
Activity of Complex Multifunctional Organic Components in Common Solvents

By

Bruce Moller

[B.Sc. (Eng.), M.Sc. (Eng.)]

University of KwaZulu-Natal, Howard College Campus

For the degree Doctor of Philosophy (Chemical Engineering)

Progress Report – October 2008

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

ABSTRACT

The models used in prediction of the activity coefficients are important tools when designing major unit operations (distillation columns, liquid-liquid extractors etc). In the petrochemical and chemical industry well established methods such as UNIFAC and ASOG are routinely employed for the prediction of the activity coefficient. These methods are however reliant on binary group interaction parameters which need to be fitted to reliable experimental data. It is for this reason that these methods are very often not applicable to systems which involve complex molecules. In these systems typically solid-liquid equilibria are of interest where the solid is some pharmaceutical product or intermediate or a molecule of similar complexity (The term complex here refers to situations where molecules contains several functional groups which are either polar, hydrogen bonding or lead to mesomeric structures in equilibrium. In many applications due to economic considerations no more than 30 solvents are usually employed.

It is for this reason that the objective of this work is to develop a method for predicting the activity coefficient of complex multifunctional compounds in common solvents. The segment activity coefficient approaches proposed by Hansen and the NRTL-SAC model show that it should be possible to “interpolate” between solvents if suitable reference solvents are available (e.g. non-polar, polar and hydrogen bonding). Therefore it is useful to classify the different solvents into suitable categories inside which analogous behaviour should be observed. To accomplish this, a significant amount of data needs to be collected for the common solvents. A novel method to achieve this is proposed in the report. The four main methods for activity coefficient prediction are reviewed and a possible method for estimating the missing binary interaction parameters in the group contribution methods: UNIFAC and mod. UNIFAC is examined.

The final chapter in this progress report outlines the further steps of this research and the work involved in each.

PREFACE

The work presented in this progress report was performed at the University of KwaZulu-Natal, Durban from January 2008 to October 2008. The work was supervised by Prof. D. Ramjugernath and Prof. Dr. J. Rarey.

This progress report is presented as a partial requirement for the degree of PhD in Chemical Engineering. All the work presented in this report is original unless otherwise stated and has not (in whole or part) been submitted previously to any tertiary institute as part of a degree.

Bruce Moller (203502126)

TABLE OF CONTENTS

Abstract.....	i
Preface	ii
Table of Contents	iii
List of Figures	v
List of Tables	vii
Nomenclature	ix
1. Introduction	1
2. Solvent Applications / Effects / Selection	4
2.1. Viscosity	4
2.2. Volatility	5
2.3. Toxicity and environmental impact.....	5
2.4. Selectivity and capacity	6
2.5. Liquidus range.....	6
2.6. Computer Aided Molecular Design (CAMD)	6
3. Classification of Solvents.....	8
3.1. Trouton's rule	8
3.2. Cohesive pressure	9
3.3. Internal pressure	9
3.4. Acid-base classification	10
3.5. Solvent classification	11
4. Theoretical Approaches to Solute Activity.....	12
4.1. Thermodynamic framework.....	12
4.2. Empirical and semi-empirical models for G^E	16

4.2.1.	Local composition concept (Wilson, NRTL, UNIQUAC).....	16
4.2.2.	Functional group activity coefficients.....	19
4.2.3.	Segment activity coefficients.....	23
4.2.4.	Combinatorial contributions.....	32
5.	Experimental Determination of Solute Activity.....	35
5.1.	Solid-liquid equilibria.....	35
5.2.	Partition coefficients.....	38
6.	Reduction of Infinite Dilution Data to a Common Solvent.....	39
6.1.	Alkanes.....	39
6.2.	n-Alcohols.....	44
6.3.	Ketones.....	47
6.4.	NRTL-SAC solvent interpolation.....	49
7.	NRTL-SAC UNIFAC Matrix Interpolation.....	55
8.	Further Work.....	57
9.	References.....	58

LIST OF FIGURES

Fig. 1 The four phase splitting example of Hildebrand [15].....	11
Fig. 2 The molecular formula and thermo physical properties required for aspirin (all data taken from Beilstein [32]).....	22
Fig. 3 The state of the UNIFAC matrix as of September 2008, data from DDBST [33].....	23
Fig. 4 Hansen solubility sphere for sucrose (\blacktriangle – position of solvents relative to the solubility sphere, all solubility parameters taken from Barton [34] and all have units of MPa ^{0.5} . Red symbolizes a poor solvent and blue a good one).....	26
Fig. 5 Sigma profiles for 3 solvents and an example solute (aspirin) (profiles from the VT database [42]).....	30
Fig. 6 Thermodynamic cycle for calculating the fugacity of a pure sub-cooled liquid, taken from Prausnitz [16].....	36
Fig. 7 The ideal solubility as a function of the melting temperature and the heat of fusion at 273.15 K.	37
Fig. 8 γ_2^∞ vs. r_1 for ethanol(2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB, - - - - SG combinatorial, mod-UNIFAC combinatorial, — GK-FV combinatorial).....	40
Fig. 9 γ_2^∞ vs. r_1 for benzene(2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB, - - - - SG combinatorial, mod-UNIFAC combinatorial, — GK-FV combinatorial).....	40
Fig. 10 γ_2^∞ vs. r_1 for butanone(2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB, - - - - SG combinatorial, mod-UNIFAC combinatorial, — GK-FV combinatorial).....	41
Fig. 11 γ_2^∞ vs. r_1 for ethylcyclohexane (2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB [33], - - - - SG combinatorial, mod-UNIFAC combinatorial, — GK-FV combinatorial).....	42
Fig. 12 γ_2^∞ vs. r_1 for methanol (2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB [33], — Eqn. (6-3), Eqn. (6-5)).....	43
Fig. 13 γ_2^∞ vs. r_1 for cyclohexane (2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB [33], — Eqn. (6-3), Eqn. (6-5)).....	43

Fig. 14 γ_2^∞ vs. r_1 for tetramethylstannane (2) in n-alcohol solvents(1) at 298 K (x – data from DDB [33], — Eqn. (6-9)).....	45
Fig. 15 γ_2^∞ vs. r_1 for pentane (2) in n-alcohol solvents(1) at 298 K (x – data from the DDB [33], — Eqn. (6-9)).	45
Fig. 16 γ_2^∞ vs. r_1 for dichloromethane (2) in n-alcohol solvents(1) at 298 K (x – data from the DDB [33], — Eqn. (6-9)).....	45
Fig. 17 γ_2^∞ vs. r_1 for dimethyl ether (2) in n-alcohol solvents(1) at 330 K (x – data from the DDB [33], — Eqn. (6-9)).....	46
Fig. 18 γ_2^∞ vs. r_1 for toluene (2) in n-alcohol solvents(1) at 330 K (x – data from the DDB [33], — Eqn. (6-9)).	46
Fig. 19 γ_2^∞ vs. r_1 for tetrachloromethane (2) in n-alcohol solvents(1) at 293.15 K (x – data from the DDB [33], — Eqn.(6-10) with squalane and methanol as the 2 reference solvents).	47
Fig. 20 γ_2^∞ vs. r_1 for hexane (2) in ketone solvents(1) at 298.15 K (x – data from the DDB [33], — Eqn. (6-11)).	48
Fig. 21 γ_2^∞ vs. r_1 for water (2) in ketone solvents(1) at 333.15 K (x – data from the DDB [33], — Eqn. (6-11)).	48
Fig. 22 γ_2^∞ vs. r_1 for ethanol (2) in ketone solvents(1) at 313.15 K (x – data from the DDB [33], — Eqn. (6-11)).	49
Fig. 23 γ_2^∞ vs. $1/T$ for ethanol (2) – 2-butanone (1) (x – data from the DDB [33], Δ – experimental data at 313.15 K, \square – value predicted from Eqn. (6-11) at 313.15 K).....	49

LIST OF TABLES

Table 1 Examples of the entropy of vaporization at the normal boiling point data from Myrdal et al. [12].....	9
Table 2 Internal and cohesive pressure for some sample compounds, data from Reichardt [11].....	10
Table 3 Examples of two more complex solvent classifications.....	11
Table 4 Group frequency illustration for the mixture hexane - ethanol	19
Table 5 Comparison of the experimental and predicted solubility of aspirin when using group contribution methods (experimental data from Frank [19]).....	22
Table 6 Hansen solubility parameters for sucrose and some solvents, data from Barton [34].....	25
Table 7 Comparison of the experimental and predicted data for aspirin solubility using Hansen solubility parameters from Barton [34] (* - the four reference solvents, ^a - Hansen's parameters, ^b - Hoy's parameters, experimental data from Frank [19])	27
Table 8 Segment interaction parameters for NRTL-SAC taken from Chen et al. [38]	29
Table 9 Comparison of the experimental and predicted data for aspirin solubility using NRTL-SAC [38] (* - the four reference solvents, experimental data from Frank [19])	29
Table 10 Comparison of the experimental and predicted data for aspirin solubility using COSMO-SAC and COSMO-RS (experimental data from Frank [19], COSMO profiles from the DDB [33])	31
Table 11 Comparison of the prediction for a selected group of compounds when using Eqn. (6-2) with the various combinatorial expressions	41
Table 12 Comparison of the experimental and fitted predictions for infinite dilution activity coefficient using Eqn. (6-13) for 4 different solutes at 298.15 K (* - the 4 reference solvents) using the NRTL-SAC parameters from [37]. All data from the DDB [33].	50
Table 13 Comparison of the experimental and fitted predictions for infinite dilution activity coefficient using NRTL-SAC for 4 different solutes at 298.15 K (* - the 4 reference solvents) using the NRTL-SAC parameters from [37]. All data from the DDB [33].	51
Table 14 The results of the solvent interpolation for the fitted values of the NRTL-SAC analogous parameters (* - the 4 reference solvents for which the original NRTL-SAC parameters are used)	53

Table 15 The results of the solvent interpolation for the fitted values of the NRTL-SAC analogous parameters (* - the 4 reference solvents for which the original NRTL-SAC parameters are used)53

Table 16 The model parameters for the interpolations in Table 14 and Table 1554

Table 17 Matrix of the infinite dilution activity coefficients of one main group in another at 298.15 K (Grp num – main group number, Grp – main group name)55

Table 18 The regressed surface segments for some of the UNIFAC groups56

NOMENCLATURE

a_i	-	Activity of component i .
c	-	Cohesive pressure (MPa).
F	-	Degrees of freedom.
f	-	Fugacity (kPa)
\hat{f}_i	-	Fugacity in solution of component i (kPa).
G	-	Gibbs free energy ($\text{J}\cdot\text{mol}^{-1}$).
H	-	Enthalpy ($\text{J}\cdot\text{mol}^{-1}$).
n	-	Number of moles (mol).
P	-	Pressure (kPa).
R	-	Ideal gas constant ($= 8.314472 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$).
S	-	Entropy ($\text{J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$).
T	-	Absolute temperature (K).
U	-	Internal energy ($\text{J}\cdot\text{mol}^{-1}$).
V	-	Specific volume ($\text{cm}^3\cdot\text{mol}^{-1}$).
x_i	-	Mole fraction of component i .
Greek symbols		
β	-	Selectivity.
δ	-	The Hildebrand solubility parameter ($\text{MPa}^{0.5}$)
ϕ_i	-	Volume fraction of component i .
γ_i	-	Activity coefficient of component i .
μ_i	-	Chemical potential of component i .
π	-	Internal pressure (MPa).

Subscripts

b	-	Normal boiling point.
d	-	Dispersive.
fus	-	Fusion.
h	-	Hydrogen bonding.
m	-	Mixing.
p	-	Polar.
v	-	Vaporization.

Superscript

∞	-	Infinite dilution.
0	-	Standard state.
C	-	Combinatorial.
E	-	Excess property.
FV	-	Free-Volume.
L	-	Liquid.
r	-	Reference state.
R	-	Residual.
S	-	Solid.

All other symbols used have been explained in the text and unless otherwise stated SI units are used.

1. INTRODUCTION

In non-ideal solutions the real mixture behaviour has a significant effect on the activity of the solute. Knowledge of the solute activity is required for the solution of a large variety of technical and environmental problems. Solute activity strongly influences the solubility of a liquid or solid solute, its partial vapour pressure above the solution, the chemical equilibrium composition, and reaction rates in the liquid phase. Activities in two immiscible solvents allow the calculation of distribution coefficients required for extraction processes and the distribution of chemicals between environmental compartments. Solute activities usually depend on solute concentration but may also be nearly constant over a significant concentration range.

Every year there is an ever increasing number of new compounds that are produced by chemical companies and unfortunately the experimental data for these compounds are generally not available. Experimental measurements are generally costly and time consuming and practically impossible to undertake for all pure compounds and mixtures. Even with the potentially huge selection of possible solvents most chemical companies only use a select list (a large pharmaceutical company that we know only uses a list of 16 preferred solvents). The aim of this work is therefore to develop a group contribution method to predict the activities of complex multifunctional organic compounds in various common solvents. This tool will especially support chemical engineers in both small and large companies dealing with the design, simulation and optimization of chemical processes. The method should be suitable for complex multifunctional compounds (biological compounds, pharmaceuticals), which are not covered well by existing models.

Current estimation techniques include simple group contribution approaches for e.g. water solubility or thermodynamic models like UNIFAC (based on group interactions) or COSMO-RS (based on quantum-mechanical calculations). In group contribution methods the number of parameters increases quadratic with the number of groups and the available experimental information usually does not allow a strong group differentiation. COSMO-RS methods on the other hand contain only very few parameters tuned to experimental results. In the case of complex multifunctional compounds like pharmaceuticals and their intermediates, pesticides or food components, a great need exists for a quantitative estimation technique with a broad parameterization.

Literature offers a large amount of experimental data like solubility information for complex solutes in common solvents (water, ethanol, acetone, benzene, etc.), that can be used for the implementation of an estimation method. Modelling the (infinite dilution) activity coefficient of a wide range of compounds in a single solvent mainly requires only single group contribution parameters (i.e. it is treated as a pure component property) and allows for sufficient group differentiation. Methods of this type, although very useful and required for solving a wide range of technical and environmental problems have only been developed to a minor extend up to now (e.g. for water solubility).

Specific Objectives of this work are:

- To select sufficient and suitable experimental information on solute activities for complex molecules from various commercial sources available to the partner institutions (Dortmund Data Bank, Beilstein, chemical provider catalogues, etc.).
- To compile or estimate the required auxiliary information like melting points, heat of fusion, etc. for the components of interest from the available sources.
- To generate or compile further pure component information that could prove useful to improve the reliability of the method, preferably from quantum mechanical calculations (polarisabilities, shielding charges in ideal conductors, etc.)
- To develop a group contribution method for solute activities in several common solvents based on the experimental information available.
- To verify, whether the activity in further solvents can be estimated by interpolation of results for similar solvents (e.g. activities in ethylbenzene from activities in hexane and benzene).
- To compare results of the new method with available methods from literature (UNIFAC, COSMO-RS, etc.).
- To develop an estimation method for the probable estimation error depending on the chemical type of component.

The methodology should be based on that developed during previous work, in which a variety of computer tools have been developed for the fast and efficient development of group contribution estimation methods.

This work had resulted in the development of group contribution estimation methods for

- Normal boiling temperatures (Nannoolal et al. [1])
- Vapour pressure as function of temperature (Nannoolal et al. [2])
- Critical property data (Nannoolal et al. [3])
- Viscosities of liquids (Nannoolal et al. [4])
- Vapour pressure as function of temperature (improved, Moller et al. [5])

Compared to available methods from literature, all these developments resulted in both a significantly broader range of applicability and greater reliability as well as smaller estimation errors.

In the case of solute activities, more effort is required for data reduction. Solute activities depend on the solvent, temperature and solute concentration. As the amount of data for solubilities and activity coefficients at infinite dilution is largest close to room temperature estimation will be limited to 25°C. For the reduction of the concentration effect a suitable mixture model will be employed. In the case of solid solutes the heat of fusion and melting temperature is required for the calculation of the hypothetical liquid activity. In the case of components with a melting point far away from room temperature, a reasonable estimate of the difference of solid and hypothetical liquid heat capacity will also be required.

2. SOLVENT APPLICATIONS / EFFECTS / SELECTION

A solution can be defined as a mixture of at least 2 substances which has uniform chemical and physical properties throughout (Mellan [6]). There are 2 parts to every solution, the solvent and the solute. The component(s) in excess is(are) usually called the solvent(s) while the component in deficiency is usually referred to as the solute. When a solvent can no longer dissolve a solute it is said to be saturated with the solute. Solvents are often used to convert substances to a form which is suitable for a particular use. Many substances exhibit the greatest usefulness when they are in solution.

When the solute (solid, liquid or gas) dissolves it can either break up into separate molecules when it enters solution (as happens when sugar is dissolved in water) or the molecules can dissociate into ions (as happens when sodium chloride is dissolved in water). In both cases the process is a physical one since the solid can be reclaimed from the solution (boiling the NaCl solution and crystallizing the sugar out the sugar solution). The tendency of the solute to dissolve in the solvent depends on how the solute and the solvent interact, solvents therefore need to be classified in order to understand how a solvent may interact with a prospective solute. This interaction is represented by the activity coefficient (see section 4.1). The lower the activity coefficient the better the solute will dissolve (as seen in section 5.1, this is sometimes not the only factor).

Some examples of the broad range of solvent applications are listed in the next paragraphs. The list is not meant to be an exhaustive one but rather to give an idea of the broad scope of solvent application (the list contains desirable properties for the solvent, the ability to dissolve the active ingredient is taken as a given - and is given much attention in later chapters).

2.1. Viscosity

The viscosity of a substance can be defined as a measure of the fluids resistance to flow. For most applications it is useful to have a fluid which has a low viscosity because of the operational costs associated with using a viscous fluid (high pumping heads needed). Solvents are very often used to lower the viscosity of certain substances so that they are in a usable form. Some examples of this are:

- Polymer adhesives which contain polymers which need to be dissolved in a solvent so that they can be applied to the surfaces,
- Pharmaceuticals which need to be dispensed in liquid form (capsule or syrup),
- Pesticides which need to be applied to crops in a (usually) liquid form and

- When a reactant in very viscous solvents can be used as a reaction medium (the purpose of this is very often twofold, firstly to alter the viscosity and secondly to alter the reaction equilibrium)

For this reason it is almost always desirable to have a solvent with a low viscosity.

2.2. Volatility

The volatility of a substance can be defined as the rate at which it evaporates. The volatility of a substance is usually gauged from the vapour pressure. A substance with a high vapour pressure would have a high volatility and *visa versa*. For solvents the volatility can either be a desirable or an undesirable property. A solvent with a high relative volatility is useful when the solvent needs to be removed so that the substance can dry. Examples of this are paints, coatings (varnishes etc.) and adhesives. Also, when solvent extraction is used for a low-volatility solute it is useful to have a solvent with a sufficiently higher volatility so that downstream separation is simplified.

Highly volatile solvents are undesirable when the substance is harmful to the environment. It is for this reason that more and more attention has been given to green solvents which are more eco-friendly. It is obviously undesirable to have a highly volatile solvent when the solvent is flammable. In order to determine how safe the solvent is the flash point is used as a constraint. The flash point is the lowest temperature at which there will be sufficient vapours above the mixture to ignite.

2.3. Toxicity and environmental impact

Toxicity is a measure of how harmful (toxic) a substance is. Two common measures of toxicity are LC₅₀ (lethal concentration which causes a 50% mortality rate) and LD₅₀ (lethal dose which causes a 50% mortality rate). In most applications it is desirable to have a solvent with a low toxicity. For applications in the pharmaceutical industries the toxicity of the solvents used is obviously of paramount importance; for both the preparation and dispensing of the pharmaceuticals the toxicity of the solvent needs to be known. In more industrial applications the toxicity is not such a limiting factor.

In recent times the toxicity of solvents (as well as other chemicals) has been given special attention and the branch of green chemistry has been formed to find alternatives to the hazardous substances in use. Green solvents are solvents which have very little environmental impact. Ionic liquids are often touted as green solvents due to their negligible vapour pressures; this is however not always true since they can still be toxic. Another consideration is the corrosivity of a solvent, since this determines what the solvent can and can't be used with.

2.4. Selectivity and capacity

The selectivity of a solvent is defined as the ratio of the “affinity” of the solvent for some solute and some undesired compound. The selectivity (β) can be defined in terms of the infinite dilution activity coefficients (the reciprocal of the infinite dilution activity coefficient is used as a measure of the affinity - see section 4.1) as:

$$\beta_{ij} = \frac{\gamma_{i,S}^{\infty}}{\gamma_{j,S}^{\infty}} \quad (2-1)$$

where subscript j refers to the solute, i to some undesired compound and S to the solvent. In applications when solvents are used for extraction (e.g. liquid-liquid extraction) the selectivity is a very important property that the solvent needs to have. Another application where a selective solvent is especially desirable is detergents where the cleaning product must dissolve the dirt and grime but not damage what it is cleaning.

In applications involving liquid-liquid extraction, another useful quantity is the solvent capacity. The capacity of the solvent (m) relative to the other carrier phase is:

$$m = \frac{\gamma_{i,A}^{\infty}}{\gamma_{i,S}^{\infty}} \quad (2-2)$$

where the superscript i refers to the solute, A to the carrier phase and S to the solvent.

