



**A Universal Segment Approach for the Prediction of the Activity Coefficient**

by

Kuveneshan Moodley

B.Sc.Eng., M.Sc.Eng. (*UKZN*)

In fulfilment of the degree of Doctor of Philosophy

In the School of Engineering, Discipline of Chemical Engineering

University of KwaZulu-Natal

July 2016

Supervisors: Prof. D. Ramjugernath and Prof. J. Rarey

## ABSTRACT

This study comprised an investigation into solid-liquid equilibrium prediction, measurement and modelling for active pharmaceutical ingredients, and solvents, employed in the pharmaceutical industry. Available experimental data, new experimental data, and novel measuring techniques, as well as existing predictive thermodynamic activity coefficient model revisions, were investigated. Thereafter, and more centrally, a novel model for the prediction of activity coefficients, at solid-liquid equilibrium, which incorporates global optimization strategies in its training, is presented.

The model draws from the segment interaction (via segment surface area), approach in solid-liquid equilibrium modelling for molecules, and extends this concept to interactions between functional groups. Ultimately, a group-interaction predictive method is proposed that is based on the popular UNIFAC-type method (Fredenslund et al. 1975). The model is termed the Universal Segment Activity Coefficient (UNISAC) model.

A detailed literature review was conducted, with respect to the application of the popular predictive models to solid-liquid phase equilibrium (SLE) problems, involving structurally complex solutes, using experimental data available in the literature (Moodley et al., 2016 (a)). This was undertaken to identify any practical and theoretical limitations in the available models. Activity coefficient predictions by the NRTL-SAC ((Chen and Song 2004), Chen and Crafts, 2006), UNIFAC (Fredenslund et al., 1975), modified UNIFAC (Dortmund) (Weidlich and Gmehling, 1987), COSMO-RS (OL) (Grensemann and Gmehling, 2005), and COSMO-SAC (Lin and Sandler, 2002), were carried out, based on available group constants and sigma profiles, in order to evaluate the predictive capabilities of these models.

The quality of the models is assessed, based on the percentage deviation between experimental data and model predictions. The NRTL-SAC model is found to provide the best replication of solubility rank, for the cases tested. It, however, was not as widely applicable as the majority of the other models tested, due to the lack of available model parameters in the literature. These results correspond to a comprehensive comparison conducted by Diedrichs and Gmehling (2011).

After identifying the limitations of the existing predictive methods, the UNISAC model is proposed (Moodley et al, 2015 (b)). The predictive model was initially applied to solid-liquid systems containing a set of 18 structurally diverse, complex pharmaceuticals, in a variety of

solvents, and compared to popular qualitative solubility prediction methods, such as NRTL-SAC and the UNIFAC based methods. Furthermore, the Akaike Information Criterion (AIC) (Akaike, 1974) and Focused Information Criterion (FIC) (Claeskens and Hjort, 2003) were used to establish the relative quality of the solubility predictions. The AIC scores recommend the UNISAC model for over 90% of the test cases, while the FIC scores recommend UNISAC in over 75% of the test cases.

The sensitivity of the UNISAC model parameters was highlighted during the initial testing phase, which indicated the need to employ a more rigorous method of determining parameters of the model, by optimization to the global minimum. It was decided that the Krill Herd algorithm optimization technique (Gandomi and Alavi, 2012), be employed to accomplish this. To verify the suitability of this decision, the algorithm was applied to phase stability (PS) and phase equilibrium calculations in non-reactive (PE) and reactive (rPE) systems, where global minimization of the total Gibbs energy is necessary. The results were compared to other methods from the literature (Moodley et al., 2015 (c)). The Krill Herd algorithm was found to reliably determine the desired global optima in PS, PE and rPE problems. The algorithm outperformed or matched all other methods considered for comparison, including swarm intelligence and genetic algorithms, with an average success rate of 89.5 %, and with an average number of function evaluations of 1406.

The UNISAC model was then reviewed, and extended, to incorporate the significantly more detailed group fragmentation scheme of Moller et al. (2008), to improve the range of application of the model. New UNISAC segment group area parameters that were obtained by data fitting, using the Krill Herd Algorithm as an optimization tool, were calculated. This Extended UNISAC model was then used to predict SLE compositions, or temperatures, of a large volume of experimental binary and ternary system data, available in the literature, (over 4000 data points), and was compared to predictions by the UNIFAC-based and COSMO-based models (Moodley et al., 2016 (d)).

The AIC scores suggest that the Extended UNISAC model is superior to the original UNIFAC, modified UNIFAC (Dortmund) (2013), COSMO-RS(OL), and COSMO-SAC models, with relative AIC scores of 1.95, 4.17, 2.17 and 2.09. In terms of percentage deviations alone between experimental and predicted values, the modified UNIFAC (Dortmund) model, and original UNIFAC models, proved superior at 21.03% and 29.03% respectively; however, the Extended UNISAC model was a close third at 32.99%.

As a conservative measure to ensure that inter-correlation of the training set did not occur, previously unmeasured data was desired as a test set, to verify the ability of the Extended UNISAC model to estimate data outside of the training set. To accomplish this, SLE measurements were conducted for the systems diosgenin/ estriol/ prednisolone/ hydrocortisone/ betulin and estrone. These measurements were undertaken in over 10 diverse organic solvents, and water, at atmospheric pressure, within the temperature range 293.2-328.2 K, by employing combined digital thermal analysis and thermal gravimetric analysis, to determine compositions at saturation (Moodley et al., 2016 (e), Moodley et al., 2016 (f), Moodley et al., 2016 (g)).

This previously unmeasured test set data was compared to predictions by the Extended UNISAC, UNIFAC-based and COSMO-based methods. It was found that the Extended UNISAC model can qualitatively predict the solubility in the systems measured (where applicable), comparably to the other popular methods tested. The desirable advantage is that the number of model parameters required to describe mixture activities is far lower than for the group contribution and COSMO-based methods.

Future developments of the Extended UNISAC model were then considered, which included the preliminary testing of alternate combinatorial expressions, to better account for size-shape effects on the activity coefficient. The limitations of the Extended UNISAC model are also discussed.



## DECLARATION 1: Statement of Original Work

The work presented in this thesis was carried out within the Thermodynamic Research Unit in the School of Engineering at the University of KwaZulu-Natal, Durban, from March 2012 to July 2016 under the supervision of Professor D. Ramjugernath and Professor J. Rarey.

This thesis is submitted as the full requirement for the award of a Ph.D. degree in Chemical Engineering.

I, Kuveneshan Moodley, therefore declare that:

- (i) The research reported in this thesis except where otherwise indicated, is my original work.
- (ii) This thesis has not been submitted for any degree or examination at any other university.
- (iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
- (iv) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
  - a) Their words have been re-written but the general information attributed to them has been referenced;
  - b) Where their exact words have been used, their writing has been placed inside quotation marks, and referenced.
- (v) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source has been detailed in the dissertation and in the *References* sections.
- (vi) As this thesis is submitted in the journal manuscript format, under Rule DR9 c) and d) of the University of KwaZulu-Natal, only manuscript versions of published or unpublished work are presented.

---

Kuveneshan Moodley

As the candidate's supervisor, I, Prof. D. Ramjugernath, approved this thesis for submission.

---

Professor D. Ramjugernath

## DECLARATION 2: Contribution to publications

Details of contribution to publications and manuscripts

1. Moodley, K., Rarey, J., Ramjugernath, D., (2015) (a), Model Evaluation for the Prediction of Solubility of Active Pharmaceutical Ingredients (APIs). *Manuscript in preparation*.

**Contribution:** I designed the study with suggestions from Prof. Ramjugernath and Prof Rarey. I performed the calculations and wrote the manuscript with suggestions from Prof. Ramjugernath and Prof Rarey.

2. Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (b), A Universal Segment Approach for the Prediction of the Activity Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents. *Fluid Phase Equilibria*, 396, pp. 98–110.

**Contribution:** The segment surface concept of a functional group was hypothesized by Prof Rarey. I the designed the study with numerous supporting ideas and strong guidance from Prof Rarey. I performed the calculations and wrote the manuscript with suggestions from Prof. Ramjugernath and Prof Rarey.

3. Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (c), Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations. *Computers and Chemical Engineering*, 74, pp. 75–88.

**Contribution:** I designed the study. I performed the calculations and wrote the manuscript with suggestions from Prof. Ramjugernath and Prof Rarey.

4. Moodley, K., Rarey, J., Ramjugernath, D., (2016) (d), An Extended UNISAC model for the Prediction of Solubility of Complex Pharmaceutical Ingredients in Non-electrolyte Pure Solvents and Solvent Mixtures. *Manuscript in preparation*.

**Contribution:** I designed the study with suggestions from Prof. Ramjugernath and Prof Rarey. I performed the calculations and wrote the manuscript with suggestions from Prof. Ramjugernath and Prof Rarey.

I performed the calculations and wrote the manuscript with suggestions from Prof. Ramjugernath and Prof Rarey.

5. Moodley, K., Rarey, J., Ramjugernath, D., (2016) (e), Experimental solubility for betulin and estrone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2)$

K. *Journal of Chemical Thermodynamics*, 98, pp. 42–50.

**Contribution:** I designed the study and performed all experimental measurements. I performed the calculations and wrote the manuscript with suggestions from Prof. Ramjugernath and Prof Rarey.

6. Moodley, K., Rarey, J., Ramjugernath, D., (2016) (f), Experimental solubility for diosgenin and estriol in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . *Manuscript submitted for publication*.

**Contribution:** I designed the study and performed all experimental measurements. I

7. Moodley, K., Rarey, J., Ramjugernath, D., (2016) (g), Experimental solubility for prednisolone and hydrocortisone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . *Manuscript submitted for publication*.

**Contribution:** I designed the study and performed all experimental measurements. I performed the calculations and wrote the manuscript with suggestions from Prof. Ramjugernath and Prof Rarey.

## ACKNOWLEDGEMENTS

The author would like to acknowledge the following people:

- The author's supervisors Professor D. Ramjugernath and Professor J. Rarey for their guidance and support during this research.
- The National Research Foundation through the Research and Innovation, Support and Advancement (RISA) program and the DST/NRF South African Research Chair in Fluorine Process Engineering and Separation for financial support for this project.
- The revisions and comments of the anonymous journal peer-reviewers and the editors that were incorporated into improving the manuscripts presented here.
- Prof. P. Naidoo and Dr C. Narasigadu for their invaluable mentorship and support.
- Mr Eugene Olivier and Dr Brian Satola for comprising the un-official personal tech-support team.
- Colleagues from the Thermodynamics Research Unit at the University of KwaZulu-Natal, as well as the thermodynamics research team at the Carl Von Ossietzky University of Oldenburg for their ideas, assistance and encouragement.
- The author's parents and siblings for a lifetime of support and motivation.
- And Deneb, the most luminous, in the darkest of days.

## TABLE OF CONTENTS

ABSTRACT .....	i
DECLARATION 1: Statement of Original Work .....	iv
DECLARATION 2: Contribution to publications.....	v
ACKNOWLEDGEMENTS.....	vii
TABLE OF CONTENTS .....	viii
LIST OF FIGURES .....	xi
LIST OF TABLES.....	xxii
NOMENCLATURE .....	xxv
CHAPTER ONE .....	1
<u>Introduction and Background .....</u>	<u>1</u>
1.1. Outline and rationale.....	1
1.2 Thermodynamics of phase equilibrium .....	6
1.3 Predictive Methods for the calculation of solubility or activity coefficient and their application.....	14
1.4 Development of the Universal Segment Activity Coefficient (UNISAC) model .....	23
1.5 Global optimization .....	27
1.6 Experimental Equipment and Procedure for Solubility Measurements.....	28
1.6.1 Experimental Techniques for Solubility Measurements.....	28
1.6.2 Experimental Apparatus.....	33
References.....	36
CHAPTER TWO .....	41
<u>Model Evaluation for the Prediction of Solubility of Active Pharmaceutical Ingredients (APIs).....</u>	<u>41</u>
Abstract.....	41
2.1 Introduction.....	42
2.2 Theory .....	44
2.3 Experimental solubility and pure component property data .....	44
2.4 Results and Discussion .....	66
2.5 Conclusion .....	72
References.....	73

CHAPTER THREE .....	77
<u>A Universal Segment Approach for the Prediction of the Activity Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents</u> .....	77
Abstract .....	77
3.1 Introduction.....	78
3.2 Theory .....	80
3.3 Universal segment approach to activity coefficient modelling.....	82
3.4 Results and Discussion .....	91
3.5 Conclusion .....	106
References.....	108
CHAPTER FOUR .....	113
<u>Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations</u> .....	113
Abstract .....	113
4.1 Introduction.....	113
4.2 Methods: Krill Herd Algorithm .....	116
4.3 Theory: Optimization problems associated with phase stability and equilibrium.....	120
4.4 Results and discussion .....	129
4.5 Conclusion .....	147
References.....	148
CHAPTER FIVE .....	153
<u>An Extended UNISAC model for the Prediction of Solubility of Complex Pharmaceutical Ingredients in Non-electrolyte Pure Solvents and Solvent Mixtures</u> .....	153
Abstract .....	153
5.1 Introduction.....	154
5.2 Theory .....	158
5.3 Error Estimation.....	168
5.4 Results and discussion .....	171
5.5 Conclusion .....	180
References.....	181
CHAPTER SIX.....	184

<u>Experimental Solubility Data For Betulin/ Estrone/ Diosgenin/ Estriol/ Prednisolone/ Hydrocortisone In Various Solvents In The Temperature Range T = (293.2 To 328.2) K.....</u>	184
Abstract.....	184
6.1 Introduction.....	184
6.2 Theory.....	186
6.3 Experimental methods.....	189
6.4 Results and discussion.....	193
6.5 Conclusions.....	226
References.....	227
CHAPTER SEVEN.....	230
<u>Culminating Discussion.....</u>	230
References.....	248
CHAPTER EIGHT.....	251
<u>Conclusions.....</u>	251
CHAPTER NINE.....	252
<u>Future Work and Recommendations.....</u>	252
References.....	257
Appendix A.....	258
<u>Supplementary data for Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations.....</u>	258
A1. Derivations.....	258
Appendix B.....	261
<u>Extended UNISAC: Supplementary data.....</u>	261
B1. Example Solid-liquid Equilibrium example system calculations and plots for binary and ternary systems:.....	261
B2. Components and data sets.....	273
Appendix C.....	277
<u>Supplementary data: Experimental Solubility Measurement.....</u>	277
C1. Calibration.....	277

## LIST OF FIGURES

Figure 1.1: Schematic showing possible hydrogen bonding sites between functional groups of 2 components. ....	24
Figure 1.2: Schematic showing the assumption of identical hydrogen bonding sites between functional groups comprising any 2 components .....	24
Figure 1.3: Schematic showing functional group surface fraction exhibiting hydrogen bonding. ....	25
Figure 1.4: Schematic showing fractional contributions to hydrogen bonding from different functional groups. ....	26
Figure 1.5 A schematic of a typical differential thermal analyser. ....	30
Furnace chamber (1), measurement sample pan (2), reference sample pan (3), heating element (4), temperature difference recorder using temperature probes (5), heat input/ temperature programmer (6). ....	30
Figure 1.6 A combined TG-DTA plot for the system NaCl-Water adapted from (Endoh and Suga, 1999); (—) TG, (···) DTA.....	31
Figure 1.7 A schematic of the furnace and balance chamber of the Shimadzu DTG-60AH used in this work. Process gas inlet (1), cleaning gas inlet (2), line to vacuum pump (3), inert gas inlet (4), balance chamber (5), Roberval balance mechanism (6), heating chamber (7). ....	34
Figure 1.8 Experimental Setup. A- Grant GD120 temperature controller; B- Sonication bath; C- Glass equilibrium vessel; D- Gas tight sampling syringe; E- Shimadzu DTG-60AH apparatus; F- Shimadzu Auto-sampler.....	35
Figure 2.1 Comparison of the natural logarithms of experimental and model calculated solubility composition ( $x_1$ ). $\diamond$ , UNIFAC, $\Delta$ , modified UNIFAC (Dortmund), $\circ$ , COSMO-RS(OL), $\square$ , COSMOSAC. ....	67
Figure 2.2 Correlation of model percentage deviations with molecular mass of solute. $\diamond$ , UNIFAC, $\Delta$ , modified UNIFAC (Dortmund), $\circ$ , COSMO-RS(OL), $\square$ , COSMOSAC. ....	68
Figure 2.3 Correlation of model percentage deviations with van der Waals area parameter ( $q_1$ ). $\diamond$ , UNIFAC, $\Delta$ , modified UNIFAC (Dortmund), $\circ$ , COSMO-RS(OL), $\square$ , COSMOSAC. ....	68
.....	69
Figure 2.4 Correlation of model percentage deviations with number of different functional groups present in solute. $\diamond$ , UNIFAC, $\Delta$ , modified UNIFAC (Dortmund), $\circ$ , COSMO-RS(OL), $\square$ , COSMOSAC. ....	69



Figure 2.5 Comparison of the natural logarithms of experimental and model calculated solubility composition ( $x_1$ ) with the NRTL-SAC model. ....	72
Figure 3.1 Principal Component Analysis (PCA) to show the heterogeneity of the selected pharmaceutical components for model testing. (a) Principal Component F1 and F2, (b) Principal Component F1 and F3. ....	98
Figure 3.2 (a) Comparison of experimental solubility data for benzoic acid to predictive models. ● – UNISAC, × - NRTL-SAC, Δ – UNIFAC, (b) Comparison of experimental solubility data for various solutes in water, to predictive models. ● – UNISAC, Δ – UNIFAC. ....	103
Figure 3.3 (a) Akaike Information Criterion Score for the solubility prediction of the models tested. ■ – UNISAC, □ – NRTL-SAC, ▣ – UNIFAC, (b) Focused Information Criterion Score for the solubility prediction of the models tested. ■ – UNISAC, □ – NRTL-SAC, ▣ – UNIFAC. ....	104
Figure 3.4 (a) Plot of Experimental and Predicted Solubility ( $x^s$ ) from UNISAC of Methylparaben in Methanol (■,—), 1-Propanol (×,— - -) and 1-Butanol (◆,—), at various temperatures, (b) Plot of Experimental vs. Predicted Solubility ( $x_s$ ) from UNISAC of Naproxen in Acetone (●, - - -), Methanol (◆,— - -), Ethanol (■,—) and 2-Propanol (×,—), at various temperatures. Experimental data are represented as symbols, and model predictions are represented as lines. ....	105
Figure 4.1 Pseudo code for (a) Krill Herd Algorithm and (b) Lévy Flight Krill Herd Algorithm, implemented in this work. ....	119
Figure 4.2 Global Success Rates of phase stability problems with increasing herd size, ■KH, ■LKH. ....	132
Figure 4.3 A convergence profile of PS-1 with a herd size of 20, ■KH, ◆LKH. ....	133
Figure 4.4 A convergence profile of PS-3 with a herd size of 20, ■KH, ◆LKH. ....	133
Figure 4.5 A convergence profile of PS-7 with a herd size of 20, ■KH, ◆LKH. ....	134
Figure 4.6 Global Success Rates of phase equilibrium problems with increasing herd size, ■KH, ■LKH. ....	136
Figure 4.7 A convergence profile of PE-2 with a herd size of 50, ■KH, ◆LKH. ....	137
Figure 4.8 A convergence profile of PE-5 with a herd size of 50, ■KH, ◆LKH. ....	137
Figure 4.9 A convergence profile of PE-8 with a herd size of 50, ■KH, ◆LKH. ....	138
Figure 4.10 Global Success Rates of reactive phase equilibrium problems with increasing herd size, ■KH, ....	141

Figure 4.11 A convergence profile of rPE-1 with a herd size of 50, ■KH, ◆LKH.....	141
Figure 4.12 A convergence profile of rPE-3 with a herd size of 50, ■KH, ◆LKH.....	142
Figure 4.13 A convergence profile of rPE-7 with a herd size of 50, ■KH, ◆LKH.....	142
Figure 5.1 Change in the influence of temperature on the experimental liquid-phase composition in SLE.....	170
Figure 5.2 Calculation of unbiased predicted solubility.....	171
Figure 5.3 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in alcohol solvents. ....	175
Figure 5.4 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in alkane solvents. ....	175
Figure 5.5 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in aromatic solvents. ....	176
Figure 5.6 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in ether solvents. ....	176
Figure 5.7 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in ketone solvents. ....	177
Figure 5.8 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in water as a solvent. ....	177
Figure 5.9 Fraction of the data with deviations in composition larger than a given composition for various binary non-aqueous solute-solvent systems, by different models. - · - ·, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ···, UNIFAC; —, COSMO-SAC; - · - ·, COSMO-RS (OL).....	178
Figure 5.10 Fraction of the data with deviations in temperature larger than a given temperature for various binary non-aqueous solute-solvent systems, by different models. - · - ·, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ···, UNIFAC; —, COSMO-SAC; - · - ·, COSMO-RS (OL).....	178

Figure 5.11 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition for various ternary solubility systems.....	179
Figure 6.1 A generic combined TG-DTA plot; (—) TG, (---) DTA.....	192
Figure 6.2 T-x plot for the systems betulin (1) + pentan-1-ol (2), ●, This work, ○, (Cao et al., (2006)), systems estrone (1) + water (2), ▲, This work, Δ, (Domańska et al., (2010)), ×, (Shareef et al., (2006)), estrone (1) + octan-1-ol (2), ◆, This work, ◇, (Domańska et al., (2010)), diosgenin (1) + pentan-1-ol (2), ■, This work, □, (Chen et al., (2012)).	194
Figure 6.3 T-x plot for the systems betulin + (experimental, correlation, model), (+, ---, T-K-Wilson), acetonitrile/ (x, —, T-K-Wilson), nonan-1-ol/ (○, -.-.-, NRTL), octan-1-ol/ (◇, - · -, T-K-Wilson), pentan-1-ol/ (■, - - -, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.....	210
Figure 6.4 Dilute range T-x plot for the systems betulin + (experimental, correlation, model), (+, ---, T-K-Wilson), acetonitrile/ (x, —, T-K-Wilson), nonan-1-ol/ (○, -.-.-, NRTL), octan-1-ol/ (◇, - · -, T-K-Wilson), pentan-1-ol/ (■, - - -, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.....	210
Figure 6.5 T-x plot for the systems betulin + (experimental, correlation, model), (▲, ---, NRTL), N,N dimethylformamide/ (■, —, NRTL), 1-methyl-2-pyrrolidone/ (x, - · -, NRTL), n-hexadecane/ (○, - - -, T-K-Wilson), n-dodecane/ (+, -.-.-, T-K-Wilson), water. Experimental data are represented as symbols, and model predictions are represented as lines. ....	211
Figure 6.6 Dilute region T-x plot for the systems betulin + (experimental, correlation, model), (▲, ---, NRTL), N,N dimethylformamide/ (■, —, NRTL), 1-methyl-2-pyrrolidone/ (x, - · -, NRTL), N-Hexadecane/ (○, - - -, T-K-Wilson), n-dodecane/ (+, -.-.-, T-K-Wilson), water. Experimental data are represented as symbols, and model predictions are represented as lines. ....	211
Figure 6.7 T-x plot for the systems estrone + (experimental, correlation, model), (+, ---, T-K-Wilson), 1-methyl-2-pyrrolidone/ (▲, - · -, T-K-Wilson), 2-aminoethanol/ (x, —, T-K-Wilson), nonan-1-ol/ (○, -.-.-, T-K-Wilson), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/ (■, - - -, NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines. ....	212
Figure 6.8 Dilute region T-x plot for the systems estrone + (experimental, correlation, model), (+, ---, T-K-Wilson), 1-methyl-2-pyrrolidone/ (▲, - · -, T-K-Wilson), 2-aminoethanol/ (x, —, T-K-Wilson), nonan-1-ol/ (○, -.-.-, T-K-Wilson), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/	

(■, - - -, NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.....212

Figure 6.9 T-x plot for the systems estrone + (experimental, correlation, model), (▲, ···, NRTL), N,N dimethylformamide/ (■, —, T-K-Wilson), n-octane/ (x, - · -, T-K-Wilson ), n-hexadecane/ (○, — — —, NRTL), n-dodecane/ (+, - - -, NRTL), water/ (●, - · · -, NRTL), morpholine-4-carbaldehyde/(◇, -.-., NRTL), acetonitrile. Experimental data are represented as symbols, and model predictions are represented as lines. ....213

Figure 6.10 Dilute region T-x plot for the systems estrone + (experimental, correlation, model), (▲, ···, NRTL), N,N dimethylformamide/ (■, —, T-K-Wilson), n-octane/ (x, - · -, T-K-Wilson ), N-Hexadecane/ (○, — — —, NRTL), n-dodecane/ (+, - - -, NRTL), water/ (●, - · · -, NRTL), morpholine-4-carbaldehyde/(◇, -.-., NRTL), acetonitrile. Experimental data are represented as symbols, and model predictions are represented as lines.....213

Figure 6.11 T-x plot for the systems diosgenin (1) + (experimental, NRTL-model), /(○, ···), octan-1-ol (2) / (Δ, -), 2-aminoethanol (2) / (◇, - - -), pentan-1-ol (2) / (+, - - -), butan-2-ol (2) / (□, - · -), nonan-1-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines. ....214

Figure 6.12 Dilute region T-x plot for the systems diosgenin (1) + (experimental, NRTL-model), /(○, ···), octan-1-ol (2) / (Δ, -), 2-aminoethanol (2) / (◇, - - -), pentan-1-ol (2) / (+, - - -), butan-2-ol (2) / (□, - · -), nonan-1-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.....214

Figure 6.13 T-x plot for the systems diosgenin (1) + (+, - - -), acetonitrile (2) / (■, -.-.), morpholine-4-carbaldehyde (2) / (Δ, - · -), n-dodecane (2) / (□, -), water/ (○, ...), dimethylformamide (2), / (▲, - · · -), 1-methyl-2-pyrrolidone (2) / (x, -.-.), N-Hexadecane (2) / (◇, ···), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines. ....215

Figure 6.14 T-x plot for the systems diosgenin (1) + (+, - - -), acetonitrile (2) / (■, -.-.) morpholine-4-carbaldehyde (2) / (Δ, - · -), n-dodecane (2) / (□, -), water/ (○, ...), dimethylformamide (2), / (▲, - · · -), 1-methyl-2-pyrrolidone (2) / (x, -.-.), N-Hexadecane (2) / (◇, ···), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines. ....215

Figure 6.15 T-x plot for the systems estriol (1) + (experimental, NRTL-model), (■, - · · -) morpholine-4-carbaldehyde (2) / (○, ···), octan-1-ol (2) / (▲, -), 2-aminoethanol (2) / (◇, - · -), pentan-1-ol (2), (+, - - -), nonan-1-ol (2) / (x, - · -), butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.....216

Figure 6.16 Dilute region T-x plot for the systems estriol (1) + (experimental, NRTL-model), (■, - · -) morpholine-4-carbaldehyde (2) / (○, ···), octan-1-ol (2) / (▲, -), 2-aminoethanol (2) / (◇, ···), pentan-1-ol (2), (+, - - -), nonan-1-ol (2) / (x, - · -), butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.....216

Figure 6.17 T-x plot for the systems estriol (1) + (experimental, NRTL-model), (Δ, -.-) n-dodecane (2) / (●, - - -), dimethylformamide (2) / (+, - · -), water / (▲, - · -), 1-methyl-2-pyrrolidone (2) / (x, ---), n-hexadecane (2) / (◇, ···), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines.....217

Figure 6.18 Dilute region T-x plot for the systems estriol (1) + (experimental, NRTL-model), (Δ, -.-) n-dodecane (2) / (●, - - -), dimethylformamide (2) / (+, - · -), water / (▲, - · -), 1-methyl-2-pyrrolidone (2) / (x, ---), n-hexadecane (2) / (◇, ···), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines.....217

Figure 6.19 T-x plot for the systems prednisolone + (experimental, model, correlation), (+, ···, T-K-Wilson), acetonitrile/ (x, -, NRTL), nonan-1-ol/ (○, -.-, NRTL), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/ (□, - - -, NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines. ....218

Figure 6.20 Dilute region T-x plot for the systems prednisolone + (experimental, model, correlation), (+, ···, T-K-Wilson), acetonitrile/ (x, -, NRTL), nonan-1-ol/ (○, -.-, NRTL), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/ (□, - - -, NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.....218

Figure 6.21 T-x plot for the systems prednisolone + (experimental, model), (Δ, ···, T-K-Wilson), N,N dimethylformamide/ (□, -, T-K-Wilson), 1-methyl-2-pyrrolidone/ (x, - · -, NRTL), n-hexadecane/ (○, - - -, NRTL), n-dodecane/ (+, -.-, NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines.....219

Figure 6.22 Dilute region T-x plot for the systems prednisolone + (experimental, model), (Δ, ···, T-K-Wilson), N,N dimethylformamide/ (□, -, T-K-Wilson), 1-methyl-2-pyrrolidone/ (x, - · -, NRTL), n-hexadecane/ (○, - - -, NRTL), n-dodecane/ (+, -.-, NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines. ....219

Figure 6.23 T-x plot for the systems hydrocortisone + (experimental, model), (+, ···, T-K-Wilson), acetonitrile/ (x, -, NRTL), nonan-1-ol/ (○, -.-, T-K-Wilson), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/ (□, - - -, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines. ....220

Figure 6.24 Dilute region T-x plot for the systems hydrocortisone + (experimental, model), (+, ..., T-K-Wilson), acetonitrile/ (×, -, NRTL), nonan-1-ol/ (○, -.-, T-K-Wilson), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/ (□, - - -, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.....220

Figure 6.25 T-x plot for the systems hydrocortisone + (experimental, model), (Δ, ..., T-K-Wilson), N,N dimethylformamide/ (□, -, NRTL), 1-methyl-2-pyrrolidone/ (×, - · -, NRTL), N-Hexadecane/ (○, - - -, NRTL), n-dodecane/ (+, -.-, NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines.....221

Figure 6.26 Dilute region T-x plot for the systems hydrocortisone + (experimental, model), (Δ, ..., T-K-Wilson), N,N dimethylformamide/ (□, -, NRTL), 1-methyl-2-pyrrolidone/ (×, - · -, NRTL), N-Hexadecane/ (○, - - -, NRTL), n-dodecane/ (+, -.-, NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines. ....221

Figure 7.1 Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models for the system solute (1) + benzene (2). ×, Experimental Data (McCloughlin and Zainal, 1959); - · - ·, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-SAC; - · - ·, COSMO-RS (OL). Solute:(a) biphenyl, (b) anthracene, (c) phenanthrene, (d) 1,2-diphenylbenzene, (e) bicyclo[4.4.0]deca-1,3,5,7,9-pentene, (f) 1,3,5-triphenylbenzene. Experimental data are represented as symbols, and model predictions are represented as lines. ....232

Figure 7.2 Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models for the system solute (1) + benzene (2). ×, Experimental Data (McCloughlin and Zainal, 1959); — · —, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-SAC; — · —, COSMO-RS (OL). Solute: (a) pyrene, (b) 1,3-diphenylbenzene, (c) fluoranthene, (d) 1,2-dihydroacenaphthylene, (e) 9H-fluorene, (f) chrysene. Experimental data are represented as symbols, and model predictions are represented as lines.....233

Figure 7.3 Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models for the system solute (1) + benzene (2). ×, Experimental Data (McCloughlin and Zainal, 1959); — · —, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-SAC; — · —, COSMO-RS (OL). Solute: (a) triphenylene, (b) 1,4-diphenylbenzene. Experimental data are represented as symbols, and model predictions are represented as lines. ....234

Figure 7.4 Experimental and predicted isobaric/isothermal vapour-liquid equilibrium data by different models. ×, Experimental Data; — · —, Extended UNISAC; - - -, mod. UNIFAC

(Dortmund); ..., UNIFAC. Systems:(a) n-hexane (1) + butan-1-ol (2), 332.53 K, (Berro et al., 1982); (b) water (1) + butan-1-ol (2), 323.23 K, (Fischer and Gmehling, 1994); (c) acetonitrile (1) + butan-1-ol (2), 101.3 kPa, (Kovac et al., 1985) ; (d) acetone (1) + butan-1-ol (2), 99.45 kPa, (Michalski et al., 1961); (e) tetrachloromethane (1) + butan-1-ol (2), 99.33 kPa, (Doniec et al., 1965); (f) chloroform (1) + butan-1-ol (2), 303 kPa, (Chen et al., 1995). Experimental data are represented as symbols, and model predictions are represented as lines. ....236

Figure 7.5 Experimental and predicted isobaric/isothermal vapour-liquid equilibrium data by different models. ×, Experimental Data; —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC. Systems:(a) butanone (1) + butan-1-ol (2), 323.15 K, (Garriga et al., 1996); (b) dibutyl ether (1) + butan-1-ol (2) 101.3 kPa, (Lladosa et al., 2006); (c) benzene (1) + butanone (2), 313.15 K, (Van Nhu and Kholer, 1989) ; (d) benzene (1) + butan-1-ol (2), 303 kPa, (Chen et al., 1995); (e) benzene (1) +dibutyl ether (2), 308.15 K, (Ott et al., 1981); (f) heptanone (1) + dibutyl ether (2), 393.22 K, (Wu and Sandler, 1988). Experimental data are represented as symbols, and model predictions are represented as lines.....237

Figure 7.6 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (e)); —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL) ; —·—, COSMO-SAC. Systems: (a) betulin (1) + nonan-1-ol (2); (b) betulin (1) + octan-1-ol (2); (c) betulin (1) + pentan-1-ol (2); (d) betulin (1) + butan-2-ol; (e) betulin (1) + n-hexadecane (2); (f) betulin (1) + n-dodecane (2).....240

Figure 7.7 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (e, f)); —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) betulin (1) + acetonitrile (2); (b) betulin (1) + water (2); (c) estriol (1) + nonan-1-ol (2); (d) estriol (1) + octan-1-ol (2); (e) estriol (1) + pentan-1-ol (2); (f) estriol (1) + butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.....241

Figure 7.8 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (e, f)); —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) estriol (1) + n-octane (2); (b) estriol (1) + n-dodecane (2); (c) estriol (1) + n-hexadecane (2); (d) estrone (1) + water (2); (e) estrone (1) + nonan-1-ol (2); (f) estrone (1) + octan-1-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.....242

Figure 7.9 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (e)); —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) estrone (1) + pentan-1-ol (2); (b) estrone (1) + butan-2-ol (2); (c) estrone (1) + n-hexadecane (2); (d) estrone (1) + n-dodecane (2); (e) estrone (1) + n-octane (2); (f) estrone (1) + water (2). Experimental data are represented as symbols, and model predictions are represented as lines.....243

Figure 7.10 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (e, f)); —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) estrone (1) + acetonitrile (2); (b) diosgenin (1) + acetonitrile (2); (c) diosgenin (1) + nonan-1-ol (2); (d) diosgenin (1) + octan-1-ol (2); (e) diosgenin (1) + pentan-1-ol (2); (f) diosgenin (1) + butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.....244

Figure 7.11 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (f, g)); —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) diosgenin (1) + n-hexadecane (2); (b) diosgenin (1) + n-dodecane (2); (c) diosgenin (1) + n-octane (2); (d) diosgenin (1) + water (2); (e) prednisolone (1) + n-dodecane (2); (f) prednisolone (1) + n-hexadecane (2). Experimental data are represented as symbols, and model predictions are represented as lines.....245

Figure 7.12 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (g)); —·—, Extended UNISAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) prednisolone (1) + nonan-1-ol (2); (b) prednisolone (1) + octan-1-ol (2); (c) prednisolone (1) + pentan-1-ol (2); (d) prednisolone (1) + butan-2-ol (2); (e) prednisolone (1) + acetonitrile (2); (f) hydrocortisone (1) + acetonitrile (2). Experimental data are represented as symbols, and model predictions are represented as lines.....246

Figure 7.13 Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (g)); —·—, Extended UNISAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) hydrocortisone (1) + nonan-1-ol (2); (b) hydrocortisone (1) + octan-1-ol (2); (c) hydrocortisone (1) + pentan-1-ol (2); (d) hydrocortisone (1) + butan-2-ol (2); (e) hydrocortisone (1) + n-dodecane (2); (f)



hydrocortisone (1) + n-hexadecane (2). Experimental data are represented as symbols, and model predictions are represented as lines.....	247
Figure 9.1 Natural log of the activity coefficient for mestanolone-alkane systems at 298 K. ○ – Experimental data (Gharavi et al, 1983), Extended UNISAC residual with combinatorial expression of; —, Flory-Huggins with Staverman-Guggenheim correction, ∙∙∙, Flory-Huggins, - - - Moller et al. (2014). Experimental data are represented as symbols, and model predictions are represented as lines. ....	255
Figure 9.2 Natural log of the activity coefficient for testosterone-alkane systems at 298 K. ○ – Experimental data (Gharavi et al, 1983), Extended UNISAC residual with combinatorial expression of; —, Flory-Huggins with Staverman-Guggenheim correction, ∙∙∙, Flory-Huggins, - - - Moller et al. (2014). Experimental data are represented as symbols, and model predictions are represented as lines. ....	255
Figure B1. Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models. ×, Experimental Data (Dortmund Data Bank, 2012); ∙∙∙, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ∙∙∙, UNIFAC; —, COSMO-SAC; ∙∙∙, COSMO-RS (OL). Systems: (a) 2,6-dimethyl naphthalene (1) + butan-1-ol (2); (b) tetradecane (1) + benzene (2); (c) pentacosane (1) + methyl tert- ether (2);(d) 1,4-dioxane (1) + cyclohexane (2). Experimental data are represented as symbols, and model predictions are represented as lines. ....	269
Figure B2. Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models. ×, Experimental Data (Dortmund Data Bank, 2012); ∙∙∙, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ∙∙∙, UNIFAC; —, COSMO-SAC; ∙∙∙, COSMO-RS (OL). Systems: (a) naphthalene (1) + butan-2-one (2); (b) phenanthrene (1) + water (2); (c) phenol (1) + water (2);(d) p-dichlorobenzene (1) + water (2). Experimental data are represented as symbols, and model predictions are represented as lines.....	270
Figure B3. Experimental and predicted solid-liquid equilibrium temperatures for the system 1-acetyl-2-naphthol (1) + cyclohexane (2) + ethanol. ○, Experimental Data (Dortmund Data Bank, 2012); surface, Extended UNISAC. Experimental data are represented as symbols, and model predictions are represented as a surface.....	271
Figure B4. Experimental and predicted solid-liquid equilibrium temperatures for the system naphthalene (1) + benzene (2) + ethanol. ○, Experimental Data (Dortmund Data Bank, 2012); surface, Extended UNISAC. Experimental data are represented as symbols, and model predictions are represented as a surface.....	272

Figure C1. Calibration curve of equilibrium cell bath temperature probe. **Error! Bookmark not defined.**

Figure C2. Plot of deviations of measured temperature form actual temperature. .... **Error! Bookmark not defined.**

Figure C3.  $\ln(x_1)$  vs  $1/T$  plot for the systems betulin + (experimental, 2<sup>nd</sup> order polynomial), (+, ---), acetonitrile/ (x, —) , nonan-1-ol/ (o, - - -), octan-1-ol/ ( $\diamond$ , - · -), pentan-1-ol/ ( $\blacksquare$ , - - -), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines..... 278

Figure C4. (a) Plot of  $\ln(x_1^{\text{exp}})$  vs  $\ln(x_1^{\text{calc}})$  for the solubility systems measured. (b) Scatter plot of  $((x_1^{\text{exp}} - x_1^{\text{calc}})/x_1^{\text{exp}})$  vs  $x_1^{\text{exp}}$  for the solubility systems measured. +, betulin, ×, prednisolone,  $\blacklozenge$ , hydrocortisone,  $\bullet$ , diosgenin,  $\Delta$  , estriol,  $\square$ , estrone.....278

## LIST OF TABLES

Table 2.1 Physical properties of the solutes used in this study.....	46
Table 2.2 Experimental and calculated solubility for various APIs. ....	50
Table 2.3 Mean Percentage Deviations of various solutes in benzene and water. ....	70
Table 2.4 Calculated segment area parameters for NRTL-SAC.....	71
Table 3.1 Structure and Classification of selected base segments BMRR-1 .....	87
Table 3.2 Binary interaction parameters for base segments B-MRR1 for non-aqueous solutions. .....	92
Table 3.3 Binary interaction parameters for base segments B-MRR1 for aqueous solutions.	92
Table 3.4 Segment area parameters for base segments B-MRR1 and non- base segments NB- MRR1.....	93
Table 3.5 Physical Properties of Solutes Used in this Work. ....	95
Table 3.6 Comparison between experimental and predicted values of the three models tested .....	100
Table 3.7 Comparison between experimental and predicted values of Aspirin solubility at 298.15 K by various models. ....	102
Table 3.8 Comparison of the performance of the UNISAC model to the modified UNIFAC (Dortmund) and Pharma-UNIFAC model from literature (Diedrichs and Gmehling, 2011), for alkanes, alcohol and water as a solvent. ....	106
Table 4.1 Benchmark test cases for phase stability (PS) and phase equilibria (PE) problems with literature sources .....	125
Table 4.2 Benchmark test cases for reactive phase equilibria (rPE) problems with literature sources.....	128
Table 4.3 Summary of fixed and variable parameters for the three types of optimization problems.....	129
Table 4.4 Performance of the Krill Herd and Levy Flight Krill Herd Algorithms for problems involving phase stability. ....	135
Table 4.5 Performance of the Krill Herd and Levy Flight Krill Herd Algorithms for problems involving phase equilibrium. ....	139

Table 4.6 Performance of the Krill Herd and Levy Flight Krill Herd Algorithms for problems involving reactive phase equilibrium.....	143
Table 4.7 Comparison of the performances of the KH and LKH with other methods in the literature.....	146
Table 5.1 Binary interaction parameters for base segments (B-MRR1) the Extended UNISAC model.....	162
Table 5.2 Moodley-Rarey-Ramjugernath Group Fragmentation Scheme with Segment area parameters.....	163
Table 5.3 Relative Akaike Information Criterion scores and percentage deviations.....	179
Table 6.1 Chemical suppliers and purities.....	190
Table 6.2 Experimental melting point and enthalpy of fusion data of solutes used at 0.101 MPa <sup>a</sup> . .....	195
Table 6.3 Experimental solid-liquid equilibrium data of betulin mixtures at various temperatures and 0.101 MPa <sup>a</sup> . ....	196
Table 6.4 Experimental solid-liquid equilibrium data of estrone mixtures at various temperatures and 0.101 MPa <sup>a</sup> . ....	197
Table 6.5 Experimental solid-liquid equilibrium data of diosgenin mixtures at various temperatures and 0.101 MPa <sup>a</sup> . ....	198
Table 6.6 Experimental solid-liquid equilibrium data of estriol mixtures at various temperatures and 0.101 MPa <sup>a</sup> . ....	199
Table 6.7 Experimental solid-liquid equilibrium data of prednisolone mixtures at various temperatures and 0.101 MPa <sup>a</sup> . ....	200
Table 6.8 Experimental solid-liquid equilibrium data of hydrocortisone mixtures at various temperatures and 0.101 MPa <sup>a</sup> . ....	201
Table 6.9 Regressed model parameters for the NRTL and T-K-Wilson models with betulin as a solute for use in equations (6.4), (6.5), (6.8) and (6.9). ....	204
Table 6.10 Regressed model parameters for the NRTL and T-K-Wilson models with estrone as a solute for use in equations (6.4), (6.5), (6.8) and (6.9). ....	205
Table 6.11 Regressed model parameters for the NRTL and T-K-Wilson models with diosgenin as a solute for use in equations (6.4), (6.5), (6.8) and (6.9). ....	206
Table 6.12 Regressed model parameters for the NRTL and T-K-Wilson models with estriol as a solute for use in equations (6.4), (6.5), (6.8) and (6.9). ....	207
Table 6.13 Regressed model parameters for the NRTL and T-K-Wilson models with prednisolone as a solute for use in equations (6.4), (6.5), (6.8) and (6.9). ....	208

Table 6.14 Regressed model parameters for the NRTL and T-K-Wilson models with hydrocortisone as a solute for use in equations (6.4), (6.5), (6.8) and (6.9).....	209
Table 6.15 Estimated infinite dilution activity coefficients of solute (1) in solvent (2) at 298.2 K 0.101 MPa from correlated experimental data. ....	222
Table 6.15 Estimated infinite dilution activity coefficients of solute (1) in solvent (2) at 298.2 K 0.101 MPa from correlated experimental data (continued).....	223
Table 6.15 Estimated infinite dilution activity coefficients of solute (1) in solvent (2) at 298.2 K 0.101 MPa from correlated experimental data (continued).....	224
Table 6.16 Experimental Solubility ranking (highest solubility in solvent to lowest) at 328.2 K and 0.101 MPa. ....	225
Table B1. Sample activity coefficient predictions by the Extended UNISAC model to be used for verification. ....	268

## NOMENCLATURE

<b>Symbols</b>	<b>Description</b>
$a_i$	Activity of component $i$
$A_{ij}$	Binary interaction parameter for the NRTL model
$a_{mn}$	Binary interaction parameter between group/segments $m$ and $n$
$b$	Bias for the Focused Information Criterion
$B_{ij}$	Binary interaction parameter for the NRTL model
$c$	Total number of components in a mixture
$C_{best}$	Effective coefficient of best krill
$C_{food}$	Food attraction coefficient
$c_i$	Cohesive energy density
$C_{ij}$	Binary interaction parameter for the T-K-Wilson model
$C_p$	Heat capacity at constant pressure ( $J. mol^{-1}$ )
$D_i$	Random Diffusion variable of $i$ th individual
$D_{ij}$	Binary interaction parameter for the T-K-Wilson model
$f_i$	Fugacity of component $i$ (kPa)
$F_i$	Foraging variable of $i$ th individual
$G$	Molar Gibbs free energy ( $J.mol^{-1}$ )
$G_{ij}$	NRTL model parameter
$g_{ij} - g_{ii}$	NRTL model fit parameter ( $J.mol^{-1}$ )
GSR	Global Success Rate
$H$	Krill generation
$H$	Molar enthalpy ( $J.mol^{-1}$ )
$I$	Total number of components for the Focused Information Criterion
$k$	Number of model parameters used in the Akaike Information Criterion
$k^{max}$	Maximum number of allowable iterations for the calculation
$K_{eq}$	Reaction equilibrium constant
$\mathbf{K}^{eq}$	Vector of reaction equilibrium constants
$K_i$	Value of objective function of $i$ th individual
$K_{best}$	Optimum value of the objective function $K$
$\hat{K}_{i,j}$	Relative fitness of the evaluated objective function, $K$ of the $i$ th and $j$ th individuals.
$\hat{K}_{i,ibest}$	Relative fitness of the current evaluated objective function, $K$ of the $i$ th individual with its previous best value
$K_{worst}$	Worst value of the objective function $K$
$L$	Likelihood function used in the Akaike Information Criterion

<i>LB</i>	<i>Abbreviation of lower parameter bound</i>
<i>m</i>	<i>Mean square deviation for the Focused Information Criterion</i>
<i>M</i>	<i>Generic thermodynamic property</i>
$\hat{M}_i$	<i>Property in solution</i>
$\bar{M}_i$	<i>Partial Property</i>
$\Delta_{fus}M$	<i>Property change of fusion</i>
$\Delta M^{mix}$	<i>Property change of mixing</i>
$\Delta_{rxn}M$	<i>Property change of reaction</i>
$\Delta_{vap}M$	<i>Property change of vaporization</i>
$M^E$	<i>Excess property</i>
<i>MM</i>	<i>Molar mass (g.mol<sup>-1</sup>)</i>
$n_i$	<i>Number of moles of component (moles)</i>
$n_i$	<i>Number of points used in the Akaike Information Criterion</i>
$n_{ij}$	<i>Number of moles of component i in phase j</i>
<b>N</b>	<i>Matrix of stoichiometric coefficients</i>
<i>NFE</i>	<i>Abbreviation of Number of function evaluations</i>
$N_i$	<i>Krill herd distribution</i>
<i>NN</i>	<i>Herd size</i>
<i>NP</i>	<i>Number of points for the Focused Information Criterion</i>
$\mathbf{n}_{ref}$	<i>Vector of the molar compositions in terms of the reference components</i>
$p(\sigma_m)$	<i>frequency of surface charge density of component i</i>
<i>P</i>	<i>Total Pressure (Pa)</i>
<i>PE</i>	<i>Abbreviation of phase equilibrium</i>
<i>PS</i>	<i>Abbreviation of phase Stability</i>
$q_i$	<i>Group/component area parameter</i>
$Q_i$	<i>van der Waals group area</i>
<i>R</i>	<i>Universal gas constant (8.314 J. mol<sup>-1</sup>. K<sup>-1</sup>)</i>
<i>rand</i>	<i>A randomly selected number in the range i to j.</i>
$\in [i, j]$	
$r_i$	<i>Group/component volume parameter</i>
$R_i$	<i>van der Waals group volume</i>
<i>rPE</i>	<i>Abbreviation of reactive phase equilibrium</i>
<i>RSS</i>	<i>Residual Sum of Squares used in the Akaike Information Criterion</i>
<i>S</i>	<i>Molar entropy (J.mol<sup>-1</sup>. K<sup>-1</sup>)</i>
<i>SC</i>	<i>Abbreviation of stopping condition</i>
<i>SR</i>	<i>Abbreviation of success rate</i>
<i>T</i>	<i>time</i>
$\Delta t$	<i>Scaling parameter of speed vector</i>
<i>T</i>	<i>Temperature ( K)</i>
<i>tol</i>	<i>Tolerance</i>
<i>TPDF</i>	<i>Tangent Plane Distance Function</i>

$U$	<i>Lévy flight distribution</i>
$U$	<i>Standard uncertainty</i>
$UB$	<i>Abbreviation of upper parameter bound</i>
$u_r$	<i>Standard relative uncertainty</i>
$V$	<i>Total volume of vapour (<math>m^3</math>)/ Volts</i>
$V_i$	<i>Molar Volume of component <math>i</math> (<math>m^3 \cdot mol^{-1}</math>)/UNIFAC, UNISAC, mod. UNIFAC volume fraction</i>
$x_i$	<i>Liquid phase mole fraction</i>
$X, Y, Z^+, Z$	<i>Segment area parameters for the NRTL-SAC model</i>
$X_i$	<i>Position of the <math>i</math>th krill individual</i>
$X_{best}$	<i>Optimal Krill position</i>
$\hat{X}_{i,j}$	<i>Relative attraction between krill individuals <math>i</math> and <math>j</math></i>
$\hat{X}_{i,ibest}$	<i>Relative position of the current individual, <math>X_i</math> to its previous best value</i>
$y_i$	<i>Experimental data point used in the Akaike Information Criterion</i>
$z_i$	<i>Overall composition/Solid phase mole fraction</i>
$Z$	<i>Coordination number</i>

**Greek letters**

$A$	<i>Alpha phase</i>
$\alpha_{12}$	<i>Non-randomness parameter for the NRTL model/ Relative volatility</i>
$B$	<i>Beta phase</i>
$\gamma_i$	<i>Activity coefficient of species <math>i</math></i>
$\Gamma_k$	<i>Contribution to activity coefficient of group/segment <math>k</math></i>
$\Delta$	<i>Residual of objective function</i>
$\Delta$	<i>Change in</i>
$\delta^2$	<i>Variance for Focused Information Criterion</i>
$\delta_i$	<i>Hildebrand/Hansen Solubility Parameter</i>
$\delta_{ij}$	<i>Cross coefficient for virial equation of state (<math>m^3 \cdot mol^{-1}</math>)</i>
$\epsilon$	<i>Tolerance of optimization</i>
$\epsilon_i$	<i>Fraction of component <math>i</math> in a phase</i>
$\zeta_{ij}$	<i>Fraction of component <math>i</math> in phase <math>j</math></i>
$\zeta_{k,i}$	<i>Segment area of segment <math>k</math> component <math>i</math> for UNISAC model</i>
$\theta_i$	<i>Area Fraction</i>
$\Phi_i$	<i>Volume fraction</i>
$\Theta_{m,i}$	<i>Segment area fraction of segment <math>m</math> in component <math>i</math> for UNISAC model</i>
$\Lambda_{ij}$	<i>T-K Wilson model parameter</i>
$\lambda$	<i>Lévy flight distribution exponent</i>
$\lambda_{ij}-\lambda_{ii}$	<i>T-K Wilson model fit parameter (<math>J \cdot mol^{-1}</math>)</i>



$\mu_i$	<i>Chemical potential of component i in a mixture</i>
$\nu_k$	<i>Frequency of a functional group/segment</i>
$\pi$	<i>Total number of phases</i>
$\sigma_m$	<i>charge density of component i/predicted uncertainty deviation uncertainty in parameter m</i>
$\tau_{ij}$	<i>NRTL model parameter</i>
$\Phi_i$	<i>Volume Fraction</i>
$\phi_i$	<i>Fugacity coefficient</i>
$\phi_i$	<i>Fugacity coefficient</i>
$\psi_{k,m}$	<i>Binary interaction parameter for the UNIFAC/UNISAC/mod. UNIFAC models</i>
$\Omega_{k,i}$	<i>Total segment area fraction of segment k in component i for UNISAC model</i>
$\infty$	<i>Property at infinite dilution</i>
$\mathbb{O}$	<i>The big O</i>

**Subscripts/  
Superscripts**

'	<i>Prime symbol to indicate a variation to original parameter</i>
◦	<i>Reference state, standard property</i>
0	<i>Hypothetical reference state</i>
A,B,i,j,k,l,m,n,1,2	<i>Reference to a particular component/group/segment</i>
Calc/calculated	<i>Calculated property</i>
Cav	<i>Cavity volume for the Moller et al. (2014) combinatorial expression</i>
Cav-Corr	<i>Cavity volume correction for the Moller et al. (2014) combinatorial expression</i>
comb	<i>Combinatorial expression</i>
d,h,p+,p-	<i>Dispersion, hydrogen bonding, positive polarity, negative polarity</i>
exp	<i>Experimental property</i>
F	<i>Parameter evaluated at feed condition</i>
F-H	<i>Flory-Huggins related property</i>
global	<i>A global optima</i>
FV	<i>Denotes a Free-Volume correction</i>
initial	<i>An initial guess of a parameter</i>
k	<i>candidate model for the Focused Information Criterion</i>
l	<i>Liquid phase</i>
pred	<i>Predicted property</i>
res	<i>Residual Expression</i>
s	<i>Solid phase</i>

$s \rightarrow l$	<i>Solid to liquid transition</i>
$S-G$	<i>Staverman-Guggenheim related property</i>
$T$	<i>A total parameter</i>

## CHAPTER ONE

### Introduction and Background

#### 1.1. Outline and rationale

An activity coefficient is an essential parameter for the accurate description of the phase behaviour of low and moderate pressure non-ideal systems. It can also be used in conjunction with equations of state to describe the phase behaviour of high pressure systems using Gibbs Excess energy mixing rules. This parameter cannot be directly measured, but is generally calculated by the fitting of activity coefficient models to phase equilibrium data. The most popular models include the Wilson (1964), Non-Random Two-Liquid (NRTL, Renon and Prausnitz, 1968) and Universal Quasi-Chemical (UNIQUAC, Anderson and Prausnitz, 1978) models. In order for a model to provide a fair representation of the experimental behaviour, a significant number of experimental data points must be measured comprising the entire composition range if possible. Additionally, if it is desired that temperature effects be replicated, non-isothermal or excess enthalpy ( $H^E$ ) data must also be measured. In this work these types of correlative models that require specific experimental data of the system being considered, are referred to as “*non-predictive*”, and contrasts with “*purely predictive*” models such as the group contribution and COSMO based models discussed below. However, in general it is of course possible to perform a certain degree of prediction using correlative activity coefficient models.

Common estimates by Hoffmann (Hoffmann, 1982) suggest that there are over 150 000 chemicals currently used in industry, with over 2000 new chemicals being introduced to the market annually. The binary combinations of these components amounts to some 11 billion possible systems, with tertiary and quaternary systems concededly yielding innumerable

combinations. In this work, focus is placed on chemical mixtures containing at least one solid phase at the experimental temperature and pressure (solid-liquid/ solid-liquid-liquid systems), and more specifically to pharmaceutical and long-chained components that are generally solid at ambient conditions. Experimental data of such systems is limited in the literature in comparison to vapour-liquid and liquid-liquid data due especially to the increasing number of pharmaceuticals being developed by the rapidly growing pharmaceutical industry. In the absence of experimental data, the non-predictive activity coefficient models such as those stated above are rendered essentially useless.

In the last 50 years a great deal of research has focused on the prediction of the activity coefficient, applying numerous theories, in order to provide useful estimates of the phase behaviour of chemical mixtures. In this work, a purely predictive model is defined as a model which does not require any experimental data of the system being considered in order to perform a prediction. These predictive activity coefficient models are useful for several reasons: they can, in many cases, provide an exceptionally accurate *a priori* replication of the experimental behaviour; less accurate predictions can still provide reasonable estimates for preliminary process simulations; predictions can be used to determine the design limits of an experimental setup for subsequent measurement; and they can be applied to the prediction of high-pressure phase behaviour via a  $G^E$  mixing rule as in the case of the Predictive Soave-Redlich-Kwong equation of state (Holderbaum and Gmehling, 1991).

The most common theories employed in the prediction of the activity coefficient are the group contribution approach, such as that of UNIFAC (Fredenslund et al., 1975), and modified UNIFAC (Dortmund) (Weidlich and Gmehling, 1987), and the segment contribution approach, such as that of COSMO-RS (Klamt, 1995), COSMO-SAC (Lin and Sandler, 2002) and more specifically to cases involving solid-liquid systems, NRTL-SAC (Chen and Song, 2004, Chen and Crafts, 2006). The NRTL-SAC model however cannot be considered “*purely-predictive*”- a deficiency which is discussed below.

All predictive models require a number of system-specific constants that improve the quality of model prediction from case to case. For instance, in the group contribution method, group interaction parameters, as well as group areas and volumes, are required in order to perform an activity coefficient prediction. The COSMO-based model requires component specific sigma profiles, while NRTL-SAC requires four component specific segment area parameters.

In the order of an increasing number of required parameters, the common models are ranked UNIFAC, modified UNIFAC Dortmund, COSMO-RS and COSMO-SAC, and NRTL-SAC. This, of course, assumes that a large number of components are being considered (greater than the number of functional groups comprising the components). If, for instance, two components are being considered, which are comprised of four functional groups, then the COSMO based methods will obviously require fewer parameters than the group based methods.

In this work the modified UNIFAC Dortmund Consortium (2013) version of the modified UNIFAC Dortmund model is used.

Chapter Two of this work is mainly comprised of a manuscript titled “*Model Evaluation for the Prediction of Solubility of Active Pharmaceutical Ingredients (APIs)*”, In this manuscript, the common predictive activity coefficient models available in the literature were reviewed, and then applied to a data set of active pharmaceutical ingredients, including complex polycyclic steroidal molecules from the literature, to highlight, first-hand, the potential strengths and weaknesses of the available models. The models were assessed on their ability to replicate the experimental solubilities, based on a percentage deviation. Furthermore, the effects of assuming negligible, or estimated heat capacity changes of the solute, were considered.

This paper shows the effectiveness of the segment approach of activity coefficient prediction (NRTL-SAC), for solid-liquid systems, and also highlights the fact that a model cannot be employed if the system specific model parameters are not available. Additionally, it is also emphasized that the group contribution-based methods cannot be employed if the molecules cannot be fragmented into the groups, considered by the relevant group contribution model. Furthermore, it is reiterated that the NRTL-SAC model, although effective, requires the largest set of model specific parameters, in comparison to all other models tested, in order to be successfully applied. The focus of this doctoral research is to address these shortcomings in the development of a new predictive activity coefficient model.

The initial version of the newly developed model is the UNiversal Segment Activity Coefficient (UNISAC) model. The model infers that the concept of segment surface area can be extended to functional groups. Details of the initial development and testing are outlined in the published article, “*A Universal Segment Approach for the Prediction of the Activity*

*Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents*”, (Moodley et al. 2015 (b)) and is presented in Chapter Three of this work.

It was found that the UNISAC model predictions are highly sensitive to the group specific parameters that are obtained by pseudo-experimental data regression (explained later). In the UNISAC model, the Nelder-Mead Simplex (Nelder and Mead, 1965), method was used to obtain parameter estimates at locally encountered minima. In order to improve these parameter estimates, global optimization techniques were considered. The newly developed Krill Herd Algorithm (Gandomi and Alavi, 2012), global optimization technique, was selected, based on the recent favourable results of the method, available in the literature (Gandomi and Alavi, 2012).

The Krill Herd algorithm, however, had not been previously tested on thermodynamically related optimization problems, and more specifically, not in problems involving phase equilibria. To test the method’s performance in such situations, in relation to other global optimization techniques, a set of complex phase stability, phase equilibrium and reactive phase equilibrium problems, were selected from the literature, and the algorithm was applied. The results of these tests were published and are presented as a manuscript in Chapter Four, titled, “*Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations*”, (Moodley et al., 2014 (c)). The Krill Herd Algorithm generally outperforms several other common stochastic methods from the literature (Fateen et al., 2012 and Bhargava et al., 2013).

After further considerations, shortcomings of the UNISAC model were found and addressed. For instance, in the original work, interaction parameters for base segment groups (defined later) were fitted to experimental solid-liquid equilibrium data. Due to this, very poor performances in the cases of vapour-liquid and liquid-liquid systems were yielded for the UNISAC model. Although the model was not developed for these applications, realistic, albeit qualitative predictions, in these alternate types of systems, should be possible, and so it was attempted to remedy this in an extended version of the UNISAC model.

Further, in the UNISAC model, (Moodley et al. 2015 (b)) the fragmentation scheme of the modified UNIFAC (Dortmund) method was used in the residual expression of the model, and

proved to be unsuitable in some cases dealing with complex pharmaceutical molecules. If a molecule cannot be fragmented into relevant groups, the UNISAC method cannot be employed.

A new fragmentation scheme, based on the work of Moller et al. (2008), was further developed and employed, which was able to fragment the majority of the components considered in this work. This scheme is slightly more detailed with approximately 130 different functional groups, and is termed the Moodley-Rarey-Ramjugernath fragmentation. The Krill Herd Algorithm was then used to determine the UNISAC model parameters for these new group allocations. Since the groups considered, and model parameters calculated, in this more elaborate version of the UNISAC model, differ from those of the original prototype of the UNISAC model, the further developed model is termed the Extended UNISAC model. It must be mentioned that currently, Extended UNISAC model parameters are only available for a portion, and not all, of the 130 Moodley-Rarey-Ramjugernath groups.

The Extended UNISAC model's ability was tested in great detail, and was used to perform infinite dilution activity coefficient (as a preliminary calculation), and ultimately, solid-liquid equilibrium predictions, in binary and multi-component systems, for a large number of experimental data points from the literature, obtained through the Dortmund Data Bank (2012). The predictive capabilities of the new model were found to be competitive with other popular based models (UNIFAC, modified UNIFAC Dortmund, COSMO-RS and COSMO-SAC), currently employed in the literature and in industry. This work is presented in manuscript format in Chapter Five titled, "*An Extended UNISAC model for the Prediction of Solubility of Complex Pharmaceutical Ingredients in Non-electrolyte Pure Solvents and Solvent Mixtures*". (Moodley et al, 2015 (d)).

After sufficient development and testing of the Extended UNISAC model, it was decided to perform predictions for systems, for which measurements have not been previously reported in the literature, in order to ensure that the favourable results, obtained from the work of Chapter Five, are also valid outside the training set. A further advantage is that the quality of the experimental data, used for comparison, is within the author's control. Hence, solid-liquid equilibrium data was measured, using combined thermal gravimetry and differential thermal analysis (DTA/TGA), to determine liquid phase compositions at saturation, of various polycyclic pharmaceutical-solvent systems. The systems considered were diosgenin/ estriol/ prednisolone/ hydrocortisone/ betulin/ estrone in organic solvents, and water, in the

temperature range of 293.2-328.2 K. These results, that comprise a series of three manuscripts (Moodley et al. 2015 (e), Moodley et al. 2015 (f), Moodley et al. 2015 (g)), are presented in a single chapter, (Chapter Six), for convenience. It is titled “*Betulin/Estrone/Diosgenin/Estriol/Prednisolone/Hydrocortisone in Various Solvents in the Temperature Range  $T = (293.2 \text{ To } 328.2) \text{ K}$ ”*. For comprehensiveness, this data was modelled by means of the classic correlative activity coefficient models.

The Extended UNISAC model was found to provide competitively accurate predictions of the solubility in these systems, in comparison to other predictive methods from the literature, which is shown in the concluding discussion of this work.

In the further sections of this chapter a brief literature review is presented which includes the theoretical approach to solid-liquid equilibrium modelling and the existing predictive methods for the activity coefficient from the literature. Subsequently the development of the new UNISAC model, optimization strategies and experimental techniques and apparatus are introduced.

## 1.2 Thermodynamics of phase equilibrium

A brief description of the equilibrium thermodynamics pertinent to this work is presented, with further details available in the work of Smith et al. (2005) for instance.

### 1.2.1 Criterion for equilibrium

The Gibbs energy,  $G$ , is a function of the measurable quantities; pressure,  $P$ , temperature,  $T$ , and number of moles,  $n$ :

$$nG = f(T, P, n_1, \dots, n_k) \quad (1.1)$$

The total differential of equation (1.1) yields an expression for the change in Gibbs energy of an open system:

$$d(nG) = \left[ \frac{\partial(nG)}{\partial P} \right]_{T,n} dP + \left[ \frac{\partial(nG)}{\partial T} \right]_{P,n} dT + \sum_{i=1}^k \left[ \frac{\partial(nG)}{\partial n_i} \right]_{T,P,n_j} dn_i \quad (1.2)$$

Where 
$$\left[ \frac{\partial(nG)}{\partial P} \right]_{T,n} = nV \quad (1.3)$$

And 
$$\left[ \frac{\partial(nG)}{\partial T} \right]_{P,n} = -nS \quad (1.4)$$

The chemical potential,  $\mu_i$ , of component,  $i$ , is defined as the partial differential of Gibbs energy, with respect to the number of moles of component, in the mixture  $i$ , at constant temperature, pressure and molar composition of all other constituents:

$$\mu_i = \left[ \frac{\partial(nG)}{\partial n_i} \right]_{T,P,n_j} = \bar{G}_i \quad (1.5)$$

Equation (1.2) can then equally be expressed as:

$$d(nG) = \left[ \frac{\partial(nG)}{\partial P} \right]_{T,n} dP + \left[ \frac{\partial(nG)}{\partial T} \right]_{P,n} dT + \sum_{i=1}^k \mu_i dn_i \quad (1.6)$$

For a system composed of phases,  $\alpha, \beta \dots \pi$ , the following can be written, if each phase is treated as an open system:

$$d(nG)^\alpha = (nV)^\alpha dP - (nS)^\alpha dT + \sum \mu_i^\alpha dn_i^\alpha \quad (1.7)$$

$$d(nG)^\beta = (nV)^\beta dP - (nS)^\beta dT + \sum \mu_i^\beta dn_i^\beta \quad (1.8)$$

$$d(nG)^\pi = (nV)^\pi dP - (nS)^\pi dT + \sum \mu_i^\pi dn_i^\pi \quad (1.9)$$

Summation of equations (1.7-1.9), for all phases according to the relation



$$nM = (nM)^\alpha + (nM)^\beta + \dots + (nM)^\pi \quad (1.10)$$

yields total changes for the system:

$$d(nG) = (nV)dP - (nS)dT + \sum \mu_i^\alpha dn_i^\alpha + \sum \mu_i^\beta dn_i^\beta + \dots + \sum \mu_i^\pi dn_i^\pi \quad (1.11)$$

However for a closed system, where no changes in the number of moles of a particular component occurs:

$$d(nG) = (nV)dP - (nS)dT \quad (1.12)$$

Substitution of equation (1.12) into equation (1.11) results, after simplification, in:

$$\sum \mu_i^\alpha dn_i^\alpha + \sum \mu_i^\beta dn_i^\beta + \dots + \sum \mu_i^\pi dn_i^\pi = 0 \quad (1.13)$$

Now consider for instance that only two phases are formed, i.e. the pair  $\alpha$  and  $\beta$ , then the conservation of mass for non-reactive closed systems dictates that  $dn_i^\alpha = -dn_i^\beta$

Substituting into equation (1.13) yields:

$$\sum (\mu_i^\alpha - \mu_i^\beta) dn_i^\alpha = 0 \quad (1.14)$$

Since the quantity  $dn_i^\alpha$  is independent and arbitrary, the only solution to equation (1.14) is:

$$\mu_i^\alpha - \mu_i^\beta = 0$$

Hence 
$$\mu_i^\alpha = \mu_i^\beta \quad (1.15)$$

Similarly, considering any combinations of pairs of phases ( $\alpha$  till  $\pi$ ) in equation (1.13) for,  $i$ , to,  $N$ , species yields:

$$\mu_i^\alpha = \mu_i^\beta \dots = \mu_i^\pi \quad (i = 1, 2, \dots, N) \quad (1.16)$$

### 1.2.2 The activity coefficient

For a closed system, from equation (1.12) it is evident that

$$dG = d\mu = VdP - SdT \quad (1.17)$$

Now if one considers an ideal gas at constant temperature and replaces volume as a function of pressure one obtains:

$$\int_{\mu^0}^{\mu} d\mu = RT \int_{P^0}^P \frac{dP}{P} \quad (1.18)$$

Which simplifies to

$$\mu - \mu^\circ = RT \ln \frac{P}{P^\circ} \quad (1.19)$$

Where the superscript,  $^\circ$ , is used to identify a selected reference state. When non-ideal gas mixtures are considered, then in equation (1.19), the chemical potential of component  $i$ ,  $\mu_i$ , replaces the pure species chemical potential,  $\mu$ , the reference pressure  $P^\circ$  is replaced by the pure component fugacity of component  $i$  at the system temperature and pressure,  $f_i^\circ$  and the ideal total pressure is replaced by the fugacity in solution of component  $i$ ,  $\hat{f}_i$ , yielding:

$$\mu_i - \mu_i^\circ = RT \ln \frac{\hat{f}_i}{f_i^\circ} \quad (1.20)$$

When considering non-ideal mixtures involving liquids at low to moderate pressures, the reference state selected is often the ideal liquid mixture. A new parameter is then defined, termed the activity coefficient,  $\gamma_i$ , which describes the departure of the real fugacity in solution of component  $i$ , from the ideal fugacity in solution of component  $i$ , as follows:

$$\gamma_i = \frac{\hat{f}_i}{x_i f_i^\circ} = \frac{\hat{f}_i}{f_i^\circ} \quad (1.21)$$

Substitution of equation (1.21) into equation (1.20) for a non-ideal mixture yields:

$$\mu_i - \mu_i^\circ = RT \ln(x_i \gamma_i) \quad (1.22)$$

Or equivalently, using equation (1.5):

$$\bar{G}_i - \bar{G}_i^\circ = RT [\ln \hat{f}_i - \ln f_i^\circ] \quad (1.23)$$

And by the definition of an excess property, equation (1.23) becomes:

$$\frac{\bar{G}_i^E}{RT} = [\ln \hat{f}_i - \ln f_i^\circ] = \ln \gamma_i \quad (1.24)$$

A consequence of equation (1.24) at constant,  $T$ , and,  $P$ , is the iso-fugacity condition of equilibrium, analogous to equation (1.15):

$$\hat{f}_i^\alpha = \hat{f}_i^\beta = \dots \hat{f}_i^\pi \quad (1.25)$$

### 1.2.3 Solid-liquid phase equilibrium

At solid-liquid phase equilibrium the solvent is saturated with the solute. This state can be described thermodynamically as the equality of chemical potentials, and equally fugacity in solution, of a component,  $i$ , in each phase i.e. solid and saturated liquid according to the equation (1.15) and (1.25):

$$\mu_i^s = \mu_i^l \quad (1.26)$$

$$\hat{f}_i^s = \hat{f}_i^l \quad (1.27)$$

Where the superscripts,  $s$ , denotes the solid phase, and,  $l$ , denotes the saturated liquid phase. The chemical potential of the solid,  $i$ , in the saturated liquid is given by equation (1.22) as:

$$\mu_i^l = \mu_i^0 + RT \ln (\gamma_i^l x_i^l) \quad (1.28)$$

Where,  $\mu_i^0$ , is the chemical potential of component,  $i$ , in the reference state,  $T$ , is the temperature in Kelvin,  $R$ , is the Universal Gas constant in,  $J.kmol^{-1}. K^{-1}$ , and,  $\gamma_i^l$ , is the activity coefficient of component,  $i$ , in the liquid phase.

The activity of the solute can be determined by combining equations (1.26) and (1.28), yielding:

$$\ln(\gamma_i^l x_i^l) = \frac{\mu_i^l - \mu_i^0}{RT} \quad (1.29)$$

Equally it was shown that:

$$\frac{\mu_i^l - \mu_i^0}{RT} = \frac{\bar{G}_i^l - \bar{G}_i^0}{RT} \quad (1.30)$$

The triple point can be used as the reference state in the case of solid-liquid equilibrium. However since solid-liquid equilibrium systems are not sensitive to low to moderate pressures, the fusion point at 1 atmosphere is often used as a substitute due mainly to the greater abundance of this data. The Gibbs energy change for the transition from solid to liquid is thus estimated as  $\frac{\Delta_{fus}\bar{G}_i}{RT}$ . Where,  $\Delta_{fus}\bar{G}_i$ , is the partial molar Gibbs energy change from the solid state to the reference state at constant pressure and temperature. Combination with equations (1.29) and (1.30) yields:

:

$$\ln(\gamma_i^l x_i^l) = \frac{\Delta_{fus}\bar{G}_i}{RT} \quad (1.31)$$

Equally from equation (1.27) at solid-liquid equilibrium:

$$\gamma_i^s z_i^s f_i^s = \gamma_i^l x_i^l f_i^l \quad (1.32)$$

Where,  $z_i^s$ , is the solid phase composition of component,  $i$ ,. If solid-solid transitions are not considered the product:  $\gamma_i^s z_i^s = 1$ . Taking logarithms yields:

$$\ln\left(\frac{f_i^s}{f_i^l}\right) = \ln(\gamma_i^l x_i^l) = \frac{\Delta\bar{G}_i^{s \rightarrow l}}{RT} \quad (1.33)$$

Where,  $\Delta\bar{G}_i^{s \rightarrow l}$ , is the Gibbs energy change from the solid phase to the liquid phase, defined by:

$$\Delta\bar{G}_i^{s \rightarrow l} = \Delta\bar{H}_i^{s \rightarrow l} - T\Delta\bar{S}_i^{s \rightarrow l} \quad (1.34)$$

And because ideal gas reference states are eliminated by difference, equation (1.34) is equally given by:

$$\bar{G}_i^l - \bar{G}_i^s = \bar{H}_i^l - \bar{H}_i^s - T(\bar{S}_i^l - \bar{S}_i^s) \quad (1.35)$$

To evaluate  $\Delta \bar{G}_i^{s \rightarrow l}$ , the enthalpic ( $\Delta \bar{H}_i^{s \rightarrow l}$ ) and entropic changes  $\Delta \bar{S}_i^{s \rightarrow l}$  must be determined:

$$\Delta \bar{H}_i^{s \rightarrow l} = \Delta_{fus} \bar{H}_i \Big|_{fus T_i} + \Delta \bar{H}_i^{fus T_i \rightarrow T} = \Delta_{fus} \bar{H}_i \Big|_{fus T_i} + \int_{fus T_i}^T \Delta_{fus} \bar{C}_{p_i} dT \quad (1.36)$$

$$\Delta \bar{S}_i^{s \rightarrow l} = \frac{\Delta_{fus} \bar{H}_i}{T} \Big|_{fus T_i} + \frac{\Delta \bar{H}_i^{fus T_i \rightarrow T}}{T} = \frac{\Delta_{fus} \bar{H}_i}{T} \Big|_{fus T_i} + \int_{fus T_i}^T \frac{\Delta_{fus} \bar{C}_{p_i}}{T} dT \quad (1.37)$$

Combining and evaluating equations (1.33-1.37), and assuming a mean heat capacity change in the temperature range yields:

$$\ln \left( \frac{f_i^s}{f_i^l} \right) = \left[ \frac{\Delta_{fus} \bar{H}_i}{R_{fus T_i}} \left( 1 - \frac{fus T_i}{T} \right) - \frac{\Delta_{fus} \bar{C}_{p_i}}{R} \left( 1 - \frac{fus T_i}{T} \right) - \frac{\Delta_{fus} \bar{C}_{p_i}}{R} \ln \left( \frac{fus T_i}{T} \right) \right] \quad (1.38)$$

A simplification yields:

$$\ln(\gamma_i^l x_i^l) = \frac{\Delta_{fus} \bar{G}_i}{RT} = \frac{\Delta_{fus} \bar{H}_i}{R} \left( \frac{1}{fus T_i} - \frac{1}{T} \right) - \frac{\Delta_{fus} \bar{C}_{p_i}}{R} \left[ \ln \left( \frac{fus T_i}{T} \right) - \frac{fus T_i}{T} + 1 \right] \quad (1.39)$$

Where  $\Delta_{fus} \bar{H}_i$  is the enthalpy of fusion,  $_{fus} T_i$  is the temperature of fusion in Kelvin and  $\Delta_{fus} \bar{C}_{p_i}$  is the difference in heat capacity between the subcooled solute melt and the solid. Often the  $\Delta_{fus} \bar{C}_{p_i}$  in equation (1.39) is very small in comparison to the other terms, and is thus considered negligible and is omitted.

In this case equation (1.39) reduces to:

$$\ln(\gamma_i^l x_i^l) = \frac{\Delta_{fus} \bar{G}_i}{RT} = \frac{\Delta_{fus} \bar{H}_i}{R} \left( \frac{1}{fus T_i} - \frac{1}{T} \right) \quad (1.40)$$

Again, this derivation ignores the pressure influence on the solubility of the solid as the difference between the system pressure and triple point pressure is small enough that a Poynting correction factor is not required.

Hildebrand and Scott (1962, 1964) have however recommend estimating the  $\Delta_{fus}\bar{C}_{p_i}$  as  $\frac{\Delta_{fus}\bar{H}_i}{fusT_i}$  yielding:

$$\ln(\gamma_i^l x_i^l) = \frac{\Delta_{fus}\bar{G}_i}{RT} = \frac{\Delta_{fus}\bar{H}_i}{R_{fus}T_i} \ln\left(\frac{T}{fusT_i}\right) \quad (1.41)$$

This improvement has been supported by Neau et al. (1997). The consequence of using the assumption of equation (1.40) and (1.41) is explored in systems containing active pharmaceutical ingredients of varying molecule size in Chapter 2, with various models from the literature.

### 1.3 Predictive Methods for the calculation of solubility or activity coefficient and their application

A brief overview of the common methods employed for solubility calculations follows. The reader is referred to the original publications for greater detail.

#### 1.3.1 The Hansen Solubility Model

Mutual solubility can only be achieved by a negative Gibbs energy change upon mixing,  $\Delta G^{mix}$ . That is, two substances will only mix if the resultant solution has a lower Gibbs energy than that of the pure components comprising the combined mixture.  $\Delta G^{mix}$ , is a function of the changes in enthalpy and entropy upon mixing and is given by:

$$\Delta G^{mix} = \Delta H^{mix} - T\Delta S^{mix} \quad (1.42)$$

For spontaneous mixing,  $\Delta S^{mix}$ , is always positive, hence, mixing occurs when  $\Delta H^{mix} < T\Delta S^{mix}$ . Hildebrand, (Hildebrand, 1916), Scatchard (Scatchard, 1931) and Hildebrand and Scott (Hildebrand and Scott, 1962, 1964) proposed a method to estimate the enthalpy change

of mixing for regular solutions, in terms of component volumes, and component specific cohesive energy densities. The resulting expression for  $\Delta H^{mix}$  in binary mixtures, is:

$$\Delta H^{mix} = (x_1V_1 + x_2V_2)\phi_1\phi_2(\delta_1 - \delta_2)^2 \quad (1.43)$$

Where,  $x_i$ , is the component mole fraction,  $V$ , is the liquid molar volume and,  $\phi_i$ , is the volume fraction given by:

$$\phi_i = \frac{x_iV_i}{\sum_j x_jV_j} \quad (1.44)$$

$\delta_i$ , is the component specific Hildebrand solubility parameter, given by the square root of the cohesive energy density,  $c_i$ :

$$\delta_i = \sqrt{c_i} = \sqrt{\left[\frac{\Delta_{vap}H_i - RT}{V_i}\right]} \quad (1.45)$$

Where  $\Delta_{vap}H_i$  is the enthalpy of vaporization,  $R$ , is the Universal Gas Constant, and,  $T$ , is the temperature in Kelvin.

Large values of  $c_i$  indicate components that may exhibit significant intermolecular forces such as dispersion, polarity or hydrogen bonding. This method however offers no distinction between these common intermolecular forces that may have varying degrees of magnitude from component to component. Hansen (Hansen, 2007) proposed separating the Hildebrand solubility parameter into contributions due to the three common intermolecular forces, i.e. dispersion,  $\delta_i^d$ , hydrogen bonding,  $\delta_i^h$ , and polarity,  $\delta_i^p$ , along with methods for calculating each parameter, that has had some success in solubility estimation (Hansen (2007), Hoy (1985)).

The heat of mixing expression, using the method of Hansen is given by:

$$\Delta H^{mix} = (x_1V_1 + x_2V_2)\phi_1\phi_2 \left[ (\delta_1^d - \delta_2^d)^2 + 0.25(\delta_1^h - \delta_2^h)^2 + 0.25(\delta_1^p - \delta_2^p)^2 \right] \quad (1.46)$$



Combination of equation (1.42) with the Flory-Huggins (Flory, 1941, Huggins, 1941) entropy of mixing term yields an expression for the Gibbs excess energy change of mixing.

$$\Delta G^{mix} = \Delta H^{mix} + RT \sum_{i=1}^J x_i \ln \phi_i \quad (1.47)$$

The Gibbs excess energy, and hence activity coefficient, can be calculated since:

$$G^E = \Delta G^{mix} - RT \sum_{i=1}^J x_i \ln x_i \quad (1.48)$$

The Hansen model is often overlooked, as more complex, seemingly reliable and generalized methods are being employed in the literature, due mainly to the advancement in computer-aided simulation. The model however introduces the concept of intermolecular forces being directly linked to solute activity, which forms the bases of the segment approach to activity modelling discussed later.

### 1.3.2 The UNIFAC and Modified UNIFAC (Dortmund) model

The UNIFAC activity coefficient model introduced by Fredenslund et al. (Fredenslund et al., 1975) is comprised of two contributions to the activity coefficient. Namely a combinatorial (accounting for size shape interactions), and residual (accounting for energetic interactions), component.

$$\ln \gamma_i = \ln \gamma_i^{comb} + \ln \gamma_i^{res} \quad (1.49)$$

Where,  $\ln \gamma_i^{comb}$ , and,  $\ln \gamma_i^{res}$ , are the combinatorial and residual contributions respectively and are given by the following expressions:

$$\ln \gamma_i^{comb} = \ln \left( \frac{\Phi_i}{x_i} \right) + 1 - \left( \frac{\Phi_i}{x_i} \right) - \frac{Z}{2} q_i \left( \ln \frac{\Phi_i}{\vartheta_i} + 1 - \frac{\Phi_i}{\vartheta_i} \right) \quad (1.50)$$

Where

$$\vartheta_i = \frac{x_i q_i}{x_i q_i + x_j q_j} = \frac{x_i q_i}{q} \quad (1.51)$$

And

$$\Phi_i = \frac{x_i r_i}{x_i r_i + x_j r_j} = \frac{x_i r_i}{r} \quad (1.52)$$

Where,  $r_i$ , and,  $q_i$ , are the molecular volume and surface area, and  $Z$  is the coordination number. For the original UNIFAC model, the molecular volume and surface area are estimated from the group contribution values of Bondi (Bondi, 1964).

The residual term,  $\ln\gamma_i^{res}$ , is evaluated from group contributions.

$$\ln\gamma_i^{res} = \sum_k v_k^{(i)} (\ln\Gamma_k - \ln\Gamma_k^{(i)}) \quad (1.53)$$

$v_k^{(i)}$ , is the number of functional groups of the type,  $k$ , in a molecule of component,  $i$ , and  $\ln\Gamma_k^{(i)}$  is the residual contribution to the activity coefficient by the functional group,  $k$ , in the pure fluid,  $i$ . Since the pure fluid,  $i$ , is also a mixture of groups, the term  $\ln\Gamma_k^{(i)}$  is incorporated to reduce the residual term of the pure fluid to zero.

The contribution to the residual portion of the activity by the functional group,  $k$ , is given by the following relationship:

$$\Gamma_k = \exp \left( Q_k \left[ 1 - \ln(\sum_m \theta_m \Psi_{mk}) - \sum_m \frac{\theta_m \Psi_{km}}{\sum_n \theta_n \Psi_{nm}} \right] \right) \quad (1.54)$$

Where,  $\theta_m$ , is the surface area fraction of the functional group,  $m$ , in the mixture. The binary interaction parameter between groups,  $m$ , and,  $n$ , while,  $a_{mn}$ , is accounted for through the parameter,  $\Psi_{mn}$ , where:

$$\Psi_{mn} = \exp \left( -\frac{a_{mn}}{T} \right) \quad (1.55)$$

$T$ , is the system temperature in Kelvin.

As mentioned above, the expression for,  $\Gamma_k$ , presented in equation (1.54), includes the functional group,  $k$ , contributions to the activity of both the mixture and the pure fluid.

Several modifications of the original UNIFAC model have been proposed with the most significant modifications made to the expression for the temperature dependence of binary interaction parameters, as well as the introduction of different combinatorial expressions, with unique group volume and area parameters, as well as component group fragmentations.

In the modified UNIFAC (Dortmund) (Gmehling and co-workers, 1987, 1993, 1998, 2002) a quadratic temperature dependence of the binary interaction parameter,  $a_{mn}$ , is proposed:

$$a_{mn} = a_{mn,0} + a_{mn,1}T + a_{mn,2}T^2 \quad (1.56)$$

Additionally, the combinatorial expression is given by:

$$\ln\gamma_i^{comb} = \ln\left(\frac{\varphi'_i}{x_i}\right) + 1 - \frac{\varphi'_i}{x_i} - \frac{z}{2}q_i\left(\ln\frac{\Phi_i}{\vartheta_i} + 1 - \frac{\Phi_i}{\vartheta_i}\right) \quad (1.57)$$

Where

$$\varphi'_i = \frac{x_i r_i^{3/4}}{\sum_j x_j r_j^{3/4}} \quad (1.58)$$

$$\Phi_i = \frac{x_i r_i}{\sum_j x_j r_j} \quad (1.59)$$

The parameters of,  $r$ , and,  $q$ , are determined by data fitting, and not from the method of Bondi (Bondi, 1964).

The modified UNIFAC (Dortmund) model was adapted further for application to pharmaceutical systems by Diedrichs (Diedrichs, 2010). This model was termed Pharma Modified UNIFAC. It was assumed that certain functional group contributions become irrelevant in solutions of pharmaceutical molecules in common solvents if the solubility is low, and can therefore be omitted. A unique group-fragmentation scheme is used in this model. Promising results for limited classes of solvents were obtained, (Diedrichs and Gmehling, 2011). The model is however limited in applicability to solute mole fractions of 0.1.

### 1.3.1 The COSMO-RS, COSMO-SAC and COSMO-RS (OL) models

Generally the activity coefficient of a mixture is determined through the Gibbs excess energy function. Klamt (Klamt, 1995), proposed a means of determining the activity coefficient, using

chemical potentials from surface shielding charge densities determined by quantum-mechanical calculations. The Conductor-like Screening MOdel for Real Solvents (COSMO-RS) was introduced, as an *a priori* predictive model, and an alternative to the traditional group contribution-based models.

In COSMO-RS molecules of a solute/solvent system are treated as a combination of molecular shaped cavity surface segments. The concept involves modelling the placement of a “cavity” that is a replica of a molecule of the solute with zero charge, inside the homogeneous theoretical solvent with a particular fixed dielectric constant,  $\epsilon$ . The energy change involved in this placement represents a component of the total Gibbs energy change of solvation. The replica molecule charges are then replaced, yielding a realistic solute. The energy change associated with this is the second contributor to the Gibbs energy change of solvation. In order to know how charges must be replaced, each shielding charge density ( $\sigma$ ) must be characterized by a “sigma profile”.

COSMO-RS (OL) is the in-built Dortmund Data Bank modified version of the COSMO-RS model. The most significant modification to the model in this version includes an empirical correction term for hydrogen bonding, which is suggested to be over-compensated for in non-hydrogen bonding mixtures in the original COSMO-RS model. The specifics of this modification is outlined in the original publication (Grensemann and Gmehling, 2005).

Lin and Sandler (Lin and Sandler, 2002) have proposed some modifications to the original COSMO-RS model. The authors have stated that the expression for the chemical potential given by Klamt (Klamt, 1995) does not converge to certain boundary conditions, and that the expression for the activity coefficient presented does not satisfy certain thermodynamic consistency tests.

The modifications of Lin and Sandler (2002) result in the COnductor-like Screening MOdel-Segment Activity Coefficient model (COSMO-SAC), which is reviewed here.

The derivation of the expression of the activity coefficient using the COSMO-SAC model is extensive and beyond the scope of this work, but the reader is referred to the original publications for both the COSMO-RS and COSMO-SAC models for further details. The final

expression for the activity coefficient of solute,  $i$ , in solvent S,  $\ln\gamma_{i/S}$ , using the COSMO-SAC model is given by:

$$\ln\gamma_{i/S} = n_i \sum_{\sigma_m} p_i(\sigma_m) [\ln\Gamma_S(\sigma_m) - \ln\Gamma_i(\sigma_m)] + \ln\gamma_{i/S}^{SG} \quad (1.60)$$

$n_i$ , is the total number of segments contributed by molecule,  $i$ .  $\sigma_m$ , is the surface charge density of segment,  $m$ , and,  $p_i(\sigma_m)$ , is the frequency of surface charge density,  $m$ , of component,  $i$ , given by:

$$p_i(\sigma_m) = \frac{n_i(\sigma_m)}{n_i} \quad (1.61)$$

Where,  $n_i(\sigma_m)$ , is the total number of segments in component,  $i$ , with charge density,  $\sigma_m$ .  $\ln\Gamma_S(\sigma_m)$ , is the segment activity coefficient in the mixture for segments with charge density,  $\sigma_m$ , given by:

$$\ln\Gamma_S(\sigma_m) = -\ln \left\{ \sum_{\sigma_n} p_s(\sigma_n) \Gamma_S(\sigma_n) \exp \left[ \frac{-\Delta W(\sigma_m, \sigma_n)}{kT} \right] \right\} \quad (1.62)$$

Where,  $\Delta W$ , is the exchange energy and,  $k$ , is the Boltzmann constant.  $\ln\Gamma_i(\sigma_m)$ , is the segment activity coefficient in the pure component,  $i$ , for segments with charge density,  $\sigma_m$ .  $\ln\gamma_{i/S}^{SG}$ , is the Staverman-Guggenheim (Staverman 1950, Guggenheim 1952) combinatorial term given by:

$$\ln\gamma_{i/S}^{SG} = \ln \frac{\phi_i}{x_i} + \frac{Z}{2} q_i \ln \frac{\vartheta_i}{\phi_i} + l_i - \frac{\phi_i}{x_i} \sum_{j=1}^n x_j l_j \quad (1.63)$$

Where,  $\vartheta_i$ , is the surface area fraction given by:

$$\vartheta_i = \frac{q_i x_i}{\sum_j q_j x_j} \quad (1.64)$$

$\phi_i$ , is the volume fraction parameter given by:

$$\phi_i = \frac{r_i x_i}{\sum_j r_j x_j} \quad (1.65)$$

And

$$l_i = \frac{z}{2}(r_i - q_i) - (r_i - l) \quad (1.66)$$

### 1.3.2 Non-Random Two Liquid Segment Activity Coefficient model (NRTL-SAC)

The NRTL-SAC (Chen and co-workers, 2004, 2006) model is based on the polymer NRTL model by Chen (Chen, 1993) and was developed specifically for the use in the modelling of the activity of complex molecules such as pharmaceuticals. The non-ideality is accounted for, based on “contributions” from four different conceptual segments that make up a particular component. These segments include polar-positive, polar-negative, hydrophobic and hydrophilic segments. Each molecular surface is conceptually divided into these segments in different proportions of the molecular surface area. Every molecule is thus designated a conceptual segment surface “composition”. The surface interactions between pairs of segments are accounted for through constant binary interaction parameters only.

The main differences between the original NRTL model of Renon and Prausnitz (1968) and the NRTL-SAC model include the concept of segment interaction, and the addition of a combinatorial term, as size/shape interactions become considerable in larger complex molecules. Additionally the NRTL-SAC model has no in-built temperature dependency.

The combinatorial term of Flory and Huggins (Flory, 1941, Huggins, 1941) is used in the model. The subscripts,  $A$ , and,  $B$ , are used to denote pure components, whereas the subscripts,  $i, j, k, m$ , and,  $m'$ , are used to represent segment-based species indices.

$$\ln \gamma_A^{Comb} = \ln \frac{\phi_A}{x_A} + 1 - r_A \sum_B \frac{\phi_B}{r_B} \quad (1.67)$$

Where

$$r_A = \sum_i r_{i,A} \quad (1.68)$$

$$\phi_A = \frac{x_A r_A}{\sum_B x_B r_B} \quad (1.69)$$

Where,  $r_A$ , is the total number of segments,  $i$ , in component,  $A$ , and,  $\phi_A$ , is the segment mole fraction of component,  $A$ .

The residual term is identical to that of the polymer NRTL model (Chen, 1993) where:

$$\ln\gamma_A^{Res} = \ln\gamma_A^{lc} = \sum_m r_{m,A} [\ln\Gamma_m^{lc} - \ln\Gamma_m^{lc,A}] \quad (1.70)$$

Where,  $\ln\Gamma_m^{lc}$ , is the segment activity coefficient of species,  $m$ , in the mixture, and,  $\ln\Gamma_m^{lc,A}$ , is the segment activity coefficient of species,  $m$ , in the pure component,  $A$ , and are calculated from the following relations:

$$\ln\Gamma_m^{lc} = \frac{\sum_j x_j G_{jm} \tau_{jm}}{\sum_k x_k G_{km}} + \sum_{m'} \frac{x_{m'} G_{mm'}}{\sum_k x_k G_{km'}} \left( \tau_{mm'} - \frac{\sum_j x_j G_{jm'} \tau_{jm'}}{\sum_k x_k G_{km'}} \right) \quad (1.71)$$

$$\ln\Gamma_m^{lc,A} = \frac{\sum_{j,A} x_j G_{jm} \tau_{jm}}{\sum_{k,A} x_k G_{km}} + \sum_{m'} \frac{x_{m',A} G_{mm'}}{\sum_{k,A} x_k G_{km'}} \left( \tau_{mm'} - \frac{\sum_j x_{j,A} G_{jm'} \tau_{jm'}}{\sum_{k,A} x_k G_{km'}} \right) \quad (1.72)$$

Where

$$x_j = \frac{\sum_B x_B r_{j,B}}{\sum_A \sum_i x_A r_{i,A}} \quad (1.73)$$

$$x_{j,A} = \frac{r_{j,A}}{\sum_i r_{i,A}} \quad (1.74)$$

And

$$G_{jm} = \exp(-\alpha \tau_{jm}) \quad (1.75)$$

Where,  $r_{m,A}$ , is the number of each segment of type,  $m$ , in component,  $A$ .  $x_j$ , is the segment mole fraction of segment,  $j$ .  $x_B$ , is the mole fraction of component,  $B$ .  $G_{jm}$ ,  $\tau_{jm}$ , and,  $\alpha$ , are the regular NRTL parameters with,  $\tau_{jm}$ , being the binary interaction energy parameter between segment,  $j$ , and,  $m$ .

### 1.3.3 Assessing the performance of popular models from the literature

Prior to any further model development, it was decided that solubility predictions by the various models available in the literature be carried out by using published experimental data as a

comparison. Data sets including complex polycyclic steroidal molecules from the literature were selected, to highlight first-hand, the potential strengths and weaknesses of the available predictive models in the literature that included the UNIFAC, modified UNIFAC (Dortmund), COSMO-RS (OL), COSMO-SAC and NRTL-SAC. The models were assessed based on their percentage deviations between the experimental and predicted solubilities. The results are available in detail in the form of a manuscript in Chapter Two.

The manuscript shows the effectiveness of the segment approach of activity coefficient prediction (NRTL-SAC) for solid-liquid systems, and also highlights the fact that a model cannot be employed if the system specific model parameters are not available. Additionally it is also emphasized that the group contribution based method cannot be employed, if the molecules considered cannot be fragmented into the groups considered by the relevant group contribution model. Furthermore it is reiterated that the NRTL-SAC model, although effective, generally requires the largest set of model specific parameters (four per component), in comparison to all other models tested, in order to be successfully applied. The model's range of applicability is thus limited.

The focus of this doctoral research is to address these shortcomings in the development of a new predictive activity coefficient model.

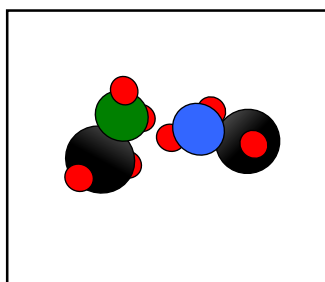
#### *1.4 Development of the Universal Segment Activity Coefficient (UNISAC) model*

The new model, introduced by this work, aims to integrate the most favourable aspects of the most common and successful predictive activity coefficient models. The segment surface interaction concept of NRTL-SAC forms the theoretical basis of the model, and the reader is therefore encouraged to be familiar with the work of Chen and co-workers (2004, 2006).

The segment surface interaction concept infers that the energetic interactions between any two components in a mixture can be attributed to only four intermolecular forces, namely dispersion, hydrogen bonding, and positive or negative polarity. Only these four intermolecular interactions are considered in the UNISAC model. However it is further postulated in this work that these interactions do not only exist between molecular surfaces, but alternatively can be modelled to exist between the functional groups that comprise the molecule, analogous to group interactions in UNIFAC.

