Prediction of Partition Coefficients and Solubilities of Active Pharmaceutical Ingredients with the SAFT-\(\gamma\) Mie Group-contribution Approach

Panatpong Hutacharoen

A thesis submitted for the Degree of Doctor of Philosophy and the Diploma of Imperial College London.

Centre for Process Systems Engineering
Department of Chemical Engineering
Imperial College London
London SW7 2AZ, United Kingdom

17th May 2017
Declaration of Originality

I, Panatpong Hutacharoen, certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution-Non Commercial-No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or distribution, researchers must make clear to others the licence terms of this work.
Abstract

Partition coefficient and solubility are very useful properties in a variety of product and process design problems. Especially in the octanol-water system, partition coefficients ($K_{i,OW}$) are used as indicators for a drug's lipophilicity which is a key physicochemical property in drug design. Solid phase solubility is a fundamental parameter in the design of crystallisation processes commonly used in the pharmaceutical and agrochemical industries. The ability to predict these properties from the molecular structure of compounds is therefore highly desirable. In this thesis, the recently developed SAFT-$\gamma$ Mie group-contribution (GC) equation of state is used as a predictive framework to study the thermodynamic properties of multifunctional compounds. The SAFT-$\gamma$ Mie approach allows one to determine the thermo-physical properties of molecules in terms of the constituent functional groups that represent their unique molecular identity. The parameters for each functional group are developed from fluid-phase equilibrium data for simple compounds and, once estimated, they are applied to the study of more complex molecules in a predictive manner. Novel SAFT-$\gamma$ models are developed for fundamental systems such as alkane + water and alcohol + water mixtures, which are typically involved in various chemical and biological applications. These GC models are able to describe accurately mutual solubilities of water and hydrocarbons which span more than ten orders of magnitude, and are also transferable to the modelling of multifunctional compounds. As a result, a quantitative prediction of $K_{i,OW}$ and solubility is achieved for several active pharmaceutical ingredients (API) including ibuprofen, ketoprofen, lovastatin, and simvastatin. We find that an important factor that needs to be taken into account in modelling these complex APIs is the formation of intramolecular hydrogen bonds (IMHB). IMHB have a pronounced effect on molecular structure and thermodynamic properties, but are often overlooked by other predictive approaches. Modelling complex organic molecules with the consideration of IMHB is challenging, especially for GC approaches which do not take into account details of molecular conformation. In this thesis, an effective treatment for IMHB is developed within the SAFT-$\gamma$ Mie framework and proven to improve the property prediction of molecules with IMHB, especially in highly associated solvents. The findings in this thesis validate the applicability of the SAFT-$\gamma$ Mie approach in modelling complex multifunctional molecules and demonstrate its broad relevance for the pharmaceutical industry.
Acknowledgements

The completion of my PhD would not have been possible without the support of many people. First and foremost, I wish to express my profound gratitude to my supervisors Professors Amparo Galindo, Claire Adjiman, and George Jackson for giving me the opportunity to pursue a PhD in the Molecular Systems Engineering (MSE) Group in the first place, and for all their tireless guidance and support throughout the last four years. I thank them for their valuable advice, invariable contribution of knowledge, and kind motivation. I also thank Dr. Ravi Shanker from Pfizer for sharing his invaluable experiences and insights, both in life and science. I consider the experience of working with all of them as a real privilege of my life.

A special thanks goes to my MSE friends, Sadia, Georgia, Gabriel, Katerina, Christianna, Niko, Sara, Smitha, Daniel, Tom, Ed(ward), Maz, and Hajar for every discussion, good time and memories we have had together in and outside C610. I would like to particularly thank Vassilis, Majid, Simon, and Andrew for sharing their knowledge on SAFT and gPROMS which were crucial at the beginning of my PhD. I would also like to thank my Thai friends, P'Por and P'Mhee, for their good company during my first year in London. My thanks also go to my friends back home: Ice, Korn, Fah, Ja, Nam, Camay, P’Yee; I am very fortunate to have all of you in my life. It is quite impossible to forget many great memories that I have shared with my squad: Neab, N'Game, N'Eye, N'Nutt, Au, and N'Im, I am very grateful for every moment we spend together. The two incredible people whom I owe many thanks to are P’Kong and Tan, especially for their constant motivation and tireless support.

Finally, I would like to thank my family, most importantly for the endless support and encouragement from my dearest parents, who, despite being far away, have always been by my side and made me feel loved. I hope I have made you proud.
For My Dearest Parents,

Grandmothers,

and Aunt

And

In Remembrance of His Majesty

King Bhumibol Adulyadej

(1927 - 2016)
Contents

Abstract iii

Content vi

List of Figures xvi

List of Tables xix

1 Foreword 1

2 Challenges and Opportunities for the Thermodynamic Prediction of Octanol-Water Partition Coefficients 5
   2.1 Introduction ................................................. 5
   2.2 Experimental determination of partition coefficients ......................... 9
   2.3 Thermodynamic routes to partition coefficients ................................ 11
   2.4 Predictive thermodynamic tools for partition coefficients ................. 14
       2.4.1 Activity coefficient models ................................... 16
       2.4.2 Statistical Associating Fluid Theory (SAFT) ....................... 23
       2.4.3 Models based on quantum mechanical (QM) calculations ............. 26
   2.5 Conclusions ................................................................ 31

3 Modelling the solvation of alkanes and alcohols in aqueous environment 33
   3.1 Introduction ................................................................ 33
   3.2 Theory and molecular models ......................................... 37
       3.2.1 The SAFT-γ Mie group-contribution approach .................... 37
       3.2.2 Combining Rules ................................................... 42
       3.2.3 Property calculations .............................................. 43

vi
3.3 Model development and phase equilibrium calculations ........................................ 45
  3.3.1 The CH₂OH group and its binary interaction parameters CH₂OH - CH₃ and CH₂OH - CH₂ ........................................ 47
  3.3.2 CH₃ - H₂O and CH₂ - H₂O binary interaction parameters .......................... 52
  3.3.3 CH₂OH - H₂O binary interaction parameters ........................................ 57
3.4 Prediction of solvation properties ................................................................. 61
3.5 Conclusions .................................................................................................... 65

4 Development of the SAFT-γ Mie group interaction parameters characterising
pharmaceutical molecules .................................................................................... 66
  4.1 Introduction ..................................................................................................... 66
  4.2 Parameter estimation of functional groups .................................................... 70
  4.3 Obtaining H₂O unlike group interaction parameters .................................... 72
  4.4 Obtaining CH₂OH unlike group interaction parameters .............................. 78
  4.5 Thermodynamic prediction with the SAFT-γ Mie approach ....................... 83
  4.6 Conclusions .................................................................................................. 88

5 Predicting the partition coefficients and solubilities of organic compounds ...... 89
  5.1 Introduction ..................................................................................................... 89
  5.2 Property calculations ..................................................................................... 92
  5.3 Results ........................................................................................................... 94
    5.3.1 Partition coefficients .............................................................................. 94
    5.3.2 Solubility .............................................................................................. 104
  5.4 Conclusions .................................................................................................. 106

6 The competition between inter- and intramolecular association in modelling
partition coefficients and solubilities of organic molecules ............................... 107
  6.1 Introduction ..................................................................................................... 107
  6.2 Definition of groups ..................................................................................... 111
  6.3 Effective treatment for intramolecular association and gauche conformation ... 114
  6.4 Results .......................................................................................................... 116
    6.4.1 Partition coefficients of alkylbenzenes, phenylalkanoic acids, and dicarboxylic
          acids ........................................................................................................ 116
    6.4.2 Octanol-water partition coefficients of APIs ........................................... 122
    6.4.3 Solubility prediction of APIs .................................................................. 124
6.5 Conclusions ................................................................. 134

7 Conclusions and future work ............................................. 135
  7.1 Summary of the key contribution of this thesis ................. 136
  7.2 Directions for future work ......................................... 137

Bibliography ................................................................. 140
List of Figures

2.1 The model description of the liquid-liquid equilibrium of the binary mixture of 1-octanol + water in water-rich phase (WR) and octanol-rich phase (OR) at a pressure of 0.101 MPa. The open circles [90] and the open triangles [142] are the experimental data. The dashed-green, dash-dot purple, dotted blue curves represent, UNIFAC, KOW UNIFAC, and modified UNIFAC calculations, respectively. 21

2.2 Octanol-water partition coefficient predictions of n-alkanes (top) and 1-alcohols (bottom) with UNIFAC (dashed green), KOW UNIFAC (dash-dot purple), modified UNIFAC (dotted blue), and Pharma Mod. UNIFAC (continuous red) approach. The open squares [143] and the open circles [104] represent the experimental data, and the error bars the associated uncertainty. . . . . . . . . . . 22

3.1 Example of the decomposition of a molecule into functional groups: 1-butanol is composed of one CH$_3$ group (shaded red), two CH$_2$ groups (shaded gray) and one CH$_3$OH group composed of two fused spherical segments (shaded green, with association sites in brown and yellow to mimic hydrogen-bonding interactions). . 38

3.2 Isothermal pressure–mole fraction (P–x) phase diagrams of the vapour-liquid equilibria of n-alkanes and 1-alcohols: (a) n-heptane + 1-pentanol mixture at temperatures of $T = 348.15$ K (circles) [259] and $T = 368.15$ K (triangles) [259], and (b) undecane + tetradecanol mixture at temperatures of $T = 393.15$ K (circles) [260] and $T = 413.15$ K (triangles) [260]. The symbols represent the experimental data, the dashed curves the SAFT-γ Mie predictions obtained with a CH$_2$OH model based on pure-component data alone, and the continuous curves the SAFT-γ Mie predictions obtained with a CH$_2$OH model developed from pure and mixture data (the parameters are reported in Tables 3.1 and 3.2). . . . . . . 51
3.3 Isothermal pressure-mole fraction ($P$-$x$) phase diagrams of the vapour-liquid equilibria of $n$-alkanes and short 1-alcohols: (a) $n$-butane + ethanol mixture at a temperatures of $T = 373.27$ K [261], and (b) $n$-heptane + 1-butanol mixture at a temperature of $T = 313.15$ K [262]. The symbols represent the experimental data and the curves the prediction with the SAFT-$\gamma$ Mie approach. The descriptions of the curves are given in Fig. 3.2. ......................................................... 52

3.4 Isobaric temperature-mole fraction ($T$-$x$) phase diagram of the liquid-liquid equilibria of $n$-tetradecane + ethanol at a pressure of $p = 0.101$ MPa. The circles [258] represent the experimental data and the curves the prediction with the SAFT-$\gamma$ Mie approach. The descriptions of the curves are given in Fig. 3.2. ........... 52

3.5 Solubilities of water in the $n$-alkane-rich liquid phase (dashed curves) and alkanes (C$_2$-C$_{10}$) in the water-rich liquid phase (continuous curves) at conditions of three-phase equilibria as a function of temperature. The diamonds and squares denote experimental correlated data for solubilities of water in $n$-pentane-rich and $n$-decane-rich phases, respectively. The circles correspond to experimental and experimental correlated data for various alkane solubilities in water, ranging from ethane to $n$-decane. The filled symbols represent the mutual solubility data employed in the determination of the unlike CH$_3$ – H$_2$O and CH$_2$ – H$_2$O interaction parameters. The data for ethane + water, propane + water and $n$-butane + water are from Mokraoui et al [269]. The data for $n$-pentane + water to $n$-decane + water are from the IUPAC-NIST solubility data series [263–268]. .......... 54

3.6 Solubilities of $n$-alkanes in the water-rich liquid phase at conditions of three-phase equilibria at $T = 298.00$ K. The filled circles represent the experimental correlated data from IUPAC-NIST [263–268]; the circles [269], triangles [270] and diamonds [271] the experimental data; the crosses [196] the MD Simulation data; the continuous curve the SAFT-$\gamma$ Mie calculations using parameters reported in Tables 3.1 and 3.2, and the dash curve the SAFT-$\gamma$ Mie calculations using previously reported parameters [169].......................... 55
3.7 Isothermal pressure–mole fraction \((P-x)\) phase diagrams of (a) \(n\)-butane + water mixture at a temperature of \(T = 477.59\) K [272], above the critical point of \(n\)-butane \((T_{c, c_{4}H_{10}} = 425.12\) K) [277]; (b) \(n\)-hexane + water mixture at a temperature of \(T = 473.15\) K (VLE [273] and LLE) [274, 275]; (c) \(n\)-hexadecane + water mixture at a temperature of \(T = 523.15\) K (VLE [276] and LLE) [274]. The symbols represent the experimental data, the continuous curves the prediction with the SAFT-\(\gamma\) Mie approach, the horizontal lines the calculated three phase line, the inset image in (a) corresponds to a magnified view of the water-rich phase, and the inset images in (b) and (c) correspond to a magnified view of the VLE region.

3.8 Asymmetric association scheme between \(\text{CH}_{2}\text{OH}\) and \(\text{H}_{2}\text{O}\) groups

3.9 (a) Isobaric temperature–mole fraction \((T-x)\) phase diagram of the vapour-liquid-liquid equilibria of \(1\)-octanol + water at a pressure of \(p = 0.101\) MPa; (b) saturated densities of the water-rich (right) and octanol-rich phase (left). The circles [90] and triangles [142] represent the experimental data, the continuous curves the description with the SAFT-\(\gamma\) Mie approach, and the horizontal line the calculated three phase line. The inset image in figure (a) corresponds to a magnified view of the water-rich boundary.

3.10 Isobaric temperature–mole fraction \((T-x)\) phase diagrams of the vapour-liquid-liquid equilibria at \(p = 0.101\) MPa of (a) \(1\)-hexanol + water (triangles [279] and circles [280]) and (b) \(1\)-butanol + water (triangles [281] and circles [282]). The symbols represent the experimental data, the continuous curves the predictions with the SAFT-\(\gamma\) Mie approach and the horizontal lines the predicted three phase lines. The inset image in figure (a) corresponds to a magnified view of the water-rich boundary.

3.11 Henry’s law constants \(K_{\text{H}_{2} \text{O}}\) for hydrocarbons in water as a function of temperature at the corresponding vapour pressure of the mixture. The symbols represent the experimental data and the continuous curves the predictions with the SAFT-\(\gamma\) Mie approach: ethane (triangles [283] and diamonds [284]), \(n\)-butane (circles) [269], \(n\)-hexane (squares) [269] and \(n\)-octane (pentagons) [285].
3.12 Gibbs free energies of solvation $\Delta G_{i,j}^{\text{sol}}$ for (a) $n$-alkanes and (b) and 1-alcohols in water at $T = 298.15$ K and $p = 0.101$ MPa. The circles [286] represent the experimental data, the error bars the standard deviations, and the continuous curves the predictions with the SAFT-$\gamma$ Mie approach. 63

3.13 Infinite dilution activity coefficients $\gamma_{i,j}^\infty$ of $n$-alkanes (top line) and 1-alcohols (bottom line) in aqueous solution at $T = 298.15$ K and $p = 0.101$ MPa (except for ethane-butane, where the calculations are done at 5.00 MPa). The circles [288], triangles [270], diamonds [271] and squares [288] represent the experimental data, the crosses [196] the MD simulation data, and the continuous curves the predictions with the SAFT-$\gamma$ Mie approach. 64

3.14 Infinite dilution activity coefficients $\gamma_{i,j}^\infty$ of $n$-alkanes (hexane and pentane) [269] and 1-alcohols (octanol [289], heptanol–butanol [290]) in aqueous solution at $p = 0.101$ MPa (except for $n$-pentane and $n$-hexane, where the calculations were carried out at 7.00 MPa). The symbols represent the experimental data and the continuous curves the predictions with the SAFT-$\gamma$ Mie approach. 64

4.1 The SAFT-$\gamma$ Mie heterosegment molecular model of ibuprofen, where the constituent methyl (CH$_3$), methanetriyl group (CH), aromatic methine (aCH), aCCH$_2$, aCCH, and carboxylic acid (COOH) functional groups that characterise the molecule are highlighted. 67

4.2 Mutual solubilities of (a) isopentane (2-methylbutane) [263] and (b) isoctane (2,2,4-trimethylpentane) [264] in the water-rich liquid phase (continuous curves) and water in the alkane-rich phase liquid (dotted curves) at conditions of three-phase equilibria as a function of temperature. The squares are experimental data for the mutual solubilities of water in the alkane-rich phase and the circles are the mutual solubilities of alkane in the water-rich phase. The dotted and continuous curves represent the description with the SAFT-$\gamma$ Mie approach. 73
4.3 Mutual solubilities of water and alkyl benzenes [303] at conditions of three-phase equilibria as a function of temperature. The alkyl benzenes include (a) methylbenzene or toluene (squares), 1,2,3-trimethylbenzene (circles, not used in fitting), 1,2,4-trimethylbenzene (triangles), and 1,3,5-trimethylbenzene (diamonds, not used in fitting); (b) ethylbenzene (squares) and butylbenzene (circles); (c) (1-methylethyl)benzene or cumene (circles). The continuous curves represent the mutual solubilities of alkyl benzenes in the water-rich liquid phase and the dotted curves the mutual solubilities of water in the hydrocarbon-rich liquid phase.

