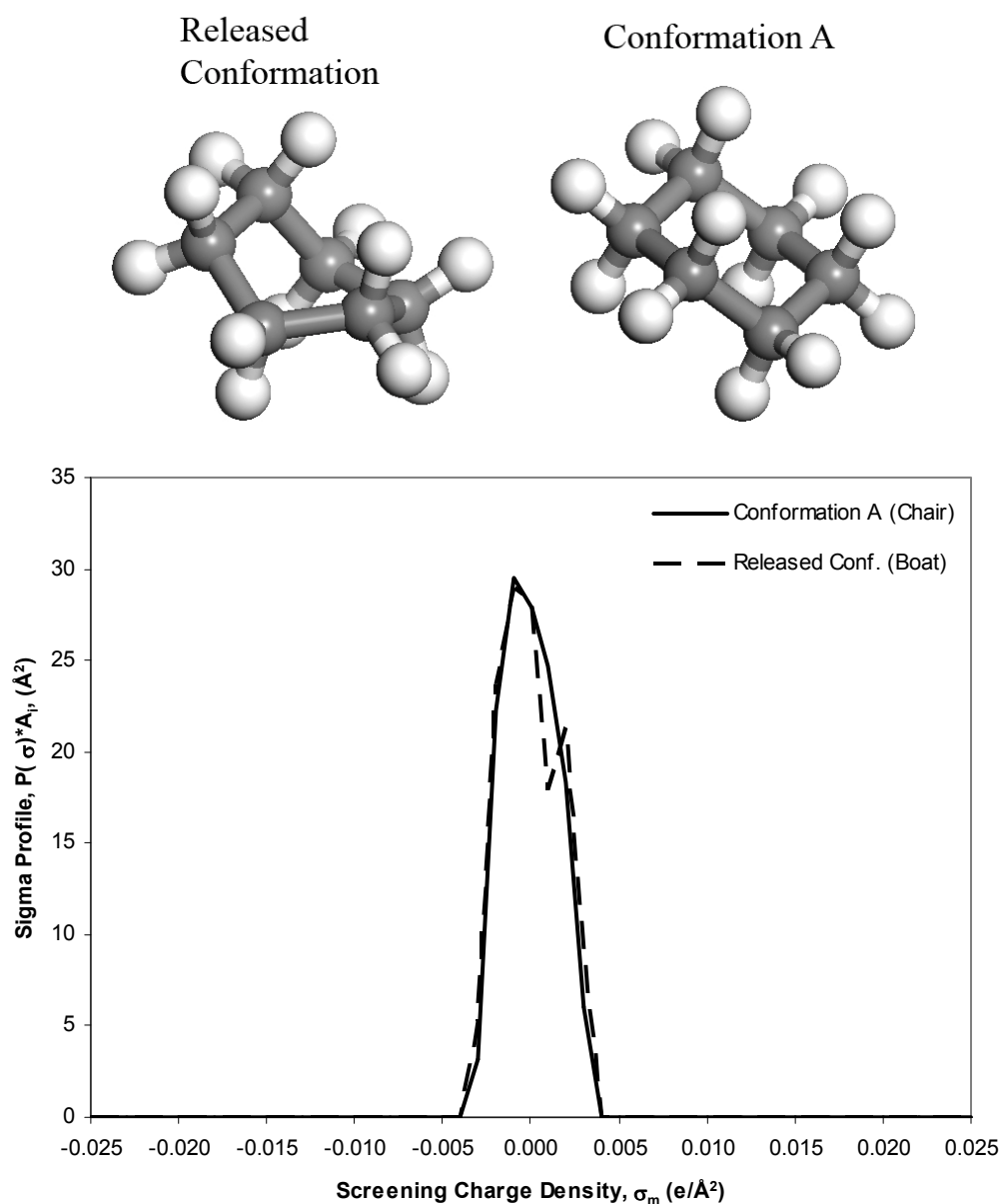


Figure 4.31: Variations in the released conformation of “boat” cyclohexane (VT-0099) and the “chair” conformation and the effects on their respective sigma profiles.

Figure 4.32 through Figure 4.37 illustrate the differences in VLE predictions for the two cyclohexane conformations for several representative examples. The second components include a representative alcohol, alkane, aromatic ring, ether, halogenated alkane, and nitrogen-containing compound.

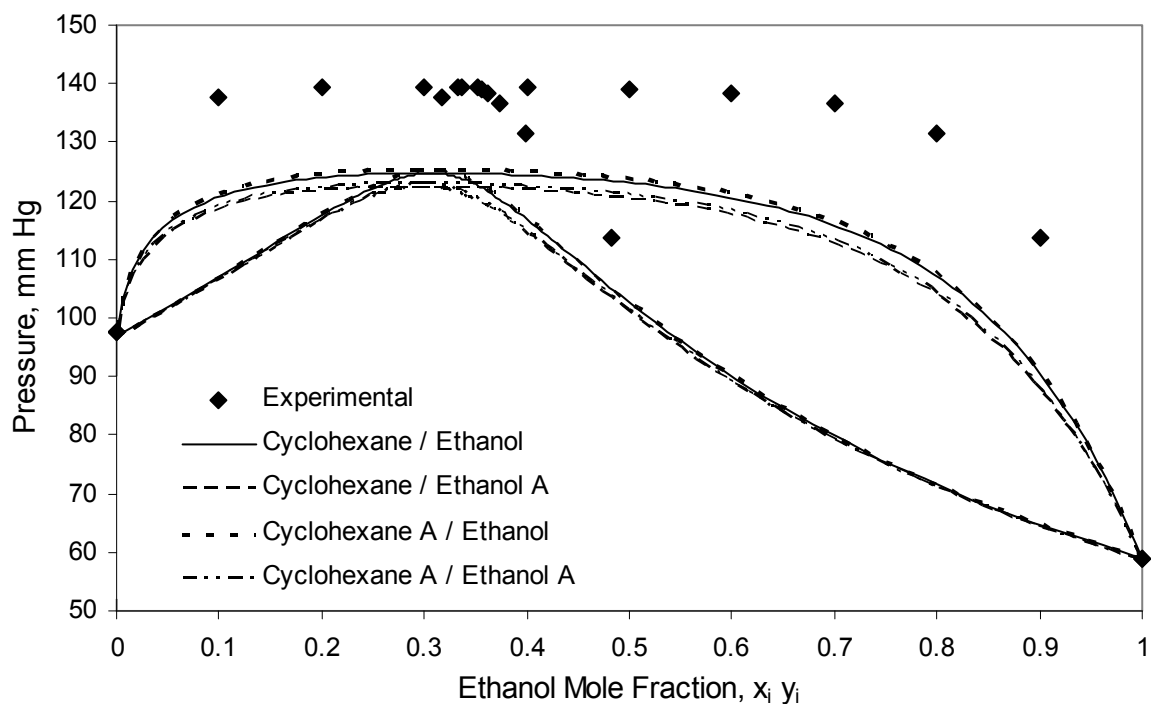


Figure 4.32: COSMO-SAC predicted pressure for (1) ethanol (2) cyclohexane at 298.15 K. Cyclohexane, cyclohexane A, ethanol, and ethanol A refer to the released conformation and conformation A for their respective species.⁵⁷

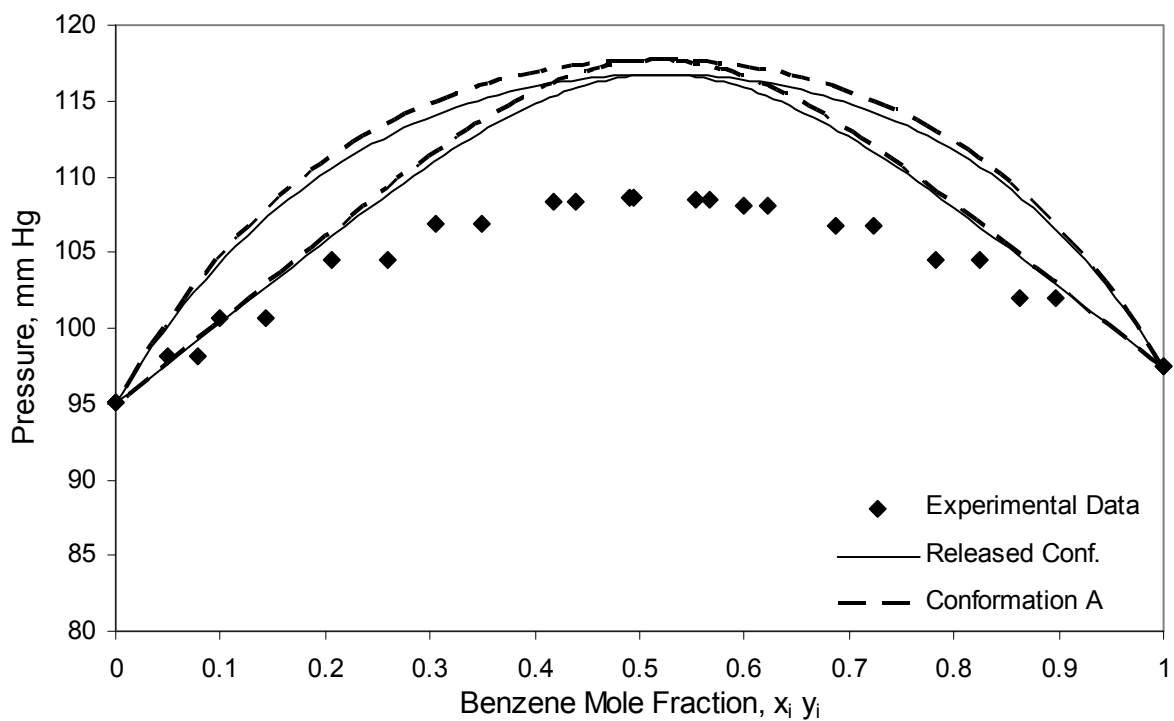


Figure 4.33: COSMO-SAC predicted pressure two cyclohexane conformations for (1) benzene (2) cyclohexane at 298.15 K.⁵⁷

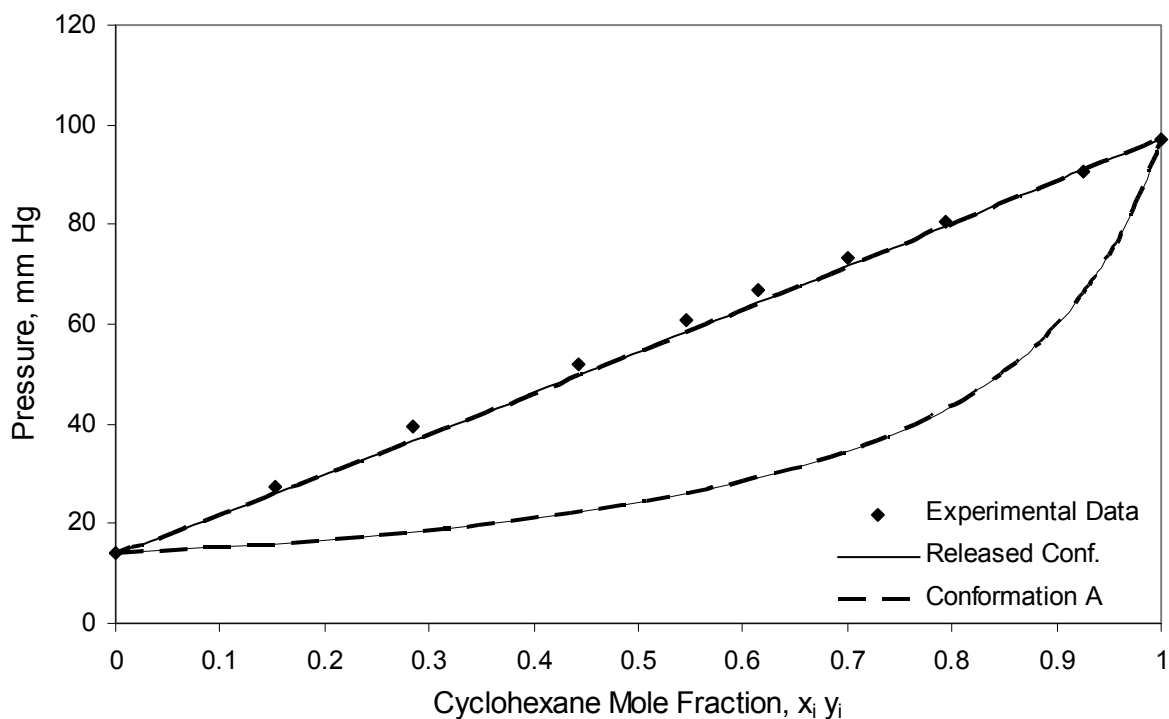


Figure 4.34: COSMO-SAC predicted pressure two cyclohexane conformations for (1) octane (2) cyclohexane at 298.15 K.⁵⁷ The vapor phase is a prediction only.

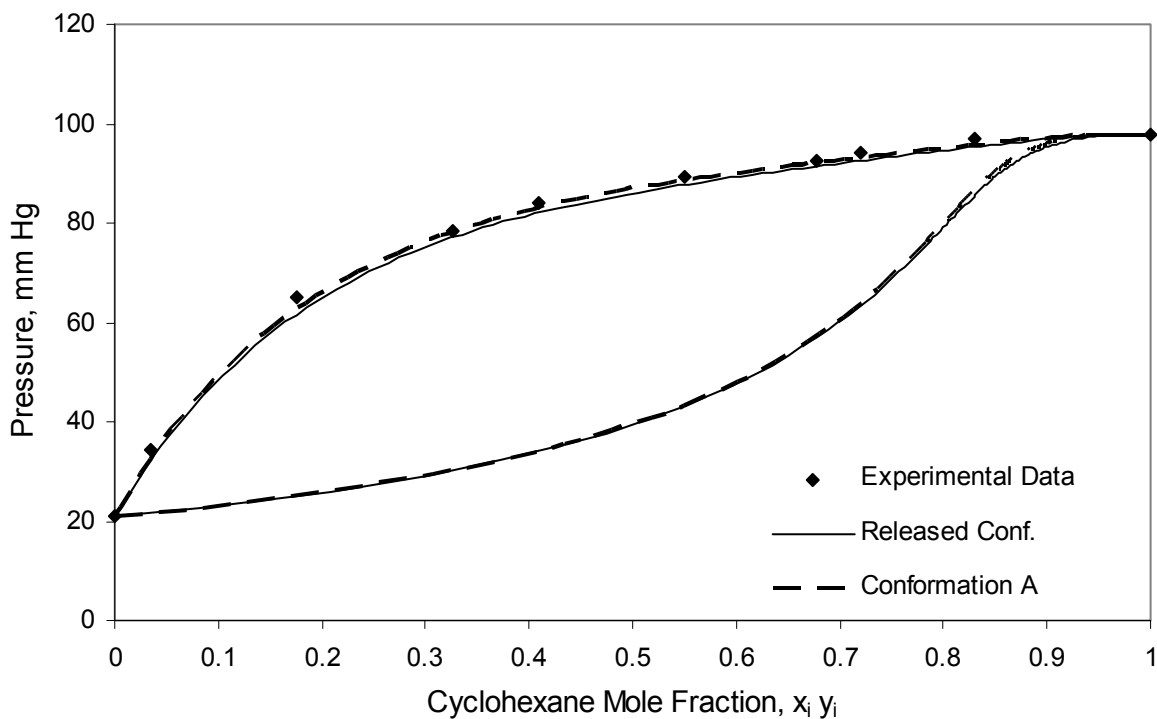


Figure 4.35: COSMO-SAC predicted pressure two cyclohexane conformations for (1) pyridine (2) cyclohexane at 298.15 K.⁵⁷ The vapor phase is a prediction only.

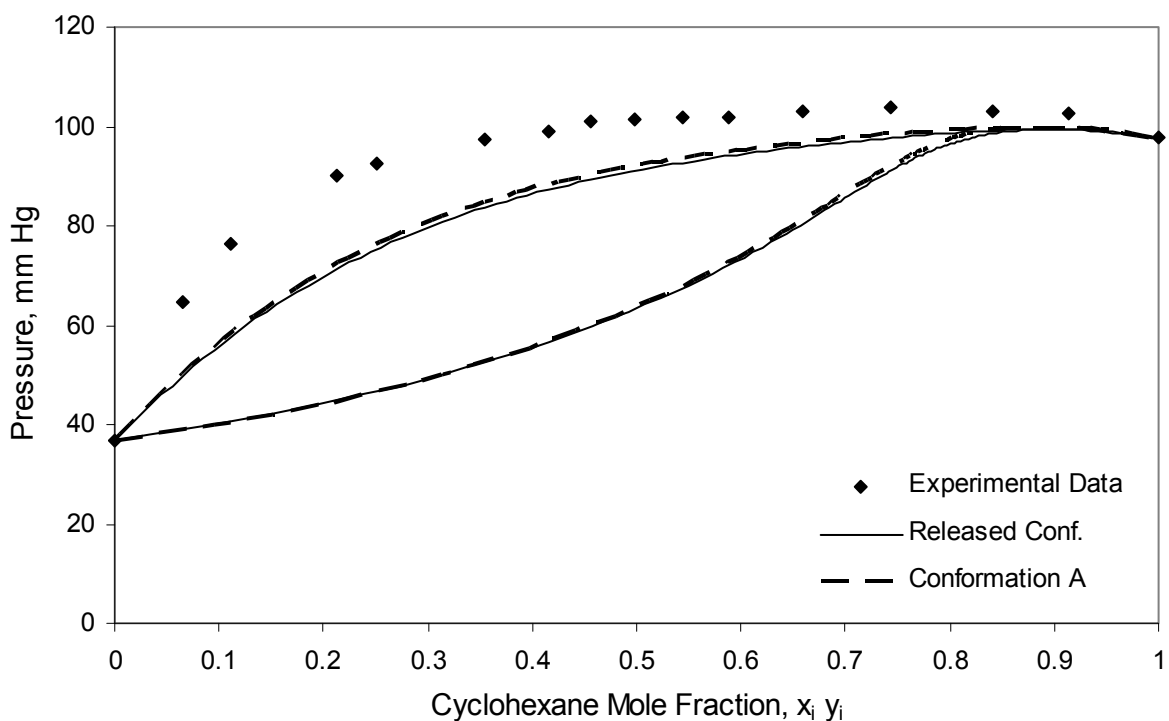


Figure 4.36: COSMO-SAC predicted pressure two cyclohexane conformations for (1) 1,4-dioxane (2) cyclohexane at 298.15 K.⁵⁷ The vapor phase is a prediction only.

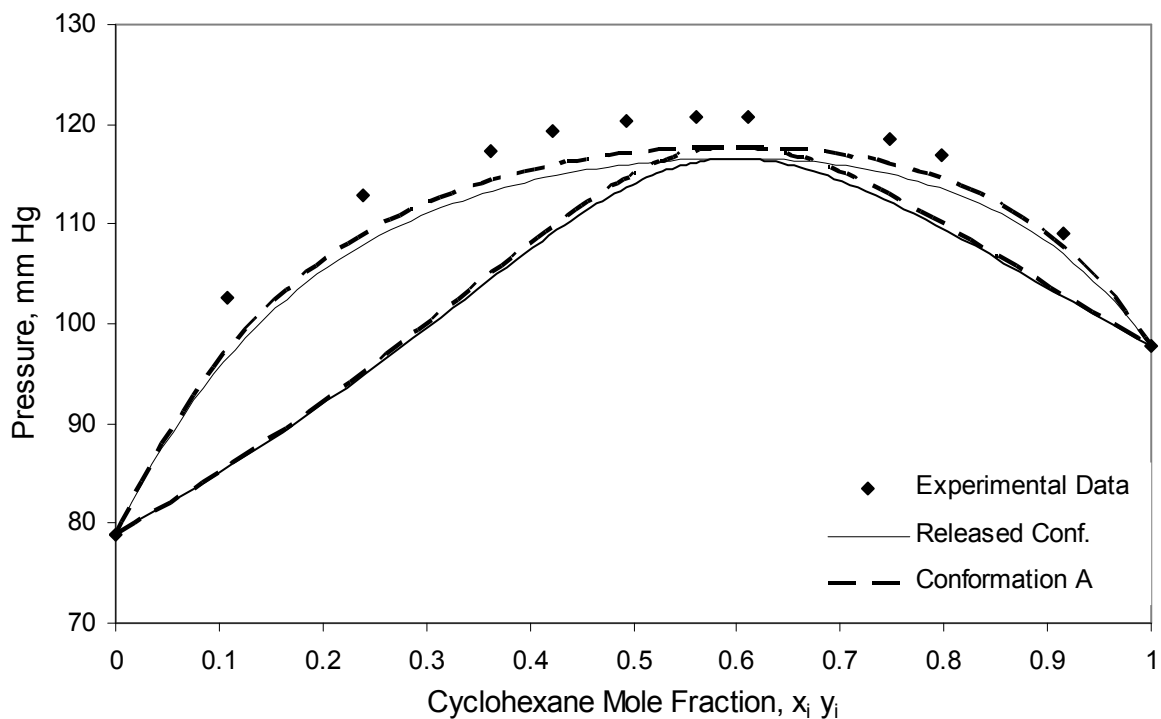


Figure 4.37: COSMO-SAC predicted pressure two cyclohexane conformations for (1) 1,2-dichloroethane (2) cyclohexane at 298.15 K.⁵⁷ The vapor phase is a prediction only.

Table 4.6: AA%E of system pressure, mm Hg, for multiple systems with various solvents at various temperatures for both conformations of cyclohexane.

Solvent	T, K	Released Conformation	Conformation A
		AA%E, mm Hg	AA%E, mm Hg
1,2-Dichloroethane	286.15	3.87	3.02
1,2-Dichloroethane	293.15	3.47	2.66
1,2-Dichloroethane	298.15	3.60	2.71
1,2-Dichloroethane	305.53	2.64	1.89
1,4-Dioxane	288.15	5.59	4.51
1,4-Dioxane	293.15	15.01	14.09
1,4-Dioxane	298.15	12.24	11.31
1,4-Dioxane	313.15	9.69	8.82
Benzene	283.15	6.46	7.30
Benzene	298.15	5.94	6.66
Benzene	313.15	5.48	6.17
Benzene	333.15	4.62	5.19
Benzene	343.15	4.94	5.51
Ethanol	278.15	13.24	12.68
Ethanol	293.15	12.95	12.45
Ethanol	298.15	14.03	13.59
Ethanol	308.15	13.02	12.56
Ethanol	323.15	12.28	11.85
Ethanol	338.15	11.84	11.44
Ethanol-A	278.15	15.12	14.64
Ethanol-A	293.15	14.71	14.27
Ethanol-A	298.15	15.65	15.26
Ethanol-A	308.15	14.66	14.26
Ethanol-A	323.15	13.82	13.45
Ethanol-A	338.15	13.29	12.94
Octane	298.15	3.57	3.72
Octane	308.15	2.93	3.09
Octane	318.15	2.66	2.82
Octane	328.15	2.21	2.37
Pyridine	298.15	2.69	1.66
Pyridine	303.15	2.38	1.38
Pyridine	318.15	2.58	1.68
Pyridine	338.15	0.86	0.68

We study six binary systems with cyclohexane, each at multiple temperatures. Table 4.6 summarizes the AA%E, mm Hg, calculated by equation (4.1) for each system. We use literature values for pure component vapor pressure when available and Antoine predictions otherwise. We observe a small increase in accuracy for most (72%) systems with the exception of the

octane and benzene systems. We also see a similar trend from the ethanol case study where the accuracy of the predicted total system pressure improves with increasing temperature. We note that the cyclohexane/ethanol system is most accurately described by the prediction with the released ethanol sigma profile and conformation A cyclohexane sigma profile. The released sigma profiles for ethanol and cyclohexane generate the second best overall prediction over all temperatures studied.

4.6 VLE Prediction for Multicomponent Systems

After examining binary system VLE predictions, we now use the COSMO-SAC model to predict the vapor-liquid equilibrium behavior of fifteen multicomponent systems. We study ternary VLE predictions for several representative systems, including aqueous/alcohol, alcohol/aromatic/halogenated alkane, aromatic alkane/alkane, ether/halogenated alkane, nitrogen-containing compound/aromatic/alkane, etc. We show the average absolute relative error of the predicted total pressure P and the vapor fractions of components 1 and 2 in relation to their literature values in Table 4.7. When we compare systems with similar components to those systems in Table 4.6 and Table 4.5, we observe higher error values than systems with only two components. From the selected systems, we see that COSMO-SAC model accuracy decreases for ternary system VLE predictions relative to activity-coefficient predictions for the binary system propagates as the number of components considered increases. COSMO-SAC most accurately predicts the system of ethanol, acetonitrile, and benzene and of cyclohexane, benzene, and aniline. COSMO-SAC also poorly predicts the total pressure of the acetonitrile/benzene/tetrachloromethane system. Finally, the COSMO-SAC model does not predict ternary VLE behavior as well as the NRTL or UNIQUAC model for these systems.⁵⁷

Table 4.7: AA%E of predicted pressure and vapor composition compared with literature values.⁵⁷ We use both conformations for cyclohexane and ethanol from Section 4.5 for applicable systems.

System	Component 1	Component 2	Component 3	T, K	AA%E P, mm Hg	AA%E y ₁ , m.f.	AA%E y ₂ , m.f.
1	Acetone	Chloroform	Ethanol	328.15	26.1	25.9	39.4
2	Acetone	Chloroform	Ethanol-A	328.15	26.6	22.1	36.3
3	Ethanol	Benzene	Tetrachloromethane	323.15	40.1		
4	Ethanol-A	Benzene	Tetrachloromethane	323.15	42.9		
5	Ethanol	Acetonitrile	Benzene	318.15	6.7	19.0	7.1
6	Ethanol-A	Acetonitrile	Benzene	318.15	7.5	19.3	6.8
7	Water	Methanol	Ethanol	298.15	11.9	16.3	7.0
8	Water	Methanol	Ethanol-A	298.15	13.2	13.0	9.1
9	Benzene	Ethanol	Water	308.15	13.2	29.7	35.5
10	Benzene	Ethanol-A	Water	273.15	10.1	31.5	40.1
11	Hexane	Cyclohexane	Benzene	298.15	19.5	17.2	29.7
12	Hexane	Cyclohexane-A	Benzene	273.15	22.8	18.3	28.7
13	Cyclohexane	Benzene	Aniline	343.15	7.0	20.9	28.3
14	Cyclohexane-A	Benzene	Aniline	343.15	6.3	21.4	28.8
15	Chloroform	Tetrahydrofuran	1,2-Dichloroethane	323.15	20.1	44.5	58.7
16	Acetonitrile	Benzene	Tetrachloromethane	318.15	82.2	42.0	31.1
17	Benzene	Toluene	Cyclohexane	318.15		18.9	27.1
18	Benzene	Toluene	Cyclohexane-A	318.15		24.6	29.6
19	Dichloromethane	Chloroform	Tetrachloromethane	318.15	16.5	11.2	41.4
20	Pyridine	Benzene	Cyclohexane	333.95		19.8	15.6
21	Pyridine	Benzene	Cyclohexane-A	333.95		19.9	15.7
22	Pyridine	Benzene	Cyclohexane	353.95		17.0	14.5
23	Pyridine	Benzene	Cyclohexane-A	353.95		17.1	14.6
24	Isoprene	2-Butyne	Acetonitrile	274.65		5.3	16.1
25	Isoprene	3-Methyl-1-Butyne	Acetonitrile	273.95		5.3	19.2
26	Isoprene	3-Methyl-1-Butyne	Acetonitrile	285.55		6.5	16.2
27	Isoprene	3-Methyl-1-Butyne	Acetonitrile	297.95		5.7	18.1
28	Thiophene	Benzene	Cyclohexane	333.95	13.4	12.8	15.4
29	Thiophene	Benzene	Cyclohexane-A	333.95	13.5	12.8	15.5

4.7 Application Heuristics for COSMO-Based Methods

The literature identifies molecules and systems for which the COSMO approach is not suited or suggested without proper research. We review many reported experiences from the literature up to November 2006 to provide the reader with a guide for COSMO model applications. Our experience is mainly concerned with COSMO-SAC, and our references to COSMO-RS refer only to cited literature observations. Several groups are researching these methods, and it should broaden its applicability and improve the model's accuracy in the future.

4.7.1 Applicability of General COSMO-Based Methods

One should use the COSMO-SAC and COSMO-RS models to predict binary and multicomponent vapor-liquid equilibria (VLE), and solid-liquid equilibrium (SLE). The COSMO-RS and COSMO-SAC methods are generally applicable and show promise in predicting equilibrium behavior.^{4-6,14,58,59} Factors such as conformational isomerism of all components present, which affect the final sigma profile, model parameters, and the exchange energy expression are major contributing factors to the model's overall accuracy in predicting thermodynamic properties. Theoretically, the only input required comes from the molecular structure; however, in the case of VLE predicted pressure, the predicted activity coefficients from the COSMO-SAC model are more meaningful when we use literature values for pure component vapor pressure, although it is possible to generate a component's vapor pressure using the COSMO-SAC-BP model. The COSMO-RS model^{1,3,15} and the COSMO-SAC-BP model,^{51,60} which is a variation of the COSMO-SAC model developed by Lin and Sandler⁶, predict pure component vapor pressures and enthalpies of vaporization.

4.7.2 Sensitivity Effects of Sigma Profiles

COSMO-SAC and COSMO-RS predictions are sensitive to variations in sigma profiles, which stems from conformational freedom and, therefore, one should use a consistent procedure or algorithm when generating sigma profiles. We find some systematic differences, in the energy calculations, between DFT calculations in DMol3 and Jaguar⁶¹ that could lead to different sigma profiles. However, the extent of error introduced by these variations appears small in comparison to the overall model error.⁹ Because certain molecules may exist in multiple stable conformations, one should carefully analyze the system to determine the effects of conformational variations on thermodynamic property predictions.

4.7.3 COSMO-Based Methods and Nitrogen-Containing Compounds

One should exercise caution when using COSMO-based models with amines. COSMO calculations generally show poor results for amines, especially trialkylamines, because of the unpaired electron in the cavity.⁶ Klamt and co-workers^{1,2,15} reported problems in applying

COSMO calculations to highly polar groups with a small surface area, such as amines. We discuss nitrogen-containing compound solubility in further detail in Section 5.6.3.

4.7.4 Polymer Applications

One should exercise caution when using COSMO-based models with polymers systems. COSMO-based models are not well proven for polymer-system thermodynamics (Panayiotou⁷ has commented on “...its inability to properly account for the thermodynamics of polymer systems except for some rather limited cases”). Klamt⁴⁹ reports successful modeling of gas solubility in polymers by creating partial sigma profiles based on the repeating unit. We note that COSMO-RS models predict solubilities well but with the use of a fitted polymer-specific correction constant.⁴⁹ Gaglione⁶² reports successful modeling the system of water, caprolactam, and nylon-6 systems using a revised COSMO-SAC model, which also uses a modified sigma profile analogous to a polymer repeat unit, with comparable accuracy to the POLYNRTL model.

4.7.5 Model Accuracy Regarding Chloroform Mixtures with Ketones or Alcohols

One should avoid COSMO-SAC with mixtures of chloroform and ketones or alcohols. Lin and Sandler⁶ report difficulties in accurately predicting interactions between chloroform and either ketones or alcohols. They suggest that “it is necessary to refine the COSMO calculation to provide a better σ -profile for chloroform.” This is identical to the behavior that Klamt and co-workers^{1,2} report, regarding largely polar groups that are similar to nitriles and carbonyl groups.

4.7.6 Selection of Appropriate Model Parameters

One should review which COSMO approach is best for the chemical system. For example, COSMO-RS provides better results for alkanes in water in comparison to COSMO-SAC.⁶ Klamt and others optimize the COSMO-RS model parameters with water/alkane solutions, whereas Lin and Sandler simply use the values from COSMO-RS without an independent optimization for use with COSMO-SAC.^{1,5,6}

4.7.7 Limitations of Validated Atomic Radii

One should use COSMO predictions for molecules that contain hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine. Klamt et al^{2,4} optimize the atomic radii and van der Waals coefficients and suggest using an approximation

rule of 117% of the van der Waals radii for predicting the atomic radii of other atoms. Klamt⁴ validates this rule for nine of the ten above stated atoms, excluding phosphorus. Klamt⁶³ currently uses atomic radii for hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, iodine, and silicon.

4.7.8 Applicability of UNIFAC and COSMO-Based Methods

One should use COSMO-SAC or COSMO-RS for property predictions of systems with chemicals for which UNIFAC cannot properly handle, for example, benzene/*n*-methylformamide.⁶ Kolar et al⁵⁹ show that UNIFAC only works for 82% of the 221 systems in their study. We find that UNIFAC is more accurate than COSMO-SAC for most of the systems studied in Chapter 4, but COSMO-SAC is unique in that it is capable of handling complex molecules which may not be modeled appropriately due to the proximity of functional groups. Grensemann and Gmehling⁶⁴ compare COSMO-RS with classical group-contribution methods, specifically modified UNIFAC. They find that the modified UNIFAC model is more accurate than COSMO-RS in predicting VLE behavior, but they also recognize COSMO-based method applicability in cases where group-contribution models fail.

4.7.9 High Temperature and High Pressure Applications

D. Constantinescu et al⁶⁵ expands COSMO-based models to predict vapor-liquid equilibrium behavior at high temperatures and pressures from infinite-dilution activity coefficients using the Huron-Vidal mixing rule.

5 VT-2006 Solute Sigma Profile Database

In this section, we discuss the newly created VT-2006 Solute Sigma Profile Database. We describe this database by its contents and deliverables, validation criteria, and error quantification. We present examples of the effect of conformational changes on solubility predictions for pure and mixed solvent systems. We also compare COSMO-SAC as an *a priori* tool for predicting solubility to other solubility prediction methods. In previous work concerning solubility modeling, researchers present a variety of model comparison studies and other solubility studies focusing on specific compounds.^{3,39,41,42,64,66-68} In the majority of these publications, the authors use different methods to quantify the prediction errors. In some cases, the authors do not publish a listing of compounds by name, but by classification, thus preventing duplication of their work. Still other researchers use smaller sets of compounds which do not lend themselves to generalizations about a model's overall accuracy. Chen and Song³⁹ correlate and predict solid solubility using the NRTL-SAC model, develop and report parameters for several solute and solvent compounds, and thus provide the means for a meaningful comparison. Therefore, we compare solubility predictions in pure solvents using both the NRTL-SAC and COSMO-SAC models. Finally, we discuss our findings as a list of general heuristics and application guidelines.

5.1 Description of VT-2006 Solute Sigma Profile Database

The VT-2006 Solute Sigma Profile Database includes 238 compounds, 206 solutes and 32 solvents. We include 181 new solutes and 32 new solvents since the original release of the VT-2005 Sigma Profile Database. We optimize each compound's molecular structure following the procedure explained in Chapter 3, using Amber8 as a pre-optimization tool for each structure as well. With the exception of five compounds in case studies of the effects of conformational variations, we present one optimized conformation and its sigma profile for each compound. We organize the database by CAS-RN, compound name (common names where applicable), chemical formula, and molecular weight. The VT-2006 Solute Sigma Profile Database uses values for the normal melting point temperatures T_m and latent heats of fusion compiled in the Chemistry Data Series by Marrero and Abildskov²⁷ and other available pure component data

literature sources.^{69,70} We include a listing of all compounds with pure solvent literature data and a measure of each solute's average error in Table A.1 in Appendix A. A complete listing of all compounds and their overall prediction error is available on our website. See Chapter 6.

5.2 Validation of VT-2006 Solute Sigma Profile Database

Here, we present several representative cases and a summary of the validation effort regarding the application of COSMO-SAC to solubility modeling. We predict solubility in pure solvents using the COSMO-SAC model for 2434 literature solubility values, which includes 194 solutes, 160 solvents, and 1356 solute/solvent pairings. We compare COSMO-SAC solubility predictions for mixed (binary and ternary) solvents to literature values for 39 systems and compare the effects of using different definitions for the exchange energy. In this chapter, we compare the exchange energies defined by Lin and Sandler⁶ and by Mathias et al¹³ by the solubility prediction error. Because the hydrogen-bonding term differs in the two exchange energy definitions, we use shorthand notation, W_{hb} [Lin 2002] and W_{hb} [Mathias 2002], to designate which of these two exchange energies we used for a particular COSMO-SAC prediction. We also study conformational effects of solutes and solvents on COSMO-SAC solubility predictions for both pure and mixed solvent solubilities.

5.2.1 Error Calculation Methodology and Quantification

We quantify prediction error using the root-mean-squared error, similar to Chen and Song,³⁹ of $\log_{10}(x_{sol})$, RMSE, comparing experimental and predicted solute mole fractions on a logarithmic base 10 scale. We also calculate the absolute average relative error, AA%E, to compare predicted and experimental solute mole fractions in % mole fraction (% M.F.). We use the definitions in equations (5.1) and (5.2) to calculate prediction error, where x_{sol}^{exp} represents the experimental solute mole fraction, x_{sol}^{pred} is the predicted solute mole fraction, and n is the number of solubility points.

$$\text{RMSE} = \left[\frac{1}{n} \sum_i^n \left(\log_{10}(x_{sol_i}^{exp}) - \log_{10}(x_{sol_i}^{pred}) \right)^2 \right]^{1/2} \quad (5.1)$$

$$\begin{aligned}
 \text{AA}\%E &= \frac{1}{n} \sum_i^n \left| \frac{x_{sol_i}^{exp} - x_{sol_i}^{pred}}{x_{sol_i}^{exp}} \right| \times 100\% = \frac{1}{n} \sum_i^n \left| 1 - \frac{x_{sol_i}^{pred}}{x_{sol_i}^{exp}} \right| \times 100\% \\
 &[\equiv] \% \text{ M.F.}
 \end{aligned}
 \tag{5.2}$$

Because we have access to more data for some compounds, we use several schemes to insure that we give equal weight for each calculation. When calculating the overall error for a given solute, we weight each solute/solvent pair equally by essentially taking an average of the average error. We calculate the average error, either RMSE or AA%E, for a given system over the full range of temperatures, then average all systems for the given solute. When we calculate the error for all systems, we first calculate the average error per solute, weighting each solute/solvent pair equally, and then average the error over all solutes. This method does magnify the effects of outlying predictions; however, we believe it provides a fair analysis.

Because the physical constraints of the composition domain [0:1], we observe very large and very small error measurements for which we discuss a proper frame of reference. Simply put, a model may do one of two things when predicting a single value, such as solute mole fraction, under-predict or over-predict. In this case, under-prediction is a result of an over-prediction of the solute activity coefficient by the model, and vice versa. It follows that when we want to compare an over-prediction with an under-prediction, we need a method for systematic quantification. Over-prediction and under-prediction generate very different AA%E values. The question becomes which model prediction is better if one model over-predicts and the other model under-predicts solute mole fractions? Secondly, how do we determine which prediction is best from these error measurements? The root-mean-squared error of the logarithm of the solute mole fraction informs us of the order of magnitude by which the prediction differs from the experimental value. For example, if the experimental solute mole fraction equals 0.01, regardless of whether the model predicts a value of 0.001 m.f. or 0.1 m.f., the RMSE is 1.0. Essentially, RMSE is a tool for comparing over-prediction to under-prediction. However, the user must decide which prediction to use.

We see a different behavior as a result of under-prediction or over-prediction when calculating AA%E. The standard formula for the absolute average relative percent error is on the left-hand side of equation (5.2), but when we rearrange this formula to the form on the right-hand-side, we see a different picture. In the limit of AA%E, as the ratio of the predicted mole

fraction to the experimental mole fraction $x_{sol}^{pred} / x_{sol}^{exp}$ approaches zero, is 100%. This represents the severe case of under-prediction. In the opposite case, as the ratio of predicted solute mole fraction to experimental solute mole fraction increases due to model over-prediction, the AA%E rapidly increases, especially if the solute mole fraction is less than 0.1, which is the case for 80% of the systems in this study.

Figure 5.1 gives a graphical representation of this discussion. This discussion hopefully places our error analysis of COSMO-SAC solubility modeling in proper perspective. The RMSE values are unaffected by the differences of under- or over-prediction, whereas AA%E values differ greatly. In the under-prediction case, AA%E is limited to a maximum of 100%, but in the over-prediction case, AA%E is unbounded. The RMSE value is linearly proportional to the ratio of the predicted to experimental solute mole fraction $x_{sol}^{pred} / x_{sol}^{exp}$ and is equivalent in value to the order of magnitude difference of the predicted and experimental solubility. We use AA%E to compare model error when model behavior is consistent, for instance, a comparison of two solubility predictions which both over-predict the solute mole fraction.

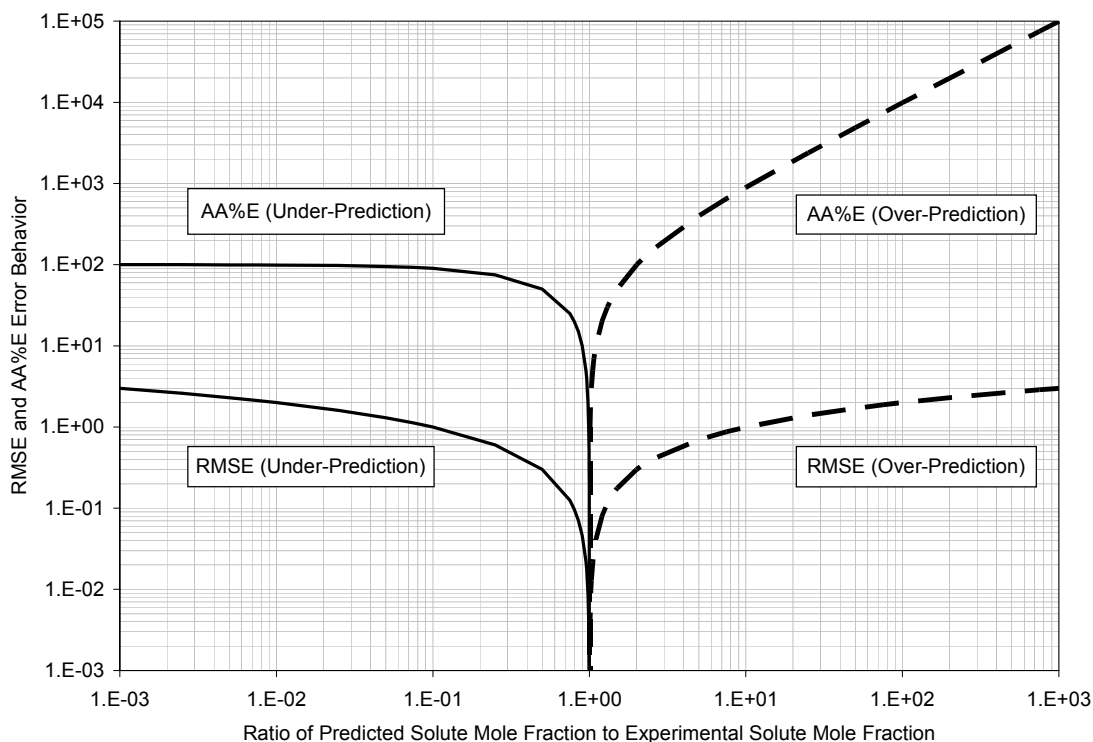


Figure 5.1: RMSE and AA%E behavior in relation to solubility modeling over-prediction and under-prediction on a logarithmic base 10 scale.

5.2.2 Solid Solubility Predictions in Pure Solvents

We use the solubility equation (2.30) and the COSMO-SAC model to predict solubility in pure solvents for two example systems at a given temperature T . We predict the activity coefficient of the solute using the COSMO-SAC model and published pure solute properties (heat of fusion and melting point)²⁷ to generate our predicted values. Figure 5.2 compares COSMO-SAC solubility predictions with published pure-solvent experimental values with solute mole fractions for benzoic acid in 50 different solvents at 25 °C. Benzoic acid is one of 25 of compounds listed in both databases, VT-0610 and VTSOL-044 in the VT-2005 Sigma Profile Database and the VT-2006 Solute Sigma Profile Database, respectively. The difference in the sigma profiles between the two databases is that we generate the sigma profiles in the VT-2006 Solute Sigma Profile Database from energy minimized structures from Amber8 output, where in the VT-2005 Sigma Profile Database we do not use a pre-optimization tool for benzoic acid.

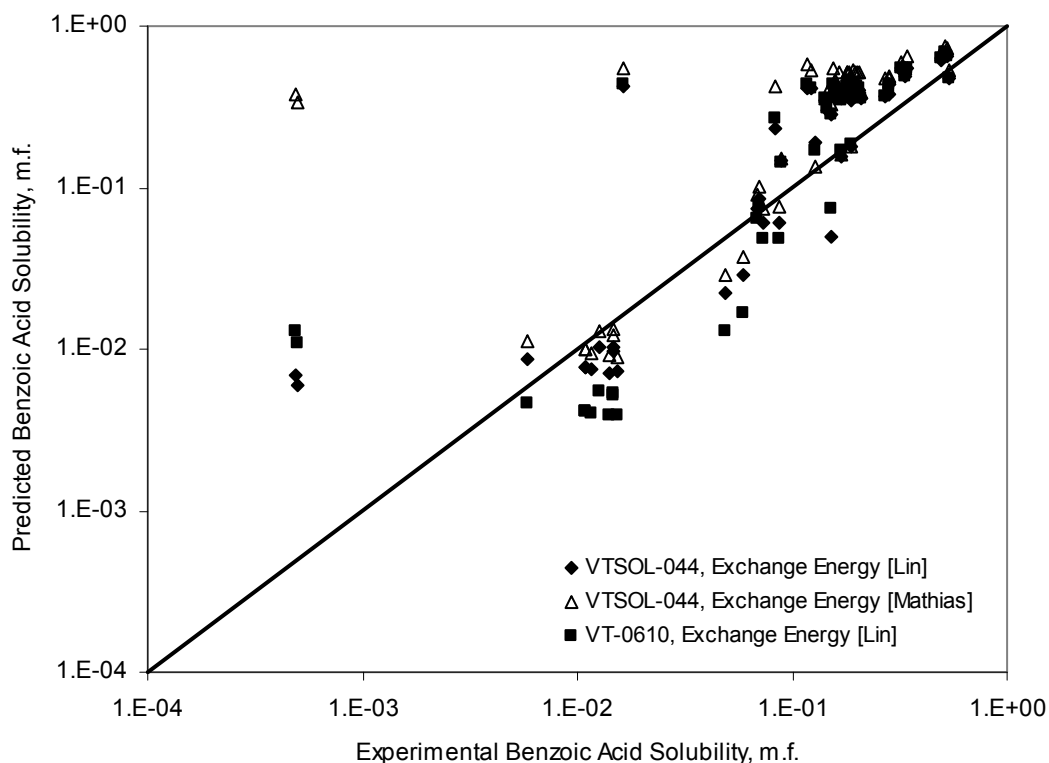


Figure 5.2: Predicted benzoic acid (VTSOL-044, VT-0610) solubility in 50 pure solvents at 298.15 K using both definitions for the exchange energy compared with experimental values on a logarithmic base 10 scale.²⁷

Weighting each solvent equally, we calculate the RMSE and AA%E for predicted benzoic acid solubility using the sigma profiles from both databases (VTSOL-044, VT-0610) and both exchange energy expressions in Table 5.1. Using the Mathias et al¹³ exchange energy expression shows a small improvement in error for 12 of the 50 solvents, but the overall error is greater than the predictions using the Lin and Sandler⁶ exchange energy expression. We see an increase in error when we use the sigma profile from the VT-2005 Sigma Profile Database (VT-0610) over using the VTSOL-044 sigma profile, but this increase in overall error is smaller than the increase in overall error from using the Mathias et al¹³ exchange energy expression in this case.