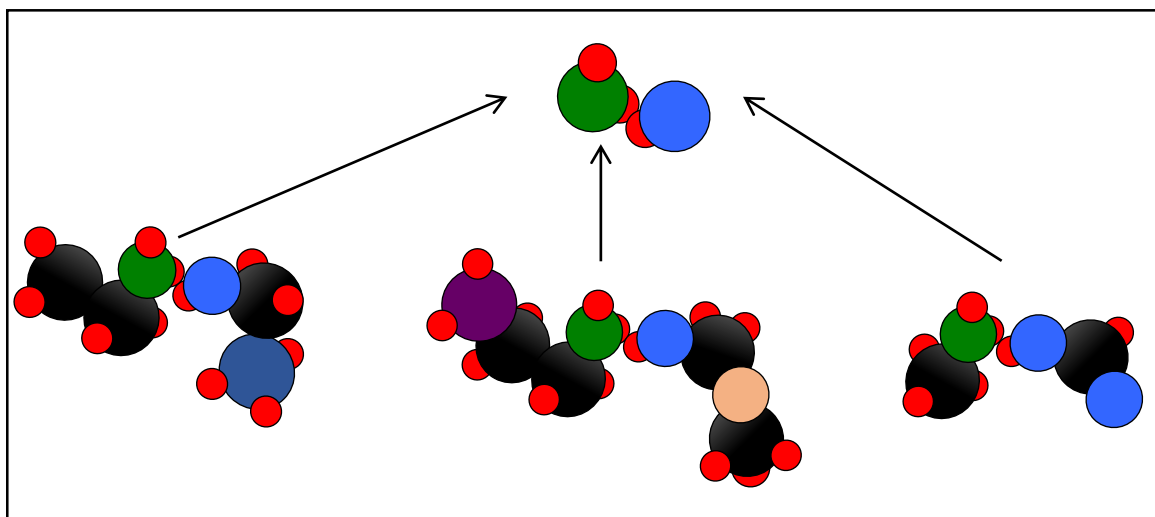


The conventional intermolecular interaction of hydrogen bonding between two molecules is given as an example, with the hydrogen bond potential existing between the green and blue generic functional groups as in Figure 1.1:



**Figure 1.1: Schematic showing possible hydrogen bonding sites between functional groups of 2 components.**

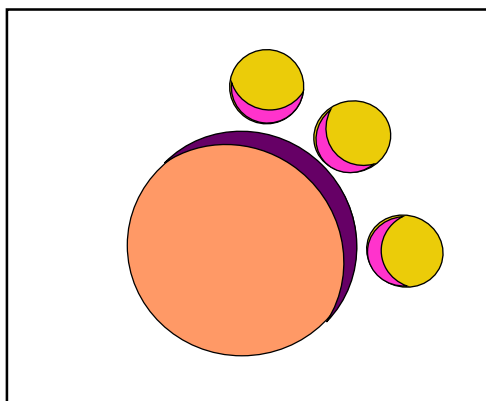
In the group contribution approach, this interaction is considered to only exist between the two groups (green and blue), comprising the molecule directly involved in the interaction, and is an identical contribution, regardless of the other groups comprising the molecules, as seen in Figure 1.2:



**Figure 1.2: Schematic showing the assumption of identical hydrogen bonding sites between functional groups comprising any 2 components**

In the segment contribution approach employed in NRTL-SAC, every molecule exhibits some degree of hydrogen bonding (sometimes virtually zero). This can be conceptualised using generic molecules (orange and yellow spheres) that do not have an obvious hydrogen bonding site, such as in an alcohol (R-O-H), for instance. Those surface sites (a theoretical construct

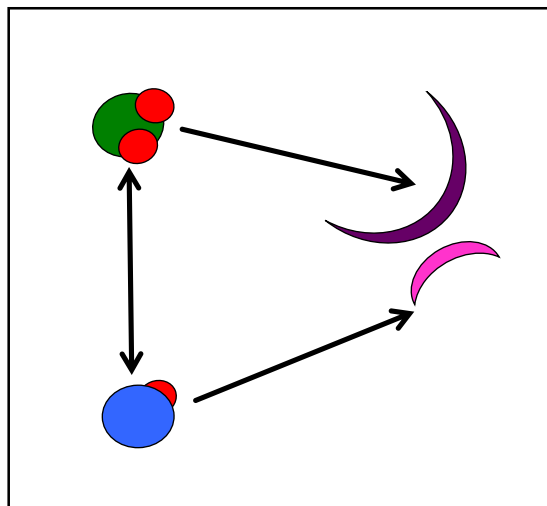
only), that exhibit a hydrogen bond-like interaction, are shown in purple and pink in Figure 1.3:



**Figure 1.3: Schematic showing functional group surface fraction exhibiting hydrogen bonding.**

In the case above, the hydrogen bond interaction occurs between the hydrogen bonding surface segment of each molecule (the pink and purple surfaces), with a larger hydrogen segment surface indicative of a higher hydrogen bonding potential for that molecule.

In the new model, a universal segment approach is proposed, which infers that the surface of a functional group can be divided into segments, each exhibiting one of the four basic intermolecular interactions considered by NRTL-SAC. As in the earlier example (green and blue), the two functional groups exhibiting hydrogen bonding potential, with each other, contribute to the total molecular hydrogen bonding potential. The total hydrogen bonding potential of the molecule is determined, by a scaled summation of the hydrogen bond potential, between each pair of functional groups, both in the pure molecule, and in the mixture, as shown in Figure 1.4.



**Figure 1.4: Schematic showing fractional contributions to hydrogen bonding from different functional groups.**

The new model, termed the Universal Segment Activity Coefficient (UNISAC) model, is presented in Chapter Three as a publication titled “*A Universal Segment Approach for the Prediction of the Activity Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents*” (Moodley et al., 2015 (b)). There, the concept of the UNISAC model is explained in detail, along with preliminary testing, and with comparison to experimental data from the literature. This model will henceforth be referred to as the UNISAC model. The promising results of the testing revealed:

- The segment contribution of functional groups was confirmed to be additive.
- The number of model parameters required to apply the UNISAC model is much lower than that of the COSMO-based, UNIFAC-based and NRTL-SAC models.

These results encouraged further development and testing.

The main challenges faced during development of the UNISAC model, as well as the shortcomings of the model, were also identified. They included:

- Obtaining good estimates of the global minimum of desired objective functions, used to determine the UNISAC model parameters, during the fitting procedures.

- The fragmentation scheme employed in the modified UNIFAC (Dortmund) method, which was also used to represent the UNISAC functional groups, was not comprehensive enough to handle complex pharmaceutical and polymer molecules.
- A four parameter segment list may not be sufficient to describe more complex behaviours, such as molecules that can both accept and donate hydrogen bonds.
- Fitting binary interaction parameters to pharmaceutical solid-liquid equilibrium data only, caused a loss of predictive power in vapour-liquid and liquid-liquid systems.

### *1.5 Global optimization*

All predictive activity coefficient models require some degree of model training in order to determine model specific parameters for usage. During this training process a fitting procedure is often used, with a particular objective function. Model equations are however generally highly complex, and consequently non-linear objective functions are a common occurrence in problems involving phase equilibrium.

Non-linear objective functions often give rise to numerous local extrema solution points. As mentioned earlier, it was found, during the training of the UNISAC model, that solutions and local minima were often obtained, which significantly affected the suitability and accuracy of the model parameters determined by regression for the UNISAC model. It was therefore decided to employ a global optimization tool that can be proven to be efficient in performing global optimization in systems involving phase equilibrium.

Many global optimization methods, which are generally classified as either deterministic or stochastic, are available in the literature: (Land and Doig, 1960, Kirkpatrick et al., 1983, Dorigo, 1992, Duan et al., 1992, Hansen and Ostermeier, 2001, Mordecai, 2003, Walster and Hansen, 2004, Srinivas and Rangaiah, 2007, Yang, 2010, Yang and Deb, 2010, Walton et al., 2011, Gandomi and Alavi, 2012, Wang et al., 2013). Deterministic methods often require a large amount of computational time, as well as restrictions on the continuity and convexity of the objective function, such as with cutting plane (Mordecai, 2003), branch and bound (Land and Doig, 1960), and interval analysis algorithms (Walster and Hansen, 2004).

However the stochastic methods require limited information on the nature of the optimization problem, and are able to handle issues pertaining to discontinuity and convexity. The

computing time is generally reasonable and convergence to the global optimum is highly probable.

Metaheuristic optimization techniques are a sub category of the stochastic methods, and involve an intelligent selection of random variables, often modelled around natural phenomena. The Krill Herd algorithm, introduced by Gandomi and Alavi (2012), is a metaheuristic, based on the simulation of the behaviour of a herd of Antarctic krill crustacean, and its response to certain environmental and biological processes such as predation, general movement, foraging for food and natural drifting. The method has been successfully applied to several benchmark optimization problems (Gandomi and Alavi, 2012), with application of the Krill Herd algorithm to problems involving complex phase stability and equilibria being performed as a part of this study to support its use (Moodley et al. (2015 (c)).

### *1.6 Experimental Equipment and Procedure for Solubility Measurements*

The primary focus of this work is the development of the novel UNISAC model for the prediction of the activity coefficient in solid-liquid systems at equilibrium. It was however recognised that after sufficient development and trials of any new model, it is prudent to perform tests on system data, which assuredly does not form part of the training set. Employing previously unmeasured data for comparison allows for this assurance. A further advantage is that the quality of the experimental data used for comparison, during testing, is within the author's control. Hence solid-liquid equilibrium data was measured, using combined thermal gravimetry and differential thermal analysis (DTA/TGA), to determine liquid phase compositions, at saturation, of various polycyclic pharmaceutical-solvent systems. This data was then compared to the UNISAC model predictions, to further confirm the results of the new predictive model. The systems considered were diosgenin/ estriol/ prednisolone/ hydrocortisone/ betulin/ estrone in organic solvents and water in the temperature range of 293.2-328.2 K

#### *1.6.1 Experimental Techniques for Solubility Measurements*

Experimental methods for solid-liquid equilibrium (SLE) measurements, of pure components or mixtures, are generally categorized as either analytical or synthetic (Weir and De Loos, 2005). Analytic methods usually involve liquid-phase sampling and composition analysis, usually by gas chromatography or equivalent means. The solid phase composition is then determined by performing mass balances, and using equilibrium relationships.

Synthetic methods, involving the preparation of a solute-solvent mixture of known composition and equilibrium, can be determined by thermal signatures, visual observation, density, or volume changes, as well as ultrasonic, or dielectric measurements. For example, Domańska and co-workers: (Domańska et al., 2009, Domańska et al., 2011), Xhang et al., (Xhang et al., 2010) and Yu et al., (Yu et al., 2013).

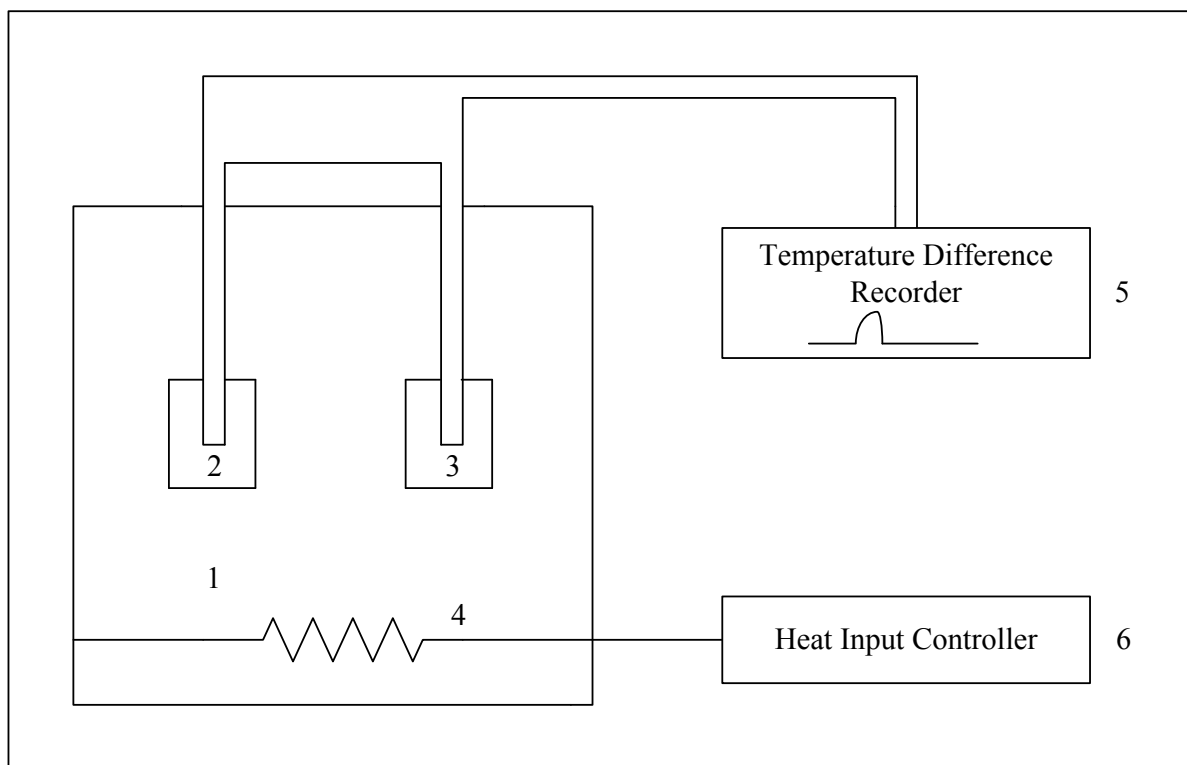
In the thermal signature analysis methods, solubility is determined by the differences in thermo-physical properties of the solution, at different temperatures and compositions. Common methods of this nature include differential scanning calorimetry (DSC), thermo-gravimetric analysis (TGA) and differential thermal analysis (DTA).

In DSC, a solution sample of known composition and a reference component, are subjected to an increasing, or decreasing, temperature profile. The heat flow rate through the solution is measured and compared to the heat flow rate through the reference component. The resulting heat flux profile can be used to determine the temperature at which saturation of the synthetic solution occurs.

The DTA method operates similarly to the DSC method. The temperature, however, of the synthetic solution and the reference component, is monitored while applying varying heating and cooling cycles. Lately, combined differential scanning and thermo-gravimetric analyses have become popular (Klančnik et al., 2009). This is mainly due to the improvement in the accuracy of DTA, as the differential changes in mass, as a function of temperature, can be measured with great precision in a well-designed apparatus.

In recent years, micro-calorimetry methods, using commercially available thermal signal analysers, have also become popular (Yu et al., 2013). Small material volumes, and low measurement times, make micro-calorimetry an attractive option for SLE, melting point, and heat of fusion data measurements.

A schematic of a typical generic differential thermal analyser is shown in Figure 1.5. The apparatus consists of six main components that include a furnace chamber (1), housing for a measurement sample and a reference sample (2 and 3), a heat source (4), a temperature difference recorder, employing temperature measuring devices (usually platinum transducers) (5), and a temperature programmer to control the heat input (6). In combined TG-DTA applications the addition of a mass measuring device is standard.



**Figure 1.5 A schematic of a typical differential thermal analyser.**

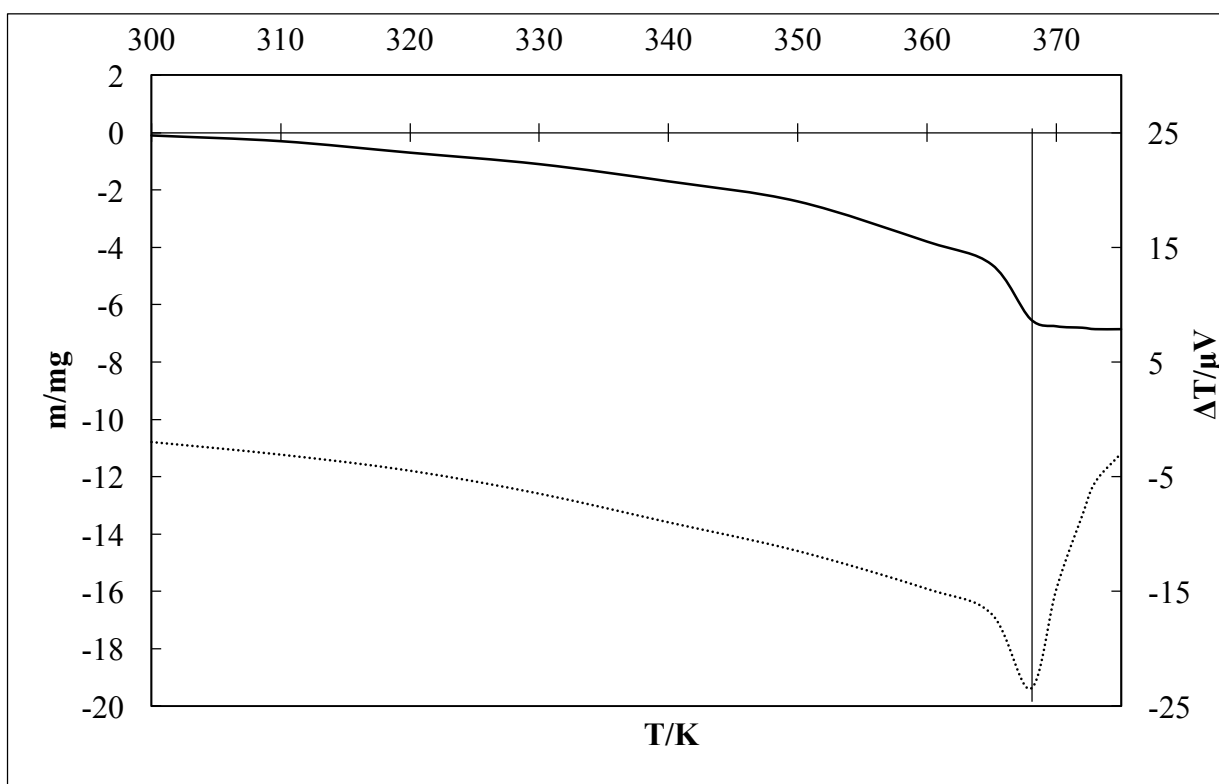
**Furnace chamber (1), measurement sample pan (2), reference sample pan (3), heating element (4), temperature difference recorder using temperature probes (5), heat input/ temperature programmer (6).**

In reference to Figure 1.5, in order to perform a phase transition measurement of a mixture, a synthetic sample, of known composition, is placed in sample pan 1 (2). The reference component, which has no phase transition in the temperature range of interest, is placed in sample pan 2 (3). The temperature programmer (6), is then set to vary the temperature of the furnace uniformly by 1-10 °C/minute, for a sample size of 5 to 10 milligrams. The temperature difference between the measurement sample and the furnace temperature, can be plotted. Similarly, in combined TG/DTA, a simultaneous plot of the change of the sample mass, with temperature, can be recorded.

A peak in the temperature/ $\Delta T$  plot indicates the occurrence of a phase transition, with a corresponding change in mass, of the sample, indicated on a TG plot of mass vs. temperature, to show whether the transition involves the loss of a volatile component. An example for the system NaCl + water is shown in Figure 1.6, using the data of Endoh and Suga (1999). It is evident that a phase transition occurs at 368 K. The authors, however, employ the non-isochoric



method, making use of an unsealed sample pan for the test component. This implies that during the heating process, the solvent escapes by evaporation. Therefore at equilibrium, the solution composition is not the composition of the synthesized input sample, but a new composition that can be simply calculated by a mass balance.



**Figure 1.6** A combined TG-DTA plot for the system NaCl-Water adapted from (Endoh and Suga, 1999); (—) TG, (···) DTA.

The temperature, at which dissolution of the solute sample occurs, is strongly influenced by the heat of solution, the mass of solute, the rate of dissolution, and the temperature effect on solubility. Consequently, the rate of temperature change has a strong effect on DTA results. An infinitely slow temperature increase rate would ensure that the solution remains saturated during the dissolution process, and equilibrium will be established at  $time = \infty$ .

When employing a finite temperature increase rate, it is possible to cause dissolution too rapidly, leading to erroneously large measurements of the equilibrium temperature. Therefore, an optimal temperature increase rate exists, which corresponds to a dissolution rate, which allows for reasonable measurement times, while still minimizing rapid dissolution rate effects.

A fundamental aspect of all experimental work is ensuring the accuracy and precision of the measured data. Mass transfer effects are a significant contributor to reduced accuracy and precision in solubility measurements (Mohan et al., 2002). Since there is no method of stirring the sample, introduced to the thermal signal analyser, dissolution is often not uniform. During the heating process, the solute eventually dissolves into the solvent. When cooled gradually, the solute will recrystallize. Generally, however, the solid will not return to its original size, distribution, or dispersion. Consequently, a repeat experiment will typically not yield an identical equilibrium temperature.

Mohan et al. (Mohan et al., 2002) found that after repeating a solubility experiment a minimum of four times, that the resulting size distribution and dispersion, within the solvent, tends to become uniform, with results comparable to literature data. However, in order to reduce extensive measurement times, these authors had employed a pre-treatment technique. The pre-treatment involves heating the solution sample to well beyond the estimated equilibrium temperature, but below the decomposition temperature, and then gradually cooling the sample back to the ambient temperature, at a controlled rate. The authors found that the pre-treatment step improves the accuracy and precision of the measured solubility data, without the need for numerous repeated measurement cycles.

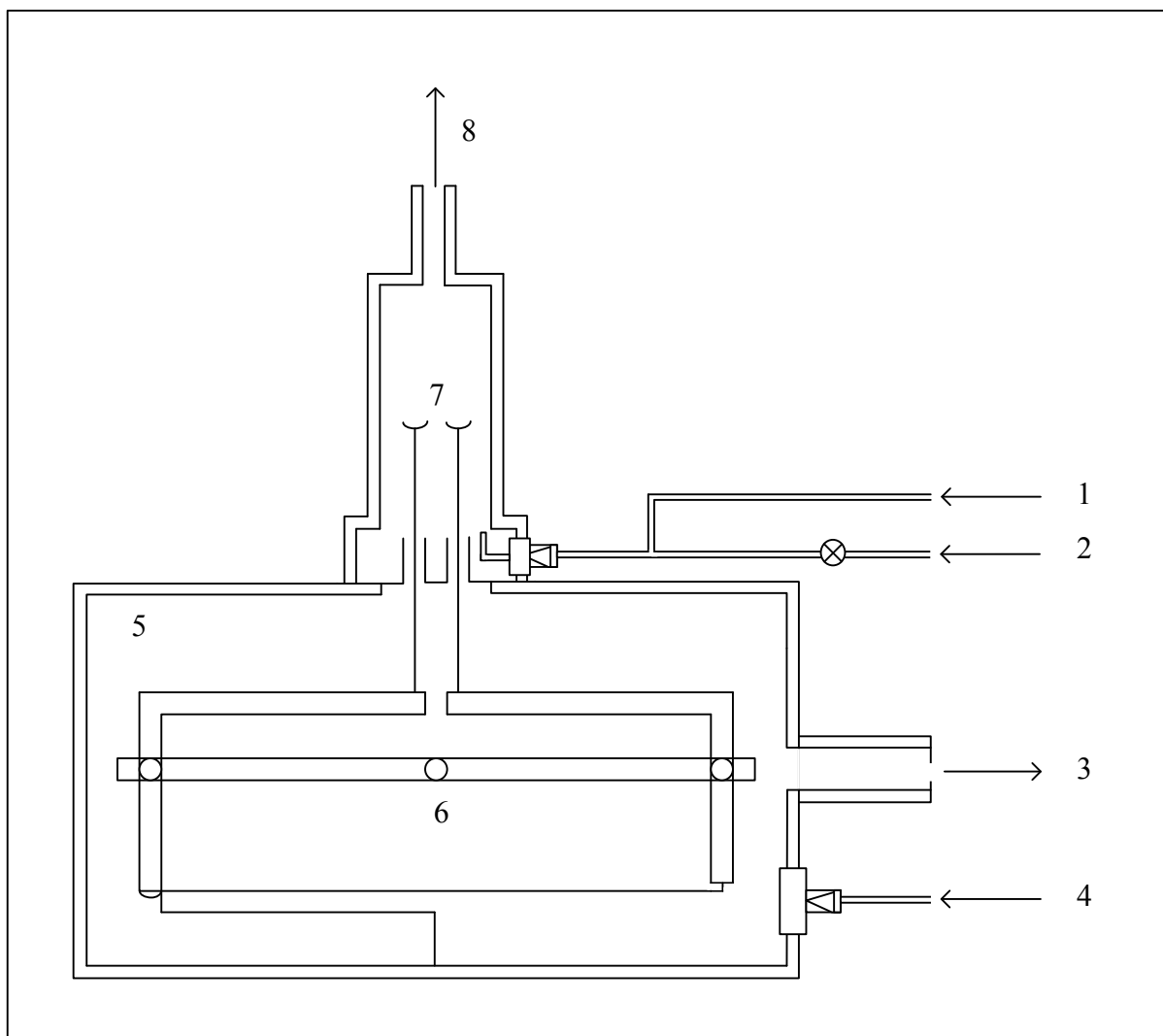
The ideas proposed in these techniques were combined, into an alternate approach, to perform solubility measurements. As mentioned, the DTG/ DTA or DSC apparatus is used, to simultaneously establish equilibrium, and measure the equilibrium composition, of a solute/solvent mixture. In this work, equilibrium was first established separately, in a closed isothermal system. A combined DTG/ DTA apparatus, supplied by Shimadzu, was then used to measure the solid-liquid equilibrium composition of complex pharmaceutical steroids (betulin, diosgenin, estriol, estrone, hydrocortisone, and mestanolone), in a variety of solvents, at several temperatures, by determining the mass of solute present, after evaporation of a mixture at solid-liquid equilibrium. Test measurements were performed to confirm the technique used. The pure solute melting point temperature, and enthalpy of fusion data, was also measured.

The majority of the data measured in this work constitutes previously unmeasured data. The experimental data at various temperatures was used in the development, and testing, of the newly introduced predictive activity coefficient model, UNISAC.

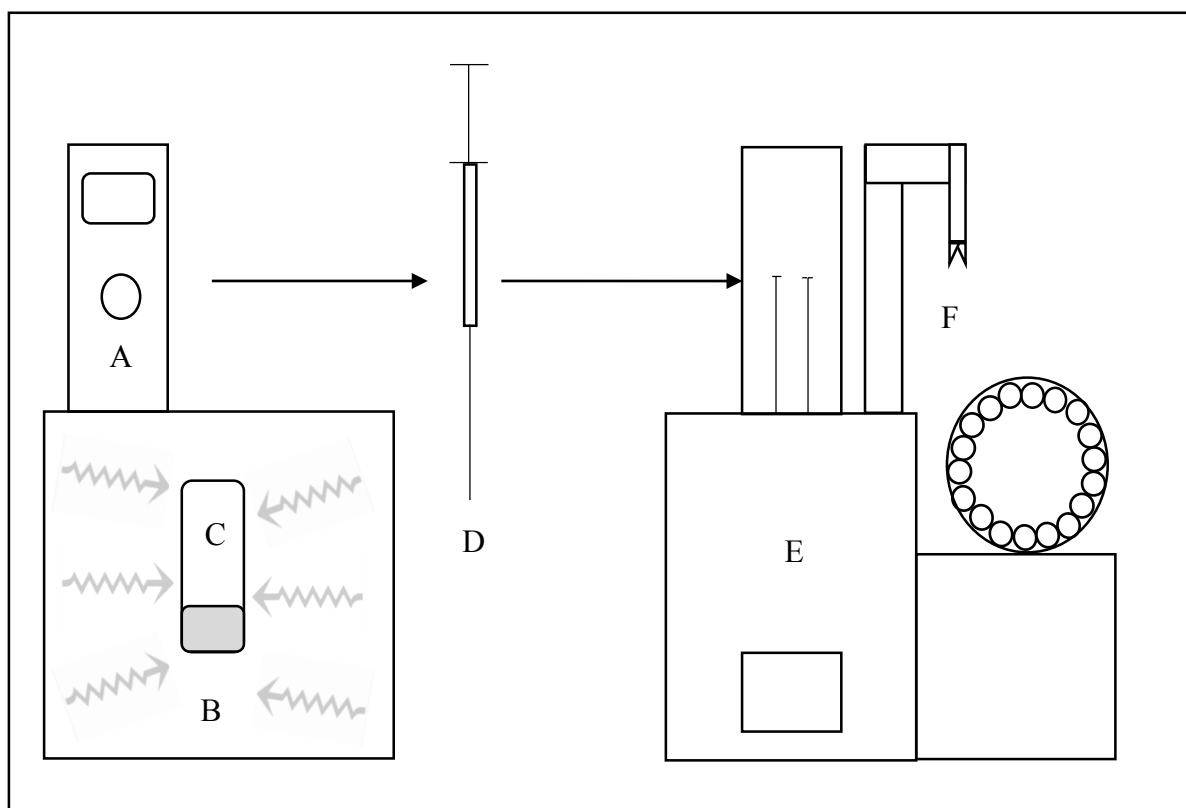
### 1.6.2 *Experimental Apparatus*

The experimental apparatus used for SLE, melting point, and heat of fusion measurements, was the DTG-60AH by Shimadzu. The apparatus provides a means to perform simultaneous measurements of thermo-gravimetry and differential thermal analysis. A schematic of the furnace and balance chamber, inlets and outlets, is provided in Figure 1.7. The chamber houses the sample and reference pans, and the parallel guide differential top pan Roberval balance mechanism, for high precision thermo-gravimetric measurements. A swivel configuration reduces vibrational effects of mass measurements.

To reduce the effects of contaminants on DTA results, the apparatus allows for inert gas purging prior to measurements, as well as to the balance chamber, using an FC-60A flow controller by Shimadzu. The measurable range for the DTA and TG components are  $\pm 1$  mV and  $\pm 500$ mg respectively, with a maximum operational temperature of 1500 °C. The mass readability of the TG is 0.1 $\mu$ g. A schematic of the experimental set-up is provided in Figure 1.8. The experimental procedure and results are discussed in further detail in Chapter Six.



**Figure 1.7** A schematic of the furnace and balance chamber of the Shimadzu DTG-60AH used in this work. Process gas inlet (1), cleaning gas inlet (2), line to vacuum pump (3), inert gas inlet (4), balance chamber (5), Roberval balance mechanism (6), heating chamber (7).



**Figure 1.8 Experimental Setup. A- Grant GD120 temperature controller; B- Sonication bath; C- Glass equilibrium vessel; D- Gas tight sampling syringe; E- Shimadzu DTG-60AH apparatus; F- Shimadzu Auto-sampler.**

*References*

- Abildskov, J. and O'Connell, J.P., (2003), Predicting the Solubilities of Complex Chemicals I. Solutes in Different Solvents. *Industrial and Engineering Chemistry Research*, 42, pp.5622–5634.
- Akaike, H., (1974), A new look at the statistical model identification. *IEEE Transactions on Information Technology in Biomedicine*, 19, pp.716–723.
- Anderson, T.F. and Prausnitz, J.M., (1978), Application of the UNIQUAC Equation to Calculation of Multicomponent Phase Equilibria. 1. Vapor-Liquid Equilibria. *Industrial and Engineering Chemical Process Design and Development*, 17, pp.552–561.
- Bhargava, V., Fateen, S.E.K. and Bonilla-Petriciolet, A., (2013), Cuckoo Search: A new nature-inspired optimization method for phase equilibrium calculations. *Fluid Phase Equilibria*, 337, pp.191–200.
- Bondi, A., (1964), van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68, pp.441–451.
- Chen, C.C. and Crafts, P.A., (2006), Correlation and prediction of drug molecule solubility in mixed solvent systems with the Nonrandom Two-Liquid Segment Activity coefficient (NRTL-SAC) model. *Industrial and Engineering Chemistry Research*, 45, pp.4816–4824.
- Chen, C.-C. and Song, Y., (2004), Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Industrial and Engineering Chemistry Research*, 43, pp.8354–8362.
- Claeskens, G. and Hjort, N.L., (2003), The Focused Information Criterion. *Journal of the American Statistical Association*, 98, pp.900–916.
- DDBST Software and Separation Technology GmbH, (2012), Dortmund Data Bank (DDB).
- Diedrichs, A., (2010), *Evaluation und Erweiterung thermodynamischer Modelle zur Vorhersage von Wirkstofflöslichkeiten*. Carl von Ossietzky Universität.
- Diedrichs, A. and Gmehling, J., (2011), Solubility calculation of active pharmaceutical ingredients in alkanes, alcohols, water and their mixtures using various activity coefficient models. *Industrial and Engineering Chemistry Research*, 50, pp.1757–1769.
- Dorigo, M., (1992), *Optimisation, learning and natural algorithms*. Politecnico di Milano.
- Duan, Q., Sorooshian, S., Gupta, H. V. and Gupta, V., (1992), Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28, pp.1015–1031.
- Endoh, K. and Suga, H., (1999), Phase diagram of salt ± water system determined by TG-DTA. *Thermochimica Acta*, 327, pp.133–137.

- Fateen, S.-E.K., Bonilla-Petriciolet, A. and Rangaiah, G.P., (2012), Evaluation of Covariance Matrix Adaptation Evolution Strategy, Shuffled Complex Evolution and Firefly Algorithms for Phase Stability, Phase Equilibrium and Chemical Equilibrium Problems. *Chemical Engineering Research and Design*, 90, pp.2051–2071.
- Flory, P.J., (1942), Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics*, 10, p.51.
- Fredenslund, A., Jones, R.L. and Prausnitz, J.M., (1975), Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21, pp.1086–1099.
- Gandomi, A.H. and Alavi, A.H., (2012), Krill herd: A new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, 17, pp.4831–4845.
- Gmehling, J., Li, J. and Schiller, M., (1993), A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Industrial and Engineering Chemistry Research*, 32, pp.178–193.
- Gmehling, J., Lohmann, J., Jakob, A., Li, J. and Joh, R., (1998), A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension. *Industrial and Engineering Chemistry Research*, 37, pp.4876–4882.
- Gmehling, J., Wittig, R., Lohmann, J. and Joh, R., (2002), A Modified UNIFAC (Dortmund) Model. 4. Revision and Extension. *Industrial and Engineering Chemistry Research*, 41, pp.1678–1688.
- Grensemann, H. and Gmehling, J., (2005), Performance of a Conductor-Like Screening Model for Real Solvents Model in Comparison to Classical Group Contribution Methods. *Industrial and Engineering Chemistry Research*, 44, pp.1610–1624.
- Guggenheim, E.A., (1952), *Mixtures*, Oxford, U.K.
- Hansen, C.M., (2007), Solubility Parameters — An Introduction. In *Hansen Solubility Parameters: A User's Handbook*. pp. 1–24.
- Hansen, E. and Walster, G. W., (2004), *Global Optimization Using Interval Analysis*. 2nd ed., Marcel Dekker, New York.
- Hansen, N. and Ostermeier, A., (2001), Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9, pp.159–195.
- Hildebrand, J.H., (1916), Solubility. *Journal of the American Chemical Society*, 38, pp.1452–1473.
- Hildebrand, J.H. and Scott, R.L., (1962), *Regular solutions*. Prentice-Hall, New Jersey.
- Hildebrand, J.H. and Scott, R.L., (1964), *The solubility of nonelectrolytes*. 3rd ed., Dover, New York.

- Hoffmann, G.R., (1982), Mutagenicity testing in environmental toxicology. *Environmental Science and Technology*, 16, p.560A–574A.
- Holderbaum, T. and Gmehling, J., (1991), PSRK: A Group Contribution Equation of State Based on UNIFAC. *Fluid Phase Equilibria*, 70, pp.251–265.
- Hoy, K.L., (1985), *Tables of Solubility Parameters*, Union Carbide Corp., Research and Development Dept., South Charleston, WV.
- Huggins, M.L., (1941), Solutions of long chain compounds. *The Journal of Chemical Physics*, 9, p.440.
- Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P., (1983), Optimization by Simulated Annealing. *Science*, 220, pp. 671–680.
- Klamt, A., (1995), Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *Journal of Physical Chemistry*, 99(7), pp.2224–2235.
- Klancnik, G., Medved, J., Mrvar, P., (2010), Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) as a method of material investigation. *Materials and Geoenvironment*, 57, pp.127–142.
- Land A H. and Doig, A.G., (1960), An Automatic Method of Solving Discrete Programming Problems. *Econometrica*, 28, pp.497–520.
- Lin, S.-T. and Sandler, S.I., (2002), A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Industrial and Engineering Chemistry Research*, 41, pp.899–913.
- Mohan, R., Lorenz, H. and Myerson, A.S., (2002), Solubility Measurement Using Differential Scanning Calorimetry. *Industrial and Engineering Chemistry Research*, 41, pp.4854–4862.
- Moller, B., Rarey, J. and Ramjugernath, D., (2008), Estimation of the vapour pressure of non-electrolyte organic compounds via group contributions and group interactions. *Journal of Molecular Liquids*, 143, pp.52–63.
- Moller, B., Rarey, J. and Ramjugernath, D., (2014), Extrapolation/interpolation of infinite dilution, activity coefficient as well as liquid and solid solubility between solvents: Part 1. Alkane solvents. *Fluid Phase Equilibria*, 361, pp.69–82.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (a), Model Evaluation for the Prediction of Solubility of Active Pharmaceutical Ingredients (APIs). *Manuscript in preparation*.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (b), A Universal Segment Approach for the Prediction of the Activity Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents. *Fluid Phase Equilibria*, 396, pp.98–110.



- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (c), Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations. *Computers and Chemical Engineering*, 74, pp.75–88.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (d), An Extended UNISAC model for the Prediction of Solubility of Complex Pharmaceutical Ingredients in Non-electrolyte Pure Solvents and Solvent Mixtures. *Manuscript in preparation*.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (e), Experimental solubility for betulin and estrone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . *Journal of Chemical Thermodynamics*, 98, pp. 42–50.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (f), Experimental solubility for diosgenin and estriol in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . Manuscript submitted for publication.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (g), Experimental solubility for prednisolone and hydrocortisone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . Manuscript submitted for publication.
- Mordecai, A., (2003), *Nonlinear Programming: Analysis and Methods*, Dover Publications.
- Neau, S.H., Bhandarkar, S. V and Hellmuth, E.W., (1997), Differential molar heat capacities to test ideal solubility estimations. *Pharmaceutical research*, 14, pp.601–5.
- Nelder, J.A. and Mead, R., (1965), A Simplex Algorithm for Function Minimization. *Computer Journal*, 7, pp.308–313.
- Renon, H. and Prausnitz, J., (1968), Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE journal*, 14, pp.135–144.
- Scatchard, G., (1931), Equilibria in Non-electrolyte Solutions in Relation to the Vapor Pressures and Densities of the Components. *Chemical Reviews*, 8, pp.321–333.
- Smith, J.M., Van Ness, H.C. and Abbott, M.M., (2005), *Introduction to Chemical Engineering Thermodynamics*,
- Srinivas, M. and Rangaiah, G.P., (2007), A study of differential evolution and tabu search for benchmark, phase equilibrium and phase stability problems. *Computers and Chemical Engineering*, 31, pp.760–772.
- Staverman, A.J., (1950), The entropy of high polymer solutions. Generalization of formulae. *Recueil des Travaux Chimiques des Pays-Bas*, 69, pp.163–174.
- Wang, G., Guo, L., Gandomi, A.H., Cao, L., Alavi, A.H., Duan, H. and Li, J., (2013), Lévy-Flight Krill Herd Algorithm. *Mathematical Problems in Engineering*, 2013, pp.1–14.
- Weidlich, U. and Gmehling, J., (1987), A modified UNIFAC Model. 1. Prediction of VLE,  $h^E$ , and  $\gamma^\infty$ . *Industrial and Engineering Chemistry Research*, 26, pp.1372–1381.

- Weir, R.D. and De Loos, T.W. ed., (2005), *Measurement of the Thermodynamic Properties of Multiple Phases*, Elsevier.
- Wilson, G.M., (1964), Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing. *Journal of the American Chemical Society*, 86, pp.127–130.
- Yang, X.-S., (2010), *Nature-inspired Metaheuristic Algorithms*, Luniver Press.
- Yang, X.S. and Deb, S., (2010), Engineering optimisation by cuckoo search. *International Journal of Mathematical Modelling and Numerical Optimisation*, pp.1–17.
- Yu, Q., Ma, X. and Xu, L., (2013), TG-assisted determination of pyrene solubility in 1-pentanol. *Journal of Thermal Analysis and Calorimetry*, 112, pp.1553–1557.
- Zhang, X., Yin, Q., Gong, J., Liu, Z. and Section, E., (2010), Solubility of 5-Amino-N,N'-bis(2,3-dihydroxypropyl)-2,4,6-triiodobenzene- 1,3-dicarboxamide in Ethanol + Water Mixtures. *J Chem. Eng. Data*, 55, pp.2355–2357.

## CHAPTER TWO

### Model Evaluation for the Prediction of Solubility of Active Pharmaceutical Ingredients (APIs)

#### *Abstract*

The predictive capabilities of the various approaches to solubility modelling for complex polycyclic steroidal and triterpene pharmaceuticals, for which experimental data is extremely limited in the literature was analysed. The solutes selected included structurally diverse APIs, such as polycyclic hydrocarbons, solid acids and complex triterpenes. The solvents considered included alkanes, alcohols, ketones, esters, aromatics and water where data was available in the literature. New NRTL-SAC segment area parameters were determined for the selected pharmaceuticals, by the regression of solubility data available in the literature. The UNIFAC, modified UNIFAC (Dortmund), COSMO-RS (OL) and COSMO-SAC activity coefficient predictions were then carried out, based on the availability of group constants and sigma profiles in the literature, in order to assess the predictive capabilities of these models, in terms of solubility. The predictive qualities of the models were assessed based on the percentage deviation. Furthermore the assumptions concerning heat capacity effects (negligibility or approximation) are explored for the components considered here, using benzene and water as a reference solvents. The Staverman-Guggenheim combinatorial term, with the modified UNIFAC (Dortmund) model exhibited a superior predictive capability, with the approximation of  $\Delta_{fus}C_{pi} = 0$  when benzene was used as a solvent. In aqueous systems, the original UNIFAC model with the approximation of  $\Delta_{fus}C_{pi} = \Delta_{fus}S_i$  provided a superior predictive capability. In those cases where NRTL-SAC model predictions were possible, a superior solubility replication was observed. This model is however semi-predictive.

### 2.1 Introduction

The separation and purification of pharmaceutical products or intermediates is arguably the most important and cost intensive process step in the pharmaceutical industry. The method, degree and efficiency of the separation process are generally dictated by the phase behaviour of the solute. Kolar et al. (2002) state that experience shows that over 30% of the efforts of industrial property modellers and experimentalists deal with solvent selection. It is therefore imperative that appropriate solvents are analytically selected, based on a broad information source that may include phase equilibrium experimental data, reliable predictions, experience and solute theory (e.g. structure, bonds and physical properties).

Most often it is not possible to determine the phase behaviour of these systems experimentally, as very small amounts of each pharmaceutical product and the respective intermediates are manufactured in the initial stages of pharmaceutical design and synthesis. Due to this constraint many thermodynamic models have been applied to predict the phase behaviour (solubility), via predictive Gibbs excess energy models. These models include functional group approaches such as UNIFAC (Fredenslund et al., 1975), modified UNIFAC (Dortmund) (Weidlich and Gmehling, 1987), and surface segment approach models such as COSMO-RS (OL) (Grensemann and Gmehling, 2005), COSMO-SAC (Lin and Sandler, 2002) and NRTL-SAC (Chen and Song, 2004).

These models have exhibited varying degrees of success in predicting the solubility of common pharmaceutical compounds with relatively simple molecular structures (Chen and Song, 2004, Gmehling, 1978, Gracin et al., 2002, Mota et al., 2012, and Bouillot et al., 2011).

Gmehling (1978) and Gracin et al. (2002) have explored the ability of the UNIFAC model to predict solid-liquid equilibria. Gmehling (1978) considered relatively simple ring structured solutes such as naphthalene and anthracene. The authors were able to provide good estimates by UNIFAC predictions for the systems considered. Gracin et al. (2002) used the UNIFAC model to predict solubilities of single-ring pharmaceuticals such as ibuprofen and aspirin. The authors concluded that accurate predictions were not achievable, and suggested the use of the UNIFAC model in the cases of the solutes considered in their study, as only a rough guide for initial estimates.

Hahnenkamp et al. (2010) have evaluated and compared the predictive capabilities of the UNIFAC, modified UNIFAC (Dortmund), and COSMO-RS models for systems containing ibuprofen and aspirin. The authors determined that the predictions of the modified UNIFAC (Dortmund) model provided the lowest deviations from the experimental data, when compared to the UNIFAC and COSMO-RS models. Diedrichs and Gmehling (2011) conducted a detailed model comparison but only systems with alcohol, alkane or water as a solvent were considered. Furthermore systems with solute mole fractions greater than 0.1 were excluded in the comparison.

Little work on the abilities of predictive models for the solubility of complex pharmaceuticals such as poly-cyclic aromatics, specifically steroids and triterpenes, is available in the literature. This is mainly due to the lack of experimental data in the literature for such systems. Such data is imperative as this data is required to generate model specific parameters that are usually essential for the application of the majority of the predictive models. It is however important that accurate predictions can be made without an extensive set of experimental data, as this would obviously limit the practicality of the predictive model. Abildskov et al. (2000) have provided some satisfactory predictions for a limited set of steroidal molecules by conducting sensitivity tests on UNIFAC model parameters; however this data is limited and not readily available in the open literature.

In this work the various aforementioned predictive models and methods were tested in order to determine the most accurate predictive method for solubility modelling for each of the solutes considered. The solutes selected are provided in Table 2.1.

The models evaluated in this work were chosen based on the different variations in the approach to solubility modelling (functional group based, segment based, reference solvent based). The differences in combinatorial and residual expressions are distinguished.

The results of the predictions are intended to provide qualitative estimates of solubility data as the predictive models generally yield poor quantitative results in the cases of solid-liquid equilibria. The model performances are correlated with the molecular surface area, molecular weight and functional group diversity. In addition the work Mishra and Yalkowsky (1990) and Neau et al. (1997) is further explored for complex steroidal systems in benzene or water as reference solvents to determine the effect of the assumption of zero or non-zero-approximates for changes in heat capacity upon fusion in systems exhibiting ideal solubility in the solid phase. Neau et al. (1997) showed that the assumption of negligible heat capacity changes can

cause large errors in the calculated solubility during modelling for solutes of melting points exceeding 420 K. However an ideal liquid phase was assumed in their work, hence the effect of the activity coefficient was not considered. Nonetheless, these findings are important, yet many popular process simulators and literature sources still present and employ the older, disproved assumptions when considering pharmaceutical solid-liquid equilibrium design.

The tests of Neau et al. (1997) has been limited to solute melting points of 470 K where the different assumptions for changes in heat capacity can result in deviations from experimental data up to 27 %. These dissimilarities are normally assumed to become more distinct as the difference between the experimental solubility temperature and fusion temperature increase, and is tested in this work.

## 2.2 Theory

The activity coefficient is a measure of the non-ideality of solutions. The parameter is a strong function of composition and of temperature to a degree, but is weakly dependent on pressure at low to moderate pressures. In most cases the activity coefficient is greater than 1, however values below 1 are common in solvating systems, such as solutions of alkanes and polymers. Usually, the degree of dissimilarity between component sizes comprising a mixture, is proportional to the activity coefficient of those components.

The theoretical treatment of solid-liquid phase equilibria, as well as the associated predictive activity coefficient models are provided in section 1.2 and 1.3. To avoid repetition they are omitted here.

## 2.3 Experimental solubility and pure component property data

### 2.3.1 Pure component thermodynamic data

Pure component property data (melting temperature, enthalpy of fusion and heat capacity) of the active pharmaceutical ingredients selected for modelling in this work, is limited in the literature. Bouillot et al. (2011) state that *thermodynamic properties of the solids are scarcely accurate*, when referring to experimentally determined heat of fusion and melting temperature data of pharmaceutical products. Bouillot et al. (2011) have proposed using average values of the available physical property data. In this work the pure component data was used where available, for the calculation of the activity coefficient from solubility measurements.

However, in the case of mestanolone, the enthalpy of fusion was predicted by the method of Chickos and Acree (2003). The literature pure component properties used are presented in Table 2.1, along with molecular masses, van der Waals molecular surface area and functional group diversity. A principal component analysis was conducted on the test set using the solute solubility in an alcohol/alkane and in water, temperature of fusion, enthalpy of fusion and molecular mass as input descriptors. The sample set of components selected were found to be heterogeneous with a minimum of 80% of the data sets described by all combinations of input descriptors.

Table 2.1 Physical properties of the solutes used in this study.

Name	IUPAC name	Formula	CAS-RN	MM (g.mol <sup>-1</sup> )	$f_{us}T_i$ (K) <sup>a</sup>	$\Delta_{fus}H_i$ (J.mol <sup>-1</sup> ) <sup>b</sup>	No. of different functional groups	$q_1$
1,2-benzophenanthrene	Chrysene	C <sub>18</sub> H <sub>12</sub>	218-01-9	228.29	528.15	26135.40	2.00	5.52
1,3,5-triphenylbenzene	1,3,5-triphenylbenzene	C <sub>24</sub> H <sub>18</sub>	612-71-5	306.41	443.15	33377.40	2.00	7.92
2,3-benzindene	9H-fluorene	C <sub>13</sub> H <sub>10</sub>	86-73-7	166.22	389.15	19563.50	3.00	4.22
2-furancarboxylic acid	furan-2-carboxylic acid	C <sub>5</sub> H <sub>4</sub> O <sub>3</sub>	88-14-2	112.085	402.5 (Gracin and Rasmuson, 2002)	22600 (Roux et al., 2004)	3.00	2.892
3-nitrobenzoic acid	3-nitrobenzoic acid	C <sub>7</sub> H <sub>5</sub> NO <sub>4</sub>	121-92-6	167.121	414.15 (Chacko et al., 2005)	21400 (Chacko et al., 2005)	4.00	4.048
9,10-benzophenanthrene	triphenylene	C <sub>18</sub> H <sub>12</sub>	217-59-4	228.29	471.15	25086.00	2.00	5.52
acenaphthene	1,2-dihydroacenaphthylene	C <sub>12</sub> H <sub>10</sub>	83-32-9	154.21	367.15	21522.50	3.00	3.56
adippic acid	hexanedioic acid	C <sub>6</sub> H <sub>10</sub> O <sub>4</sub>	124-04-9	146.143	419 (Roux et al., 2004)	33700.00 (Roux et al., 2004)	2.00	4.608
anthracene	anthracene	C <sub>14</sub> H <sub>10</sub>	120-12-7	178.23	489.60	28840.30	2.00	4.48
ascorbic acid	(R)-3,4-dihydroxy-5-((S)-1,2-dihydroxyethyl)furan-2(5H)-one	C <sub>6</sub> H <sub>8</sub> O <sub>6</sub>	50-81-7	176.126	465.15 (DIPPR)	29200.00	-	-
azelaic acid	nonanedioic acid	C <sub>9</sub> H <sub>16</sub> O <sub>4</sub>	123-99-9	188.224	372.4 (Roux et al., 2004)	30400.00 (Roux et al., 2004)	2.00	6.228
betulin	lup-20(29)-ene-3 $\beta$ ,28-diol	C <sub>30</sub> H <sub>50</sub> O <sub>2</sub>	473-98-3	442.73	528.22 (Zhao and Yan, 2008)	55169.00 (Zhao and Yan, 2008)	6.00	14.55
biphenyl	Biphenyl	C <sub>12</sub> H <sub>10</sub>	92-52-4	154.21	341.95	18580.00	2.00	4.24
citric acid	2-hydroxypropane-1,2,3-tricarboxylic acid	C <sub>6</sub> H <sub>8</sub> O <sub>7</sub>	77-92-9	192.125	426.15 (DIPPR)	26700.00	4.00	5.336
diglycolic acid	2-(carboxymethoxy)acetic acid	C <sub>4</sub> H <sub>6</sub> O <sub>5</sub>	110-99-6	134.089	421.15 (DIPPR)	26400.00	3.00	3.768
diosgenin	(3 $\beta$ ,25R)-spirost-5-en-3-ol	C <sub>27</sub> H <sub>42</sub> O <sub>3</sub>	512-04-9	414.63	474.35 Moodley et al.(2016)	52105.00 Moodley et al.(2016)	7.00	12.68
estrone	(8R,9S,13S,14S)-3-hydroxy-13-methyl-6,7,8,9,11,12,13, 14,15,16-decahydrocyclopenta[a]phenanthren- 17- one	C <sub>18</sub> H <sub>22</sub> O <sub>2</sub>	53-16-7	270.37	527.62 Domańska et al. (2010)	45101.00 Domańska et al. (2010)	9.00	7.53
fluoranthene	Fluoranthene	C <sub>16</sub> H <sub>10</sub>	206-44-0	202.26	380.95	18858.10	2.00	4.72
glutaric acid	pentanedioic acid	C <sub>5</sub> H <sub>8</sub> O <sub>4</sub>	110-94-1	132.116	363.9 (Roux et al., 2004)	21100.00 (Roux et al., 2004)	2.00	4.068
hydrocortisone	(11 $\beta$ )-11,17,21-trihydroypregn-4-ene-3,20-dione	C <sub>21</sub> H <sub>30</sub> O <sub>5</sub>	50-23-7	362.47	485.15 Hagen and G.L. Flynn (1983)	33890.40 Hagen and G.L. Flynn (1983)	-	-



levulinic acid	4-oxopentanoic acid	C <sub>5</sub> H <sub>8</sub> O <sub>3</sub>	123-76-2	116.117	306.15 (CRC Handbook)	9220.00 (CRC Handbook)	3.00	3.792
malic acid	hydroxybutanedioic acid	C <sub>4</sub> H <sub>6</sub> O <sub>5</sub>	6915-15-7	134.089	403.15 (DIPPR)	25300.00 (DIPPR)	4.00	3.8
malonic acid	propanedioic acid	C <sub>3</sub> H <sub>4</sub> O <sub>4</sub>	141-82-2	104.062	407.95 (DIPPR)	25480.00	2.00	2.988
mestanolone	(5 $\alpha$ ,17 $\beta$ )-17-hydroxy-17-methylandrostan-3-one	C <sub>20</sub> H <sub>32</sub> O <sub>2</sub>	521-11-9	304.47	465.65 Hill and Makin (1991)	21504 +/-7600 <sup>c</sup>	6.00	9.54
m-hydroxybenzoic acid	3-hydroxybenzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>	99-06-9	138.123	474.8 (Nordström and Rasmuson, 2006) (b)	35920.00 (Nordström and Rasmuson, 2006) (b)	4.00	3.624
m-terphenyl	1,3-diphenylbenzene	C <sub>18</sub> H <sub>14</sub>	33-76-3	230.31	362.15	24073.50	2.00	6.08
naphthalene	bicyclo[4.4.0]deca-1,3,5,7,9-pentene	C <sub>10</sub> H <sub>8</sub>	91-20-3	128.17	353.35	19110.00	2.00	3.44
o-terphenyl	1,2-Diphenylbenzene	C <sub>18</sub> H <sub>14</sub>	84-15-1	230.31	331.15	17179.10	2.00	6.08
oxalic acid	ethanedioic acid	C <sub>2</sub> H <sub>2</sub> O <sub>4</sub>	144-62-7	90.035	465.26 (Omar and Ulrich, 2006)	58158.00 (Omar and Ulrich, 2006)	1.00	2.448
phenanthrene	Phenanthrene	C <sub>14</sub> H <sub>10</sub>	85-01-8	178.23	369.40	18627.20	2.00	4.48
phtalic acid	benzene-1,2-dicarboxylic acid	C <sub>8</sub> H <sub>6</sub> O <sub>4</sub>	88-99-3	166.133	463.45 (Sabbah and Perez, 1999)	36500.00 (Sabbah and Perez, 1999)	3.00	4.288
p-hydroxybenzoic acid	4-hydroxy benzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>	99-96-7	138.123	487.15 (Gracin and Rasmuson, 2002)	31400.00 (Gracin and Rasmuson, 2002)	4.00	3.624
p-hydroxyphenylacetic acid	2-(4-hydroxyphenyl)acetic acid	C <sub>8</sub> H <sub>8</sub> O <sub>3</sub>	156-38-7	152.15	422.85 (Gracin and Rasmuson, 2002)	28000.00 (Gracin and Rasmuson, 2002)	4.00	4.164
pimelic acid	heptanedioic acid	C <sub>7</sub> H <sub>12</sub> O <sub>4</sub>	111-16-0	160.17	368.2 (Roux et al., 2004)	25200.00 (Roux et al., 2004)	2.00	5.148
prednisolone	(11 $\beta$ )-11,17,21-Trihydroxypregna-1,4-diene-3,20-dione	C <sub>21</sub> H <sub>28</sub> O <sub>5</sub>	50-24-8	360.45	506.00 Cai et al. (1997)	59303.20 Cai et al. (1997)	-	-
p-terphenyl	1,4-diphenylbenzene	C <sub>18</sub> H <sub>14</sub>	92-94-4	230.31	486.15	35476.10	2.00	6.08
pyrene	Pyrene	C <sub>16</sub> H <sub>10</sub>	129-00-0	202.26	422.15	17100.00	2.00	4.72
salicylic acid	2-hydroxybenzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>	69-72-7	138.123	431.35 (Nordström and Rasmuson, 2006) (a))	27090.00 (Nordström and Rasmuson, 2006) (a))	4.00	3.624
suberic acid	octanedioic acid	C <sub>8</sub> H <sub>14</sub> O <sub>4</sub>	505-48-6	174.197	413.2 (Roux et al., 2004)	41800.00 (Roux et al., 2004)	2.00	5.688
succinic acid	butanedioic acid	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	110-15-6	118.089	455.2 (Roux et al., 2004)	34000.00 (Roux et al., 2004)	2.00	3.528

tartaric acid	2,3-dihydroxybutanedioic acid	C <sub>4</sub> H <sub>6</sub> O <sub>6</sub>	133-37-9	150.088	479.15 (DIPPR)	30100.00 (DIPPR)	3.00	4.072
testosterone	(8R,9S,10R,13S,14S,17S)- 17-hydroxy-10,13-dimethyl- 1,2,6,7,8,9,11,12,14,15,16,17-dodecahydrocyclopenta[a]phenanthren-3-one	C <sub>19</sub> H <sub>28</sub> O <sub>2</sub>	58-22-0	288.43	424.40 Kosal et al. (1992)	27946.20 Kosal et al. (1992)	7.00	8.83
1,2-benzophenanthrene	Chrysene	C <sub>18</sub> H <sub>12</sub>	218-01-9	228.29	528.15	26135.40	2.00	5.52
1,3,5-triphenylbenzene	1,3,5-triphenylbenzene	C <sub>24</sub> H <sub>18</sub>	612-71-5	306.41	443.15	33377.40	2.00	7.92
2,3-benzindene	9H-fluorene	C <sub>13</sub> H <sub>10</sub>	86-73-7	166.22	389.15	19563.50	3.00	4.22
2-furancarboxylic acid	furan-2-carboxylic acid	C <sub>5</sub> H <sub>4</sub> O <sub>3</sub>	88-14-2	112.085	402.5 (Gracin and Rasmuson, 2002)	22600 (Roux et al., 2004)	3.00	2.892
3-nitrobenzoic acid	3-nitrobenzoic acid	C <sub>7</sub> H <sub>5</sub> NO <sub>4</sub>	121-92-6	167.121	414.15 (Chacko et al., 2005)	21400 (Chacko et al., 2005)	4.00	4.048

<sup>a</sup> Obtained from the Dortmund Data Bank (2012) unless otherwise stated, <sup>b</sup> Obtained from the Dortmund Data Bank (2012) unless otherwise stated, <sup>c</sup> Predicted by the method of Chickos and Acree with reported uncertainty (2009)

### 2.3.2 *API selection and experimental solubility data*

Solubility data for the APIs selected here (specifically steroids and triterpenes) are extremely limited in the literature. It is therefore important that preliminary predictions of the solubility of these solutes can be made, in order to provide, at the very least, initial estimates for later use in the design and optimization of separation processes such as crystallization.

While all triterpenes examined here contain a similar basic structure, they differ by the number of ester, ketone and alcohol groups in the molecule, which should be the major cause of the dependence of the solubilities on the solvent. The major differences in solubility between the solutes are due to the differences in melting temperature and heat of fusion.

The components and literature sources for the experimental solubility data are presented in Table 2.2.

Table 2.2 Experimental and calculated solubility for various APIs.

Solute	Solvent	T <sup>exp</sup> (K)	ln(x <sub>1</sub> <sup>exp</sup> )	ln(x <sub>1</sub> <sup>calc</sup> )								Reference
				<sup>a</sup> M1	M2	M3	M4	M5	M6	M7	M8	
mestanolone	hexane	298.15	-7.78	-5.27	-7.59	-5.53	-5.96	-2.75	-6.90	-2.88	-2.98	Gharavi et al.(1983)
mestanolone	heptane	298.15	-7.99	-5.51	-7.42	-5.64	-6.22	-2.98	-6.73	-2.95	-3.18	Gharavi et al.(1983)
mestanolone	octane	298.15	-8.02	-5.68	-7.30	-5.71	-6.41	-3.15	-6.61	-3.01	-3.33	Gharavi et al.(1983)
mestanolone	nonane	298.15	-7.71	-5.80	-7.22	-5.76	-6.54	-3.27	-6.53	-3.05	-3.44	Gharavi et al.(1983)
mestanolone	decane	298.15	-7.75	-5.88	-7.15	-5.80	-6.64	-3.37	-6.47	-3.08	-3.51	Gharavi et al.(1983)
mestanolone	undecane	298.15	-7.58	-5.94	-7.10	-5.83	-6.71	-3.43	-6.42	-3.10	-3.57	Gharavi et al.(1983)
mestanolone	dodecane	298.15	-7.68	-5.99	-7.06	-5.85	-6.76	-3.48	-6.38	-3.11	-3.61	Gharavi et al.(1983)
mestanolone	hexadecane	298.15	-7.60	-6.06	-6.93	-5.88	-6.83	-3.58	-6.26	-3.14	-3.65	Gharavi et al.(1983)
mestanolone	pentan-1-ol	298.15	-3.73	-3.41	-3.42	-3.12	-3.00	-2.54	-2.72	-2.37	-2.30	Gharavi et al.(1983)
mestanolone	hexan-1-ol	298.15	-3.65	-3.38	-3.41	-3.12	-3.09	-2.55	-2.72	-2.40	-2.38	Gharavi et al.(1983)
mestanolone	heptan-1-ol	298.15	-3.54	-3.38	-3.41	-3.15	-3.18	-2.57	-2.72	-2.43	-2.45	Gharavi et al.(1983)
mestanolone	octan-1-ol	298.15	-3.51	-3.38	-3.42	-3.17	-3.26	-2.58	-2.74	-2.45	-2.51	Gharavi et al.(1983)
mestanolone	nonan-1-ol	298.15	-3.65	-3.40	-3.44	-3.22	-3.18	-2.60	-2.76	-2.48	-2.51	Gharavi et al.(1983)
mestanolone	decan-1-ol	298.15	-3.82	-3.41	-3.46	-3.27	-3.27	-2.62	-2.78	-2.50	-2.56	Gharavi et al.(1983)
estrone	hexane	303.15	-12.53	-11.80	-14.59	-11.30	-13.73	-5.95	-7.59	-6.62	-7.63	Ruchelman (1967)
estrone	cyclohexane	298.15	-10.52	-11.64	-15.79	-11.82	-13.97	-5.67	-7.35	-6.81	-7.53	Ruchelman (1967)
estrone	benzene	298.15	-6.94	-9.41	-10.47	-9.08	-10.70	-5.39	-6.38	-5.90	-6.33	Ruchelman (1967)
estrone	toluene	298.15	-7.21	-9.34	-10.57	-9.35	-11.17	-5.58	-5.95	-6.06	-6.67	Ruchelman (1967)
estrone	dichloromethane	298.15	-5.13	-9.82	b	-7.34	-7.96	-5.47	-6.23	-5.37	-5.41	Ruchelman (1967)
estrone	chloroform	298.15	-4.51	-7.71	b	-6.73	-5.26	-5.63	-4.77	-5.56	-5.20	Ruchelman (1967)
estrone	tetrahydrofuran	298.15	-3.08	-7.84	b	-7.63	-4.23	-6.39	-6.03	-5.49	-4.06	Ruchelman (1967)
estrone	1,4-dioxane	298.15	-3.61	-9.82	b	-7.71	-5.63	-8.61	-8.34	-5.55	-4.58	Ruchelman (1967)
estrone	diethyl ether	298.15	-6.50	-8.15	-10.28	-8.28	-6.56	-6.06	-8.24	-5.78	-5.39	Ruchelman (1967)
estrone	acetone	298.15	-4.34	-6.63	-6.48	-6.73	-5.79	-6.99	-4.55	-5.33	-4.49	Ruchelman (1967)
estrone	methanol	298.15	-5.50	-13.82	b	-8.03	-6.56	-8.74	-5.07	-5.42	-4.72	Ruchelman (1967)
estrone	ethanol	298.15	-5.10	-7.93	-9.37	-7.71	-6.91	-8.65	-7.32	-5.47	-5.04	Ruchelman (1967)

estrone	water	298.15	-10.68	-20.60	-20.54	-15.73	-14.27	-20.69	-12.00	-8.47	-10.51	Budavari (1989)
prednisolone	2,2,4-trimethylpentane	298.15	-20.63	-	-	-21.46	-28.76	-	-	-10.73	-14.29	Lin and Nash (1993)
prednisolone	chloroform	298.15	-5.45	-	-	-8.80	-4.20	-	-	-7.65	-7.05	Martin et al. (1982)
prednisolone	acetone	298.15	-4.20	-	-	-9.49	-13.82	-	-	-6.88	-6.83	Martin et al. (1982)
prednisolone	ethanol	298.15	-3.69	-	-	-8.91	-8.29	-	-	-6.45	-5.89	Martin et al. (1982)
prednisolone	octan-1-ol	298.15	-5.18	-	-	-11.78	-11.92	-	-	-8.43	-8.77	Yalkowsky et al. (1983)
prednisolone	water	298.15	-8.42	-	-	-18.42	-18.42	-	-	-7.37	-8.30	Lin and Nash (1993)
hydrocortisone	hexane	298.15	-17.39	-	-	-12.60	-16.30	-	-	-5.96	-7.69	Domańska et al.(2010)
hydrocortisone	cyclohexane	298.15	-14.31	-	-	-12.74	-15.98	-	-	-5.95	-7.21	Domańska et al.(2010)
hydrocortisone	tetrachloromethane	298.15	-11.70	-	-	-10.28	-14.33	-	-	-5.26	-6.78	Domańska et al.(2010)
hydrocortisone	benzene	298.15	-9.21	-	-	-7.92	-10.58	-	-	-4.33	-5.10	Domańska et al.(2010)
hydrocortisone	toluene	298.15	-9.46	-	-	-8.50	-11.45	-	-	-4.63	-5.75	Domańska et al.(2010)
hydrocortisone	chloroform	298.15	-5.71	-	-	-3.30	-2.14	-	-	-3.73	-3.02	Domańska et al.(2010)
hydrocortisone	acetone	298.15	-4.89	-	-	-4.21	-3.41	-	-	-3.42	-2.83	Budavari (1989)
hydrocortisone	methyl acetate	298.15	-5.36	-	-	-4.78	-4.22	-	-	-3.59	-3.22	Domańska et al.(2010)
hydrocortisone	ethyl acetate	298.15	-6.07	-	-	-4.98	-4.45	-	-	-3.79	-3.54	Domańska et al.(2010)
hydrocortisone	isopropyl acetate	298.15	-6.81	-	-	-5.36	-4.87	-	-	-4.01	-3.89	Domańska et al.(2010)
hydrocortisone	methanol	298.15	-5.30	-	-	-4.50	-3.45	-	-	-3.14	-2.77	Budavari (1989)
hydrocortisone	ethanol	298.15	-4.42	-	-	-4.56	-3.97	-	-	-3.34	-2.97	Budavari (1989)
hydrocortisone	octan-1-ol	298.15	-5.88	-	-	-6.07	-5.89	-	-	-4.43	-4.44	Domańska et al.(2010)
hydrocortisone	1,2-propanediol	298.15	-4.34	-	-	-5.29	-4.73	-	-	-3.88	-3.58	Domańska et al.(2010)
hydrocortisone	water	298.15	-8.52	-	-	-11.69	-10.20	-	-	-5.34	-7.64	Domańska et al.(2010)
testosterone	hexane	298.15	-8.47	-5.77	-9.31	-6.02	-7.11	-3.39	-3.84	-3.35	-3.95	Lin and Nash (1993)
testosterone	heptane	298.15	-8.18	-5.97	-9.03	-6.11	-7.34	-3.60	-3.81	-3.41	-4.12	Gharavi et al.(1983)
testosterone	octane	298.15	-8.11	-6.11	-8.84	-6.18	-7.50	-3.75	-3.79	-3.46	-4.24	Gharavi et al.(1983)
testosterone	nonane	298.15	-8.08	-6.21	-8.69	-6.22	-7.61	-3.85	-3.78	-3.49	-4.32	Gharavi et al.(1983)
testosterone	decane	298.15	-8.05	-6.28	-8.58	-6.25	-7.69	-3.93	-3.78	-3.51	-4.37	Gharavi et al.(1983)
testosterone	n-undecane	298.15	-7.96	-6.32	-8.49	-6.27	-7.75	-3.98	-3.79	-3.52	-4.41	Gharavi et al.(1983)
testosterone	dodecane	298.15	-7.99	-6.36	-8.41	-6.29	-7.79	-4.02	-3.80	-3.53	-4.44	Gharavi et al.(1983)
testosterone	cyclohexane	298.15	-7.37	-5.30	-8.64	-6.08	-6.58	-2.97	-3.67	-3.31	-3.46	Martin et al. (1982)
testosterone	2,2,4-trimethylpentane	298.15	-8.62	-6.18	-8.83	-5.93	-6.99	-3.81	-7.73	-3.39	-3.96	Bowen et al. (1970)

testosterone	benzene	298.15	-3.10	-2.71	-4.78	-3.62	-3.41	-2.09	-6.34	-2.67	-2.59	Martin et al. (1982)
testosterone	dichloromethane	298.15	-0.58	-2.15	-6.24	-2.16	-1.69	-2.25	-6.17	-2.33	-1.97	Martin et al. (1982)
testosterone	chloroform	298.15	-0.69	-2.12	-2.55	-1.80	-1.29	-2.06	-5.24	-2.35	-1.44	Martin et al. (1982)
testosterone	tetrahydrofuran	298.15	-1.97	-2.82	-3.76	-3.30	-2.27	-2.25	-5.80	-2.56	-2.16	Lin and Nash (1993)
testosterone	ethyl acetate	298.15	-3.30	-3.33	-5.84	-3.37	-3.37	-2.50	-2.30	-2.67	-2.66	Lin and Nash (1993)
testosterone	methanol	298.15	-1.47	-3.44	-4.37	-3.62	-2.72	-2.48	-4.50	-2.58	-2.22	Lin and Nash (1993)
testosterone	propan-1-ol	298.15	-1.97	-3.83	-3.87	-3.07	-2.84	-2.99	-4.23	-2.59	-2.45	Lin and Nash (1993)
testosterone	pentan-1-ol	298.15	-2.30	-3.62	-3.85	-3.05	-3.06	-2.93	-3.43	-2.67	-2.69	Gharavi et al.(1983)
testosterone	hexan-1-ol	298.15	-2.30	-3.60	-3.88	-3.06	-3.15	-2.94	-3.30	-2.71	-2.78	Gharavi et al.(1983)
testosterone	heptan-1-ol	298.15	-2.38	-3.59	-3.93	-3.10	-3.24	-2.95	-3.23	-2.75	-2.86	Gharavi et al.(1983)
testosterone	octan-1-ol	298.15	-2.38	-3.60	-3.98	-3.14	-3.31	-2.97	-3.20	-2.78	-2.92	Gharavi et al.(1983)
testosterone	nonan-1-ol	298.15	-2.60	-3.61	-4.02	-3.16	-3.20	-2.99	-3.20	-2.80	-2.91	Gharavi et al.(1983)
testosterone	decan-1-ol	298.15	-2.80	-3.63	-4.07	-3.22	-3.29	-3.01	-3.20	-2.83	-2.96	Gharavi et al.(1983)
testosterone	1,2-propanediol	298.15	-2.88	-6.63	-5.94	-4.06	-3.27	-5.47	-7.02	-3.11	-2.79	Lin and Nash (1993)
testosterone	1,3-propanediol	298.15	-2.98	-6.64	-6.13	-5.30	-4.85	-5.48	-7.87	-3.88	-4.18	Rytting et al. (1989)
testosterone	water	298.15	-11.04	-17.86	-14.71	-12.40	-9.21	-14.60	-16.73	-5.94	-7.63	Chen et al. (2012)
diosgenin	ethanol	298.15	-6.52	-9.26	-9.76	-8.57	-7.44	b	-12.45	-5.77	-5.36	Chen et al. (2012)
diosgenin	propan-1-ol	298.15	-5.88	-8.69	-9.03	-8.15	-7.52	b	-10.72	-5.77	-5.58	Chen et al. (2012)
diosgenin	butan-1-ol	298.15	-5.26	-8.80	-8.80	-7.98	-7.66	b	-9.70	-5.83	-5.81	Chen et al. (2012)
diosgenin	2-methyl-propan-1-ol	298.15	-6.05	-8.80	-8.80	-7.88	-7.41	b	-9.93	-5.79	-5.63	Chen et al. (2012)
diosgenin	pentan-1-ol	298.15	-5.17	-8.23	-8.52	-7.88	-7.75	b	-9.06	-5.88	-5.97	Chen et al. (2012)
diosgenin	6-methyl-1-heptanol	298.15	-4.82	-8.52	-8.29			b	-7.11	-5.32	-	Chen et al. (2012)
diosgenin	acetone	298.15	-7.01	-7.55	-9.13	-7.39	-7.01	b	-14.63	-5.51	-4.82	Chen et al. (2012)
diosgenin	n-propyl acetate	298.15	-5.94	-7.33	-9.28	-9.43	-7.60	b	-12.20	-5.86	-5.66	Chen et al. (2012)
betulin	methanol	298.15	-8.02	-9.98	-13.82	-18.42	-10.91	-5.98	-15.51	-6.69	-6.54	Cao et al. (2007)
betulin	propan-1-ol	298.15	-6.53	-9.71	-10.17	-9.43	-9.62	-6.73	-12.23	-6.61	-6.89	Cao et al. (2007)
betulin	pentan-1-ol	298.15	-6.55	-9.34	-8.80	-18.42	-9.48	-6.61	-9.79	-6.80	-7.31	Cao et al. (2007)
betulin	methyl formate	298.15	-3.45	-11.51	-13.82	-18.42	-18.42	-6.09	-18.43	-7.87	-8.46	Cao et al. (2007)
betulin	methyl acetate	298.15	-7.13	-10.35	-13.82	-18.42	-9.77	-5.64	-16.44	-6.69	-6.53	Cao et al. (2007)
betulin	chloroform	298.15	-5.43	-8.29	-10.13	-7.07	-5.13	-4.31	-15.56	-6.05	-5.15	Cao et al. (2007)
betulin	cyclohexane	298.15	-9.67	-11.11	-16.17	-18.42	-18.42	-6.31	-9.63	-7.28	-7.38	Cao et al. (2007)

betulin	ethanol	298.15	-7.02	-10.27	-13.82	-9.32	-9.93	-7.05	-14.60	-6.55	-6.67	Cao et al. (2007)
betulin	butan-1-ol	298.15	-6.04	-9.45	-9.84	-9.43	-9.52	-6.63	-10.76	-6.71	-7.14	Cao et al. (2007)
betulin	hexan-1-ol	298.15	-6.89	-9.30	-9.63	-9.43	-9.49	-6.65	-9.12	-6.88	-7.46	Cao et al. (2007)
betulin	ethyl methanoate	298.15	-8.06	-10.26	-13.82	-18.42	-11.31	-5.97	-14.82	-7.26	-7.73	Cao et al. (2007)
betulin	ethyl acetate	298.15	-6.18	-9.92	-13.82	-9.21	-9.45	-5.81	-13.75	-6.69	-6.73	Cao et al. (2007)
betulin	dichloromethane	298.15	-6.92	-7.71	-13.82	-8.80	-8.77	-3.95	-14.55	-5.96	-5.72	Cao et al. (2007)
betulin	acetone	298.15	-6.45	-9.59	-13.82	-9.21	-8.99	-5.34	-15.47	-6.30	-5.99	Cao et al. (2007)
biphenyl	benzene	310.15	-0.67	-0.64	-0.65	-0.67	-0.64	-0.61	-0.62	-0.63	-0.61	McLaughlin and Zainal (1959)
biphenyl	benzene	320.75	-0.43	-0.42	-0.42	-0.43	-0.42	-0.41	-0.41	-0.42	-0.41	McLaughlin and Zainal (1959)
biphenyl	benzene	332.35	-0.20	-0.19	-0.19	-0.19	-0.19	-0.18	-0.18	-0.19	-0.18	McLaughlin and Zainal (1959)
biphenyl	benzene	336.35	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	McLaughlin and Zainal (1959)
o-terphenyl	benzene	301.15	-0.54	-0.57	-0.59	-0.61	-0.58	-0.54	-0.56	-0.58	-0.55	McLaughlin and Zainal (1959)
o-terphenyl	benzene	305.55	-0.44	-0.48	-0.50	-0.51	-0.49	-0.47	-0.48	-0.49	-0.47	McLaughlin and Zainal (1959)
o-terphenyl	benzene	317.95	-0.21	-0.25	-0.25	-0.26	-0.25	-0.24	-0.25	-0.25	-0.24	McLaughlin and Zainal (1959)
o-terphenyl	benzene	323.55	-0.10	-0.14	-0.14	-0.15	-0.14	-0.14	-0.14	-0.14	-0.14	McLaughlin and Zainal (1959)
m-terphenyl	benzene	309.95	-1.26	-1.17	-1.25	-1.30	-1.18	-1.09	-1.15	-1.20	-1.10	McLaughlin and Zainal (1959)
m-terphenyl	benzene	320.15	-0.96	-0.93	-0.98	-1.01	-0.94	-0.87	-0.92	-0.96	-0.88	McLaughlin and Zainal (1959)
m-terphenyl	benzene	333.95	-0.60	-0.61	-0.64	-0.66	-0.62	-0.59	-0.61	-0.63	-0.60	McLaughlin and Zainal (1959)
m-terphenyl	benzene	340.55	-0.45	-0.47	-0.48	-0.50	-0.47	-0.46	-0.47	-0.48	-0.46	McLaughlin and Zainal (1959)
m-terphenyl	benzene	347.35	-0.28	-0.32	-0.33	-0.34	-0.32	-0.32	-0.32	-0.33	-0.32	McLaughlin and Zainal (1959)
m-terphenyl	benzene	350.75	-0.18	-0.25	-0.25	-0.26	-0.25	-0.25	-0.25	-0.25	-0.25	McLaughlin and Zainal (1959)
p-terphenyl	benzene	311.15	-4.95	-4.35	-4.74	-4.80	-4.42	-3.38	-3.73	-3.79	-3.44	McLaughlin and Zainal (1959)
p-terphenyl	benzene	333.35	-4.16	-3.48	-3.83	-3.90	-3.54	-2.82	-3.13	-3.20	-2.88	McLaughlin and Zainal (1959)
p-terphenyl	benzene	337.35	-4.03	-3.33	-3.68	-3.75	-3.40	-2.73	-3.02	-3.09	-2.78	McLaughlin and Zainal (1959)
p-terphenyl	benzene	341.15	-3.89	-3.20	-3.54	-3.61	-3.26	-2.64	-2.93	-3.00	-2.69	McLaughlin and Zainal (1959)
p-terphenyl	benzene	350.75	-3.58	-2.89	-3.20	-3.27	-2.95	-2.43	-2.69	-2.76	-2.47	McLaughlin and Zainal (1959)
naphthalene	benzene	308.15	-0.98	-0.94	-0.96	-0.95	-0.93	-0.88	-0.89	-0.89	-0.87	McLaughlin and Zainal (1959)
naphthalene	benzene	318.15	-0.73	-0.71	-0.72	-0.72	-0.71	-0.67	-0.68	-0.68	-0.67	McLaughlin and Zainal (1959)
naphthalene	benzene	320.55	-0.67	-0.66	-0.67	-0.66	-0.65	-0.62	-0.63	-0.63	-0.62	McLaughlin and Zainal (1959)
naphthalene	benzene	336.35	-0.34	-0.33	-0.33	-0.33	-0.32	-0.32	-0.32	-0.32	-0.32	McLaughlin and Zainal (1959)
naphthalene	benzene	348.95	-0.09	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	McLaughlin and Zainal (1959)

anthracene	benzene	308.95	-4.58	-3.98	-4.22	-4.09	-3.92	-3.11	-3.33	-3.22	-3.07	McLaughlin and Zainal (1959)
anthracene	benzene	315.55	-4.34	-3.74	-3.98	-3.86	-3.69	-2.96	-3.17	-3.07	-2.92	McLaughlin and Zainal (1959)
anthracene	benzene	323.75	-4.06	-3.47	-3.70	-3.58	-3.41	-2.78	-2.99	-2.89	-2.75	McLaughlin and Zainal (1959)
anthracene	benzene	332.75	-3.79	-3.18	-3.41	-3.29	-3.13	-2.59	-2.79	-2.70	-2.56	McLaughlin and Zainal (1959)
anthracene	benzene	343.35	-3.46	-2.87	-3.08	-2.97	-2.82	-2.38	-2.56	-2.48	-2.35	McLaughlin and Zainal (1959)
phenanthrene	benzene	305.15	-1.50	-1.21	-1.28	-1.25	-1.19	-1.10	-1.16	-1.14	-1.09	McLaughlin and Zainal (1959)
phenanthrene	benzene	313.35	-1.26	-1.03	-1.08	-1.07	-1.02	-0.95	-0.99	-0.99	-0.94	McLaughlin and Zainal (1959)
phenanthrene	benzene	314.95	-1.21	-0.99	-1.04	-1.03	-0.99	-0.92	-0.96	-0.96	-0.91	McLaughlin and Zainal (1959)
phenanthrene	benzene	323.35	-0.98	-0.82	-0.86	-0.85	-0.82	-0.77	-0.80	-0.80	-0.77	McLaughlin and Zainal (1959)
phenanthrene	benzene	331.15	-0.78	-0.67	-0.69	-0.69	-0.67	-0.64	-0.66	-0.66	-0.63	McLaughlin and Zainal (1959)
pyrene	benzene	305.55	-2.61	-1.79	-1.99	-1.81	-1.71	-1.51	-1.67	-1.54	-1.46	McLaughlin and Zainal (1959)
pyrene	benzene	331.75	-1.89	-1.27	-1.38	-1.30	-1.23	-1.13	-1.21	-1.15	-1.10	McLaughlin and Zainal (1959)
pyrene	benzene	339.95	-1.66	-1.13	-1.22	-1.15	-1.09	-1.01	-1.08	-1.04	-0.99	McLaughlin and Zainal (1959)
pyrene	benzene	349.35	-1.41	-0.97	-1.04	-0.99	-0.95	-0.89	-0.94	-0.91	-0.87	McLaughlin and Zainal (1959)
pyrene	benzene	357.75	-1.20	-0.84	-0.89	-0.86	-0.82	-0.78	-0.82	-0.79	-0.76	McLaughlin and Zainal (1959)
9,10-benzophenanthrene	benzene	312.55	-4.27	-2.95	-3.34	-3.13	-2.93	-2.37	-2.68	-2.54	-2.37	McLaughlin and Zainal (1959)
9,10-benzophenanthrene	benzene	320.75	-4.01	-2.72	-3.08	-2.89	-2.70	-2.22	-2.51	-2.38	-2.22	McLaughlin and Zainal (1959)
9,10-benzophenanthrene	benzene	336.95	-3.54	-2.30	-2.60	-2.45	-2.28	-1.93	-2.17	-2.07	-1.93	McLaughlin and Zainal (1959)
9,10-benzophenanthrene	benzene	342.55	-3.38	-2.16	-2.44	-2.31	-2.15	-1.83	-2.05	-1.97	-1.84	McLaughlin and Zainal (1959)
9,10-benzophenanthrene	benzene	355.95	-3.00	-1.86	-2.09	-1.99	-1.85	-1.61	-1.79	-1.73	-1.62	McLaughlin and Zainal (1959)
1,2-benzophenanthrene	benzene	308.75	-6.17	-3.90	-4.35	-4.10	-3.84	-2.90	-3.28	-3.10	-2.88	McLaughlin and Zainal (1959)
1,2-benzophenanthrene	benzene	318.95	-5.74	-3.59	-4.02	-3.78	-3.53	-2.72	-3.08	-2.91	-2.70	McLaughlin and Zainal (1959)
1,2-benzophenanthrene	benzene	333.75	-5.26	-3.16	-3.57	-3.35	-3.12	-2.47	-2.79	-2.64	-2.45	McLaughlin and Zainal (1959)
1,2-benzophenanthrene	benzene	345.35	-4.84	-2.86	-3.24	-3.04	-2.82	-2.28	-2.58	-2.44	-2.26	McLaughlin and Zainal (1959)
2,3-benzindene	benzene	306.75	-1.83	-1.55	-1.63	-1.61	-1.53	-1.38	-1.44	-1.42	-1.36	McLaughlin and Zainal (1959)
2,3-benzindene	benzene	327.55	-1.28	-1.09	-1.13	-1.13	-1.08	-1.00	-1.03	-1.03	-0.99	McLaughlin and Zainal (1959)
2,3-benzindene	benzene	331.55	-1.17	-1.01	-1.04	-1.04	-1.00	-0.93	-0.96	-0.96	-0.92	McLaughlin and Zainal (1959)
2,3-benzindene	benzene	342.55	-0.90	-0.79	-0.81	-0.82	-0.78	-0.74	-0.76	-0.77	-0.74	McLaughlin and Zainal (1959)
2,3-benzindene	benzene	345.95	-0.82	-0.73	-0.75	-0.75	-0.72	-0.69	-0.70	-0.71	-0.68	McLaughlin and Zainal (1959)
acenaphthene	benzene	303.75	-1.71	-1.42	-1.48	-1.47	-1.42	-1.30	-1.33	-1.33	-1.28	McLaughlin and Zainal (1959)
acenaphthene	benzene	314.55	-1.37	-1.14	-1.18	-1.18	-1.14	-1.06	-1.08	-1.08	-1.05	McLaughlin and Zainal (1959)



acenaphthene	benzene	336.35	-0.75	-0.63	-0.64	-0.64	-0.63	-0.60	-0.61	-0.62	-0.60	McLaughlin and Zainal (1959)
acenaphthene	benzene	342.55	-0.59	-0.50	-0.50	-0.51	-0.49	-0.48	-0.48	-0.49	-0.48	McLaughlin and Zainal (1959)
fluoranthene	benzene	317.95	-1.53	-1.13	-1.22	-1.15	-1.09	-1.03	-1.11	-1.06	-1.00	McLaughlin and Zainal (1959)
fluoranthene	benzene	329.15	-1.20	-0.90	-0.96	-0.92	-0.87	-0.84	-0.88	-0.86	-0.81	McLaughlin and Zainal (1959)
fluoranthene	benzene	337.55	-0.96	-0.74	-0.77	-0.75	-0.72	-0.69	-0.73	-0.71	-0.68	McLaughlin and Zainal (1959)
fluoranthene	benzene	350.35	-0.63	-0.50	-0.52	-0.51	-0.49	-0.48	-0.50	-0.49	-0.48	McLaughlin and Zainal (1959)
1,3,5-triphenylbenzene	benzene	298.35	-3.51	-3.43	-4.05	-	-	-2.75	-3.25	-	-	McLaughlin and Zainal (1959)
1,3,5-triphenylbenzene	benzene	301.75	-3.35	-3.30	-3.90	-	-	-2.67	-3.15	-	-	McLaughlin and Zainal (1959)
1,3,5-triphenylbenzene	benzene	313.55	-3.03	-2.88	-3.40	-	-	-2.40	-2.81	-	-	McLaughlin and Zainal (1959)
1,3,5-triphenylbenzene	benzene	319.35	-2.88	-2.69	-3.18	-	-	-2.27	-2.65	-	-	McLaughlin and Zainal (1959)
1,3,5-triphenylbenzene	benzene	332.55	-2.53	-2.30	-2.69	-	-	-2.00	-2.30	-	-	McLaughlin and Zainal (1959)
1,3,5-triphenylbenzene	benzene	339.75	-2.34	-2.11	-2.45	-	-	-1.86	-2.12	-	-	McLaughlin and Zainal (1959)
ethanedioic acid	water	278.15	-4.68	-10.62	-10.62	-9.70	-9.86	-8.24	-8.24	-7.32	-7.48	Apelblat and Manzurola (1987)
ethanedioic acid	water	283.15	-4.49	-10.18	-9.82	-9.26	-9.41	-7.98	-7.61	-7.06	-7.21	Apelblat and Manzurola (1987)
ethanedioic acid	water	288.15	-4.21	-9.76	-9.36	-8.83	-8.98	-7.72	-7.32	-6.80	-6.94	Apelblat and Manzurola (1987)
ethanedioic acid	water	293.15	-3.98	-9.35	-8.91	-8.43	-8.57	-7.46	-7.03	-6.55	-6.68	Apelblat and Manzurola (1987)
ethanedioic acid	water	298.15	-3.77	-8.95	-8.47	-8.03	-8.16	-7.21	-6.74	-6.30	-6.43	Apelblat and Manzurola (1987)
ethanedioic acid	water	303.15	-3.62	-8.57	-8.05	-7.65	-7.77	-6.96	-6.45	-6.06	-6.18	Apelblat and Manzurola (1987)
ethanedioic acid	water	308.15	-3.38	-8.19	-7.64	-7.28	-7.40	-6.72	-6.17	-5.82	-5.93	Apelblat and Manzurola (1987)
ethanedioic acid	water	313.15	-3.19	-7.84	-7.24	-6.93	-7.03	-6.48	-5.89	-5.58	-5.68	Apelblat and Manzurola (1987)
ethanedioic acid	water	318.15	-3.01	-7.49	-6.84	-6.58	-6.68	-6.24	-5.61	-5.35	-5.44	Apelblat and Manzurola (1987)
ethanedioic acid	water	323.15	-2.83	-7.15	-6.46	-6.25	-6.34	-6.01	-5.33	-5.12	-5.20	Apelblat and Manzurola (1987)
ethanedioic acid	water	328.15	-2.67	-6.82	-6.09	-5.92	-6.00	-5.78	-5.06	-4.90	-4.97	Apelblat and Manzurola (1987)
ethanedioic acid	water	333.15	-2.55	-6.50	-5.73	-5.61	-5.68	-5.55	-4.79	-4.68	-4.74	Apelblat and Manzurola (1987)
ethanedioic acid	water	338.15	-2.32	-6.19	-5.37	-5.30	-5.37	-5.33	-4.52	-4.46	-4.51	Apelblat and Manzurola (1987)
propanedioic acid	water	278.15	-1.78	-4.31	-3.57	-3.05	-3.22	-3.59	-2.93	-2.49	-2.59	Apelblat and Manzurola (1987)
propanedioic acid	water	283.15	-1.72	-4.09	-3.32	-2.88	-3.02	-3.44	-2.75	-2.38	-2.46	Apelblat and Manzurola (1987)
propanedioic acid	water	288.15	-1.66	-3.89	-3.08	-2.71	-2.83	-3.28	-2.57	-2.27	-2.32	Apelblat and Manzurola (1987)
propanedioic acid	water	293.15	-1.59	-3.68	-2.85	-2.56	-2.60	-3.12	-2.40	-2.16	-2.14	Apelblat and Manzurola (1987)
propanedioic acid	water	298.15	-1.53	-3.47	-2.63	-2.40	-2.43	-2.96	-2.23	-2.05	-2.02	Apelblat and Manzurola (1987)
propanedioic acid	water	303.15	-1.47	-3.27	-2.42	-2.26	-2.26	-2.80	-2.07	-1.96	-1.89	Apelblat and Manzurola (1987)

propanedioic acid	water	308.15	-1.40	-3.06	-2.22	-2.09	-2.09	-2.64	-1.91	-1.86	-1.77	Apelblat and Manzurola (1987)
propanedioic acid	water	313.15	-1.33	-2.86	-2.03	-1.93	-1.93	-2.47	-1.77	-1.77	-1.65	Apelblat and Manzurola (1987)
propanedioic acid	water	318.15	-1.28	-2.65	-1.85	-1.78	-1.78	-2.30	-1.62	-1.67	-1.53	Apelblat and Manzurola (1987)
propanedioic acid	water	323.15	-1.23	-2.38	-1.68	-1.63	-1.63	-2.12	-1.49	-1.58	-1.41	Apelblat and Manzurola (1987)
propanedioic acid	water	328.15	-1.16	-2.26	-1.52	-1.48	-1.48	-1.96	-1.36	-1.49	-1.29	Apelblat and Manzurola (1987)
propanedioic acid	water	333.15	-1.10	-2.30	-1.37	-1.34	-1.34	-1.80	-1.24	-1.39	-1.18	Apelblat and Manzurola (1987)
propanedioic acid	water	338.15	-1.04	-2.94	-1.23	-1.20	-1.21	-1.66	-1.12	-1.30	-1.07	Apelblat and Manzurola (1987)
butanedioic acid	water	278.15	-5.22	-7.10	-5.86	-5.37	-5.37	-5.78	-4.55	-4.24	-4.08	Apelblat and Manzurola (1987)
butanedioic acid	water	283.15	-4.99	-6.84	-5.52	-5.11	-5.11	-5.62	-4.31	-4.07	-3.92	Apelblat and Manzurola (1987)
butanedioic acid	water	288.15	-4.76	-6.59	-5.19	-4.87	-4.86	-5.46	-4.07	-3.91	-3.76	Apelblat and Manzurola (1987)
butanedioic acid	water	293.15	-4.52	-6.35	-4.87	-4.63	-4.62	-5.30	-3.84	-3.76	-3.60	Apelblat and Manzurola (1987)
butanedioic acid	water	298.15	-4.31	-6.11	-4.55	-4.40	-4.38	-5.14	-3.61	-3.60	-3.45	Apelblat and Manzurola (1987)
butanedioic acid	water	303.15	-4.14	-5.87	-4.20	-4.29	-4.25	-4.98	-3.38	-3.45	-3.36	Apelblat and Manzurola (1987)
butanedioic acid	water	308.15	-3.95	-5.65	-3.94	-4.07	-4.03	-4.82	-3.16	-3.30	-3.21	Apelblat and Manzurola (1987)
butanedioic acid	water	313.15	-3.74	-5.42	-3.65	-3.86	-3.81	-4.66	-2.94	-3.16	-3.06	Apelblat and Manzurola (1987)
butanedioic acid	water	318.15	-3.52	-5.20	-3.37	-3.65	-3.60	-4.50	-2.73	-3.02	-2.91	Apelblat and Manzurola (1987)
butanedioic acid	water	323.15	-3.34	-4.99	-3.10	-3.45	-3.39	-4.34	-2.53	-2.88	-2.76	Apelblat and Manzurola (1987)
butanedioic acid	water	328.15	-3.16	-4.79	-2.84	-3.26	-3.19	-4.18	-2.34	-2.74	-2.62	Apelblat and Manzurola (1987)
butanedioic acid	water	333.15	-3.02	-4.67	-2.59	-3.08	-2.99	-4.02	-2.16	-2.61	-2.48	Apelblat and Manzurola (1987)
butanedioic acid	water	338.15	-2.81	-5.69	-2.37	-2.90	-2.80	-3.86	-1.99	-2.48	-2.34	Apelblat and Manzurola (1987)
hexanedioic acid	water	278.15	-6.70	-7.48	-5.34	-4.50	-3.61	-6.52	-4.33	-3.59	-2.79	Apelblat and Manzurola (1987)
hexanedioic acid	water	283.15	-6.55	-7.21	-4.95	-4.24	-3.38	-6.34	-4.02	-3.42	-2.66	Apelblat and Manzurola (1987)
hexanedioic acid	water	288.15	-6.25	-6.96	-4.57	-3.99	-3.17	-6.16	-3.70	-3.25	-2.52	Apelblat and Manzurola (1987)
hexanedioic acid	water	293.15	-6.04	-6.70	-4.18	-3.75	-2.96	-5.98	-3.38	-3.09	-2.40	Apelblat and Manzurola (1987)
hexanedioic acid	water	298.15	-5.79	-6.46	-3.80	-3.52	-2.77	-5.81	-3.07	-2.93	-2.27	Apelblat and Manzurola (1987)
hexanedioic acid	water	303.15	-5.57	-6.22	-3.42	-3.30	-2.59	-5.63	-2.77	-2.78	-2.16	Apelblat and Manzurola (1987)
hexanedioic acid	water	308.15	-5.35	-5.98	-3.06	-3.09	-2.41	-5.45	-2.50	-2.63	-2.04	Apelblat and Manzurola (1987)
hexanedioic acid	water	313.15	-5.08	-5.75	-2.72	-2.89	-2.24	-5.27	-2.25	-2.48	-1.93	Apelblat and Manzurola (1987)
hexanedioic acid	water	318.15	-4.78	-5.52	-2.41	-2.70	-2.08	-5.09	-2.04	-2.34	-1.82	Apelblat and Manzurola (1987)
hexanedioic acid	water	323.15	-4.53	-5.22	-2.15	-2.52	-1.92	-4.91	-1.84	-2.21	-1.72	Apelblat and Manzurola (1987)
hexanedioic acid	water	328.15	-4.19	-4.73	-1.92	-2.35	-1.77	-4.72	-1.67	-2.07	-1.62	Apelblat and Manzurola (1987)

hexanedioic acid	water	333.15	-3.85	-3.82	-1.73	-2.19	-1.62	-4.50	-1.51	-1.95	-1.52	Apelblat and Manzurola (1987)
hexanedioic acid	water	338.15	-3.71	-2.71	-1.57	-2.03	-1.48	-4.17	-1.37	-1.82	-1.43	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	278.15	-2.19	-3.74	-2.65	-2.95	-2.64	-3.05	-2.14	-2.43	-2.09	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	283.15	-2.11	-3.50	-2.38	-2.78	-2.46	-2.87	-1.95	-2.31	-1.97	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	288.15	-2.03	-3.25	-2.15	-2.61	-2.29	-2.70	-1.79	-2.20	-1.86	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	293.15	-1.93	-3.01	-1.96	-2.45	-2.13	-2.52	-1.64	-2.09	-1.76	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	298.15	-1.85	-2.78	-1.81	-2.31	-1.98	-2.35	-1.51	-1.98	-1.65	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	303.15	-1.78	-2.56	-1.70	-2.16	-1.84	-2.18	-1.41	-1.88	-1.55	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	308.15	-1.71	-2.35	-1.62	-2.02	-1.70	-2.02	-1.34	-1.77	-1.46	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	313.15	-1.62	-2.15	-1.55	-1.89	-1.58	-1.87	-1.29	-1.67	-1.37	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	318.15	-1.56	-1.97	-1.52	-1.76	-1.46	-1.73	-1.27	-1.57	-1.28	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	323.15	-1.46	-1.80	-1.45	-1.64	-1.35	-1.59	-1.24	-1.47	-1.19	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	328.15	-1.39	-1.64	-1.40	-1.52	-1.24	-1.47	-1.21	-1.38	-1.11	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	333.15	-1.31	-1.51	-1.33	-1.41	-1.14	-1.35	-1.17	-1.28	-1.03	Apelblat and Manzurola (1987)
hydroxybutanedioic acid	water	338.15	-1.22	-1.55	-1.25	-1.29	-1.04	-1.24	-1.11	-1.19	-0.96	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	278.15	-1.70	-4.05	-3.07	-2.91	-2.57	-3.08	-2.36	-2.32	-1.68	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	283.15	-1.66	-3.79	-2.74	-2.76	-2.47	-2.89	-2.16	-2.24	-1.68	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	288.15	-2.40	-3.52	-2.49	-2.65	-1.68	-2.71	-1.97	-2.16	-1.03	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	293.15	-2.31	-3.26	-2.28	-2.52	-1.56	-2.54	-1.80	-2.08	-0.95	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	298.15	-2.21	-3.02	-2.13	-2.40	-1.45	-2.38	-1.66	-1.99	-0.88	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	303.15	-2.12	-2.79	-2.04	-2.28	-1.34	-2.24	-1.58	-1.92	-0.82	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	308.15	-2.02	-2.58	-1.98	-2.16	-1.26	-2.10	-1.54	-1.84	-0.78	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	313.15	-1.90	-2.39	-1.93	-2.05	-1.22	-1.97	-1.53	-1.76	-0.77	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	318.15	-1.82	-2.22	-1.94	-1.95	-1.14	-1.86	-1.60	-1.69	-0.75	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	323.15	-1.78	-2.06	-1.98	-1.85	-1.04	-1.75	-1.72	-1.61	-0.70	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	328.15	-1.61	-1.91	-1.86	-1.74	-1.07	-1.64	-1.65	-1.54	-0.77	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	333.15	-1.56	-1.77	-1.86	-1.65	-0.99	-1.54	-1.70	-1.47	-0.75	Apelblat and Manzurola (1987)
2-hydroxypropane-1,2,3-tricarboxylic acid	water	338.15	-1.50	-1.85	-1.82	-1.55	-0.93	-1.45	-1.68	-1.39	-0.74	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	278.15	-1.76	-5.34	-	-4.34	-3.35	-3.90	-	-3.16	-2.15	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	283.15	-1.70	-5.10	-	-4.13	-3.22	-3.75	-	-3.06	-2.07	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	288.15	-2.11	-4.87	-	-3.94	-3.05	-3.61	-	-2.96	-1.99	Apelblat and Manzurola (1987)