4.4 Association schemes for COOH - COOH (dimerisation) and COOH - H₂O groups (asymmetric association).

4.5 Fluid-phase equilibria of (a) n-butanoic acid + water binary mixture (VLE) [304] at pressures of \( p = 0.013 \) MPa (circles) and \( p = 0.101 \) MPa (squares); (b) n-hexanoic acid + water binary mixture at ambient and elevated pressure, the symbols: crosses [305], circles [306], triangles [296], and squares [307] represent the experimental data; (c) n-octanoic acid + water binary mixture at conditions of three-phase existence (circles) [306] and pressures of 5.00 - 15.0 MPa (triangles) [308]. The red curves represent the VLE description in (a) and the description at conditions of three-phase equilibria in (b) and (c). The inserts in (b) and (c) magnify n-hexanoic acid and n-octanoic acid solubilities in the water-rich phase.

4.6 Isobaric temperature–mole fraction \((T - x)\) phase diagrams of the vapour-liquid equilibria of: (a) 1-hexanol + benzene (squares) [314] and 1-butanol + benzene (circles) [315] at pressure \( p = 0.101 \) MPa; (b) 1-butanol + 1,3-dimethylbenzene (triangles) [316], 1-butanol + 1,4-dimethylbenzene (squares) [316], and 1-butanol + methylbenzene (circles) [317] at \( p = 0.101 \) MPa; (c) 1-butanol + ethylbenzene (squares) [316] at \( p = 0.101 \) MPa and ethanol + ethylbenzene (circles) [318] at \( p = 0.096 \) MPa; (d) 1-butanol + (1-methylethyl)benzene at \( p = 0.101 \) MPa [319]. The symbols represent experimental data and the continuous and dotted curves the description with the SAFT-\( \gamma \) Mie approach (the dotted curves represent 1-butanol mixtures).

4.7 Isobaric temperature–mole fraction \((T - x)\) phase diagram of the vapour-liquid equilibria of n-butanoic acid + 1-butanol at pressures of \( p = 0.053 \) MPa (squares) and 0.027 MPa (circles). The symbols represent the experimental data [320], the continuous curves the description with the SAFT-\( \gamma \) Mie approach.
4.8 Mutual solubilities of water and hydrocarbons at conditions of three-phase equilibria as a function of temperature of: (a) water + isobutane (squares [321] and circles [269]) and water + 2,3-dimethylbutane (pentagons and triangles [263]); (b) water + 1,2-dimethylbenzene (crosses), water + 1,3-dimethylbenzene (squares), water + 1,4-dimethylbenzene (circles) [303]; (c) water + 1,3-diethylbenzene [303]; (d) water + (1-methylpropyl)benzene (squares) and water + 1,4-diisopropylbenzene (triangle) [303]. The continuous curves represent the mutual solubilities of hydrocarbons in the water-rich liquid phase and the dotted curves the mutual solubilities of water in the hydrocarbon-rich liquid phase.

4.9 a) Isobaric temperature–mole fraction \((T - x)\) phase diagram of the vapour-liquid equilibrium of propanoic acid + water at a pressure of \(p = 0.101\) MPa; symbols represent the experimental data [323]. b) Infinite dilution activity coefficients of \(n\)-carboxylic acids [198, 288, 324, 325] in aqueous solution at \(T = 298.15\) K and \(p = 0.101\) MPa, the red dotted curve the prediction based on the published parameters [240], and the blue dashed curves the prediction based on the current model.

4.10 Isobaric temperature–mole fraction \((T - x)\) phase diagram of the vapour-liquid equilibria of (a) 1-propanol + benzene (green triangles) [326] and ethanol + benzene (blue diamonds) [327] at pressure \(p = 0.101\) MPa; (b) 1-octanol + benzene [314] at \(p = 0.100\) MPa; (d) \(n\)-propanoic acid + ethanol at \(p = 0.101\) MPa [331]. (c) Isothermal–isobaric excess molar enthalpy of mixing of ethanol + benzene at \(T = 303.15\) K and \(p = 0.101\) MPa [330]. The symbols represent the experimental data and the dashed curves the prediction with the SAFT-\(\gamma\) Mie approach.

5.1 Partition coefficients of \(n\)-alkanes, \(1\)-alcohols, and \(n\)-alkanoic acids in (a) hexane-water, (b) hexadecane-water, and (c) octanol-water systems at \(T = 298.15\) K and \(p = 0.101\) MPa. The symbols [104, 143, 346, 347] represent the experimental data, the error bars the associated uncertainty [104], and the lines the predictions with the SAFT-\(\gamma\) Mie approach.
5.2 Comparison between experimental and predicted (a) \( K_{i,C_8W} \), (b) \( K_{i,C_{16}W} \), and (c) \( K_{i,OW} \) at \( T = 298.15 \) K and \( p = 0.101 \) MPa. The experimental data can be found in Refs. [104, 143, 346, 347]. The blue circles represent the \( n \)-alkanes, the purple circles the branched alkanes, the red squares the 1-alcohols, the orange squares the branched alcohols, the dark-green diamonds the \( n \)-alkanoic acids, the light-green diamonds the branched acids, the brown diamonds the dicarboxylic acids, the light-blue triangles the alkylbenzenes, and the yellow pentagons the phenylalkanoic acids (also see Chapter 6).

5.3 Solubility predictions (solid-liquid equilibrium) of (a) azelaic acid (structure shown in the figure) and stearic acid at ambient pressure [353] and (b) azelaic acid in water (triangles [355], diamonds [356] and circles [357]) at \( p = 0.101 \) MPa. The symbols represent experimental data, the continuous curves the solubility predictions, and the dashed curve the solidus line of the mixture predicted by the SAFT-\( \gamma \) Mie approach.

6.1 Intramolecular association via a \( \text{CH}/\pi \) hydrogen bond of (I) \( n \)-propylbenzene and an \( \text{OH}/\pi \) hydrogen bond of (II) phenylacetic acid.

6.2 The effective treatment for gauche conformations of (a) \( n \)-propylbenzene, (b) phenylacetic acid, (c) propanedioic acid, and (d) ethylene glycol within the SAFT-\( \gamma \) Mie framework.

6.3 The octanol-water partition coefficient (\( K_{i,OW} \)) of phenyl-\( n \)-alkanoic acids at \( T = 298.15 \) K and \( p = 0.101 \) MPa. The symbols represent experimental data [104, 351, 387, 388], the error bars the experimental uncertainty, the dashed curve the SAFT-\( \gamma \) Mie prediction based on the fully associated model, and the continuous curve the prediction after the IMHB assessment (when the appropriate model is used).

6.4 Octanol-water partition coefficients (\( K_{i,OW} \)) of dicarboxylic acids at \( T = 298.15 \) K and \( p = 0.101 \) MPa. The symbols represent experimental data [348–350], the dashed curve the SAFT-\( \gamma \) Mie prediction based on the fully associated model, and the continuous curve the prediction after the IMHB assessment (when the appropriate model is used).
6.5 Prediction of the solid-liquid equilibria in binary systems of APIs in different solvents as a function of temperature at $p = 0.101$ MPa. The symbols represent the experimental data and the continuous and dashed curves the prediction with the SAFT-γ Mie approach with and without effective intramolecular treatment, respectively: (a) phenylacetic acid in ethanol (circle [395] and purple curves) and in acetone (triangle [395] and orange curves); (b) ibuprofen (triangles [406] and green curves) and ketoprofen (squares [403] and blue curves) in 1-butanol; (c) ibuprofen (diamonds [397], star [398], plus [399], circle [400], squares [401], triangles [402], and green curves) and ketoprofen (cross [404], pentagons [405], and blue curves) in water; (d) ibuprofen (triangles [406] and green curves) and ketoprofen (circles [405] and blue curves) in acetone. Chemical structures for phenylacetic acid, ibuprofen, and ketoprofen are shown in figure (a), (b), and (c), respectively.

6.6 The SAFT-γ Mie group-contribution model for lovastatin highlighting the IMHB treatment and the plausible conformational equilibrium of the lactone ring showing the IMHB (right).

6.7 customise caption here for TOC

6.8 Solubilities of (a) lovastatin and (b) simvastatin in ethyl acetate at pressure $p = 0.101$ MPa as a function of temperature. The squares [417], circles [416], triangles [413], diamonds [418] represent experimental data, the continuous curves the SAFT-γ Mie prediction with effective IMHB treatment, and the dashed curves the prediction without the treatment.
List of Tables

2.1 Summary of thermodynamic approaches for the prediction of partition coefficients. 30

3.1 Like group parameters for use within the SAFT-\(\gamma\) Mie group-contribution approach: \(n_k^*\) is the number of segments constituting group \(k\), \(S_k\) the shape factor, \(\lambda_{kk}^\lambda\) the repulsive exponent, \(\lambda_{kk}^\sigma\) the attractive exponent, \(\sigma_{kk}\) the segment diameter of group \(k\), and \(\varepsilon_{kk}\) the dispersion energy of the Mie potential characterising the interaction of two \(k\) groups (\(k_B\) is the Boltzmann constant); \(N_{ST,k}\) represents the number of association site types on group \(k\), with \(n_{k,H}\) and \(n_{k,e1}\) denoting the number of association sites of type \(H\) and \(e_1\), respectively. 46

3.2 Group-group dispersion interaction energy \(\varepsilon_{kl}\), repulsive exponent \(\lambda_{kl}^\lambda\), and site-site association energy \(\varepsilon_{kl,ab}^{HB}\) and bonding volume \(K_{kl,ab}\) for use within the SAFT-\(\gamma\) Mie group-contribution approach. The unlike segment diameter \(\sigma_{kl}\) is also obtained from combining rules [22] and all unlike attractive exponents \(\lambda_{kl}^\sigma = 6.0000\). CR denotes that a combining rule is used [22]. 47

3.3 Percentage average absolute deviations (\%AAD) for the vapour pressures \(p_{vap,i}(T)\) and the saturated liquid densities \(\rho_{sat,i}(T)\) obtained with the SAFT-\(\gamma\) Mie group-contribution approach compared to experiment [256] (where \(N_{p_{vap,i}}\) and \(N_{\rho_{sat,i}}\) are the number of vapour pressure and saturated liquid density data points used for each of the 1-alcohols considered in the parameter estimation). 49

4.1 Groups developed for use within the SAFT-\(\gamma\) Mie approach 68
4.2 Like group parameters for use in the SAFT-$\gamma$ Mie group-contribution approach: $n_k^*$ is the number of segments constituting group $k$, $S_k$ the shape factor, $\lambda_{kk}^*$ the repulsive exponent, $\lambda_{kk}^a$ the attractive exponent, $\sigma_{kk}$ the segment diameter of group $k$, and $\varepsilon_{kk}$ the dispersion energy of the Mie potential characterising the interaction of two $k$ groups; $N_{ST,k}$ represents the number of association site types on group $k$, with $n_{k,H}$, $n_{k,e_1}$, and $n_{k,e_2}$ denoting the number of association sites of type H, $e_1$, and $e_2$ respectively. Unpublished† parameters can be found in Dufal and Papaioannou personal communications.

4.3 Absolute deviation (AD = $|R_{\text{exp}} - R_{\text{calc}}|$) and percentage relative deviation ($\%$RD = $\frac{|R_{\text{exp}} - R_{\text{calc}}|}{R_{\text{exp}}} \times 100$) for solubilities of branched alkanes ($i$) in water-rich (liquid) phase $x_i$ and water in alkane-rich (liquid) phase $y_{\text{water}}$ at three-phase equilibria at 298.15 K used for CH - H$_2$O and C - H$_2$O parameter estimation.

4.4 Group dispersion interaction energies $\varepsilon_{kl}$ and repulsive exponent $\lambda_{kl}^*$ for use in the SAFT-$\gamma$ Mie group-contribution approach. The unlike segment diameter $\sigma_{kl}$ is obtained from the arithmetic combining rule and all unlike attractive exponents $\lambda_{kl}^a = 6.0000$; these are not shown in the table. CR indicates that the unlike repulsive exponent $\lambda_{kl}^*$ is obtained from a combining rule [22]. *indicates parameters obtained in this work and †unpublished parameters (Dufal and Papaioannou personal communications).

4.5 Group association energies $\varepsilon_{kl,ab}^{\text{HB}}$ and bonding volume parameters $K_{kl,ab}$ for use within the SAFT-$\gamma$ Mie group-contribution approach.

5.1 Hexane-water partition coefficients $K_{i,C_6W}$ of alkanes, alcohols, alkanolic acids, and alkylbenzenes (at $T = 298.15$ K and $p = 0.101$ MPa). The deviation is defined as $|\log K_{i,C_6W}^{\text{exp}} - \log K_{i,C_6W}^{\text{calc}}|$. The experimental data are taken from Ref. [347].

5.2 Hexadecane-water partition coefficients $K_{i,C_{16}W}$ of alkanes, alcohols, alkanolic acids, and alkylbenzenes (at $T = 298.15$ K and $p = 0.101$ MPa). The deviation is defined as $|\log K_{i,C_{16}W}^{\text{exp}} - \log K_{i,C_{16}W}^{\text{calc}}|$. The experimental data are taken from Ref. [346, 347].
5.3 Octanol-water partition coefficients \( (K_{i,OW}) \) of alkanes, alcohols, alkanoic acids, alkylbenzenes, and phenylalkanoic acids (at \( T = 298.15 \) K and \( p = 0.101 \) MPa). The deviation is defined as \( |\log K_{i,OW}^{exp} - \log K_{i,OW}^{calc}| \). The experimental data are taken from Ref. [104] unless stated otherwise.

6.1 Active pharmaceutical ingredients mentioned in this chapter and their decomposition into SAFT-\( \gamma \) Mie groups.

6.2 Hexane-water partition coefficient \( K_{i,C_6W} \) of alkylbenzene (at \( T = 298.15 \) K and \( p = 0.101 \) MPa) with and without intramolecular association treatment. The deviation is defined as \( |\log K_{i,C_6W}^{exp} - \log K_{i,C_6W}^{calc}| \).

6.3 Hexadecane-water partition coefficient \( K_{i,C_{16}W} \) of alkylbenzenes (at \( T = 298.15 \) K and \( p = 0.101 \) MPa) with and without intramolecular association treatment. The deviation is defined as \( |\log K_{i,C_{16}W}^{exp} - \log K_{i,C_{16}W}^{calc}| \).

6.4 Octanol-water partition coefficients \( (K_{i,OW}) \) of dicarboxylic acids, alkylbenzenes, and phenylalkanoic acids (at \( T = 298.15 \) K and \( p = 0.101 \) MPa) with and without intramolecular association treatment. The deviation is defined as \( |\log K_{i,OW}^{exp} - \log K_{i,OW}^{calc}| \).

6.5 Average absolute error (AAE) for alkane-water (\( \log K_{i,C_6W} \) and \( \log K_{i,C_{16}W} \)) and octanol-water partition coefficients (\( \log K_{i,OW} \)) of alkylbenzenes, phenylalkanoic, and dicarboxylic acids with and without intramolecular association treatment.

6.6 Octanol-water partition coefficients \( (K_{i,OW}) \) of pharmaceutical-related molecules (at \( T = 298.15 \) K and \( p = 0.101 \) MPa). The deviation is defined as \( |\log K_{i,OW}^{exp} - \log K_{i,OW}^{calc}| \).

6.7 Literature data on lovastatin and simvastatin pure component and solubility in alcohols and alkyl acetates.

6.8 \( \log_{10} \) squared error between the reported experimental values [412, 413, 417], \( N_p \) being the total number of experimental points, and the SAFT-\( \gamma \) Mie predictions for the solubility of lovastatin and simvastatin in alcohols and alkyl acetates, with and without the effective treatment of the IMHB.
Chapter 1

Foreword

Drug discovery and development is a notoriously challenging process in which poor pharmacokinetics remains one of the main obstacles to higher success rates [1]. In 1997, Lipinski et al. [2] defined four simple physicochemical parameter ranges (the ‘rule-of-five’: molecular weight ≤ 500, octanol-water partition coefficient, log$K_{ow}$, ≤ 5, H-bond donors ≤ 5, H-bond acceptors ≤ 10) associated with acceptable bioavailability of 90% of orally active drugs that have achieved phase II clinical status. These ‘rule-of-five’ [2] can serve as a guideline to reduce attrition in forms of physicochemical properties. Computational approaches are an attractive option to obtain this information at the design stage; accordingly a considerable amount of effort has been devoted to developing predictive tools to find the ‘sweet spot’ in drug development [3]. The need for novel predictive tools has become a pressing issue over the years because the pharmaceutical industry is under growing pressure to develop innovative and cost-effective medicines, largely due to major losses of revenue owing to patent expirations and increasingly cost-constrained healthcare budgets. Lipitor, one of Pfizer’s best-selling drugs, is a characteristic example: its revenue fell from $17.19 billion to $14 billion primarily due to the loss of its exclusivity on November 2011 [4]. From this example, if the R&D process can be shortened, allowing the company to acquire a longer patent protection period even for just a day, an immense amount of profit can be saved. Without a dramatic increase in R&D productivity, today’s pharmaceutical industry will not be able to provide sufficient innovation to overcome the loss of revenue due to patent expiration. Thus, having to improve the R&D productivity has been posed as the industry’s grand challenge [5].