2.5. Liquidus range

The liquidus range is the temperature range for which a fluid is a liquid. The 2 bounds of the liquidus range are the melting (lower bound) and the boiling point (upper bound). The liquidus range is usually considered in crystallization applications. The liquidus range is important because the solvent needs to be liquid when the solute is dissolved at higher temperatures and also when the solute is crystallized out at lower temperatures (for cooling crystallization).

2.6. Computer Aided Molecular Design (CAMD)

CAMD is the process whereby a compound is selected based on a certain property. The properties are not measured but are obtained only from the molecular structure. CAMD is basically the reverse of property estimation, the main property estimation technique used with CAMD is the group contribution methods. A wide range of group contribution methods (of varying accuracy) are available for the properties mentioned above.

CAMD has received much attention in literature for crystallization [7] and liquid extraction [8, 9] solvent selection. There has also been some attention given in the literature to green solvent selection [10], whereby special attention is given to the environmental impact of the

solvent. When using CAMD it is usually desirable to specify a set of groups that can be used in order to reduce the problem variables. For example one may specify that only saturated aliphatic compounds with oxygen groups may be used. When using certain groups there should obviously be some restrictions placed on the group selection (e.g. when using aromatic groups enough groups should be present to form the rings). When running the algorithm, specific values are not usually set, rather ranges are specified (for example a viscosity < 1 cP).

Therefore in theory it would be possible to obtain a structure of a hypothetical solvent which has all of the desirable properties. The problem is that even if a hypothetical structure could be found it would more than likely not be a simple, readily available compound and would therefore need to be synthesised which is time consuming, expensive and impractical. It is therefore very often that, for large scale use, the economic considerations limit the solvent select to a list of probably less than 50 different solvents.

3. CLASSIFICATION OF SOLVENTS

Solvents are usually liquids, however supercritical fluids have received some attention since they have no surface tension and low viscosities but will not be considered in this work. Water is the most common solvent. Solvents can be broadly classified into 3 groups:

- molecular liquids (e.g. water, hexane etc)
- atomic liquids (e.g. liquid mercury)
- ionic liquids (e.g. 1-butyl-3-methylimidazolium tetrafluoroborate)

Solvents can either fall into one or a mixture of the three classes. Of the three classes, atomic liquids are the least common solvents, especially in “everyday use”. Ionic liquids have received some attention in recent times due to their environmental friendliness (virtually no vapour pressure) and ability to almost be designed specifically for any purpose (Reichardt [11]). Molecular liquids are by far the most common solvents in use; they find application in many areas from adhesives to pharmaceuticals (see section 2 for some applications).

Solvents can be classified in terms of their physical behaviour. For example they can be split into solvents which are high, medium and low boilers or solvents with high, medium or low viscosity. These classifications make a lot of sense when considering operational costs and design specification but make little sense when trying to classify solute-solvent interactions. It is clear that in order to account for these a chemical classification is necessary.

The basic principle when determining solubility is “like dissolves like”, where molecules with similar intermolecular forces will generally dissolve. The intermolecular forces which make up molecules are hydrogen bonding; polar forces (dipole-dipole) and dispersive forces (van der Waal’s forces). All compounds have dispersive interactions between the molecules but not all contain hydrogen bonding and polar forces. The problem is therefore to determine which intermolecular forces predominate in the molecule. The various approaches for determining these forces are given in the following paragraphs.

3.1. Trouton’s rule

Trouton’s rule states that the entropy of vaporization ($\Delta_b S$) at the normal boiling point (T_b) has a constant value of $88 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ for all compounds.

$$\Delta_b S = \frac{\Delta_b H}{T_b} \approx 88 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1} \quad (3-1)$$

This rule holds for a fair number of compounds but deviates especially for hydrogen bonding compounds (see Table 1). As shown in Table 1 there is a positive deviation from Trouton’s

rule for all compounds which hydrogen bond. The reason for this is that the hydrogen bonding increases the heat of vaporization and therefore causes the entropy of vaporization to rise. This deviation means that depending on the entropy of vaporization the compounds can either be classified as hydrogen bonding or non-hydrogen bonding.

Unfortunately this deviation is not always so clear cut, for example acetic acid, which is hydrogen bonding, has an entropy of vaporization of $60.5 \text{ J.mol}^{-1}.\text{K}^{-1}$. The reason for this is that the small carboxylic acids form dimmers, the dimmers behave like non-polar molecules and therefore decrease the heat of vaporization (the vapour phase also forms dimmers) and therefore a lower entropy of vaporization. Also large molecules tend to deviate from Trouton's rule for example the entropy of vaporization of eicosane is $93.2 \text{ J.mol}^{-1}.\text{K}^{-1}$. Therefore Trouton's rule is a useful means of classification but cannot be blindly applied.

Table 1 Examples of the entropy of vaporization at the normal boiling point data from Myrdal et al. [12]

Name	$\Delta S_b \text{ (J.mol}^{-1}.\text{K}^{-1})$
Water	109.0
Ethanol	110.1
n-Octane	86.3
Acetone	88.4
Pentanoic acid	96.2
Diethyl ether	86.5
Benzene	87

3.2. Cohesive pressure

The cohesive pressure (also called the cohesive energy density) of a substance is defined by the following relationship:

$$c = \frac{\Delta_v U}{V} \approx \frac{\Delta_v H - RT}{V} \quad (3-2)$$

The cohesive pressure is used as a measure of the total intermolecular forces in a molecule. The reason for this is that in order to vaporize a fluid all the intermolecular forces need to be broken, and hence the heat of vaporization will contain this information. Large values of cohesive pressure indicate compounds which are polar/H-bonding, while lower values indicate compounds which are non-polar. The problems associated with dumping all these intermolecular forces into a single parameter are discussed in section 4.2.3.1.

3.3. Internal pressure

The internal pressure of a substance is defined by the following relationship:

$$\pi = \left(\frac{\partial U}{\partial V} \right)_T \quad (3-3)$$

The internal pressure is used as a measure of the non-hydrogen-bonding forces present (i.e. polar and dispersive forces). This is possibly due to the fact that polar and dispersive forces

are “always on” while hydrogen bonding forces are more “on/off” (i.e. polar and dispersive forces may get stronger or weaker depending on molecule proximity while hydrogen bonds are either on or off). Therefore when the molar volume changes the hydrogen bonding forces will not be disrupted, while the others will. The internal and cohesive pressures are often used together as a measure of the types of forces present in a molecule:

$$n = \frac{c}{\pi} \quad (3-4)$$

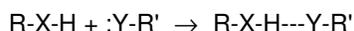
where values of n much greater than unity indicate hydrogen bonding molecules and those which are close to unity are non-polar molecules. Consider the examples shown in Table 2, water clearly has a large amount of hydrogen bonding while toluene does not.

Table 2 Internal and cohesive pressure for some sample compounds, data from Reichardt [11]

Name	π (MPa)	c (MPa)	n
Perfluoroheptane	215	151	0.70
Toluene	355	337	0.95
Ethanol	293	687	2.34
Water	151	2294	15.19

3.4. Acid-base classification

A common method to classify solvents is by their acid-base behaviour. According to the Brønsted/Lowry acid-base theory acids are proton donors and bases are proton acceptors. This is more or less what happens with hydrogen bonding, except that the proton is not transferred but shared:



where XH is some acidic group and Y is some basic group. The strength of an acid is measured by the pK_a , where for some acid HA:

$$pK_a = -\log \frac{[A^-][H^+]}{[HA]} \quad (3-5)$$

Therefore using the pK_a substances can be classified as protic or non-protic.

According to the Lewis acid-base theory, an acid is an electron pair acceptor while a base is an electron donor. A Lewis base is a Brønsted/Lowry base but a Lewis acid is not necessarily a Brønsted/Lowry acid. For further classification the acids and bases are classified as hard or soft, where a hard acid or base is usually derived from small atoms with high electronegativities and low polarizabilities. These are typically polar or hydrogen bonding compounds, for example water. A soft acid or base is usually derived from larger atoms with low electronegativities and are usually polarizable. These are typically non-polar or slightly polar compounds for example alkanes.

3.5. Solvent classification

The advantage of a good solvent classification is that solvents in the same categories should behave in a similar way; qualitative predictions of solvent behaviour can therefore be gleaned from analogous solvents. Organic solvents can be very generally classified into 3 categories:

- Apolar, aprotic
- Polar, aprotic
- Protic

This classification is however probably too broad to observe any meaningful trends. Examples of two, more complex classifications are shown in Table 3.

Table 3 Examples of two more complex solvent classifications

Chastrette et al. [13] classification	Gramatica et al. [14] classification
POLAR	
aprotic dipolar (AD) aprotic highly dipolar (AHD) aprotic highly dipolar and polarizable (AHDP)	aprotic polar (AP)
APOLAR	
aromatic apolar (ARA) aromatic relatively polar (ARP) electron pair donors (EPD)	aromatic apolar or lightly polar (AALP) electron pair donors (EPD)
PROTIC	
hydrogen bonding (HB) hydrogen bonding strongly associated (HBSA)	hydrogen bonding donors (HBD)
MISC	
miscellaneous (MISC)	aliphatic aprotic apolar (AAA)

The solvent classification should be able to explain most of the extraordinary phase splitting that occurs. For instance the famous four phase mixture of Hildebrand [15], as shown in Fig. 1 in which all four fluids are sufficiently different to form 4 distinct phases.

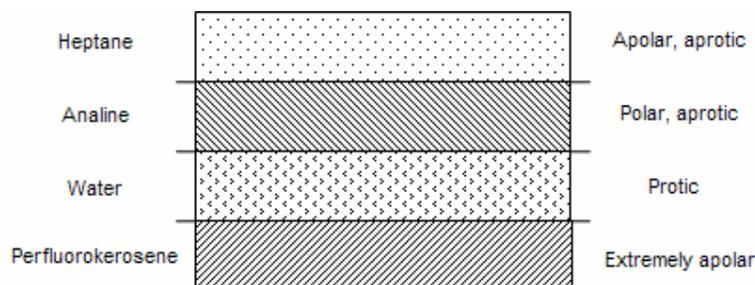


Fig. 1 The four phase splitting example of Hildebrand [15]

4. THEORETICAL APPROACHES TO SOLUTE ACTIVITY

4.1. Thermodynamic framework

For a single phase fluid in a closed system where no reaction occurs, the Gibbs energy can be related to temperature and pressure by:

$$d(nG) = (nV)dP - (nS)dT \quad (4-1)$$

where

$$\left[\frac{\partial(nG)}{\partial P} \right]_{T,n} = nV \quad (4-2)$$

$$\left[\frac{\partial(nG)}{\partial T} \right]_{P,n} = -nS \quad (4-3)$$

A more general case is where there is a single phase open system where mass can be transferred across the system boundary, the Gibbs energy for this case is then not only a function of temperature and pressure but also of the number of moles of each species (for m species):

$$nG = f(P, T, n_1, \dots, n_m) \quad (4-4)$$

Taking the total differential of Eqn. (4-4) and inserting Eqn's (4-2) & (4-3) the following is obtained:

$$d(nG) = (nV)dP - (nS)dT + \sum_{i=1}^m \left[\frac{\partial(nG)}{\partial n_i} \right]_{P,T,n_j} dn_i \quad (4-5)$$

where the chemical potential of species i is defined as (the overbar represents a partial molar property):

$$\mu_i \equiv \left[\frac{\partial(nG)}{\partial n_i} \right]_{P,T,n_j} = \bar{G}_i \quad (4-6)$$

Now consider a closed system of two phases (α and β) in equilibrium, where each phase is an open system:

$$d(nG)^\alpha = (nV)^\alpha dP - (nS)^\alpha dT + \sum_{i=1}^m \mu_i^\alpha dn_i^\alpha \quad (4-7)$$

$$d(nG)^\beta = (nV)^\beta dP - (nS)^\beta dT + \sum_{i=1}^m \mu_i^\beta dn_i^\beta \quad (4-8)$$

In Eqn's (4-7) and (4-8) thermal and mechanical equilibrium are assumed and therefore temperature and pressure are uniform throughout the whole system. The Gibbs energy differential for the whole system is simply the sum of the Gibbs energy in each phase (where $(nV) = (nV)^\alpha + (nV)^\beta$ and $(nS) = (nS)^\alpha + (nS)^\beta$):

$$d(nG) = (nV)dP - (nS)dT + \sum_{i=1}^m \mu_i^\alpha dn_i^\alpha + \sum_{i=1}^m \mu_i^\beta dn_i^\beta \quad (4-9)$$

Since the whole system is closed, Eqn. (4-1) holds, also since the system is closed mass lost by one phase must be equal to mass gained by another and therefore at equilibrium:

$$\sum_{i=1}^m (\mu_i^\alpha - \mu_i^\beta) dn_i^\alpha = 0 \quad (4-10)$$

Therefore the chemical potentials in each phase must be equal at equilibrium; this relation can be easily extended to systems with π phases to yield:

$$T^{(1)} = T^{(2)} = \dots = T^{(\pi)} \quad (4-11)$$

$$P^{(1)} = P^{(2)} = \dots = P^{(\pi)} \quad (4-12)$$

$$\mu_i^{(1)} = \mu_i^{(2)} = \dots = \mu_i^{(\pi)} \quad i = 1..m \quad (4-13)$$

For each phase there will be $m+2$ (T , P and μ 's) intensive variables, however only $m+1$ of these variables are independent. Therefore for the total system there will be $\pi(m+1)$ independent variables. The number of equilibrium relations given by Eqn's (4-11)-(4-13) is $(\pi-1)(m+2)$. Therefore the number of intensive variables that can be set (known as the degrees of freedom of the system) can be given by:

$$F = \pi(m+1) - (\pi-1)(m+2) = m+2 - \pi \quad (4-14)$$

The chemical potential of a pure species is related to the temperature and pressure through an equation analogous to Eqn. (4-1):

$$d\mu_i = v_i dP - s_i dT \quad (4-15)$$

This equation can be integrated by using some reference state (superscript r):

$$\mu_i(T, P) = \mu_i(T^r, P^r) - \int_{T^r}^T s_i dT + \int_{P^r}^P v_i dP \quad (4-16)$$

The reference state, often called the standard state, can be arbitrarily specified, but is usually specified so as to simplify the required calculations (see section 5.1 for an example).

The chemical potential is somewhat of an abstract concept and doesn't have a direct equivalent in the real world. It is therefore desirable to express the chemical potential in terms of some function which does have a real world equivalent. Assuming a pure ideal gas ($Pv = RT$), and integrating Eqn. (4-15) under isothermal conditions (using superscript 0 to denote the standard state):

$$\mu_i - \mu_i^0 = RT \ln \frac{P}{P^0} \quad (4-17)$$

The advantage of Eqn. (4-17) is that it simply relates temperature and pressure to the more abstract chemical potential. The disadvantage is that it is only suitable for pure ideal gases. This problem was overcome by Lewis by defining a function called fugacity (f , when used in solution it is given the symbol \hat{f}_i), where fugacity replaces the pressure for real solids, liquids and gases:

$$\mu_i - \mu_i^0 = RT \ln \frac{\hat{f}_i}{f_i^0} \quad (4-18)$$

The ratio \hat{f}_i / f_i^0 was called the activity (a_i) by Lewis (the fugacity in solution is not used for the standard state since it is more convenient to choose the standard state as the pure liquid, the superscript 0 is therefore often left off). Eqn. (4-18) represents an isothermal change in chemical potential when going from f_i^0 to \hat{f}_i . Therefore the standard states for all components in all phases must be at the same temperatures; the pressure and composition of the standard states, however, do not necessarily have to be the same. The phase equilibrium criterion given in Eqn. (4-13) can therefore be rewritten in terms of fugacity as:

$$\hat{f}_i^{(1)} = \hat{f}_i^{(2)} = \dots = \hat{f}_i^{(x)} \quad i = 1..m \quad (4-19)$$

The activity coefficient is defined as the ratio of the fugacity in solution to the ideal fugacity in solution, it can also be defined in terms of the activity relative to some measure of composition:

$$\gamma_i \equiv \frac{\hat{f}_i}{\hat{f}_i^{id}} = \frac{a_i}{x_i} = \frac{\hat{f}_i}{x_i f_i^0} \quad (4-20)$$

The activity coefficient is a measure of the deviation from ideality. For ideal solutions the activity coefficient is unity. The two important limits when working with activity coefficients are:

$$\lim_{x_i \rightarrow 1} \gamma_i \equiv 1 \quad (4-21)$$

$$\lim_{x_i \rightarrow 0} \gamma_i \equiv \gamma_i^\infty \quad (4-22)$$

where γ_i^∞ is called the infinite dilution activity coefficient. The activity coefficient can be defined in terms of Gibbs energy by using the following procedure. Taking the difference Eqn (4-18) for the real and the ideal chemical potential (partial molar Gibbs energy is used for the sake of notation – see Eqn. (4-6)).

$$\bar{G}_{i(\text{real})} - \bar{G}_{i(\text{ideal})} = RT \left[\ln \hat{f}_{i(\text{real})} - \ln \hat{f}_{i(\text{ideal})} \right] \quad (4-23)$$

This can be rewritten in terms of excess partial molar Gibbs energy (where an excess property is simply the difference between the real and the ideal solution denoted by superscript E).

$$\bar{G}_i^E = RT \ln \left[\frac{\hat{f}_i}{\hat{f}_{i(\text{ideal})}} \right] \quad (4-24)$$

To simplify the notation the subscript real was dropped. Substituting Eqn. (4-20) (recalling that $a_i = \hat{f}_i / f_i^0$) into Eqn. (4-24) results in the relationship between excess Gibbs energy and activity:

$$\frac{\bar{G}_i^E}{RT} = \ln \gamma_i \quad (4-25)$$

Therefore if an analytical expression for the concentration dependence of the excess Gibbs energy is known then it is a fairly simple process to find an analytical expression for the activity coefficient. Some of the more popular expressions for excess Gibbs energy are shown in the sections that follow.

The temperature and pressure dependence of the activity coefficient can be shown using the following procedure. Taking the total differential of the excess partial Gibbs energy yields the following:

$$d \left(\frac{\bar{G}_i^E}{RT} \right) = \left(\frac{\partial \bar{G}_i^E}{\partial P} \right)_{P,x} \frac{dP}{RT} + \left(\frac{\partial \bar{G}_i^E}{\partial T} \right)_{T,x} \frac{dT}{RT} + \sum_i \left(\frac{\partial \left(\sum_j \ln \gamma_j dn_j \right)}{\partial n_i} \right) dn_i \quad (4-26)$$

Eqn. (4-5) can be rewritten in terms of excess properties as:

$$d \left(\frac{nG^E}{RT} \right) = \frac{nV^E}{RT} dP - \frac{nH^E}{RT^2} dT + \sum_i \ln \gamma_i dn_i \quad (4-27)$$

Combining Eqn. (4-26) and the partial derivative of Eqn. (4-27) leads to the following temperature and pressure dependencies of the activity coefficient:

$$\left(\frac{\partial \ln \gamma_i}{\partial P}\right)_{T,x} = \frac{\bar{V}_i^E}{RT} \quad (4-28)$$

$$\left(\frac{\partial \ln \gamma_i}{\partial T}\right)_{P,x} = -\frac{\bar{H}_i^E}{RT^2} \quad (4-29)$$

At temperatures far removed from the critical point the pressure dependence is usually ignored, however the temperature dependence is very often non-trivial and needs to be given some attention when developing activity coefficient models.

4.2. Empirical and semi-empirical models for G^E

The following section outlines the main models that are available in the literature for the representation of excess Gibbs free energy. Polynomial type expansions of G^E such as Redlich-Kister, Margules and van Laar are not covered in the following sections since they are fairly well known (and widely covered in the literature [16-18]) and the local composition models have been shown to be more suitable [19]. The models covered in the following sections are by no means an exhaustive list; rather they are the ones which find the widest use in the prediction/modelling of complex molecule solubility.