2,3-dihydroxybutanedioic acid	water	293.15	-2.06	-4.65	-	-3.76	-2.92	-3.46	-	-2.86	-1.92	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	298.15	-2.02	-4.43	-	-3.58	-2.81	-3.32	-	-2.76	-1.86	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	303.15	-1.99	-4.21	-	-3.42	-2.70	-3.18	-	-2.66	-1.80	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	308.15	-1.94	-4.00	-	-3.26	-2.59	-3.04	-	-2.57	-1.75	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	313.15	-1.90	-3.79	-	-3.11	-2.49	-2.91	-	-2.48	-1.70	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	318.15	-1.84	-3.59	-	-2.97	-2.39	-2.77	-	-2.39	-1.65	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	323.15	-1.81	-3.42	-	-2.83	-2.29	-2.65	-	-2.31	-1.61	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	328.15	-1.65	-3.43	-	-2.70	-2.25	-2.52	-	-2.23	-1.59	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	333.15	-1.61	-6.34	-	-2.57	-2.15	-2.40	-	-2.15	-1.54	Apelblat and Manzurola (1987)
2,3-dihydroxybutanedioic acid	water	338.15	-1.57	-6.19	-	-2.45	-2.06	-2.28	-	-2.07	-1.51	Apelblat and Manzurola (1987)
2-hydroxybenzoic acid	water	339.15	-6.51	-5.11	-0.68	-2.53	-3.68	-4.78	-0.49	-2.19	-3.23	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	334.65	-6.64	-5.25	-0.92	-2.72	-3.87	-4.90	-0.73	-2.35	-3.42	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	330.65	-6.81	-5.37	-1.13	-2.90	-4.04	-5.01	-0.92	-2.49	-3.59	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	328.15	-6.94	-5.45	-1.25	-3.01	-4.15	-5.07	-1.02	-2.58	-3.69	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	327.15	-6.98	-5.48	-1.29	-3.05	-4.19	-5.10	-1.05	-2.62	-3.72	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	325.15	-7.07	-5.55	-1.38	-3.14	-4.27	-5.15	-1.12	-2.69	-3.80	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	324.15	-7.15	-5.58	-1.42	-3.18	-4.32	-5.18	-1.16	-2.73	-3.84	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	323.15	-7.21	-5.62	-1.46	-3.22	-4.36	-5.21	-1.19	-2.77	-3.88	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	320.15	-7.31	-5.72	-1.58	-3.36	-4.48	-5.29	-1.28	-2.87	-3.98	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	318.15	-7.30	-5.79	-1.65	-3.44	-4.57	-5.34	-1.34	-2.95	-4.05	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	317.15	-7.47	-5.83	-1.69	-3.49	-4.61	-5.36	-1.37	-2.98	-4.09	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	307.15	-7.73	-6.16	-2.23	-3.93	-5.02	-5.62	-1.70	-3.34	-4.43	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	298.15	-8.01	-6.59	-3.03	-4.34	-5.40	-5.86	-2.09	-3.66	-4.73	Apelblat and Manzurola (1989)
2-hydroxybenzoic acid	water	283.15	-8.52	-7.17	-4.34	-5.04	-6.04	-6.25	-3.16	-4.19	-5.21	Apelblat and Manzurola (1989)
ascorbic acid	water	280.15	-3.92	-	-	-4.09	-3.62	-	-	-3.08	-2.67	Apelblat and Manzurola (1989)
ascorbic acid	water	281.15	-3.88	-	-	-4.05	-3.58	-	-	-3.05	-2.65	Apelblat and Manzurola (1989)
ascorbic acid	water	283.15	-3.84	-	-	-3.97	-3.51	-	-	-3.01	-2.61	Apelblat and Manzurola (1989)
ascorbic acid	water	285.15	-3.79	-	-	-3.89	-3.44	-	-	-2.96	-2.57	Apelblat and Manzurola (1989)
ascorbic acid	water	289.15	-3.66	-	-	-3.73	-3.30	-	-	-2.88	-2.49	Apelblat and Manzurola (1989)
ascorbic acid	water	294.15	-3.51	-	-	-3.55	-3.13	-	-	-2.77	-2.39	Apelblat and Manzurola (1989)
ascorbic acid	water	298.15	-3.40	-	-	-3.41	-3.00	-	-	-2.69	-2.32	Apelblat and Manzurola (1989)

ascorbic acid	water	302.15	-3.26	-	-	-3.28	-2.88	-	-	-2.61	-2.25	Apelblat and Manzurola (1989)
ascorbic acid	water	307.15	-3.12	-	-	-3.12	-2.73	-	-	-2.51	-2.16	Apelblat and Manzurola (1989)
ascorbic acid	water	309.15	-3.08	-	-	-3.06	-2.68	-	-	-2.47	-2.13	Apelblat and Manzurola (1989)
ascorbic acid	water	312.15	-3.00	-	-	-2.97	-2.59	-	-	-2.41	-2.08	Apelblat and Manzurola (1989)
ascorbic acid	water	316.15	-2.88	-	-	-2.85	-2.49	-	-	-2.34	-2.01	Apelblat and Manzurola (1989)
ascorbic acid	water	317.15	-2.87	-	-	-2.82	-2.46	-	-	-2.32	-2.00	Apelblat and Manzurola (1989)
ascorbic acid	water	318.15	-2.84	-	-	-2.80	-2.44	-	-	-2.30	-1.98	Apelblat and Manzurola (1989)
ascorbic acid	water	320.85	-2.79	-	-	-2.72	-2.37	-	-	-2.25	-1.94	Apelblat and Manzurola (1989)
ascorbic acid	water	321.65	-2.75	-	-	-2.70	-2.35	-	-	-2.24	-1.93	Apelblat and Manzurola (1989)
ascorbic acid	water	323.15	-2.71	-	-	-2.66	-2.31	-	-	-2.21	-1.90	Apelblat and Manzurola (1989)
ascorbic acid	water	323.55	-2.70	-	-	-2.65	-2.30	-	-	-2.20	-1.90	Apelblat and Manzurola (1989)
ascorbic acid	water	324.15	-2.68	-	-	-2.63	-2.29	-	-	-2.19	-1.89	Apelblat and Manzurola (1989)
ascorbic acid	water	325.65	-2.61	-	-	-2.59	-2.25	-	-	-2.17	-1.86	Apelblat and Manzurola (1989)
ascorbic acid	water	327.15	-2.60	-	-	-2.56	-2.22	-	-	-2.14	-1.84	Apelblat and Manzurola (1989)
ascorbic acid	water	327.25	-2.58	-	-	-2.55	-2.22	-	-	-2.14	-1.84	Apelblat and Manzurola (1989)
ascorbic acid	water	328.15	-2.56	-	-	-2.53	-2.20	-	-	-2.12	-1.83	Apelblat and Manzurola (1989)
ascorbic acid	water	333.18	-2.51	-	-	-2.41	-2.08	-	-	-2.03	-1.75	Apelblat and Manzurola (1989)
ascorbic acid	water	335.15	-2.48	-	-	-2.36	-2.03	-	-	-2.00	-1.72	Apelblat and Manzurola (1989)
ascorbic acid	water	338.15	-2.42	-	-	-2.29	-1.96	-	-	-1.95	-1.68	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	279.15	-5.49	-6.20	-	-3.82	-4.41	-5.65	-	-3.19	-3.75	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	280.15	-5.49	-6.30	-	-3.49	-4.06	-5.74	-	-2.90	-3.44	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	283.65	-5.45	-6.30	-	-3.34	-3.92	-5.76	-	-2.79	-3.32	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	284.15	-5.38	-6.40	-	-3.32	-3.90	-5.85	-	-2.77	-3.31	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	285.15	-5.36	-6.48	-	-3.28	-3.86	-5.94	-	-2.74	-3.27	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	289.15	-5.33	-6.43	-	-3.11	-3.69	-5.92	-	-2.62	-3.14	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	294.15	-5.06	-6.34	-	-2.91	-3.48	-5.86	-	-2.46	-2.97	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	298.15	-4.83	-6.29	-	-2.75	-3.32	-5.83	-	-2.34	-2.83	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	301.15	-4.73	-6.26	-	-2.64	-3.19	-5.82	-	-2.25	-2.72	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	302.15	-4.61	-6.32	-	-2.60	-3.15	-5.88	-	-2.22	-2.68	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	307.15	-4.38	-6.21	-	-2.41	-2.94	-5.80	-	-2.07	-2.49	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	309.15	-4.34	-6.21	-	-2.33	-2.85	-5.81	-	-2.01	-2.42	Apelblat and Manzurola (1989)

furan-2-carboxylic acid	water	311.15	-4.17	-6.21	-	-2.26	-2.77	-5.82	-	-1.95	-2.34	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	315.15	-3.77	-6.12	-	-2.11	-2.59	-5.75	-	-1.84	-2.18	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	317.15	-3.66	-6.10	-	-2.04	-2.50	-5.76	-	-1.78	-2.10	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	319.80	-3.47	-6.06	-	-1.95	-2.38	-5.73	-	-1.71	-2.00	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	323.15	-2.97	-5.99	-	-1.84	-2.22	-5.67	-	-1.62	-1.86	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	325.15	-2.71	-5.97	-	-1.77	-2.11	-5.66	-	-1.57	-1.78	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	327.15	-2.63	-5.95	-	-1.70	-2.02	-5.65	-	-1.51	-1.70	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	327.65	-2.60	-5.99	-	-1.69	-2.00	-5.69	-	-1.50	-1.68	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	328.15	-2.40	-6.01	-	-1.67	-1.96	-5.74	-	-1.49	-1.66	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	330.65	-2.05	-5.95	-	-1.59	-1.82	-5.69	-	-1.42	-1.57	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	333.20	-1.94	-5.66	-	-1.51	-1.70	-5.64	-	-1.36	-1.47	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	334.55	-1.94	-5.60	-	-1.47	-1.65	-5.65	-	-1.33	-1.42	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	336.15	-1.61	-4.27	-	-1.43	-1.56	-5.63	-	-1.29	-1.37	Apelblat and Manzurola (1989)
furan-2-carboxylic acid	water	341.15	-1.34	-4.09	-	-1.28	-1.35	-5.45	-	-1.17	-1.20	Apelblat and Manzurola (1989)
pentanedioic acid	water	279.15	-2.61	-3.63	-2.08	-1.84	-1.33	-3.23	-1.78	-1.62	-1.26	Apelblat and Manzurola (1989)
pentanedioic acid	water	281.15	-2.55	-3.66	-2.09	-1.87	-1.36	-3.29	-1.80	-1.65	-1.29	Apelblat and Manzurola (1989)
pentanedioic acid	water	284.15	-2.46	-3.63	-2.04	-1.86	-1.37	-3.29	-1.77	-1.65	-1.29	Apelblat and Manzurola (1989)
pentanedioic acid	water	286.15	-2.30	-3.62	-2.04	-1.88	-1.40	-3.34	-1.78	-1.67	-1.31	Apelblat and Manzurola (1989)
pentanedioic acid	water	289.15	-2.13	-3.56	-1.98	-1.86	-1.41	-3.32	-1.74	-1.66	-1.31	Apelblat and Manzurola (1989)
pentanedioic acid	water	291.15	-2.05	-3.49	-1.97	-1.87	-1.43	-3.34	-1.74	-1.68	-1.33	Apelblat and Manzurola (1989)
pentanedioic acid	water	293.15	-1.98	-3.43	-1.95	-1.88	-1.45	-3.36	-1.73	-1.69	-1.34	Apelblat and Manzurola (1989)
pentanedioic acid	water	298.15	-1.83	-3.26	-1.79	-1.78	-1.38	-3.19	-1.61	-1.62	-1.29	Apelblat and Manzurola (1989)
pentanedioic acid	water	302.15	-1.67	-3.12	-1.67	-1.71	-1.35	-3.06	-1.52	-1.56	-1.25	Apelblat and Manzurola (1989)
pentanedioic acid	water	304.15	-1.61	-3.02	-1.64	-1.70	-1.35	-3.04	-1.50	-1.56	-1.25	Apelblat and Manzurola (1989)
pentanedioic acid	water	307.15	-1.53	-2.90	-1.57	-1.66	-1.32	-2.94	-1.44	-1.53	-1.23	Apelblat and Manzurola (1989)
pentanedioic acid	water	310.15	-1.45	-2.74	-1.49	-1.61	-1.29	-2.82	-1.38	-1.49	-1.20	Apelblat and Manzurola (1989)
pentanedioic acid	water	312.15	-1.39	-2.64	-1.46	-1.59	-1.28	-2.76	-1.35	-1.47	-1.19	Apelblat and Manzurola (1989)
pentanedioic acid	water	316.15	-1.27	-2.48	-1.34	-1.49	-1.22	-2.53	-1.25	-1.40	-1.14	Apelblat and Manzurola (1989)
pentanedioic acid	water	317.15	-1.25	-2.35	-1.34	-1.50	-1.23	-2.51	-1.25	-1.40	-1.15	Apelblat and Manzurola (1989)
pentanedioic acid	water	318.15	-1.24	-2.27	-1.34	-1.50	-1.24	-2.51	-1.25	-1.41	-1.15	Apelblat and Manzurola (1989)
pentanedioic acid	water	320.55	-1.18	-2.72	-1.28	-1.46	-1.21	-2.35	-1.20	-1.37	-1.12	Apelblat and Manzurola (1989)

pentanedioic acid	water	323.15	-1.12	-2.59	-1.21	-1.40	-1.16	-2.18	-1.14	-1.32	-1.09	Apelblat and Manzurola (1989)
pentanedioic acid	water	327.65	-1.04	-2.70	-1.07	-1.27	-1.06	-2.12	-1.02	-1.21	-1.00	Apelblat and Manzurola (1989)
pentanedioic acid	water	329.15	-0.92	-2.60	-1.04	-1.24	-1.06	-2.03	-0.99	-1.19	-0.99	Apelblat and Manzurola (1989)
pentanedioic acid	water	332.15	-0.87	-3.04	-0.96	-1.16	-0.99	-1.96	-0.92	-1.11	-0.93	Apelblat and Manzurola (1989)
pentanedioic acid	water	334.15	-0.86	-2.92	-0.91	-1.11	-0.95	-1.82	-0.87	-1.07	-0.90	Apelblat and Manzurola (1989)
pentanedioic acid	water	336.15	-0.74	-4.31	-0.85	-1.06	-0.92	-1.93	-0.83	-1.02	-0.86	Apelblat and Manzurola (1989)
pentanedioic acid	water	342.15	-0.62	-4.04	-0.68	-0.85	-0.74	-1.42	-0.66	-0.82	-0.71	Apelblat and Manzurola (1989)
heptanedioic acid	water	279.15	-6.19	-5.72	-2.71	-2.59	-2.13	-5.36	-2.21	-2.21	-1.81	Apelblat and Manzurola (1989)
heptanedioic acid	water	281.15	-6.16	-5.61	-2.54	-2.50	-2.05	-5.29	-2.09	-2.14	-1.76	Apelblat and Manzurola (1989)
heptanedioic acid	water	284.15	-6.01	-5.47	-2.31	-2.37	-1.95	-5.18	-1.93	-2.05	-1.68	Apelblat and Manzurola (1989)
heptanedioic acid	water	289.15	-5.48	-5.26	-1.99	-2.17	-1.79	-5.00	-1.71	-1.89	-1.56	Apelblat and Manzurola (1989)
heptanedioic acid	water	291.15	-5.35	-5.14	-1.88	-2.09	-1.73	-4.92	-1.64	-1.83	-1.52	Apelblat and Manzurola (1989)
heptanedioic acid	water	293.15	-5.17	-5.03	-1.78	-2.01	-1.67	-4.84	-1.56	-1.77	-1.47	Apelblat and Manzurola (1989)
heptanedioic acid	water	297.15	-4.91	-4.79	-1.61	-1.86	-1.55	-4.68	-1.43	-1.66	-1.38	Apelblat and Manzurola (1989)
heptanedioic acid	water	298.15	-4.88	-4.61	-1.57	-1.83	-1.52	-4.64	-1.40	-1.63	-1.36	Apelblat and Manzurola (1989)
heptanedioic acid	water	302.15	-4.70	-4.16	-1.42	-1.69	-1.41	-4.46	-1.28	-1.52	-1.28	Apelblat and Manzurola (1989)
heptanedioic acid	water	304.15	-4.61	-3.04	-1.35	-1.62	-1.36	-4.35	-1.23	-1.47	-1.23	Apelblat and Manzurola (1989)
heptanedioic acid	water	307.15	-4.49	-2.45	-1.25	-1.53	-1.28	-4.14	-1.15	-1.39	-1.17	Apelblat and Manzurola (1989)
heptanedioic acid	water	310.15	-4.33	-2.08	-1.17	-1.44	-1.21	-3.30	-1.07	-1.32	-1.11	Apelblat and Manzurola (1989)
heptanedioic acid	water	313.15	-3.98	-1.87	-1.09	-1.35	-1.13	-2.51	-1.01	-1.24	-1.05	Apelblat and Manzurola (1989)
heptanedioic acid	water	316.15	-3.31	-3.87	-1.04	-1.26	-1.05	-1.99	-0.94	-1.17	-1.00	Apelblat and Manzurola (1989)
heptanedioic acid	water	318.15	-3.07	-3.69	-1.01	-1.20	-1.01	-1.75	-0.91	-1.12	-0.96	Apelblat and Manzurola (1989)
heptanedioic acid	water	320.65	-2.84	-3.82	-0.97	-1.14	-0.95	-3.48	-0.86	-1.06	-0.91	Apelblat and Manzurola (1989)
heptanedioic acid	water	323.15	-2.56	-3.64	-0.92	-1.07	-0.89	-3.26	-0.82	-1.00	-0.86	Apelblat and Manzurola (1989)
heptanedioic acid	water	327.98	-1.97	-3.97	-0.83	-0.94	-0.79	-3.08	-0.73	-0.89	-0.77	Apelblat and Manzurola (1989)
heptanedioic acid	water	333.15	-1.67	-3.79	-0.72	-0.81	-0.68	-2.67	-0.64	-0.77	-0.68	Apelblat and Manzurola (1989)
heptanedioic acid	water	337.65	-1.34	-4.94	-0.62	-0.70	-0.60	-2.70	-0.55	-0.67	-0.60	Apelblat and Manzurola (1989)
heptanedioic acid	water	342.15	-1.12	-4.80	-0.53	-0.59	-0.52	-2.39	-0.47	-0.57	-0.52	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	283.65	-8.17	-10.18	-9.22	-6.10	-6.10	-8.84	-7.85	-4.73	-4.67	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	296.15	-7.25	-9.52	-8.38	-5.40	-5.42	-8.41	-7.26	-4.27	-4.22	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	298.15	-7.18	-9.41	-8.25	-5.30	-5.32	-8.35	-7.16	-4.20	-4.14	Apelblat and Manzurola (1989)

benzene-1,2-dicarboxylic acid	water	302.65	-6.97	-9.19	-7.93	-5.06	-5.09	-8.19	-6.92	-4.05	-3.98	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	307.15	-6.84	-8.97	-7.68	-4.84	-4.86	-8.05	-6.73	-3.89	-3.81	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	309.15	-6.73	-8.87	-7.56	-4.74	-4.76	-7.98	-6.64	-3.82	-3.74	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	311.15	-6.68	-8.78	-7.44	-4.64	-4.66	-7.92	-6.54	-3.75	-3.66	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	315.15	-6.50	-8.60	-7.19	-4.44	-4.46	-7.79	-6.35	-3.62	-3.51	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	317.15	-6.41	-8.51	-7.07	-4.35	-4.36	-7.73	-6.25	-3.55	-3.44	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	318.15	-6.41	-8.46	-7.01	-4.30	-4.31	-7.69	-6.21	-3.52	-3.40	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	319.15	-6.31	-8.41	-6.95	-4.26	-4.26	-7.66	-6.16	-3.49	-3.36	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	323.15	-6.19	-8.24	-6.71	-4.07	-4.06	-7.54	-5.96	-3.35	-3.21	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	323.15	-6.11	-8.20	-6.71	-4.07	-4.06	-7.54	-5.96	-3.35	-3.21	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	327.15	-6.00	-8.02	-6.47	-3.89	-3.87	-7.41	-5.76	-3.23	-3.06	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	334.15	-5.74	-6.79	-6.05	-3.59	-3.53	-7.19	-5.41	-3.01	-2.80	Apelblat and Manzurola (1989)
benzene-1,2-dicarboxylic acid	water	338.15	-5.52	-6.64	-5.81	-3.42	-3.34	-7.08	-5.20	-2.88	-2.65	Apelblat and Manzurola (1989)
4-hydroxy benzoic acid	water	278.15	-8.16	-8.96	-6.58	-6.04	-5.83	-7.45	-4.94	-4.53	-4.27	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	284.15	-7.80	-8.68	-6.18	-5.73	-5.54	-7.29	-4.63	-4.35	-4.09	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	289.15	-7.53	-8.46	-5.85	-5.49	-5.30	-7.16	-4.35	-4.20	-3.94	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	293.15	-7.30	-8.28	-5.59	-5.30	-5.12	-7.06	-4.11	-4.08	-3.82	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	298.15	-7.04	-8.07	-5.25	-5.07	-4.89	-6.93	-3.79	-3.93	-3.67	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	303.15	-6.78	-7.87	-4.90	-4.84	-4.67	-6.81	-3.45	-3.78	-3.52	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	307.15	-6.53	-7.71	-4.61	-4.67	-4.50	-6.71	-3.15	-3.67	-3.39	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	312.15	-6.29	-7.51	-4.22	-4.46	-4.28	-6.58	-2.77	-3.53	-3.25	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	316.15	-6.00	-7.36	-3.88	-4.29	-4.11	-6.49	-2.46	-3.42	-3.13	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	321.15	-5.73	-7.17	-3.41	-4.09	-3.91	-6.37	-2.12	-3.28	-2.98	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	326.15	-5.43	-6.99	-2.91	-3.89	-3.70	-6.25	-1.88	-3.14	-2.83	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	330.15	-5.12	-6.85	-2.52	-3.74	-3.54	-6.15	-1.74	-3.04	-2.71	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	335.15	-4.87	-6.68	-2.12	-3.55	-3.33	-6.03	-1.56	-2.91	-2.57	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	340.15	-4.59	-6.52	-1.85	-3.37	-3.14	-5.91	-1.38	-2.78	-2.43	Apelblat and Manzurola (1997)
4-hydroxy benzoic acid	water	345.15	-4.35	-6.32	-1.64	-3.19	-2.94	-5.79	-1.21	-2.65	-2.29	Apelblat and Manzurola (1997)
3-hydroxybenzoic acid	water	342.15	-4.78	-6.69	-1.97	-3.58	-3.28	-6.13	-1.51	-3.00	-2.59	Nordström and Rasmuson (2006) (b)
3-hydroxybenzoic acid	water	352.75	-4.26	-6.28	-1.47	-3.16	-2.82	-5.84	-1.10	-2.69	-2.25	Nordström and Rasmuson (2006) (b)
3-hydroxybenzoic acid	water	357.75	-3.82	-6.11	-1.22	-2.97	-2.60	-5.70	-0.85	-2.55	-2.10	Nordström and Rasmuson (2006) (b)



3-hydroxybenzoic acid	water	366.45	-3.17	-5.72	-1.19	-2.66	-2.25	-5.46	-0.82	-2.31	-1.86	Nordström and Rasmuson (2006) (b)
3-hydroxybenzoic acid	water	371.45	-2.77	-5.66	-1.82	-2.49	-2.06	-5.33	-2.00	-2.18	-1.73	Nordström and Rasmuson (2006) (b)
3-hydroxybenzoic acid	water	382.95	-2.16	-4.92	-1.58	-2.12	-1.69	-5.00	-1.96	-1.89	-1.47	Nordström and Rasmuson (2006) (b)
3-hydroxybenzoic acid	water	407.15	-1.43	-5.43	-1.00	-1.44	-1.11	-4.28	-1.58	-1.33	-1.03	Nordström and Rasmuson (2006) (b)
3-hydroxybenzoic acid	water	474.45	0.00	-3.92	-0.69	-0.01	-0.01	-0.88	-0.54	-0.01	-0.01	Nordström and Rasmuson (2006) (b)
2-(4-hydroxyphenyl)acetic acid	water	283.15	-5.66	-7.85	-5.22	-4.09	-4.75	-7.09	-4.20	-3.31	-3.91	Gracin and Rasmuson (2002)
2-(4-hydroxyphenyl)acetic acid	water	288.15	-5.43	-7.64	-4.84	-3.86	-4.52	-6.95	-3.80	-3.15	-3.74	Gracin and Rasmuson (2002)
2-(4-hydroxyphenyl)acetic acid	water	293.15	-5.21	-7.45	-4.42	-3.64	-4.28	-6.82	-3.34	-2.98	-3.56	Gracin and Rasmuson (2002)
2-(4-hydroxyphenyl)acetic acid	water	298.15	-4.94	-7.26	-3.93	-3.42	-4.06	-6.68	-2.78	-2.83	-3.39	Gracin and Rasmuson (2002)
3-nitrobenzoic acid	water	278.15	-8.66	-	-	-3.78	-4.91	-	-	-3.07	-4.20	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	280.65	-8.64	-	-	-3.67	-4.81	-	-	-2.99	-4.12	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	283.15	-8.46	-	-	-3.56	-4.71	-	-	-2.90	-4.03	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	285.35	-8.46	-	-	-3.46	-4.62	-	-	-2.82	-3.96	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	288.15	-8.30	-	-	-3.34	-4.50	-	-	-2.73	-3.86	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	290.15	-8.28	-	-	-3.25	-4.42	-	-	-2.66	-3.79	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	293.15	-8.15	-	-	-3.12	-4.29	-	-	-2.56	-3.68	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	295.15	-8.09	-	-	-3.04	-4.21	-	-	-2.49	-3.61	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	298.15	-7.97	-	-	-2.91	-4.09	-	-	-2.39	-3.49	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	300.15	-7.90	-	-	-2.82	-4.01	-	-	-2.33	-3.41	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	303.15	-7.80	-	-	-2.70	-3.88	-	-	-2.23	-3.29	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	305.15	-7.70	-	-	-2.61	-3.79	-	-	-2.17	-3.20	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	308.15	-7.57	-	-	-2.49	-3.66	-	-	-2.08	-3.07	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	310.65	-7.48	-	-	-2.39	-3.56	-	-	-2.00	-2.95	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	313.15	-7.36	-	-	-2.29	-3.45	-	-	-1.93	-2.83	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	315.65	-7.27	-	-	-2.20	-3.33	-	-	-1.86	-2.70	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	318.15	-7.17	-	-	-2.11	-3.22	-	-	-1.79	-2.57	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	320.95	-7.03	-	-	-2.00	-3.10	-	-	-1.71	-2.41	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	323.15	-6.93	-	-	-1.93	-3.00	-	-	-1.65	-2.28	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	326.05	-6.77	-	-	-1.83	-2.86	-	-	-1.58	-2.10	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	328.15	-6.67	-	-	-1.76	-2.76	-	-	-1.53	-1.97	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	333.15	-6.44	-	-	-1.60	-2.53	-	-	-1.41	-1.66	Manzurola and Apelblat (2002)

3-nitrobenzoic acid	water	335.65	-6.29	-	-	-1.53	-2.40	-	-	-1.35	-1.51	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	338.15	-6.17	-	-	-1.46	-2.28	-	-	-1.30	-1.38	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	341.15	-5.95	-	-	-1.38	-2.14	-	-	-1.23	-1.24	Manzurola and Apelblat (2002)
3-nitrobenzoic acid	water	343.15	-5.85	-	-	-1.32	-2.04	-	-	-1.19	-1.16	Manzurola and Apelblat (2002)
octanedioic acid	water	280.15	-9.26	-9.69	-6.69	-6.36	-5.87	-8.65	-5.59	-5.27	-4.75	Apelblat and Manzurola (1990)
octanedioic acid	water	285.85	-8.96	-9.32	-6.15	-5.96	-5.49	-8.40	-5.13	-4.98	-4.47	Apelblat and Manzurola (1990)
octanedioic acid	water	291.85	-8.64	-8.95	-5.57	-5.55	-5.10	-8.13	-4.63	-4.67	-4.18	Apelblat and Manzurola (1990)
octanedioic acid	water	298.15	-8.29	-8.58	-4.96	-5.14	-4.70	-7.86	-4.08	-4.35	-3.87	Apelblat and Manzurola (1990)
octanedioic acid	water	305.65	-7.98	-8.15	-4.20	-4.66	-4.23	-7.54	-3.39	-3.97	-3.50	Apelblat and Manzurola (1990)
octanedioic acid	water	307.65	-7.90	-8.03	-3.99	-4.53	-4.10	-7.46	-3.20	-3.87	-3.41	Apelblat and Manzurola (1990)
octanedioic acid	water	313.15	-7.77	-7.73	-3.41	-4.18	-3.76	-7.23	-2.72	-3.59	-3.14	Apelblat and Manzurola (1990)
octanedioic acid	water	317.65	-7.63	-7.49	-2.95	-3.91	-3.49	-7.04	-2.40	-3.36	-2.93	Apelblat and Manzurola (1990)
octanedioic acid	water	323.35	-7.45	-7.18	-2.47	-3.56	-3.15	-6.81	-2.09	-3.08	-2.67	Apelblat and Manzurola (1990)
octanedioic acid	water	325.85	-7.35	-7.06	-2.30	-3.40	-3.00	-6.71	-1.97	-2.96	-2.56	Apelblat and Manzurola (1990)
octanedioic acid	water	334.65	-6.90	-6.67	-1.82	-2.89	-2.52	-6.35	-1.60	-2.55	-2.20	Apelblat and Manzurola (1990)
octanedioic acid	water	337.15	-6.75	-6.54	-1.72	-2.75	-2.40	-6.25	-1.52	-2.44	-2.11	Apelblat and Manzurola (1990)
octanedioic acid	water	343.25	-6.51	-7.08	-1.57	-2.43	-2.11	-5.99	-1.36	-2.18	-1.89	Apelblat and Manzurola (1990)
octanedioic acid	water	349.65	-6.15	-6.79	-0.96	-2.12	-1.85	-5.73	-0.90	-1.93	-1.69	Apelblat and Manzurola (1990)
nonanedioic acid	water	280.15	-9.73	-7.79	-3.95	-3.81	-3.29	-7.38	-3.16	-3.21	-2.64	Apelblat and Manzurola (1990)
nonanedioic acid	water	285.85	-9.42	-7.50	-3.19	-3.42	-2.93	-7.16	-2.45	-2.89	-2.38	Apelblat and Manzurola (1990)
nonanedioic acid	water	291.85	-9.06	-7.18	-2.42	-3.02	-2.56	-6.94	-1.97	-2.56	-2.13	Apelblat and Manzurola (1990)
nonanedioic acid	water	298.15	-8.68	-6.86	-1.92	-2.61	-2.21	-6.70	-1.68	-2.24	-1.89	Apelblat and Manzurola (1990)
nonanedioic acid	water	307.85	-8.16	-6.38	-1.47	-2.06	-1.77	-6.34	-1.33	-1.83	-1.58	Apelblat and Manzurola (1990)
nonanedioic acid	water	316.15	-8.00	-2.16	-1.17	-1.68	-1.47	-6.02	-1.09	-1.54	-1.35	Apelblat and Manzurola (1990)
nonanedioic acid	water	325.75	-7.64	-1.64	-0.97	-1.33	-1.19	-5.65	-0.89	-1.24	-1.10	Apelblat and Manzurola (1990)
nonanedioic acid	water	330.15	-7.27	-1.42	-0.94	-1.18	-1.07	-1.88	-0.84	-1.11	-0.99	Apelblat and Manzurola (1990)
nonanedioic acid	water	334.65	-6.98	-1.19	-0.90	-1.04	-0.96	-1.39	-0.79	-0.98	-0.89	Apelblat and Manzurola (1990)
nonanedioic acid	water	338.15	-6.66	-1.04	-0.86	-0.93	-0.87	-1.16	-0.74	-0.89	-0.80	Apelblat and Manzurola (1990)
nonanedioic acid	water	344.15	-6.63	-0.83	-0.66	-0.76	-0.72	-0.86	-0.59	-0.73	-0.67	Apelblat and Manzurola (1990)
nonanedioic acid	water	347.65	-6.48	-0.71	-0.39	-0.66	-0.63	-0.73	-0.42	-0.64	-0.59	Apelblat and Manzurola (1990)
nonanedioic acid	water	353.05	-6.26	-0.55	1.71	-0.51	-0.50	-0.55	-0.35	-0.50	-0.47	Apelblat and Manzurola (1990)

nonanedioic acid	water	357.65	-6.36	-0.41	-0.28	-0.39	-0.38	-0.41	-0.29	-0.38	-0.36	Apelblat and Manzurola (1990)
4-oxopentanoic acid	water	280.15	-1.71	-0.35	-0.33	-0.34	-0.33	-0.33	-0.31	-0.32	-0.31	Apelblat and Manzurola (1990)
4-oxopentanoic acid	water	283.15	-1.54	-0.30	-0.29	-0.29	-0.29	-0.29	-0.28	-0.28	-0.27	Apelblat and Manzurola (1990)
4-oxopentanoic acid	water	287.65	-1.28	-0.41	-0.23	-0.23	-0.23	-0.23	-0.22	-0.22	-0.22	Apelblat and Manzurola (1990)
4-oxopentanoic acid	water	293.95	-0.98	-0.28	-0.15	-0.15	-0.15	-0.15	-0.14	-0.15	-0.14	Apelblat and Manzurola (1990)
4-oxopentanoic acid	water	298.15	-0.80	-1.04	-0.10	-0.10	-0.10	-0.10	-0.09	-0.10	-0.09	Apelblat and Manzurola (1990)
4-oxopentanoic acid	water	303.25	-0.66	-0.94	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	278.25	-3.31	-4.72	-4.23	-	-	-3.92	-3.43	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	284.15	-3.06	-4.43	-3.92	-	-	-3.70	-3.21	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	288.75	-2.68	-4.17	-3.69	-	-	-3.53	-3.03	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	293.75	-2.59	-3.91	-3.44	-	-	-3.34	-2.84	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	296.65	-2.50	-3.63	-3.30	-	-	-3.22	-2.73	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	301.25	-2.27	-3.41	-3.08	-	-	-3.04	-2.57	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	310.65	-2.05	-2.96	-2.64	-	-	-2.65	-2.24	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	313.15	-1.97	-2.82	-2.53	-	-	-2.54	-2.15	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	321.15	-1.83	-2.49	-2.20	-	-	-2.21	-1.90	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	323.15	-1.79	-2.36	-2.12	-	-	-2.13	-1.84	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	327.65	-1.67	-2.17	-1.95	-	-	-1.96	-1.71	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	332.65	-1.58	-1.99	-1.77	-	-	-1.80	-1.57	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	338.15	-1.51	-1.81	-1.59	-	-	-1.63	-1.42	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	343.15	-1.43	-1.65	-1.44	-	-	-1.49	-1.30	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	348.15	-1.37	-1.63	-1.30	-	-	-1.37	-1.18	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	353.05	-1.32	-1.38	-1.17	-	-	-1.25	-1.07	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	357.15	-1.26	-1.86	-1.07	-	-	-1.16	-0.99	-	-	Apelblat and Manzurola (1990)
2-(carboxymethyloxy)acetic acid	water	361.35	-1.22	-1.17	-0.98	-	-	-1.08	-0.90	-	-	Apelblat and Manzurola (1990)

<sup>a</sup> Model definitions are provided in Table 2,3, and 4, <sup>b</sup> Failure to converge

## 2.4 Results and Discussion

### 2.4.1 Modelling and model comparison

In order to quantify the quality of the predictions for the various models tested, a Percentage Deviation (PD) was defined:

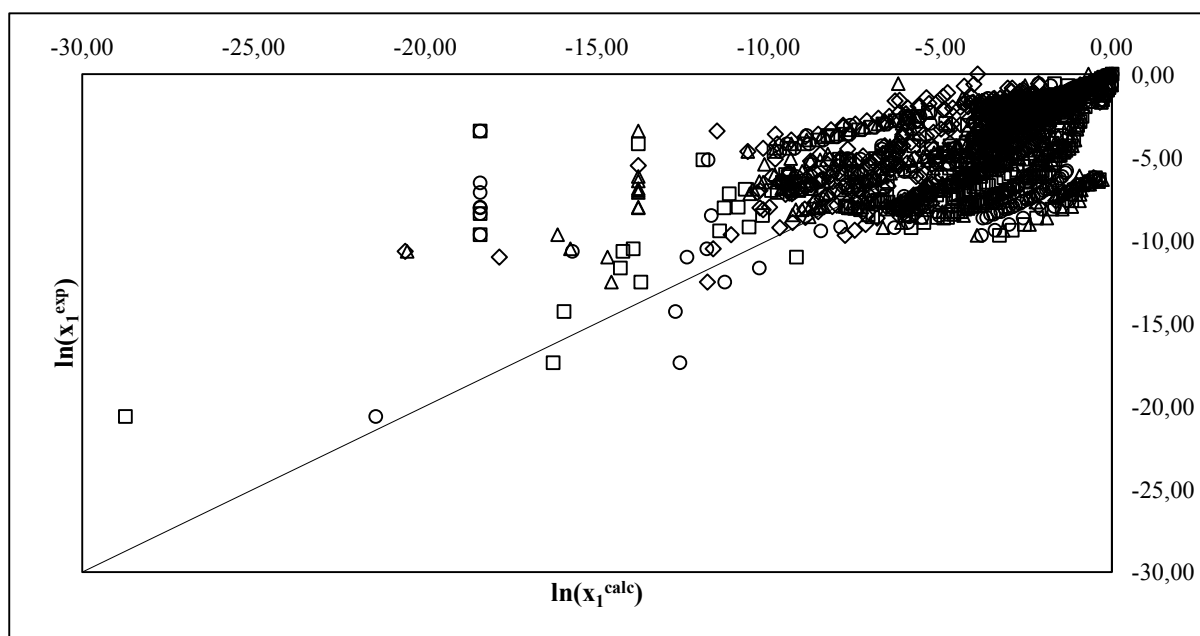
$$PD = 100 \left( \frac{\sum_{i=1}^N (x_i^{pred} - x_i^{exp})^2}{\sum_{i=1}^N (x_i^{exp} - \bar{x}^{exp})^2} \right)^{1/2} \quad (2.1)$$

Where  $x_i^{pred}$ ,  $x_i^{exp}$  are the calculated and experimental solute compositions and  $N$  is the total number of data points considered.  $\bar{x}^{exp}$  is the average experimental composition for a particular set.

Essentially all predictive models require certain information about the solute in order to be utilized. For the UNIFAC-based models, group volume and surface as well as group interaction parameters represent the functional groups and their energetic interactions. The COSMO-based models require so-called sigma-profiles that characterize the shielding charge distribution as well as the cavity volume and surface. In this work the Oldenburg version of COSMO-RS (COSMO-RS (OL)) was used. Unfortunately group interaction parameters and segment area parameters were not available for all groups or solutes and solvents, respectively, considered for prediction. Hence not all solubilities could be described by all of the predictive methods. These systems are indicated by a dash in Table 2.2. The sigma profiles of the solutes, used in the COSMO-RS(OL) and COSMO-SAC methods were determined by Gaussian 03 calculations with the hybrid density function theory type B3LYP and basis sets 6-311G(d,p) (Mu et al., 2007). These profiles were obtained from the Dortmund Data Bank software package (2012).

The mean percentage deviations between experimental data and the model predictions are presented in Table 2.2. These results are presented graphically in Figure 2.1 for ease of comparison. It is clear that based on prediction power alone, there isn't a strictly superior activity coefficient model. In the majority of the systems tested, all the predictive models tend to underestimate the solubility. Furthermore very large discrepancies are apparent for sparingly soluble solute-solvent mixtures. In Figures 2.2-2.4 an attempt to correlate the prediction

capabilities of each model considered with; molecular weight, van der Waals molecular surface area and functional group diversity. The van der Waals molecular surface area were determined using the method of Bondi (1964). It is confirmed from the presented figures for benzene as a solvent that virtually no correlation of these parameters to solubility exists in the systems considered here. A similar result was obtained in aqueous systems.



**Figure 2.1 Comparison of the natural logarithms of experimental and model calculated solubility composition ( $x_1$ ).  $\diamond$ , UNIFAC,  $\Delta$ , modified UNIFAC (Dortmund),  $\circ$ , COSMO-RS(OL),  $\square$ , COSMOSAC.**

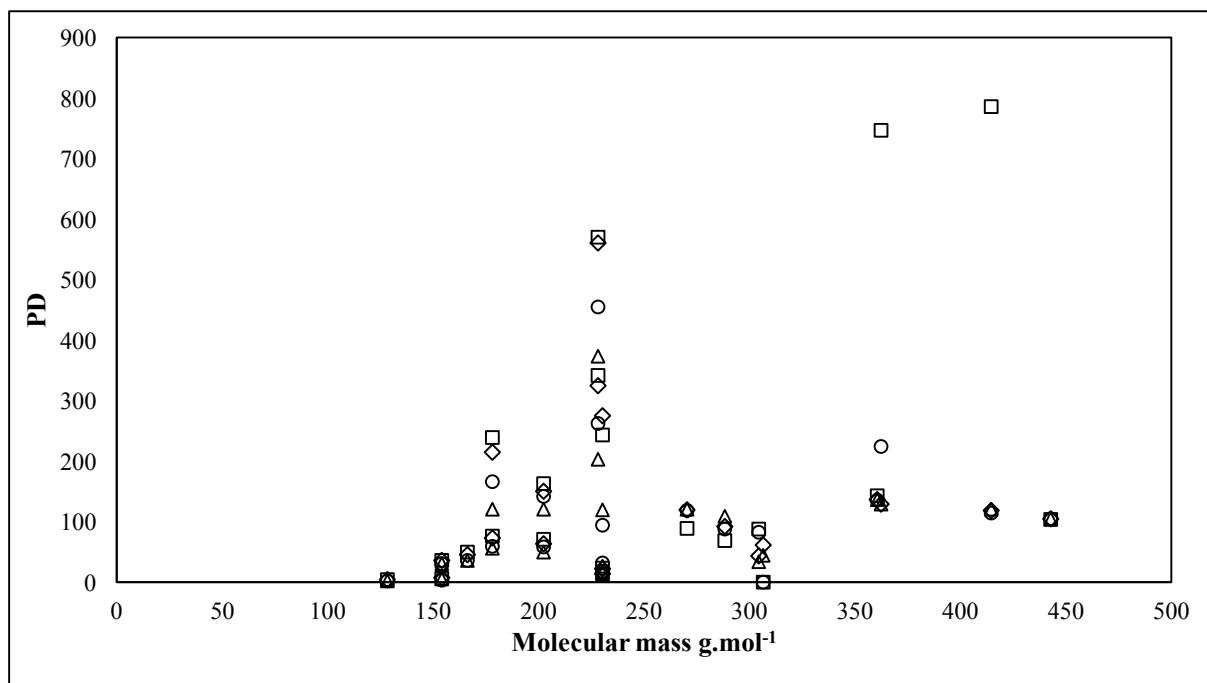


Figure 2.2 Correlation of model percentage deviations with molecular mass of solute.  $\diamond$ , UNIFAC,  $\Delta$ , modified UNIFAC (Dortmund),  $\circ$ , COSMO-RS(OL),  $\square$ , COSMOSAC.

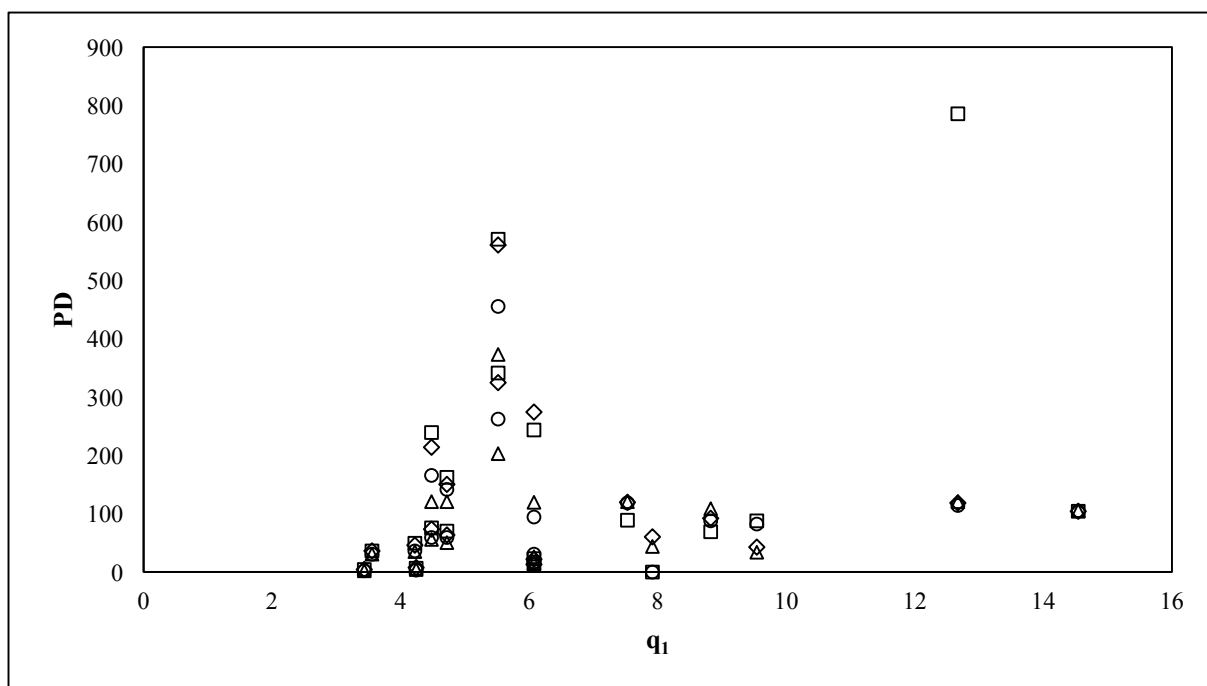
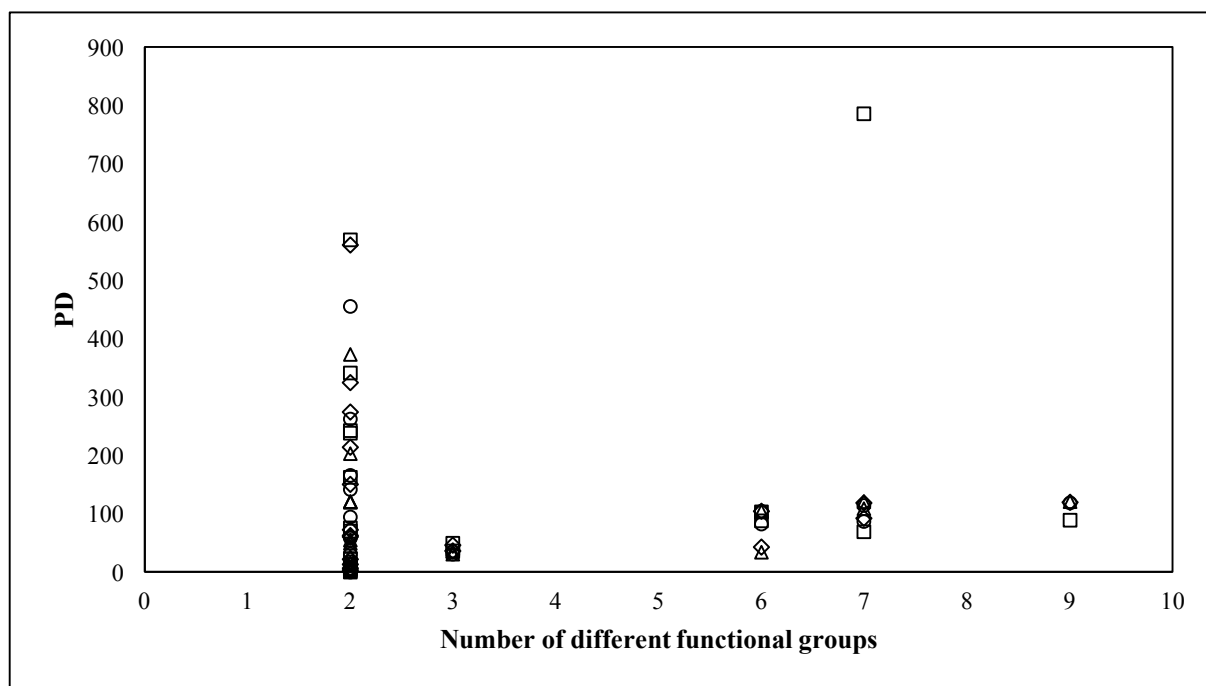


Figure 2.3 Correlation of model percentage deviations with van der Waals area parameter ( $q_1$ ).  $\diamond$ , UNIFAC,  $\Delta$ , modified UNIFAC (Dortmund),  $\circ$ , COSMO-RS(OL),  $\square$ , COSMOSAC.



**Figure 2.4 Correlation of model percentage deviations with number of different functional groups present in solute. ◇, UNIFAC, △, modified UNIFAC (Dortmund), ○, COSMO-RS(OL), □, COSMOSAC.**

As mentioned above, physical property data for the solutes considered are limited in the literature. The standard state used in the calculation of these properties is a pure hypothetical liquid, at a temperature much lower than the actual melting point. In order to calculate the change of heat of fusion with temperature, the difference of the heat capacities of the solid and the subcooled liquid is required. This calculation is often simplified by assuming a negligible heat capacity difference in this range (given by equation 1.40). An alternative assumption is to approximate the heat capacity change as the entropy of fusion (given by equation 1.41). Uncertainties can thus be introduced in the calculation of the activity coefficient from solubility data, and vice versa. The effect of these two assumptions is considered in this work using benzene and water as reference solvents. These results are compared in Table 2.3. Mishra and Yalkowsky (1990) have analysed this behaviour for similar solutes to those considered in this work, in benzene. In their work, for APIs in benzene, employing the UNIFAC combinatorial term with the Scatchard-Hildebrand residual term with the assumption of zero heat capacity changes provided the best prediction of solubility. For benzene as solvent, the modified UNIFAC (Dortmund) model with the Staverman-Guggenheim combinatorial term with free-volume correction is recommended from the results of this work, and the original UNIFAC

model with the Staverman-Guggenheim combinatorial term is recommended when water is used as a solvent.

**Table 2.3 Mean Percentage Deviations of various solutes in benzene and water.**

Solvent	Model	Heat capacity	Combinatorial	Residual	PD	Reference
Benzene	M1	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim Staverman-Guggenheim with modified UNIFAC parameters and free volume correction	UNIFAC mod UNIFAC (Dortmund)	20.24	This work
	M2	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim	COSMO-RS (OL)	18.33	This work
	M3	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim	COSMO-SAC	21.56	This work
	M4	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim	UNIFAC	29.09	This work
	M5	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim with modified UNIFAC parameters	mod UNIFAC (Dortmund)	23.79	This work
	M6	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim	COSMO-RS (OL)	25.60	This work
	M7	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim	COSMO-SAC	29.67	This work
	M8	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim	Scatchard- Hildebrand	20.00	Mishra and Yalkowsky (1990)
	r1	$\Delta_{fus}C_{pi} = 0$	Flory-Huggins	Scatchard- Hildebrand	31.62	Mishra and Yalkowsky (1990)
	r2	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Flory-Huggins	Scatchard- Hildebrand	37.42	Mishra and Yalkowsky (1990)
	r3	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim	UNIFAC	53.85	Mishra and Yalkowsky (1990)
	r4	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim	UNIFAC	40.00	Mishra and Yalkowsky (1990)
	r5	$\Delta_{fus}C_{pi} = 0$	Flory-Huggins	UNIFAC	56.57	Mishra and Yalkowsky (1990)
	r6	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Flory-Huggins	UNIFAC Scatchard- Hildebrand	17.32	Mishra and Yalkowsky (1990)
	r7	$\Delta_{fus}C_{pi} = 0$	UNIFAC	Scatchard- Hildebrand	28.28	Mishra and Yalkowsky (1990)
	r8	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	UNIFAC	Hildebrand		
Water	M1	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim Staverman-Guggenheim with modified UNIFAC parameters and free volume correction	UNIFAC mod UNIFAC (Dortmund)	116.40	This work
	M2	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim	COSMO-RS (OL)	107.09	This work
	M3	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim	COSMO-SAC	113.19	This work
	M4	$\Delta_{fus}C_{pi} = 0$	Staverman-Guggenheim	UNIFAC	104.59	This work
	M5	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim with modified UNIFAC parameters	mod UNIFAC (Dortmund)	141.61	This work
	M6	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim	COSMO-RS (OL)	114.95	This work
	M7	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim	COSMO-SAC	130.26	This work
	M8	$\Delta_{fus}C_{pi} = \Delta_{fus}S_i$	Staverman-Guggenheim	COSMO-SAC		



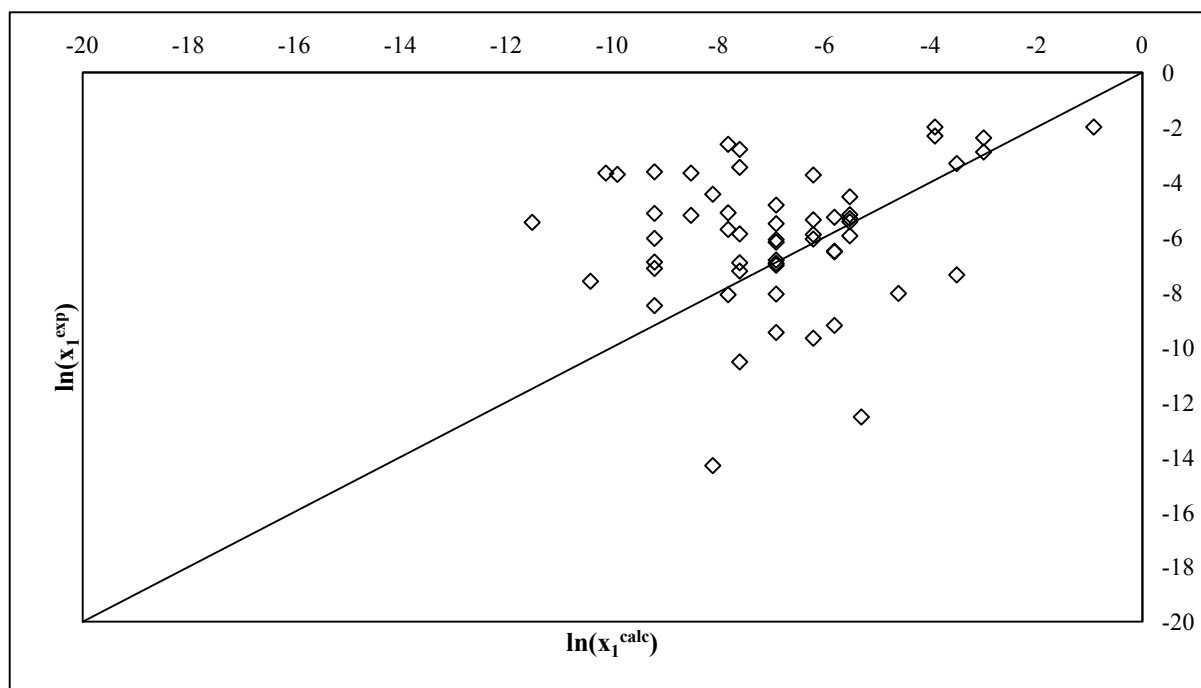
The NRTL-SAC model was applied to a subset of the dataset considered here. Comparisons are only made to experimental data, as the model is semi-correlative, and would not offer a fair comparison to the purely predictive models discussed above. In order to apply the NRTL-SAC model to solubility predictions, the segment area parameters (X, Y+, Y- and Z) must be known for the solutes and solvents considered. If these parameters are not available in the literature, they can be regressed from solubility data via the calculation of the activity coefficient and using pure component property data. Some of the NRTL-SAC model parameters for the solutes considered in this work were not available in the literature, and were therefore determined by the regression of the solubility data provided in Table 2.2. These new parameters are available in Table 2.4 along with literature sources where available.

Solubility predictions were then performed using the new segment area parameters and solvent parameters provided by Chen and Song (2004) and are shown in Figure 2.5. The results show that the NRTL-SAC model generally does not exhibit any tendency to over or under predict the experimental solubility. Again the predictive capability of the model is qualitative in most cases of the systems of steroidal APIs that were tested. This is a significant deficiency, as the model is semi-correlative.

**Table 2.4** Calculated segment area parameters for NRTL-SAC.

Solute	This work <sup>a</sup>				Literature <sup>b</sup>			
	X	Y+	Y-	Z	X	Y+	Y-	Z
<b>betulin</b>	0.0441	0.0743	0.0189	0.0024	-	-	-	-
<b>diosgenin</b>	0.1651	0.0112	0.1696	0.0183	-	-	-	-
<b>mestanolone</b>	0.3224	1.1220	0.7231	0.1953	-	-	-	-
<b>hydrocortisone</b>	0.4130	1.3020	0.9420	0.7110	0.4010	1.2480	0.9700	1.2480
<b>estrone</b>	0.4822	1.4240	0.710	0.1973	0.4990	1.5210	0.6790	0.1960
<b>prednisolone</b>	0.3945	1.1039	1.8975	0.3290	-	-	-	-
<b>testosterone</b>	1.041	0.2290	0.5460	0.7010	1.0510	0.2330	0.7710	0.6690

<sup>a</sup> Objective function of Chen and Song (2004) used:  $\delta = \left( \frac{\sum_{i=1}^N (\ln x_i^{calc} - \ln x_i^{exp})^2}{N} \right)^{1/2}$ , <sup>b</sup> Taken from Chen and Song (2004)



**Figure 2.5 Comparison of the natural logarithms of experimental and model calculated solubility composition ( $x_1$ ) with the NRTL-SAC model.**

### 2.5 Conclusion

Solubility predictions were carried out for the polycyclic steroidal and triterpene solutes considered in this work using various predictive models where model parameters were available in the literature. It has been found that no single model tested provides a superior solubility prediction for all of the systems considered. The UNIFAC-based and COSMO-based models tend to underestimate the solubility in the solutes considered, while the NRTL-SAC model shows no appreciable under or overestimating tendencies. New NRTL-SAC segment area parameters have been determined for some of the solutes considered in this work. The Staverman-Guggenheim combinatorial term with correction for free-volume with the modified UNIFAC (Dortmund) model residual term, with the approximation of  $\Delta_{fus}C_{pi} = 0$  provided the lowest percentage deviation for the solutes considered here in benzene. In aqueous systems the original UNIFAC model, with the approximation of  $\Delta_{fus}C_{pi} = \Delta_{fus}S_i$  provided a superior fit. This information can be used as a subsidiary guide for the selection of solvents in crystallization processes involving the studied solutes, however experimental results will be required if quantitative data is desired.

*References*

- Apelblat, A. and Manzurola, E., (1990), Solubility of suberic, azelaic, levulinic, glycolic, and diglycolic acids in water from 278.25 K to 361.35 K. *The Journal of Chemical Thermodynamics*, 22, pp.289-292.
- Apelblat, A. and Manzurola, E., (1987), Solubility of oxalic, malonic, succinic, adipic, maleic, malic, citric, and tartaric acids in water from 278.15 to 338.15 K. *The Journal of Chemical Thermodynamics*, 19, pp.317-320.
- Apelblat, A. and Manzurola, E., (1989), Solubility of ascorbic, 2-furancarboxylic, glutaric, pimelic, salicylic, and o-phthalic acids in water from 279.15 to 342.15 K, and apparent molar volumes of ascorbic, glutaric, and pimelic acids in water at 298.15 K. *The Journal of Chemical Thermodynamics*, 21, pp.1005-1008.
- Abildskov, J., Gani, R., Nielsen, M.B., Kolar, P. and Tsuboi, A., (2001), Solvent Selection for Drug Development. *Proceedings of the AIChE Annual Meeting*, Los Angeles.
- Bondi, A., (1964), van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68, pp.441-451.
- Bouillot, B., Teychené, S. and Biscans, B., (2011), An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents. *Fluid Phase Equilibria*, 309, pp.36-52.
- Bowen, D.B., James, K.C. and Roberts, M., (1970), An investigation of the distribution coefficients of some androgen esters using paper chromatography. *Journal of Pharmacy and Pharmacology*, 22, pp.518-522.
- Budavari, S. ed., *The Merck Index, An Encyclopedia of Chemicals, Drugs, and Biologicals*. 11th ed., Merck and Co., Rahway, New Jersey.
- Cai, X., Grant, D.J. and Wiedmann, T.S., (1997), Analysis of the solubilisation of steroids by bile salt micelles. *Journal of pharmaceutical sciences*, 86, pp.372-377.
- Cao, D., Zhao, G. and Yan, W., (2007), Solubilities of betulin in 14 solvents. *Journal of Chemical and Engineering Data*, 52, pp.1366-1368.
- Chacko, A., Devi, R., Abraham, S. and Mathew, B., (2005), A comparison of the oxidizing ability of polystyrene-supported linear and cyclic polyoxyethylene bound permanganates. *Journal of Applied Polymer Science*, 96, pp.1897-1905.
- Chen, C.-C. and Song, Y., (2004), Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Industrial and Engineering Chemistry Research*, 43, pp.8354-8362.
- Chen, F.X., Zhao, M.R., Liu, C.C., Peng, F.F. and Ren, B.Z., (2012), Determination and correlation of the solubility for diosgenin in alcohol solvents. *Journal of Chemical Thermodynamics*, 50, pp.1-6.

Chickos, J.S. and Acree, W.E., (2009), Total phase change entropies and enthalpies. An update on fusion enthalpies and their estimation. *Thermochimica Acta*, 495, pp.5–13.

CRC Handbook of Chemistry and Physics 87th edition, Taylor and Francis, 2007.

DDBST Software and Separation Technology GmbH, (2012), Dortmund Data Bank (DDB).

Diedrichs, A. and Gmehling, J., (2011), Solubility calculation of active pharmaceutical ingredients in alkanes, alcohols, water and their mixtures using various activity coefficient models. *Industrial and Engineering Chemistry Research*, 50, pp.1757–1769.

DIPPR DIADEM Public v. 1.2.

Domańska, U., Pobudkowska, A., Pelczarska, A., Winiarska-Tusznio, M. and Gierycz, P., (2010), Solubility and pKa of select pharmaceuticals in water, ethanol, and 1-octanol. *The Journal of Chemical Thermodynamics*, 42, pp.1465–1472.

Fredenslund, A., Jones, R.L. and Prausnitz, J.M., (1975), Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21, pp.1086–1099.

Gharavi, M., James, K.C. and Sanders, L.M., (1983), Solubilities of mestanolone, methandienone, methyltestosterone, nandrolone and testosterone in homologous series of alkanes and alkanols. *International Journal of Pharmaceutics*, 14, pp.333–341.

Gmehling, J.G., Anderson, T.F. and Prausnitz, J.M., (1978), Solid-Liquid Equilibria Using UNIFAC. *Industrial and Engineering Chemistry Fundamentals*, 17, pp.269–273.

Gracin, S., Brinck, T. and Rasmuson, Å.C., (2002), Prediction of solubility of solid organic compounds in solvents by UNIFAC. *Industrial and Engineering Chemistry Research*, 41, pp.5114–5124.

Gracin, S. and Rasmuson, Å.C., (2002), Solubility of Phenylacetic Acid, p-Hydroxyphenylacetic Acid, p-Aminophenylacetic Acid, p-Hydroxybenzoic Acid, and Ibuprofen in Pure Solvents. *Journal of Chemical and Engineering Data*, 47, pp.1379–1383.

Grensemann, H. and Gmehling, J., (2005), Performance of a Conductor-Like Screening Model for Real Solvents Model in Comparison to Classical Group Contribution Methods. *Industrial and Engineering Chemistry Research*, 44, pp.1610–1624.

Hagen, T. A and Flynn, G.L., (1983), Solubility of hydrocortisone in organic and aqueous media: evidence for regular solution behavior in apolar solvents. *Journal of pharmaceutical sciences*, 72, pp.409–414.

Hahnenkamp, I., Graubner, G. and Gmehling, J., (2010), Measurement and prediction of solubilities of active pharmaceutical ingredients. *International Journal of Pharmaceutics*, 388, pp.73–81.

- Hill, R.A., Kirk, D.N., Makin, H.L.J. and Murphy, G.M., (1991), *Dictionary of steroids: Chemical data, structures and bibliographies*. H. L. J. Hill, R.A. and Makin, ed., Chapman and Hall, London.
- Kolář, P., Shen, J.W., Tsuboi, A. and Ishikawa, T., (2002), Solvent selection for pharmaceuticals. *Fluid Phase Equilibria*, 194-197, pp.771–782.
- Kosal, E., Lee, C.H. and Holder, G.D., (1992), Solubility of Progesterone, Testosterone, and Cholesterol in Supercritical Fluids. *Journal of supercritical fluids*, 5, pp.169–179.
- Lin, H.M. and Nash, R. A, (1993), An experimental method for determining the Hildebrand solubility parameter of organic nonelectrolytes. *Journal of pharmaceutical sciences*, 82, pp.1018–26.
- Lin, S.-T. and Sandler, S.I., (2002), A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Industrial and Engineering Chemistry Research*, 41, pp.899–913.
- Manzurola, E. and Apelblat, A., (2002), Solubilities of l-glutamic acid, 3-nitrobenzoic acid, p-toluic acid, calcium-l-lactate, calcium gluconate, magnesium-dl-aspartate, and magnesium-l-lactate in water. *The Journal of Chemical Thermodynamics*, 34, pp.1127-1136.
- Martin, A., Wu, P.L., Adjei, A., Mehdizadeh, M., James, K.C. and Metzler, C., (1982), Extended Hildebrand solubility approach: Testosterone and testosterone propionate in binary solvents. *Journal of Pharmaceutical Sciences*, 71, pp.1334–1340.
- McLaughlin, E. and Zainal, H.A., (1959), 177. The solubility behaviour of aromatic hydrocarbons in benzene. *Journal of the Chemical Society (Resumed)*, p.863-867.
- Mishra, D. and Yalkowsky, S., (1990), Solubility of organic compounds in non-aqueous systems: polycyclic aromatic hydrocarbons in benzene. *Industrial and Engineering Chemistry*, 29, pp.2278–2283.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (f), Experimental solubility for diosgenin and estriol in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . Manuscript submitted for publication.
- Mota, F.L., Queimada, A.J., Andreatta, A.E., Pinho, S.P. and Macedo, E.A., (2012), Calculation of drug-like molecules solubility using predictive activity coefficient models. *Fluid Phase Equilibria*, 322-323, pp.48–55.
- Mu, T., Rarey, J. and Gmehling, J., (2007), Performance of COSMO-RS with Sigma Profiles from Different Model Chemistries. *Industrial and Engineering Chemistry Research*, 46, pp.6612–6629.
- Neau, S.H., Bhandarkar, S. V. and Hellmuth, E.W., (1997), Differential molar heat capacities to test ideal solubility estimations. *Pharmaceutical Research*, 14, pp.601–605.

- Nordström, F.L. and Rasmuson, Å.C., (2006) (a), Solubility and Melting Properties of Salicylic Acid. *Journal of Chemical and Engineering Data*, 51, pp.1668–1671.
- Nordström, F.L. and Rasmuson, Å.C., (2006) (b), Polymorphism and thermodynamics of m-hydroxybenzoic acid. *European Journal of Pharmaceutical Sciences*, 28, pp.377–384.
- Omar, W. and Ulrich, J., (2006), Solid Liquid Equilibrium, Metastable Zone, and Nucleation Parameters of the Oxalic Acid–Water System. *Crystal Growth and Design*, 6, pp.1927–1930.
- Roux, M.V., Temprado, M., Jiménez, P., Foces-Foces, C., García, M.V. and Redondo, M.I., (2004), 2- and 3-furancarboxylic acids: a comparative study using calorimetry, IR spectroscopy and X-ray crystallography. *Thermochimica Acta*, 420, pp.59–66.
- Roux, M.V., Temprado, M. and Chickos, J.S., (2005), Vaporization, fusion and sublimation enthalpies of the dicarboxylic acids from C4 to C14 and C16. *The Journal of Chemical Thermodynamics*, 37, pp.941–953.
- Ruchelman, M.W., (1967), Solubility studies of estrone in organic solvents using gas-liquid chromatography. *Analytical biochemistry*, 19, pp.98–108.
- Rytting, J.H., Braxton, B.K., Xia, J., Breimer, Crommelin, D.J.A., and Midha, K.K. ed., (1989), Topics in Pharmaceutical Sciences 1989. In *Proceedings of the 49th International Congress of Pharmaceutical Sciences of FIP*. The Hague, pp. 447–457.
- Sabbah, R. and Perez, L., (1999), Étude thermodynamique des acides phtalique, isophtalique et téréphtalique. *Canadian Journal of Chemistry*, 77, pp.1508–1513.
- Weidlich, U. and Gmehling, J., (1987), A modified UNIFAC Model. 1. Prediction of VLE,  $h^E$ , and  $y^\infty$ . *Industrial and Engineering Chemistry Research*, 26, pp.1372–1381.
- Yalkowsky, S.H., Valvani, S.C. and Roseman, T.J., (1983), Solubility and partitioning VI: octanol solubility and octanol-water partition coefficients. *Journal of pharmaceutical sciences*, 72, pp.866–870.
- Zhao, G. and Yan, W., (2008), Solubilities of betulin in chloroform + methanol mixed solvents at T = (278.2, 288.2, 293.2, 298.2, 308.2 and 313.2) K. *Fluid Phase Equilibria*, 267, pp.79–82.

## CHAPTER THREE

### **A Universal Segment Approach for the Prediction of the Activity Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents**

#### *Abstract*

A novel method for the prediction of the activity coefficient in pharmaceutical-solvent systems is introduced. The method infers that the concept of segment interactions applies not only to components, but to the functional groups that comprise the component; more specifically, the popular UNIFAC based functional groups. In the present work, four basic segment interactions are considered, that include hydrophobic (dispersive), hydrophilic, as well as positive and negative polarity. The predictive model was applied to a set of structurally diverse, complex pharmaceuticals, and compared to popular qualitative solubility prediction methods such as NRTL-SAC (Chen and Song, 2004) and the UNIFAC based methods. Furthermore, the Akaike Information Criterion (Akaike, 1974) and Focused Information Criterion (Claeskens and Hjort, 2003) were used to establish the relative quality of the solubility predictions of various predictive models, with the new model exhibiting favourable results. The temperature dependence provided by the new model was also investigated.

### 3.1 Introduction

The interactions between the components of a mixture contribute considerably to the activity and hence the solubility of a solute in a solvent. The improvement and development of industrial technologies in sectors such as the petro-chemical, pharmaceutical and polymer industries requires the description of new mixtures of components for which no experimental data are available. Experimental measurements are costly and time consuming and this especially restricts rapid evaluation of different process alternatives by process simulation. Generally phase equilibrium data of mixtures containing complex organic molecules, such as pharmaceuticals are not readily available in literature. It is therefore important that accurate predictive models for the activity coefficient be developed for use in solvent selection for the modelling and design of separation processes.

Popular predictive methods that have been widely applied to solubility calculations include the UNIFAC group contribution methods (UNIFAC (Fredenslund et al., 1975), Gracin et al., (2002), Mota et al., (2012), mod. UNIFAC (Dortmund), Hahnenkamp et al., (2010), Bouillot et al., (2011), PHARM-UNIFAC, Diedrichs, (2010), Diedrichs and Gmehling, (2011) and the use of Hansen solubility parameters (Modearresi et al., (2008), Lindvig et al., (2002)). Additionally, recent methods employing solutions of surfaces or segment interactions (Non-Random Two-Liquid Segment Activity Coefficient model (NRTL-SAC) (Chen and Song, 2004), Conductor-like Screening Model for Realistic Solvation (COSMO-RS) (Klamt, 1995), Conductor-like Screening Model Segment Activity Coefficient (COSMO-SAC) (Lin and Sandler, 2002)) have become available. Other options are advanced generalized van der Waals equations of state like PC-SAFT (Ruether and Sadowski, 2009), and lattice theory approaches such as Non-Random Hydrogen Bonding (NRHB) (Tsivintzelis et al., 2009), which require a large number of parameters that cannot easily be obtained from experimental phase equilibria data alone.

Regrettably UNIFAC and the method of Hansen yield unsatisfactory results for solubility prediction in cases of large molecules with molecular masses approximately greater than 200 g/mol (such as pharmaceuticals and polymers), (Frank et al., 1999, Chen and Song, 2004), in comparison to the performance of UNIFAC in VLE prediction for example. The COSMO-based methods are computationally expensive and require component specific sigma profile information. NRTL-SAC has proven very useful for qualitative predictions, but again requires a large data bank of component specific segment area parameters for activity coefficient



predictions. Moodley et al. (2015) (a), for example, have shown the weakness (in accuracy or applicability) of UNIFAC, COSMO-based and NRTL-SAC models for solubility prediction in cases of complex poly-cyclic steroidal pharmaceuticals with large molecular masses. However these methods have been successfully applied to solubility predictions for structurally more simple pharmaceuticals (Gracin et al., (2002), Mota et al., (2012)).

The group contribution methods base the prediction of activity coefficients on the group-group interactions between the various functional groups in a mixture. Both the component mixture and the pure components are then treated as mixtures of these groups. As of 2011 the most widely applicable method (mod. UNIFAC Consortium Version, Gmehling et al., 2011) contains 91 main functional groups. This amounts to over 8000 temperature dependent group-group interaction parameters required, of which about 1400 have been determined so far. It is therefore desirable that an alternate approach for the prediction of the activity coefficient be developed, that reduces the number of interaction parameters required for an application.

Hansen (2007) described the enthalpy change of mixing,  $\Delta H_m$ , as a function of the volume fraction of a component and various “contributors” to solubility. Each solubility contribution is due to a surface segment (dispersion, hydrogen bonding and polar). The entropy change of mixing,  $\Delta S_m$ , was calculated by the Flory-Huggins (Flory, 1941, Huggins, 1941) expression. Together with the Gibbs energy change upon mixing of an ideal solution, the Gibbs excess energy can be calculated.

The segment contribution concept suggests that all energetic interactions that contribute to the activity coefficient, result from the interaction of different types of segments on the “surface” of the molecule (the surface of a molecule is a hypothetical construct only). The segment can also be viewed as the surface of a molecular shaped cavity, as used in COSMO calculations or the intermolecular “contact area” in a dense fluid. Three basic types of interactions in non-electrolyte liquid mixtures of organic (and some inorganic) components are hydrophobic (dispersion), hydrophilic (hydrogen bonding) and polar. The polar segment can be further separated into positive and negative polarity, which may also act in part as hydrogen bond donors or acceptors, as is the case in this work.

It is therefore proposed in this work, that the structural groups used in group contribution approaches like UNIFAC, can be viewed as being made up of the four basic segment types given above. Instead of group-group interaction parameters with all other groups, only the segment sizes for each group are required. This reduces the number of required parameters in

the example given above (Modified UNIFAC Consortium Version), from 8190 to 364 if temperature dependent model parameters are not employed. The UNiVersal Segment Activity Coefficient model (UNISAC) as presented here, has been developed specifically for the prediction of activity and hence solubility of complex pharmaceuticals. UNISAC could in principle also replace UNIFAC in all its other applications but it will have to be analysed as to whether the drastic reduction in the number of parameters has a serious effect on the quality of the results. Potential shortcomings will be discussed here.

### 3.2 Theory

#### 3.2.1 Solubility Modelling

When a solid-liquid mixture is in phase equilibrium, the solvent is saturated with the solute. For eutectic mixtures, the solubility of the solvent in the solid solute is neglected and the chemical potential of the solute,  $i$ , in the pure solid phase  $\mu_i^s$  is equal to the chemical potential of the solute in the liquid solution  $\mu_i^{sat}$ :

$$\mu_i^s = \mu_i^{sat} \quad (3.1)$$

The chemical potential of the solute in the liquid solution is given by:

$$\mu_i^{sat} = \mu_i^0 + RT \ln (\gamma_i^{sat} x_i^{sat}) \quad (3.2)$$

Where,  $\mu_i^0$ , is the chemical potential of the hypothetical pure liquid solute at the system temperature,  $R$  is the Universal Gas constant in  $J.kmol^{-1}.K^{-1}$ ,  $T$  is the temperature in Kelvin, and  $\gamma_i^{sat}$  is the activity coefficient of the solute in the saturated solution.

An expression for the activity ( $a_i^{sat} = \gamma_i^{sat} x_i^{sat}$ ) of the solute can be obtained by combining equations (3.1) and (3.2), yielding:

$$\ln(\gamma_i^{sat} x_i^{sat}) = \frac{\mu_i^s - \mu_i^0}{RT} \quad (3.3)$$

At constant pressure and temperature, the chemical potential is equal to the partial molar Gibbs energy, hence:

$$\frac{\mu_i^{sat} - \mu_i^0}{RT} = \frac{\bar{G}_i^s - \bar{G}_i^0}{RT} \quad (3.4)$$

And therefore

$$\ln(\gamma_i^{sat} x_i^{sat}) = \frac{\Delta_{fus} \bar{G}_i}{RT} \quad (3.5)$$

Where  $\Delta_{fus} \bar{G}_i$  is the hypothetical partial molar Gibbs energy of fusion at the system pressure and temperature which is zero for the pure solute at its triple point. The following expression can be derived, assuming a constant difference in heat capacity between the solid and the subcooled liquid solute between the triple point and the system temperature (Gmehling et al., 2012):

$$\frac{\Delta_{fus} \bar{G}_i(T)}{RT} = \frac{\Delta_{fus} \bar{H}_i(T_{tr})}{R} \left( \frac{1}{T_{tr}} - \frac{1}{T} \right) - \frac{\Delta_{fus} \bar{C}_{pi}}{R} \left[ \ln \left( \frac{T_{tr}}{T} \right) - \frac{T_{tr}}{T} + 1 \right] \quad (3.6)$$

Where  $\Delta_{fus} \bar{H}_i(T_{tr})$  is the enthalpy of fusion at the triple point (usually approximated at the fusion point ( $_{fus}T_i$ )),  $\Delta_{fus} \bar{C}_{pi}$  is the partial difference in heat capacity between the subcooled liquid solute and the solid and  $T_{tr}$  is the triple point temperature in Kelvin. Generally the effect of  $\Delta_{fus} \bar{C}_{pi}$  is considered negligible in comparison to the other term.

In this case equation (3.6) reduces to:

$$\frac{\Delta_{fus} \bar{G}_i}{RT} = \frac{\Delta_{fus} \bar{H}_i(_{fus}T_i)}{R} \left( \frac{1}{_{fus}T_i} - \frac{1}{T} \right) \quad (3.7)$$

This derivation also ignores the pressure influence on solid solubility as the difference between system pressure and triple point pressure is regarded as sufficiently small so that a Poynting correction term is not necessary.

Alternatively, Hildebrand and Scott (1962) suggest approximating  $\Delta_{fus}\overline{C_{pi}}$  as  $\frac{\Delta_{fus}\overline{H}_i}{f_{us}T_i}$ . This assumption is not exceptionally accurate, however Neau et al. (1997) have shown that the true value of  $\Delta_{fus}\overline{C_{pi}}$  is generally much closer to  $\frac{\Delta_{fus}\overline{H}_i}{f_{us}T_i}$  than it is to 0. Equation (3.6) then reduces to:

$$\frac{\Delta_{fus}\overline{G}_i}{RT} = -\frac{\Delta_{fus}\overline{H}_i(f_{us}T)}{R f_{us}T} \ln\left(\frac{f_{us}T}{T}\right) \quad (3.8)$$

Moller (2009) has reported a mean difference of 10% between the results of equation (3.7) and equation (3.8), when  $T > \frac{f_{us}T_i}{1.45}$ . This improvement was confirmed by Moodley et al. (2015)(a) for aqueous systems.

Equation (3.8) was used in this work, and substitution into equation (3.5) yields the expression for solubility modelling.

$$\ln(\gamma_i^{sat} x_i^{sat}) = -\frac{\Delta_{fus}\overline{H}_i(f_{us}T_i)}{R f_{us}T_i} \ln\left(\frac{f_{us}T_i}{T}\right) \quad (3.9)$$

At a fixed temperature,  $\gamma_i^{sat}$  is a function of composition only, and evidently strongly influences the solubility of the solute in the solvent.

### 3.3 Universal segment approach to activity coefficient modelling

#### 3.3.1 Segmented Functional Group Concept

In group contribution methods for the prediction of the activity coefficient it is assumed that interactions between two molecules can be described by the interactions between the functional groups that make up the molecule. Methods such as NRTL-SAC that utilize the “*conceptual segment concept*” impose that interactions between two molecules can be attributed to interactions between conceptual segments on a hypothetical molecular surface. It stands to reason that if a molecule can be divided into functional groups *or* into “smaller” segments, that the molecule can firstly be divided into its functional groups, and then further into segments.

Hence, the conceptual surface segments of a functional group can be identified based on the functional groups in the molecule.

In this work the existing original UNIFAC groups are expressed as combinations of four basic segments as sources of the different molecular interactions similar to those considered by Chen and Song (Chen and Song, 2004) in NRTL-SAC; namely, dispersion (A), hydrogen bonding (B) and positive and negative polarity (C and D respectively) to represent energetic interactions in the new model. To represent these segment types, four existing original UNIFAC main groups were selected that include C-CH<sub>3</sub>, H<sub>2</sub>O, C-CN and C-Cl. These groups are termed base segments (B-MRR1), and it must be stressed here that these base segments do not physically exist in any of the remaining original UNIFAC groups (non-base segments), termed NB-MRR1, but are merely used to represent the energetic interactions exhibited by these groups. Specifically, the assumption is that all other original UNIFAC groups can be represented as mere combinations of these four base segments, rationed by the segment areas of each group that, in turn, “mimic” the unique group interactions.

### 3.3.2 *UNiversal Segment Activity Coefficient (UNISAC) model*

The activity coefficient is expressed in two parts, a combinatorial component for size/shape interactions, and a residual component for energetic interactions. The newly proposed model can be viewed as a combination of two popular activity coefficient models used for solubility prediction, namely original UNIFAC and NRTL-SAC for the residual term, with the modified UNIFAC combinatorial term for size/shape considerations:

$$\ln \gamma_i = \ln \gamma_i^{comb} + \ln \gamma_i^{res} \quad (3.10)$$

The combinatorial term is identical to the Flory-Huggins contribution with the Staverman-Guggenheim (Staverman, 1950, Guggenheim, 1952) correction term and a further modification to better describe asymmetrical systems, given by (Weidlich and Gmehling, (1987), Gmehling et al., (1998)). That is:

$$\ln \gamma_i^{comb} = 1 - V_i' + \ln(V_i') - 5q_i \left[ 1 - \frac{V_i}{F_i} + \ln \left( \frac{V_i}{F_i} \right) \right] \quad (3.11)$$

Where

$$V_i' = \frac{r_i^{3/4}}{\sum_j x_j r_j^{3/4}} \quad (3.12)$$

$$V_i = \frac{r_i}{\sum_j x_j r_j} \quad (3.13)$$

$$F_i = \frac{q_i}{\sum_j x_j q_j} \quad (3.14)$$

And

$$r_i = \sum_j v_j^{(i)} R_j \quad (3.15)$$

$$q_i = \sum_j v_j^{(i)} Q_j \quad (3.16)$$

Where  $r_i$  and  $q_i$  are the molecular van der Waals volumes and surface areas, estimated from the group contribution values ( $R_j, Q_j$ ) of modified UNIFAC by (Weidlich and Gmehling, (1987), Gmehling et al., (1998)) and  $v_j^{(i)}$  is the frequency of group  $j$  in component  $i$ .

The residual part incorporates the group-related segment interaction concept, and is given by:

$$\ln \gamma_i^{res} = \sum_k^N \Omega_{k,i} (\ln \Gamma_k - \ln \Gamma_{k,i}) \quad (3.17)$$

Where  $\Omega_{k,i}$  is the total segment area of a particular segment,  $k$ , in component  $i$  given by:

$$\Omega_{k,i} = \sum_{l=1}^N \nu_{l,i} \zeta_{k,l} \quad (3.18)$$

Where  $\nu_{l,i}$  is the number of groups of type  $l$  in a component,  $i$  and  $\zeta_{k,l}$  is the segment area of segment,  $k$ , in group  $l$ .

The natural logarithm of the segment activity coefficient of a unit surface segment,  $k$ , in the pure component,  $i$ , is given by:

$$\ln \Gamma_{k,i} = (1 - \ln(\sum_{m=1}^N \theta_{m,i} \psi_{m,k})) - \sum_{m=1}^N \frac{\theta_{m,i} \psi_{k,m}}{\sum_{n=1}^N \theta_{n,i} \psi_{n,m}} \quad (3.19)$$

Where  $\theta_{m,i}$  is the segment area fraction of segment  $m$  in the pure component,  $i$ , given by:

$$\theta_{m,i} = \frac{\Omega_{m,i}}{\sum_{m=1}^N \Omega_{m,i}} \quad (3.20)$$

The natural logarithm of the segment activity coefficient of a unit surface of segment,  $k$ , in the mixture  $\ln \Gamma_k$  is given by:

$$\ln \Gamma_k = (1 - \ln(\sum_{m=1}^N \theta_m \psi_{m,k})) - \sum_{m=1}^N \frac{\theta_m \psi_{k,m}}{\sum_{n=1}^N \theta_n \psi_{n,m}} \quad (3.21)$$

Where

$$\theta_m = \frac{\sum_{i=1}^I \Omega_{m,i} x_i}{\sum_{m=1}^N \sum_{i=1}^I \Omega_{m,i} x_i} \quad (3.22)$$

In equations (3.19) to (3.22),  $m$  and  $n$  are the segment based indices,  $N$  and  $I$ , are the total number of segments and components, respectively, while  $x_i$  is the mole fraction of component,  $i$ .  $\psi_{u,v}$  are the segment specific interactions, set equivalent to the dimensionless original UNIFAC group interactions between the base segments given by equation (1.55).

### 3.3.3 Selection of the base segments and interaction parameters

In the present work four basic segment interactions were selected in order to test the viability of the new model with the selection based on group chemistry, available experimental data for systems containing the groups, as well as classifications from literature. The molecular structures of B-MRR1, indicating conceptual segment interaction sites, are presented in Table 3.1.

Israelachvili (1985) states that the dispersive force can perhaps be considered as the most significant of the van der Waals forces, as it acts between all atoms and molecules regardless of the molecular properties. Dispersion is a short range force ( $(\frac{1}{r^6})$  for spheres and a long range force ( $(\frac{1}{r})$  for planes) that may be attractive or repulsive, and is often difficult to correlate with molecular size. The CH<sub>3</sub> group was selected to represent dispersion, since the group has a very low dipole moment, and is hydrophobic. This implies that polar and hydrogen bonding interactions to other groups are negligible.

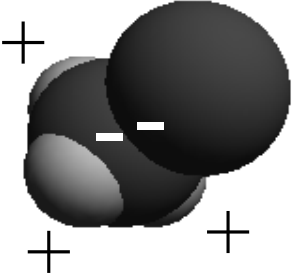
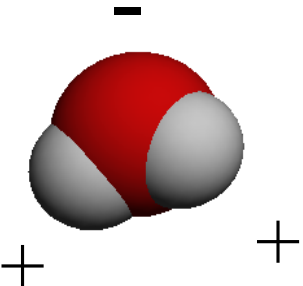
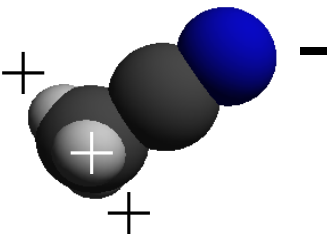
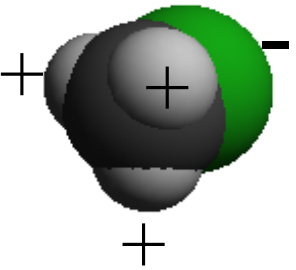
Hydrophilic molecules are able to act as hydrogen bond donors and/or acceptors. Hydrophilicity of a molecule increases with decreasing hydrophobic portions of the molecular surface, which explains the decreasing hydrophilicity of alcohols with increasing hydrocarbon portion, (methanol, ethanol, propanol...). The original UNIFAC H<sub>2</sub>O group was chosen to represent the hydrophilic segment.

Positive and negative polar segments are considered separately in this work, but both exhibit dipole-dipole interactions, polarizability, and induced dipole-dipole interactions. The C-CN and C-Cl groups were selected to represent the positive and negative polar segments respectively.

The C-CN group is polar with a strong positive polar site, due to the permanent positive dipole created by the relatively strongly negative N<sup>3-</sup> atom, on the adjacent side of the molecule. This site also acts as a hydrogen bond acceptor. The dipole moment of C-CN is a substantial 3.92 Debye. Similarly a strong polar-negative site is present in C-Cl, as a permanent negative dipole is induced by the Cl molecule, with a dipole moment of 1.92. The polarizability of H<sub>3</sub>C-CN is 4.4, and is 4.53 for H<sub>3</sub>C-Cl (Israelachvili, 1985). Due to the size of the chlorine atom, the negative charge density is not sufficient for hydrogen bond acceptance.



Table 3.1 Structure and Classification of selected base segments BMRR-1

BMRR-1	Structure	Dipole moment/ Debye (MOPAC, 2009)	Classification
C-CH <sub>3</sub>		0.4	Non-polar, Strongly hydrophobic (Solely dispersive)
H <sub>2</sub> O		1.85	Strongly hydrophilic Hydrogen bond donor and acceptor
C-CN		3.92	Polar-aprotic with strong positive polar site, and hydrogen bond acceptor
C-Cl		1.92	Polar with strong negative polar site

In order to apply the UNISAC model, all required model parameters must be available. These include the binary interaction parameters between the four base segments (B-MRR1), constituting 12 parameters, as well as the segment area parameters for the remaining  $\pm 60$  existing UNIFAC groups. These segment area parameters are unique to each non-base segment and must be determined by regression of experimental phase equilibria data. This may include a variety of data, e.g. vapour-liquid, liquid-liquid, solid-liquid equilibrium, and infinite dilution activity coefficient data. The binary interaction parameters between the four base segments were obtained by optimizing the existing base group interactions provided in the Dortmund Data Bank, in order to be specifically applied to solid-liquid equilibrium using regression. The binary interaction parameters for B-MRR1 are presented in Table 3.2. The data used in the training set of the UNISAC model to determine segment areas was obtained solely by original UNIFAC vapour-liquid equilibrium prediction i.e. experimental data was not directly used to obtain the non-base segment compositions for the UNISAC model, but rather pseudo-experimental data by original UNIFAC prediction was used. However it must be mentioned that the binary interaction parameters available in the literature used for the UNIFAC predictions, are the result of several decades of studies centred on data collection, processing and fitting.

The NB-MRR1 segment area parameters for a set of 30 of the most common original UNIFAC groups that appear in the majority of common solvents and solutes are presented in this work. The segment area parameters for the base segments (B-MRR1) are identical to the group area parameter  $q$ , defined in the UNIFAC models.

It is highly probable that not all original UNIFAC group interactions can be simply represented by the initial four base segments selected. Indeed, it is expected that the division of segment interactions into a more refined set (such as considering hydrogen bond donor and acceptor separately) would improve the prediction quality of the model to an extent. Furthermore, the addition of a possibly new type of segment interaction may also assist in representing some of the more complex original UNIFAC non-base segments, in terms of base segments and segment areas. However, the UNISAC model can easily be extended by the addition of further base segments.

In this work, the segment composition of an original UNIFAC group is assumed to be constant, regardless of whether the group is part of the solute or solvent. In most cases this assumption

is valid; however for some groups such as the acid group (COOH), a different behaviour is exhibited when the group comprises part of the solute as opposed to the solvent due to dimerization in the solid phase. This dimerization effect is relatively weak in the liquid phase.

The selection of appropriate models is of great importance for describing thermodynamic phase behaviour. Generally selection is based heavily on the ability of the tested models to represent experimental data via the root mean square deviation (RMSD), or some equivalent. Often little consideration is given to the number of model fitting parameters involved, the variance of the RMSDs or the bias of the model. Information criterion methods provide a means of quantifying the trade-off between the major pros and cons of a particular model from a set of candidate models, and are especially useful when selecting predictive models that usually consist of multiple empirical/semi-empirical parameters.

In this work, the Akaike Information Criterion and Focused Information Criterion were applied to a list of candidate models in order to determine the proficiency of the novel UNISAC model in comparison to popular existing methods i.e. UNIFAC and NRTL-SAC. It must be clarified that the models are not fitted to the experimental solubility data, but comparisons are made between the experimental values and the values predicted by the models.

#### 3.3.4 Akaike Information Criterion

The Akaike Information Criterion (AIC), (Akaike, 1974), quantifies the trade-off between model complexities, vs. the “goodness of fit” of a statistical model. The method provides a relative scale to express the “information entropy”, or information lost when a candidate model is used to represent (predict) a set of experimental data. The AIC gives no indication of the absolute fit or complexity of a single model, but provides a rank of the suitability of the candidate models relative to one another. The model with the lowest AIC score is the most favourable.

The AIC is given by:

$$AIC = 2k - 2\ln(L) \quad (3.23)$$

Where  $k$  is the number of the parameters in the particular candidate model and,  $L$ , is the maximised value of the likelihood function. If it is assumed that the differences between the

experimental and predicted values of each of the candidate models is normally distributed, with similar standard deviations, then a special case of a chi-squared ( $\chi^2$ ) fit is yielded.

The likelihood function then becomes:

$$L = \left(\frac{RSS}{n}\right)^{-n/2} \quad (3.24)$$

Where RSS is the residual sum of squares, given by:

$$RSS = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (3.25)$$

$n$  is the number of data points considered,  $y_i$  is an experimental data point, and  $f(x_i)$  is the corresponding predicted data point.

Combination of equation (3.23) and (3.24) yields:

$$AIC = 2k + n \ln \left(\frac{RSS}{n}\right) \quad (3.26)$$

### 3.3.5 Focused Information Criterion

The Focused Information Criterion (FIC), (Claeskens and Hjort, 2003) does not directly assess the quality of the fit of a particular model like the AIC, but rather the precision of the model in replicating the experimental data. More specifically, the precision in replicating the parameter of interest or *focus parameter* is calculated.

The *actual* precision is quantified by the *actual* mean square error ( $m$ ) in terms of bias and variance, where a low  $m$  indicates a high precision:

$$m_k = b_k^2 + \sigma_k^2 \quad (3.27)$$

Where the subscript,  $k$ , denotes the candidate model,  $b_k$ , is the bias of the predicted data in comparison to the experimental data for a set, and  $\sigma_k^2$  is the variance of the deviation between the experimental and predicted data.

The actual precision is then used to obtain an estimated precision, ( $\hat{m}$ ), using numerous input-specific FIC formula, that include the number of focus parameters considered etc. outlined in detail in the original publication. A FIC score is then generated, that indicates the ability of the model to replicate the focus parameter. A low FIC score indicates a better suited model.

### 3.4 Results and Discussion

#### 3.4.1 Model Parameters

The difference in solubility of the majority of pharmaceuticals in organic solvents as compared to water is widely acknowledged (Chen and Song, (2004), (Diedrichs and Gmehling, (2011), Moller, (2009), (Abildskov and O'Connell, (2003), (Abildskov et al., (2000), Pelczarska et al., (2013)). As such, it was found that the use of a single set of interaction parameters for B-MRR1, to encompass solute behaviour in aqueous and non-aqueous solutions, was not possible with four base segments. Hence, two sets of interaction parameters for B-MRR1 were used for non-aqueous and aqueous solutions, presented in Tables 3.2 and 3.3 respectively. In order to determine the segment interaction parameters of B-MRR1 the following objective function ( $\delta$ ) was minimized, while adjusting the parameter  $\psi_{u,v}$  in equations (3.19) and (3.21) for the calculated activity coefficient by equation (3.10):

$$\delta = \sum_{i=1}^{NP} \sum_{i=1}^I \text{abs}(\ln(\gamma_i^{sat})^{exp} - \ln(\gamma_i^{sat})^{calc}) \quad (3.28)$$

Where  $\gamma_i^{sat}$  is the activity coefficient at saturation,  $I$  is the number of components in the mixture, and  $NP$  is the total number of points considered. The superscripts *exp* and *calc* represent experimental (pseudo-experimental) and UNISAC model calculated values respectively.

The segment area parameters for NB-MRR1 are presented in Table 3.4. For comprehensiveness, the segment areas for (B-MRR1) are also included in Table 3.4. These parameters were determined by the regression of an extensive collection of vapour-liquid phase

equilibria data of mixtures containing the base segment components, predicted by the original UNIFAC model using parameters available in the Dortmund Data Bank (2011). The optimization required for determining segment area parameters, involved complex multi-parameter regression. In order to determine the segment area parameters of NB-MRR1 the objective function presented in equation (3.28) was minimized, while adjusting the parameter  $\zeta_{k,l}$  in equation (3.18) for the calculated activity coefficient by equation (3.10).

It must be reiterated that the applicability of the UNISAC model presented in this work is limited by the UNIFAC group fragmentation and the available model parameters. Systems with components that cannot be fragmented by the UNIFAC model or cases where model parameters for the UNISAC model are not available cannot be predicted.

**Table 3.2 Binary interaction parameters for base segments B-MRR1 for non-aqueous solutions.**

<b>B-MRR1</b>	<b>CH<sub>3</sub></b>	<b>H<sub>2</sub>O</b>	<b>CCN</b>	<b>CCI</b>
<b>CH<sub>3</sub></b>	0	400.1	540.1	186.8
<b>H<sub>2</sub>O</b>	1518.1	0	292.8	-1698.2
<b>CCN</b>	690.0	79.6	0	69.9
<b>CCI</b>	-248.3	-325.4	-100.0	0

**Table 3.3 Binary interaction parameters for base segments B-MRR1 for aqueous solutions.**

<b>B-MRR1</b>	<b>CH<sub>3</sub></b>	<b>H<sub>2</sub>O</b>	<b>CCN</b>	<b>CCI</b>
<b>CH<sub>3</sub></b>	0	1391	401	-326.55
<b>H<sub>2</sub>O</b>	-17	0	-634.1	86
<b>CCN</b>	-65	509	0	176
<b>CCI</b>	248.3	813	-368	0

**Table 3.4 Segment area parameters for base segments B-MRR1 and non- base segments NB-MRR1.**

Group	Segment Areas ( $\zeta$ )			
	A	B	C	D
AC	0.3404	0.0068	0.0000	0.0000
AC2H2	2.0157	1.2024	0.8000	0.0000
ACCH	0.0985	0.0000	0.0000	0.0008
ACCH2	0.0404	0.0006	0.0000	0.0000
ACCH3	0.3586	0.1406	0.0000	0.0000
ACH	0.6423	0.0134	0.1200	0.1000
ACNH2	26.0100	0.5900	2.6000	0.7100
ACOH	0.4627	0.0955	0.0000	0.0000
ACRY	1.2000	0.6000	0.7161	0.5978
C	0.0656	0.0000	0.0000	0.0000
CCl	0.0000	0.0000	0.0000	0.7240
CCl4	0.0093	0.6195	0.2447	0.0024
CH	0.1937	0.0000	0.0000	0.0000
CH=C	0.1497	0.0002	0.0728	0.0000
CH2	0.5400	0.0000	0.0000	0.0000
CH2CO	0.4318	0.0491	0.1550	0.0000
CH2NH	0.9044	0.0232	0.0027	0.0000
CH2O	0.6838	0.0000	1.2609	0.0004
CH3	0.8480	0.0000	0.0000	0.0000
CCN	2.5090	0.0000	1.7240	0.0000
CH3CO	0.4337	0.0624	0.2687	0.0000
CH3COO	0.7981	0.0739	0.1011	0.1220
CH3N	1.2409	0.0407	0.0943	0.3860
CH3NH	0.5863	0.3000	0.0100	0.0000
CH3O	1.0930	0.6195	0.2447	0.0024
CH3OH	0.4627	0.0955	0.0000	0.0000
CHCl3	1.3301	0.3343	0.0000	0.0379
CHNH	1.2554	0.0000	0.5347	0.0000
COO	0.6425	0.4822	0.0000	0.0341
COOH	1.8618	0.1985	0.6914	0.0000
DMF	0.2982	0.2124	0.0672	0.0000
DMSO	2.5047	1.1800	0.3860	0.0170
H2O	0.0000	1.4000	0.0000	0.0000
OH	0.7863	0.3000	0.0100	0.0000
THF	0.4751	0.2958	0.0000	0.0000

A - Dispersion, B - hydrogen bonding, C - polarity (+), D - polarity (-).

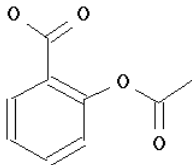
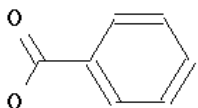
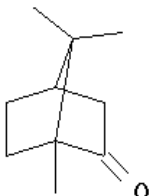
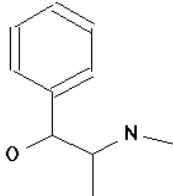
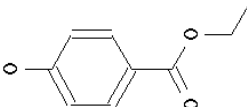
### 3.4.2 Model Test Set

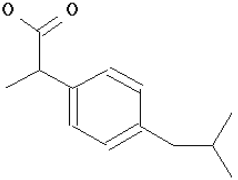
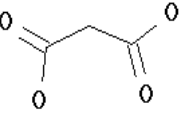
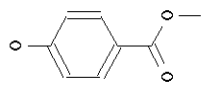
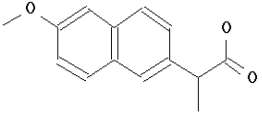
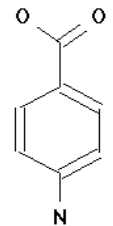
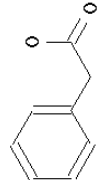
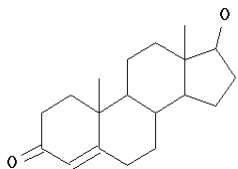
The predictive capability of the UNISAC model for solubility modelling was tested and compared to experimental data. Twelve solutes were considered in a variety of solvents, limited by the available experimental data for the selected solutes, and available NB-MRR1 parameters. The structure and relevant physical properties of the test solutes considered are presented in Table 3.5, along with literature sources of the solubility data used.