Moving down along the pipeline of the drug discovery and development, manufacturing efficiency in the pharmaceutical industry lags behind that of most other speciality chemical manufacturers [6]. A mechanistic understanding of its manufacturing processes is essential for
the design of modern production processes in lieu of traditional batch processes with laborious testing of samples. A viable option is to increase the systematic use of process systems engineering (PSE) concepts in the development, optimisation, and control of continuous pharmaceutical manufacturing processes and green pharmaceutical production processes, two major trends in pharmaceutical manufacturing [7, 8]. This leads to an opportunity in the development of predictive methods that can be used for both molecular and process modelling to facilitate the use of PSE methodologies, ultimately to foster drug development and manufacturing processes.

In parallel, thermodynamic tools are being continuously developed and improved in order to meet the need for accurate property prediction required commonly to many industrial sectors. Predictive approaches play a vital role in the design of processes and the accuracy of their output can significantly affect process-design decisions [9, 10]. In addition, advances in thermodynamic modelling have led to novel applications, such as the mixture/blend design [11, 12] and the integrated design of solvents and processes [13], where molecular characteristics of solvents are determined as part of the process optimisation. Despite the wide variety of thermodynamic methods that are available, industrial requirements in property prediction highlight that there is a great need for reliable predictive, rather than correlative, methods which are capable of covering a wide range of thermodynamic conditions and predicting properties such as phase equilibria, enthalpy, and heat capacity [14].

Among the methods for property prediction from molecular structure, group-contribution (GC) methods are the oldest and most widely used ones [15]. GC methods allow one to determine the properties of molecules in terms of an appropriate sum of contributions of functional groups that represent the unique molecular identity under the underlying assumption that the contributions of a given group are the same, regardless of where it appears. Each functional group is characterised by a set of group-specific parameters which are typically developed from experimental data of simple compounds and, once estimated, they are applied to the study of more complex molecules in a predictive manner. The GC concept has earned a great deal of success by having been implemented in various tools, ranging from empirical to more theoretically sound models [16]. Nevertheless the concept is, obviously, an assumption and some natural questions may arise such as: Can a GC method describe a broad range of states, properties, and phases within one methodology? and How accurate can a GC method be for a complex multifunctional molecule? The latter statement questions the very fundamental concept of a GC method whether the proposition of a group having the same characteristic parameters regardless
of the molecule still holds in highly functional molecules, which sometime even have intramolecular interaction. The following question is: Can a GC method be extended to account, even if effectively, for the presence of intramolecular interaction?

One of the most widely employed GC methods in industrial applications is the universal quasi-chemical functional group activity coefficient (UNIFAC) method [17]. The popularity of UNIFAC stems mainly from its extensive parameter table, which now covers more than 80 functional groups and 1,200 group-group interaction parameters [18], and also from its good performance in the description of fluid phase equilibria [19]. Despite its popularity, the UNIFAC method is subject to a number of limitations: the simultaneous description of different types of phase behaviour (vapour-liquid and liquid-liquid) with this method requires a specific set of parameters to describe each type of the phase behaviour. Moreover, UNIFAC has possibly exhausted its applicability to account for highly asymmetric systems, especially aqueous solutions of complex polyfunctional compounds as evidenced by the inaccurate prediction of octanol-water partition coefficient of these compounds [20]. In order to satisfy the industrial needs mentioned earlier [21], GC equations of state (EoS) [16] present themselves as a viable option given their predictive power and compatibility with molecular and process simulation. In this work, the recently developed SAFT-\(\gamma\) Mie group-contribution equation of state [22] is used to study thermodynamic properties of a variety of organic molecules. The method stems from the statistical associating fluid theory for potentials of variable range (SAFT-VR) [23], with the implementation of a Mie potential [24] and a heteronuclear model which allows a straightforward recasting as a group-contribution method [25, 26]. Several functional groups within the SAFT-\(\gamma\) Mie framework have been developed and used to predict physicochemical properties, i.e., partition coefficients and solubilities of active pharmaceutical ingredients (API), which are complex multifunctional molecules. These functional groups can be used in the formulation of an integrated product and process design method for future applications.

This thesis is structured as follows: in Chapter 2, a critical literature review on octanol-water partition coefficient prediction from thermodynamic approaches is presented. The main focus is placed on the applicability of the popular UNIFAC GC activity coefficient method and the opportunity for the recent formulation of GC EoS within the SAFT framework. In Chapter 3, the performance of the SAFT-\(\gamma\) Mie EoS for the description of fluid-phase behaviour of aqueous solution of alkanes and alcohols, as well as their solvation properties are examined. In Chapter 4, an extensive group parameter development for use within the SAFT-\(\gamma\) Mie framework is demonstrated. Subsequently in Chapter 5, these groups are used to create a model for the
partition coefficient and solubility prediction. Chapter 6, a quantitative prediction for partition coefficients and solubilities of complex organic molecules including several APIs is demonstrated. Thermodynamic properties of some of these molecules are governed by a competition between inter- and intramolecular hydrogen bonds (IMHB), which can be taken into account effectively by our novel approach. The IMHB have profound effects on the molecular structure and physical properties of the compound [27], however, they are often under-recognised and overlooked by many predictive tools. The significant impact of the IMHB and the necessity to incorporate such interaction in the predictive model used for drug design are highlighted in Chapter 6. The work presented in the thesis is summarised in Chapter 7, highlighting the contributions that have been made. Recommendations and directions for future work are briefly discussed.
Chapter 2

Challenges and Opportunities for the Thermodynamic Prediction of Octanol-Water Partition Coefficients

2.1 Introduction

The partition coefficient is a very useful property in a variety of product and process design problems. It commonly refers to the distribution of a solute, \( i \), in two mutually saturated liquid phases at equilibrium and usually of great difference in polarity (e.g., a hydrocarbon-rich and a water-rich phases), and is defined as the ratio of the molar concentration of the solute in the two phases in coexistence. As it is a property of a ternary system, the partition coefficient is a function of temperature, pressure, and global composition of the system. Nevertheless, as the composition of the solute goes to an infinitesimally small value at the limit of infinite dilution, the partition coefficient goes to a constant value, independent of the composition of the solute in the system. It is worth noting that the term partition coefficient is restricted to the equilibrium ratio of the consistent molecular species throughout all phases, as first pointed out by Nernst (Nernst Distribution Law) and explained in Ref. [28]. In the case of ionisable compounds, where speciation affects the molecular species present, such distribution ratios (which are now pH-dependent) are usually referred to as distribution coefficients or apparent partition coefficients.

Among all pairs of solvents that can be used to form the two liquid phases, the 1-octanol + water binary mixture is the most commonly used because it serves as an ideal surrogate in modelling biological activity [29]. The octanol-water partition coefficient of a solute \( i \), \( K_{i,\text{OW}} \) (or \( P_{i,\text{OW}} \)) is defined as the ratio of the molar concentration of solute \( i \) in the organic phase
over the molar concentration of solute \( i \) in the aqueous phase: \( K_{i,\text{OW}} = C_{i,\text{OR}} / C_{i,\text{WR}} \). Since the values of \( K_{i,\text{OW}} \) for different substances can vary over many orders of magnitude for solutes of different chemical nature, usually its logarithm of base 10 is reported, i.e., \( \log K_{i,\text{OW}} \).

For orally active drugs two crucial counter-balancing parameters, solubility and permeability, determine the fate of the drug after delivery. Solubility and dissolution are the first important properties to consider, given that a drug must be soluble in the aqueous contents of the gastrointestinal lumen to be orally absorbed [30]. Poor solubility leads to the drug being quickly excreted through faeces, while too high a solubility results in rapid urine excretion. Both lead to reduced bioavailability of the drug. Permeability is one of the key determinants in the pharmacokinetic (PK) properties (absorption, distribution, metabolism, excretion or ADME) of the drug. It is usually associated with a physicochemical property, lipophilicity, because measurement and prediction of physicochemical properties are relatively easy compared to those for PK properties, where biological factors come into play [30]. In general, high lipophilicity means that the drugs are likely to partition through the cell membranes, since they are essentially phospholipids. Too high a lipophilicity, however, could give rise to non specific binding with plasma proteins through van der Waals interactions. On the contrary, low lipophilicity makes it difficult for the molecules to cross biological membranes and cellular compartments. The exception is extracellular or cell-surface targets where lower lipophilicity is favoured to limit the amount of drugs that enters the cells.

In drug research, \( K_{i,\text{OW}} \) is considered as an important indicator for the determination of the lipophilicity of the compound, as first introduced by Hansch [31], primarily since the polarity difference between octanol and water is close to that between the aqueous medium and living cell membranes. Due to the underlying relation between lipophilicity and permeability, \( K_{i,\text{OW}} \) is widely used to estimate membrane penetration and permeability for the gastrointestinal track and the blood-brain barrier [30, 32]. In addition to its significance in quantitative structure-activity relationships (QSAR) of drug lipophilicity, \( K_{i,\text{OW}} \) has been used to study the ADME properties of drugs as well as their toxicity [29]. Many studies have documented the importance of \( K_{i,\text{OW}} \) in predicting ADME properties [1, 2, 33, 34]. In addition to a significant impact in pharmaceutical applications, \( K_{i,\text{OW}} \) plays an important role in toxicology [31]. For instance, values of \( K_{i,\text{OW}} \) have been found to correlate with the bioaccumulation phenomena in the aquatic food chain [35, 36]. It was found that the chemical uptake efficiency from water, excretion rate, and chemical assimilation efficiency are a function of the \( K_{i,\text{OW}} \). In particular, biomagnification (accumulation up the food chain) is likely to occur when \( \log K_{i,\text{OW}} \) of the contaminated
compounds is greater than five [37].

Given that the partition coefficient is a useful property in a range of applications, its reliable prediction is expected to have a great impact in such areas. Especially in the pharmaceutical industry, having information on the lipophilicity of a drug at the early discovery stage would enable medicinal chemists to design/modify novel compounds, their formulations, and delivery approaches, and would also enable biologists to better interpret screening results. The current $K_{i,OW}$ prediction methods are mainly empirical, based on regression to experimental values of $K_{i,OW}$ that have been measured via solubility measurements in two partially miscible liquids. The first widely accepted methodology for $K_{i,OW}$ calculation from chemical structures was developed by Fujita et al. in 1964 [38]. The authors considered $K_{i,OW}$ as an additive-constitutive, free-energy-related property which is numerically equal to the sum of the $K_{i,OW}$ of the “parent” structure, whose $K_{i,OW}$ was already known, plus a “π-constant” of a hydrogen-replacement atom. The method was, of course, limited to the compounds whose $K_{i,OW}$ of their parent structures was known. This led to the development of the first fragmental contribution method by Rekker et al. [39–42]. The Rekker method provided more general procedures that involved assigning the “fragmental constants” to different structural moieties. These studies subsequently led to many other substructure-based approaches, which include fragmental methods and atom-based methods [43–45].

All substructure-based methods are developed from a regression analysis of experimental $K_{i,OW}$ values of diverse organic compounds. The main difference is that in fragmental methods molecules are broken down into fragments and correction factors are applied in order to compensate for intramolecular interactions, while in atom-based methods molecules are broken into their component atoms (generally without correction rules) [46]. Examples of the fragmental methods include CLOGP [28, 47–49], $\Sigma f$ System [42], KLOGP [50], KOWWIN [51–53], ACD/LogP [54], and AB/LogP [55]; while the Ghose-Crippen Approach [56] and XLOGP [57, 58] are examples of the atom-based methods. Full details of the substructure-based methods can be found in Ref. [45]. The main drawbacks of these methods are, for instance, bias in fragmentation, simplification of steric effects (proximity effects), and lacking differentiation between structural isomers [44].

To overcome shortcomings of the substructure-based methods, property-based methods have been developed. In these methods, partition coefficients are calculated on the level of the solute
as a molecular/chemical species, rather than the contributions of its ubiquitous chemical moieties. Examples of property-based methods include conductor-like screening model (COSMO)-based approaches \cite{59}, *ab initio* methods \cite{60-62}, and models based on molecular dynamics (MD) simulations \cite{63-66}. These property-based methods, however, are usually time-consuming and their advantage over the simple approaches is not yet evident. The approaches to predict octanol-water partition coefficients \cite{44,48} and comparative studies of different models are greatly documented in extensive reviews \cite{43,46,67-70}.

Despite the number of $K_{i,OW}$ prediction methods available, the development of a reliable predictive model for $K_{i,OW}$ is still a challenge and highly desired for the pharmaceutical industry. The main drawback of existing substructure-based methods is the limited reliability in the prediction of $K_{i,OW}$ for compounds with chemical structures substantially different from those included in the regression of the models \cite{71}. As an example of the limited reliability of common methods, Mannhold *et al.* \cite{46} performed a comprehensive comparative study of 30 calculation methods on a public dataset of 266 molecules and 18 methods on two industrial datasets (882 compounds from Nycomed and 95,809 compounds from Pfizer). Most methods tended to do well in the public dataset but only a few produced an acceptable prediction in industrial datasets. In fact, a simple algebraic equation, proposed by the authors, solely based on number of carbon atoms and number of heteroatoms outperformed a large number of tools benchmarked in the study. The failure of many models to reliably predict $K_{i,OW}$ for the industrial dataset suggests that these models have been trained primarily for small molecules rather than drug-like compounds, especially novel classes of compounds which are the real pursuit of the pharmaceutical industry.

While current prediction tools are restricted to octanol-water partition coefficient, partition coefficients of other systems such as oil-water and liposome-water systems are also of interest to the pharmaceutical industry. To this end, general predictive approaches, which are not restricted to the prediction of a single property, could overcome this limitation. In addition, the development of more generic methodologies could broaden the investigation of drug molecules, *e.g.*, the study of temperature dependence and modelling of the drug partition coefficients which are also a research avenue of interest for scientists \cite{72}. Recent advances in thermodynamic models have made the modelling of complex molecules (including, but not limited to, pharmaceutical compounds) possible in a variety of solvents and over a wide range of thermodynamic conditions \cite{73-75}. This development allows for the prediction of various thermodynamic properties including partition coefficients in different systems and the phase behaviour of complex mixtures.
Such information is of critical importance for the downstream processing of drug molecules. Despite the relatively limited application of such methodologies in systems of complexity similar to what is typically encountered in the pharmaceutical industry in the literature, these seem to be promising approaches with great potential for future applications.

In this chapter, we provide an overview of the thermodynamic tools that are typically employed for the prediction of octanol-water partition coefficients. Direct experimental methods to determine $K_{i,OW}$ that capture partitioning of a solute in two mutually saturated liquids will be briefly mentioned in the next section. A general thermodynamic derivation of the partition coefficient, with a focus on how this is calculated by means of an equation of state, is presented next. The final section documents the thermodynamic tools ranging from activity coefficient models, e.g., the nonrandom two liquid segment activity coefficient model (NRTL-SAC) [76, 77] and the universal quasichemical functional group activity coefficients (UNIFAC) [17], to the more advanced frameworks for the modelling of continuous associating fluid based on perturbation theories, i.e., statistical associating fluid theory (SAFT) approaches [78, 79]. Models based on quantum mechanical (QM) calculations will also be presented. Finally, the conclusions of the review and a future perspective on the potential of the thermodynamic approaches for pharmaceutical applications are discussed.

### 2.2 Experimental determination of partition coefficients

There are several ways to determine partition coefficients experimentally. The international organisation for economic co-operation and development (OECD) has accredited three guidelines, namely the shake-flask method (OECD 107 [80]), the HPLC method (OECD 117 [81]), and the slow-stirring method (OECD 123 [82]), which can be classified into two different groups: direct and indirect methods. For the direct methods, simultaneous thermodynamic equilibrium of 1-octanol, water and the substance of interest should be attained before the composition of the phases is analysed. Therefore from the OECD guidelines, the shake-flask and the slow-stirring methods are considered to be direct methods. The shake flask is the most widely used method due to its classic extraction procedure. This traditional method is generally used for a log$K_{i,OW}$ range from -2 to 4 (occasionally up to 5) [80]. For highly lipophilic compounds (log$K_{i,OW} > 4$), the slow-stirring method is used as an alternative to the shake flask with reduced artifacts of micro-droplet formation in the aqueous phase. This method is applicable for compounds with log$K_{i,OW}$ up to 8.2 [82].
In indirect methods, the concentrations of the solute in two liquid phases are not determined explicitly. Instead, an empirical correlation with other partitioning phenomena, such as the compound's retention time in chromatography, is established from the values of reference substances for which the $K_{i,OW}$ is known. Indirect methods are therefore suitable for estimating $K_{i,OW}$ for structural homologues to the reference compounds, but become unreliable when applied to structurally unrelated compounds [83]. The HPLC method is considered to be an indirect method and typically used to measure log$K_{i,OW}$ in the range from 0 to 6 [81]. This method is suitable for high-throughput screening due to the minimal time and sample consumption. High-throughput screening is a favourable technique particularly in drug development work, thus several chromatographic techniques to determine partition coefficients have been extensively reviewed [84].