4.2.1. Local composition concept (Wilson, NRTL, UNIQUAC)

Wilson is widely credited with developing the local composition concept, however Guggenheim was probably the first, nevertheless Wilson was the first to provide a practical application of the concept. The concept of local composition assumes that the composition in the region of some molecule is different from the composition in the bulk liquid. Wilson expressed the local compositions (ξ_{11} is the volume fraction of molecules 1 around a molecule of 1, similarly ξ_{21} is the volume fraction of molecule 2 around molecule 1) for a binary system as:

$$\xi_{11} = \frac{x_1}{x_1 + x_2 \left(\frac{v_2}{v_1}\right) \exp\left(-\frac{\lambda_{21} - \lambda_{11}}{RT}\right)} \quad (4-30)$$

$$\xi_{11} + \xi_{21} = 1 \quad (4-31)$$

$$\xi_{22} = \frac{x_2}{x_2 + x_1 \left(\frac{v_1}{v_2}\right) \exp\left(-\frac{\lambda_{12} - \lambda_{22}}{RT}\right)} \quad (4-32)$$

$$\xi_{12} + \xi_{22} = 1 \quad (4-33)$$

where λ_{12} is the interaction between component 1 and 2. Then by using an equation analogous to the Flory-Huggins expression for athermal solutions he defined the excess Gibbs energy of mixing as:

$$\frac{G^E}{RT} = x_1 \ln \left(\frac{\xi_{11}}{x_1} \right) + x_2 \ln \left(\frac{\xi_{22}}{x_2} \right) \quad (4-34)$$

By substitution of Eqn's (4-30) & (4-32) into Eqn. (4-34) the following results:

$$\frac{G^E}{RT} = -x_1 \ln(x_1 + x_2 A_{21}) - x_2 \ln(x_1 A_{12} + x_2) \quad (4-35)$$

where

$$A_{12} = \frac{v_1}{v_2} \exp \left(-\frac{\lambda_{12} - \lambda_{22}}{RT} \right) = \frac{v_1}{v_2} \exp \left(-\frac{\Delta\lambda_{12}}{RT} \right) \quad (4-36)$$

$$A_{21} = \frac{v_2}{v_1} \exp \left(-\frac{\lambda_{21} - \lambda_{11}}{RT} \right) = \frac{v_2}{v_1} \exp \left(-\frac{\Delta\lambda_{21}}{RT} \right) \quad (4-37)$$

(Note: this derivation is the one given by Wilson [20]; there are however different ways to arrive at the same result - see Tsuboka and Katayama [21]) The derivation above is only shown for a binary system; however it can be easily extended to multi-component systems. This is the great advantage of the local composition equations; they can predict multi-component behaviour from binary data. The Wilson equation is still widely used today, its biggest flaw is that it cannot account for 2 liquid phases (since when the phase stability criterion is investigated it always results in a positive number and therefore only a single stable phase [18]). The general form of the activity coefficient for the Wilson equation of some component i in a mixture of m components is:

$$\ln \gamma_i = -\ln \left(\sum_{j=1}^m x_j A_{ji} \right) + 1 - \frac{\sum_{k=1}^m x_k A_{ik}}{\sum_{j=1}^m x_j A_{jk}} \quad (4-38)$$

Since the initial work of Wilson, many local composition models have been developed (including modifications to the Wilson equation which allow phase splitting [21, 22]). The two most popular and widely used are the NRTL (Non-Random Two Liquid) and UNIQUAC (UNiversal QUasi Chemical) equations. The NRTL [23] equation introduced a 3rd parameter call the non-randomness parameter (α_{ij} , where $\alpha_{ij} = \alpha_{ji}$), local mole fractions (as with local volume fractions $x_{11} + x_{21} = 1$ and $x_{12} + x_{22} = 1$) are used and the excess Gibbs energy is defined differently:

$$x_{21} = \frac{x_2 G_{21}}{x_1 + x_2 G_{21}} \quad (4-39)$$

$$x_{12} = \frac{x_1 G_{12}}{x_2 + x_1 G_{12}} \quad (4-40)$$

$$G_{ij} = \exp(-\alpha_{ij} \tau_{ij}) \quad (4-41)$$

$$\tau_{ij} = \frac{g_{ij} - g_{jj}}{RT} \quad (4-42)$$

$$\frac{G^E}{RT} = x_1 x_{21} \frac{g_{21} - g_{11}}{RT} + x_2 x_{12} \frac{g_{12} - g_{22}}{RT} \quad (4-43)$$

As with the Wilson equation, the NRTL equation is widely used, it is superior to the Wilson equation in that it can account for 2 (or more) liquid phases. A disadvantage of the NRTL equation is that for LLE (liquid-liquid equilibria) α_{12} must be set and cannot be solved for. The general form of the activity coefficient for the NRTL equation of some component i in a mixture of m components is:

$$\ln \gamma_i = \frac{\sum_{j=1}^m \tau_{ji} G_{ji} x_j}{\sum_{l=1}^m G_{li} x_l} + \sum_{j=1}^m \frac{x_j G_{ij}}{\sum_{l=1}^m G_{lj} x_l} \left(\tau_{ij} - \frac{\sum_{r=1}^m x_r \tau_{rj} G_{rj}}{\sum_{l=1}^m G_{lj} x_l} \right) \quad (4-44)$$

The UNIQUAC equation is somewhat of a compromise between Wilson and NRTL in that it has only 2 model parameters but can account for phase splitting. The derivation of the UNIQUAC equation is given by Maurer [24], the basic idea is that the excess Gibbs free energy is split into 2 parts. One part for the energetic interactions between molecules (called the residual) and one for the effect of size/shape interaction (called the combinatorial):

$$\frac{G^E}{RT} = \left(\frac{G^E}{RT} \right)_{\text{Residual}} + \left(\frac{G^E}{RT} \right)_{\text{Combinatorial}} \quad (4-45)$$

Following the derivation given by Maurer [24] the following expressions result for the residual and combinatorial expressions (for the binary case):

$$\left(\frac{G^E}{RT} \right)_{\text{Residual}} = -x_1 q_1 \ln(\theta_1 + \theta_2 \psi_{21}) - x_2 q_2 \ln(\theta_1 \psi_{12} + \theta_2) \quad (4-46)$$

$$\left(\frac{G^E}{RT} \right)_{\text{Combinatorial}} = x_1 \ln \left(\frac{\Phi_1}{x_1} \right) + x_2 \ln \left(\frac{\Phi_2}{x_2} \right) + \frac{z}{2} \left[q_1 x_1 \ln \left(\frac{\theta_1}{\Phi_1} \right) + q_2 x_2 \ln \left(\frac{\theta_2}{\Phi_2} \right) \right] \quad (4-47)$$

$$\psi_{ij} = \exp\left[-\frac{u_{ij} - u_{jj}}{RT}\right] = \exp\left[-\frac{\Delta u_{ij}}{RT}\right] \quad (4-48)$$

where r and q are the size and surface area parameters (defined in terms of van der Waals volume and surface area) respectively, ψ_{ij} is the binary interaction parameter, Φ is the volume fraction and θ is the surface fraction (r and q and defined by Eqn's (4-60) and (4-61)):

$$\Phi_i = \frac{x_i r_i}{\sum_j x_j r_j} \quad (4-49)$$

$$\theta_i = \frac{x_i q_i}{\sum_j x_j q_j} \quad (4-50)$$

Taking the partial derivative of Eqn. (4-45) yields the following relationship for activity coefficient:

$$\ln \gamma_i = \ln \gamma_i^R + \ln \gamma_i^C \quad (4-51)$$

$$\ln \gamma_i^R = q_i \left[1 - \ln \left(\sum_j \theta_j \psi_{ji} \right) - \sum_j \frac{\theta_j \psi_{ij}}{\sum_k \theta_k \psi_{kj}} \right] \quad (4-52)$$

$$\ln \gamma_i^C = 1 - \frac{\Phi_i}{x_i} + \ln \frac{\Phi_i}{x_i} - 5q_i \left(1 - \frac{\Phi_i}{\theta_i} - \ln \frac{\Phi_i}{\theta_i} \right) \quad (4-53)$$

where the superscripts C and R refer to the combinatorial and residual parts respectively.

4.2.2. Functional group activity coefficients

Probably the most widely known and used of the activity coefficient prediction methods is the group contribution methods. The prediction of activity coefficients by the concept of functional group contribution (also known as the solution of groups concept) is generally attributed to the work of Wilson and Deal [25]. As the name suggests solution of groups entails breaking a molecule up in to functional groups and considering mixtures as solutions of these groups and not compounds. For example consider a mixture of ethanol and hexane, the mixture would be made up of only alkane (CH₂) and alcohol (OH) groups (see Table 4).

Table 4 Group frequency illustration for the mixture hexane - ethanol

Groups	Ethanol	Hexane	Mixture
CH ₂	2	6	8
OH	1	0	1

Deal and Derr [26, 27] were the first to develop a model for the solution of groups; it was called the ASOG (Analytical Solution Of Groups) method. Later Fredenslund et al. [28]

developed the UNIFAC (UNIversal Functional group Activity Coefficient) model; both models are the same in principle but differ in the details. The UNIFAC model has subsequently had many modifications and these are developed for specific applications [29, 30].

The great advantage of the solution of groups approach is that the number of possible functional groups is much less than the amount of compounds, this means that the number of binary interaction parameters needed is much less. For example if the CH₂-OH interactions (model parameters in the ASOG and UNIFAC equations) are known then any alkane - aliphatic alcohol mixture can be predicted, drastically reducing the number of experimental measurements needed.

As mentioned above, ASOG was the first method to use the principle of the solution of groups. In a similar way to the UNIQAUC equation the activity coefficient is assumed to be made up of a part for molecular forces interactions ($\ln \gamma_i^G$) and a part for size/shape interactions ($\ln \gamma_i^{FH}$).

$$\ln \gamma_i = \ln \gamma_i^G + \ln \gamma_i^{FH} \quad (4-54)$$

The size/shape term is calculated from an equation which is analogous to the Flory-Huggins equation for athermal solutions (see section 4.2.4):

$$\ln \gamma_i^{FH} = \ln \frac{v_i^{FH}}{\sum_j v_j^{FH} x_j} + 1 - \frac{v_i^{FH}}{\sum_j v_j^{FH} x_j} \quad (4-55)$$

where v_i^{FH} is the (total) number of size groups in molecule i , and x_i is the mole fraction. The group interactions are related to the molecular interactions by the following equation:

$$\ln \gamma_i^G = \sum_k v_{ki} (\ln \Gamma_k - \ln \Gamma_k^{(i)}) \quad (4-56)$$

where v_{ki} is the frequency of group k in molecule i , Γ_k is the activity coefficient of group k in the mixture and $\Gamma_k^{(i)}$ is the activity coefficient of group k in the pure component i . For compounds with only one group (e.g. hexane), $\Gamma_k^{(i)}$ will be zero, however for compounds with multiple groups (e.g. ethanol) it will have a nonzero value. The group activity coefficients are calculated from the Wilson equation (Eqn. (4-38)):

$$\ln \Gamma_k = -\ln \sum_m X_m a_{mk} + \left(1 - \frac{X_m a_{km}}{\sum_n X_n a_{nm}} \right) \quad (4-57)$$

$$X_m = \frac{\sum_j X_j v_{mj}}{\sum_j X_j \sum_n v_{nj}} \quad (4-58)$$

where X_m is the group fraction and a_{km} is given by equations similar to Eqn (4-36), but for the interaction between groups and not compounds. $\Gamma_k^{(i)}$ is calculated from the same equation but instead of X_m , $X_m^{(i)}$ is used, where:

$$X_m^{(i)} = \frac{v_m^{(i)}}{\sum_n v_n^{(i)}} \quad (4-59)$$

UNIFAC follows the same scheme as UNIQUAC (Eqn.(4-51)) and depending on which UNIFAC is being used (original, modified Dortmund etc.) the combinatorial expression changes (see section 4.2.4). For original UNIFAC the combinatorial expression of UNIQUAC (Eqn. (4-53) is used, where the surface and volume parameters are calculated from the following relations:

$$r_i = \sum_k v_k^{(i)} R_k \quad \text{and} \quad q_i = \sum_k v_k^{(i)} Q_k \quad (4-60)$$

$$R_k = \frac{V_{wk}}{15.17} \quad \text{and} \quad Q_k = \frac{A_{wk}}{2.5 \times 10^9} \quad (4-61)$$

where R_k and Q_k are the group volume and surface parameters respectively, and V_{mk} and A_{mk} are the group van der Waals volume and surface area as given by Bondi [31].

The expression for the residual is analogous to Eqn. (4-56), given as:

$$\ln \gamma_i^R = \sum_k v_k^{(i)} (\ln \Gamma_k - \ln \Gamma_k^{(i)}) \quad (4-62)$$

However unlike with ASOG the group activity coefficients are calculated from the UNIQUAC residual term:

$$\ln \Gamma_k = Q_k \left[1 - \ln \left(\sum_m \theta_m \Psi_{mk} \right) - \sum_m \frac{\theta_m \Psi_{km}}{\sum_n \theta_n \Psi_{nm}} \right] \quad (4-63)$$

$$\theta_m = \frac{Q_m X_m}{\sum_n Q_n X_n} \quad (4-64)$$

where Ψ_{mk} is given by an equation analogous to Eqn. (4-48) but is for the interaction between groups and not molecules, and as with ASOG, $\Gamma_k^{(i)}$ is calculated from the same expression with X_m being replaced by $X_m^{(i)}$. It should be noted that with both methods the residual is

calculated from the solution of groups while the combinatorial must be calculated from the molecule properties.

As mentioned above, these group contribution methods have found wide use in industry. The applications for which they work the best are for molecules of similar sizes. The reason for this is that most of the data to which the binary interaction data was fitted to is VLE data. Nevertheless they can be applied to the prediction of solutions containing large molecules. This is typically the case when a solid is dissolved in a solvent (see section 5.1 for the calculation details). As a means of comparison the methods in this and the following sections will be used to generate predictions of the solubility of aspirin (2-acetoxybenzoic acid) in a variety of solvents at 298.15 K. The molecular structure of aspirin and the various thermo physical properties required are shown in Fig. 2.

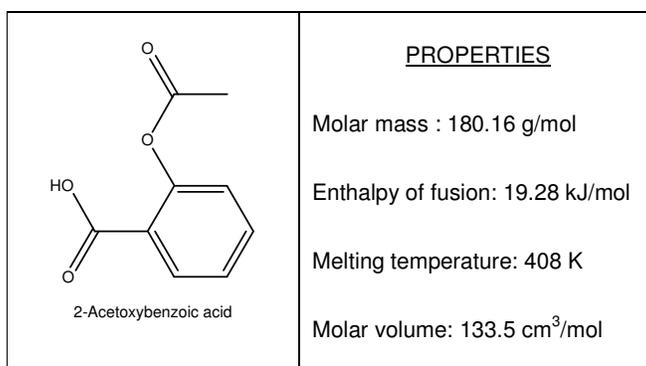


Fig. 2 The molecular formula and thermo physical properties required for aspirin (all data taken from Beilstein [32])

The results for the aspirin test set data are shown in Table 5. The results for the ASOG method are not included as there are many missing groups and application to SLE is very uncommon. In many instances the predicted results are very good, only for cyclohexane and chloroform are there extremely large deviations. As mentioned much above, the great power in the group contribution type methods is that the results are purely predictive and do not require any raw data (fusion data is obviously still required) but still provide very good predictions.

Table 5 Comparison of the experimental and predicted solubility of aspirin when using group contribution methods (experimental data from Frank [19])

Solvent	wt% _{experimental}	wt% _{UNIFAC}	wt% _{mod-UNIFAC}
1,1,1-Trichloroethane	0.5	8.7	-
1,2-Dichloroethane	3	12.9	7.7
1,4-Dioxane	19	19.0	35.9
2-Ethylhexanol	10	2.9	2.3
Acetic acid	12	34.6	22.9
Acetone	29	34.5	34.7
Chloroform	6	22.2	40.3
Cyclohexane	0.005	0.3	0.2
Diacetone alcohol	10	15.1	16.1
Diethyl ether	5	7.4	5.2
Ethanol	20	10.5	11.8

Solvent	wt% experimental	wt% UNIFAC	wt% mod-UNIFAC
Isopropanol	10	7.8	8.1
Methanol	33	46.7	36.8
Methyl ethyl ketone	12	24.1	25.9
n-Octanol	3	2.9	2.3
Propylene glycol	9	3.6	4.0
Tetrachloroethylene	3	2.0	8.6

Quite frequently when trying to predict the activity coefficients of systems with large molecules the problem is not so much that the parameters are erroneous but rather that they simply do not exist. The reason why many more “exotic” interactions are not present is that there is simply no VLE data with the appropriate groups. [Possible “fixes” for this problem are discussed in section 7]. Fig. 3 shows the state of the UNIFAC matrix as of September 2008, the common binary interaction (left side of triangle) are fairly well known while the others are fairly sparse.

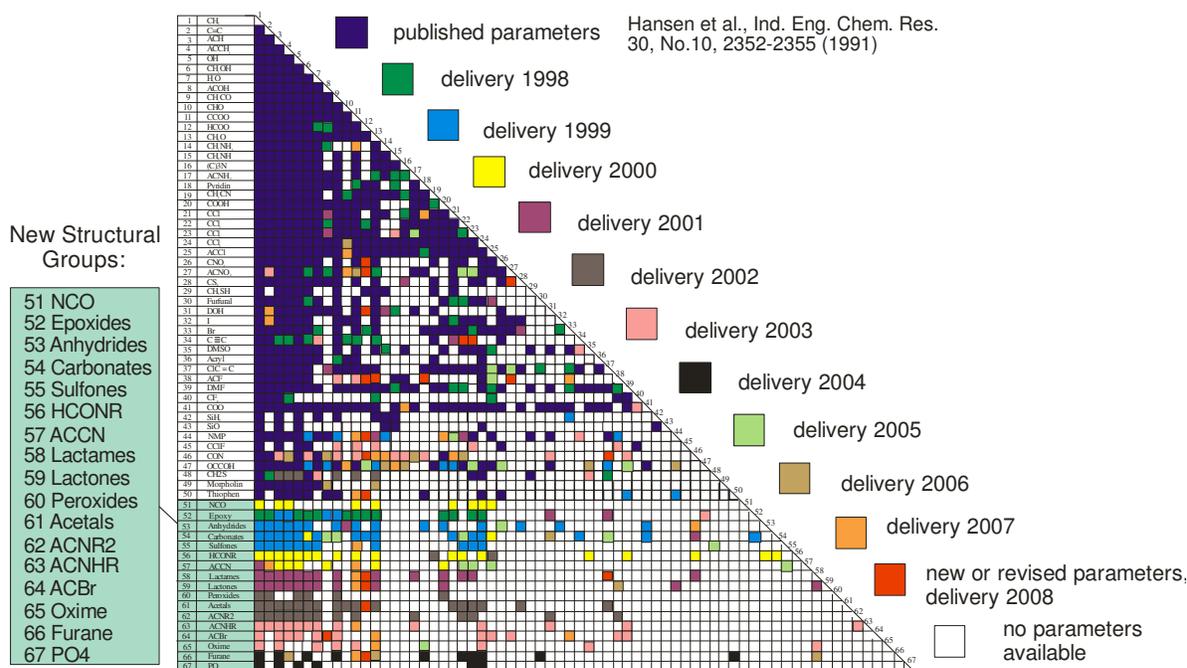


Fig. 3 The state of the UNIFAC matrix as of September 2008, data from DDBST [33]

4.2.3. Segment activity coefficients

4.2.3.1. Hansen solubility parameters

Thermodynamically, two substances will be mutually soluble when the change in Gibbs energy upon mixing is negative. The change in Gibbs energy upon mixing ($\Delta_m G$) is defined in terms of the enthalpy ($\Delta_m H$) and entropy ($\Delta_m S$) change upon mixing as follows:

$$\Delta_m G = \Delta_m H - T\Delta_m S \quad (4-65)$$

Since in the mixture there will usually be a greater degree of randomness in the molecule distribution $\Delta_m S$ will usually be positive. Therefore if we consider Eqn.(4-65), substances will be mutually soluble if $\Delta_m H < T\Delta_m S$. Hildebrand [15] proposed the Scatchard-Hildebrand equation for the change in enthalpy upon mixing (Eqn. (4-66)). This equation was developed by using regular solution theory, where a regular solution has an ideal change of volume and entropy upon mixing but unlike an ideal solution it has a non-zero enthalpy of mixing.

$$\Delta_m H = (x_1 V_1 + x_2 V_2) \phi_1 \phi_2 (\delta_1 - \delta_2)^2 \quad (4-66)$$

$$\phi_i = \frac{x_i V_i}{\sum_j x_j V_j} \quad (4-67)$$

where V is the molar volume, x is the mole fraction, ϕ is the volume fraction and δ is the Hildebrand solubility parameter which is defined as the square root of the cohesive pressure (see section 3.2):

$$\delta = \sqrt{c} = \left[\frac{\Delta_v H - RT}{V} \right]^{0.5} \quad (4-68)$$

It is therefore clear from Eqn. (4-66) that when the Hildebrand solubility parameters are similar (i.e. minimizing $\Delta_m H$) then the substances will be mutually soluble. The Hildebrand solubility parameter is only suitable for apolar or slightly polar compounds (i.e. approaching regular solution behaviour).

One of the reasons why the Hildebrand solubility parameter doesn't work for polar and hydrogen bonding compounds is that all the molecular interactions are simply "dumped" into one parameter and no differentiation is made between them. This problem was noted by many authors and many multi-parameter solubility parameters were developed [34]. The approach which has found the widest use is the one developed by Hansen [35] where he split the cohesive pressure (and therefore the solubility parameter) into a dispersive (δ_d), polar (δ_p) and hydrogen bonding (δ_h) part, given mathematically as:

$$\delta_t^2 = \delta_d^2 + \delta_p^2 + \delta_h^2 \quad (4-69)$$

where δ_t should equal the Hildebrand solubility parameter but may differ slightly in some cases. The three parameters can be calculated in various means; these are outlined by Barton [34] and Hansen [35]. The two methods which have found the widest use are those of Hansen and Hoy. If the three components ($\delta_d, \delta_p, \delta_h$) are considered as three mutually orthogonal axis then each compound is a point in space. It is therefore the distance between these points that will determine whether compounds will be mutually soluble or not. The

distance equation is similar to the normal equation for the distance between vectors but the dispersive axis is doubled to give:

$$R_{12} = \left[4(\delta_{d,1} - \delta_{d,2})^2 + (\delta_{p,1} - \delta_{p,2})^2 + (\delta_{h,1} - \delta_{h,2})^2 \right]^{0.5} \quad (4-70)$$

Every solute (1) is experimentally assigned a sphere of solubility with radius R_1 and therefore if $R_{12} < R_1$ the compounds will be mutually soluble. Consider the example of sucrose shown in Table 6 and Fig. 4, the sphere of solubility only encompasses water and therefore we would expect water to be a good solvent and the others to be poor (which can be verified in the literature [36]).