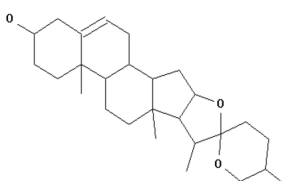
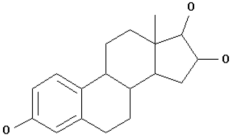
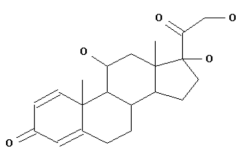
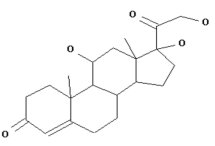
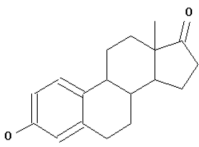
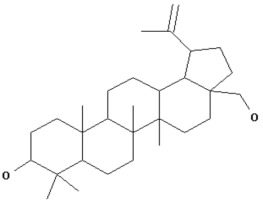
Solutes were selected based on the variety of structure and application. Principal component analysis was performed to describe the heterogeneity of the selected solutes, with the predicted activity coefficient at saturation in water, hexane or ethanol, melting point temperature and molecular mass, as input descriptors. The results of the first two descriptors (F1 and F2) describe approximately 80% (shown in Figure 3.1 (a)) of the diversity of the test set while F1 and F3 (shown in Figure 3.1 (b)) describe approximately 58% of the test set. The solutes selected are considered heterogeneous as they cover all four quadrants of all PCA plots. Again it must be mentioned that the test data were not used for the regression of the UNISAC segment area parameters that were used for prediction. For this test case 136 UNISAC segment area parameters were required for prediction while original UNIFAC required 378 binary interaction parameters.

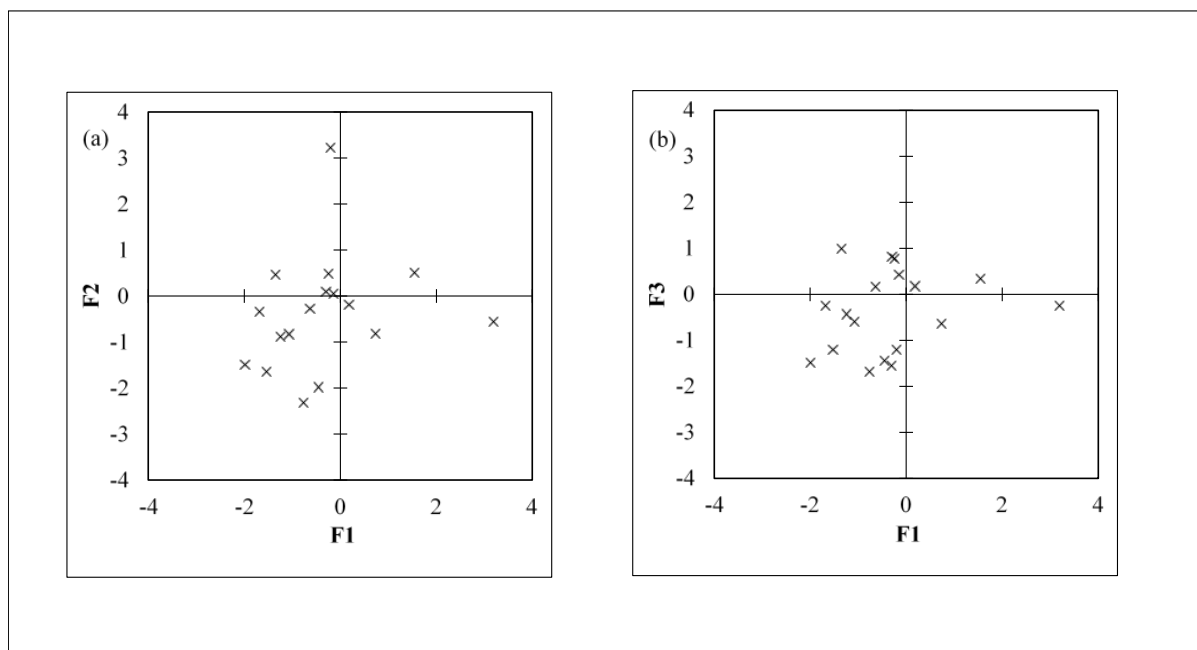


Table 3.5 Physical Properties of Solutes Used in this Work.

Solute	Structure	Molecular Mass/ g.mol <sup>-1</sup>	$\Delta_{fus}H$ (J.mol <sup>-1</sup> )	$T_{fus}$ (K)	Solubility data
Aspirin		180.16	25900 Górniak et al. (2011)	415 Górniak et al. (2011)	Perlovich and Bauer-Brandl (2003)
Benzoic Acid		122.12	18700 Murray et al. (1980)	396 Murray et al. (1980)	Perlovich and Bauer-Brandl (2003)
Camphor		152.23	5837 Lin and Nash (1993)	451 Lin and Nash (1993)	Lin and Nash (1993)
Ephedrine		165.23	20355 Lin and Nash (1993)	313 Lin and Nash (1993)	Lin and Nash (1993)
Ethylparaben		166.17	26400 Perlovich et al. (2005)	389 Perlovich et al. (2005)	Kenneth et al. (1977)

<b>Ibuprofen</b>		206.29	25500 Gracin et al. (2002)	347 Gracin et al. (2002)	Wang et al. (2010)
<b>Malonic Acid</b>		104.06	18739 Booth et al. (2010)	406 Booth et al. (2010)	Daneshfar (2012)
<b>Methylparaben</b>		152.15	25300 Perlovich et al. (2005)	399 Perlovich et al. (2005)	Kenneth et al. (1977)
<b>Naproxen</b>		230.26	31400 Perlovich et al. (2004)	428 Perlovich et al. (2004)	Daniels et al. (2004), (Yan et al. (2009)
<b>p-Aminobenzoic Acid</b>		137.14	24016 Lin and Nash (1993)	460 Lin and Nash (1993)	Lin and Nash (1993)
<b>Phenylacetic Acid</b>		136.15	15500 Gracin and Rasmuson (2002)	349 Gracin and Rasmuson (2002)	Gracin and Rasmuson (2002)
<b>Testosterone</b>		288.42	26167 Lin and Nash (1993)	429 Lin and Nash (1993)	Lin and Nash (1993)

<b>Diosgenin</b>		414.629	52105	474 Moodley et al. (2015)(f)	Moodley et al. (2015)(f) Chen et al., (2012)
<b>Estriol</b>		288.387	42718 Moodley et al. (2015)(f)	553 Moodley et al. (2015)(f)	Moodley et al. (2015)(f)
<b>Prednisolone</b>		360.450	59296 Moodley et al. (2015)(g)	507 Moodley et al. (2015)(g)	Moodley et al. (2015)(g)
<b>Hydrocortisone</b>		362.466	33900 Moodley et al. (2015)(g)	486 Moodley et al. (2015)(g)	Moodley et al. (2015)(g)
<b>Estrone</b>		270.371	45107 Moodley et al. (2015)(e)	528 Moodley et al. (2015)(e)	Moodley et al. (2015)(e), Ruchelman (1967)
<b>Betulin</b>		442.726	55169 Moodley et al. (2015)(e)	528 Moodley et al. (2015)(e)	Moodley et al. (2015)(e), Cao et al. (2007)



**Figure 3.1 Principal Component Analysis (PCA) to show the heterogeneity of the selected pharmaceutical components for model testing. (a) Principal Component F1 and F2, (b) Principal Component F1 and F3.**

### 3.4.3 Model Comparison

The original UNIFAC (publication version) and NRTL-SAC models were applied to the test set, for a comparison of performance to UNISAC, first for organic solvents using B-MRR1 interaction parameters from Table 3.2 and in water using interaction parameters from Table 3.3. The Root Mean Square Deviations (RMSD) in  $\log x^s$  between experimental data and the predictive methods are presented in Table 3.6, with the lowest RMSDs in bold. The UNISAC model provides a superior prediction in over two-thirds of the systems investigated.

The solvent parameters of the NRTL-SAC model available in the literature were used for the predictions conducted here. The experimental and predicted solubility data of aspirin is shown as an example in Table 3.7, and highlights one of the limitations of the NRTL-SAC model, specifically, a prediction cannot be made if the molecule specific area parameters are not available i.e. those for 1-hexanol and 1-heptanol.

The resultant predicted data for benzoic acid are presented graphically in Figure 3.2 (a), and indicates the non-bias of the new model, evident from the scattering of predictions around the reference  $y = x$ .

Figure 3.2 (b) presents the solubility prediction in water by UNISAC and UNIFAC. With the two clear outliers (camphor and p-aminobenzoic acid) removed, the RMSDs for UNISAC and UNIFAC are 2.13 and 2.90 respectively.

The Akaike and Focused Information Criterion scores are presented in Figures 3.3 (a) and 3.3 (b) respectively. The AIC scores recommend the UNISAC model for over 90% of the test cases, while the FIC scores recommend UNISAC in over 75% of the test cases.

The temperature dependence of the model is tested for the case of methylparaben in three low molecular weight alcohols. Figure 3.4 (a) shows that UNISAC can qualitatively predict the effect of increasing alkane chain length of alkanols, in relation to solubility at various temperatures.

From the systems tested, UNISAC offers the poorest prediction to the case of Naproxen, albeit superior to the prediction by original UNIFAC. However, interestingly, Figure 3.4 (b) shows that although quantitative results are poor, UNISAC still provides a suitable qualitative prediction, as the solubility rank of the solvents tested are maintained at various temperatures.

In Table 3.8 the UNISAC model is compared to results of the Pharma-UNIFAC and modified UNIFAC Dortmund model predictions for solid-liquid equilibrium calculations from the literature (Diedrichs and Gmehling, 2011). Although in the work of Diedrichs and Gmehling (2011), only alkane, alcohol and water as solvents were considered, it is evident that the UNISAC model is competitive with these existing models and has potential for future development. In future work it is intended to improve and expand the model parameter list of the UNISAC model and to apply it to a large data set to further explore its performance.

Table 3.6 Comparison between experimental and predicted values of the three models tested

Solute	Model	Number of Solvents	RMSD in $\log_{10}x^{s,a,b}$
<b>Aspirin</b>	UNISAC	11	<b>0.158</b>
	NRTL-SAC	9	0.351
	UNIFAC	11	0.559
<b>Benzoic Acid</b>	UNISAC	20	<b>0.208</b>
	NRTL-SAC	18	0.240
	UNIFAC	20	0.240
<b>Betulin</b>	UNISAC	10	<b>1.174</b>
	UNIFAC	9	1.294
	mod-UNIFAC <sup>c</sup>	7	1.913
<b>Camphor</b>	UNISAC	7	<b>0.201</b>
	NRTL-SAC	6	0.214
	UNIFAC	7	0.233
<b>Diosgenin</b>	UNISAC	10	<b>0.618</b>
	UNIFAC	9	1.246
	mod-UNIFAC <sup>c,d</sup>	8	1.386
<b>Ephedrine</b>	UNISAC	7	<b>0.117</b>
	NRTL-SAC	5	0.126
	UNIFAC	7	0.078
<b>Estriol</b>	UNISAC <sup>d</sup>	10	0.343
	NRTL-SAC	-	-
	UNIFAC	-	-
<b>Estrone</b>	UNISAC	12	<b>1.382</b>
	UNIFAC	11	1.898
	mod-UNIFAC <sup>c</sup>	9	2.117

**Ethylparaben**

UNISAC	7	0.359
NRTL-SAC	-	-
mod-UNIFAC <sup>e</sup>	7	<b>0.113</b>

**Hydrocortisone**

UNISAC <sup>d</sup>	10	0.615
NRTL-SAC	-	-
UNIFAC	-	-

**Ibuprofen**

UNISAC	9	0.097
NRTL-SAC	8	0.204
UNIFAC	9	<b>0.051</b>

**Malonic Acid**

UNISAC	5	0.290
NRTL-SAC	-	-
UNIFAC	5	0.425

**Methylparaben**

UNISAC	7	0.420
NRTL-SAC	6	0.187
mod-UNIFAC <sup>e</sup>	7	<b>0.167</b>

**Naproxen**

UNISAC	12	0.444
NRTL-SAC	-	-
UNIFAC	12	0.460

**p-Aminobenzoic Acid**

UNISAC	6	0.579
NRTL-SAC	6	0.694
UNIFAC	6	0.908

**Phenylacetic Acid**

UNISAC	7	0.103
NRTL-SAC	7	0.066
UNIFAC	7	<b>0.065</b>

**Prednisolone**

UNISAC <sup>d</sup>	10	2.250
NRTL-SAC	-	-
UNIFAC	-	-

**Testosterone**

UNISAC	6	<b>0.326</b>
NRTL-SAC	6	0.380
UNIFAC	6	0.882

$${}^a \text{RMSD} = \left[ \frac{\sum_{i=1}^N (\ln x_i^{s,exp} - \ln x_i^{s,calc})^2}{N} \right]^{1/2}, \quad {}^b \text{ bold values indicate lowest RMSD, } {}^c \text{ modified UNIFAC}$$

(Dortmund) was used as NRTL-SAC parameters were not available, <sup>d</sup> UNIFAC/mod UNIFAC fragmentation was not possible, a manual fragmentation was employed, <sup>e</sup> modified UNIFAC (Dortmund) was used as original UNIFAC parameters were not available.

**Table 3.7 Comparison between experimental and predicted values of Aspirin solubility at 298.15 K by various models.**

Solvent	Experimental Solubility/ $x^{s,exp}$	Predicted Solubility/ $x^{s,calc}$		
		UNISAC	NRTL-SAC	UNIFAC
<b>1, 4-Dioxane</b>	0.0516	0.0961	0.0961	0.0366
<b>Acetone</b>	0.0828	0.0588	0.0834	0.0610
<b>Acetonitrile</b>	0.0185	0.0114	0.0700	0.0496
<b>Ethyl Acetate</b>	0.0448	0.0235	0.0287	0.0259
<b>Ethanol</b>	0.0855	0.0681	0.0339	0.0105
<b>1-Propanol</b>	0.0418	0.0536	0.0161	0.0103
<b>1-Butanol</b>	0.0453	0.0438	0.014	0.0100
<b>1-Pentanol</b>	0.0395	0.0372	0.0432	0.0096
<b>1-Hexanol</b>	0.0393	0.0325	-	0.0093
<b>1-Heptanol</b>	0.0386	0.0290	-	0.0090
<b>1-Octanol</b>	0.0341	0.0265	0.0186	0.0087



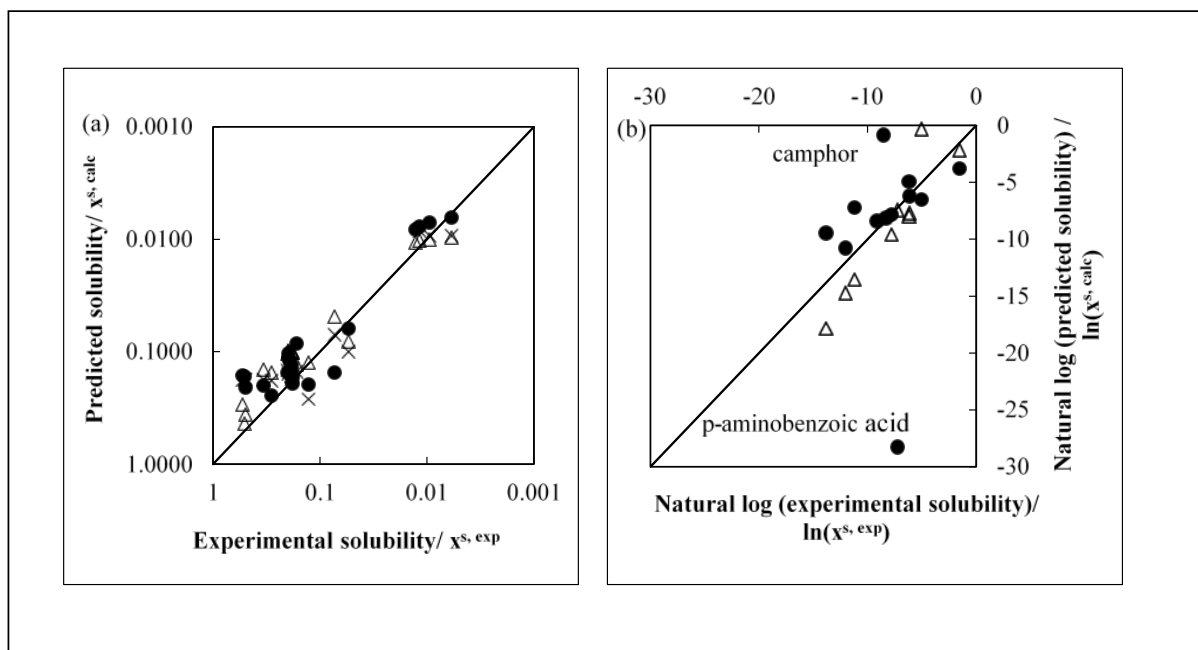
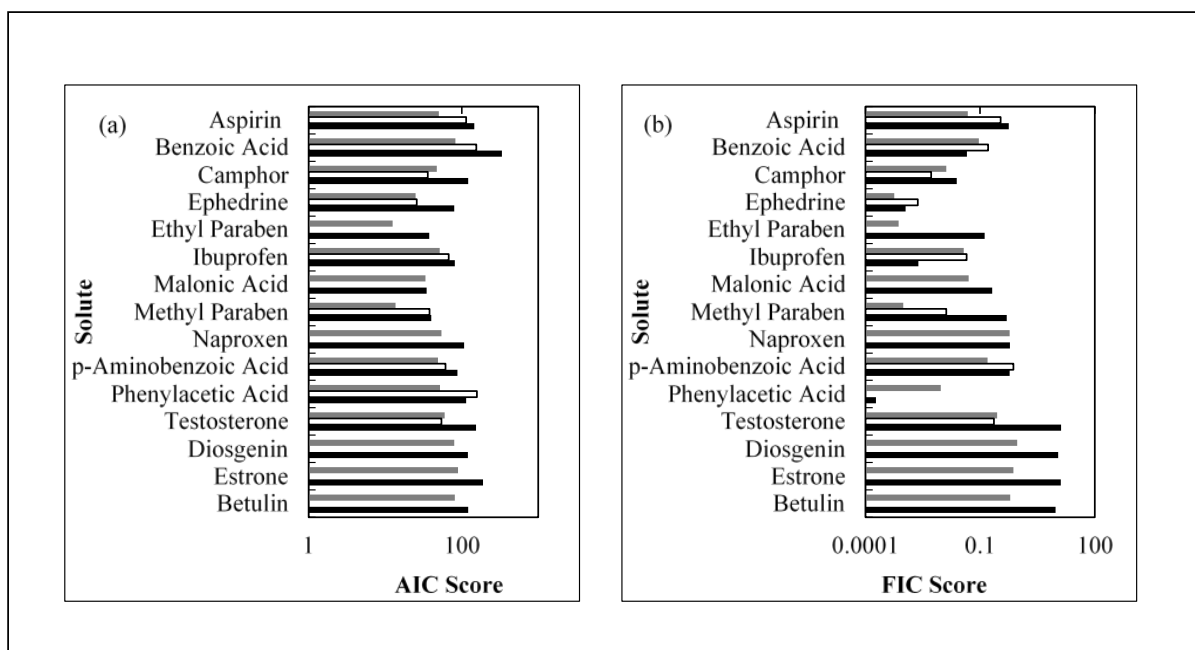
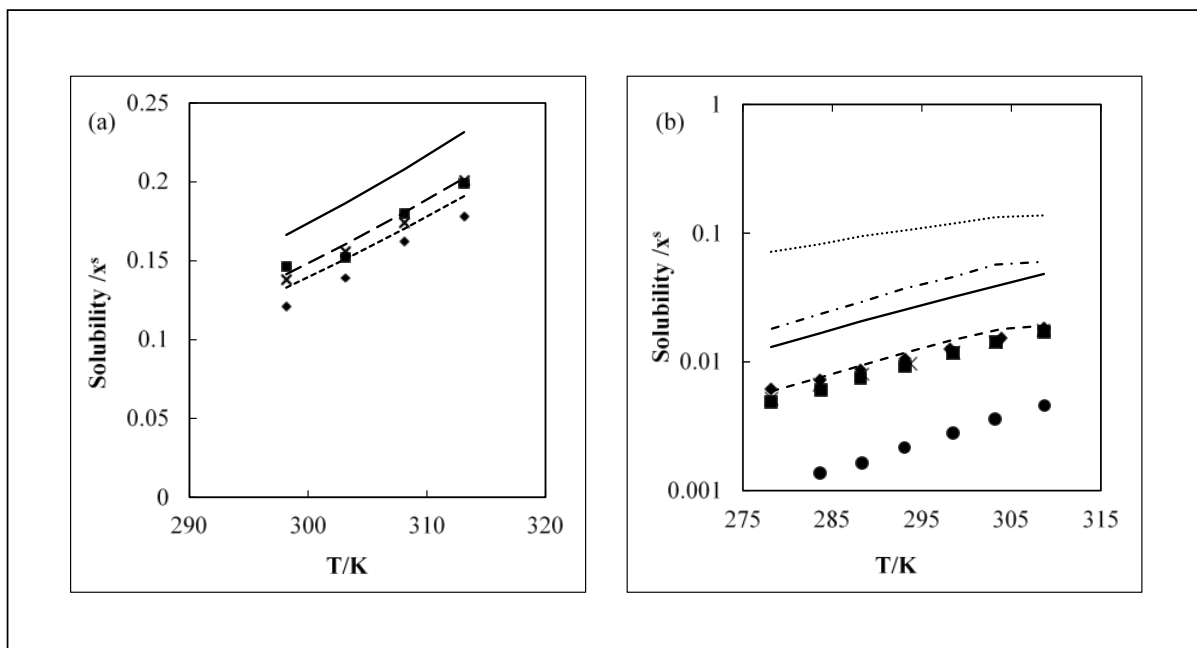


Figure 3.2 (a) Comparison of experimental solubility data for benzoic acid to predictive models.

● – UNISAC, × - NRTL-SAC, Δ – UNIFAC, (b) Comparison of experimental solubility data for various solutes in water, to predictive models. ● – UNISAC, Δ – UNIFAC.



**Figure 3.3 (a) Akaike Information Criterion Score for the solubility prediction of the models tested. ■ – UNISAC, □ – NRTL-SAC, ■ – UNIFAC, (b) Focused Information Criterion Score for the solubility prediction of the models tested. ■ – UNISAC, □ – NRTL-SAC, ■ – UNIFAC.**



**Figure 3.4 (a) Plot of Experimental and Predicted Solubility ( $x^s$ ) from UNISAC of Methylparaben in Methanol (■,—), 1-Propanol (×,— —) and 1-Butanol (◆,— · ·), at various temperatures, (b) Plot of Experimental vs. Predicted Solubility ( $x^s$ ) from UNISAC of Naproxen in Acetone (●,— —), Methanol (◆,— · —), Ethanol (■,—) and 2-Propanol (×,— · ·), at various temperatures. Experimental data are represented as symbols, and model predictions are represented as lines.**

**Table 3.8 Comparison of the performance of the UNISAC model to the modified UNIFAC (Dortmund) and Pharma-UNIFAC model from literature (Diedrichs and Gmehling, 2011), for alkanes, alcohol and water as a solvent.**

Model name	Number of solutes	Number of solvents	RMSD in $\log_{10}x^s$ <sup>a,b</sup>
<b>Alkanes</b>			
mod. UNIFAC (Diedrichs and Gmehling, 2011).	35	27	<b>0.797</b>
Pharma- UNIFAC (Diedrichs and Gmehling, 2011).	57	26	0.817
UNISAC	9	5	0.801
<b>Alcohols</b>			
mod. UNIFAC (Diedrichs and Gmehling, 2011).	36	17	0.582
Pharma- UNIFAC (Diedrichs and Gmehling, 2011).	63	20	0.620
UNISAC	17	15	<b>0.198</b>
<b>Water</b>			
mod. UNIFAC (Diedrichs and Gmehling, 2011).	63	1	1.553
Pharma- UNIFAC (Diedrichs and Gmehling, 2011).	115	1	1.414
UNISAC	17	1	<b>1.342</b>

$${}^a \text{RMSD} = \left[ \frac{\sum_{i=1}^N (\ln x_i^{s,exp} - \ln x_i^{s,calc})^2}{N} \right]^{1/2}, \quad {}^b \text{ bold values indicate lowest RMSD}$$

### 3.5 Conclusion

The UNISAC model provides a means of performing qualitative predictions of solubility for complex pharmaceutical components. The model only requires functional-group specific

segment area parameters (hydrophobic, hydrophilic, polar positive and polar negative), that can be determined from VLE, LLE or SLE data. The Akaike Information Criterion and Focused Information Criterion tests recommend UNISAC over the original UNIFAC and NRTL-SAC models in the majority of the cases tested in this work.

*References*

- Abildskov, J., Gani, R., Nielsen, M.B., Kolar, P. and Tsuboi, A., (2001), Solvent Selection for Drug Development. *Proceedings of the AIChE Annual Meeting*, Los Angeles
- Abildskov, J. and O'Connell, J.P., (2003), Predicting the Solubilities of Complex Chemicals I. Solutes in Different Solvents. *Industrial and Engineering Chemistry Research*, 42, pp.5622–5634.
- Akaike, H., (1974), A new look at the statistical model identification. *IEEE Transactions on Information Technology in Biomedicine*, 19, pp.716–723.
- Alexander, K.S., Mauger, J.W., Petersen, H. and Paruta, A.N., (1977), Solubility profiles and thermodynamics of parabens in aliphatic alcohols. *Journal of Pharmaceutical Sciences*, 66, pp.42–48.
- Booth, A.M., Barley, M.H., Topping, D.O., McFiggans, G., Garforth, a. and Percival, C.J., (2010), Solid state and sub-cooled liquid vapour pressures of substituted dicarboxylic acids using Knudsen Effusion Mass Spectrometry (KEMS) and Differential Scanning Calorimetry. *Atmospheric Chemistry and Physics*, 10, pp.4879–4892.
- Bouillot, B., Teychené, S. and Biscans, B., (2011), An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents. *Fluid Phase Equilibria*, 309, pp.36–52.
- CAChe Research Stewart Computational Chemistry LLC, (2009), *Molecular Orbital Package (MOPAC)*.
- Cao, D., Zhao, G. and Yan, W., (2007), Solubilities of betulin in 14 solvents.pdf. *J. Chem. Eng. Data*, 52, pp.1366–1368.
- Chen, C.-C. and Song, Y., (2004), Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Industrial and Engineering Chemistry Research*, 43, pp.8354–8362.
- Chen, F.X., Zhao, M.R., Liu, C.C., Peng, F.F. and Ren, B.Z., (2012), Determination and correlation of the solubility for diosgenin in alcohol solvents. *Journal of Chemical Thermodynamics*, 50, pp.1–6.
- Claeskens, G. and Hjort, N.L., (2003), The Focused Information Criterion. *Journal of the American Statistical Association*, 98, pp.900–916.
- Daneshfar, A., Baghlani, M., Sarabi, R.S., Sahraei, R., Abassi, S., Kaviyan, H. and Khezeli, T., (2012), Solubility of citric, malonic, and malic acids in different solvents from 303.2 to 333.2K. *Fluid Phase Equilibria*, 313, pp.11–15.
- Daniels, C.R., Charlton, A.K., Wold, R.M., Pustejovsky, E., Furman, A.N., Bilbrey, A.C., Love, J.N., Garza, J.A., Acree, W.E., et al., (2004), Mathematical correlation of naproxen

- solubilities in organic solvents with the abraham solvation parameter model. *Physics and Chemistry of Liquids*, 42, pp.481–491.
- DDBST Software and Separation Technology GmbH, (2012), Dortmund Data Bank (DDB).
- Diedrichs, A., (2010), *Evaluation und Erweiterung thermodynamischer Modelle zur Vorhersage von Wirkstofflöslichkeiten*/. Carl von Ossietzky Universität.
- Diedrichs, A. and Gmehling, J., (2011), Solubility calculation of active pharmaceutical ingredients in alkanes, alcohols, water and their mixtures using various activity coefficient models. *Industrial and Engineering Chemistry Research*, 50, pp.1757–1769.
- Flory, P.J., (1942), Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics*, 10, pp.51.
- Frank, T. C. Downey, J. R. Gupta, S. K., (1999) Quickly Screen Solvents for Organic Solids. *Chemical Engineering Progress*, Dec, pp. 41
- Fredenslund, A., Jones, R.L. and Prausnitz, J.M., (1975), Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21, pp.1086–1099.
- Gmehling, J., Kolbe, B., Kleiber, M. and Rarey, J., (2012), *Chemical Thermodynamics for Process Simulation*, Wiley, Germany.
- Gmehling, J., Lohmann, J., Jakob, A., Li, J. and Joh, R., (1998), A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension. *Industrial and Engineering Chemistry Research*, 37, pp.4876–4882.
- Gmehling, J., Rarey, J., Menke, J., (2011), *Dortmund Data Bank*, Oldenburg. [www.unifac.org](http://www.unifac.org)
- Górnaiak, A., Wojakowska, A., Karolewicz, B. and Pluta, J., (2011), Phase diagram and dissolution studies of the fenofibrate–acetylsalicylic acid system. *Journal of Thermal Analysis and Calorimetry*, 104, pp.1195–1200.
- Gracin, S., Brinck, T. and Rasmuson, Å.C., (2002), Prediction of solubility of solid organic compounds in solvents by UNIFAC. *Industrial and Engineering Chemistry Research*, 41, pp.5114–5124.
- Gracin, S. and Rasmuson, Å.C., (2002), Solubility of Phenylacetic Acid, p - Hydroxyphenylacetic Acid, p -Aminophenylacetic Acid, p -Hydroxybenzoic Acid, and Ibuprofen in Pure Solvents. *Journal of Chemical and Engineering Data*, 47, pp.1379–1383.
- Guggenheim, E.A., (1952), *Mixtures*, Oxford, U.K.
- Hahnenkamp, I., Graubner, G. and Gmehling, J., (2010), Measurement and prediction of solubilities of active pharmaceutical ingredients. *International Journal of Pharmaceutics*, 388, pp.73–81.

- Hansen, C.M., (2007), Solubility Parameters — An Introduction. In *Hansen Solubility Parameters: A User's Handbook*. pp. 1–24.
- Hildebrand, J.H. and Scott, R.L., (1964), *The solubility of nonelectrolytes*. 3rd ed., Dover, New York.
- Huggins, M.L., (1941), Solutions of long chain compounds. *The Journal of Chemical Physics*, 9, p.440.
- Israelachvili, J.N., (1985), *Intermolecular and Surface Forces, with applications to Colloidal and Biological Systems*, Academic Press, London.
- Klamt, A., (1995), Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *Journal of Physical Chemistry*, 99, pp.2224–2235.
- Lin, H.M. and Nash, R. A, (1993), An experimental method for determining the Hildebrand solubility parameter of organic nonelectrolytes. *Journal of pharmaceutical sciences*, 82, pp.1018–26.
- Lin, S.-T. and Sandler, S.I., (2002), A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Industrial and Engineering Chemistry Research*, 41, pp.899–913.
- Lindvig, T., Michelsen, M.L. and Kontogeorgis, G.M., (2002), A Flory-Huggins model based on the Hansen solubility parameters. *Fluid Phase Equilibria*, 203, pp.247–260.
- Modarresi, H., Conte, E., Abildskov, J., Gani, R. and Crafts, P., (2008), Model-based calculation of solid solubility for solvent selection - A review. *Industrial and Engineering Chemistry Research*, 47, pp.5234–5242.
- Moller, B., (2009), *Activity of complex multifunctional organic compounds in common solvents*. University of KwaZulu-Natal.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (a), Model Evaluation for the Prediction of Solubility of Active Pharmaceutical Ingredients (APIs). *Manuscript in preparation*.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (e), Experimental solubility for betulin and estrone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . *Journal of Chemical Thermodynamics*, 98, pp. 42–50.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (f), Experimental solubility for diosgenin and estriol in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . Manuscript submitted for publication.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (g), Experimental solubility for prednisolone and hydrocortisone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . Manuscript submitted for publication.



- Mota, F.L., Queimada, A.J., Andreatta, A.E., Pinho, S.P. and Macedo, E.A., (2012), Calculation of drug-like molecules solubility using predictive activity coefficient models. *Fluid Phase Equilibria*, 322-323, pp.48–55.
- Murray, J.P., Cavell, K.J. and Hill, J.O., (1980), A DSC study of benzoic acid: a suggested calibrant compound. *Thermochimica Acta*, 36, pp.97–101.
- Neau, S.H., Bhandarkar, S. V and Hellmuth, E.W., (1997), Differential molar heat capacities to test ideal solubility estimations. *Pharmaceutical research*, 14, pp.601–605.
- Pelczarska, A., Ramjugernath, D., Rarey, J. and Domańska, U., (2013), Prediction of the solubility of selected pharmaceuticals in water and alcohols with a group contribution method. *The Journal of Chemical Thermodynamics*, 62, pp.118–129.
- Perlovich, G.L. and Bauer-Brandl, A., (2003), Thermodynamics of solutions I: Benzoic acid and acetylsalicylic acid as models for drug substances and the prediction of solubility. *Pharmaceutical Research*, 20, pp.471–478.
- Perlovich, G.L., Kurkov, S. V, Kinchin, A.N. and Bauer-Brandl, A., (2004), Thermodynamics of solutions III: comparison of the solvation of (+)-naproxen with other NSAIDs. *European journal of pharmaceuticals and biopharmaceutics: official journal of Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik e.V*, 57, pp.411–20.
- Perlovich, G.L., Rodionov, S. V. and Bauer-Brandl, A., (2005), Thermodynamics of solubility, sublimation and solvation processes of parabens. *European Journal of Pharmaceutical Sciences*, 24, pp.25–33.
- Ruchelman, M.W., (1967), Solubility studies of estrone in organic solvents using gas-liquid chromatography. *Analytical biochemistry*, 19, pp.98–108.
- Ruether, F. and Sadowski, G., (2009), Modeling the solubility of pharmaceuticals in pure solvents and solvent mixtures for drug process design. *Journal of pharmaceutical sciences*, 98, pp.4205–15.
- Staverman, A.J., (1950), The entropy of high polymer solutions. Generalization of formulae. *Recueil des Travaux Chimiques des Pays-Bas*, 69, pp.163–174.
- Tsivintzelis, I., Economou, I.G. and Kontogeorgis, G.M., (2009), Modeling the Phase Behavior in Mixtures of Pharmaceuticals with Liquid or Supercritical Solvents. *The Journal of Physical Chemistry B*, 113, pp.6446–6458.
- Wang, S., Song, Z., Wang, J., Dong, Y. and Wu, M., (2010), Solubilities of ibuprofen in different pure solvents. *Journal of Chemical and Engineering Data*, 55, pp.5283–5285.
- Weidlich, U. and Gmehling, J., (1987), A modified UNIFAC Model. 1. Prediction of VLE, hE, and  $y^\infty$ . *Industrial and Engineering Chemistry Research*, 26, pp.1372–1381.
- Yan, F., Chen, L., Liu, D., Si Ma, L.-F., Chen, M.-J., Shi, H., Zhu, J.-X. and College, (2009), Solubility of (+)-(S)-2-(6-Methoxynaphthalen-2-yl) Propanoic Acid in Acetone,

Methanol, Ethanol, Propan-2-ol, and Ethyl Ethanoate at Temperatures between (278 and.  
*Journal of Chemical and Engineering Data*, 54, pp.1117–1119.

## CHAPTER FOUR

### **Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations**

#### *Abstract*

The Krill Herd optimization technique, which is based on the simulated herding behaviour of the krill crustacean, is applied to calculations involving phase equilibrium and phase stability, as the application of this emerging technique is extremely limited in the literature. In this work, the Krill Herd algorithm (KH) and the modified Lévy-Flight Krill Herd algorithm (LKH) has been applied to phase stability (PS) and phase equilibrium calculations in non-reactive (PE) and reactive (rPE) systems, where global minimization of the total Gibbs energy is necessary. Several phase stability and phase equilibrium systems were considered for the analysis of the performance of the technique that includes both vapour and liquid phase conditions.

The Krill Herd algorithm was found to reliably determine the desired global optima in PS, PE and rPE problems with generally higher success rates and lower computing time requirements than previously applied metaheuristic techniques such as those involving swarm intelligence and genetic and evolutionary algorithms.

#### *4.1 Introduction*

The behaviour of a particular system of components at a fixed temperature ( $T$ ), pressure ( $P$ ) and overall composition ( $z_i$ ) is essential for the design and simulation of most processes involving the system. Particularly from a chemical engineering stand point, the knowledge of the phase behaviour of a system of components is imperative for the design and simulation of the majority of separation processes. Such processes can include complex techniques such as reactive distillation and supercritical extraction, where the accuracies and reliabilities of phase behaviour calculations have a strong influence on the simulation results (Seider and Widagdo, 1996).

The phase behaviour of a mixture at a fixed  $T, P, z_i$ , is generally characterized by phase stability (PS) and phase equilibrium (PE) (Wakeham and Stateva, 2004). The phase stability (PS) problem requires the calculation of the number of phases a particular closed system at a given

$T$ ,  $P$  and  $z_i$  will form in order to achieve the state of its lowest total Gibbs energy ( $G^t$ ). A system is stable if, for instance, a mixture of two liquids forms a single liquid phase at a fixed temperature, pressure and overall composition.

A consequence of instability is of course the formation of two or more phases that equilibrate at the closed system temperature and pressure. This leads to the phase equilibrium (PE) calculation where the composition and volume of each phase must be determined to characterize the phase behaviour of the entire system.

It is also possible that a chemical reaction may occur between the constituents of the mixture under consideration. In such cases additional species are generated by the chemical reaction. The phase behaviour is thus influenced by the reaction kinetics, and conversely, the reaction equilibrium is influenced by the phase behaviour (composition and volume at phase equilibrium). These type of systems exhibit simultaneous chemical (reactive) and physical (phase) equilibrium, (rPE).

The phase stability problem requires the global minimization of the Tangent Plane Distance Function (TPDF), (discussed in section 3 of this chapter), which is used to indicate the stability of a system for a given  $T$ ,  $P$  and  $z_i$ . The PE and rPE problems require the global minimization of the total Gibbs free energy. These optimizations have proven to be challenging (Zhang et al., 2011). This is due to several factors; firstly, the number and types of phases (vapour/liquid) are not known before the optimization procedure. Secondly, the non-linearity of the various thermodynamic models usually applied to PE modelling such as cubic equations of state and complex activity coefficient models, infer local minima with objective function values very close to the global minima, especially near the critical region and phase boundaries. Furthermore, non-physical and trivial solutions are often possible at local minima. The consequences of erroneously considering local minima as the global minimum can lead to, for instance, prediction of false phase splits, as discussed by Gau et al. (2000) and Ohanomah and Thompson (1984). Thirdly, in systems exhibiting multiple phases such as vapour-liquid-liquid at equilibrium, a large difference in the Gibbs energy exists between the vapour and liquid phases, with much smaller differences in the Gibbs energy between the two liquid phases. Consequently optimization techniques struggle to locate the global minimum due to the large variance in the orders of magnitude of the terms that comprise the objective function.

Numerous global optimization methods are available in the literature (Land and Doig, 1960, Kirkpatrick et al., 1983, Dorigo, 1992, Duan et al., 1992, Hansen and Ostermeier, 2001, Mordecai, 2003, Walster and Hansen, 2004, Srinivas and Rangaiah, 2007, Yang, 2010, Yang and Deb, 2010, Walton et al., 2011, Gandomi and Alavi, 2012, Wang et al., 2013), and are generally classified as either deterministic or stochastic. Deterministic methods often require a large amount of computational time, as well as restrictions on the continuity and convexity of the objective function such as with cutting plane (Mordecai, 2003), branch and bound (Land and Doig, 1960) and interval analysis algorithms (Walster and Hansen, 2004).

Conversely the stochastic methods require very limited information on the nature of the optimization problem, and are able to handle issues pertaining to discontinuity and convexity. The computing time is generally reasonable and convergence to the global optimum is highly probable.

Metaheuristic optimization techniques are a sub category of the stochastic methods and involve an intelligent selection of random variables, often modelled around natural activities such as the cooling and heating of metal (simulated annealing, Kirkpatrick et al., 1983), the evolution of a species (differential evolution, Srinivas and Rangaiah, 2007), the swarm intelligence of insects (ant colony Dorigo, 1992, and firefly algorithms Yang, 2010), or the reproduction strategy of cuckoos (Cuckoo Search) (Walton et al., 2011, Yang and Deb, 2010). Each technique has its own strengths and weaknesses, the revision of which is beyond the scope of this work. Rashedi et al. (2009) state that, to date (2009), no single stochastic technique is capable of solving all optimization problems of different types and structures. The Krill Herd algorithm introduced by Gandomi and Alavi (2012) is a metaheuristic based on the simulation of the behaviour of a herd of the Antarctic krill crustacean, and its response to certain environmental and biological processes such as predation, general movement, foraging for food and natural drifting. The method has been successfully applied to several benchmark optimization problems (Gandomi and Alavi, 2012), however to the knowledge of the authors in this study; the application of the Krill Herd algorithm to problems involving complex phase stability and equilibria is not available in the literature.

## 4.2 Methods: Krill Herd Algorithm

### 4.2.1 Original Krill Herd Algorithm (KH)

The propensity of an individual Antarctic krill to thrive is dependent on its location relative to the bulk krill herd. A large herd density increases the chance of finding food locations as more individuals are searching the same target location. The behaviour of krill herds has been studied for several decades and numerous models describing the motion of the krill herd and krill individuals are available in the literature, e.g. Miller and Hampton (1989) and Hofmann et al. (2004). Gandomi and Alavi (2012) have developed an optimization algorithm that simulates the movement of a krill organism along a path that will eventually lead the organism to an optimal location within a krill herd, where it has the best chance of survival.

The location of a krill individual relative to the krill herd is dependent on three factors; the motion of the other individuals within the herd, the motion of an individual while foraging for food, and the physical diffusion of a krill individual. The position of the krill at any given time is thus given by the Lagrangian model:

$$\frac{dX_i}{dt} = N_i + F_i + D_i \quad (4.1)$$

Where  $X_i$  is the krill position,  $t$  is the time,  $N_i$  is the position of the other krill individuals (herd distribution),  $F_i$  is the foraging motion and  $D_i$  is the random diffusion. Equation (4.1) is solved iteratively during the optimization procedure.

The krill herd distribution,  $N_i$ , at the iteration  $k + 1$  is given by the following expression:

$$N_i^{k+1} = 0.01 \left[ \left( \sum_{j=1}^{NN} \hat{K}_{i,j} \hat{X}_{i,j} \right) + C_{best} \hat{K}_{i,best} \hat{X}_{i,best} \right] + rand \in [0,1] \times N_i^k \quad (4.2)$$

Where  $\hat{X}_{i,j}$  is used to quantify the relative attractive or repulsive tendencies between two krill individuals,  $i$  and  $j$  given by:

$$\hat{X}_{i,j} = \frac{X_j - X_i}{\|X_j - X_i\| + \sigma_1} \quad (4.3)$$

Where  $X$  is the relative position of a krill individual  $i$  or  $j$  in the herd, and  $\sigma_1$  is a small positive parameter, used to avoid singularities. The parameter  $X$  is essentially the manipulated parameter or vector of parameters for the optimization, and the intention of the algorithm is to determine the optimum position  $X$ , for the  $i$ th krill individual, to maximize its chance of survival.

$\widehat{K}_{i,j}$  represents the relative fitness of the evaluated objective function,  $K$  at krill positions  $X_i$  and  $X_j$ , with respect to the overall worst ( $K_{worst}$ ) and best ( $K_{best}$ ) objective function solutions obtained, given by:

$$\widehat{K}_{i,j} = \frac{K_i - K_j}{K_{worst} - K_{best}} \quad (4.4)$$

Upon termination of the optimization procedure, the parameter  $K_{best}$  is calculated from the optimal krill position,  $X_{best}$ , and represents the magnitude of the objective function that is closest to the optimal solution of the problem.

The parameter  $C_{best}$  is the effective coefficient of the krill individual yielding the best objective function to the  $i$ th individual and is calculated by:

$$C_{best} = 2(rand \in [0,1] + \frac{k}{k^{max}}) \quad (4.5)$$

Where  $k^{max}$  is the maximum number of allowable iterations for the calculation.  $\widehat{K}_{i,best}$  and  $\widehat{X}_{i,best}$  are the relative objective function and krill position of the  $i$ th individual with respect to the overall best location of any individual in the herd at the current iteration. The parameter  $NN$  is the herd size.

The foraging motion of a krill individual,  $F_i$ , for iteration  $k + 1$  is given by the following expression:

$$F_i^{k+1} = 0.02(C_{food}\widehat{K}_{i,food}\widehat{X}_{i,food} + \widehat{K}_{i,ibest}\widehat{X}_{i,ibest}) + rand \in [0,1] \times F_i^k \quad (4.6)$$

Where  $\hat{X}_{i,food}$  is the relative position of a krill individual  $i$ , in relation to the food location  $X_{food}$ , given by:

$$X_{food} = \frac{\sum_{i=1}^{NN} \frac{X_i}{K_i}}{\sum_{i=1}^{NN} \frac{1}{K_i}} \quad (4.7)$$

$\hat{K}_{i,food}$  is the fitness of the evaluated objective function of a krill individual  $i$ , relative to the food location, and can be calculated by equation (4.3), using  $X_{food}$  to evaluate  $K_{food}$ .

The parameter  $C_{food}$  is the food attraction coefficient, given by:

$$C_{food} = 2\left(1 - \frac{k}{k_{max}}\right) \quad (4.8)$$

The parameters  $\hat{K}_{i,ibest}$  and  $\hat{X}_{i,ibest}$  are the relative objective function and krill position of the  $i$ th individual with respect to the overall best location of the same individual,  $i$ , in the herd at any of its previous iterations,  $k$ .

The random diffusive motion,  $D_i$ , is given by:

$$D_i = rand \in [0.002,0.01] \times \left(1 - \frac{k}{k_{max}}\right) \times rand \in [-1,1] \quad (4.9)$$

The foraging and herd position term for the krill individuals both contain dual local and global strategies for optimization, while the diffusive motion allows for random search. In order to improve the performance of the KH algorithm, genetic operators were introduced by the original authors. These include *crossover*, where a new krill individual is “reproduced” as a result of a combination of two other krill, and *mutation*, which allows for genetic diversity in the next generation of the krill. The application of these parameters are explained in detail by Gandomi and Alavi (2012). A pseudo code of the KH algorithm developed for this work is provided in Figure 4.1(a).



**(a) Begin**

Initialize variables: herd size, solution boundaries, lower bound (LB), upper bound (UB) stopping conditions (SC), tolerance (tol)

Create data structures: initial individual krill position, ( $X_{initial}$ ), initial foraging motions,  $F_{initial}$ , initial herd distribution,  $N_{initial}$ , initial random diffusion,  $D_{initial}$

Induce predation  $\rightarrow K_{best} = 1E10$ ; % best solution set % to high value

$$X_{initial} = LB + rand \in [0,1](UB - LB)$$

(randomly selected between parameter bounds)

$$X = X_{initial}$$

while  $K > tol$  &  $K^{k+1} - K^k > 0$  for SC iterations

for  $i = 1 : herd\ size$

$$\frac{dX_i}{dt} = N_i + F_i + D_i$$

% Update krill position

$$X(i)^{k+1} = X(i)^k + \Delta t \times \frac{dX_i}{dt}$$

Calculate  $K(X(i)^{k+1})$

end

\*\*\*

$$k = k + 1$$

end

**(b) Begin**

\*\*\*

$u = h^{-2}$  Induce Pareto distribution using the survival function

Determine step size  $dx$  by performing Lévy flight

$$r = [rand \in [0,1] * herd\ size]$$

for  $i = 1 : herd\ size$

if  $rand \in [0,1] < 0.5$

$$L^{k+1}(i) = u \times dx(i) + X(i)_{herd\ size-r+1}^{k+1}$$

else

$$L^{k+1}(i) = u \times dx(i) - X(i)_{herd\ size-r+1}^{k+1}$$

end

end

if  $K(L^{k+1}(i)) < K(X^{k+1}(i))$

$$X(i)^{k+1} = L^{k+1}(i)$$

else

$$X(i)^{k+1} = X(i)^{k+1}$$

end

**End**

**Figure 4.1 Pseudo code for (a) Krill Herd Algorithm and (b) Lévy Flight Krill Herd Algorithm, implemented in this work.**

#### 4.2.2 Lévy Flight Krill Herd Algorithm (LKH)

Many biologists believe that the selection of the path of movement of certain animal species towards optimum locations such as food centres must be due to an efficiency strategy. Due to natural selection, only those species that follow efficient paths to optimum locations will survive and reproduce. One such efficient path is that characterized by a Lévy flight, walk or indeed swim.

Wang et al. (2012) have introduced a modification to the original Krill Herd algorithm that introduces a Lévy flight efficiency into the iteration process. The Lévy flight distribution is given by the survivor function ( $u$ ) which has infinite variance and infinite mean:

$$u = \mathcal{O}h^{-\lambda} \quad \text{Where } 1 < \lambda \leq 3 \quad (4.10)$$

Here  $h$  is the generation of the krill organism, and  $\mathcal{O}$ , is the big O. The Lévy flight is principally a power-law distribution with a heavy tail. The introduction of the Lévy flight in the Krill Herd algorithm allows for a selection of new solutions in the area around a current best solution (local minimum), while still allowing for random searches in areas far away from the current best solution. Wang et al. (2012) have stated that the LKH algorithm performs better than or as well as the KH algorithm for the benchmark problems considered in their work. A pseudo code of the incorporation of Lévy Flights into the KH algorithm developed for this work is provided in Figure 4.1(b).

In this study, both the original KH and LKH algorithms were tested for their performance in calculating the global minima in several benchmark phase equilibria and stability problems that are prevalent in the literature.

#### 4.3 Theory: Optimization problems associated with phase stability and equilibrium

A brief description of the formulation of the optimization problems involved in stability and phase equilibrium calculations follows. A more detailed description of the various formulations and objective functions that have been explored in the literature is available for example in (Hua et al., 1998, Jalali and Seader, 1999, Lee et al. 1999, Harding and Floudas, 2000, Wasykiewicz and Ung, 2000, Rangaiah, 2001, Bonilla-Petriciolet, 2006., Bonilla-Petriciolet

et al., 2008a, Bonilla-Petriciolet et al., 2008b, Bonilla-Petriciolet and Segovia-Hernández, 2010, Bonilla-Petriciolet et al., 2011, Bhargava et al., 2013).

#### 4.3.1 Phase Stability (PS)

The phase stability (PS) problem is generally the precursor to phase equilibrium calculations, as it specifies the condition and number of phases that occur at equilibrium. Gibbs (1873) states that if the tangent plane generated at the feed composition lies below the molar Gibbs energy surface for all compositions, then the phase is stable. Michelsen (1982) quantified the difference between the tangent plane and molar Gibbs energy surface, as the Tangent Plane Distance Function (TPDF). The TPDF is defined as the vertical distance between the molar Gibbs energy surface and the tangent plane at the initial composition  $z_i$ . The TPDF is given by:

$$TPDF = \sum_{i=1}^c y_i (\mu_i|_y - \mu_i|_z) \quad (4.11)$$

Where  $c$  is the number of components in the mixture,  $z$  and  $y$  are the initial and trial compositions respectively and  $\mu_i|_{\Theta}$  is the chemical potential calculated at the composition  $\Theta$ . In order to determine if instability exists, TPDF must be globally minimised; if the resulting TPDF is less than zero, than a phase split will occur, as the system is not stable. Phase split calculations must then be performed in order to determine the composition of the phases generated.

In cases of two phases the number of moles of component  $i$  in phase  $y$  ( $n_{iy}$ ) can be expressed as some fraction of the initial number of moles of component  $i$ , ( $z_i n_T$ ). Hence:

$$n_{iy} = \varepsilon_i z_i n_T \quad (4.12)$$

Where  $\varepsilon_i$  is the fraction of component  $i$  in phase  $y$ , and  $n_T$  is the total number of moles in the feed mixture. In the case of optimization by the Krill Herd Algorithm, the parameters  $\varepsilon$  are represented by the krill position vector  $X$ , defined earlier, with the optimal solution of  $\varepsilon$  given by  $X_{best}$  upon termination of a succesful optimization.

The composition of phase  $y$  can then be calculated from:

$$y_i = \frac{n_{iy}}{\sum_{i=1}^c n_{iy}} \quad (4.13)$$

$\varepsilon_i$ , which is  $\varepsilon[0,1]$ , then becomes the decision variable for the optimization, where the minimization required is:

$$\begin{aligned} \min_{\varepsilon_i} TPDF & \quad \text{where } i=1, \dots, c & (4.14) \\ 0 \leq \varepsilon_i \leq 1 & \end{aligned}$$

In the case of optimization by the Krill Herd-based Algorithms the minimization of the TPDF is represented by the objective function  $K$ , defined earlier, with the global minimum of the TPDF corresponding to  $K_{best}$  upon termination of a successful optimization.

The chemical potential of a mixture component relative to the pure component in the saturated state at the system temperature can be calculated using one of the following two relations for either vapour-liquid equilibrium or liquid-liquid equilibrium calculations

$$\frac{\mu_i - \mu_i^0}{RT} = \ln \left( \frac{x_i \hat{\phi}_i}{\phi_i} \right) \quad (4.15)$$

$$\frac{\mu_i - \mu_i^0}{RT} = \ln(x_i \gamma_i) \quad (1.22)$$

Where  $\mu_i^0$  is the pure component chemical potential,  $\hat{\phi}_i$  is the fugacity coefficient in solution of component  $i$ ,  $\phi_i$  is the fugacity coefficient of component  $i$ , and  $\gamma_i$  is the activity coefficient of component  $i$ . Equation (4.15) applies to both vapour and liquid phases for vapour-liquid equilibrium calculations. The reference state is the pure component  $i$  in a perfect gas state at a given temperature and pressure. The fugacity coefficient and fugacity coefficients in solution are determined by applying an appropriate equation of state in this work. Equation (1.22) applies to both liquid phases in a liquid-liquid equilibrium problem. The reference state is the pure component  $i$  in the liquid state at a given temperature and pressure, and is applied to the calculation of liquid-liquid equilibrium via a relevant activity coefficient model in this work.

#### 4.3.2 Phase Equilibrium (PE)

If a mixture at a fixed temperature and pressure separates into two or more phases, that is, it initially proved to be unstable, it becomes necessary to calculate the phase composition of each

of the phases generated. This is usually performed via the constrained minimization of the reduced molar Gibbs free energy, where the conservation of mass must of course be adhered to. The reduced molar Gibbs free energy  $G$  of subcritical components relative to the pure components is given by:

$$G = \sum_{j=1}^{\pi} \sum_{i=1}^c n_{ij} \ln(x_{ij}\gamma_{ij}) = \sum_{j=1}^{\pi} \sum_{i=1}^c n_{ij} \ln\left(\frac{x_{ij}\hat{\phi}_{ij}}{\phi_i}\right) \quad (4.16)$$

Where  $\pi$  is the number of phases present at equilibrium and  $c$  is the total number of components in the system. The subscript  $ij$  denotes a property of component  $i$  in phase  $j$ .

The minimization of  $G$  is constrained by mass, which for each species  $i$  is:

$$\sum_{j=1}^{\pi} n_{ij} = z_i n_T \quad \text{Where } i = 1, \dots, c \quad (4.17)$$

With 
$$0 \leq n_{ij} \leq z_i n_T \quad \text{Where } i = 1, \dots, c \text{ and } j = 1, \dots, \pi \quad (4.18)$$

The optimization can be converted into an unconstrained optimization if, ( $\zeta_{ij} \in [0,1]$ ) decision variables for the optimization algorithm are used, that describe the fraction of component  $i$  present in each phase  $j$ . That is for instance,  $\zeta_{11}$  is the fraction of the total number of moles of component 1 present in phase 1 and  $\zeta_{12}$  is the fraction of the remaining number of moles (the total number of moles of component 1 less the number of moles of component 1 in phase 1) of component 1 present in phase 2. By introducing this term it is assured that all candidate solutions of the number of moles of each component in a phase will be physically realistic and hence computation time is decreased. The decision variables impose the mass balance restriction, eliminating the need for a constrained optimization. The number of moles of each component in each of the phases can then be given by:

$$n_{i1} = \zeta_{i1} z_i n_T \quad \text{Where } i = 1, \dots, c \quad (4.19)$$

For any number of phases, this leads to:

$$n_{ij} = \zeta_{ij} (z_i n_T - \sum_k^{j-1} n_{ik}) \quad \text{Where } i = 1, \dots, c \text{ and } j = 2, \dots, \pi - 1 \quad (4.20)$$

$$n_{i\pi} = z_i n_T - \sum_k^{\pi-1} n_{ik} \quad \text{Where } i = 1, \dots, c \quad (4.21)$$

The minimization problem for PE is then defined as:

$$\begin{aligned} \min_{\zeta_{ij}} G \quad & \text{Where } i=1, \dots, c \text{ and } j = 1, \dots, \pi - 1 \quad (4.22) \\ 0 \leq \zeta_{ij} \leq 1 \end{aligned}$$

The summation of equation (4.20) over all phases is shown in the supplemental data (Appendix A), along with the equivalence of equation (4.18) and (4.20).

In order to solve this optimization problem the number and type of each phase present at equilibrium must be known *a priori*. With Krill Herd Algorithm, the decision variables,  $\zeta$ , are represented by the krill position vector  $X$ , while the minimization of  $G$  is represented by the objective function  $K$ . For a successful optimization, the global minimum of  $G$  is obtained upon termination, when  $G = K_{best} f(X_{best})$ . The benchmark test cases of phase stability and phase equilibrium considered in this work, are presented in Table 4.1.

**Table 4.1 Benchmark test cases for phase stability (PS) and phase equilibria (PE) problems with literature sources**

No.	System	Feed conditions	Thermodynamic models	Global Optimum		Reference
				PE	PS	
1	n-Butyl Acetate + Water (Liquid-liquid equilibrium)	$n_F = (0.5, 0.5)$ at 298 K and 101.325 kPa	NRTL model parameters from (Heideman and Mandhane, 1973)	-0.020198	-0.032466	(McDonald and Floudas, 1997)
2	Toluene + Water + Aniline (Liquid-liquid equilibrium)	$n_F = (0.29989, 0.20006, 0.50005)$ at 298 K and 101.325 kPa	NRTL model parameters from (Bender and Block, 1975)	-0.352957	-0.294540	(McDonald and Floudas, 1995, Rahman et al., 2009)
3	Nitrogen + Methane + Ethane (Vapour-liquid equilibrium)	$n_F = (0.3, 0.1, 0.6)$ at 270 K and 7600 kPa	SRK EoS with classical mixing rules model parameters from (Hua et al., 1998)	-0.547791	-0.015767	(Nichita et al., 2002)
4	Methane + Hydrogen Sulphide (Vapour-liquid equilibrium)	$n_F = (0.9813, 0.0187)$ at 190 K and 4053 kPa	SRK EoS with classical mixing rules model parameters from (Hua et al., 1998)	-0.019892	-0.003932	(Sun and Seider, 1995, Balogh et al. 2003)
5	Ethane + Propane + n-Butane + n-Pentane + n-Hexane (Vapour-liquid equilibrium)	$n_F = (0.401, 0.293, 0.199, 0.0707, 0.0363)$ at 390 K and 5583 kPa	SRK EoS with classical mixing rules (all interactions set to 0)	-1.183653	-0.000002	(Ammar, and Renon, 1987)
6	Methane + Ethane + Propane + n-Butane + n-Pentane + n-Hexane + n-Heptane - n-Hexadecane + >n-Heptadecane (Vapour-liquid equilibrium)	$n_F = (0.7212, 0.09205, 0.04455, 0.03123, 0.01273, 0.01361, 0.07215, 0.01248)$ at 353 K and 38,500 kPa	SRK EoS with classical mixing rules (all interactions set to 0)	-0.838783	-0.002688	(Harding and Floudas, 2000)
7	Methane + Ethane + Propane + i-Butane + n-Butane + i-Pentane + n-Pentane + n-Hexane + i-Pentadecane (Vapour-liquid equilibrium)	$n_F = (0.614, 0.10259, 0.04985, 0.008989, 0.02116, 0.00722, 0.01187, 0.01435, 0.16998)$ at 314 K and 2010.288 kPa	SRK EoS with classical mixing rules (all interactions set to 0)	-0.769772	-1.486205	(Bonilla-Petriciolet et al., 2006)
8	Methane + Ethane + Propane + n-Butane + n-Pentane + n-Hexane + n-Heptane + n-Octane+ n-Nonane + n-Decane (Vapour-liquid equilibrium)	$n_F = (0.6436, 0.0752, 0.0474, 0.0412, 0.0297, 0.0138, 0.0303, 0.0371, 0.0415, 0.0402)$ at 435.35 K and 19,150 kPa	SRK EoS with classical mixing rules (all interactions set to 0)	-1.121176	-0.000021	(Bonilla-Petriciolet et al., 2006)

### 4.3.3 Reactive Phase Equilibrium (rPE)

In reactive phase equilibrium (rPE) problems, the Gibbs free energy change is due to the equilibrium of phases as well as chemical reaction. The constrained minimization of the Gibbs free energy is carried out, which is again subject to mass balance and reaction equilibrium limitations. Several authors have applied the constrained approach, using equilibrium constants to describe the Gibbs free energy change of reaction that include for example (Bonilla-Petriciolet and Segovia-Hernández, 2010, Bonilla-Petriciolet et al., 2011, Zhang et al., 2011, Bhargava et al., 2013). For multi-component and multiphase reactive systems of  $r$  reactions, with simultaneous phase equilibrium over  $\pi$  phases, the total reduced molar Gibbs free energy ( $G_T$ ) is given by:

$$G_T = G - \sum_{j=1}^{\pi} \ln K^{eq} \mathbf{N}^{-1} \mathbf{n}_{ref,j} \quad (4.23)$$

Where  $G$  is given by equation (4.16),  $\mathbf{K}^{eq}$  is a vector of equilibrium constants for the  $r$  independent reactions,  $\mathbf{N}$  is a square matrix comprised of the stoichiometric coefficients of the  $r$  reference components in the  $r$  reactions and  $\mathbf{n}_{ref}$  is a vector of the molar compositions in terms of the reference components.

From the mass balance it can be shown that the number of moles of component  $i$  in a particular phase,  $\pi$ , at equilibrium is equal to the initial number of moles of component  $i$ , less the moles consumed by reaction, and the total moles of  $i$  present in the other phases:

$$n_{i\pi} = z_i n_T - \mathbf{v}_i \mathbf{N}^{-1} (\mathbf{n}_{ref,F} - \mathbf{n}_{ref,\pi}) - \sum_{j=1}^{\pi-1} (n_{ij} - \mathbf{v}_i \mathbf{N}^{-1} \mathbf{n}_{ref,j}) \quad (4.24)$$

Where  $i = 1, \dots, c - r$

Equation (4.24) constitutes the optimization constraint on the decision variables. For rPE problems  $n_{i\pi}$  are directly implemented as the decision variables with the obvious constraint of  $n_{i\pi} \geq 0$  and are represented by the krill position vector  $X$ , in the case of the Krill Herd Algorithm.

The minimization problem for rPE is then defined as:



$$\begin{aligned} \min_{n_{i\pi}} G_T \quad & \text{Where } i=1, \dots, c & (4.25) \\ n_{i\pi} \geq 0 \end{aligned}$$

The minimization of  $G_T$  via the Krill Herd Algorithm corresponds to the objective function  $K$ , with the global optimum of a successful optimization of  $G_T$  occurring at  $K_{best}f(X_{best})$ . The benchmark test cases for reactive phase equilibrium problems considered in this work are provided in Table 4.2.

**Table 4.2 Benchmark test cases for reactive phase equilibria (rPE) problems with literature sources**

	System	Feed conditions	Thermodynamic models	Global optimum	Reference
1	Ethanol+ Acetic acid ↔ Ethyl acetate + Water (Reactive with liquid-liquid equilibrium)	$n_F = (0.5, 0.5, 0.0, 0.0)$ at 355 K and 101.325 kPa	NRTL model and ideal gas. $K_{eq} = 18.670951$	-2.058120	(Lee et al., 1999)
2	Isobutene + Methanol ↔ Methyl tert-butyl ether with n-Butane as an inert component (Reactive with liquid-liquid equilibrium)	$n_F = (0.3, 0.3, 0.0, 0.4)$ at 373.15 K and 101.325 kPa	Wilson model and ideal gas. $\Delta G_{rxs}^\circ/R = -4205.05 + 10.0982T - 0.2667T \ln T$ $\ln K_{eq} = -\Delta G_{rxs}^\circ/R$ where T is in K	-1.434267	McDonald and Floudas, 1997)
3	2-Methyl-1-butene + 2-Methyl-2-butene + 2*Methanol ↔ 2*Tert-amyl methyl ether (Reactive with liquid-liquid equilibrium)	$n_F = (0.354, 0.183, 0.463, 0.0)$ at 355 K and 151.95 kPa	Wilson model and ideal gas. $K_{eq} = 1.057 \times 10^{-04} e^{4273.5/T}$ where T is in K	-1.226367	(Ung and Doherty, 1995a)
4	Acetic acid + n-Butanol ↔ Water + n-Butyl acetate (Reactive with liquid-liquid equilibrium)	$n_F = (0.3, 0.4, 0.3, 0.0)$ at 298.15 K and 101.325 kPa	UNIQUAC model and ideal gas. $\ln K_{eq} = 450/T + 0.8$	-1.106300	(Wasykiewicz and Ung, 2000)
5	A1 + A2 ↔ A3 (Reactive with liquid-liquid equilibrium)	$n_F = (0.6, 0.4, 0.0)$	Margules solution model $G^E/RT = 3.6x_1x_2 + 2.4x_1x_3 + 2.3x_2x_3$ . $K_{eq} = 0.9825$	-0.144508	(Bonilla-Petriciolet et al., 2008a)
6	2-Methyl-1-butene + 2-Methyl-2-butene+ 2* Methanol ↔ 2*Tert-amyl methyl ether with n-Pentane as an inert component (Reactive with liquid-liquid equilibrium)	$n_F = (0.1, 0.15, 0.7, 0.0, 0.05)$ at 335 K and 151.9875 kPa	Wilson model and ideal gas. $K_{eq} = 1.057 \times 10^{-04} e^{4273.5/T}$ where T is in K	-0.872577	(Bonilla-Petriciolet et al., 2008a)
7	A1 + A2 ↔ A3 (Reactive with liquid-liquid equilibrium)	$n_F = (0.52, 0.48, 0.0)$ at 323.15 K and 101.325 kPa	Margules solution model. $K_{eq} = 3.5$	-0.653756	(Ung and Doherty, 1995b)
8	Propene + Water ↔ 2-Propanol (Reactive with vapour-liquid equilibrium)	$n_F = (0.37, 0.63)$ at 353.15K and 100 kPa	SRK EoS with conventional mixing rules and all interaction parameters equal to zero. $K_{eq} = 23$	-1.347900	(Stateva and Wakeham, 1997)

The fixed parameters required for optimization in each of the problem types considered in this work, and those determined by the Krill Herd or Levy Flight Krill Herd Algorithm, are summarized in Table 4.3.

**Table 4.3 Summary of fixed and variable parameters for the three types of optimization problems.**

Problem	Fixed parameters	Variable parameters determined by KH/LKH algorithm, $X_i$	Objective Function $K_i$
PS	$n_F, T, P$ , thermodynamic model parameters	$\varepsilon_i$ , the fraction of a component $i$ in a phase	$\min_{\varepsilon_i} TPDF$ where $i=1, \dots, c$ $0 \leq \varepsilon_i \leq 1$
PE	$n_F, T, P$ , thermodynamic model parameters	$\zeta_{ij}$ , the fraction of component $i$ in phase $j$	$\min_{\zeta_{ij}} G$ where $i=1, \dots, c$ and $j=1, \dots, \pi - 1$ $0 \leq \zeta_{ij} \leq 1$
rPE	$n_F, T, P$ , thermodynamic model parameters, reaction constants	$n_{i\pi}$ , the number of moles of a component $i$ in phase $\pi$ at equilibrium	$\min_{n_{i\pi}} G_T$ where $i=1, \dots, c$ $n_{i\pi} \geq 0$

#### 4.4 Results and discussion

##### 4.4.1 Algorithm development

The programs used for the Krill Herd and Lévy Flight Krill Herd optimization techniques were developed for this study using the MATLAB® (R2012b) programming language. No built-in optimization functions were used at any stage during the performance assessment of the KH and LKH algorithms. The parameter,  $\Delta t$ , defined in the original work of Gandomi and Alavi (2012), serves as a scaling parameter for the speed vector,  $\frac{dX_i}{dt}$ , as shown in Figure 4.1(a). This parameter is generally case-dependent, lying between the solution bounds of the individual problem, but the authors suggest selecting low values between 0 and 2 to encourage a thorough search of the solution space by a krill individual. In this work  $\Delta t$  was set prudently to 0.001 which guarantees a thorough search, at the expense of computation time.

Each of the benchmark phase analysis problems were solved 100 times, each with a random selection of the initial distribution of the krill organisms. This was done to ensure a reliable analysis of the performance of the algorithm. The performance of the algorithm was quantified

and assessed based on two factors; The success rate of achieving the global optimum (SR) and the computing time required to do so, quantified by the number of objective function evaluations (NFE). A constant stopping condition of an additional 50 iterations in the PS and PE problems and 25 in the rPE problems, within the specified tolerance, were used to ensure that the global minimum was successfully reached in each test case, for both methods tested. These stopping conditions have been shown to provide good convergence success rates for the benchmark test cases used in this work, as shown by (Zhang et al., 2011, Fateen et al., 2012 and Bhargava et al., 2013).

The effect of the krill herd size on the NFE required, and SR of obtaining the global minimum was tested. The size of the krill herd has an obvious implication on the success rate and ability of the herd as a whole finding optimum food locations and passing these characteristics on to future generations. The success rate is defined as the number of times the objective function has been satisfied for each of the 100 trials of each benchmark test system, where the objective function is given by:

$$|f^{Global} - f^{Calculated}| \leq \epsilon \quad (4.26)$$

Where  $f^{Global}$  is the global minimum of the benchmark test, as provided in Tables 4.1 and 4.2,  $f^{Calculated}$  is the calculated minimum, determined by the optimization algorithm and  $\epsilon$ , is the case specific tolerance.  $\epsilon$  was set to  $10^{-7}$  for the cases PS-5 and PS-7, as the global minima are of lower orders, while all other  $\epsilon$  were set to  $10^{-5}$ .

Due to the dual termination criteria, a successful trial is signified by convergence within the tolerance, as well as no further change after the stopping condition of an additional 50 (25 in rPE) iterations is satisfied.

In order to determine the efficiency of the methods for each type of phase problem, a Global Success Rate (GSR) was defined:

$$GSR = \sum_i^{NP} \frac{SR_i}{NP} \quad (4.27)$$

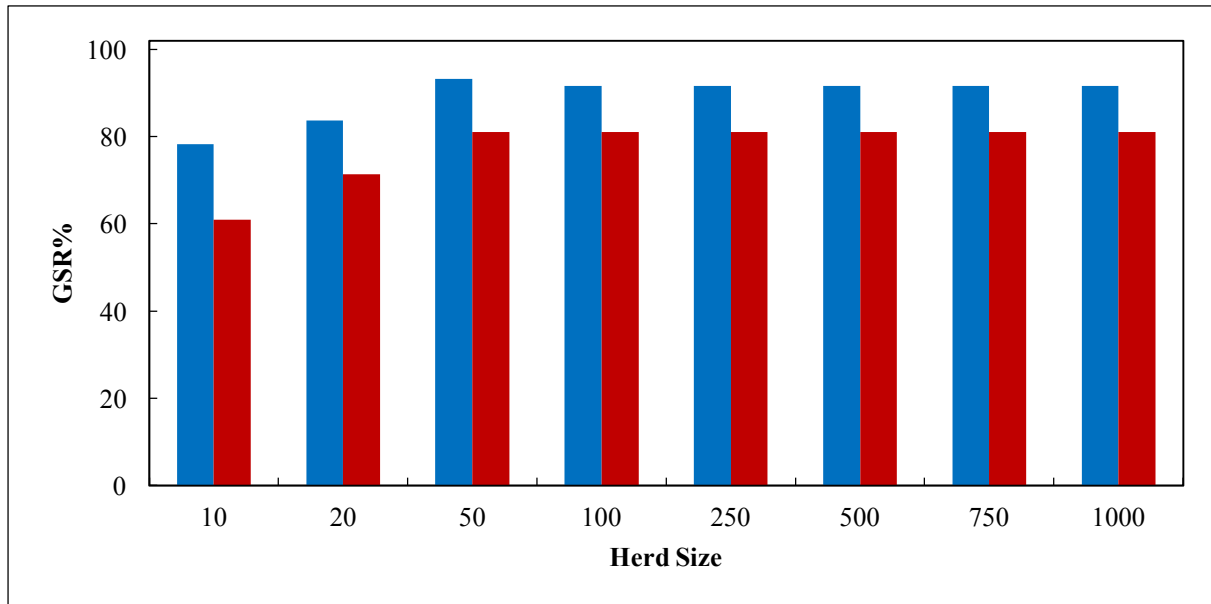
Where  $SR_i$  is the percentage success rate of each test problem, and  $NP$  is the total number of optimizations of each class of phase problem.

#### 4.4.2 Problems involving phase stability

Eight phase stability problems were considered in this work, which include systems forming liquid-liquid and vapour-liquid phases. The Non-Random Two-liquid (NRTL) (Renon, and Prausnitz, 1968) activity coefficient model and Soave-Redlich-Kwong (SRK) (Soave, 1972) equation of state with quadratic mixing were used to calculate activity and fugacity coefficients. These models are reviewed in detail by, for instance, Raal and Mühlbauer, (1998). The case of PS-2 (toluene + water + aniline at 298 K and 101.325 kPa) consists of three components with the formation of a maximum of three phases, according to the phase rule and can be used as a good example to illustrate the importance of the stability test. Paules and Floudas (1989) have reported the difficulties experienced with the convergence to the global minimum of this system, and the existence of false non-negative TPDFs (incorrectly indicating phase stability) for trivial trial compositions. KH and LKH do however correctly reveal the physical presence of multiple liquid phases (instability) at the specified temperature, pressure, and initial composition, which is indicated by a negative minimum TPDF.

Problem PS-7 provides a difficult test case, with a system of nine hydrocarbons, studied intensively by Castillo and Grossman (1981). The treatment of the KH-based algorithms of this problem with a substantial number of unknowns provides a good indication of the technique's proficiency.

The global success rates for the phase stability problems are shown in Figure 4.2. It is evident that the original KH algorithm outperforms the LKH method for all herd sizes tested.



**Figure 4.2 Global Success Rates of phase stability problems with increasing herd size, ■KH, ■LKH.**

Figures 4.3 to 4.5 show examples of convergence profiles for the PS problems. It is clear that the LKH algorithm converges within the tolerance much faster than the KH algorithm and with a fewer number of function evaluations. However, it would seem that due to the stopping criteria of 50 additional iterations after convergence within the tolerance, the KH algorithm tends to explore further for objective function values significantly below the specified tolerance.

Table 4.4 provides individual success rates for each stability problem for herd sizes of 10, 20 and 50. Beyond the herd size of 50, GSRs did not improve substantially. The reduced NFE required for LKH is evident in cases such as PS-1 and PS-6 where success rates very similar to the original KH algorithm were obtained with much lower computational times. The KH algorithm provides favourable success rates for the complex cases of PS-2 (81) and PS-7 (88) for a herd size of 50.

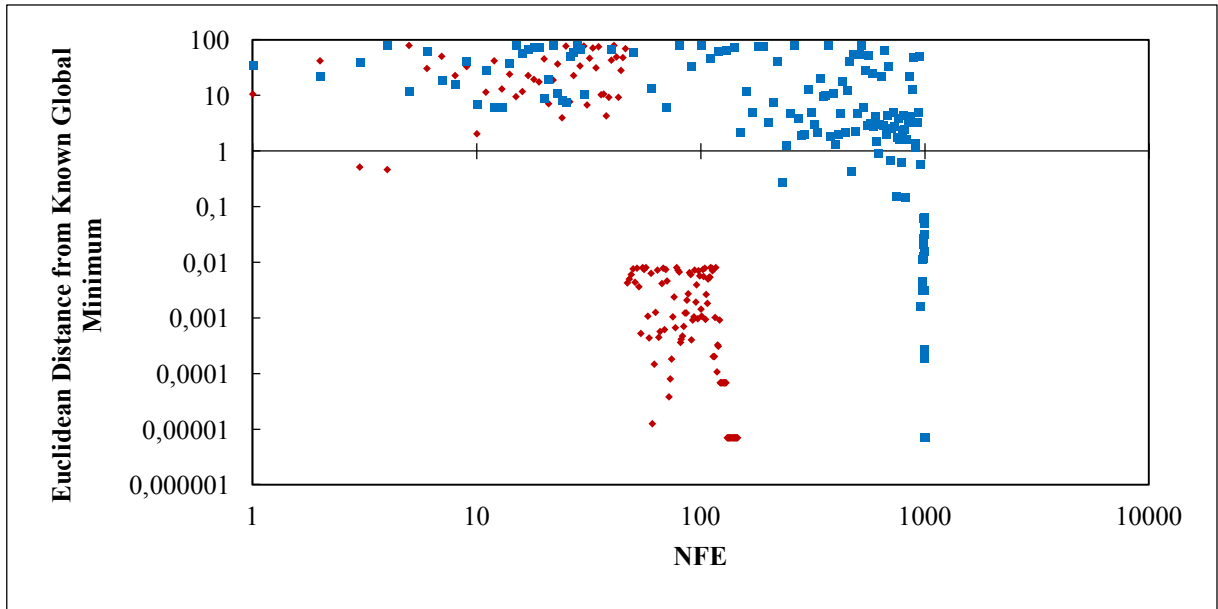


Figure 4.3 A convergence profile of PS-1 with a herd size of 20, ■KH, ◆LKH.

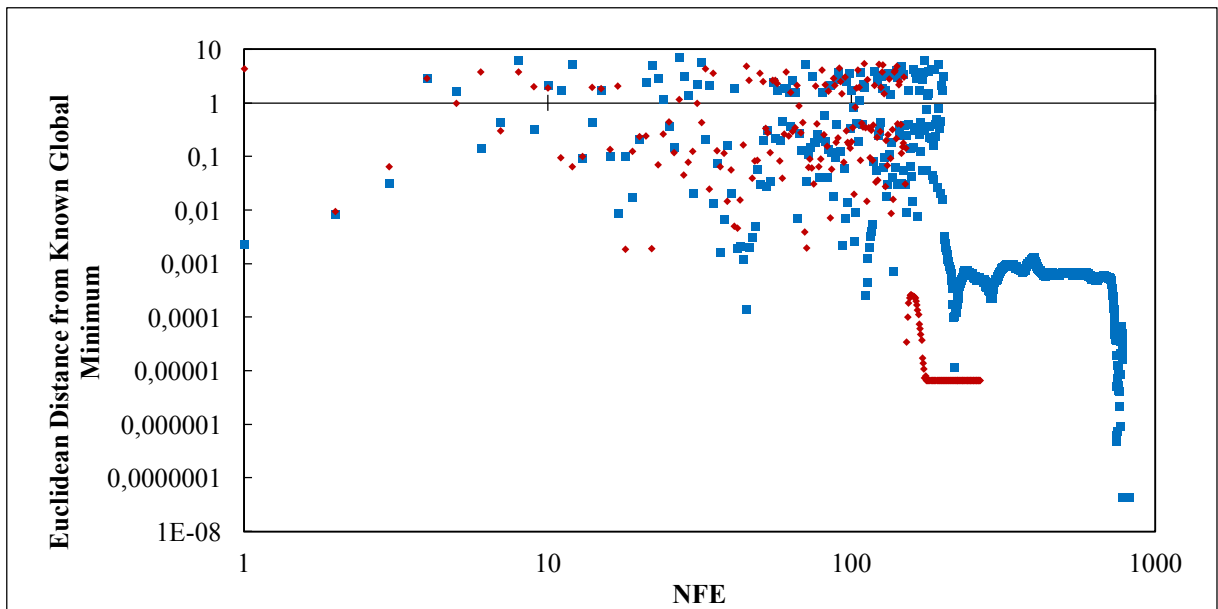


Figure 4.4 A convergence profile of PS-3 with a herd size of 20, ■KH, ◆LKH.

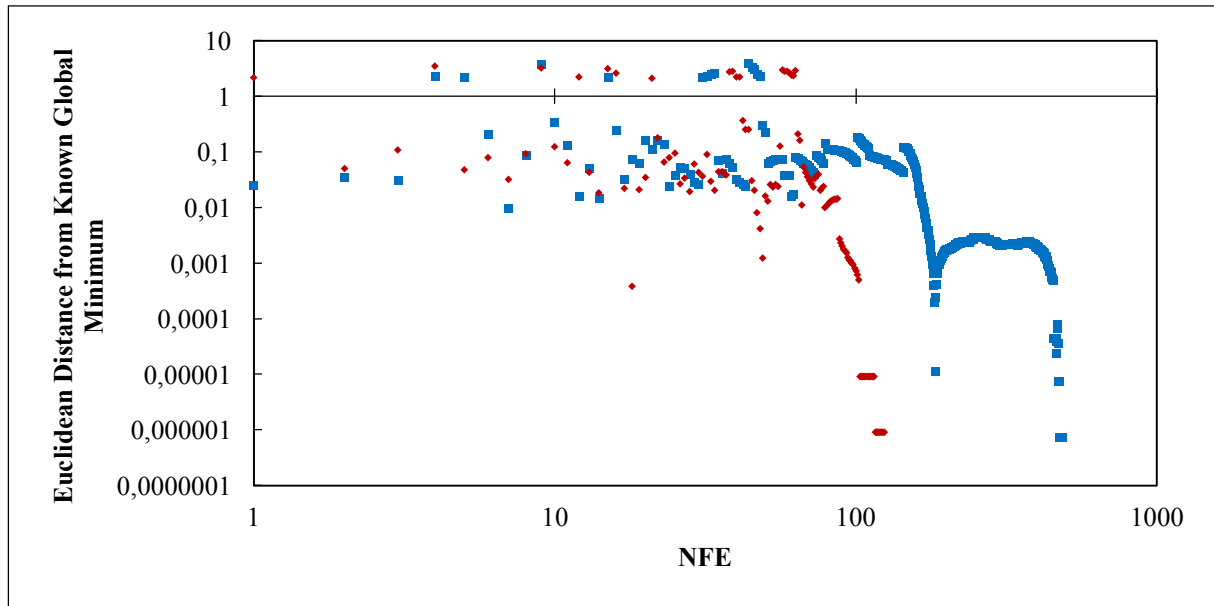


Figure 4.5 A convergence profile of PS-7 with a herd size of 20, ■KH, ◆LKH.



Table 4.4 Performance of the Krill Herd and Levy Flight Krill Herd Algorithms for problems involving phase stability.

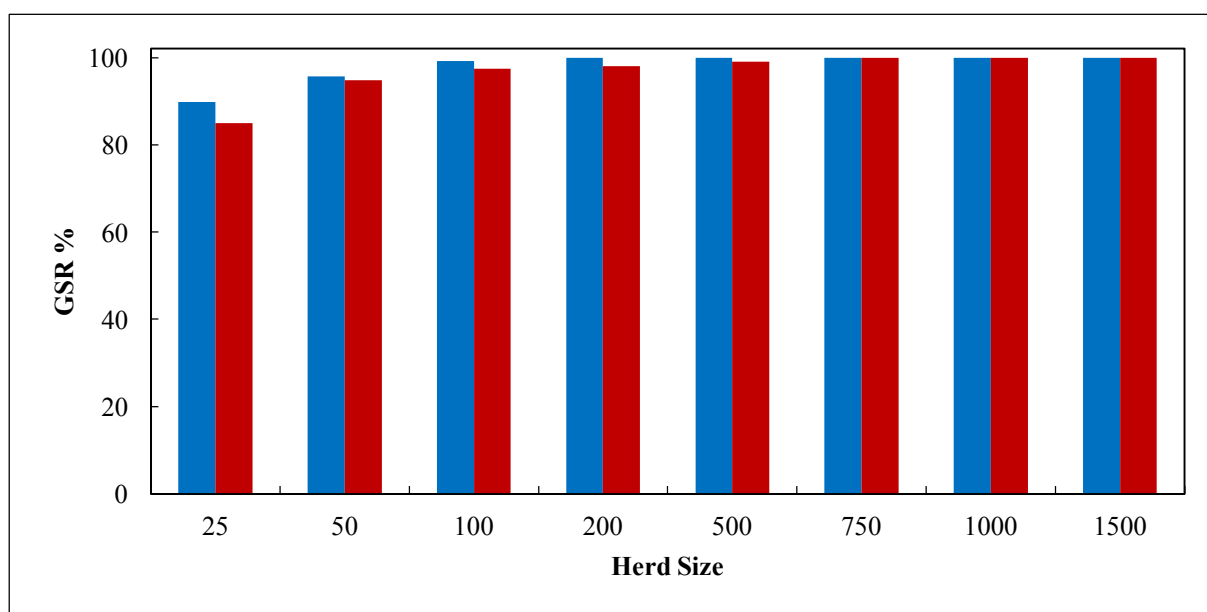
Problem	Herd Size	Algorithm			
		Krill Herd		Levy Flight Krill Herd	
		Success Rate (%)	NFE	Success Rate (%)	NFE
<b>PS-1</b>	10	90	1999	81	878
	20	94	1769	92	267
	50	95	377	94	106
<b>PS-2</b>	10	71	7689	64	3159
	20	80	4457	77	2739
	50	81	2487	84	1261
<b>PS-3</b>	10	100	12601	82	1687
	20	100	1124	87	466
	50	100	710	92	324
<b>PS-4</b>	10	89	14476	63	7845
	20	94	8187	88	3216
	50	97	4488	89	1567
<b>PS-5</b>	10	39	6784	14	4785
	20	42	3744	20	2415
	50	80	1398	54	996
<b>PS-6</b>	10	97	15492	96	8773
	20	99	6563	99	5877
	50	99	864	99	425
<b>PS-7</b>	10	59	1845	21	784
	20	62	561	29	112
	50	88	214	48	89
<b>PS-8</b>	10	70	5441	66	1465
	20	77	987	79	672
	50	78	712	88	143
<b>GSR %</b>		<b>83</b>		<b>71</b>	

#### 4.4.3 Problems involving phase equilibrium

The phase equilibrium problems considered are comprised of the same problems as the stability tests, with the NRTL model and SRK equation of state with quadratic mixing to calculate activity and fugacity coefficients. The binary interaction and model parameters used are available in the literature cited for each system in Table 4.1.

Global minima are of course now calculated for the equilibrium condition. The KH-based algorithms handle complex PE cases such as PE-2 and PE-8 with high success rates, by calculating the correct minimum Gibbs energy for each case, indicating the physical occurrence of phase equilibrium.

The global success rates of the phase equilibrium problems are shown in Figure 4.6. The KH algorithm outperforms the LKH algorithm, albeit narrowly. Again an increasing herd size proves to increase GSRs to a point.



**Figure 4.6 Global Success Rates of phase equilibrium problems with increasing herd size, ■KH, ■LKH.**

The convergence profiles for PE-2, 5 and 8 are shown as examples in Figures 4.7 to 4.9. For the majority of the PE test problems considered, LKH converged within the tolerance in a lower NFE, with PE-5 being the exception. In the example given, PE-8 has failed to converge to the desired tolerance for LKH.

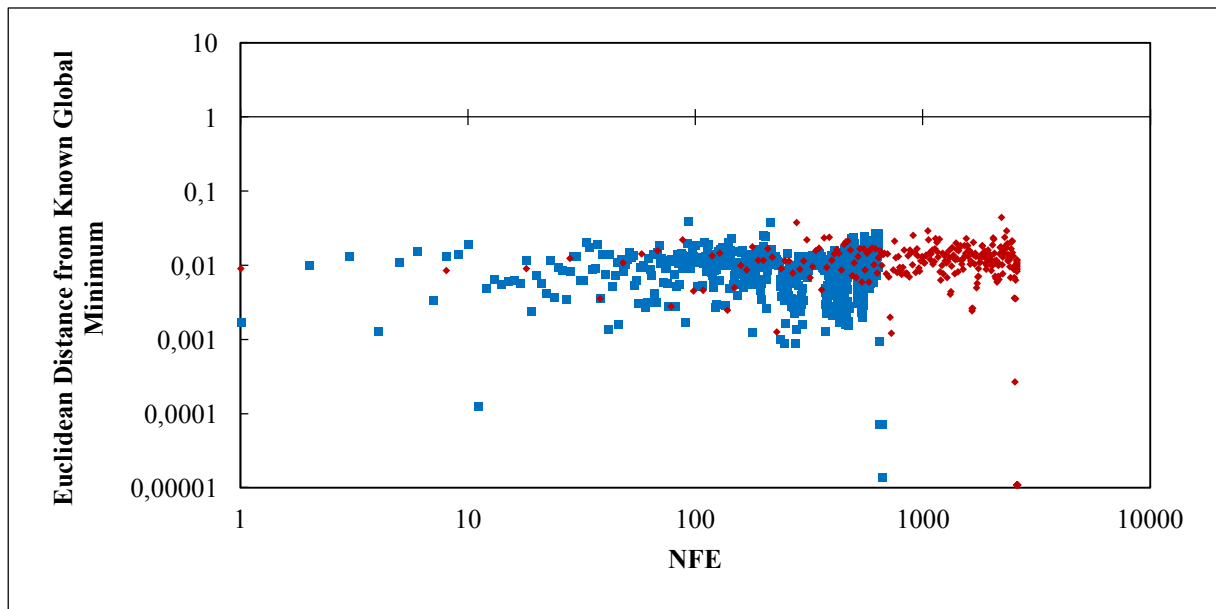


Figure 4.7 A convergence profile of PE-2 with a herd size of 50, ■KH, ◆LKH.

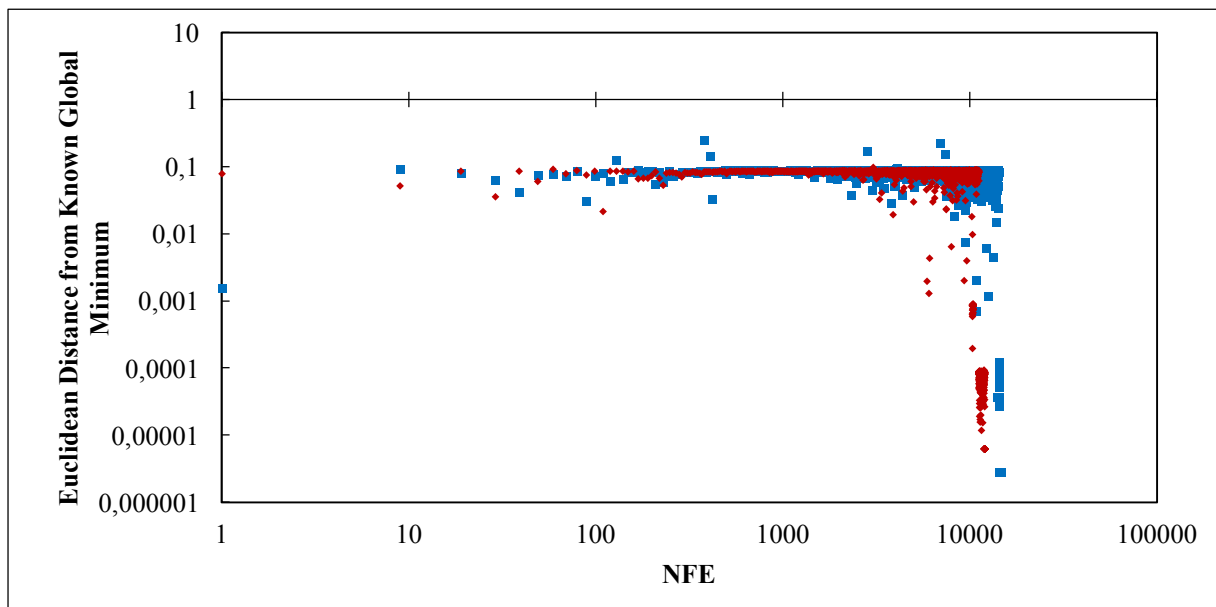


Figure 4.8 A convergence profile of PE-5 with a herd size of 50, ■KH, ◆LKH.

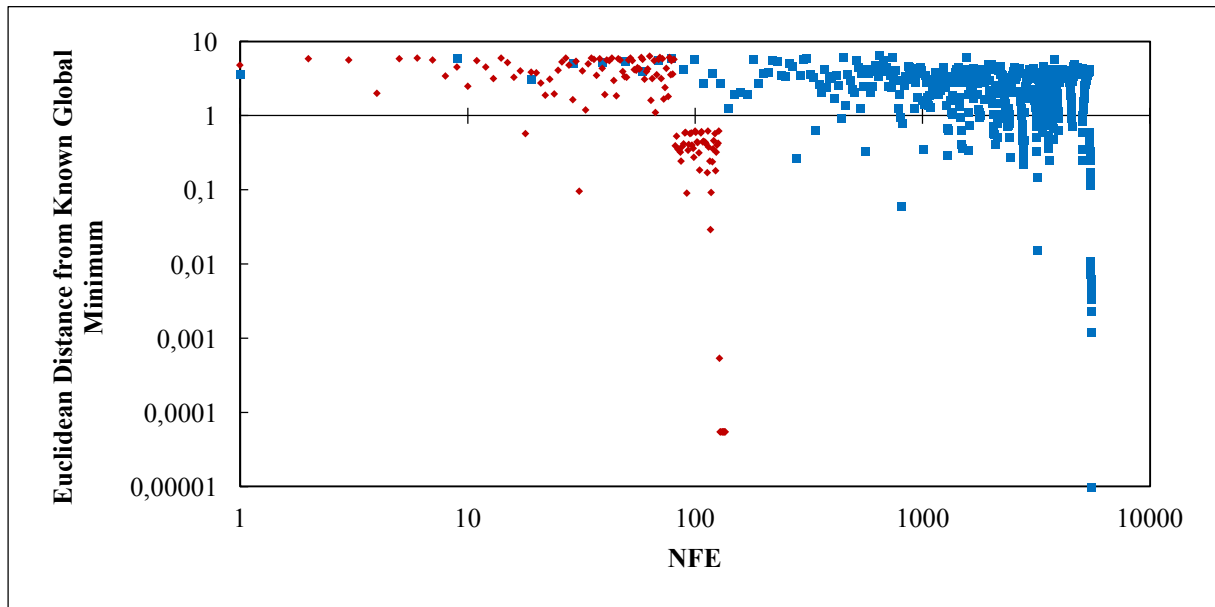


Figure 4.9 A convergence profile of PE-8 with a herd size of 50, ■KH, ◆LKH.

The individual case success rates are provided in Table 4.5. It was found that an increased herd size was required in order to achieve convergence in the PE problems, as herd sizes below 25 failed to converge for usually simple problems such as PE-4. A maximum herd size of 100 was chosen, as success rates did not improve substantially beyond this point. It is obvious that the KH algorithm requires a much greater computing time (almost double in some cases) than LKH, but yields only slightly higher GSRs.

**Table 4.5 Performance of the Krill Herd and Levy Flight Krill Herd Algorithms for problems involving phase equilibrium.**

Problem	Herd Size	Algorithm			
		Krill Herd		Levy Flight Krill Herd	
		Success Rate (%)	NFE	Success Rate (%)	NFE
<b>PE-1</b>	25	84	13647	82	10245
	50	88	9647	97	6334
	100	100	5452	99	2114
<b>PE-2</b>	25	96	15478	92	13324
	50	100	11345	98	10478
	100	100	7884	99	4613
<b>PE-3</b>	25	99	3241	92	11146
	50	100	1224	97	9008
	100	100	875	100	511
<b>PE-4</b>	25	77	16877	63	9745
	50	92	12544	82	8777
	100	98	5546	85	8044
<b>PE-5</b>	25	98	5329	97	8747
	50	100	842	100	4687
	100	100	477	100	3113
<b>PE-6</b>	25	99	3455	98	2113
	50	100	1546	99	977
	100	100	799	100	301
<b>PE-7</b>	25	94	14475	92	8477
	50	97	8148	93	5784
	100	99	1687	99	1245
<b>PE-8</b>	25	72	12458	64	2879
	50	89	8874	93	112
	100	97	5624	98	106
<b>GSR (%)</b>		<b>95</b>		<b>92</b>	

#### 4.4.4 Problems involving phase equilibrium in reacting systems

The NRTL, Wilson, (Wilson, 1964), UNIQUAC, (Abrams and Prausnitz, 1975) and Margules activity coefficient models and SRK equation of state with quadratic mixing were used to calculate activity and fugacity coefficients for the rPE problems, with the model and binary interaction parameters taken from the literature cited for each system in Table 4.2. The reaction equilibrium constants used, are also presented in Table 4.2.

The KH-based algorithms handle the rPE cases with high success rates, and correctly calculate the minimum total Gibbs energy for each case, indicating the physical occurrence of simultaneous chemical reaction and phase equilibrium.

Among the reactive cases, rPE-2 and rPE-6 are special in that inert components are introduced into a reacting mixture. With respect to the Gibbs energy minimization, the inert component does not affect the reaction directly, but does affect phase equilibrium. However in doing so, the inert component influences the composition of a phase, and thus the chemical equilibrium. In case rPE-5 a hypothetical reaction is considered with the formation of a heterogeneous ternary system at equilibrium. This system was proposed by Ung and Doherty (1995), and was used to illustrate that phase compositions at equilibrium need only be calculated from the reduced solution set of compositions that satisfy the reaction equilibrium constraints.

The global success rates for the reactive phase equilibrium problems are shown in Figure 4.10. The KH algorithm marginally outperforms the LKH algorithm. A herd size beyond 100 showed no significant improvement in the GSRs. Again LKH converged substantially faster than KH for some cases with the exception of rPE-1, 7 and 8.

Convergence profiles of cases rPE-1, 3 and 7 are shown in Figures 4.11 to 4.13. Again, exploration while within the convergence limit is more evident in KH than LKH. Figure 4.12 shows an example of a case for which rPE-3 has failed to converge for a herd size of 50.

The individual case success rates are provided in Table 4.6. The NFEs for LKH are generally only slightly lower than KH. KH does however outperform LKH in the difficult test cases of rPE-4 and rPE-7. The KH algorithm yields success rates of 100 for the interesting test cases involving inert components (rPE-2 and rPE-6) with herd sizes of 100.