As no single method can precisely cover the whole range of partition coefficients, additional methods, e.g., counter-current chromatography (CCC), capillary electrophoresis, have been introduced as methods of choice for partition coefficient determination. Reviews and discussions about the features of each method can be found in Refs. [83, 85]. In light of the focus of this chapter on the thermodynamic prediction of $K_{i,OW}$, only the direct methods will be discussed since the experimental factors, i.e., phase equilibrium, solute concentration, and temperature, directly relate to the theory.

The direct methods: Shake-flask and Slow-stirring method

Shake-flask method serves as a fundamental method in $K_{i,OW}$ determination and yields extensive and useful partition coefficient data [28]. The method involves solubilisation of the compound in a mutually presaturated water and 1-octanol, agitation until equilibrium is reached, and direct measurement of the solute concentrations in both phases [29, 86]. The partition coefficient is then obtained as a quotient of these two concentrations, in direct correspondence with its definition. Although the method is seemingly simple, there are several factors that can affect the accuracy of the reported values of partition coefficient, e.g., temperature, mutual phase saturation, phase miscibility, solute concentration, and failure to reach equilibrium. The variation of $K_{i,OW}$ with temperature is small, approximately 0.01 log unit/K around room temperature, but temperature control is essential for highest accuracy [28]. By contrast, mutual phase saturation has a significant impact on the partition coefficient; $K_{i,OW}$ values determined from mutually saturated octanol-water and from pure solvents (no mutual solubilities) can differ by as much as 0.4 log unit, depending on the solute [87]. Dearden et al. [88] examined the factors
that can affect the accuracy of the partition coefficient, set forth the requirements for reliable measurement, and made recommendations for good practice.

Despite being the direct determination of partition coefficients, the shake-flask and slow-stirring methods also possess some shortcomings. For example, it can take several hours to days to reach equilibrium and large amounts of sample are required. The sample consumption in the shake-flask experiment is about 20 mg and can be up to 3000 mg for the slow-stirring method, while the HPLC method generally requires less than 1 mg of sample [83]. As mentioned earlier, \( K_{i,OW} \) is an infinite dilution property, thus, another challenge associated with its determination is the accuracy of the measurement of infinitely low solute concentration in both phases. Ideally, several experiments should be carried out with decreasing solute concentrations and \( K_{i,OW} \) extrapolated to a constant quantity to ensure that the “infinite dilution” condition of \( K_{i,OW} \) is reached.

To overcome the problem of having to conduct manually many time-consuming experiments, an automated high-throughput measurement technology for the partition coefficient has been developed and used in modern pharmaceutical industries. The automated shake-vial or the “modified shake-flask” is operated by transferring the traditional shake-flask method onto 96-well plate using a robotic liquid handler for sample preparation, and analysing the concentrations by reverse phase high performance liquid chromatography (RP-HPLC) [89].

2.3 Thermodynamic routes to partition coefficients

As mentioned in the previous section, the partition coefficient is measured when thermodynamic equilibrium is reached. At equilibrium, water and 1-octanol are partially miscible, forming an octanol-rich phase (OR) with mole fraction of octanol (oct) \( x_{\text{oct,OR}} = 0.726 \), and a water-rich phase (WR) with \( x_{\text{oct,WR}} = 5.60 \times 10^{-5} \), at ambient temperature and pressure [90].

The chemical potential of component \( i \) in the liquid mixture is expressed in terms of its activity coefficient as:

\[
\mu_i(T, p, x) = \mu_i^\Theta(T, p) + RT \ln x_i \gamma_i(T, p, x),
\]  

(2.1)

where \( \mu_i \) and \( \mu_i^\Theta \) are the chemical potentials of compound \( i \) in solution and of the pure substance \( i \), respectively. \( x \) is the set of mole fractions of all species, \( x_i \) is the mole fraction of solute \( i \), and \( \gamma_i \) is its activity coefficient, which accounts for the non-ideality of the solution. At equilibrium,
the chemical potentials of every species \(i\) in the two phases (octanol-rich and water-rich phases) are equal and as the reference chemical potentials cancel out the following relation is obtained:

\[
x_i,\text{WR} \gamma_i,\text{WR}(T, p, x_{\text{WR}}) = x_i,\text{OR} \gamma_i,\text{OR}(T, p, x_{\text{OR}}) .
\] (2.2)

The partition coefficient is defined as an equilibrium constant of the species of interest as:

\[
K_{i,\text{OW}}(T, p, z) = \frac{C_{i,\text{OR}}}{C_{i,\text{WR}}} = \frac{\nu_{\text{WR}} x_i,\text{OR}}{\nu_{\text{OR}} x_i,\text{WR}} ,
\] (2.3)

where \(z\), \(C\), \(\nu\) denote global composition, molar concentration, and molar volume, respectively. Here, we have generalised the thermodynamic derivation of partition coefficient to be functions of temperature, pressure, and compositions although, typically, \(K_{i,\text{OW}}\) should be independent of the solute concentration, i.e., in the special case of infinite dilution. Combining Eqs. (2.2) and (2.3), the partition coefficient can be written in terms of activity coefficients of the solute \(i\) in the two coexisting phases as:

\[
K_{i,\text{OW}}(T, p, z) = \frac{\nu_{\text{WR}} \gamma_i,\text{WR}(T, p, x_{\text{WR}})}{\nu_{\text{OR}} \gamma_i,\text{OR}(T, p, x_{\text{OR}})} .
\] (2.4)

Eq. (2.4) serves as a fundamental equation to obtain \(K_{i,\text{OW}}\) from activity coefficient models, e.g., UNIFAC [20, 91–94], NRTL [95], and COSMO-SAC (conductor-like screening model-segment activity coefficient) [96–99]. Note that \(\nu_{\text{WR}}\) and \(\nu_{\text{OR}}\) can be obtained experimentally and their ratio is constant at saturation for given \(T\) and \(p\) (the third component \(i\) is ignored due to its minimal presence). The activity coefficient models simply use the experimental ratio (e.g., 0.151 [20] or 0.148 [100]) as a coefficient for Eq. (2.4) since the models cannot provide volumes. This dependence on the experimental data presents a shortcoming of activity coefficient models, especially when other solvent systems come into play. Alternatively, partition coefficients can be calculated via an equation of state where both volumes and activity coefficients can be determined.

**Partition coefficients from an equation of state**

A number of equation of state models have been developed for predicting phase equilibria of fluid mixtures and these can therefore be used for partition coefficient calculation. These models are often expressed in terms of the Helmholtz free energy \((A)\), with the other thermodynamic
properties obtained through standard relations. In this section, we show how partition coefficients can be expressed in terms of the free energy and chemical potential following standard thermodynamic relations [101, 102].

The activity coefficient which accounts for the non-ideality of a solution is defined in terms of the activity \((a)\) as:

\[
\gamma_i(T, p, x) = \frac{a_i(T, p, x)}{x_i} = \frac{\hat{f}_i(T, p, x)}{x_i \hat{f}_i(T, p)} = \frac{\hat{\varphi}_i(T, p, x)}{\varphi_i(T, p)}, \tag{2.5}
\]

where \(\hat{f}_i\) and \(\hat{\varphi}_i\) are the fugacity and the fugacity coefficient of component \(i\) in solution (the \(^\ddagger\) sign denotes the solution phase), respectively. The fugacity coefficient is, in turn, related to the residual chemical potential \((\mu_i^{\text{res}})\) as:

\[
RT \ln \hat{\varphi}_i(T, p, x) = \mu_i^{\text{res}}(T, p, x). \tag{2.6}
\]

The residual chemical potential is the change in free energy referenced to the ideal gas mixture (IGM) state, i.e., \(\mu_i(T, p, x) - \mu_i^{\text{IGM}}(T, p, x)\). By combining Eqs. (2.4), (2.5), and (2.6), the partition coefficient is written as a function of chemical potential:

\[
\ln K_{i,\text{OW}}(T, p, z) = \ln \frac{\nu_{\text{WR}}}{\nu_{\text{OR}}} + \frac{1}{RT} \left[ \mu_i^{\text{res}}(T, p, x_{\text{WR}}) - \mu_i^{\text{res}}(T, p, x_{\text{OR}}) \right]. \tag{2.7}
\]

It is also common to write \(K_{i,\text{OW}}\) in terms of the Gibbs free energy of solvation \((\Delta G_i^{\text{sol}})\). The Gibbs free energy of solvation is the change in free energy in transferring one solute particle \(i\) from an ideal gas phase to a solution at infinite dilution at the same temperature \(T\) and pressure \(p\). By definition, the change in Gibbs free energy of solvation is equal to the residual chemical potential of the solute at infinite dilution [103], written as:

\[
\Delta G_i^{\text{sol}}(T, p, x) = \mu_i^{\infty}(T, p, x) - \mu_i^{\text{IGM}}(T, p, x) = \mu_i^{\text{res,\infty}}(T, p, x), \tag{2.8}
\]

where the superscript \(\infty\) denotes the condition at infinite dilution. As the partition coefficient is also a property at infinite dilution, Eq. (2.7) can be rewritten as:

\[
\ln K_{i,\text{OW}}(T, p, z) = \ln \frac{\nu_{\text{WR}}}{\nu_{\text{OR}}} + \frac{1}{RT} \left[ \Delta G_i^{\text{sol}}(T, p, x_{\text{WR}}) - \Delta G_i^{\text{sol}}(T, p, x_{\text{OR}}) \right], \tag{2.9}
\]

where \(\Delta G_i^{\text{sol}}(T, p, x_{\text{WR}})\) and \(\Delta G_i^{\text{sol}}(T, p, x_{\text{OR}})\) are the Gibbs free energies of solvation of solute \(i\) in the water-rich phase and octanol-rich phase, respectively at temperature \(T\) and pressure \(p\).
Often, this equation is written by omitting the volume fraction term [28, 104]. Nevertheless, it is crucial to emphasize that the quantity obtained in this way is a partition coefficient based on the mole fractions of solute \( i \) (i.e., \( x_{i,\text{OR}} / x_{i,\text{WR}} \)) denoted by \( K'_{i,\text{OW}} \) which differs from the ratio of molar concentrations, \( K_{i,\text{OW}} \). Its definition reads:

\[
RT \ln K'_{i,\text{OW}}(T, p, z) = \Delta G_{i}^{\text{sol}}(T, p, x_{\text{WR}}) - \Delta G_{i}^{\text{sol}}(T, p, x_{\text{OR}}) = \Delta u_{i} G_{i,\text{OW}}(T, p, z),
\]

(2.10)

where \( \Delta u_{i} G_{i,\text{OW}} \) is the energy of transfer of solute \( i \) from octanol-rich to water-rich phase. One needs to be attentive not to use this relation interchangeably with Eq. (2.9).

Since many equations of state [79, 105–107] express chemical potential as a function of \( T, V, \) and \( x \), the transformation of the chemical potential as a function of pressure, \( \mu^{\text{res}}(T, p, x) \), is done based on a standard thermodynamic relation using the compressibility factor [102]:

\[
\mu^{\text{res}}(T, p, x) = \mu^{\text{res}}(T, V, x) - RT \ln Z,
\]

(2.11)

which \( \mu^{\text{res}}_{i}(T, V, x) = \frac{\partial A^{\text{res}}}{\partial n_{i}} \bigg|_{T,V,n_{j\neq i}} \) is the residual chemical potential of component \( i \) at \( T \) and \( V \) obtained from an equation of state and \( Z = \frac{p v_{p}}{RT} \) is the compressibility factor, where \( v_{p} = \frac{V_{p}}{N} \) is the molar volume corresponding to the specified pressure. The partition coefficient is then written as:

\[
K_{i,\text{OW}}(T, p, z) = \frac{v_{\text{WR}}}{v_{\text{OR}}} \cdot \frac{\frac{1}{Z} \exp \left[ \frac{\mu^{\text{res,\infty}}_{i}(T, V, x_{\text{WR}})}{RT} \right]}{\frac{1}{Z} \exp \left[ \frac{\mu^{\text{res,\infty}}_{i}(T, V, x_{\text{OR}})}{RT} \right]}.
\]

(2.12)

The thermodynamic formulations of \( K_{i,\text{OW}} \) in terms of activity coefficients (Eq. (2.4)), Gibbs free energy of solvation (Eq. (2.9)), and chemical potential (Eq. (2.12)) are now written out. These expressions are used as the basis of calculation with the thermodynamic approaches, which are detailed in the next section.

### 2.4 Predictive thermodynamic tools for partition coefficients

The octanol-water partition coefficient is a thermodynamic quantity, yet its prediction via thermodynamic tools is still not a common practice. This is because modelling the complex structure of pharmaceutical molecules, which are composed of diverse functional groups and atoms, is challenging. Nevertheless, the fast-moving development of theoretical approaches leads to opportunities for thermodynamic predictions of drug-like molecules. In this section, the thermodynamic tools that have been employed to predict \( K_{i,\text{OW}} \) and models with the potential to
do so, which have been applied to phase equilibrium calculations and solubility studies, will now be discussed.

Typically, it is difficult to develop model parameters for thermodynamic modelling of pharmaceuticals because limited data are available. A number of studies lean toward a property with more data to correlate the model parameters. For instance, the prediction of solid solubility has been studied more extensively \([76, 77, 108-117]\) than the partition coefficient prediction due to the amount of the data available, despite the more complex thermodynamic relation for solid-liquid equilibrium (SLE) calculation. In SLE calculations, the experimental values of the heat of fusion, melting point temperature, and difference between the heat capacity of the supercooled melt and the heat capacity of the solid are usually used as inputs. For \(K_{\text{ow}}\) predictions, the application is rather early in drug discovery timeline, i.e., the prediction is done usually without any experimental data. Therefore, fully predictive models based on the chemical structure of APIs alone are desired. Group contribution (GC) approaches provide a promising route to the development of such methods. In GC approaches, the model parameters describing the groups can be obtained from data for any compound that contains such groups, hence overcoming the problem of limited experimental data of the APIs.

Group contribution approaches are based on the concept of solution-of-groups postulated by Langmuir in 1925 \([118]\), indicating that the force field around an atom group or radical which can interact with other molecules is considered as the characteristic of that group or radical and is largely independent of the nature of the rest of the molecule of which it forms part \([16]\). GC methods, therefore, perceive the properties of the systems of interest (pure components or mixtures) as the appropriate contributions of the chemically-distinct functional groups making up the molecules. For example, 1-octanol can be described as one \(\text{CH}_3\) group, six \(\text{CH}_2\) groups and one \(\text{CH}_2\text{OH}\) group. The groups are then characterised by a set of parameters, which can be obtained from any thermodynamic data of the compounds containing such groups, i.e., not restricted to 1-octanol data, under the assumption that the combination (e.g., addition) of the contribution of these groups lead to the thermodynamic properties of the molecules or mixtures of interest. GC approaches have proven numerous advantages especially in terms of correlating the properties of large number of chemical compounds into a much smaller set of parameters and allowing the property prediction to be done independently from the experimental data of the compounds or mixtures of interest. Despite all their appealing features, most group contribution methods have some common difficulties that limit their applicability. The definition of groups, for instance whether group \(\text{CH}_2\text{OH}\) or two groups \(\text{CH}_2\) and \(\text{OH}\) should be used, is
usually determined empirically and varies between GC methods. It is not trivial to determine the set of building models that gives the best trade-off between accuracy and breadth of predictions [119]. The major problems are the difficulty in distinguishing between isomers and the inability to account for proximity effects when two or more polar functional groups are in close proximity within the same molecule. We might argue that differentiation of more subgroups could alleviate the problem, but this will also deteriorate the benefit of the GC concepts. Nevertheless, the success of GC methodologies is evidenced by the inclusion of the scheme into many chemical engineering thermodynamic models such as analytical solution of group (ASOG) [120], many incarnations of UNIFAC [17]: KOW UNIFAC [93], modified UNIFAC [121], Pharma Mod. UNIFAC [112], functional-segment activity coefficient model (F-SAC) [122, 123], group contribution with association equation of state (GCA EoS) [106], and SAFT-based methods [22, 25, 124–126]. Readers are redirected to Refs. [16, 19, 127] for reviews of the prediction of thermodynamic properties by GC methodologies.