Table 6 Hansen solubility parameters for sucrose and some solvents, data from Barton [34]

Solute (MPa ^{0.5})				
Name	δ_d	δ_p	δ_h	R_j
Sucrose	21.7	26.3	29.6	20.4
Solvent (MPa ^{0.5})				
Name	δ_d	δ_p	δ_h	R_{ij}
Water	15.6	16.0	42.3	17.5
Benzene	18.4	0.0	2.0	38.3
Acetone	15.5	10.4	7.0	28.3
Ethanol	15.8	8.8	19.4	21.1
Cyclohexanol	17.4	4.1	13.5	27.8

The Hansen model is not only useful for a qualitative estimation of solubility but it can also be used to provide a quantitative one. The change in the enthalpy upon mixing in terms of the Hansen solubility parameters is analogous to Eqn. (4-66) and is given by:

$$\Delta_m H = (x_1 V_1 + x_2 V_2) \phi_1 \phi_2 \left[(\delta_{d,1} - \delta_{d,2})^2 + b_1 (\delta_{p,1} - \delta_{p,2})^2 + b_1 (\delta_{h,1} - \delta_{h,2})^2 \right] \quad (4-71)$$

Since it is not desirable to only be able to calculate properties of regular solutions we can assume that the entropy change upon mixing is given by the Flory-Huggins term:

$$\Delta_m S = -R(x_1 \ln \phi_1 + x_2 \ln \phi_2) \quad (4-72)$$

The excess change in Gibbs energy upon mixing is therefore given by:

$$\frac{G^E}{RT} = \frac{\Delta_m H}{RT} - \frac{\Delta_m S}{R} - \frac{\Delta_m G^{id}}{RT} \quad (4-73)$$

Substituting in all the terms on the right hand side of Eqn. (4-73) and taking the partial derivative with respect to number of moles of component 1 the following results:

$$\ln \gamma_1 = \frac{V_1}{RT} \phi_2^2 \left[(\delta_{d,1} - \delta_{d,2})^2 + b_1 (\delta_{p,1} - \delta_{p,2})^2 + b_1 (\delta_{h,1} - \delta_{h,2})^2 \right] + \ln \frac{\phi_1}{x_1} + 1 - \frac{\phi_1}{x_1} \quad (4-74)$$

This is frequently written in a simplified form as follows:

$$\ln \gamma_1 = \frac{V_1}{RT} \left[(\delta_{d,1} - \bar{\delta}_d)^2 + b_1 (\delta_{p,1} - \bar{\delta}_p)^2 + b_1 (\delta_{h,1} - \bar{\delta}_h)^2 \right] + \ln \frac{\phi_1}{x_1} + 1 - \frac{\phi_1}{x_1} \quad (4-75)$$

where $\bar{\delta}_x$ is the “effective” solubility parameter for either the dispersive, polar or hydrogen bonding part, for the binary case $\bar{\delta}_x = \delta_{x,1}\phi_1 + \delta_{x,2}\phi_2$.

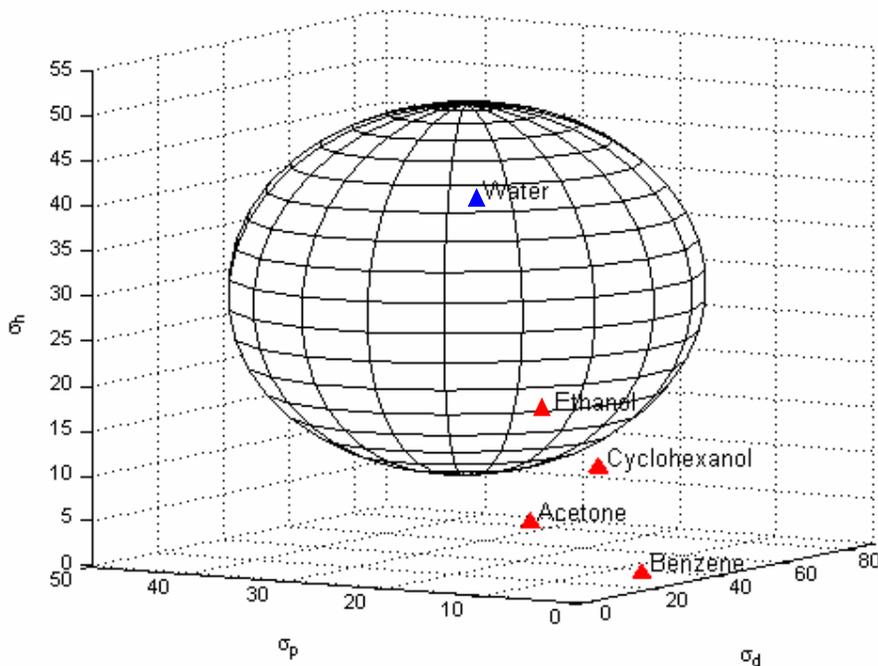


Fig. 4 Hansen solubility sphere for sucrose (▲ – position of solvents relative to the solubility sphere, all solubility parameters taken from Barton [34] and all have units of $\text{MPa}^{0.5}$. Red symbolizes a poor solvent and blue a good one)

When modelling the test solubility data (for aspirin) there are two different methods of approach. Firstly if it is assumed that no solubility data is known, then the solubility parameters can be predicted from group contribution methods (e.g. van Krevelen's [34]). On the other hand if it is assumed that solubility data is known in three or four solvents then the parameters can be fitted to the data. For the latter approach the solvents available need to be sufficiently different so that they contain enough information about the three segments.

Table 7 shows both the fitted and the pure predicted results for the solubility of aspirin in mixed solvents. The predicted results are fairly poor, and in many cases are more than a factor of two off. The reason for this is probably that the group contribution method used was not designed for complex molecules such as aspirin. For the fitted approach the four solvents that were used were: cyclohexane, methyl ethyl ketone, acetone and methanol. The fitted results are in most cases very good, and in most cases reproducing the experimental data

very well. In some cases there are two values for a single compound, this is because when large deviations were found the parameters of Hoy were tried. Remarkably with no further fitting the results were dramatically improved in all three cases. This is not because Hoy's parameters are generally better but rather that for some polar compounds the Hansen parameters have a polar parameter which seems too small to be correct (1,4-dioxane is a case in point).

Table 7 Comparison of the experimental and predicted data for aspirin solubility using Hansen solubility parameters from Barton [34] (* - the four reference solvents, ^a – Hansen's parameters, ^b – Hoy's parameters, experimental data from Frank [19])

Solvent	δ_d	δ_p	δ_h	wt% ^{experimental}	†wt% ^{fitted}	‡wt% ^{predicted}
1,1,1-Trichloroethane	17.0	4.3	2.1	0.5	1.1	4.4
1,2-Dichloroethane	19.0	7.4	4.1	3	2.9	13.0
1,4-Dioxane ^a	19.0	1.8	7.4	19	1.3	19.3
1,4-Dioxane ^b	16.3	10.1	7.9	19	18.2	10.5
2-Ethylhexanol	16.0	3.3	11.9	10	4.1	8.9
Acetic acid ^a	14.5	8.0	13.5	12	31.6	14.7
Acetic acid ^b	13.9	12.2	18.9	12	23.6	2.7
Acetone	15.5	10.4	7.0	29	27.0*	11.2
Chloroform ^a	17.8	3.1	5.7	6	1.9	12.5
Chloroform ^b	11.0	13.7	6.3	6	7.1	0.1
Cyclohexane	16.8	0.0	0.2	0.005	0.2*	2.6
Diacetone alcohol	15.8	8.2	10.8	10	13.5	8.3
Diethyl ether	14.5	2.9	5.1	5	3.5	4.7
Ethanol	15.8	8.8	19.4	20	21.7	13.7
Isopropanol	15.8	6.1	16.4	10	15.3	15.4
Methanol	15.1	12.3	22.3	33	32.9*	7.9
Methyl ethyl ketone	16.0	9.0	5.1	12	13.6*	8.9
n-Octanol	17.0	3.3	11.9	3	3.4	11.9
Propylene glycol	16.8	9.4	23.3	9	2.7	2.9
Tetrachloroethylene	19.0	6.5	2.9	3	0.9	5.8

4.2.3.2. NRTL-SAC

Just as UNIFAC and ASOG use the concept of the solution of groups, NRTL-SAC [37, 38] (NRTL – Segmented Activity Coefficient) uses the principle of the “solution of surfaces”. Instead of considering molecules as been broken down into functional groups the molecules are broken down into surface segments. The four surfaces that are used are hydrophobic (X), hydrophilic (Z), polar positive (Y+) and polar negative(Y-). So for example hexane would only have a hydrophobic surface, whereas something like acetic acid would have a bit of all 4 surfaces.

† The 4 parameters that were fitted are: $\delta_{d,1}$, $\delta_{p,1}$, $\delta_{h,1}$ and b_1 from Eqn. (4-75). In the article by Frank the Hansen fit is quite a bit worse than this one, the reason is twofold, firstly b_1 was not used as a fitting parameter by Franks and secondly he used an estimated value of the heat of fusion. The fitted parameters ($\delta_{d,1}$, $\delta_{p,1}$, $\delta_{h,1}$, b_1) = (14.80,12.47,11.51,0.311).

‡ The predicted values were obtained from the group contribution method of van Krevelen given by Barton. The predicted parameters ($\delta_{d,1}$, $\delta_{p,1}$, $\delta_{h,1}$) = (19.55,4.90,11.28), b_1 was set to 0.25.

As with UNIFAC and ASOG it is assumed that the activity coefficient consists of a combinatorial and residual part. The combinatorial expression used is analogous to the Flory-Huggins expression used in ASOG.

$$\ln \gamma_i^c = \ln \frac{\phi_i}{x_i} + 1 - r_i \sum_J \frac{\phi_J}{r_J} \quad (4-76)$$

$$r_i = \sum_l r_{i,l} \quad (4-77)$$

$$\phi_i = \frac{r_i x_i}{\sum_J r_J x_J} \quad (4-78)$$

where $r_{i,l}$ is the number (analogous to group frequency but a real number rather than a whole number) of segments i in molecule l . The lower case indices (i, j, k, m and m') run from 1 to 4 where 1 to 4 are assigned segments (usually 1 = X; 2 = Y-; 3 = Y+ and 4 = Z) and the uppercase indices (I and J) are for the number of components (for examples of the types of values found for the segment numbers, see Table 9). The expression for the residual contribution to the activity coefficient is analogous to those of UNIFAC and ASOG:

$$\ln \gamma_i^R = \sum_m r_{m,l} [\ln \Gamma_m - \ln \Gamma_m^l] \quad (4-79)$$

where Γ_m is the activity coefficient of segment m in the mixture and Γ_m^l is the activity coefficient of segment m in the pure component l . As the name suggests, NRTL is used to compute the segment activity coefficients:

$$\ln \Gamma_m = \frac{\sum_j x_j G_{jm} \tau_{jm}}{\sum_k x_k G_{km}} + \sum_{m'} \frac{x_{m'} G_{mm'}}{\sum_k x_k G_{km'}} \left(\tau_{mm'} - \frac{\sum_j x_j G_{jm'} \tau_{jm'}}{\sum_k x_k G_{km'}} \right) \quad (4-80)$$

$$x_j = \frac{\sum_l x_l r_{j,l}}{\sum_l \sum_i x_l r_{i,l}} \quad (4-81)$$

The segment activity coefficient in the pure component (Γ_m^l) is calculated from the same equation except $x_{j,l}$ is used in place of x_j , where:

$$x_{j,l} = \frac{r_{j,l}}{\sum_i r_{i,l}} \quad (4-82)$$

The NRTL-SAC segment interaction parameters were obtained by considering large amounts of VLE and LLE data. The table of the interaction parameters are shown in Table 8.

The values for α_{ij} were set by considering the typical values that are found in systems containing the appropriate surfaces.

Table 8 Segment interaction parameters for NRTL-SAC taken from Chen et al. [38]

i \ j	τ_{ij}				α_{ij}			
	X	Y-	Y+	Z	X	Y-	Y+	Z
X	0	1.643	1.643	6.547	0	0.2	0.2	0.2
Y-	1.834	0	0	-2	0.2	0	0	0.3
Y+	1.834	0	0	2	0.2	0	0	0.3
Z	10.949	1.787	1.787	0	0.2	0.3	0.3	0

Table 9 shows the results for the aspirin solubility data prediction. Quite surprisingly the results are of a comparable accuracy to the Hansen fit; it is unclear whether this is just the case for this example or if the results are generally comparable. The same four reference solvents were used for the Hansen equation for the sake of continuity. As with the Hansen approach the data for acetic acid is quite a large outlier which indicates that the data point is probably wrong.

The rationale behind using equations such as this and Hansen's to fit to some reference solvents to predict the general behaviour is that during the manufacture of drugs these measurements are undertaken anyway. Nevertheless this does mean that in the absence of experimental data no pure predictions can be undertaken (in the case of NRTL-SAC since there are predictive methods available for Hansen parameters as shown above).

Table 9 Comparison of the experimental and predicted data for aspirin solubility using NRTL-SAC [38] (* - the four reference solvents, experimental data from Frank [19])

Solvent	r_x	r_{Y-}	r_{Y+}	r_z	wt% ^{experimental}	\S wt% ^{fit}	**wt% ^{fit}
1,1,1-Trichloroethane	0.548	-	0.287	-	0.5	1.6	1.1
1,2-Dichloroethane	0.394	-	0.691	-	3	6.8	5.2
1,4-Dioxane	0.154	0.086	0.401	-	19	20.9	21.6
2-Ethylhexanol	-	-	-	-	10	-	-
Acetic acid	0.045	0.164	0.157	0.217	12	34.8	37.2
Acetone	0.131	0.109	0.513	-	29	28.5*	28.8
Chloroform	0.278	-	0.039	-	6	4.4	4.7
Cyclohexane	0.892	-	-	-	0.005	0.2*	0.09
Diacetone alcohol	-	-	-	-	10	-	-
Diethyl ether	0.448	0.041	0.165	-	5	5.2	4.1
Ethanol	0.256	0.081	-	0.507	20	18.0	18.6
Isopropanol	0.351	0.070	0.003	0.353	10	11.2	11.1
Methanol	0.088	0.149	0.027	0.562	33	35.1*	37.1
Methyl ethyl ketone	0.247	0.036	0.480	-	12	15.3*	14.3
n-Octanol	0.766	0.032	0.624	0.335	3	8.5	7.4
Propylene glycol	-	-	-	-	9	-	-
Tetrachloroethylene	-	-	-	-	3	-	-

[§] The 4 surface parameters were fitted using the same 4 solvents as with the Hansen fit. The results obtained are a bit worse than those reported by Chen but this is because they fitted a solubility parameter (Eqn. (5-12)) since they did not have experimental data heat of fusion data. The surface segments (X,Y-,Y+,Z) = (0,0.132,0.800,0.347).

** The fit using all the solvents except for acetic acid (because the data seems erroneous), surface segments (X,Y-,Y+,Z) = (0,0.167,0.869,0.428).

4.2.3.3. COSMO-RS and COSMO-SAC

The huge potential with quantum chemical methods such as COSMO-RS [39] (CONductor like Screening MOdels for Real Solvents) and COSMO-SAC [40, 41] (COSMO – Segmented Activity Coefficient) is that they could supply *ab initio* predictions of activity coefficients (and other properties) eliminating any need to undertake measurements. Both methods are based on the so called sigma profiles that are obtained from quantum chemical calculations. These profiles are generated by breaking the surface of a molecule into segments and calculating the distribution of the charge density across the molecular surface. The segment charge density is denoted by σ , and the frequency of each segment charge density (or simply just segment) is given by $p(\sigma)$. An example of some sigma profiles is shown in Fig. 5, quite a bit of qualitative data can be gleaned by examining the sigma profiles. For example hexane is clearly non-polar as all its segments lie in the low charge density range, while acetone is clearly a polar molecule since it has a much wider profile.

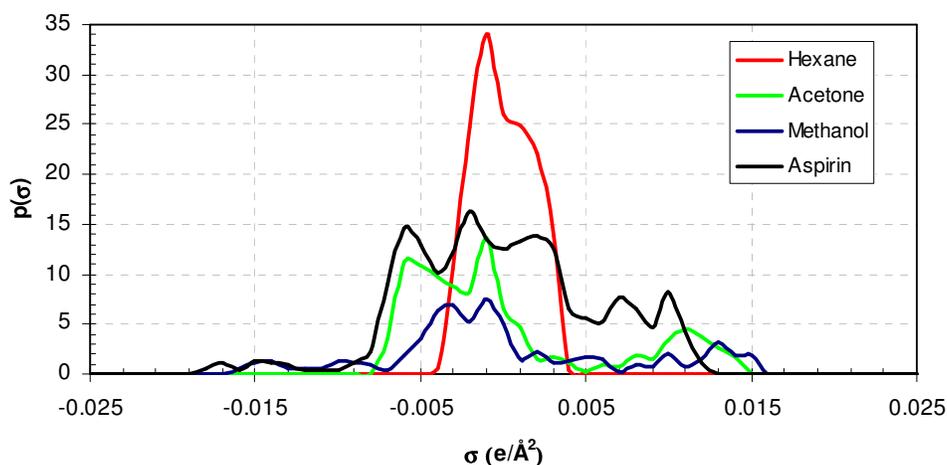


Fig. 5 Sigma profiles for 3 solvents and an example solute (aspirin) (profiles from the VT database [42])

Where COSMO-RS and COSMO-SAC differ is that COSMO-SAC sums up the activity coefficient of the surface segment in the mixture (NRTL-SAC is based on the same principle) while COSMO-RS calculates the chemical potential of each segment and then calculates activity coefficient. Lin and Sandler [41] report certain anomalies in the method of COSMO-RS which are not present in the method of COSMO-SAC. Both methods use the Staverman-Guggenheim (denoted by superscript SG) combinatorial term (see section 4.2.4) in the calculation of the activity coefficient. The COSMO-SAC expression is:

$$\ln \gamma_{i|S} = n_i \sum_{\sigma_m} p_i(\sigma_m) [\ln \Gamma_S(\sigma_m) - \ln \Gamma_i(\sigma_m)] + \ln \gamma_{i|S}^{SG} \quad (4-83)$$

where $\ln\Gamma_s(\sigma_m)$ is the segment activity coefficient of segment σ_m in the mixture and $\ln\Gamma_i(\sigma_m)$ is the activity coefficient of segment σ_m in pure component i (the exact details of the calculations are not in the scope of this work). The expression for the COSMO-RS activity coefficient is:

$$\ln \gamma_{i|S} = n_i \sum_{\sigma_m} p_i(\sigma_m) \exp \left[\frac{\mu'_s(\sigma_m) - \mu'_i(\sigma_m)}{kT} \right] - \lambda \ln \left(\frac{A_s}{A_i} \right) + \ln \gamma_{i|S}^{SG} \quad (4-84)$$

where $\mu'_s(\sigma_m)$ is the chemical potential of segment σ_m in the mixture, $\mu'_i(\sigma_m)$ is the chemical potential of segment σ_m in pure component i , A_s is the mole fraction weighted surface area of all the species in the mixture, similarly A_i is for pure component i and λ is a solvent specific adjustable parameter.

As mentioned above the huge advantage of using COSMO calculations for calculating phase equilibria is that results are obtained from first principles. Table 10 shows the results for the prediction of the aspirin solubility test set. The results for both methods are fairly poor, this is to be expected as out of all the methods analysed in this section these have the “seen” the least amount of experimental data. In almost every instance the COSMO-RS (Oldenburg) is superior to that of COSMO-SAC.

As mentioned above the huge advantage of the COSMO methods is that they can provide results (almost) solely from first principles, with very little need for experimental data. However the results for the test set suggest that these methods should be used with caution and not simply blindly applied. The test results show that, unsurprisingly, the “reference solvent” methods are superior, however UNIFAC is still competitive in many instances and in the absence of experimental data should be the method of choice (out of the methods analysed in this section).