Table 5.1: Overall error for predicted benzoic acid solubility in 50 pure solvents.

Exchange Energy Error/Solute	W_{hb} [Lin 2002] VTSOL-044	W_{hb} [Lin 2002] VT-0610	W_{hb} [Mathias 2002] VTSOL-044
RMSE	0.2811	0.3554	0.3502
AA%E	125.2	222.7	1604.2

In Figure 5.3, we apply the COSMO-SAC model to predict dibenzofuran (VTSOL-120, VT-0770) solubility in pure solvents, thiophene (VT-0991), cyclohexane (VT-0099), pyridine (VT-0962), and benzene (VT-0242), over a range of temperatures (314 - 345 K) using both exchange energy expressions.

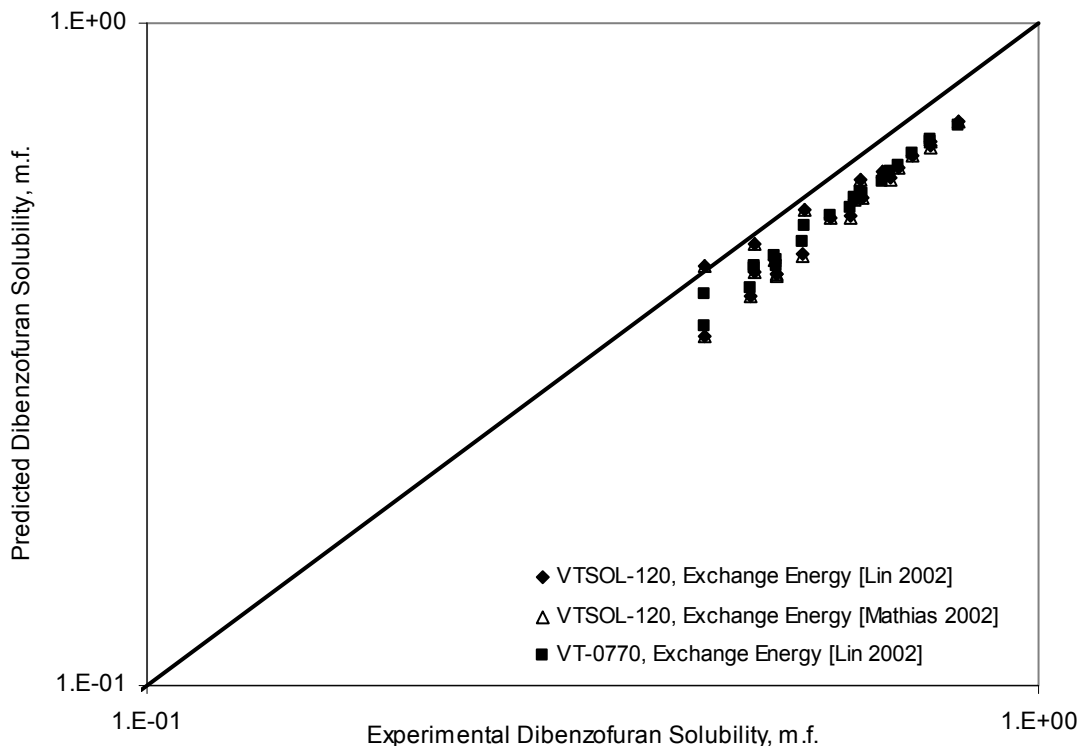


Figure 5.3: Predicted dibenzofuran (VTSOL-120, VT-0770) solubility in four pure solvents at various temperatures using both exchange energy expressions compared with experimental solubility on a logarithmic base 10 scale.²⁷

We calculate the AA%E, weighting each solvent equally, from experimental solubility using the dibenzofuran sigma profiles from both databases and both exchange energy expressions as 12.7%. The error difference between using the two exchange energy expressions and using the two different sigma profiles is statistically insignificant in this case. However, the difference in the two sigma profiles is significant, which leads us to conclude that predicted pure-solvent dibenzofuran solubility is insensitive to changes in its sigma profile. We discuss conformational effects in more detail in Section 5.3.

As stated above, we predict solubility in pure solvents for comparison with 2434 literature solubility points. This sample size includes 1356 solute/solvent pairs, 194 solutes, and 160 solvents. The types of solutes in this study include a wide range of functional groups and combination of functional groups, which we break down into several subsets in this chapter. To calculate the overall prediction error in comparison to all literature values, we weight each solute/solvent pair equally. We calculate the RMSE and AA%E for the entire sample set and several smaller sample sets, which exclude select solutes and solute/solvent pairings. We set a

cutoff value for exclusion at 5000% AA%E to remove outliers. We summarize these error calculations in Table 5.2.

Table 5.2: Error Summary, RMSE and AA%E, of predicted solubility in pure solvents in comparison to experimental values for the entire sample set and select solutes which exclude outliers.

Sample Set	Literature Points	Solute - Solvent Pairs	No. Solutes	W _{hb} [Lin2002]		W _{hb} [Mathias 2002]	
				RMSE	AA%E	RMSE	AA%E
1	2434	1356	194	0.9054	5314.7	1.1139	21215.8
2	2366	1295	171	0.7421	669.3	0.6346	2743.7
3	2205	1180	166	0.7487	711.8	0.9426	9499.4
4	2053	1114	149	0.7181	664.1	0.8254	980.9

We divide the entire data set into several sample sets to demonstrate the changes in overall error as we remove specific outlying solute/solvent pairs and outlying solutes as a whole from the error calculation. Sample set no. 1 includes all literature solubility points and has the greatest values for RMSE and AA%E. Sample set no. 2 excludes all solute/solvent pairs (61) with an error greater than the error cutoff value. Removing these outliers drastically reduces the overall error to 669% AA%E and 0.742 RMSE. Sample sets no. 3 excludes all solute/solvent pairs for 28 solutes whose overall error exceeds cutoff value when using the exchange energy defined by Lin and Sandler⁶ and sample set no. 4 excludes all solute/solvent pairs for 45 solutes whose overall error exceeds the cutoff value when using the exchange energy defined by Mathias et al¹³.

When we sort the predicted solubilities by their respective literature values, we see that the AA%E value rapidly increases as the literature solute composition in a pure solvent drops below 10 mole percent due to model over-prediction. We also see that, on average, the exchange energy defined by Lin and Sandler⁶ generates more accurate solubility predictions in pure solvents for the majority of the systems. Based on the exclusions in sample sets 2 – 4 in Table 5.2, we caution users when using the COSMO-SAC model to predict solubility with the 19 solutes listed in Table 5.3, because each solute generates an error greater than the cutoff value for more than half of the solvents studied without further model improvements or in depth research. However, we must note that literature values for only one solvent are available for most of these solutes. In our experience, the COSMO-SAC model generally over-predicts solute mole fraction; however, there are exceptions. We may be able to correct this over-prediction with a re-parameterization of the COSMO-SAC model parameters, incorporating SLE predictions into the parameterization. We discuss this potential improvement further in Chapter 8.

Table 5.3: Nineteen solutes with an overall AA%E exceeding the error cutoff value of 5000%.

Index No.	Solute Name	Formula	Solvents	RMSE	AA%E
6	3-Hydroxy-5-Nitropyridine	C5H4N2O3	1	2.8851	76661.0
7	5-Chloro-3-Pyridinol	C5H4CLNO	1	1.7924	6100.8
11	2-Amino-2-Nitropyridine	C5H5N3O2	1	1.9094	8017.9
15	2-Aminopyridine	C5H6N2	1	2.2439	17433.8
20	2,4,6-Triiodophenol	C6H3I3O	2	1.6793	6517.5
25	2-Hydroxynicotinic Acid	C6H5NO3	1	2.1987	15701.8
47	3-Hydroxybenzoic Acid	C7H6O3	1	2.8961	78620.1
52	P-Hydroxybenzamide	C7H7NO2	1	2.7791	60034.2
54	4-Aminosalicylic Acid	C7H7NO3	1	1.8747	7633.0
59	4-Hydroxybenzyl Alcohol	C7H8O2	1	3.0487	111776.5
62	2,6-Pyridinedimethanol	C7H9NO2	1	3.2640	183567.1
77	Phenoxyacetic Acid	C8H8O3	1	2.0036	9982.3
79	3-Hydroxy-4-Methyloxybenzoic Acid	C8H8O4	1	1.9305	8420.3
85	Acyclovir	C8H11N5O3	2	1.6588	6144.1
108	Sulfapyridine	C11H11N3O2S	2	1.6406	22005.2
146	Mitotane	C14H10CL4	4	2.4328	101776.2
155	Morphine	C17H19NO3	8	1.5885	25827.0
191	Desoxycorticosterone Acetate	C23H32O4	1	2.5366	34306.0
196	Norethindrone Enanthate	C27H38O3	1	1.9127	8078.3

5.3 Effects of Conformational Isomerism on Solubility Predictions

As we discuss in Chapter 4, molecular conformation plays a large part in the accuracy of any property prediction with the COSMO-SAC model. We quantify the effect of changes in molecular conformations on solubility predictions. For our case study, we use five examples of common pharmaceuticals and bio-related compounds, acetaminophen, aspirin, ibuprofen, lidocaine, and cholesterol. We present anywhere from three to six different conformations in each case, depending on the complexity of each compound. We also follow the geometry optimization procedure, whether we draw the conformation manually or we import the conformation from Amber8 output. We generate conformation A, in each case study, from Amber8 output and we draw the remaining conformations manually. We outline the optimization procedure in Chapter 3. We compare and rank each conformation based on several criteria: (1) the condensed phase energy, (2) the relative difference in the peak height of each conformation's sigma profile using the conformation with the lowest condensed phase energy as a reference, (3) error calculations, RMSE and AA%E, calculated by equations (5.1) and (5.2), from experimental mole fraction weighting each solvent equally, and (4) the number of solvents

for which a conformation predicts the most accurate results relative to the other conformations. We refer to the number of solvents (criterion no. 4) as “Best Case” in the following summary tables in this section. We also compare both exchange energy definitions, equation (2.7) defined by Lin and Sandler⁶ and equations (2.11) and (2.12) defined by Mathias et al.¹³ We summarize the criteria in an abbreviated table in the following sections, but we include an unabridged table itemizing the error for each solute/solvent pairing in Appendix A for each case.

5.3.1 Acetaminophen (VTSOL-082, CAS-RN: 103-90-2, C₈H₉NO₂)

Acetaminophen, or paracetamol, is a common over-the-counter analgesic and antipyretic. Acetaminophen is not a non-steroidal anti-inflammatory drug (NSAID), but it is common used in conjunction with NSAIDs. We present several optimized conformations of acetaminophen and the relative differences in their sigma profiles and condensed phase energies. As we see in Figure 5.4 and Figure 5.5, these three conformations differ by the hydroxyl hydrogen bond angle only. The calculated condensed phase energies for these three acetaminophen conformations are ($- 515.6557945 E_h$), ($- 515.6561944 E_h$), ($- 515.6492785 E_h$) for conformation A, B, and C, respectively.

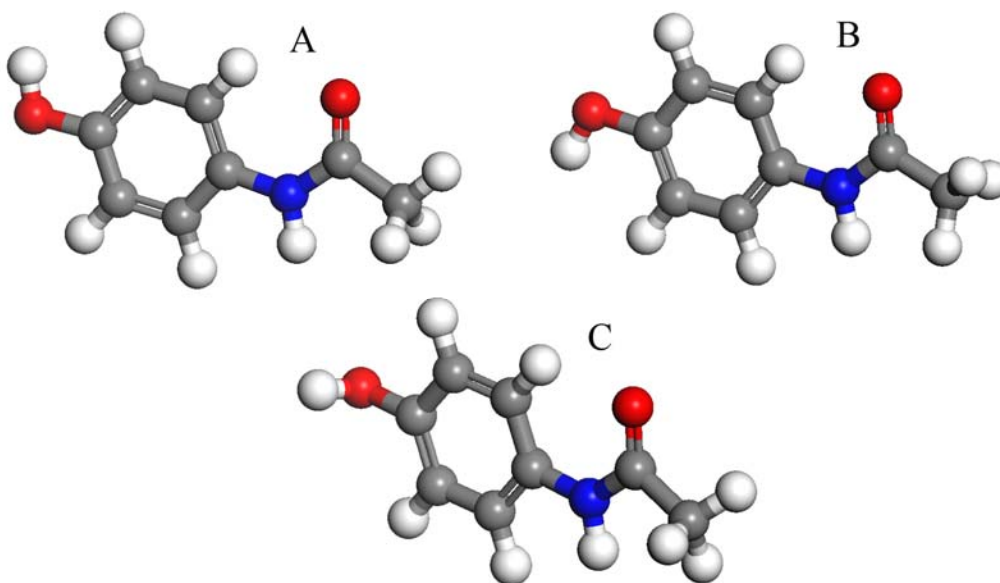


Figure 5.4: Three acetaminophen conformations with variations in the position of the hydroxyl group hydrogen atom relative to the aromatic ring.

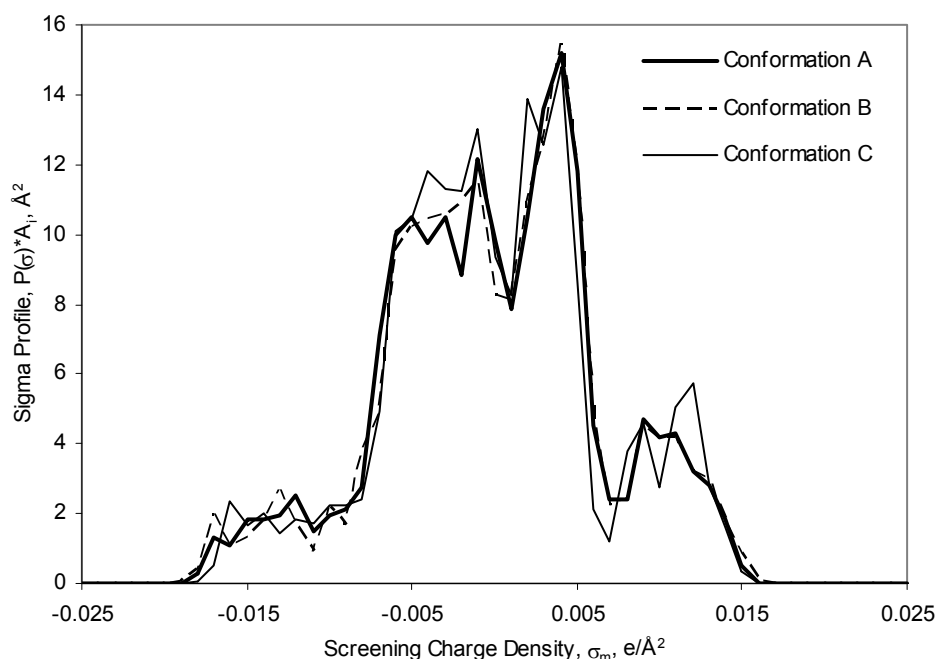


Figure 5.5: Sigma Profiles for their respective acetaminophen conformations, A, B, and C.

Relative to conformation B which has the lowest condensed phase energy, the sigma profile of conformation A differs in peak height by 15%, and conformation C differs by 30%. We compare literature values for 95 pure-solvent solubility points for 26 solvents with COSMO-SAC predictions using both exchange energy expressions. Table 5.4 summarizes the error calculations and remaining evaluation criteria.

Table 5.4: Error summary, RMSE and AA%E, for each acetaminophen conformation in 26 pure solvents at various temperatures comparing both exchange energy expressions.

Exchange Energy Conformation	W_{hb} [Lin 2002]			W_{hb} [Mathias 2002]		
	RMSE	AA%E	Best Case	RMSE	AA%E	Best Case
A	0.5329	310.7	1	0.5753	290.2	0
B	0.5291	262.2	5	0.5865	267.8	4
C	0.4835	325.1	20	0.5385	259.6	22

Conformation C has both the lowest RMSE and predicts the most accurate solubility for the majority of solvents. On average, the exchange energy equation defined by Lin and Sandler⁶ is more accurate, but for seven solvents, toluene, ethyl acetate, 1,4-dioxane, dichloromethane, chloroform, diethyl amine, and acetonitrile, the exchange energy defined by Mathias et al¹³ generates more accurate results, which explains the discrepancy between the AA%E values.

Figure 5.6 shows the COSMO-SAC predicted pure-solvent acetaminophen solubilities using the exchange energy expression defined by Lin and Sandler⁶ compared with literature solubilities.²⁷

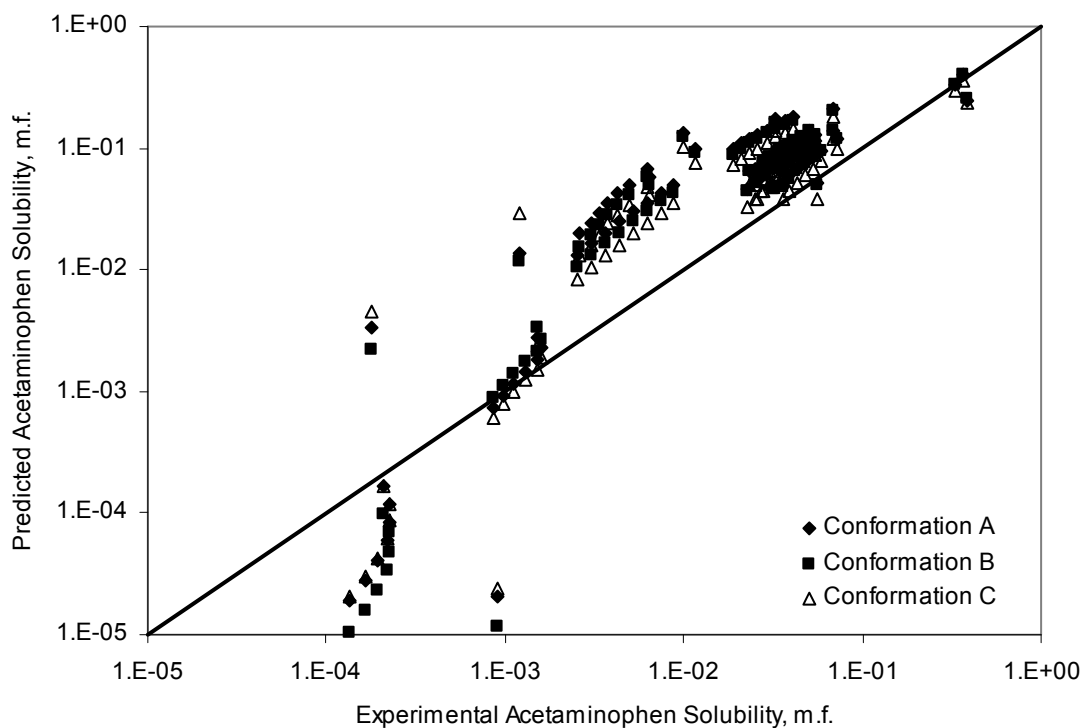


Figure 5.6: Predicted acetaminophen solubility in 26 pure solvents compared with experimental values²⁷ at various temperatures for each conformation using the exchange energy defined by Lin and Sandler⁶ on a logarithmic base 10 scale.

We include a detailed table with error calculations for each conformation and solvent in Table A.2 in Appendix A. Finally, after evaluating each conformation based on our criteria, Conformation C generates the most accurate results on average with exchange energy expression defined by Lin and Sandler,⁶ despite having the highest calculated condensed phase energy. From the difference in the error between each conformation per solvent, we find that acetaminophen is mildly sensitive to changes in its conformation with COSMO-based methods.

5.3.2 Aspirin (VTSOL-088, VT-1422, CAS-RN: 50-78-2, C₉H₈O₄)

Aspirin is another common pharmaceutical with similar analgesic properties, but with different functional groups than acetaminophen, an ester and a carboxyl group. Aspirin is a NSAID and a derivative of benzoic acid. Shown in Figure 5.7, we rotate the ester group by 45°

between conformation A and C about the C-O bond, and we rotate the carboxylic acid group by 180° between conformations A and B and rotate the hydroxyl hydrogen atom by 180° between conformation B and C. We also note that the functional groups in aspirin are much closer in proximity than the functional groups in acetaminophen. Conformations A and B show how these functional groups might align when unaffected by other functional groups, whereas conformation C may better describe the interaction of these functional groups when in close proximity.

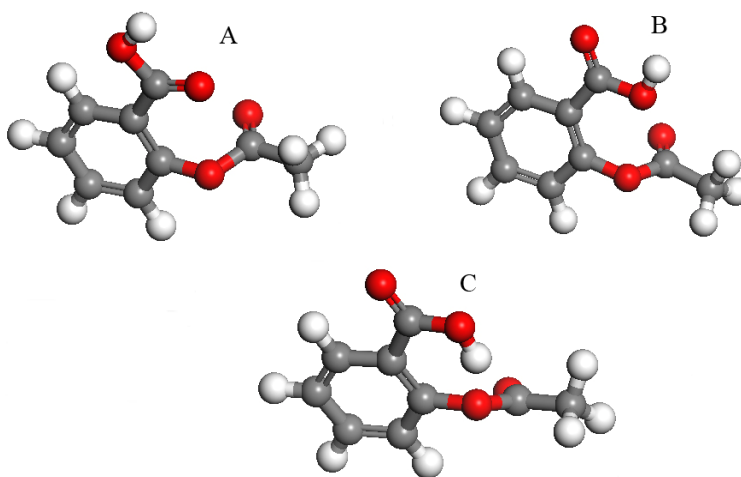


Figure 5.7: Three optimized aspirin (VTSOL-088) conformations with variations in the relative positions of the carboxyl group in proximity to the ester linkage.

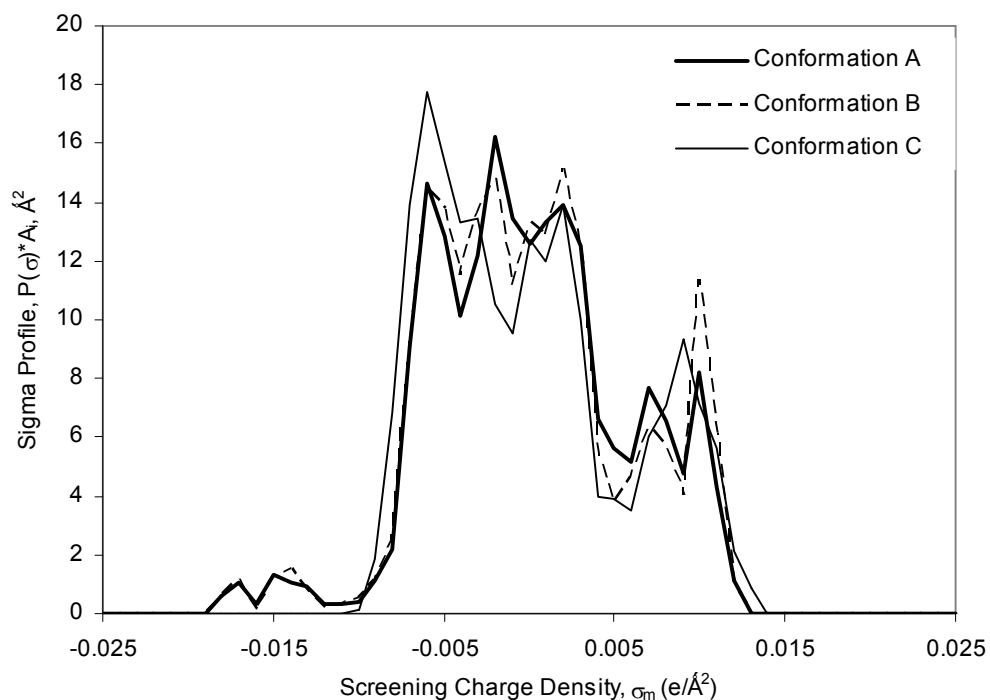


Figure 5.8: Sigma Profiles for their respective aspirin conformations, A, B, and C.

We calculate the condensed phase energy for each conformation in the COSMO calculation, and the energies of each conformation are $(-648.9150609 E_h)$, $(-648.9149189 E_h)$, $(-648.9148635 E_h)$, for conformation A, B, and C, respectively. Conformation A has the lowest condensed phase energy, and the resulting sigma profiles of conformations B and C differ in peak height by 17% and 56% from the sigma profile of conformation A. The sigma profile of conformation C does not have peaks below $-0.010 \text{ e}/\text{\AA}^2$, where both conformation A and B have two peaks past this point. Table 5.5 summarizes the calculated errors for each conformation.

Table 5.5: Error summary, RMSE and AA%E, for each aspirin conformation in 15 pure solvents at various temperatures comparing both exchange energy expressions.

Exchange Energy Conformation	W_{hb} [Lin 2002]			W_{hb} [Mathias 2002]		
	RMSE	AA%E	Best Case	RMSE	AA%E	Best Case
A	0.5857	496.8	1	0.7586	1216.3	2
B	0.5487	383.0	3	0.7346	1203.5	3
C	0.4947	895.4	11	0.4759	903.1	10

We find that conformation C generates predictions with the lowest RMSE regardless of which exchange energy expression we use. Conformation C generates better predictions for 11 of 15 when using the exchange energy defined by Lin and Sandler⁶ and 10 of 15 solvents when using the exchange energy defined by Mathias et al.¹³ From the difference in the RMSE between each conformation per solvent, we find that aspirin is sensitive to conformational effects. Figure 5.9 illustrates the sensitivity to conformational effects of the COSMO-SAC predicted solubilities.

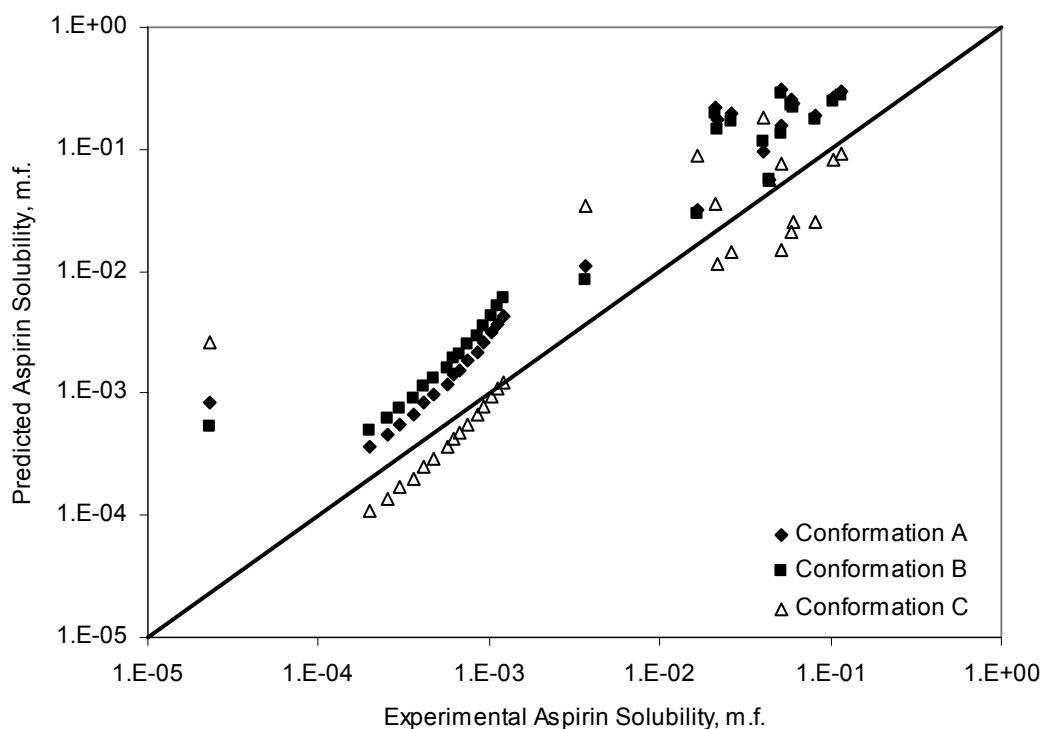


Figure 5.9: Predicted aspirin solubility in 15 pure solvents compared with experimental values²⁷ at various temperatures for each conformation using the exchange energy defined by Lin and Sandler⁶ on a logarithmic base 10 scale.

We summarize the error calculations for aspirin solubility predictions in each solvent Table A.3 in Appendix A. From this table, we observe that for 9 of the 15 solvents, the exchange energy defined by Mathias et al¹³ generates lower RMSE values. We see a significant improvement when using the exchange energy defined by Mathias et al¹³ with water and chloroform. If we exclude one outlying prediction, cyclohexane, from our error calculation, we see a significant reduction in overall error as summarized in Table 5.6.

Table 5.6: Error summary for each aspirin conformation in 14 pure solvents, excluding cyclohexane, comparing both exchange energy expressions to experimental values.²⁷

Exchange Energy Conformation	W_{hb} [Lin 2002]			W_{hb} [Mathias 2002]		
	RMSE	AA%E	Best Case	RMSE	AA%E	Best Case
A	0.5169	286.9	1	0.7011	1048.9	2
B	0.4911	256.1	2	0.6882	1123.5	2
C	0.3831	151.9	11	0.3623	141.4	10

5.3.3 Ibuprofen (VTSOL-141, VT-1205, CAS-RN: 15687-27-1, C₁₃H₁₈O)

Both acetaminophen and aspirin are relatively smaller pharmacological compounds, so here we study conformational effects on a relatively medium-sized molecule, ibuprofen, which is also an analgesic. Ibuprofen has one aromatic ring similar to acetaminophen and aspirin, but more single bonds. The change in the relative positions of the para-substituted carboxyl and isobutyl functional groups has little effect on the sigma profile. We calculate the condensed phase energy as $(-656.8755021 E_h)$, $(-656.8754566 E_h)$, and $(-656.8740487 E_h)$, for conformation A, B and C respectively. The relative difference in peak height in the resultant sigma profiles is 2.5% for conformation B and 15.7% for conformation C using the conformation A sigma profile as a reference. Figure 5.10 shows the variations in the relative positions of the atoms in three conformations.

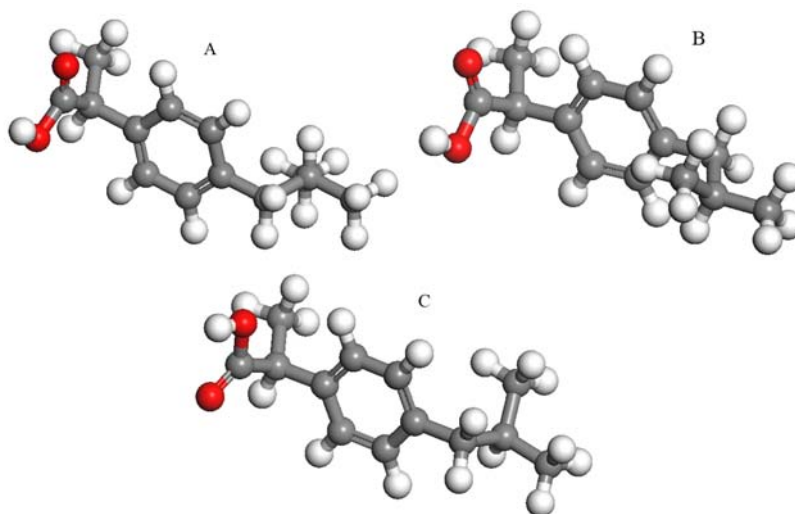


Figure 5.10: Three optimized ibuprofen conformations with variations in the position of the carboxyl group in relation to the isobutyl functional group.

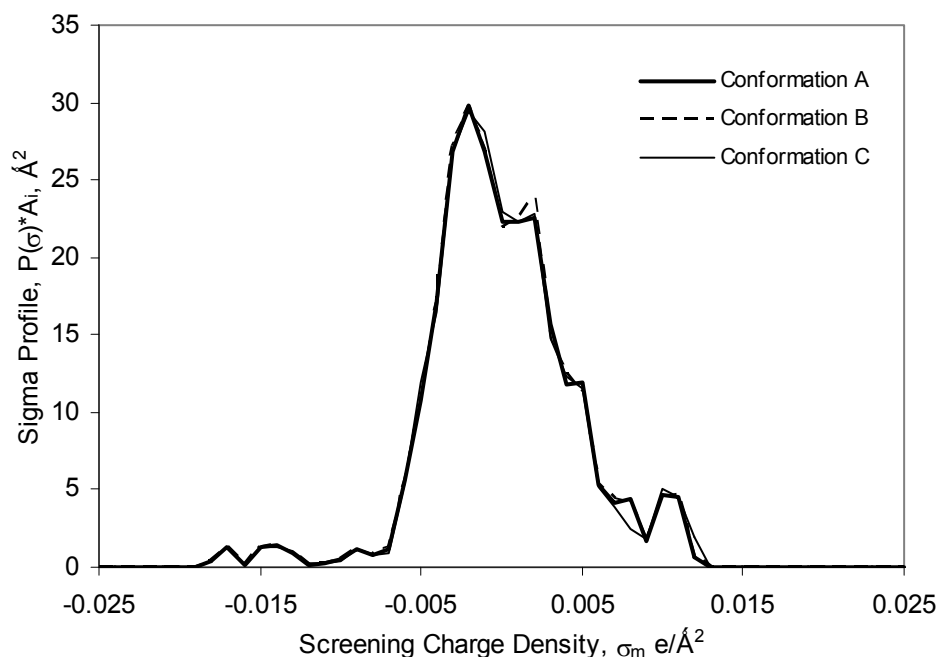


Figure 5.11: Sigma Profiles for their respective ibuprofen conformations, A, B, and C.

Despite the similarities in the sigma profiles in Figure 5.11, all three conformations grossly over predict pure-solvent solubilities for the majority, 16 of 23, of the solvents we study. Only seven solvents, acetic acid, acetone, chloroform, cyclohexane, ethyl acetate, and heptane, have AA%E values below 100%. We identify the worst three solvents, acetophenone, 1,2-propanediol, diethyl ether, and exclude them from the error summary in Table 5.7. We also include Table A.4 in Appendix A which itemizes the error calculations for each conformation and solvent.

Table 5.7: Error summary for each ibuprofen conformation in 20 pure solvents, excluding acetophenone, 1,2-propanediol, and diethyl ether, comparing both exchange energy expressions to experimental values.²⁷

Exchange Energy Conformation	RMSE	W_{hb} [Lin 2002]		W_{hb} [Mathias 2002]		
		AA%E	Best Case	RMSE	AA%E	Best Case
A	0.5961	639.7	1	0.7130	1976.6	1
B	0.5933	630.3	2	0.7109	1922.6	1
C	0.5538	539.9	17	0.6742	1438.7	18

Figure 5.12 shows the predicted ibuprofen solubilities versus their literature values for all 23 pure solvents.

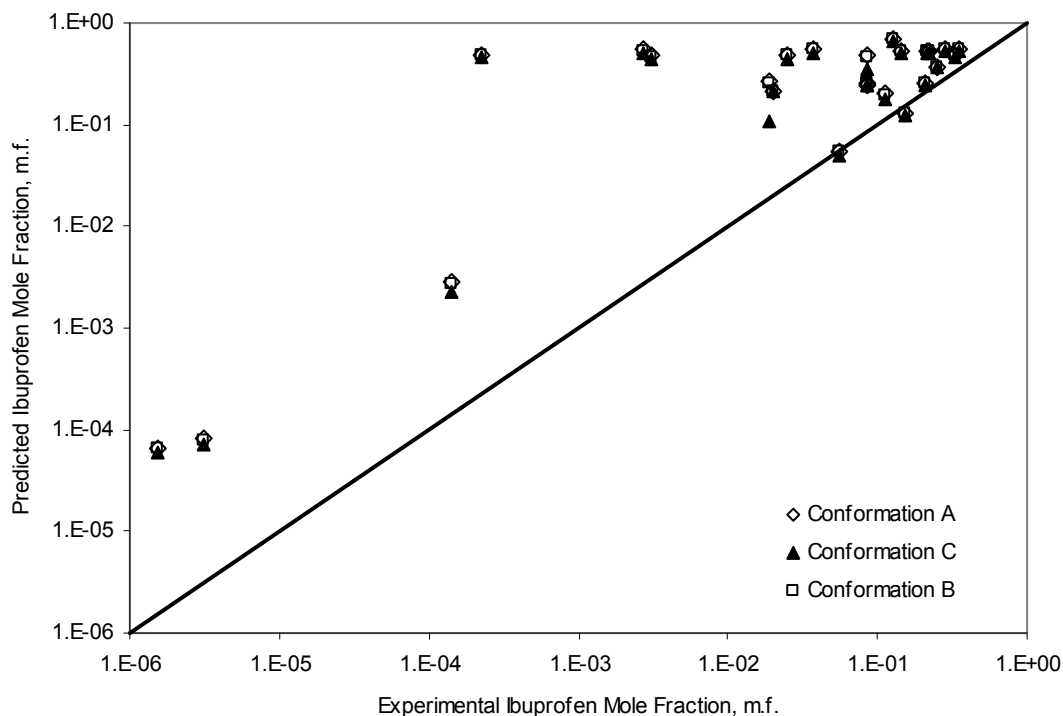


Figure 5.12: Predicted ibuprofen solubility in 25 pure solvents²⁷ for each conformation using the exchange energy defined by Lin and Sandler⁶ at various temperatures on a logarithmic base 10 scale.

The most accurate overall set of solubility predictions comes from using conformation C with the exchange energy expression defined by Lin and Sandler.⁶ This exchange energy definition is more accurate for all but two solvents, cyclohexane and chloroform. Ultimately, ibuprofen pure-solvent solubility predictions are mildly sensitive to the effects on sigma profiles by conformational variations, but the COSMO-SAC model does not accurately predict ibuprofen solubility in pure solvents, with a few exceptions. We see that conformation C, while having the highest condensed phase energy, predicts the most accurate ibuprofen pure-solvent solubilities.

5.3.4 Lidocaine (VTSOL-148, CAS-RN: 137-58-6, C₁₄H₂₂N₂O)

Lidocaine is a local anesthetic and anti-arrhythmic drug. It functions by limiting nervous signals by blocking the sodium ion channels in the cellular membrane. Lidocaine also consists of more atoms in the molecules than the cases above, and we calculate that a lidocaine molecule

cavity occupies slightly more volume than the ibuprofen molecule cavity. We classify lidocaine as a medium-sized molecule, similarly to ibuprofen. We compare COSMO-SAC pure-solvent solubility predictions of four slightly different conformations in twenty solvents with experimental solubilities.²⁷ We optimize each conformation following the procedure outlined in Chapter 3. Conformation A results from using Amber8 output as an initial structure, and we manually draw the remaining conformations by rotating the tertiary amine into different positions relative to the aromatic ring and the amide. Figure 5.13 and Figure 5.14 illustrate the differences in the four conformations and the differences in the sigma profiles. We rank each conformation by lowest to highest condensed phase energy. Conformation B has the lowest condensed phase energy ($-731.6647081 E_h$), followed by conformation A ($-731.6594249 E_h$), D ($-731.6588885 E_h$), and C ($-731.6559627 E_h$) in that order. The relative differences of the peak height in the sigma profiles of A, D and C are 13.6%, 13.5%, and 15.9%, respectively.

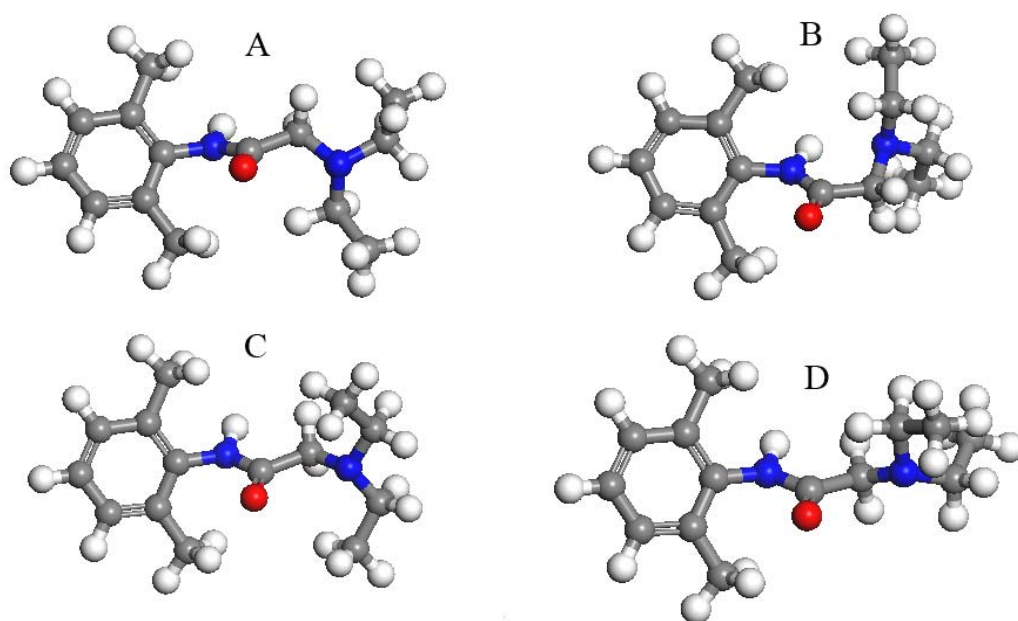


Figure 5.13: Four lidocaine conformations and their variations in structure concerning the relative positions of the amide and amine functional groups.

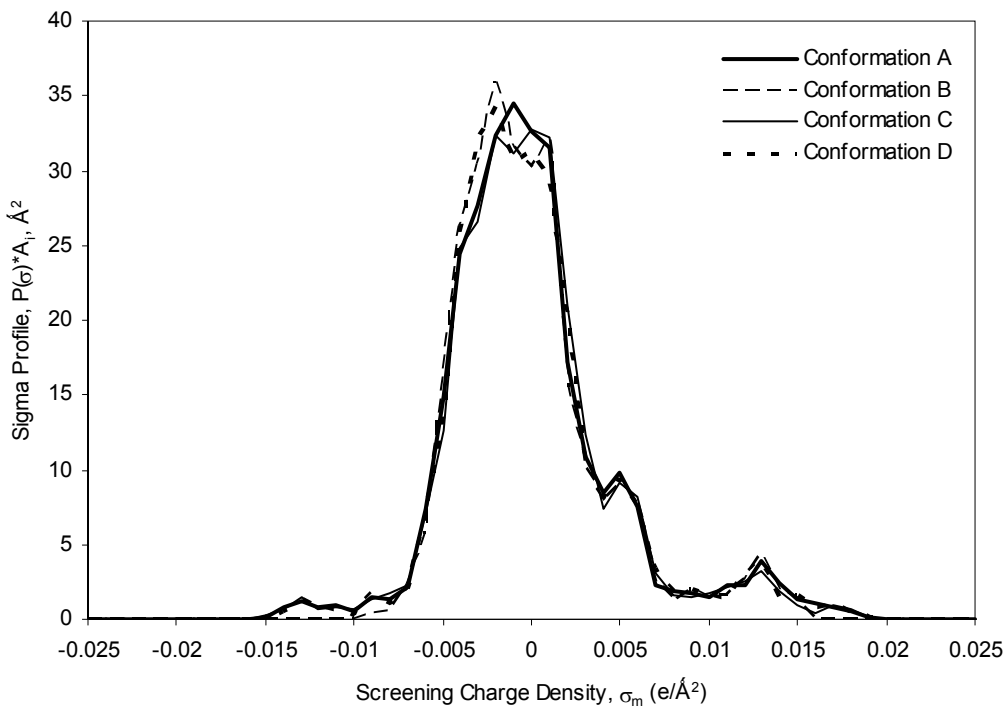


Figure 5.14: Sigma Profiles for their respective lidocaine conformations.

We compare lidocaine solubility for twenty pure solvents, and summarize the resulting error from experimental values in Table 5.8. We also include an itemized listing of calculated errors for each conformation and solvent in Table A.5 in Appendix A. The exchange energy expression defined by Mathias et al¹³ generates improved predictions for 12 of the 20 solvents studied, some of which are statistically significant.

Table 5.8: Error summary for each lidocaine conformation in 20 pure solvents comparing both exchange energy expressions to experimental values.²⁷

Exchange Energy Conformation	W_{hb} [Lin 2002]			W_{hb} [Mathias 2002]		
	RMSE	AA%E	Best Case	RMSE	AA%E	Best Case
A	0.1241	37.5	7	0.1265	42.5	4
B	0.1355	32.7	10	0.1214	33.3	11
C	0.1357	34.9	3	0.1386	40.9	3
D	0.1284	38.9	0	0.1293	44.4	2

We find that conformation B has the lowest prediction error for both exchange energy expressions, and the most accurate prediction for at least half of the solvents when using either exchange energy expression. We also find that pure-solvent lidocaine solubility predictions are significantly affected by variations in sigma profiles due to conformational changes for some

solvents, such as 1,2-propanediol, 1,3-propanediol, and triacetin, while being fairly insensitive to conformational changes for other solvents. Figure 5.15 compares COSMO-SAC predicted lidocaine pure-solvents solubilities for each solvent studied to their experimentally determined values.²⁷

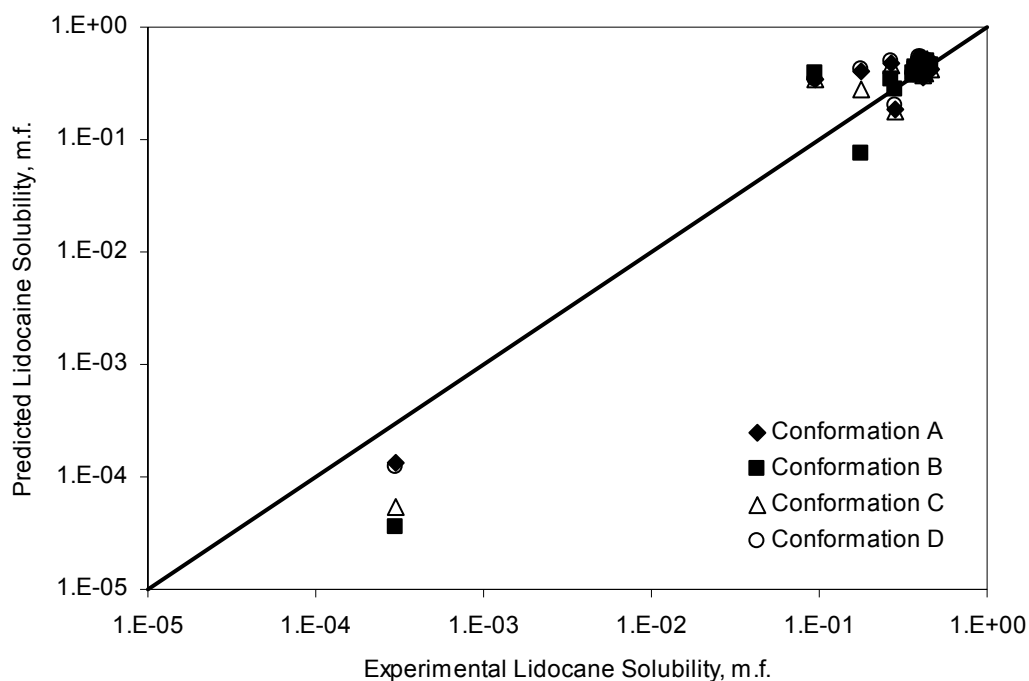


Figure 5.15: Predicted lidocaine solubility in 20 pure solvents using the exchange energy expression defined by Lin and Sandler⁶ compared to their experimental value at 298.15 K on a logarithmic base 10 scale.²⁷

5.3.5 Cholesterol (VTSOL-197, CAS-RN: 57-88-5, C₂₇H₄₆O)

Cholesterol is an important bio-molecule and comes from several sources. Much research has gone into reducing and controlling the cholesterol levels in the body. Cholesterol is also a relatively large molecule when compared to the previous four cases. Figure 5.16 illustrates the relative differences in six optimized cholesterol conformations. Conformation F has the lowest condensed phase energy, followed by conformations A, E, B, C, and D ranked from lowest to highest condensed phase energy. However, conformations F and D are only separated by $0.025 E_h$. By using conformation F as a reference, the remaining sigma profiles differ in peak height by 25%, 17.5%, 20%, 52%, and 20% for conformations A, B, C, D, and E,

respectively. We study the COSMO-SAC solubility predictions of cholesterol in sixty pure solvents at various temperatures. Conformation A is the resulting structure from using Amber8 as a pre-optimization tool.