For all the above reasons, the thermodynamic tools that offer the possibility of predicting partition coefficients and are of interest for this study are mostly GC-based approaches. We also discuss briefly non-GC approaches (e.g., NRTL-SAC, the universal quasichemical (UNIQUAC), and cubic-plus-association equation of state (CPA EoS)) as they have been used to “predict” partition coefficients. The tools are categorised based on their nature of prediction. First, the activity coefficient models, i.e., NRTL-SAC and UNIFAC, are discussed. Next, the molecular-based equations of state for associating fluids, e.g., the SAFT-based approaches are reviewed. As \textit{ab initio} calculations are another important set of techniques for $K_{i, OW}$ prediction, in this chapter, studies based on quantum mechanical (QM) calculation, e.g., the COSMO-based approaches, as well as QM studies based on the GC framework are also reviewed. A summary of the approaches is reported in Table 2.1.

2.4.1 Activity coefficient models

As partition coefficient can be expressed in terms of activity coefficients as shown in Eq. (2.4), it is therefore convenient to use “activity coefficient models” to predict $K_{i, OW}$ as a class of thermodynamic models that have been developed to represent the thermodynamic behaviour of liquid mixtures only. It is assumed in such models that the fluids are incompressible and the effects of density or pressure are therefore not included so that the experimentally-measured volume ratio of octanol-rich phase and water-rich phase must be used in calculating $K_{i, OW}$. The volume ratio is approximated to be a constant since the amount of the third component
in octanol-water partitioning experiment is minimal so its volume can be neglected. Since the experimental conditions are assumed to be at ambient temperature and pressure, most studies use a constant number of 0.149 [109] or 0.151 [20] as the volume ratio. The most popular activity coefficient approaches applied for \( K_{i,ow} \) prediction seem to be UNIFAC-type models: UNIFAC [17], KOW UNIFAC [93], and modified UNIFAC (Dortmund) [121]. The NRTL-SAC model, a tool to predict SLE that has gained popularity for solvent screening in crystallisation process for pharmaceutical product isolation, and the UNIQUAC model have also been studied, but not as comprehensively as the UNIFAC model due to their inability to predict \( K_{i,ow} \) without some experimental data for the compounds of interest.

**Nonrandom two liquid segment activity coefficient model (NRTL-SAC)**

The nonrandom two liquid segment activity coefficient model (NRTL-SAC) was first introduced by Chen and Song et al. [76] in 2004 based on the polymer nonrandom two-liquid model [91], a derivative of the original NRTL model of Renon et al [128]. The model was designed for the purpose of fast, qualitative estimation of the solubilities of neutral organic compounds in common solvents to aid the solvent-selection procedure during pharmaceutical process design. In the NRTL-SAC model, the surface interaction of molecules are characterised into four types of conceptual segments: hydrophobic (\( X \)), hydrophilic (\( Z \)), repulsive (\( Y^- \)), and attractive polar (\( Y^+ \)). The model parameters describing solvents are obtained by regression to experimental vapor-liquid equilibrium (VLE) and liquid-liquid equilibrium (LLE) data for the solvents of interest and the reference solvents: hexane (for the hydrophobic segment), water (for the hydrophilic segment), and acetonitrile (for the polar segments). The parameters for each solvent are then stored in the model database. The conceptual segment values of a solute are determined by regression of experimental solubility data (SLE) in at least four solvents: one hydrophobic (i.e., n-alkane), one hydrophilic (i.e., water), one polar donor (i.e., ketone), one polar attractor (i.e., alcohol) [109]. Once the segment numbers of solute and solvents are known, NRTL-SAC can then be used to predict solubility in other solvents or solvent mixtures as well as \( K_{i,ow} \) of the solute. Readers are referred to Refs. [77, 108, 112, 116] for the application of NRTL-SAC to solubility prediction.

As experimental solubility data for the compound of interest are needed to calibrate the model, NRTL-SAC is in some ways a correlative approach for \( K_{i,ow} \) calculation. Most \( K_{i,ow} \) studies based on NRTL-SAC required experimental partition coefficient data (in octanol-water and other solvent systems at different temperatures and compositions) to demonstrate that the
model can reasonably represent LLE and partition coefficient data at different temperatures [129] or is capable of selecting a suitable solvent system for chromatographic separation [130]. This limits the applicability of NRTL-SAC for $K_{i,OW}$ prediction based on the molecular structures of lead compounds. Nevertheless, chemical compounds or APIs employed to study with this correlative approach are generally more complex, e.g., propranolol, atenolol [129], magnolol and honokiol [130], than those studied by the GC methods.

Universal quasichemical functional-group activity coefficients model (UNIFAC)

Since the first publication of UNIFAC by Fredenslund et al. [17] in 1975, considerable attention has been devoted to model improvement and group parameter development. UNIFAC is probably the most widespread activity coefficient GC method and is used across industry and academia to predict the non-ideal phase behaviour of multicomponent systems. As its name suggests, UNIFAC is an extension of the quasichemical theory of liquid mixtures, UNIQUAC [131], combined with the solution-of-groups concept of Wilson [132]. Both models are based primarily on Guggenheim’s quasichemical lattice model [131]. In the lattice theory, a liquid can be represented by a three-dimensional lattice which simplifies the statistical mechanics of the fluid. In UNIQUAC, parameters are a measure of the interactions between components instead of chemical groups, making the approach unsuitable for $K_{i,OW}$ prediction since experimental data of the component mixtures is needed for parameter fitting. In the case of the UNIFAC, parameters are used to characterise each functional group (and group-group interactions) which are then used to make up a molecule. The activity coefficient in UNIFAC is expressed as the sum of two contributions: a combinatorial and a residual part, given by:

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R.\quad (2.13)$$

The combinatorial part ($\gamma_i^C$) represents the contribution of the excess entropy and takes into account the differences in the size and shape of the molecule via relative van der Waals surface areas and volumes. The residual part ($\gamma_i^R$) takes into account the excess enthalpy from molecular interactions. Despite its success in modelling wide ranges of non-electrolyte mixtures, the original UNIFAC model possesses some weaknesses such as the inability to accurately model LLE and the phase behaviour of mixtures at high (>400 K) and low (<290 K) temperatures [133]. The major disadvantage of the original UNIFAC for the prediction of partition coefficients seems to be its poor performance in predicting infinite dilution activity coefficients ($\gamma_i^\infty$) [133] since these quantities provide a direct link to the partition coefficient (see Eq. (2.4)).
Despite the poor ability of the original UNIFAC to describe phase behaviour at the low solubility limit, the model has been employed to study $K_{i,OW}$ prediction. The original model, in which group-interaction parameters were obtained from binary vapor-liquid equilibria in the whole concentration range, was found to be quite inaccurate in predicting the property at infinite dilution region [91, 134]. Banerjee et al. [94] and Park et al. [135] used the original UNIFAC in correlation with the experimental $K_{i,OW}$ data and generated empirical equations relating $K_{i,OW}$ with the infinite dilution activity coefficients of the compounds in pure octanol and pure water. The constants in the empirical equations were intended to compensate for group nonadditivity and to improve the accuracy of the predictions. In spite of the availability of different parameter sets derived from VLE and LLE data, the original UNIFAC model was found to be useful for the estimation of partition coefficients only when used in conjugation with the experimental values of the aqueous-phase activity coefficients [136]. The revision of the set of UNIFAC group interaction parameters on the basis of the infinite dilution activity coefficients of organic compounds in aqueous solution by Hansen et al. [137] has enhanced the accuracy of the UNIFAC model especially for the systems containing highly hydrophobic compounds [91]. The deviations, however, increase rapidly for $\log K_{i,OW}$ values greater than six. This deviation indicates both the reliability of the UNIFAC model for highly hydrophobic compounds and may also be symptomatic of high levels of uncertainty for compounds with large $K_{i,OW}$ values.

Another notable development of the UNIFAC model for infinite dilution property prediction was the KOW UNIFAC model, which was targeted specifically for the estimation of octanol-water partition coefficients [93]. The KOW UNIFAC is distinguished from other UNIFAC models in the sense that the group interaction parameters were entirely estimated from infinite dilution data ($K_{i,OW}$ and the infinite dilution activity coefficients) and solubility data of various substances in water and in 1-octanol or in their equilibrium mixture.

In spite of tremendous attempts to refit the original UNIFAC parameters, it is the underlying theory in the model that needs to be modified to resolve the aforementioned limitations. Out of different modifications proposed by various authors, the most popular modification is to the modified UNIFAC (Dortmund) by Gmehling et al. [121]. In the modified UNIFAC, the combinatorial part of the original UNIFAC was changed for an empirical component that permits a better description of the influence of the shape and size of the molecules, and the residual part was modified with the more advanced temperature-dependent interaction parameters to describe phase behaviour over a wide range of temperature. In addition, the interaction parameter estimation for chemical groups in the modified UNIFAC was done by including the activity
coefficients at infinite dilution ($\gamma_i^{\infty}$) and excess enthalpies ($h^E$) in order to improve the accuracy of these properties, leading to a series of papers to revise and extend the parameter matrix [18, 138–140]. In parallel with the idea for the specific version of original UNIFAC for $K_{i,OW}$ prediction, the modified UNIFAC for solubility prediction of APIs was developed and referred to as Pharma Mod. UNIFAC [112]. In Pharma Mod. UNIFAC, a number of SLE data were also included to obtain the model’s parameters. Interactions between groups that occur in the API were neglected, so that the interaction parameter matrix can be reduced.

After various UNIFAC-type models were proposed, comprehensive analyses of several UNIFAC-type models have been carried out directly for the $K_{i,OW}$ prediction [20, 92, 93]. It was found that the quality of $K_{i,OW}$ predictions depends strongly on the parameters used in the model. The UNIFAC model performed well when using parameters that were fitted to the infinite dilution property data, i.e., partition coefficients and the activity coefficients at infinite dilution. In addition, it is suggested that additional modifications accounting for electrostatic interactions and proximity effects have to be considered for the prediction of $K_{i,OW}$ by UNIFAC models for biochemicals [92].

In this chapter, we show the comparison of octanol-water LLE calculations (Fig. 2.1) and partition coefficient predictions of n-alkanes and 1-alcohols (Fig. 2.2) using different versions of UNIFAC: original UNIFAC [141], KOW UNIFAC [93], modified UNIFAC [138], and Pharma Mod. UNIFAC [112]. It is noted that the Pharma Mod. UNIFAC calculation for octanol-water LLE is worse than all other UNIFAC-type models and out of scale for Fig. 2.1. As can be seen in the figure, KOW UNIFAC gives fairly good predictions of octanol solubilities in the water-rich phase since the solubilities and infinite dilution activity coefficients in the aqueous phase were used for the model regression. In the octanol-rich phase, the original UNIFAC model yields a better prediction on average. Although, the mutual solubilities of octanol + water mixture is not predicted accurately using modified UNIFAC, the method tends to reproduce the dependence of the solubility on temperature better than the original version. This ability is due to the introduction of more advanced temperature-dependent interaction parameters in the residual part of the modified UNIFAC to describe phase behaviour over a wide range temperature [138]. As expected for $K_{i,OW}$ predictions, the KOW UNIFAC provides significantly better results compared to the other versions since $K_{i,OW}$ data were included to obtain the model parameters (cf. Fig. 2.2). It is interesting to observe that the level of agreement to the experimental data given by UNIFAC and modified UNIFAC is different between $K_{i,OW}$ for alkanes and alcohols, despite the fact that these two systems are composed of the same functional groups: alkyl,
alcohol, and water. While the $K_{i,\text{OW}}$ predictions for alkanes, Fig. 2.2 (top), with UNIFAC and modified UNIFAC are quite inaccurate, their predictions for 1-alcohols, Fig. 2.2 (bottom), are significantly better. This inconsistency in the predictions might be due to the alcohol-water interaction parameters compensating for the error prevailed form the $K_{i,\text{OW}}$ of alkanes as these interactions dominate in the $K_{i,\text{OW}}$ of alcohols.

![Graph showing liquid-liquid equilibrium](image)

**Figure 2.1:** The model description of the liquid-liquid equilibrium of the binary mixture of 1-octanol + water in water-rich phase (WR) and octanol-rich phase (OR) at a pressure of 0.101 MPa. The open circles [90] and the open triangles [142] are the experimental data. The dashed-green, dash-dot purple, dotted blue curves represent, UNIFAC, KOW UNIFAC, and modified UNIFAC calculations, respectively.
In the literature, a comparative study of the UNIFAC models and the substructure-based (correlation) method, i.e., the atom/fragment contribution (AFC) method [53], which derives from a multiple linear regression of $K_{i,OW}$ data, was also presented. The correlation method was shown to be superior to all UNIFAC models studied [20]. The AFC correlation model, however, is limited to the octanol-water system. The authors of the study commented that the
GC concept in UNIFAC has possibly exhausted its applicability to account for highly asymmetric systems, especially aqueous solutions with complex polyfunctional chemicals [20]. Despite some degrees of applicability of the UNIFAC models to the $K_{i,OW}$ prediction, the models still suffer from a number of limitations that prevail for all type of $g^E$-models such as the inability to describe critical conditions and some thermodynamic properties such as densities, enthalpies, heat capacities, etc., which are also important for process design. This leaves a need for other predictive methods, such as SAFT approaches, as an alternative to the classical UNIFAC approaches.

2.4.2 Statistical Associating Fluid Theory (SAFT)

The Statistical Associating Fluid Theory (SAFT) [78, 79] is an equation of state that stems from the first-order thermodynamic perturbation theory (TPT1) of Wertheim in the 1980s [144-147]. SAFT was introduced by Chapman et al. [78, 79]; its many incarnations have constituted a major advancement towards a theoretical framework for modelling complex fluids. Recently, an investigation on the industrial requirement for thermodynamics and transport properties [14] suggested that the use of new methods, such as SAFT, is increasing in addition to the use of the traditional methods, e.g., cubic equation of state or the UNIFAC group contribution approach.

In the SAFT approaches, molecules are modelled as associating chains of segments (referred to also as monomers) interacting via attractive and repulsive potentials (e.g., square-well [148], Lennard-Jones [149], and Mie potentials [126]), with short ranged attractive sites on the segments to mediate association interactions. The SAFT equation of state is written in terms of the Helmholtz free energy ($A$) as the sum of four separate contributions:

$$A = A^{\text{ideal}} + A^{\text{mono}} + A^{\text{chain}} + A^{\text{assoc}},$$

(2.14)

where $A^{\text{ideal}}$ is the free energy of an ideal gas, $A^{\text{mono}}$ the contribution to the free energy due to the segments’ repulsion and dispersion interactions, $A^{\text{chain}}$ the free energy change due to the formation of chains of monomeric segments, and $A^{\text{assoc}}$ the contribution due to association.

The explicit consideration of different contributions in the SAFT equations allows the models to provide an accurate description of the fluid phase behaviour as well as thermodynamic properties of a broad range of compounds and mixtures over wide range of thermodynamic conditions [22]. More details about the theory, development, and applications of the SAFT approaches can be found in these comprehensive reviews [73, 127, 150, 151].
Originally, SAFT was designed as a molecular (homonuclear) approach with molecules constituted of identical segments [23, 78]. The group contribution formalism was introduced later to attain the predictive features of the approach [152, 153]. Here we discuss the relatively new versions of SAFT which are based on a group contribution framework with an interest in $K_{i,OW}$ prediction. It is noted that there are a number of group contribution equations of state, such as GC-SAFT [153, 154], GC-SAFT-VR [125], predictive Soave-Redlich-Kwong (PSRK EoS) [155], GC EoS [156], GCA EoS [106, 157–161], and the group contribution Peng-Robinson cubic plus association EoS (GC-PR-CPA) [162] which, in principle, can also be used to predict partition coefficients. It is worth noting that the equation of state which is generally used in oil and gas industry, the CPA EoS, has also been applied to the prediction of $K_{i,OW}$ but the approach is non-GC [163].

**Group contribution polar perturbed-chain SAFT (GC-PPC-SAFT)**

GC-PPC-SAFT [124] is an extension of the PC (perturbed-chain)-SAFT model proposed by Gross and Sadowski [105] with a group contribution formulation and a polar perturbation term. In PC-SAFT, molecules are considered as hard chains (its reference system) to which the dispersion interaction is added. This is in contrast to original versions of SAFT equations where the molecules are considered as hard spheres to which the dispersion interactions are added and are then grouped into chain molecules. PC-SAFT has been used to “predict” partition coefficient of two APIs (nicotinamide and salicylamide) in different solvent mixtures as functions of temperature and pH for extraction purposes [164]. The solid solubility data for the APIs in different solvents as well as binary LLE data of organic solvents and water were used to optimise the model parameters.