Table 10 Comparison of the experimental and predicted data for aspirin solubility using COSMO-SAC and COSMO-RS (experimental data from Frank [19], COSMO profiles from the DDB [33])

Solvent	wt% ^{experimental}	wt% ^{COSMO-SAC}	wt% ^{COSMO-RS}
1,1,1-Trichloroethane	0.5	1.2	2.0
1,2-Dichloroethane	3	15.3	11.2
1,4-Dioxane	19	68.0	27.9
2-Ethylhexanol	10	25.1	10.6
Acetic acid	12	30.2	25.6
Acetone	29	71.7	48.0
Chloroform	6	37.7	17.5
Cyclohexane	0.005	0.7	0.2
Diacetone alcohol	10	64.6	37.6
Diethyl ether	5	57.4	12.9
Ethanol	20	67.3	42.7
Isopropanol	10	64.4	35.5
Methanol	33	75.2	52.4
Methyl ethyl ketone	12	66.6	39.8
n-Octanol	3	25.1	10.6
Propylene glycol	9	45.4	29.0
Tetrachloroethylene	3	0.1	0.3

4.2.4. Combinatorial contributions

As mentioned above the combinatorial contribution refers to the fact the change in the thermodynamic property is from size/shape interactions. Thermodynamically this is when the mixture is athermal (i.e. $H^E = \Delta_m H = 0$), therefore for an athermal mixture (from Eqn. (4-65)):

$$\left(\frac{\Delta_m G}{RT}\right)_c = -\frac{\Delta_m S}{R} \quad (4-85)$$

where the subscript c refers to the combinatorial contribution. One of the simplest expressions for the entropy change upon mixing is given the expression proposed by Flory and Huggins (see Eqn. (4-72)). Rearrangement of Eqn. (4-73) yields Eqn. (4-86) which can be combined with the Flory-Huggins entropy expression to give Eqn. (4-87), where the van der Waals volume fraction (Eqn. (4-49)) is used in place of the molar volume fraction.

$$\frac{G^E}{RT} = \frac{\Delta_m G}{RT} - \sum_i x_i \ln x_i \quad (4-86)$$

$$\frac{G^E}{RT} = \sum_i x_i \ln \frac{\Phi_i}{x_i} \quad (4-87)$$

Taking the partial derivative of Eqn. (4-87) yields the following expression for the Flory-Huggins combinatorial contribution to the activity coefficient:

$$\ln \gamma_i^C = 1 - \frac{\Phi_i}{x_i} + \ln \frac{\Phi_i}{x_i} \quad (4-88)$$

For athermal mixtures of chain molecules Guggenheim [43] proposed the following expression for the entropy change upon mixing:

$$\frac{-\Delta_m S}{R} = \sum_i x_i \ln \Phi_i + \frac{1}{2} \sum_i \left(x_i z q_i \ln \frac{\theta_i}{\Phi_i} \right) \quad (4-89)$$

where z is the coordination number of the lattice (number of atoms which touch) also θ_i and Φ_i are defined by Eqn's (4-50) and (4-49) respectively. Combining Eqn's (4-85), (4-86), (4-89) and (4-25) yields the Guggenheim-Staverman combinatorial expression:

$$\ln \gamma_i^C = \left(1 - \frac{\Phi_i}{x_i} + \ln \frac{\Phi_i}{x_i} \right) - \frac{z q_i}{2} \left(1 - \frac{\Phi_i}{\theta_i} + \ln \frac{\Phi_i}{\theta_i} \right) \quad (4-90)$$

In many applications (such as UNIFAC and UNIQUAC) the value of z is set to 10. The Guggenheim-Staverman equation reduces to the Flory-Huggins equation when $r_i = q_i$ for all i. These two expressions were used among others by Kikic et al. [44] to predict the infinite dilution activity coefficients for systems where the residual will fall away (alkane-alkane

systems – see section 6.1 for the analysis proposed by this work). Their analysis showed that the Flory-Huggins expression actually performs better than the Guggenheim-Staverman (SG) equation in all the examples tried. They proposed the following empirical modification for the SG equation:

$$\ln \gamma_i^C = \ln \frac{\Phi'_i}{x_i} + 1 - \frac{\Phi'_i}{x_i} - \frac{1}{2} z q_i \left(\ln \frac{\Phi_i}{\theta_i} + 1 - \frac{\Phi_i}{\theta_i} \right) \quad (4-91)$$

where Φ'_i is the empirically modified volume fraction, for which two different expressions were proposed:

$$\Phi'_i = \frac{x_i r_i^{2/3}}{\sum_j x_j r_j^{2/3}} \quad (4-92)$$

$$\Phi'_i = \frac{x_i V_i^{2/3}}{\sum_j x_j V_j^{2/3}} \quad (4-93)$$

where r and V are the van der Waals and molar volumes respectively. Both the expressions provided much improved predictions of the infinite dilution activity coefficients. Then Gmehling et al. [40] proposed another empirical modification of Φ'_i which provided better representation of mixtures of large and small molecules. The modification is given by Eqn. (4-94) and is used in the combinatorial expression of the modified UNIFAC (Dortmund) equation (from this point onwards this will be referred to as the mod-UNIFAC combinatorial).

$$\Phi'_i = \frac{x_i V_i^{3/4}}{\sum_j x_j V_j^{3/4}} \quad (4-94)$$

When these expressions are applied to systems which have massive differences in component size such as polymer solutions they tend to perform quite poorly. It was theorised that this deviation was due to the “free volume” in the solution, which is the volume which is not occupied by molecules and is only really noticeable in polymer solutions. One of the first of these free-volume equations, called the ELBRO-FV expression, was proposed [45] :

$$\ln \gamma_i^{C-FV} = 1 - \frac{\Phi_i^{FV}}{x_i} + \ln \frac{\Phi_i^{FV}}{x_i} \quad (4-95)$$

where the superscript C-FV refers to the combinatorial free-volume part and the free volume fraction is defined as:

$$\Phi_i^{FV} = \frac{x_i V_i^{FV}}{\sum_j x_j V_j^{FV}} \quad (4-96)$$

$$V_i^{FV} = V_i - V_{w,i} \quad (4-97)$$

The mathematical definition of free volume is not distinct and varies from author to author [45]. The ELBRO-FV expression was modified by Kontogeorgis et al. [45] (named the GK-FV expression) by introducing the free volume into the Guggenheim-Staverman equation instead of the Flory-Huggins:

$$\ln \gamma_i^{C-FV} = \ln \frac{\Phi_i^{FV}}{x_i} + 1 - \frac{\Phi_i^{FV}}{x_i} - \frac{1}{2} z q_i \left(\ln \frac{\Phi_i}{\theta_i} + 1 - \frac{\Phi_i}{\theta_i} \right) \quad (4-98)$$

These expressions provide much improved results with polymer systems indicating that there is some realism behind the idea of free volume. These are by no means the only expressions available, many others for example UNIFAC-FV can be found in literature [45-48].

5. EXPERIMENTAL DETERMINATION OF SOLUTE ACTIVITY

The chemical potential or fugacity of solute can usually not be measured directly. The experimental techniques for extracting solute fugacity are, among others: VLE (Vapour-Liquid Equilibrium) measurements, LLE (Liquid-Liquid Equilibrium) measurements and SLE (Solid-Liquid Equilibrium) measurements. Since this work deals with the prediction of complex multifunctional compounds the sources of data are somewhat limited. VLE data for such systems is more or less nonexistent, similarly with LLE there will be very few if any data for the systems of interest in this work. Therefore the two main sources of data will be:

1. SLE data
2. Partition coefficient data

The following sections will outline the applications and limitations of these two sources of data.

5.1. Solid-liquid equilibria

The phase equilibrium criterion given in Eqn. (4-19) for equilibrium between a solid and a liquid phase becomes (superscripts S and L represent solid and liquid phase):

$$\hat{f}_i^S = \hat{f}_i^L \quad (5-1)$$

$$x_i \gamma_i^L f_i^L = z_i \gamma_i^S f_i^S \quad (5-2)$$

where x_i and z_i are the mole fractions in the liquid and solid phases respectively. If it is assumed that there is negligible solubility of the liquid in the solid (i.e. pure solid, $z_i \rightarrow 1$, $\gamma_i^S \rightarrow 1$) then:

$$x_i = \frac{f_i^S}{\gamma_i^L f_i^L} \quad (5-3)$$

As mentioned above (see section 4.1) the only requirement of the standard state is that it be at the same temperature as the solution, therefore the sub-cooled liquid is chosen as the standard state. Therefore the solubility (x_i) is not only dependant on the way the solute and the solvent interact (γ_i) but also on the properties of the pure solute (since f_i^S and f_i^L are both properties of the pure solute). The ratio f_i^S / f_i^L can be related to the change in the Gibbs energy when going from the solid (point a) to the sub cooled liquid (point d).

$$\Delta G_{a \rightarrow d} = RT \ln \frac{f_i^L}{f_i^S} \quad (5-4)$$

This process can be represented by the thermodynamic cycle given by Prausnitz [16] and shown in Fig. 6.

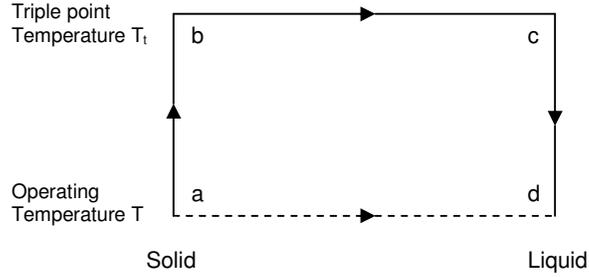


Fig. 6 Thermodynamic cycle for calculating the fugacity of a pure sub-cooled liquid, taken from Prausnitz [16].

The Gibbs free energy is related to the enthalpy and the entropy through Eqn. (5-5). Since the enthalpy and the entropy are both state functions, and are therefore not dependant on the path taken, the alternate path $a \rightarrow b \rightarrow c \rightarrow d$ can be used.

$$\Delta G_{a \rightarrow d} = \Delta H_{a \rightarrow d} - T \Delta S_{a \rightarrow d} \quad (5-5)$$

The path $a \rightarrow b \rightarrow c \rightarrow d$ can be calculated by using the appropriate relations for changing the temperature of a solid ($a \rightarrow b$), melting the solid ($b \rightarrow c$) and changing the temperature of the liquid ($c \rightarrow d$). This is represented by the following 4 equations:

$$\Delta H_{a \rightarrow d} = \Delta H_{a \rightarrow b} + \Delta H_{b \rightarrow c} + \Delta H_{c \rightarrow d} \quad (5-6)$$

$$\Delta H_{a \rightarrow d} = \Delta_{fus} H \Big|_{T_t} + \int_{T_t}^T \Delta C_p dT \quad (5-7)$$

$$\Delta S_{a \rightarrow d} = \Delta S_{a \rightarrow b} + \Delta S_{b \rightarrow c} + \Delta S_{c \rightarrow d} \quad (5-8)$$

$$\Delta S_{a \rightarrow d} = \frac{\Delta_{fus} H}{T_t} \Big|_{T_t} + \int_{T_t}^T \frac{\Delta C_p}{T} dT \quad (5-9)$$

where $\Delta C_p \equiv C_p^l - C_p^s$ and T_t is the triple point temperature. Since data for the triple point is not abundant we can reasonably assume that the triple point temperature is well approximated by the melting temperature (T_m). Assuming that ΔC_p is constant over the temperature range and combining Eqn's (5-4), (5-5), (5-7) and (5-9):

$$\ln \frac{f_i^L}{f_i^S} = \frac{\Delta_{fus} H}{RT_m} \left(\frac{T_m}{T} - 1 \right) - \frac{\Delta C_p}{R} \left(\frac{T_m}{T} - 1 \right) + \frac{\Delta C_p}{R} \ln \frac{T_m}{T} = \ln x_i \gamma_i \quad (5-10)$$

For temperatures not too remote from the melting point the heat capacity term may be neglected. This results in the following simplified relation:

$$\ln x_i = -\ln \gamma_i - \frac{\Delta_{fus} H}{RT_m} \left(\frac{T_m}{T} - 1 \right) \quad (5-11)$$

Fig. 7 shows how the ideal solubility ($\gamma_i = 1$) changes with the melting temperature and the heat of fusion of the solute in a solvent at 273.15 K. This shows that the solubility is quite a strong function of the pure component properties of the solute.

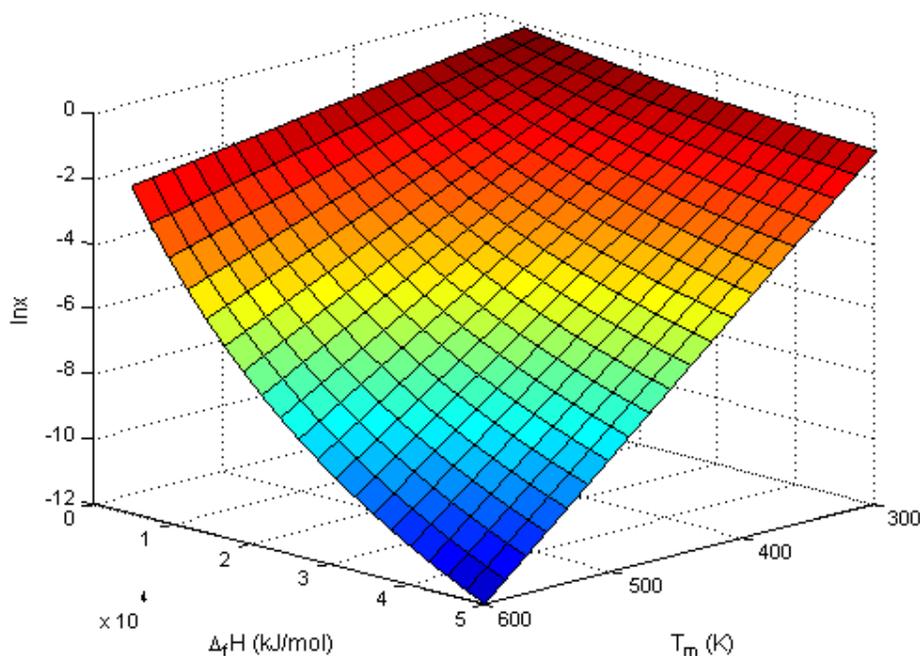


Fig. 7 The ideal solubility as a function of the melting temperature and the heat of fusion at 273.15 K.

Since heat capacity data is often unavailable for solutes Eqn. (5-11) is often used even when the temperature is fairly far from the triple point. When heat of fusion and melting temperature data are not available then the solubility product can be used instead (where the solubility parameter is an adjustable parameter used in the regression):

$$\ln x_i = \ln K_{sp} - \ln \gamma_i \quad (5-12)$$

For temperatures far removed from the melting point Eqn. (5-11) may not be sufficient as the contribution of the heat capacity change may be significant. This may require some approximations because it is unlikely that there are models (certainly not data) for some of the complex multifunctional compounds being dealt with in this work.

5.2. Partition coefficients

Partition coefficients are widely used in the pharmaceutical and biological sciences. They are used as a measure of how compounds partition (or distribute) between 2 phases. The most common partition coefficient is the octanol-water partition coefficient. This is a measure of the partitioning between water-saturated octanol and octanol-saturated water:

$$K_{ow} = \frac{C_i^{ow}}{C_i^{wo}} \quad (5-13)$$

where C is the concentration and the superscripts refer to the water saturated octanol (ow) and the octanol saturated water (wo) phases. Using the molar volumes of octanol ($V_{oct} = 6.313 \text{ mol.l}^{-1}$) and water ($V_{water} = 55.343 \text{ mol.l}^{-1}$) and the equilibrium relationships (essentially pure water phase; 27.5 mol% water in the octanol phase) the following relationship results (the superscript *wo* is replaced by *w* since the water is essentially pure):

$$K_{ow} = 0.1508 \frac{X_i^{ow}}{X_i^w} \quad (5-14)$$

If it is assumed that the solute concentration is small then, using the criterion for phase equilibrium (isofugacity criterion Eqn. (4-19)) the following results:

$$K_{ow} = 0.1508 \frac{\gamma_i^{\infty,w}}{\gamma_i^{\infty,ow}} \quad (5-15)$$

This relation is only useful if the infinite dilution activity coefficient is known in water saturated octanol as well as pure water. However Tse and Sandler [49] have shown that the octanol water partition coefficient can be related, fairly simply, to the pure phase infinite dilution activity coefficients:

$$\log K_{ow} = 0.1 + 0.91 \cdot \log \left(0.1508 \frac{\gamma_i^{\infty,w}}{\gamma_i^{\infty,o}} \right) \quad (5-16)$$

Where this correlation should only be applied to hydrophobic species ($1.0 < \log K_{ow} < 5.0$).

The great advantage of a correlation such as this is that if K_{ow} and $\gamma_i^{\infty,w}$ are known then the value for the octanol infinite dilution activity can be extracted.

6. REDUCTION OF INFINITE DILUTION DATA TO A COMMON SOLVENT

Since this work is concerned with the prediction of the activity coefficient of complex compounds in common solvents it is imperative that there are sufficient data for the solvents that are chosen. Of all the solvents water is going to have the largest amount of solubility data available since it is the most readily available of all solvents. However when trying to compile a set of data it soon became clear that large amounts of data in other single solvents was not that easy to come by. It is for this reason that one of our objectives was to be able to grow our dataset in a simple reliable way; the following sections outline our proposed method for alkanes, n-alcohols and ketones.

6.1. Alkanes

The success of methods like UNIFAC has shown that the principle of solutions of groups does have practical application. It can therefore be assumed that a solute molecule at infinite dilution will only “see” the solvent groups around it. Since alkanes are all made up of sp^3 carbons (CH_3 UNIFAC group) it would make sense that the ratio of a solutes infinite dilution activity coefficient in two different alkane solvents should only in some way depend on the size and shape of the solvent molecules. (This reasoning applies to infinite dilution data since the solute molecule will only “see” solvent groups and not other solute groups). If it is assumed that the activity coefficient is only made up a combinatorial (size/shape interactions) and a residual (energetic interactions) contribution then, since the residual contributions are equivalent, the following expression would result:

$$\frac{\gamma_{i,sol1}^{\infty}}{\gamma_{i,sol2}^{\infty}} = \frac{\gamma_{i,sol1}^{C,\infty}}{\gamma_{i,sol2}^{C,\infty}} \quad (6-1)$$

where the superscript *C* denotes the combinatorial expression and the subscripts *sol1* and *sol2* differentiate between any two alkane solvents. This expression can therefore be rearranged to give the infinite dilution activity coefficient in any alkane solvent relative to a known solvent:

$$\gamma_{i,sol1}^{\infty} = \gamma_{i,sol2}^{\infty} \frac{\gamma_{i,sol1}^{C,\infty}}{\gamma_{i,sol2}^{C,\infty}} \quad (6-2)$$

Different combinatorial expressions have been discussed in section 4.2.4. As was expected the Guggenheim-Staverman expression showed quite a large deviation when going from small to large solvents (see section 4.2.4). The modified UNIFAC expression performed a bit better but tended to under-predict the activity coefficient when going from a large reference to a small one (and *visa versa*). The free volume expression provides a very good prediction of

the experimental data and can go from very big to very small solvents with a fair level of accuracy. This success is somewhat surprising since free volume is almost never considered in non-polymer applications. The success of this term in this method could indicate that free volume does have an impact of the combinatorial contribution. Some examples of the usage of Eqn. (6-2) are shown in Fig. 8 to Fig. 10 and Table 11. In all the examples the biggest solvent was chosen as the reference solvent (for no reason other than consistency) and where multiple data points were available GLC (Gas-Liquid Chromatography) measurements were considered as superior.

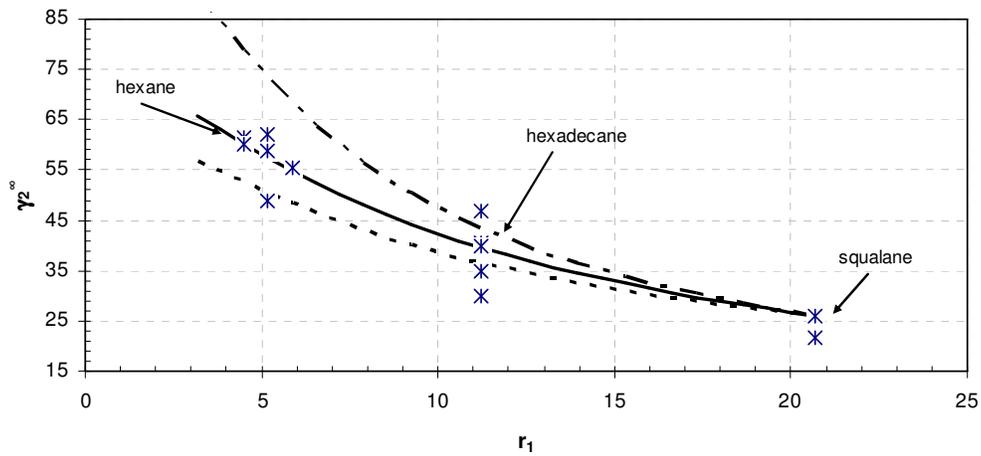


Fig. 8 γ_2^∞ vs. r_1 for ethanol(2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB, - - - SG combinatorial, mod-UNIFAC combinatorial, — GK-FV combinatorial).

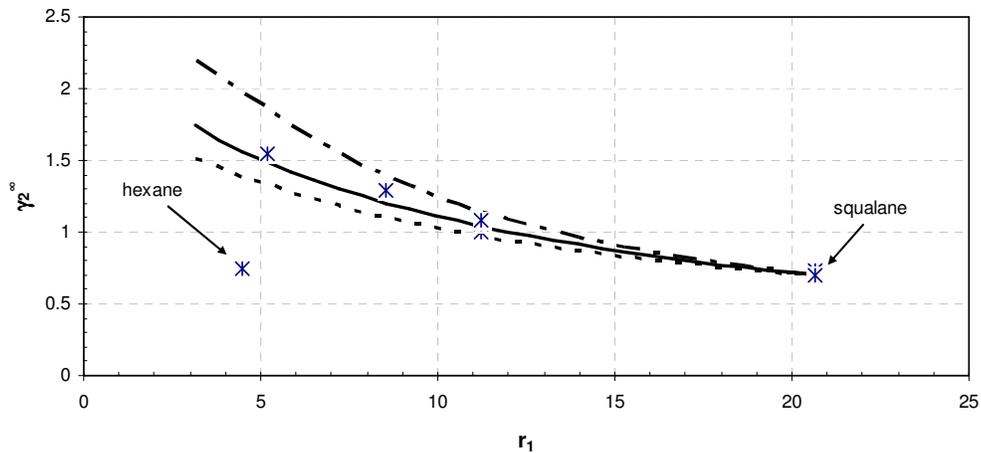


Fig. 9 γ_2^∞ vs. r_1 for benzene(2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB, - - - SG combinatorial, mod-UNIFAC combinatorial, — GK-FV combinatorial).