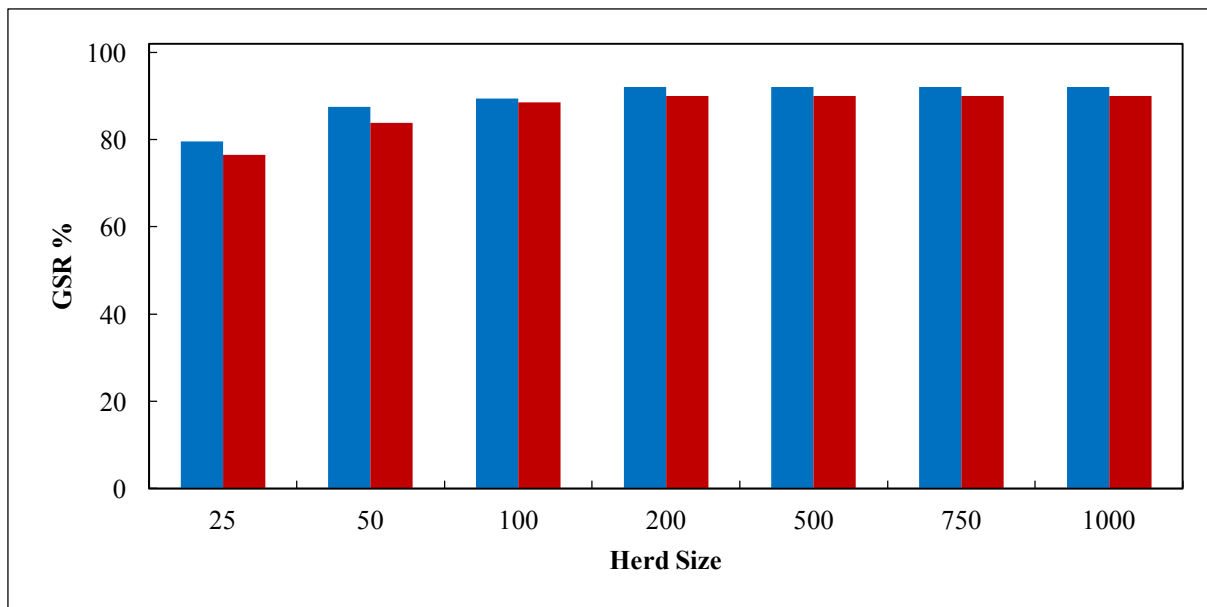


Figure 4.10 Global Success Rates of reactive phase equilibrium problems with increasing herd size, ■KH, ◆LKH.

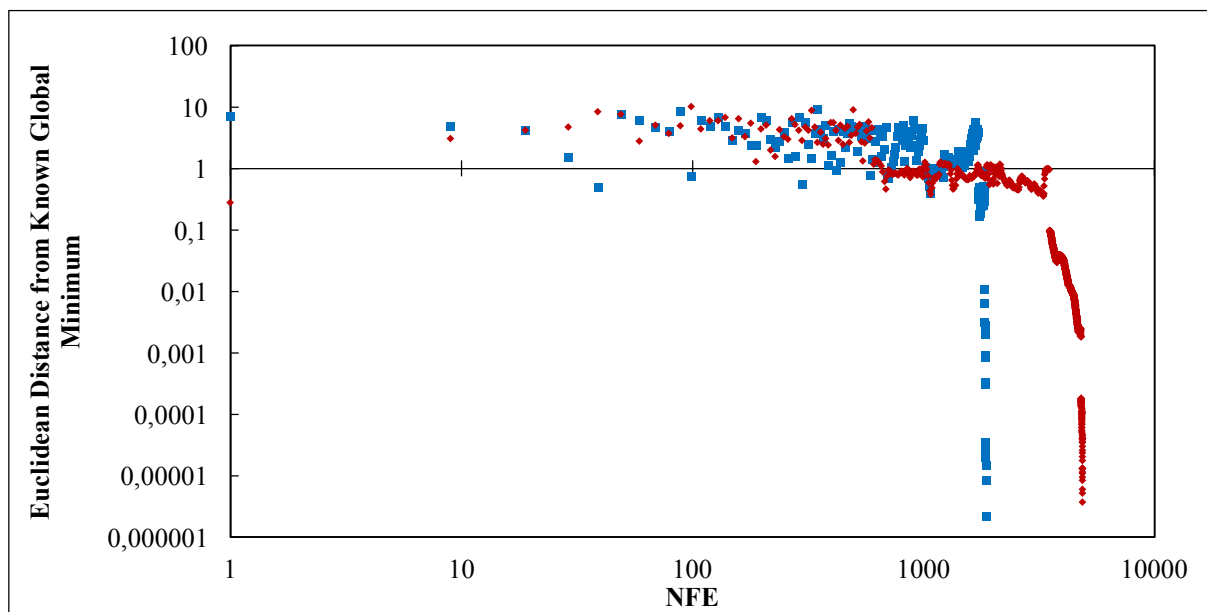


Figure 4.11 A convergence profile of rPE-1 with a herd size of 50, ■KH, ◆LKH.

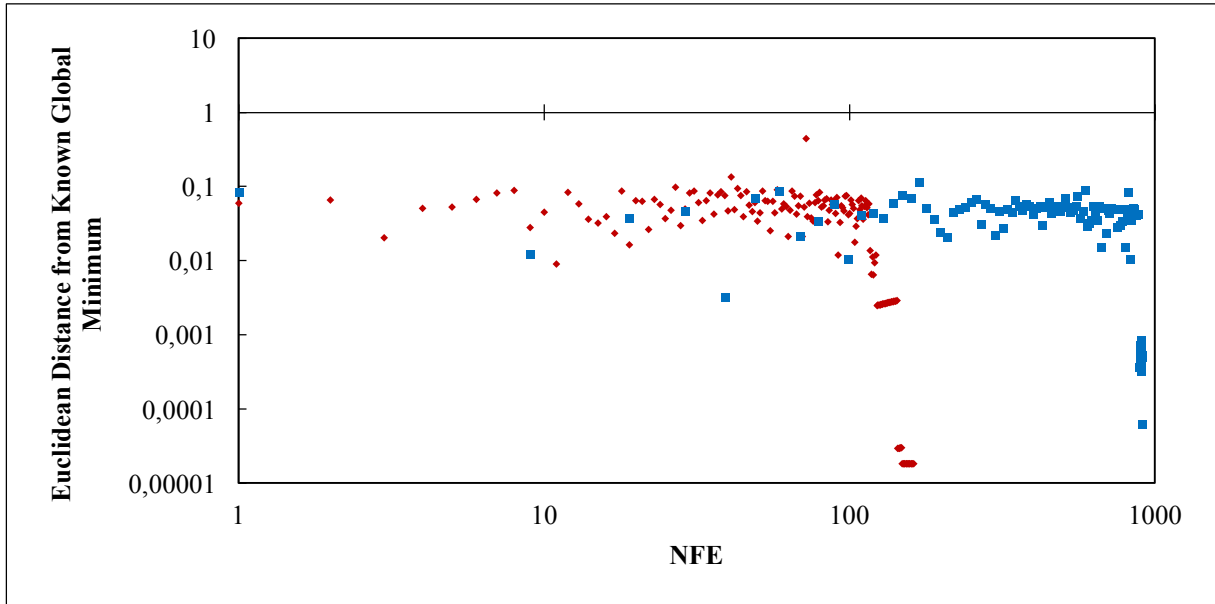


Figure 4.12 A convergence profile of rPE-3 with a herd size of 50, ■KH, ◆LKH.

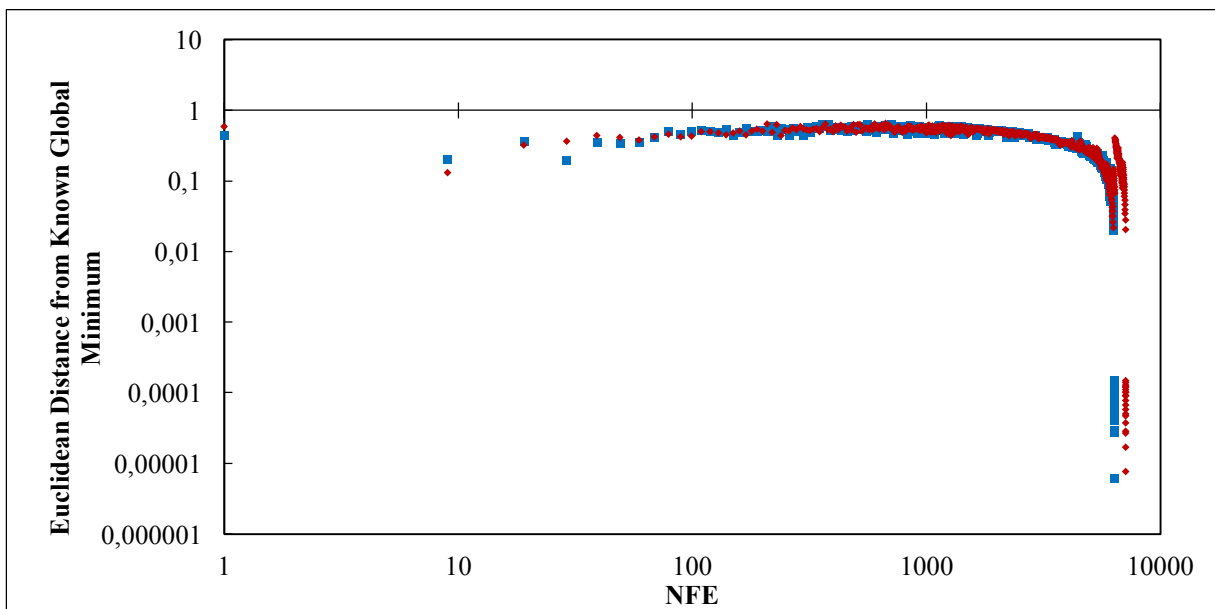


Figure 4.13 A convergence profile of rPE-7 with a herd size of 50, ■KH, ◆LKH.



**Table 4.6 Performance of the Krill Herd and Levy Flight Krill Herd Algorithms for problems involving reactive phase equilibrium.**

Problem	Herd Size	Algorithm			
		Krill Herd		Levy Flight Krill Herd	
		Success Rate (%)	NFE	Success Rate (%)	NFE
<b>rPE-1</b>	25	100	8744	96	12664
	50	100	3496	97	7482
	100	100	912	100	2441
<b>rPE-2</b>	25	100	8974	99	5778
	50	100	6478	100	4771
	100	100	1125	100	906
<b>rPE-3</b>	25	99	1124	100	477
	50	100	980	100	204
	100	100	319	100	174
<b>rPE-4</b>	25	23	3887	19	2664
	50	57	1977	51	1143
	100	66	446	64	348
<b>rPE-5</b>	25	99	13436	87	10322
	50	100	2114	90	1339
	100	100	1933	96	1778
<b>rPE-6</b>	25	84	15141	89	11475
	50	99	10447	94	7889
	100	100	3324	99	2554
<b>rPE-7</b>	25	34	16558	23	18554
	50	45	13394	39	10551
	100	56	9884	47	9770
<b>rPE-8</b>	25	98	13554	99	14997
	50	99	9246	100	9783
	100	100	6554	100	7001
<b>GSR %</b>		<b>86</b>		<b>83</b>	

#### 4.4.5 *Comparison between the performances of the KH based algorithms and to other methods in the literature*

Although phase equilibrium (PE) and reactive phase equilibrium (rPE) problems are physically more complex, the literature suggests that the minimization of the Tangent Plane Distant Function for the phase stability (PS) cases considered in this work, are computationally more complex. This is evidenced by the lower success rates of optimization of PS problems from other stochastic methods as presented in Table 4.7, in comparison to PE and rPE problems. Consequently lower average number of function (NFE) evaluations are observed for these cases, since self-termination of the program seems to occur at local minima.

In the case of the PE and rPE problems, the literature generally shows high success rates, indicating less complex computations. This may possibly be due to fewer local minima being encountered on the path to the global optimum for these cases, in comparison to the PS cases. The performance of the KH and LKH algorithms are therefore generally similar as the simplest path to the global minimum is easily found and followed.

The performance of the KH and LKH algorithms for calculating global optima in phase problems in this work were compared to alternate metaheuristic algorithms using data available in the literature. These alternate algorithms include genetic algorithms (GA), (modified and applied by Rangaiah, 2001, Teh and Rangaiah, 2003, and Bonilla-Petriciolet et al., 2011) Covariant Matrix Adaptation Evaluation Strategy (CMA-ES) (Hansen and Ostermeier, 2001), Shuffled Complex Evolution (SCE) (Duan et al., 1992), Firefly Algorithm (FA) (Yang, 2010), applied by Fateen et al. (2012), Cuckoo Search (CS) (Yang and Deb, 2010) and Modified Cuckoo Search (MCS) (Walton et al., 2011) applied by Bhargava et al. (2013). A detailed review of the merits of these methods, and comparisons to the KH and LKH algorithms, is beyond the scope of this work. The reader is referred to the original literature for further details. The phase problems considered for comparison are the same as those explored in this work, with the exception of rPE-8 in all cases excluding the genetic algorithms, where the phase stability and phase equilibrium problems considered by Rangaiah (2001) and Teh and Rangaiah (2003) form a subset of those considered in this work.

Trials with the maximum stopping condition of 50 iterations with no change in the objective function were used as a basis for the comparison. The average success rates and average

number of function evaluations for each stability problem are presented in Table 4.7, along with literature sources.

In the case of phase stability, the KH algorithm outperforms all other algorithms considered for comparison with the exception of the GA algorithm of Rangaiah (2001), with an average success rate of 89.5 % and with an average number of function evaluations of 1406. It must however be mentioned that the GA algorithm employed by Rangaiah (2001) also includes a Nelder-Mead simplex (Nelder and Mead, 1965) optimization step subsequent to the GA algorithm optimization, which is not performed in any of the other methods presented here for comparison. The LKH algorithm yields a much lower 81% success rate, however an average NFE of 614 is required. For the phase equilibrium and reactive phase equilibrium problems, the KH algorithm performs very well, matching the success rates of the most efficient method tested in the literature (the CS algorithm). LKH outperforms the alternate algorithms considered for comparison in terms of NFE, for both the PS and PE problems considered. However in the case of rPE problems the original KH algorithm is superior in both success rate and NFE.

Table 4.7 Comparison of the performances of the KH and LKH with other methods in the literature.

Problem	Method	Average SR (%)	Average NFE	Reference
<b>PS</b>	KH	<b>89.5</b>	1406	This work
	LKH	81	<b>614</b>	This work
	GA	100	<b>3524</b>	(Rangaiah, 2001)
	CMA-ES	86	7223	(Fateen et al., 2012)
	SCE	69	11752	(Fateen et al., 2012)
	FA	68	8461	(Fateen et al., 2012)
	CS	81	18373	(Bhargava et al., 2013)
	MCS	76	41645	(Bhargava et al., 2013)
<b>PE</b>	KH	<b>99</b>	3543	This work
	LKH	98	<b>2506</b>	This work
	GA	93	<b>20179</b>	(Teh and Rangaiah, 2003)
	CMA-ES	97	10007	(Fateen et al., 2012)
	SCE	95	13902	(Fateen et al., 2012)
	FA	85	7538	(Fateen et al., 2012)
	CS	<b>99</b>	35209	(Bhargava et al., 2013)
	MCS	71	61166	(Bhargava et al., 2013)
<b>rPE</b>	KH	<b>90</b>	<b>3062</b>	This work
	LKH	89	3122	This work
	GA	<b>70</b>	8130	(Bonilla-Petriciolet et al., 2011)
	CMA-ES	83	14474	(Fateen et al., 2012)
	SCE	76	7529	(Fateen et al., 2012)
	FA	74	3206	(Fateen et al., 2012)
	CS	<b>90</b>	12162	(Bhargava et al., 2013)
	MCS	80	30337	(Bhargava et al., 2013)

#### *4.5 Conclusion*

The recently developed Krill Herd optimization technique was applied to phase equilibrium and stability problems, for the calculation of global minima of the total Gibbs energy. The original Krill Herd algorithm was found to outperform the Lévy flight Krill Herd Algorithm, when considering the success rates of obtaining the global minimum. The KH algorithm was compared to other stochastic algorithms in the literature (Genetic Algorithm (GA), Covariant Matrix Adaptation Evaluation Strategy (CMA-ES), Shuffled Complex Evolution (SCE), Firefly Algorithm (FA), Cuckoo Search (CS) and Modified Cuckoo Search (MCS)), where the same phase problems were considered, and outperforms one of the leading technique (CS) in phase stability problems, while matching the performance of one of the leading technique (CS) in both non-reactive and reactive phase equilibria problems when considering success rate and number of function evaluations. The algorithm is seemingly outperformed by a combined Nelder-Mead Simplex + GA in the case of phase stability, in terms of success rate, but not in the number of function evaluations. In all cases (PS, PE and rPE), KH and LKH required by far the smallest number of function evaluations, with respect to success rate, compared to the alternate techniques considered.

*References*

- Abrams, D.S. and Prausnitz, J.M., (1975), Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems. *AIChE Journal*, 21, pp.116–128.
- Ammar, M.N. and Renon, H., (1987), The isothermal flash problem: New methods for phase split calculations. *AIChE Journal*, 33, pp.926–939.
- Balogh, J., Csendes, T. and Stateva, R.P., (2003), Application of a stochastic method to the solution of the phase stability problem: cubic equations of state. *Fluid Phase Equilibria*, 212, pp.257–267.
- Bender, E. and Block, U., (1975), Thermodynamische Berechnung der Flüssig Extraktion. *Verfahrenstechnik*, 9, p.106.
- Bhargava, V., Fateen, S.E.K. and Bonilla-Petriciolet, A., (2013), Cuckoo Search: A new nature-inspired optimization method for phase equilibrium calculations. *Fluid Phase Equilibria*, 337, pp.191–200.
- Bonilla-Petriciolet, A., Bravo-Sanchez, U.I., Castillo-Borja, F., Frausto-Hernandez, S. and Segovia-Hernandez, J.G., (2008), Gibbs Energy Minimization Using Simulated Annealing for Two-phase Equilibrium Calculations in Reactive Systems. *Chemical and Biochemical Engineering Quarterly*, 22, pp.285–298.
- Bonilla-Petriciolet, A., Iglesias-Silva, G.A. and Hall, K.R., (2008), An effective calculation procedure for two-phase equilibria in multireaction systems. *Fluid Phase Equilibria*, 269, pp.48–55.
- Bonilla-Petriciolet, A., Rangaiah, G.P. and Segovia-Hernández, J.G., (2011), Constrained and unconstrained Gibbs free energy minimization in reactive systems using genetic algorithm and differential evolution with tabu list. *Fluid Phase Equilibria*, 300, pp.120–134.
- Bonilla-Petriciolet, A. and Segovia-Hernández, J.G., (2010), A comparative study of particle swarm optimization and its variants for phase stability and equilibrium calculations in multicomponent reactive and non-reactive systems. *Fluid Phase Equilibria*, 289, pp.110–121.
- Bonilla-Petriciolet, A., Vázquez-Román, R., Iglesias-Silva, G. A and Hall, K.R., (2006), Performance of Stochastic Global Optimization Methods in the Calculation of Phase Stability Analyses for Nonreactive and Reactive Mixtures. *Industrial and Engineering Chemistry Research*, 45, pp.4764–4772.
- Castillo, J. and Grossmann, I.E., (1981), Computation of phase and chemical equilibria. *Computers and Chemical Engineering*, 5, pp.99–108.
- Dorigo, M., (1992), *Optimisation, learning and natural algorithms*. Politecnico di Milano.

- Duan, Q., Sorooshian, S., Gupta, H. V. and Gupta, V., (1992), Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28, pp.1015–1031.
- Fateen, S.-E.K., Bonilla-Petriciolet, A. and Rangaiah, G.P., (2012), Evaluation of Covariance Matrix Adaptation Evolution Strategy, Shuffled Complex Evolution and Firefly Algorithms for Phase Stability, Phase Equilibrium and Chemical Equilibrium Problems. *Chemical Engineering Research and Design*, 90, pp.2051–2071.
- Gandomi, A.H. and Alavi, A.H., (2012), Krill herd: A new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, 17, pp.4831–4845.
- Gau, C.Y., Brennecke, J.F. and Stadtherr, M.A., (2000), Reliable nonlinear parameter estimation in VLE modeling. *Fluid Phase Equilibria*, 168, pp.1–18.
- Hansen, E. and Walster, G. W., (2004), *GLOBAL OPTIMIZATION Using Interval Analysis* 2nd ed., Marcel Dekker, New York.
- Hansen, N. and Ostermeier, A., (2001), Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9, pp.159–195.
- Harding, S.T. and Floudas, C.A., (2000), Phase stability with cubic equations of state: Global optimization approach. *AIChE Journal*, 46, pp.1422–1440.
- Heidemann, R.A. and Mandhane, J.M., (1973), Some properties of the NRTL equation in correlating liquid–liquid equilibrium data. *Chemical Engineering Science*, 28, pp.1213–1221.
- Hofmann, E.E., Haskell, A.G.E., Klinck, J.M. and Lascara, C.M., (2004), Lagrangian modelling studies of Antarctic krill (*Euphausia superba*) swarm formation. *Ices Journal of Marine Science*, 61, pp.617–631.
- Hua, J.Z., Brennecke, J.F. and Stadtherr, M.A., (1998), Enhanced Interval Analysis for Phase Stability: Cubic Equation of State Models. *Industrial and Engineering Chemistry Research*, 37, pp.1519–1527.
- Jalali, F. and Seader, J.D., (1999), Homotopy continuation method in multi-phase multi-reaction equilibrium systems. *Computers and Chemical Engineering*, 23, pp.1319–1331.
- Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P., (1983), Optimization by Simulated Annealing. *Science*, 220, pp. 671–680.
- Land, A.H. and Doig, A.G., (1960), An Automatic Method of Solving Discrete Programming Problems. *Econometrica*, 28, pp.497–520.
- Lee, P.Y., Rangaiah, G.P. and Luus, R., (1999), Phase and chemical equilibrium calculations by direct search optimization. *Computers and Chemical Engineering*, 23, pp.1183–1191.
- McDonald, C.M. and Floudas, C.A., (1995), Global optimization for the phase stability problem. *AIChE Journal*, 41, pp.1798–1814.

- McDonald, C.M. and Floudas, C. A., (1996), GLOPEQ: A new computational tool for the phase and chemical equilibrium problem. *Computers and Chemical Engineering*, 21, pp.1–23.
- Michelsen, M.L., (1982), The isothermal flash problem. Part I. Stability. *Fluid Phase Equilibria*, 9, pp.1–19.
- Miller, D.G.M. and Hampton, I., (1989), Krill aggregation characteristics: Spatial distribution patterns from hydroacoustic observations. *Polar Biology*, 10 pp. 125-134.
- Mordecai, A., (2003), *Nonlinear Programming: Analysis and Methods*. Dover Publications.
- Nelder, J.A. and Mead, R., (1965), A Simplex Algorithm for Function Minimization. *Computer Journal*, 7, pp.308–313.
- Nichita, D.V., Gomez, S. and Luna, E., (2002), Phase stability analysis with cubic equations of state by using a global optimization method. *Fluid Phase Equilibria*, 194-197, pp.411–437.
- Ohanomah, M.O. and Thompson, D.W., (1984), Computation of multicomponent phase equilibria—Part III. Multiphase equilibria. *Computers and Chemical Engineering*, 8, pp.163–169.
- Paules, G. E. IV, Floudas, C.A., (1989), A New Optimization Approach for Phase and Chemical Equilibrium Problems. In *Proceedings of the American Institute of Chemical Engineers Meeting*.
- Raal, J.D. and Mühlbauer, A.L., (1998), *Phase Equilibria: Measurement and Computation*. Taylor and Francis, Bristol.
- Rahman, I., Das, A.K., Mankar, R.B. and Kulkarni, B.D., (2009), Evaluation of repulsive particle swarm method for phase equilibrium and phase stability problems. *Fluid Phase Equilibria*, 282, pp.65–67.
- Rangaiah, G.P., (2001), Evaluation of genetic algorithms and simulated annealing for phase equilibrium and stability problems. *Fluid Phase Equilibria*, 187-188, pp.83–109.
- Rashedi, E., Nezamabadi-pour, H. and Saryazdi, S., (2009), GSA: A Gravitational Search Algorithm. *Information Sciences*, 179, pp.2232–2248.
- Renon, H. and Prausnitz, J., (1968), Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE journal*, 14, pp.135–144.
- Seider, W., (1996), Multiphase equilibria of reactive systems. *Fluid Phase Equilibria*, 123, pp.283–303.
- Soave, G., (1972), Equilibrium constants from a modified Redlich-Kwong equation of state. *Chemical Engineering Science*, 27, pp.1197–1203.



- Srinivas, M. and Rangaiah, G.P., (2007), A study of differential evolution and tabu search for benchmark, phase equilibrium and phase stability problems. *Computers and Chemical Engineering*, 31, pp.760–772.
- Stateva, R.P. and Wakeham, W.A., (1997), Phase Equilibrium Calculations for Chemically Reacting Systems. *Ind. Eng. Chem. Res.*, 36, pp.5474–5482.
- Sun, A.C. and Seider, W.D., (1995), Homotopy-continuation method for stability analysis in the global minimization of the Gibbs free energy. *Fluid Phase Equilibria*, 103, pp.213–249.
- Teh, Y.S. and Rangaiah, G.P., (2003), Tabu search for global optimization of continuous functions with application to phase equilibrium calculations. *Computers and Chemical Engineering*, 27, pp.1665–1679.
- Ung, S. and Doherty, M.F., (1995), Vapor-liquid phase equilibrium in systems with multiple chemical reactions. *Chemical Engineering Science*, 50, pp.23–48.
- Ung, S. and Michael, F., (1995), Theory of Phase Equilibria in Multireaction Systems. *Chemical Engineering Science*, 50, pp.3201–3216.
- Wakeham, W.A. and Stateva, R.P., (2004), Numerical Solution of the Isothermal, Isobaric Phase Equilibrium Problem. *Reviews in Chemical Engineering*, 20, pp.1–56.
- Walton, S., Hassan, O., Morgan, K. and Brown, M.R., (2011), Modified cuckoo search: A new gradient free optimisation algorithm. *Chaos, Solitons and Fractals*, 44, pp.710–718.
- Wang, G., Guo, L., Gandomi, A.H., Cao, L., Alavi, A.H., Duan, H. and Li, J., (2013), Lévy-Flight Krill Herd Algorithm. *Mathematical Problems in Engineering*, 2013, pp.1–14.
- Wasylikiewicz, S.K. and Ung, S., (2000), Global phase stability analysis for heterogeneous reactive mixtures and calculation of reactive liquid–liquid and vapor–liquid–liquid equilibria. *Fluid Phase Equilibria*, 175, pp.253–272.
- Wilson, G.M., (1964), Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing. *Journal of the American Chemical Society*, 86, pp.127–130.
- Xiao, W., Zhu, K., Yuan, W. and Chien, H.H., (1989), An algorithm for simultaneous chemical and phase equilibrium calculation. *AIChE Journal*, 35, pp.1813–1820.
- Yang, X.S., (2009), Firefly algorithms for multimodal optimization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5792 LNCS, pp.169–178.
- Yang, X.S. and Deb, S., (2010), Engineering optimisation by cuckoo search. *Mathematical Modelling and Numerical Optimisation*, pp.1–17.
- Zhang, H., (2011), A Review on Global Optimization Methods for Phase Equilibrium Modeling and Calculations. *The Open Thermodynamics Journal*, 5, pp.71–92.

Zhang, H., Fernández-Vargas, J.A., Rangaiah, G.P., Bonilla-Petriciolet, A. and Segovia-Hernández, J.G., (2011), Evaluation of integrated differential evolution and unified bare-bones particle swarm optimization for phase equilibrium and stability problems. *Fluid Phase Equilibria*, 310, pp.129–141.

## CHAPTER FIVE

### **An Extended UNISAC model for the Prediction of Solubility of Complex Pharmaceutical Ingredients in Non-electrolyte Pure Solvents and Solvent Mixtures**

#### *Abstract*

The previously published Universal Segment Activity Coefficient Model (UNISAC) has been extended with a focus on the solubility of structurally complex pharmaceutical ingredients. The extension involves the addition of three new base segment types and the incorporation of a significantly more detailed group fragmentation scheme that has been successfully applied to the estimation of pure component properties by group contribution methods. Extended UNISAC segment group area parameters for 24 unique non-base segments (59 structural groups) are presented, that were obtained during model training using the recently developed global optimization technique, Krill Herd Algorithm. Due to the physical significance of the segment area parameters, instead of 168 parameters for 24 non-base groups, only 49 values had to be regressed. 30 interaction parameter values and 6 group surface parameters were taken from original UNIFAC without modification. The Extended UNISAC model is applied to the prediction of solid-liquid equilibrium calculations for solute-pure solvent and solute-solvent-solvent/solute-solute-solvent systems. Comparisons to existing popular models such as UNIFAC, modified UNIFAC (Dortmund), COSMO-RS(OL) and COSMO-SAC are made and show that the Extended UNISAC model is competitive with these existing models. A major advantage of the Extended UNISAC model is that a similar quality as with UNIFAC or modified UNIFAC (Dortmund) can be achieved using a drastically lower number of physically realistic group area parameters, and no additional group specific interaction parameters. Further development of the method especially with respect to new structural groups for complex molecules is therefore not hindered by a radical increase in the number of model parameters.

### 5.1 Introduction

Besides pure component properties like vapour pressure and critical data, binary parameters for the description of the real liquid mixture behaviour are vital and most often missing in chemical process simulation. These binary interaction parameters (BIP) for the different  $G^E$ -models and equations of state are best obtained by regression to binary mixture data, typically vapour-liquid equilibria in the case of volatile compounds.

Although a large amount of such data is available in the open literature or from data banks like the Dortmund Data Bank (2012) or NIST-TDE (2015), for the majority of the possible binary combinations of components of practical interest no experimental information can be found. In many cases, the specific mixture behaviour can then be deduced from data for very similar mixtures and new or unexpected results are highly improbable. If the mixture behaviour of ethanol with n-octane and n-decane is available, the parameters for the system ethanol and n-nonane can be interpolated. In many other cases, no sufficiently similar data have been published.

To estimate the urgently needed BIP, mixture data are usually predicted using a reliable estimation method and the BIP are regressed to these estimations.

It is for this reason that vast efforts were made into the development of such methods.

#### 5.1.1 Solubility parameter methods

One of the earliest methods was based on the Hildebrand (1916) solubility parameters, which can be calculated from the heat of vaporisation of the components. The energetic interaction between unlike components was then estimated as a volumetric mean value of the like-like interaction. It quickly became apparent, that this approach is not generally applicable as the energetic interaction between like molecules may be of different degrees. If one component is strongly interacting via dispersive forces but the other component shows weak dispersion but is instead strongly polar, the energetic interactions between unlike molecules will be rather weak.

Hansen (1967) incorporated this concept in his method so that the interaction between unlike molecules is a measure of the distance of these molecules in an “interaction space” with the three orthogonal axis dispersive, polar and hydrogen bonding strength. In the case of polymer-

solvent systems, for which this method is most often used, the location of a polymer in the interaction space can rather easily be obtained by observing the solution behaviour in a few common solvents.

### 5.1.2 UNIFAC *et al.*

In a separate development,  $G^E$ -models like Wilson (Wilson, 1964) or UNIQUAC (Abrams and Prausnitz, 1975) were reformulated to describe pure components and mixtures in term of the interactions between structural groups (ASOG (Derr and Deal, 1969), UNIFAC (Fredenslund *et al.*, 1975), modified UNIFAC (Dortmund) (Weidlich, J. Gmehling, 1987)).

The activity coefficient  $\gamma_i$  of a component in a mixture is then the ratio of the products of all group activity coefficients  $\Gamma_k$  of the groups  $k$  constituting compound  $i$  in the mixture to that in the pure component  $\Gamma_{k,i}$  (Hala, 1978). In logarithmic form and summing over the different groups with  $\nu_{k,i}$  as the frequency of group  $k$  in component  $i$ , this can be written as

$$\ln\gamma_i = \sum_k^N \nu_{k,i} (\ln\Gamma_k - \ln\Gamma_{k,i}) \quad (5.1)$$

Significant manpower has been invested into the further development of these methods, namely in the collection of literature data, measurement of new data and the regression of group interaction parameters (e.g. UNIFAC-Consortium ([unifac.org](http://unifac.org))). Due to the importance of these methods, a large part of these efforts were funded by industry.

Latest versions of modified UNIFAC are often able to reproduce and predict vapour-liquid equilibria, solid-liquid equilibria, infinite dilution behaviour, and azeotropic composition, close to typical experimental uncertainties.

One of the downsides of these methods is the large number of required group-group interaction parameters, which grows quadratically with the number of different main groups. In UNIFAC, main groups are chemically different entities that differ by their energetic interactions with other groups. There are often several subgroups that only differ by surface and volume.

In order to restrict the number of parameters, methods like UNIFAC are generally employing a simple as possible fragmentation scheme. The consortium version of modified UNIFAC currently uses 99 main groups and about one third of the required BIP are available.

This severely limits the application in case of special components that cannot realistically be separated into smaller groups and larger multifunctional components like pharmaceuticals and their intermediates that often contain unusual structural elements.

As a result of this, there is an interest in novel predictive methods that also yield acceptable results outside the scope of the highly trained UNIFAC methods.

### 5.1.3 *COSMO-RS and COSMO-SAC*

One such approach are the different versions of COSMO-RS (Klamt, (1995) Grensemann and Gmehling, (2005)) and COSMO-SAC (Lin and Sandler, (2002), Hsieh et al., (2010)). In contrast to UNIFAC, these methods construct a molecular shaped cavity and optimize the charge-distribution on the cavity wall in order to optimally shield the charges inside the molecule via a quantum-chemical calculation. Molecules are then said to interact via this shielding charge surface. The surface is divided into a large number of surface segments, each of which has a different shielding charge. The mixture is now viewed as mixture of surface segments and the activity coefficient of the component is again calculated by equation (5.1), only now the summation takes place over the surface segments. All information about the original surface and the relative position of the charges is lost in this process.

The need for binary segment-segment interaction parameters is avoided by calculating the segment-segment interaction from a simple electrostatic relationship. Hydrogen bonding is introduced by assuming an especially strong interaction between segments of very low and very high shielding charge.

Multiple adjustments and improvements to this basic concept have been added (Grensemann and Gmehling, (2005), Hsieh et al., (2010)). In order to avoid the very time consuming quantum-chemical calculation in the case of larger molecules, group contribution methods for the shielding charge distribution profiles (sigma-profiles) have been developed (Mu et al., 2007). Thorough tests of two popular versions of these methods in the past, Grensemann and Gmehling, (2005) showed several weaknesses and inconsistencies, especially in case of some hydrogen bonding mixtures. In addition, results can vary strongly for different conformers of the same molecule.

### 5.1.4 *NRTL-SAC*

In a different approach, Chen and Song (2004) applied the idea of Hansen (1967), who considered three molecular interactions types to pharmaceutical-solvent mixtures. In this

method, there are only four basic surface segments that represent the different types of interactions, namely dispersive, polar-plus, polar-minus and hydrogen-bonding. NRTL binary interaction parameters between these 4 basic types are available in the method.

Each component is then viewed to interact via a specific combination of four surface types. The activity coefficient of a unit segment of each type is calculated by the NRTL model and the activity coefficient of the molecule is again calculated from the activity coefficients of the constituting segments in the mixture and in the pure component (see equation. (5.1)). Segment compositions for a number of common solvents were given by the authors. The values for a new component can then be regressed to data for mixtures with at least 4 sufficiently different solvents. The method is therefore well suited for solvent selection in case of pharmaceuticals and their intermediates. Nevertheless, data for at least 4 mixtures are required to characterize a new solute which is usually not available.

### 5.1.5 UNISAC

Recently Moodley et al. (Moodley et al., 2015(b)) published a method, Universal Segment Activity Coefficient (UNISAC), which successfully combined the group contribution approach in UNIFAC with the segment-contribution principle in NRTL-SAC. In this method, group interactions for only four of the UNIFAC main groups were used, all other groups were viewed as mixtures of these four basic groups. Compared to the UNIFAC-method, UNISAC requires only four parameters for each new group, thus greatly reducing the need for experimental data to regress group interactions.

The original UNISAC method requires the group segment designation of each UNIFAC group that constitutes the molecule to be known, with the corresponding segment area parameters, each representing the effect of one of four unique intermolecular forces (dispersion, hydrogen bonding, +/- polarity). Based on those results, the work presented here resulted in the development of a more detailed UNISAC model.

In this work, the UNISAC model is extended to consider an additional three intermolecular forces which are discussed later. A major disadvantage of the common predictive activity coefficient models such as the original UNIFAC (Fredenslund et al., 1975) and modified UNIFAC (Dortmund) (Weidlich and Gmehling, 1987) models is that structurally complex molecules such as various pharmaceuticals, cannot be fragmented into the groups considered by the respective models, thus predictions by the models become impossible. In the original UNISAC model, this issue was not addressed, as the fragmentation scheme of the UNIFAC

and modified UNIFAC (Dortmund) models were used. Hence a more detailed group fragmentation scheme based on the work of Moller et al. (2008) was developed, which allows the application to a wider range of components. Since only seven base groups are still considered, this increase in non-base groups does not cause a quadratic increase in the number of binary interaction parameters required. However the additional segment area parameters are still required, and in order to calculate the new segment area parameters to include these groups, regression to phase equilibrium data of systems with components comprised of all seven groups, was performed using a global optimization technique; the Krill Herd Algorithm (Gandomi and Alavi, 2012). The performance of the Krill Herd Algorithm for optimization problems involving phase equilibrium was investigated in an earlier publication (Moodley et al., 2015 (c)).

The applicability of henceforth termed Extended UNISAC model to the prediction of solid-liquid equilibria in binary and ternary eutectic mixtures was verified by performing predictions for over 4000 experimental data points from literature.

## 5.2 Theory

### 5.2.1 Thermodynamic Relationship

For solid-liquid equilibria in eutectic mixtures Hildebrand and Scott (1964) proposed:

$$\ln(\gamma_i^{sat} x_i^{sat}) = \frac{\Delta_{fus}H_i}{R_{fus}T_i} \ln\left(\frac{T}{T_i}\right) \quad (1.41)$$

Where  $\gamma_i^{sat}$  is the activity coefficient of the solute in the saturated solution,  $x_i^{sat}$  is the mole fraction of the solute at saturation,  $\Delta_{fus}H_i$  is the enthalpy of fusion (melting) at the melting temperature  $T_i$ ,  $T$  is the temperature at saturation and  $R$  is the Universal Gas Constant. Equation (1.41) was derived in Chapter 1, and approximates the change of  $\Delta_{fus}H_i$  with temperature via the entropy of melting and performs mostly better than the calculation with constant heat of melting if the UNIFAC combinatorial term is used. This was confirmed for aqueous systems in Chapter Two and additionally by Neau et al., (Neau et al., 1997) and Moller et al. (Moller et al., 2014).



## 5.2.2 Extension of the Universal Segment Activity Coefficient Model

As in the original UNISAC model (Moodley et al., 2015 (b)), the activity coefficient in the Extended UNISAC model is expressed as a combination of a combinatorial and a residual term:

$$\ln\gamma_i = \ln\gamma_i^{comb} + \ln\gamma_i^{res} \quad (1.49)$$

As the combinatorial term in the Extended UNISAC model, the Flory-Huggins (Flory, 1941, Huggins, 1941) model with the Staverman-Guggenheim (Staverman, 1950, Guggenheim, 1952) correction term for molecules of different shape is used. The modification of Weidlich and Gmehling (Weidlich and Gmehling, 1987) is no longer used, as the improvements in the combinatorial expression of the model will be considered in future work, and is briefly outlined in Chapter Nine. Hence the original UNIFAC combinatorial expression is employed, which is equivalent to equation (1.50):

$$\ln\gamma_i^{comb} = 1 - V_i + \ln(V_i) - 5q_i \left[ 1 - \frac{V_i}{F_i} + \ln\left(\frac{V_i}{F_i}\right) \right] \quad (5.2)$$

Where

$$V_i = \frac{r_i}{\sum_j x_j r_j} \quad (3.13)$$

$$F_i = \frac{q_i}{\sum_j x_j q_j} \quad (3.14)$$

Where

$$r_i = \sum_j v_{j,i} R_j \quad (3.15)$$

$$q_i = \sum_j v_{j,i} Q_j \quad (3.16)$$

Where  $r_i$  and  $q_i$  are the van der Waals volume and surface area, estimated by the method of Bondi (Bondi, 1964).  $v_{j,i}$  denotes the frequency of group  $j$  in component  $i$ .

The residual term employs the segment or group interaction concept (Hala, 1978):

$$\ln\gamma_i^{res} = \sum_k^N \Omega_{k,i} (\ln\Gamma_k - \ln\Gamma_{k,i}) \quad (3.17)$$

Where  $\Omega_{k,i}$  is the total segment area of segment  $k$  in component  $i$ :

$$\Omega_{k,i} = \sum_{l=1}^N v_{l,i} \zeta_{k,l} \quad (3.18)$$

Where  $N$  is the number of different groups in the mixture and  $\zeta_{k,l}$  is the segment area of segment  $k$  in group  $l$ .

$\ln\Gamma_{k,i}$  is the natural logarithm of the segment activity coefficient of segment  $k$  in the pure component  $i$ :

$$\ln\Gamma_{k,i} = 1 - \ln\left(\sum_{m=1}^N \Theta_{m,i} \psi_{m,k}\right) - \sum_{m=1}^N \frac{\Theta_{m,i} \psi_{k,m}}{\sum_{n=1}^N \Theta_{n,i} \psi_{n,m}} \quad (3.19)$$

Where  $\Theta_{m,i}$  is the segment area fraction of segment  $m$  in the pure component  $i$ :

$$\Theta_{m,i} = \frac{\Omega_{m,i}}{\sum_{m=1}^N \Omega_{m,i}} \quad (3.20)$$

$\ln\Gamma_k$  is the natural logarithm of the segment activity coefficient for the mixture given by:

$$\ln \Gamma_k = 1 - \ln \left( \sum_{m=1}^N \theta_m \psi_{m,k} \right) - \sum_{m=1}^N \frac{\theta_m \psi_{k,m}}{\sum_{n=1}^N \theta_n \psi_{n,m}} \quad (3.21)$$

Where

$$\theta_m = \frac{\sum_{i=1}^I \Omega_{m,i} x_i}{\sum_{m=1}^N \sum_{i=1}^I \Omega_{m,i} x_i} \quad (3.22)$$

$x_i$  is the mole fraction of component  $i$ ,  $I$  is the total number of components.  $\psi_{u,v}$  is the segment interaction between segment  $u$  and  $v$ .

Moodley et al. (Moodley et al., 2015 (b)) have reported segment area parameters for approximately 30 of the original UNISAC groups using the BMRR1 base groups, (C-CH<sub>3</sub>, H<sub>2</sub>O, C-CN and C-Cl) to represent dispersion, hydrogen bonding, positive polarity and negative polarity respectively. The binary interaction parameters between the base segments in the original UNISAC model were fitted to experimental data and form a unique set. In the Extended UNISAC model presented in this work, additional base segment areas are now considered. The hydrogen bonding area is separated into a hydrogen bond donor and acceptor. The “water” group segment area is however retained to assist in modelling the unique behaviour of many components in aqueous mixtures, as well as mixtures that exhibit aqueous-like characteristics. The seventh additional segment area incorporated is termed an “empty” segment, and represents a segment surface area which contributes a zero intermolecular force in the mixture. The binary interactions for this base segment are zero. This segment does however contribute to the total group surface area. Furthermore, it was decided in the Extended UNISAC model, to revert the binary interaction parameters between base segments to those of the original UNIFAC model, as unrealistic predictions were sometimes obtained for systems of vapour-liquid and liquid-liquid when the original UNISAC model was tested. Consequently predictions between base segments for the Extended UNISAC model are identical to those of the UNIFAC model. Water is therefore no longer treated with separate interaction parameters. The binary interaction parameters between the seven base segments of the Extended UNISAC model are provided in Table 5.1. Structural groups for individual components (e.g. water, methanol) are not present in the fragmentation scheme of Moller et al. (2008) and were added from the original UNISAC model, hence group numbers >300 are now included. The new Moodley-Rarey-Ramjugernath group list is presented in Table 5.2. It must be reiterated that the Moodley-Rarey-Ramjugernath fragmentation scheme only applies to the residual

expression. The original UNIFAC fragmentation is still used for the combinatorial expression. The updated seven parameter segment area list is also presented in Table 5.2. These parameters have been improved due to an expanded literature search and enhanced regression scheme (Krill Herd Algorithm) and now also include additional segment area parameters, determined by regression to pseudo-experimental data generated by UNIFAC VLE prediction, similarly to the procedure used in the original UNISAC model development. For this fitting procedure the following objective function was used:

$$\delta = \sum_{i=1}^{NP} \sum_{i=1}^I \text{abs}(\ln(\gamma_i^{sat})^{exp} - \ln(\gamma_i^{sat})^{calc}) \quad (3.28)$$

Where  $\gamma_i^{sat}$  is the activity coefficient at saturation,  $I$  is the number of components in the mixture, and  $NP$  is the total number of points considered. The superscripts *exp* and *calc* represent experimental (pseudo-experimental) and UNISAC model calculated values respectively.

An attempt was made to restrict the fitting parameters for segment area parameters to realistic values. For instance, a CH group in a ring should ideally broadly exhibit only dispersive forces. This was not always possible, and in some cases the segment area parameters determined may only represent fitting parameters and not have a strictly realistic significance.

**Table 5.1 Binary interaction parameters for base segments (B-MRR1) the Extended UNISAC model.**

	Dispersive (CH <sub>3</sub> )	Polar <sup>+</sup> (CCN)	Polar <sup>-</sup> (C-Cl) <sup>a</sup>	H donor (C-Cl <sub>3</sub> )	H acceptor (CH <sub>3</sub> CO)	Water (H <sub>2</sub> O)	Empty
Dispersive (CH <sub>3</sub> )	0.00	597.00	104.30	24.90	476.40	1318.00	0.00
Polar <sup>+</sup> (CCN)	24.82	0.00	-54.86	-15.62	-287.50	242.80	0.00
Polar <sup>-</sup> (C-Cl) <sup>a</sup>	-78.45	491.95	0.00	51.90	372.00	1201.00	0.00
H donor (C-Cl <sub>3</sub> )	36.70	74.04	-30.10	0.00	552.10	826.76	0.00
H acceptor (CH <sub>3</sub> CO)	26.76	481.70	-39.20	-354.55	0.00	472.50	0.00
Water (H <sub>2</sub> O)	300.00	112.60	497.54	353.68	-195.40	0.00	0.00
Empty	0.00	0.00	0.00	0.00	0.00	0.00	0.00

<sup>a</sup>CCl<sub>4</sub> was used

Table 5.2 Moodley-Rarey-Ramjugernath Group Fragmentation Scheme with Segment area parameters

No	Name	Description	Example	Priority	Segment Area						
					Dispersive	Polar <sup>+</sup>	Polar <sup>-</sup>	H donor	H acceptor	Water	Empty
Aliphatic carbon groups											
1	CH3-(ne)	CH3 attached to a non-aromatic non-electronegative atom	Pentane	151	0.8480	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	CH3-(e)	CH3 group attached to a non-aromatic electronegative atom	N,N-Dimethylformamide (DMF)	148	0.8480	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	-C(c)H2-	CH2 in a chain	Octane	157	0.5400	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	>C(c)H-	CH in a chain	2-Methylhexane	160	0.2280	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	>C(c)<	C in a chain	2,2,4-Trimethylhexane	162	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	-CH2(c)-(e)	CH2 in a chain attached to an electronegative atom	Propan-1-ol	152	0.5400	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	>CH(c)-(e)	CH in a chain attached to an electronegative atom	Pentan-3-ol	153	0.2280	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	>C(c)<(e)	C in a chain attached to an electronegative atom	tert-Pentanol	154	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	-C(r)H2-	CH2 in a ring	Cyclopentane	159	0.5200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
11	>C(r)H-	CH in a ring	Methylcycloheptane	161	0.2280	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	>C(r)<	C in a ring	1,1- Dimethylcyclohexane	163	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	-C(r)=C(r)-	Double bond between carbon in a ring with two other neighbours	1-Methyl cycloheptene	136	2.4547	0.2728	0.0781	0.0000	0.0000	0.1833	0.0029
14	>CH(r)-(e,c)	CH in a ring attached to an electronegative atom	Cyclooctanol	155	0.1008	0.3500	0.0000	0.0000	0.0000	0.0000	0.0000
15	>C(r)<(e,c)	C in a ring attached to an electronegative atom	Perfluorocyclohexane	156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	-C(c)H=C(c)-	Double bonded carbon in a chain with only 1 carbon neighbour	2-Heptene	137	2.4547	0.2728	0.0781	0.0000	0.0000	0.1833	0.0029
21	-C(r)H=C(r)H-	Double bonded C in a ring with 1 C neighbours	Cycloheptene	139	2.4547	0.2728	0.0781	0.0000	0.0000	0.1833	0.0029
22	-C(c)#C(c)-	Triple bond between 2 Cs in a chain	2-Hexyne	141	-	-	-	-	-	-	-

24	-C(r)H2-(en)	CH2 in a ring attached to electronegative neighbour	1,4-Dioxane	158	0.5400	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
25	C(c)H#C(c)-	C triple bonded to another C at the end of a chain	1-Nonyne	140	-	-	-	-	-	-	-
26	C(c)H2=C(na)<	Double bonded C at the end of a chain/ring	Methylenecyclohexane	133	2.4547	0.2728	0.0781	0.0000	0.0000	0.1833	0.0029
27	-C(c)=C(c)<	Double bonded C in a chain with 2 C neighbours	2-Methyl-2-pentene	134	2.4547	0.2728	0.0781	0.0000	0.0000	0.1833	0.0029
29	CH3-C(r)-	Methyl group connected to a C in a ring	2-Methylcyclooctane	138	0.8480	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
131	CH(r1)-C(r2)	CH in a ring bonded to a C in a different ring	(1-Propyl)-1,1'-bicyclohexyl	126	0.5200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
134	C(r)_C(k)	C in a ring double bonded to a C outside the chain	Beta-Pinene	122	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
136	C-(C(r))<-C(r)	C in a ring bonded to 3 ring Cs and 1 chain	(1-Methylethyl)-1,1'-bicyclohexyl	124	1.8004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
137	C-4C(r)	Ring C attached to 4 other ring Cs	Spiro[4.5]decane	125	0.5200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
139	C(r)_3C(r)_en	Ring C bonded to 3 other ring Cs and an en atom	1-Nitroadamantane	123	0.2009	0.0159	0.1155	0.0063	0.0474	0.0000	0.3284

## Aromatic Carbon groups

164	3	CH3-(a)	CH3 group attached to an aromatic atom	Toluene	149	0.2280	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	16	-CH(a)<	CH in an aromatic ring	Benzene	144	0.2012	0.0162	0.0198	0.0070	0.0580	0.0000	0.0041
	17	>C(a)<	C in an aromatic ring	o-Xylene	147	0.2280	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	18	>C(a)<(e)	C in an aromatic ring attached to an electronegative atom	Aniline	135	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	19	(a)=C(a)<2(a)	Aromatic C attached to three aromatic neighbours	Naphthalene	132	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	132	-C(ac)-C(ac)-	Aromatic C bonded to a C in a ring	1,2,3,4-Tetrahydronaphthalene	145	0.2009	0.0159	0.1155	0.0063	0.0474	0.0000	0.3284
	133	-C(ac)-C(r)-	Aromatic C chain bonded to a C chain in an aromatic ring	Benzidine	146	0.2009	0.0159	0.1155	0.0063	0.0474	0.0000	0.3284
	135	C(a)-r-C(a)	Aromatic C bonded to an aromatic C in a ring	9H-Fluorene	131	0.2009	0.0159	0.1155	0.0063	0.0474	0.0000	0.3284
	138	-C(ac)-C(r)=	Aromatic C attached to a double bonded C	trans-1,2-Diphenylethene	143	0.2009	0.0159	0.1155	0.0063	0.0474	0.0000	0.3284

## Fluorine groups

	35	F-C(na)	F attached to non-aromatic C	2-Fluoropentane	108	-	-	-	-	-	-	-
	37	F-C(a)	F attached to aromatic C	Fluorobenzene	91	0.0100	0.0000	0.0000	1.0200	0.0000	0.0000	0.0000

38	F-C(na)-1Halo	F attached to a C in a chain with one other halogen atom	1,1-Difluoroethane	71	-	-	-	-	-	-	-
39	F-C(na)-2Halo	F attached to a C in a chain with at least two other halogen atoms	1,1,1- Trifluoroheptane	70	-	-	-	-	-	-	-
Chlorine groups					-	-	-	-	-	-	-
40	Cl-C(na)	Cl attached to nonaromatic C	2-Chloroheptacosane <sup>a</sup>	92	0.1530	0.0000	0.0000	0.3004	0.0000	0.0000	0.0000
41	Cl-C(a)	Cl attached to aromatic C	Dichloronaphthalene <sup>a</sup>	93	0.1530	0.0000	0.0000	0.3004	0.0000	0.0000	0.0000
43	Cl-C(na)-1Halo	Cl attached to a C in a chain with one other halogen atom	1,1,Dichlorohexacosane <sup>a</sup>	67	0.0000	0.0000	0.0000	1.1940	0.0000	0.0000	0.0000
44	Cl-C(na)-2Halo	Cl attached to a C with at least two other halogen atoms	1,1,1-Trichloroeicosane <sup>a</sup>	66	0.0000	0.0000	0.0000	0.8840	0.0000	0.0000	0.0000
Bromine groups											
45	Br-C(na)	Br attached to a non-aromatic C	-	94	-	-	-	-	-	-	-
144	Br-C(na)-1Halo	Br attached to a C with one other halogen atom	-	91	-	-	-	-	-	-	-
145	Br-C(na)-2Halo	Br attached to a C with at least two other halogen atom	-	90	-	-	-	-	-	-	-
46	Br-C(a)	Br attached to aromatic C	-	95	-	-	-	-	-	-	-
Iodine groups											
47	I-C(na)	I attached to a non-aromatic C	-	77	-	-	-	-	-	-	-
Oxygen groups											
48	C(a)-COOH	Aromatic COOH	Benzoic acid <sup>a</sup>	40	0.0080	0.1985	0.6914	0.0000	0.0000	0.0000	0.0000
49	C(c)-OH	OH Group attached to a chain C	1-Octanol	109	0.1511	0.0000	0.0000	0.0000	0.6155	0.8110	0.0000
50	C(a)-OH	Aromatic OH	2-Methyl phenol	110	0.0045	0.0000	0.0000	0.0000	0.0000	0.7910	0.0000
51	C(na)-O-C(na)	Ether O	Diethyl ether	112	0.0094	0.0016	0.0000	0.0915	0.4123	0.1087	0.0000
52	C(a)-O(a)-C(a)	Aromatic O	5-Methyl Furfural	111	0.0094	0.0016	0.0000	0.0915	0.4123	0.1087	0.0000
53	C(na)-COOH	COOH Group attached to a C	Acetic acid	41	0.008	0.1985	0.6914	0	0	0	0
54	C(c)-COO-C(c)	Ester in a chain	Butyl acetate	42	0.008	0.1985	0.6914	0	0	0	0
55	HCOO-C(c)	Formic acid ester	Methyl formate	44	-	-	-	-	-	-	-
56	C(r)-C(r)OO-C(r)	Ester in a ring (lactone)	gamma-Caprolactone	43	-	-	-	-	-	-	-
57	O=C(a)<	Ketone bonded to aromatic ring	p-Methylacetophenone	74	0.0000	0.0000	0.0000	0.0000	1.1800	0.0000	0.0000
58	O=C(na)<	Ketone	Butanone	75	0.0000	0.0000	0.0000	0.0000	1.4880	0.0000	0.0000

59	HCO-C(na)	Aldehyde in Chain	Acetaldehyde	73	0.0000	0.0861	0.0000	0.0001	0.0398	0.6191	0.0008
60	HCO-C(a)	Aldehyde attached to an aromatic ring	Benzaldehyde	72	0.0000	0.1961	0.0000	0.0001	0.4980	0.0000	0.1000
63	(-C=O-O-C=O-)r	Cyclic anhydrides with double or aromatic bond	-	25	-	-	-	-	-	-	-
64	>(OC2)<	Epoxide	-	70	-	-	-	-	-	-	-
68	>N-(C=O)-N<	Urea	-	23	-	-	-	-	-	-	-
69	-OCON<	Carbamate Methylidimethylcarbamate	-	22	-	-	-	-	-	-	-
163	C(r)-OH	OH Group attached to a ring C	Cyclononalol	46	0.0114	0.0000	0.0000	0.0265	0.7155	0.0110	0.0000
Nitrogen groups											
70	-CONH2	Amide with no substituents	Acetamide	45	-	-	-	-	-	-	-
71	-CO0NH-	Amide with one substituent attached to the nitrogen	-	27	-	-	-	-	-	-	-
72	-CON<	Amide with two substituents attached to the nitrogen	N,N-Dimethylformamide (DMF)	28	-	-	-	-	-	-	-
74	ON=C-	Oxime	-	48	-	-	-	-	-	-	-
75	NO2-C(na)	Nitro group attached to a non-aromatic C	1-Nitroheptane	37	-	-	-	-	-	-	-
76	NO2-C(a)	Nitro group attached to an aromatic C	Nitrobenzene	38	-	-	-	-	-	-	-
78	-ON=C	Isoxazole O-N=C	-	39	-	-	-	-	-	-	-
80	NH2-C(na)	Primary amine attached to non-aromatic C	Octylamine	67	-	-	-	-	-	-	-
81	NH2-C(a)	Primary amine attached to an aromatic C	Benzidine	30	-	-	-	-	-	-	-
82	-N(na)H-	Secondary amines	N,N-Diethylamine	114	-	-	-	-	-	-	-
86	(C,Si)=N-	Secondary amines attached to one Cs via a double bond	-	113	-	-	-	-	-	-	-
84	>N(na)-	Tertiary amine	Triethylamine	118	-	-	-	-	-	-	-
83	-N=N-	Azene N=N	Azobenzene	66	-	-	-	-	-	-	-
140	N<-C(c)	Tertiary amine attached to C/silicons in a ring	Tricyclohexylamine	119	-	-	-	-	-	-	-
142	N-N<	Hydrazine with 2 C neighbours	-	63	-	-	-	-	-	-	-
87	N(a)-(Q5)	Aromatic nitrogen in a five-membered ring	Pyrrole	116	-	-	-	-	-	-	-
88	N(a)-(Q6)	Aromatic nitrogen in a six-membered ring	Pyridine	115	-	-	-	-	-	-	-
89	(C)-CtN	CN Group attached to a C	Acetonitrile	76	0.0000	1.7240	0.0000	0.0000	0.0000	0.0000	0.0000
94	N<C(r)	Secondary amines attached to aromatic Cs/silicons	Diarylamine	49	-	-	-	-	-	-	-
Sulfur groups											



99	C(a,na)-SH	Thiol/mercaptane attached to C	-	96	-	-	-	-	-	-	-	
100	-S(na)-	Thioether	-	97	-	-	-	-	-	-	-	
101	-S(a)-	Aromatic thioether	-	98	-	-	-	-	-	-	-	
102	-SO2-	Sulfolane O=S=O	-	34	0.0000	0.0059	0.0000	0.0000	0.0206	0.0394	0.0176	
104	-SO2N<	Sulfon amides, attached to N and to S with 2 double bond O	-	54	-	-	-	-	-	-	-	
105	>S=O	Sulfoxide	-	56	0.0000	0.0000	0.0000	0.0000	0.0000	1.8100	0.0000	
Phosphorous groups												
97	O=P	O double bonded to a phosphorus	-	61	-	-	-	-	-	-	-	
Metal groups												
110	(C)2>Sn<(C)2	Stannane with four C neighbours	-	78	-	-	-	-	-	-	-	
Special groups												
156	Ortho	(NH2, OH, COOH or NO2) in the ortho position on the ring	-	-	-	-	-	-	-	-	-	
157	Meta	(NH2, OH, COOH or NO2) in the meta position on the ring	-	-	-	-	-	-	-	-	-	
158	Para	(NH2, OH, COOH or NO2) in the para position on the ring	-	-	-	-	-	-	-	-	-	
300	C-H4	Methanol Component	Methanol	1	0.0000	0.2612	0.0000	0.0806	0.1622	0.6230	0.3630	
301	H2O	Water Component	Water	2	0.0000	0.0000	0.0000	0.0000	0.0000	1.4000	0.0000	
303	(C)-CtN	Acetonitrile component	Acetonitrile	3	0.0000	1.7240	0.0000	0.0000	0.0000	0.0000	0.0000	
304	CCl3	Trichloromethane Component	Trichloromethane	4	0.0000	0.0000	0.0000	2.4100	0.0000	0.0000	0.0000	
305	CCl4	Tetrachloromethane Component	Tetrachloromethane	5	0.0000	0.0000	2.9100	0.0000	0.0000	0.0000	0.0000	
306	4C(r)-O	Tetrahydrofuran Group	Tetrahydrofuran	6	-	-	-	-	-	-	-	
307	5C(r)-N	Pyridine Group	Pyridine	7	-	-	-	-	-	-	-	
308	Additional	-	-	8	-	-	-	-	-	-	-	
309	Additional	-	-	9	-	-	-	-	-	-	-	
310	O=S<2(CH3)	Dimethyl Sulfoxide Component	Dimethyl Sulfoxide	10	-	-	-	-	-	-	-	
311	C(c)=C(c)-C(c)#N	Acrylonitrile Group	Acrylonitrile	11	-	-	-	-	-	-	-	
312	HCF3	Trifluoromethane Component	Trifluoromethane	12	-	-	-	-	-	-	-	
313	H3C-O-COOH	Methyl Methanoate Component	Methyl Methanoate	13	-	-	-	-	-	-	-	
314	R-O-C(c)-C(c)-O-R'	Oxy-ethanol with branch on 2 Os	2-Methoxyethanol Nitrate	14	-	-	-	-	-	-	-	

---

315	$\begin{array}{c} \text{R-O-C(c)-C(c)-O-} \\ \text{H} \end{array}$	Oxy-ethanol with branch on 1 O	2-Ethoxyethanol	15	-	-	-	-	-	-
-----	--------------------------------------------------------------------	--------------------------------	-----------------	----	---	---	---	---	---	---

<sup>a</sup> Parameters apply to solid solubility calculations only

---

### 5.3 Error Estimation

From equation (5.2), it is evident that the solubility temperature is strongly dependent on the composition and activity coefficient at saturation. The dependence of the activity coefficient  $\ln \gamma_i$  on the compositions  $x_i$  and  $x_j$  of a multicomponent system can be expressed by the following relationships for the Extended UNISAC model:

$$\begin{aligned}
 \frac{\partial \ln \gamma_i}{\partial x_i} &= \frac{\partial \ln \gamma_i^{Comb}}{\partial x_i} + \frac{\partial \ln \gamma_i^{Res}}{\partial x_i} \\
 &= -V'_i (V'_Z - V'_i) + (V'_Z - V'_i) - 5q_i \left\{ - \left( \frac{V_i (V_Z - V_i) - V_i \times (F_Z - F_i)}{F_i} \right) + \right. \\
 &\quad \left. (V_Z - V_i) - (F_Z - F_i) \right\} + \sum_{k=1}^N X_{k,i} \left[ \frac{-1}{\sum_{m=1}^N \Theta_m \psi_{m,k}} \times \right. \\
 &\quad \left. \sum_{m=1}^N \psi_{m,k} \left\{ \frac{((X_{m,i} - X_{m,Z}) \sum_{k=1}^N \sum_{i=1}^I X_{m,i} x_i) - (\sum_{i=1}^I X_{m,i} x_i) (\sum_{m=1}^N (X_{m,i} - X_{m,Z}))}{(\sum_{m=1}^N \sum_{i=1}^I X_{m,i} x_i)^2} \right\} - \right. \\
 &\quad \left. \sum_{m=1}^N \left\{ \frac{\psi_{k,m} \left( \frac{((X_{m,i} - X_{m,Z}) \sum_{m=1}^N \sum_{i=1}^I X_{m,i} x_i) - (\sum_{i=1}^I X_{m,i} x_i) (\sum_{m=1}^N (X_{m,i} - X_{m,Z}))}{(\sum_{m=1}^N \sum_{i=1}^I X_{m,i} x_i)^2} \right) (\sum_{n=1}^N \Theta_n \psi_{n,m}) -}{\theta_m \psi_{k,m} (\sum_{n=1}^N \left( \frac{((X_{n,i} - X_{n,Z}) \sum_{n=1}^N \sum_{i=1}^I X_{n,i} x_i) - (\sum_{i=1}^I X_{n,i} x_i) (\sum_{n=1}^N (X_{n,i} - X_{n,Z}))}{(\sum_{n=1}^N \sum_{i=1}^I X_{n,i} x_i)^2} \right))} \right\} \right] \quad (5.3)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial \ln \gamma_i}{\partial x_j} &= \frac{\partial \ln \gamma_i^{Comb}}{\partial x_j} + \frac{\partial \ln \gamma_i^{Res}}{\partial x_j} \\
 &= -V'_i (V'_Z - V'_j) + (V'_Z - V'_j) - 5q_i \left\{ - \left( \frac{V_i (V_Z - V_j) - V_i \times (F_Z - F_j)}{F_i} \right) + \right. \\
 &\quad \left. (V_Z - V_j) - (F_Z - F_j) \right\} + \sum_{k=1}^N X_{k,i} \left[ \frac{-1}{\sum_{m=1}^N \Theta_m \psi_{m,k}} \times \right. \\
 &\quad \left. \sum_{m=1}^N \psi_{m,k} \left\{ \frac{((X_{m,j} - X_{m,Z}) \sum_{k=1}^N \sum_{i=1}^I X_{m,i} x_i) - (\sum_{i=1}^I X_{m,i} x_i) (\sum_{m=1}^N (X_{m,j} - X_{m,Z}))}{(\sum_{m=1}^N \sum_{i=1}^I X_{m,i} x_i)^2} \right\} - \right.
 \end{aligned}$$

$$\sum_{m=1}^N \left\{ \begin{array}{l} \psi_{k,m} \left( \frac{((x_{m,j}-x_{m,Z}) \sum_{i=1}^N x_{m,i} x_i) - (\sum_{i=1}^N x_{m,i} x_i) (\sum_{i=1}^N (x_{m,j}-x_{m,Z}))}{(\sum_{i=1}^N x_{m,i} x_i)^2} \right) (\sum_{n=1}^N \theta_n \psi_{n,m}) - \\ \theta_m \psi_{k,m} (\sum_{n=1}^N \left( \frac{((x_{n,j}-x_{n,Z}) \sum_{i=1}^N x_{n,i} x_i) - (\sum_{i=1}^N x_{n,i} x_i) (\sum_{i=1}^N (x_{n,j}-x_{n,Z}))}{(\sum_{i=1}^N x_{n,i} x_i)^2} \right)) \end{array} \right\} \quad (5.4)$$

Where the symbols used follow the convention of those used in equations (5.2) to (5.14). The subscript  $Z$  used here represents the identity of the  $i^{th}$  component in the mixture, the composition of which is dependent on the other  $i-1$  components.

Equations (5.16) and (5.17) were used to determine the influence of the composition uncertainty of the solubility data tested on the calculated solubility temperatures, i.e.:

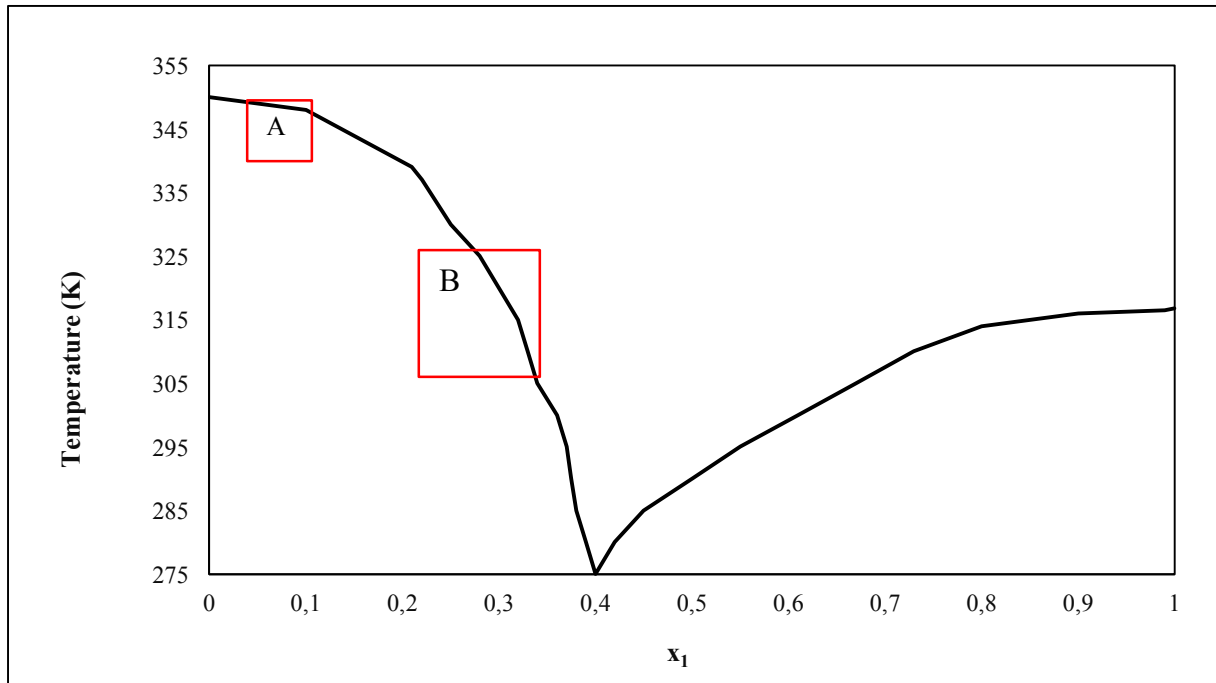
$$\frac{\partial(\ln T)}{\partial x_i} = \frac{R_{fus} T}{\Delta_{fus} H_i} \left( \frac{\partial(\ln \gamma_i)}{\partial x_i} + \frac{\partial(\ln x_i)}{\partial x_i} \right) \quad (5.5)$$

These relationships were then used to determine the accuracy and acceptable error of prediction of the solubility temperature determined from the Extended UNISAC model.

Similarly, the influence of temperature uncertainty on the equilibrium composition can be expressed by:

$$\frac{\partial(\ln x_i)}{\partial T} = \frac{\Delta_{fus} H_i}{R_{fus} T} \left( \frac{\partial(\ln T)}{\partial T} \right) - \frac{\partial(\ln \gamma_i)}{\partial T} \quad (5.6)$$

These influences were used to compensate the predictions of solubility composition or temperature for all the models tested in this work. This was done to ensure that the apparent differences in for instance the predicted composition is not due to a steep gradient in the temperature/composition profile. This is illustrated for a generic solubility profile in Figure 5.1. The influence of temperature on composition is clearly greater in region B, than in region A.



**Figure 5.1 Change in the influence of temperature on the experimental liquid-phase composition in SLE.**

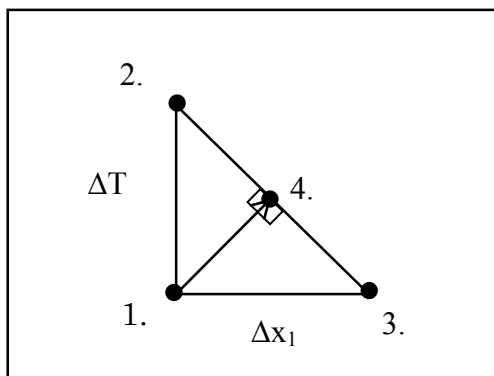
The standard experimental uncertainty in temperature ( $\sigma_T$ ) for all data points considered was conservatively assumed to be 0.01 K while the experimental uncertainty in composition ( $\sigma_{x_i}$ ) was assumed to be  $0.01 \cdot x_i$ . The combined uncertainty in temperature taking into account the effect of composition uncertainty is then:

$$u(T) = \Delta T = (\sigma_T^2 + (\sigma_{x_i} \frac{\partial T}{\partial x_i})^2)^{1/2} \quad (5.7)$$

And similarly in the case of composition:

$$u(x_i) = \Delta x_i = (\sigma_{x_i}^2 \sigma_{x_i} + (\sigma_T \frac{\partial x_i}{\partial T})^2)^{1/2} \quad (5.8)$$

Then for an experimental data point and prediction the following can be identified:f



**Figure 5.2 Calculation of unbiased predicted solubility.**

Referring to Figure 5.2., if for instance the experimental data point is located at point 1, then a large  $\Delta T$  yielding a temperature biased prediction would be generated at point 2, due to a region of high temperature gradient. Similarly at point 3, a composition biased prediction is generated. A temperature and composition unbiased prediction is assumed to occur at point 4 i.e. the orthogonal intersection of the prediction extrema and the experimental data point.

#### *5.4 Results and discussion*

In order to test the performance of the Extended UNISAC model, with respect to solid solubility predictions, experimental data for a diverse set of solid-liquid equilibrium measurements were compared to results of Extended UNISAC, original UNIFAC (public, Dortmund Data Bank version, 2012), modified UNIFAC (Dortmund) (public, Dortmund Data Bank version, 2012), COSMO-RS(OL) and COSMO-SAC. The NRTL-SAC model was not used for comparison as it is not purely predictive, and a search for the large number of suitable NRTL-SAC model parameters that would be required for testing was not feasible or within the scope of this work for the large number of components tested here. An excellent model comparison, including NRTL-SAC is provided by Diedrichs and Gmehling (2011).

The experimental data used here are primarily obtained from the Dortmund Data Bank (2012), and are comprised of pharmaceutical ingredient or complex molecule in a single solvent as well as multicomponent solvent/solute systems. The useable data was carefully selected to remove duplicate measurements with considerable differences in the experimental data points, clearly erroneous experimental data points (above pure component melting points), and

components for which Extended UNISAC segment area parameters are not available. Components for which UNIFAC or modified UNIFAC (Dortmund) group fragmentation could not be performed or cases where group interaction parameters were unavailable were omitted from the common database. Similarly, components for which COSMO sigma profiles were not available, were also excluded from the common database used for comparison. A principal component analysis was performed using solubility in water, alkane, alcohol and fusion (melting) temperature as descriptive factors, and the data set was confirmed to be heterogeneous.

In total, solubility predictions were performed for over 4000 experimental data points from 655 data sets and systems made up of 385 different components. The complete set of components tested and experimental data sets used are provided in the supplementary information using the Dortmund Data Bank (2012) data set identification numbers in Table B1 and Table B2 of Appendix B. For the common database used, all models were applicable.

The squared natural logarithmic deviations in solubility prediction of pharmaceutical components in pure organic solvent classes and water, are presented in Figures 5.3-5.8.

$$SLD = \sum_{i=1}^N (\ln x_i^{pred} - \ln x_i^{exp})^2 \quad (5.9)$$

Where  $x_i^{pred}$ ,  $x_i^{exp}$  are the calculated and experimental solute compositions and  $N$  is the total number of data points considered

The results (Figures 5.3 to 5.8) are compared to predictions from popular literature models as a measure of currently available prediction quality. In all cases the temperature/composition discrepancy is accounted for as shown in Figure 5.2. It can be seen that the Extended UNISAC model provides satisfactory qualitative predictions of solubility for the cases tested. It must be noted that the fragmentation failure of the UNIFAC model also affects the combinatorial expression of the Extended UNISAC model, but can be easily overcome using pure component volumes and surface areas such as in the UNIQUAC model. In nearly every case differences in the combinatorial expression has a rather small effect on the solubility, so uncertainties in

the R and Q values have limited effect on the result. Some system specific plots of predictions and experimental data are presented in Appendix B for the common solvent classes. Additionally a step by step calculation, along with additional reference calculations is provided

In Figure 5.9 and 5.10, the five models tested are compared by calculating either the composition (Figure 5.9) or temperature (Figure 5.10), by the various models and expressing the portion of the data set with a fraction deviation larger than a given value. It is evident that the Extended UNISAC model is competitive with the other models tested. The modified UNIFAC (Dortmund) model is however superior to all models tested, and predicts the solubility within 10 mol % for approximately 90% of systems tested. It is also one of the worst performers in the remaining 10%, which likely constitutes the aqueous mixtures. The UNIFAC predictions are within 15 mol % and the Extended UNISAC model is within 18 mol % for 90% of the data tested. Similar results are obtained when the temperature is predicted for a given solubility. Although the PHARM modified UNIFAC (Diedrichs and Gmehling, 2011) model was not tested for comparison in this work, comparisons to Extended UNISAC can be made indirectly. In the work of (Diedrichs and Gmehling, 2011) PHARM modified UNIFAC is stated to perform very similarly to UNIFAC and modified UNIFAC (Dortmund) but slightly outperforms the UNIFAC model overall based on root mean square deviation only. This implies that it will outperform Extended UNISAC. PHARM modified UNIFAC is however limited in application to alkane, alcohol and water as solvent, and has only been applied to systems with solute compositions less than 0.1 mole fraction.

As in the case of the original UNISAC model (Moodley et al., 2015 (b)), the Akaike Information Criterion (AIC) (Akaike, 1975) was used to evaluate the relative effectiveness of the five models tested in this work, considering accuracy of prediction, number of calculable data points, and number of model parameters used. The Extended UNISAC model yielded the lowest (most favourable) AIC score for the common data base systems tested, with relative AIC scores of 1.95, 4.17, 2.17 and 2.09 for the UNIFAC, modified UNIFAC (Dortmund), COSMO-RS(OL) and COSMO-SAC models respectively. The overall percentage deviations (PD) between experimental and predicted values, calculated by the relation below are also presented.

$$PD = 100 \left( \frac{\sum_{i=1}^N (x_i^{pred} - x_i^{exp})^2}{\sum_{i=1}^N (x_i^{exp} - \bar{x}^{exp})^2} \right)^{1/2} \quad (5.10)$$

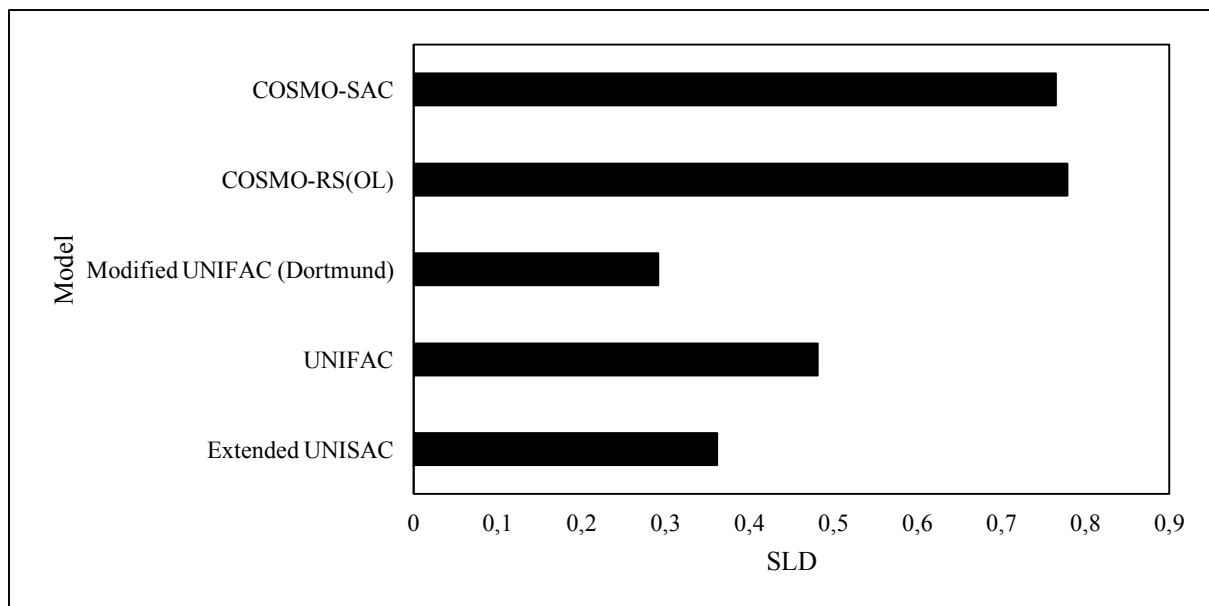


Where  $x_i^{pred}$ ,  $x_i^{exp}$  are the calculated and experimental solute compositions and  $N$  is the total number of data points considered.  $\bar{x}^{exp}$  is the average experimental composition for a particular set.

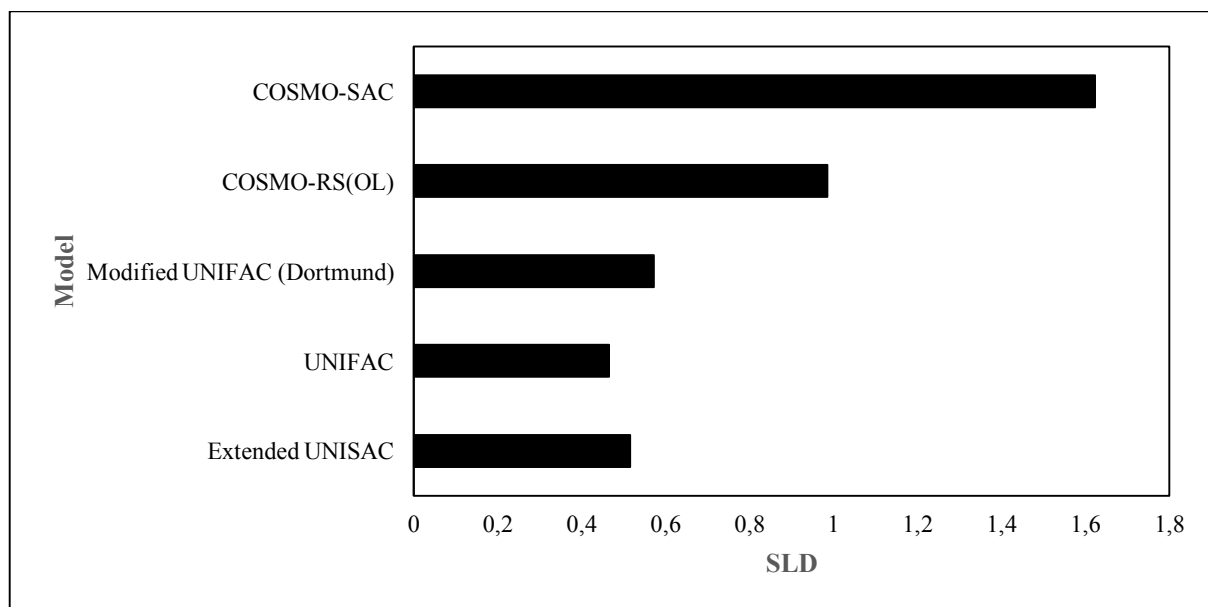
This representation was chosen over the root mean square deviation (RMSD) used in the original UNISAC work presented earlier, as the experimental solubility data considered here was of largely varying magnitudes. These results are presented in Table 5.3 along with the relative AIC scores.

The PD for the modified UNIFAC (Dortmund) model is the lowest in this case as PD = 21.03%. This gives an indication of the reliability of the modified UNIFAC (Dortmund) model in this test set. In terms of the percentage deviation (PD), the Extended UNISAC model performs slightly poorer (PD = 32.99%) than original UNIFAC (PD = 29.03 %). The COSMO-based models yielded PDs greater than 60%.

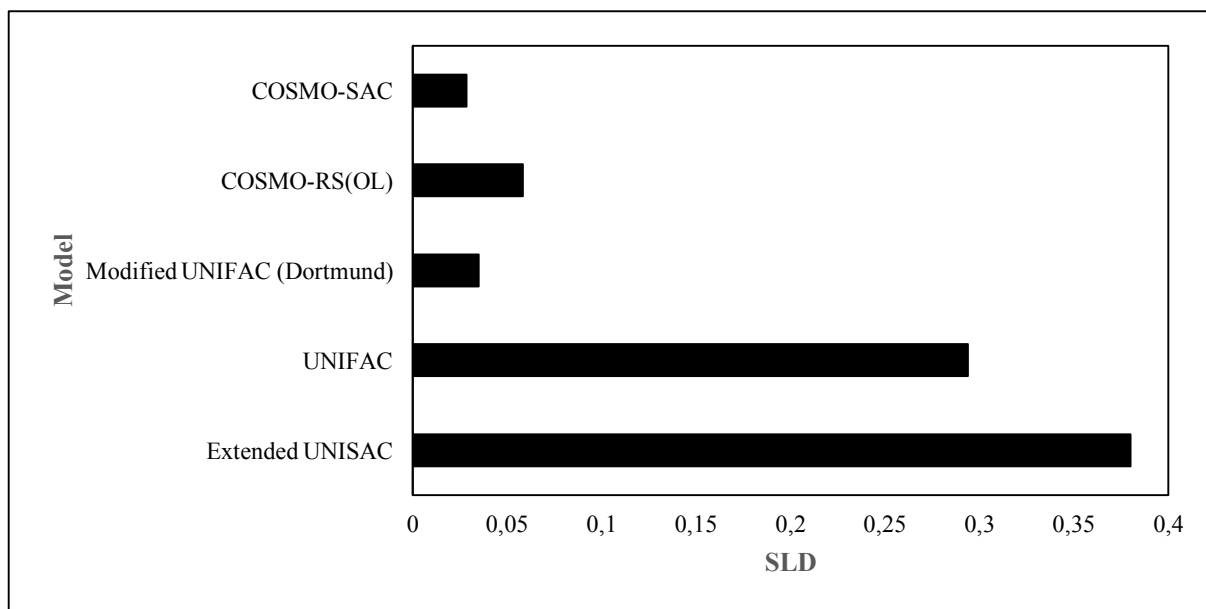
To test the ability of Extended UNISAC to predict solubility in ternary systems results for Extended UNISAC, UNIFAC, modified UNIFAC (Dortmund), COSMO-RS(OL) and COSMO-SAC are presented for a common database of multicomponent SLE systems in Figure 11. This demonstrates that the model has the potential to be versatile and applied to these types of systems. In the majority of the systems presented, adequate SLE temperatures can be predicted, and the model competes well with the existing models in the literature. The prediction quality of all the models is however qualitative in many instances. Further work in this area will be required to assess the model for application in multi-component systems and also assessing the quality of the experimental data used for comparison. Some example system specific plots of predictions and experimental data for ternary systems are presented in Appendix B.



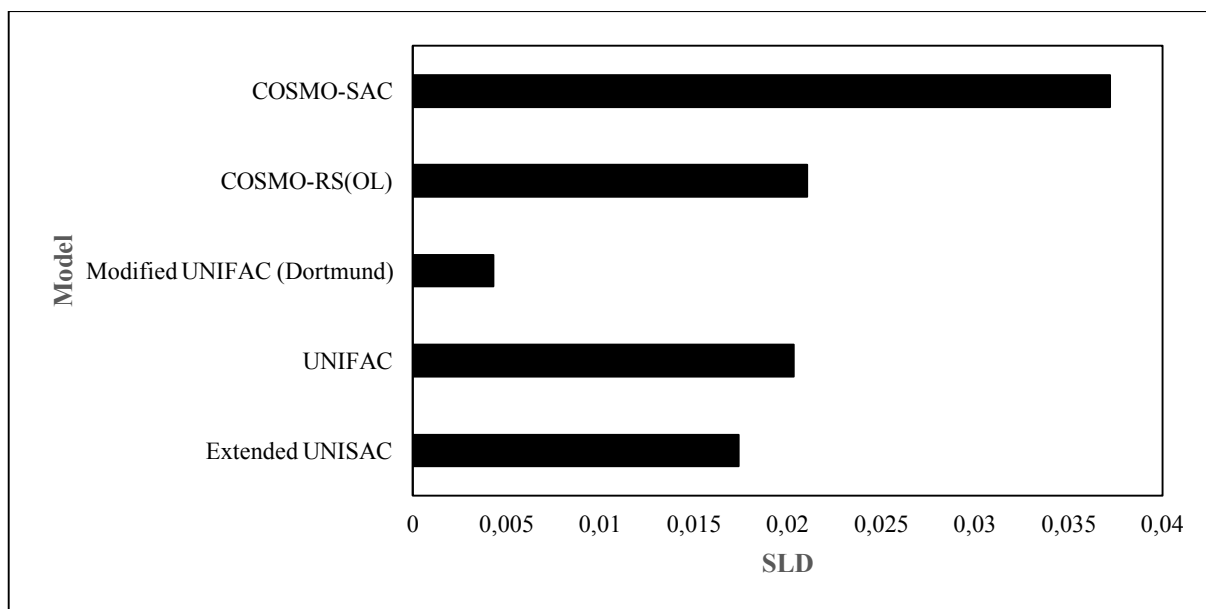
**Figure 5.3** Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in alcohol solvents.



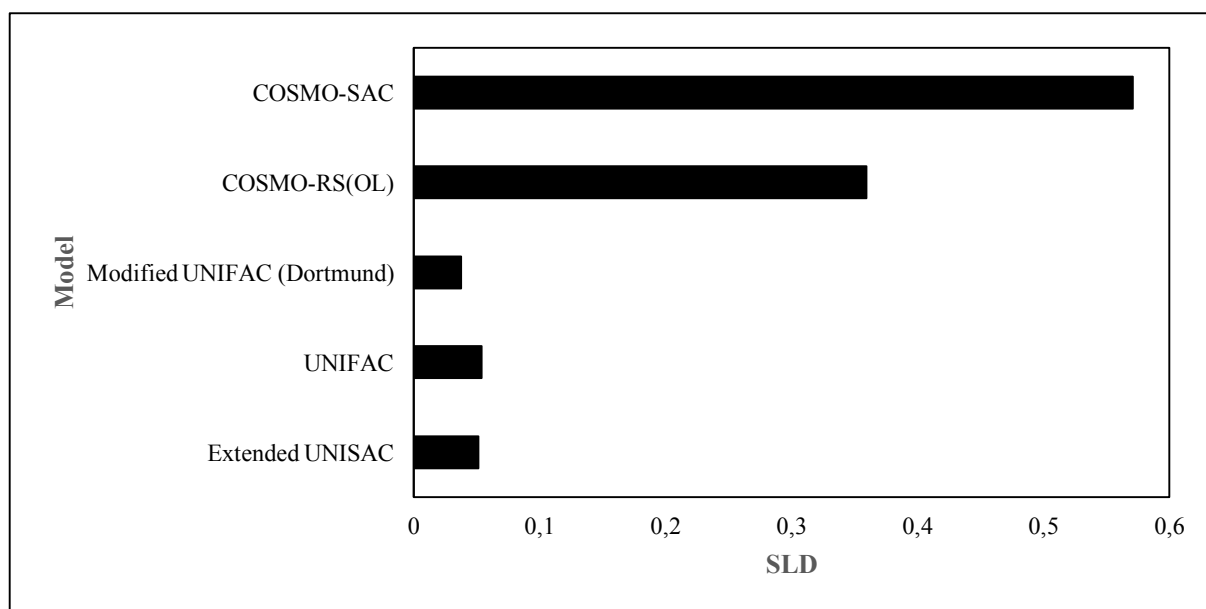
**Figure 5.4** Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in alkane solvents.



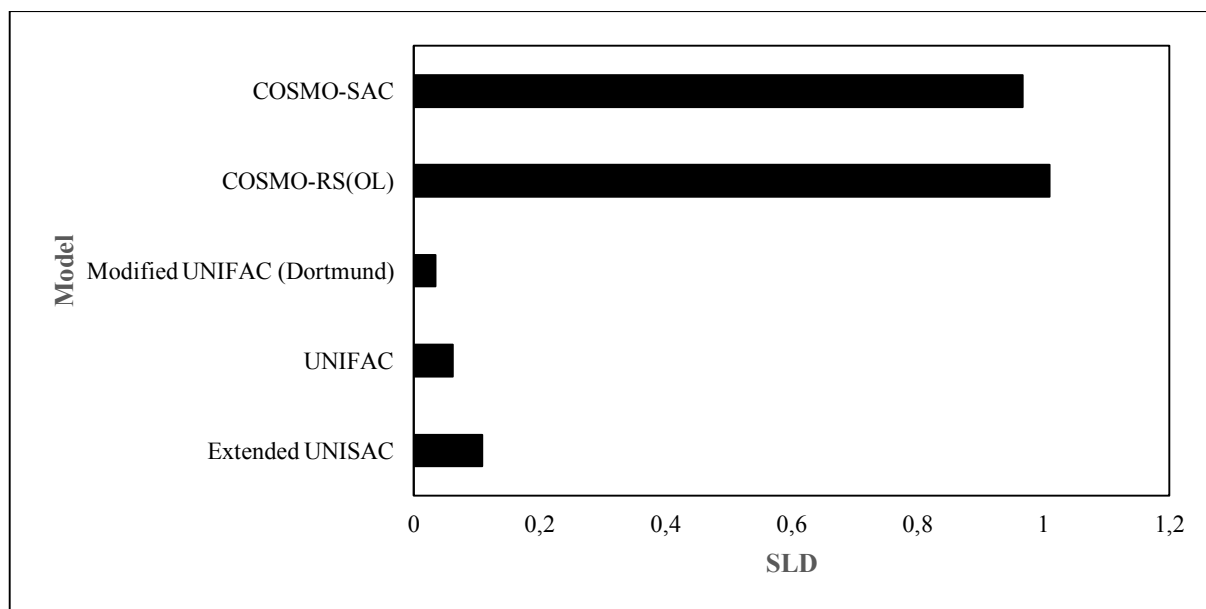
**Figure 5.5 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in aromatic solvents.**



**Figure 5.6 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in ether solvents.**



**Figure 5.7** Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in ketone solvents.



**Figure 5.8** Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition in water as a solvent.

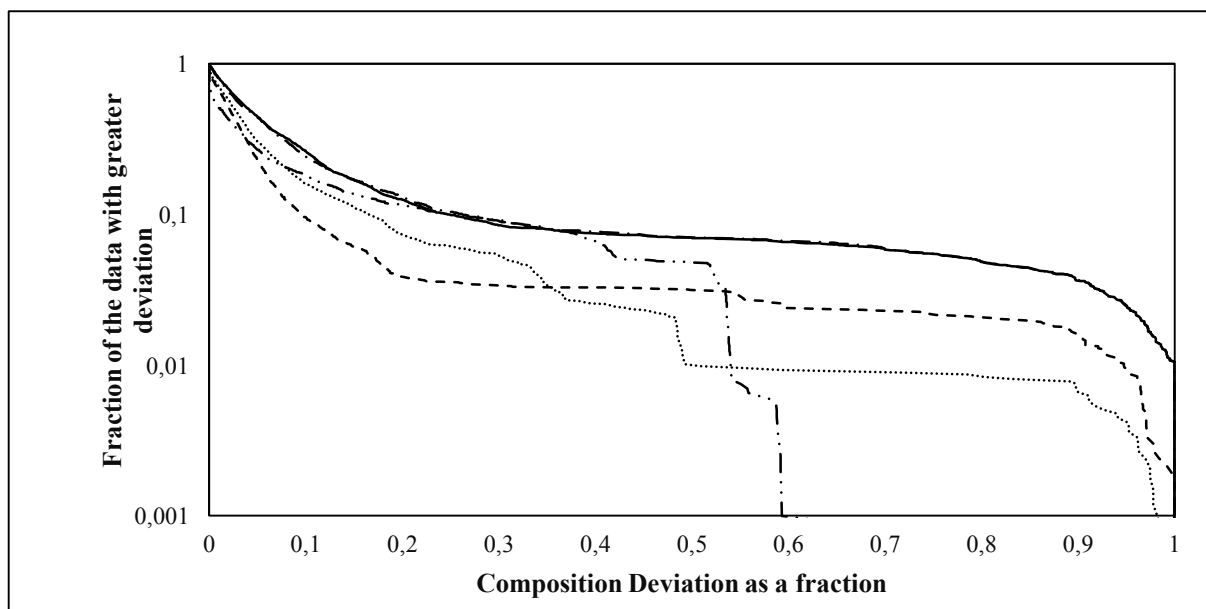


Figure 5.9 Fraction of the data with deviations in composition larger than a given composition for various binary non-aqueous solute-solvent systems, by different models.  $- \cdot -$ , Extended UNISAC;  $- - -$ , mod. UNIFAC (Dortmund);  $\cdots$ , UNIFAC;  $-$ , COSMO-SAC;  $- \cdot -$ , COSMO-RS (OL).

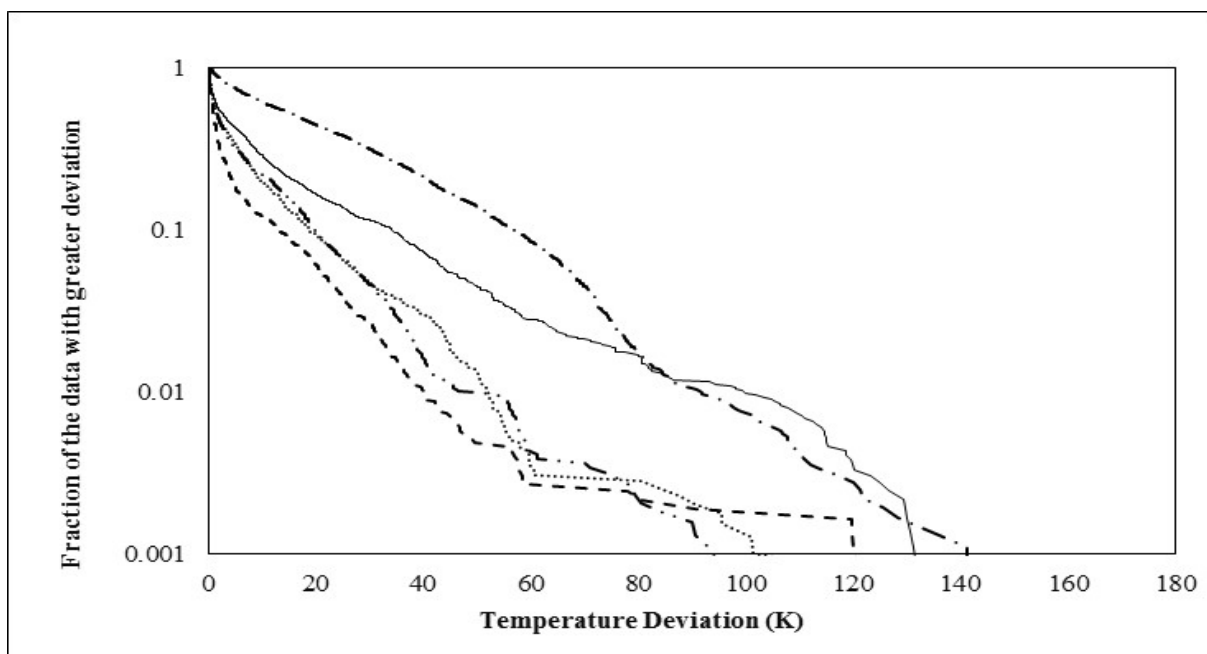


Figure 5.10 Fraction of the data with deviations in temperature larger than a given temperature for various binary non-aqueous solute-solvent systems, by different models.  $- \cdot -$ , Extended UNISAC;  $- - -$ , mod. UNIFAC (Dortmund);  $\cdots$ , UNIFAC;  $-$ , COSMO-SAC;  $- \cdot -$ , COSMO-RS (OL).

Table 5.3 Relative Akaike Information Criterion scores and percentage deviations.

Model	Relative AIC score	PD
Extended UNISAC	0	32.99
UNIFAC	1.95	29.03
Mod. UNIFAC (Dortmund)	4.17	21.03
COSMO-RS (OL)	2.17	65.97
COSMO-SAC	2.09	61.60

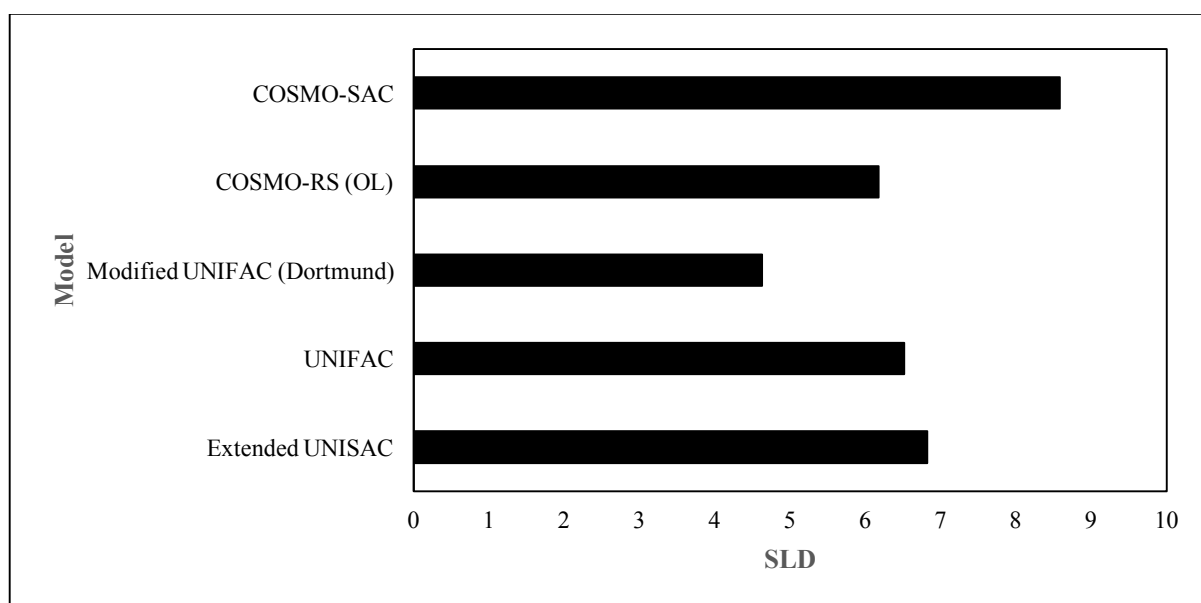


Figure 5.11 Comparison of the Square natural logarithmic deviations (SLD) of experimental (Dortmund Data Bank, 2012) and model calculated solubility composition for various ternary solubility systems.