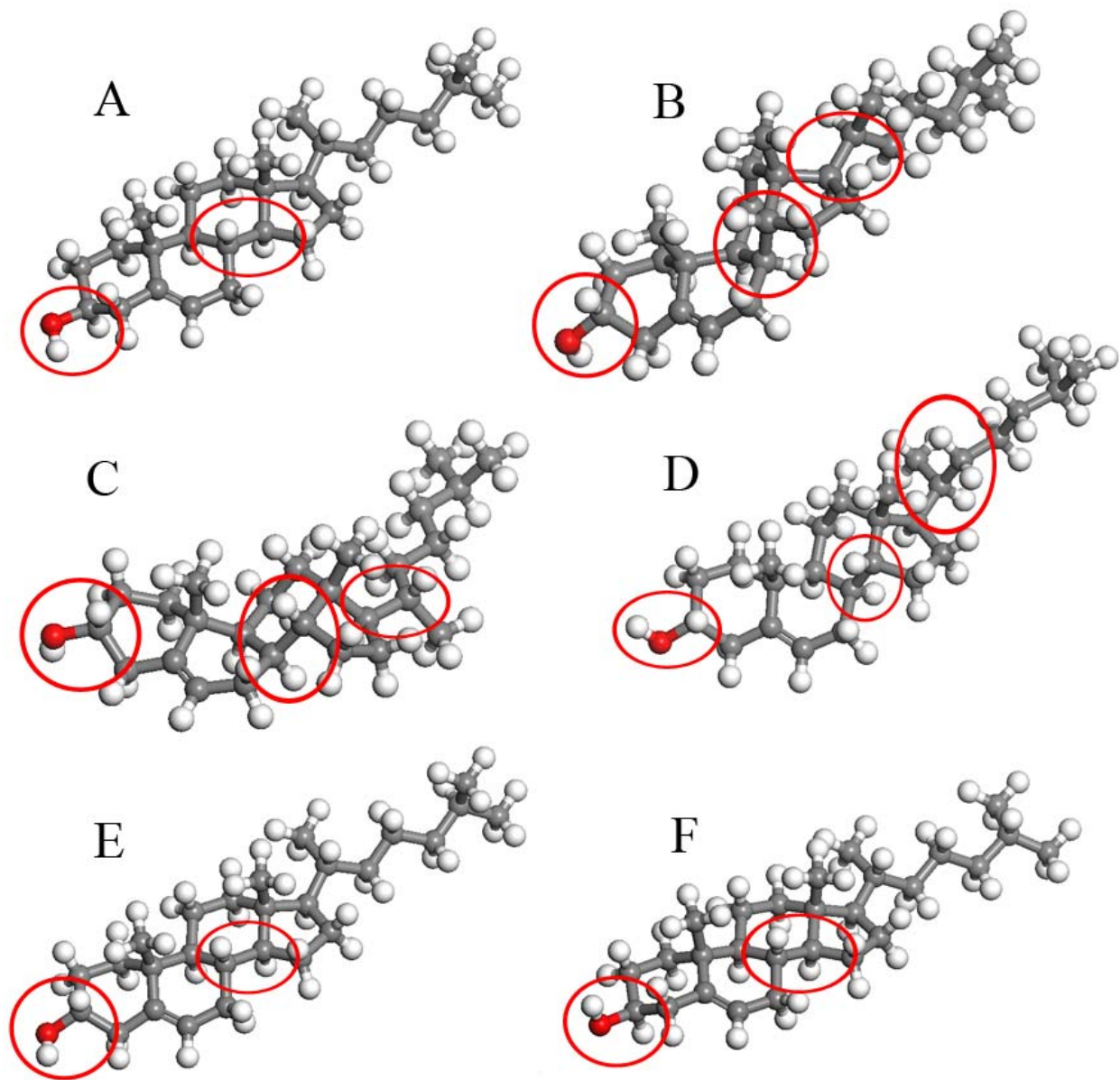


Figure 5.16: Six cholesterol conformations with circled regions to emphasize their relative differences. Conformations A and F have similar atomic positions for all atoms except the hydroxyl hydrogen placement.

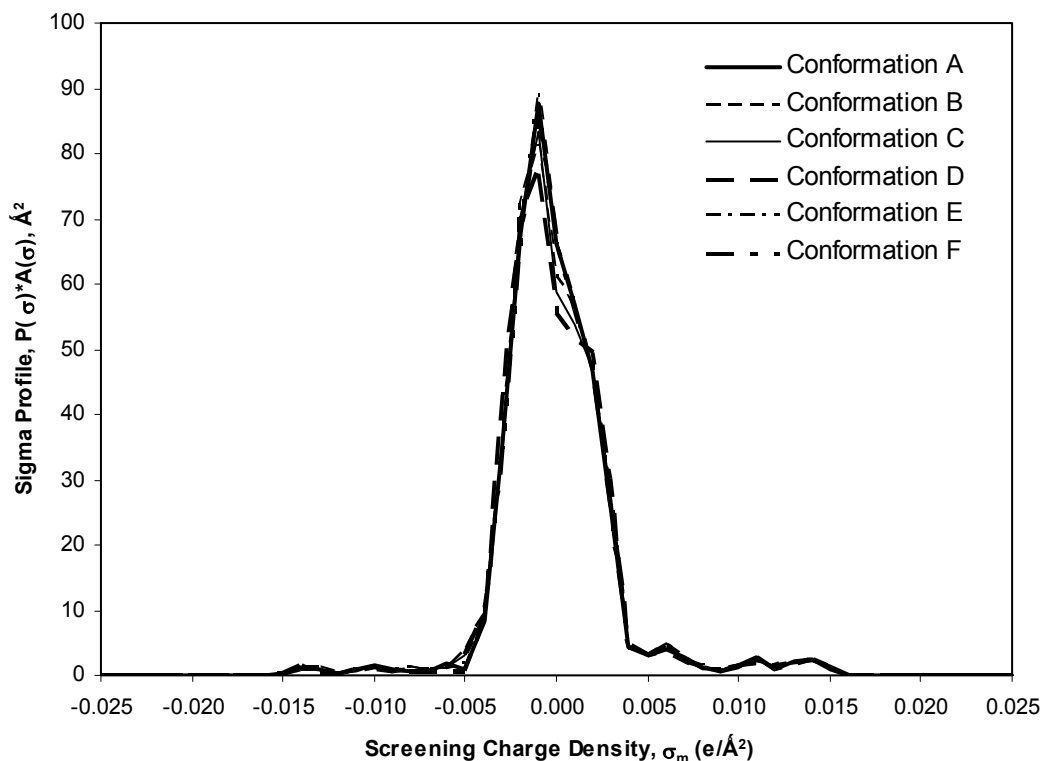


Figure 5.17: Sigma Profile for six cholesterol conformations.

We observe relatively small variations in the sigma profile which is likely the result of the conformational freedom of the molecule. Although cholesterol is the largest molecule we use to study conformational effects, it also has the largest fused-ring structure, thus inhibiting its conformational freedom. Despite the cholesterol conformations exhibiting several differences in structure, the variations in their sigma profiles are not significant. We see in Table 5.9 that using conformation A with the exchange energy defined by Lin and Sandler⁶ generates the lowest overall RMSE and the most accurate solubility prediction for the majority of solvents. However, there is only a small variation between the best and worst predictions, which leads us to the conclusion that cholesterol is not sensitive to conformational changes. We also include a supplemental table, Table A.6 in Appendix A, to summarize the error calculations for each conformation and each solvent.

Table 5.9: Error summary for each cholesterol conformation in 60 pure solvents comparing both exchange energy expressions to experimental values.²⁷

Exchange Energy Conformation	W_{hb} [Lin 2002]			W_{hb} [Mathias 2002]		
	RMSE	AA%E	Best Case	RMSE	AA%E	Best Case
A	0.4602	214.5	31	0.5140	245.9	19
B	0.4760	206.4	7	0.5294	228.7	4
C	0.4771	207.4	3	0.5315	228.5	6
D	0.4911	202.9	6	0.5554	228.8	3
E	0.4658	208.0	5	0.5207	243.5	8
F	0.4650	213.5	8	0.5168	236.7	20

Noting that conformations A, E, and F are similar with the exception of the hydroxyl group hydrogen atom position and the position of a methyl group on the tail of the molecule, there is still an uneven distribution when we compare the number of solvents for which each conformation generates the lowest RMSE. Figure 5.18 illustrates the relative insensitivity in cholesterol solubility predictions as a function of conformational changes.

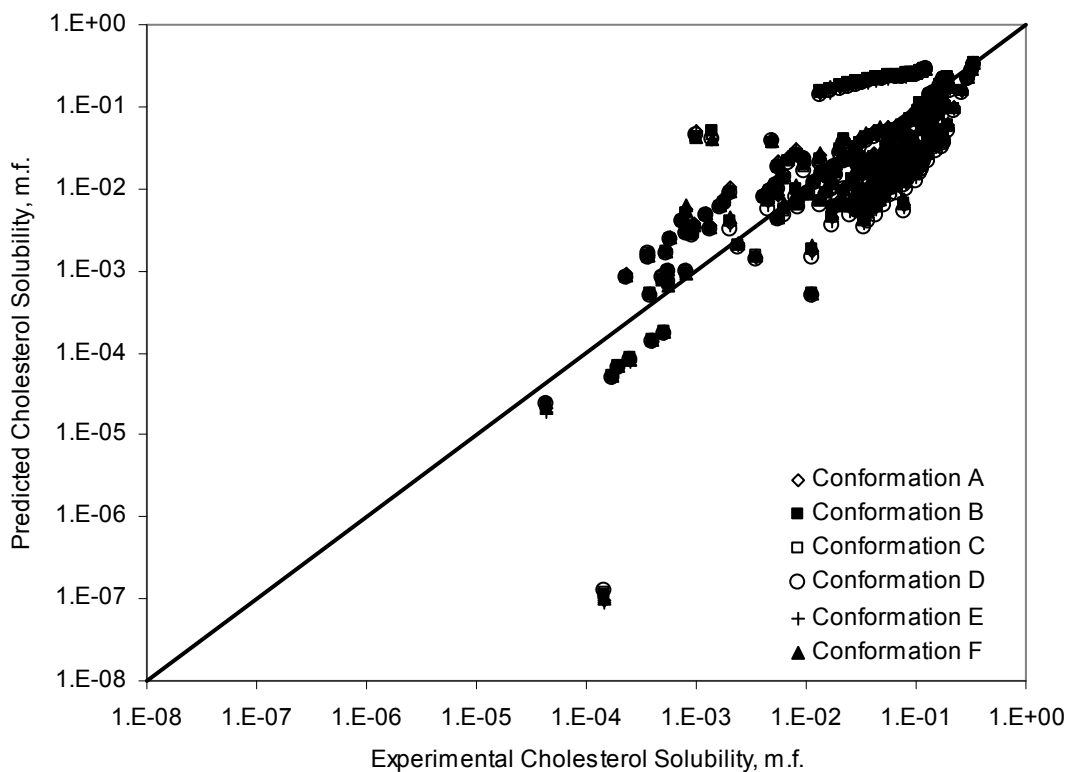


Figure 5.18: Predicted cholesterol solubility in 60 pure solvents for each conformation using the exchange energy equation defined by Lin and Sandler⁶ at various temperatures compared with their experimental values on a logarithmic base 10 scale.²⁷

5.3.6 Summary of Conformational Effects

In the five cases above, we look at different sized molecules with different degrees of conformational freedom, from cholesterol and ibuprofen which show little variation in their sigma profiles as a result of conformational changes, to acetaminophen and aspirin which have the most conformational freedom. We find that conformation C for acetaminophen, aspirin, and ibuprofen generates the best pure-solvent solubility predictions, but conformation C for each molecule has the highest calculated condensed phase energy. We also note that the energy difference between the highest and lowest condensed phase energy is small, usually less than $5.0E-02 E_h$. Acetaminophen and aspirin are the smallest molecules in the above cases, and their predicted solubilities are also most sensitive to conformational variations. Ibuprofen is a larger molecule, but its predicted solubilities are relatively less sensitive as well as less accurate than acetaminophen and aspirin.

We find that lidocaine conformation B, which has the lowest calculated condensed phase energy, generates the best overall predicted solubilities. Lidocaine is a medium-sized molecule, relative to the others in the above cases, and its predicted solubility sensitivity depends on the solute/solvent pair. Lidocaine solubility predictions are only sensitive for a small number of the solvents we study. Cholesterol is the largest molecule and also the least conformationally flexible. We find that cholesterol conformation A generates the best overall solubility predictions. This conformation has the second lowest calculated condensed phase energy to conformation F, but only by a small margin. Conformations A and F are very similar, except for the hydroxyl hydrogen atom placement and the arrangement of the isopropyl group on the tail of the molecule. However, cholesterol is not sensitive to conformational effects. The overall error, when excluding outliers for the best conformation for each molecule, generates an error less than the overall error for all solute/solvent pairs excluding outliers.

5.4 Solid Solubility in Binary and Mixed Solvents

We discuss solubility in pure solvents in the sections above, and now we discuss COSMO-SAC predicted solubility for selected cases in mixed solvents. Because the COSMO-SAC model is not limited to a fixed number of compounds, we quantify its accuracy as a function of the number of compounds present. First, we discuss the model predictions for 14

solutes in 14 binary solvents, overall 37 systems⁷⁰⁻⁷⁴. Several of the systems use ethanol and cyclohexane as one or both components in the binary solvents, and in these cases, we look at all possible combinations of conformations for these solvents. Second, we look at solubility predictions for three ternary solvent systems and one quaternary solvent system. During our iterative calculation scheme, we keep the ratio of solvent composition constant to that reported in the experimental data for all systems.

For the binary solvent systems, all of the solutes are either small or medium-sized molecules. The largest solute is sulfamethazine (VTSOL-130), which has two separate aromatic rings and a molecular weight of 278.33 g/mol. We exclude the system of water (1) ethanol (2) and propyl-4 hydrobenzoate (3) from our error calculations because its AA%E exceeds the error cutoff value of 5000%. We calculate an RMSE of 0.8839 and an AA%E of 986.7% using exchange energy expression defined by Lin and Sandler⁶ for 36 binary solvent systems using the best prediction in each case involving multiple conformations. The calculated RMSE and AA%E when using the exchange energy expression defined by Mathias et al¹³ for 36 binary solvent systems are 1.0418 and 1779.3%, respectively. The Mathias et al¹³ exchange energy expression presents a significant improvement in accuracy for only 4 of the 36 systems. The overall RMSE error for these systems is greater than the overall RMSE error for all solubility predictions in pure solvents. Our sample size is too small to make a general statement regarding the model accuracy in relation to solubility predictions in binary solvents; however, for the systems studied, we find that the error in COSMO-SAC solubility predictions in binary solvents is greater than the overall model error for solubility in pure solvents, excluding outliers.

Acetaminophen, cyclohexane, and ethanol are the only compounds with multiple conformations in this study. We predict solubility in a binary solvent system of (1) cyclohexane and (2) ethanol for 15 of the 37 systems. We use the two conformations for ethanol and cyclohexane from Section 4.5 and the acetaminophen conformations from Section 5.3.1 to compare the conformational effects of the solvent on solubility predictions. We observe that acetaminophen conformation C generates the best results in all but one binary solvent, 1,4-dioxane and water, which is consistent with our findings for predicting solubility in pure solvents. The binary solvent system of cyclohexane and ethanol is the only case involving cyclohexane, and ethanol is used in two other binary solvent systems of ethanol and ethyl acetate, and of ethanol and water. In the system of cyclohexane and ethanol, we see a moderate

decrease in error as a result of using ethanol conformation A in place of the released ethanol conformation, and the cyclohexane conformation has little or no effect on the prediction accuracy. We see a similar increase as a result of using ethanol conformation A in conjunction with ethyl acetate. On the contrary, all systems with a binary solvent system of water and ethanol show no significant increase in prediction accuracy when using ethanol conformation A. In Table A.7, Appendix A, we list the error measurements, RMSE and AA%E, for each binary solvent system, including multiple conformations, using both exchange energy expressions.

Literature Data for solubilities in ternary and quaternary solvents are limited, but we predict mixed solvent solubility for acetaminophen over a range of temperatures⁷¹ and naphthalene at 298.15K⁷² comparing both exchange energy expressions. As with pure and binary solvents, the original exchange energy expression generates more accurate predictions. We summarize the calculated error for each system in Table 5.10.

Table 5.10: Mixed solvent solubility error for acetaminophen⁷¹ and naphthalene⁷² using both exchange energy expressions from experimental solute mole fraction.

Solvent Name(s)					W _{hb} [Lin 2002]		W _{hb} [Mathias et al 2002]	
					RMSE	AA%E	RMSE	AA%E
Water	Acetone	Toluene	Acetaminophen-A	0.2431	67.2	0.2825	86.6	
Water	Acetone	Toluene	Acetaminophen-B	0.2265	60.8	0.2820	87.2	
Water	Acetone	Toluene	Acetaminophen-C	0.1792	42.3	0.2308	62.7	
Methanol	1-Propanol	Water	Naphthalene	1.2985	1889.5	1.6321	4188.4	
Methanol	1-Butanol	Water	Naphthalene	1.2714	1768.4	1.6049	3927.6	
Methanol	1-Propanol	1-Butanol	Water	Naphthalene	4.4427	2903815.5	4.4537	2978292.0

COSMO-SAC grossly over-predicts the naphthalene mole fraction for the quaternary solvent, but predicts solute mole fractions with comparable accuracy to pure and binary solvent systems for the ternary solvent systems. The sample size for these systems is not large enough to make general statements regarding model accuracy as a function of the number of components.

5.5 Comparison of COSMO-SAC and NRTL-SAC Solubility Predictions

We compare predicted COSMO-SAC^{5,6} and NRTL-SAC³⁹ solubilities in pure solvents by both models, which can serve as a point of reference for the relative accuracy when compared with other solubility models. For this comparison, we use NRTL-SAC binary parameters (τ_{12} , τ_{21} , α_{12} , and α_{21}) and molecular segment parameters (X , Y , Y^+ , and Z) for a set of fifteen solutes

published in Chen and Song³⁹. These solutes include benzoic acid, 4-aminobenzoic acid, theophylline, methyl paraben, acetaminophen, aspirin, sulfadiazine, ephedrine, camphor, lidocaine, piroxicam, morphine, estrone, estriol, testosterone, haloperidol, and hydrocortisone with various solvents. We cannot compare the solubility predictions for both models for the same set of solvents for each solute due to incomplete literature data sets. Compliments of Dr. Chen,⁷⁵ we obtain the necessary NRTL-SAC molecular segment parameters for 128 solvents, all of which are included in the VT databases. From these sixteen solutes and 128 solvents, we create a comparison set of 258 solubility points from literature.²⁷ For acetaminophen and lidocaine, we use the best conformation from Section 5.3 with the COSMO-SAC model for this comparison study. We also use the Lin and Sandler⁶ definition for the exchange energy in this comparison.

Finally, we use literature values of $\Delta H_{fus}^{T_m}$ and T_m from Marrero and Abildskov²⁷ to calculate K_{sp} from equation (2.31); however, we duplicate the results in Chen³⁹ using their published K_{sp} values to insure consistency. All literature values in this comparison are at 298.15 K, except acetaminophen at 303.15 K, which is the temperature at which Chen and Song³⁹ regress the binary interaction parameters. We predict solubilities for several solute/solvent systems not included in the regression data set by Chen and Song, therefore, we present different overall error values. For our comparison, we use literature values of the latent heat of fusion and the melting-point temperature²⁷ for both model predictions. By doing so, we evaluate the accuracy of each model, not the accuracy of the literature values for the physical properties.

Figure 5.19 shows the predicted solubilities from both models for the entire data set relative to their respective experimental values. We see the COSMO-SAC model over-predicts solubility for the majority of systems as a result of an under-prediction of the solute activity coefficient, and NRTL-SAC predicts a more even scattering of values. One major difference to point out is that NRTL-SAC is a correlative and predictive model, whereas the COSMO-SAC is only predictive. NRTL-SAC requires binary interaction parameters for every compound, but once the parameters in the COSMO-SAC model are set for all compounds; there is no further need for fitting.

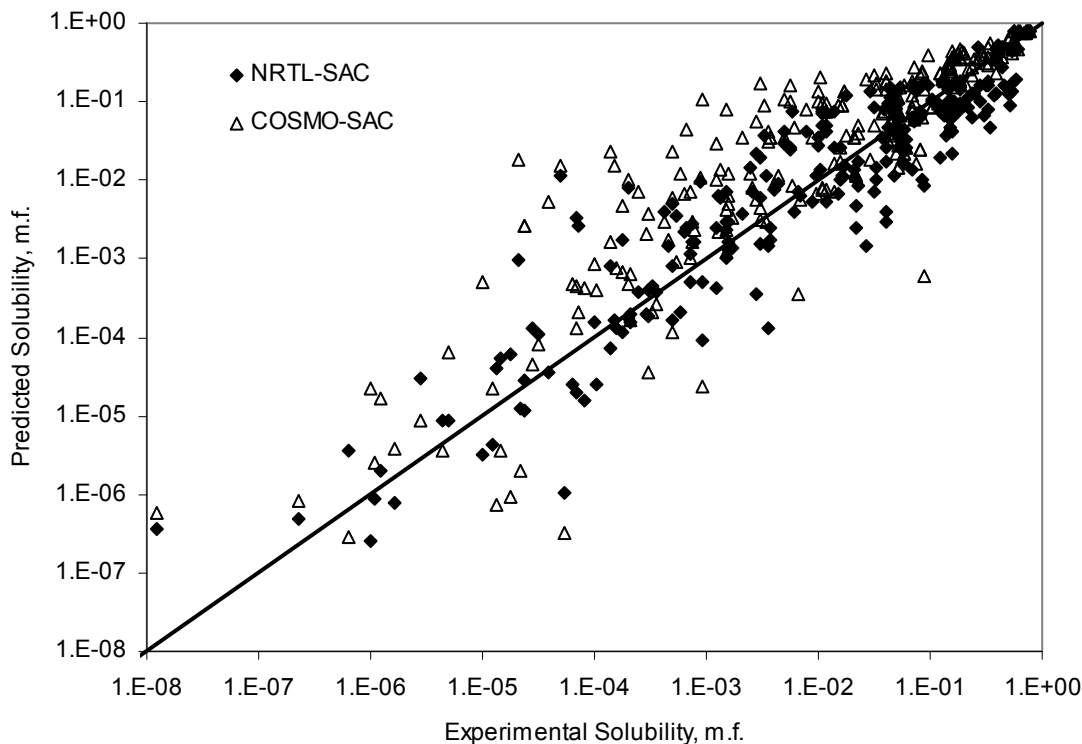


Figure 5.19: NRTL-SAC and COSMO-SAC predicted solubilities for 17 solutes and 258 experimental solubility points at 298.15 K, except acetaminophen solubility at 303.15 K, on a logarithmic base 10 scale.²⁷

For this set of points, the NRTL-SAC model is more accurate at predicting solubilities in pure solvents than the COSMO-SAC with an average RMSE of 0.43 and 0.74, respectively, using equation (5.1), which is greater than the average RMSE value Chen and Song³⁹ report for NRTL-SAC, 0.37 for 14 solutes, but still in general agreement considering the larger comparison set. The COSMO-SAC model generally over-predicts the solute mole fraction with an RMSE value of 0.74, which is comparable to the overall RMSE we report in Section 5.2 when excluding outliers. Table 5.11 summarizes the error calculations for each solute and model, using the Lin and Sandler⁶ definition for the exchange energy with the COSMO-SAC model.

Table 5.11: Comparison Error Summary for NRTL-SAC and COSMO-SAC predicted solubilities for 17 solutes and 258 solute/solvent pairings at 298.15 K, except acetaminophen at 303.15 K.

VT-2006			No. Pure Solvents	NRTL-SAC		COSMO-SAC	
Index No.	Solute Name	CAS-RN		RMSE	AA%E	RMSE	AA%E
44	Benzoic Acid	65-85-0	40	0.5025	133.7	0.3138	96.6
53	4-Aminobenzoic Acid	150-13-0	7	0.3545	49.5	0.7323	551.9
60	Theophylline	58-55-9	13	0.6306	161.0	0.8945	396.6
74	Methylparaben	99-76-3	29	0.5004	65.7	0.4616	181.3
82	Acetaminophen	103-90-2	24 ^a	0.4501	88.6	0.6709	370.4
88	Aspirin-1 ^b	50-78-2	14	0.3887	50.5	0.7169	957.3
88	Aspirin-2 ^c	50-78-2	14	0.4850	59.2	0.7169	957.3
100	Sulfadiazine	68-35-9	2	0.2492	76.8	1.1812	2287.5
105	Ephedrine	299-42-3	9	0.0636	12.4	0.4280	23.9
106	Camphor	76-22-2	8	0.1411	21.9	0.2156	50.1
148	Lidocaine	137-58-6	9	0.1365	33.2	0.3747	57.7
149	Piroxicam	36322-90-4	22	1.0149	2004.5	1.4503	7432.0
155	Morphine	57-27-2	1 ^d	0.2410	42.6	1.0325	90.7
158	Estrone	53-16-7	11	0.3033	73.9	0.5187	246.6
160	Estriol	50-27-1	10	0.5916	50.8	1.1082	2304.8
163	Testosterone	58-22-0	16	0.3731	119.0	0.4645	200.1
173	Haloperidol	52-86-8	16	0.6629	368.5	1.2222	2067.5
181	Hydrocortisone	50-23-7	13	0.7106	510.9	0.8267	1381.8
Average ^e				0.4303	227.3	0.7419	1099.8

^aAcetaminophen solubility data taken at 303.15 K

^bParameterized in Chen⁴⁰ with 14 solvents.

^cParameterized in Chen³⁹ with 4 solvents.

^dOnly one solubility value available for morphine at 298.15 K in Marrero and Abildskov²⁷

^eAverage excludes Aspirin-2

5.6 Application Guidelines and Heuristics of COSMO-SAC Solid Solubility

Predictions

We discuss several trends and factors we find with COSMO-SAC solubility predictions. We look at aqueous and n-octanol solubilities and their affecting factors such as size, solvent-solute similarity, etc., nitrogen-containing compound solubility, general accuracy, and model sensitivity factors. We see some similarities between the heuristics discussed in Section 4.7 and our findings here, which apply only to solubility.

5.6.1 Aqueous Solubility

Klamt⁷⁶ reports satisfactory solubility predictions of hydrocarbons in aqueous binary systems using the COSMOtherm program developed by COSMOlogic and released in 2002. The test set included aqueous systems of alkanes, alkylbenzenes, alkylcyclohexanes, and alkenes. Given this observation, we study aqueous pure-solvent solubility for 147 of the 206 solutes in the VT-2006 Solute Sigma Profile Database. This includes 458 aqueous pure-solvent solubility literature values. We find that all but two experimental values for aqueous solubility are below 10 solute mole percent. As one might expect from this composition range, the COSMO-SAC model predicts aqueous solubility less accurately than that the average accuracy for all predicted solubilities. Using the exchange energy defined by Lin and Sandler⁶, we calculate a RMSE of 1.0409 and an AA%E of 6324%, which is slightly less than a two-fold increase in average error. We observe an even greater increase in error when we use the Mathias et al¹³ exchange energy, RMSE = 1.4303, AA%E = 44273%. We should also point out that we predict aqueous solubility for 14 of 19 solutes listed in Table 5.3. Figure 5.20 shows the COSMO-SAC predicted aqueous solubilities compared with their experimental values.

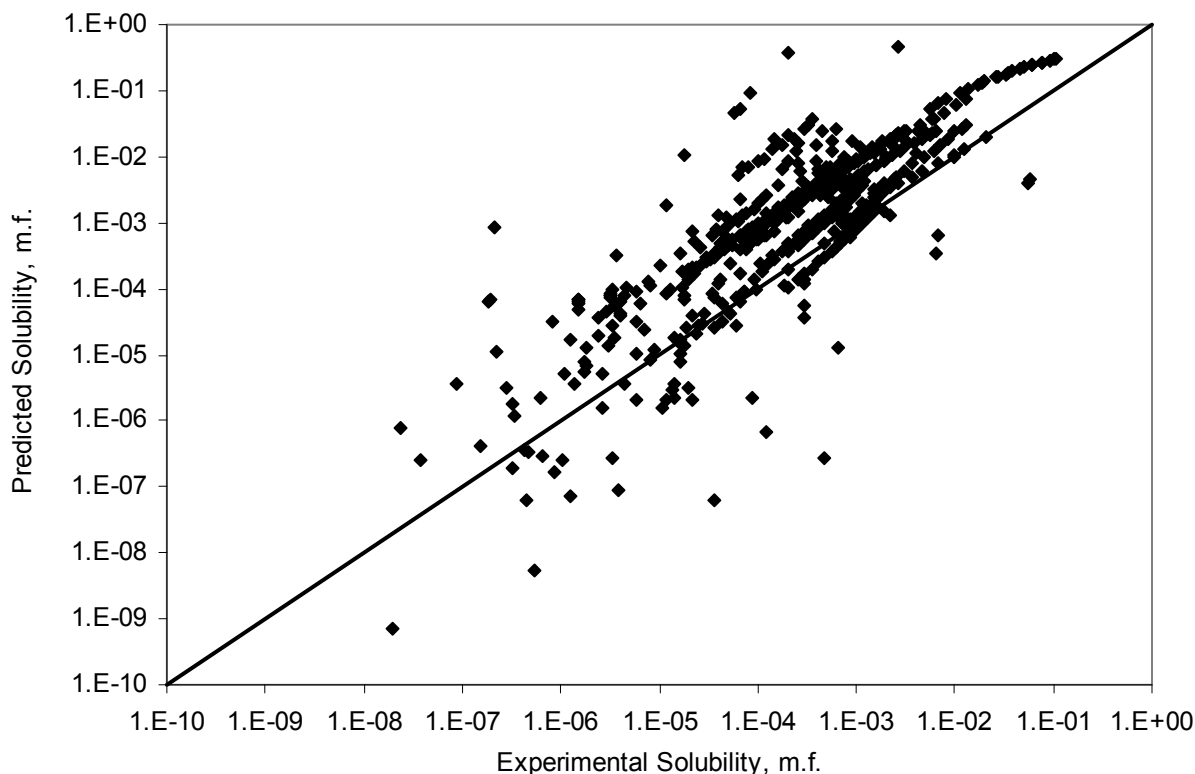


Figure 5.20: Predicted solubility in pure water using the exchange energy defined by Lin and Sandler⁶ compared with literature values for 144 solutes and 438 solubility points on a logarithmic base 10 scale.²⁷

Most of the solutes in this database are large, hydrophobic, and contain both polar and non-polar functional groups. Therefore, it is difficult to discern trends solely based on one factor without the influence of another. For instance, the number of pure hydrocarbon solutes with literature values for aqueous solubility in our database is too small to make general statements about accuracy. Table A.8 in Appendix A details the predicted aqueous solubility error for each solute.

5.6.2 n-Octanol Solubility

We present a brief summary of predicted COSMO-SAC solubilities for 52 solutes in n-octanol. We have a smaller sample size for n-octanol solubility than aqueous solubility, but we observe far fewer outliers, only one solute, niflumic acid (VTSOL-136), has an AA%E error exceeding the cutoff value. Summarizing the prediction error of the other solutes in n-octanol, we find that the RMSE and AA%E values are comparable to the overall model error when excluding outliers. We show these error measurements in Table 5.12.

Table 5.12: Summary and comparison of solubility prediction error in n-octanol using both exchange energy expressions for 52 solutes at various temperatures.²⁷

Exchange Energy	W_{hb} [Lin 2002]	W_{hb} [Mathias 2002]
RMSE	0.5640	0.6605
AA%E	312.4	582.5

We find that COSMO-SAC predicts solubility in n-octanol more accurately than aqueous solubility, but we do study fewer solutes for n-octanol than for water. N-octanol also has both hydrophilic and hydrophobic segments, which may explain the improvement in error. Ultimately, we cannot differentiate the source of the improvement in model accuracy from its potential factors, molecular weight, molecular rigidity, hydrophobicity, and hydrophilicity. We show COSMO-SAC predicted n-octanol pure-solvent solubilities compared with their respective literature values in Figure 5.21. In Appendix A, Table A.9 summarizes the COSMO-SAC prediction errors for all solutes in n-octanol as a pure solvent.

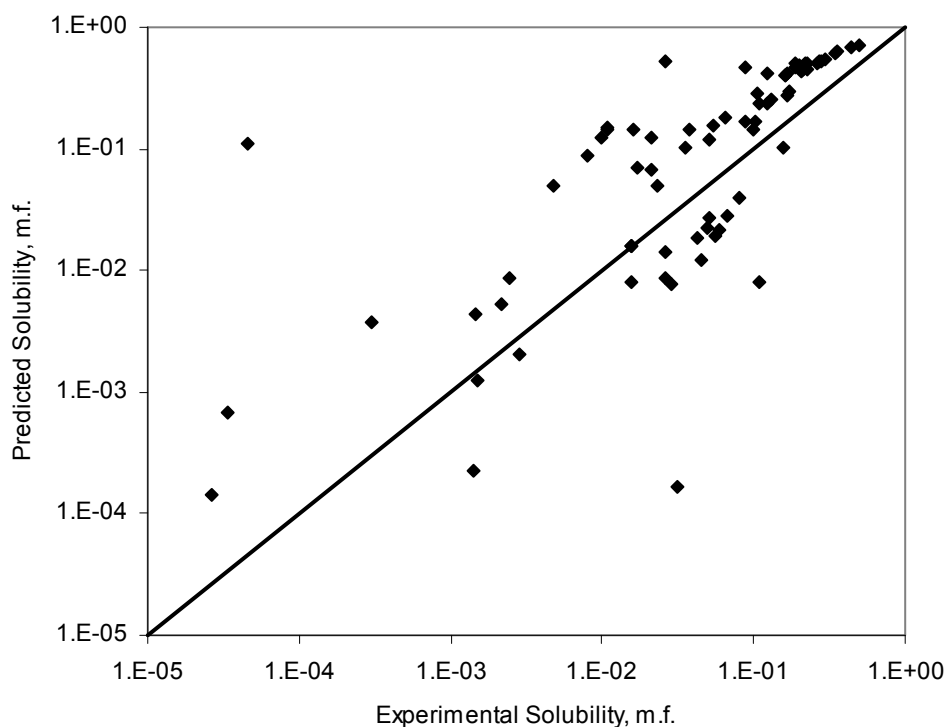


Figure 5.21: Predicted solubility in pure n-octanol using the exchange energy defined by Lin and Sandler⁶ compared with literature values for 52 solutes and 73 solubility points on a logarithmic base 10 scale.²⁷

5.6.3 Solvent Conformational Effects on Solubility in Ethanol and Cyclohexane

Using the same conformations of ethanol and cyclohexane from Section 4.5 for this comparison, we examine the effect of conformational isomerism of the solvent on predicted solid-liquid equilibrium behavior. We use the COSMO-SAC model and the solubility equation (2.30) to predict ethanol solubility for 50 solutes at various temperatures and to predict cyclohexane solubility for 39 solutes at various temperatures. Prediction accuracy improves for 39 of 50 solutes in ethanol and 29 of 39 solutes in cyclohexane, but solubility predictions in cyclohexane are much less sensitive to conformational changes than solubility predictions in ethanol. We see the disparity between predictions using the two conformations of ethanol and cyclohexane in Figure 5.22 and Figure 5.23, respectively.

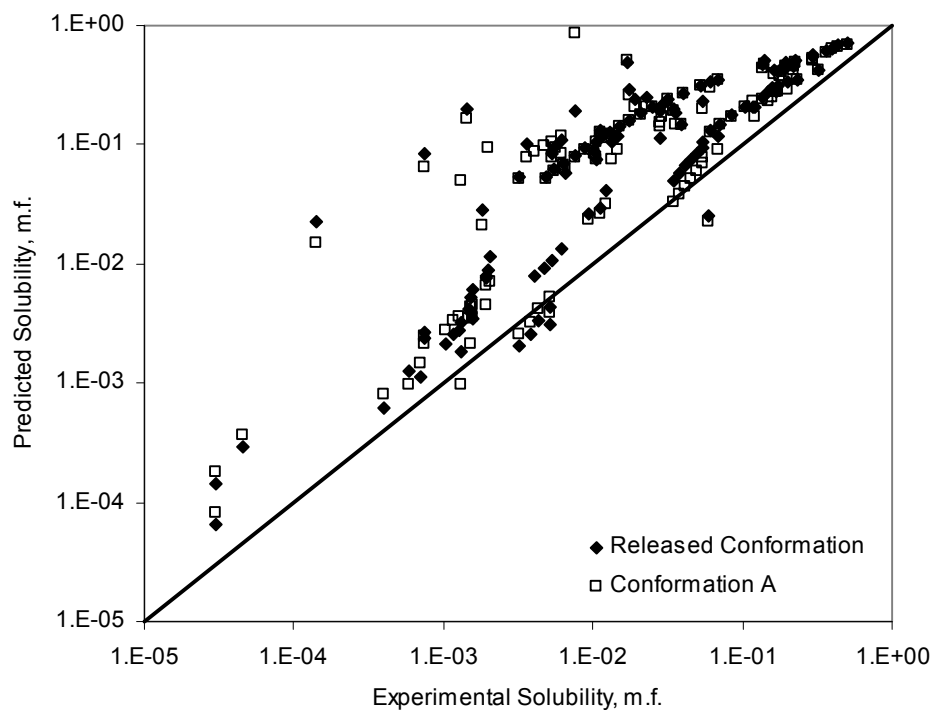


Figure 5.22: Predicted solubility in pure ethanol for 50 solutes at various temperatures using Lin and Sandler exchange energy expression compared with their experimental values on a logarithmic base 10 scale.²⁷

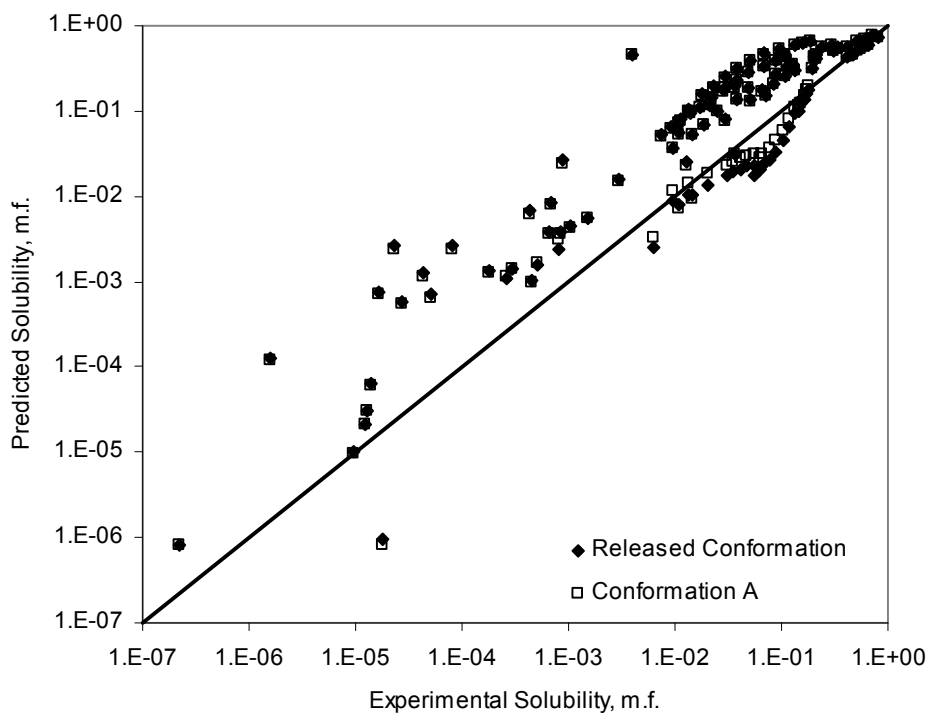


Figure 5.23: Predicted solubility in pure cyclohexane for 39 solutes at various temperatures using Lin and Sandler exchange energy expression compared with literature values on a logarithmic base 10 scale.²⁷

We also see that solubility predictions in cyclohexane are fairly insensitive to using different exchange energy expressions, but solubility predictions in ethanol improve when we use the exchange energy expression defined by Lin and Sandler⁶. COSMO-SAC predicts solubility in cyclohexane and in ethanol with similar accuracy, and we see an improvement for the majority of solutes when using conformation A for ethanol instead of the released ethanol conformation. The overall RMSE improves from 0.791 to 0.778 when using the ethanol conformation A instead of the released ethanol conformation, and the overall RMSE also improves from 0.711 to 0.708 when using the cyclohexane conformation A, the “chair” conformation, instead of the released cyclohexane conformation, the “boat” conformation. Table A.10 and Table A.11 in Appendix A give more detailed information regarding the calculated error for each solute in ethanol and cyclohexane.

5.6.4 Nitrogen-Containing Compound Solubility

Ninety-three compounds in the VT-2006 Solute Sigma Profile Database contain nitrogen in some form, whether as an amide, amine, pyridine-derivative, nitrile, or nitro functional group or some combination thereof. Other researchers document issues concerning the accuracy of nitrogen-containing compound predictions with COSMO-based methods.^{1,6,7} We analyze the effect of the presence of nitrogen atoms on the accuracy of pure-solvent COSMO-SAC solubility predictions. Nearly all solutes in the VT-2006 Solute Sigma Profile Database have more than one functional group, and approximately 60% of the 93 solutes have multiple nitrogen-containing functional groups as well as other functional groups. We compare the pure-solvent solubility for these solutes to experimental solubilities using both exchange energy expressions in Figure 5.24.

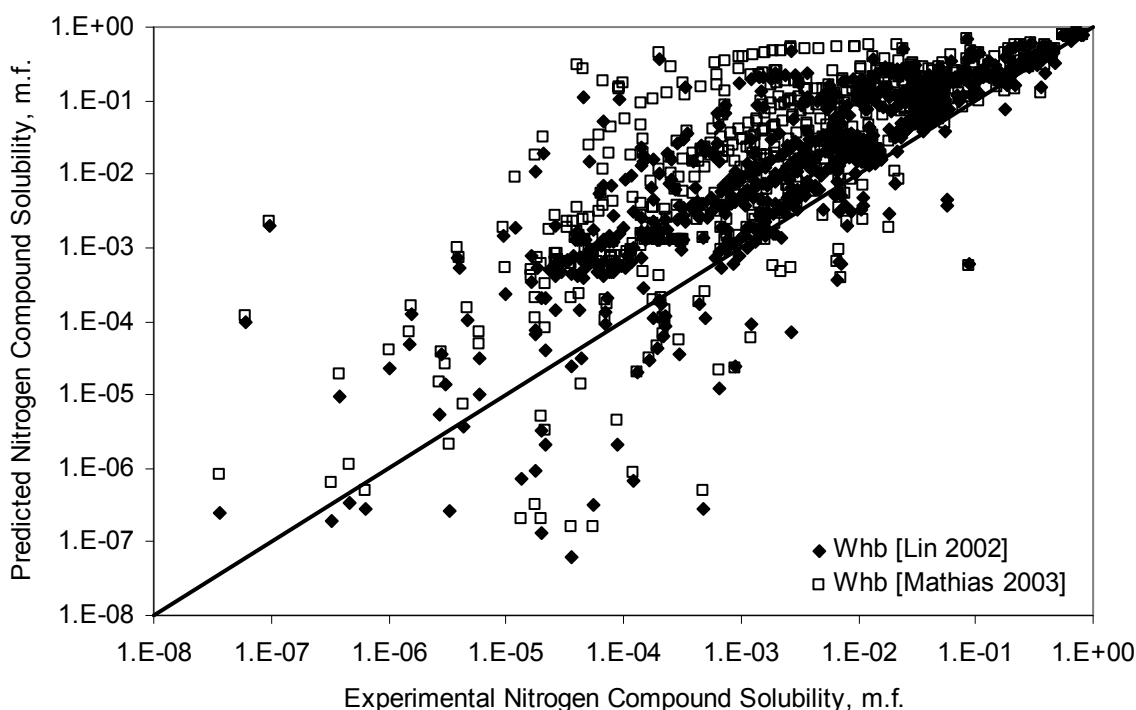


Figure 5.24: Predicted nitrogen-containing solubility for both exchange energy expressions compared with literature values for 93 solutes (664 solubility points) on a logarithmic base 10 scale.²⁷

When we compare the RMSE and AA%E of the nitrogen-containing and nitrogen-free solutes, and we find that COSMO-SAC predicts solubility for nitrogen-containing solutes significantly less accurate than solubility of nitrogen-free solutes. In Table 5.13 and Figure 5.24, we see an increase in error, similar to the error difference in Table 5.2, when using the Mathias et al¹³ defined exchange energy for most solutes. Using the exchange energy expression defined by Lin and Sandler⁶ generally improves solubility predictions for the majority of the nitrogen-containing solutes, 73 of 93. However, there are exceptions, like most of the solutes containing a sulfur atom, 7 of these 10 sulfur-containing compounds favor using the exchange energy expression defined by Mathias et al.¹³

Table 5.13: COSMO-SAC pure-solvent predicted solubility error summary for nitrogen-containing and nitrogen free solutes. Compare with sample set no. 1 in Table 5.2.

Sample Set	Literature Points	Solute Solvent Pairs	No. Solutes	W _{hb} [Lin 2002]		W _{hb} [Mathias 2002]	
				RMSE	AA%E	RMSE	AA%E
Nitrogen-Free	1770	919	101	0.7851	4461.9	0.9843	24610.1
Nitrogen-Containing	664	437	93	1.0360	6240.7	1.2547	17529.5

There are significantly more outliers in the nitrogen-containing solute set, which contributes to its greater than average calculated error. To further determine how a specific functional nitrogen functional group behaves, we categorize the 93 solutes by their nitrogen-containing functional group, amides, amines, nitriles, nitro, pyridines, aminosulfonyls, and multiple nitrogen functional groups. We show generic structures for these nitrogen functional groups in Figure 5.25.

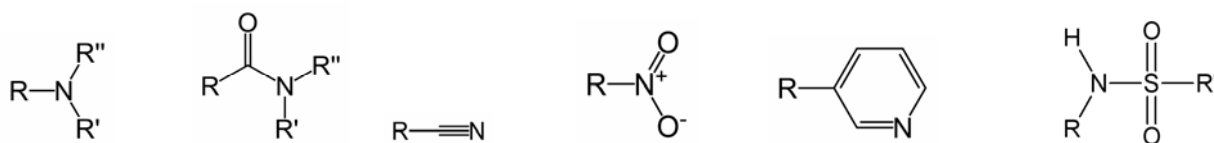


Figure 5.25: General chemical structures for the following functional groups in this order: Amine, Amide, Nitrile, Nitro, Pyridine, Aminosulfonyl

After calculating the RMSE and AA%E for each functional group category, we find that the prediction error for solutes with a single amide, amine, or nitro functional group is comparable to the average error for all literature values, see Table 5.14. We calculate the greatest errors for pyridine derivatives, aminosulfonyl groups, and multiple nitrogen functional group molecules.