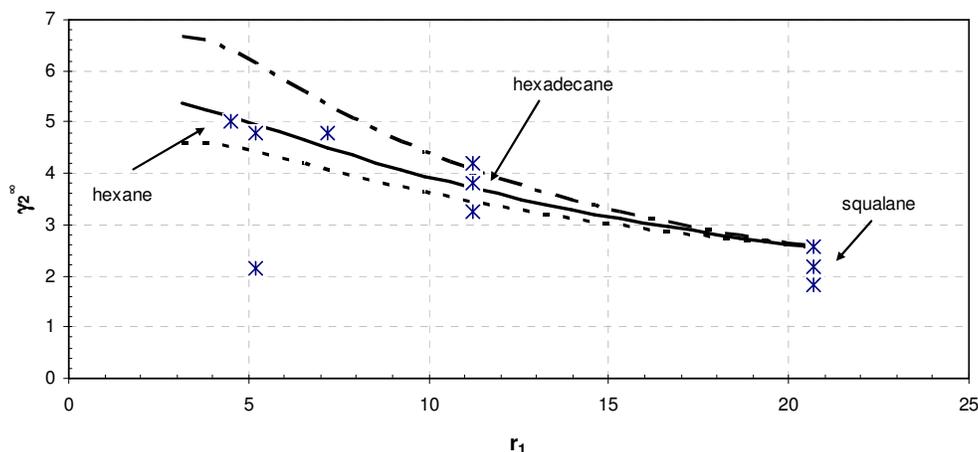


Fig. 10 γ_2^∞ vs. r_1 for butanone(2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB, --- SG combinatorial, mod-UNIFAC combinatorial, — GK-FV combinatorial).

Table 11 Comparison of the prediction for a selected group of compounds when using Eqn. (6-2) with the various combinatorial expressions

Solute/Solvent	T (K)	γ_2^{exp}	Prediction		
			SG	mod-UNIFAC	GK-FV
Naphthalene					
Heptane	403.15	2.4	2.064	1.233	2.268
19,24-Dioctadecyldotetracontane (ref.)	403.15	0.4	0.400	0.400	0.400
1-Heptene					
Hexadecane*	373.15	0.825	1.161	0.768	1.143
19,24-Dioctadecyldotetracontane (ref.)	373.15	0.354	0.354	0.354	0.354
Chlorobenzene					
Octadecane	303.15	1.15	1.243	1.089	1.155
Squalane (ref.)	303.15	0.84	0.840	0.840	0.840
1,3-Cyclohexadiene					
Hexadecane*	303.15	0.737	0.932	0.607	0.978
19,24-Dioctadecyldotetracontane (ref.)	303.15	0.251	0.251	0.251	0.251

* - data are from the same source

Initially when testing the method, quite accidentally, only solutes which were smaller than all of the solvents used were tested (this is due to the high availability of such data). However when the size of the solute became larger than the size of the solvent, the predictions started to show very large negative deviations. This is quite well illustrated in the example of ethylcyclohexane shown in Fig. 11. All three combinatorial expressions show much worse predictions as soon as the size of the solute (r_2) becomes larger than the size of the solvent (r_1). The larger the size of the solute the worse this error becomes.

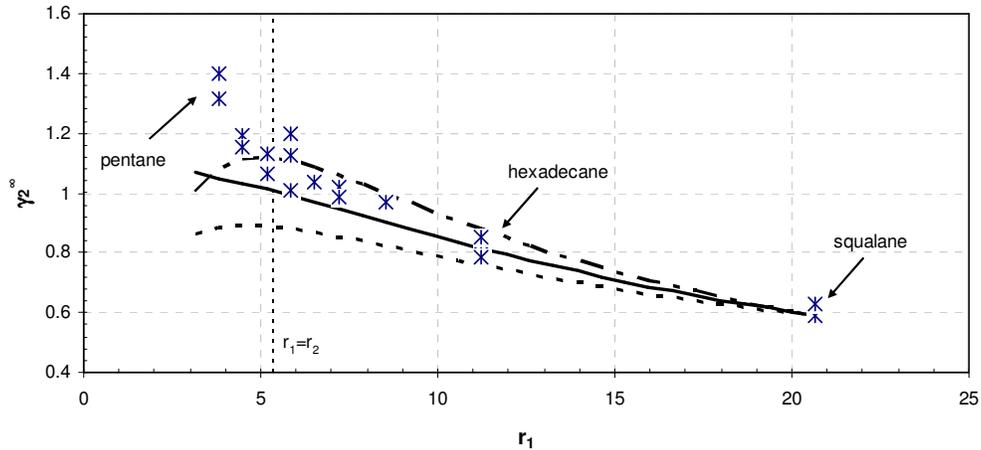


Fig. 11 γ_2^∞ vs. r_1 for ethylcyclohexane (2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB [33], - · - · - SG combinatorial, ····· mod-UNIFAC combinatorial, — GK-FV combinatorial).

This problem was solved by empirically modifying the GK-FV expression as follows:

$$\gamma_2^{FV,\infty} = \exp \left(1 - \frac{V_2^{FV}}{V_1^{FV}} + \ln \left(\frac{V_2^{FV}}{V_1^{FV}} \right) - 5q_2 \left(1 - \frac{r_2/r_1}{q_2/q_1} + \ln \left(\frac{r_2/r_1}{q_2/q_1} \right) \right) \right) \quad (6-3)$$

$$V_i^{FV} = (V_i)^{2/3} - (V_i)^{2/3} \quad (6-4)$$

This expression provides a far better prediction for the large solutes in small solvents while still providing almost identical results for small solutes in larger solvents. Another solution is to use the following approximation of the combinatorial ratio:

$$\gamma_{i,sol1}^\infty = \gamma_{i,sol2}^\infty \left(\frac{r_{sol2}}{r_{sol1}} \right)^{0.6} \quad (6-5)$$

While this expression is very simple it approximates the use of the modified GK-FV expression quite well in many instances, see Fig. 12 and Fig. 13 for good and bad approximations respectively. The great advantage with this approach is that only the molar volumes of the solvents are needed; no structural information about the solute is required.

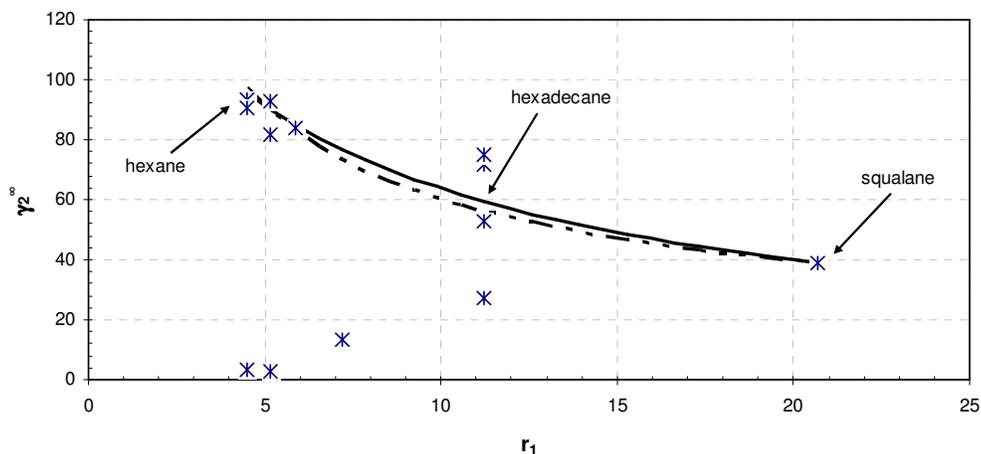


Fig. 12 γ_2^∞ vs. r_1 for methanol (2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB [33], — Eqn. (6-3), - - - Eqn. (6-5)).

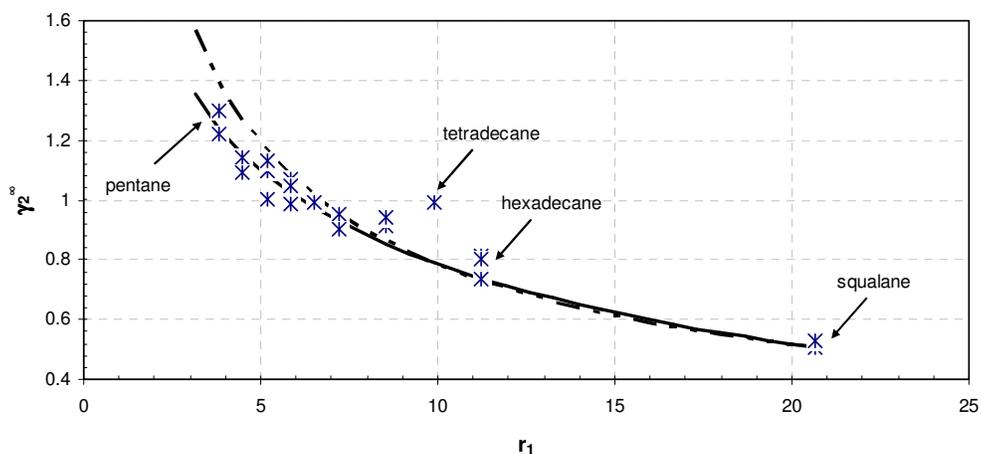


Fig. 13 γ_2^∞ vs. r_1 for cyclohexane (2) in alkane solvents(1) using squalane as the reference solvent at 298.15 K (x – data from the DDB [33], — Eqn. (6-3), - - - Eqn. (6-5)).

As mentioned above the purpose of this method was to expand the available data for the list of common solvents that have been chosen for this work. This is however only one possible application of such a method. If the data for some solute in any alkane is known then the data for all the other alkanes can be estimated. This is especially useful for example with GLC (Gas Liquid Chromatography) where the non-polar stationary phase is very often squalane or some other large alkane (e.g. 19,24-dioctadecyldotetracontane or tridecane). Another application is data validation; if one or two accurate data points are known then it would be possible to reject any outliers depending on how far away they are from the predicted points. An example of this is shown in Fig. 13 where the data for tetradecane seems to be erroneous.

This method also provides a good method to test or develop combinatorial expressions. As mentioned in section 4.2.4 often combinatorial expressions are only developed by considering

solutions of alkanes where the residual falls away. However if this approach is applied it provides a much broader scope and applicability while still remaining realistic.

For the molar volumes and the r and q values of the alkane solvents, the following empirical correlations can be used:

$$r_1 = \frac{6.88 + 10.23 \times n_c}{15.17} \quad (6-6)$$

$$q_1 = \frac{1.54 + 1.35 \times n_c}{2.5} \quad (6-7)$$

$$V = 24.026 \times r_1 + 23.102 \quad (6-8)$$

where n_c is the number of carbons, V is the molar volume and the subscript 1 refers to solvent.

6.2. *n*-Alcohols

For the *n*-alcohols a similar procedure as the one for alkanes can be applied. However since there are now two different groups (-OH and -CH₂-) there needs to be two reference solvents. Therefore if it is assumed that the infinite dilution activity coefficient is made up of an OH part (infinite dilution of the solute in a solution of OH groups) and a CH₂ part, we can write (where the combinatorial expression is obtained from Eqn. (6-3)):

$$\gamma_i^\infty = (\gamma_i^{\infty, CH_2})^{a_{i, CH_2}} (\gamma_i^{\infty, OH})^{a_{i, OH}} \quad (6-9)$$

where the exponents $a_{i, OH}$ and a_{i, CH_2} are the surface fractions of OH and CH₂ groups in the solvent, given as: $a_{i, OH} = q_{OH} / q_i$ and $a_{i, CH_2} = (q_i - q_{OH}) / q_i$ (the value for q_{OH} is taken from UNIFAC) and the values γ_i^{∞, CH_2} and $\gamma_i^{\infty, OH}$ are the model parameters found from the two reference solvents. This expression works best for interpolation but with good data can provide good extrapolation. Fig. 14 to Fig. 16 and Fig. 17 to Fig. 18 show results for a typical interpolation and a good extrapolation respectively.

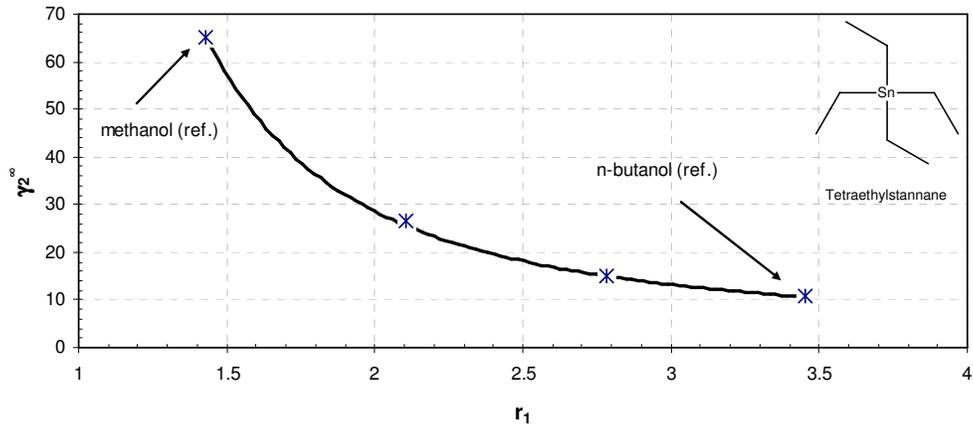


Fig. 14 γ_2^∞ vs. r_1 for tetramethylstannane (2) in n-alcohol solvents(1) at 298 K (x – data from DDB [33], — Eqn. (6-9)).

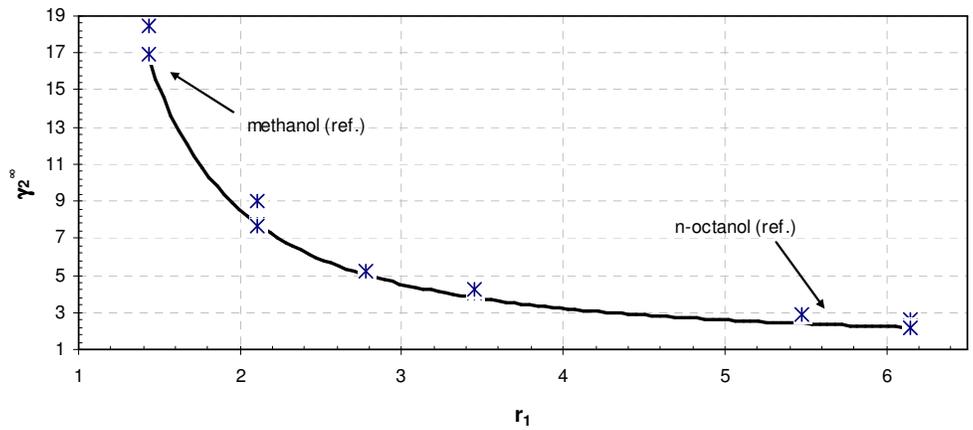


Fig. 15 γ_2^∞ vs. r_1 for pentane (2) in n-alcohol solvents(1) at 298 K (x – data from the DDB [33], — Eqn. (6-9)).

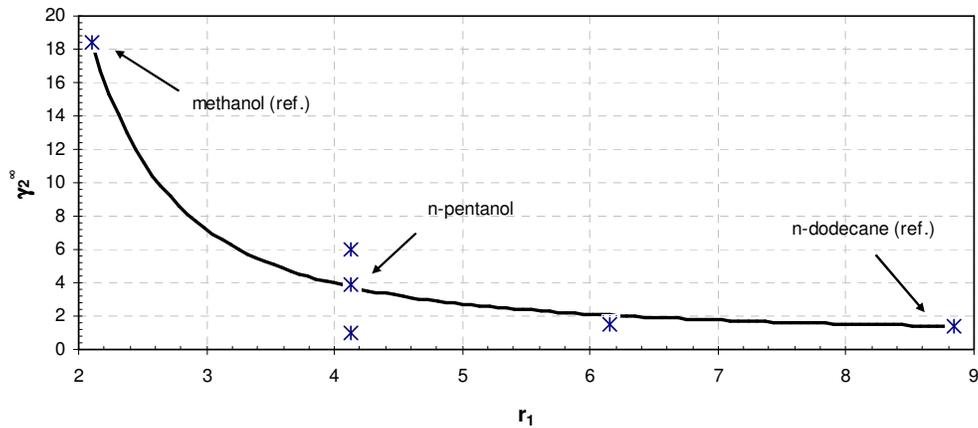


Fig. 16 γ_2^∞ vs. r_1 for dichloromethane (2) in n-alcohol solvents(1) at 298 K (x – data from the DDB [33], — Eqn. (6-9)).

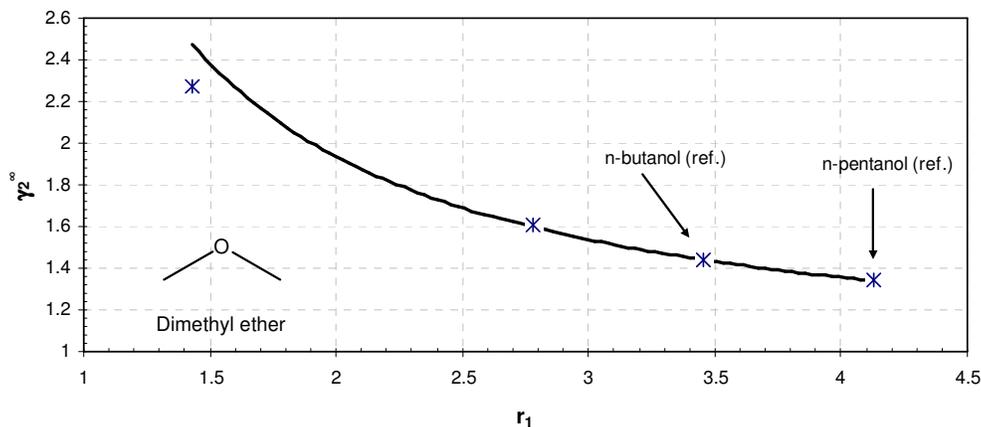


Fig. 17 γ_2^∞ vs. r_1 for dimethyl ether (2) in n-alcohol solvents(1) at 330 K (x – data from the DDB [33], — Eqn. (6-9)).

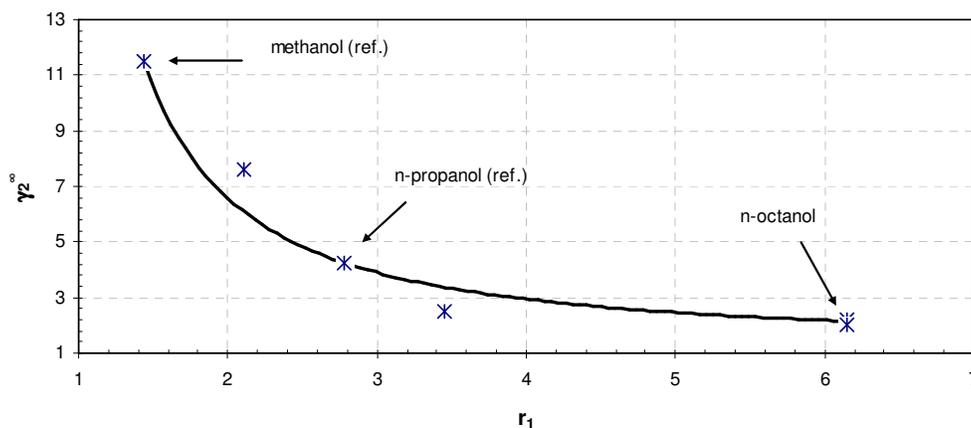


Fig. 18 γ_2^∞ vs. r_1 for toluene (2) in n-alcohol solvents(1) at 330 K (x – data from the DDB [33], — Eqn. (6-9)).

This approach is very reliable (especially for interpolation) but suffers in that it requires two n-alcohol solvents. This could potentially be overcome by only splitting the residual part of the activity coefficient into OH and CH₂ contributions as follows:

$$\gamma_i^\infty = \left(\gamma_i^{\infty,CH_2,R}\right)^{a_{i,CH_2}} \left(\gamma_i^{\infty,OH,R}\right)^{a_{i,OH}} \gamma_i^{\infty,C} \quad (6-10)$$

If Eqn. (6-10) holds then it would be possible to predict the behaviour in all n-alcohol solvents from one n-alcohol and one alkane reference solvent (for the alkane we can directly calculate $\gamma_i^{\infty,CH_2,R}$ since $a_{i,OH} = 0$). Fig. 19 shows a good example of where this works well. This n-alcohol -alkane reference method is not as reliable as the n-alcohol - n-alcohol reference method. It was shown to be suitable for all solutes where the $\gamma_i^{\infty,CH_2,R}$ parameter is less than about two (typically alkanes, aromatics and halo-alkanes) but failed in most (if not all) other

cases. It was also noted that for cases when the n-alcohol -alkane method works, the best results are obtained when the smallest n-alcohol is used as the reference solvent.

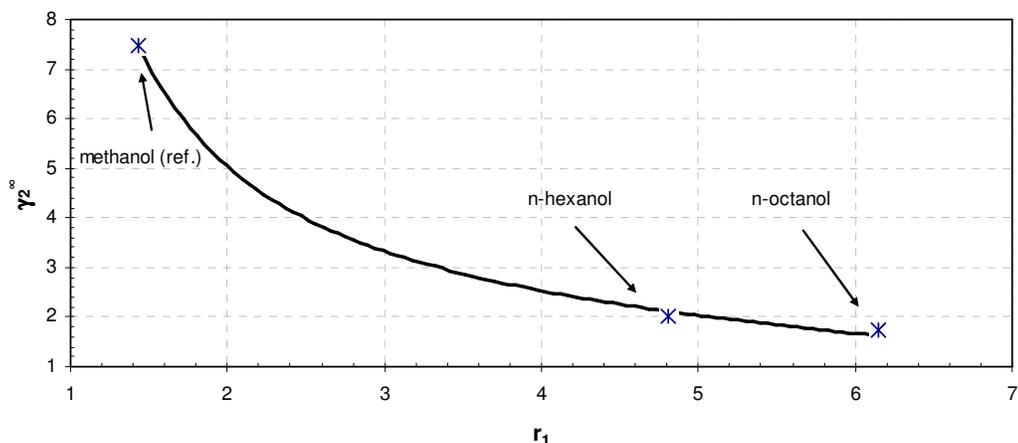


Fig. 19 γ_2^∞ vs. r_1 for tetrachloromethane (2) in n-alcohol solvents(1) at 293.15 K (x – data from the DDB [33], — Eqn.(6-10) with squalane and methanol as the 2 reference solvents).