For an interest of the $K_{i,OW}$ prediction from chemical structures, the GC version of PC-SAFT is more applicable. In addition to the GC formalism, the main difference between the GC-PPC-SAFT and the original PC-SAFT is the additional multipolar term. Hence, the Helmholtz free energy in (GC-)PPC-SAFT reads:

$$ A = A^{\text{ideal}} + A^{\text{hc}} + A^{\text{pc}} + A^{\text{assoc.}} + A^{\text{multipolar}}, $$

(2.15)

where $A^{\text{ideal}}$ represents the free energy of an ideal gas, $A^{\text{hc}}$ corresponds to the free energy of a hard-chain fluid, $A^{\text{pc}}$ describes the contribution due to dispersion, $A^{\text{assoc.}}$ corresponds to the contribution due to association, and $A^{\text{multipolar}}$ describes the polar interactions of some segments.
Chapter 2: Challenges and opportunities for $K_{i,ow}$ prediction

The $A^{\text{multipolar}}$ term has been shown to improve the representation of systems containing polar compounds such as aromatics (quadrupolar) and water (dipolar) [165].

Generally, the group contribution formulation in SAFT can be classified into homonuclear and heteronuclear approaches based on molecular model. In homonuclear GC approaches, molecules are modelled as chains of identical segments. By contrast, molecules in heteronuclear approaches are considered as chains of different segment types, thus a realistic representation of a molecule is preserved [166]. Recently, the homonuclear GC-PPC-SAFT approach has been used to predict infinite dilution activity coefficients and partition coefficients of hydrocarbons and monofunctional oxygenated compounds [167, 168]. It was found that such properties were highly sensitive to the model parameters. Importantly, the binary interaction parameters were required to be fitted to the mutual solubilities of water and organic compounds to improve the prediction. This observation is not surprising because modelling the infinite dilution properties, especially for highly non-ideal systems such as hydrocarbon + water, is considered as a challenging task. Some models often have to use a specific set of parameters to describe the infinite dilution data [93, 158, 161].

SAFT-γ Mie

SAFT-γ Mie approach [22, 169] is the most recent formulation of the SAFT-γ equation of state [25, 26]. As a heteronuclear GC approach, molecules in SAFT-γ are subdivided into distinct functional groups chosen to represent the various chemical moieties of a molecule. In the SAFT-γ Mie approach, monomeric segments are modelled to interact via Mie (generalised Lennard-Jones) potentials of variable repulsive and attractive range. The Mie intermolecular potentials allow the attractive and repulsive exponents, which characterise the softness/hardness and the range of interactions, to vary freely therefore increasing the degree of freedom of the system. The use of Mie potentials to describe the group-group interaction has been shown to allow accurate simultaneous descriptions of the fluid-phase behaviour and second-order thermodynamic derivative properties [22]. The Helmholtz free energy of a fluid or a fluid mixture in SAFT-γ Mie is written as the sum of different contributions, cf. Eq. (2.14).

Recently, Dufal et al. [169] have reported the parameter table for various functional groups for use within the SAFT-γ Mie approach and shown accurate prediction of fluid-phase behaviour as well as other thermodynamic derivative properties of several families of compounds. The quantitative prediction of various thermodynamic properties sheds light on the possibility of using SAFT-γ Mie method for $K_{i,ow}$ prediction. As having group parameters are the heart of
group contribution approach, therefore, the important undertaking of this work is to expand the group parameter matrix. The parameter development within the SAFT-\(\gamma\) Mie approach is carried out with the aim to describe many types of phase behaviour, e.g., VLE, LLE and VLLE, over the entire range of composition with a single set of parameters for each chemical group. In the next chapters of this thesis, we will demonstrate extensively group parameter development based on functional groups reported in Ref. [169] for the accurate prediction of \(K_{i,OW}\) for a number of organic solutes including several APIs.

### 2.4.3 Models based on quantum mechanical (QM) calculations

Several models based on semiempirical quantum mechanical calculations were proposed for octanol-water partition coefficient calculation. Generally, correlations between \(K_{i,OW}\) and quantum mechanical variables, for example the atomic charges, dipole moments and molecular shapes are used for the prediction; readers are redirected to Refs. [46, 170] for further information. Here our interest is in thermodynamic models that are independent from regression to experimental \(K_{i,OW}\) data, therefore, theoretical models that allow the calculation to be done directly from Gibbs free energy will be discussed. The main factor that hinders the application of QM for \(K_{i,OW}\) prediction is that QM calculations for systems comprising of multiple molecules are generally difficult to perform in a reasonable computational time. Methods that require the explicit presence of solvent molecules in solvation processes are accurate but computationally expensive; however several continuum solvation models have been proposed in the literature to overcome this difficulty [171]. Despite the computationally expensive nature of QM calculations, the major advantage of these approaches over the typical GC methods seems to be that the 3D structure and proximity effects can be taken into account.

**Group contribution solvation model for octanol-water partition coefficient (GC-SKOW)**

GCSKOW is a group contribution solvation model (GCS) [172] for the \(K_{i,OW}\) calculation developed by Lin and Sandler [100]. The GCS model [172] was originally developed to determine the infinite dilution activity coefficient of a solute in a single solvent from the Gibbs free energy of solvation calculation. In the GCS model, the solvation process can be thought as two consecutive steps: the cavity formation step and the charging step. The Gibbs free energy of solvation
of solute \( i \) in solvent \( j \) can then be written as:

\[
\Delta G_{i,j}^{sol} = \Delta G_{i,j}^{cav} + \Delta G_{i,j}^{*\text{chg}}.
\]  
(2.16)

In the first step, the charges on the solute are turned off and the remaining hard particle is inserted into the solvent, creating a cavity of the necessary volume in the solvent. In this step, the free energy change is a result of the repulsive interactions between the solute \( i \) and the solvent \( j \) and is referred to as the cavity formation free energy (\( \Delta G_{i,j}^{cav} \)). Subsequently, charges are turned on thus the free energy change is a result of the attractive interactions and is called the charging free energy (\( \Delta G_{i,j}^{*\text{chg}} \)). In this context, the change of molecular motions upon solvation, i.e., the rotational and vibrational motions of the solute when it is transferred from an ideal gas to a solution has been neglected since the corresponding free energy change usually has a negligible effect on the total solvation energy.

Within the GCS model, the quantum mechanical calculation is used to determine the charging free energy (\( \Delta G_{i,j}^{*\text{chg}} \)) and the combinatorial part (\( \gamma_i^C \)) of the UNIQUAC model is employed to derive a realistic cavity formation free energy (\( \Delta G_{i,j}^{cav*} \)). The activity coefficient is then expressed as:

\[
\ln \gamma_i = \ln \gamma_i^C + \frac{\Delta G_{i,j}^{*\text{chg}} - \Delta G_{i,i}^{*\text{chg}}}{RT},
\]  
(2.17)

In the GCS model, the group scale factor \( \alpha \), which accounts for the dielectric behaviour of the solvent in vicinity of the solute, was obtained from fitting the computational chemistry calculations to experimental data for infinite dilution activity coefficients. This group scale factor \( \alpha \) is solvent specific, i.e., the scale factor for functional group \( \text{NH}_2 \) in octanol and in other alcohols are not the same. This means that the aforementioned procedure needs to be performed for every single solvent, with the functional group of interest. As the original GCS model [172] only applies to the infinite dilution activity coefficient of a solute in a single solvent, an empirical correlation between \( K_{i,OW} \) and the ratio of infinite dilution activity coefficients in pure water and pure 1-octanol must be used since the actual \( \gamma_i^\infty \) should be from the mutually saturated octanol-water (see Eq. (2.4)).

For the case of GCSKOW [100], the infinite dilution activity coefficient from the original GCS model was extended to multicomponent solvent mixtures. It is noted that, however, the infinite dilution activity coefficients in water-rich phase are approximated to that in the pure water phase since the octanol content in the water-rich phase is minimal. The extended model [100] was also developed to be less dependent on QM calculations by implementing the GC concept. This
was driven by the detailed analysis of the structure and energy parameters obtained from QM calculations which suggested that each functional group incurs a fixed amount of contribution to the total value of each molecular parameter. A group contribution method was then applied for both the molecular structure and the energy parameters based on the results of the previous QM calculations. The predictions from the GCSKOW model for both linear and nonlinear monofunctional compounds agreed well with the experimental data given that the model only uses three parameters: molecular volume, surface area, and charging free energy. The model, however, still possessed some limitations in case of difunctional compounds (compounds with two polar functional groups at two ends of the alkyl chain) due to the proximity effects [100].

Lin and Sandler later published an additional approach [173] to account for the structure and proximity effects which are not generally considered in group contribution approaches. The approach was referred as multipole corrections, in which molecular orbital calculations for isolated molecules are performed to obtain the net charge and dipole moment on each group within the molecule. The GCSKOW model with the multipole corrections have overcome the challenge associated with the proximity effects of multifunctional compounds, including a few APIs. The model provided good agreement with experimental $K_{i,OW}$ data, despite the dependence on QM calculations to determine the group charges and dipole moments for every isolated molecule. Lin and Sandler finally pointed out that there may not be a universal way to make such corrections for all physical and thermodynamic properties since different properties have different dependence on the multipole moment properties [173].

**Conductor-like screening model (COSMO)**

COSMO [174] is a quantum mechanical calculation method based on a molecular structural information and a small set of universal parameters. The method has been employed for a priori prediction of partition coefficients and other thermodynamic properties. The first attempt to predict $K_{i,OW}$ via COSMO-based approaches was the COSMO-RS (conductor-like screening model for real solvent) approach developed by Klamt [175]. In COSMO-RS, molecules are treated as interacting molecular surface segments dissolved in a perfect conductor where they are completely screened for charge density on the surface. Subsequently, the screening charge is removed to restore the real solvent state. The surface charge densities are then averaged over the typical contact segments. This probability function is referred to as the “σ-profile”, which has to be calculated only once for a given substance or system and can be stored in a database. With the σ-profile, the chemical potential of a solute in a real solvent can be calculated from
the chemical potential of surface segments via statistical thermodynamics. Chemical potentials then allow the calculation of many other related thermodynamic properties including partition coefficients (Eq. (2.7)). In the original study [175], the difference in the chemical potentials of solutes in pure octanol and water was regressed to the experimental $K_{i,OW}$ data to generate an empirical equation (also with a correlation factor for the structures) for $K_{i,OW}$ prediction. Later, a much larger set of partition coefficient predictions in different hydrocarbon-water and octanol-water systems by COSMO-RS was reported [176]. This highlights a benefit of the predictive thermodynamic approaches over the conventional empirical models: thermodynamic properties of more than one system can be obtained. Notably, the COSMO-RS method has also been studied to model the pH dependent partition coefficient (the distribution coefficient) of ionisable multifunctional pharmaceuticals containing one acidic group [177, 178].

A prominent implementation based on the COSMO-RS is the COSMO-SAC approach proposed by Lin and Sandler [96]. The starting point of the COSMO-SAC (COSMO-segment activity coefficient) model was the previously developed group contribution solvation (GCS) model [172]. In the GCS model, the Gibbs free energy of solvation was obtained from the traditional continuum solvation models which require the dielectric properties of solvents that are generally not available especially for mixed solvents. These properties are no longer necessary when using ideal solvation from the COSMO-RS model. Lin and Sandler related the solvation process in the GCS model to the charge screening process in COSMO-RS. In particular, the charging free energy ($\Delta G^{*\text{chS}}$) in the GCS model can be replaced by the free energy required to remove the screening charges on the solute in the restoring process ($\Delta G^{*\text{restore}}$) in COSMO-RS. Based on Eq. (2.17) of the GCS model, the activity coefficient in COSMO-SAC is expressed as:

$$\ln \gamma_i = \ln \gamma^C_i + \frac{\Delta G^{*\text{restore}}_{i/s} - \Delta G^{*\text{restore}}_{i/i}}{RT}.$$  \hspace{1cm} (2.18)

The chemical potentials of segments calculated by the COSMO-SAC model are then used to calculate the activity coefficient at infinite dilution, which in turn, used to determine partition coefficients [96]. The COSMO-SAC model was later used to predict $K_{i,OW}$ of larger molecules including pharmaceuticals [98], biofuel-related compounds [179], and nanoscale building blocks [99].

A further implementation of COSMO-SAC was a combination of the approach to the Peng-Robinson equation of state (PR EoS), denoted as PR+COSMOSAC approach [97]. The utilisation of an equation of state allows the interrelationship between temperature, pressure, volume,
and compositions of fluid mixtures. In addition, EoS models can be used in all types of fluid phase equilibrium [180] and various thermophysical properties. In the PR+COSMOSAC approach, two parameters in PR EoS are determined by a COSMO-SAC solvation calculation instead of using an experimental input. In brief, the molecular volume parameter, $b(x)$, is approximated by the solvation cavity volume and the energetic parameter, $a(T, x)$, is determined by the charging component of the Gibbs free energy of solvation. It was found that the liquid-liquid equilibrium of 1-octanol and water at ambient pressure, from 280 to 390 K, determined from the PR+COSMOSAC approach had a better agreement with the experiments over the whole temperature range in comparison to the calculations from the COSMO-SAC model. The octanol-water partition coefficients and infinite dilution activity coefficients were also calculated to illustrate the applicability of the model [97].

In spite of the appealing applicability of the COSMO-based approaches, comprehensive examinations of performance of both COSMO-RS and COSMO-SAC (on solubility [109] and vapor-liquid equilibrium [181, 182]) suggest that the models still have weaknesses which are mainly caused by an inadequate description of dispersive forces and hydrogen bonding [183]. A comprehensive study [182] of the performance of the COSMO-RS model in comparison to a classical group contribution method, modified UNIFAC, indicated that poor results for the COSMO-RS model were obtained for aqueous systems. Moreover, the study stated that it could not be confirmed that the weaknesses of group contribution methods (i.e., isomer and proximity effects) can be addressed by using the COSMO-RS approach. The study, however, recommended the application of the approach when the required group interaction parameters of the group contribution method are missing.

<table>
<thead>
<tr>
<th>Table 2.1: Summary of thermodynamic approaches for the prediction of partition coefficients.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>model name</strong></td>
</tr>
<tr>
<td>NRTL-SAC</td>
</tr>
<tr>
<td>UNIQUAC</td>
</tr>
<tr>
<td>UNIFAC</td>
</tr>
<tr>
<td>PC-SAFT</td>
</tr>
<tr>
<td>GC-PPC-SAFT</td>
</tr>
<tr>
<td>SAFT-$\gamma$ Mie</td>
</tr>
<tr>
<td>GUSKOW</td>
</tr>
<tr>
<td>COSMO</td>
</tr>
</tbody>
</table>
2.5 Conclusions

A reliable prediction of octanol-water partition coefficients is highly desirable in early stages of drug discovery and development since it provides a meaningful insight into the lipophilicity as well as the ADME properties of the molecule. Current prediction models are mostly empirical and fail to provide good prediction for compounds which their chemical families were not included in the regression of the models. Moreover, they are usually restricted to the prediction of only a single property. On the other end, theoretical approaches are becoming more capable of modelling complex fluids and multifunctional molecules. These approaches are able to predict a broad range of physicochemical properties, including partition coefficients, distribution coefficients, and solubilities in various systems. In addition, these models can be incorporated within process modelling tools and used for pharmaceutical process design and optimisation, for example for solvent screening, which is also a challenging and resource-intensive task.

Some thermodynamic tools, for example NRTL-SAC and UNIQUAC, are of general use for the pharmaceutical industry for the application like solvent screening but are not ideally applicable for partition coefficient prediction due to their dependence on experimental data of the compounds of interest. The methods that are of greatest interest are mainly the QM-based and GC-based methods. The QM-based methods such as GCSKOW, COSMO-RS, and COSMO-SAC calculate molecular properties based on molecular orbital theory. The calculation can be computationally expensive but the structure and proximity effects of the molecules can be considered. In GC approaches, e.g., UNIFAC, GC-PPC-SAFT, and SAFT-γ, molecules are decomposed into different functional groups with a unique set of parameters describing each group. The parameters are obtained from regression to experimental data of any compounds pertaining to the groups. The property prediction with GC approaches are usually very fast once the group parameters are developed but most approaches suffer from an insufficient number of chemical groups to describe drug molecules. The general shortcomings of the GC approaches are difficulty in distinguishing between isomers and inefficiency in accounting for proximity effect and intramolecular association.