Table 5.14: Error summary of the predicted solubility for each nitrogen-containing solute categorized by functional group. Each error measurement weights each solute/solvent pair equally, and each solute equally. Compare with the sample set no. 1 error values from Table 5.2.

Functional Group	Solute	Literature Points	W _{hb} [Lin 2002]		W _{hb} [Mathias 2003]	
			RMSE	AA%E	RMSE	AA%E
Amine	12	151	0.7557	1091.2	1.0184	5204.2
Amide	5	120	0.7787	845.2	1.3008	21636.1
Nitrile	1	3	0.1606	37.7	0.4517	183.2
Nitro	5	27	0.7908	1030.2	1.4277	6954.6
Pyridine	14	42	1.3167	15532.0	1.6651	31300.1
Aminosulfone	11	42	0.9376	3448.9	0.9320	3222.4
Multiple Groups	57	324	0.9782	4869.5	1.1334	17204.9

We include a detailed list of the prediction error for each nitrogen-containing solute in Table A.12 in Appendix A.

5.6.5 Comparison of Exchange Energy Expressions of Lin and Sandler⁶ and Mathias et al¹³

Of the 2434 pure-solvent solubility literature points and 1356 solute/solvent pairs, we see an improvement in predicting solute mole fractions for 444 solute/solvent pairs. These pairings include 106 of the 160 total solvents and 102 of 194 total solutes. We identify several solvents and solutes which have a higher probability for consistent improvement for predicted pure-solvent solubility using the Mathias et al¹³ exchange energy definition. Each solvent generates improved predictions for more than 50% of the solute/solvent pairings and constitutes a 5% improvement or greater in the overall RMSE per solvent or solute. We recommend using the Mathias et al¹³ exchange energy definition for 11 solutes and 14 solvents, listed in Table 5.15.

Table 5.15: Recommended solutes and solvents for use with the Mathias et al¹³ exchange energy expression.

VT-2006 Solute Sigma Profile Database			VT-2005 Sigma Profile Database		
Index No.	Solute Name	CAS-RN	Index No.	Solvent Name	CAS-RN
088	Acetylsalicylic Acid	50-78-2	0242	Benzene	71-43-2
089	Sulfamethoxazole	723-46-6	0243	Toluene	108-88-3
101	Ephedrine	299-42-3	0583	Acetic-Acid	64-19-7
105	Sulfisomidine	515-64-0	0584	Propionic-Acid	79-09-4
115	Sulfamethazine	57-68-1	0638	Methyl-Acetate	79-20-9
126	Thioxanthone	492-22-8	0639	Ethyl-Acetate	141-78-6
129	9,10-Anthraquinone	84-65-1	0641	N-Butyl-Acetate	123-86-4
130	Lidocaine	137-58-6	0728	1,4-Dioxane	123-91-1
134	Prostaglandin	363-24-6	0749	Triethylene-Glycol-Dimethyl-Ether	112-49-2
135	Progesterone	57-83-0	0785	Dichloromethane	75-09-2
138	Hydrocortisone	50-23-7	0786	Chloroform	67-66-3
			0807	Chlorobenzene	108-90-7
			0962	Pyridine	110-86-1
			1388	Formamide	75-12-7

5.6.6 Effect of Multicomponent Systems on Accuracy of Solubility Predictions

With the given sample set of 37 binary solvent systems, COSMO-SAC predicts solute mole fractions for 36 systems within comparable accuracy to the model predictions for solubility in a pure solvent. The model accuracy is also similar for ternary solvent systems, but we only study 3 systems, which is not a large enough sample to make general statement concerning accuracy. See Table 5.10 for the summarized error values of solubility predictions in mixed solvents. We find that conformational effects of both solutes and solvents play a role in the

overall accuracy of the model for binary and ternary solvent systems, similar to their effects on pure-solvent solubility predictions.

5.6.7 Overall Accuracy of COSMO-SAC as Solubility Predictor

The COSMO-SAC model systematically over-predicts the solute mole fraction, but this model is an improvement over using ideal solubility. Refer to Table 5.2 for an overall error summary of COSMO-SAC solubility predictions in pure solvents. We recommend looking at a similar representative system to the system of interest before using COSMO-SAC to model solubility and evaluating this model for individual solute/solvent pairs. For example, if a new compound contains an ester and amine functional group, find a similar molecule from our databases to serve as a representative chemical for how solubility predictions with this new compound may behave. Although the overall average error for a particular solute may be poor, particular solute solvent pairs may generate predictions with acceptable accuracy, or vice versa.

5.6.8 Sensitivity to Melting-point Temperature and Latent Heat of Fusion

We derive that error in ideal solubility is proportional to the exponential of the change in the latent heat of fusion $\Delta H_{fus}^{T_m}$ and the solute melting temperature T_m as shown in equation (5.3). However, this relationship does not hold for non-ideal solubility when using COSMO-SAC activity coefficients.

$$\begin{aligned} E_{x_{sol}}^{ID} &\propto e^{\Delta(\Delta H_{fus}^{T_m})} \\ E_{x_{sol}}^{ID} &\propto e^{\Delta T_m} \end{aligned} \tag{5.3}$$

When we examine the overall absolute average deviation of the predicted solute mole fraction from the literature value, we see a small linear increase in accuracy as a function of the latent heat of fusion over the entire sensitivity range, and a similar linear increase in accuracy for estimated melting temperatures greater than the reported melting temperature. However, when the estimated melting temperature is less than the literature value, we see a rapid decrease in overall accuracy. Figure 5.26 shows these trends. One potential reason for the increase in error

may stem from the system temperature exceeding the estimated normal melting temperature and therefore invalidating the solubility equation, $T > T_m$. We calculate that this condition affects roughly 17% of the pure-solvent literature points with a 15% decrease in the estimated melting temperature. With a 10% decrease in the estimated melting temperature, we see that the system temperature still exceeds the melting temperature for 5.5% of the pure-solvent literature points, and only 1.4% of the pure-solvent points for a 5% decrease in estimated melting temperature. The reference state is 0% change in both properties, which signifies the use of the literature values.²⁷ These RMSE values include all solubility points from our set of validation points.

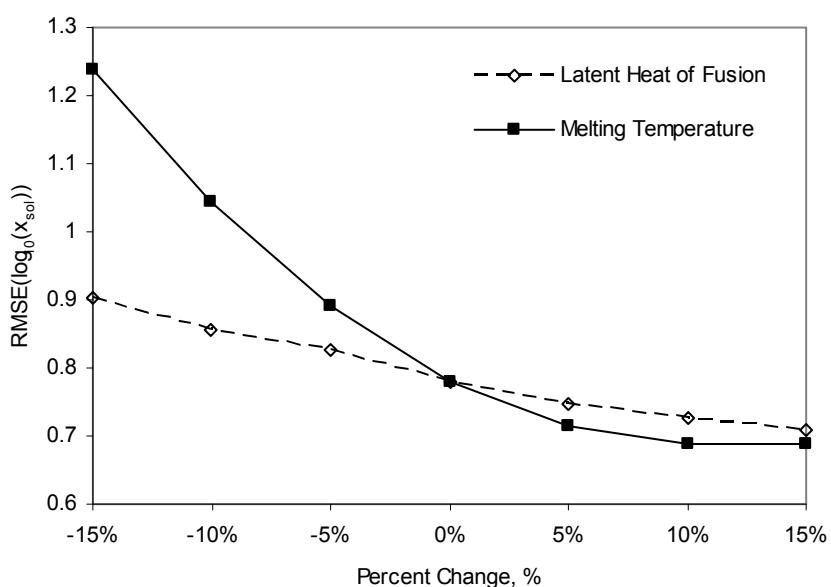


Figure 5.26: RMSE sensitivity of predicted solubilities as a function of normal melting-point temperature and the latent heat of fusion.

5.6.9 Effect of Conformation Pre-optimization with Amber 8 on Solubility

Predictions

We recommend Amber8 for generating initial molecular geometries for smaller molecules with conformational freedom for COSMO-SAC solubility modeling applications. In a contrary finding, only one Amber8 optimized conformation of the five conformational variation case studies presented in Section 5.3 generates the most accurate pure-solvent solubility predictions. The best conformation does not necessarily have the lowest condensed phase energy for these cases as well. In general, we recommend that users consider manually drawn

conformations as well as optimized initial structures from Amber8, MS Forcite Plus, or other pre-optimization tools when studying molecules with greater conformational freedom. We find that these tools may not converge on a structure which accurately describes the electronic signature of more than one functional group in close proximity to each other as we see in the aspirin conformational effects study.

6 Resources

We offer several resources on our research group website, www.design.che.vt.edu, which we list below.

1. Indices of the VT-2005 Sigma Profile Database, containing 1432 compounds, and VT-2006 Solute Sigma Profile Database, containing 206 compounds, which are searchable by CAS-RN, chemical formula, and name. The VT-2005 Sigma Profile Database index also includes the normal boiling point, and the pure component vapor pressure as predicted a revised COSMO-SAC-BP model⁵¹.
2. Procedures for generating sigma profiles using Accelrys' MS and using the FORTRAN programs to predict vapor-liquid and solid-liquid equilibria. These procedures include screen captures and sample outputs for reference.
3. Calculation outputs from the DMol3 module geometry optimization and energy calculation tasks and the COSMO calculation. These files include atomic coordinates, surface charges, etc. in *.OUTMOL and *.COSMO files and are viewable as simple text.
4. Executable FORTRAN90 programs and source code for calculation programs for sigma profile averaging and activity coefficient prediction for use with VLE and SLE systems.^{8,9}

Appendix B contains the FORTAN source code for executable programs listed above that use the COSMO-SAC model to calculate vapor-liquid and solid-liquid equilibrium behavior. Appendix C illustrates example OUTMOL and COSMO output and sigma profile for water (VT-1076) from the VT-2005 Sigma Profile database. We use the same format for all database entries.

7 Conclusions

With the completion of this work, we provide the means to predict vapor-liquid and solid-liquid (VLE and SLE) equilibrium behavior using the COSMO-SAC model developed by Lin and Sandler.^{5,6} We produce molecular-specific sigma profiles for 1670 organic compounds, including 1464 solvents and 206 pharmacologically-related solutes, and validate the compound's sigma profile by predicted pure component vapor pressure in the case of VLE and predicted solubility in pure solvent in the case of SLE. We study conformational isomerism, or the existence of alternative, low-energy, stable variations in molecular structure, and their effects on thermodynamic property prediction. We separate these compounds into two databases, VT-2005 Sigma Profile Database and the VT-2006 Solute Sigma Profile Database. We publish the VT-2005 Sigma Profile Database in *Industrial & Engineering Chemical Research* in 2006.⁹

The COSMO-SAC model predicts VLE behavior more accurately than SLE behavior with the current model parameters, but this model's strength lies in its ability to generate *a priori* property predictions. Excluding 23 of the total 194 solutes for which we compare solubility data in pure solvents to literature values as outliers, COSMO-SAC predicts solute mole fractions with an overall absolute error of 670% in mole fraction and a RMS error of 0.742, $\log_{10}(x_{\text{sol}})$ units, which is greater than the RMS error of the NRTL-SAC model, a regressed parameter model, developed by Chen and Song.³⁹ The accuracy of the COSMO-SAC model depends largely on the molecular conformation and the exchange energy expression. The use of pre-optimization tools, such as Amber8 and MS Forcite Plus, does improve COSMO-SAC solubility predictions by improving the initial molecular structure in most cases, especially with small, flexible molecules. However, a compound's sensitivity to changes in its sigma profile as a result of conformational variations is system-specific.

We confirm the issues other researchers observe regarding COSMO-SAC model applications involving nitrogen-containing compounds, and we study the effects of individual nitrogen functional groups on predicted solubility. We also discuss COSMO-SAC applicability as a solubility predictor in several common solvents, water, ethanol, cyclohexane, and n-octanol. Ultimately, the application guidelines and resources we provide should help future researchers improve the applicability and accuracy of the COSMO-SAC model.

8 Improvements and Future Work

We encourage further refinement of this work and its applications in several areas after its initial publication. First, we suggest improving molecular conformations by utilizing the energy minimization tools Amber8 or MS Forcite Plus for all molecules in the databases, beginning with alcohols and diols. We improve the hydroxyl hydrogen bond angles and lengths of several alcohols, including ethanol, by using Amber8 output as an initial guess for an optimized geometry. Secondly, we suggest developing an averaging algorithm for molecules with conformational flexibility, essentially incorporating the multiple conformations of a single molecule into a thermodynamic property prediction. There are specific issues that arise, such as the potential improvement of such an algorithm if all conformations produce a predicted property value less than (or greater than) the experimental value. The practicality of determining the ratios of actual conformations in nature is a second issue that arises.

In the area of solubility modeling, we recommend further investigation into modeling solubility for electrolytes, since this is a common method of drug delivery. Accelrys' DMol3 module now includes the capacity to perform geometry optimization tasks for charged molecules. We also suggest improving on the current model parameters. The original optimized parameter set by Klamt¹ is based on 642 data points for various properties, which includes vapor pressure and partition coefficients, but not solubility. This set of 642 data points includes small to medium-sized molecules,² which may or may not apply to many of the compounds listed in the VT-2006 Solute Sigma Profile Database. Such model parameters include the COSMO-SAC model parameters regarding hydrogen-bonding and segment interactions and the parameters which affect the Staverman Guggenheim contribution term, equation (2.3). For example, the coordination number ranges from 2 – 16, but is generally set to 10.0, the same value used by Lin and Sandler.^{5,6} Klamt sets the coordination number to 7.2 in COSMO-RS.⁴ If it is possible to compute a coordination number based on the components actually present in the system, this may improve the overall model accuracy for solubility modeling by improving the Staverman – Guggenheim contribution term.

9 Nomenclature

English Symbols

A	Coulomb Interaction Energy matrix
$AA\%E$	Absolute Average Relative Percent Error
a_{av}	Average segment surface area, Å^2
a_{eff}	Effective segment surface area, Å^2
A_i	Total molecular cavity surface area, Å^2
$A_i(\sigma_m)$	Area of segments with charge density σ , $e/\text{Å}^2$
a.u.	Atomic unit, Bohr radius, $5.2918 \times 10^{-11} \text{ Å}$
c_{hb}	Hydrogen bonding Constant, $\text{kcal Å}^4 \text{ mol}^{-1} e^{-2}$
ΔC_p	Constant Pressure Heat Capacity, J/mol K
d_{mn}	Distance between surface segment m and n , Å
e	Elementary charge, 1.6022×10^{-19} coulomb
E_h	Hartree, an atomic unit of energy, $4.35974417 \times 10^{-18} \text{ J}$
f_o	Reference State Fugacity
f_{pol}	Polarizability factor, 0.64
f^S	Fugacity of pure solid
f^L	Fugacity of a pure, subcooled liquid
ΔG	Gibbs Free Energy Change, kcal/mol
ΔG^{IS}	Gibbs Ideal Solvation Energy, kcal/mol
ΔG^{*cav}	Cavity Formation Free Energy, kcal/mol
ΔG^{*chg}	Charging Free Energy, kcal/mol
ΔG^{*res}	Restoring Free Energy, kcal/mol
ΔG^{*sol}	Solvation Free Energy, kcal/mol
ΔH	Enthalpy Change, kJ/mol
$\Delta H_{fus}^{T_i}$	Heat of Fusion at the Triple Point Temperature, kJ/mol
$\Delta H_{fus}^{T_m}$	Heat of Fusion at the Normal Melting Point Temperature, kJ/mol
l_i	SG combinatorial term parameter
m_i	Group-contribution parameter for estimating melting temperature
$M_j(tfp2j)$	2 nd order group-contribution parameter
n_i	Total number of segments on the surface of the molecular cavity
$n_i(\sigma)$	Number of segments with charge density σ
n_i^t	Group-contribution parameter for estimating melting temperature
$N_k(tfpk)$	1 st order group-contribution parameter for estimating melting point temperatures
$N_k(hmk)$	1 st order group-contribution parameter for estimating latent heats of fusion

$p_i(\sigma)$	Sigma profile, probability of segment i having a charge density σ
$p_i'(\sigma)$	Area-weighted Sigma Profile of component i , \AA^2
$p_S(\sigma)$	Sigma Profile of Mixture S
q	Standard area parameter, 79.53 \AA^2
q^*	Ideal screening charge, e
q_{avg}	Average screening charge, e
q_i	Normalized surface area parameter for SG combinatorial term
r	Standard volume parameter, 66.69 \AA^2
r_{av}	Average segment radius, \AA
r_{eff}	Effective segment radius, \AA
r_i	Normalized volume parameter for SG combinatorial term
r_n	Circular segment radius, \AA
R	Ideal Gas Constant, $0.001987 \text{ kcal mol}^{-1} \text{ K}^{-1}$, $8.314 \text{ kJ kmol}^{-1} \text{ K}^{-1}$
$RING$	Number of monocyclic fused ring systems
ΔS	Entropy change, J/K
$\Delta S_{fus}^{T_t}$	Entropy of Fusion at the Triple Point Temperature, J/K
$\Delta S_{fus}^{T_m}$	Entropy of Fusion at the Normal Melting Point Temperature, J/K
ΔS_m^{conf}	Conformational Entropy Contribution to a solid-liquid phase change, J/K
ΔS_m^{pos}	Positional Entropy Contribution to a solid-liquid phase change, J/K
ΔS_m^{rot}	Rotational Entropy Contribution to a solid-liquid phase change, J/K
ΔS_m^{tot}	Total Entropy change associated with a solid-liquid phase change, J/K
$SP3$	Number of SP3 hybridized atoms
$SP2$	Number of SP2 hybridized atoms
T	System Temperature, K
T_m	Normal Melting Point Temperature, K
T_t	Triple Point Temperature, K
V_i	Molecular cavity volume, \AA^3
$\Delta W(\sigma_m, \sigma_n)$	Exchange Energy between segments σ_m and σ_n , kcal mol^{-1}
ΔW_{hb}	Hydrogen-bonding contribution of the Exchange Energy, kcal mol^{-1}
W	Group-contribution parameter set to 1 in presence of 2 nd order parameter $M_j(tfp2.j)$
x_{sol}	Solute mole fraction, m.f.
x_i	Mole fraction of component i
$x_{j,I}$	Segment mole fraction of component i
z	Coordination Number

Greek Symbols

α	Model constant, $\text{\AA}^4 \text{ kcal } e^{-2} \text{ mol}^{-2}$
α'	Misfit energy constant, $\text{\AA}^4 \text{ kcal } e^{-2} \text{ mol}^{-2}$
γ_i	Activity coefficient of solute i
γ_I^C	Combinatorial contribution to NRTL-SAC activity coefficient
γ_I^R	Residual contribution to NRTL-SAC activity coefficient
$\gamma_{i/S}^{SG}$	Staverman-Guggenheim activity coefficient of solute i in a solvent S
γ_{sol}	Solute activity coefficient
$\Gamma_i(\sigma_m)$	Segment activity coefficient of segment σ_m in a pure liquid i
$\Gamma_m^{lc,I}$	Segment local composition interaction contribution of component I
Γ_m^{lc}	Segment local composition interaction contribution
$\Gamma_s(\sigma_m)$	Segment activity coefficient of segment σ_m in a solvent S
ε	Dielectric constant
ϵ_o	Permittivity of free space, 2.395E-04
θ_i	Composition-weighted volume fraction
σ	charge density, $e/\text{\AA}^2$
σ^*	surface segment charge density from COSMO calculation output
σ^r	Molecular rotational symmetry number
σ_{acc}	Hydrogen acceptor segment
σ_{don}	Hydrogen donor segment
σ_{hb}	Hydrogen-bonding cutoff value, $0.0084 e/\text{\AA}^2$
σ_{hb}^n	New hydrogen-bonding cutoff value from Mathias defined exchange energy, $0.0084 e/\text{\AA}^2$
σ_m	segment charge density of segment m
τ	Number of torsional angles
ϕ	Molecular flexibility number
ϕ_i	Composition-weighted surface area fraction
Φ_i	Potential due to the charge distribution of the solute i
$\Phi(q^*)$	Potential as a function of the ideal screening charge q^*
Φ_{tot}	Total potential on the cavity surface

10 References

1. Klamt, A., Conductor-Like Screening Model for Real Solvents: A New Approach to the Quantitative Calculations of Solvation Phenomena. *J. Phys. Chem.* **1995**, 99, 2224.
2. Klamt, A., COSMO and COSMO-RS. In *Encyclopedia of Computational Chemistry*, Schleyer, P. v. R., Ed. Chichester, 1998.
3. Eckert, F.; Klamt, A., Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AICHE J.* **2002**, 48, 369.
4. Klamt, A.; Eckert, F., COSMO-RS: A Novel and Efficient Method for the *a priori* Prediction of Thermophysical Data of Liquids. *Fluid Phase Equilibria* **2000**, 172, 43.
5. Lin, S. T. Quantum Mechanical Approaches to the Prediction of Phase Equilibria: Solvation Thermodynamics and Group Contribution Methods. PhD. Dissertation, University of Delaware, Newark, DE, 2000.
6. Lin, S. T.; Sandler, S. I., *A Priori* Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.* **2002**, 41, 899.
7. Panayiotou, C., Equation of State Models and Quantum Mechanical Calculations. *Ind. Eng. Chem. Res.* **2003**, (42), 1495.
8. Oldland, R. Predicting Phase Equilibria Using COSMO-Based Thermodynamic Models and the VT-2004 Sigma-Profile. M.S., Virginia Polytechnic Institute and State University, Blacksburg, VA, 2004.
9. Mullins, E.; Oldland, R.; Liu, Y. A.; Wang, S.; Sandler, S. I.; Chen, C. C.; Zwolak, M.; Seavey, K. C., Sigma-profile database for using COSMO-based thermodynamic methods. *Industrial & Engineering Chemistry Research* **2006**, 45, (12), 4389-4415.
10. Klamt, A.; Schüürmann, 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.
11. Lin, S. T.; Sandler, S. I., Infinite Dilution Activity Coefficients from Ab Initio Solvation Calculations. *AICHE J.* **1999**, 45, 2606.
12. Lin, S. T.; Sandler, S. I., Prediction of Octanol-Water Partition Coefficients Using Group Contribution Solvation Model. *Ind. Eng. Chem. Res.* **1999**, 38, 4081.
13. Mathias, P.; Lin, S. T.; Song, Y.; Chen, C.-C.; Sandler, S. I. In *Phase-Equilibrium Predictions for Hydrogen-Bonding Systems from a New Expression for COSMO Solvation Models*, AIChE Annual Meeting, Indianapolis, IN, 2002; Indianapolis, IN, 2002.
14. Eckert, F.; Klamt, A., Validation of the COSMO-RS Method: Six Binary Systems. *Ind. Eng. Chem. Res.* **2001**, 40, 2371.
15. Putnam, R.; Klamt, A.; Taylor, R.; Eckert, F.; Schiller, M., Prediction of Infinite Dilution Activity Coefficients Using COSMO-RS. *Ind. Eng. Chem. Res.* **2003**, 42, 3635.
16. Wang, S.; Stubbs, J. M.; Siepman, J. I.; Sandler, S. I., Effects of Conformational Distributions on Sigma Profiles in COSMO Theories. *J. Phys. Chem.* **2005**, 109, 11285-11294.
17. Koch, W.; Holthausen, M. C., *Chemist's Guide to Density Functional Theory*. 2nd ed.; Weinheim, Germany, 2001.
18. Delley, B., An All-Electron Numerical Method for Solving the Local Density Functional for Polyatomic Molecules. *J. Chem. Phys.* **1990**, 92, 508.
19. Delley, B., Analytic Energy Derivatives in the Numerical Local-Density-Functional Approach. *J. Chem. Phys.* **1991**, 94, 7245.
20. Delley, B., Modern Density Functional Theory: A Tool for Chemistry. In *Theoretical and Computational Chemistry*, Seminario, J. M.; Politzer, P., Eds. Elsevier Science Publisher: Amsterdam, 1995; Vol. 2.
21. Andzelm, J.; Kolmel, C.; Klamt, A., Incorporation of Solvent Effects into Density Functional Calculations of Molecular Energies and Geometries. *J. Chem. Phys.* **1995**, 103, 9312.
22. Thijssen, J. M., *Computational physics*. Cambridge University Press: Cambridge, U.K. ; New York, 1999; p xiii, 546 p.
23. Levine, I. N., *Quantum chemistry*. 5th ed.; Prentice Hall: Upper Saddle River, N.J., 2000; p x, 739 p.
24. Harrison, N. M., An Introduction to Density Functional Theory. In *Computational Materials Science*, Catlow; Kotomin, Eds. IOS Press: 2003; Vol. 187.
25. Ben-Naim, A., *Solvation Thermodynamics*. Plenum Press: New York, 1987.
26. Prausnitz, J. M.; Lichtenthaler, R. N.; Azevedo, E. G. d., *Molecular Thermodynamics of Fluid-Phase Equilibria*. 3rd ed.; Prentice Hall: New Jersey, 1999.

27. Marrero, J.; Abildskov, J., *Solubility and Related Properties of Large Complex Chemicals, Part 1: Organic Solutes ranging from C4 to C40*. DECHEMA: 2003; Vol. XV.
28. Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P., *The properties of gases and liquids*. 5th ed.; McGraw-Hill: New York, 2001; p 1 v. (various pagings).
29. Joback, K. G.; Massachusetts Institute of Technology. Dept. of Chemical Engineering. A unified approach to physical property estimation using multivariate statistical techniques. Thesis M.S. --Massachusetts Institute of Technology Dept. of Chemical Engineering 1984., 1984.
30. Joback, K. G.; Reid, R. C., Estimation of Pure-Component Properties from Group-Contributions. *Chemical Engineering Communications* **1987**, 57, (1-6), 233-243.
31. Constantinou, L.; Gani, R., New Group-Contribution Method for Estimating Properties of Pure Compounds. *Aiche Journal* **1994**, 40, (10), 1697-1710.
32. Constantinou, L.; Gani, R.; Oconnell, J. P., Estimation of the Acentric Factor and the Liquid Molar Volume at 298-K Using a New Group-Contribution Method. *Fluid Phase Equilibria* **1995**, 103, (1), 11-22.
33. Yalkowsky, S. H.; Dannenfelser, R. M.; Myrdal, P.; Simamora, P.; Mishra, D., Unified Physical Property Estimation Relationships (Upper). *Chemosphere* **1994**, 28, (9), 1657-1673.
34. Krzyzaniak, J. F.; Myrdal, P. B.; Simamora, P.; Yalkowsky, S. H., Boiling-Point and Melting-Point Prediction for Aliphatic, Non-Hydrogen-Bonding Compounds. *Industrial & Engineering Chemistry Research* **1995**, 34, (7), 2530-2535.
35. Zhao, L. W.; Yalkowsky, S. H., A combined group contribution and molecular geometry approach for predicting melting points of aliphatic compounds. *Industrial & Engineering Chemistry Research* **1999**, 38, (9), 3581-3584.
36. Dannenfelser, R. M.; Yalkowsky, S. H., Estimation of entropy of melting from molecular structure: A non-group contribution method. *Industrial & Engineering Chemistry Research* **1996**, 35, (4), 1483-1486.
37. Chickos, J. S.; Acree, W. E., Estimating solid-liquid phase change enthalpies and entropies. *Journal of Physical and Chemical Reference Data* **1999**, 28, (6), 1535-1673.
38. Chickos, J. S.; Nichols, G.; Ruelle, P., The estimation of melting points and fusion enthalpies using experimental solubilities, estimated total phase change entropies, and mobile order and disorder theory. *Journal of Chemical Information and Computer Sciences* **2002**, 42, (2), 368-374.
39. Chen, C. C.; Song, Y. H., Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Industrial & Engineering Chemistry Research* **2004**, 43, (26), 8354-8362.
40. Chen, C. C., A Segment-Based Local Composition Model for the Gibbs Energy of Polymer-Solutions. *Fluid Phase Equilibria* **1993**, 83, 301-312.
41. Chen, C. C.; Song, Y. H., Extension of nonrandom two-liquid segment activity coefficient model for electrolytes. *Industrial & Engineering Chemistry Research* **2005**, 44, (23), 8909-8921.
42. Chen, C. C.; Crafts, P. A., Correlation and prediction of drug molecule solubility in mixed solvent systems with the Nonrandom Two-Liquid Segment Activity Coefficient (NRTL-SAC) model. *Industrial & Engineering Chemistry Research* **2006**, 45, (13), 4816-4824.
43. Case, D. A.; Darden, T. A.; III, T. E. C.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Godlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *Amber 8*, University of California, San Francisco: San Francisco, CA, 2004.
44. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **1995**, 91, (1-3), 1-41.
45. Accelrys, *MS Modeling Getting Started, Release 4.0*. Accelrys Software Inc.: San Diego, CA, January 2007.
46. Becke, A. D., Density Functional Calculations of Molecular Bond Energies. *J. Chem. Phys.* **1986**, 84, 4525.
47. Perdew, J. P., Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev.* **1986**, 33, 8822.
48. Vosko, S. J.; Wilk, L.; Nusair, M., Accurate Spin-Dependant Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, 58, 1200.
49. Klamt, A., *COSMO-RS, From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*. Elsevier B.V.: Amsterdam, 2005.
50. Wang, S., Private Communication. In Mullins, E., Ed. University of Delaware: Newark, DE, 2005.

51. Wang, S.; Lin, S. T.; Chang, J.; Goddard, W. A.; Sandler, S. I., Application of the COSMO-SAC-BP solvation model to predictions of normal boiling temperatures for environmentally significant substances. *Industrial & Engineering Chemistry Research* **2006**, 45, (16), 5426-5434.
52. Giles, N. F.; Wilson, G. M., Phase equilibria on seven binary mixtures. *Journal of Chemical and Engineering Data* **2000**, 45, (2), 146-153.
53. Carmona, F. J.; Bhethanabotla, V. R.; Campbell, S. W.; Gonzalez, J. A.; de la Fuente, I. G.; Cobos, J. C., Thermodynamic properties of (n-alkoxyethanols plus organic solvents). XII. Total vapour pressure measurements for (n-hexane, n-heptane or cyclohexane+2-methoxyethanol) at different temperatures. *Journal of Chemical Thermodynamics* **2001**, 33, (1), 47-59.
54. Martin, M. C.; Cocero, M. J.; Mato, R. B., Vapor-Liquid-Equilibrium Data at 298.15-K for Binary-Systems Containing Methyl Acetate or Methanol with 2-Methoxyethanol or 2-Ethoxyethanol. *Journal of Chemical and Engineering Data* **1994**, 39, (3), 535-537.
55. Jonasson, A.; Savoia, M.; Persson, O.; Fredenslund, A., Isothermal Vapor-Liquid-Equilibrium Data for Ether Plus Glycol, Chloroalkene Plus Glycol, Epoxy Ether Plus Alkane, Epoxy Ether Plus Alkene, and Epoxy Ether Plus Chloroalkane Systems. *Journal of Chemical and Engineering Data* **1994**, 39, (1), 134-139.
56. Nienhaus, B.; Wittig, R.; Bolts, R.; de Haan, A. B.; Niemann, S. H.; Gmehling, J., Vapor-liquid equilibria at 453.25 K and excess enthalpies at 363.15 K and 413.15 K for mixtures of benzene, toluene, phenol, benzaldehyde, and benzyl alcohol with benzyl benzoate. *Journal of Chemical and Engineering Data* **1999**, 44, (2), 303-308.
57. Gmehling, J.; Onken, U.; Arlt, W., *Vapor-liquid equilibrium data collection*. Dechema; Distributed by Scholium International: Frankfurt/Main Flushing, N.Y., 1977; p v.
58. Eckert, F.; Klamt, A., Prediction of Halocarbon Thermodynamics with COSMO-RS. *Fluid Phase Equilibria* **2003**, 210, 117-141.
59. Kolar, P.; Shen, J.-W.; Tsuboi, A.; Ishikawa, T., Solvent Selection for Pharmaceuticals. *Fluid Phase Equilibria* **2002**, 194-197, 771-782.
60. Lin, S. T.; Chang, J.; Wang, S.; III, W. A. G.; Sandler, S. I., Prediction of Vapor Pressures and Enthalpies of Vaporization Using a COSMO Solvation Model. *J. Phys. Chem.* **2004**, 108, 7429-7439.
61. Sandler, S. I., Private Communication. In Mullins, E., Ed. University of Delaware: Newark, DE, 2004.
62. Gaglione, A. Multiscale Modeling of an Industrial Nylon-6 Leacher. Thesis M.S., Virginia Polytechnic Institute and State University, Blacksburg, VA, 2007.
63. Klamt, A., Comments on "Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods". *Industrial & Engineering Chemistry Research* **2005**, 44, (17), 7042-7042.
64. Grensemann, H.; Gmehling, J., Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. *Industrial & Engineering Chemistry Research* **2005**, 44, (5), 1610-1624.
65. Constantinescu, D.; Klamt, A.; Geană, D., Vapor-liquid equilibrium prediction at high pressure using activity coefficients at infinite dilution from COSMO-type methods. *Fluid Phase Equilibria* **2005**, 231, 231-238.
66. Klamt, A., Prediction of the mutual solubilities of hydrocarbons and water with COSMO-RS. *Fluid Phase Equilibria* **2003**, 206, (1-2), 223-235.
67. Klamt, A.; Eckert, F.; Hornig, M.; Beck, M.; Bürger, T., Prediction of Aqueous Solubility of Drugs and Pesticides with COSMO-RS. *J. Comput. Chem.* **2001**, 23, 275-281.
68. Oleszek-Kudlak, S.; Grabda, M.; Shibata, E.; Eckert, F.; Nakamura, T., Application of the conductor-like screening model for real solvents for prediction of the aqueous solubility of chlorobenzenes depending on temperature and salinity. *Environmental Toxicology and Chemistry* **2005**, 24, (6), 1368-1375.
69. Bustamante, P.; Romero, S.; Pena, A.; Escalera, B.; Reillo, A., Enthalpy-entropy compensation for the solubility of drugs in solvent mixtures: Paracetamol, acetanilide, and nalidixic acid in dioxane-water. *Journal of Pharmaceutical Sciences* **1998**, 87, (12), 1590-1596.
70. Manzo, R. H.; Ahumada, A. A., Effects of Solvent Medium on Solubility .5. Enthalpic and Entropic Contributions to the Free-Energy Changes of Disubstituted Benzene-Derivatives in Ethanol Water and Ethanol Cyclohexane Mixtures. *Journal of Pharmaceutical Sciences* **1990**, 79, (12), 1109-1115.
71. Granberg, R. A.; Rasmuson, A. C., Solubility of paracetamol in binary and ternary mixtures of water plus acetone plus toluene. *Journal of Chemical and Engineering Data* **2000**, 45, (3), 478-483.
72. Dickhut, R. M.; Andren, A. W.; Armstrong, D. E., Naphthalene Solubility in Selected Organic Solvent-Water Mixtures. *Journal of Chemical and Engineering Data* **1989**, 34, (4), 438-443.

73. Romero, S.; Reillo, A.; Escalera, B.; Bustamante, P., The behavior of paracetamol in mixtures of amphiprotic and amphiprotic-aprotic solvents. Relationship of solubility curves to specific and nonspecific interactions. *Chemical & Pharmaceutical Bulletin* **1996**, 44, (5), 1061-1064.
74. Bustamante, P.; Ochoa, R.; Reillo, A.; Escalera, J. B., Chameleonic Effect of Sulfanilamide and Sulfamethazine in Solvent Mixtures - Solubility Curves with 2 Maxima. *Chemical & Pharmaceutical Bulletin* **1994**, 42, (5), 1129-1133.
75. Chen, C. C., Private Communication. In Mullins, E., Ed. Aspen Technology, Inc.: Cambridge, MA, 2006.
76. Klamt, A., Prediction of the mutual solubilities of hydrocarbons and water with COSMO-RS. *Fluid Phase Equilibria* **2003**, 206, 223-235.

Appendices

A: Supplemental Tables

B: FORTRAN Code and Scripts

C: Sample Database Entry

Appendix A: Supplemental Tables

Table A.1: VT-2006 Solute Sigma Profile Database compound and property full listing.

Index		CAS-RN	Solvents	Points	W _{hb} [Lin 2002]		W _{hb} [Mathias 2002]	
No.	Solute Name				RMSE	AA%E	RMSE	AA%E
1	Maleic Acid Anhydride	108-31-6	6	6	0.7389	792.1	0.7472	863.1
2	Fumaric Acid	110-17-8	2	2	1.1823	1455.7	1.8831	12319.8
3	DL-Aspartic Acid	617-45-8	1	18	0.6645	363.2	1.3649	2249.7
4	L-Aspartic Acid	56-84-8	1	16	0.9229	737.4	1.6221	4101.8
5	3,5-Dichloropyridine	2457-47-8	1	1	1.0168	90.4	0.8553	86.0
6	3-Hydroxy-5-Nitropyridine	5418-51-9	1	1	2.8851	76661.0	3.4453	278724.7
7	5-Chloro-3-Pyridinol	74115-12-1	1	1	1.7924	6100.8	3.0440	110559.2
8	2-Amino-3,5-Dichloropyridine	4214-74-8	1	1	0.1000	25.9	0.2178	65.1
9	2-Hydroxypyridine	142-08-5	1	1	0.9331	757.3	1.2785	1799.1
10	3-Hydroxypyridine	109-00-2	1	1	1.4876	2973.1	1.8528	7025.0
11	2-Amino-2-Nitropyridine	4214-76-0	1	1	1.9094	8017.9	2.1962	15609.7
12	3-Amino-2-Chloropyridine	6298-19-7	1	1	0.9518	795.0	1.0777	1096.0
13	2-Amino-5-Chloropyridine	1072-98-6	1	1	0.8165	555.4	0.9394	769.7
14	Itaconic Acid	97-65-4	1	22	0.7490	477.2	1.1355	1508.5
15	2-Aminopyridine	504-29-0	1	1	2.2439	17433.8	2.3042	20044.5
16	3-Aminopyridine	504-24-5	1	1	1.0854	1117.3	1.3147	1964.0
17	2-Amino-3-Hydroxypyridine	16867-03-1	1	1	1.5752	3660.0	1.6616	4487.4
18	DL-Glutamic Acid	617-65-2	1	16	0.2945	95.4	1.0335	984.5
19	Picric Acid	88-89-1	7	7	1.1975	2173.9	1.2075	2248.0
20	2,4,6-Triiodophenol	609-23-4	2	2	1.6793	6517.5	1.6757	6519.1
21	6-Chloronicotinic Acid	5326-23-8	1	1	1.2147	1539.6	1.9365	8539.0
22	Isonicotinic Acid	55-22-1	1	1	1.3606	2194.2	1.7078	5002.4
23	Picolinic Acid	98-98-6	1	1	1.0410	999.0	1.2568	1706.4
24	Nicotinic Acid	59-67-6	1	1	1.2395	1635.8	1.5410	3375.3
25	2-Hydroxynicotinic Acid	609-71-2	1	1	2.1987	15701.8	2.8712	74229.9
26	4-Nitrophenol	100-02-7	2	2	0.9303	1636.1	0.8887	1215.9
27	4-Iodophenol	540-38-5	2	2	0.7155	857.4	0.7146	776.9

28	Isonicotinamide	1453-82-3	1	1	1.3479	2127.7	1.3647	2215.7
29	Nicotinamide	98-92-0	1	1	1.5924	3812.0	1.5903	3793.4
30	4-Nitroaniline	100-01-6	8	8	0.9559	1887.7	0.9708	2700.8
31	2-Methoxy-5-Nitropyridine	5446-92-4	1	1	0.1152	23.3	0.1302	34.9
32	2-Amino-4-Methyl-3-Nitropyridine	6635-86-5	1	1	0.4721	196.5	0.5995	297.7
33	2-Amino-4-Methylpyridine	695-34-1	1	1	0.7041	405.9	0.9393	769.7
34	Sulfanilamide	63-74-1	4	4	1.1466	1698.2	1.1371	1810.2
35	3,5-Dinitrosalicylic Acid	609-99-4	1	16	0.1277	30.6	2.2364	22200.7
36	Tetrachloroguaiacol	2539-17-5	1	5	0.6766	153.9	1.1170	599.4
37	M-Nitrobenzaldehyde	99-61-6	1	1	0.6895	389.2	0.8733	646.9
38	4-Iodobenzoic Acid	619-58-9	1	1	0.5372	244.5	1.2968	1880.6
39	4-Fluorobenzoic Acid	456-22-4	1	1	0.7778	499.6	1.4804	2922.8
40	4-Chlorobenzoic Acid	74-11-3	1	1	0.2987	98.9	0.9825	860.6
41	3,4,5-Trichloroguaiacol	57057-83-7	1	5	0.1230	33.9	0.5052	220.8
42	4,5,6-Chloroguaiacol	2668-24-8	1	5	0.9272	711.8	1.3773	2179.9
43	4-Bromobenzoic Acid	585-76-2						
44	Benzoic Acid	65-85-0	50	57	0.2811	125.2	0.3502	1604.2
45	Salicylic Acid	69-72-7	34	41	0.7871	1042.1	0.8696	6959.9
46	4-Hydroxybenzoic Acid	99-96-7	31	46	0.3460	162.6	0.4669	642.7
47	3-Hydroxybenzoic Acid	99-06-9	1	1	2.8961	78620.1	3.7196	524259.4
48	4,5-Dichloroguaiacol	2460-49-3	1	5	0.0827	0.9	0.4088	150.8
49	4,6-Dichloroguaiacol	16766-31-7	1	5	0.4100	155.2	0.7292	428.4
50	Benzamide	27208-38-4	1	1	1.5034	3087.2	1.5882	3774.1
51	Methylnicotinate	93-60-7	1	1	0.4464	179.5	0.6994	400.5
52	P-Hydroxybenzamide	619-57-8	1	1	2.7791	60034.2	3.0188	104319.9
53	4-Aminobenzoic Acid	150-13-0	18	19	0.4022	285.1	0.4662	504.7
54	4-Aminosalicylic Acid	65-49-6	1	9	1.8783	7633.0	2.6861	49883.0
55	2-Hydroxy-6-Methylnicotinic Acid	38116-61-9	1	1	1.0090	921.0	1.4779	2905.3
56	4-Chlorobenzyl Alcohol	873-76-7	1	1	0.5093	223.1	0.7943	522.8
57	4-Chloroguaiacol	16766-30-6	1	1	0.4367	173.3	0.8398	591.5
58	5-Chloroguaiacol	3743-23-5	1	1	0.0343	8.2	0.4141	159.5
59	4-Hydroxybenzyl Alcohol	623-05-2	1	1	3.0487	111776.5	3.5363	343705.4
60	Theophylline	58-55-9	23	24	0.5417	269.9	0.7417	705.4
61	Theobromine	83-67-0	5	5	0.6363	55.4	0.7676	130.0

62	2,6-Pyridinedimethanol	1195-59-1	1	1	3.2640	183567.1	3.3461	221759.4
63	1,2-Dimethyl-3-Hydroxy-4-Pyridone	30652-11-0						
64	2-Amino-4,6-Dimethylpyridine	5407-87-4	1	1	0.2802	90.6	0.4430	177.3
65	Simazine	122-34-9	1	3	1.2885	1984.9	1.4652	3030.5
66	Phthalic Acid Anhydride	85-44-9	9	9	0.9133	844.0	0.9028	842.1
67	2-Methyl-3-Nitrobenzoic Acid	1975-50-4	1	1	1.0090	921.0	1.9325	8461.4
68	2-Chlorophenoxyacetic Acid	614-61-9	1	1	1.5376	3348.2	3.6715	469224.6
69	5-Chlorovanillin	19463-48-0	1	14	0.1956	43.4	0.2682	75.0
70	P-Toluic Acid	99-45-5	6	123	0.4158	226.2	0.5140	625.8
71	4-Hydroxyacetophenone	99-93-4	1	1	1.6771	4654.4	2.9283	84674.7
72	M-Toluic Acid	99-04-7	3	3	0.3990	169.3	1.2010	20107.9
73	P-Anisic Acid	100-09-4	1	18	0.8704	645.1	1.3789	2305.6
74	Methyl-P-Hydroxybenzoate	99-76-3	44	66	0.3004	144.2	0.3996	360.7
75	M-Anisic Acid	586-38-9	1	1	1.3753	2273.0	3.4213	263693.9
76	4-Cyclohexene-1,2-Dicarboxylic Anhydride	85-43-8	6	6	0.8075	680.6	0.7913	638.4
77	Phenoxyacetic Acid	122-59-8	1	1	2.0036	9982.3	3.4066	254933.4
78	O-Anisic Acid	579-75-9	1	20	0.9256	751.2	1.5905	5869.1
79	3-Hydroxy-4-Methoxybenzoic Acid	645-08-9	1	1	1.9305	8420.3	2.5279	33620.4
80	Acetanilide	103-84-4	9	9	0.8927	1182.5	1.0400	1871.0
81	4-Aminoacetophenone	99-92-3	1	1	0.9381	767.1	0.9372	765.3
82A	Acetaminophen	103-90-2	26	95	0.5329	310.7	0.5753	290.2
82B	Acetaminophen	103-90-2	26	95	0.5291	262.2	0.5865	267.8
82C	Acetaminophen	103-90-2	26	95	0.4835	325.1	0.5385	259.6
83	Methyl-P-Aminobenzoate	619-45-4	3	3	0.8003	585.3	0.8351	650.4
84	Caffeine	58-08-2	6	7	0.7589	3164.3	0.8305	4243.4
85	Acyclovir	59277-89-3	2	4	1.6588	6144.1	1.9323	8808.7
86	Barbital	57-44-3	3	3	0.8219	585.3	1.3963	3709.7
87	Atrazine	1912-24-9	1	4	0.5501	257.8	0.7583	485.6
88A	Acetylsalicylic Acid	50-78-2	15	30	0.5857	496.8	0.7586	1216.3
88B	Acetylsalicylic Acid	50-78-2	15	30	0.5487	383.0	0.7346	1203.5
88C	Acetylsalicylic Acid	50-78-2	15	30	0.4947	895.4	0.4759	903.1
89	P-Tolylacetic Acid	622-47-9	5	5	1.0313	991.1	1.0313	991.1
90	Ethyl-4-Hydroxybenzoate	120-47-8	10	32	0.4953	349.6	0.7988	38520.6
91	Sulfamethizole	144-82-1						