### 5.5 Conclusion

The Extended UNISAC model provides a practical and simple method of performing qualitative solubility predictions for systems containing complex pharmaceutical ingredients, often in cases where existing popular prediction models fail. The model requires a marginally larger number of segment area parameters than the original UNISAC model, but is able to handle a larger range of structurally diverse components. A significant number of updated Extended UNISAC segment area parameters are available, which facilitate predictions for a substantial number of the available SLE systems in the Dortmund Data Bank (2012), for a total of 59 groups and 30 unique segments. The Akaike Information Criterion suggests that the Extended UNISAC model is favourable to the original UNIFAC, modified UNIFAC (Dortmund), COSMO-RS(OL) and COSMO-SAC models. Percentage deviations between experimental and model calculated Extended UNISAC predictions, are competitive with the existing popular predictive methods.

*References*

- Abrams, D.S. and Prausnitz, J.M., (1975), Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems. *AIChE Journal*, 21, pp.116–128.
- Akaike, H., (1974), A new look at the statistical model identification. *IEEE Transactions on Information Technology in Biomedicine*, 19, pp.716–723.
- Bondi, A., (1964), van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68, pp.441–451.
- Chen, C.-C. and Song, Y., (2004), Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Industrial and Engineering Chemistry Research*, 43, pp.8354–8362.
- DDBST Software and Separation Technology GmbH, (2012), Dortmund Data Bank (DDB).
- Derr, E.L. and Deal, C.H., (1969), Analytical solution of groups: correlation of activity coefficients through structural group parameters. *Institute of Chemical Engineers Symposium Series*, 32, pp.44–51.
- Diedrichs, A. and Gmehling, J., (2011), Solubility calculation of active pharmaceutical ingredients in alkanes, alcohols, water and their mixtures using various activity coefficient models. *Industrial and Engineering Chemistry Research*, 50, pp.1757–1769.
- Flory, P.J., (1942), Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics*, 10, p.51.
- Fredenslund, A., Jones, R.L. and Prausnitz, J.M., (1975), Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21, pp.1086–1099.
- Gandomi, A.H. and Alavi, A.H., (2012), Krill herd: A new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, 17, pp.4831–4845.
- Grensemann, H. and Gmehling, J., (2005), Performance of a Conductor-Like Screening Model for Real Solvents Model in Comparison to Classical Group Contribution Methods. *Industrial and Engineering Chemistry Research*, 44, pp.1610–1624.
- Guggenheim, E.A., (1952), *Mixtures*, Oxford, U.K.
- Hala, E., (1978), Some Notes on The Group Solution Concept. *Collection Czechoslov. Vhem. Commun.*, 43, pp.10–15.
- Hansen, C.M., (1967), *The Three Dimensional Solubility Parameter and Solvent Diffusion Coefficient and Their Importance in Surface Coating Formulation*, Danish Technical Press, Copenhagen.



- Hildebrand, J.H., (1916), Solubility. *Journal of the American Chemical Society*, 38, pp.1452–1473.
- Hildebrand, J.H. and Scott, R.L., (1964), *The solubility of nonelectrolytes*. 3rd ed., Dover, New York.
- Hsieh, C.-M., Sandler, S.I. and Lin, S.-T., (2010), Improvements of COSMO-SAC for vapor–liquid and liquid–liquid equilibrium predictions. *Fluid Phase Equilibria*, 297, pp.90–97.
- Huggins, M.L., (1941), Solutions of long chain compounds. *The Journal of Chemical Physics*, 9(1941), p.440.
- Klamt, A., (1995), Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *Journal of Physical Chemistry*, 99, pp.2224–2235.
- Lin, S.-T. and Sandler, S.I., (2002), A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Industrial and Engineering Chemistry Research*, 41, pp.899–913.
- Moller, B., Rarey, J. and Ramjugernath, D., (2008), Estimation of the vapour pressure of non-electrolyte organic compounds via group contributions and group interactions. *Journal of Molecular Liquids*, 143, pp.52–63.
- Moller, B., Rarey, J. and Ramjugernath, D., (2014), Extrapolation/interpolation of infinite dilution, activity coefficient as well as liquid and solid solubility between solvents: Part 1. Alkane solvents. *Fluid Phase Equilibria*, 361, pp.69–82.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (a), A Universal Segment Approach for the Prediction of the Activity Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents. *Fluid Phase Equilibria*, 396, pp.98–110.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (b), Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations. *Computers and Chemical Engineering*, 74, pp.75–88.
- Mu, T., Rarey, J. and Gmehling, J., (2007), Performance of COSMO-RS with Sigma Profiles from Different Model Chemistries. *Industrial and Engineering Chemistry Research*, 46, pp.6612–6629.
- Neau, S.H., Bhandarkar, S. V and Hellmuth, E.W., (1997), Differential molar heat capacities to test ideal solubility estimations. *Pharmaceutical research*, 14, pp.601–5.
- NIST, NIST TDE. Available at: <http://trc.nist.gov/tde.html>.
- Staverman, A.J., (1950), The entropy of high polymer solutions. Generalization of formulae. *Recueil des Travaux Chimiques des Pays-Bas*, 69, pp.163–174.
- UNIFAC Consortium, UNIFAC Consortium (unifac.org).

Weidlich, U. and Gmehling, J., (1987), A modified UNIFAC Model. 1. Prediction of VLE,  $h^E$ , and  $\gamma^\infty$ . *Industrial and Engineering Chemistry Research*, 26, pp.1372–1381.

Wilson, G.M., (1964), Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing. *Journal of the American Chemical Society*, 86, pp.127–130.

## CHAPTER SIX

### **Experimental Solubility Data For Betulin/ Estrone/ Diosgenin/ Estriol/ Prednisolone/ Hydrocortisone In Various Solvents In The Temperature Range T = (293.2 To 328.2) K**

#### *Abstract*

The solubility of complex triterpenes Lup-20(29)-ene-3 $\beta$ ,28-diol (betulin), (8R,9S,13S,14S)-3-hydroxy-13-methyl-6,7,8,9,11,12,13,14,15,16-decahydrocyclopenta[a]phenanthren-17-one (estrone), systems (3 $\beta$ ,25R)-spirost-5-en-3-ol (diosgenin), (16 $\alpha$ ,17 $\beta$ )-Estra-1,3,5(10)-triene-3,16,17-triol (estriol), (11 $\beta$ )-11,17,21-trihydroxypregna-1,4-diene-3,20-dione (prednisolone) and (11 $\beta$ )-11,17,21-trihydroxypregn-4-ene-3,20-dione (hydrocortisone) in numerous common solvents were determined by combined thermal gravimetry and digital thermal analysis (TGA/DTA). The measurements were conducted at atmospheric pressure and within the temperature range of T = (293.2 to 328.2) K. Melting point and enthalpy of fusion data of the solutes considered were also determined by DTA/TGA. The measured data were modelled using the Non-Random Two Liquid (NRTL) and Tsuboka-Katayma modified Wilson activity coefficient models which incorporated a polynomial temperature dependence. The models provided a very good correlation of the experimental data in most cases.

#### *6.1 Introduction*

The majority of separation processes employed in industry are based on promoting a mixture to form two phases, the compositions of which are generally different from the feed mixture. The conditions of temperature, composition, and pressure at which a certain phase will appear is governed by thermodynamic principles and the equilibrium condition. Crystallization is generally a solid-liquid separation process commonly employed in the petrochemical and pharmaceutical industries, in the production of waxes, polymers and pharmaceutical ingredients. Novel research, development, and optimization is common in such industries, as the need for the exploration of alternate process routes and new product development is often required. To accomplish this, it is essential that the physical behaviour of the mixtures being considered is well described by either experimental data or accurate model predictions.

Betulin is a naturally occurring pentacyclic triterpene found in the bark of birch wood (Green et al., (2007)). Recent studies have shown the effectiveness of betulin derivatives in the treatment of numerous cancers including melanoma, neuroblastoma, lung carcinoma, and as a Human Immunodeficiency Virus inhibitor, (Alakurtti et al. (2006), Pisha et al. (1995), Fulda (1999), Evers et al. (1996)). There is, however, limited research available in the literature on the solid-liquid behaviour of betulin in common solvents at equilibrium, (Cao et al. (2007), Zhao and Yan (2006). Zhao and Yan (2007)), which is imperative for the design of efficient betulin extraction processes.

Estrone is a common estrogenic hormone used mainly as a supplement drug in post-hysterectomy and post-menopausal women, to assist in regulating normal endocrine behaviour. Nevertheless, the physical behaviour of estrone in common solvents is not readily available in the literature, (Ruchelman, (1967), Shareef et al. (2006), Sanghvi et al. (2008)).

Diosgenin is a naturally occurring steroidal sapogenin found in the roots of legumes and yams. The additive is being generally employed in the pharmaceutical industry for the production of common steroids such as cortisone and progesterone, but recent studies have explored the effectiveness of the component as a cancer chemo-preventative, (Malisetty et al. 2005), (Raju and Mehta, 2008). Although the interest in diosgenin is recently increasing, limited information about the solubility of diosgenin in common solvents is available in the literature (Chen et al. (2012), Chen et al. (2012)).

Estriol is an estrogenic hormone naturally produced by the human body. Studies have shown the effectiveness of estriol in the treatment of arthritis, encephalomyelitis and multiple sclerosis, (Jansson et al., (1994), Kim et al., (1999), Bebo et al., (2001), Sicotte et al., (2002)). A limited amount of solid-liquid equilibrium data is available in the literature for estriol in common solvents (Ruchelman and Howe, 1969).

Prednisolone is a polycyclic hydrocarbon steroid that is used extensively for the treatment of a variety of auto-immune diseases in humans and other mammals Czock et al. (2005). Although commonly produced in the pharmaceutical industry, little research has been carried out to explore possible improvements to the separation techniques involved in the isolation of prednisolone (Kabasakalian et al., (1966), Yalkowsky and Valvani, (1980), Regosz et al., (1994)).

Hydrocortisone is a steroidal polycyclic hydrocarbon used primarily for immunosuppression in humans. Limited solubility data of hydrocortisone-solvent systems are available in the

literature with the majority of measurements undertaken at approximately 298 K (Kabasakalian et al., (1966), Yalkowsky and Valvani, (1980), Regosz et al., (1994), Hagen and Flynn, (1983), Li et al., (2002)).

The solubility behaviour of these six pharmaceutical components was determined in more than 10 solvents at atmospheric pressure in the temperature range of  $T = (293.2 \text{ to } 328.2) \text{ K}$  in this study. The solvents used included: nonan-1-ol, pentan-1-ol, acetonitrile, 2-aminoethanol, n-dodecane, butan-2-ol, N,N-dimethylformamide, water, n-octane, octan-1-ol, morpholine-4-carbaldehyde, n-hexadecane and 1-methyl-2-pyrrolidone. Furthermore, the experimental data measured were modelled using the Non-Random Two Liquid (NRTL) (Renon and Prausnitz, 1968) and Tsuboka-Katayama modified Wilson (Tsuboka and Katayama, 1975) activity coefficient models which incorporated a polynomial temperature dependence.

## 6.2 Theory

### 6.2.1 Thermodynamic relationship

In the case of a eutectic system not undergoing solid-solid phase transitions, i.e. a single solid phase exists, Hildebrand and Scott (1964) have shown that:

$$\ln(x_i^l \gamma_i^l) = -\frac{\Delta_{fus}H_i(f_{us}T_i)}{R f_{us}T_i} \ln\left(\frac{T}{f_{us}T_i}\right) \quad (1.41)$$

In equation (1.41),  $x_i^l$  is the liquid mole fraction of the solute at the saturation condition,  $\gamma_i^l$  is the activity coefficient of the solute in the saturated solution,  $\Delta_{fus}H_i$  is the enthalpy of fusion at the melting temperature  $f_{us}T_i$ ,  $R$  is the Universal Gas Constant and  $T$  is the temperature of the system. In this expression, the difference between the solid heat capacity and sub-cooled liquid heat capacity is estimated by the entropy of fusion at the triple or melting point.

$\gamma_i^l$  is a function of composition at a given temperature and can be calculated using a Gibbs excess energy model.

### 6.2.2 Activity coefficient models

The NRTL (Renon and Prausnitz, 1968) and Tsuboka-Katayama modified Wilson (Tsuboka and Katayama, 1975) models were used to correlate the experimental data. A brief description of the models is presented below.

#### 6.2.2.1 The Non-Random Two-Liquid (NRTL) model (Renon and Prausnitz, 1968)

The activity coefficient expression of the NRTL model for a binary mixture is given by:

$$\ln \gamma_i = x_j^2 \left[ \tau_{ji} \left( \frac{G_{ji}}{x_i + x_j G_{ji}} \right)^2 + \left( \frac{\tau_{ij} G_{ij}}{(x_j + x_i G_{ij})^2} \right) \right] \quad (6.1)$$

where

$$G_{ij} = \exp(-\alpha_{ij} \tau_{ij}) \quad (6.2)$$

and

$$\tau_{ij} = \frac{g_{ij} - g_{jj}}{RT} \quad (6.3)$$

The interaction parameters  $(g_{ij} - g_{jj})$  are the fitting parameters and are determined by regression. The non-randomness parameter  $\alpha_{12}$ , is usually set to a constant value but can also be obtained by data regression, should it provide a better quality fit to the experimental data.

In order to incorporate a stronger temperature dependence of the model fitting parameters,  $\tau_{ij}$  is usually modified. The following polynomial expression for the temperature dependence of the NRTL model parameters was employed in this study:

$$\tau_{ij} = \frac{A_{ij} + \frac{B_{ij}}{T}}{RT} \quad (6.4)$$

where  $A_{ij}$  and  $B_{ij}$  are adjustable parameters determined by data regression.

### 6.2.2.2 The Tsuboka-Katayama (T-K) Wilson model

The activity coefficient using the T-K-Wilson model (1975) for a binary mixture is given by:

$$\ln \gamma_i = \ln \left[ \frac{x_i + V_{ij}x_j}{x_i + \Lambda_{ij}x_j} \right] + x_j \left[ \frac{\Lambda_{ij}}{x_i + \Lambda_{ij}x_j} - \frac{\Lambda_{ji}}{x_j + \Lambda_{ji}x_i} + \frac{V_{ji}}{x_j + V_{ji}x_i} - \frac{V_{ji}}{x_i + V_{ij}x_j} \right] \quad (6.5)$$

where  $V_{ij}$  are the ratios of the molar volume given by:

$$V_{ij} = \frac{V_j}{V_i} \quad (6.6)$$

and

$$\Lambda_{ij} = \frac{V_j}{V_i} \exp \left[ \frac{-(\lambda_{ij} - \lambda_{ii})}{RT} \right] \quad (6.7)$$

The cross interaction parameters ( $\lambda_{ij} - \lambda_{ii}$ ) are determined by reduction of  $x_i$  and  $T$  data in the case of SLE and represent the different interactions between the molecules in the mixture.

The expression ( $\lambda_{ij} - \lambda_{ii}$ ) was modified in order to provide a stronger dependence of temperature, in which case:

$$\Lambda_{ij} = \frac{V_j}{V_i} \exp \left[ \frac{-(C_{ij} + \frac{D_{ij}}{T})}{RT} \right] \quad (6.8)$$

where  $C_{ij}$  and  $D_{ij}$  are now the variable parameters determined by data regression.

### 6.3 Experimental methods

#### 6.3.1 Materials

The betulin, estrone, estriol, prednisolone and hydrocortisone crystals were purchased from Sigma-Aldrich and had a supplier stated purity of greater than 0.99 mass fraction, hence no further purification procedures were performed. The diosgenin crystals used in this work were purchased from Sigma-Aldrich with a stated purity of  $>0.93$  mass fraction. This solute was recrystallized twice with acetone. The acetone/diosgenin solution was heated till the solute dissolved completely, and was filtered while heated to remove any un-dissolved impurities. Recrystallization of the solute was allowed to occur naturally. The crystals were filtered and washed repeatedly with water (conductivity of  $27.6 \mu\text{S}\cdot\text{m}^{-1}$ ) and re-filtered. The collected crystals were then dried in a desiccator at an ambient temperature not exceeding 300 K for 24 hours. Prior to use, the crystals of all solutes used were dried in a desiccator for 24 hours, and the final purities were subsequently confirmed by gas chromatography mass spectroscopy (GC-MS) analysis using a pyrolizer prior to measurement. A GC 2010 Plus in combination with a GCMS-QP2010 Ultra and EGA/PY-330D Multi-Shot Pyrolizer supplied by Shimadzu was used. A Zebron<sup>TM</sup> 7HG-G010-11 capillary column (30m x 0.25mm x 0.25 $\mu\text{m}$  film thickness) was used. The GC method incorporated a temperature ramp of  $5^\circ\text{C min}^{-1}$  from  $70^\circ\text{C}$  to  $280^\circ\text{C}$ . Helium was used as the carrier gas with a flowrate of  $0.91 \text{ ml}\cdot\text{min}^{-1}$ . The organic solvents used were reagent grade with stated purities  $> 0.99$  mass fraction. This was confirmed by gas chromatography using a thermal conductivity detector. The refractive indices of the solvents used were determined at 293.15 K using an ATAGO RX-7000 $\alpha$  refractometer with a reported standard uncertainty in refractive index of 0.0001. The water used was ultra-pure water with a conductivity of  $27.3 \mu\text{S}\cdot\text{m}^{-1}$ . All alcohols were dehydrated using molecular sieve for 24 hours prior to usage. The purities, refractive indices and chemical suppliers of the components used are provided in Table 6.1.



Table 6.1 Chemical suppliers and purities.

Component	CAS No.	Supplier	Refractive index (RI) at 293.15 K and 0.101 MPa. <sup>†</sup>		Minimum stated mass fraction purity	GC peak relative area (mass fraction purity)
			Exptl.	Lit.		
<b>Solvents</b>						
Nonan-1-ol	143-08-8	Sigma-Aldrich	1.4334	1.4333 <sup>a</sup>	≥0.99	0.9999 <sup>d</sup>
Pentan-1-ol	71-41-0	Sigma-Aldrich	1.4098	1.4101 <sup>a</sup>	≥0.99	0.9970 <sup>d</sup>
Acetonitrile	75-05-8	Sigma-Aldrich	1.3439	1.3442 <sup>b</sup>	≥0.999	0.9999
2-Aminoethanol	141-43-5	Sigma-Aldrich	1.4542	1.4541 <sup>a</sup>	≥0.99	0.9903 <sup>d</sup>
n-Dodecane	112-40-3	Merck	1.4218	1.4210 <sup>a</sup>	≥0.99	0.9999
Butan-2-ol	78-92-2	Sigma-Aldrich	1.3983	1.3988 <sup>a</sup>	≥0.99	0.9916 <sup>d</sup>
N,N Dimethylformamide	68-12-2	Aldrich	1.4304	1.4305 <sup>a</sup>	≥0.999	0.9999
n-Octane	111-65-9	Merck	1.3978	1.3974 <sup>b</sup>	≥0.99	0.9999
Octan-1-ol	111-87-5	Reidel-de Haën	1.4291	1.4295 <sup>a</sup>	≥0.995	0.9950 <sup>d</sup>
Morpholine-4-carbaldehyde	4394-85-8	Merck	1.4849	1.4845 <sup>a</sup>	≥0.99	0.9999
N-Hexadecane	544-76-3	Sigma-Aldrich	1.4346	1.4345 <sup>a</sup>	≥0.99	0.9999
1-Methyl-2-pyrrolidone	872-50-4	Merck	1.4690	1.4684 <sup>a</sup>	≥0.995	0.9999
Water <sup>c</sup>	7732-18-5	-	1.3329	1.3330 <sup>a</sup>	-	-
<b>Solutes</b>						
Lup-20(29)-ene-3β,28-diol (Betulin)	473-98-3	Sigma-Aldrich	-	-	≥0.99	≥0.99 <sup>e</sup>
(8R,9S,13S,14S)-3-hydroxy-13-methyl-6,7,8,9,11,12,13,14,15,16-decahydrocyclopenta[a]phenanthren-17-one (Estrone)	53-16-7	Sigma-Aldrich	-	-	≥0.99	≥0.99 <sup>e</sup>
(3β,25R)-Spirost-5-en-3-ol (Diosgenin)	512-04-9	Sigma-Aldrich	-	-	≥0.93 <sup>f</sup>	≥0.99 <sup>e</sup>
(16α,17β)-Estra-1,3,5(10)-triene-3,16,17-triol (Estriol)	50-27-1	Sigma-Aldrich	-	-	≥0.99	≥0.99 <sup>e</sup>
(11β)-11,17,21-trihydroxypregna-1,4-diene-3,20-dione (Prednisolone)	50-24-8	Sigma-Aldrich	-	-	≥0.99	≥0.99 <sup>e</sup>
(11β)-11,17,21-trihydroxypregn-4-ene-3,20-dione (Hydrocortisone)	50-23-7	Sigma-Aldrich	-	-	≥0.99	≥0.99 <sup>e</sup>

1. <sup>†</sup>Standard uncertainties  $u$  are  $u(RI) = 0.0001$ ,  $u(T) = 0.01K$ ,  $u(P) = 0.002 MPa$

2. <sup>a</sup>Lide (2005) at 293.15 K, <sup>b</sup>James and Lord (1992) at 293.15K, <sup>c</sup>Water conductivity = 27.6  $\mu S.m^{-1}$ , <sup>d</sup>Purified by molecular sieving, <sup>e</sup>GC-MS fraction relative abundance, <sup>f</sup>Purified by recrystallization with acetone

### 6.3.2 *Experimental procedure*

In this study combined thermal gravimetry and digital thermal analysis was used to determine saturated compositions of a liquid phase at solid-liquid equilibrium. A similar experimental procedure has been employed by Zhang et al. (2010) and Yu et al. (2013). The solute and solvent were prepared in a sealable glass vessel. Excess solute was used to ensure a saturated solution in equilibrium with excess solid was formed. The vessel was then placed in an ultrasonic thermostated bath at the experimental temperature (within 1 K) to induce thorough mixing by vibration. This was conducted for four hours. The vessel was then removed and inspected to insure that the mixture formed was still saturated.

The glass vessel was then placed in a large thermostated bath, at the experimental temperature, controlled by a Grant GD120 controller with an uncertainty of 0.1 K, confirmed by calibration using a WIKA CTH6500 standard shown in Appendix C. This controlled temperature represents the solid-liquid equilibrium temperature. Mild mechanical vibration was employed to induce agitation, and it was conservatively assumed that the mixture reached equilibrium after 24 hours.

After 24 hours the liquid phase that forms was carefully extracted using a gas tight syringe to eliminate partial evaporation during transfer, with care not to disturb the underlying solid layer. DTA/TGA was used to determine the composition of the liquid phase by mass difference using a Shimadzu DTG-60AH with a mass readability of 0.0001mg for mass and 0.1 K for temperature. The liquid sample (90mg -950mg) from the gas tight syringe was placed in a small sample pan and immediately weighed in the DTA/TGA apparatus. The sample was then control heated in the DTA/TGA apparatus at 10 K/min to 10 K below the boiling point of the solvent (or 10 K below the melting point of solute if it is lower than the boiling point of the solvent), under a N<sub>2</sub> flow of 1 cm<sup>3</sup>.s<sup>-1</sup>, which guaranteed a constant evaporation rate of the solvent. Boiling of the solvent is undesirable as it can cause entrainment of the solid out of the pan, resulting in erroneous masses being measured. Hence N<sub>2</sub> purging is necessary. Aluminium oxide (Al<sub>2</sub>O<sub>3</sub>) was used as a reference material in a sealed crucible.

The liquid phase equilibrium composition of the original mixture at the experimental temperature is determined by the difference in mass of the initial sample in the pan (solute + solvent mass) and the final mass in the pan (solute mass). In Figure 6.1 a generic thermogram is shown. It can be seen that changes in the differential temperature of a sample under investigation correlate with changes in the percentage mass loss of the sample, relative to the

starting point. For example, at time A, it is evident that the percentage mass loss of the sample begins to stabilize, indicating that all the solvent in the mixture has evaporated, and that only the solute remains. This corresponds to a change in the differential temperature of the sample as indicated in Figure 6.1. It is also useful that combined DTA/TGA allows for the observance of any chemical reaction or impurity, signalled by an additional peak, which would not be easily observed by TGA alone. These measurements were conducted in triplicate for each temperature measured. The average of the three values are presented as the molar composition at equilibrium. The maximum standard deviation between each repeatable run was calculated to be less than 0.19 %. Since there was no detectable indication of evaporation of the solutes in any of the thermograms generated for the mixture, (only solvent evaporation and boiling is indicated), and for the pure solute (virtually zero change in mass until the melting point) it was assumed that partial vaporization of the solutes did not occur in any of the measurements performed.

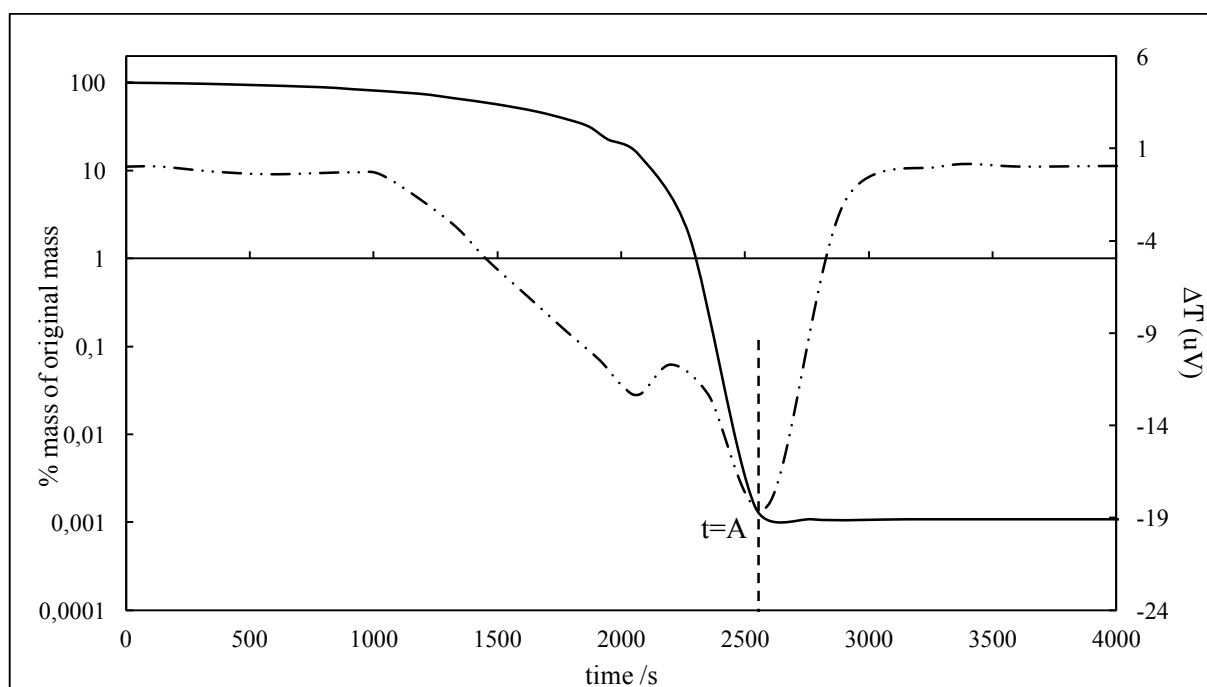


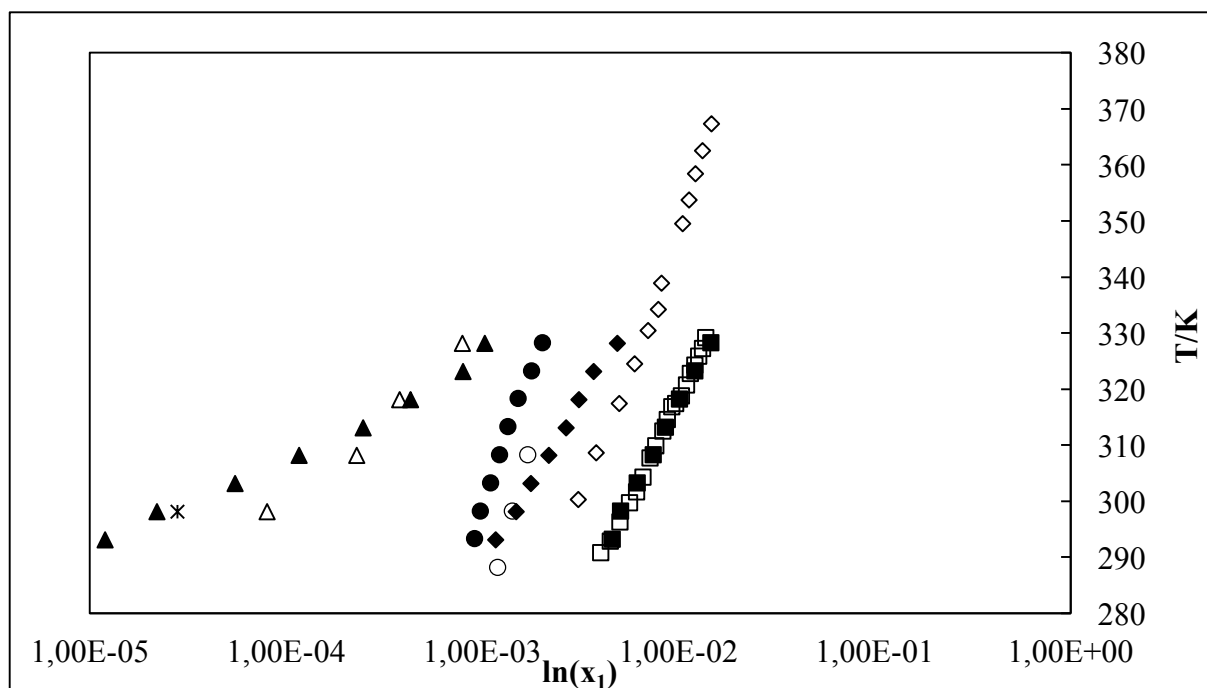
Figure 6.1 A generic combined TG-DTA plot; (—) TG, (---) DTA.

#### 6.4 Results and discussion

In Figure 6.2, comparisons of the systems measured in this work to those in the literature are shown.

The system of betulin + pentan-1-ol has been measured by Cao et al. (2006) in an overlapping temperature range to this study. A comparison with the data measured in this work reveals that although the solubility trend is very similar in both data sets, a minor difference that is not within the experimental uncertainty of either data sets is apparent. In this work this system was measured in triplicate and the authors are confident in the results presented. The differences in the data sets are attributed to the chemical purities and differences in experimental procedure. Specifically, in the work of Cao et al. (2006), the procurement and recrystallization procedure for the betulin crystals, as well as the mixing time and mechanism prior to equilibration, which is not mentioned in the text.

Furthermore the systems of estrone + water has been measured by Domańska et al. (2010) in an overlapping temperature range, and Shareef et al. (2006) at 298.15 K. At 298.2 it is evident that the data measured in this work lies in the same range of the data from the two references. At 298.2 and 308.2 K a significant difference between the data presented in this work and that of Domańska et al. (2010) is evident. There is however good agreement between the data sets at higher temperatures. Again it is reinstated that the measurements performed in this work were triplicated. The authors attribute differences in results here to the experimental method used to measure SLE compositions, (DTA-TGA in this work vs. UV-Vis spectroscopy in Domańska et al. (2010)). Furthermore the differences of the purity of the water used (established by conductivity in this work) may have been a factor. Domańska et al. (2010) also measured the system estrone + octan-1-ol, in a similar temperature range to this work. A significant difference between the data measured in this work, and Domańska et al. (2010) is evident at the lower temperatures. There is reasonable agreement at  $T > 300$  K. The authors are confident in the measurements performed in this work. Differences in this system are attributed to the measurement techniques used (DTA-TGA in this work vs. a visual technique in Domańska et al. (2010)). Because of these results, it was decided to perform a further test measurement to confirm the technique and procedures used in this work. The system of diosgenin + pentan-1-ol was measured. The agreement between the data of Chen et al. (2012) and that determined in this work is excellent, and serves as a further confirmation of the experimental procedures used in this study.



**Figure 6.2** T-x plot for the systems betulin (1) + pentan-1-ol (2), ●, This work, ○, (Cao et al., (2006)), systems estrone (1) + water (2), ▲, This work, △, (Domańska et al., (2010)), ×, (Shareef et al., (2006)), estrone (1) + octan-1-ol (2), ◆, This work, ◇, (Domańska et al., (2010)), diosgenin (1) + pentan-1-ol (2), ■, This work, □, (Chen et al., (2012)).

The enthalpy of fusion and melting point temperature data of betulin and estrone were measured using combined DTA/TGA, and are compared to literature data as listed in Table 6.2. Good agreement, within the experimental uncertainty, between experimental and literature values is observed.

Table 6.2 Experimental melting point and enthalpy of fusion data of solutes used at 0.101 MPa<sup>a</sup>.

Compound	Experimental		Literature	
	$\Delta_{fus}H$ (J.mol <sup>-1</sup> )	$fusT$ (K)	$\Delta_{fus}H$ (J.mol <sup>-1</sup> )	$fusT$ (K)
Betulin	55169 ± 551	528.1 ± 0.1	55160 (Zhao and Yan, 2008)	527.9 (Zhao and Yan, 2008)
Estrone	45107 ± 451	527.6 ± 0.1	45100 (Domańska et al., 2010)	527.1 (Domańska et al., 2010)
Diosgenin	52105 ± 521	474.4 ± 0.1	34064.2 (Chen et al., 2014)	480.24 (Chen et al., 2014)
Estriol	42718 ± 427	553.2 ± 0.1	42695 (Ruchelman, Howe, 1969)	555 (Ruchelman, Howe, 1969)
Prednisolone	59296 ± 592	506.8 ± 0.1	59303.2 (Cai, 1997)	506 (Cai, 1997)
Hydrocortisone	33900 ± 339	486.1 ± 0.1	33890.4 (Hagen and Flynn, 1983)	485 (Hagen and Flynn, 1983)

<sup>a</sup> Standard uncertainties  $u$  are  $u(T) = 0.1K$ ,  $u(P) = 0.002 MPa$  and the standard relative uncertainties are  $u_r(\Delta H^{fus}) = 0.01$ .

The solid-liquid equilibrium data that was measured in this study are presented in Tables 6.3 to 6.8 and in Figures 6.3 to 6.26, along with the correlative activity coefficient model that provides the best description of the data. The uncertainties in measured variables were estimated according to the procedures outlined by the National Institute of Standards and Technology (NIST, 2000). The standard relative composition uncertainty of the measurements is due to the uncertainty in the measurement of mass.

$$u_r = \frac{f(u_{MM})}{x_i^{sat}} = \frac{(u_c)}{x_i^{exp}} \quad (6.9)$$

where  $u_r$  is the standard relative uncertainty in composition and  $u_{MM}$  is the uncertainty in the measurement of mass,  $u_c$  is the standard uncertainty in composition, and  $x_i^{exp}$  is the measured composition at saturation. Since there is a small change in solubility within the standard experimental temperature uncertainty (0.1 K) in comparison to the change due to the uncertainty in mass, it is assumed that all experimental composition uncertainties result from uncertainties in mass measurement. This also includes the contribution of repeatability to uncertainty. The uncertainty in temperature and mass exhibited by the DTA/TGA apparatus

was also used to determine the uncertainty in the experimental solute melting points and enthalpies of fusion.

**Table 6.3 Experimental solid-liquid equilibrium data of betulin mixtures at various temperatures and 0.101 MPa<sup>a</sup>.**

	$x_1^{exp} \times 10^3$				
	T/K	328.2 ± 0.1	323.2 ± 0.1	318.2 ± 0.1	313.2 ± 0.2
<b>Betulin</b>					
<b>Solvent</b>					
Nonan-1-ol		5.1	4.31	3.71	3.24
Pentan-1-ol		1.98	1.75	1.49	1.32
Acetonitrile		3.41	2.44	1.93	1.37
n-Dodecane <sup>b</sup>		0.045	0.035	0.03	0.025
Butan-2-ol		1.48	1.2	1.08	0.941
N,N Dimethylformamide		5.18	4.61	4.26	3.68
Water		0.033	0.025	0.018	0.014
Octan-1-ol		3.41	2.69	2.29	2.09
N-Hexadecane		0.117	0.098	0.083	0.076
1-Methyl-2-pyrrolidone		6.97	6.76	6.53	6.35
<b>Diosgenin</b>					
<b>Solvent</b>					
Pentan-1-ol		14.8	12.2	10.2	8.66
		<b>308.2 ± 0.1</b>	<b>303.2 ± 0.1</b>	<b>298.2 ± 0.1</b>	<b>293.2 ± 0.2</b>
<b>Betulin</b>					
<b>Solvent</b>					
Nonan-1-ol		2.92	2.54	2.34	2.17
Pentan-1-ol		1.2	1.08	0.96	0.894
Acetonitrile		1.04	0.697	0.473	0.269
n-Dodecane <sup>b</sup>		0.023	0.019	0.016	0.015
Butan-2-ol		0.845	0.813	0.728	0.701
N,N Dimethylformamide		3.34	2.94	2.44	2.08
Water		0.009 <sup>c</sup>	0.006 <sup>c</sup>	0.005 <sup>c</sup>	0.004 <sup>c</sup>
Octan-1-ol		1.86	1.79	1.62	1.6
n-Hexadecane <sup>d</sup>		0.065	0.057	0.053	0.046
1-Methyl-2-pyrrolidone		6.15	5.9	5.69	5.49
<b>Diosgenin</b>					
<b>Solvent</b>					
Pentan-1-ol		7.5	6.21	5.1	4.62

<sup>a</sup> Standard uncertainties  $u$  are  $u(T) = 0.1K$ ,  $u(P) = 0.002 MPa$  and the standard relative uncertainties are  $u_r(x_1^{exp}) = 0.04$ .

<sup>b</sup> $u_r(x_1^{exp}) = 0.2$ , <sup>c</sup> $u_r(x_1^{exp}) = 6.5$ , <sup>d</sup> $u_r(x_1^{exp}) = 0.05$ .

**Table 6.4 Experimental solid-liquid equilibrium data of estrone mixtures at various temperatures and 0.101 MPa<sup>a</sup>.**

Solvent	$x_1^{exp} \times 10^3$				
	T/K	328.2 ± 0.1	323.2 ± 0.1	318.2 ± 0.1	313.2 ± 0.1
Nonan-1-ol		6.17	5.08	4.45	3.96
Pentan-1-ol		3.22	2.42	2.04	1.8
2-Aminoethanol		7.93	5.11	3.08	1.84
Acetonitrile		18.5	9.98	6.8	3.45
n-Dodecane		1.24	0.975	0.663	0.459
Butan-2-ol		2.57	2.13	1.88	1.62
N,N Dimethylformamide		19.5	14	10.1	6.32
n-Octane		0.626	0.483	0.377	0.277
Water		1.03	0.797	0.431	0.247
Octan-1-ol		4.89	3.69	3.11	2.68
Morpholine-4-carbaldehyde		70.9	57	45.9	37.4
N-Hexadecane		9.52	6.59	4.64	3.03
1-Methyl-2-pyrrolidone		11.5	9.98	7.4	5.39
	T/K	308.2 ± 0.1	303.2 ± 0.1	298.2 ± 0.1	293.2 ± 0.1
Nonan-1-ol		3.47	3.06	2.43	2.13
Pentan-1-ol		1.6	1.26	1.04	0.823
2-Aminoethanol		1.39	0.897	0.634	0.419
Acetonitrile		2.17	1.2	0.755	0.423
n-Dodecane		0.368	0.275	0.236	0.184
Butan-2-ol		1.34	1.11	0.893	0.712
N,N Dimethylformamide		4.3	2.8	1.84	1.12
n-Octane		0.235	0.189	0.16	0.137
Water		0.117	0.055	0.022 <sup>b</sup>	0.012 <sup>b</sup>
Octan-1-ol		2.19	1.77	1.49	1.17
Morpholine-4-carbaldehyde		31.7	29.1	24.2	20.3
N-Hexadecane		2.09	1.29	0.96	0.717
1-Methyl-2-pyrrolidone		4.38	3.51	2.89	2.16

<sup>a</sup> Standard uncertainties  $u$  are  $u(T) = 0.1K$ ,  $u(P) = 0.002 MPa$  and the standard relative uncertainties are  $u_r(x_1^{exp}) = 0.04$ ,

<sup>b</sup>  $u_r(x_1^{exp}) = 0.13$ .



Table 6.5 Experimental solid-liquid equilibrium data of diosgenin mixtures at various temperatures and 0.101 MPa<sup>a</sup>.

Solvent	$x^{exp} \times 10^3$				
	T/K	328.2 ± 0.1	323.2 ± 0.1	318.2 ± 0.1	313.2 ± 0.1
Nonan-1-ol		15.5	14.5	13.2	12.3
2-Aminoethanol		13.2	11.0	9.14	8.07
Acetonitrile		22.1	18.8	15.1	12.5
n-Dodecane		1.05	0.898	0.743	0.631
Butan-2-ol		3.66	2.53	1.70	1.19
N,N Dimethylformamide		11.6	7.94	5.87	4.61
n-Octane		0.915	0.680	0.466	0.302
Water		2.12	1.16	0.353	0.0884
Octan-1-ol		11.6	10.4	9.47	8.42
Morpholine-4-carbaldehyde		19.8	18.1	16.2	14.7
N-Hexadecane		2.44	1.71	1.31	0.949
1-Methyl-2-pyrrolidone		21.9	19.6	17.8	16.3
	T/K	308.2 ± 0.1	303.2 ± 0.1	298.2 ± 0.1	293.2 ± 0.1
Nonan-1-ol		11.0	9.65	8.25	6.70
2-Aminoethanol		6.25	4.20	2.18 <sup>b</sup>	1.42 <sup>b</sup>
Acetonitrile		8.94	6.75	5.29	3.59
n-Dodecane		0.542	0.439	0.351	0.266
Butan-2-ol		0.784	0.436	0.229	0.128
N,N Dimethylformamide		3.69	2.32	1.59	1.13
n-Octane		0.210	0.160	0.130 <sup>c</sup>	0.0810 <sup>c</sup>
Water		0.045 <sup>d</sup>	0.015 <sup>e</sup>	0.006 <sup>e</sup>	0.002 <sup>f</sup>
Octan-1-ol		6.60	5.16 <sup>c</sup>	4.23 <sup>c</sup>	3.48 <sup>c</sup>
Morpholine-4-carbaldehyde		12.0	9.76	8.09	6.63
N-Hexadecane		0.626	0.465	0.373	0.256
1-Methyl-2-pyrrolidone		14.8	13.8	12.8	11.8

<sup>a</sup> Standard uncertainties  $u$  are  $u(T) = 0.1K$ ,  $u(P) = 0.002 MPa$  and the standard relative uncertainties are  $u_r(x_1^{exp}) = 0.04$ ,  
<sup>b</sup> $u_r(x_1^{exp}) = 0.18$ , <sup>c</sup> $u_r(x_1^{exp}) = 0.1$ , <sup>d</sup> $u_r(x_1^{exp}) = 0.5$ , <sup>e</sup> $u_r(x_1^{exp}) = 4.2$ , <sup>f</sup> $u_r(x_1^{exp}) = 12$

**Table 6.6 Experimental solid-liquid equilibrium data of estriol mixtures at various temperatures and 0.101 MPa<sup>a</sup>.**

Solvent	$x^{exp} \times 10^3$				
	T/K	328.2 ± 0.1	323.2 ± 0.1	318.2 ± 0.1	313.2 ± 0.1
Nonan-1-ol		19.9	16.8	14.7	13.1
Pentan-1-ol		12.1	9.67	8.09	5.94
2-Aminoethanol		36.3	31.4	25.4	21.4
n-Dodecane		2.85	2.43	1.79	1.28 <sup>c</sup>
Butan-2-ol		8.12	6.68	5.33 <sup>c</sup>	4.53
N,N Dimethylformamide		18.0	13.0	9.58	5.81 <sup>c</sup>
n-Octane		1.86	1.52	1.38	1.23
Water <sup>b</sup>		0.803	0.577	0.422	0.306
Octan-1-ol		17.1	14.7	12.1	10.4
Morpholine-4-carbaldehyde		10.7	7.26	4.95	3.43 <sup>d</sup>
N-Hexadecane		3.82	3.35	2.99	2.51
1-Methyl-2-pyrrolidone		7.18	4.91	3.64	2.75 <sup>c</sup>
	T/K	308.2 ± 0.1	303.2 ± 0.1	298.2 ± 0.1	293.2 ± 0.1
Nonan-1-ol		11.6	10.4	8.63	7.40
Pentan-1-ol		4.68	3.67	2.85	2.43
2-Aminoethanol		18.0	14.8	12.7	11.1
n-Dodecane		0.956 <sup>c</sup>	0.920 <sup>c</sup>	0.779 <sup>c</sup>	0.720 <sup>c</sup>
Butan-2-ol		4.03 <sup>c</sup>	3.20	2.50	1.89
N,N Dimethylformamide		4.26 <sup>c</sup>	2.58 <sup>c</sup>	1.58 <sup>c</sup>	1.12 <sup>c</sup>
n-Octane		1.03	0.873	0.735	0.621
Water		0.191	0.168	0.135	0.115
Octan-1-ol		8.68	7.47	6.78	5.93
Morpholine-4-carbaldehyde		1.47 <sup>d</sup>	0.652 <sup>d</sup>	0.230 <sup>d</sup>	0.062 <sup>d</sup>
N-Hexadecane		2.19	1.75	1.47	1.14
1-Methyl-2-pyrrolidone		1.74 <sup>c</sup>	1.05 <sup>c</sup>	0.677 <sup>c</sup>	0.415 <sup>c</sup>

<sup>a</sup> Standard uncertainties  $u$  are  $u(T) = 0.1K$ ,  $u(P) = 0.002 MPa$  and the standard relative uncertainties are  $u_r(x_1^{exp}) = 0.04$   
<sup>b</sup>  $u_r(x_1^{exp}) = 0.15$ , <sup>c</sup>  $u_r(x_1^{exp}) = 0.08$ , <sup>d</sup>  $u_r(x_1^{exp}) = 0.09$ .

**Table 6.7** Experimental solid-liquid equilibrium data of prednisolone mixtures at various temperatures and 0.101 MPa<sup>a</sup>.

Solvent	$x^{exp} \times 10^3$				
	T/K	328.2 ± 0.1	323.2 ± 0.1	318.2 ± 0.1	313.2 ± 0.1
Nonan-1-ol		2.07	1.79	1.51	1.26
Pentan-1-ol		1.29	1.1	0.881	0.679
Acetonitrile		4.19	3.64	3.34	3.05
n-Dodecane		0.428	0.365	0.279	0.240
Butan-2-ol		0.845	0.750	0.658	0.530
N,N-Dimethylformamide		14.1	12.5	11.6	10.8
Water <sup>b</sup>		0.021	0.018	0.017	0.015
Octan-1-ol		1.74	1.36	1.12	0.966
n-Hexadecane		0.579	0.541	0.447	0.339
1-Methyl-2-pyrrolidone		11.4	9.81	8.81	7.84
	T/K	308.2 ± 0.1	303.2 ± 0.1	298.2 ± 0.1	293.2 ± 0.1
Nonan-1-ol		1.07	0.833	0.643	0.507
Pentan-1-ol		0.559	0.459	0.338	0.283
Acetonitrile		2.80	2.51	2.36	2.19
n-Dodecane		0.187	0.144	0.114	0.0839
Butan-2-ol		0.414	0.342	0.284	0.248
N,N-Dimethylformamide		10.1	9.63	8.69	8.18
Water <sup>b</sup>		0.013	0.012	0.011	0.010
Octan-1-ol		0.760	0.608	0.409	0.316
n-Hexadecane		0.283	0.216	0.179	0.130
1-Methyl-2-pyrrolidone		6.88	6.10	5.58	4.54

<sup>a</sup> Standard uncertainties  $u$  are  $u(T) = 0.1K$ ,  $u(P) = 0.002 MPa$  and the standard relative uncertainties are  $u_r(x_1^{exp}) = 0.04$ ,  $u_r(x_1^{exp}) = 2$

Table 6.8 Experimental solid-liquid equilibrium data of hydrocortisone mixtures at various temperatures and 0.101 MPa<sup>a</sup>.

Solvent	$x^{exp} \times 10^3$				
	T/K	328.2 ± 0.1	323.2 ± 0.1	318.2 ± 0.1	313.2 ± 0.1
Nonan-1-ol		16.2	12.6	10.1	8.32
Pentan-1-ol		8.26	6.04	3.54	2.36
Acetonitrile		4.16	3.03	2.36	1.64
n-Dodecane <sup>b</sup>		0.042	0.038	0.031	0.026
Butan-2-ol		4.48	2.97	2.2	1.55
N,N Dimethylformamide		14.0	12.8	11.6	10.9
Water		2.10	1.29	0.507 <sup>b</sup>	0.230 <sup>b</sup>
Octan-1-ol		14.3	9.19	6.27	3.78
N-Hexadecane		0.143	0.129	0.116	0.107
1-Methyl-2-pyrrolidone		11.4	11.0	10.7	10.6
	T/K	308.2 ± 0.1	303.2 ± 0.1	298.2 ± 0.1	293.2 ± 0.1
Nonan-1-ol		6.47	4.19	2.86	2.05
Pentan-1-ol		1.59	1.04	0.547	0.409
Acetonitrile		1.27	0.894	0.578	0.410
n-Dodecane <sup>b</sup>		0.023	0.021	0.018	0.016
Butan-2-ol		1.03	0.803	0.565	0.386
N,N Dimethylformamide		10.1	9.57	8.65	8.13
Water		0.133 <sup>b</sup>	0.060 <sup>b</sup>	0.014 <sup>c</sup>	0.004 <sup>c</sup>
Octan-1-ol		2.76	1.82	1.50	1.20
N-Hexadecane		0.104	0.098	0.083	0.076
1-Methyl-2-pyrrolidone		10.2	10.1	10.0	9.77

<sup>a</sup> Standard uncertainties  $u$  are  $u(T) = 0.1K$ ,  $u(P) = 0.002 MPa$  and the standard relative uncertainties are  $u_r(x_1^{exp}) = 0.04$   
<sup>b</sup> $u_r(x_1^{exp}) = 0.2$ , <sup>c</sup> $u_r(x_1^{exp}) = 5.3$ .

By calculating and minimizing the Root Mean Square Deviation (RMSD) between the experimental composition and the calculated composition by the activity coefficient models, the experimental data were fitted to each activity coefficient model. The RMSD and objective function is given by:

$$\text{RMSD} = \left[ \frac{\sum_{i=1}^N (x_i^{\text{exp}} - x_i^{\text{calc}})^2}{N} \right] \quad (6.10)$$

where  $x_i^{\text{exp}}$  and  $x_i^{\text{calc}}$  are the experimental and calculated compositions at the saturated condition and  $N$  is the total number of experimental points measured for the particular system.

The calculated data fitting parameters for all systems measured are presented in Tables 6.9 to 6.14. It was found that the model parameter expressions with polynomial temperature dependence (equations (6.5) and (6.9)) are required in all cases in order to provide an accurate representation of the experimental data.

Similarly to regular solution theory, the excess entropy  $S^E$  is neglected in the NRTL model formulation. Stemming from this, Kontogeorgis and Folas (2010) have shown that the NRTL non-randomness parameter ( $\alpha_{12}$ ) physically represents  $(2/Z)$  where  $Z$  is the coordination number, and therefore negative values of the parameter are unrealistic. Renon and Prausnitz (1968) have presented some guidelines to use when the parameter is fixed during data regression that do include proposed negative values for ( $\alpha_{12}$ ). Physical meaning is compromised for practicality and increased applicability of the model. These standards however are more suitable to vapour-liquid and liquid-liquid mixtures. Regardless they are most often not used and the non-randomness parameter is regressed if superior data fitting is obtained as was the case in this work.

It is apparent that the T-K-Wilson and NRTL model are both able to provide a good representation of the experimental data in most systems, as there is no distinctly superior model. The model with the lowest overall RMSD cannot be easily identified as each model performs better in about half of the cases considered. There is no significant bias (over or under predicted solubility) with regards to the fitting of the experimental data to the NRTL model, as shown in Appendix C Figure C4. Similar results were apparent for the T-K- Wilson model. Any differences in RMSD achieved by the two models for a particular system is generally

within the experimental uncertainty in composition. It must however be mentioned that the regression of the alpha parameter (a third fitting parameter) in the case of the NRTL model of course offers a degree of bias towards the model. In most cases with extremely dilute solubilities (e.g. alkane and water as a solvent), the T-K-Wilson model proved superior to NRTL. This is a generally accepted shortfall and is discussed by Vetere (2000) for instance. The poor correlation of these systems is probably due to the lack of an explicit entropic term in the NRTL model, as entropic effects are significant in the dilute region especially in systems composed of molecules with such large size differences, as those considered here.

In some cases the model parameters determined by regression seem rather large with  $\Delta g_{ij}$  and  $\Delta \lambda_{ij}$  values of orders  $10^5$ - $10^6$ . These ranges are generally not consistent with vapour-liquid and liquid-liquid systems, but are consistent with those determined for similar systems considered in this work, in the literature (Veranda et al. (2006), Domańska et al. (2010)).

In Table 6.15, the calculated infinite dilution activity coefficients ( $\gamma_1^\infty$ ) (IDAC) of each solute in each solvent at 298.2 K is presented. These values were calculated from the fitted activity coefficient model parameters, when the composition of the solute tends to 0. The results show that the activity coefficient at infinite dilution tends to decrease with increasing solvent chain length. This is noticed in the series of butan-2-ol, pentan-1-ol, octan-1-ol, nonan-1-ol for most solutes and n-octane, n-dodecane, n-hexadecane in the case of the estrone systems. These values are however merely extrapolations and do not serve as definitive experimental IDAC data, which should be determined by other direct means i.e. gas-liquid chromatography or gas stripping. Consequently it was not always possible to estimate reasonable IDAC values from the models. In Table 6.16 the solvents used in each of the systems are ranked according to the degree of solubility observed experimentally for convenience.

Table 6.9 Regressed model parameters for the NRTL and T-K-Wilson models with betulin as a solute for use in equations (6.4), (6.5), (6.8) and (6.9).

Solvent	Model														
	NRTL								T-K-Wilson						
	A <sub>12</sub> /J.mol <sup>-1</sup>	A <sub>21</sub> /J.mol <sup>-1</sup>	B <sub>12</sub> /J.K.mol <sup>-1</sup>	B <sub>21</sub> / J.K.mol <sup>-1</sup>	α <sub>12</sub>	<sup>a</sup> Δg <sub>12</sub> /J.mol <sup>-1</sup>	<sup>a</sup> Δg <sub>21</sub> /J.mol <sup>-1</sup>	RMSD <sup>b</sup>	C <sub>12</sub> /J.mol <sup>-1</sup>	C <sub>21</sub> /J.mol <sup>-1</sup>	D <sub>12</sub> /J.K.mol <sup>-1</sup>	D <sub>21</sub> / J.K.mol <sup>-1</sup>	<sup>a</sup> Δλ <sub>12</sub> /J.mol <sup>-1</sup>	<sup>a</sup> Δλ <sub>21</sub> /J.mol <sup>-1</sup>	RMSD <sup>b</sup>
Nonan-1-ol	-58830.88	68648.59	-34422651.04	4565198.92	0.00	174285.02	83960.34	3.60E-07	1205.92	2283.28	-504.95	-1445355.71	1204.23	-2564.47	6.75E-07
Pentan-1-ol	-508450.25	515799.38	-332329171.02	71710832.66	0.00	-654.78	756318.69	1.30E-07	-1646.57	3253.20	1989456.56	-12055516.17	2.03	-37181.20	2.03E-07
Acetonitrile	-1060.76	4143.59	-327120.41	-89647.13	3.60	-0.87	3842.91	7.46E-06	6820.29	2.90	2112513.88	-2935936.18	5.61	-9844.28	1.61E-07
n-Dodecane	0.97	-1.14	-1226.88	7376.36	0.87	0.01	23.60	2.45E-08	-16107.70	-1657.06	82349717.67	2208731.35	104.93	5751.06	1.42E-08
Butan-2-ol	-239.34	287.79	2252836.06	-2877170.46	4.00	2.95	-9362.29	7.91E-07	12957.02	-1639.12	-8907.41	2531.94	5.22	-1630.63	1.80E-08
N,N Dimethylformamide	52523.78	-37396.29	-115330.78	-666188.78	0.03	21.03	-39630.70	2.48E-06	-1805.91	2006.72	8015817.48	-8823318.07	10.12	-27586.83	1.80E-08
Water	0.18	-0.52	4479.58	8250.71	0.76	0.01	27.15	2.56E-08	2008.95	-840.34	452289.03	3343907.75	1.42	10375.18	1.00E-10
Octan-1-ol	2900.68	-325.21	-189233951.15	109413985.90	0.30	-254.88	366651.10	2.00E-08	-1224.63	3375.75	-16005.95	-12741918.81	-0.52	-39360.85	3.04E-08
N-Hexadecane	-206.71	-95.84	-6719258.35	13501624.60	4.00	-9.18	45188.83	1.05E-08	-96329.63	2053.99	427506727.70	-11962684.43	539.59	-38069.05	1.90E-08
1-Methyl-2- Pyrrolidone	76876.71	-60099.25	-11411672.71	7647000.74	4.00	15.57	-34451.09	3.76E-06	0.10	2416.04	2990.73	-10604606.56	0.00	-33151.99	3.03E-06

<sup>a</sup> Reported at a reference of 298.15 K, <sup>b</sup> RMSD =  $\sqrt{\frac{\sum_{i=1}^N (x_i^{exp} - x_i^{calc})^2}{N}}$

Table 6.10 Regressed model parameters for the NRTL and T-K-Wilson models with estrone as a solute for use in equations (6.4), (6.5), (6.8) and (6.9).

Solvent	Model														
	NRTL								T-K-Wilson						
	A <sub>12</sub> /J.mol <sup>-1</sup>	A <sub>21</sub> /J.mol <sup>-1</sup>	B <sub>12</sub> /J.K.mol <sup>-1</sup>	B <sub>21</sub> /J.K.mol <sup>-1</sup>	α <sub>12</sub>	<sup>a</sup> Δg <sub>12</sub> /J.mol <sup>-1</sup>	<sup>a</sup> Δg <sub>21</sub> /J.mol <sup>-1</sup>	RMSD <sup>b</sup>	C <sub>12</sub> /J.mol <sup>-1</sup>	C <sub>21</sub> /J.mol <sup>-1</sup>	D <sub>12</sub> /J.K.mol <sup>-1</sup>	D <sub>21</sub> /J.K.mol <sup>-1</sup>	<sup>a</sup> Δλ <sub>12</sub> /J.mol <sup>-1</sup>	<sup>a</sup> Δλ <sub>21</sub> /J.mol <sup>-1</sup>	RMSD <sup>b</sup>
Nonan-1-ol	26327.16	-19019.04	-380538.30	5448.40	0.05	10.11	-19000.77	2.21E-06	1750.98	-654.59	10386445.65	-7304438.48	36587.29	-25153.80	6.30E-07
Pentan-1-ol	16083.22	-13139.55	-770801.86	481422.82	0.06	5.45	-11524.85	1.16E-07	2401.62	38.79	-8955839.15	-1309816.47	-27636.41	-4354.36	1.24E-07
2-Aminoethanol	35613.38	-29361.11	-98948377.54	26496772.30	0.02	-119.52	59509.50	1.10E-05	97324.15	-2831.31	-49422458.10	2577311.12	-68439.59	5813.03	1.27E-06
Acetonitrile	-18691.04	24693.03	-377176.45	6842147.98	0.07	-8.05	47641.70	9.15E-06	82108.80	-1971.67	-1040793.48	213215.68	78617.96	-1256.54	3.49E-03
n-Dodecane	55081.73	-57113.13	4858089.59	-1244550.41	0.00	28.79	-61287.37	2.00E-10	-1740.50	-366.25	7573735.57	1655519.54	23661.94	5186.39	4.00E-10
Butan-2-ol	-68074.74	69922.45	15904780.17	-62295699.05	0.00	-5.94	-139018.35	1.22E-07	5434.68	-1611.13	-60660.90	-362029.07	5231.23	-2825.38	1.46E-07
N,N-Dimethylformamide	38865.66	-26920.21	-489328.62	-92979.26	0.04	15.02	-27232.06	1.01E-05	10842.78	-1818.85	-1339349.98	-2091015.76	6350.58	-8832.15	1.13E-05
n-Octane	985.04	-2872.95	2411673.80	-898502.88	-3.57	3.66	-5886.54	1.20E-09	-14268.94	4804.96	57899549.11	-17502476.65	179927.10	-53898.63	0.00E+00
Water	2689.65	-2907.49	32256.95	16231.52	-0.13	1.13	-2853.05	4.36E-08	2876.82	-555.19	-639622.20	1592370.00	731.52	4785.65	1.12E-04
Octan-1-ol	173950.64	-169683.34	-6406662.32	2480859.21	0.00	61.51	-161362.50	9.07E-08	1243.29	-3284.86	-20108572.85	28507759.36	-66201.20	92330.63	3.00E-09
Morpholine-4-carbaldehyde	80229.13	-62522.66	594540.12	-770605.99	0.01	33.17	-65107.29	3.55E-03	36575.79	-3673.21	-2011.66	2795.13	36569.04	-3663.83	5.73E-03
N-Hexadecane	-198381.64	204271.43	-62455920.57	-197561293.32	0.00	-164.54	-458352.39	1.98E-06	227.95	138865.53	-16202366.04	12510716.27	-54115.06	180826.68	1.88E-08
1-Methyl-2-Pyrrolidone	72347.12	-64125.93	-1638165.63	-155092.37	0.01	26.97	-64646.11	1.69E-06	89186.57	-4360.83	-10109421.55	3617437.45	55279.40	7772.11	8.16E-07

<sup>a</sup> Reported at a reference of 298.15 K, <sup>b</sup>  $\text{RMSD} = \left[ \frac{\sum_{i=1}^N (x_i^{\text{exp}} - x_i^{\text{calc}})^2}{N} \right]^{1/2}$



**Table 6.11 Regressed model parameters for the NRTL and T-K-Wilson models with diosgenin as a solute for use in equations (6.4), (6.5), (6.8) and (6.9).**

Solvent	Model														
	NRTL								T-K-Wilson						
	$A_{12}/\text{J.mol}^{-1}$	$A_{21}/\text{J.mol}^{-1}$	$B_{12}/\text{J.K.mol}^{-1}$	$B_{21}/\text{J.K.mol}^{-1}$	$\alpha_{12}$	${}^a\Delta g_{12}/\text{J.mol}^{-1}$	${}^a\Delta g_{21}/\text{J.mol}^{-1}$	RMSD <sup>b</sup>	$C_{12}/\text{J.mol}^{-1}$	$C_{21}/\text{J.mol}^{-1}$	$D_{12}/\text{J.K.mol}^{-1}$	$D_{21}/\text{J.K.mol}^{-1}$	${}^a\Delta\lambda_{12}/\text{J.mol}^{-1}$	${}^a\Delta\lambda_{21}/\text{J.mol}^{-1}$	RMSD <sup>b</sup>
Nonan-1-ol	1.39	-1.38	0.03	-0.03	0.00	1.39	-1.38	6.52E-06	-5030.93	312087.90	-4258914.88	3140343.64	19315.40	322620.67	1.91E-06
2-Aminoethanol	0.21	-0.28	0.00	-0.01	4.00	0.21	-0.28	3.99E-06	40553.68	-1259.91	-231329.03	-279826.71	-39777.80	2198.45	1.05E-05
Acetonitrile	52258.86	11943.59	-523706.90	-944987.26	2.56	50502.34	8774.09	2.26E-05	-	-	-	-	-	-	-
n-Dodecane	90721.08	-85315.40	-7823.86	5377.94	0.00	90694.84	-85297.36	3.22E-09	8800.99	-3435.05	-120443.61	-64589.34	-8397.02	3651.68	4.00E-09
Butan-2-ol	-62.77	64.14	-4.30	-44.49	1.06	-62.79	63.99	8.40E-07	14694.56	-2278.01	30955.15	-39095.43	-14798.38	2409.14	6.19E-08
N,N-Dimethylformamide	0.61	-0.65	-0.01	0.01	4.00	0.61	-0.65	2.94E-06	117032.82	-2014.30	-6914770.34	-139621.32	-93840.57	2482.59	5.95E-06
n-Octane	-984257.41	988255.51	-24181.48	-21809.19	0.00	-984338.52	988182.36	1.36E-08	-10303.26	9886.81	91597.92	-171333.19	9996.04	-9312.16	3.20E-09
Water	367285.60	-362200.13	-12751.63	12094.05	0.00	367242.83	-362159.57	9.60E-08	12264.37	-154.52	53159.80	-116649.44	-12442.67	545.76	7.45E-08
Octan-1-ol	303829.78	-292263.01	6356.02	-10920.14	0.00	303851.10	-292299.64	4.38E-06	-7686.88	616304.89	-3252602.26	-3385288.69	18596.16	604950.57	2.34E-06
Morpholine-4-carbaldehyde	508423.80	-495628.73	-872.75	-1264.52	0.00	508420.87	-495632.97	1.56E-05	111114.98	-2737.20	-1309867.07	-143772.38	106721.66	3219.41	2.29E-05
N-Hexadecane	1.10	-1.10	-0.02	0.00	4.00	1.10	-1.10	5.85E-07	9356.32	-4665.35	-20514.50	17713.25	-9287.51	4605.94	2.07E-07
1-Methyl-2-Pyrrolidone	0.47	-0.61	0.00	0.00	0.18	0.47	-0.61	7.87E-06	111259.95	-2526.30	-161668.80	-200176.45	110717.71	3197.70	1.53E-05

<sup>a</sup> Reported at a reference of 298.15 K, <sup>b</sup>  $\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (x_i^{\text{exp}} - x_i^{\text{calc}})^2}{N}}$

Table 6.12 Regressed model parameters for the NRTL and T-K-Wilson models with estriol as a solute for use in equations (6.4), (6.5), (6.8) and (6.9).

Solvent	Model														
	NRTL								T-K-Wilson						
	$A_{12}/J.mol^{-1}$	$A_{21}/J.mol^{-1}$	$B_{12}/J.K.mol^{-1}$	$B_{21}/J.K.mol^{-1}$	$\alpha_{12}$	${}^a\Delta g_{12}/J.mol^{-1}$	${}^a\Delta g_{21}/J.mol^{-1}$	RMSD <sup>b</sup>	$C_{12}/J.mol^{-1}$	$C_{21}/J.mol^{-1}$	$D_{12}/J.K.mol^{-1}$	$D_{21}/J.K.mol^{-1}$	${}^a\Delta\lambda_{12}/J.mol^{-1}$	${}^a\Delta\lambda_{21}/J.mol^{-1}$	RMSD <sup>b</sup>
Nonan-1-ol	0.02	-0.76	0.00	0.00	0.00	0.02	-0.76	5.22E-05	112983.88	-33022.00	-1149.34	481.62	-	33020.38	2.21E-05
Pentan-1-ol	-23056.30	29013.65	-9239.51	-5621.78	0.04	-23087.29	28994.79	5.22E-01	11097.48	-3234.63	2599.04	-19609.00	-11106.20	3300.40	2.42E-06
2-Aminoethanol	-58630.21	68558.77	480.87	-144.19	-0.01	-58628.59	68558.29	2.93E-05	91882.48	-2137.21	-149657.03	-139625.79	-91380.53	2605.52	2.89E-05
n-Dodecane	13221.78	-10678.86	-312.39	479.09	-0.07	13220.73	-10677.26	7.35E-09	4691.09	-3098.31	-1477.01	-4226.39	-4686.14	3112.48	1.92E-08
Butan-2-ol	455143.68	448949.42	1964.64	-15817.79	0.00	455150.27	-449002.47	4.03E-07	16277.16	-3927.89	-453234.88	340888.67	0.00	2784.54	1.50E-09
N,N-Dimethylformamide	143673.48	150961.62	1593.56	-3480.73	0.00	-143668.13	150949.95	6.91E-06	-13804.96	156621.18	115443.43	344739.13	13417.76	157777.4	5.22E-06
n-Octane	469938.13	470955.07	9503.89	-79071.91	0.00	-469906.25	470689.87	7.75E-08	3706.38	-2061.57	848.36	-2369.96	-3709.23	2069.52	9.19E-08
Water	339071.26	335961.59	10472.22	23015.19	0.00	-339036.14	336038.79	9.00E-10	4415.27	-632.47	-169211.45	197032.48	-3847.74	-28.38	7.85E-08
Octan-1-ol	-0.07	-0.53	0.00	-0.01	4.00	-0.07	-0.53	4.27E-05	17255.90	-6253.31	-281978.19	581661.39	-16310.14	4302.41	3.59E-08
Morpholine-4-carbaldehyde	196005.29	201614.38	-1110.85	961.74	0.00	-196009.01	201617.61	3.15E-06	94061.55	-5080.39	-2777660.75	654653.25	-84745.23	2884.68	3.06E-06
N-Hexadecane	280032.07	276223.27	2335.62	-4036.14	0.00	280039.91	-276236.81	1.84E-07	-1981.65	332579.11	-905708.52	1818092.87	5019.42	338677.0	2.18E-07
1-Methyl-2-Pyrrolidone	-1.06	0.90	0.00	-0.01	-4.00	-1.06	0.90	1.10E-05	-12499.03	161062.49	873118.56	405354.44	9570.58	162422.0	8.28E-07

<sup>a</sup> Reported at a reference of 298.15 K, <sup>b</sup>  $RMSD = \sqrt{\frac{\sum_{i=1}^N (x_i^{exp} - x_i^{calc})^2}{N}}$

**Table 6.13 Regressed model parameters for the NRTL and T-K-Wilson models with prednisolone as a solute for use in equations (6.4), (6.5), (6.8) and (6.9).**

Solvent	Model															
	NRTL								T-K-Wilson							
	$A_{12}/\text{J.mol}^{-1}$	$A_{21}/\text{J.mol}^{-1}$	$B_{12}/\text{J.K.mol}^{-1}$	$B_{21}/\text{J.K.mol}^{-1}$	$\alpha_{12}$	${}^a\Delta g_{12}/\text{J.mol}^{-1}$	${}^a\Delta g_{21}/\text{J.mol}^{-1}$	RMSD <sup>b</sup>	$C_{12}/\text{J.mol}^{-1}$	$C_{21}/\text{J.mol}^{-1}$	$D_{12}/\text{J.K.mol}^{-1}$	$D_{21}/\text{J.K.mol}^{-1}$	${}^a\Delta\lambda_{12}/\text{J.mol}^{-1}$	${}^a\Delta\lambda_{21}/\text{J.mol}^{-1}$	RMSD <sup>b</sup>	
Nonan-1-ol	33025.48	-18571.49	-81855.98	43825.37	0.08	32750.94	-18424.50	6.88E-12	801.80	1384.24	857554.21	-1370371.38	-4.64	-3678.05	3.00E-15	
Pentan-1-ol	4960.73	-2334.59	-2499.14	-7631.12	0.60	4952.35	-2360.18	5.73E-10	-1075.20	53.42	-2605117.02	1847363.45	-0.18	9812.81	1.80E-14	
Acetonitrile	1668.11	-768.52	1194100.68	-1015835.12	0.04	5673.14	-4175.65	5.90E-12	3915.29	684.63	819394.72	-780694.21	-2.30	-6663.55	6.66E-12	
n-Dodecane	46.21	-1.34	20769.21	-17094.37	3.46	115.87	-58.67	4.18E-09	11248.91	4971.77	5657019.49	-2488055.22	-16.68	-7724.82	0.00E+00	
Butan-2-ol	12222.86	-2651.10	61615.51	-37604.72	0.45	12429.52	-2777.23	1.74E-12	1740.63	-992.90	2489453.85	-346562.49	3.33	-10090.30	2.70E-14	
N,N-Dimethylformamide	-797.93	-1065.99	-43357.75	337372.03	2.45	-943.35	65.56	4.03E-12	6700.97	732068.38	-8376248.29	372751636.00	2455.37	21393.10	0.00E+00	
Water	1875.01	-534.55	1276398.09	944060.55	-0.05	-2406.05	2631.85	9.20E-14	455.66	-1538.58	-1261741.46	850491.26	5.16	3776.24	6.60E-14	
Octan-1-ol	3767.84	-1061.98	336091.86	-277969.77	0.23	4895.10	-1994.30	3.27E-13	-799.55	2864.72	595725.28	-1555737.41	-9.61	-1198.53	0.00E+00	
N-Hexadecane	14.18	-108.21	72167.13	-47223.19	1.08	256.23	-266.59	1.66E-07	905.92	488.93	5808.66	-690284.03	-1.64	-925.40	1.17E-08	
1-Methyl-2-Pyrrolidone	-4041.08	5875.30	5089104.32	5187090.86	0.00	21110.02	23272.89	7.25E-12	24581.42	-2782.91	11279380.84	1043485.53	9.33	13249.80	2.00E-15	

<sup>a</sup> Reported at a reference of 298.15 K, <sup>b</sup>  $\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (x_i^{\text{exp}} - x_i^{\text{calc}})^2}{N}}$

Table 6.14 Regressed model parameters for the NRTL and T-K-Wilson models with hydrocortisone as a solute for use in equations (6.4), (6.5), (6.8) and (6.9).

Solvent	Model														
	NRTL								T-K-Wilson						
	$A_{12}/\text{J.mol}^{-1}$	$A_{21}/\text{J.mol}^{-1}$	$B_{12}/\text{J.K.mol}^{-1}$	$B_{21}/\text{J.K.mol}^{-1}$	$\alpha_{12}$	${}^a\Delta g_{12}/\text{J.mol}^{-1}$	${}^a\Delta g_{21}/\text{J.mol}^{-1}$	RMSD <sup>b</sup>	$C_{12}/\text{J.mol}^{-1}$	$C_{21}/\text{J.mol}^{-1}$	$D_{12}/\text{J.K.mol}^{-1}$	$D_{21}/\text{J.K.mol}^{-1}$	${}^a\Delta\lambda_{12}/\text{J.mol}^{-1}$	${}^a\Delta\lambda_{21}/\text{J.mol}^{-1}$	RMSD <sup>b</sup>
Nonan-1-ol	-4.49	4.23	30.52	-51.42	-4.00	-4.38	4.06	4.84E-05	-10285.27	437337.68	1850557.24	281063262.33	4078.48	-505353.12	8.80E-06
Pentan-1-ol	1920.81	-4819.66	-2473.67	4096.21	-1.76	1912.51	-4805.92	9.83E-07	-1119.66	-1722.77	1427703.13	520803.73	-3668.88	-24.01	8.27E-07
Acetonitrile	119.79	-143.52	408917.58	-512777.68	-0.13	1491.31	-1863.39	1.04E-06	986.56	-1960.56	829548.89	728038.40	-3768.88	-481.29	1.95E-08
n-Dodecane	-24.72	-34.67	-580.67	4313.59	-4.00	-26.67	-20.20	4.66E-04	-66660.90	7396.27	29867319.36	4951352.60	-33514.58	-24003.19	0.00E+00
Butan-2-ol	-1.39	2.13	1168115.72	1011033.56	-0.03	-3919.27	3393.15	1.50E-08	-18464.20	4297.49	5765378.96	-546388.52	-872.97	-2464.90	1.24E-06
N,N-Dimethylformamide	-81.38	68.68	32.75	359.59	-3.92	-81.27	69.89	1.03E-05	-21239.55	890.77	11520781.12	-736592.94	-17401.34	1579.78	0.00E+00
Water	-417.85	-1207.70	67279.27	-23576.24	-4.00	-192.19	-1286.78	4.60E-08	1125.71	-1315.78	407542.14	581656.16	-2492.62	-635.11	4.64E-08
Octan-1-ol	1563.13	537.05	-221517.66	225796.27	-0.15	-2306.10	1294.38	3.77E-06	-14695.95	160801.80	3681034.86	-9280997.92	2349.70	-129673.18	2.48E-06
N-Hexadecane	-0.01	-0.01	11.50	501.46	4.00	0.03	1.67	7.09E-04	-59548.56	6354.69	27147918.20	2314837.81	-31506.00	-14118.69	1.90E-07
1-Methyl-2-Pyrrolidone	5375.43	133.05	32677.83	5553.60	-2.57	-5265.82	151.68	7.38E-06	109122.34	2116.33	53546617.28	-1224549.36	-70473.90	1990.83	0.00E+00

<sup>a</sup> Reported at a reference of 298.15 K, <sup>b</sup>  $\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (x_i^{\text{exp}} - x_i^{\text{calc}})^2}{N}}$

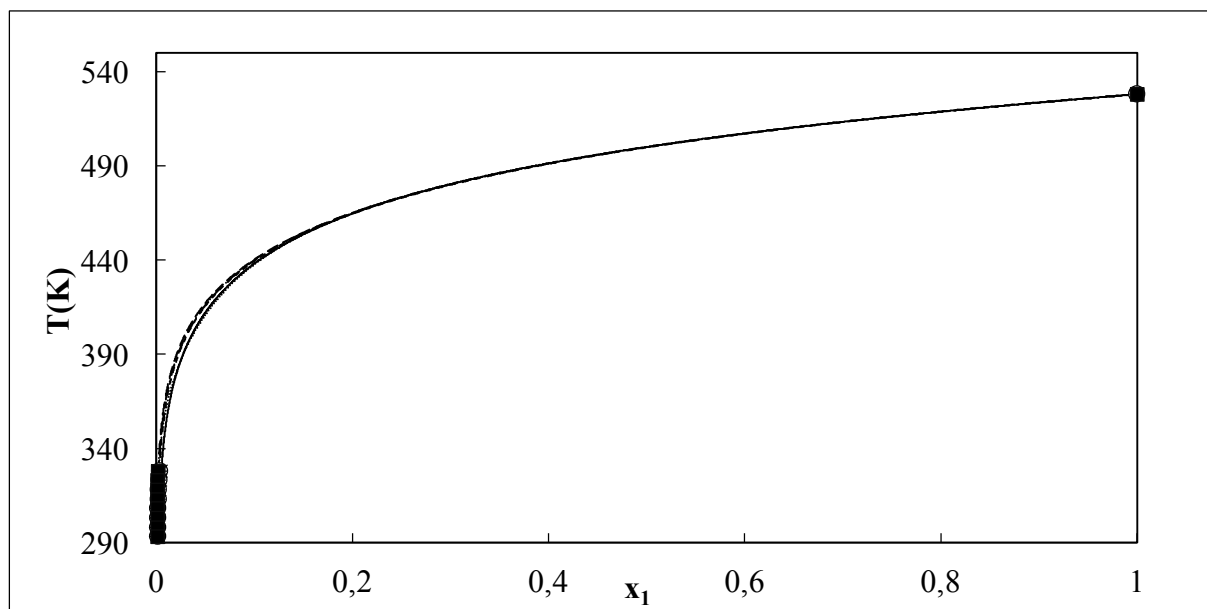


Figure 6.3 T-x plot for the systems betulin + (experimental, correlation, model), (+,  $\dots$ , T-K-Wilson), acetonitrile/ (x, —, T-K-Wilson), nonan-1-ol/ ( $\circ$ , - - -, NRTL), octan-1-ol/ ( $\diamond$ , -  $\cdot$  -, T-K-Wilson), pentan-1-ol/ ( $\blacksquare$ , - - -, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

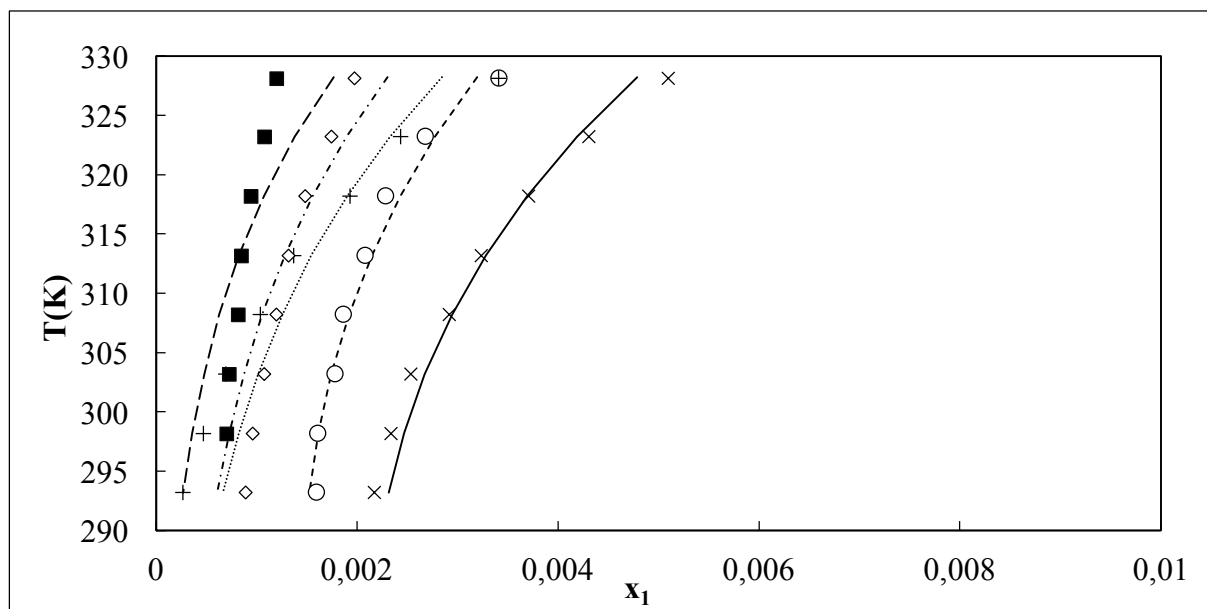


Figure 6.4 Dilute range T-x plot for the systems betulin + (experimental, correlation, model), (+,  $\dots$ , T-K-Wilson), acetonitrile/ (x, —, T-K-Wilson), nonan-1-ol/ ( $\circ$ , - - -, NRTL), octan-1-ol/ ( $\diamond$ , -  $\cdot$  -, T-K-Wilson), pentan-1-ol/ ( $\blacksquare$ , - - -, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

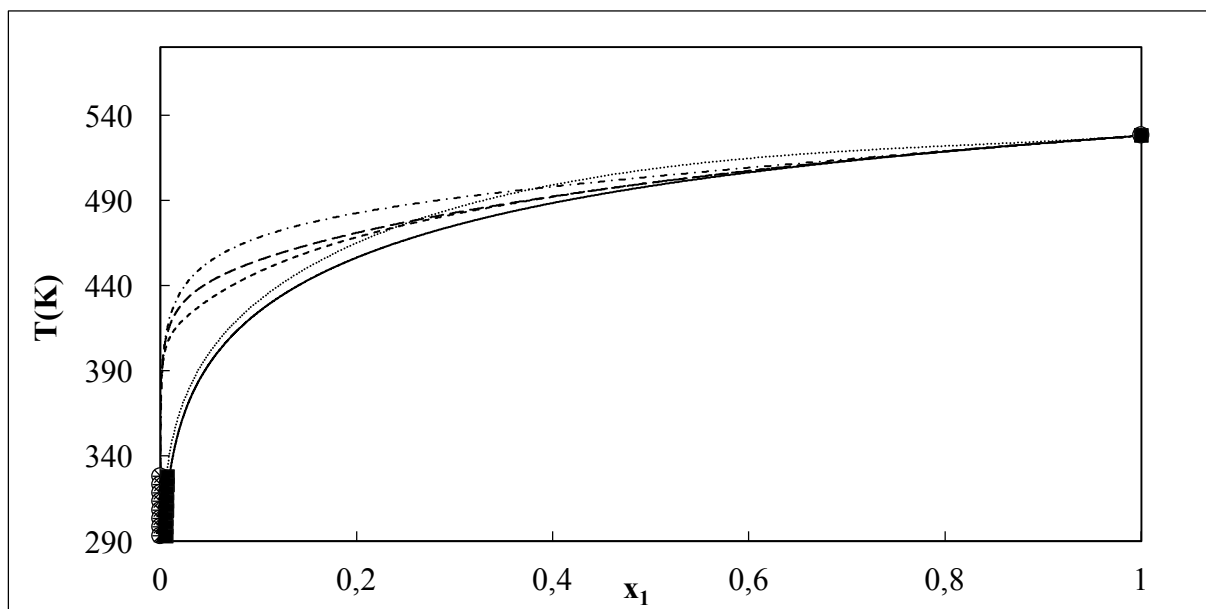


Figure 6.5 T-x plot for the systems betulin + (experimental, correlation, model), ( $\blacktriangle$ ,  $\cdots$ , NRTL), N,N dimethylformamide/ ( $\blacksquare$ ,  $—$ , NRTL), 1-methyl-2-pyrrolidone/ ( $\times$ ,  $-\cdot-$ , NRTL), n-hexadecane/ ( $\circ$ ,  $---$ , T-K-Wilson), n-dodecane/ ( $+$ ,  $---$ , T-K-Wilson), water. Experimental data are represented as symbols, and model predictions are represented as lines.

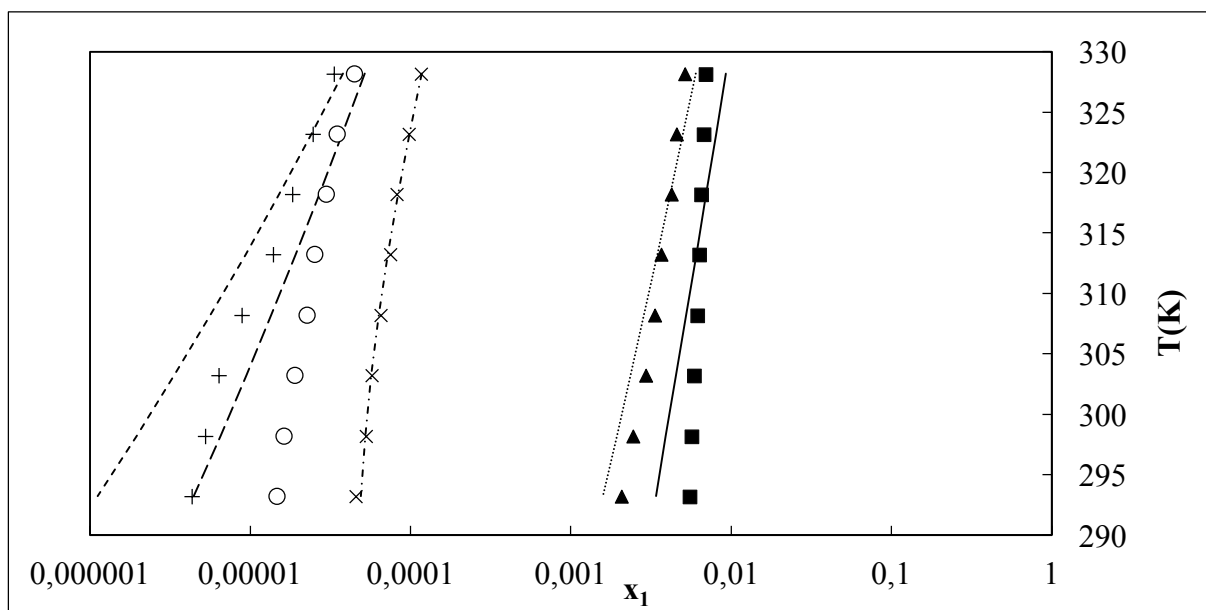


Figure 6.6 Dilute region T-x plot for the systems betulin + (experimental, correlation, model), ( $\blacktriangle$ ,  $\cdots$ , NRTL), N,N dimethylformamide/ ( $\blacksquare$ ,  $—$ , NRTL), 1-methyl-2-pyrrolidone/ ( $\times$ ,  $-\cdot-$ , NRTL), N-Hexadecane/ ( $\circ$ ,  $---$ , T-K-Wilson), n-dodecane/ ( $+$ ,  $---$ , T-K-Wilson), water. Experimental data are represented as symbols, and model predictions are represented as lines.

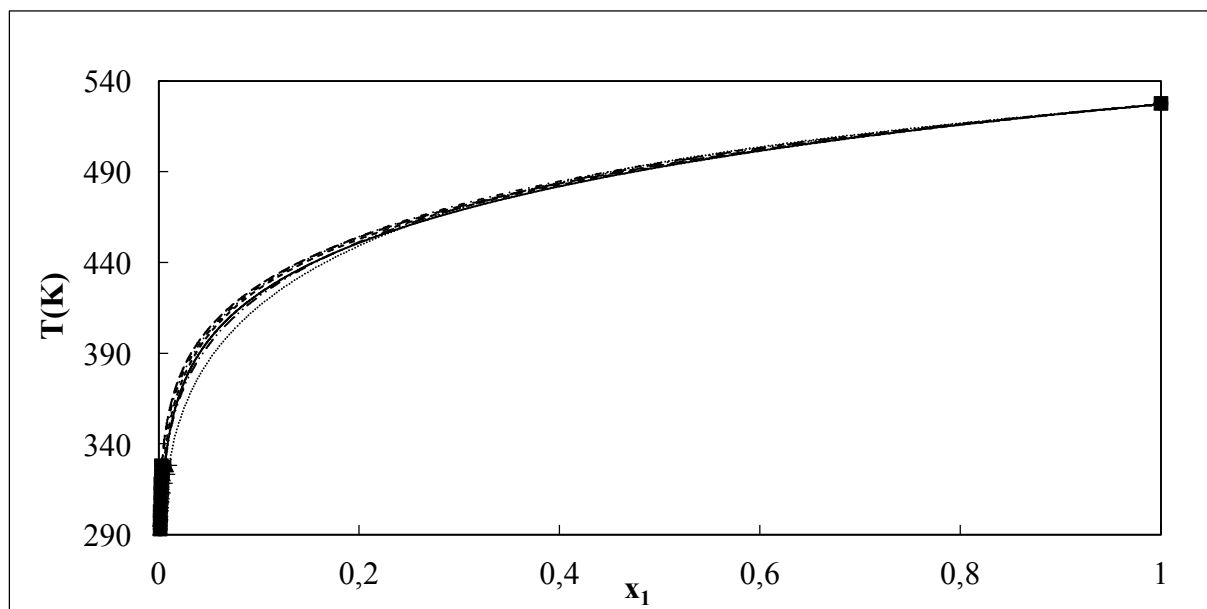


Figure 6.7 T-x plot for the systems estrone + (experimental, correlation, model), (+,  $\cdots$ , T-K-Wilson), 1-methyl-2-pyrrolidone/ ( $\blacktriangle$ ,  $-\cdots-$ , T-K-Wilson), 2-aminoethanol/ (x,  $-$ , T-K-Wilson), nonan-1-ol/ ( $\circ$ ,  $-\cdots-$ , T-K-Wilson), octan-1-ol/ ( $\diamond$ ,  $-\cdot-$ , NRTL), pentan-1-ol/ ( $\blacksquare$ ,  $-\cdots-$ , NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

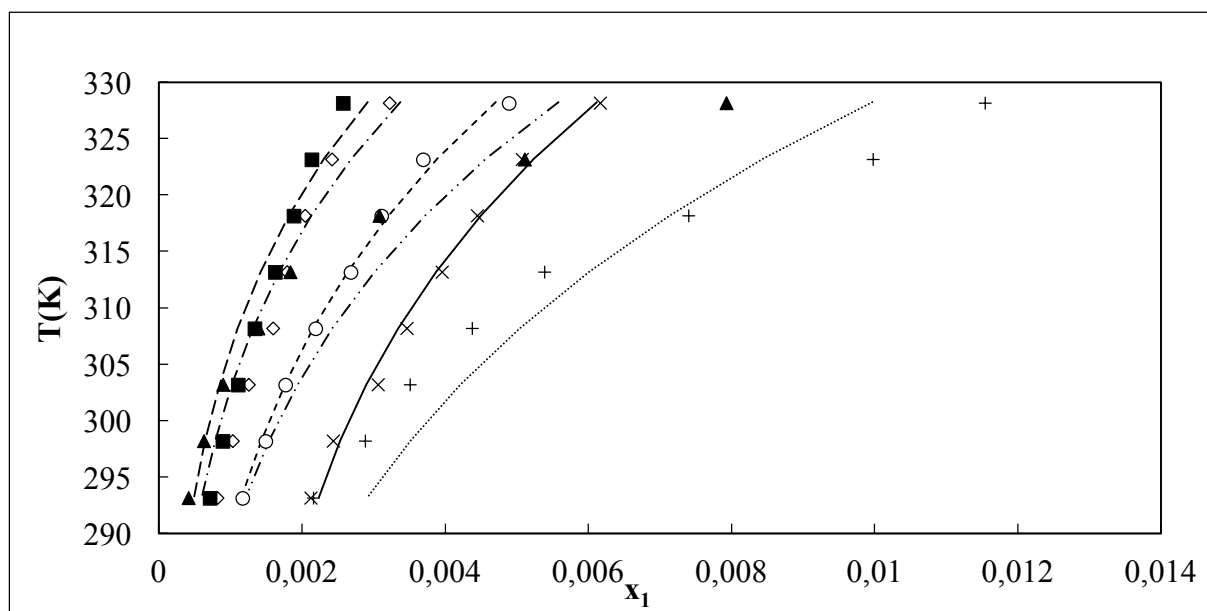


Figure 6.8 Dilute region T-x plot for the systems estrone + (experimental, correlation, model), (+,  $\cdots$ , T-K-Wilson), 1-methyl-2-pyrrolidone/ ( $\blacktriangle$ ,  $-\cdots-$ , T-K-Wilson), 2-aminoethanol/ (x,  $-$ , T-K-Wilson), nonan-1-ol/ ( $\circ$ ,  $-\cdots-$ , T-K-Wilson), octan-1-ol/ ( $\diamond$ ,  $-\cdot-$ , NRTL), pentan-1-ol/ ( $\blacksquare$ ,  $-\cdots-$ , NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

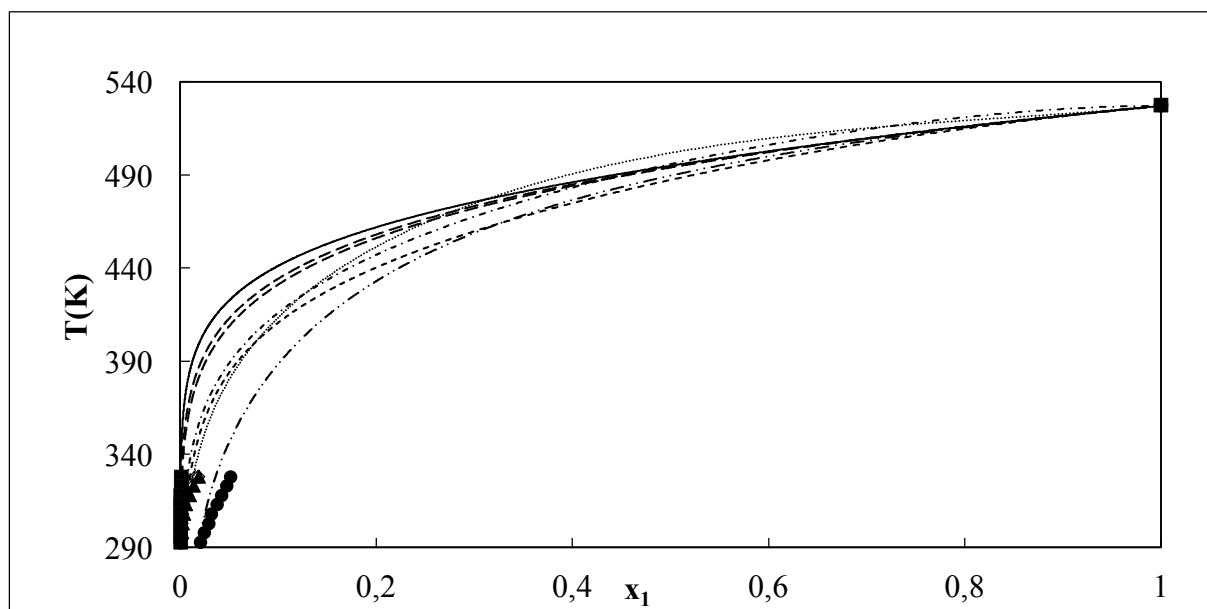


Figure 6.9 T-x plot for the systems estrone + (experimental, correlation, model), ( $\blacktriangle$ ,  $\cdots$ , NRTL), N,N dimethylformamide/ ( $\blacksquare$ , —, T-K-Wilson), n-octane/ ( $\times$ , - · -, T-K-Wilson), n-hexadecane/ ( $\circ$ , — — —, NRTL), n-dodecane/ ( $+$ , - - -, NRTL), water/ ( $\bullet$ , - · · -, NRTL), morpholine-4-carbaldehyde/ ( $\diamond$ , - - -, NRTL), acetonitrile. Experimental data are represented as symbols, and model predictions are represented as lines.