Up to this point, we can say that the prediction of partition coefficients is still a challenging task regardless of the number of years that they have been introduced. Great opportunities are still open for advanced computational tools to play a part in property prediction and pharmaceutical process modelling. Given an extensive undergoing research for the versatile and predictive thermodynamic approaches especially on their abilities to model complex molecules and systems, these theoretically advanced approaches will be increasingly implemented by many
industrial sectors, including pharmaceuticals, and will facilitate the drug development process to a great extent. As examples, in the next chapters of this thesis, we will demonstrate the use of SAFT-γ Mie approach for the prediction of thermodynamic properties that are of interest for pharmaceutical application.
Chapter 3

Modelling the solvation of alkanes and alcohols in aqueous environment

3.1 Introduction

The complex intermolecular interactions between non-polar moieties and water, and in particular the hydrophobic effect [186–188], play a central role in determining the structure and properties of biomolecules, their compartmentalisation and subsequent organisation in the living cell [189]. The impact of the hydrophobic effect can be seen in the macroscopic properties of simple mixtures of water and alkanes which exhibit extreme non-ideal phase behaviour, with very limited miscibility over broad ranges of thermodynamic conditions [190]. The mutual solubilities in the two coexisting phases (water-rich and alkane-rich) are highly asymmetric; the solubility of an alkane in the water-rich phase is several orders of magnitude lower than the solubility of water in the alkane-rich phase. Interestingly, while the solubility of the alkanes in the water-rich phase at conditions of three-phase coexistence presents a minimum at a temperature of around 303 K [191] (a phenomenon believed to be related to the hydrophobic effect), the water solubility in the alkane-rich phase increases monotonically with temperature [192]. In mixtures including more functionalised molecules, such as alcohols and larger organic molecules, a delicate balance between the hydrophilic and the hydrophobic interactions of these molecules with water determines unexpected changes in phase behaviour, self-organisation and segregation that lead to membrane and micelle formation, protein folding and ligand-protein binding. This balance is also relevant in industrial applications, in refineries, petrochemical, pharmaceutical and biotechnological processes given the ubiquitous presence of water.
In the quest to understand and control the macroscopic behaviour of complex aqueous mixtures, the development of theoretical approaches that, grounded in statistical mechanics, capture the balance of the different intermolecular interactions and deliver bulk properties predictively and accurately remains a key challenge. This can be approached by starting with simpler systems, such as mixtures of water and alkanes, in which the hydrophobic effect plays a central role, and mixtures of water and alcohols where the impact of hydrophilicity can be felt, then progressing to multifunctional compounds. Phase behaviour and solvation properties (e.g., solubility, partition coefficient, free energy of solvation) offer a way to determine and probe the effectiveness of proposed models and theories and provide a direct link to applications such as drug development [28, 193, 194]. A number of approaches can be adopted, such as molecular simulation methods [195–197], activity coefficient models [198–201], the conductor-like screening model for real solvent (COSMO-RS) [202] and equations of state [203, 204]. From a theoretical perspective equations of state (EoSs) are a particularly useful tool that can access a breadth of states and properties in a computationally efficient way. Classical cubic equations of state [205, 206] can be used to describe the phase behaviour of aqueous systems. Though a degree of success has been achieved with appropriate mixing rules [207–209], these methods either yield limited accuracy or cannot be applied to multicomponent mixtures. Their limited success has been ascribed to their inability to take into account association and solvation effects explicitly. A breakthrough in modelling of aqueous systems was the development of equations of state that explicitly take hydrogen bond formation into account, including widely employed methods such as the statistical associating fluid theory (SAFT) [78, 79] and the cubic plus association (CPA) [107] EoSs; both stem from the first-order thermodynamic perturbation theory for associating fluids (TPT1) of Wertheim [144–147]. The CPA model has been used to study many types of phase equilibria of alkane + water mixtures [210–213] and comparative studies with the SAFT approach for the modelling of aqueous systems have been documented [127, 192, 210, 214]. Both SAFT and CPA have been used to describe low temperature hydrocarbon solubilities and can capture the solubility minima of alkanes in the aqueous phase with models specifically designed for this purpose [149, 213]. The SAFT and CPA EOs have also been employed to investigate the phase behaviour of alcohol + water mixtures [211, 214–217]. A recent review on the application of molecular-based equations of state for water and aqueous solutions can be found in Ref. [204].

A more challenging task involves the development of group-contribution (GC) equations of state [16] to capture the complex behaviour and subtleties of aqueous systems. GC approaches
are based on the premise that the properties of a molecule can be determined from the appropriate contributions of the chemically-distinct functional groups that the compound comprises. A major advantage of GC approaches lies in the possibility of describing the properties of a large number of chemical compounds with a small number of group parameters and in allowing property prediction to be carried out without the need for experimental data of the target compounds or mixtures. The GC concept was applied early on to obtain activity coefficients of a broad range of solutes and solvents in semi-empirical methods based on the number of carbon atoms in the solute [218]. Subsequent works, such as the analytical solution of group (ASOG) [199, 200] and the universal functional activity coefficient (UNIFAC) models [198–200, 219], continued delivering activity coefficient approaches with more theoretically sound models. Mengarelli et al. [220] extended the traditional UNIFAC model by incorporating an association term, based on Wertheim's theory for fluids [144–147] and presented a good description of the VLE of ethanol + water mixture. Recently, Soares et al. [122] have proposed a functional-segment activity coefficient (F-SAC) model that is based on the concept of functional groups as in UNIFAC, but with the interaction energy between groups derived from the COSMO-RS theory. The model was used to model several mixtures of hydrocarbons and water, and was shown to provide an improvement over UNIFAC-type models in the prediction of infinite dilution activity coefficient ($\gamma^\infty$) and LLE of most systems considered [123, 221]. Interestingly, the F-SAC model with an additional parameter for the computation of the association energy is able to capture the minimum solubility points of hydrocarbons in the water-rich phase [221].

Our focus, however, is devoted to the use of GC EoSs to model aqueous systems. A further advantage of EoSs such as SAFT formulated with the GC framework is that they can be used to develop force-field parameters that can be used in molecular simulations [222]. Given these benefits, there is a growing body of work applying GC EoSs to aqueous systems. The GC-CPA [162] and group contribution plus association (GCA) [223] EoSs have been used to describe the mutual solubilities of hydrocarbons and water [160, 162] as well as the phase behaviour of several alcohol + water mixtures [158, 161]. A group contribution form of a SAFT-type approach including multipolar interactions has been used by Nguyen-Huynh et al. [165] to model aqueous mixtures. This approach was found to provide a good overall description of the LLE and VLE of alkane + water and alcohol + water mixtures, although large deviations from the experimental data were found in the calculated aqueous solubilities of both alkanes and alcohols at ambient temperature. The GC-polar perturbed chain (PPC)-SAFT EoS has also been applied to the prediction of infinite dilution properties, e.g., Henry's law constants and octanol-water partition
coefficients of hydrocarbons and oxygenated compounds [167, 168]. Especially relevant to the current work is the recasting of the SAFT-variable range (VR) equation of state [23] as a group-contribution method in the SAFT-$\gamma$ EoS [25, 26]. The fluid-phase behaviour of aqueous solutions of alkanes and alcohols over wide ranges of thermodynamic conditions has been studied in this framework based on the use of square-well (SW) potentials [148]. The method was found to provide a simultaneous prediction of both vapour-liquid and liquid-liquid equilibria using a unique set of transferable group interaction parameters, although the accuracy was seen to deteriorate in the prediction of the alkane solubility in the water-rich phase. Recently, Chremos et al. [224] have revised the SAFT-$\gamma$ SW group parameters required to model these mixtures leading to a better description of the properties of the water-rich phase.

Despite this progress, there is, as yet, no general model that can access the infinitely dilute regime as well as the properties of concentrated solutions. In attempting to develop such models in the framework of an EoS, the need to incorporate explicitly dipolar interactions to treat water, and other polar molecules, has been the focus of debate. The original SAFT [192, 210, 225, 226], hard-sphere SAFT [227], SAFT-variable range (VR) [228], soft-SAFT [149] and perturbed-chain SAFT (PC-SAFT) [229-231] have been successful in the study of the phase behaviour of alkane + water mixtures, especially in the correlation of low-pressure phase equilibria (VLE), the prediction of critical conditions [103, 227] and the description of water solubility in the hydrocarbon-rich phase at high pressure (LLE) [192, 226]. The models are however noticeably less accurate for the calculation of hydrocarbon solubilities in the water-rich phase, which can be extremely small. These larger errors have been attributed to the use of inappropriate water models and to the inadequacies of some SAFT models in accounting for the high polarity of the water molecule [149, 232, 233]. The argument, however, remains controversial as a number of studies have demonstrated that SAFT models that do not incorporate an explicit dipole (e.g. Liang et al. [214] and Chremos et al. [224]) can perform well in the description of the $n$-alkane solubility in water; it is worth noting that this is often at the expense of accuracy in the calculated concentration of water in the alkane-rich phase. Al-Saif et al. [234] and de Villiers et al. [235] have used PC-SAFT with the incorporation of dipole interactions, while Folas et al. [203], Kontogeorgis et al. [236] and Liang et al. [214] have used the original PC-SAFT and CPA without the explicit consideration for such interactions to model the VLE and VLLE of alcohol + water mixtures. Interestingly, the results for the phase equilibrium calculations are comparable for both approaches, with and without the consideration of the polar interaction in the theories.
In this chapter, we develop novel models for the accurate prediction of vapour-liquid, liquid-liquid and vapour-liquid-liquid equilibria (VLE, LLE, and VLLE) of aqueous solutions of alkanes and alcohols within the SAFT-\(\gamma\) Mie group contribution approach [22]. We show that the SAFT-\(\gamma\) Mie EoS is able to deliver an accurate description of the phase equilibria over a broad range of compositions of the mixtures, including the infinite dilution regime. Finally, our work provides new insights into the modelling of polarity and hydrophobicity in aqueous systems.

3.2 Theory and molecular models

3.2.1 The SAFT-\(\gamma\) Mie group-contribution approach

In the SAFT-\(\gamma\) Mie approach [22, 169] molecules are represented as heteromolecular chains of fused spherical segments which correspond to the various chemical moieties of a molecule and which have short-range associating sites if necessary to mediate directional interactions. The Mie intermolecular potential [24] is implemented together with a high-temperature expansion to third order [237] to provide a high level of accuracy in the description of fluid-phase equilibria and thermodynamic derivative properties. An example of a molecule represented in the SAFT-\(\gamma\) Mie treatment is given in Fig. 3.1, where the CH\(_3\), CH\(_2\), and CH\(_2\)OH functional groups that characterise 1-butanol are shown. A given group \(k\) is formed by a number \(n_k^s\) of spherical segments and a shape factor \(S_k\) is used to characterise the contribution of each segment to the overall free energy of the molecule. Two segments \(k\) and \(l\) are modelled as interacting via Mie [24] (generalised Lennard-Jones) potentials of variable repulsive and attractive range:

\[
\Phi_{kl}^{Mie}(r_{kl}) = C_{kl}\varepsilon_{kl} \left[ \left( \frac{\sigma_{kl}}{r_{kl}} \right)^{\lambda_{kl}^r} - \left( \frac{\sigma_{kl}}{r_{kl}} \right)^{\lambda_{kl}^a} \right],
\]

(3.1)

where \(r_{kl}\) is the distance between the centres of the segments, \(\sigma_{kl}\) the segment diameter, \(\varepsilon_{kl}\) the depth of the potential well (the dispersion energy), and \(\lambda_{kl}^r\) and \(\lambda_{kl}^a\) the repulsive and attractive exponents of the segment-segment interactions, respectively. It is noted that the \(\lambda_{kl}^a\) is kept equal to six. The prefactor \(C_{kl}\) is a function of the exponents and ensures the minimum interaction to be \(-\varepsilon_{kl}\):

\[
C_{kl} = \frac{\lambda_{kl}^r}{\lambda_{kl}^r - \lambda_{kl}^a} \left( \frac{\lambda_{kl}^r}{\lambda_{kl}^r - \lambda_{kl}^a} \right)^{\lambda_{kl}^r - \lambda_{kl}^a}.
\]

(3.2)

In common with other SAFT approaches, hydrogen bonding and strongly polar interactions can be treated through the incorporation of short-range square-well association sites, which are placed on any given segment. A segment may have a number \(N_{ST,k}\) of different site types, with
Figure 3.1: Example of the decomposition of a molecule into functional groups: 1-butanol is composed of one CH$_3$ group (shaded red), two CH$_2$ groups (shaded gray) and one CH$_2$OH group composed of two fused spherical segments (shaded green, with association sites in brown and yellow to mimic hydrogen-bonding interactions).

$n_{k,a}$ sites of type $a = 1, ..., N_{ST,k}$. The association interaction between two square-well sites, one of type $a$ placed on segment $k$, and a second of type $b$ on segment $l$ is given by:

$$
\Phi_{kl,ab}^{HB}(r_{kl,ab}) = \begin{cases} 
-\varepsilon_{kl,ab}^{HB} & \text{if } r_{kl,ab} \leq r_{kl,ab}^c \\
0 & \text{if } r_{kl,ab} > r_{kl,ab}^c
\end{cases}, \quad (3.3)
$$

where $r_{kl,ab}$ is the centre-centre distance between sites $a$ and $b$, $-\varepsilon_{kl,ab}^{HB}$ is the association energy, and $r_{kl,ab}^c$ the cut-off range of the association interaction. The cut-off range $r_{kl,ab}^c$ can be equivalently described as a bonding volume $K_{kl,ab}$ [238]. Each site is positioned at a distance $r_{kk,aa}^{d}$ or $r_{ll,bb}^{d}$ from the centre of the segment on which it is placed.

Once the relevant parameters are determined the total Helmholtz free energy $A$ of a mixture of associating heteronuclear chains that interact via Mie potentials can be obtained from the appropriate contributions of the different groups that make up the molecules as the sum of four separate contributions [22], as in other SAFT approaches [78, 79]:

$$
A = A^{\text{ideal}} + A^{\text{mono}} + A^{\text{chain}} + A^{\text{assoc}},
$$

where $A^{\text{ideal}}$ is the free energy of an ideal gas, $A^{\text{mono}}$ the contribution to the free energy due to the segment-segment repulsion and dispersion interactions, $A^{\text{chain}}$ the free energy change due to the formation of chains from Mie segments, and $A^{\text{assoc}}$ the contribution due to association. The detailed expressions of the theory and each of the four contributions are given in the original publication [22] and in Ref. [239] for details of the expressions used in the association term. In this thesis, we quote the concise version of the theory from Refs. [169, 240].
Chapter 3: Alkane and alcohol in aqueous systems

Ideal term

The free energy corresponding to an ideal mixture of molecules is given by [241]:

$$\frac{A_{\text{ideal}}}{N k_B T} = \left( \sum_{i=1}^{N_C} x_i \ln \left( \frac{\rho_i \Lambda_i^3}{N_i} \right) \right) - 1,$$

(3.5)

where $x_i$ is the mole fraction of component $i$ in the mixture, $\rho_i = N_i/V$ is the number density of component $i$, $N_i$ being the number of molecules of component $i$, $V$ the total volume, $N$ the total number of molecules, $k_B$ the Boltzmann constant, and $T$ the absolute temperature. The summation in Eq. (3.5) is over all of the components $N_C$ present in the mixture (i.e., $N = \sum_{i}^{N_C} N_i$). The ideal free energy incorporates the effects of the translational, rotational and vibrational contributions to the kinetic energy implicitly in the thermal de Broglie volume, $\Lambda_i^3$.

Monomer term

The free-energy contribution due to repulsive and attractive interactions for the monomeric fluid characterized by the Mie potential is obtained following a Barker-Henderson [242, 243] high-temperature perturbation expansion up to third order [237] which can be expressed as:

$$\frac{A_{\text{mono.}}}{N k_B T} = \frac{A_{\text{HS}}}{N k_B T} + \frac{A_1}{N k_B T} + \frac{A_2}{N k_B T} + \frac{A_3}{N k_B T},$$

(3.6)

where $A_{\text{HS}}$ is the free energy of the hard-sphere reference system of diameter $d_{kk}$ which is temperature-dependent [243]. The remaining terms correspond to the mean-attractive energy, which is calculated from a knowledge of the contact radial distribution function of the reference hard-sphere fluid at an effective density; the fluctuation term, calculated following the improved macroscopic compressibility approximation proposed by Zhang [244] and a correction of the type proposed by Paricaud [245]; and the third-order term which is represented with an empirical expression based on simulation phase-coexistence data for representative pure Mie fluids (see Refs. [22, 169] for detailed information about the development of the each of these terms).