92	Diuron	330-54-1	49	49	0.6767	44172.0	0.7401	49283.9
93	Ethyl-P-Aminobenzoate	94-09-7	3	4	0.6381	374.3	0.6704	419.8
94	Monuron	150-68-5	25	25	0.5915	452.9	0.6559	580.3
95	Chloropyrifos	2921-88-2	2	6	0.8735	992.2	1.0414	1320.6
96	Ganciclovir	82410-32-0	2	4	0.9110	787.1	1.2898	1896.2
97	Cyanazine	21725-46-2	1	3	0.1606	37.7	0.4517	183.2
98	Metharbital	50-11-3	2	2	0.3786	141.7	0.4839	397.5
99	Naphthalene	91-20-3	11	20	0.5622	347.2	0.6435	558.6
100	Sulfadiazine	68-35-9	3	3	0.8351	1678.0	0.8296	1010.6
101	Sulfamethoxazole	723-46-6	3	3	0.7704	912.3	0.9198	2650.0
102	Propyl-4-Hydroxybenzoate	94-13-3	10	32	0.4771	317.3	0.7175	24533.3
103	Phenacetin	62-44-2	3	14	0.9955	970.9	1.0565	1091.0
104	Propyl-P-Aminobenzoate	94-12-2	2	2	0.8595	725.9	0.8921	828.8
105	Ephedrine	299-42-3	13	13	0.1487	20.6	0.1333	21.4
106	Camphor	76-22-2	16	16	0.1419	39.4	0.1323	32.8
107	Butabarbital	125-40-6	2	2	1.1538	1377.6	1.7887	13207.0
108	Sulfapyridine	144-83-2	2	2	1.6406	22005.2	1.8027	17791.5
109	Antipyrine	60-80-0	4	5	0.5218	105.4	0.4166	175.3
110	Sulfamerazine	127-79-7	1	1	0.4228	164.7	0.2498	77.8
111	Sulfamer	651-06-9	1	1	0.2519	78.6	0.0674	16.8
112	Sulfamethoxypyridazine	80-35-3	5	17	1.5601	5449.2	1.6893	6566.7
113	Sulfisoxazole	127-69-5	1	1	0.1557	43.1	0.0407	9.8
114	Butyl-4-Hydroxybenzoate	94-26-8	10	32	0.5005	20092.5	0.6174	178099.4
115	Butyl-P-Aminobenzoate	94-25-7	3	4	0.8008	770.4	0.8266	897.5
116	Thiopental	76-75-5	2	2	0.5614	373.2	1.1267	1247.8
117	Amobarbital	57-43-2	2	2	0.7859	545.9	1.2619	2940.0
118	Pentobarbital	76-74-4	2	2	1.1239	1789.3	2.3539	373571.2
119	Dibenzothiophene	132-65-0	4	23	0.1372	44.8	0.1370	44.8
120	Dibenzofuran	132-64-9	4	22	0.0627	13.0	0.0631	13.1
121	Carbazole	86-74-8	29	46	0.7203	737.1	0.7534	928.0
122	Biphenyl	92-52-4	32	33	0.4872	334.2	0.5303	553.2
123	Acenaphthene	83-32-9	36	36	0.4612	235.5	0.4889	270.6
124	P-Hydroxybiphenyl	92-69-3	5	5	0.9546	1042.0	0.9552	1021.7
125	1-Acetyl-2-Naphthol	574-19-6	5	50	0.6404	467.3	0.6532	472.6

126	2-Acetyl-1-Naphthol	711-79-5	5	64	0.8042	554.2	0.8211	586.4
127	Diphenyl Sulfone	127-63-9	19	19	1.2712	2064.0	1.3151	2213.7
128	Phenobarbital	50-06-6	3	4	1.0850	1200.3	1.8902	55358.6
129	Sulfisomidine	515-64-0	3	3	1.0278	347.9	0.8678	170.9
130	Sulfamethazine	57-68-1	5	5	1.3592	5303.3	1.3114	4422.3
131	4-Hexylresorcinol	136-77-6	3	3	0.7312	5048.2	0.7845	7254.2
132	Thiamylal	77-27-0	2	2	0.4359	225.4	0.8381	597.0
133	Tolbutamide	64-77-7	2	2	1.1843	840.4	1.3243	1066.6
134	Thioxanthone	492-22-8	35	35	0.2717	104.9	0.2661	103.0
135	Acridine	260-94-6	4	24	0.2223	100.2	0.2213	100.0
136	Niflumic Acid	4294-00-7	21	25	1.1974	19189.7	1.2632	40556.8
137	Fluorene	86-73-7	2	2	1.0827	2611.9	1.3078	6661.3
138	Thioxanthene	261-31-4	4	24	0.1645	53.0	0.1624	52.4
139	Xanthene	92-83-1	4	24	0.0943	22.7	0.0932	22.5
140	Aminopyrine	58-15-1						
141A	Ibuprofen	15687-27-1	22	25	0.8461	11761.4	0.9536	13823.7
141B	Ibuprofen	15687-27-1	22	25	0.8437	11739.9	0.9516	13757.8
141C	Ibuprofen	15687-27-1	22	25	0.8086	11158.4	0.9166	12784.8
142	9,10-Anthraquinone	84-65-1	3	3	0.8159	651.2	0.6507	459.8
143	Phenanthrene	85-01-8	41	41	0.6721	618.9	0.7030	701.7
144	Anthracene	120-12-7	44	65	0.4481	250.2	0.4711	279.8
145	Diphenylglyoxal	134-81-6	50	50	0.5494	356.5	0.5678	385.2
146	Mitotane	53-19-0	4	4	2.4328	101776.2	2.5232	94296.9
147	N,N-Dimethylindoaniline	2150-58-5						
148A	Lidocaine	137-58-6	20	20	0.1241	37.5	0.1265	42.5
148B	Lidocaine	137-58-6	20	20	0.1355	32.7	0.1214	33.3
148C	Lidocaine	137-58-6	20	20	0.1357	34.9	0.1386	40.9
148D	Lidocaine	137-58-6	20	20	0.1284	38.9	0.1293	44.4
149	Piroxicam	36322-90-4	21	22	1.2854	7753.1	1.4334	13101.3
150	Flurbiprofen	5104-49-4	2	2	1.0918	1636.4	1.6184	13837.1
151	Mefenamic Acid	61-68-7	3	3	1.0565	742.8	0.9593	1927.9
152	Meperidine	57-42-1	1	1	3.2347	99.9	2.9802	99.9
153	Pyrene	129-00-0	25	29	0.7733	752.8	0.8082	877.3
154	Triazolam	28911-01-5						

155	Morphine	57-27-2	8	9	1.5885	25827.0	1.6233	31815.4
156	Hydromorphinone	466-99-9	2	2	1.6772	607.2	1.6345	641.1
157	Codeine	76-57-3	2	2	1.4493	764.2	1.3337	772.6
158	Estrone	53-16-7	12	51	0.4665	227.2	0.4928	290.4
159	Estradiol	50-28-2	1	1	0.5665	268.6	1.0051	911.8
160	Estriol	50-27-1	10	46	0.9333	2132.2	0.8836	3215.7
161	Nandrolone	434-22-0	12	12	0.3258	104.4	0.3861	140.3
162	Morniflumate	65847-85-0	4	4	1.0209	975.1	1.0604	1074.6
163	Testosterone	58-22-0	33	46	0.3016	141.2	0.3335	152.8
164	Androstanolone	521-18-6	2	2	0.6880	392.9	0.6559	363.0
165	Ethinyl Estradiol	57-63-6	1	1	0.5681	269.9	1.2550	1699.0
166	Norethindrone	68-22-4	2	2	0.3129	140.3	0.3854	203.2
167	Methanediene	72-63-9	12	12	0.2786	57.5	0.3103	67.4
168	Nandrolone Acetate	1425-10-1	1	1	0.0786	16.6	0.0786	16.6
169	Methyltestosterone	58-18-4	17	19	0.3680	144.1	0.4144	171.6
170	Mestanolone	521-11-9	14	14	0.2852	98.8	0.3502	135.1
171	Prostaglandin	363-24-6	2	2	0.2600	44.2	0.0378	8.6
172	Prostaglandin F2Alpha	551-11-1	2	2	0.4074	218.3	0.5374	426.7
173	Haloperidol	52-86-8	16	16	1.1333	2067.5	1.1476	2279.3
174	Ethisterone	434-03-7	1	1	1.5128	3157.1	1.5018	3075.7
175	Cortisone	53-06-5	2	3	0.7511	82.1	0.6979	79.0
176	Prednisolone	50-24-8	2	3	1.5469	92.2	1.4936	93.9
177	Progesterone	57-83-0	2	2	0.9910	1144.0	0.8472	1024.6
178	Desoxycorticosterone	64-85-7	2	2	0.2622	38.3	0.3408	49.4
179	Testosterone Acetate	1045-69-8	8	8	1.0466	6121.9	1.0249	4594.4
180	Nandrolone Propionate	7207-92-3	1	1	0.3921	146.7	0.3921	146.7
181	Hydrocortisone	50-23-7	13	13	0.5827	1381.8	0.4950	883.4
182	Prednisone	53-03-2	1	1	0.4406	63.7	0.5334	70.7
183	Norethindrone Acetate	51-98-9	2	2	0.9695	851.8	0.9891	875.2
184	Fentanyl	437-38-7	2	2	0.7312	733.4	0.9557	912.2
185	Betamethasone	378-44-9	1	1	0.2187	39.6	0.2170	39.3
186	Dexamethasone	50-02-2	1	1	1.6450	97.7	1.6170	97.6
187	Sufentanil	56030-54-7	2	2	2.0299	997.8	1.8194	997.8
188	Testosterone Propionate	57-85-2	29	29	0.4638	330.0	0.4594	286.5

189	Prednisone Acetate	125-10-0	1	1	0.6047	75.2	0.6880	79.5
190	Cortisone Acetate	50-04-4	1	1	0.6990	80.0	0.8154	84.7
191	Desoxycorticosterone Acetate	56-47-3	1	1	2.5366	34306.0	2.3683	23251.2
192	Hydrocortisone Acetate	50-03-3	1	1	0.8546	86.0	0.8635	86.3
193	Dexamethasone Acetate	1177-87-3	1	1	1.9991	99.0	1.9784	98.9
194	Betamethasone Acetate	987-24-6	1	1	1.2364	94.2	1.1696	93.2
195	Estradiol Benzoate	50-50-0	1	1	1.4340	96.3	1.1661	93.2
196	Norethindrone Enanthate	3836-23-5	1	1	1.9127	8078.3	1.9174	8167.1
197A	Cholesterol	57-88-5	60	221	0.4602	214.5	0.5140	245.9
197B	Cholesterol	57-88-5	60	221	0.4760	206.4	0.5294	228.7
197C	Cholesterol	57-88-5	60	221	0.4771	207.4	0.5315	228.5
197D	Cholesterol	57-88-5	60	221	0.4911	202.9	0.5554	228.8
197E	Cholesterol	57-88-5	60	221	0.4658	208.0	0.5207	243.5
197F	Cholesterol	57-88-5	60	221	0.4667	214.6	0.5177	237.5
198	Cholesterol Acetate	604-35-3	4	21	0.2815	108.4	0.3070	122.5
199	Sitosterol	83-46-5	11	23	0.8230	77.4	0.8216	81.0
200	Cholesteryl Benzoate	604-32-0	7	18	0.7422	117.1	0.7648	142.9
201	Beta Carotene	7235-40-7						
202	Vinbarbital	125-42-8						
205	Methyl-4-Dimethylaminobenzoate	1202-25-1						
206	Methacetin	51-66-1						
207	Nalidixic Acid	389-08-2						

Table A.2: Error summary of predicted COSMO-SAC solubility from experimental mole fraction²⁷ for each acetaminophen conformation in 26 pure solvents at various temperatures comparing of both exchange energy expressions (Part 1: RMSE, Part 2: AA%E).

Part 1: RMSE

Exchange Energy Conformation Solvent (Literature Points)	W _{hb} [Lin 2002]			W _{hb} [Mathias 2002]		
	A RMSE	B RMSE	C RMSE	A RMSE	B RMSE	C RMSE
Toluene (7)	0.5809	0.8129	0.5645	0.5499	0.8898	0.5406
Acetone (8)	0.6843	0.6538	0.5732	0.7103	0.7002	0.6475
Methyl Ethyl Ketone (1)	0.7268	0.7057	0.6390	0.7573	0.7524	0.7116
Methyl Isobutyl Ketone (1)	0.9305	0.8957	0.8175	0.9972	0.9873	0.9459
Methanol (8)	0.2071	0.2016	0.1135	0.3108	0.3224	0.2433
Ethanol (8)	0.3224	0.3174	0.2225	0.4759	0.4890	0.4209
N-Propanol (8)	0.3692	0.3627	0.2582	0.5456	0.5606	0.4841
Isopropyl Alcohol (8)	0.4364	0.4348	0.3326	0.6432	0.6592	0.5965
N-Butanol (8)	0.3705	0.3643	0.2535	0.5882	0.6047	0.5265
1-Pentanol (1)	0.3764	0.3748	0.2737	0.5585	0.5754	0.5065
1-Hexanol (1)	0.4036	0.4022	0.2966	0.6183	0.6358	0.5667
1-Heptanol (1)	0.3972	0.3954	0.2850	0.6215	0.6407	0.5650
1-Octanol (1)	0.4458	0.4446	0.3311	0.6975	0.7173	0.6419
Ethylene-Glycol (1)	0.0417	0.0487	0.1686	0.0574	0.0682	0.0468
Acetic-Acid (1)	0.1944	0.1688	0.1835	0.2590	0.2349	0.2869
Ethyl Acetate (8)	0.9652	0.8741	0.7954	0.7962	0.7118	0.6962
1,4-Dioxane (2)	0.8478	0.8157	0.7438	0.8041	0.7830	0.7357
Tetrahydrofuran (1)	0.4770	0.4706	0.4183	0.5566	0.5607	0.5323
Carbon-Tetrachloride (1)	1.6473	1.8830	1.5754	1.6695	2.0389	1.5911
Dichloromethane (1)	1.2678	1.0920	1.4083	1.0152	0.7598	1.0623
Chloroform (1)	1.0559	0.9779	1.3762	0.4124	0.2276	0.6413
Diethyl Amine (1)	0.1950	0.1892	0.2180	0.0896	0.0833	0.1035
Acetonitrile (8)	0.7478	0.6635	0.5705	0.5971	0.5095	0.4697
Dimethyl-Sulfoxide (1)	0.0347	0.0333	0.0069	0.1169	0.1217	0.1029
N,N-Dimethylformamide (1)	0.0083	0.0032	0.0415	0.0681	0.0712	0.0447
Water (7)	0.1220	0.1707	0.1030	0.4417	0.5424	0.2910
Average	0.5329	0.5291	0.4835	0.5753	0.5865	0.5385

Part 2: AA%E

Exchange Energy Conformation Solvent (Literature Points)	W _{hb} [Lin 2002]			W _{hb} [Mathias 2002]		
	A AA%E	B AA%E	C AA%E	A AA%E	B AA%E	C AA%E
Toluene (7)	64.6	79.5	63.6	58.4	81.1	58.6
Acetone (8)	383.8	350.7	274.3	414.0	402.1	344.4
Methyl Ethyl Ketone (1)	433.0	407.8	335.5	471.8	465.4	414.7
Methyl Isobutyl Ketone (1)	752.1	686.5	556.9	893.5	871.1	782.9
Methanol (8)	60.6	58.4	27.3	104.4	110.0	74.7
Ethanol (8)	110.0	107.6	66.4	199.2	208.4	163.6
N-Propanol (8)	133.9	130.4	80.3	251.3	263.7	204.9
Isopropyl Alcohol (8)	173.1	172.2	115.0	340.1	356.8	295.1
N-Butanol (8)	134.5	131.1	78.0	287.6	302.6	236.2
1-Pentanol (1)	137.9	137.0	87.8	261.8	276.2	221.0
1-Hexanol (1)	153.3	152.5	98.0	315.3	332.3	268.7
1-Heptanol (1)	149.6	148.6	92.8	318.3	337.3	267.3
1-Octanol (1)	179.1	178.3	114.3	398.4	421.6	338.4
Ethylene-Glycol (1)	9.2	10.6	32.2	14.1	17.0	10.2
Acetic-Acid (1)	56.5	47.5	52.6	81.6	71.8	93.6
Ethyl Acetate (8)	825.8	651.9	526.9	528.6	418.2	399.1
1,4-Dioxane (2)	701.2	638.4	512.3	616.3	578.4	499.1
Tetrahydrofuran (1)	199.9	195.5	162.0	260.3	263.6	240.6
Carbon-Tetrachloride (1)	97.7	98.7	97.3	97.9	99.1	97.4
Dichloromethane (1)	1752.7	1136.0	2460.4	935.5	475.2	1054.2
Chloroform (1)	1037.5	850.3	2277.8	158.4	68.9	337.8
Diethyl Amine (1)	36.2	35.3	39.5	18.7	17.5	21.2
Acetonitrile (8)	459.7	361.1	272.1	295.8	223.5	195.0
Dimethyl-Sulfoxide (1)	8.3	8.0	1.6	30.9	32.3	26.7
N,N-Dimethylformamide (1)	1.9	0.7	9.1	17.0	17.8	10.8
Water (7)	25.5	41.3	19.8	176.9	251.2	92.5
Average	310.7	262.2	325.1	290.2	267.8	259.6

Table A.3: Error summary of predicted COSMO-SAC solubility from experimental mole fraction²⁷ for each aspirin conformation in 26 pure solvents at various temperatures comparing of both exchange energy expressions (Part 1: RMSE, Part 2: AA%E).

Part 1: RMSE

Exchange Energy Conformation Solvent	W_{hb} [Lin 2002]			W_{hb} [Mathias 2002]		
	A RMSE	B RMSE	C RMSE	A RMSE	B RMSE	C RMSE
Cyclohexane (1)	1.5485	1.3542	2.0571	1.5634	1.3844	2.0670
Acetone (1)	0.4146	0.3816	0.1007	0.5023	0.4826	0.1099
Methyl Ethyl Ketone (1)	0.7744	0.7385	0.1729	0.8684	0.8481	0.1637
Methanol (1)	0.3693	0.3320	0.5094	0.5801	0.5561	0.4556
Ethanol (1)	0.6023	0.5630	0.3732	0.7961	0.7757	0.3429
2-Propanol (1)	0.6303	0.5885	0.4427	0.8249	0.8057	0.4195
Isoamyl Alcohol (1)	0.4858	0.4122	0.5440	0.7493	0.7116	0.5365
N-Octanol (2)	0.8950	0.8128	0.2767	1.1674	1.1337	0.2742
Acetic Acid (1)	0.1173	0.1186	0.1088	0.1433	0.1390	0.1670
Diethyl Ether (1)	1.0222	0.9680	0.2223	1.1516	1.1237	0.2135
1,4-Dioxane (1)	0.4144	0.3731	0.0911	0.4939	0.4660	0.0993
Chloroform (1)	0.3777	0.4570	0.6530	0.1075	0.1400	0.4646
1,2-Dichloroethane (1)	0.2807	0.2544	0.7216	0.3416	0.3386	0.7162
1,1,1-Trichloroethane (1)	0.4724	0.3642	0.9676	0.5495	0.4732	0.9760
Water (15)	0.3803	0.5120	0.1795	1.5393	1.6410	0.1330
Average	0.5857	0.5487	0.4947	0.7586	0.7346	0.4759

Part 2: AA%E

Exchange Energy Conformation Solvent	W _{hb} [Lin 2002]			W _{hb} [Mathias 2002]		
	A AA%E	B AA%E	C AA%E	A AA%E	B AA%E	C AA%E
Cyclohexane (1)	3435.7	2160.2	11304.1	3559.7	2323.1	11567.4
Acetone (1)	159.8	140.8	20.7	217.9	203.8	22.4
Methyl Ethyl Ketone (1)	494.8	447.7	48.9	638.7	604.8	45.8
Methanol (1)	134.1	114.8	69.1	280.3	259.9	65.0
Ethanol (1)	300.2	265.6	57.7	525.3	496.7	54.6
2-Propanol (1)	326.9	287.7	63.9	568.1	539.4	61.9
Isoamyl Alcohol (1)	206.0	158.3	71.4	461.5	414.7	70.9
N-Octanol (2)	685.5	549.9	47.1	1372.2	1262.0	46.8
Acetic Acid (1)	31.0	31.4	28.5	39.1	37.7	46.9
Diethyl Ether (1)	952.5	828.9	66.8	1317.7	1229.6	63.5
1,4-Dioxane (1)	159.6	136.1	18.9	211.8	192.4	20.4
Chloroform (1)	138.6	186.4	349.7	28.1	38.0	191.5
1,2-Dichloroethane (1)	90.8	79.6	426.8	119.6	118.1	420.3
1,1,1-Trichloroethane (1)	196.8	131.3	828.1	254.4	197.3	846.3
Water (15)	139.3	226.6	28.8	8649.8	10135.1	23.1
Average	496.8	383.0	895.4	1216.3	1203.5	903.1

Table A.4: Error summary of predicted COSMO-SAC solubility from experimental mole fraction²⁷ for each ibuprofen conformation in 26 pure solvents at various temperatures comparing of both exchange energy expressions (Part 1: RMSE, Part 2: AA%E).

Part 1: RMSE

Exchange Energy Conformation Solvent (Literature Points)	W _{hb} [Lin 2002]			W _{hb} [Mathias 2002]		
	A RMSE	B RMSE	C RMSE	A RMSE	B RMSE	C RMSE
Heptane (1)	0.0208	0.0148	0.0561	0.0594	0.0651	0.0432
Cyclohexane (1)	0.0727	0.0679	0.0799	0.0422	0.0380	0.0447
Benzene (1)	0.4719	0.4724	0.4652	0.4908	0.4912	0.4873
Chlorobenzene (1)	1.0291	1.0301	1.0170	1.0564	1.0572	1.0496
Acetone (1)	0.2052	0.2045	0.1816	0.2337	0.2329	0.2102
Acetophenone (1)	2.1844	2.1834	2.1500	2.2143	2.2132	2.1789
Methanol (1)	1.2880	1.2870	1.2601	1.3632	1.3623	1.3381
Ethanol (1)	0.5696	0.5689	0.5470	0.6345	0.6338	0.6149
1-Pentanol (1)	0.3903	0.3896	0.3669	0.4565	0.4558	0.4365
1-Octanol (2)	0.3374	0.3368	0.3149	0.4012	0.4006	0.3823
1,2 Ethanediol (1)	1.1454	1.1309	0.7553	1.4682	1.4661	1.4041
1,2-Propanediol (1)	1.7045	1.7030	1.6532	1.7853	1.7845	1.7603
Glycerol (1)	0.5527	0.5498	0.4496	0.6534	0.6520	0.6159
Acetic Acid (1)	0.2448	0.2425	0.1990	0.2741	0.2710	0.2096
Propionic Acid (1)	0.0899	0.0888	0.0659	0.1052	0.1036	0.0705
Ethyl Acetate (1)	0.1659	0.1651	0.1349	0.1875	0.1864	0.1527
Diethyl Ether (1)	3.3449	3.3445	3.3259	3.3751	3.3746	3.3560
1,4-Dioxane (1)	1.1642	1.1634	1.1383	1.1936	1.1927	1.1665
Chloroform (1)	0.1516	0.1522	0.1664	0.1137	0.1142	0.1216
1,2-Dichloroethane (1)	0.4641	0.4644	0.4619	0.4828	0.4830	0.4828
N,N-Dimethylformamide (1)	0.7351	0.7347	0.7215	0.7665	0.7661	0.7559
Water (2)	1.5318	1.5206	1.4856	2.2855	2.2704	2.1296
Formamide (1)	1.2907	1.2820	1.2091	1.9915	1.9745	1.7691
Average	0.8328	0.8303	0.7915	0.9406	0.9387	0.9035

Part 2: AA%E

Exchange Energy Conformation Solvent (Literature Points)	W _{hb} [Lin 2002]			W _{hb} [Mathias 2002]		
	A AA%E	B AA%E	C AA%E	A AA%E	B AA%E	C AA%E
Heptane (1)	4.7	3.3	12.1	14.7	16.2	10.5
Cyclohexane (1)	15.4	14.5	16.8	9.3	8.4	9.8
Benzene (1)	196.4	196.7	191.9	209.6	209.9	207.1
Chlorobenzene (1)	969.4	971.7	939.9	1038.7	1040.9	1020.9
Acetone (1)	60.4	60.1	51.9	71.3	71.0	62.2
Acetophenone (1)	15191.2	15154.8	14026.9	16279.9	16236.3	14995.7
Methanol (1)	1841.0	1836.6	1720.2	2207.7	2202.8	2078.0
Ethanol (1)	271.2	270.6	252.3	331.1	330.3	312.0
1-Pentanol (1)	145.6	145.2	132.7	186.1	185.6	173.2
1-Octanol (2)	117.2	116.9	106.1	151.8	151.4	141.0
1,2 Ethanediol (1)	1297.8	1251.7	469.3	2838.8	2824.8	2435.9
1,2-Propanediol (1)	10002.9	9973.9	9172.0	11884.9	11860.8	11227.9
Glycerol (1)	452.5	447.9	301.7	662.9	660.9	600.5
Acetic Acid (1)	75.7	74.8	58.1	88.0	86.6	62.0
Propionic Acid (1)	23.0	22.7	16.4	27.4	26.9	17.6
Ethyl Acetate (1)	46.5	46.2	36.4	54.0	53.6	42.1
Diethyl Ether (1)	221176.3	220950.0	211700.4	237097.0	236793.2	226865.8
1,4-Dioxane (1)	1359.5	1356.9	1274.8	1461.8	1458.6	1367.3
Chloroform (1)	41.8	42.0	46.7	29.9	30.1	32.3
1,2-Dichloroethane (1)	191.1	191.3	189.7	203.9	204.1	203.9
N,N-Dimethylformamide (1)	443.3	442.8	426.7	484.1	483.6	470.1
Water (2)	3388.8	3300.0	3035.0	19754.8	19075.5	13751.0
Formamide (1)	1852.8	1814.4	1518.4	9705.8	9330.4	5776.6
Average	11268.0	11247.2	10682.5	13251.9	13188.8	12254.9

Table A.5: Error summary of predicted COSMO-SAC solubility from experimental mole fraction²⁷ for each lidocaine conformation in 26 pure solvents at various temperatures comparing of both exchange energy expressions (Part 1: RMSE, Part 2: AA%E).

Part 1: RMSE

Exchange Energy Conformation Solvent Name	W _{hb} [Lin 2002]				W _{hb} [Mathias 2002]			
	A	B	C	D	A	B	C	D
	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
Hexane	0.5525	0.6158	0.5550	0.5635	0.5578	0.6160	0.5608	0.5678
Gamma-Butyrolactone	0.0300	0.0332	0.0246	0.0317	0.0148	0.0330	0.0061	0.0168
Methanol	0.1308	0.0708	0.1166	0.1360	0.1673	0.1074	0.1488	0.1764
1-Propanol	0.0882	0.0337	0.0754	0.0920	0.1118	0.0506	0.0920	0.1205
3-Methyl-3-Pentanol	0.0325	0.0118	0.0238	0.0357	0.0402	0.0083	0.0259	0.0472
1,2-Propanediol	0.2577	0.1163	0.2266	0.2647	0.3101	0.1887	0.2788	0.3224
1,3 Propanediol	0.3536	0.3814	0.1915	0.3673	0.4745	0.1272	0.3826	0.4961
Ethyl Acetate	0.0643	0.0705	0.0644	0.0658	0.0597	0.0704	0.0585	0.0613
Triglyme	0.0120	0.0477	0.0173	0.0175	0.0008	0.0475	0.0027	0.0049
1,2-Dimethoxyethane	0.0491	0.0530	0.0518	0.0501	0.0441	0.0529	0.0454	0.0454
Diglycol Methyl Ether	0.0281	0.0443	0.0323	0.0309	0.0193	0.0441	0.0216	0.0223
Tetraglyme	0.0589	0.0030	0.0522	0.0508	0.0761	0.0032	0.0706	0.0678
Tetrahydrofuran	0.0656	0.0552	0.0725	0.0644	0.0637	0.0551	0.0729	0.0616
1-Methoxy-2-Propyl Acetate	0.0470	0.0675	0.0481	0.0503	0.0413	0.0674	0.0412	0.0445
Water	0.3495	0.9214	0.7430	0.3905	0.0164	0.7326	0.4465	0.0033
Triacetin	0.1756	0.0076	0.2024	0.1541	0.2259	0.0062	0.2522	0.2049
2-Ethoxyethanol	0.1158	0.0702	0.0986	0.1220	0.1613	0.1266	0.1421	0.1707
Diacetin	0.0085	0.0704	0.0436	0.0220	0.0963	0.0557	0.0599	0.1106
Dimethyl Isosorbide	0.0278	0.0257	0.0398	0.0266	0.0080	0.0253	0.0228	0.0057
N-Butyl Lactate	0.0345	0.0114	0.0345	0.0312	0.0401	0.0111	0.0402	0.0367
Average	0.1241	0.1355	0.1357	0.1284	0.1265	0.1214	0.1386	0.1293

Part 2: AA%E

Exchange Energy Conformation Solvent Name	W _{hb} [Lin 2002]				W _{hb} [Mathias 2002]			
	A AA%E	B AA%E	C AA%E	D AA%E	A AA%E	B AA%E	C AA%E	D AA%E
Hexane	256.8	312.9	258.9	266.0	261.3	313.0	263.7	269.6
Gamma-Butyrolactone	7.1	7.9	5.8	7.6	3.5	7.9	1.4	3.9
Methanol	35.2	17.7	30.8	36.8	47.0	28.0	40.9	50.1
1-Propanol	22.5	8.1	19.0	23.6	29.4	12.4	23.6	32.0
3-Methyl-3-Pentanol	7.8	2.8	5.6	8.6	9.7	1.9	6.1	11.5
1,2-Propanediol	81.0	30.7	68.5	83.9	104.2	54.4	90.0	110.1
1,3 Propanediol	125.7	58.4	55.4	133.0	198.2	25.4	141.3	213.4
Ethyl Acetate	16.0	17.6	16.0	16.4	14.7	17.6	14.4	15.2
Triglyme	2.8	11.6	4.1	4.1	0.2	11.6	0.6	1.1
1,2-Dimethoxyethane	12.0	13.0	12.7	12.2	10.7	12.9	11.0	11.0
Diglycol Methyl Ether	6.7	10.7	7.7	7.4	4.5	10.7	5.1	5.3
Tetraglyme	12.7	0.7	11.3	11.0	16.1	0.7	15.0	14.4
Tetrahydrofuran	16.3	13.5	18.2	16.0	15.8	13.5	18.3	15.2
1-Methoxy-2-Propyl Acetate	11.4	16.8	11.7	12.3	10.0	16.8	9.9	10.8
Water	55.3	88.0	81.9	59.3	3.8	81.5	64.2	0.7
Triacetin	33.3	1.7	37.2	29.9	40.6	1.4	44.0	37.6
2-Ethoxyethanol	30.5	17.5	25.5	32.4	45.0	33.9	38.7	48.1
Diacetin	2.0	15.0	9.5	5.2	24.8	13.7	14.8	29.0
Dimethyl Isosorbide	6.6	6.1	9.6	6.3	1.8	6.0	5.4	1.3
N-Butyl Lactate	7.6	2.6	7.6	6.9	8.8	2.5	8.8	8.1
Average	37.5	32.7	34.9	38.9	42.5	33.3	40.9	44.4

Table A.6: Error summary of predicted COSMO-SAC solubility from experimental mole fraction²⁷ for each cholesterol conformation in 26 pure solvents at various temperatures comparing of both exchange energy expressions (Part 1: RMSE, Part 2: AA%E).

Part 1: RMSE

Exchange Energy Conformation Solvent Name	Points	W _{hb} [Lin 2002]						W _{hb} [Mathias 2002]					
		A RMSE	B RMSE	C RMSE	D RMSE	E RMSE	F RMSE	A RMSE	B RMSE	C RMSE	D RMSE	E RMSE	F RMSE
Cyclohexane	22	0.2233	0.2076	0.2038	0.1930	0.2169	0.2332	0.2524	0.2305	0.2265	0.2110	0.2424	0.2588
Benzene	24	0.6186	0.6186	0.6267	0.6973	0.6464	0.5703	0.5543	0.5581	0.5642	0.6440	0.5875	0.5126
Toluene	21	0.6749	0.6796	0.6874	0.7490	0.6979	0.6337	0.6138	0.6222	0.6281	0.6967	0.6418	0.5792
Acetone	4	0.5878	0.5224	0.5248	0.5176	0.5781	0.5518	0.4647	0.3880	0.3838	0.3630	0.4587	0.4271
Methanol	11	0.5875	0.5509	0.5565	0.5516	0.5566	0.5774	0.6661	0.6461	0.6489	0.6388	0.6385	0.6725
Ethanol	6	0.4517	0.3943	0.3979	0.4040	0.4316	0.4238	0.4918	0.4357	0.4362	0.4447	0.4804	0.4680
1-Propanol	4	0.0489	0.0660	0.0640	0.0606	0.0487	0.0504	0.0614	0.0524	0.0523	0.0510	0.0569	0.0512
2-Propanol	1	0.4106	0.3511	0.3536	0.3667	0.3977	0.3771	0.4650	0.3962	0.3959	0.4183	0.4642	0.4251
1-Butanol	5	0.3164	0.3423	0.3412	0.3374	0.3225	0.3285	0.3084	0.3315	0.3317	0.3269	0.3105	0.3182
2-Methyl-1-Propanol	1	0.2450	0.2404	0.2387	0.2131	0.2210	0.2592	0.3095	0.3288	0.3258	0.2875	0.2808	0.3408
2-Butanol	1	0.0766	0.0446	0.0442	0.0378	0.0600	0.0688	0.1000	0.0783	0.0765	0.0672	0.0848	0.1021
T-Butanol	1	0.9214	0.8697	0.8719	0.8894	0.9166	0.8870	0.9721	0.9062	0.9058	0.9350	0.9789	0.9265
1-Pentanol	4	0.3866	0.4363	0.4352	0.4285	0.3987	0.4105	0.3730	0.4222	0.4231	0.4144	0.3783	0.3945
1-Hexanol	4	0.3893	0.4427	0.4419	0.4338	0.4005	0.4154	0.3711	0.4275	0.4286	0.4169	0.3743	0.3981
1-Heptanol	4	0.5461	0.5978	0.5974	0.5904	0.5569	0.5702	0.5341	0.5873	0.5887	0.5789	0.5382	0.5578
1-Octanol	4	0.4989	0.5511	0.5510	0.5429	0.5084	0.5236	0.4850	0.5412	0.5427	0.5308	0.4870	0.5118
1-Nonanol	4	0.5099	0.5606	0.5608	0.5537	0.5192	0.5328	0.5028	0.5554	0.5570	0.5473	0.5059	0.5259
1-Decanol	4	0.6525	0.7035	0.7039	0.6966	0.6610	0.6755	0.6418	0.6969	0.6986	0.6879	0.6435	0.6673
1-Undecanol	1	0.5807	0.6304	0.6313	0.6236	0.5883	0.6023	0.5700	0.6250	0.6268	0.6150	0.5713	0.5953
1-Dodecanol	1	0.3139	0.3633	0.3643	0.3569	0.3212	0.3348	0.3041	0.3589	0.3608	0.3493	0.3054	0.3290
1,2-Ethanediol	1	0.3136	0.2812	0.2645	0.2646	0.3418	0.3024	0.1728	0.1187	0.1045	0.1140	0.2027	0.1432
1,2-Propanediol	1	0.6430	0.6275	0.6368	0.6430	0.6257	0.6301	0.6962	0.6895	0.6963	0.6998	0.6845	0.6934
Dipropylene Glycol	1	0.0010	0.0429	0.0381	0.0297	0.0011	0.0360	0.0048	0.0519	0.0522	0.0444	0.0105	0.0386
1,3-Butanediol	1	0.0915	0.0829	0.0944	0.1069	0.0761	0.0763	0.1498	0.1476	0.1563	0.1686	0.1406	0.1425
1,4-Butanediol	1	0.2221	0.2017	0.2127	0.2268	0.2062	0.2017	0.2773	0.2637	0.2717	0.2851	0.2678	0.2656
Oleic Acid	1	0.0405	0.0124	0.0093	0.0551	0.0815	0.0204	0.2639	0.3155	0.3136	0.2638	0.2300	0.3115
Ethyl Acetate	2	0.0972	0.1458	0.1470	0.1714	0.1083	0.1121	0.2626	0.3149	0.3210	0.3737	0.2779	0.2706
3-Methyl-Acetate 1-Butanol	1	0.1836	0.2325	0.2349	0.2546	0.1907	0.2010	0.3186	0.3689	0.3750	0.4162	0.3279	0.3296

Ethyl Propionate	1	0.0935	0.1473	0.1500	0.1757	0.1045	0.1093	0.2519	0.3035	0.3102	0.3625	0.2665	0.2565
Ethyl-N-Butyrate	1	0.1450	0.1996	0.2028	0.2274	0.1550	0.1612	0.2954	0.3485	0.3555	0.4057	0.3090	0.3012
Ethylene Glycol Diacetate	1	0.0455	0.0693	0.0658	0.0912	0.0578	0.0525	0.2390	0.2627	0.2648	0.3249	0.2547	0.2396
Methyl Benzoate	1	0.6122	0.6268	0.6296	0.6741	0.6259	0.6010	0.7560	0.7632	0.7687	0.8493	0.7768	0.7326
Isopropyl Myristate	1	0.4072	0.4571	0.4625	0.4877	0.4123	0.4177	0.5236	0.5732	0.5811	0.6311	0.5340	0.5274
Carbon Tetrachloride	17	0.2924	0.3133	0.3223	0.3566	0.3071	0.2715	0.2521	0.2751	0.2816	0.3138	0.2698	0.2384
Chloroform	20	0.6607	0.6803	0.6787	0.6547	0.6458	0.6843	0.5939	0.6186	0.6169	0.5860	0.5728	0.6209
1,2-Dichloroethane	14	0.2318	0.2259	0.2300	0.2795	0.2539	0.2055	0.2077	0.2065	0.2086	0.2551	0.2286	0.1935
Acetonitrile	5	0.4781	0.4809	0.4708	0.4959	0.4979	0.4754	0.6015	0.6033	0.5989	0.6584	0.6235	0.5922
N,N-Dimethylformamide	1	1.7107	1.6331	1.6400	1.6661	1.7096	1.6471	1.8340	1.7323	1.7333	1.7781	1.8469	1.7509
Propylene Carbonate	2	0.9473	0.9429	0.9398	0.9648	0.9567	0.9392	1.0710	1.0631	1.0635	1.1257	1.0857	1.0565
Triacetin	1	0.3720	0.3721	0.3675	0.3986	0.3815	0.3686	0.5657	0.5630	0.5641	0.6359	0.5799	0.5547
Formamide	1	3.1736	3.1023	3.0769	3.0716	3.2079	3.1514	2.9840	2.8951	2.8722	2.8743	3.0191	2.9484
N-Methyl-Formamide	1	0.1083	0.0556	0.0683	0.0896	0.0938	0.0650	0.1262	0.0742	0.0819	0.1031	0.1213	0.0881
2-Methoxyethanol	1	0.1857	0.2454	0.2564	0.2544	0.1852	0.1969	0.2404	0.3269	0.3355	0.3192	0.2382	0.2756
2-Ethoxyethanol	1	0.5069	0.5277	0.5236	0.5350	0.5307	0.5084	0.4623	0.4577	0.4557	0.4805	0.4880	0.4410
2-(2-Ethoxyethoxy)-Ethanol	1	0.4595	0.4731	0.4688	0.4828	0.4814	0.4567	0.4220	0.4081	0.4057	0.4354	0.4476	0.3952
Diethylene Glycol Butyl Ether	1	0.6254	0.6432	0.6405	0.6544	0.6454	0.6241	0.5968	0.5883	0.5873	0.6164	0.6203	0.5726
2-Ethoxyethyl Acetate	1	0.1747	0.2227	0.2233	0.2482	0.1866	0.1898	0.3520	0.3993	0.4048	0.4591	0.3669	0.3594
2-(2-Ethoxyethoxy)-Acetate-Ethanol	1	0.2483	0.2904	0.2905	0.3159	0.2595	0.2607	0.4254	0.4670	0.4719	0.5287	0.4401	0.4310
Ethyl Ester Octanoic Acid	1	0.3597	0.4091	0.4143	0.4449	0.3687	0.3678	0.4869	0.5326	0.5403	0.5980	0.5018	0.4836
Ethyl Ester Dodecanoic Acid	1	0.4576	0.5058	0.5115	0.5421	0.4651	0.4645	0.5776	0.6227	0.6307	0.6886	0.5912	0.5745
Ethyl Ester Decanoic Acid	1	0.4009	0.4501	0.4556	0.4855	0.4089	0.4088	0.5240	0.5701	0.5779	0.6349	0.5379	0.5216
2-Methoxy-Acetate Ethanol	1	0.0834	0.1254	0.1245	0.1494	0.0956	0.0971	0.2707	0.3116	0.3159	0.3723	0.2857	0.2765
Ethyl Ester Nonanoic Acid	1	0.3902	0.4390	0.4443	0.4747	0.3986	0.3979	0.5153	0.5605	0.5682	0.6258	0.5296	0.5120
Ethyl Ester Hexanoic Acid	1	0.2630	0.3130	0.3177	0.3476	0.2726	0.2722	0.3939	0.4402	0.4477	0.5041	0.4091	0.3913
Ethyl Myristate	1	0.4929	0.5406	0.5465	0.5774	0.5000	0.4991	0.6104	0.6551	0.6632	0.7217	0.6237	0.6070
2,3-Butanediol	1	1.4949	1.5550	1.5505	1.4651	1.4490	1.5677	1.4385	1.5122	1.5071	1.4047	1.3877	1.5163
Glycerol Tricaprylate	1	0.5740	0.5981	0.6020	0.6410	0.5818	0.5702	0.7117	0.7294	0.7360	0.8100	0.7261	0.6965
Ethyl Ester Pentanoic Acid	1	0.2468	0.2969	0.3014	0.3316	0.2571	0.2561	0.3803	0.4265	0.4338	0.4907	0.3961	0.3773
Ethyl Ester Undecanoic Acid	1	0.4985	0.5476	0.5531	0.5828	0.5061	0.5064	0.6203	0.6665	0.6744	0.7310	0.6338	0.6183
Ethyl Ester Hexadecanoic Acid	1	0.4928	0.5398	0.5459	0.5772	0.4995	0.4982	0.6075	0.6517	0.6599	0.7191	0.6206	0.6037
Average		0.4645	0.4805	0.4818	0.4957	0.4697	0.4650	0.5188	0.5345	0.5368	0.5606	0.5251	0.5168

Part 2: AA%E

Conformation Solvent Name	Points	W _{hb} [Lin 2002]						W _{hb} [Mathias 2002]					
		A AA%E	B AA%E	C AA%E	D AA%E	E AA%E	F AA%E	A AA%E	B AA%E	C AA%E	D AA%E	E AA%E	F AA%E
Cyclohexane	22	53.8	46.9	45.1	40.1	51.2	58.3	67.7	56.8	54.8	48.1	62.6	71.0
Benzene	24	70.3	70.5	71.0	74.6	71.9	67.5	66.3	66.8	67.2	71.7	68.4	63.6
Toluene	21	68.6	69.0	69.4	71.8	69.6	66.8	65.7	66.4	66.7	69.6	67.1	64.1
Acetone	4	287.3	233.1	234.9	229.5	278.7	256.5	191.6	144.3	142.0	130.6	187.6	167.3
Methanol	11	288.9	256.9	261.6	257.5	261.7	279.8	367.7	346.3	349.2	338.6	338.4	374.7
Ethanol	6	185.0	148.1	150.3	154.0	171.5	166.5	213.6	174.2	174.6	180.2	205.2	196.3
1-Propanol	4	10.7	10.8	10.5	10.1	10.3	9.9	12.7	10.3	10.2	10.6	11.6	11.1
2-Propanol	1	157.4	124.4	125.7	132.7	149.8	138.3	191.7	149.0	148.8	162.0	191.2	166.1
1-Butanol	5	25.5	26.9	26.8	26.2	25.4	25.2	27.1	25.3	25.2	25.5	26.6	25.6
2-Methyl-1-Propanol	1	75.8	73.9	73.3	63.3	66.3	81.6	103.9	113.2	111.8	93.9	90.9	119.2
2-Butanol	1	19.3	10.8	10.7	9.1	14.8	17.2	25.9	19.8	19.3	16.7	21.6	26.5
T-Butanol	1	734.5	640.9	644.6	675.1	725.2	670.9	837.7	705.8	705.0	760.9	852.6	744.3
1-Pentanol	4	53.0	58.4	58.3	57.6	54.3	55.7	51.2	56.9	57.0	56.0	51.9	53.8
1-Hexanol	4	58.2	63.0	63.0	62.3	59.2	60.6	56.3	61.7	61.8	60.7	56.6	59.0
1-Heptanol	4	70.5	73.8	73.8	73.4	71.2	72.1	69.7	73.2	73.3	72.7	69.9	71.3
1-Octanol	4	67.2	71.0	70.9	70.4	67.9	69.0	66.1	70.3	70.4	69.5	66.3	68.2
1-Nonanol	4	68.2	71.7	71.7	71.3	68.9	69.8	67.7	71.4	71.5	70.8	67.9	69.3
1-Decanol	4	76.7	79.3	79.3	79.0	77.2	77.9	76.1	79.0	79.1	78.5	76.2	77.5
1-Undecanol	1	73.7	76.6	76.6	76.2	74.2	75.0	73.1	76.3	76.4	75.7	73.2	74.6
1-Dodecanol	1	51.5	56.7	56.8	56.0	52.3	53.7	50.4	56.2	56.4	55.3	50.5	53.1
1,2-Ethanediol	1	51.4	47.7	45.6	45.6	54.5	50.2	32.8	23.9	21.4	23.1	37.3	28.1
1,2-Propanediol	1	339.5	324.2	333.3	339.6	322.4	326.7	396.8	389.2	396.9	401.0	383.7	393.7
Dipropylene Glycol	1	0.2	9.4	8.4	6.6	0.2	8.0	1.1	11.3	11.3	9.7	2.4	8.5
1,3-Butanediol	1	23.4	21.0	24.3	27.9	19.1	19.2	41.2	40.5	43.3	47.4	38.2	38.8
1,4-Butanediol	1	66.8	59.1	63.2	68.6	60.8	59.1	89.4	83.5	86.9	92.8	85.3	84.3
Oleic Acid	1	8.9	2.9	2.2	11.9	17.1	4.8	83.6	106.8	105.9	83.6	69.8	104.9
Ethyl Acetate	2	16.7	26.4	26.7	30.9	19.1	19.9	44.2	50.7	51.3	56.9	46.2	45.2
3-Methyl-Acetate 1-Butanol	1	34.5	41.5	41.8	44.4	35.5	37.1	52.0	57.2	57.8	61.6	53.0	53.2
Ethyl Propionate	1	19.4	28.8	29.2	33.3	21.4	22.3	44.0	50.3	51.0	56.6	45.9	44.6
Ethyl-N-Butyrate	1	28.4	36.8	37.3	40.8	30.0	31.0	49.3	55.2	55.9	60.7	50.9	50.0
Ethylene Glycol Diacetate	1	9.9	14.8	14.1	18.9	12.5	11.4	42.3	45.4	45.7	52.7	44.4	42.4

Methyl Benzoate	1	75.6	76.4	76.5	78.8	76.3	74.9	82.5	82.8	83.0	85.9	83.3	81.5
Isopropyl Myristate	1	60.8	65.1	65.5	67.5	61.3	61.8	70.0	73.3	73.8	76.6	70.8	70.3
Carbon Tetrachloride	17	39.7	41.5	42.3	45.1	40.9	37.9	36.2	38.3	38.9	41.7	37.8	34.9
Chloroform	20	379.6	404.3	402.2	372.0	361.6	409.7	304.5	330.7	328.8	296.2	283.5	333.3
1,2-Dichloroethane	14	40.1	41.0	41.2	42.0	40.8	40.5	39.1	40.5	40.4	39.4	39.5	40.2
Acetonitrile	5	66.6	66.9	66.1	68.0	68.1	66.4	74.9	75.0	74.7	78.0	76.1	74.3
N,N-Dimethylformamide	1	5037.2	4196.6	4265.3	4535.7	5024.0	4337.5	6723.5	5299.4	5311.2	5898.8	6929.8	5534.9
Propylene Carbonate	2	65.7	66.8	67.6	61.7	63.5	67.7	53.6	52.4	52.5	60.8	55.7	51.4
Triacetin	1	57.5	57.5	57.1	60.1	58.5	57.2	72.8	72.6	72.7	76.9	73.7	72.1
Formamide	1	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
N-Methyl-Formamide	1	28.3	13.7	17.0	22.9	24.1	16.1	33.7	18.6	20.7	26.8	32.2	22.5
2-Methoxyethanol	1	53.3	76.0	80.5	79.6	53.2	57.4	73.9	112.3	116.5	108.6	73.1	88.6
2-Ethoxyethanol	1	68.9	70.3	70.0	70.8	70.5	69.0	65.5	65.1	65.0	66.9	67.5	63.8
2-(2-Ethoxyethoxy)-Ethanol	1	65.3	66.4	66.0	67.1	67.0	65.1	62.2	60.9	60.7	63.3	64.3	59.7
Diethylene Glycol Butyl Ether	1	76.3	77.3	77.1	77.8	77.4	76.2	74.7	74.2	74.1	75.8	76.0	73.2
2-Ethoxyethyl Acetate	1	33.1	40.1	40.2	43.5	34.9	35.4	55.5	60.1	60.6	65.3	57.0	56.3
2-(2-Ethoxyethoxy)-Acetate-Ethanol	1	43.6	48.8	48.8	51.7	45.0	45.1	62.4	65.9	66.3	70.4	63.7	62.9
Ethyl Ester Octanoic Acid	1	56.3	61.0	61.5	64.1	57.2	57.1	67.4	70.7	71.2	74.8	68.5	67.2
Ethyl Ester Dodecanoic Acid	1	65.1	68.8	69.2	71.3	65.7	65.7	73.6	76.2	76.6	79.5	74.4	73.4
Ethyl Ester Decanoic Acid	1	60.3	64.5	65.0	67.3	61.0	61.0	70.1	73.1	73.6	76.8	71.0	69.9
2-Methoxy-Acetate Ethanol	1	17.5	25.1	24.9	29.1	19.8	20.0	46.4	51.2	51.7	57.6	48.2	47.1
Ethyl Ester Nonanoic Acid	1	59.3	63.6	64.0	66.5	60.1	60.0	69.5	72.5	73.0	76.3	70.5	69.2
Ethyl Ester Hexanoic Acid	1	45.4	51.4	51.9	55.1	46.6	46.6	59.6	63.7	64.3	68.7	61.0	59.4
Ethyl Myristate	1	67.9	71.2	71.6	73.5	68.4	68.3	75.5	77.9	78.3	81.0	76.2	75.3
2,3-Butanediol	1	3025.4	3489.2	3452.2	2818.0	2712.1	3595.5	2644.6	3152.5	3114.2	2439.2	2341.8	3183.5
Glycerol Tricaprylate	1	73.3	74.8	75.0	77.1	73.8	73.1	80.6	81.4	81.6	84.5	81.2	79.9
Ethyl Ester Pentanoic Acid	1	43.4	49.5	50.0	53.4	44.7	44.6	58.3	62.5	63.2	67.7	59.8	58.1
Ethyl Ester Undecanoic Acid	1	68.3	71.7	72.0	73.9	68.8	68.8	76.0	78.4	78.8	81.4	76.8	75.9
Ethyl Ester Hexadecanoic Acid	1	67.9	71.1	71.5	73.5	68.3	68.2	75.3	77.7	78.1	80.9	76.0	75.1
Average		217.3	209.1	210.2	205.4	210.5	213.5	249.0	231.6	231.5	231.6	246.4	236.7

Table A.7: Error summary of predicted solubility in binary solvents for 37 systems with consideration for multiple conformations of acetaminophen, ethanol, and cyclohexane.