6.3. Ketones

As with the n-alcohols it was assumed that the infinite dilution of any solute in a (non-cyclic, non-aromatic) ketone solution may be represented by the following equation:

$$\gamma_i^\infty = (\gamma_i^{\infty, CH_2})^{a_{i, CH_2}} (\gamma_i^{\infty, CO})^{a_{i, CO}} \quad (6-11)$$

where the superscript CO refers to the ketone contribution, $a_{i, CO}$ and a_{i, CH_2} are the surface fractions of CO and CH_2 groups in the solvent, given as: $a_{i, CO} = q_{CO} / q_i$ and $a_{i, CH_2} = (q_i - q_{CO}) / q_i$ (q_{CO} is taken as the value from Bondi [31]). As with the n-alcohols two reference solvents are required since there are two different contributions. Unfortunately, unlike with the n-alcohols, there are not large amounts of data available with which to make model tests, Fig. 20 shows the example of hexane for which there was a fair amount of data available. As with the n-alcohols the interpolation is fairly good.

In all the examples that have been shown so far the infinite dilution activity coefficient decreases with the increasing size of the solute. This is, however, not necessarily the case. Fig. 21 and Fig. 22 show the curves for water and ethanol in ketone solvents; in these examples the size of the infinite dilution activity coefficient increases with increasing solvent size. The reason for this is that as the hydrophobic chain of the solvent increases the solvent becomes more hydrophobic and therefore one would expect the infinite dilution activity of hydrophilic solutes in the mixture to increase. This effect is accounted for in the model by

making the hydrophobic contribution to γ_i^∞ larger (i.e. the regressed value of γ_i^{∞,CH_2} will be larger than the value of $\gamma_i^{\infty,CO}$).

When looking at Fig. 22 it seems like the prediction for 2-butanone is quite off. However looking at Fig. 23, the predicted point seems to follow the trend of the other data while the experimental point seems off. Therefore in this case we can assume that the measured value was subject to experimental error.

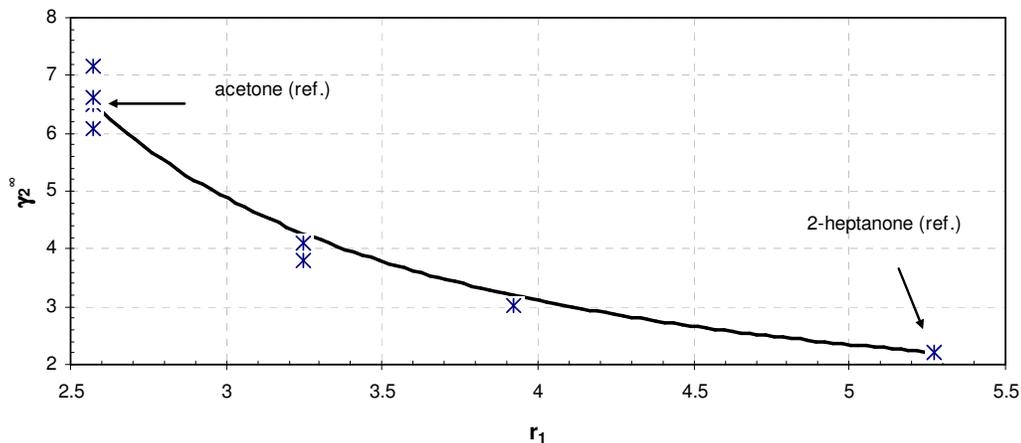


Fig. 20 γ_2^∞ vs. r_1 for hexane (2) in ketone solvents(1) at 298.15 K (x – data from the DDB [33], — Eqn. (6-11)).

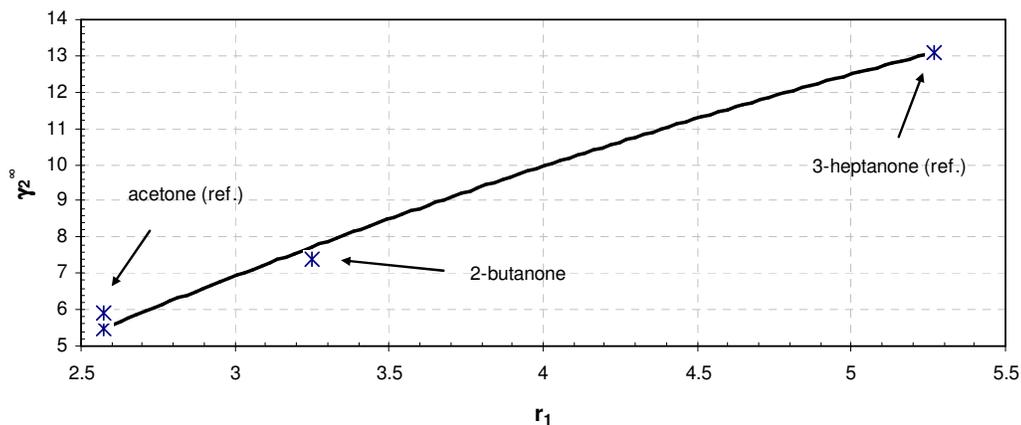


Fig. 21 γ_2^∞ vs. r_1 for water (2) in ketone solvents(1) at 333.15 K (x – data from the DDB [33], — Eqn. (6-11)).

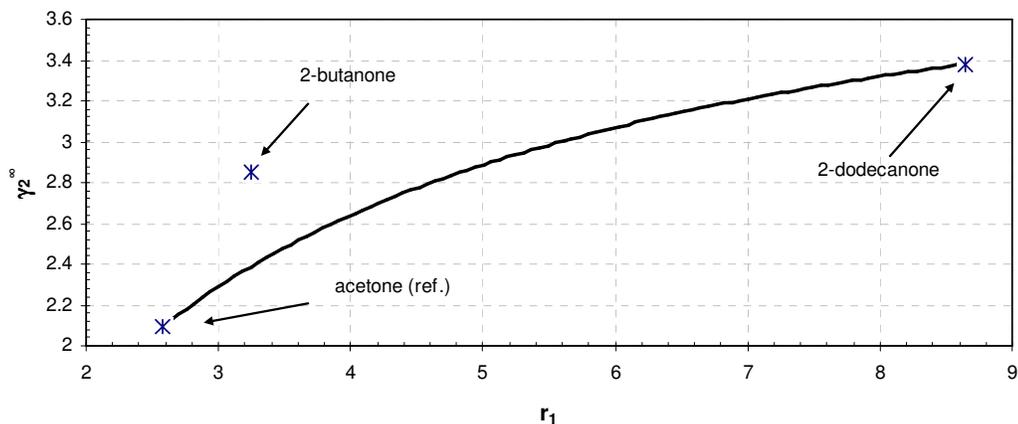


Fig. 22 γ_2^∞ vs. r_1 for ethanol (2) in ketone solvents(1) at 313.15 K (x – data from the DDB [33], — Eqn. (6-11)).

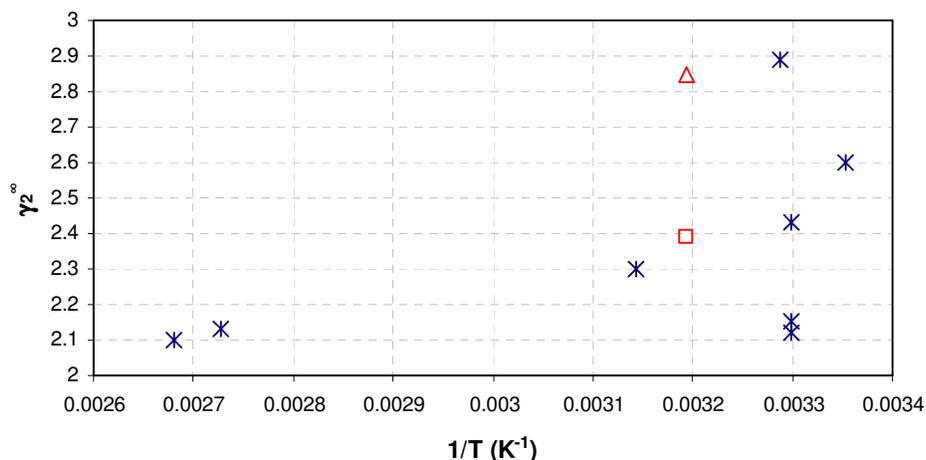


Fig. 23 γ_2^∞ vs. $1/T$ for ethanol (2) – 2-butanone (1) (x – data from the DDB [33], Δ – experimental data at 313.15 K, \square – value predicted from Eqn. (6-11) at 313.15 K)

6.4. NRTL-SAC solvent interpolation

The success of the “surface contribution” methods described in the previous sections suggests that the principle is fairly sound when the appropriate surface fraction is chosen and when a suitable chemical family is defined. It is however not always the case that there are multiple solvents from the same homologous series; in some cases there may only be one or two data from any series. Therefore it would be useful to be able to interpolate between solvents from different chemical families. This is essentially what the NRTL-SAC approach allows one to do, however the equations involved are tedious to implement and the regression of the models parameters can be a bit problematic if not carried out carefully.

One of the biggest advantages of the NRTL-SAC approach is that the model parameters make sense in that they reflect what we would expect. In the previous sections infinite dilution

activity coefficient was split into two parts a CH₂ part and functional group part. If this is extended to more general terms:

$$\ln \gamma_i^\infty = r_X \ln \gamma_i^{\infty, X} + r_{Y^-} \ln \gamma_i^{\infty, Y^-} + a_{Y^+} \ln \gamma_i^{\infty, Y^+} + a_Z \ln \gamma_i^{\infty, Z} \quad (6-12)$$

or

$$\gamma_i^\infty = (\gamma_i^{\infty, X})^{r_X} (\gamma_i^{\infty, Y^-})^{r_{Y^-}} (\gamma_i^{\infty, Y^+})^{r_{Y^+}} (\gamma_i^{\infty, Z})^{r_Z} \quad (6-13)$$

where r_X , r_{Y^-} , r_{Y^+} and r_Z are the NRTL-SAC hydrophobic, polar negative, polar positive and hydrophilic surface segments respectively; the terms in brackets are the model parameters. Therefore, as with NRTL-SAC, if data is known about four sufficiently different solvents then the model parameters can be fitted and solvent which has NRTL-SAC parameters can be predicted. Table 12 shows the results when using Eqn. (6-13) for four solutes in various solvents. Considering that the NRTL-SAC parameters are used “as is” and are not modified in any way the results are quite good. In many cases, at least qualitatively, the correct behaviour of the infinite dilution activity coefficient is observed.

For the hydrophobic solutes the behaviour in water and 1,2-ethandiol cannot be reproduced with any kind of reasonable accuracy. When comparing the fits in Table 13 with the NRTL-SAC fits the results are comparable in most cases. The NRTL-SAC prediction is slightly better in some cases, which is unsurprising since the parameters are fitted to the equation; however in many cases the NRTL-SAC prediction is significantly worse.

Considering that the results are fairly good even for the raw NRTL-SAC parameters it could mean that a new table of surface segment values specific to this method could be drawn up. The great power of a method such as this is that if a method can be developed for the infinite dilution activity coefficient in four suitably different solvents then these predictions can be extended to all other solvents with relative ease.

Table 12 Comparison of the experimental and fitted predictions for infinite dilution activity coefficient using Eqn. (6-13) for 4 different solutes at 298.15 K (* - the 4 reference solvents) using the NRTL-SAC parameters from [37]. All data from the DDB [33].

Solvent	Tetramethylstannane		Ethanol		Benzene		Ethyl iodide	
	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}
1,1,1-Trichloroethane	1.66	3.0	-	-	-	-	-	-
1,2-Dichloroethane	4.87	12.0	-	-	-	-	1.44	3.0
1,2-Ethandiol	464	76.7	1.86	1.1	44.2	91.5	-	-
1,4-Dioxane	-	-	-	-	1.08	1.5	1.34	1.8
1-Butanol	6.13	8.2	0.93	6.1	3.786	13.8	3.16	4.0
1-Octanol	4.5	10.6	1.256	28.2	2.07	20.4	-	-
1-Pentanol	-	-	8.24	6.6	48	14.3	-	-
1-Propanol	8.36	9.8	0.92	5.2	3.3	16.7	3.88	4.3
2-Butanone	-	-	2.6	3.6	1.12	1.7	1.42	2.1
2-Propanol	8	14.7	0.87	5.0	-	-	3.85	5.3
Acetic acid	-	-	-	-	4.4	3.9	-	-
Acetone	7.98	8.0*	2.44	2.4*	1.7	1.7*	2.1	2.1*

Solvent	Tetramethylstannane		Ethanol		Benzene		Ethyl iodide	
	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}
Acetonitrile	27.4	25.2	3.79	2.4	2.77	2.1	4.25	2.9
Benzene	2.28	3.0	11.09	11.3	-	-	1.07	2.1
Chlorobenzene	2.28	6.6	-	-	-	-	1.01	2.9
Chloroform	2.07	1.9	-	-	35.3	1.3	0.96	1.6
Cyclohexane	1.25	1.2	23.8	20.4	0.96	1.4	-	-
Dichloromethane	4	5.6	-	-	-	-	2.06	2.5
Diethyl ether	-	-	-	-	-	-	2.26	1.6
Dimethyl sulfoxide	62	14.6	0.53	0.1	3.05	2.1	3.11	1.4
Ethanol	15.14	15.1*	-	-	26.63	26.6*	5	5.0*
Ethyl acetate	3.88	5.7	1.39	4.9*	1.04	1.7	9.16	2.2
Heptane	-	-	49	49.0	1.55	1.5*	-	-
Hexane	1.2	1.2*	60.1	29.3	0.75	1.5	1.9	1.9*
Methanol	31	16.5	1.13	1.7	7.55	22.5	7.98	4.3
Methyl phenyl ether	-	-	-	-	4.4	2.1	-	-
N,N-Dimethylacetamide	-	-	0.57	0.6*	2.11	2.1*	-	-
N,N-Dimethylformamide	15.3	15.3*	-	-	-	-	1.92	1.9*
Nitromethane	-	-	19.1	5.7	2.931	3.1	4.81	4.3
N-Methyl-2-pyrrolidone	11	18.7	0.67	0.8	1.6	2.4	-	-
Pyridine	-	-	0.8	2.8	1.2	4.7	-	-
Sulfolane	-	-	3	3.0*	2.4	40.4	-	-
Tetrachloromethane	1.54	2.0	27.84	13.8	-	-	-	-
Tetrahydrofuran	2.04	3.5	2.23	3.1	-	-	0.93	1.7
Toluene	2.5	3.2	15.43	11.2	0.98	1.6	1.1	2.1
Triethylamine	-	-	1.96	3.7	-	-	-	-
Water	400000	61.9	3.744	2.2	2289	152.2	2165	9.8

Table 13 Comparison of the experimental and fitted predictions for infinite dilution activity coefficient using NRTL-SAC for 4 different solutes at 298.15 K (* - the 4 reference solvents) using the NRTL-SAC parameters from [37]. All data from the DDB [33].

Solvent	Tetramethylstannane		Ethanol		Benzene		Ethyl iodide	
	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}
1,1,1-Trichloroethane	1.66	0.8	-	-	-	-	-	-
1,2-Dichloroethane	4.87	3.3	-	-	-	-	1.44	1.2
1,2-Ethanediol	464	8868.9	1.86	763.6	44.2	2.E+06	-	-
1,4-Dioxane	-	-	-	-	1.08	0.1	1.34	1.2
1-Butanol	6.13	5.1	0.93	0.5	3.786	7.2	3.16	3.5
1-Octanol	4.5	3.5	1.256	0.7	2.07	9.7	-	-
1-Pentanol	-	-	8.24	0.6	48	7.2	-	-
1-Propanol	8.36	6.5	0.92	0.6	3.3	9.5	3.88	3.8
2-Butanone	-	-	2.6	1.4	1.12	0.1	1.42	1.2
2-Propanol	8	10.0	0.87	0.9	-	-	3.85	4.5
Acetic acid	-	-	-	-	4.4	7.5	-	-
Acetone	7.98	8.0*	2.44	2.3*	1.7	0.3*	2.1	1.7*
Acetonitrile	27.4	49.1	3.79	19.5	2.77	14.6	4.25	3.7
Benzene	2.28	0.8	11.09	5.3	-	-	1.07	1.0
Chlorobenzene	2.28	1.4	-	-	-	-	1.01	1.0
Chloroform	2.07	0.4	-	-	35.3	0.0	0.96	0.7
Cyclohexane	1.25	0.8	23.8	31.2	0.96	0.7	-	-
Dichloromethane	4	1.6	-	-	-	-	2.06	1.0
Diethyl ether	-	-	-	-	-	-	2.26	0.7
Dimethyl sulfoxide	62	66.9	0.53	2.5	3.05	25.6	3.11	4.1
Ethanol	15.14	15.1*	-	-	26.63	26.6*	5	5.0*
Ethyl acetate	3.88	2.1	1.39	1.6	1.04	0.1	9.16	1.0
Heptane	-	-	49	49.0*	1.55	1.7*	-	-
Hexane	1.2	0.9*	60.1	39.0	0.75	1.1	1.9	1.7*
Methanol	31	181.3	1.13	11.8	7.55	711.2	7.98	12.4

Solvent	Tetramethylstannane		Ethanol		Benzene		Ethyl iodide	
	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}	γ^{exp}	γ^{fit}
Methyl phenyl ether	-	-	-	-	4.4	0.3	-	-
N,N-Dimethylacetamide	-	-	0.57	0.8*	2.11	2.4*	-	-
N,N-Dimethylformamide	15.3	15.3*	-	-	-	-	1.92	2.1*
Nitromethane	-	-	19.1	35.4	2.931	9.3	4.81	2.7
N-Methyl-2-pyrrolidone	11	11.7	0.67	0.8	1.6	2.0	-	-
Pyridine	-	-	0.8	1.6	1.2	1.2	-	-
Sulfolane	-	-	3	3.0	2.4	93.6	-	-
Tetrachloromethane	1.54	0.7	27.84	3.5	-	-	-	-
Tetrahydrofuran	2.04	1.2	2.23	0.6	-	-	0.93	0.8
Toluene	2.5	0.9	15.43	5.2	0.98	0.1	1.1	1.0
Triethylamine	-	-	1.96	0.2	-	-	-	-
Water	400000	1.E+09	3.744	2.E+08	2289	8.E+15	2165	26456

Since the solvent interpolation showed such great promise with the “un-tweaked” NRTL-SAC parameters, the ideal was to see how the interpolation would behave when a new fitted set of NRTL-SAC-analogous parameters were used. The four reference solvents that were chosen were water, hexane, acetone and N-methyl-2-pyrrolidone. They were chosen since they represent the full spectrum of surface segments. The only empirical modification is that water was assigned a surface segment value of two as opposed to 1 because, as shown in Table 12 the values of water are frequently wrong. Therefore for the other three solvents the NRTL-SAC segment values were used, whether they are right or wrong is not a huge problem since the fitted parameters are just relative to these and will adjust accordingly. The solutes used were: ethanol, n-octane, benzene and tetramethylstannane (choice based only on the availability of data for the four reference solvents).

The new NRTL-SAC-analogous parameters were fitted by minimising the square of the errors for all four solutes. The results obtained are shown in Table 14 and 15 and are very promising. These results are by no means final, they should just be used as an illustration of the potential that there seems to be in such a method. The regressed parameters for the interpolation are shown in Table 16 and as with the NRTL-SAC parameters they seem to have some physical realism in that the surfaces have values which seem intuitively correct.

Table 15 shows the results for the two solutes whose data was not included in the parameter regression. In most cases the results are fairly good and where there is a large deviation (e.g. cyclohexane-methanol and n-heptane-2-butanone) it is the data which seems to be erroneous and not the prediction. This method seems to hold a lot of promise as it is very simple to do “hand calculations” and the results are far more accurate than the NRTL-SAC equivalent interpolations. This is in no way meant as a replacement for NRTL-SAC as there are severe limitations to this method (such as concentration dependence) but it does fill a useful niche and suits the purposes of this work quite nicely.