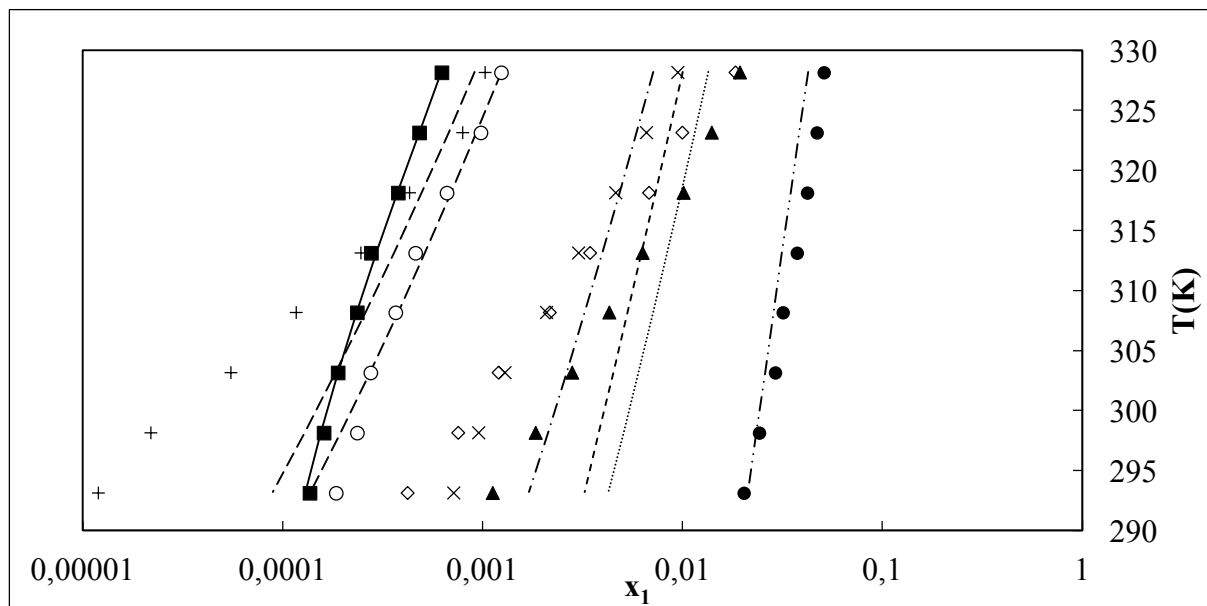
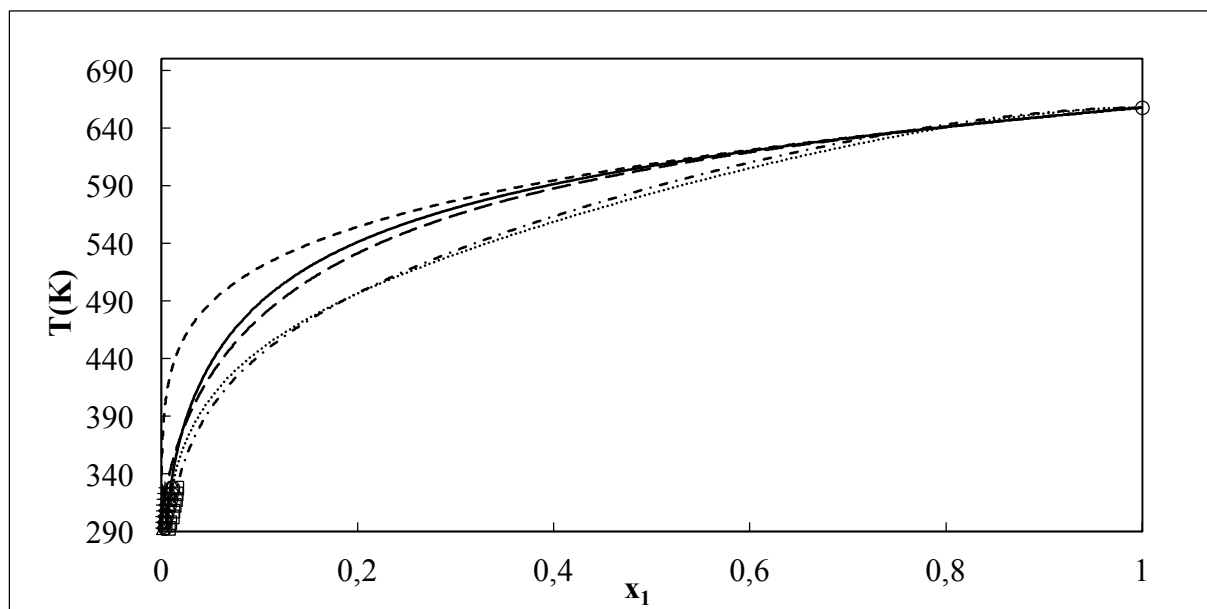
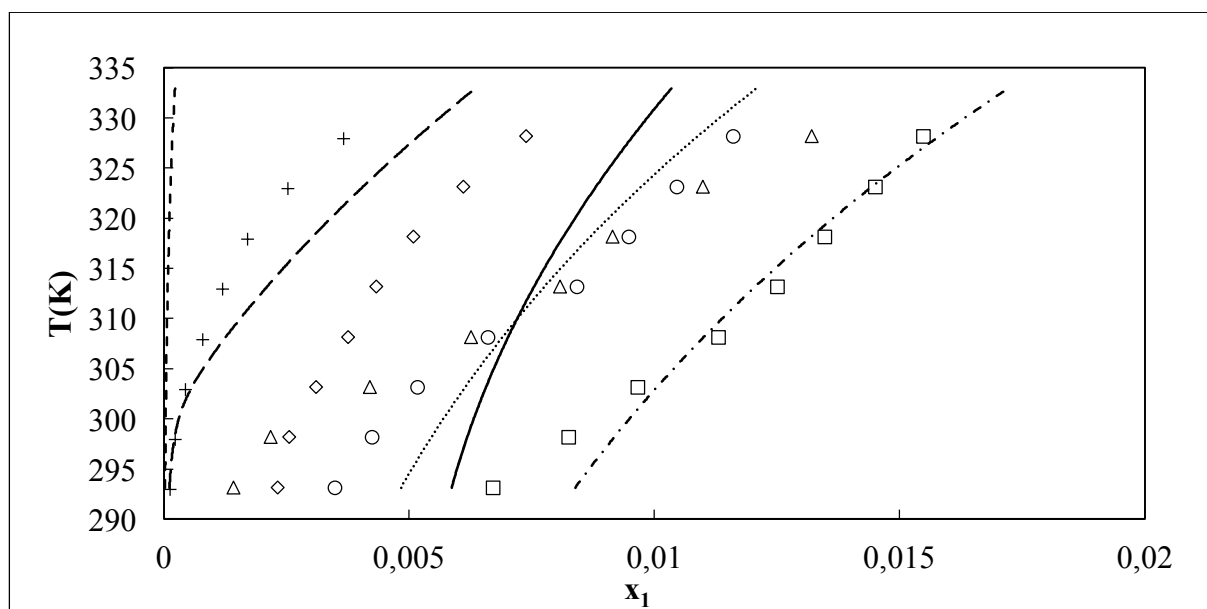


Figure 6.10 Dilute region T-x plot for the systems estrone + (experimental, correlation, model), ( $\blacktriangle$ ,  $\cdots$ , NRTL), N,N dimethylformamide/ ( $\blacksquare$ , —, T-K-Wilson), n-octane/ ( $\times$ , - · -, T-K-Wilson), N-Hexadecane/ ( $\circ$ , — — —, NRTL), n-dodecane/ ( $+$ , - - -, NRTL), water/ ( $\bullet$ , - · · -, NRTL), morpholine-4-carbaldehyde/ ( $\diamond$ , - - -, NRTL), acetonitrile. Experimental data are represented as symbols, and model predictions are represented as lines.





**Figure 6.11** T-x plot for the systems diosgenin (1) + (experimental, NRTL-model), / (○, ---), octan-1-ol (2) / (△, -), 2-aminoethanol (2) / (◇, - - -), pentan-1-ol (2) / (+, - - -), butan-2-ol (2) / (□, - · -), nonan-1-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 6.12** Dilute region T-x plot for the systems diosgenin (1) + (experimental, NRTL-model), / (○, ---), octan-1-ol (2) / (△, -), 2-aminoethanol (2) / (◇, - - -), pentan-1-ol (2) / (+, - - -), butan-2-ol (2) / (□, - · -), nonan-1-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.

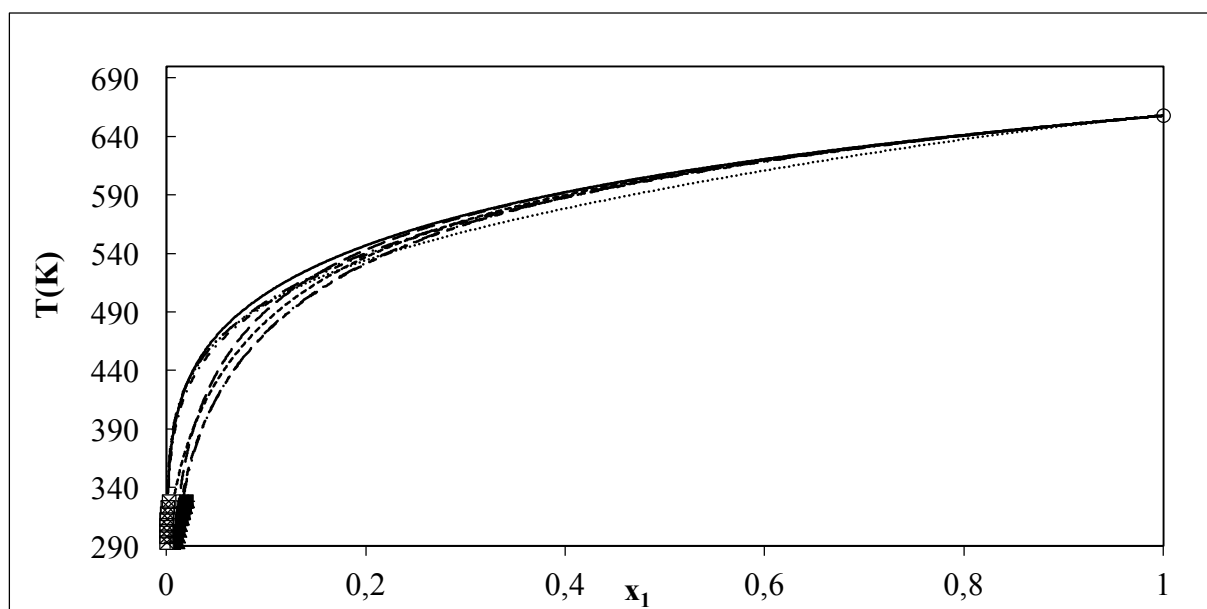


Figure 6.13 T-x plot for the systems diosgenin (1) + (+, - - -), acetonitrile (2) / (■, - - -) morpholine-4-carbaldehyde (2) / (Δ, - · -), n-dodecane (2) / (□, -), water / (○, ...), dimethylformamide (2), / (▲, - · -), 1-methyl-2-pyrrolidone (2) / (×, - - -), N-Hexadecane (2) / (◇, ···), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines.

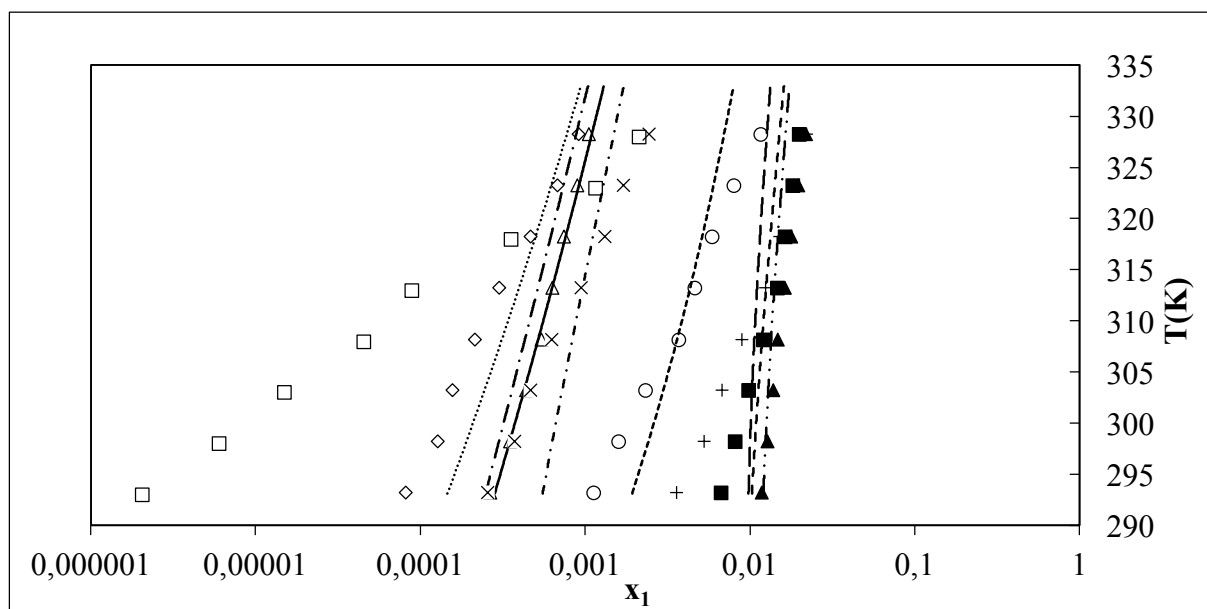


Figure 6.14 T-x plot for the systems diosgenin (1) + (+, - - -), acetonitrile (2) / (■, - - -) morpholine-4-carbaldehyde (2) / (Δ, - · -), n-dodecane (2) / (□, -), water / (○, ...), dimethylformamide (2), / (▲, - · -), 1-methyl-2-pyrrolidone (2) / (×, - - -), N-Hexadecane (2) / (◇, ···), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines.

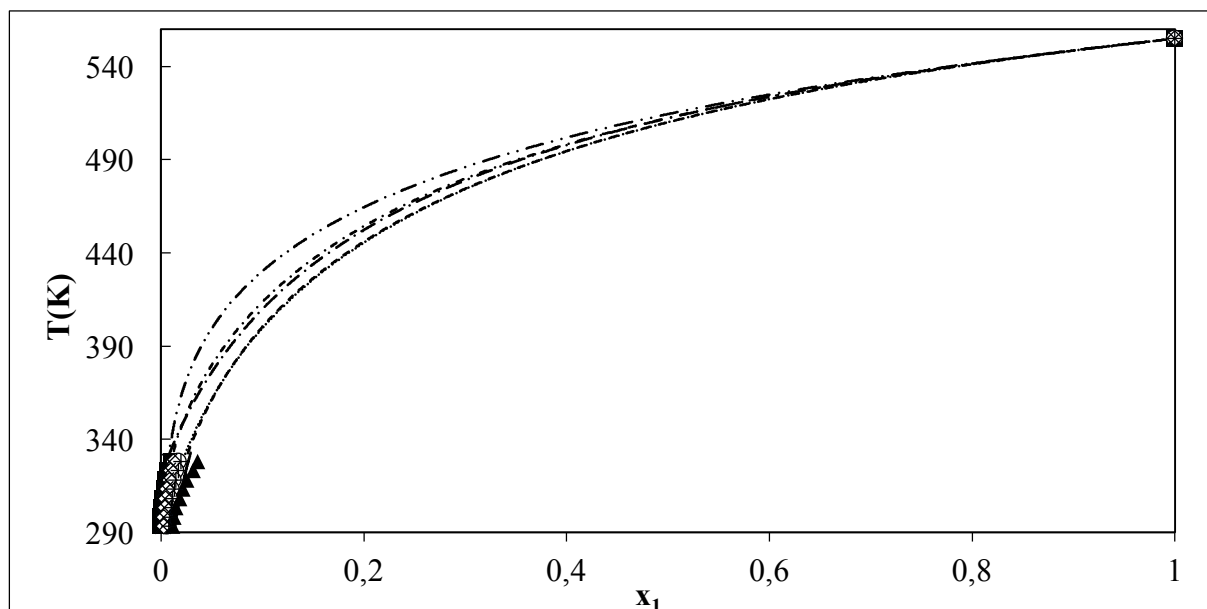


Figure 6.15 T-x plot for the systems estriol (1) + (experimental, NRTL-model), ( $\blacksquare$ , - · -) morpholine-4-carbaldehyde (2) / ( $\circ$ , ···), octan-1-ol (2) / ( $\blacktriangle$ , -), 2-aminoethanol (2) / ( $\diamond$ , ···), pentan-1-ol (2), (+, - - -), nonan-1-ol (2) / (x, - · -), butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.

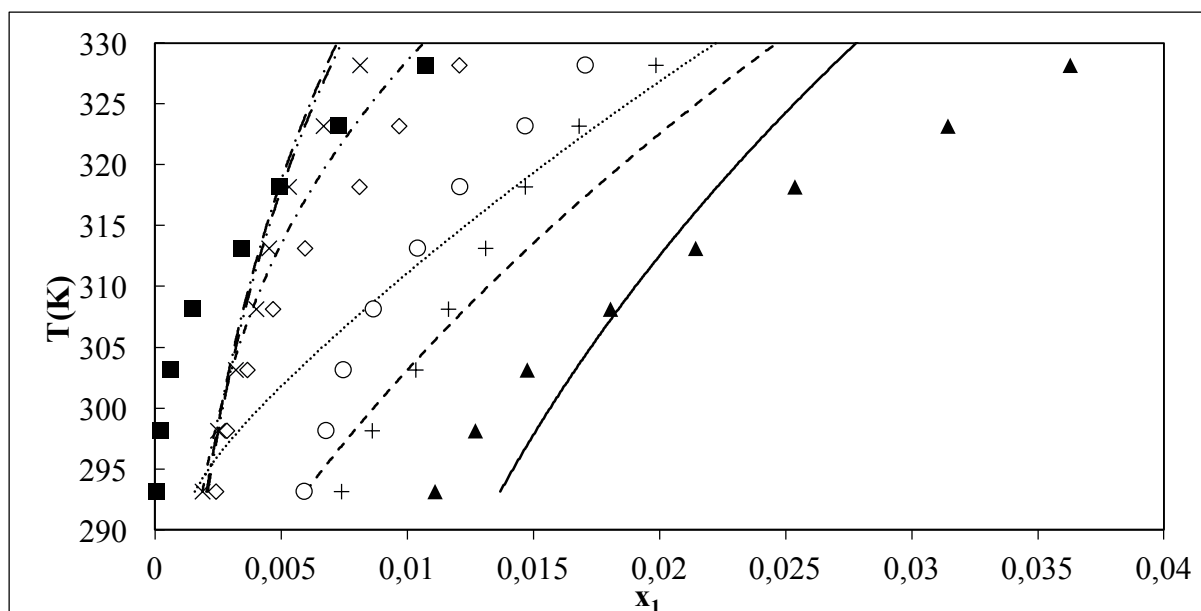
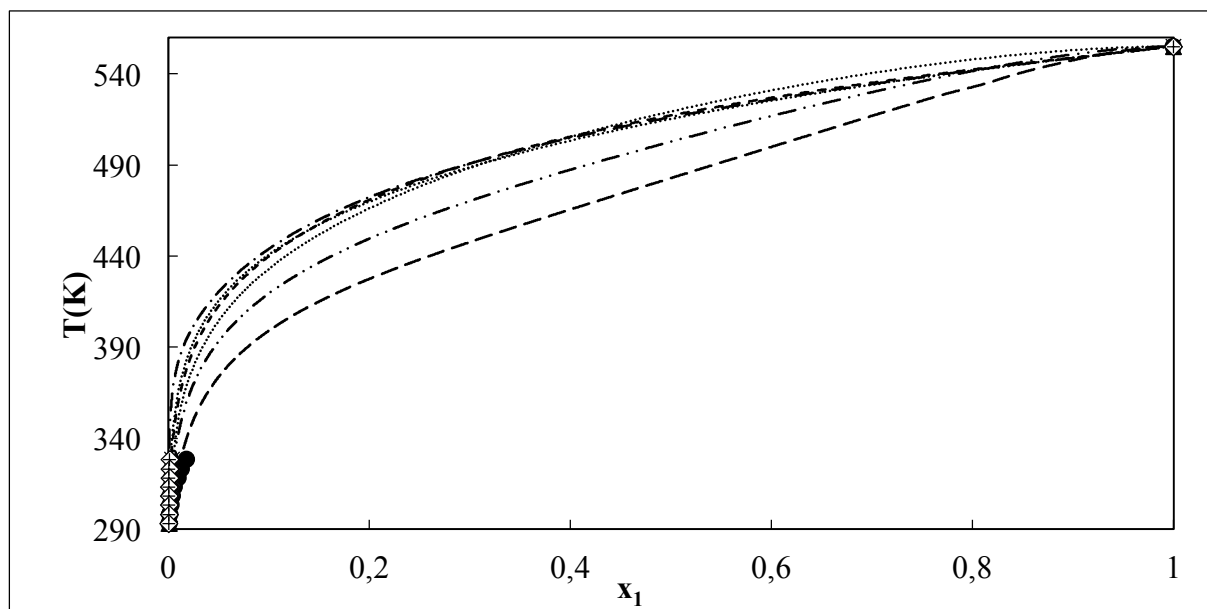
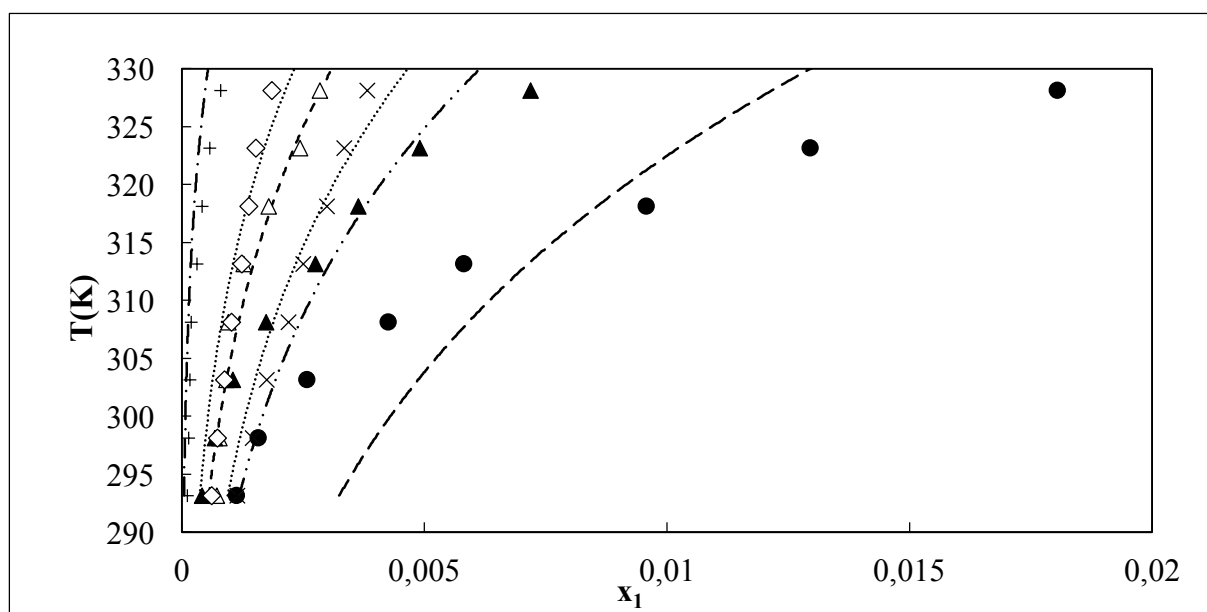


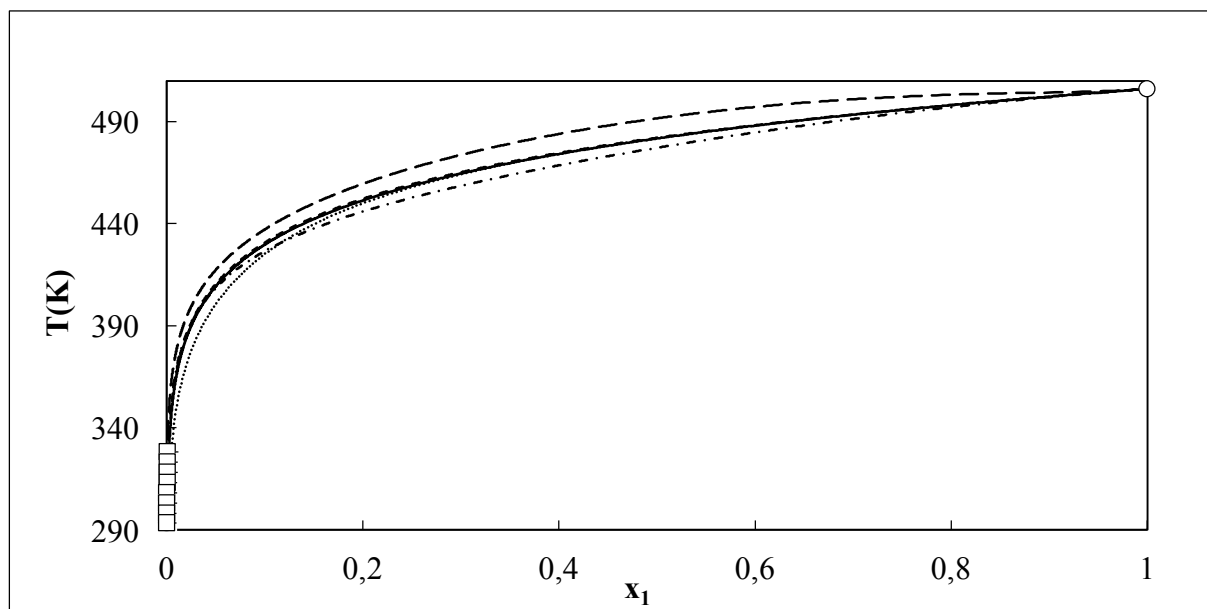
Figure 6.16 Dilute region T-x plot for the systems estriol (1) + (experimental, NRTL-model), ( $\blacksquare$ , - · -) morpholine-4-carbaldehyde (2) / ( $\circ$ , ···), octan-1-ol (2) / ( $\blacktriangle$ , -), 2-aminoethanol (2) / ( $\diamond$ , ···), pentan-1-ol (2), (+, - - -), nonan-1-ol (2) / (x, - · -), butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.



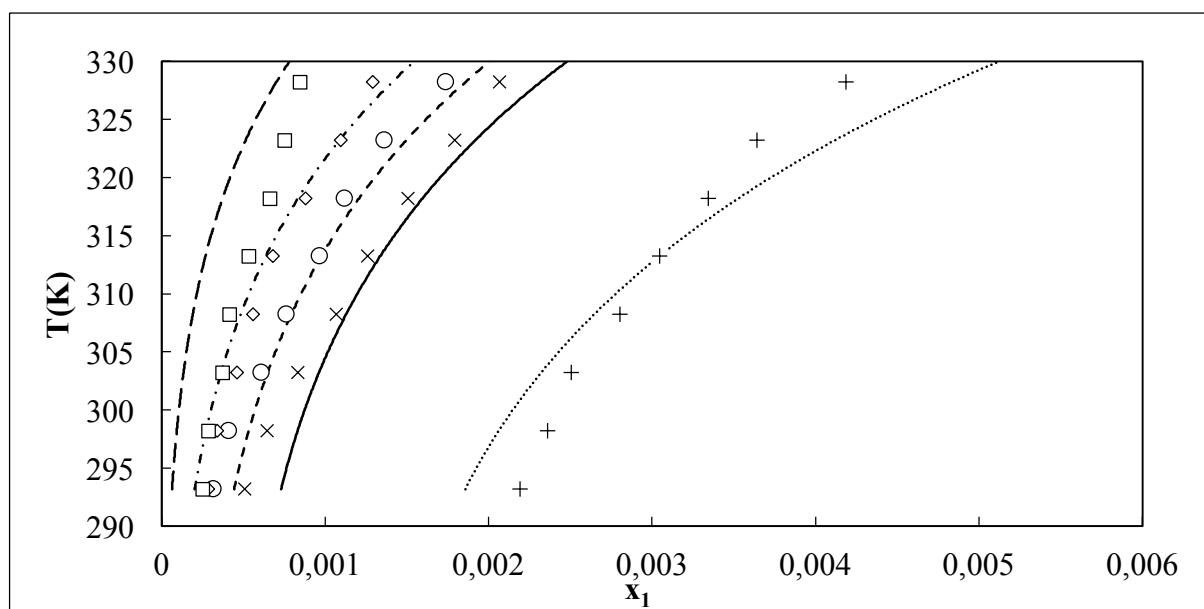
**Figure 6.17** T-x plot for the systems estriol (1) + (experimental, NRTL-model), ( $\Delta$ , - - -) n-dodecane (2) / ( $\bullet$ , - - -), dimethylformamide (2) / (+, - · -), water / ( $\blacktriangle$ , - · -), 1-methyl-2-pyrrolidone (2) / (x, ---), n-hexadecane (2) / ( $\diamond$ , ---), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 6.18** Dilute region T-x plot for the systems estriol (1) + (experimental, NRTL-model), ( $\Delta$ , - - -) n-dodecane (2) / ( $\bullet$ , - - -), dimethylformamide (2) / (+, - · -), water / ( $\blacktriangle$ , - · -), 1-methyl-2-pyrrolidone (2) / (x, ---), n-hexadecane (2) / ( $\diamond$ , ---), n-octane (2). Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 6.19** T-x plot for the systems prednisolone + (experimental, model, correlation), (+, ---, T-K-Wilson), acetone nitrile/ (x, -, NRTL), nonan-1-ol/ (o, ---, NRTL), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/ (□, ---, NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 6.20** Dilute region T-x plot for the systems prednisolone + (experimental, model, correlation), (+, ---, T-K-Wilson), acetone nitrile/ (x, -, NRTL), nonan-1-ol/ (o, ---, NRTL), octan-1-ol/ (◇, - · -, NRTL), pentan-1-ol/ (□, ---, NRTL), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

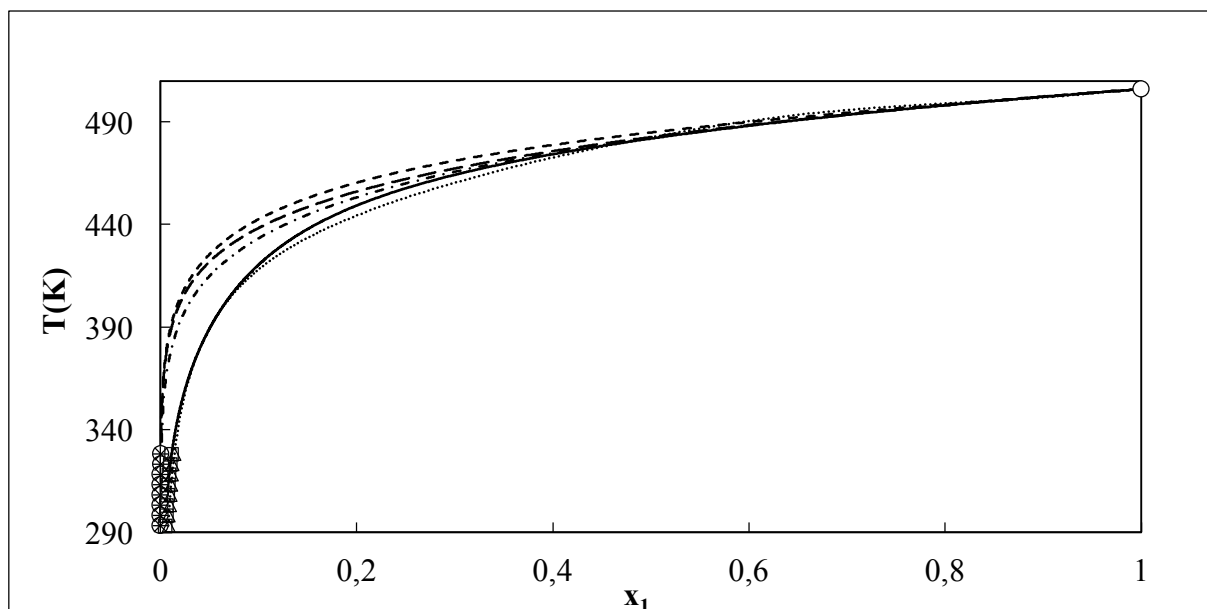


Figure 6.21 T-x plot for the systems prednisolone + (experimental, model), ( $\Delta$ ,  $\cdots$ , T-K-Wilson), N,N dimethylformamide/ ( $\square$ ,  $-$ , T-K-Wilson), 1-methyl-2-pyrrolidone/ ( $\times$ ,  $-\cdot-$ , NRTL), n-hexadecane/ ( $\circ$ ,  $---$ , NRTL), n-dodecane/ ( $+$ ,  $---$ , NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines.

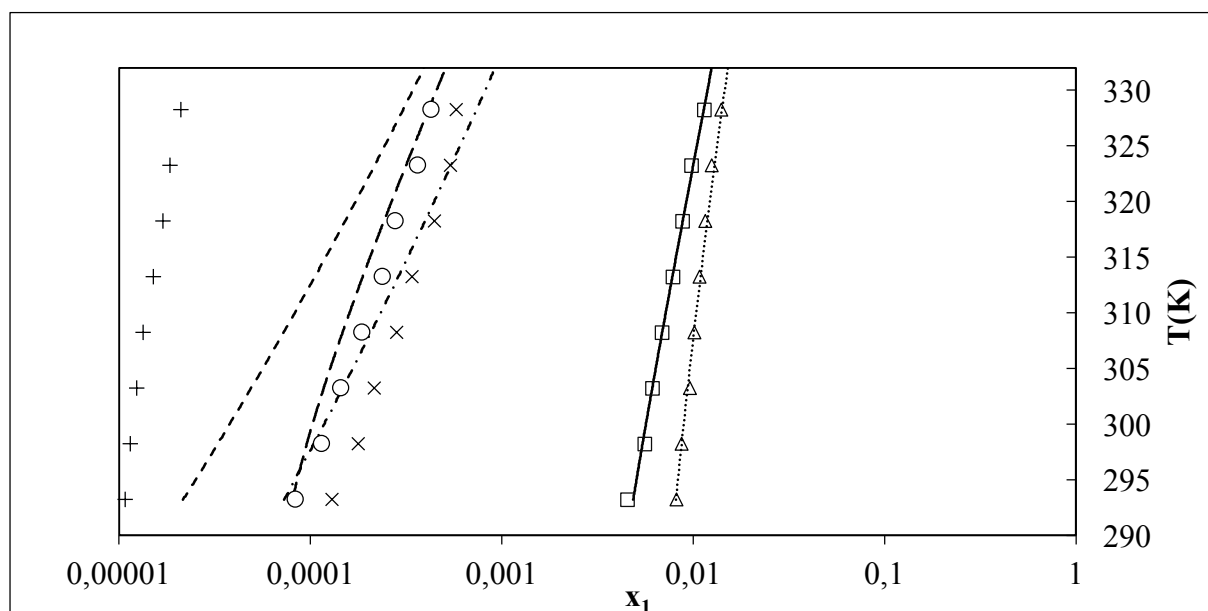


Figure 6.22 Dilute region T-x plot for the systems prednisolone + (experimental, model), ( $\Delta$ ,  $\cdots$ , T-K-Wilson), N,N dimethylformamide/ ( $\square$ ,  $-$ , T-K-Wilson), 1-methyl-2-pyrrolidone/ ( $\times$ ,  $-\cdot-$ , NRTL), n-hexadecane/ ( $\circ$ ,  $---$ , NRTL), n-dodecane/ ( $+$ ,  $---$ , NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines.

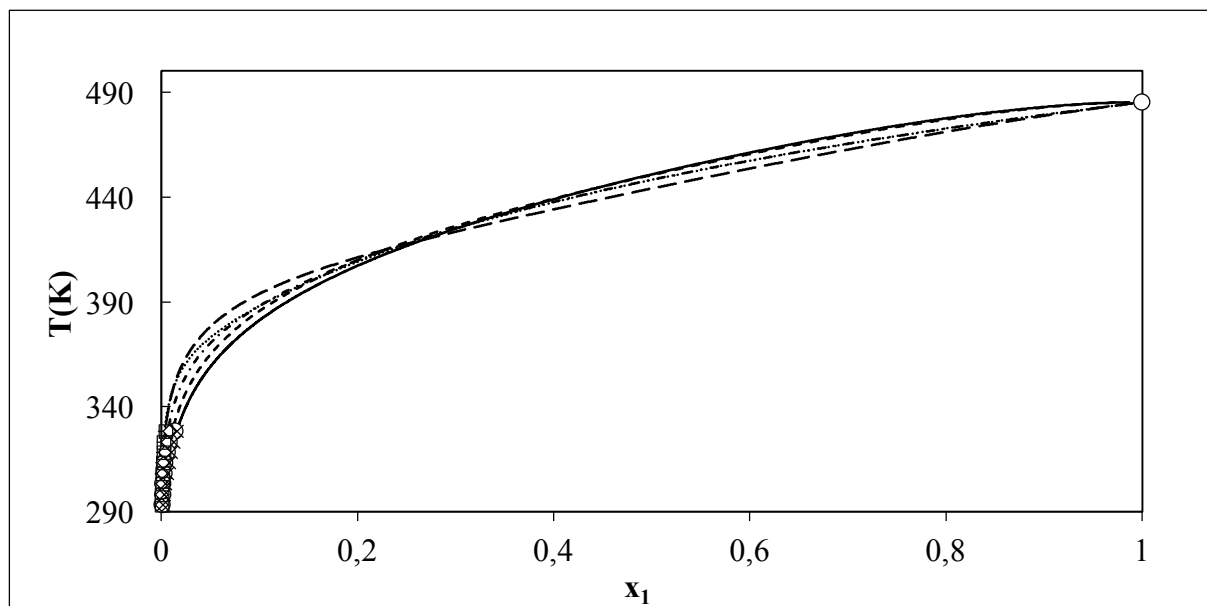


Figure 6.23 T-x plot for the systems hydrocortisone + (experimental, model), (+, ---, T-K-Wilson), acetonitrile/ ( $\times$ , -, NRTL), nonan-1-ol/ ( $\circ$ , ---, T-K-Wilson), octan-1-ol/ ( $\diamond$ , -.-, NRTL), pentan-1-ol/ ( $\square$ , ---, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

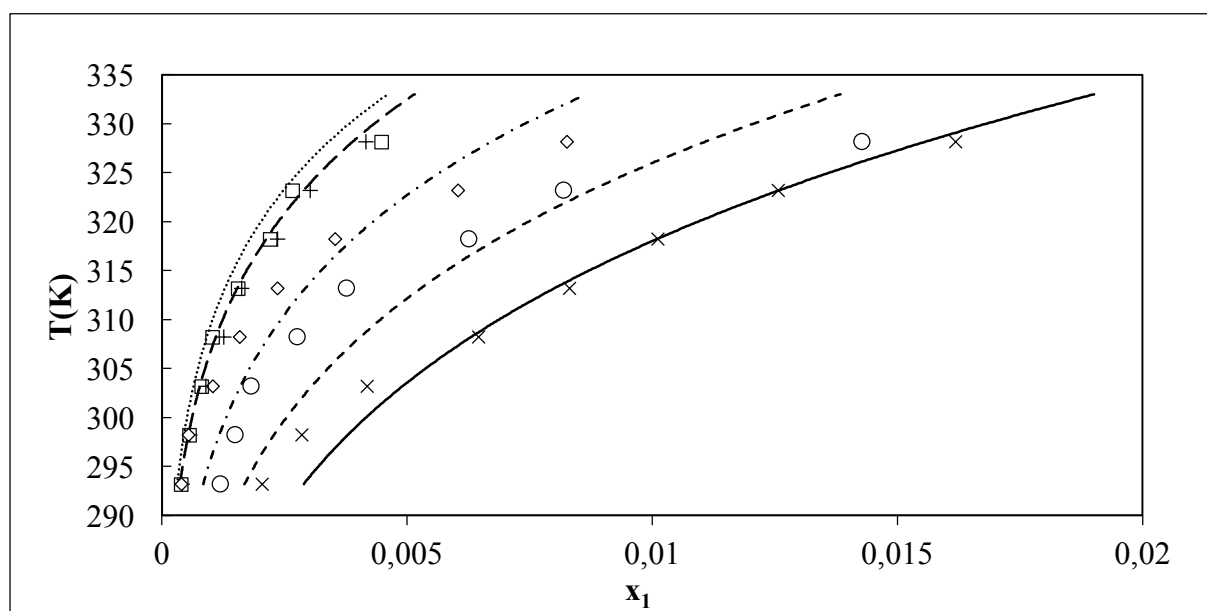


Figure 6.24 Dilute region T-x plot for the systems hydrocortisone + (experimental, model), (+, ---, T-K-Wilson), acetonitrile/ ( $\times$ , -, NRTL), nonan-1-ol/ ( $\circ$ , ---, T-K-Wilson), octan-1-ol/ ( $\diamond$ , -.-, NRTL), pentan-1-ol/ ( $\square$ , ---, T-K-Wilson), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

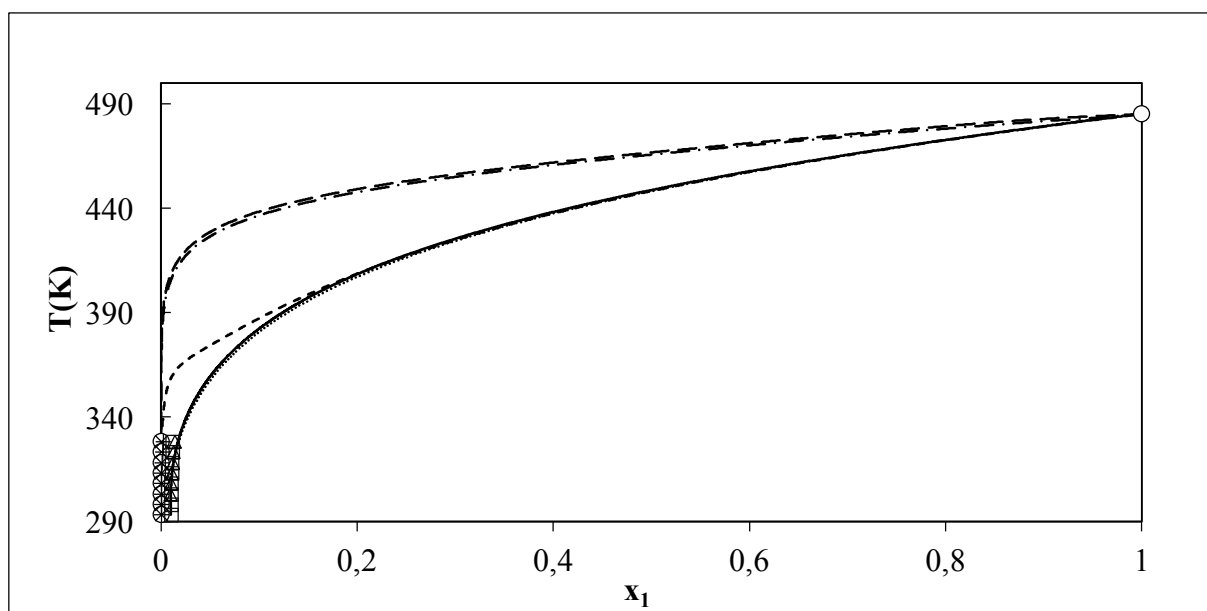


Figure 6.25 T-x plot for the systems hydrocortisone + (experimental, model), ( $\Delta$ ,  $\dots$ , T-K-Wilson), N,N dimethylformamide/ ( $\square$ ,  $-$ , NRTL), 1-methyl-2-pyrrolidone/ ( $\times$ ,  $-\cdot-$ , NRTL), N-Hexadecane/ ( $\circ$ ,  $---$ , NRTL), n-dodecane/ ( $+$ ,  $---$ , NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines.

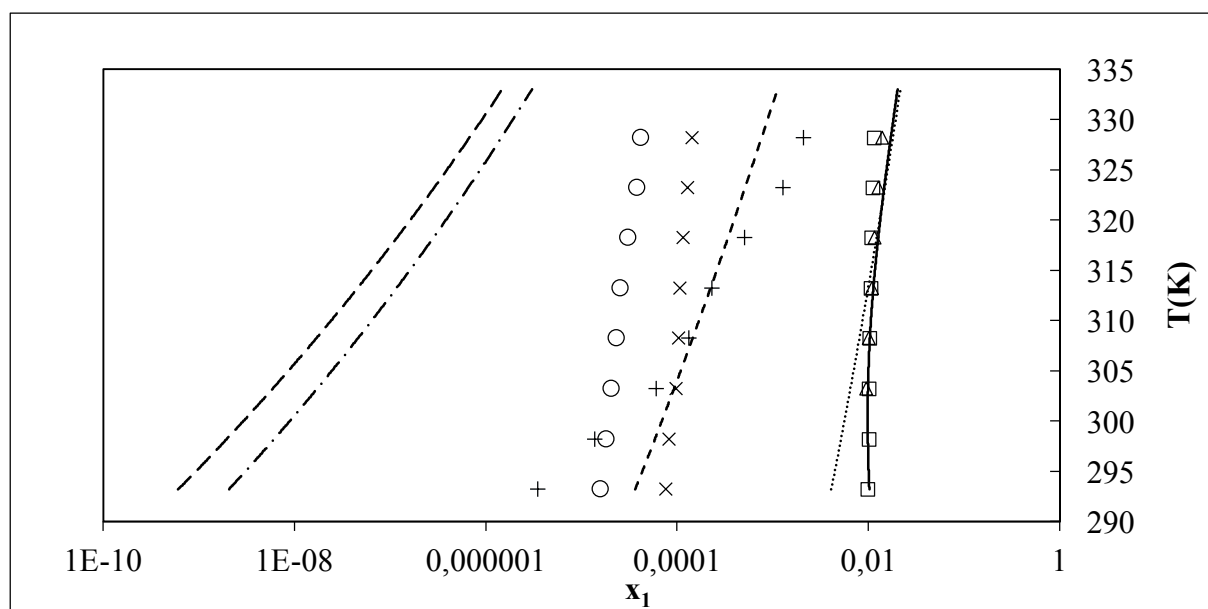


Figure 6.26 Dilute region T-x plot for the systems hydrocortisone + (experimental, model), ( $\Delta$ ,  $\dots$ , T-K-Wilson), N,N dimethylformamide/ ( $\square$ ,  $-$ , NRTL), 1-methyl-2-pyrrolidone/ ( $\times$ ,  $-\cdot-$ , NRTL), N-Hexadecane/ ( $\circ$ ,  $---$ , NRTL), n-dodecane/ ( $+$ ,  $---$ , NRTL), water. Experimental data are represented as symbols, and model predictions are represented as lines.



**Table 6.15** Estimated infinite dilution activity coefficients of solute (1) in solvent (2) at 298.2 K  
0.101 MPa from correlated experimental data.

Solute	Solvent	$\gamma_1^\infty$ at 298.2 K	
		NRTL	T-K-Wilson
<b>Betulin</b>			
	Nonan-1-ol	0.025	0.024
	Pentan-1-ol	0.091	0.060
	Acetonitrile	0.633	0.118
	n-Dodecane	1.000	656.928
	Butan-2-ol	0.995	0.115
	N,N Dimethylformamide	0.026	0.010
	Water	1.000	226.026
	Octan-1-ol	0.026	0.027
	n-Hexadecane	1.152	-
	1-Methyl-2- Pyrrolidone	0.013	0.005
<b>Estrone</b>			
	Nonan-1-ol	0.168	0.117
	Pentan-1-ol	0.450	0.441
	2-Aminoethanol	0.219	-
	Acetonitrile	0.051	0.507
	n-Dodecane	1.981	1.981
	Butan-2-ol	0.550	0.635
	N,N Dimethylformamide	0.061	0.170
	n-Octane	2.931	4.651
	Water	1.079	4.797
	Octan-1-ol	0.258	0.237
	Morpholine-4-carbaldehyde	0.010	994.847
	n-Hexadecane	0.149	0.160
	1-Methyl-2- Pyrrolidone	0.090	-

**Table 6.15 Estimated infinite dilution activity coefficients of solute (1) in solvent (2) at 298.2 K 0.101 MPa from correlated experimental data (continued)**

Solute	Solvent	$\gamma_1^\infty$ at 298.2 K	
		NRTL	T-K-Wilson
<b>Diosgenin</b>			
	Nonan-1-ol	0.002	0.003
	Pentan-1-ol	0.016	0.836
	2-Aminoethanol	1.000	1513.203
	n-Dodecane	0.078	0.092
	Butan-2-ol	1	0.510-
	N,N Dimethylformamide	1.000	-
	n-Octane	0.222	0.152
	Water	0.079	0.948
	Octan-1-ol	0.004	0.005
	Morpholine-4-carbaldehyde	0.002	-
	n-Hexadecane	1.000	0.042
	1-Methyl-2- Pyrrolidone	1.000	-
	Acetonitrile	113.345	-
<b>Estriol</b>			
	Nonan-1-ol	1.000	0.226
	Pentan-1-ol	0.093	0.223
	2-Aminoethanol	0.016	-
	n-Dodecane	0.466	0.453
	Butan-2-ol	0.126	2.145
	N,N Dimethylformamide	0.064	0.081
	n-Octane	1.528	0.694
	Water	2.105	5.860
	Octan-1-ol	1.000	0.793
	Morpholine-4-carbaldehyde	0.124	-
	n-Hexadecane	0.260	0.275
	1-Methyl-2- Pyrrolidone	1.000	0.219

**Table 6.15** Estimated infinite dilution activity coefficients of solute (1) in solvent (2) at 298.2 K  
0.101 MPa from correlated experimental data (continued)

Solute	Solvent	$\gamma_1^\infty$ at 298.2 K	
		NRTL	T-K-Wilson
<b>Prednisolone</b>			
	Nonan-1-ol	0.061	0.051
	Pentan-1-ol	0.704	0.181
	Acetonitrile	1.445	0.021
	n-Dodecane	1.017	0.465
	Butan-2-ol	0.5717	0.860
	N,N Dimethylformamide	0.321	0.005
	Water	1.665	143.829
	Octan-1-ol	1.87554	0.083
	n-Hexadecane	0.966	0.498
	1-Methyl-2- Pyrrolidone	0.061	0.008
<b>Hydrocortisone</b>			
	Nonan-1-ol	1.216	1.406
	Pentan-1-ol	3.176	4.860
	Acetonitrile	0.975	13.148
	n-Dodecane	0.977	-
	Butan-2-ol	0.975	10.921
	N,N Dimethylformamide	0.999	52.542
	Water	0.558	109.984
	Octan-1-ol	0.701	2.400
	n-Hexadecane	1.000	-
	1-Methyl-2- Pyrrolidone	1.049	-

**Table 6.16 Experimental Solubility ranking (highest solubility in solvent to lowest) at 328.2 K and 0.101 MPa.**

<b>Betulin</b>	<b>Estrone</b>
1-Methyl-2-pyrillidone	Morpholine-4-carbaldehyde
N,N Dimethylformamide	N,N Dimethylformamide
Nonan-1-ol	Acetonitrile
Octan-1-ol	1-Methyl-2-pyrillidone
Acetonitrile	N-Hexadecane
Pentan-1-ol	Ethanolamine
Butan-2-ol	1-Nonanol
n-Hexadecane	Octan-1-ol
n-Dodecane	Pentan-1-ol
Water	Butan-2-ol
	n-Dodecane
	Water
	n-Octane
<b>Diosgenin</b>	<b>Estriol</b>
Acetonitrile	2-Aminoethanol
1-Methyl-2-pyrrolidone	Nonan-1-ol
Morpholine-4-carbaldehyde	N,N Dimethylformamide
Nonan-1-ol	Octan-1-ol
2-Aminoethanol	Pentan-1-ol
Octan-1-ol	Morpholine-4-carbaldehyde
N,N Dimethylformamide	Butan-2-ol
Pentan-1-ol	1-Methyl-2-pyrrolidone
Butan-2-ol	n-Hexadecane
n-Hexadecane	n-Dodecane
Water	n-Octane
n-Dodecane	Water
n-Octane	2-Aminoethanol
<b>Prednisolone</b>	<b>Hydrocortisone</b>
N,N Dimethylformamide	Nonan-1-ol
1-Methyl-2-pyrillidone	Octan-1-ol
n-Hexadecane	N,N Dimethylformamide
Acetonitrile	1-Methyl-2-pyrillidone
Nonan-1-ol	Pentan-1-ol
Octan-1-ol	Butan-2-ol
Pentan-1-ol	Acetonitrile
Butan-2-ol	Water
n-Dodecane	n-Hexadecane
Water	n-Dodecane

*6.5 Conclusions*

Solid-liquid equilibrium data was measured for the systems of betulin, estrone, diosgenin, estriol, prednisolone and hydrocortisone in various solvents by employing combined thermal gravimetry and digital thermal analysis. The pure solute enthalpy of fusion and melting point temperatures were also measured. The hierarchy of solubility in each solvent was determined and listed. The experimental SLE data was modelled using the NRTL model and T-K-Wilson models with good representation in most cases. The T-K-Wilson model provided superior fits in regions of extremely dilute solubility such as the cases of alkane and water as a solvent.

*References*

- Alakurtti, S., Mäkelä, T., Koskimies, S. and Yli-Kauhaluoma, J., (2006), Pharmacological properties of the ubiquitous natural product betulin. *European Journal of Pharmaceutical Sciences*, 29, pp.1–13.
- Bebo, B.F. Jr., Fyfe-Johnson A., Adlard, K., Beam, A.G. Vandebark, A.A. and Offner, H., (2001), Low-Dose Estrogen Therapy Ameliorates Experimental Autoimmune Encephalomyelitis in Two Different Inbred Mouse Strains. *Journal of Immunology*, 166, pp.2080–2089.
- Cai, X., Grant, D.J. and Wiedmann, T.S., (1997), Analysis of the solubilization of steroids by bile salt micelles. *Journal of pharmaceutical sciences*, 86, pp.372–377.
- Cao, D., Zhao, G. and Yan, W., (2007), Solubilities of betulin in 14 solvents.pdf. *J. Chem. Eng. Data*, 52, pp.1366–1368.
- Chen, F.X., Zhao, M.R., Liu, C.C., Peng, F.F. and Ren, B.Z., (2012), Determination and correlation of the solubility for diosgenin in alcohol solvents. *Journal of Chemical Thermodynamics*, 50, pp.1–6.
- Chen, F.X., Zhao, M.R., Ren, B.Z., Zhou, C.R. and Peng, F.F., (2012), Solubility of diosgenin in different solvents. *Journal of Chemical Thermodynamics*, 47, pp.341–346.
- Chen, F.X., Qi, Z. L., Feng, L., Miao J. Y., Ren B.Z., (2014), Application of the NRTL method to correlate solubility of diosgenin , *Journal of Chemical Thermodynamics*, 71, pp.231-235.
- Czock, D., Keller, F., Rasche, F.M. and Häussler, U., (2005), Pharmacokinetics and pharmacodynamics of systemically administered glucocorticoids. *Clinical pharmacokinetics*, 44, pp.61–98.
- Domańska, U., Pobudkowska, A., Pelczarska, A., Winiarska-Tusznio, M. and Gierycz, P., (2010), Solubility and pKa of select pharmaceuticals in water, ethanol, and 1-octanol. *The Journal of Chemical Thermodynamics*, 42, pp.1465–1472.
- Evers, M., Poujade, C., Soler, F., Ribeill, Y., James, C., Lelièvre, Y., Gueguen, J.C., Reisdorf, D., Morize, I., et al., (1996), Betulinic acid derivatives: A new class of human immunodeficiency virus type 1 specific inhibitors with a new mode of action. *Journal of Medicinal Chemistry*, 39, pp.1056–1068.
- Fulda, S., Jeremias, I., Steiner, H.H., Pietsch, T. and Debatin, K.M., (1999), Betulinic acid: a new cytotoxic agent against malignant brain-tumor cells. *Int. J. Cancer*, 82, pp.435–441.
- Green, B., Bentley, M.D., Chung, B.Y., Lynch, N.G. and Jensen, B.L., (2007), Isolation of Betulin and Rearrangement to Allobetulin. A Biomimetic Natural Product Synthesis. *Journal of Chemical Education*, 84, p.1985.

- Hagen, T. A and Flynn, G.L., (1983), Solubility of hydrocortisone in organic and aqueous media: evidence for regular solution behavior in apolar solvents. *Journal of pharmaceutical sciences*, 72, pp.409–414.
- Hildebrand, J.H. and Scott, R.L., (1964), *The solubility of nonelectrolytes*. 3rd ed., Dover, New York.
- James, A., Lord, M., (1992), *Macmillan's Chemical and Physical Data*. Macmillan Press, London, UK.
- Jansson, L., Olsson, T. and Holmdahl, R., (1994), Estrogen induces a potent suppression of experimental autoimmune encephalomyelitis and collagen-induced arthritis in mice. *Journal of Neuroimmunology*, 53, pp.203–207.
- Kabasakalian, P., Britt, E. and Yudis, M.D., (1966), Solubility of some steroids in water. *Journal of pharmaceutical sciences*, 55, p.642.
- Kim, S., Liva, S.M., Dalal, M.A., Verity, M.A. and Voskuhl, R.R., (1999), Estriol ameliorates autoimmune demyelinating disease: Implications for multiple sclerosis. *Neurology*, 52, pp.1230–1230.
- Kontogeorgis, G.M. and Folas, G.K., (2010), *Thermodynamic Models for Industrial Applications*. Chichester, UK: John Wiley and Sons, Ltd.
- Li, J., Masso, J. and Guertin, J., (2002), Prediction of drug solubility in an acrylate adhesive based on the drug-polymer interaction parameter and drug solubility in acetonitrile. *Journal of controlled release: official journal of the Controlled Release Society*, 83, pp.211–221.
- Lide, D.R., (2005), *CRC Handbook of Chemistry and Physics 86TH Edition 2005-2006*. CRC Press, Boca Raton.
- Malisetty, V.S., Patlolla, J.M.R., Raju, J., Marcus, L.A., Choi, C.-I. and Rao, C. V., (2005), Chemoprevention of colon cancer by diosgenin, a steroidal saponin constituent of fenugreek. *Cancer Res.*, 65, p.580.
- NIST (2000) <http://physics.nist.gov/cgi-bin/cuu/Info/Uncertainty/index.html> (Accessed January 2015)
- Pisha, E., Chai, H., Lee, I.-S., Chagwedera, T.E., Farnsworth, N.H.S., Cordell, G.A., Beecher, C.W.W., Fong, H.H.S., Kinghorn, A.D., et al., (1995), Discovery of betulinic acid as a selective inhibitor of human melanoma that functions by induction of apoptosis. *Nature medicine*, 1, pp.1046–1051.
- Raju, J. and Mehta, R., (2009), Cancer chemopreventive and therapeutic effects of diosgenin, a food saponin. *Nutrition and cancer*, 61, pp.27–35.
- Regosz, A., Chmielewska, A. Pelplinska, T., Kowalski, P., (1994), Prediction of the solubility of steroid hormones. *Die Pharmazie*, 49, pp.371–373.

- Renon, H. and Prausnitz, J., (1968), Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE journal*, 14, pp.135–144.
- Ruchelman, M.W., (1967), Solubility studies of estrone in organic solvents using gas-liquid chromatography. *Analytical biochemistry*, 19, pp.98–108.
- Ruchelman, M.W. and Howe, C.D., (1969), Solubility Studies of Estriol in Organic Solvents Using Gas-Liquid Chromatography. *Journal of Chromatographic Science*, 7, pp.340–347.
- Sanghvi, R., Narazaki, R., Machatha, S.G. and Yalkowsky, S.H., (2008), Solubility improvement of drugs using N-methyl pyrrolidone. *AAPS PharmSciTech*, 9, pp.366–376.
- Shareef, A., Angove, M.J., Wells, J.D. and Johnson, B.B., (2006), Aqueous Solubilities of Estrone, 17 $\beta$ -Estradiol, 17 $\alpha$ -Ethinylestradiol, and Bisphenol A. *Journal of chemical and engineering data*, 51, pp.879–881.
- Sicotte, N.L., Liva, S.M., Klutch, R., Pfeiffer, P., Bouvier, S., Odesa, S., Wu, T.C.J. and Voskuhl, R.R., (2002), Treatment of multiple sclerosis with the pregnancy hormone estriol. *Annals of neurology*, 52, pp.421–428.
- Tsuboka, T. and Katayama, T., (1975), Modified Wilson Equation for Vapor-Liquid and Liquid-Liquid Equilibria. *Journal of Chemical Engineering of Japan*, 8, pp.181–187.
- Varanda, F., Pratas De Melo, M.J., Caço, A.I., Dohrn, R., Makrydaki, F. A., Voutsas, E., Tassios, D. and Marrucho, I.M., (2006), Solubility of antibiotics in different solvents. 1. Hydrochloride forms of tetracycline, moxifloxacin, and ciprofloxacin. *Industrial and Engineering Chemistry Research*, 45, pp.6368–6374.
- Vetere, A., (2004), The NRTL equation as a predictive tool for vapor-liquid equilibria. *Fluid Phase Equilibria*, 218(April 2003), pp.33–39.
- Yalkowsky, S.H. and Valvani, S.C., (1980), Solubility and partitioning I: Solubility of nonelectrolytes in water. *Journal of pharmaceutical sciences*, 69, pp.912–922.
- Yu, Q., Ma, X. and Xu, L., (2013), TG-assisted determination of pyrene solubility in 1-pentanol. *Journal of Thermal Analysis and Calorimetry*, 112, pp.1553–1557.
- Zhang, X., Yin, Q., Gong, J., Liu, Z. and Section, E., (2010), Solubility of 5-Amino-N,N'-bis(2,3-dihydroxypropyl)-2,4,6-triiodobenzene-1,3-dicarboxamide in Ethanol + Water Mixtures. *J Chem. Eng. Data*, 55, pp.2355–2357.
- Zhao, G. & Yan, W., (2007), Experimental Determination of Solubilities of Betulin in Acetone + Water and Ethanol + Water Mixed Solvents at T = (278.2, 288.2, 298.2, 308.2, and 318.2) K. *Journal of Chemical & Engineering Data*, 52, pp.2365–2367.
- Zhao, G. and Yan, W., (2008), Solubilities of betulin in chloroform + methanol mixed solvents at T = (278.2, 288.2, 293.2, 298.2, 308.2 and 313.2) K. *Fluid Phase Equilibria*, 267, pp.79–82.



## CHAPTER SEVEN

### Culminating Discussion

In Chapter Two, the most common predictive and semi-predictive methods for the solubility of complex, active pharmaceutical ingredients, including triterpene and steroidal pharmaceuticals, were analysed. The models tested included the UNIFAC-based (Fredenslund et al. (1975), Weidlich and Gmehling (1987)), COSMO-based (Grensemann and Gmehling (2005), Lin and Sandler (2002)), and NRTL-SAC (Chen and Song, 2004) models.

It became clear, upon analysis, that although the NRTL-SAC model provides a good replication of the experimental data, it is not easily employed, due to the large number of model parameters required for its application. The UNIFAC-based, COSMO-based, and NRTL-SAC models tend to underestimate the solubility, in the majority of the solutes considered.

It was also noted that, in many cases, predictions by each of the models could not be conducted, if the relevant model parameters were not available in the literature. For instance, sigma profiles for several of the solutes considered were not available in the literature, thus Gaussian calculations had to be performed to obtain these parameters in this work.

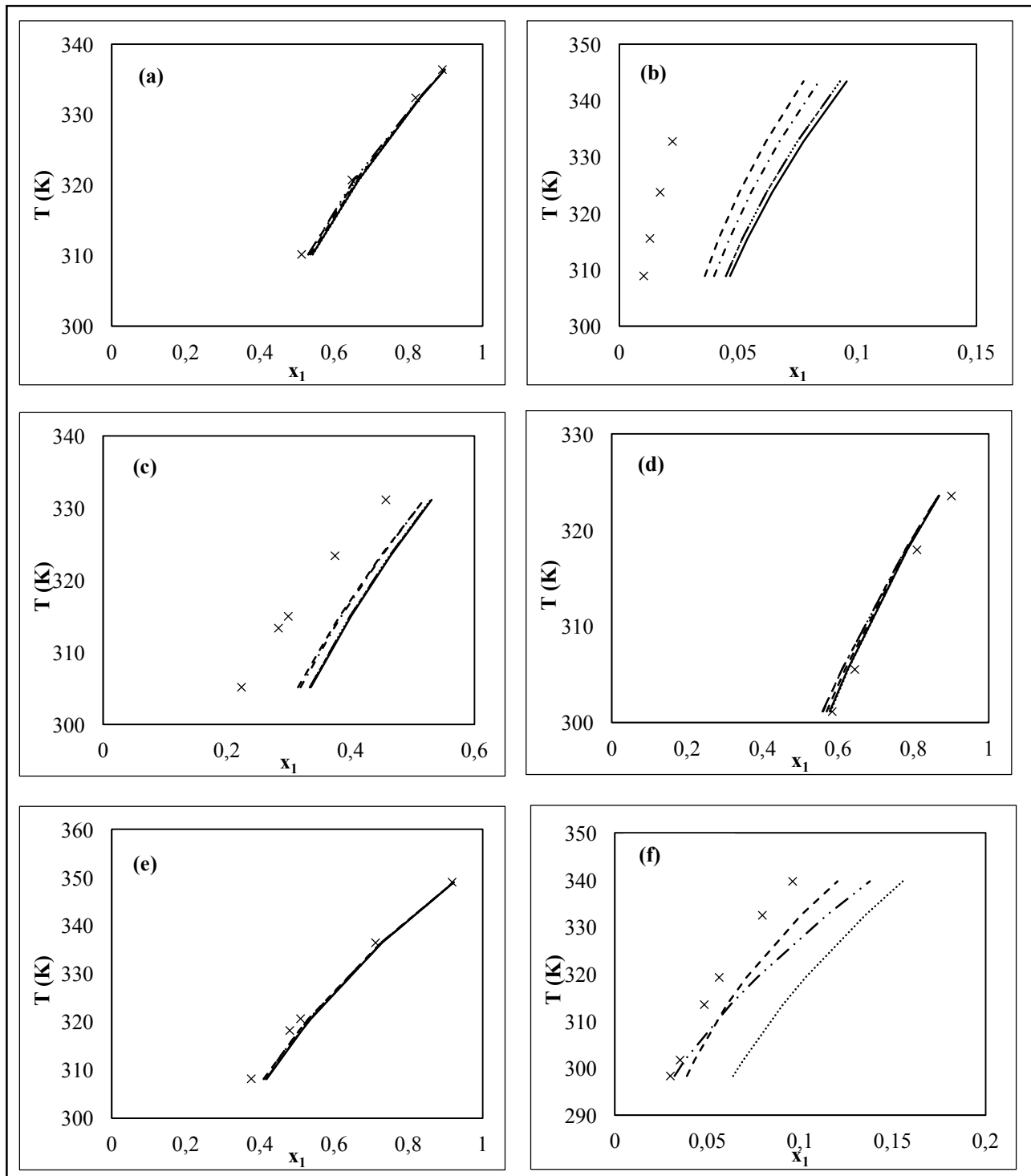
Furthermore the group contribution methods could not be applied to the cases of prednisolone and hydrocortisone, as these components could not be fragmented by the fragmentation schemes of UNIFAC, and modified UNIFAC (Dortmund).

The effect of the assumption of negligible heat capacity changes, in the solubility modelling of the systems considered, was also analysed, and it was found that the Staverman-Guggenheim combinatorial term, with the modified UNIFAC (Dortmund) model, with the approximation of  $\Delta_{fus}C_{pi} = 0$ , provided the lowest percentage deviation for the solutes considered, for benzene as a solvent. In aqueous systems the original UNIFAC model, with the approximation of  $\Delta_{fus}C_{pi} = \Delta_{fus}S_i$  provided a superior fit.

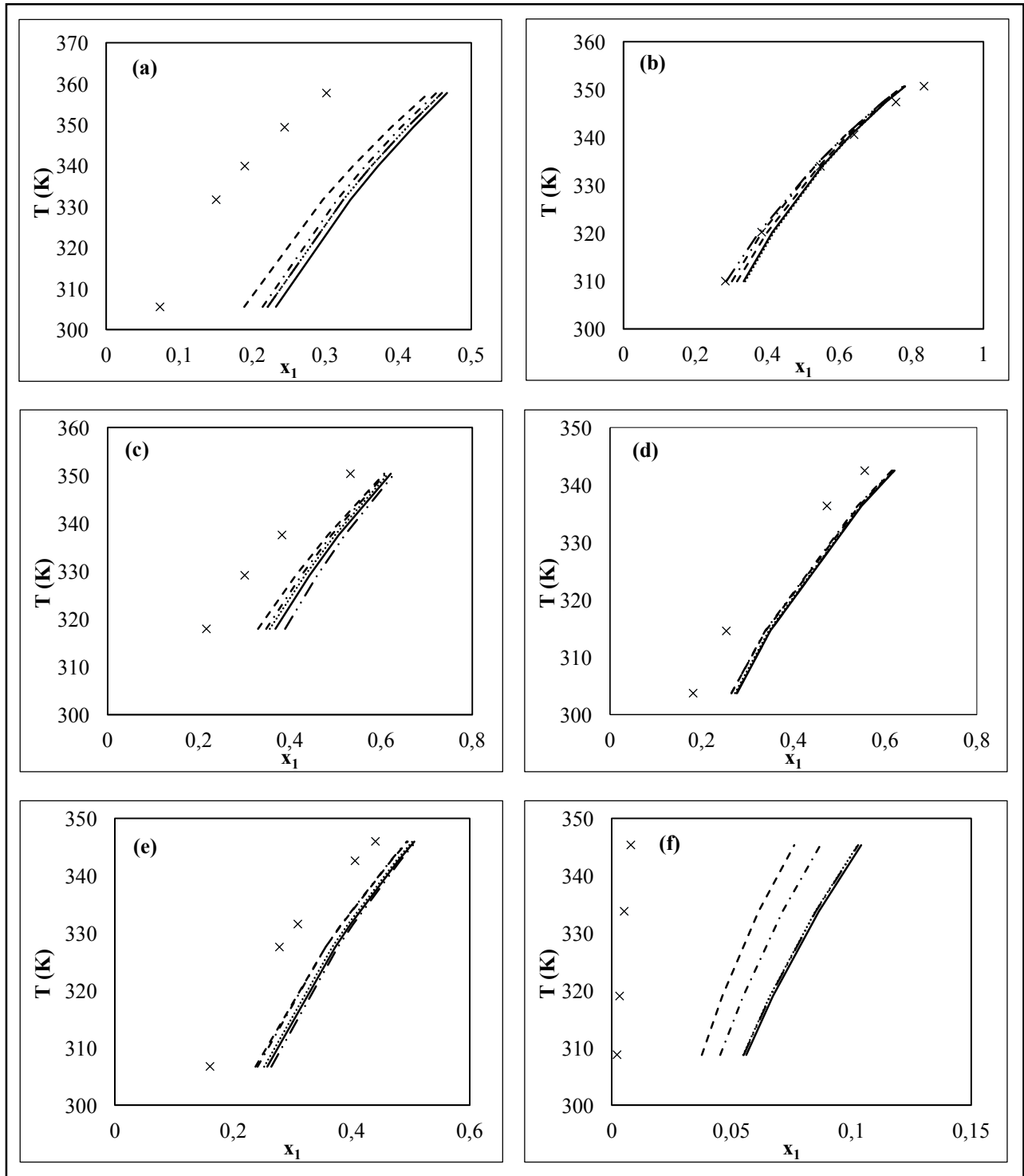
Furthermore, attempts were made to correlate the performances of the various models, with any particular parameter of the solute that may give an indication of any model bias. The molecular weight, van der Waals molecular surface area parameter, and functional group diversity were considered, with no conclusive result.

Subsequent to the development of the Extended UNISAC model, a portion of the test data set used in Chapter Two, with benzene as a solvent, was compared to the predictions by the Extended UNISAC model, to observe its performance and establish if similar predictive patterns exist to the common models from the literature. In these systems, the Extended UNISAC model with Staverman-Guggenheim combinatorial term, and with the approximation of  $\Delta_{fus}C_{pi} = \Delta_{fus}S_i$ , provided a superior prediction to alternate combinatorial expressions (discussed below), or with the approximation of  $\Delta_{fus}C_{pi} = 0$ . These results are presented in Figure 7.1-7.3.

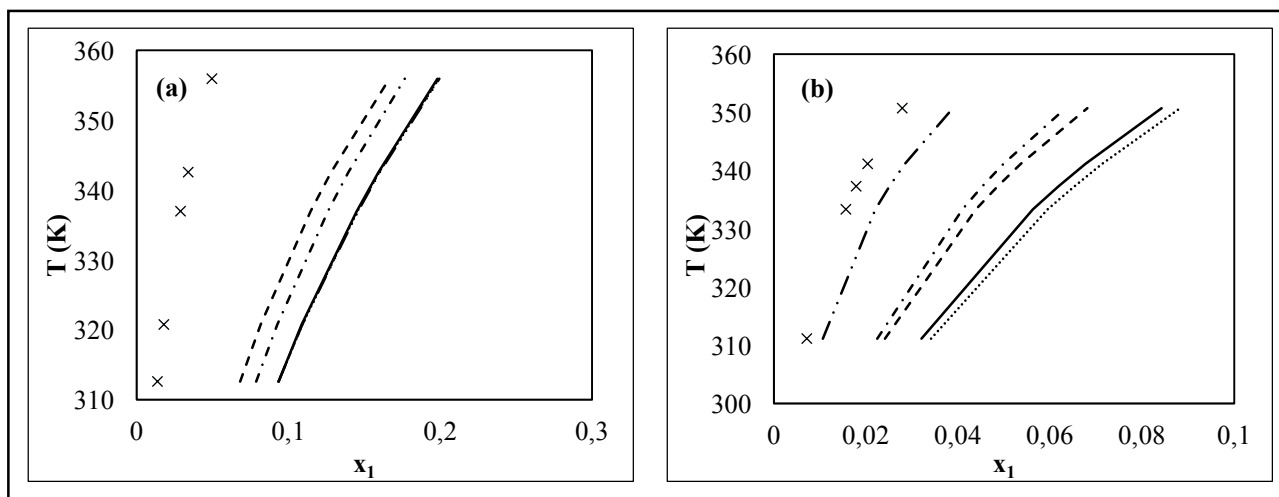
Extended UNISAC is clearly competitive with the existing predictive methods in the literature, despite the drastically reduced model parameter set required for application.



**Figure 7.1** Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models for the system solute (1) + benzene (2).  $\times$ , Experimental Data (McCloughlin and Zainal, 1959);  $\cdots$ , Extended UNISAC;  $---$ , mod. UNIFAC (Dortmund);  $\cdots$ , UNIFAC;  $—$ , COSMO-SAC;  $- \cdot -$ , COSMO-RS (OL). Solute: (a) biphenyl, (b) anthracene, (c) phenanthrene, (d) 1,2-diphenylbenzene, (e) bicyclo[4.4.0]deca-1,3,5,7,9-pentene, (f) 1,3,5-triphenylbenzene. Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 7.2** Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models for the system solute (1) + benzene (2).  $\times$ , Experimental Data (McCloughlin and Zainal, 1959);  $-\cdot-\cdot-$ , Extended UNISAC;  $-\cdot-\cdot-$ , mod. UNIFAC (Dortmund);  $\dots$ , UNIFAC;  $-$ , COSMO-SAC;  $-\cdot-$ , COSMO-RS (OL). Solute: (a) pyrene, (b) 1,3-diphenylbenzene, (c) fluoranthene, (d) 1,2-dihydroacenaphthylene, (e) 9H-fluorene, (f) chrysene. Experimental data are represented as symbols, and model predictions are represented as lines.

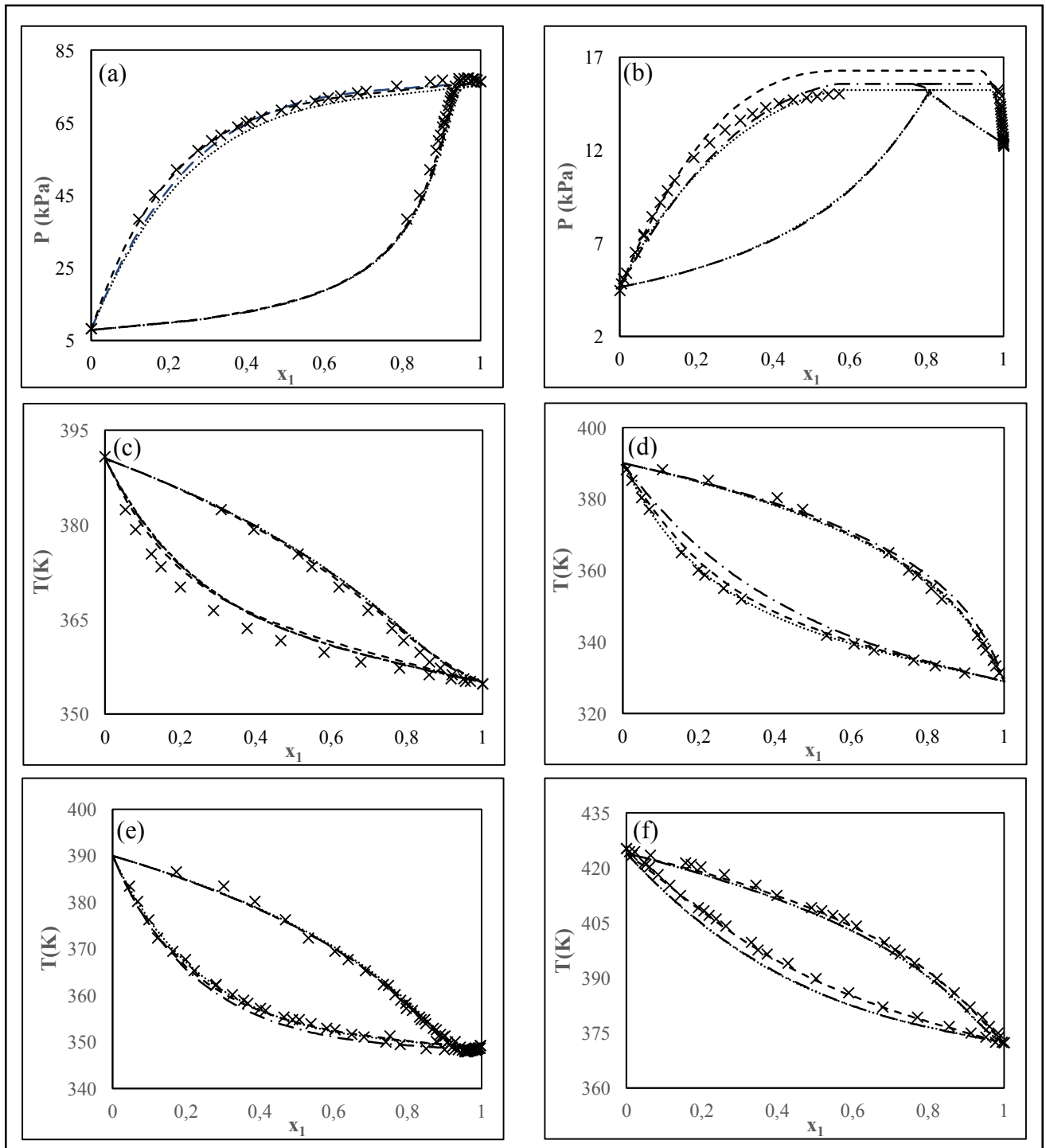


**Figure 7.3** Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models for the system solute (1) + benzene (2). ×, Experimental Data (McCloughlin and Zainal, 1959); —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-SAC; —·—, COSMO-RS (OL). Solute: (a) triphenylene, (b) 1,4-diphenylbenzene. Experimental data are represented as symbols, and model predictions are represented as lines.

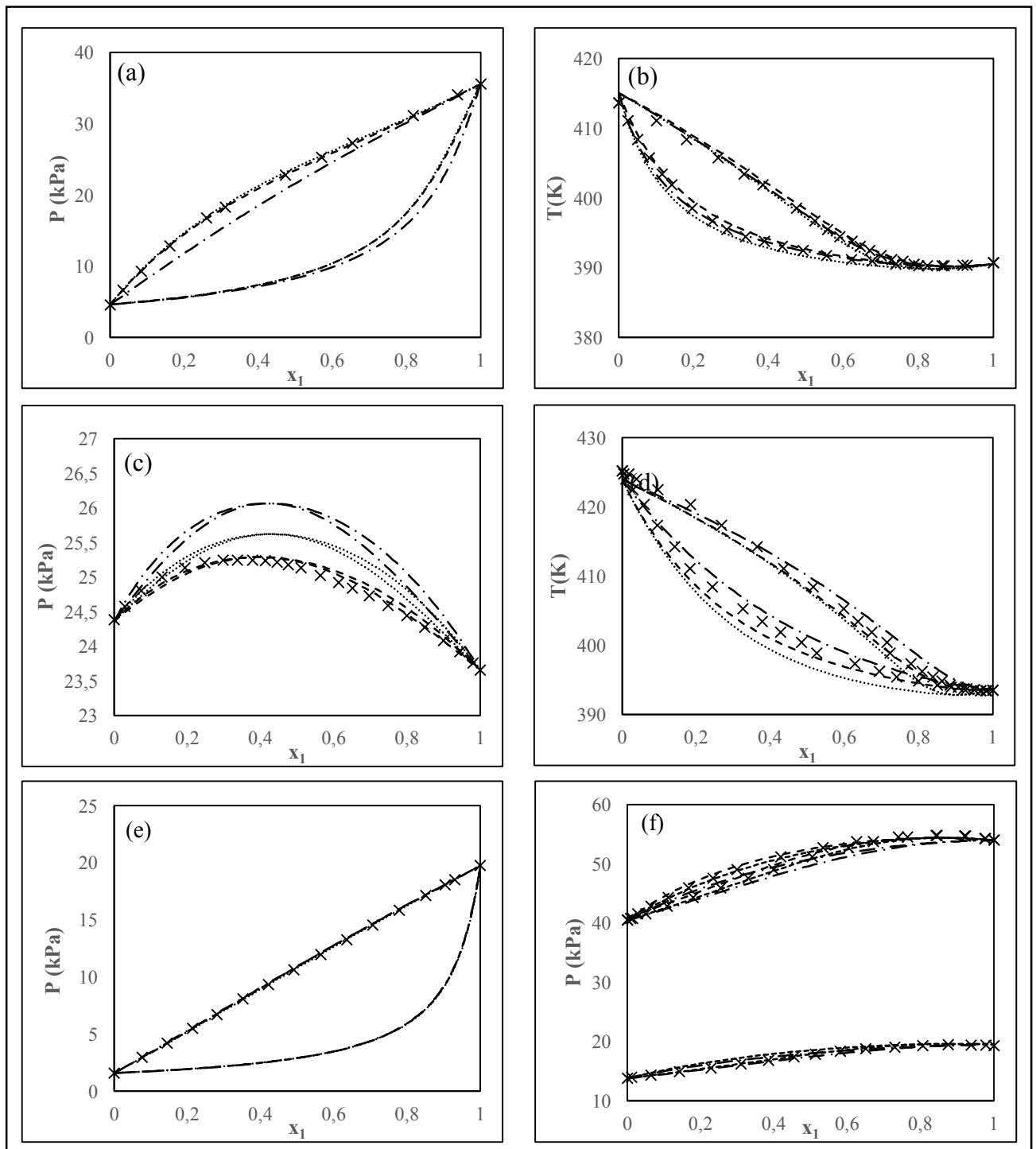
The UNISAC model, presented in Chapter Three, aimed to address the identified shortcomings of the various predictive models already available in the literature. In this paper, four main hypotheses were tested and subsequently confirmed:

1. Can the segment contribution concept of interactions between molecular surfaces, proposed by Chen and co-workers (Chen and Song, 2004, Chen and Crafts, 2006), in NRTL-SAC, be analogously applied to interactions between functional groups?
2. Does a reduction in the group-group binary interaction parameter matrix drastically affect the predictive capability of the group contribution approach, when considering solid-liquid equilibrium calculations?
3. Can surface segment areas adequately scale binary interactions between adjacent surface segments when normalized, and is the UNISAC model very sensitive to these surface segment area parameters?
4. Is the UNISAC model competitive with other popular models from the literature, in terms of range of applicability and degree of accuracy?

With regards to question (1) above, although the four parameter segment contribution concept was found to be sufficient, in describing solid-liquid behaviour, in the majority of the instances tested, due to the refitting of group interaction parameters specifically to solid-liquid systems for the UNISAC model, the vapour-liquid equilibrium (VLE) predictive capabilities of the model was significantly affected. Thus although question (2) was confirmed for solid-liquid systems, a revision of the model was conducted, and an extension implemented, with additional base segments considered. Some test VLE predictions for the most useful solvent classes (comprised of commonly occurring functional groups), were performed, and compared to experimental data, along with predictions by the UNIFAC and modified UNIFAC (Dortmund) models, and are presented in Figure 7.4-7.5. It was confirmed that qualitative, and mostly quantitative, VLE predictions can also be obtained by means of the Extended UNISAC model.



**Figure 7.4** Experimental and predicted isobaric/isothermal vapour-liquid equilibrium data by different models.  $\times$ , Experimental Data; —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC. Systems:(a) n-hexane (1) + butan-1-ol (2), 332.53 K, (Berro et al., 1982); (b) water (1) + butan-1-ol (2), 323.23 K, (Fischer and Gmehling, 1994); (c) acetonitrile (1) + butan-1-ol (2), 101.3 kPa, (Kovac et al., 1985); (d) acetone (1) + butan-1-ol (2), 99.45 kPa, (Michalski et al., 1961); (e) tetrachloromethane (1) + butan-1-ol (2), 99.33 kPa, (Doniec et al., 1965); (f) chloroform (1) + butan-1-ol (2), 303 kPa, (Chen et al., 1995). Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 7.5** Experimental and predicted isobaric/isothermal vapour-liquid equilibrium data by different models. ×, Experimental Data; —·—, Extended UNISAC; - - -, mod. UNIFAC (Dortmund); ..., UNIFAC. Systems:(a) butanone (1) + butan-1-ol (2), 323.15 K, (Garriga et al., 1996); (b) dibutyl ether (1) + butan-1-ol (2) 101.3 kPa, (Lladosa et al., 2006); (c) benzene (1) + butanone (2), 313.15 K, (Van Nhu and Kholer, 1989) ; (d) benzene (1) + butan-1-ol (2), 303 kPa, (Chen et al., 1995); (e) benzene (1) +dibutyl ether (2), 308.15 K, (Ott et al., 1981); (f) heptanone (1) + dibutyl ether (2), 393.22 K, (Wu and Sandler, 1988). Experimental data are represented as symbols, and model predictions are represented as lines.



As to question (3), essentially, in the Extended UNISAC model, the residual term of the original UNIFAC model was replaced to incorporate the segment surface interaction concept. However, the binary interaction parameters of the original UNIFAC model were retained. This decision reveals a major underlying assumption of the Extended UNISAC model concept, which is that although altering the molecular surface area parameter (albeit in the residual term only), would potentially give rise to changes in the binary interaction parameters between groups, it was assumed that these variances can be captured through fitting the “*interacting*” segment surface areas only, for non-base groups. The segment areas of the base groups, that correspond to the interaction parameters used, are identical to those employed in UNIFAC, (Bondi, (1964) values).

These newly defined interacting segment surface areas are highly sensitive, and a superior regression algorithm and method was required. Thus the Krill Herd (KH) Algorithm (Gandomi, and Alavi, 2012), and Lévy flight Krill Herd Algorithm (Wang et al., 2013), were tested in Chapter Four, with the Krill Herd Algorithm subsequently employed in this work.

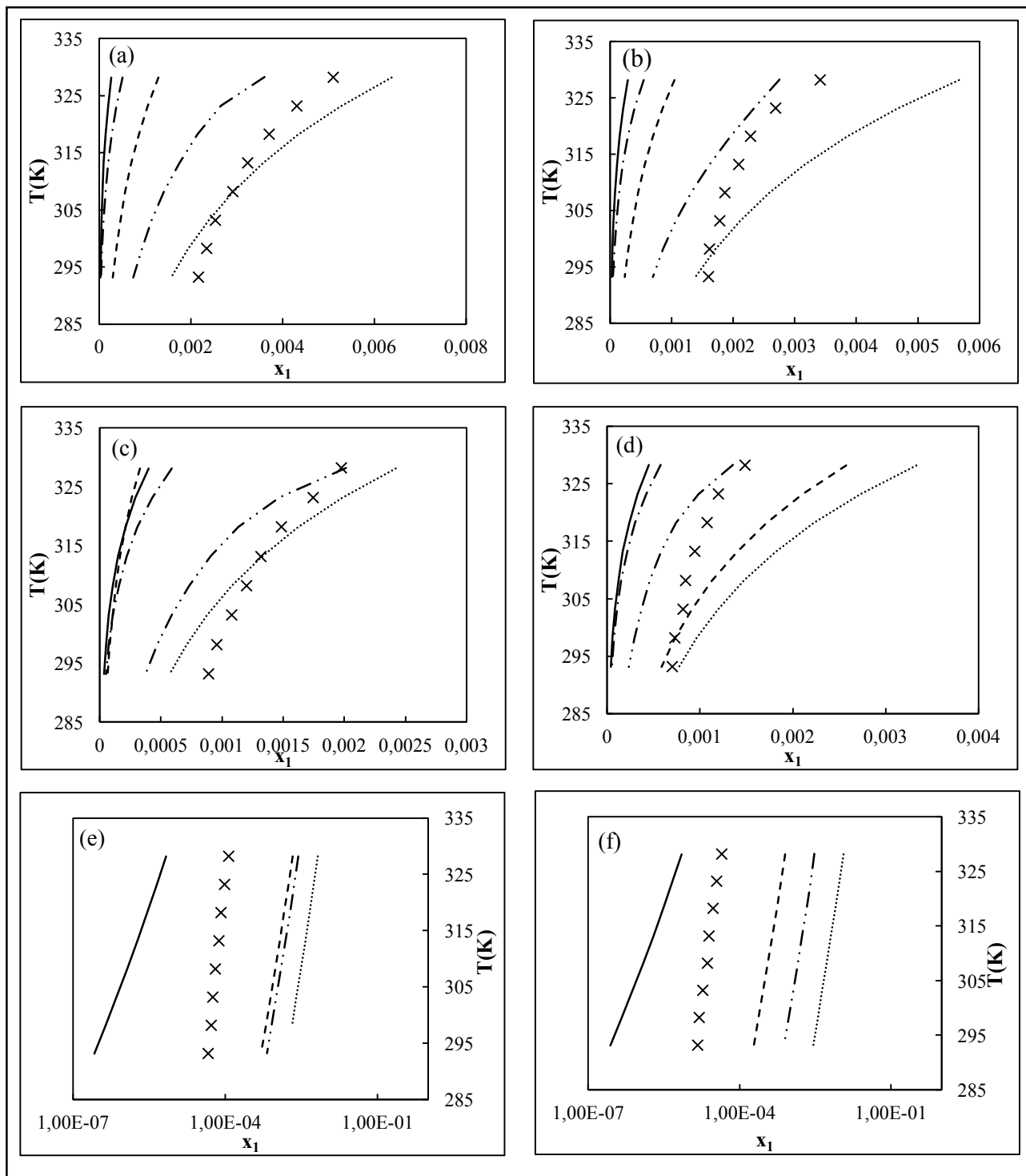
The algorithm was found to reliably calculate the global optima in phase stability, phase equilibrium, and reactive phase equilibrium problems, with lower computing times and higher, or equal success rates, than previously applied metaheuristic techniques, such as those involving swarm intelligence and genetic algorithms, from the literature. For the phase stability problems tested, a success rate of 89.5 % was obtained for the KH algorithm, in comparison to the most favourable method in the literature, Covariant Matrix Adaptation Evaluation Strategy (CMA-ES) (Fateen et al., 2012), which yields a success rate of 86 %. For the cases of phase equilibrium and reactive phase equilibrium tested, the KH algorithm matches the success rate of the most favourable method in the literature, Cuckoo Search (CS) (Bhargava et al., 2013), with success rates of 99% in both cases.

The Extended UNISAC model, presented in Chapter Five, introduced the Moodley-Rarey-Ramjugernath fragmentation scheme, which improves the applicability of the UNISAC model. Additional segment area parameters, obtained by regression using the Krill Herd Algorithm, are also presented. The extended model requires a slightly larger number of segment area parameters than the UNISAC model, but is able to handle a broader range of structurally dissimilar components. The updated Extended UNISAC segment area parameters allow predictions for a significant portion of the available SLE systems in the Dortmund Data Bank (2012), when not considering duplicate measurements.

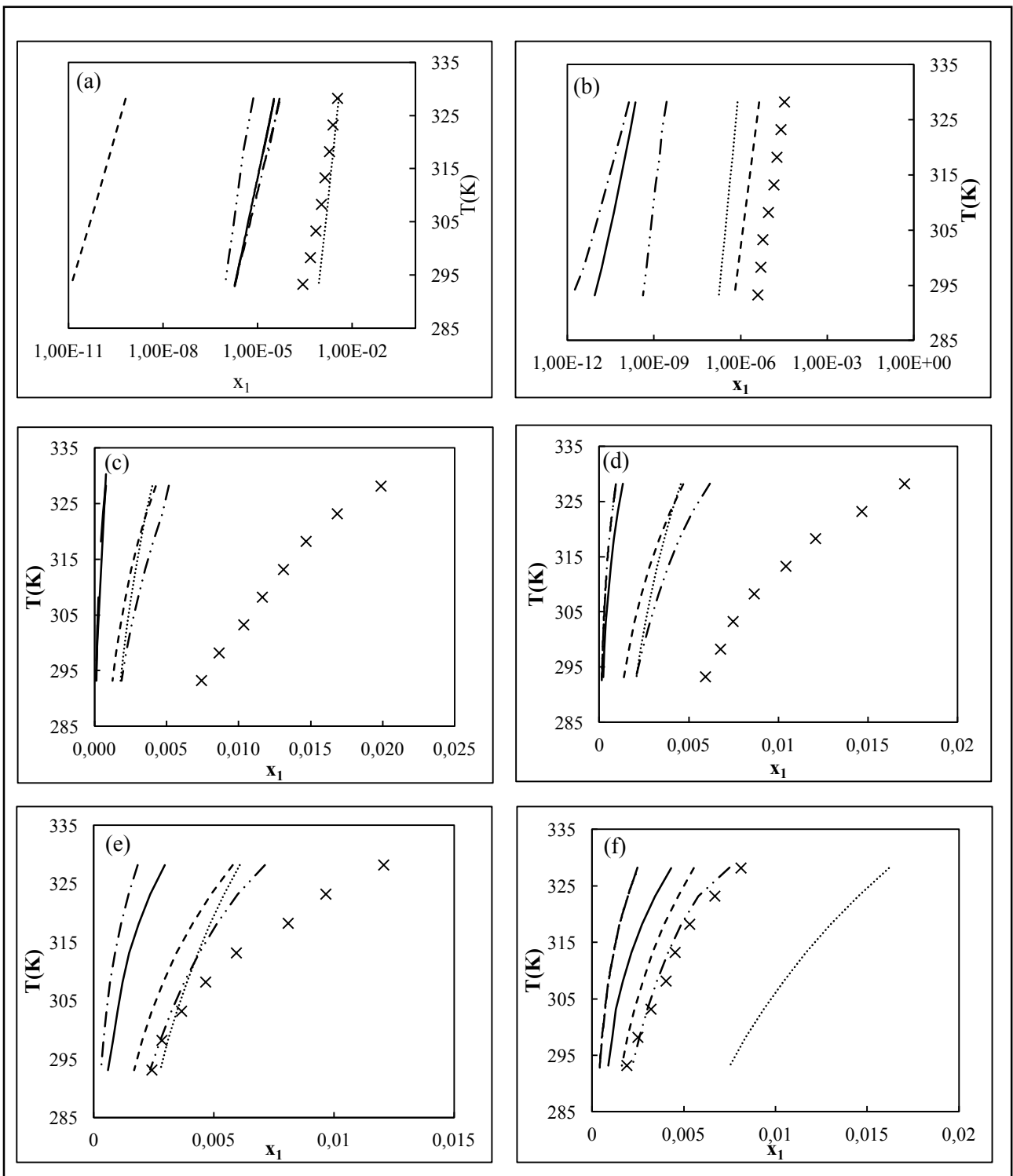
Confirming question (4), the robustness of the Extended UNISAC model was tested by application to the prediction of solid-liquid equilibrium compositions, or temperatures, in binary and ternary mixtures, which include non-isothermal solubility data for complex solutes from the literature. The Akaike Information Criterion (Akaike, 1974), suggests that the Extended UNISAC model is favourable to the original UNIFAC, modified UNIFAC (Dortmund), COSMO-RS(OL), and COSMO-SAC models, with relative AIC scores of 1.95, 4.17, 2.17 and 2.09 for UNIFAC, respectively. The model is also competitive in terms of percentage deviations alone, and is a close third to the original UNIFAC model.

A test set, of previously unmeasured data, was desired to ensure that inter-correlation of data fitting parameters did not occur. Namely, it is certain that previously unmeasured data did not form part of the training set data used to determine the Extended UNISAC segment area parameters, during the regression procedure. Hence in Chapter Six, solid-liquid equilibrium (SLE) measurements were presented, for the systems diosgenin/ estriol/ prednisolone/ hydrocortisone/ betulin and estrone, in approximately 10 diverse organic solvents, and water, at atmospheric pressure, within the temperature range 293.2-328.2 K, by employing combined digital thermal analysis, and thermal gravimetric analysis (DTA/TGA), to determine compositions at saturation.

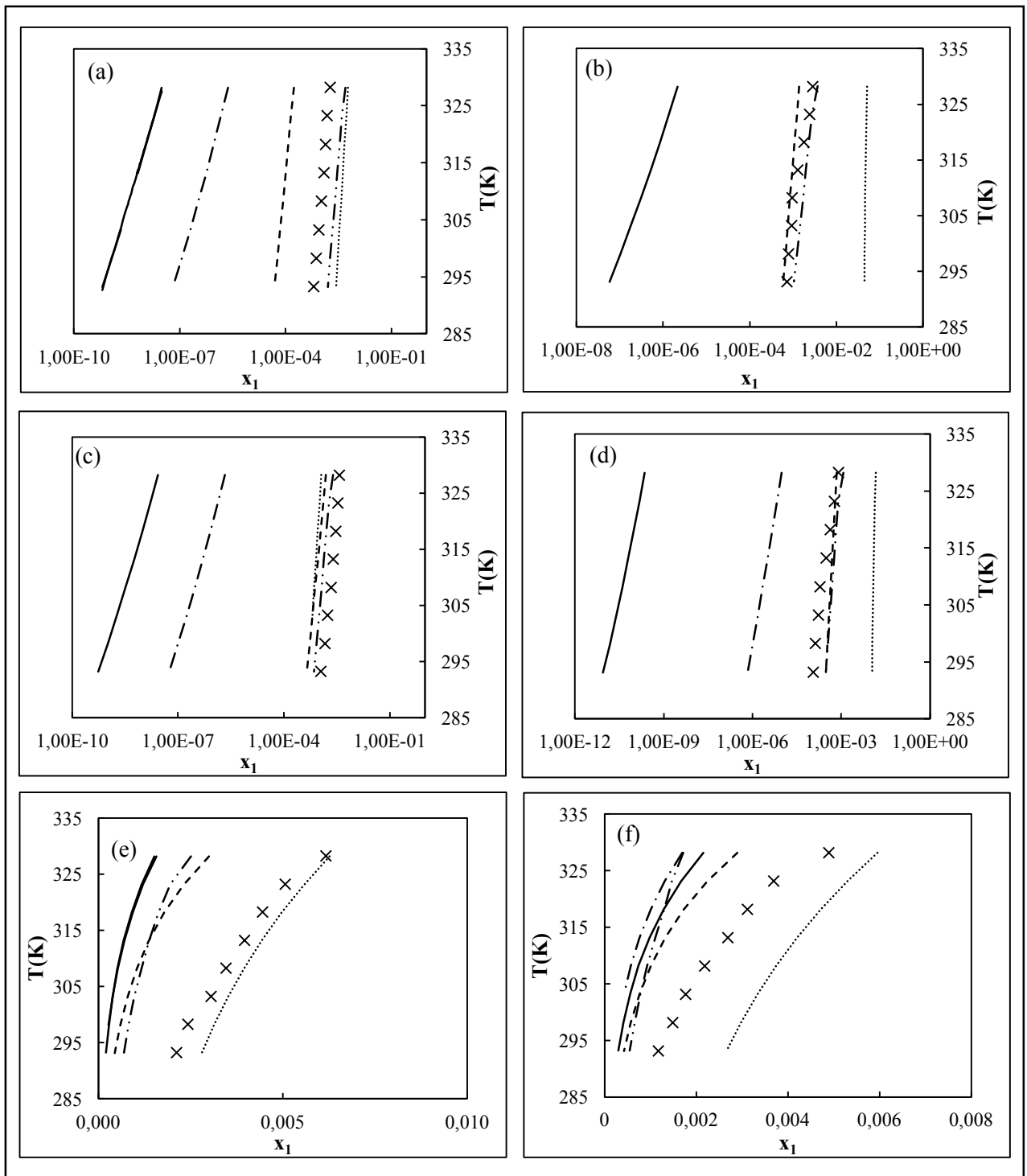
The experimentally determined solvent solubility, along with the Extended UNISAC model predictions, and other popular predictive methods, are presented in Figures 7.6-7.13, for the systems where the Extended UNISAC model can be applied. Hydrocortisone, and prednisolone molecules, cannot be group-fragmented by the original UNIFAC fragmentation scheme, which is also used in the Extended UNISAC combinatorial term. This issue can be overcome by using molecular based volumes and areas. The Extended UNISAC model proves to be competitive with other models, from the literature, and provides a superior prediction in many of these cases.