Solvent Name	Solvent Name	Solute Name	Points	W _{hb} [Lin 2002]		W _{hb} [Mathias 2002]	
				RMSE	AA%E	RMSE	AA%E
Water	Ethanol	Sulfanilamide	10	1.0880	1154.9	1.1490	1370.6
Water	Ethanol-A	Sulfanilamide	10	1.0532	1064.5	1.1079	1252.2
Cyclohexane	Ethanol	Methyl-P-Hydroxybenzoate	12	0.6617	369.8	0.8335	606.0
Cyclohexane	Ethanol-A	Methyl-P-Hydroxybenzoate	12	0.5790	285.2	0.7648	499.5
Cyclohexane-A	Ethanol	Methyl-P-Hydroxybenzoate	12	0.6601	368.0	0.8339	606.2
Cyclohexane-A	Ethanol-A	Methyl-P-Hydroxybenzoate	12	0.5769	283.1	0.7629	496.8
Water	Ethanol	Methyl-P-Hydroxybenzoate	12	0.9656	892.2	2.0892	15550.9
Water	Ethanol-A	Methyl-P-Hydroxybenzoate	12	0.9671	896.3	2.0808	15239.0
1,4-Dioxane	Water	Acetanilide	45	0.2238	48.0	0.2145	56.4
Cyclohexane	Ethanol	Acetaminophen-A	12	1.1394	1567.3	1.3945	2982.2
Cyclohexane	Ethanol	Acetaminophen-B	12	1.1411	1575.5	1.4115	3106.1
Cyclohexane	Ethanol	Acetaminophen-C	12	1.0589	1271.2	1.3582	2739.9
Cyclohexane	Ethanol-A	Acetaminophen-A	12	0.9723	999.7	1.1818	1763.5
Cyclohexane	Ethanol-A	Acetaminophen-B	12	0.9708	995.9	1.1939	1817.5
Cyclohexane	Ethanol-A	Acetaminophen-C	12	0.9018	819.7	1.1476	1621.9
Cyclohexane-A	Ethanol	Acetaminophen-A	12	1.1389	1564.9	1.3920	2963.5
Cyclohexane-A	Ethanol	Acetaminophen-B	12	1.1408	1573.9	1.4095	3090.3
Cyclohexane-A	Ethanol	Acetaminophen-C	12	1.0583	1268.5	1.3555	2720.8
Cyclohexane-A	Ethanol-A	Acetaminophen-A	12	0.9711	995.6	1.1783	1747.1
Cyclohexane-A	Ethanol-A	Acetaminophen-B	12	0.9698	992.5	1.1911	1803.3
Cyclohexane-A	Ethanol-A	Acetaminophen-C	12	0.9005	815.7	1.1439	1605.4
Acetone	Toluene	Acetaminophen-A	86	0.8582	644.8	0.9746	928.9
Acetone	Toluene	Acetaminophen-B	86	0.8093	560.5	0.9472	852.8
Acetone	Toluene	Acetaminophen-C	86	0.7409	466.0	0.9202	810.9
Ethanol	Ethyl Acetate	Acetaminophen-A	11	0.3197	106.6	0.3941	147.6
Ethanol	Ethyl Acetate	Acetaminophen-B	11	0.3049	99.5	0.3965	149.0
Ethanol	Ethyl Acetate	Acetaminophen-C	11	0.2323	66.1	0.3452	120.7
Ethanol-A	Ethyl Acetate	Acetaminophen-A	11	0.2512	71.3	0.2979	96.2
Ethanol-A	Ethyl Acetate	Acetaminophen-B	11	0.2310	63.2	0.2955	95.4
Ethanol-A	Ethyl Acetate	Acetaminophen-C	11	0.1706	36.5	0.2442	71.9
1,4-Dioxane	Water	Acetaminophen-A	50	0.1794	30.0	0.1369	26.4

1,4-Dioxane	Water	Acetaminophen-B	50	0.1531	26.7	0.1845	40.1
1,4-Dioxane	Water	Acetaminophen-C	50	0.2601	39.4	0.1061	21.5
Water	Acetone	Acetaminophen-A	87	0.1421	29.2	0.2652	79.5
Water	Acetone	Acetaminophen-B	87	0.1337	27.2	0.3194	104.5
Water	Acetone	Acetaminophen-C	87	0.1729	31.6	0.1677	39.7
Water	Ethanol	Acetaminophen-A	24	0.2017	55.5	0.5581	263.1
Water	Ethanol	Acetaminophen-B	24	0.2615	80.0	0.6538	354.6
Water	Ethanol	Acetaminophen-C	24	0.1354	33.2	0.4190	162.0
Water	Ethanol-A	Acetaminophen-A	24	0.2021	55.5	0.5541	259.8
Water	Ethanol-A	Acetaminophen-B	24	0.2615	79.9	0.6492	349.8
Water	Ethanol-A	Acetaminophen-C	24	0.1392	34.3	0.4187	161.7
Cyclohexane	Ethanol	Methyl-P-Aminobenzoate	12	1.0940	1185.7	1.1532	1367.2
Cyclohexane	Ethanol-A	Methyl-P-Aminobenzoate	12	1.0123	964.3	1.0499	1055.0
Cyclohexane-A	Ethanol	Methyl-P-Aminobenzoate	12	1.0900	1174.8	1.1470	1346.7
Cyclohexane-A	Ethanol-A	Methyl-P-Aminobenzoate	12	1.0071	952.5	1.0419	1034.4
Water	Ethanol	Methyl-P-Aminobenzoate	12	1.0729	1092.8	1.1213	1235.0
Water	Ethanol-A	Methyl-P-Aminobenzoate	12	1.0804	1113.5	1.1274	1253.8
Cyclohexane	Ethanol	Ethyl-P-Aminobenzoate	12	0.7265	447.9	0.7730	511.4
Cyclohexane	Ethanol-A	Ethyl-P-Aminobenzoate	12	0.6674	375.8	0.6975	410.7
Cyclohexane-A	Ethanol	Ethyl-P-Aminobenzoate	12	0.7240	444.9	0.7684	504.7
Cyclohexane-A	Ethanol-A	Ethyl-P-Aminobenzoate	12	0.6641	372.2	0.6916	403.5
Water	Ethanol	Ethyl-P-Aminobenzoate	12	0.9148	741.8	0.9640	844.5
Water	Ethanol-A	Ethyl-P-Aminobenzoate	12	0.9270	765.4	0.9752	868.9
Benzene	Water	Naphthalene	2	1.2946	1870.8	1.6344	4210.5
Toluene	Water	Naphthalene	2	1.2843	1824.5	1.6244	4111.0
Acetone	Water	Naphthalene	4	1.1422	1306.9	1.4771	2951.1
Methyl Ethyl Ketone	Water	Naphthalene	4	1.0982	1190.7	1.4307	2693.7
Diethyl Ketone	Water	Naphthalene	2	1.2322	1607.1	1.5710	3624.8
Methanol	Water	Naphthalene	4	1.3988	2434.4	1.7314	5348.1
Ethanol	Water	Naphthalene	7	1.7862	4490.3	1.9641	8725.5
Ethanol-A	Water	Naphthalene	7	1.7916	4583.0	1.9728	8990.1
1-Propanol	Water	Naphthalene	4	1.7966	1797.1	1.8909	3900.4
1-Butanol	Water	Naphthalene	4	1.8158	1650.1	1.9007	3596.0
1-Pentanol	Water	Naphthalene	3	2.2642	1077.6	2.1688	2344.7
Cyclohexane	Ethanol	Propyl-4-Hydroxybenzoate	12	0.6879	405.5	0.7401	473.8
Cyclohexane	Ethanol-A	Propyl-4-Hydroxybenzoate	12	0.6631	376.1	0.7149	439.6

Cyclohexane-A	Ethanol	Propyl-4-Hydroxybenzoate	12	0.6879	405.5	0.7400	473.7
Cyclohexane-A	Ethanol-A	Propyl-4-Hydroxybenzoate	12	0.6630	376.0	0.7147	439.4
Water	Ethanol	Propyl-4-Hydroxybenzoate	10	2.1534	23409.0	3.4355	386200.0
Water	Ethanol-A	Propyl-4-Hydroxybenzoate	10	2.1599	23445.2	3.4338	384834.7
Cyclohexane	Ethanol	Phenacetin	12	1.5753	3900.6	1.6160	4291.1
Cyclohexane	Ethanol-A	Phenacetin	12	1.5380	3573.7	1.5599	3762.9
Cyclohexane-A	Ethanol	Phenacetin	12	1.5713	3865.8	1.6117	4248.0
Cyclohexane-A	Ethanol-A	Phenacetin	12	1.5332	3535.1	1.5546	3715.8
Water	Ethanol	Phenacetin	12	1.2600	1862.7	1.4106	2681.9
Water	Ethanol-A	Phenacetin	12	1.2819	1970.3	1.4345	2848.8
Cyclohexane	Ethanol	Propyl-P-Aminobenzoate	10	0.8324	598.8	0.8782	674.4
Cyclohexane	Ethanol-A	Propyl-P-Aminobenzoate	10	0.7885	533.6	0.8254	587.3
Cyclohexane-A	Ethanol	Propyl-P-Aminobenzoate	10	0.8287	593.4	0.8721	663.6
Cyclohexane-A	Ethanol-A	Propyl-P-Aminobenzoate	10	0.7839	527.5	0.8178	575.6
Water	Ethanol	Propyl-P-Aminobenzoate	11	1.2432	1687.8	1.3081	1971.1
Water	Ethanol-A	Propyl-P-Aminobenzoate	11	1.2601	1758.0	1.3245	2049.7
Ethanol	Ethyl Acetate	Sulfamethazine	12	1.5751	3858.7	1.5185	3455.5
Ethanol-A	Ethyl Acetate	Sulfamethazine	12	1.5017	3184.7	1.4115	2606.9
Water	Ethanol	Sulfamethazine	9	1.0589	1357.7	1.1196	1503.5
Water	Ethanol-A	Sulfamethazine	9	0.9715	1033.8	1.0208	1109.9
Cyclohexane	Ethanol	Methyl-4-Dimethylaminobenzoate	12	0.4890	208.7	0.4467	179.8
Cyclohexane	Ethanol-A	Methyl-4-Dimethylaminobenzoate	12	0.4986	215.6	0.4461	179.4
Cyclohexane-A	Ethanol	Methyl-4-Dimethylaminobenzoate	12	0.4777	200.7	0.4355	172.6
Cyclohexane-A	Ethanol-A	Methyl-4-Dimethylaminobenzoate	12	0.4869	207.2	0.4344	171.9
Water	Ethanol	Methyl-4-Dimethylaminobenzoate	12	0.8641	637.5	0.8669	641.0
Water	Ethanol-A	Methyl-4-Dimethylaminobenzoate	12	0.8902	683.6	0.8939	688.9
Cyclohexane	Ethanol	Methacetin	12	0.8450	634.9	0.9062	749.1
Cyclohexane	Ethanol-A	Methacetin	12	0.7881	542.2	0.8174	589.1
Cyclohexane-A	Ethanol	Methacetin	12	0.8389	624.1	0.8993	735.3
Cyclohexane-A	Ethanol-A	Methacetin	12	0.7810	531.2	0.8094	575.8
Water	Ethanol	Methacetin	12	0.4094	156.9	0.5357	247.2
Water	Ethanol-A	Methacetin	12	0.4245	166.6	0.5521	260.9
1,4-Dioxane	Water	Nalidixic Acid	60	0.7943	582.3	0.7541	501.0
<hr/>							
Average				0.8839	1435.5	1.0418	9454.1

Table A.8: COSMO-SAC predicted aqueous pure-solvent solubility error summary.

Index		CAS-RN	Formula	MW, g/mol	Points	W _{hb} [Lin 2002]		W _{hb} [Mathias 2002]	
No.	Solute Name					RMSE	AA%E	RMSE	AA%E
2	Fumaric Acid	110-17-8	C4H4O4	116.07	1	1.2741	1779.8	2.3466	22111.6
3	DL-Aspartic Acid	617-45-8	C4H7NO4	133.1	18	0.6645	363.2	1.3649	2249.7
4	L-Aspartic Acid	56-84-8	C4H7NO4	133.1	16	0.9229	737.4	1.6221	4101.8
5	3,5-Dichloropyridine	2457-47-8	C5H3CL2N	147.99	1	1.0168	90.4	0.8553	86.0
6	3-Hydroxy-5-Nitropyridine	5418-51-9	C5H4N2O3	140.1	1	2.8851	76661.0	3.4453	278724.7
7	5-Chloro-3-Pyridinol	74115-12-1	C5H4CLNO	129.54	1	1.7924	6100.8	3.0440	110559.2
8	2-Amino-3,5-Dichloropyridine	4214-74-8	C5H4CL2N2	163	1	0.1000	25.9	0.2178	65.1
9	2-Hydroxypyridine	142-08-5	C5H5NO	95.1	1	0.9331	757.3	1.2785	1799.1
10	3-Hydroxypyridine	109-00-2	C5H5NO	95.1	1	1.4876	2973.1	1.8528	7025.0
11	2-Amino-2-Nitropyridine	4214-76-0	C5H5N3O2	139.11	1	1.9094	8017.9	2.1962	15609.7
12	3-Amino-2-Chloropyridine	6298-19-7	C5H5CLN2	128.56	1	0.9518	795.0	1.0777	1096.0
13	2-Amino-5-Chloropyridine	1072-98-6	C5H5CLN2	128.56	1	0.8165	555.4	0.9394	769.7
14	Itaconic Acid	97-65-4	C5H6O4	130.1	22	0.7490	477.2	1.1355	1508.5
15	2-Aminopyridine	504-29-0	C5H6N2	94.11	1	2.2439	17433.8	2.3042	20044.5
16	3-Aminopyridine	504-24-5	C5H6N2	94.11	1	1.0854	1117.3	1.3147	1964.0
17	2-Amino-3-Hydroxypyridine	16867-03-1	C5H6N2O	110.11	1	1.5752	3660.0	1.6616	4487.4
18	DL-Glutamic Acid	617-65-2	C5H9NO4	147.13	16	0.2945	95.4	1.0335	984.5
21	6-Chloronicotinic Acid	5326-23-8	C6H4CLNO2	157.55	1	1.2147	1539.6	1.9365	8539.0
22	Isonicotinic Acid	55-22-1	C6H5NO2	123.11	1	1.3606	2194.2	1.7078	5002.4
23	Picolinic Acid	98-98-6	C6H5NO2	123.11	1	1.0410	999.0	1.2568	1706.4
24	Nicotinic Acid	59-67-6	C6H5NO2	123.11	1	1.2395	1635.8	1.5410	3375.3
25	2-Hydroxynicotinic Acid	609-71-2	C6H5NO3	139.11	1	2.1987	15701.8	2.8712	74229.9
28	Isonicotinamide	1453-82-3	C6H6N2O	122.12	1	1.3479	2127.7	1.3647	2215.7
29	Nicotinamide	98-92-0	C6H6N2O	122.12	1	1.5924	3812.0	1.5903	3793.4
30	4-Nitroaniline	100-01-6	C6H6N2O2	138.12	1	2.0133	10211.8	2.2249	16683.3
31	2-Methoxy-5-Nitropyridine	5446-92-4	C6H6N2O3	154.12	1	0.1152	23.3	0.1302	34.9
32	2-Amino-4-Methyl-3-Nitropyridine	6635-86-5	C6H7N3O2	153.14	1	0.4721	196.5	0.5995	297.7
33	2-Amino-4-Methylpyridine	695-34-1	C6H8N2	108.14	1	0.7041	405.9	0.9393	769.7
35	3,5-Dinitrosalicylic Acid	609-99-4	C7H4N2O7	228.12	16	0.1277	30.6	2.2364	22200.7
36	Tetrachloroguaiacol	2539-17-5	C7H4CL4O2	261.92	5	0.6766	381.8	1.1170	1248.5
37	M-Nitrobenzaldehyde	99-61-6	C7H5NO3	151.12	1	0.6895	389.2	0.8733	646.9
38	4-Iodobenzoic Acid	619-58-9	C7H5IO2	248.02	1	0.5372	244.5	1.2968	1880.6

39	4-Fluorobenzoic Acid	456-22-4	C7H5FO2	140.11	1	0.7778	499.6	1.4804	2922.8
40	4-Chlorobenzoic Acid	74-11-3	C7H5CLO2	156.57	1	0.2987	98.9	0.9825	860.6
41	3,4,5-Trichloroguaiacol	57057-83-7	C7H5CL3O2	227.47	5	0.1230	31.5	0.5052	220.2
42	4,5,6-Chloroguaiacol	2668-24-8	C7H5CL3O2	227.47	5	0.9272	748.8	1.3773	2295.3
44	Benzoic Acid	65-85-0	C7H6O2	122.12	2	1.1180	1216.0	2.8651	73381.0
45	Salicylic Acid	69-72-7	C7H6O3	138.12	2	1.4981	3162.6	3.2896	203831.5
46	4-Hydroxybenzoic Acid	99-96-7	C7H6O3	138.12	15	0.9865	911.0	2.0523	14498.8
47	3-Hydroxybenzoic Acid	99-06-9	C7H6O3	138.12	1	2.8961	78620.1	3.7196	524259.4
48	4,5-Dichloroguaiacol	2460-49-3	C7H6CL2O2	193.03	5	0.0827	13.4	0.4088	156.1
49	4,6-Dichloroguaiacol	16766-31-7	C7H6CL2O2	193.03	5	0.4100	157.0	0.7292	437.5
50	Benzamide	27208-38-4	C7H7NO	121.14	1	1.5034	3087.2	1.5882	3774.1
51	Methylnicotinate	93-60-7	C7H7NO2	137.14	1	0.4464	179.5	0.6994	400.5
52	P-Hydroxybenzamide	619-57-8	C7H7NO2	137.14	1	2.7791	60034.2	3.0188	104319.9
53	4-Aminobenzoic Acid	150-13-0	C7H7NO2	137.14	2	1.0057	913.5	1.3947	2382.2
54	4-Aminosalicylic Acid	65-49-6	C7H7NO3	153.14	9	1.8783	7633.0	2.6861	49883.0
55	2-Hydroxy-6-Methylnicotinic Acid	38116-61-9	C7H7NO3	153.14	1	1.0090	921.0	1.4779	2905.3
56	4-Chlorobenzyl Alcohol	873-76-7	C7H7CLO	142.58	1	0.5093	223.1	0.7943	522.8
57	4-Chloroguaiacol	16766-30-6	C7H7CLO2	158.58	1	0.4367	173.3	0.8398	591.5
58	5-Chloroguaiacol	3743-23-5	C7H7CLO2	158.58	1	0.0343	8.2	0.4141	159.5
59	4-Hydroxybenzyl Alcohol	623-05-2	C7H8O2	124.14	1	3.0487	111776.5	3.5363	343705.4
60	Theophylline	58-55-9	C7H8N4O2	180.16	2	0.1931	55.3	0.2672	84.6
61	Theobromine	83-67-0	C7H8N4O2	180.16	1	0.1604	30.9	0.5031	68.6
62	2,6-Pyridinedimethanol	1195-59-1	C7H9NO2	139.15	1	3.2640	183567.1	3.3461	221759.4
64	2-Amino-4,6-Dimethylpyridine	5407-87-4	C7H10N2	122.17	1	0.2802	90.6	0.4430	177.3
65	Simazine	122-34-9	C7H12CLN5	201.66	3	1.2885	1984.9	1.4652	3030.5
67	2-Methyl-3-Nitrobenzoic Acid	1975-50-4	C8H7NO4	181.15	1	1.0090	921.0	1.9325	8461.4
68	2-Chlorophenoxyacetic Acid	614-61-9	C8H7CLO3	186.59	1	1.5376	3348.2	3.6715	469224.6
69	5-Chlorovanillin	19463-48-0	C8H7CLO3	186.69	14	0.1956	43.4	0.2682	75.0
70	P-Toluic Acid	99-45-5	C8H8O2	136.15	42	0.9372	770.1	1.5058	3112.7
71	4-Hydroxyacetophenone	99-93-4	C8H8O2	136.15	1	1.6771	4654.4	2.9283	84674.7
72	M-Toluic Acid	99-04-7	C8H8O2	136.15	1	0.6245	321.2	2.7789	60007.2
73	P-Anisic Acid	100-09-4	C8H8O3	152.15	18	0.8704	645.1	1.3789	2305.6
74	Methyl-P-Hydroxybenzoate	99-76-3	C8H8O3	152.15	2	0.9368	769.6	1.8900	7854.1
75	M-Anisic Acid	586-38-9	C8H8O3	152.15	1	1.3753	2273.0	3.4213	263693.9
77	Phenoxyacetic Acid	122-59-8	C8H8O3	152.15	1	2.0036	9982.3	3.4066	254933.4
78	O-Anisic Acid	579-75-9	C8H8O3	152.15	20	0.9256	751.2	1.5905	5869.1

79	3-Hydroxy-4-Methyloxybenzoic Acid	645-08-9	C8H8O4	168.15	1	1.9305	8420.3	2.5279	33620.4
80	Acetanilide	103-84-4	C8H9NO	135.16	9	0.3505	118.8	0.5247	245.4
81	4-Aminoacetophenone	99-92-3	C8H9NO	135.16	1	0.9381	767.1	0.9372	765.3
82C	Acetaminophen	103-90-2	C8H9NO2	151.16	7	0.1030	19.8	0.2910	92.5
83	Methyl-P-Aminobenzoate	619-45-4	C8H9NO2	151.16	1	1.0160	937.5	1.0757	1090.4
84	Caffeine	58-08-2	C8H10N4O2	194.19	2	0.1662	29.5	0.5955	74.1
85	Acyclovir	59277-89-3	C8H11N5O3	225.2	3	2.0175	10392.8	2.0491	11178.5
86	Barbital	57-44-3	C8H12N2O3	184.19	1	0.6800	378.6	1.9386	8581.5
87	Atrazine	1912-24-9	C8H14IN5	215.68	4	0.5501	257.8	0.7583	485.6
88C	Acetylsalicylic Acid	50-78-2	C9H8O4	180.16	15	0.1795	28.8	0.1330	23.1
90	Ethyl-4-Hydroxybenzoate	120-47-8	C9H10O3	166.17	2	1.3236	2007.3	3.5835	383181.4
93	Ethyl-P-Aminobenzoate	94-09-7	C9H11NO2	165.19	2	0.8787	656.6	0.9377	766.9
95	Chloropyrifos	2921-88-2	C9H11CL3NO3PS	350.59	3	0.5075	218.0	0.8282	798.6
96	Ganciclovir	82410-32-0	C9H13N5O4	255.23	3	1.0835	1126.6	1.2059	1528.6
97	Cyanazine	21725-46-2	C9H13CLN6	240.69	3	0.1606	37.7	0.4517	183.2
98	Metharbital	50-11-3	C9H14N2O3	198.22	1	0.2213	39.9	0.0181	4.3
99	Naphthalene	91-20-3	C10H8	128.17	10	1.1614	1363.3	1.5019	3106.3
100	Sulfadiazine	68-35-9	C10H10N4O2S	250.28	1	0.0899	18.7	0.2226	67.0
101	Sulfamethoxazole	723-46-6	C10H11N3O3S	253.28	1	1.3379	2077.1	1.8742	7384.6
102	Propyl-4-Hydroxybenzoate	94-13-3	C10H12O3	180.2	2	1.2354	1698.4	3.0847	243479.2
103	Phenacetin	62-44-2	C10H13NO2	179.22	8	0.7447	455.9	0.8929	681.8
104	Propyl-P-Aminobenzoate	94-12-2	C10H13NO2	179.22	1	1.0878	1124.0	1.1562	1333.0
105	Ephedrine	299-42-3	C10H15NO	165.23	1	1.2667	94.6	1.0032	90.1
106	Camphor	76-22-2	C10H16O	152.23	1	0.3857	143.1	0.1934	56.1
107	Butabarbital	125-40-6	C10H16N2O3	212.24	1	1.2714	1768.1	2.3998	25008.9
108	Sulfapyridine	144-83-2	C11H11N3O2S	249.29	1	0.6401	336.6	1.0661	1064.4
109	Antipyrine	60-80-0	C11H12N2O	188.23	2	1.1424	92.8	0.6397	336.4
112	Sulfamethoxy pyridazine	80-35-3	C11H12N4O3S	280.3	5	1.2794	1804.9	1.7320	5301.6
114	Butyl-4-Hydroxybenzoate	94-26-8	C11H14O3	194.23	2	2.7659	200165.0	3.6423	1780100.4
115	Butyl-P-Aminobenzoate	94-25-7	C11H15NO2	193.24	2	1.2570	1718.2	1.3413	2107.3
116	Thiopental	76-75-5	C11H18N2O2S	242.34	1	0.2334	71.1	1.0757	1090.4
117	Amobarbital	57-43-2	C11H18N2O3	226.27	1	0.9323	755.8	1.7379	5369.0
118	Pentobarbital	76-74-4	C11H18N2O3	226.27	1	1.5093	3130.9	3.8731	746559.0
122	Biphenyl	92-52-4	C12H10	154.21	1	1.5835	3732.2	1.9870	9605.6
127	Diphenyl Sulfone	127-63-9	C12H10O2S	218.27	1	1.0023	905.2	1.2564	1704.5
128	Phenobarbital	50-06-6	C12H12N2O3	232.23	2	1.2559	1704.8	3.2108	162547.0

129	Sulfisomidine	515-64-0	C12H14N4O2S	278.33	1	1.6181	97.6	1.3001	95.0
130	Sulfamethazine	57-68-1	C12H14N4O2S	278.33	1	0.6515	348.2	0.9284	748.1
132	Thiamylal	77-27-0	C12H18N2O2S	254.35	1	0.1553	30.1	0.7710	490.3
137	Fluorene	86-73-7	C13H10	166.22	1	1.7109	5039.0	2.1209	13110.2
141C	Ibuprofen	15687-27-1	C13H18O	206.28	2	1.4856	3035.0	2.1296	13751.0
146	Mitotane	53-19-0	C14H10CL4	320.04	1	0.4474	180.2	0.9360	763.0
148B	Lidocaine	137-58-6	C14H22N2O	234.34	1	0.9214	88.0	0.7326	81.5
150	Flurbiprofen	5104-49-4	C15H13FO2	244.26	1	1.2955	1474.3	1.7274	13588.4
151	Mefenamic Acid	61-68-7	C15H15NO2	241.28	1	1.0924	91.9	0.1931	35.9
152	Meperidine	57-42-1	C15H21NO2	247.33	1	3.2347	99.9	2.9802	99.9
155	Morphine	57-27-2	C17H19NO3	285.34	2	0.9184	87.2	0.7258	79.9
156	Hydromorphone	466-99-9	C17H19NO3	285.34	1	2.2699	99.5	2.1608	99.3
157	Codeine	76-57-3	C18H21NO3	299.36	1	1.7138	98.1	1.4774	96.7
159	Estradiol	50-28-2	C18H24O2	272.38	1	0.5665	268.6	1.0051	911.8
163	Testosterone	58-22-0	C19H28O2	288.42	1	1.1291	1246.0	0.9298	750.8
164	Androstanolone	521-18-6	C19H30O2	290.44	1	0.7523	465.3	0.5642	266.6
165	Ethinyl Estradiol	57-63-6	C20H24O2	296.4	1	0.5681	269.9	1.2550	1699.0
166	Norethindrone	68-22-4	C20H26O2	298.42	1	0.0603	13.0	0.0812	17.1
169	Methyltestosterone	58-18-4	C20H30O2	304.45	19	0.5271	236.7	0.3932	147.2
171	Prostaglandin	363-24-6	C20H32O5	352.46	1	0.3372	54.0	0.0429	9.4
172	Prostaglandin F2Alpha	551-11-1	C20H34O5	354.48	1	0.7060	408.2	0.9661	825.0
173	Haloperidol	52-86-8	C21H23CLFNO2	375.86	1	0.3509	55.4	0.1160	23.4
174	Ethisterone	434-03-7	C21H28O2	312.44	1	1.5128	3157.1	1.5018	3075.7
175	Cortisone	53-06-5	C21H28O5	360.44	2	0.7034	80.1	0.8281	85.1
176	Prednisolone	50-24-8	C21H28O5	360.44	2	0.8214	84.9	0.9460	88.7
177	Progesterone	57-83-0	C21H30O2	314.46	1	0.6786	377.1	0.3930	147.2
178	Desoxycorticosterone	64-85-7	C21H30O3	330.46	1	0.0380	9.2	0.1407	27.7
179	Testosterone Acetate	1045-69-8	C21H30O3	330.46	1	2.5277	33601.7	2.3305	21303.0
181	Hydrocortisone	50-23-7	C21H30O5	362.46	1	0.6059	75.2	0.6817	79.2
182	Prednisone	53-03-2	C21H26O4	358.55	1	0.4406	63.7	0.5334	70.7
183	Norethindrone Acetate	51-98-9	C22H28O3	340.45	1	1.0586	1044.5	0.9860	868.2
184	Fentanyl	437-38-7	C22H28N2O	336.47	1	0.2952	97.3	0.7443	455.0
185	Betamethasone	378-44-9	C22H29FO5	392.46	1	0.2187	39.6	0.2170	39.3
186	Dexamethasone	50-02-2	C22H29FO5	392.46	1	1.6450	97.7	1.6170	97.6
187	Sufentanil	56030-54-7	C22H30N2O2S	386.55	1	2.7597	99.8	2.3385	99.5
188	Testosterone Propionate	57-85-2	C22H32O3	344.49	1	1.6172	4042.1	1.4557	2755.4

189	Prednisone Acetate	125-10-0	C23H28O6	400.46	1	0.6047	75.2	0.6880	79.5
190	Cortisone Acetate	50-04-4	C23H30O6	402.48	1	0.6990	80.0	0.8154	84.7
191	Desoxycorticosterone Acetate	56-47-3	C23H32O4	372.5	1	2.5366	34306.0	2.3683	23251.2
192	Hydrocortisone Acetate	50-03-3	C23H32O6	404.5	1	0.8546	86.0	0.8635	86.3
193	Dexamethasone Acetate	1177-87-3	C24H31FO6	434.5	1	1.9991	99.0	1.9784	98.9
194	Betamethasone Acetate	987-24-6	C24H31FO6	434.5	1	1.2364	94.2	1.1696	93.2
195	Estradiol Benzoate	50-50-0	C25H28O3	376.49	1	1.4340	96.3	1.1661	93.2
Average					458	1.0624	6681.9	1.4495	46754.6

Table A.9: COSMO-SAC Predicted n-octanol pure-solvent solubility error summary for all solutes.

Index No.	Solute Name	CAS	MW, g/mol	W _{hb} [Lin 2002]		W _{hb} [Mathias 2002]	
				RMSE	AA%E	RMSE	AA%E
2	Fumaric Acid	110-17-8	116.07	1.0905	1131.6	1.4196	2528.1
44	Benzoic Acid	65-85-0	122.12	0.4655	192.4	0.5696	272.3
45	Salicylic Acid	69-72-7	138.12	0.3666	132.5	0.4613	189.4
46	4-Hydroxybenzoic Acid	99-96-7	138.12	0.2061	60.7	0.4186	162.2
53	4-Aminobenzoic Acid	150-13-0	137.14	0.7654	482.6	0.9232	737.9
60	Theophylline	58-55-9	180.16	0.4726	196.9	1.0712	1078.1
74	Methyl-P-Hydroxybenzoate	99-76-3	152.15	0.2565	80.2	0.4321	170.5
80	Acetanilide	103-84-4	135.16	0.1577	43.8	0.2489	77.4
82C	Acetaminophen	103-90-2	151.16	0.3311	114.3	0.6419	338.4
83	Methyl-P-Aminobenzoate	619-45-4	151.16	0.5797	280.0	0.6227	319.5
84	Caffeine	58-08-2	194.19	0.5491	254.1	0.4570	186.4
85	Acyclovir	59277-89-3	225.2	1.3000	1895.4	1.8155	6438.8
86	Barbital	57-44-3	184.19	0.9515	794.3	1.2308	1601.2
88C	Acetylsalicylic Acid	50-78-2	180.16	0.2694	46.2	0.2679	46.0
90	Ethyl-4-Hydroxybenzoate	120-47-8	166.17	0.3605	129.1	0.4515	182.8
92	Diuron	330-54-1	233.09	0.0013	0.3	0.0632	15.7
93	Ethyl-P-Aminobenzoate	94-09-7	165.19	0.4528	183.7	0.4877	207.4
94	Monuron	150-68-5	198.65	0.0120	2.8	0.0512	12.5
96	Ganciclovir	82410-32-0	255.23	0.7385	447.6	1.3736	2263.9
102	Propyl-4-Hydroxybenzoate	94-13-3	180.2	0.3129	105.2	0.3768	137.9
103	Phenacetin	62-44-2	179.22	1.1403	1281.4	1.2036	1498.1
109	Antipyrine	60-80-0	188.23	0.2856	93.0	0.3189	108.4
114	Butyl-4-Hydroxybenzoate	94-26-8	194.23	0.2167	63.8	0.2487	76.4
115	Butyl-P-Aminobenzoate	94-25-7	193.24	0.4221	164.3	0.4215	163.9
121	Carbazole	86-74-8	167.21	1.0308	973.6	1.2236	1573.3
122	Biphenyl	92-52-4	154.21	0.3876	143.9	0.4207	163.4
123	Acenaphthene	83-32-9	154.21	0.3653	131.9	0.4001	151.2
127	Diphenyl Sulfone	127-63-9	218.27	1.0167	939.1	1.0568	1039.7
128	Phenobarbital	50-06-6	232.23	1.1352	1265.1	1.4207	2534.6
134	Thioxanthone	492-22-8	212.67	0.1366	27.0	0.1627	31.3
136	Niflumic Acid	4294-00-7	282.22	3.3814	240579.7	3.7768	597996.2
137	Fluorene	86-73-7	166.22	0.4546	184.8	0.4947	212.4
141C	Ibuprofen	15687-27-1	206.28	0.5959	332.2	0.7450	651.1

143	Phenanthrene	85-01-8	178.23	0.4628	190.3	0.5029	218.3
144	Anthracene	120-12-7	178.23	0.3839	142.0	0.4290	168.5
145	Diphenylglyoxal	134-81-6	210.23	0.6136	310.8	0.6503	347.0
149	Piroxicam	36322-90-4	331.35	1.0927	1137.9	1.4758	2891.0
150	Flurbiprofen	5104-49-4	244.26	0.7127	416.1	0.8016	533.3
153	Pyrene	129-00-0	202.25	0.5086	222.6	0.5529	257.2
161	Nandrolone	434-22-0	274.4	0.3101	51.0	0.3468	55.0
163	Testosterone	58-22-0	288.42	0.2739	46.8	0.3124	51.3
167	Methanediene	72-63-9	300.43	0.4439	64.0	0.4705	66.2
169	Methyltestosterone	58-18-4	304.45	0.3532	55.6	0.3968	59.9
170	Mestanolone	521-11-9	304.47	0.3023	50.1	0.3288	53.1
171	Prostaglandin	363-24-6	352.46	0.1829	34.4	0.0327	7.8
175	Cortisone	53-06-5	360.44	0.7988	84.1	0.5677	72.9
176	Prednisolone	50-24-8	360.44	2.2723	99.5	2.0412	99.1
177	Progesterone	57-83-0	314.46	1.3034	1910.9	1.3015	1902.0
178	Desoxycorticosterone	64-85-7	330.46	0.4864	67.4	0.5408	71.2
181	Hydrocortisone	50-23-7	362.46	0.0759	16.0	0.0125	2.9
197A	Cholesterol	57-88-5	386.65	0.4989	67.2	0.4850	66.1
199	Sitosterol	83-46-5	414.7	1.1278	92.5	1.1133	92.3
Average				0.6233	4957.9	0.7239	12118.9

Table A.10: Comparison of COSMO-SAC predicted solute mole fraction in ethanol as a pure-solvent using the sigma profile from both ethanol conformations using the both exchange energy expressions to literature solute mole fractions.

Index No.	Ethanol Conformation Exchange Energy			Released		Conformation A		Released		Conformation A	
	Solute Name	CAS-RN	Pts.	W _{hb} [Lin 2002] RMS	AA%E	W _{hb} [Lin 2002] RMS	AA%E	W _{hb} [Mathias 2002] RMS	AA%E	W _{hb} [Mathias 2002] RMS	AA%E
30	4-Nitroaniline	100-01-6	1	1.0919	1135.6	1.0219	951.8	1.1112	1191.9	1.0226	953.4
34	Sulfanilamide	63-74-1	1	1.3979	2399.8	2.0491	11096.9	1.4111	2477.2	1.2959	1876.4
44	Benzoic Acid	65-85-0	1	0.3671	132.9	0.3212	109.5	0.4639	191.0	0.4265	167.0
45	Salicylic Acid	69-72-7	1	0.5356	243.2	0.4979	214.7	0.6259	322.6	0.6007	298.7
46	4-Hydroxybenzoic Acid	99-96-7	1	0.2341	71.4	0.1437	39.2	0.3896	145.2	0.3186	108.3
74	Methyl-P-Hydroxybenzoate	99-76-3	4	0.2592	81.5	0.1953	56.6	0.3919	146.5	0.3382	117.7
82C	Acetaminophen	103-90-2	8	0.2225	66.4	0.0896	19.7	0.4209	163.6	0.2512	78.0
86	Barbital	57-44-3	1	0.8344	582.9	0.7352	443.5	1.0197	946.4	0.9450	781.1
88C	Acetylsalicylic Acid	50-78-2	1	0.3732	57.7	0.4161	61.6	0.3429	54.6	0.3888	59.1
90	Ethyl-4-Hydroxybenzoate	120-47-8	4	0.3670	132.7	0.3320	114.6	0.4417	176.5	0.4113	157.8
98	Diuron	330-54-1	1	0.4513	182.7	0.3955	148.6	0.4974	214.4	0.4137	159.2
94	Monuron	150-68-5	1	0.4188	162.3	0.3756	137.5	0.4524	183.4	0.3889	144.8
98	Metharbital	50-11-3	1	0.5360	243.5	0.4162	160.7	0.9498	790.8	0.7833	507.2
99	Naphthalene	91-20-3	1	0.5858	285.3	0.5794	279.7	0.6426	339.1	0.6410	337.6
102	Propyl-4-Hydroxybenzoate	94-13-3	4	0.3285	112.5	0.3034	100.4	0.3832	141.3	0.3607	129.0
107	Butabarbital	125-40-6	1	1.0362	987.0	0.9622	816.7	1.1776	1405.2	1.1184	1213.6
109	Antipyrine	60-80-0	1	0.2426	74.8	0.2951	97.3	0.2918	95.8	0.3618	130.0
112	Sulfamethoxypyridazine	80-35-3	5	2.1325	13468.3	2.0504	11131.5	2.1834	15153.0	2.0846	12051.3
114	Butyl-4-Hydroxybenzoate	94-26-8	4	0.2020	58.8	0.1890	54.1	0.2309	69.7	0.2182	64.8
116	Thiopental	76-75-5	1	0.8895	675.3	0.7517	464.6	1.1776	1405.3	1.0624	1054.6
117	Amobarbital	57-43-2	1	0.6395	336.0	0.5654	267.6	0.7860	510.9	0.7251	431.1
118	Pentobarbital	76-74-4	1	0.7385	447.7	0.6903	390.1	0.8347	583.5	0.7936	521.7
121	Carbazole	86-74-8	1	1.4502	2719.9	1.3323	2049.3	1.6082	3956.9	1.4764	2894.8
122	Biphenyl	92-52-4	1	0.7579	472.6	0.7528	466.0	0.8188	558.9	0.8181	557.8
123	Acenaphthene	83-32-9	1	0.8511	609.7	0.8541	614.6	0.9187	729.2	0.9276	746.4
125	1-Acetyl-2-Naphthol	574-19-6	8	0.2678	83.6	0.2618	81.2	0.3047	99.9	0.2980	96.8
126	2-Acetyl-1-Naphthol	711-79-5	16	0.9531	808.2	0.9484	797.5	0.9997	914.9	0.9956	904.7
127	Diphenyl Sulfone	127-63-9	1	1.1961	1470.8	1.1671	1369.2	1.2618	1727.3	1.2360	1621.8
128	Phenobarbital	50-06-6	1	0.8639	630.9	0.7704	489.4	1.0391	994.3	0.9719	837.3

130	Sulfamethazine	57-68-1	1	2.0557	11269.6	1.9331	8472.2	2.0590	11355.6	1.9016	7873.4
132	Thiamylal	77-27-0	1	0.7166	420.7	0.6241	320.8	0.9051	703.7	0.8277	572.5
133	Tolbutamide	64-77-7	1	1.2274	1587.9	1.1711	1383.0	1.3300	2037.9	1.2769	1791.9
134	Thioxanthone	492-22-8	1	0.5049	219.8	0.5194	230.7	0.5039	219.1	0.5110	224.3
136	Niflumic Acid	4294-00-7	2	1.0939	1200.7	0.9796	895.2	1.4527	2903.3	1.3551	2292.7
141C	Ibuprofen	15687-27-1	1	0.5469	252.3	0.5163	228.3	0.6149	312.0	0.5884	287.6
143	Phenanthrene	85-01-8	1	1.0732	1083.6	1.0629	1055.7	1.1415	1285.0	1.1372	1271.4
145	Diphenylglyoxal	134-81-6	1	0.9318	754.8	0.9175	727.0	0.9939	886.0	0.9818	858.8
149	Piroxicam	36322-90-4	1	2.2049	15929.6	2.0260	10516.7	2.5100	32258.0	2.2644	18284.2
151	Mefenamic Acid	61-68-7	1	1.1989	1481.0	1.0554	1036.0	1.6968	4875.1	1.5222	3228.4
153	Pyrene	129-00-0	1	1.2261	1582.9	1.2168	1547.5	1.3053	1919.9	1.3037	1912.4
158	Estrone	53-16-7	5	0.5094	223.9	0.4175	160.7	0.8241	572.1	0.6419	341.1
160	Estriol	50-27-1	5	0.5367	246.1	0.3385	111.9	0.8930	705.9	0.5593	266.0
162	Morniflumate	65847-85-0	1	0.9478	786.8	0.9798	854.6	0.9970	893.1	1.0288	968.5
173	Haloperidol	52-86-8	1	1.2355	1620.0	1.2203	1560.7	1.2839	1822.8	1.2479	1669.9
179	Testosterone Acetate	1045-69-8	1	1.4549	2750.5	1.4740	2878.4	1.4658	2822.7	1.4959	3032.5
188	Testosterone Propionate	57-85-2	1	0.6093	306.7	0.7103	413.3	0.6319	328.5	0.7563	470.5
197B	Cholesterol	57-88-5	6	0.3943	148.1	1.4202	2661.2	0.4357	174.2	0.5473	257.8
198	Cholesterol Acetate	604-35-3	6	0.2973	97.1	0.4082	155.6	0.3596	128.2	0.4703	195.5
199	Sitosterol	83-46-5	5	0.1612	29.3	0.0779	13.8	0.1209	21.7	0.0525	11.4
200	Cholesteryl Benzoate	604-32-0	3	0.6346	340.5	0.7327	457.6	0.7232	445.2	0.8250	594.6
Average				0.7957	1421.3	0.7853	1374.1	0.9085	2047.2	0.8594	1508.1

Table A.11: Comparison of COSMO-SAC predicted solute mole fraction in cyclohexane as a pure-solvent using the sigma profile from both cyclohexane conformations using the exchange energy expression defined by Lin⁶ to literature solute mole fractions.