Table 14 The results of the solvent interpolation for the fitted values of the NRTL-SAC analogous parameters (* - the 4 reference solvents for which the original NRTL-SAC parameters are used)

Solvent	Ethanol		Tetramethylstannane		Benzene	
	$\gamma^{\infty \text{ exp}}$	$\gamma^{\infty \text{ fit}}$	$\gamma^{\infty \text{ exp}}$	$\gamma^{\infty \text{ fit}}$	$\gamma^{\infty \text{ exp}}$	$\gamma^{\infty \text{ fit}}$
Acetonitrile	4.1	4.0	27.4	27.5	3.1	2.1
Acetone*	2.4	2.4	8.0	8.0	1.7	1.7
1,2-Ethanediol	1.9	1.4	464.0	435.5	44.2	100.6
Ethyl acetate	1.4	1.4	3.9	3.8	1.0	1.4
Benzene	11.1	11.1	2.3	2.3	-	-
1-Butanol	0.9	1.2	6.1	5.3	3.8	3.7
2-Butanone	2.6	2.6	-	-	1.1	1.1
Cyclohexane	23.8	23.8	1.3	1.2	1.0	0.8
Hexane*	61.3	61.3	1.2	1.2	0.8	0.7
Heptane	62.1	62.1	-	-	1.5	0.7
2-Propanol	0.9	1.2	8.0	6.5	-	-
Methanol	1.1	3.4	31.0	23.6	7.6	10.8
Nitromethane	19.1	19.1	-	-	4.9	5.0
1-Propanol	0.9	1.4	8.4	6.5	3.3	4.2
Pyridine	0.8	1.9	-	-	1.2	1.8
Dimethyl sulfoxide	0.8	0.6	62.0	62.0	3.6	2.9
Tetrachloromethane	27.8	27.8	1.5	1.5	-	-
Tetrahydrofuran	2.2	2.2	2.0	2.0	-	-
Toluene	15.4	15.4	2.5	2.1	1.0	1.0
Water*	4.1	4.1	4.0E+05	4.0E+05	2.4E+04	2.4E+04
N,N-Dimethylacetamide	0.6	0.7	-	-	2.1	1.8
N-Methyl-2-pyrrolidone*	0.7	0.7	11.0	11.0	1.6	1.6
1-Octanol	1.3	1.4	4.5	3.1	2.1	2.4

Table 15 The results of the solvent interpolation for the fitted values of the NRTL-SAC analogous parameters (* - the 4 reference solvents for which the original NRTL-SAC parameters are used)

Solvent	n-Octane		2-Butanone		Methanol	
	$\gamma^{\infty \text{ exp}}$	$\gamma^{\infty \text{ fit}}$	$\gamma^{\infty \text{ exp}}$	$\gamma^{\infty \text{ fit}}$	$\gamma^{\infty \text{ exp}}$	$\gamma^{\infty \text{ fit}}$
Acetonitrile	61.3	61.3	1.2	1.9	3.1	2.8
Acetone*	11.0	11.0	1.1	1.1	1.8	1.8
1,2-Ethanediol	2250.0	2252.8	8.4	6.1	-	-
Ethyl acetate	4.8	4.9	-	-	-	-
Benzene	2.3	2.3	-	-	25.2	12.3
1-Butanol	7.5	8.0	2.1	1.7	-	-
2-Butanone	5.3	5.3	-	-	-	-
Cyclohexane	2.1	1.1	-	-	7.3	32.9
Hexane*	1.1	1.1	29.4	29.4	93.5	93.5
Heptane	1.1	1.1	4.8	29.7	92.6	94.8
2-Propanol	9.6	10.4	-	-	0.8	1.3
Methanol	48.0	50.3	34.2	5.7	-	-
Nitromethane	125.6	125.6	-	-	-	-
1-Propanol	9.5	10.2	2.3	2.0	-	-
Pyridine	9.2	9.1	1.0	0.7	-	-
Dimethyl sulfoxide	200.0	200.0	2.9	0.4	0.5	0.3
Tetrachloromethane	1.3	1.4	-	-	-	-
Tetrahydrofuran	2.3	2.3	-	-	2.5	2.2
Toluene	1.8	2.1	1.4	7.2	20.4	18.1
Water*	1.0E+07	1.0E+07	65.7	65.7	5.9	5.9
N,N-Dimethylacetamide	18.5	18.5	1.1	0.6	0.5	0.5
N-Methyl-2-pyrrolidone*	33.6	33.6	1.3	1.3	0.5	0.5
1-Octanol	3.4	4.1	2.3	1.7	1.3	1.5

Table 16 The model parameters for the interpolations in Table 14 and Table 15

Solvent	X	Y-	Y+	Z
Acetonitrile	0.385	0.443	0.685	-
Acetone	0.131	0.109	0.513	-
1,2-Ethandiol	-	0.143	-	0.905
Ethyl acetate	0.056	0.107	0.313	-
Benzene	0.529	-	0.198	-
1-Butanol	-	-	-	0.258
2-Butanone	0.423	0.405	0.103	-
Cyclohexane	0.770	-	-	-
Hexane	1.000	-	-	-
Heptane	1.003	-	-	-
2-Propanol	-	-	-	0.290
Methanol	0.214	-	-	0.484
Nitromethane	0.445	-	0.862	0.165
1-Propanol	0.032	-	-	0.288
Pyridine	-	-	0.534	0.007
Dimethyl sulfoxide	-	0.654	0.831	-
Tetrachloromethane	0.790	-	0.066	-
Tetrahydrofuran	0.190	0.071	0.153	-
Toluene	0.616	-	0.172	-
Water	-	-	-	2.000
N,N-Dimethylacetamide	-	0.390	0.436	-
N-Methyl-2-pyrrolidone	0.252	0.790	0.281	-
1-Octanol	0.049	-	-	0.175

7. NRTL-SAC UNIFAC MATRIX INTERPOLATION

As mentioned above (see section 4.2.2) one of the main reasons that UNIFAC cannot be applied to solutions containing complex molecules is that there are missing parameters in the UNIFAC matrix. In order to fill these gaps with plausible numbers there would need to be large amounts of measurements on some fairly exotic systems. This is both time consuming and very expensive. It is for this reason that a method is needed that enables the prediction of the missing parameters either from the existing parameters or from some prediction technique. The need for interaction parameters sometimes results in parameters being predicted from COSMO predictions. Even though this prediction could be way off the need for the interaction parameters is so great that almost any number will do.

Another approach is the one taken by Gani [50] which is to fit “connectivity indices” to the binary interaction parameters. In essence the binary interaction parameters are treated as pure component properties and the model parameters are fitted to them. The idea of connectivity indices has been used quite extensively by Gani [50], with moderate success, to predict pure component properties. The problem, however, with using the similar procedure with the UNIFAC parameters is that it is assumed that the parameters have some physical significance. This assumption may not always be a good one and therefore any predicted interaction parameters could be drastically affected by this.

While the interaction parameters may not be physically realistic, the infinite dilution activity coefficients in many cases are. The idea was therefore to combine the physical realism of the infinite dilution data with the “molecular surface interpolation” type idea of NRTL-SAC. The surface segments for the groups were assigned by considering hypothetical situations of an infinite dilution of one group in another (so for example the infinite dilution activity coefficient of CH₂ group in a mixture of OH groups). The obvious advantage of this is that if the binary interactions for three or four (depending on whether a hydrogen bonding group is assigned or not) groups are known then all the other interactions could be interpolated.

Table 17 shows some of the UNIFAC group-in-group matrix to which the surface segments were fitted to. The r and q values were taken as the same as the first secondary group that appears in the main group (i.e. the one with the lowest number).

Table 17 Matrix of the infinite dilution activity coefficients of one main group in another at 298.15 K (Grp num – main group number, Grp – main group name)

Grp num	Grp num	1	5	9	11	14
Grp num	Grp	CH ₃	CH ₂ =CH	ACH	ACCH ₃	OH (P)
1	CH ₃		1.39	1.22	1.76	1158.81
5	CH ₂ =CH	1.15		0.76	1.14	116.71
9	ACH	1.48	0.65		0.82	56.83
11	ACCH ₃	1.58	1.51	0.87		73.45
14	OH (P)	28.58	63.23	2.76		

As a first attempt the surface segments were found by using the same parameters as the original NRTL-SAC. The objective function (F) used in the regression was:

$$F = \sum_{i=1}^m \frac{(\gamma_i^{\infty, \text{UNIFAC}} - \gamma_i^{\infty, \text{NRTL-SAC}})}{m \times \gamma_i^{\infty, \text{UNIFAC}}} \quad (7-1)$$

where m is the total number of UNIFAC group in group infinite dilution activity coefficients. There were no constraints on the surface segments except that CH_3 was assumed to only have an X segment and water was assumed to only have a Z segment (in accordance with the NRTL-SAC model development). Table 18 shows the regressed surface segment values for some of the UNIFAC groups. Even though there were no constraints placed on the groups they seem to be quite physically realistic. The hydrocarbon groups have predominantly a hydrophobic (X) surface while for example the OH group has mainly a hydrophilic (Z) surface which is what one would expect. The physical realism of the groups is found for most cases.

Table 18 The regressed surface segments for some of the UNIFAC groups

Grp num	Grp	Surface segments			
		X	Y-	Y+	Z
1	CH_3	0.336	-	-	-
5	$\text{CH}_2=\text{CH}$	0.326	0.025	0.166	-
9	ACH	0.166	0.015	0.141	-
11	ACCH_3	0.300	0.019	0.058	-
14	OH (P)	0.001	0.063	0.000	1.071

The objective function for the first 27 main groups in the UNIFAC matrix was found to be 0.39. This moderate error and the good physical realism observed by the groups seem to indicate that this method is based on sound principles. The problem is however reproducing the infinite dilution activities when one of the values is large relative to the others. For example consider Table 17, the OH(P)-CH_3 activity coefficient is much higher than any of the others in that column. This problem is almost solely associated with the interactions involving OH groups (water included). The NRTL-SAC model parameters were also regressed as an attempted workaround but this didn't reduce the error by any appreciable extent. Many different methods are currently being tried so this is very much a work in progress.

8. FURTHER WORK

- **Evaluation of the NRTL-UNIFAC approach**

This approach looks promising but further evaluation is needed to assess whether it provides interpolations which have suitable accuracy. The time frame for this analysis is uncertain but may be along the lines of a month or two.

- **Solvent interpolation**

It has been shown in section 6.4 that this process could be possible; further work needs to go into evaluating what the best way to do it is. This would probably entail drawing up a list of solvent specific parameters for a list of common solvents, similar to the way that NRTL-SAC does it. Considering that the method in section 6.4 seems more or less usable the time needed to complete should be 4 – 6 weeks.

- **Infinite dilution data as pure component property**

It has been shown in literature with methods such as AQUAFAC [51] that the infinite dilution activity coefficient in a single solvent can be predicted via group contributions. Much of the work in this project has been associated with predicting pure component properties and therefore most, if not all, of the tools are in place to carry out infinite dilution activity as a pseudo pure component property. This could take a very long time if unforeseen problems are encountered, the unpredictability of such an approach could mean a time span of about 3 – 4 months.

- **Data compilation**

Data compilation is probably the single most time consuming aspect of this work. It is important that there is data of sufficient accuracy in order for the appropriate conclusions to be drawn from the data. At the same time it is important to have data for a fairly diverse range of compounds in order that the method is more generally applicable. It is difficult to estimate the amount of time, but with things going smoothly could be done in about 2 months as some has been done already.

- **Other approaches**

While the general aims of this work are fixed the methods that should be used in getting there are by no means certain, therefore any further work may result in a slightly different approach being taken to the general approach outlined above.

9. REFERENCES

1. Nannoolal, Y., Rarey, J., Ramjugernath, D., Cordes, W., Estimation of Pure Component Properties, Part 1: Estimation of the Normal Boiling Point of Non-Electrolyte Organic Compounds via Group Contributions and Group Interactions. *Fluid Phase Equilib.*, 2004. **226**: p. 45-63.
2. Nannoolal, Y., Rarey, J., Ramjugernath, D., Estimation of Pure Component Properties Part 2: Estimation of Critical Data by Group Contribution. *Fluid Phase Equilib.*, 2007. **252**: p. 1.
3. Nannoolal, Y., Rarey, J., Ramjugernath, D., Estimation of Pure Component Properties Part 3: Estimation of the Vapor Pressure of Non-Electrolyte Organic Compounds via Group Contributions and Group Interactions. *Fluid Phase Equilib.*, 2008. **269**: p. 117-133.
4. Nannoolal, Y., *Development and Critical Evaluation of Group Contribution Methods for the Estimation of Critical Properties, Liquid Vapour Pressure and Liquid Viscosity of Organic Compounds*, in *Chemical Engineering*. 2006, University of Kwa-Zulu Natal: Durban.
5. Moller, B., Rarey, J., Ramjugernath, D., Estimation of the Vapour Pressure of Non-Electrolyte Organic Compounds via Group Contributions and Group Interactions. *J. Mol. Liq.*, 2008. **143**: p. 52-63.
6. Mellan, I., *Industrial Solvents Handbook*. 2nd ed. 1977, New Jersey: Noyes Data Corporation.
7. Karunanithi, A. T., Achenie, L. E. K., Gani, R., A Computer-Aided Molecular Design Framework for Crystallization Solvent Design. *Chem. Eng. Sci.*, 2006. **61**: p. 1247-1260.
8. Pretel, E. J., Lopez, P. A., Bottini, S. B., Brignole, E. A., Computer-Aided Molecular Design of Solvents for Separation Processes. *AIChE J.*, 1994. **40**(8): p. 1349-1360.
9. Yang, X., Song, H., Computer Aided Molecular Design of Solvents for Separation Processes. *Chem. Eng. Technol.*, 2006. **29**(1): p. 33-43.
10. Song, J., Song, H., Computer-Aided Molecular Design of Environmentally Friendly Solvents for Separation Processes. *Chem. Eng. Technol.*, 2008. **31**: p. 177-187.
11. Reichardt, C., *Solvents and Solvent Effects in Organic Chemistry*. 3rd ed. 2003, Weinheim: Wiley-VCH.
12. Myrdal, P. B., Krzyzaniak, J. F., Yalkowsky, S. H., Modified Trouton's Rule for Predicting the Entropy of Boiling. *Ind. Eng. Chem. Res.*, 1996. **35**: p. 1788-1792.
13. Chastrette, M., Rajzmann, M., Chanon, M., Purcell, K. F., Approach to a General Classification of Solvents Using a Multivariate Statistical Treatment of Quantitative Solvent Parameters. *J. Am. Chem. Soc.*, 1985. **107**(1): p. 1-11.
14. Gramatica, P., Navas, N., Todeschini, R., Classification of Organic Solvents and Modelling of their Physico-Chemical Properties by Chemometric Methods using Different Sets of Molecular Descriptors. *Trends Anal. Chem.*, 1999. **18**(7): p. 461-471.
15. Hildebrand, J. H., Scott, R. L., *Solubility of Nonelectrolytes*. 3rd ed. 1964, New York: Dover.

16. Prausnitz, J. M., Lichtenhaler, R. N., de Azevedo, E. G., Molecular Thermodynamics of Fluid-Phase Equilibria. 3rd ed. 1999, New Jersey: Prentice-Hall.
17. Reid, R. C., Prausnitz, J. M., Poling, B. E., The Properties of Gases and Liquids. 4th ed. 1987, New York: McGraw-Hill.
18. Smith, J. M., Van Ness, H. C., Abbott, M. M., Introduction to Chemical Engineering Thermodynamics. 6th ed. 2001: McGraw-Hill.
19. Frank, T. C., Downey, J. R., Gupta, S. K., Quickly Screen Solvents for Organic Solids. *Chem. Eng. Progr.*, 1999: p. 41.
20. Wilson, G. M., A New Expression for the Excess Free Energy of Mixing. *J. Am. Chem. Soc.*, 1963. **86**: p. 127-130.
21. Tsuboka, T., Katayama, T., Modified Wilson Equation for Vapor-Liquid and Liquid-Liquid Equilibria. *J. Chem. Eng. Jpn.*, 1975. **8**(3): p. 181-187.
22. Huang, J.-F., Lee, L.-S., A Modification of the Wilson Equation With Only Binary Parameters for Ternary Liquid-Liquid Equilibria. *Fluid Phase Equilib.*, 1994. **101**: p. 27-51.
23. Renon, H., Prausnitz, J. M., Local Composition in Thermodynamic Excess Functions for Liquid Mixtures. *AIChE J.*, 1968. **14**(1): p. 135-144.
24. Maurer, G., Prausnitz, J. M., On the Derivation and Extension of the UNIQUAC Equation. *Fluid Phase Equilib.*, 1978. **2**: p. 91-99.
25. Wilson, G. M., Deal, C. H., Activity Coefficients in Changing Environments - Solutions of Groups. *Ind. Eng. Chem. Fund.*, 1962. **1**(1): p. 20-23.
26. Deal, C. H., Derr, E. L., Group Contributions in Mixtures. *Ind. Eng. Chem.*, 1968. **60**(4): p. 28-38.
27. Kojima, K., Tochigi, K., Prediction of Vapor-Liquid Equilibria by the ASOG Method. 1979, Tokyo: Kodansha Ltd.
28. Fredenslund, A., Jones, R. L., Prausnitz, J. M., Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE J.*, 1975. **21**(6): p. 1086-1099.
29. Larsen, B. L., Rasmussen, P., Fredenslund, A., A Modified UNIFAC group-Contribution Model for Prediction of Phase Equilibria and Heats of Mixing. *Ind. Eng. Chem. Res.*, 1987. **26**: p. 2274-2286.
30. Weidlich, U., Gmehling, J., A Modified UNIFAC Model. 1. Prediction of VLE, h^E and γ^E . *Ind. Eng. Chem. Res.*, 1987. **26**: p. 1372-1381.
31. Bondi, A., van der Waals Volumes and Radii. *J. Phys. Chem.*, 1964. **68**(3): p. 441-451.
32. Beilstein, F. K., CrossFire Beilstein, Frankfurt (2008) <http://info.crossfiredatabases.com/>
33. Gmehling, J., Rarey, J., Menke, J., Dortmund Data Bank, Oldenburg (2008) <http://www.ddbst.com>
34. Barton, A. F. M., CRC Handbook of Solubility Parameters and Other Cohesion Parameters. 2nd ed. 1991, London: CRC Press.
35. Hansen, C. M., Hansen Solubility Parameters: A User's Handbook. 2nd ed. 2007: CRC Press.

-
36. Macedo, E. A., Solubility of Amino Acids, Sugars, and Proteins. *Pure Appl. Chem.*, 2005. **77**(3): p. 559-568.
 37. Chen, C.-C., Crafts, P. A., Correlation and Prediction of Drug Molecule Solubility in Mixed Solvent Systems with the Nonrandom Two-Liquid Segment Activity Coefficient (NRTL-SAC) Model. *Ind. Eng. Chem. Res.*, 2006. **45**: p. 4816-4824.
 38. Chen, C.-C., Song, Y., Solubility Modelling with a Nonrandom Two-Liquid Segment Activity Coefficient Model. *Ind. Eng. Chem. Res.*, 2004. **43**: p. 8354-8362.
 39. Klamt, A., Eckert, F., COSMO-RS: A Novel and Efficient Method for the A Priori Prediction of Thermophysical Data of Liquids. *Fluid Phase Equilib.*, 2000. **172**: p. 43-72.
 40. Lei, Z., Chen, B., Li, C., Lui, H., Predictive Molecular Thermodynamic Models for Liquid Solvents, Solid Salts, Polymers and Ionic Liquids. *Chem. Rev.*, 2008. **108**: p. 1419-1455.
 41. Lin, S.-T., Sandler, S. I., A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.*, 2002. **41**: p. 899-913.
 42. Mullins, E., Oldland, R., Liu, Y. A., Wang, S., Sandler, S. I., Chen, C.-C., Zwolak, M., Seavey, K. C., Sigma-Profile Database for Using COSMO-Based Thermodynamic Methods. *Ind. Eng. Chem. Res.*, 2006. **45**(12): p. 4389-4415.
 43. Vera, J. H., Sayegh, S. G., Ratcliff, G. A., A Quasi Lattice Local Composition Model For The Excess Gibbs Free Energy of Liquid Mixtures. *Fluid Phase Equilib.*, 1977. **1**: p. 113-135.
 44. Kikic, I., Alessi, P., Rasmussen, P., Fredenslund, A., On the Combinatorial Part of the UNIFAC and UNIQUAC Models. *Can. J. Chem. Eng.*, 1980. **58**: p. 253-258.
 45. Kontogeorgis, G. M., Fredenslund, A., Tassios, D. P., Simple Activity Coefficient Model for the Prediction of Solvent Activities in Polymer Solutions. *Ind. Eng. Chem. Res.*, 1993. **32**: p. 362-372.
 46. Bekker, A. Y., Knox, D. E., Sund, S. E., Prediction of Solvent Activities Using the UNIFAC-FV Model. *J. Solution Chem.*, 1986. **16**(8): p. 635-639.
 47. Pappa, G. D., Kontogeorgis, G. M., Tassios, D. P., Prediction of Ternary Liquid-Liquid Equilibria in Polymer-Solvent-Solvent Systems. *Ind. Eng. Chem. Res.*, 1997. **36**: p. 5461-5466.
 48. Wibawa, G., Takishima, S., Sato, Y., Masuoka, H., An Improved Prediction Result of Entropic-FV Model for Vapor-Liquid Equilibria of Solvent-Polymer Systems. *J. Appl. Polym. Sci.*, 2005. **97**: p. 1145-1153.
 49. Tse, G., Sandler, S. I., Determination of Infinite Dilution Activity Coefficients and 1-Octanol/Water Partition Coefficients of Volatile Organic Pollutants. *J. Chem. Eng. Data.*, 1994. **39**: p. 354-357.
 50. Gonzalez, H. E., Abildskov, J., Gani, R., A Method for Prediction of UNIFAC Group Interaction Parameters. *AIChE J.*, 2007. **53**(6): p. 1620-1632.
 51. Lee, Y.-C., Myrdal, P. B., Yalkowsky, S. H., Aqueous Functional Group Activity Coefficients (AQUAFAC) 4: Application to Complex Organic Compounds. *Chemosphere*, 1996. **33**(11): p. 2129-2144.
-