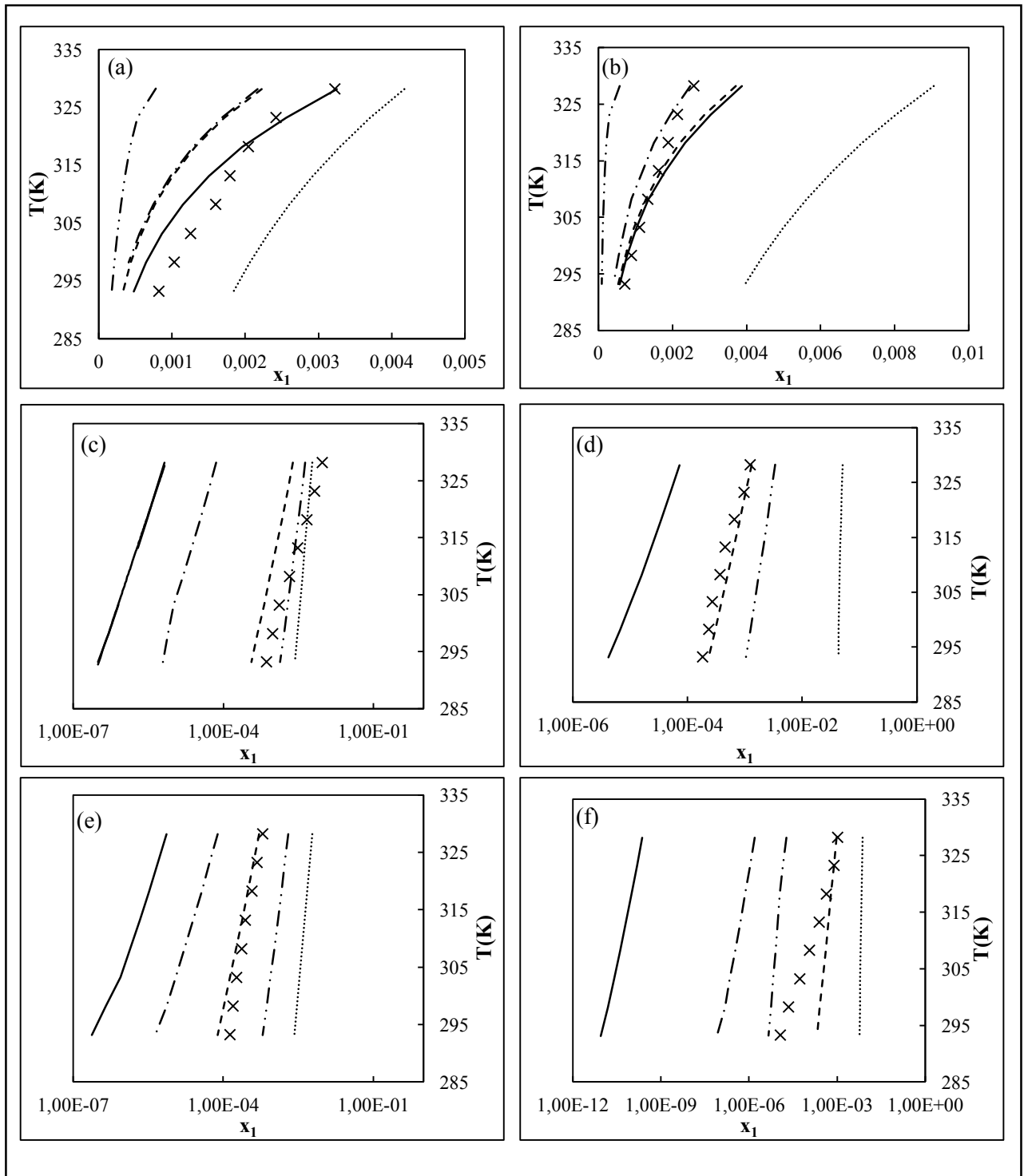
**Figure 7.6** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K.  $\times$ , Experimental Data (Moodley et al., 2015 (e));  $- \cdot -$ , Extended UNISAC;  $- - -$ , mod. UNIFAC (Dortmund);  $\cdots$ , UNIFAC;  $-$ , COSMO-RS(OL);  $- \cdot - \cdot -$ , COSMO-SAC. Systems: (a) betulin (1) + nonan-1-ol (2); (b) betulin (1) + octan-1-ol (2); (c) betulin (1) + pentan-1-ol (2); (d) betulin (1) + butan-2-ol; (e) betulin (1) + n-hexadecane (2); (f) betulin (1) + n-dodecane (2).



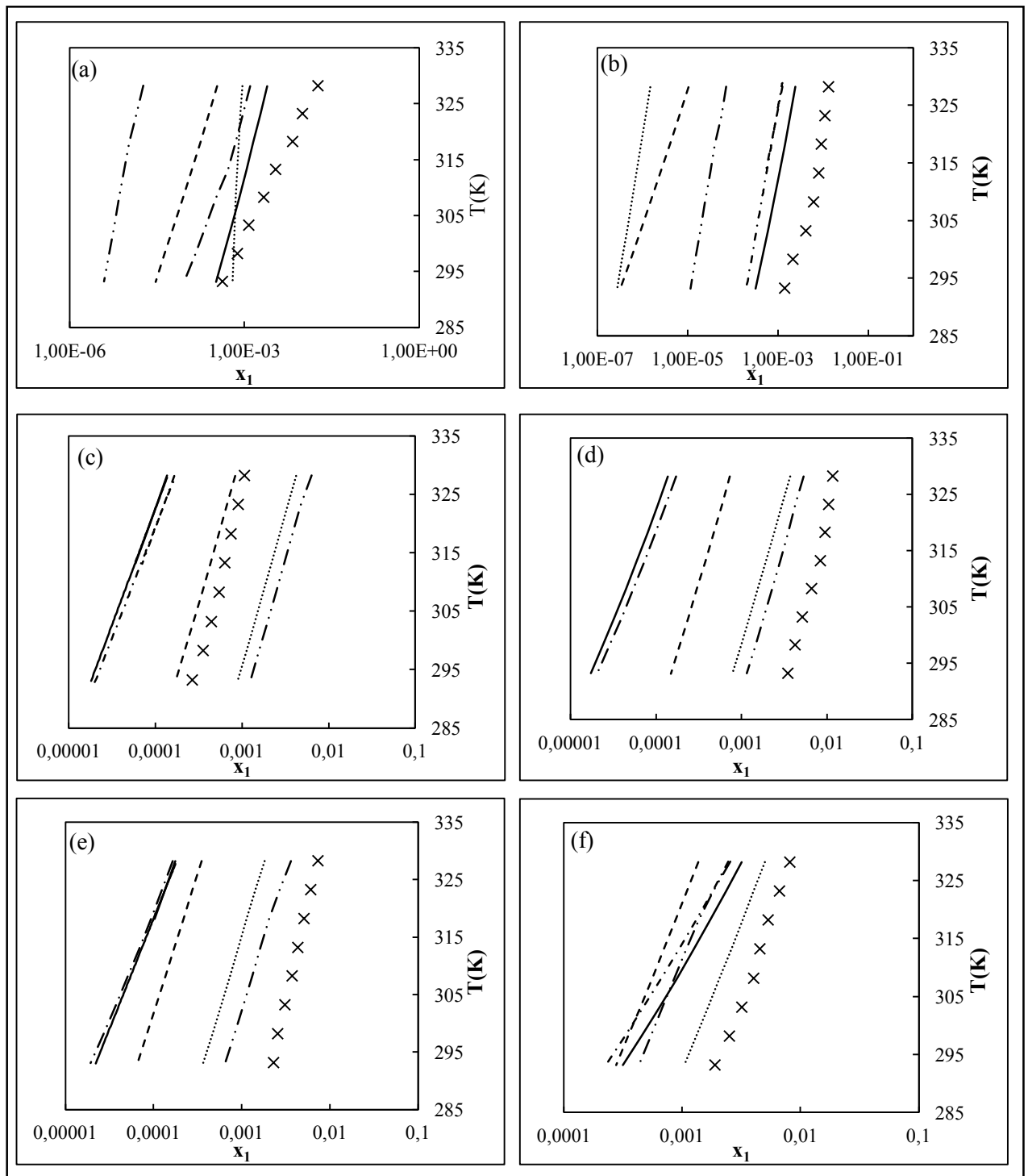
**Figure 7.7** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K.  $\times$ , Experimental Data (Moodley et al., 2015 (e, f));  $-\cdot-$ , Extended UNISAC;  $- - -$ , mod. UNIFAC (Dortmund);  $\dots$ , UNIFAC;  $-$ , COSMO-RS(OL);  $- \cdot -$ , COSMO-SAC. Systems: (a) betulin (1) + acetonitrile (2); (b) betulin (1) + water (2); (c) estriol (1) + nonan-1-ol (2); (d) estriol (1) + octan-1-ol (2); (e) estriol (1) + pentan-1-ol (2); (f) estriol (1) + butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.



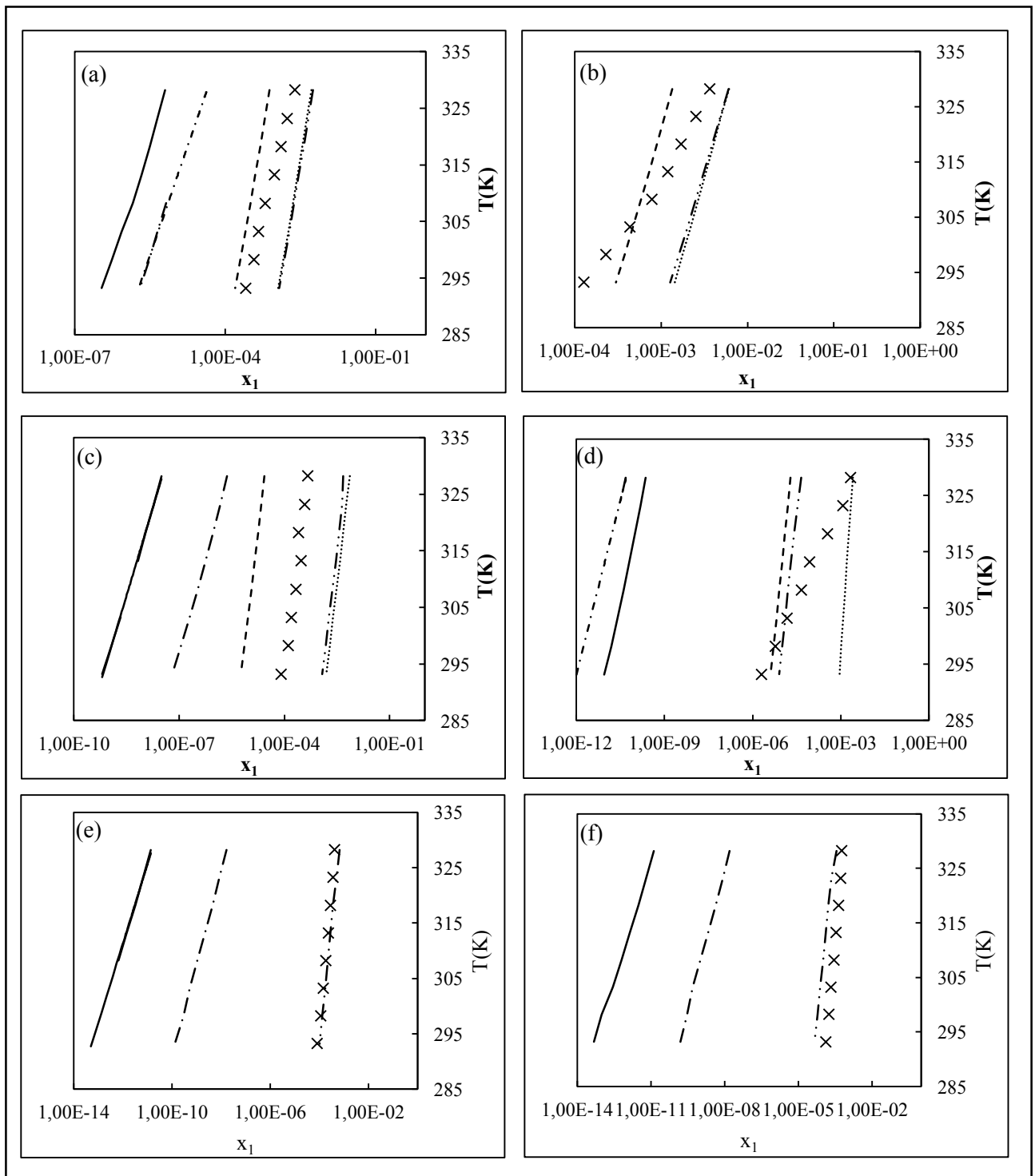
**Figure 7.8** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (e, f)); —·—, Extended UNISAC; ---, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) estriol (1) + n-octane (2); (b) estriol (1) + n-dodecane (2); (c) estriol (1) + n-hexadecane (2); (d) estrone (1) + water (2); (e) estrone (1) + nonan-1-ol (2); (f) estrone (1) + octan-1-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 7.9** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (e)); —·—, Extended UNISAC; ---, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); — —, COSMO-SAC. Systems: (a) estrone (1) + pentan-1-ol (2); (b) estrone (1) + butan-2-ol (2); (c) estrone (1) + n-hexadecane (2); (d) estrone (1) + n-dodecane (2); (e) estrone (1) + n-octane (2); (f) estrone (1) + water (2). Experimental data are represented as symbols, and model predictions are represented as lines.

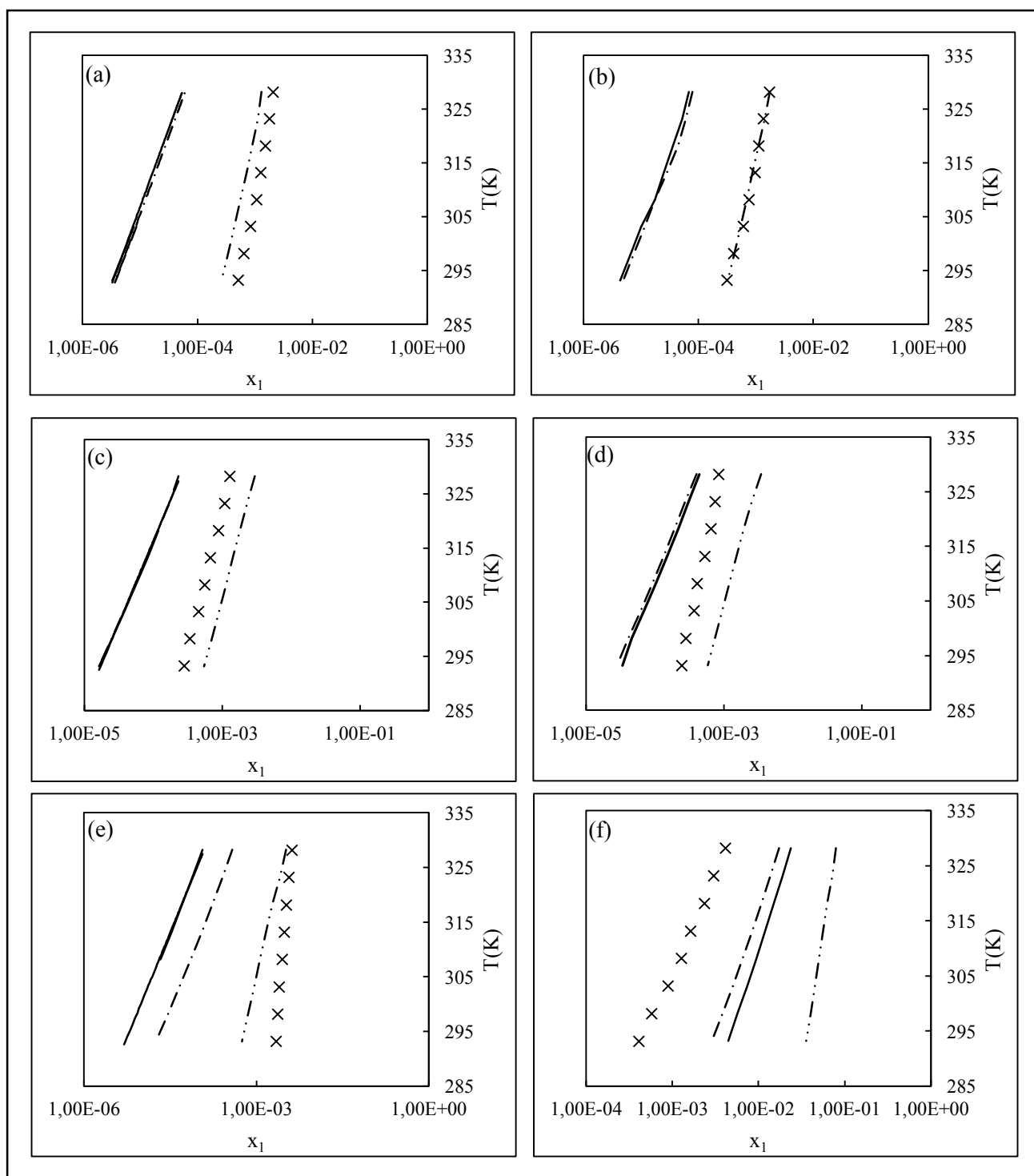


**Figure 7.10** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K.  $\times$ , Experimental Data (Moodley et al., 2015 (e, f));  $-\cdot-$ , Extended UNISAC;  $---$ , mod. UNIFAC (Dortmund);  $\dots$ , UNIFAC;  $—$ , COSMO-RS(OL);  $-\cdot-$ , COSMO-SAC. Systems: (a) estrone (1) + acetonitrile (2); (b) diosgenin (1) + acetonitrile (2); (c) diosgenin (1) + nonan-1-ol (2); (d) diosgenin (1) + octan-1-ol (2); (e) diosgenin (1) + pentan-1-ol (2); (f) diosgenin (1) + butan-2-ol (2). Experimental data are represented as symbols, and model predictions are represented as lines.

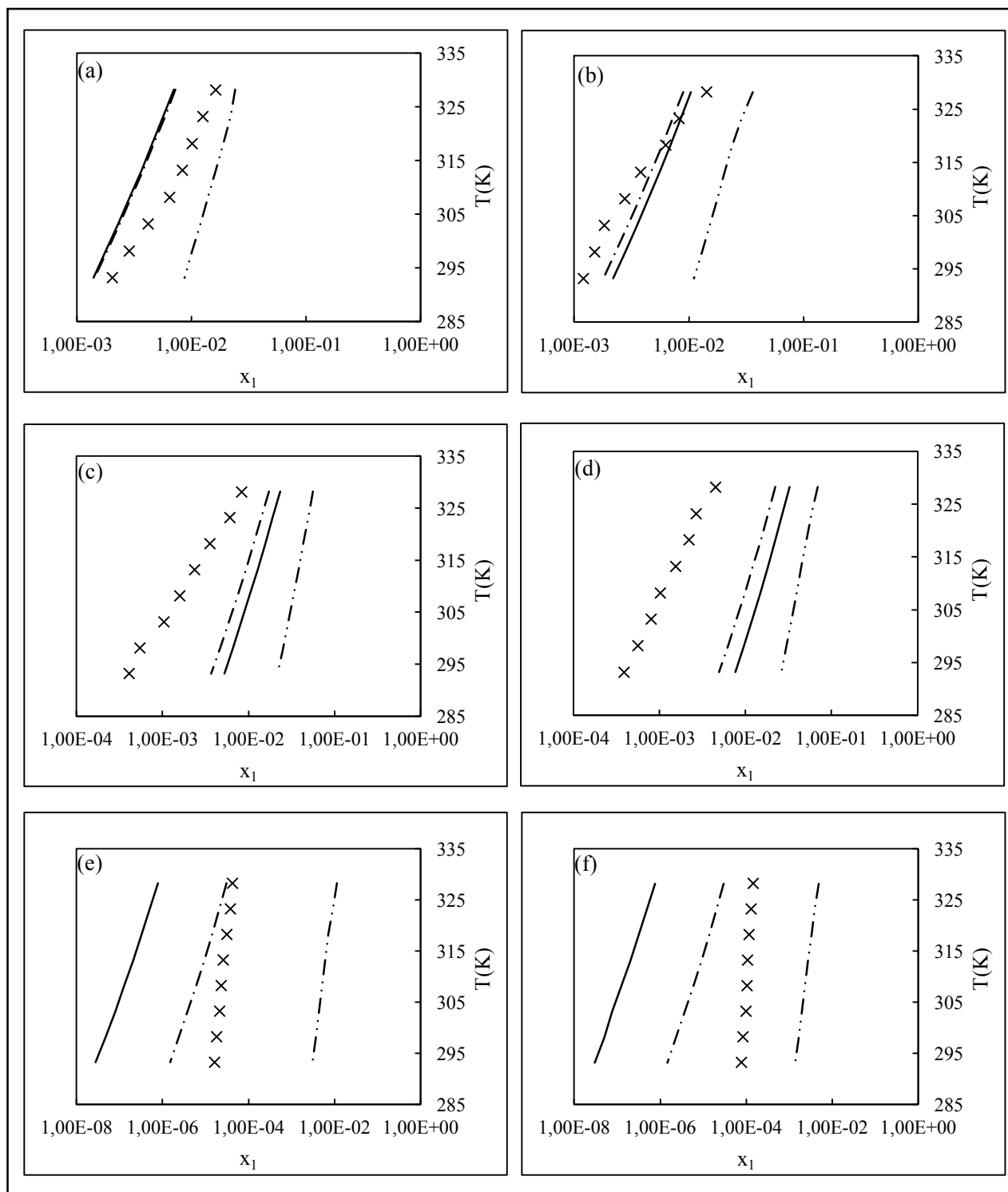


**Figure 7.11** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K. ×, Experimental Data (Moodley et al., 2015 (f, g)); —·—, Extended UNISAC; ---, mod. UNIFAC (Dortmund); ..., UNIFAC; —, COSMO-RS(OL); —·—, COSMO-SAC. Systems: (a) diosgenin (1) + n-hexadecane (2); (b) diosgenin (1) + n-dodecane (2); (c) diosgenin (1) + n-octane (2); (d) diosgenin (1) + water (2); (e) prednisolone (1) + n-dodecane (2); (f) prednisolone (1) + n-hexadecane (2). Experimental data are represented as symbols, and model predictions are represented as lines.





**Figure 7.12** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K.  $\times$ , Experimental Data (Moodley et al., 2015 (g));  $-\cdot-$ , Extended UNISAC;  $-$ , COSMO-RS(OL);  $- -$ , COSMO-SAC. Systems: (a) prednisolone (1) + nonan-1-ol (2); (b) prednisolone (1) + octan-1-ol (2); (c) prednisolone (1) + pentan-1-ol (2); (d) prednisolone (1) + butan-2-ol (2); (e) prednisolone (1) + acetonitrile (2); (f) hydrocortisone (1) + acetonitrile (2). Experimental data are represented as symbols, and model predictions are represented as lines.



**Figure 7.13** Experimental and predicted solid-liquid equilibrium data by different models between 293.2 K and 328.2 K.  $\times$ , Experimental Data (Moodley et al., 2015 (g));  $-\cdot-$ , Extended UNISAC;  $-$ , COSMO-RS(OL);  $\cdots$ , COSMO-SAC. Systems: (a) hydrocortisone (1) + nonan-1-ol (2); (b) hydrocortisone (1) + octan-1-ol (2); (c) hydrocortisone (1) + pentan-1-ol (2); (d) hydrocortisone (1) + butan-2-ol (2); (e) hydrocortisone (1) + n-dodecane (2); (f) hydrocortisone (1) + n-hexadecane (2). Experimental data are represented as symbols, and model predictions are represented as lines.

*References*

- Akaike, H., (1974), A new look at the statistical model identification. *IEEE Transactions on Information Technology in Biomedicine*, 19, pp.716–723.
- Berro, C., Rogalski, M. and Peneloux, A., (1982), Excess Gibbs energies and excess volumes of 1-butanol-n-hexane and 2-methyl-1-propanol-n-hexane binary systems. *Journal of Chemical and Engineering Data*, 27, pp.352–355.
- Bhargava, V., Fateen, S.E.K. and Bonilla-Petriciolet, A., (2013), Cuckoo Search: A new nature-inspired optimization method for phase equilibrium calculations. *Fluid Phase Equilibria*, 337, pp.191–200.
- Bondi, a., (1964), van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68, pp.441–451.
- Chen, C.C. and Crafts, P.A., (2006), Correlation and prediction of drug molecule solubility in mixed solvent systems with the Nonrandom Two-Liquid Segment Activity coefficient (NRTL-SAC) model. *Industrial and Engineering Chemistry Research*, 45, pp.4816–4824.
- Chen, C.-C. and Song, Y., (2004), Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Industrial and Engineering Chemistry Research*, 43, pp.8354–8362.
- Chen, G.-H., Wang, Q., Ma, Z.-M., Yan, X.-H. and Han, S.-J., (1995), Phase equilibria at superatmospheric pressures for systems containing halohydrocarbon, aromatic hydrocarbon, and alcohol. *Journal of Chemical and Engineering Data*, 40, pp.361–366.
- Doniec, A., Krauze, R., Michalowski, S., Serwinski, M., (1965), Vapor-Liquid Equilibrium for Carbon Tetrachloride-n-Butanol System. *Zesz.Nauk.Politech.Lodz.Chem.*, 16, pp.33–43.
- Fateen, S.-E.K., Bonilla-Petriciolet, A. and Rangaiah, G.P., (2012), Evaluation of Covariance Matrix Adaptation Evolution Strategy, Shuffled Complex Evolution and Firefly Algorithms for Phase Stability, Phase Equilibrium and Chemical Equilibrium Problems. *Chemical Engineering Research and Design*, 90, pp.2051–2071.
- Fischer, K. and Gmehling, J., (1994), P-x and  $\gamma$  infinite. Data for the Different Binary Butanol-Water Systems at 50.degree.C. *Journal of Chemical and Engineering Data*, 39, pp.309–315.
- Fredenslund, A., Jones, R.L. and Prausnitz, J.M., (1975), Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21, pp.1086–1099.
- Fuente, J.D. La, Santiago, J., Román, A., Dumitrache, C. and Casasanto, D., (2014), When you think about it, your past is in front of you: How culture shapes spatial conceptions of time. *Psychological Science*, 25, pp.1682–1690.
- Gandomi, A.H. and Alavi, A.H., (2012), Krill herd: A new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, 17, pp.4831–4845.

- Garriga, R., Martínez, S., Pérez, P. and Gracia, M., (2001), Isothermal (vapour + liquid) equilibrium at several temperatures of (1-chlorobutane + 1-butanol, or 2-methyl-2-propanol). *The Journal of Chemical Thermodynamics*, 33, pp.523–534.
- Garriga, R., Sánchez, F., Pérez, P. and Gracia, M., (1996), Isothermal Vapor–Liquid Equilibrium of Butanone + Butan-1-ol at Eight Temperatures between 278.15 and 323.15 K. *Journal of Chemical and Engineering Data*, 41, pp.451–454.
- Gharavi, M., James, K.C. and Sanders, L.M., (1983), Solubilities of mestanolone, methandienone, methyltestosterone, nandrolone and testosterone in homologous series of alkanes and alkanols. *International Journal of Pharmaceutics*, 14, pp.333–341.
- Grensemann, H. and Gmehling, J., (2005), Performance of a Conductor-Like Screening Model for Real Solvents Model in Comparison to Classical Group Contribution Methods. *Industrial and Engineering Chemistry Research*, 44, pp.1610–1624.
- Kováč, A., Svoboda, J., and Ondruš, I., (1985), Vapor-liquid-equilibrium of some binary-systems containing 1,1,2-trichloroethane. *Chemical Papers*, 39, pp.737–742.
- Lewandowski, C.M., (2015), *The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference*. PhD Thesis, Southern Illinois University Carbondale.
- Lin, S.-T. and Sandler, S.I., (2002), A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Industrial and Engineering Chemistry Research*, 41, pp.899–913.
- Lladosa, E., Montón, J.B., Burguet, M.C. and Muñoz, R., (2006), Isobaric vapor–liquid equilibria for the binary systems 1-propyl alcohol+dipropyl ether and 1-butyl alcohol+dibutyl ether at 20 and 101.3kPa. *Fluid Phase Equilibria*, 247, pp.47–53.
- Martin, A., Wu, P.L., Adjei, A., Mehdizadeh, M., James, K.C. and Metzler, C., (1982), Extended Hildebrand solubility approach: Testosterone and testosterone propionate in binary solvents. *Journal of Pharmaceutical Sciences*, 71, pp.1334–1340.
- McLaughlin, E. and Zainal, H.A., (1959), 177. The solubility behaviour of aromatic hydrocarbons in benzene. *Journal of the Chemical Society (Resumed)*, p.863-867.
- Michalski, H. Michalowski, S. Serwinski, M. Strumillo, C., (1961), Determination of vapor-liquid equilibrium for the system acetone - n-butanol. *Zesz. Nauk. - Politech. Lodz., Chem.*, 10, pp.73–84.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (a), Model Evaluation for the Prediction of Solubility of Active Pharmaceutical Ingredients (APIs). *Manuscript in preparation*.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (b), A Universal Segment Approach for the Prediction of the Activity Coefficient of Complex Pharmaceuticals in Non-electrolyte Solvents. *Fluid Phase Equilibria*, 396, pp.98–110.

- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (c), Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations. *Computers and Chemical Engineering*, 74, pp.75–88.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2015) (d), An Extended UNISAC model for the Prediction of Solubility of Complex Pharmaceutical Ingredients in Non-electrolyte Pure Solvents and Solvent Mixtures. *Manuscript in preparation*.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (e), Experimental solubility for betulin and estrone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . *Journal of Chemical Thermodynamics*, 98, pp. 42–50.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (f), Experimental solubility for diosgenin and estriol in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . Manuscript submitted for publication.
- Moodley, K., Rarey, J. and Ramjugernath, D., (2016) (g), Experimental solubility for prednisolone and hydrocortisone in various solvents within the temperature range  $T = (293.2 \text{ to } 328.2) \text{ K}$ . Manuscript submitted for publication.
- van Nhu, N. and Kohler, F., (1989), Excess properties of mixtures of polar components and hydrocarbons of varying local polarisability. *Fluid Phase Equilibria*, 50, pp.267–296.
- Ott, J.B., Marsh, K.N. and Richards, A.E., (1981), Excess enthalpies, excess Gibbs free energies, and excess volumes for (di-n-butyl ether + benzene) and excess Gibbs free energies and excess volumes for (di-n-butyl ether + tetrachloromethane) at 298.15 and 308.15 K. *The Journal of Chemical Thermodynamics*, 13, pp.447–455.
- Wang, G., Guo, L., Gandomi, A.H., Cao, L., Alavi, A.H., Duan, H. and Li, J., (2013), Lévy-Flight Krill Herd Algorithm. *Mathematical Problems in Engineering*, 2013, pp.1–14.
- Weidlich, U. and Gmehling, J., (1987), A modified UNIFAC Model. 1. Prediction of VLE,  $hE$ , and  $y^\infty$ . *Industrial and Engineering Chemistry Research*, 26, pp.1372–1381.
- Wu, H.S. and Sandler, S.I., (1988), Vapor-liquid equilibria for binary mixtures of butyl ether with 2-furaldehyde and with 2-, 3-, and 4-heptanone. *Journal of Chemical and Engineering Data*, 33, pp.316–321.

## CHAPTER EIGHT

### Conclusions

- The segment contribution concept of interactions, between molecular surfaces, can be analogously applied to interactions, between functional groups.
- A reduction in the group-group binary interaction parameter matrix does not drastically affect the predictive capability of the group contribution approach, when considering solid-liquid equilibrium calculations.
- The Extended UNISAC model is capable of scaling binary interactions, between adjacent surface segments, when normalized. The model is sensitive to surface segment area parameters, and a global optimization technique provides a superior estimation of model parameters.
- The recently developed Krill Herd Algorithm was selected for global optimization in Extended UNISAC model training, as it was found to outperform, or match, the performance of other stochastic algorithms, in the literature (Genetic Algorithm (GA), Covariant Matrix Adaptation Evaluation Strategy (CMA-ES), Shuffled Complex Evolution (SCE), Firefly Algorithm (FA), Cuckoo Search (CS), and Modified Cuckoo Search (MCS)), in test cases of phase equilibrium calculations.
- The Extended UNISAC model provides a means of performing qualitative predictions of solubility for complex pharmaceutical components, and is competitive with other popular models, from the literature (UNIFAC, modified UNIFAC (Dortmund), COSMO-RS(OL), and COSMO-SAC), in terms of range of applicability, and degree of accuracy.
- The Extended UNISAC model performance was also tested, on a set of previously unmeasured data of complex steroidal triterpenes, in various solvents. The performance of the model was again found to be competitive with the popular models from the literature.

## CHAPTER NINE

### Future Work and Recommendations

A challenge faced in the development of the Extended UNISAC model, is the calculation and usage of group parameters (group volumes, areas and segment areas), that simultaneously represent the behaviour of the conceptual group, in all phases. In the UNISAC model versions presented here, UNIFAC volume and area parameters ( $R$  and  $Q$ ) were used, since they are generalized group contribution values, and are not model specific, such as the case in modified UNIFAC (Dortmund), where  $R$  and  $Q$  are fit simultaneously with group interaction parameters.

When considering energetic contributions to activity (that are governed by the segment area parameters in the case of the Extended UNISAC model), it was found that some groups, for instance, COOH and OH, exhibit varying degrees of hydrogen bonding, depending not only on the other groups present in the mixture, but also on the phase of the component. This is due to changes in molecular arrangements, such as dimerization effects, and leads to the possible requirement of liquid or solid state specific segment area parameters for the Extended UNISAC model. For example, the COOH and Cl segment area parameters, presented earlier, are specific to the solid phase only. Although this would increase the databank of component parameters required for the precise application of Extended UNISAC, the number of required parameters will still be much lower than those of current popular predictive methods such as UNIFAC, and modified UNIFAC (Dortmund), and will improve the accuracy of the method.

When considering size/shape contributions to activity, varying due to increasing solute molecule size, and by default, state, many authors recommend considering the contribution of free-volume and cavity formation (Kontogeorgis et al. (1993), Bekker et al. (1986), Pappa et al. (1997), Wibawa et al. (2005), Moller et al. (2014)). To provide a preliminary test for the effect of free-volume considerations, on the predictions by the Extended UNISAC model, the reportedly-improved combinatorial expression of Moller et al. (2014) for alkane solvents, was incorporated into the Extended UNISAC model, for application at finite concentrations, and tested on systems of alkane solvents only. The revision and development of the combinatorial expression of Moller et al. (2014) is beyond the scope of this work, however the reader is referred to the original publication for further details. A brief description follows.

The Moller et al. (2014) ( $\ln\gamma_{i,M}^{comb,\infty}$ ) combinatorial term at infinite dilution in alkane solvents is composed of three parts, including an empirical correction factor.

$$\begin{aligned} \ln\gamma_{i,M}^{comb,\infty} &= \ln\gamma_i^{comb\,FV,\infty} + \ln\gamma_i^{cav,\infty} + \ln\gamma_i^{cav-corr,\infty} \\ &= 1 - \frac{V_i^{FV}}{V_j^{FV}} + \ln \frac{V_i^{FV}}{V_j^{FV}} - 5q'_i \left(1 - \frac{r'_i/r'_j}{q'_i/q'_j} + \ln \left( \frac{r'_i/r'_j}{q'_i/q'_j} \right)\right) + \frac{V_i^{FV}}{V_j^{FV}} \\ &\quad - \frac{V_i^{FV'}}{V_j^{FV'}} + \ln (1.15 - 0.15 \times \exp[-0.5(q'_i - q'_j)^2 - 0.5(r'_i - r'_j)^2]) \end{aligned} \quad (9.1)$$

Where 
$$V_i^{FV} = V_i - V_i^* \quad (9.2)$$

And 
$$V_i^{FV'} = (V_i)^{2/3} - (V_i^*)^{2/3} \quad (9.3)$$

Where  $V_i$  is the molar volume of component  $i$  in the mixture and  $V_i^*$  is the van der Waals volume of component  $i$  and  $r'_i$  and  $q'_i$  are the UNIQUAC (Anderson and Prausnitz, 1978) volume and area parameters.

In Figures 7.14 and 7.15, the test cases of mestanolone and testosterone in alkane solvents, of increasing chain length, are presented, comparing the experimental and calculated activity coefficients, using the combinatorial expression of Flory-Huggins (Flory 1941, Huggins 1941), with Staverman-Guggenheim (Staverman, 1950, Guggenheim 1952), and Moller et al. (2014), at finite concentrations. It is evident from these predictions that the combinatorial expression of Moller et al. (2014) does not compensate for cavity formation, when extended to finite concentrations in the two systems presented here. This is probably due to the empirically-derived cavity correction term ( $\ln\gamma_i^{cav-corr,\infty}$ ), tuned specifically, by infinite, dilution activity coefficient data.

It may prove necessary that a composition-dependent, empirical, cavity-correction term be included, in order to improve the effect of the correction of the cavity formation combinatorial expression at finite concentrations. In fact, the Flory-Huggins combinatorial expression, with Staverman-Guggenheim correction terms, provides the more precise prediction, in the case of



the Extended UNISAC model. Although the predictions are not quantitative in the two cases presented, this exercise does show the significant effect, of the selected combinatorial expression, on the predicted activity coefficient, in systems comprised of molecules of vastly different sizes and shapes.

In future work, a combinatorial expression, more suited to API's and other systems of vastly different molecule sizes, may be incorporated into the Extended UNISAC model.

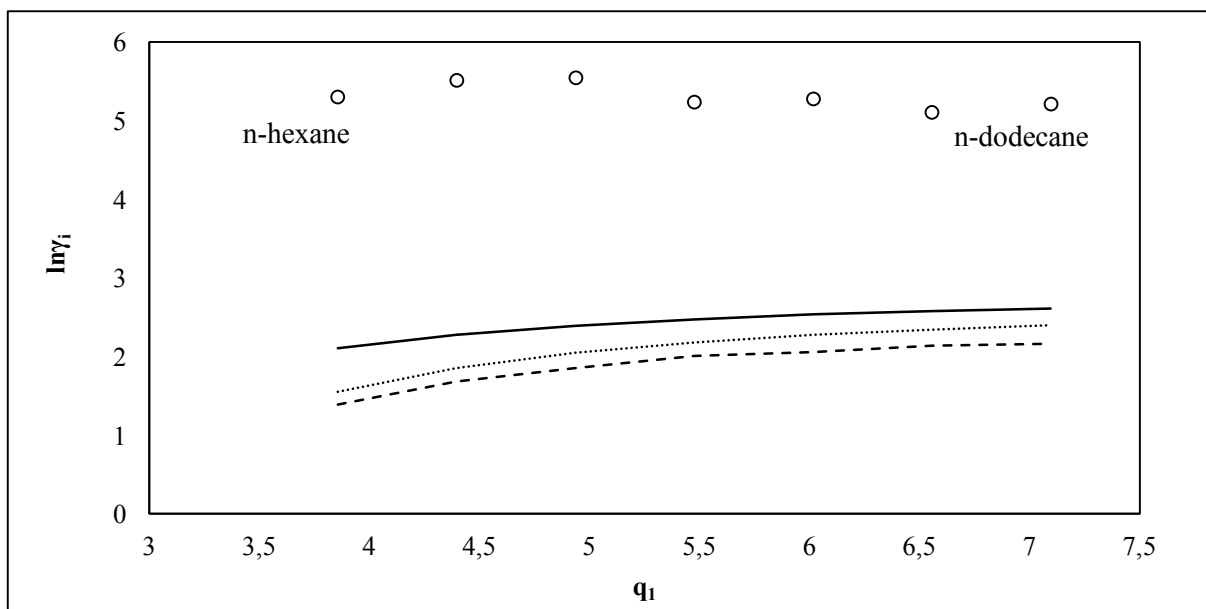


Figure 9.1 Natural log of the activity coefficient for mestanolone-alkane systems at 298 K. ○ – Experimental data (Gharavi et al, 1983), Extended UNISAC residual with combinatorial expression of; —, Flory-Huggins with Staverman-Guggenheim correction, ···, Flory-Huggins, - - - Moller et al. (2014). Experimental data are represented as symbols, and model predictions are represented as lines.

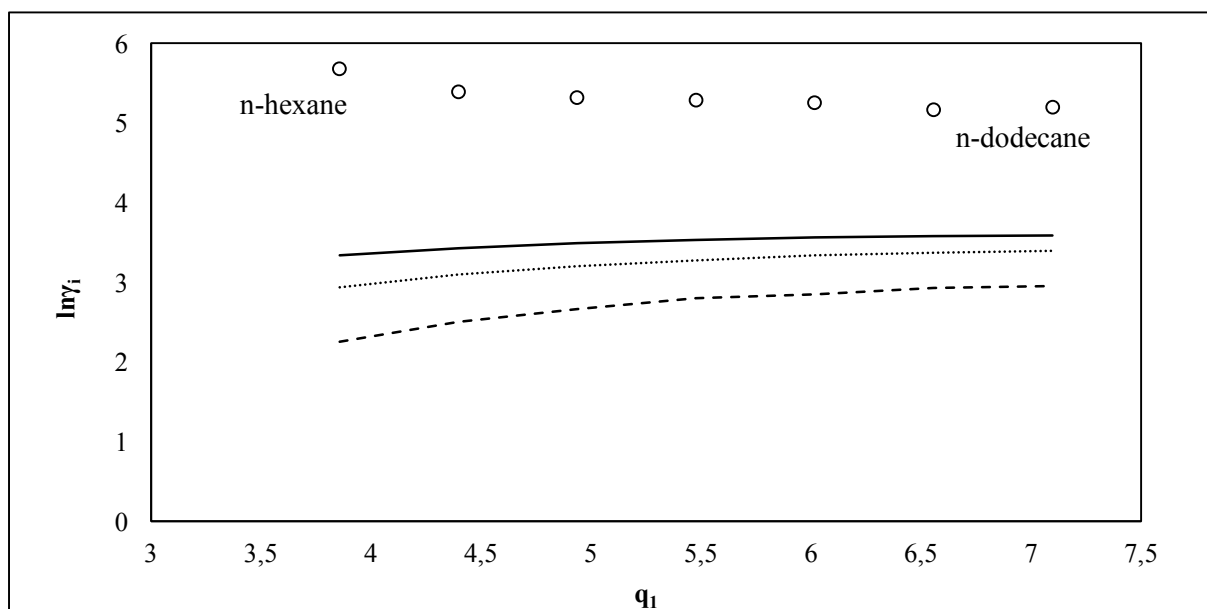


Figure 9.2 Natural log of the activity coefficient for testosterone-alkane systems at 298 K. ○ – Experimental data (Gharavi et al, 1983), Extended UNISAC residual with combinatorial expression of; —, Flory-Huggins with Staverman-Guggenheim correction, ···, Flory-Huggins, - - - Moller et al. (2014). Experimental data are represented as symbols, and model predictions are represented as lines.

As previously mentioned, UNIFAC VLE predictions were treated as pseudo-experimental data, as part of the training set, in order to determine the Extended UNISAC model parameters. This decision does, however, restrict the performance of the Extended UNISAC model, as the model is trained to behave similarly to UNIFAC. Consequently, for cases where UNIFAC offers a poor representation of the experimental behaviour, Extended UNISAC is predisposed to also provide a poor prediction. In many instances however, Extended UNISAC does outperform the UNIFAC model. This is most probably due to the suspected “*overfitting*” of the UNIFAC model, due to the numerous group contribution parameters being accounted for in a mixture, which may not always be necessary in the cases of solid-liquid equilibrium.

It is therefore recommended for future work, that the Extended UNISAC model be trained only with purely experimental data, of a high quality, in order to improve the predictive capability of the model. This can include VLE, LLE, SLE and excess enthalpy experimental data. After this rigorous training, it is recommended that all outstanding Extended UNISAC model parameters be determined by data regression.

This work intends to provide the framework for the universal-functional segment approach to solubility prediction, and provides constructive evidence of the strength and applicability of the concepts involved. Further research and development of the model is essential to improve its range of application and accuracy.

*References*

- Bekker, A.Y., Knox, D.E. and Sund, S.E., (1987), Prediction of solvent activities using the UNIFAC-FV model. *Journal of Solution Chemistry*, 16, pp.635–639.
- Berro, C., Rogalski, M. and Peneloux, A., (1982), Excess Gibbs energies and excess volumes of 1-butanol-n-hexane and 2-methyl-1-propanol-n-hexane binary systems. *Journal of Chemical and Engineering Data*, 27, pp.352–355.
- Flory, P.J., (1942), Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics*, 10, p.51.
- Guggenheim, E.A., (1952), *Mixtures*, Oxford, U.K.
- Huggins, M.L., (1941), Solutions of long chain compounds. *The Journal of Chemical Physics*, 9, p.440.
- Kontogeorgis, G.M., Fredenslund, A. and Tassios, D.P., (1993), Simple activity coefficient model for the prediction of solvent activities in polymer solutions. *Industrial and Engineering Chemistry Research*, 32, pp.362–372.
- Moller, B., Rarey, J. and Ramjugernath, D., (2008), Estimation of the vapour pressure of non-electrolyte organic compounds via group contributions and group interactions. *Journal of Molecular Liquids*, 143, pp.52–63.
- Pappa, G.D., Kontogeorgis, G.M. and Tassios, D.P., (1997), Prediction of ternary liquid–liquid equilibria in polymer–solvent–solvent systems. *Industrial and Engineering Chemistry Research*, 36, pp.5461–5466.
- Staverman, A.J., (1950), The entropy of high polymer solutions. Generalization of formulae. *Recueil des Travaux Chimiques des Pays-Bas*, 69, pp.163–174.
- Wibawa, G., Takishima, S., Sato, Y. and Masuoka, H., (2005), An Improved Prediction Result of Entropic-FV Model for Vapor – Liquid Equilibria of Solvent – Polymer Systems. *Journal of Applied Polymer Science*, 97, pp.1145–1153.

## Appendix A

### Supplementary data for Application of the bio-inspired Krill Herd optimization technique to phase equilibrium calculations

#### *A1. Derivations*

The optimization can be converted into an unconstrained optimization if, ( $\zeta_{ij} \in [0,1]$ ) decision variables for the optimization algorithm are used, that describe the fraction of component  $i$  present in each phase  $j$ . That is for instance,  $\zeta_{11}$  is the fraction of the total number of moles of component 1 present in phase 1 and  $\zeta_{12}$  is the fraction of the remaining number of moles (the total number of moles of component 1 less the number of moles of component 1 in phase 1) of component 1 present in phase 2. By introducing this term it is assured that all candidate solutions of the number of moles of each component in a phase will be physically realistic and hence computation time is decreased. The decision variables impose the mass balance restriction, eliminating the need for a constrained optimization. The number of moles of each component in each of the phases can then be given in general by:

$$n_{ij} = \zeta_{ij} \left( z_i n_T - \sum_{k=1}^{j-1} n_{ik} \right) \text{ Where } i = 1, \dots, c \quad (\text{A1})$$

If the first phase is considered, then  $\sum_{k=1}^{j-1} n_{ik} = 0$  and

$$n_{i1} = \zeta_{i1} z_i n_T \text{ Where } i = 1, \dots, c \text{ and } j = 2, \dots, \pi - 1 \quad (\text{A2})$$

Since  $\zeta_{ij} = 1$  for the final phase the number of moles of component  $i$  in the final phase is given

:

$$n_{i\pi} = z_i n_T - \sum_{k=1}^{k=\pi-1} n_{ik} \quad \text{Where } i = 1, \dots, c \quad (\text{A3})$$

The summation over all phases gives the total number of moles of component  $i$  in the mixture and is given by:

$$\sum_{j=1}^{j=\pi} n_{ij} = [\zeta_{i1} z_i n_T] + [\zeta_{ij} (z_i n_T - \sum_{k=1}^{k=j-1} n_{ik})] + [z_i n_T - \sum_{k=1}^{k=\pi-1} n_{ik}] \quad (\text{A4})$$

Where the first parenthesis represents the number of moles of component  $i$  in the first phase, the second parenthesis represents the number of moles of component  $i$  in phases 2 to  $\pi - 1$  and the third parenthesis represents the number of moles of component  $i$  in the last phase.

Since

$$\sum_{k=1}^{k=\pi-1} n_{ik} = \zeta_{i1} z_i n_T + \zeta_{ij} (z_i n_T - \sum_{k=1}^{k=j-1} n_{ik}) \quad (\text{A5})$$

Substitution of equation (A5) into equation (A4) yields:

$$\sum_{j=1}^{\pi} n_{ij} = z_i n_T \quad (\text{A6})$$

Which is equivalent to equation (4.17).

The minimization constraint on  $g$  imposed by the mass balance is given by equation (4.18) as:

$$0 \leq n_{ij} \leq z_i n_T \quad \text{Where } i = 1, \dots, c \text{ and } j = 1, \dots, \pi - 1 \quad (\text{A7})$$

For a component  $i$  the limiting constraints on equation (A1) yield the following:

For the case of  $\zeta_{ij} = 0$

$$n_{ij} = 0 \quad (\text{A8})$$

And for the case  $\zeta_{ij} = 1$

$$n_{ij} = \left( z_i n_T - \sum_k^{j-1} n_{ik} \right) \quad (\text{A9})$$

Since  $0 \leq \sum_k^{j-1} n_{ik} \leq z_i n_T$  by definition, then it is always true that  $n_{ij} \leq z_i n_T$

Hence the minimization constraint on  $g$  given by:

$$0 \leq \zeta_{ij} \leq 1 \quad (\text{A10})$$

is equivalent to equation (A7).

## Appendix B

### Extended UNISAC: Supplementary data

*B1. Example Solid-liquid Equilibrium example system calculations and plots for binary and ternary systems:*

**Example 1.** aspirin (2-(acetyloxy)benzoic acid) (1) + ethyl acetate at T= 298.15 K and  $x_1 = 0.0448$

The activity coefficient is calculated using equation (1.49)

$$\ln\gamma_i = \ln\gamma_i^{comb} + \ln\gamma_i^{res} \quad (1.49)$$

For the combinatorial (identical to the original UNIFAC model):

UNIFAC Sub groups	aspirin	ethyl acetate	R	Q
	$v_{j,i}$			
ACH	4	0	0.5313	0.4
AC	2	0	0.3652	0.12
CC3COO	1	1	1.9031	1.7280
COOH	1	0	1.3013	1.2240
CH3	0	1	0.9011	0.848
CH2	0	1	0.6744	0.54

Substituting into equations (3.13) to (3.16)

$$r_i = \sum_j v_{j,i} R_j \quad (3.15)$$

$$r_1 = 6.06, r_2 = 3.479$$

$$q_i = \sum_j v_{j,i} Q_j \quad (3.16)$$

$$q_1 = 4.792, q_2 = 3.116$$

$$V_i = \frac{r_i}{\sum_j x_j r_j} \quad (3.13)$$

$$V_1 = 1.686, V_2 = 0.968$$



$$F_i = \frac{q_i}{\sum_j x_j q_j} \tag{3.14}$$

$$F_1 = 1.502, F_2 = 0.976$$

And substituting into equation (5.2):

$$\ln \gamma_i^{comb} = 1 - V_i + \ln(V_i) - 5q_i \left[ 1 - \frac{V_i}{F_i} + \ln\left(\frac{V_i}{F_i}\right) \right] \tag{5.2}$$

$$\ln \gamma_1^{comb} = 0.0033457, \ln \gamma_2^{comb} = 0.0000853$$

For the Extended UNISAC residual:

Ex. UNISAC groups	aspirin	ethyl acetate								
				$v_{l,i}$						
CH3	1	2	$\zeta_{k,l}$	0.8480	0	0	0	0	0	
CH	4	0		0.2012	0.0162	0.0198	0.007	0.058	0	0.0041
AC	1	0		0.2280	0	0	0	0	0	0
AC-	1	0		0	0	0	0	0	0	0
COOH	1	0		0.008	0.1985	0.6914	0	0	0	0
CCOO	1	1		0.008	0.1985	0.6914	0	0	0	0
CH2	0	1		0.5400	0	0	0	0	0	0

$$\Omega_{k,i} = \sum_{l=1}^N v_{l,i} \zeta_{k,l} \tag{3.18}$$

$$\begin{aligned} \Omega_{1,1} &= 1 \times 0.8480 + 4 \times 0.2012 + 1 \times 0.2280 + 1 \times 0 + 1 \times 0.008 + 1 \times 0.008 + 0 \times 0.5400 \\ &= 1.8970 \end{aligned}$$

$$\begin{aligned} \Omega_{1,1} &= 2 \times 0.8480 + 0 \times 0.2012 + 0 \times 0.2280 + 0 \times 0 + 0 \times 0.008 + 1 \times 0.008 + 1 \times 0.5400 \\ &= 2.244 \end{aligned}$$

Similarly for 7 segments in the groups:

$k$	$\Omega_{k,1}$	$\Omega_{k,2}$
1	1.897	2.244
2	0.462	0.199
3	1.462	0.691
4	0.028	0
5	0.232	0
6	0	0
7	0.016	0

$\sum_{k=1}^7 \Omega_{k,i}$	4.097	3.134
-----------------------------	-------	-------

To calculate the segment area fraction in the pure components  $\theta_{k,i}$ :

$$\theta_{k,i} = \frac{\Omega_{k,i}}{\sum_{k=1}^N \Omega_{k,i}} \tag{3.20}$$

$$\theta_{1,1} = \frac{1.897}{4.097} = 0.463, \theta_{1,2} = \frac{2.244}{3.134} = 0.716$$

Similarly for all segments:

$k$	$\theta_{k,1}$	$\theta_{k,2}$
1	0.463	0.716
2	0.113	0.063
3	0.357	0.221
4	6.834E-3	0
5	0.057	0
6	0	0
7	4.003E-3	0

To calculate the psi values for binary interaction parameters:

$$\psi_{k,m} = \exp\left(-\frac{a_{k,m}}{T}\right)$$

$a$	$m$	1	2	3	4	5	6	7
$k$								
1		0.00	597.00	104.30	24.90	476.40	1318.00	0.00
2		24.82	0.00	-54.86	-15.62	-287.50	242.80	0.00
3		-78.45	491.95	0.00	51.90	372.00	1201.00	0.00
4		36.70	74.04	-30.10	0.00	552.10	826.76	0.00
5		26.76	481.70	-39.20	-354.55	0.00	472.50	0.00
6		300.00	112.60	497.54	353.68	-195.40	0.00	0.00
7		0.00	0.00	0.00	0.00	0.00	0.00	0.00

$$\psi_{1,2} = \exp\left(-\frac{597.00}{298.15}\right) = 0.135$$

$$\psi_{2,1} = \exp\left(-\frac{24.82}{298.15}\right) = 0.92$$

$\psi_{k,m}$	$m$	1	2	3	4	5	6	7
$k$								
1		1	0.135	0.705	0.92	0.202	0.012	1
2		0.92	1	1.202	1.054	2.623	0.443	1
3		1.301	0.192	1	0.84	0.287	0.018	1
4		0.884	0.78	1.106	1	0.157	0.062	1
5		0.914	0.199	1.141	3.284	1	0.205	1
6		0.366	0.685	0.188	0.305	1.926	1	1
7		1	1	1	1	1	1	1

Similarly for all segments:

Now to calculate  $\ln\Gamma_{k,i}$ :

$$\ln\Gamma_{k,i} = 1 - \ln\left(\sum_{m=1}^N \theta_{m,i} \psi_{m,k}\right) - \sum_{m=1}^N \frac{\theta_{m,i} \psi_{k,m}}{\sum_{n=1}^N \theta_{n,i} \psi_{n,m}} \quad (3.19)$$

$$\begin{aligned} \sum_{m=1}^7 \theta_{m,1} \psi_{m,1} &= 0.463 \times 1 + 0.113 \times 0.92 + 0.357 \times 1.301 + 6.83E - 3 \times 0.884 + 0.057 \times 0.914 + 0 \times 0.366 \\ &\quad + 4.00E - 3 \times 1 \\ &= 1.093559 \end{aligned}$$

$$\begin{aligned} \sum_{m=1}^7 \theta_{m,1} \psi_{m,1} &= 0.716 \times 1 + 0.063 \times 0.92 + 0.221 \times 1.301 + 0 \times 0.884 + 0 \times 0.914 + 0 \times 0.366 + 0 \times 1 \\ &= 1.061481 \end{aligned}$$

$$\begin{aligned} &\sum_{m=1}^7 \frac{\theta_{m,1} \psi_{1,m}}{\sum_{n=1}^N \theta_{n,1} \psi_{n,m}} \\ &= \left[ \frac{0.463 \times 1}{(0.463 \times 1 + 0.113 \times 0.92 + 0.357 \times 1.301 + 6.83E - 3 \times 0.884 + 0.057 \times 0.914 + 0 \times 0.366 + 4.00E - 3 \times 1)} \right] \\ &+ \left[ \frac{0.135 \times 1}{(0.463 \times 0.135 + 0.113 \times 1 + 0.357 \times 0.192 + 6.83E - 3 \times 0.78 + 0.057 \times 0.199 + 0 \times 0.685 + 4.00E - 3 \times 1)} \right] \dots \\ &= 0.793 \end{aligned}$$

Similarly for all segments

$\sum_{m=1}^7 \theta_{m,1} \psi_{m,1}$	$\sum_{m=1}^7 \theta_{m,2} \psi_{m,2}$	$\sum_{m=1}^N \frac{\theta_{m,1} \psi_{k,m}}{\sum_{n=1}^N \theta_{n,1} \psi_{n,m}}$	$\sum_{m=1}^N \frac{\theta_{m,2} \psi_{k,m}}{\sum_{n=1}^N \theta_{n,2} \psi_{n,m}}$
1.093	1.061	0.793	0.911
0.264	0.202	1.575	1.265
0.895	0.801	1.071	1.213
1.041	0.911	1.175	1.145
0.553	0.374	1.055	0.993
0.078	0.041	0.725	0.513
1	1	1.362	1.263

$\ln\Gamma_k$  is then evaluated:

$$\begin{aligned} \ln\Gamma_{1,1} &= 1 - \ln(1.093) - 0.793 \\ &= 0.118074 \end{aligned}$$

$$\begin{aligned} \ln\Gamma_{1,2} &= 1 - \ln(1.061) - 0.911 \\ &= 0.03 \end{aligned}$$

Similarly for all segments:

$\ln\Gamma_{1,1}$	$\ln\Gamma_{1,2}$
0.118	0.03
0.756	1.333
0.04	8.26E-03
-0.216	-0.052
0.537	0.99
2.827	3.691
-0.362	-0.263

For the mixture  $\ln\Gamma_k$  a similar procedure is followed. Firstly, the mixture segment areas are calculated:

$$\sum_{i=1}^I \Omega_{m,i} x_i = 0.0448 \times 1.897 + 0.9552 \times 2.244 = 2.228$$

Similarly for all segments:

$m$	$\sum_{i=1}^I \Omega_{m,i} x_i$	$\theta_m$
1	2.228	0.701
2	0.21	0.066

3	0.726	0.228
4	1.25E-03	3.95E-04
5	0.01	3.27E-03
6	0	0
7	7.35E-04	2.31E-04
$\sum_{m=1}^N \sum_{i=1}^I \Omega_{m,i} x_i$	3.177	

$\theta_m$  can thus be calculated using equation (3.22) and is given in the table above.

$$\theta_m = \frac{\sum_{i=1}^I \Omega_{m,i} x_i}{\sum_{m=1}^N \sum_{i=1}^I \Omega_{m,i} x_i} \tag{3.22}$$

$\sum_{m=1}^N \theta_m \psi_{m,k}$ ,  $\sum_{m=1}^N \frac{\theta_m \psi_{k,m}}{\sum_{n=1}^N \theta_n \psi_{n,m}}$ , and  $\ln \Gamma_k$  are calculated for the mixture in a similar manner to the pure components using equations (3.21):

$$\ln \Gamma_k = 1 - \ln \left( \sum_{m=1}^N \theta_m \psi_{m,k} \right) - \sum_{m=1}^N \frac{\theta_m \psi_{k,m}}{\sum_{n=1}^N \theta_n \psi_{n,m}} \tag{3.21}$$

And yields

$\sum_{m=1}^N \theta_m \psi_{m,k}$	$\sum_{m=1}^N \frac{\theta_m \psi_{k,m}}{\sum_{n=1}^N \theta_n \psi_{n,m}}$	$\ln \Gamma_k$
1.063	0.905	0.034
0.206	1.292	1.288
0.807	1.206	8.373E-3
0.918	1.149	-0.064
0.385	1	0.955
0.043	0.532	3.621
1	1.273	-0.273

Equation (3.17) is used to determine the residual term:

$$\ln \gamma_i^{res} = \sum_k^N \Omega_{k,i} (\ln \Gamma_k - \ln \Gamma_{k,i}) \tag{3.17}$$

$\Omega_{k,1}$
1.897
0.462
1.462
0.028
0.232
0
0.016

$$\begin{aligned}
 \ln\gamma_1^{res} &= 1.897 \times (0.034 - 0.118) + 0.462 \times (1.288 - 0.756) \\
 &\quad + 1.462 \times (8.373E - 3 - 0.04) + 0.028 \times (-0.064 + 0.216) \\
 &\quad + 0.232 \times (0.955 - 0.537) + 0 \times (3.621 - 2.827) + 0.016 \times (-0.273) \\
 &\quad + 0.362 \\
 &= 0.142
 \end{aligned}$$

Similarly  $\ln\gamma_2^{res}$  is calculated as  $\ln\gamma_1^{res} = 4.789E - 4$

From equation (3.10),  $\ln\gamma_i$  is calculated:

$$\ln\gamma_1 = 0.0033457 + 0.142 =$$

$$\ln\gamma_2 = 0.0000853 + 4.789E - 4 = 5.642E - 4$$

Therefore  $\gamma_1 = 1.15587$  and  $\gamma_2 = 1.00056$

**Table B1. Sample activity coefficient predictions by the Extended UNISAC model to be used for verification.**

Component 1	Component 2	Temperature (K)	$x_1$	$\gamma_1$	$x_2$	$\gamma_2$
n-hexane	ethanol	298.15	0	7.6646	1	1
n-hexane	ethanol	298.15	0.1	5.45131	0.9	1.01801
n-hexane	ethanol	298.15	0.2	4.04051	0.8	1.07323
n-hexane	ethanol	298.15	0.3	3.09963	0.7	1.17244
n-hexane	ethanol	298.15	0.4	2.42198	0.6	1.34045
n-hexane	ethanol	298.15	0.5	1.93759	0.5	1.60854
n-hexane	ethanol	298.15	0.6	1.61465	0.4	2.01067
n-hexane	ethanol	298.15	0.7	1.38399	0.3	2.6809
n-hexane	ethanol	298.15	0.8	1.21099	0.2	4.02135
n-hexane	ethanol	298.15	0.9	1.07644	0.1	8.04269
n-hexane	ethanol	298.15	1	1	0	45.06792
toluene	butan-1-ol	298.15	0	1.88609	1	1
toluene	butan-1-ol	298.15	0.1	1.73573	0.9	1.00443
toluene	butan-1-ol	298.15	0.2	1.60338	0.8	1.01865
toluene	butan-1-ol	298.15	0.3	1.48665	0.7	1.04474
toluene	butan-1-ol	298.15	0.4	1.3835	0.6	1.08615
toluene	butan-1-ol	298.15	0.5	1.29218	0.5	1.14884
toluene	butan-1-ol	298.15	0.6	1.21127	0.4	1.24387
toluene	butan-1-ol	298.15	0.7	1.13974	0.3	1.39402
toluene	butan-1-ol	298.15	0.8	1.07739	0.2	1.65412
toluene	butan-1-ol	298.15	0.9	1.02648	0.1	2.19428
toluene	butan-1-ol	298.15	1	1	0	3.87599
benzene	dibutyl ether	298.15	0	0.9889	1	1
benzene	dibutyl ether	298.15	0.1	1.00472	0.9	0.99918
benzene	dibutyl ether	298.15	0.2	1.01734	0.8	0.997
benzene	dibutyl ether	298.15	0.3	1.02635	0.7	0.99412
benzene	dibutyl ether	298.15	0.4	1.0314	0.6	0.99155
benzene	dibutyl ether	298.15	0.5	1.03232	0.5	0.99094
benzene	dibutyl ether	298.15	0.6	1.02919	0.4	0.9948
benzene	dibutyl ether	298.15	0.7	1.02252	0.3	1.00724
benzene	dibutyl ether	298.15	0.8	1.01351	0.2	1.03504
benzene	dibutyl ether	298.15	0.9	1.00453	0.1	1.09039
benzene	dibutyl ether	298.15	1	1	0	1.19713

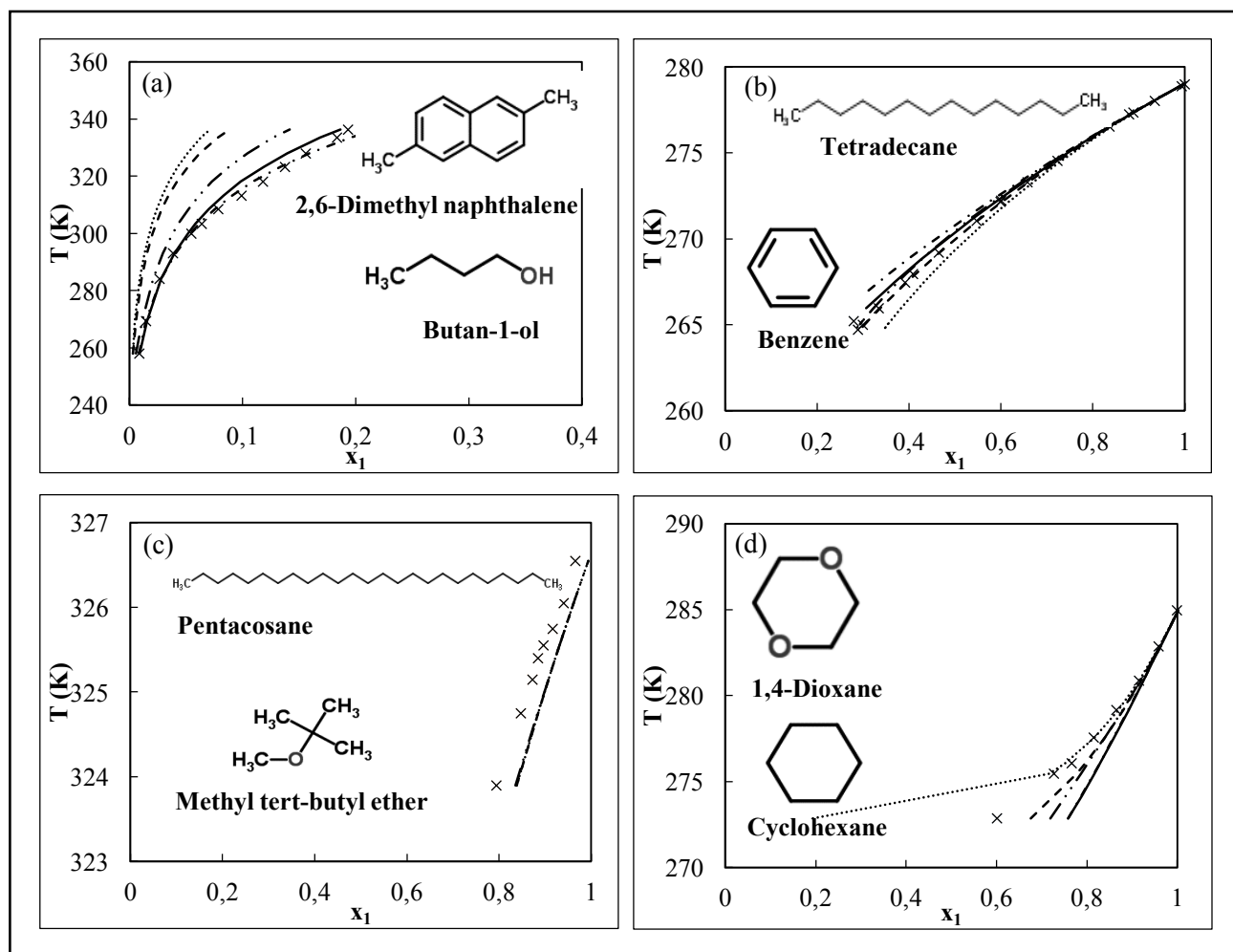
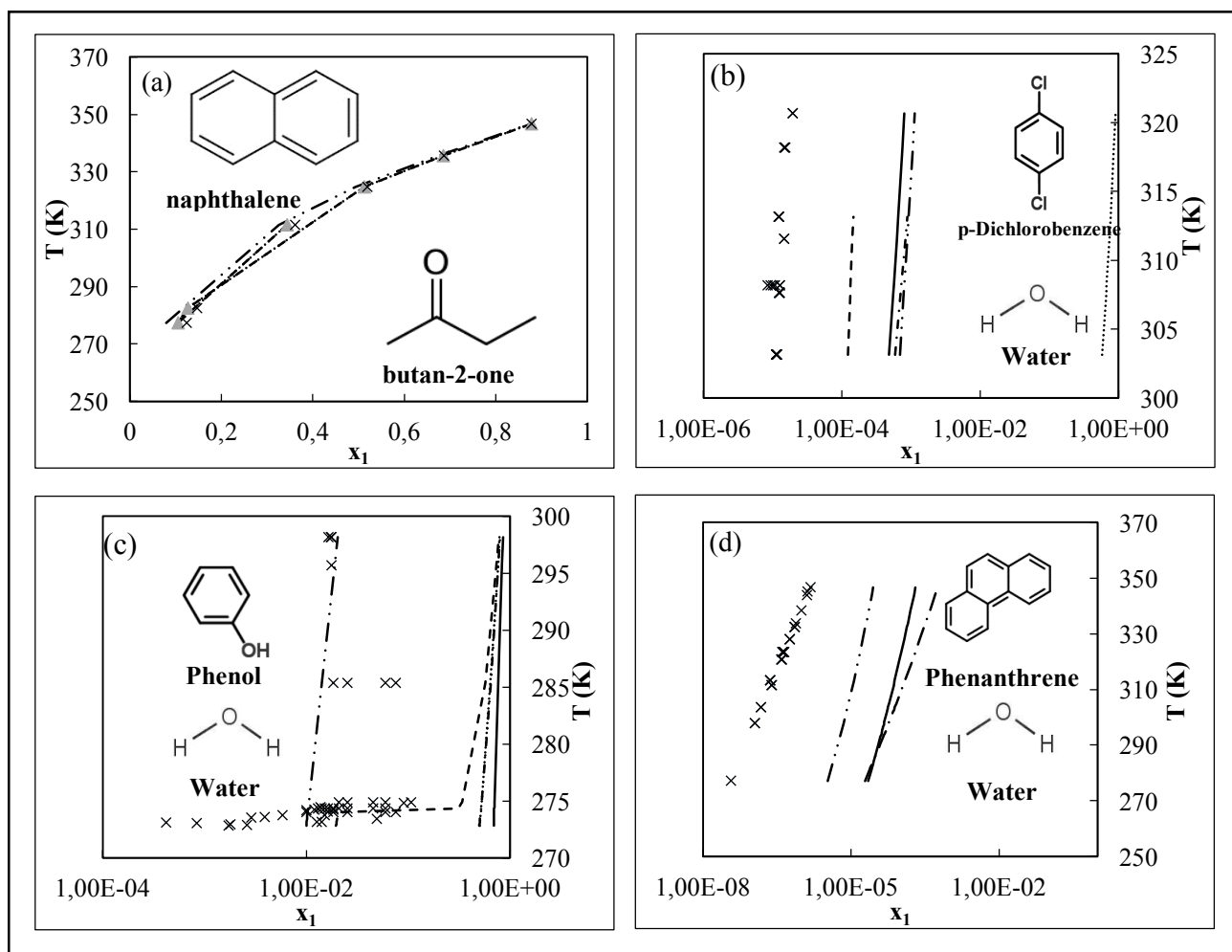
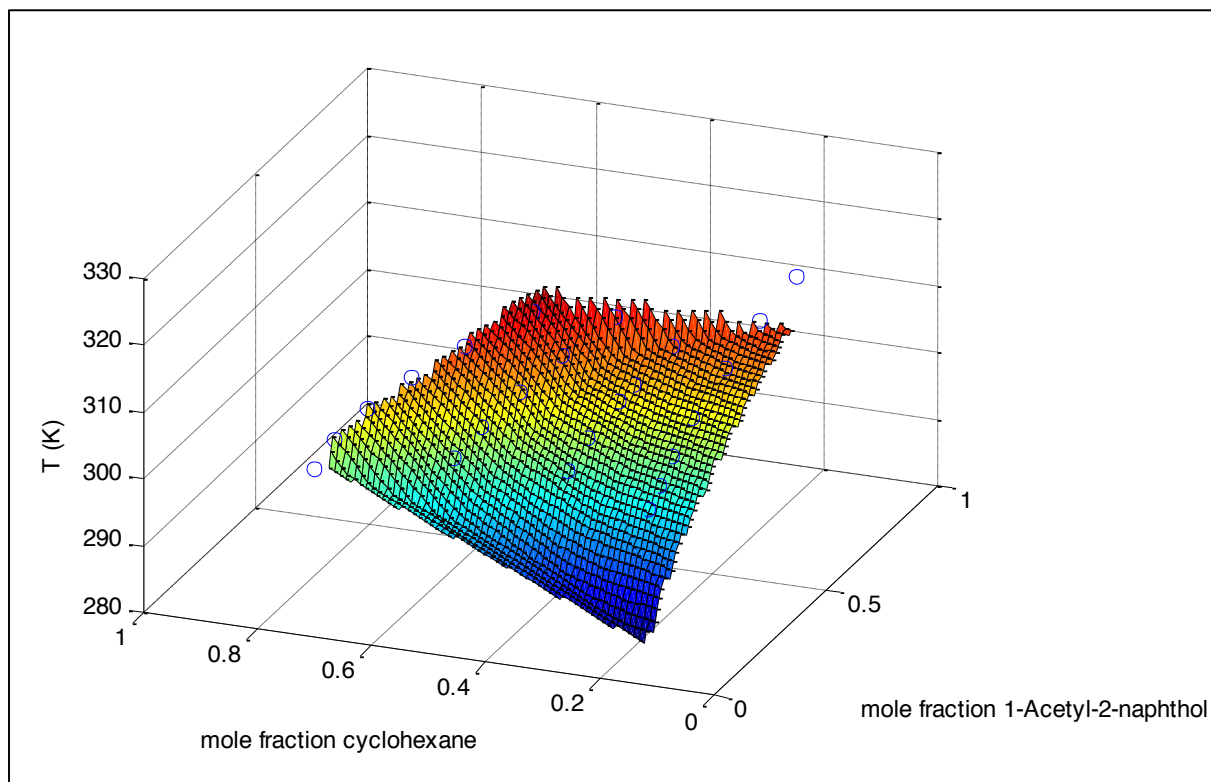


Figure B1. Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models.  $\times$ , Experimental Data (Dortmund Data Bank, 2012);  $\text{---}\cdot\text{---}$ , Extended UNISAC;  $\text{---}$ , mod. UNIFAC (Dortmund);  $\text{---}\cdot\text{---}$ , UNIFAC;  $\text{---}$ , COSMO-SAC;  $\text{---}\cdot\text{---}$ , COSMO-RS (OL). Systems: (a) 2,6-dimethyl naphthalene (1) + butan-1-ol (2); (b) tetradecane (1) + benzene (2); (c) pentacosane (1) + methyl tert- ether (2);(d) 1,4-dioxane (1) + cyclohexane (2). Experimental data are represented as symbols, and model predictions are represented as lines.

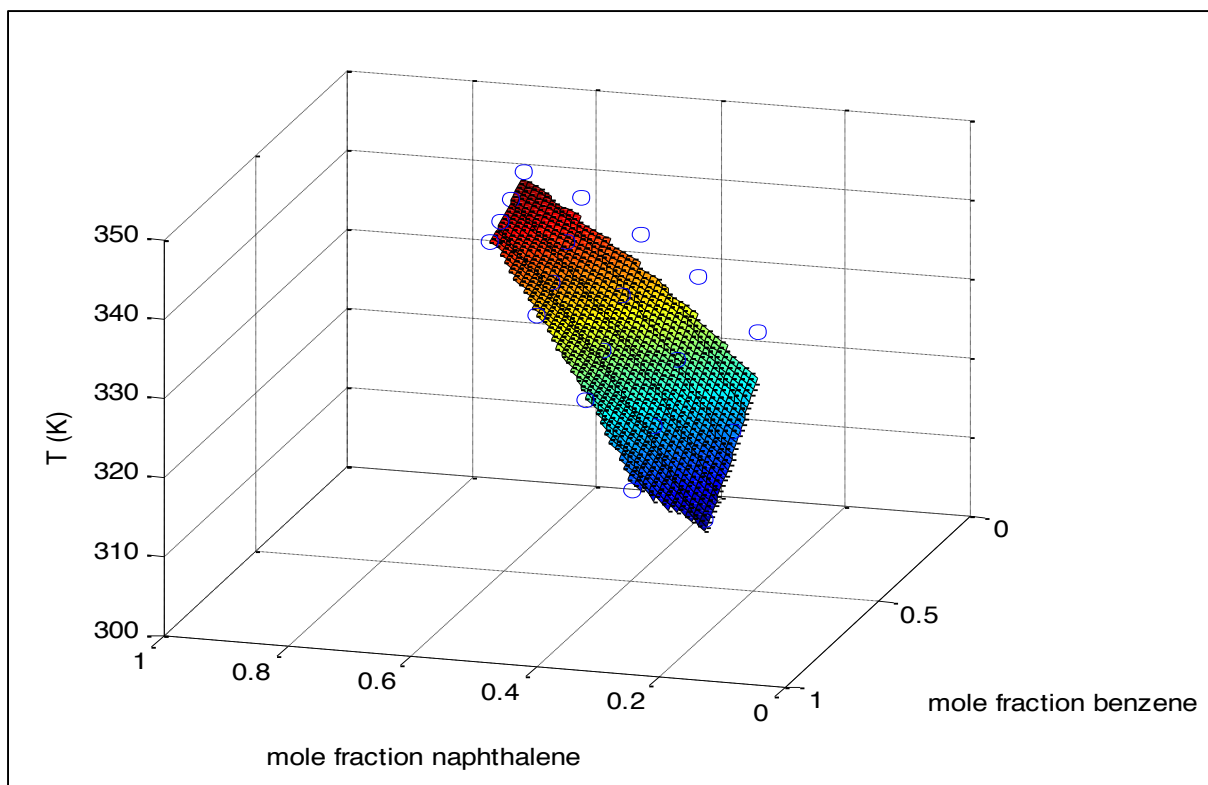




**Figure B2. Experimental and predicted solid-liquid equilibrium composition at various temperatures by different models.  $\times$ , Experimental Data (Dortmund Data Bank, 2012);  $- \cdot -$ , Extended UNISAC;  $- - -$ , mod. UNIFAC (Dortmund);  $\cdots$ , UNIFAC;  $-$ , COSMO-SAC;  $- \cdot -$ , COSMO-RS (OL). Systems: (a) naphthalene (1) + butan-2-one (2); (b) phenanthrene (1) + water (2); (c) phenol (1) + water (2); (d) p-dichlorobenzene (1) + water (2). Experimental data are represented as symbols, and model predictions are represented as lines.**



**Figure B3. Experimental and predicted solid-liquid equilibrium temperatures for the system 1-acetyl-2-naphthol (1) + cyclohexane (2) + ethanol. ○, Experimental Data (Dortmund Data Bank, 2012); surface, Extended UNISAC. Experimental data are represented as symbols, and model predictions are represented as a surface.**



**Figure B4. Experimental and predicted solid-liquid equilibrium temperatures for the system naphthalene (1) + benzene (2) + ethanol.  $\circ$ , Experimental Data (Dortmund Data Bank, 2012); surface, Extended UNISAC. Experimental data are represented as symbols, and model predictions are represented as a surface.**

## B2. Components and data sets

Table B1. Component list.

1,1,1-Trichloroethane [r140a]	2,3-Butanediol	Betulin	1,2,4,5-Tetramethylbenzene	2-Butanol	Chrysene	M-Terphenyl	Trans-Decahydronaphthalene
1,2,3,4-Tetrachlorobenzene	2,3-Pentanedione	Biphenyl	1,2-benzophenanthrene	2-Butanone	Cis-Decahydronaphthalene	N,N-Dimethylacetamide	Tricosane
1,2,3,4-Tetrahydronaphthalene	2,4,6-Trichlorobiphenyl	Biphenyl	1,2-Dichloroethane	2-Butoxyethanol	Cyclohexane	Nadolol	Tridecane
1,2,3-Trihydroxybenzene	2,5-Dimethylphenol	Bis(2-chloroethyl) ether	1,2-Dihydro-acenaphthylene	2-Heptanone	Cyclohexanol	Naphthalene	Triphenylene
1,2,4,5-Tetramethylbenzene	2,6-Dimethylnaphthalene	Butylbenzene	1,2-Ethandiol	2-Methylnaphthalene	Cyclohexanone	N-Butane	Water
1,2-benzophenanthrene	2,6-Dinitrotoluene	Carbazole	1,2-Propylene oxide	2-Methyl-1-propanol	Cyclooctane	N-Methyl-2-pyrrolidone	
1,2-Dichloroethane	2,6-Di-tert.butyl-4-methylphenol	Chloroform	1,3,5-triphenylbenzene	2-Methylbutane	Decane	Nonadecane	
1,2-Dihydro-acenaphthylene	2,7-Dimethylnaphthalene	Cholest-5-en-3-ol	1,4-Dihydroxybenzene	2-Methylnaphthalene	Dibutyl ether	Nonane	
1,2-Ethandiol	2-Butanol	Chrysene	1,4-Dioxane	2-Methylpentane	Diethyl ether	N-Tridecanol	
1,2-Propylene oxide	2-Butanone	Cis-Decahydronaphthalene	1-Acetyl-2-naphthol	2-Methylphenol	Dimethyl sulfoxide	N-Undecane	
1,3,5-Trinitrobenzene	2-Butoxyethanol	Components list	1-Butanol	2-Methylpyridine	Dimethyl terephthalate	Octacosane	
1,3,5-triphenylbenzene	2-Heptanone	Cyclohexane	1-Cyclohexylnonadecane	2-Octanone	Di-n-propyl ether	Octadecane	
1,4-Dihydroxybenzene	2-Methylnaphthalene	Cyclohexanol	1-Decanol	2-Pentanol	Diosgenin	Octane	
1,4-Dioxane	2-Methyl-1-propanol	Cyclohexanone	1-Dodecanol	2-Propanol	DL-Camphor	O-Terphenyl	
1-Acetyl-2-naphthol	2-Methylbutane	Cyclooctane	1-Eicosanol	3,4-Dimethylphenol	Docosane	P-Dichlorobenzene	
1-Butanol	2-Methylnaphthalene	Decane	1-Heptadecanol	3-Methyl-1-butanol	Dodecane	Pentacosane	
1-Cyclohexylnonadecane	2-Methylpentane	Dibutyl ether	1-Heptanol	3-Methylphenol	Dotriacontane	Pentadecane	
1-Decanol	2-Methylphenol	Diethyl ether	1-Hexadecanol	3-Pentanol	Eicosane	Pentadecylcyclohexane	
1-Dodecanol	2-Methylpyridine	Dimethyl sulfoxide	1-Hexanol	4-Chlorobiphenyl	Estrone	Pentane	

1-Eicosanol	2-Octanone	Dimethyl terephthalate	1-Nonanol	4-Methyl-2-pentanone	Ethanol	Phenanthrene	
1-Heptadecanol	2-Pentanol	Di-n-propyl ether	1-Octadecanol	4-Methylphenol	Ethyl acetate	Phenol	
1-Heptanol	2-Propanol	Dioctylamine	1-Octanol	9,10-benzophenanthrene	Ethylbenzene	Prednisolone	
1-Hexadecanol	3,4-Dimethylphenol	Diosgenin	1-Pentadecanol	9H-Fluorene	Ethylene oxide	Propane	
1-Hexanol	3-Methyl-1-butanol	Diphenylamine	1-Pentanol	Acenaphthene	Fluoranthene	Propionitrile	
1-Naphthylamine	3-Methylphenol	DL-Camphor	1-Propanol	Acetone	Heneicosane	Propyl 4-hydroxybenzoate	
1-Nonanol	3-Pentanol	Docosane	1-Tetradecanol	Acetonitrile	Heptacosane	P-Terphenyl	
1-Octadecanol	4-Chlorobiphenyl	Dodecane	1-Undecanol	Aniline	Heptadecane	P-Xylene	
1-Octanamine	4-Methyl-2-pentanone	Dotriacontane	2,2',3,3',4,4'-Hexachlorobiphenyl	Anthracene	Heptane	Pyrene	
1-Octanol	4-Methylphenol	Eicosane	2,2',4,4',6,6'-Hexachlorobiphenyl	Anthracene	Hexachlorobenzene	Pyridine	
1-Pentadecanol	9,10-benzophenanthrene	Estrone	2,2',4,5'-Tetrachlorobiphenyl	Atropine	Hexacosane	Tert-Butanol	
1-Pentanol	9H-Fluorene	Ethanol	2,2,4-Trimethylpentane	Benzene	Hexadecane	Tert-Pentanol	
1-Propanol	Acenaphthene	Ethyl acetate	2,2-Bis-(4-methoxyphenyl)-1,1,1-trichloroethane	Benzyl alcohol	Hexane	Testosterone	
1-Tetradecanol	Acetone	Ethylbenzene	2,3-benzindene	Betulin	Hydrocortisone	Testosterone valerate	
1-Undecanol	Acetonitrile	Ethylene oxide	2,3-Butanediol	Biphenyl	Ibuprofen	Tetrachloromethane	
2,2',3,3',4,4'-Hexachlorobiphenyl	Aniline	Fluoranthene	2,3-Pentanedione	Bis(2-chloroethyl) ether	Indole	Tetracosane	
2,2',4,4',6,6'-Hexachlorobiphenyl	Anthracene	Heneicosane	2,4,6-Trichlorobiphenyl	Butylbenzene	Mestanolone	Tetradecane	
2,2',4,5'-Tetrachlorobiphenyl	Anthracene	1,1,1-Trichloroethane [r140a]	2,5-Dimethylphenol	Carbazole	Methanol	Tetrahydrofuran	
2,2,4-Trimethylpentane	Atropine	1,2,3,4-Tetrachlorobenzene	2,6-Dimethyl naphthalene	Chloroform	Methyl p-hydroxybenzoate	Tetratriacontane	
2,2-Bis-(4-methoxyphenyl)-1,1,1-trichloroethane	Benzene	1,2,3,4-Tetrahydronaphthalene	2,6-Di-tert.butyl-4-methylphenol	Cholest-5-en-3-ol	Methyl tert-butyl ether (MTBE)	Toluene	
2,3-benzindene	Benzyl alcohol	1,2,3-Trihydroxybenzene	2,7-Dimethyl naphthalene		Methylcyclohexane	Trans-1,2-Diphenylethene	

Table B2. Dortmund Data Bank (2012) set numbers. \*, \$

1	274	706	1084	1434	2266	3082	5729	6773	7748	10413	11526	13794	17103	19772	32847	2002491
2	275	708	1085	1436	2267	3087	5730	6799	7750	10414	11616	14097	17104	19839	32852	2002539
4	276	717	1086	1437	2268	3115	5734	6883	7771	10415	11617	14098	17105	20394	32869	2002540
6	277	718	1087	1438	2272	3363	5781	6885	7772	10416	11899	14099	17296	20395	35226	2002541
10	280	720	1100	1439	2273	3365	6009	6922	7773	10417	11900	14220	17297	21336	35227	2003630
16	282	777	1104	1450	2274	3583	6071	6923	7775	10418	11901	14221	17671	21554	35536	2004277
17	283	778	1109	1494	2275	3626	6118	6984	7794	10419	11902	14231	17686	21556	35539	2004278
32	290	797	1111	1567	2293	3631	6136	6985	7798	10547	11903	14232	17695	21559	36344	2004279
33	462	836	1162	1573	2294	3694	6245	6986	7854	10548	11904	14249	17751	21712	36348	2004280
38	463	865	1163	1604	2295	3695	6249	6987	7857	10549	11905	14250	17752	22910	37372	2004281
40	464	880	1164	1605	2297	3830	6252	6988	7858	10550	11907	14251	17753	24211	37374	2004282
114	465	883	1165	1607	2300	3831	6590	6989	7859	10652	11931	14252	17755	24215	1000020	2004283
115	466	918	1166	1608	2318	3834	6624	6990	7860	10655	12383	14253	17756	24216	2000108	2004284
123	467	920	1167	1637	2319	3835	6625	6991	7861	11010	12948	14525	17830	24886	2000592	2004285
124	468	921	1173	1656	2415	3841	6626	6992	7991	11015	13528	14666	17832	25059	2000594	2004286
150	469	922	1178	1687	2416	3849	6655	6993	8049	11016	13529	14715	17835	25060	2000595	2004287
152	470	923	1179	1771	2418	3883	6656	6994	8063	11017	13530	15151	17873	25296	2000606	2004288
154	505	924	1185	1773	2419	3937	6657	6995	8067	11099	13531	15278	18287	27593	2000620	2004397
156	511	925	1187	1800	2422	4004	6661	7098	8298	11113	13532	15281	18313	27595	2000625	2004399
175	572	926	1226	1891	2448	4005	6671	7099	8380	11114	13533	15299	18323	27597	2001295	2004402
180	576	927	1227	1892	2449	4006	6741	7100	8391	11206	13534	15303	18324	27598	2001304	2004403
181	581	928	1318	1916	2467	4007	6742	7101	9010	11207	13536	15311	18325	27643	2001337	2004463
184	585	929	1372	1917	2627	4146	6746	7143	9199	11208	13540	15333	18326	28555	2001379	2004769
188	590	930	1375	1918	2769	4223	6748	7254	9260	11209	13562	15843	18327	30248	2001768	2004783
217	594	932	1411	1960	2778	4228	6749	7538	9423	11210	13564	15957	18328	30253	2001794	2005168
228	610	933	1412	1976	2780	4928	6751	7539	9429	11211	13594	16318	18331	30303	2001796	2005182

**APPENDIX B**

**Extended UNISAC**

231	612	934	1414	1977	2781	5342	6752	7594	9432	11212	13610	16623	18430	31943	2001798	2005304
233	613	990	1415	1978	2782	5461	6753	7595	9435	11213	13660	16742	18681	32695	2001799	2005989
235	616	1053	1416	2157	2784	5474	6754	7596	9438	11214	13706	16783	18726	32696	2001800	2006257
237	679	1057	1417	2167	2785	5477	6755	7597	9444	11215	13743	16813	18727	32705	2001805	2006271
241	680	1058	1418	2251	2854	5571	6764	7598	9501	11470	13784	16814	18728	32706	2001806	2006273
266	681	1062	1422	2252	2947	5572	6765	7599	9502	11472	13786	16815	18729	32731	2001938	
267	682	1063	1423	2255	2953	5573	6766	7600	9503	11474	13787	16816	18731	32732	2001940	
268	683	1064	1425	2256	2984	5574	6767	7601	9556	11476	13788	16817	18732	32741	2002351	
269	684	1068	1426	2258	3000	5581	6768	7602	10064	11484	13789	16819	18920	32742	2002352	
270	692	1071	1427	2259	3058	5702	6769	7603	10104	11486	13790	16820	19137	32827	2002487	
271	702	1074	1428	2260	3060	5703	6770	7743	10194	11488	13791	17100	19141	32832	2002488	
272	703	1082	1429	2261	3061	5705	6771	7744	10211	11490	13792	17101	19477	32837	2002489	
273	705	1083	1433	2265	3062	5728	6772	7745	10412	11524	13793	17102	19725	32842	2002490	

\*For more detailed information of the individual data sets, please refer to the free DDBST Explorer Edition ([www.ddbst.com](http://www.ddbst.com)), <sup>5</sup> Set numbers greater than 2000000 form part of a private database

## Appendix C

### Supplementary data: Experimental Solubility Measurement

#### C1. Calibration

The temperature sensor (Pt-100) used in this work to measure the equilibrium temperature was calibrated using a WIKA CTH6500 kit. The temperature calibration plot is shown in Figure C1. The plot of temperature deviation from those of standard values (Shown in Figure C2) reveals a maximum deviation in temperature of 0.1 K.

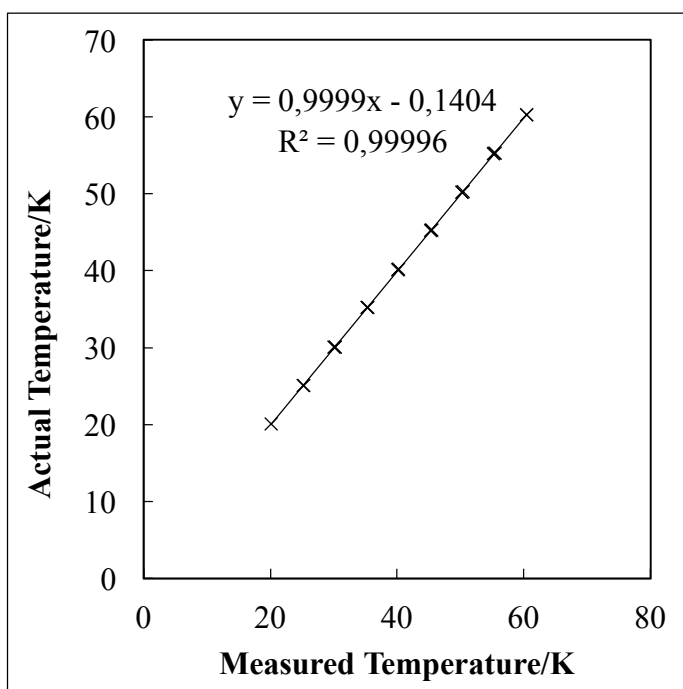


Figure C1. Calibration curve of equilibrium cell bath temperature probe.

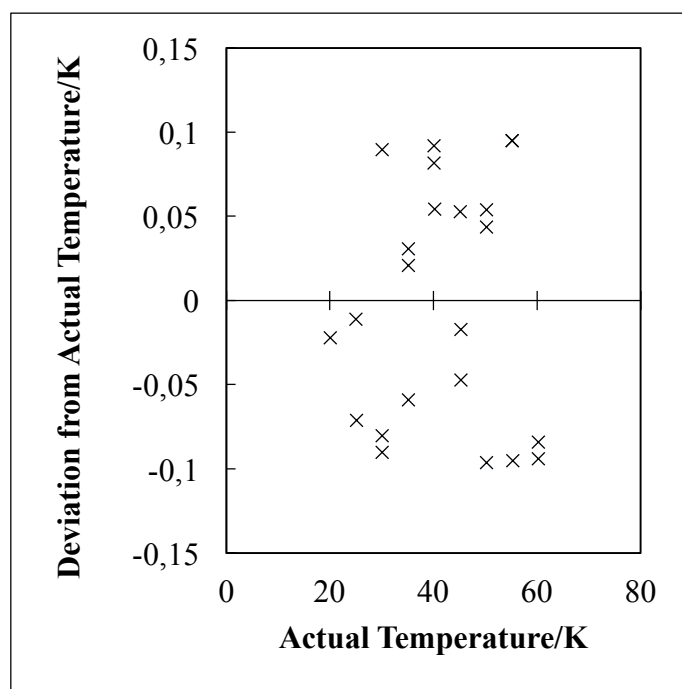


Figure C2. Plot of deviations of measured temperature from actual temperature.



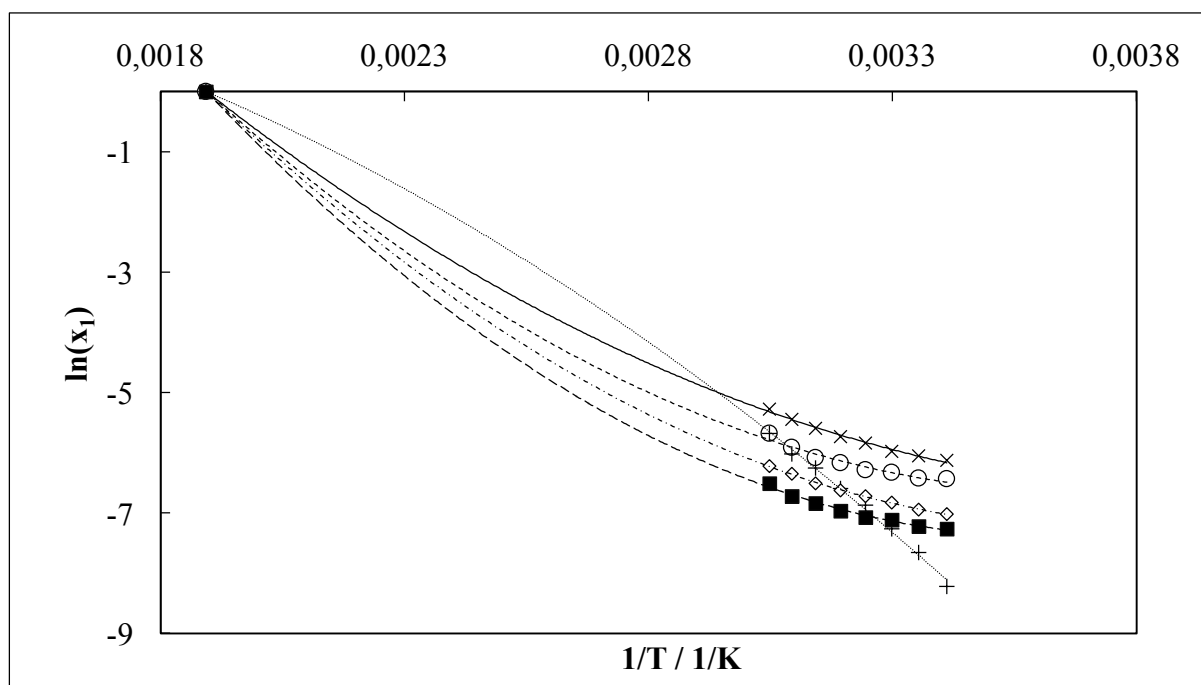


Figure C3.  $\ln(x_1)$  vs  $1/T$  plot for the systems betulin + (experimental, 2<sup>nd</sup> order polynomial), (+, ---), acetoneitrile/ (x, —), nonan-1-ol/ (o, ---), octan-1-ol/ (◇, - · -), pentan-1-ol/ (■, - - -), butan-2-ol. Experimental data are represented as symbols, and model predictions are represented as lines.

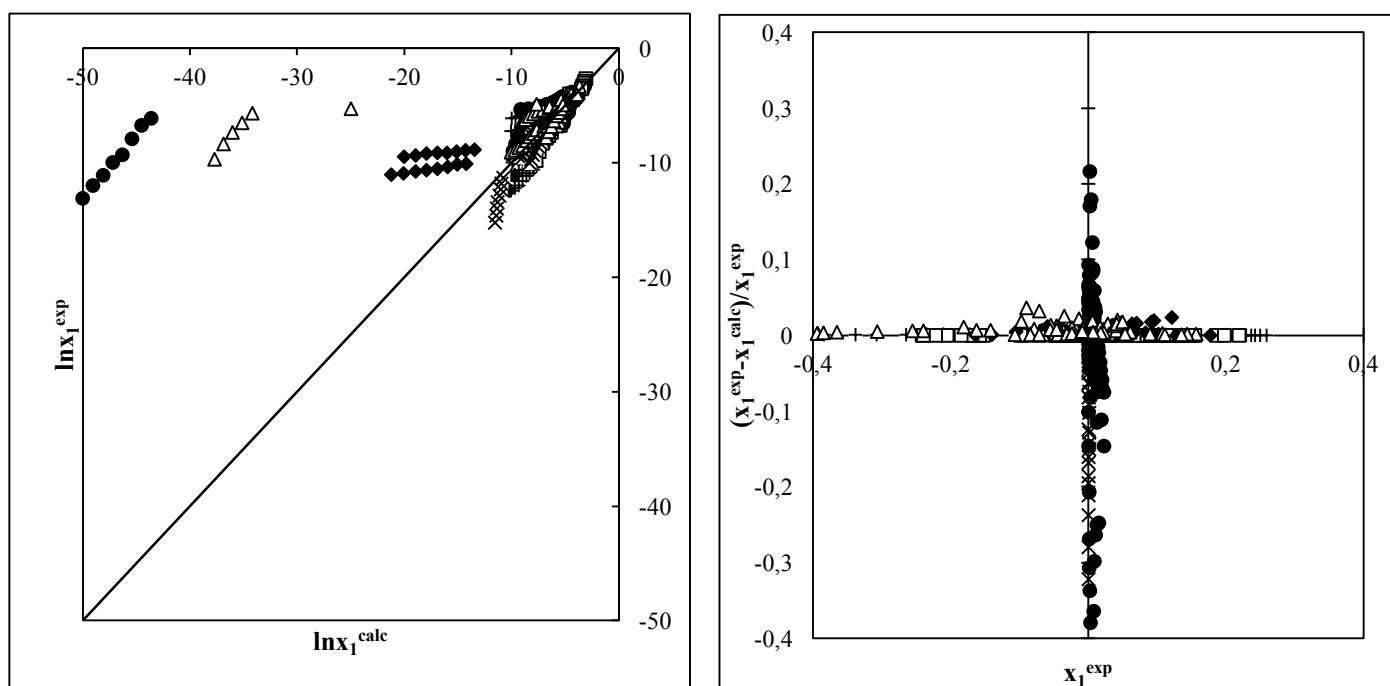


Figure C4. (a) Plot of  $\ln(x_1^{\text{exp}})$  vs  $\ln(x_1^{\text{calc}})$  for the solubility systems measured. (b) Scatter plot of  $((x_1^{\text{exp}} - x_1^{\text{calc}})/x_1^{\text{exp}})$  vs  $x_1^{\text{exp}}$  for the solubility systems measured. +, betulin, ×, prednisolone, ◇, hydrocortisone, ●, diosgenin, Δ, estriol, □, estrone.