Cyclohexane Conformation			Released		Conformation A		Released		Conformation A	
Exchange Energy			W _{hb} [Lin 2002]		W _{hb} [Lin 2002]		W _{hb} [Mathias 2002]		W _{hb} [Mathias 2002]	
Solute Name	CAS-RN	Lit. Pts.	RMSE	AA%E	RMSE	AA%E	RMSE	AA%E	RMSE	AA%E
Picric Acid	88-89-1	3	1.4602	2785.3	1.4104	2472.6	1.4743	2880.6	1.4245	2557.6
Benzoic Acid	65-85-0	2	0.1488	29.0	0.1893	35.3	0.0356	7.8	0.0757	16.0
Salicylic Acid	69-72-7	1	1.1965	1472.0	1.1586	1340.9	1.0957	1146.5	1.0576	1041.7
Phthalic Acid Anhydride	85-44-9	1	1.0922	1136.6	1.0550	1034.9	1.0940	1141.8	1.0568	1039.7
M-Toluic Acid	99-04-7	1	0.2939	96.7	0.2548	79.8	0.4180	161.8	0.3802	140.0
Acetanilide	103-84-4	1	1.5210	3219.0	1.4735	2875.0	1.7200	5147.6	1.6723	4602.2
Acetylsalicylic Acid	50-78-2	1	2.0571	11304.1	2.0124	10188.8	2.0670	11567.4	2.0223	10426.3
P-Tolylacetic Acid	622-47-9	1	0.8993	693.1	0.8730	646.4	0.8993	693.1	0.8730	646.4
Diuron	330-54-1	1	1.3375	2075.1	1.3040	1913.6	1.4816	2931.2	1.4482	2706.5
Monuron	150-68-5	1	1.1412	1284.3	1.1019	1164.5	1.2646	1738.9	1.2251	1579.1
Dibenzothiophene	132-65-0	6	0.3806	138.7	0.3790	137.8	0.3806	138.7	0.3790	137.8
Dibenzofuran	132-64-9	6	0.0357	6.7	0.0360	6.8	0.0357	6.7	0.0360	6.8
Carbazole	86-74-8	5	0.7358	448.4	0.7112	417.9	0.7349	447.3	0.7103	416.8
Biphenyl	92-52-4	1	0.2113	62.7	0.2068	61.0	0.2113	62.7	0.2068	61.0
Acenaphthene	83-32-9	1	0.3401	118.8	0.3336	115.6	0.3401	118.8	0.3336	115.6
1-Acetyl-2-Naphthol	574-19-6	10	0.8130	564.0	0.8086	556.5	0.8130	564.1	0.8086	556.5
2-Acetyl-1-Naphthol	711-79-5	11	0.7747	499.6	0.7640	484.3	0.7747	499.6	0.7640	484.3
Diphenyl Sulfone	127-63-9	1	1.4796	2917.2	1.4497	2716.6	1.4796	2917.2	1.4497	2716.6
Thioxanthone	492-22-8	1	0.3645	131.5	0.3423	119.9	0.3697	134.3	0.3475	122.6
Acridine	260-94-6	6	0.6287	333.6	0.6236	328.1	0.6285	333.4	0.6234	327.9
Niflumic Acid	4294-00-7	1	1.6722	4601.0	1.6461	4326.8	1.4899	2989.4	1.4644	2813.1
Thioxanthene	261-31-4	6	0.3858	143.0	0.3828	141.4	0.3858	143.0	0.3828	141.4
Xanthene	92-83-1	6	0.1664	41.1	0.1659	41.0	0.1664	41.1	0.1659	41.0
Ibuprofen	15687-27-1	1	0.0869	18.1	0.0825	17.3	0.0506	11.0	0.0485	10.6
Phenanthrene	85-01-8	1	0.7254	431.3	0.7189	423.5	0.7254	431.3	0.7189	423.5
Anthracene	120-12-7	7	0.5731	274.2	0.5677	269.6	0.5731	274.2	0.5677	269.6
Diphenylglyoxal	134-81-6	1	0.7256	431.6	0.7022	403.7	0.7256	431.6	0.7022	403.7
Piroxicam	36322-90-4	1	1.2775	94.7	1.3347	95.4	1.7575	98.3	1.8190	98.5
Morphine	57-27-2	1	1.9039	7915.6	1.8917	7692.8	2.0047	10007.8	1.9945	9774.0
Estrone	53-16-7	4	0.3981	143.0	0.3862	134.8	0.3942	140.4	0.3850	135.7

Testosterone	58-22-0	1	0.6827	381.6	0.6915	391.5	0.8067	540.8	0.8173	556.6
Androstanolone	521-18-6	1	0.6238	320.5	0.6359	332.4	0.7477	459.3	0.7612	477.0
Nandrolone Acetate	1425-10-1	1	0.0786	16.6	0.0800	16.8	0.0786	16.6	0.0800	16.8
Methyltestosterone	58-18-4	1	0.4875	207.3	0.5014	217.2	0.5701	271.6	0.5848	284.4
Testosterone Acetate	1045-69-8	1	2.0693	11629.7	2.0718	11697.7	2.0693	11629.7	2.0718	11697.7
Nandrolone Propionate	7207-92-3	1	0.3921	146.7	0.3992	150.7	0.3921	146.7	0.3992	150.7
Hydrocortisone	50-23-7	1	0.5728	273.9	0.5505	255.2	0.5213	232.1	0.5000	216.2
Testosterone Propionate	57-85-2	1	0.8048	537.9	0.8175	556.8	0.8048	537.9	0.8175	556.8
Cholesterol	57-88-5	22	0.3184	49.4	0.2409	42.3	0.2792	47.3	0.2242	44.0
Average			0.7912	1460.9	0.7783	1382.1	0.8169	1566.4	0.8051	1482.4

Table A.12: COSMO-SAC Predicted nitrogen-containing solute pure-solvent solubility error summary.

Index		Literature		Whb [Lin 2002]		Whb [Mathias 2002]		
No.	Solute Name (Solvent Pairs)	CAS-RS	MW, g/mol	Points	RMSE	AA%E	RMSE	AA%E
3	DL-Aspartic Acid (1)	617-45-8	133.10	18	0.6645	363.2	1.3649	2249.7
4	L-Aspartic Acid (1)	56-84-8	133.10	16	0.9229	737.4	1.6221	4101.8
5	3,5-Dichloropyridine (1)	2457-47-8	147.99	1	1.0168	90.4	0.8553	86.0
6	2-Hydroxy-5-Nitropyridine (1)	5418-51-9	140.10	1	2.8851	76661.0	3.4453	278724.7
7	5-Chloro-3-Pyridinol (1)	74115-12-1	129.54	1	1.7924	6100.8	3.0440	110559.2
8	2-Amino-3,5-Dichloropyridine (1)	4214-74-8	163.00	1	0.1000	25.9	0.2178	65.1
9	2-Hydroxypyridine (1)	142-08-5	95.10	1	0.9331	757.3	1.2785	1799.1
10	3-Hydroxypyridine (1)	109-00-2	95.10	1	1.4876	2973.1	1.8528	7025.0
11	2-Amino-2-Nitropyridine (1)	4214-76-0	139.11	1	1.9094	8017.9	2.1962	15609.7
12	3-Amino-2-Chloropyridine (1)	6298-19-7	128.56	1	0.9518	795.0	1.0777	1096.0
13	2-Amino-5-Chloropyridine (1)	1072-98-6	128.56	1	0.8165	555.4	0.9394	769.7
15	2-Aminopyridine (1)	504-29-0	94.11	1	2.2439	17433.8	2.3042	20044.5
16	4-Aminopyridine (1)	504-24-5	94.11	1	1.0854	1117.3	1.3147	1964.0
17	2-Amino-3-Hydroxypyridine (1)	16867-03-1	110.11	1	1.5752	3660.0	1.6616	4487.4
18	DL-Glutamic Acid (1)	617-65-2	147.13	16	0.2945	95.4	1.0335	984.5
19	Picric Acid (7)	88-89-1	229.10	7	1.1975	2173.9	1.2075	2248.0
21	6-Chloronicotinic Acid (1)	5326-23-8	157.55	1	1.2147	1539.6	1.9365	8539.0
22	Isonicotinic Acid (1)	55-22-1	123.11	1	1.3606	2194.2	1.7078	5002.4
23	Picolinic Acid (1)	98-98-6	123.11	1	1.0410	999.0	1.2568	1706.4
24	Nicotinic Acid (1)	59-67-6	123.11	1	1.2395	1635.8	1.5410	3375.3
25	2-Hydroxynicotinic Acid (1)	609-71-2	139.11	1	2.1987	15701.8	2.8712	74229.9
26	4-Nitrophenol (2)	100-02-7	139.11	2	0.9303	1636.1	0.8887	1215.9
28	Isonicotinamide (1)	1453-82-3	122.12	1	1.3479	2127.7	1.3647	2215.7
29	Nicotinamide (1)	98-92-0	122.12	1	1.5924	3812.0	1.5903	3793.4
30	4-Nitroaniline (8)	100-01-6	138.12	8	0.9559	1887.7	0.9708	2700.8
31	2-Methoxy-5-Nitropyridine (1)	5446-92-4	154.12	1	0.1152	23.3	0.1302	34.9
32	2-Amino-4-Methyl-3-Nitropyridine (1)	6635-86-5	153.14	1	0.4721	196.5	0.5995	297.7
33	2-Amino-4-Methylpyridine (1)	695-34-1	108.14	1	0.7041	405.9	0.9393	769.7
34	Sulfanilamide (4)	63-74-1	172.20	4	1.1466	1698.2	1.1371	1810.2
35	3,5-Dinitrosalicylic Acid (1)	609-99-4	228.12	16	0.1277	30.6	2.2364	22200.7
37	M-Nitrobenzaldehyde (1)	99-61-6	151.12	1	0.6895	389.2	0.8733	646.9
50	Benzamide (1)	27208-38-4	121.14	1	1.5034	3087.2	1.5882	3774.1

51	Methylnicotinate (1)	93-60-7	137.14	1	0.4464	179.5	0.6994	400.5
52	P-Hydroxybenzamide (1)	619-57-8	137.14	1	2.7791	60034.2	3.0188	104319.9
53	4-Aminobenzoic Acid (18)	150-13-0	137.14	19	0.4022	285.1	0.4662	504.7
54	4-Aminosalicylic Acid (1)	65-49-6	153.14	9	1.8783	7633.0	2.6861	49883.0
55	2-Hydroxy-6-Methylnicotinic Acid (1)	38116-61-9	153.14	1	1.0090	921.0	1.4779	2905.3
60	Theophylline (23)	58-55-9	180.16	24	0.5417	269.9	0.7417	705.4
61	Theobromine (5)	83-67-0	180.16	5	0.6363	55.4	0.7676	130.0
62	2,6-Pyridinedimethanol (1)	1195-59-1	139.15	1	3.2640	183567.1	3.3461	221759.4
64	2-Amino-4,6-Dimethylpyridine (1)	5407-87-4	122.17	1	0.2802	90.6	0.4430	177.3
65	Simazine (1)	122-34-9	201.66	3	1.2885	1984.9	1.4652	3030.5
67	2-Methyl-3-Nitrobenzoic Acid (1)	1975-50-4	181.15	1	1.0090	921.0	1.9325	8461.4
80	Acetanilide (9)	103-84-4	135.16	9	0.8927	1182.5	1.0400	1871.0
81	4-Aminoacetophenone (1)	99-92-3	135.16	1	0.9381	767.1	0.9372	765.3
82C	Acetaminophen (25)	103-90-2	151.16	95	0.4835	325.1	0.5385	259.6
83	Methyl-P-Aminobenzoate (3)	619-45-4	151.16	3	0.8003	585.3	0.8351	650.4
84	Caffeine (6)	58-08-2	194.19	7	0.7589	3164.3	0.8305	4243.4
85	Acyclovir (2)	59277-89-3	225.20	4	1.6588	6144.1	1.9323	8808.7
86	Barbital (3)	57-44-3	184.19	3	0.8219	585.3	1.3963	3709.7
87	Atrazine (1)	1912-24-9	215.68	4	0.5501	257.8	0.7583	485.6
92	Diuron (49)	330-54-1	233.09	49	0.6767	44172.0	0.7401	49283.9
93	Ethyl-P-Aminobenzoate (3)	94-09-7	165.19	4	0.6381	374.3	0.6704	419.8
94	Monuron (25)	150-68-5	198.65	25	0.5915	452.9	0.6559	580.3
95	Chloropyrifos (2)	2921-88-2	350.59	6	0.8735	992.2	1.0414	1320.6
96	Ganciclovir (2)	82410-32-0	255.23	4	0.9110	787.1	1.2898	1896.2
97	Cyanazine (3)	21725-46-2	240.69	3	0.1606	37.7	0.4517	183.2
98	Metharbital (2)	50-11-3	198.22	2	0.3786	141.7	0.4839	397.5
100	Sulfadiazine (3)	68-35-9	250.28	3	0.8351	1678.0	0.8296	1010.6
101	Sulfamethoxazole (3)	723-46-6	253.28	3	0.7704	912.3	0.9198	2650.0
103	Phenacetin (3)	62-44-2	179.22	14	0.9955	970.9	1.0565	1091.0
104	Propyl-P-Aminobenzoate (2)	94-12-2	179.22	2	0.8595	725.9	0.8921	828.8
105	Ephedrine (13)	299-42-3	165.23	13	0.1487	20.6	0.1333	21.4
107	Butabarbital (2)	125-40-6	212.24	2	1.1538	1377.6	1.7887	13207.0
108	Sulfapyridine (2)	144-83-2	249.29	2	1.6406	22005.2	1.8027	17791.5
109	Antipyrine (4)	60-80-0	188.23	5	0.5218	105.4	0.4166	175.3
110	Sulfamerazine (1)	127-79-7	264.30	1	0.4228	164.7	0.2498	77.8
111	Sulfameter (1)	651-06-9	280.30	1	0.2519	78.6	0.0674	16.8

112	Sulfamethoxypyridazine (5)	80-35-3	280.30	17	1.5601	5449.2	1.6893	6566.7
113	Sulfisoxazole (1)	127-69-5	267.30	1	0.1557	43.1	0.0407	9.8
115	Butyl-P-Aminobenzoate (3)	94-25-7	193.24	4	0.8008	770.4	0.8266	897.5
116	Thiopental (2)	76-75-5	242.34	2	0.5614	373.2	1.1267	1247.8
117	Amobarbital (2)	57-43-2	226.27	2	0.7859	545.9	1.2619	2940.0
118	Pentobarbital (2)	76-74-4	226.27	2	1.1239	1789.3	2.3539	373571.2
121	Carbazole (29)	86-74-8	167.21	46	0.7203	737.1	0.7534	928.0
128	Phenobarbital (3)	50-06-6	232.23	4	1.0850	1200.3	1.8902	55358.6
129	Sulfisomidine (3)	515-64-0	278.33	3	1.0278	347.9	0.8678	170.9
130	Sulfamethazine (5)	57-68-1	278.33	5	1.3592	5303.3	1.3114	4422.3
132	Thiamylal (2)	77-27-0	254.35	2	0.4359	225.4	0.8381	597.0
133	Tolbutamide (2)	64-77-7	270.35	2	1.1843	840.4	1.3243	1066.6
135	Acridine (4)	260-94-6	179.22	24	0.2223	100.2	0.2213	100.0
136	Niflumic Acid (21)	4294-00-7	282.22	25	1.1974	19189.7	1.2632	40556.8
148B	Lidocaine (20)	137-58-6	234.34	20	0.1355	32.7	0.1214	33.3
149	Piroxicam (21)	36322-90-4	331.35	22	1.2854	7753.1	1.4334	13101.3
151	Mefenamic Acid (3)	61-68-7	241.28	3	1.0565	742.8	0.9593	1927.9
152	Meperidine (1)	57-42-1	247.33	1	3.2347	99.9	2.9802	99.9
155	Morphine (8)	57-27-2	285.34	9	1.5885	25827.0	1.6233	31815.4
156	Hydromorfinone (2)	466-99-9	285.34	2	1.6772	607.2	1.6345	641.1
157	Codeine (2)	76-57-3	299.36	2	1.4493	764.2	1.3337	772.6
162	Morniflumate (4)	65847-85-0	395.37	4	1.0209	975.1	1.0604	1074.6
173	Haloperidol (16)	52-86-8	375.86	16	1.1333	2067.5	1.1476	2279.3
184	Fentanyl (2)	437-38-7	336.47	2	0.7312	733.4	0.9557	912.2
187	Sufentanil (2)	56030-54-7	386.55	2	2.0299	997.8	1.8194	997.8

Appendix B: FORTRAN Code and Scripts

B1: COSMO-SAC FORTRAN .f90 Code for validating SLE equilibrium literature data.

Input: components, COSMO volumes, temperature, solute melting temperature(s), solute latent heat(s) of fusion, sigma profiles, and initial guess for mole fractions from input file

Output: solvent and solute mole fractions, activity coefficients

Notes:

- (1) Easily modified from pure-solvent to mixed (binary and higher) solvents and full array of mole fractions
- (2) Easily modifiable to calculate solubility at multiple temperatures
- (3) Iterative calculation scheme for segment activity coefficient and solute mole fraction

```
PROGRAM GAMMASOLUBILITY
!*****
!   This program, GAMMASOLUBILITY, calculates activity coefficients for a solute and
!   solvent(s) for pure solvent systems and binary or mixed solvent systems.  The
!   activity coefficients are then used to predict a solubility value for the solute
!   in that mixture.
!
!   This program uses the COSMO-SAC model as published (Lin, S.T.,
!   S.I. Sandler, Ind. Eng. Chem. Res. 41, (2002), 899-913).
!
!   PROGRAM WRITTEN BY:
!   ERIC MULLINS (PMULLINS@VT.EDU)
!   DEPARTMENT OF CHEMICAL ENGINEERING
!   VIRGINIA TECH
!   BLACKSBURG, VA 24061
!
!   ORIGINAL "COSMO-SAC-VT-2004" PROGRAM WRITTEN BY:
!   RICHARD OLDLAND (roldland@vt.edu)   MIKE ZWOLAK (zwolak@caltech.edu)
!   DEPARTMENT OF CHEMICAL ENGINEERING  PHYSICS DEPARTMENT
!   VIRGINIA TECH                       CALIFORNIA INSTITUTE OF TECHNOLOGY
!   BLACKSBURG, VA 24061                 PASADENA, CA 91125
!
!   PHYSICAL CONSTANTS AND PARAMETERS:
!   EO = PERMITTIVITY IN A VACUUM (e**2*mol/Kcal*Angstrom)
!   AEFFPRIME = EFFECTIVE SURFACE AREA (ANGSTROMS**2) --FROM LIN
```

```

!      RGAS = IDEAL GAS CONSTANT (Kcal/mol*K)
!      VNORM = VOLUME NORMALIZATION CONSTANT (A**3) --FROM LIN
!      ANORM = AREA NORMALIZATION CONSTANT (A**2) --FROM LIN
!      COORD = THE COORIDINATION NUMBER --FROM LIN
!      CHB = HYDROGEN BONDING COEFFICIENT (Kcal/mole*Angstroms**4/e**2)
!      SIGMAHB = CUTOFF VALUE FOR HYDROGEN BONDING (e/Angstrom**2)
!      EPS = RELATIVE PERMITTIVITY --FROM LIN
!      ALPHAPRIME = A CONSTANT USED IN THE MISFIT ENERGY CALCULATION
!
! INPUT PARAMETERS:
!      SYSTEMP = THE SYSTEM TEMPERATURE (K)
!      COMP = NUMBER OF COMPONENTS IN THE SYSTEM --SET TO 2 FOR BINARY
!      SYSCOMP = NAMES OF COMPONENTS IN THE SYSTEM
!      VCSMO = CAVITY VOLUME FROM COSMO OUTPUT (A**3)
!      ACOSMO = MOLECULAR SURFACE AREA FROM COSMO OUTPUT (A**2) --THE SUM
!              OF THE INDIVIDUAL PROFILE.
!
! LITERATURE CITED:
! 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.
! Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. Refinement and Parameterization of
!   COSMO-RS. J. Phys. Chem A 1998, 102, 5074.
! Klamt, A.; Eckert, F.; COSMO-RS: A Novel and Efficient Method for the a Priori
!   Prediction of Thermophysical Data of Liquids. Fluid Phase Equilibria 2000,
!   172, 43.
! Lin, S.T.; Sandler, S. A Priori Phase Equilibrium Prediction from a Segment
!   Contribution Solvation Model. Ind. Eng. Chem. Res, 2002, 41, 899
! Lin, S.T.; Quantum Mechanical Approaches to the Prediction of Phase Equilibria:
!   Solvation Thermodynamics and Group Contribution Methods, PhD. Dissertation,
!   University of Delaware, Newark, DE, 2000
!
! PROGRAM CURRENTLY SETUP FOR PURE SOLVENT SOLUBILITY PREDICTIONS ONLY
! TO PREDICT BINARY SOLVENT SOLUBILITY, MODIFY THE MIXTURE SIGMA PROFILE (LINE 245)
! MODIFY READ INPUT FILE (LINE 130)
! MODIFY NUMBER OF COMPONENTS, VARIABLE: [COMP], CURRENTLY SET TO 2
!*****
IMPLICIT NONE
REAL*8, PARAMETER :: EO = 0.0002395D0 ! PERMITTIVITY OF FREE SPACE (e^2*MOL/KCAL/ANG^2)

```



```

REAL*8, PARAMETER :: AEFFPRIME = 7.5D0 ! ANGSTROMS^2
REAL*8, PARAMETER :: RGAS = 0.001987D0 ! KCAL/(MOL K)
REAL*8, PARAMETER :: VNORM = 66.69D0 ! NORMALIZED CAVITY VOLUME, ANGSTROMS^3
REAL*8, PARAMETER :: ANORM = 79.53D0 ! NORMALIZED SURFACE AREA, ANGSTROMS^2
REAL*8, PARAMETER :: COORD = 10.0D0 ! COORDINATE NUMBER, Z, KLAMT SET TO 7.2
INTEGER, PARAMETER :: COMPSEG = 51 ! NUMBER OF INTERVALS IN SIGMA PROFILE (-0.025 TO 0.025)

REAL*8, DIMENSION (:,:), ALLOCATABLE :: SIGMA, DELTAW, WHB, SEGGAMMAPURE, SEGGAMMAPUREOLD, CONVPURE
REAL*8, DIMENSION (:,:), ALLOCATABLE :: SIGMASOLUTE, SIGMA1SOLVENT, COUNTERSOLUTE !SIGMA2SOLVENT,
REAL*8, DIMENSION (:,:), ALLOCATABLE :: COUNTER1SOLVENT, COUNTER2SOLVENT, CONPR
REAL*8, DIMENSION (:), ALLOCATABLE :: SEGGAMMA, SEGGAMMAOLD, CONVERG, X1DATA, X2DATA, X !, X3DATA,
REAL*8, DIMENSION (:), ALLOCATABLE :: VCOSMO, ACOSMO, RNORM, QNORM, COUNTER, MIXPROFILE, DELTAHF
REAL*8, DIMENSION (:), ALLOCATABLE :: THETA, PHI, LSG, LNGAMMASG, SUMGAMMA, GAMMA, LNGAMMA, TM, SYSTEMP
REAL*8, DIMENSION (:), ALLOCATABLE :: VCOSMOSOLUTE, VCOSMO1SOLVENT !, VCOSMO2SOLVENT
REAL*8, DIMENSION (:), ALLOCATABLE :: ACOSMOSOLUTE, ACOSMO1SOLVENT !, ACOSMO2SOLVENT
REAL*8 :: FPOL, ALPHA, ALPHAPRIME, EPS, SIGMAHB, CHB, SIGMAACC, SIGMADON, SIGMAHBNEW, SUMMATION
REAL*8 :: RELTOL, CHANGE, X2OLD, X1NEW, X2NEW, MFSUM ! X3NEW

INTEGER :: I, J, K, L, P, T, ITER, NUMPOINTS, ISTAT , MAXITS, COMP
INTEGER, DIMENSION (:), ALLOCATABLE :: UNIT1SOLV, UNITSOLU ! UNIT2SOLV,

CHARACTER (128) :: INPUTFILE, OUTPUTFILE
CHARACTER (16), DIMENSION (:), ALLOCATABLE :: SYSCOMP
CHARACTER (50), DIMENSION (:), ALLOCATABLE :: SOLVENT1FILE, SOLUTEFILE !, SOLVENT2FILE,
CHARACTER (16), DIMENSION (:), ALLOCATABLE :: SOLV1 !, SOLV2
CHARACTER (15), DIMENSION (:), ALLOCATABLE :: SOLU

EPS = 3.667D0 ! DIELECTRIC CONSTANT, LIN AND SANDLER USE A CONSTANT FPOL WHICH YIELDS EPS=3.68
SIGMAHB = 0.0084D0 ! HYDROGEN-BONDING INTERACTION CUT-OFF, e/ANG^2
SIGMAHBNEW = 0.022D0 ! FOR USE WITH NEW EXPRESSION FOR EXCHANGE ENERGY, e/ANG^2
CHB = 85580.0D0 ! HYDROGEN-BONDING INTERACTION CONSTANT, KCAL*ANG^4/MOL/e^2
FPOL = (EPS-1.0)/(EPS+0.5) !UNITLESS
ALPHA = (0.3*AEFFPRIME**(1.5))/(EO) ! KCAL*ANG^4/MOL/e^2
ALPHAPRIME = FPOL*ALPHA ! MISFIT ENERGY CONSTANT, KCAL*ANG^4/MOL/e^2
COMP = 2 !PURE SOLVENT [USER SPECIFIED]

! ALLOCATE INPUT ARRAYS
INPUTFILE = "C:\SOL\PURE_SOLVENT\DATA\SOLUTE_DATA\VTSOL-001DATA.TXT" ![USER SPECIFIED FILE LOCATION]
NUMPOINTS = 10 ![USER SPECIFIED, MAXIMUM VALUE: 10,000]

```

```

        ALLOCATE (SOLV1(NUMPOINTS), VCOSMO1SOLVENT(NUMPOINTS), & !SOLV2(NUMPOINTS),
VCOSMO2SOLVENT(NUMPOINTS), &
        SOLU(NUMPOINTS), VCOSMOSOLUTE(NUMPOINTS), TM(NUMPOINTS), DELTAHF(NUMPOINTS),
SYSTEMP(NUMPOINTS), &
        X2DATA(NUMPOINTS), UNIT1SOLV(NUMPOINTS), SOLUTEFILE(NUMPOINTS), SOLVENT1FILE(NUMPOINTS), &
        UNITSOLU(NUMPOINTS), ACOSMO1SOLVENT(NUMPOINTS), SIGMASOLUTE(COMPSEG,NUMPOINTS), &
        COUNTERSOLUTE(COMPSEG,NUMPOINTS), ACOSMOSOLUTE(NUMPOINTS), &
        SIGMA1SOLVENT(COMPSEG,NUMPOINTS), COUNTER1SOLVENT(COMPSEG,NUMPOINTS), STAT = ISTAT) !, &
        !SIGMA2SOLVENT(COMPSEG,NUMPOINTS), COUNTER2SOLVENT(COMPSEG,NUMPOINTS), &
        ACOSMO2SOLVENT(NUMPOINTS)),
        !UNIT2SOLV(NUMPOINTS),X1DATA(NUMPOINTS), X3DATA(NUMPOINTS), SOLVENT2FILE(NUMPOINTS))
WRITE (*,*) ISTAT, " ALLOCATION SUCCESSFUL - INPUT ARRAYS"

![USER SPECIFIED INPUT FILE] USE TEMPLATE, L IS COUNTER FOR NUMBER OF POINTS
!COLUMN 1: SOLUTE FILE NAME , SOLU(L)
!COLUMN 2: SOLUTE COSMO CALCULATION VOLUME, VCOSMOSOLUTE(L)
!COLUMN 3: SOLVENT FILE NAME, SOLV1(L)
!COLUMN 4: SOLVENT COSMO CALCULATION VOLUME, VCOSMO1SOLVENT(L)
!COLUMN 5: MELTING TEMPERATURE, TM(L)
!COLUMN 6: LATENT HEAT OF FUSION, DELTAHF(L)
!COLUMN 7: SYSTEM TEMPERATURE, SYSTEM(L)
!COLUMN 8: SOLUTE MOLE FRACTION, X2DATA(L)

OPEN (UNIT=12, FILE=INPUTFILE, STATUS="OLD", ACTION="READ", POSITION="REWIND")
DO L=1, NUMPOINTS
        READ (12,*) SOLU(L), VCOSMOSOLUTE(L), SOLV1(L), VCOSMO1SOLVENT(L), TM(L), DELTAHF(L), &
        SYSTEMP(L), X2DATA(L) ! , X1DATA(L), X3DATA(L), SOLV2(L), VCOSMO2SOLVENT(L)
END DO
CLOSE (12)

DO I = 1, NUMPOINTS
        UNIT1SOLV (I) = I + 20
        UNITSOLU (I) = I + 10020
        SOLUTEFILE (I) = "C:\SOL\SOLUTE\\"//SOLU(I) ![USER SPECIFIED FILE LOCATION]
        SOLVENT1FILE (I) = "C:\VT-2006\\"//SOLV1(I) ![USER SPECIFIED FILE LOCATION]
        WRITE (*,*) SOLVENT1FILE(I), SOLUTEFILE(I)
END DO

DO K = 1, NUMPOINTS

```

```

OPEN (UNIT=UNIT1SOLV(K), FILE = SOLVENT1FILE(K), STATUS="OLD", ACTION="READ", &
POSITION="REWIND")
DO J=1, COMPSEG
    READ(UNIT1SOLV(K),*) COUNTER1SOLVENT(J,K), SIGMA1SOLVENT(J,K)
    ACOSMO1SOLVENT(K) = ACOSMO1SOLVENT(K) + SIGMA1SOLVENT(J,K)
END DO
CLOSE (UNIT1SOLV(K))
END DO

DO K = 1, NUMPOINTS
    OPEN (UNIT=UNITSOLU(K), FILE = SOLUTEFILE(K), STATUS="OLD", ACTION="READ", POSITION = "REWIND")
    DO J = 1, COMPSEG
        READ(UNITSOLU(K),*) COUNTERSOLUTE(J,K), SIGMASOLUTE(J,K)
        ACOSMOSOLUTE(K) = ACOSMOSOLUTE(K) + SIGMASOLUTE(J,K)
    END DO
    CLOSE (UNITSOLU(K))
END DO

![USER SPECIFIED FILE LOCATION]
OUTPUTFILE = "C:\SOL\PURE_SOLVENT\PREDICTIONS\VTSOL_001_PREDICTION.TXT"
OPEN (UNIT = 13, FILE = OUTPUTFILE, STATUS = "NEW")

2 FORMAT (1X,A1,1X,A1, 1X, A16, 1X, A16)
3 FORMAT (1X,A5,1X,A4,1X,A16,1X,A16,1X,A4,1X,A2,1X,A2,1X,A7,1X,A7)

WRITE (13,2) "-", "-", "COMP(1)", "COMP(2)"
WRITE (13,3) "POINT", "ITER", "SOLU", "SOLV1", "TEMP", "X1", "X2", "GAMMA1", "GAMMA2"

!ALLOCATE ARRAYS USED IN MAIN ITERATIVE LOOP
ALLOCATE (SIGMA(COMPSEG,COMP), COUNTER(COMPSEG), MIXPROFILE(COMPSEG), DELTAW(COMPSEG,COMPSEG), &
SEGGAMMA(COMPSEG), SEGGAMMAOLD(COMPSEG), CONVERG(COMPSEG), SEGGAMMAPURE(COMPSEG,COMP), &
SEGGAMMAPUREOLD(COMPSEG,COMP), CONVPURE(COMPSEG,COMP), WHB(COMPSEG, COMPSEG), &
VCOSMO(COMP), ACOSMO(COMP), X(COMP), THETA(COMP), PHI(COMP), LNGAMMASG(COMP), GAMMA(COMP), &
LSG(COMP), SUMGAMMA(COMP), LNGAMMA(COMP), RNORM(COMP), QNORM(COMP), &
CONPR(COMPSEG, COMP), STAT = ISTAT)

WRITE(*,*) ISTAT, " ALLOCATION SUCCESSFUL - CALCULATION ARRAYS"

! BEGIN ITERATION LOOP FOR ALL POINTS, REASSIGNING NEW SIGMA PROFILES FOR EACH ITERATIONS

```

```
DO P = 1, NUMPOINTS
```

```
! RE-ZERO CALCULATION LOOP ARRAYS
```

```
SIGMA = 0.0D0
```

```
COUNTER = 0.0D0
```

```
MIXPROFILE = 0.0D0
```

```
DELTAW = 0.0D0
```

```
SEGGAMMA = 1.0D0
```

```
SEGGAMMAOLD = 1.0D0
```

```
CONVERG = 1.0D0
```

```
SEGGAMMAPURE = 1.0D0
```

```
SEGGAMMAPUREOLD = 1.0D0
```

```
WHB = 0.0D0
```

```
SIGMAACC = 0.0D0
```

```
SIGMADON = 0.0D0
```

```
THETA = 0.0D0
```

```
PHI = 0.0D0
```

```
LSG = 0.0D0
```

```
RNORM = 0.0D0
```

```
QNORM = 0.0D0
```

```
LNGAMMASG = 0.0D0
```

```
SUMGAMMA = 0.0D0
```

```
LNGAMMA = 0.0D0
```

```
GAMMA = 0.0D0
```

```
ACOSMO = 0.0D0
```

```
VCOSMO = 0.0D0
```

```
DO I = 1, COMPSEG
```

```
    SIGMA(I,1) = SIGMA1SOLVENT(I,P)
```

```
    SIGMA(I,2) = SIGMASOLUTE(I,P)
```

```
    COUNTER(I) = COUNTER1SOLVENT(I,P)
```

```
END DO
```

```
ACOSMO(1) = ACOSMO1SOLVENT(P)
```

```
ACOSMO(2) = ACOSMOSOLUTE(P)
```

```
VCOSMO(1) = VCOSMO1SOLVENT(P)
```

```
VCOSMO(2) = VCOSMOSOLUTE(P)
```

```
X(1) = 1.0D0 - X2DATA(P)
```

```
X(2) = X2DATA(P)
```

```

! CONVERGENCE LOOP, MAX ITERATIONS = 1000
DO ITER = 1, 1000

    X(1) = 1.0D0 - X(2)
    X2OLD = X(2)

    ! CALCULATE SIGMA PROFILE FOR MIXTURE FOR EACH DATA POINT
    ! Ps(SIGMA) = SUM(Xi*Ai*Pi(SIGMA))/SUM(Xi*Ai)
    ! SIGMA(J,K) = P'(SIGMA) = Pi(SIGMA)*Ai

    !WRITE (13,*) " MIXPROFILE"
    DO J = 1,COMPSEG ! 51 SEGMENTS IN MIXTURE SIGMA PROFILE

        ![MIXTURE PROFILE FOR PURE SOLVENT SYSTEM]
        MIXPROFILE(J) = ((X(1)*SIGMA(J,1))+(X(2)*SIGMA(J,2))) / &
            ((X(1)*ACOSMO(1)) + (X(2)*ACOSMO(2)))

        ![MIXTURE PROFILE FOR BINARY SOLVENT SYSTEM]
        !MIXPROFILE(J)=((X(1)*SIGMA(J,1))+(X(2)*SIGMA(J,2))+(X(3)*SIGMA(J,3))) / &
            !((X(1)*ACOSMO(1)) + (X(2)*ACOSMO(2))+(X(3)*ACOSMO(3)))

    END DO

    ! DETERMINE THE LIN AND SANDLER EXCHANGE ENERGY, DELTAW, KCAL/MOL (ORIGINAL MODEL)
    DO I = 1, COMPSEG
        DO K = 1, COMPSEG
            ! ASSIGN HYDROGEN BONDING DONOR AND ACCEPTOR
            IF (COUNTER(I)>=COUNTER(K)) THEN
                SIGMAACC = COUNTER(I)
                SIGMADON = COUNTER(K)
            END IF
            IF (COUNTER(I)<COUNTER(K)) THEN
                SIGMADON = COUNTER(I)
                SIGMAACC = COUNTER(K)
            END IF
            DELTAW(I,K) = (ALPHAPRIME/2.0D0)*(COUNTER(I)+COUNTER(K))**2.0D0 + CHB * &
                MAX(0.0D0,(SIGMAACC - SIGMAHB))*MIN(0.0D0,(SIGMADON + SIGMAHB))
        END DO
    END DO

```

```

! DETERMINE THE MATHIAS ET AL EXCHANGE ENERGY, DELTAW, KCAL/MOL (NEW MODEL)
! DELTAW(SIGMAm,SIGMAn) = -CHB*max[0, abs(SIGMAm - SIGMAn) - SIGMAHBNEW]^2
! SIGMAHBNEW = 0.022
!DO I = 1, COMPSEG
  !DO K= 1, COMPSEG
    ! WHB (I,K) = HYDROGEN BONDING CONTRIBUTION TO EXCHANGE ENERGY
    !WHB (I,K) = -CHB*(MAX(0.0D0,(DABS(COUNTER(I)-COUNTER(K)))-&
    SIGMAHBNEW)**2.0D0
    !DELTAW(I,K) = (ALPHAPRIME/2.0D0)*(COUNTER(I)+COUNTER(K))**2.0D0+
    WHB(I,K)
  !END DO
!END DO

! ITERATION FOR MIXTURE SEGMENT ACTIVITY COEF.
! CONVERGENCE BY DAMPING METHOD, DAMPING FACTOR = 1/2
SEGGAMMA = 1.0D0 !INTIAL VALUE FOR MIXTURE SEGMENT ACTIVITY COEF.
DO
  SEGGAMMAOLD = SEGGAMMA
  DO I = 1, COMPSEG
    SUMMATION = 0.0D0
    DO K = 1, COMPSEG
      SUMMATION = SUMMATION + MIXPROFILE(K)*SEGGAMMAOLD(K) * &
      DEXP(-DELTAW(I,K)/(RGAS * SYSTEMP(P)))
    END DO
    SEGGAMMA(I) = DEXP(-DLOG(SUMMATION))
    SEGGAMMA(I) = (SEGGAMMA(I) + SEGGAMMAOLD(I))/2.0D0
  END DO
  DO I=1, COMPSEG
    CONVERG(I) = DABS((SEGGAMMA(I) - SEGGAMMAOLD(I))/SEGGAMMAOLD(I))
  END DO
  IF (MAXVAL(CONVERG) <=0.000001) EXIT
END DO

! ITERATION FOR PURE SPECIES, L, SEGMENT ACTIVITY COEF.
! CONVERGENCE BY DAMPING METHOD, DAMPING FACTOR = 1/2
!ITERATION FOR SEGMENT ACITIVITY COEF (PURE SPECIES)
DO L = 1, COMP
  SEGGAMMAPURE (:,L) = 1.0D0
DO

```

```

      SEGGAMMAPUREOLD (:,L) = SEGGAMMAPURE (:,L)
      DO I = 1, COMPSEG
        SUMMATION = 0.0D0
        DO K = 1, COMPSEG
          SUMMATION = SUMMATION + (SIGMA(K,L)/ACOSMO(L)) * &
            SEGGAMMAPUREOLD(K,L)*DEXP(-DELTAW(I,K)/(RGAS*SYSTEMP(P)))
        END DO
        SEGGAMMAPURE(I,L)=DEXP(-DLOG(SUMMATION))
        SEGGAMMAPURE(I,L)=(SEGGAMMAPURE(I,L)+SEGGAMMAPUREOLD(I,L))/2.0D0
      END DO
      DO I=1, COMPSEG
        CONPR(I,L)=DABS((SEGGAMMAPURE(I,L)-SEGGAMMAPUREOLD(I,L))/&
          SEGGAMMAPUREOLD(I,L))
      END DO
      IF (MAXVAL(CONPR) <=0.000001) EXIT
    END DO
  END DO

! THE STAVERMAN-GUGGENHEIM EQUATION
  DO I = 1,COMP
    RNORM(I) = VCOSMO(I)/VNORM
    QNORM(I) = ACOSMO(I)/ANORM
  END DO

  DO I = 1, COMP
    THETA(I) = (X(I)*QNORM(I))/(X(1)*QNORM(1) + X(2)*QNORM(2))
    PHI(I) = (X(I)*RNORM(I))/(X(1)*RNORM(1) + X(2)*RNORM(2))
    LSG(I) = (COORD/2.0D0)*(RNORM(I)-QNORM(I))-(RNORM(I)-1.0D0)
  END DO

  ! GAMMASG1 AND GAMMASG2 ARE ACTUALLY LNGAMMASG
  DO I = 1, COMP
    LNGAMMASG(I) = DLOG(PHI(I)/X(I)) + (COORD/2.0D0)*QNORM(I)*DLOG
      (THETA(I)/PHI(I)) + LSG(I) - (PHI(I)/X(I))*(X(1)*LSG(1) + X(2)*LSG(2))
  END DO

  !CALCULATION OF GAMMAS
  SUMGAMMA = 0.0D0
  DO K = 1, COMP
    DO I = 1, COMPSEG
      SUMGAMMA(K) = SUMGAMMA(K)+((SIGMA(I,K)/AEFFPRIME)*(DLOG(SEGGAMMA(I)/ &

```

```

                                SEGGAMMAPURE(I,K)))
    END DO
    GAMMA(K) = DEXP(SUMGAMMA(K) + (LNGAMMASG(K)))
    LNGAMMA(K) = DLOG(GAMMA(K))
END DO

! RECALCULATE SOLUTE MOLE FRACTION FROM DERIVED EQUATION
X(2) = DEXP((DELTAHF(P)/8.3145D0/TM(P))*(1.0D0-TM(P)/SYSTEMP(P)) - LNGAMMA(2))

!CHECK FOR NEGATIVE MOLE FRACTIONS
IF (X(2) >= 1.0) THEN
    WRITE (*,*) "SOLUTE MOLE FRACTION CALCULATED AS GREATER THAN 1.0"
    WRITE (*,*) X(2)
END IF

!NORMALIZE MOLE FRACTIONS KEEPING SOLVENT RATIO (X1:X2) CONSTANT
! DAMPING FACTOR (1/3) USED IN CONVERGING X(2)
X(2) = (X(2) + 2.0D0*X2OLD)/3.0D0
MFSUM = X(1)+X(2) ! +X(3)
X(1) = X(1)/MFSUM
X(2) = X(2)/MFSUM

! CALCULATE THE RELATIVE CHANGE OF THE SOLUTE MOLE FRACTION
CHANGE = DABS((X(2) - X2OLD)/(X2OLD + 1.0D-16))
WRITE (*,*) "    RELATIVE CHANGE:", P, CHANGE

! COMPARE RELATIVE TOLERANCE
RELTOL = 1.E-8
IF (CHANGE < RELTOL) GOTO 10

!END ITERATIVE LOOP
END DO

! POINT, ITER, SOLU, SOLV1,SYSTEMP, X1, X2, GAMMA1, GAMMA2 (INPUT FORMAT TXT FILE)
4   FORMAT (1X,I4,2X,I4,2X,A16,2X,A16,2X,E,1X,E,1X,E,1X,E,1X,E)

10  WRITE (13,4) P,ITER,SOLU(P),SOLV1(P),SYSTEMP(P),X(1),X(2),GAMMA(1),GAMMA(2)

```



```

! END OF CALCULATIONS FOR ALL POINTS
END DO

CLOSE (13) ! CLOSE OUTPUT FILE

! DEALLOCATE ARRAYS FOR NEXT POINT
DEALLOCATE (SIGMA, COUNTER, MIXPROFILE, DELTAW, &
            SEGGAMMA, SEGGAMMAOLD, CONVERG, SEGGAMMAPURE, &
            SEGGAMMAPUREOLD, WHB, CONVPURE, &
            VCOSMO, ACOSMO, X, THETA, PHI, LNGAMMASG, GAMMA, &
            LSG, SUMGAMMA, LNGAMMA, RNORM, QNORM, CONPR, STAT = ISTAT)

WRITE (*,*) ISTAT, " DEALLOCATION SUCCESSFUL - CALCULATION ARRAYS"

DEALLOCATE (SOLV1, VCOSMO1SOLVENT, SOLU, VCOSMOSOLUTE, TM, DELTAHF, SYSTEMP, &
            X2DATA, UNIT1SOLV, SOLUTEFILE, SOLVENT1FILE, UNITSOLU, &
            SIGMASOLUTE, COUNTERSOLUTE, ACOSMOSOLUTE, SIGMA1SOLVENT, COUNTER1SOLVENT, &
            ACOSMO1SOLVENT, STAT = ISTAT)

WRITE (*,*) ISTAT, " DEALLOCATION SUCCESSFUL - INPUT ARRAYS"
!SOLV2, VCOSMO2SOLVENT, SOLVENT2FILE, SIGMA2SOLVENT, COUNTER2SOLVENT, ACOSMO2SOLVENT
!X1DATA, X3DATA, UNIT2SOLV,

END PROGRAM GAMMASOLUBILITY

```

B2: Amber Command Scripts

Amber is a Linux-based application and can be run via a graphical interface or command line prompt. We use two command scripts to run an energy minimization task.

(1) Script1: preps input .PDB file for minimization

```
***** SET ENVIRONMENT VARIABLES *****
#!/bin/csh
cd /home/Eric/vt-0045
setenv AMBERHOME /usr/local/amber8

***** RUN ANTECHAMBER AND PARMCHK COMMANDS *****
antechamber -i vt-0045.pdb -o aaa.mol2 -fi pdb -fo mol2 -c bcc
parmchk -i aaa.mol2 -f mol2 -o frcmod

cat <<EOF > leaprc
source leaprc.gaff
mods=loadamberparams frcmod
aaa=loadmol2 aaa.mol2
saveamberparm aaa aaa.top aaa.crd
quit
EOF

tleap
```

(2) Script2: creates output .PDB file

```
***** SET ENVIRONMENT VARIABLES *****
#!/bin/csh
cd ~Eric/vt-0045
setenv AMBERHOME /usr/local/amber8

if ( ! -d /scratch/$USER/min ) then
mkdir -p /scratch/$USER/min
endif
```

```
cat <<EOF> min.in

***** minimization: the whole system *****
&cntrl
  imin   = 1,
  maxcyc = 50000,
  ntpc   = 50,
  ntb    = 0,
  cut    = 10,
  ntr    = 0,
&end

EOF

***** SANDER COMMAND CREATES NEW FILE *****
sander -O -i min.in -p aaa.top -c aaa.crd \
        -o min.out \
        -r min.rst \
        -inf min.tmp

ls -l /scratch/$USER/min
mv    /scratch/$USER/min/min.* ./
rmdir /scratch/$USER/min

ambpdb -p aaa.top < min.rst > min.pdb
```

B3: NRTL-SAC FORTRAN Code

We write the NRTL-SAC model into this FORTRAN code with the specific purpose of predicting solid solubility in pure and mixed solvents. Currently, this code is set up for predicted solubility in pure solvents but is easily edited for solubility in mixed solvent.

Input: components, NRTL-SAC segment parameters, temperature, solute melting temperature(s), solute latent heat(s) of fusion, and initial guess for mole fractions from input file

Output: Initial and converged solvent and solute mole fractions, activity coefficients, and relative percent error

```
PROGRAM NRTL_SAC
!*****
! GENERAL COMMENTS:
!   Author:  Adel Ghaderi
!   Supervisor:  Eric Mullins
!   Advisor:  Y. A. Liu
!   Institution:  Virginia Polytechnic Institute and State University
!   Date:  October 27,2006
!   Code is setup to handle pure solvent systems for solubility of single solute
!   for subscripts referring to the segments
!   1 refers to X
!   2 refers to Y-
!   3 refers to Y+
!   4 refers to Z
!   Program is coded for 4 segments; simply changing value of NUMSEG is not
!   sufficient to vary number of segments.
!*****
! REFERENCES:
!   Chen, C.C.; Song, Y.; "Solubility modeling with a nonrandom two-liquid segment activity coefficient
!   model," IECR (2004), v.43, no.26, 8354-8362.
!*****
IMPLICIT NONE
INTEGER, PARAMETER :: SINGLE = 4
INTEGER, PARAMETER :: DOUBLE = 8
INTEGER :: i, j, k, m, n, y, z, ITER, ITERcount, NUMSPEC, NUMPOINTS, NUMSYS, istat
INTEGER, PARAMETER :: NUMSEG = 4          !number of segments; 4 (X, Y-, Y+, Z) as DEFINED by C. C. Chen
INTEGER, PARAMETER :: CHARLEN = 100      !maximum allowable length name of the system
REAL(KIND=DOUBLE) :: LNSACTERM1TOP, LNSACTERM1BOT, LNSACTERM2
```

```

REAL(KIND=DOUBLE) :: LNSACITERM1TOP, LNSACITERM1BOT, LNSACITERM2
REAL(KIND=DOUBLE) :: LNSACTERM2BOT, LNSACTERM4TOP, LNSACTERM4BOT
REAL(KIND=DOUBLE) :: LNSACITERM2BOT, LNSACITERM4TOP, LNSACITERM4BOT
REAL(KIND=DOUBLE) :: LNSAC, LNSACI, dummy, XOLD, MFSUM, RELTOL, CHANGE, LNRE, Rtot
REAL(KIND=DOUBLE), ALLOCATABLE, DIMENSION(:) :: T, XSEG, LNGAMMAR, LNGAMMAC, PHI, dHfus, TM
REAL(KIND=DOUBLE), ALLOCATABLE, DIMENSION(:, :) :: TAU, ALPHA, G, R, XORIG, X, XSPEC, GAMMA
CHARACTER(CHARLEN), ALLOCATABLE, DIMENSION(:) :: TITLE
CHARACTER(CHARLEN), ALLOCATABLE, DIMENSION(:, :) :: NAME
!-----
! Initialize Tau and Alpha values, from C.C. Chen, et al.
!-----
ALLOCATE( TAU(NUMSEG,NUMSEG), ALPHA(NUMSEG,NUMSEG), G(NUMSEG,NUMSEG) )
DO i = 1, 4
    ! 4 segment species (X, Y-, Y+, Z)
    DO j = 1, 4
        TAU(i,j) = 0
        ALPHA(i,j) = 0
        ! note ALPHA(i,j) = ALPHA(j,i)
    END DO
END DO
TAU(1,2) = 1.643
! 1 denotes X (hydrophobic segment)
TAU(2,1) = 1.834
! 2 denotes Y- (polar attractive segment)
ALPHA(1,2) = 0.200; ALPHA(2,1) = .200
TAU(1,4) = 6.547
! 4 denotes Z (hydrophilic segment)
TAU(4,1) = 10.949
ALPHA(1,4) = 0.200; ALPHA(4,1) = .200
TAU(2,4) = -2.000
TAU(4,2) = 1.787
ALPHA(2,4) = 0.300; ALPHA(4,2) = .300
TAU(3,4) = 2.000
! 3 denotes Y+ (polar repulsive segment)
TAU(4,3) = 1.787
ALPHA(3,4) = 0.300; ALPHA(3,2) = .300
TAU(1,3) = 1.643
TAU(3,1) = 1.834
ALPHA(1,3) = 0.200; ALPHA(1,3) = .200
!-----
! Calculate values for G
DO i = 1, 4
    DO j = 1, 4
        G(i,j) = exp( - ALPHA(i,j) * TAU(i,j) )
    END DO
END DO
! Generate input file

```

```

OPEN(UNIT = 13, FILE = "C:\NRTL-SAC\Input_Data\NRTL-SAC_all_points_01-15-07.txt", &
     STATUS="OLD", ACTION="READ", POSITION="REWIND")
! File will be closed at end of section

OPEN(UNIT=14,FILE="C:\NRTL-SAC\Input_Data\Calc_files\NRTL_pure_solvent_all_data_01-15-07.txt",&
     STATUS="NEW", ACTION="WRITE")
! File will be closed at end of section
! The input data of interest is for pure component solubility
NUMSPEC = 2
! For simplicity, every point is treated as a single system (this is usually the case anyway)
NUMPOINTS = 1

119 FORMAT( I4 )
120 FORMAT( I1, 5X, I1 )
121 FORMAT( A15 )
122 FORMAT( I1,2X,F10.2,2X,F7.2,2X,F12.10,2X,F12.10,2X,F12.10,2X,F12.10 )
123 FORMAT( F8.2, 2X, F17.15, 2X, F17.15 )

READ(13,*) NUMSYS
WRITE(14,119) NUMSYS

ALLOCATE( R(8,NUMSYS), dhfus(NUMSYS), TM(NUMSYS), T(NUMSYS), X(NUMSYS,2), NAME(NUMSYS,2), STAT = istat )
WRITE(*,*) istat , " ALLOCATION SUCCESSFUL IF 0"

DO i = 1, NUMSYS
! INPUT FILE FORMAT: SOLUTE NAME, HEAT OF FUSION, MELTING TEMPERATURE, SEGMENT PARAMETERS (X,Y-,
! Y+,Z), SOLVENT NAME, SEGMENT PARAMETERS (X,Y-,Y+,Z), SYSTEM TEMPERATURE, EXPERIMENTAL SOLUBILITY

READ(13,*) NAME(i,2), dhfus(i), TM(i), R(1,i), R(2,i), R(3,i), R(4,i), NAME(i,1), R(5,i), R(6,i), &
      R(7,i), R(8,i), T(i), X(i,2)
X(i,1) = 1.0D0 - X(i,2)

WRITE(14,120) NUMSPEC, NUMPOINTS
WRITE(14,121) NAME(i,1)
WRITE(14,121) NAME(i,2)
WRITE(14,122) 1, 0, 0, R(5,i), R(6,i), R(7,i), R(8,i)
WRITE(14,122) 2, dhfus(i), TM(i), R(1,i), R(2,i), R(3,i), R(4,i)
WRITE(14,123) T(i), X(i,1), X(i,2)
END DO
DEALLOCATE( R, dhfus, TM, T, X, NAME, STAT = istat )
WRITE(*,*) istat, " DEALLOCATION SUCCESSFUL IF 0"

```

```

CLOSE(13)
CLOSE(14)
!-----

! Open input and output files
OPEN(UNIT = 11, FILE = "C:\NRTL-SAC\Input_Data\Calc_files\&
  NRTL_pure_solvent_all_data_01-15-07.txt",STATUS="OLD", ACTION="READ", POSITION="REWIND")
! File will be closed at end of program
READ(11,*) NUMSYS

OPEN(UNIT=12, FILE = "C:\NRTL-SAC\Output\NRTL-SAC_all_points_PREDICTIONS_01-15-07.txt",&
  STATUS="NEW", ACTION="WRITE")
! File will be closed at end of program

WRITE(12,115) "ITER", "SOLUTE", "SOLVENT", "TEMP(K)", "GAMMA-1", "GAMMA-2", "X1-EXP", "X2-exp", "X1-pred",&
  "X2-pred", "Rel.-%E-(X2)", "%LN(E)"
WRITE(12,115) "====", "=====", "=====", "=====", "=====", "=====", "=====", "=====", "=====",&
  "=====", "====="
115 FORMAT(2X, A4, 3X, A6, 11X, A7, 13X, A7, 20X, A7, 20X, A7, 20X, A6, 21X, A6, 21X, A7, 20X, A7, 20X,&
  A12, 15X, A6)

! Iterate over all different multicomponent systems
!-----
DO z = 1, NUMSYS

! Some file manipulation and memory allocation
READ(11,*) NUMSPEC, NUMPOINTS
!number of system components and number of data points; used in memory allocation

WRITE(*,*) NUMSPEC, NUMPOINTS

ALLOCATE( NAME(NUMSPEC,1), TITLE(NUMSYS), dhfus(NUMSPEC), TM(NUMSPEC), &
  T(NUMPOINTS), XSEG(NUMSEG), LNGAMMAR(NUMSPEC), LNGAMMAC(NUMSPEC), PHI(NUMSPEC), &
  R(NUMSEG,NUMSPEC), XSPEC(NUMSEG,NUMSPEC), XORIG(NUMPOINTS,NUMSPEC), X(NUMPOINTS,NUMSPEC), &
  GAMMA(NUMPOINTS,NUMSPEC), STAT = istat )
WRITE(*,*) istat , " ALLOCATION SUCCESSFUL IF 0"

! Collect VLE data from file
!READ(11,*) TITLE(z)
DO i = 1, NUMSPEC
  READ(11,*) NAME(i,1)

```

```

END DO

DO i = 1, NUMSPEC
  READ(11,*) j , dHfus(i), TM(i), R(1,i), R(2,i), R(3,i), R(4,i)
  ! values for X, Y-, Y+, and Z for species i
END DO

IF ( NUMSPEC == 2 ) THEN
  DO i = 1, NUMPOINTS
    READ(11,*) T(i), XORIG(i,1), XORIG(i,2)          ! PURE Solvent
    X(i,1) = XORIG(i,1)
    X(i,2) = XORIG(i,2)
  END DO
ELSE IF ( NUMSPEC == 3 ) THEN
  DO i = 1, NUMPOINTS
    READ(11,*) T(i), XORIG(i,1), XORIG(i,2), XORIG(i,3)      ! BINARY Solvent
    X(i,1) = XORIG(i,1)
    X(i,2) = XORIG(i,2)
    X(i,3) = XORIG(i,3)
  END DO
ELSE IF ( NUMSPEC == 4 ) THEN
  DO i = 1, NUMPOINTS
    READ(11,*) T(i), XORIG(i,1), XORIG(i,2), XORIG(i,3), XORIG(i,4)  ! TERNARY Solvent
    X(i,1) = XORIG(i,1)
    X(i,2) = XORIG(i,2)
    X(i,3) = XORIG(i,3)
    X(i,4) = XORIG(i,4)
  END DO
ELSE IF ( NUMSPEC == 5 ) THEN
  DO i = 1, NUMPOINTS !QUATERNARY Solvent
    READ(11,*) T(i), XORIG(i,1), XORIG(i,2), XORIG(i,3), XORIG(i,4), XORIG(i,5)
    X(i,1) = XORIG(i,1)
    X(i,2) = XORIG(i,2)
    X(i,3) = XORIG(i,3)
    X(i,4) = XORIG(i,4)
    X(i,5) = XORIG(i,5)
  END DO
ELSE
  WRITE(*,*) "The number of components is not 2, 3, 4, or 5; please refer to source code."
  WRITE(*,*) NUMSPEC
  PAUSE

```



```

END IF

! Generate XSPEC, denoted Xj,I in C. C. Chen (note: XSPEC is independent of system composition)
!-----
DO i = 1, NUMSPEC
  dummy = 0
  DO j = 1, 4
    XSPEC(j,i) = 0
    DO k = 1, 4
      XSPEC(j,i) = XSPEC(j,i) + R(k,i)
    END DO
    XSPEC(j,i) = R(j,i) / XSPEC(j,i)
  END DO
END DO
! calculate over all points
DO n = 1, NUMPOINTS

  DO ITER = 1, 200
    ! Generate XSEG, denoted Xj in C. C. Chen
    !-----
    DO i = 1, 4
      dummy = 0
      XSEG(i) = 0
      DO j = 1, NUMSPEC
        DO k = 1, 4
          dummy = dummy + X(n,j)*R(k,j)
        END DO
        XSEG(i) = XSEG(i) + X(n,j)*R(i,j)
      END DO
      XSEG(i) = XSEG(i) / dummy
    END DO
    ! calculate over all species
    DO i = 1, NUMSPEC

      ! Calculate residual term, LNGAMMAR, from NRTL-SAC model
      !-----
      LNGAMMAR(i) = 0
      DO m = 1, 4
        LNSACTERM1TOP = 0
        LNSACTERM1BOT = 0
        LNSACTERM2 = 0
      END DO
    END DO
  END ITER
END DO

```

```

LNSACITERM1TOP = 0
LNSACITERM1BOT = 0
LNSACITERM2 = 0
DO j = 1, 4          !j is same as j, k, and m' for terms 1 and 2
  LNSACTERM2BOT = 0
  LNSACTERM4TOP = 0
  LNSACTERM4BOT = 0
  LNSACITERM2BOT = 0
  LNSACITERM4TOP = 0
  LNSACITERM4BOT = 0
  DO k = 1, 4       !k is same as k, j, and k for terms 2 and 4
    LNSACTERM2BOT = LNSACTERM2BOT + XSEG(k)*G(k,j)
    LNSACTERM4TOP = LNSACTERM4TOP + XSEG(k)*G(k,j)*TAU(k,j)
    LNSACTERM4BOT = LNSACTERM4BOT + XSEG(k)*G(k,j)
    ! note same as LNSACTERM2BOT

    LNSACITERM2BOT = LNSACITERM2BOT + XSPEC(k,i)*G(k,j)
    LNSACITERM4TOP = LNSACITERM4TOP + XSPEC(k,i)*G(k,j)*TAU(k,j)
    LNSACITERM4BOT = LNSACITERM4BOT + XSPEC(k,i)*G(k,j)
    ! note same as LNSACITERM2BOT
  END DO          ! end iteration of k
  !note that G(m,j) in LNSACTERM2 and LNSACITERM2 have indicies reversed
  !from those of other G values, but this is confirmed by lit (C.C. Chen)
  LNSACTERM1TOP = LNSACTERM1TOP + XSEG(j)*G(j,m)*TAU(j,m)
  LNSACTERM1BOT = LNSACTERM1BOT + XSEG(j)*G(j,m)
  LNSACTERM2 = LNSACTERM2 + XSEG(j) * G(m,j) / &
    LNSACTERM2BOT * ( TAU(m,j) - LNSACTERM4TOP / LNSACTERM4BOT )

  LNSACITERM1TOP = LNSACITERM1TOP + XSPEC(j,i)*G(j,m)*TAU(j,m)
  LNSACITERM1BOT = LNSACITERM1BOT + XSPEC(j,i)*G(j,m)
  LNSACITERM2 = LNSACITERM2 + XSPEC(j,i) * G(m,j) / &
    LNSACITERM2BOT * ( TAU(m,j) - LNSACITERM4TOP / LNSACITERM4BOT )
END DO          ! end iteration of j
LNSAC = LNSACTERM1TOP / LNSACTERM1BOT + LNSACTERM2
LNSACI = LNSACITERM1TOP / LNSACITERM1BOT + LNSACITERM2
LNGAMMAR(i) = LNGAMMAR(i) + R(m,i) * ( LNSAC - LNSACI )
END DO          ! end iteration of m

```

```

!-----
! Calculate combinatorial term, LNGAMMAC, from Flory-Huggins, as done by C. C. Chen
!-----

```

```

dummy = 0
DO k = 1, NUMSPEC
  PHI(k) = 0
  DO j = 1, NUMSPEC
    Rtot = 0
    DO m = 1, NUMSEG
      Rtot = Rtot + R(m,j)
    END DO
    PHI(k) = PHI(k) + X(n,j) * Rtot      ! calculate denominator
  END DO
  Rtot = 0
  DO m = 1, NUMSEG
    Rtot = Rtot + R(m,k)
  END DO
  PHI(k) = ( 1 / PHI(k) ) * X(n,k) * Rtot
  dummy = dummy + PHI(k) / Rtot
END DO
PHI(i) = 0
DO j = 1, NUMSPEC
  Rtot = 0
  DO m = 1, NUMSEG
    Rtot = Rtot + R(m,j)
  END DO
  PHI(i) = PHI(i) + X(n,j) * Rtot      ! calculate denominator
END DO
Rtot = 0
DO m = 1, NUMSEG
  Rtot = Rtot + R(m,i)
END DO
PHI(i) = ( 1 / PHI(i) ) * X(n,i) * Rtot
! model as shown by C. C. Chen
LNGAMMAC(i) = LOG( PHI(i) / X(n,i) ) + 1 - Rtot * dummy
! model as found in G. M. Kontogeorgis et al., Fluid Phase Equilibria 92 (1994) 35-66
! LNGAMMAC(i) = LOG( PHI(i) / X(n,i) ) + 1 - ( PHI(i) / X(n,i) )
! note LOG denotes natural log in FORTRAN
!-----
! Calculate overall activity coefficient, GAMMA, and predicted Y, YPRED
!-----
      GAMMA(n,i) = EXP( LNGAMMAR(i) + LNGAMMAC(i) )
!-----

```

```

END DO          ! end iteration of i
XOLD = X(n,NUMSPEC)
! RECALCULATE SOLUTE MOLE FRACTION FROM DERIVED EQUATION
X(n,NUMSPEC)=EXP(dHfus(NUMSPEC)/(TM(NUMSPEC)*8.314)*(1-TM(NUMSPEC)/T(n))- &
LOG(GAMMA(n,NUMSPEC)))
! 8.314 J/mol*K; 0.001 kJ/ 1 J conversion because dHfus is in kJ/mol

! DAMP CALCULATED SOLUTE MOLE FRACTION
X(n,NUMSPEC) = ( 2.0D0*XOLD + X(n,NUMSPEC) ) / 3.0D0
!CHECK FOR NEGATIVE MOLE FRACTIONS
IF (X(n,NUMSPEC) >= 1.0) THEN
    WRITE (*,*) "SOLUTE MOLEFRACTION CALCULATED AS GREATER THAN 1.0"
    WRITE (*,*) X(n,NUMSPEC)
END IF
!NORMALIZE MOLE FRACTIONS KEEPING SOLVENT RATIO (X1:X2) CONSTANT
MFSUM = 0
DO y = 1, NUMSPEC
    MFSUM = MFSUM + X(n,y)
END DO
DO y = 1, NUMSPEC
    X(n,y) = X(n,y) / MFSUM
END DO
! CALCULATE THE RELATIVE CHANGE OF THE SOLUTE MOLE FRACTION
CHANGE = ABS( ( X(n,NUMSPEC) - XOLD)/( XOLD + 1.0D-16 ) )
!WRITE (*,*) "    RELATIVE CHANGE:", CHANGE
WRITE(*,*) ITER

! COMPARE RELATIVE TOLERANCE
RELTOL = 1.E-8
IF (CHANGE < RELTOL) THEN
    ITERcount = ITER
    GOTO 200
END IF
!END ITERATIVE LOOP
END DO          ! end iteration of ITER
200 dummy = ABS( X(n,NUMSPEC) - XORIG(n,NUMSPEC) ) / XORIG(n,NUMSPEC) * 100    ! relative error
IF ( NUMSPEC == 2 ) THEN
    WRITE(12,111) ITER, NAME(2,1), NAME(1,1), T(n), GAMMA(n,1), GAMMA(n,2), XORIG(n,1),&
        XORIG(n,NUMSPEC), X(n,1), X(n,NUMSPEC), dummy          ! Pure Solvent
111 FORMAT(1X,I3,5X,A15,2X,A15,2X,E,2X,E,2X,E,2X,E,2X,E,2X,E,2X,E,2X,E)
ELSE

```

```

        WRITE(*,*) "The number of components is not 2 (pure solvent solubility); &
        source code must be changed for different systems."
        PAUSE
    END IF
    dummy = 0
END DO
        ! end iteration of n
DEALLOCATE( NAME, TITLE, dhfus, TM, T, XSEG, LNGAMMAR, LNGAMMAC, PHI, R, XSPEC, XORIG, X, GAMMA, STAT
= istat)
    WRITE(*,*) istat, " DEALLOCATION SUCCESSFUL IF 0"
    WRITE(*,*) "System ", z, " complete."
END DO
        ! end iteration of z
CLOSE(12)
CLOSE(11)
DEALLOCATE ( TAU, ALPHA, G )
WRITE(*,*) "DONE!"
PAUSE
END PROGRAM NRTL_SAC

```

Appendix C: Sample Database Entry

We present the Accelrys MS output and sigma profile of water (VT-1076) as a sample database entry from the VT-2005 Sigma Profile Database. In the sigma profile, the value of the screening charge density range from $-0.025 \text{ e}/\text{\AA}^2$ – $0.025 \text{ e}/\text{\AA}^2$. Each table below represents one part of the entire database entry. The files, sigma profile, COSMO calculation output, geometry optimization output, and energy calculation output, are in simple TXT file format. The format for each entry is identical for both databases.

Table C.1: Example water (VT-1076) sigma profile from the VT-2005 Sigma Profile Database.

Component Number	VT-1076
Component Name	Water
Chemical Formula	H2O
CAS #	7732-18-5
Molecular Volume, V (\AA^3)	25.73454
<u>Screening Charge Density, σ_m ($\text{e}/\text{\AA}^2$)</u>	<u>Sigma Profile, P(σ)A(σ), (\AA^2)</u>
-0.025	0
-0.024	0
-0.023	0
-0.022	0
-0.021	0
-0.020	0
-0.019	0
-0.018	0
-0.017	0
-0.016	0.633945732
-0.015	2.377619002
-0.014	2.968564534
-0.013	1.444555526
-0.012	1.822041754
-0.011	2.003554175
-0.010	2.54564325

-0.009	1.958697314
-0.008	0.996725046
-0.007	0.401363668
-0.006	0.487173311
-0.005	1.165354697
-0.004	0.803501492
-0.003	1.348667493
-0.002	0.060873008
-0.001	0.018527579
0.000	0.836695456
0.001	1.07831424
0.002	0.499516267
0.003	1.156890953
0.004	0.733891941
0.005	0.706732991
0.006	1.390222103
0.007	0.917417664
0.008	1.442660977
0.009	0.433118284
0.010	0.950353683
0.011	1.22948396
0.012	2.308847865
0.013	2.324487815
0.014	2.722118291
0.015	1.368180697
0.016	2.053392475
0.017	0.080146761
0.018	0
0.019	0
0.020	0
0.021	0
0.022	0
0.023	0
0.024	0
0.025	0

Molecular Surface Area, Å²

43.26928

Table C.2: Accelrys MS Geometry Optimization Task Output (OUTMOL file) for water, VT-1076. geometry optimization task output is iterative, therefore we truncate display only the last iteration.

```
=====
Materials Studio DMol^3 version 2.2
compiled on Oct 18 2002 07:08:48
=====
```

```
=====
Density Functional Theory Electronic Structure Program
Copyright (c) 2002, Accelrys Inc. All rights reserved.
Cite work using this program as:
B. Delley, J. Chem. Phys. 92, 508 (1990).
B. Delley, J. Chem. Phys. 113, 7756 (2000).
DMol^3 is available as part of Materials Studio.
=====
```

```
DATE:      Apr 13 10:12:21 2004
Basis set is read from file:
C:\PROGRA~1\Accelrys\MATERI~1\DMol3\..\Data\Resources\Quantum\DMol3\BASFILE_v4.0
```

```
Message: License checkout of MS_dmol successful
Message: Number of licenses checked out      1
```

```
Geometry is read from file: 1076.car
INCOOR, atomic coordinates in au (for archive):
```

>8

\$coordinates

O	-5.74870322136619	0.79792378035512	0.00027985141137
H	-6.44852978930695	2.77533371809792	-0.00026427251032
H	-3.65110722475524	0.79856792180421	-0.00034185521079

\$end

>8

```
N_atoms =      3      N_atom_types =  2
```

```
INPUT_DMOL keywords (for archive):
```

```

>8
<--
# Task parameters <--
Calculate optimize <--
Opt_energy_convergence 1.0000e-005 <--
Opt_gradient_convergence 2.0000e-003 A <--
Opt_displacement_convergence 5.0000e-003 A <--
Opt_iterations 50 <--
Opt_max_displacement 0.3000 A <--
<--
# Electronic parameters <--
Spin_polarization restricted <--
Charge 0.0000 <--
Basis dnp <--
Atom_rcut 5.5000 angstrom <--
Pseudopotential none <--
Functional vwn-bp <--
Aux_density octupole <--
Integration_grid fine <--
Scf_density_convergence 1.0000e-006 <--
Scf_charge_mixing 0.2000 <--
Scf_iterations 50 <--
Scf_diis 6 pulay <--
Occupation fermi <--
<--
# Print options <--
Print eigval_last_it <--
<--
# Calculated properties <--
>8

```

```

Density functional:
  Becke 1988 exchange
    gradient corrected exchange potential used in SCF
  P91_gradients for COSMO run
    gradient corrected correlation potential used in SCF
  VWN local correlation (p91-pwc)

```

```

Calculation is Spin_restricted

```

Molecule has been put into center of mass coordinate system
Molecule has been rotated to standard orientation

Symmetry group of the molecule: c2v

** GEOMETRY OPTIMIZATION IN DELOCALIZED COORDINATES **

Searching for a Minimum

Optimization Cycle: 0

Input Coordinates (Angstroms)

	ATOM	X	Y	Z
1	O	0.000000	0.000000	-0.427236
2	H	0.906315	0.000000	0.213618
3	H	-0.906315	0.000000	0.213618

Message: Generating delocalized internals

Total number of primitive bonds: 2
Total number of primitive angles: 1
Total number of primitive dihedrals: 0
Total number of primitive internals: 3

Message: Generation of delocalized internals is successful

Specifications for basis set selection:

atomic cutoff radius 10.39 au

Hydrogen nbas= 1 z= 1. 5 radial functions, e_ref= -0.0400364Ha

n=1	L=0	occ= 1.00	e=	-0.240078Ha	-6.5329eV	
n=1	L=0	occ= 0.00	e=	-0.845000Ha	-22.9936eV	
n=2	L=1	occ= 0.00	e=	-0.206651Ha	-5.6233eV	
n=2	L=1	occ= 0.00	e=	-2.000000Ha	-54.4228eV	eliminated
n=1	L=0	occ= 0.00	e=	-8.000000Ha	-217.6912eV	eliminated

Oxygen nbas= 2 z= 8. 10 radial functions, e_ref= -0.0566492Ha

n=1	L=0	occ= 2.00	e=	-18.912344Ha	-514.6313eV	
n=2	L=0	occ= 2.00	e=	-0.878677Ha	-23.9100eV	
n=2	L=1	occ= 4.00	e=	-0.332064Ha	-9.0359eV	

n=2	L=0	occ= 0.00	e=	-2.144989Ha	-58.3682eV	
n=2	L=1	occ= 0.00	e=	-1.592070Ha	-43.3224eV	
n=3	L=2	occ= 0.00	e=	-1.388886Ha	-37.7935eV	
n=2	L=1	occ= 0.00	e=	-3.125000Ha	-85.0356eV	eliminated
n=1	L=0	occ= 0.00	e=	-12.500000Ha	-340.1425eV	eliminated
n=3	L=2	occ= 0.00	e=	-2.722222Ha	-74.0755eV	eliminated
n=2	L=1	occ= 0.00	e=	-6.125000Ha	-166.6698eV	eliminated

Point group symmetry c2v symmetry orbital prototypes generated (SYMDEC)

Symmetry orbitals

n	norb	representation
1	11	A1.1
2	2	A2.1
3	7	B1.1
4	4	B2.1

total number of valence orbitals: 24

molecule charge=	0.0	active electron number=	10.0
including core=	10.0	(without charge=	10.0)

Integration points and checksum: 3507 9.999999 105

extra disk use on option Direct_scf off= 5.3Mbytes

 Memory Stack Information

	Number of elements	Mbytes
integer arrays	19475	0.1
real arrays	1058405	8.1
total	1077880	8.1

+++ Entering Optimization Section +++
 Message: Entering optimization section

~~~~~ Start Computing SCF Energy/Gradient ~~~~~

[3 iterations truncated]

~~~~~ End Computing SCF Energy/Gradient ~~~~~

** GEOMETRY OPTIMIZATION IN DELOCALIZED COORDINATES **

Searching for a Minimum

Optimization Cycle: 4

Input Coordinates (Angstroms)

| ATOM | X | Y | Z |
|------|-----------|----------|-----------|
| 1 O | 0.000000 | 0.000000 | -0.409364 |
| 2 H | 0.770804 | 0.000000 | 0.204682 |
| 3 H | -0.770804 | 0.000000 | 0.204682 |

| Cycle | Total Energy | Energy change | Max Gradient | Max Displacement |
|---------|--------------|---------------|--------------|------------------|
| opt== 4 | -76.4538907 | -0.0008456 | 0.004890 | 0.005503 |

~~~~~ Start Computing SCF Energy/Gradient ~~~~~

Integration points and checksum: 3507 9.999996 105  
Message: Start SCF iterations

|    | Total E (au) | Binding E (au) | Convergence | Time (m) | Iter |
|----|--------------|----------------|-------------|----------|------|
| Ef | -76.4539231  | -0.3698236     | 0.21E-02    | 0.137    | 1    |
| Ef | -76.4539253  | -0.3698258     | 0.89E-03    | 0.139    | 2    |
| Ef | -76.4539292  | -0.3698297     | 0.34E-03    | 0.141    | 3    |
| Ef | -76.4539291  | -0.3698296     | 0.26E-04    | 0.144    | 4    |
| Ef | -76.4539291  | -0.3698296     | 0.46E-05    | 0.146    | 5    |
| Ef | -76.4539290  | -0.3698295     | 0.22E-05    | 0.148    | 6    |
| Ef | -76.4539290  | -0.3698295     | 0.19E-06    | 0.150    | 7    |

Message: SCF converged

Energy of Highest Occupied Molecular Orbital -0.26466Ha -7.202eV  
HOMO is orbital number 5  
LUMO is orbital number 6

| state | eigenvalue<br>(au) | occupation<br>(ev) |
|-------|--------------------|--------------------|
|-------|--------------------|--------------------|

|    |   |   |      |            |          |       |
|----|---|---|------|------------|----------|-------|
| 1  | + | 1 | A1.1 | -18.774480 | -510.880 | 2.000 |
| 2  | + | 2 | A1.1 | -0.927317  | -25.234  | 2.000 |
| 3  | + | 1 | B1.1 | -0.482100  | -13.119  | 2.000 |
| 4  | + | 3 | A1.1 | -0.346770  | -9.436   | 2.000 |
| 5  | + | 1 | B2.1 | -0.264657  | -7.202   | 2.000 |
| 6  | + | 4 | A1.1 | -0.007476  | -0.203   | 0.000 |
| 7  | + | 2 | B1.1 | 0.048166   | 1.311    | 0.000 |
| 8  | + | 2 | B2.1 | 0.113879   | 3.099    | 0.000 |
| 9  | + | 5 | A1.1 | 0.117449   | 3.196    | 0.000 |
| 10 | + | 3 | B1.1 | 0.183667   | 4.998    | 0.000 |
| 11 | + | 6 | A1.1 | 0.196757   | 5.354    | 0.000 |
| 12 | + | 1 | A2.1 | 0.199188   | 5.420    | 0.000 |
| 13 | + | 4 | B1.1 | 0.282961   | 7.700    | 0.000 |
| 14 | + | 7 | A1.1 | 0.535401   | 14.569   | 0.000 |
| 15 | + | 8 | A1.1 | 0.725214   | 19.734   | 0.000 |
| 16 | + | 3 | B2.1 | 0.737620   | 20.072   | 0.000 |

Orbital occupation is:

3 A1( 2)  
0 A2( 2)  
1 B1( 2)  
1 B2( 2)

Total number electrons: 10.0

| df |   | ATOMIC COORDINATES (au) |          |           | DERIVATIVES (au) |          |           |
|----|---|-------------------------|----------|-----------|------------------|----------|-----------|
| df |   | x                       | y        | z         | x                | y        | z         |
| df | O | 0.000000                | 0.000000 | -0.768082 | 0.000000         | 0.000000 | -0.000723 |
| df | H | 1.454618                | 0.000000 | 0.384041  | 0.000144         | 0.000000 | 0.000361  |
| df | H | -1.454618               | 0.000000 | 0.384041  | -0.000144        | 0.000000 | 0.000361  |

|                | Total E (au) | Binding E (au) | Convergence      | Time (m)       | Iter |
|----------------|--------------|----------------|------------------|----------------|------|
| Ef             | -76.4539290  | -0.3698295     | 0.19E-06         | 0.155          | 8    |
| binding energy | -0.369830Ha  | -10.06358eV    | -232.076kcal/mol | -971.007kJ/mol |      |

~~~~~ End Computing SCF Energy/Gradient ~~~~~

** GEOMETRY OPTIMIZATION IN DELOCALIZED COORDINATES **
Searching for a Minimum

Optimization Cycle: 5

Input Coordinates (Angstroms)

| ATOM | | X | Y | Z |
|------|---|-----------|----------|-----------|
| 1 | O | 0.000000 | 0.000000 | -0.406452 |
| 2 | H | 0.769751 | 0.000000 | 0.203226 |
| 3 | H | -0.769751 | 0.000000 | 0.203226 |

| Cycle | Total Energy | Energy change | Max Gradient | Max Displacement |
|---------|--------------|---------------|--------------|------------------|
| opt== 5 | -76.4539290 | -0.0000384 | 0.000723 | 0.001411 |

~~~~~ Start Computing SCF Energy/Gradient ~~~~~

-----  
Integration points and checksum: 3507 9.999996 105  
Message: Start SCF iterations

|    | Total E (au) | Binding E (au) | Convergence | Time (m) | Iter |
|----|--------------|----------------|-------------|----------|------|
| Ef | -76.4539297  | -0.3698302     | 0.25E-03    | 0.160    | 1    |
| Ef | -76.4539306  | -0.3698311     | 0.11E-03    | 0.162    | 2    |
| Ef | -76.4539312  | -0.3698317     | 0.50E-04    | 0.164    | 3    |
| Ef | -76.4539313  | -0.3698318     | 0.51E-05    | 0.166    | 4    |
| Ef | -76.4539313  | -0.3698318     | 0.79E-06    | 0.168    | 5    |

Message: SCF converged

Energy of Highest Occupied Molecular Orbital -0.26467Ha -7.202eV  
HOMO is orbital number 5  
LUMO is orbital number 6

| state |   |   | eigenvalue |            | occupation |       |
|-------|---|---|------------|------------|------------|-------|
|       |   |   | (au)       | (ev)       |            |       |
| 1     | + | 1 | A1.1       | -18.774352 | -510.876   | 2.000 |
| 2     | + | 2 | A1.1       | -0.927427  | -25.237    | 2.000 |
| 3     | + | 1 | B1.1       | -0.482486  | -13.129    | 2.000 |
| 4     | + | 3 | A1.1       | -0.346568  | -9.431     | 2.000 |
| 5     | + | 1 | B2.1       | -0.264672  | -7.202     | 2.000 |
| 6     | + | 4 | A1.1       | -0.007346  | -0.200     | 0.000 |
| 7     | + | 2 | B1.1       | 0.048252   | 1.313      | 0.000 |
| 8     | + | 2 | B2.1       | 0.113916   | 3.100      | 0.000 |
| 9     | + | 5 | A1.1       | 0.117442   | 3.196      | 0.000 |

|    |   |   |      |          |        |       |
|----|---|---|------|----------|--------|-------|
| 10 | + | 3 | B1.1 | 0.183702 | 4.999  | 0.000 |
| 11 | + | 6 | A1.1 | 0.196730 | 5.353  | 0.000 |
| 12 | + | 1 | A2.1 | 0.199144 | 5.419  | 0.000 |
| 13 | + | 4 | B1.1 | 0.283220 | 7.707  | 0.000 |
| 14 | + | 7 | A1.1 | 0.535550 | 14.573 | 0.000 |
| 15 | + | 8 | A1.1 | 0.725372 | 19.738 | 0.000 |
| 16 | + | 3 | B2.1 | 0.737748 | 20.075 | 0.000 |

Orbital occupation is:

```

3  A1( 2)
0  A2( 2)
1  B1( 2)
1  B2( 2)

```

**Table C.3: Accelrys MS Energy Calculation Task Output (OUTMOL file) for water, VT-1076. The energy calculation task output calculates atomic coordinates, total energy, and molecular cavity dimensions. The energy calculation task runs simultaneously with the COSMO calculation for a condensed phase molecule.**

```

=====
Materials Studio DMol^3 version 2.2
compiled on Oct 18 2002 07:08:48
=====

Density Functional Theory Electronic Structure Program
Copyright (c) 2002, Accelrys Inc. All rights reserved.
Cite work using this program as:
B. Delley, J. Chem. Phys. 92, 508 (1990).
B. Delley, J. Chem. Phys. 113, 7756 (2000).
DMol^3 is available as part of Materials Studio.
=====

```

DATE: Apr 13 10:21:04 2004

Basis set is read from file:

C:\PROGRA~1\Accelrys\MATERI~1\DMol3\..\Data\Resources\Quantum\DMol3\BASFILE\_v4.0

Message: License checkout of MS\_dmol successful  
Message: Number of licenses checked out 1

Geometry is read from file: 1076.car  
INCOOR, atomic coordinates in au (for archive):

---

>8

\$coordinates

|   |                   |                  |                   |
|---|-------------------|------------------|-------------------|
| O | -5.72516870618525 | 0.83122866266686 | 0.00026022282753  |
| H | -6.25008701574016 | 2.61013929568994 | -0.00026028707821 |
| H | -3.87308451539269 | 0.93045746190045 | -0.00032621016933 |

\$end

---

>8

N\_atoms = 3 N\_atom\_types = 2  
INPUT\_DMOL keywords (for archive):

---

>8

# Task parameters

Calculate energy

# Electronic parameters

|                         |                 |
|-------------------------|-----------------|
| Spin_polarization       | restricted      |
| Charge                  | 0.0000          |
| Basis                   | dnp             |
| Atom_rcut               | 5.5000 angstrom |
| Pseudopotential         | none            |
| Functional              | vwn-bp          |
| Aux_density             | octupole        |
| Integration_grid        | fine            |
| Scf_density_convergence | 1.0000e-006     |
| Scf_charge_mixing       | 0.2000          |
| Scf_iterations          | 50              |
| Scf_diis                | 6 pulay         |
| Occupation              | fermi           |

# Environment Keywords

Cosmo on



```

Cosmo_Grid_Size          1082          <--
Cosmo_Segments           92           <--
Cosmo_Solvent_Radius     1.300000     <--
Cosmo_A-Matrix_Cutoff    7.000000     <--
Cosmo_Radius_Incr        0.000000     <--
Cosmo_A-Constant         1.882190     <--
Cosmo_B-Constant         0.010140     <--
Cosmo_RadCorr_Incr       0.150000     <--
Cosmo_Atomic_Radii      <--
  1          1.300
  6          2.000
  7          1.830
  8          1.720
  9          1.720
 15          2.120
 16          2.160
 17          2.050
 35          2.160
 53          2.320
# Print options          <--
Print                   eigval_last_it <--
                        <--
# Calculated properties <--
Plot                   density         <--
Plot                   potential       <--
Grid                   msbox  3 0.2500 0.2500 0.2500 3.0000 <--
                        >8

```

Density functional:

```

Becke 1988 exchange
  gradient corrected exchange potential used in SCF
  P91_gradients for COSMO run
  gradient corrected correlation potential used in SCF
  VWN local correlation (p91-pwc)
Warning: Current COSMO implementation can not use symmetry
Calculation is Spin_restricted
Molecule has been put into center of mass coordinate system
Molecule has been rotated to standard orientation
Symmetry group of the molecule: c2v
Molecule has been snapped to exact symmetry

```

Specifications for basis set selection:

atomic cutoff radius 10.39 au

Hydrogen nbas= 1 z= 1. 5 radial functions, e\_ref= -0.0400364Ha  
n=1 L=0 occ= 1.00 e= -0.240078Ha -6.5329eV  
n=1 L=0 occ= 0.00 e= -0.845000Ha -22.9936eV  
n=2 L=1 occ= 0.00 e= -0.206651Ha -5.6233eV  
n=2 L=1 occ= 0.00 e= -2.000000Ha -54.4228eV eliminated  
n=1 L=0 occ= 0.00 e= -8.000000Ha -217.6912eV eliminated  
Oxygen nbas= 2 z= 8. 10 radial functions, e\_ref= -0.0566492Ha  
n=1 L=0 occ= 2.00 e= -18.912344Ha -514.6313eV  
n=2 L=0 occ= 2.00 e= -0.878677Ha -23.9100eV  
n=2 L=1 occ= 4.00 e= -0.332064Ha -9.0359eV  
n=2 L=0 occ= 0.00 e= -2.144989Ha -58.3682eV  
n=2 L=1 occ= 0.00 e= -1.592070Ha -43.3224eV  
n=3 L=2 occ= 0.00 e= -1.388886Ha -37.7935eV  
n=2 L=1 occ= 0.00 e= -3.125000Ha -85.0356eV eliminated  
n=1 L=0 occ= 0.00 e= -12.500000Ha -340.1425eV eliminated  
n=3 L=2 occ= 0.00 e= -2.722222Ha -74.0755eV eliminated  
n=2 L=1 occ= 0.00 e= -6.125000Ha -166.6698eV eliminated

Symmetry orbitals

n norb representation  
1 24 a

total number of valence orbitals: 24

molecule charge= 0.0 active electron number= 10.0  
including core= 10.0 (without charge= 10.0)  
Integration points and checksum: 12357 9.999996 302  
extra disk use on option Direct\_scf off= 9.1Mbytes

-----  
Memory Stack Information  
-----

|                | Number of elements | Mbytes |
|----------------|--------------------|--------|
| integer arrays | 29854              | 0.1    |
| real arrays    | 1149436            | 8.8    |
| total          | 1179290            | 8.9    |

-----

COSMO input

In this run the solvent is an ideal conductor

Dielectric Constant = infinity  
 Basic Grid Size = 1082  
 Number of Segments = 92  
 Solvent Radius = 1.30  
 A - Matrix Cutoff = 7.00  
 Radius Increment = 0.00  
 Radius Increase for Outlying Charge corr. = 0.15  
 Non-Electrostatic Energy = A+B\*area  
 A = 1.88219 B = 0.01014

van der Waals radius

| atomic number | vdW  | vdW+raddc |
|---------------|------|-----------|
| 1 8           | 1.72 | 1.72      |
| 2 1           | 1.30 | 1.30      |
| 3 1           | 1.30 | 1.30      |

The program will read COSMO basic grid from file:

C:\PROGRA~1\Accelrys\MATERI~1\DMol3\..\Data\Resources\Quantum\DMol3\COSGRID

nps(usual segment ) = 92  
 nps(hydrogen segment) = 32  
 Total number of COSMO charges: 136

+++ Entering SCF Section +++

Message: Entering SCF section

~~~~~ Start Computing SCF Energy/Gradient ~~~~~

 Message: Start SCF iterations

| | Total E (au) | Binding E (au) | Convergence | Time (m) | Iter |
|----|--------------|----------------|-------------|----------|------|
| Ef | -76.5245896 | -0.4404901 | 0.18E+00 | 0.043 | 1 |
| Ef | -76.4680061 | -0.3839066 | 0.93E-01 | 0.054 | 2 |
| Ef | -76.4596955 | -0.3755960 | 0.55E-01 | 0.066 | 3 |
| Ef | -76.4667126 | -0.3826131 | 0.34E-01 | 0.078 | 4 |
| Ef | -76.4649490 | -0.3808495 | 0.35E-02 | 0.091 | 5 |
| Ef | -76.4650144 | -0.3809149 | 0.54E-02 | 0.103 | 6 |
| Ef | -76.4649379 | -0.3808384 | 0.23E-03 | 0.116 | 7 |
| Ef | -76.4649399 | -0.3808404 | 0.20E-03 | 0.129 | 8 |
| Ef | -76.4649388 | -0.3808393 | 0.42E-05 | 0.141 | 9 |
| Ef | -76.4649388 | -0.3808393 | 0.25E-05 | 0.153 | 10 |

| | | | | | |
|----|-------------|------------|----------|-------|----|
| Ef | -76.4649388 | -0.3808393 | 0.19E-05 | 0.165 | 11 |
| Ef | -76.4649388 | -0.3808392 | 0.53E-06 | 0.177 | 12 |

Message: SCF converged

Energy of Highest Occupied Molecular Orbital -0.26486Ha -7.207eV
HOMO is orbital number 5
LUMO is orbital number 6

| state | | | | eigenvalue | | occupation |
|-------|---|----|---|------------|----------|------------|
| | | | | (au) | (ev) | |
| 1 | + | 1 | a | -18.768737 | -510.724 | 2.000 |
| 2 | + | 2 | a | -0.920554 | -25.050 | 2.000 |
| 3 | + | 3 | a | -0.473206 | -12.877 | 2.000 |
| 4 | + | 4 | a | -0.345153 | -9.392 | 2.000 |
| 5 | + | 5 | a | -0.264860 | -7.207 | 2.000 |
| 6 | + | 6 | a | 0.017508 | 0.476 | 0.000 |
| 7 | + | 7 | a | 0.074959 | 2.040 | 0.000 |
| 8 | + | 8 | a | 0.123778 | 3.368 | 0.000 |
| 9 | + | 9 | a | 0.128144 | 3.487 | 0.000 |
| 10 | + | 10 | a | 0.192979 | 5.251 | 0.000 |
| 11 | + | 11 | a | 0.201699 | 5.489 | 0.000 |
| 12 | + | 12 | a | 0.213074 | 5.798 | 0.000 |
| 13 | + | 13 | a | 0.301004 | 8.191 | 0.000 |
| 14 | + | 14 | a | 0.546646 | 14.875 | 0.000 |
| 15 | + | 15 | a | 0.713291 | 19.410 | 0.000 |

Orbital occupation is:

5 a (2)

Total number electrons: 10.0

| | | | | | |
|----------------|-------------|-------------|------------------|----------------|----|
| Ef | -76.4649388 | -0.3808393 | 0.43E-06 | 0.192 | 13 |
| binding energy | -0.380839Ha | -10.36317eV | -238.985kcal/mol | -999.913kJ/mol | |

DMol3/COSMO Results

Total energy (au) [TE]
including COSMO solvation energy, Eq.2 = -76.464939
Dielectric energy (au) (1/2<q|U>) [DE] = -0.012563

Total energy corrected (au) [TE(corr)]
TE(corr) = TE + DE(corr) - DE = -76.465425
Dielectric energy corrected (au) [DE(corr)]

```

DE(corr) = (q+q")A(q+q") = -0.013048

(U is the potential on the inner cavity
it depends on the accurate density)

Sum of polarization charges q = -0.02223
Sum of polarization charges(corrected)
(q+q") = -0.00131

Total surface area (Angstrom**2) = 43.26923
Total volume of cavity (A**3) = 25.73454

Nonelectrostatic Solvation Energy (kcal/mol) = 2.320940

Total energy (au) [TE]
+ Non-Electrostatic Energy = -76.461241
Total energy corrected (au) [TE(corr)]
+ Non-Electrostatic Energy = -76.461727
~~~~~ End Computing SCF Energy/Gradient ~~~~~
+++ Entering Properties Section +++
Plotting output:
property: file name:
DMol3 total electron density 1076_density.grd
DMol3 electrostatic potential 1076_potential.grd

grid specifications: I_dim, origin, n_intervals to corner:
 3 -7.0865 -5.6692 -6.1416 26 -7.0865 -5.6692 6.1416
24 -7.0865 5.6692 -6.1416 30 7.0865 -5.6692 -6.1416

Message: DMol3 job finished successfully
time all done 0.20m 12.06s

```

Vita

Paul Eric Mullins was born June 6, 1981 in Richmond, Virginia to Phillip and Joalyn Mullins. He graduated Magna Cum Laude with a Bachelor of Science in Chemical Engineering from Virginia Polytechnic Institute and State University in May 2004. He completed his Masters of Science in Chemical Engineering under Dr. Y. A. Liu at Virginia Tech in the Spring 2007 semester. During his stay at Virginia Tech, he competed with the Virginia Tech Men's Club Volleyball team on an intercollegiate club sports level. In April 2006 at the NIRSA Club Volleyball National Championship Tournament in Salt Lake City, Utah, Virginia Tech finished T-9th, one of the Men's Club's top finishes in the history of Virginia Tech. After completion of his graduate studies, he accepted a position with Eastman Chemical Company as a chemical engineer in Kingsport, Tennessee.



The 2006 Virginia Tech Men's Volleyball Team in Salt Lake City, UT. 1st row, L to R: Nathan King(11), T.J. Purcell(16), Ryan Arcese(1), Chris Hazell(5), Casey Doyle(7). 2nd row, L to R: Andrew Fuller (6), Will McGough (20), Paul Eric Mullins (9), Steve Sharp (17), Steve Donaldson (3), Nate Buttermore (24), David Tatum (23)