Chain term

In the SAFT-$\gamma$ approach the change in free energy associated with the formation of a molecule from its constituting segments is obtained using average molecular parameters ($\tilde{\sigma}_{ii}$, $\bar{d}_{ii}$, $\bar{\varepsilon}_{ii}$ and $\bar{\lambda}_{ii}$) for each molecular species $i$ [25, 26], so that the formal TPT1 expression [246, 247] can be retained. The averaging of the molecular size and energy parameters is independent of the
composition of the mixture; it requires the molecular fraction \( z_{k,i} \) of a given group \( k \) in a molecule \( i \):

\[
z_{k,i} = \frac{\nu_{k,i} \nu_k^* S_k}{\sum_{l=1}^{N_G} \nu_{l,i} \nu_l^* S_l} ,
\]

(3.7)

The average molecular segment diameter \( \bar{\sigma}_{ii} \) and the effective hard-sphere diameter \( \bar{d}_{ii} \) are then defined as:

\[
\bar{\sigma}_{ii}^3 = \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} z_{k,i} z_{l,i} \sigma_{kl}^3 ,
\]

(3.8)

and

\[
\bar{d}_{ii}^3 = \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} z_{k,i} z_{l,i} d_{kl}^3 ,
\]

(3.9)

respectively. Other effective molecular parameters are obtained in the same way, so that the average interaction energy \( \bar{\varepsilon}_{ii} \) and exponents which characterize the range of the potential \( \bar{\lambda}_{ii} \) are obtained as:

\[
\bar{\varepsilon}_{ii} = \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} z_{k,i} z_{l,i} \varepsilon_{kl} ,
\]

(3.10)

and

\[
\bar{\lambda}_{ii} = \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} z_{k,i} z_{l,i} \lambda_{kl} ,
\]

(3.11)

respectively, for both the repulsive \( \bar{\lambda}_{ii}^r \), and the attractive \( \bar{\lambda}_{ii}^a \) exponents. The resulting contribution to the free energy of the mixture due to the formation of chains of tangent (or fused) segments using the effective molecular parameters is given by:

\[
\frac{A_{\text{chain}}}{N k_B T} = - \sum_{i=1}^{N_G} x_i \left( \sum_{k=1}^{N_G} \nu_{k,i} \nu_k^* S_k - 1 \right) \ln g_{ii}^\text{Mie}(\bar{\sigma}_{ii}; \zeta_x) ,
\]

(3.12)

where \( g_{ii}^\text{Mie}(\bar{\sigma}_{ii}; \zeta_x) \) is the value of the radial distribution function (RDF) evaluated at a distance \( \bar{\sigma}_{ii} \) in a hypothetical fluid of packing fraction \( \zeta_x \), defined as \( \zeta_x = \frac{\pi}{6} \rho_s \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} x_{s,k} x_{s,l} d_{kl}^3 \). It is relevant to mention here that a different expression is used by Peng et al. [125] in the GC-SAFT-VR theory. They propose a chain term summing the contributions from the different segments.
in each group and each contact between groups, thereby preserving the connectivity present in the molecule. It has been shown previously [22] that the two approaches perform similarly in the description of model systems but that the SAFT-\(\gamma\) approach is generally more accurate for the treatment of real substances.

**Association term**

The contribution to the Helmholtz free energy due to the association of molecules via short-range bonding sites follows from the original TPT1 expressions of Wertheim [144–147, 238, 247] by summing over the number of species \(N_C\), the number of groups \(N_G\), and the number of site types on each group \(N_{\text{ST},k}\), so that

\[
\frac{A_{\text{assoc.}}}{N_k B T} = \sum_{i=1}^{N_C} \sum_{k=1}^{N_G} \sum_{a=1}^{N_{\text{ST},k}} n_{k,a} \ln x_{i,k,a} + \frac{1 - X_{i,k,a}}{2},
\]

(3.13)

where \(n_{k,a}\) is the number of sites of type \(a\) on group \(k\), and \(X_{i,k,a}\) is the fraction of molecules of component \(i\) that are not bonded at a site of type \(a\) on group \(k\). \(X_{i,k,a}\) is obtained from the solution of the mass action equations as [25, 147, 238, 248, 248]:

\[
X_{i,k,a} = \left[1 + \rho \sum_{j=1}^{N_C} x_j \sum_{l=1}^{N_G} \sum_{b=1}^{N_{\text{ST},l}} n_{l,b} X_{j,l,b} \Delta_{ij,kl,ab} \right]^{-1},
\]

(3.14)

where \(\Delta_{ij,kl,ab}\) characterizes the overall strength of the association between a site of type \(a\) on a group of type \(k\) of component \(i\) and a site of type \(b\) on a group of type \(l\) of component \(j\). It is approximated as:

\[
\Delta_{ij,kl,ab} = F_{kl,ab} K_{kl,ab} I_{ij,kl,ab},
\]

(3.15)

where \(F_{kl,ab} = \exp (x_{kl,ab} B T) - 1\), \(K_{kl,ab}\) is a bonding-volume parameter, and \(I_{ij,kl,ab}\) is a polynomial temperature-density correlation of the association integral for a Lennard-Jones monomer, expressed as:

\[
I_{ij,kl,ab} = \sum_{p=0}^{10} \sum_{q=0}^{10-p} c_{pq} \left(\rho \sigma_x^3\right)^p \left(\frac{B T}{x_{ij}}\right)^q,
\]

(3.16)

where the coefficients \(c_{pq}\) for the association contribution are given in Ref. [239]; \(\sigma_x^3\) is obtained as:

\[
\sigma_x^3 = \sum_{k=1}^{N_C} \sum_{l=1}^{N_G} x_{s,k} x_{s,l} \sigma_{kl}^3,
\]

(3.17)
and $\varepsilon_{ij}$ is given by:

$$
\varepsilon_{ij} = \frac{\sqrt{\sigma_{ii} \sigma_{jj}}}{\sigma_{ij}} \sqrt{\varepsilon_{ii} \varepsilon_{jj}},
$$

(3.18)

where $\sigma_{ii}$ is given by Eq. 3.8 and

$$
\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}.
$$

(3.19)

### 3.2.2 Combining Rules

The expressions for the Helmholtz free energy presented above require the prescription of a number of unlike group parameters. These are typically determined using combining rules or estimation from experimental data. The event of determining unlike interaction using combining rules or experimental data will be discussed later in this chapter and in Chapter 4.

The unlike segment diameter $\sigma_{kl}$ is obtained using the Lorentz-like arithmetic mean of the like diameters [249]:

$$
\sigma_{kl} = \frac{\sigma_{kk} + \sigma_{ll}}{2}.
$$

(3.20)

The same combining rule is applied for the calculation of the unlike effective hard-sphere diameter $d_{kl}$:

$$
d_{kl} = \frac{d_{kk} + d_{ll}}{2}.
$$

(3.21)

Although a more rigorous approach would be to calculate the corresponding effective diameter by numerical integration using the segment diameter, as for like group interactions, this simple rule has been shown to be of comparable accuracy, at a fraction of the computational cost [237].

The unlike dispersion energy $\varepsilon_{kl}$ between groups $k$ and $l$ is obtained by applying an augmented geometric mean (Berthelot-like rule), which also accounts for asymmetries in size [250]:

$$
\varepsilon_{kl} = \frac{\sqrt{\sigma_{kk}^{3} \sigma_{ll}^{3}}}{\sigma_{kl}^{3}} \varepsilon_{kk} \varepsilon_{ll}.
$$

(3.22)

The exponents of the unlike segment-segment interaction $\lambda_{kl}^{I}$ and $\lambda_{kl}^{S}$ are obtained as:

$$
\lambda_{kl} = 3 + \sqrt{(\lambda_{kk} - 3)(\lambda_{ll} - 3)},
$$

(3.23)

which results from imposing the geometric mean of the integrated van der Waals energy (Berthelot rule) of a Sutherland fluid of range $\lambda_{kl}$ [22].
In associating mixtures, the unlike association energy $\varepsilon_{kl,ab}^{\text{HB}}$ can be obtained by using a simple geometric mean:

$$
\varepsilon_{kl,ab}^{\text{HB}} = \sqrt{\varepsilon_{kk,aa}^{\text{HB}} \varepsilon_{ll,bb}^{\text{HB}}},
$$

(3.24)

while the unlike bonding volume $K_{kl,ab}$ is obtained as:

$$
K_{kl,ab} = \left( \frac{\sqrt[3]{K_{kk,aa}} + \sqrt[3]{K_{ll,bb}}}{2} \right)^3.
$$

(3.25)

### 3.2.3 Property calculations

The SAFT-$\gamma$ Mie equation of state expressed in Eq. (3.4) provides the functional form of the Helmholtz free energy as a function of volume $V$, temperature $T$ and the composition vector $\mathbf{N}$ ($N_1, N_2, \ldots$). Other properties can be determined through standard thermodynamic relations [101, 102]. The pressure ($p$), residual chemical potential ($\mu_i^{\text{res}}$) and fugacity coefficient ($\varphi_i$) of component $i$ can be obtained from the Helmholtz free energy as:

$$
p = - \frac{\partial A(T,V,\mathbf{N})}{\partial V} \bigg|_{T,\mathbf{N}},
$$

(3.26)

$$
\mu_i^{\text{res}}(T,p,\mathbf{x}) = \frac{\partial A^{\text{res}}(T,V,\mathbf{N})}{\partial N_i} \bigg|_{T,V,N_i,\mathbf{x} = \mathbf{x}_i} - RT \ln Z(T,p,\mathbf{x}), \text{ and}
$$

(3.27)

$$
\ln \varphi_i(T,p,\mathbf{x}) = \frac{\mu_i^{\text{res}}(T,p,\mathbf{x})}{RT}, \text{ respectively}
$$

(3.28)

where $A^{\text{res}} = A - A^{\text{ideal}}$, the compressibility factor $Z = \frac{pV_p}{RT}$, where $V_p = \frac{V}{N}$ is the molar volume corresponding to the specified pressure, $N$ is the total number of molecules and $\mathbf{x} = \frac{N}{N_i}$. Through the pressure, chemical potential and fugacity coefficient, the phase behaviour and solution properties can be calculated. In this chapter, we use phase behaviour and limited excess property data to determine the groups' parameters and assess the performance of the models developed by predicting the solvation properties of $n$-alkanes and 1-alcohols in aqueous solutions.

### Fluid phase equilibria

At a given pressure, temperature and total composition, the conditions of phase equilibria are solved using the solvers available in the gPROMS software package [251] and the HELD flash
algorithm [252, 253] is used to confirm the stability of the equilibrium phases.

**Henry’s law constant**

The Henry’s law constant of solute \( i \) in solvent \( j \) \( K_{H_{i,j}} \) is obtained from measurement of partial pressures of the highly dilute solute over its solution [101]. It is often used as an estimate of solubility, as larger values of \( K_{H_{i,j}} \) usually correspond to lower solubility of the solute in the solvent and vice versa [103]. \( K_{H_{i,j}} \) can be calculated from the fugacity coefficient as:

\[
K_{H_{i,j}}(T) = \Phi_{i,j}^\infty(T) \cdot p^{sat},
\]  

where \( \Phi_{i,j}^\infty \) is the liquid-phase fugacity coefficient of the infinitely dilute solute \( i \) in the mixture at saturated (vapour) pressure \( p^{sat} \). In practice, this pressure is calculated at the fluid-phase equilibrium of the mixture at the limiting concentration of the solute \( (x_i \to 0) \).

**Solvation Gibbs free energy**

The solvation Gibbs free energy is defined as the change in Gibbs free energy in transferring a solute particle from an ideal gas phase to a solution at infinite dilution at constant temperature and pressure. By definition, it is equivalent to the residual chemical potential of a solute \( i \) (cf. Eq. (3.28)) at infinite dilution in a solvent \( j \) [103], i.e.,

\[
\Delta G_{i,j}^{\text{sol}}(T, p) = \mu_{i,j}^{\text{res},\infty}(T, p) = RT \ln \Phi_{i,j}^\infty(T, p).
\]  

In practice a solute composition \( x_i = 10^{-10} \) is used to ensure the infinite dilution condition. The calculations are performed via a single-phase calculation with specified \( T = 298.15 \) K and \( p = 0.100 \) MPa. Alternatively, based on Eqs. (3.29) and (3.30), the solvation Gibbs free energy can also be calculated directly from the Henry’s constant [254] as:

\[
\Delta G_{i,j}^{\text{sol}}(T) = RT \ln \left[ K_{H_{i,j}}(T) / p^{sat} \right].
\]  

**Infinite dilution activity coefficient**

The infinite dilution activity coefficient \( \gamma_{i,j}^\infty \) provides another measure of the behaviour of a solute molecule \( i \) in the solvent environment \( j \). It is an excess property calculated from the ratio of the fugacity coefficient of the solute in solution at infinite dilution, \( \Phi_{i,j}^\infty \) (i.e., \( x_i = 10^{-10} \)) and the fugacity coefficient of the pure solute, \( \Phi_i^\infty \) (i.e., \( x_i \to 1 \)), at the same temperature and pressure.
(via a single-phase calculation) [102]:

\[
\gamma_{i,j}^{\infty}(T, p) = \frac{\varphi_{i,j}^{\infty}(T, p)}{\varphi_{i}^{\infty}(T, p)}.
\] (3.32)

### 3.3 Model development and phase equilibrium calculations

To study the phase behaviour of alkane and alcohol aqueous mixtures, SAFT-\(\gamma\) Mie parameters characterising the interactions between CH\(_3\), CH\(_2\), CH\(_2\)OH and H\(_2\)O functional groups are needed. The parameter estimation procedure is at the very heart of the group-contribution methodology. In most cases the functional group parameters are estimated by regression to experimental data of compounds that contain the relevant groups of interest. The procedure is initiated with a chemical family containing “simple” functional groups (e.g., CH\(_3\) and CH\(_2\)) and compounds that are composed of a single group (e.g., H\(_2\)O). Once the parameters for these groups are defined, they are transferred to study other chemical families, partly composed of some of the established groups, and including also additional functional groups. Furthermore, the heteronuclear formulation of the SAFT-\(\gamma\) Mie approach allows the parameter estimation to be carried out based on pure-component data alone even when determining cross interaction parameters between unlike groups. An example for this is the characterisation of the CH\(_3\) and CH\(_2\) groups, in which pure-component data of the \(n\)-alkane series from ethane to \(n\)-decane are used [22]. Vapour pressures and saturated liquid densities in a temperature range spanning from 40% to 90% of the experimental critical temperature are used. Mixture data are employed occasionally in the characterisation in order to improve the statistical significance of the group parameters or to obtain cross-interaction parameters between a functional group (e.g., CH\(_2\)) and a molecular group (e.g., H\(_2\)O). A strength of group contribution approaches is that for compounds that can be defined from a number of functional groups (i.e., not compounds that are defined as molecular groups) it is not necessary to have experimental data for the specific compound of interest.

In order to reduce the time and complexity of the parameter estimation procedure a number of parameters, typically \(\nu^*_k\), \(N_{ST,k}\), \(n_{k,a}\), and \(\lambda_{kk}^k\) are pre-assigned fixed values based on the chemical nature of each group or, in some cases, with a trial-and-error approach, and the unlike-interaction parameters \(\sigma_{kl}^k\), \(\lambda_{kl}^k\), and often \(X_{kl}^k\), are determined by means of appropriate combining rules (Eqs. 3.20 and 3.23). In this thesis, we use our previously developed parameters for the CH\(_3\), CH\(_2\), and H\(_2\)O groups [22, 169] and apply them to obtain new CH\(_2\)OH like-interaction group parameters as well as its unlike-interaction parameters. The CH\(_3\) and CH\(_2\) functional
groups have been examined extensively for their applicability and transferability for a variety of thermodynamic properties and systems [22]. The model for H₂O used here was presented in Dufal et al. [239]. It includes four association sites (two sites of type H and two of type e₁) to meditate hydrogen bonding. In addition to the development of the CH₂OH functional group, we revise the group parameters for the interaction between the alkyl CH₃/CH₂ groups and water that reported previously [169] in order to obtain a more accurate description and better transferability in the modelling of n-alkane + water mixtures. New CH₂OH – H₂O unlike-interaction parameters are also presented here and are used to model a number of 1-alcohol + water mixtures.

**Table 3.1:** Like group parameters for use within the SAFT-γ Mie group-contribution approach: νₖ is the number of segments constituting group k, Sₖ the shape factor, λₖ the repulsive exponent, λₖ the attractive exponent, σₖ the segment diameter of group k, and εₖ the dispersion energy of the Mie potential characterising the interaction of two k groups (k₆ is the Boltzmann constant); Nₖ represents the number of association site types on group k, with nₖ,H and nₖ,e₁ denoting the number of association sites of type H and e₁, respectively.

<table>
<thead>
<tr>
<th>Group k</th>
<th>νₖ</th>
<th>Sₖ</th>
<th>λₖ</th>
<th>λₖ</th>
<th>σₖ/Å</th>
<th>(εₖ/k₆)/K</th>
<th>NST,k</th>
<th>nₖ,H</th>
<th>nₖ,e₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₃</td>
<td>1</td>
<td>0.57255</td>
<td>15.050</td>
<td>6.0000</td>
<td>4.0773</td>
<td>256.77</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH₂</td>
<td>1</td>
<td>0.22932</td>
<td>19.871</td>
<td>6.0000</td>
<td>4.8801</td>
<td>473.39</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH₂OH</td>
<td>2</td>
<td>0.58538</td>
<td>22.699</td>
<td>6.0000</td>
<td>3.4054</td>
<td>407.22</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>H₂O</td>
<td>1</td>
<td>1.0000</td>
<td>17.020</td>
<td>6.0000</td>
<td>3.0063</td>
<td>266.68</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3.2: Group-group dispersion interaction energy $\varepsilon_{kl}$, repulsive exponent $\lambda_{kl}^*$, and site-site association energy $\varepsilon_{kl,ab}^{\text{HB}}$ and bonding volume $K_{kl,ab}$ for use within the SAFT-γ Mie group-contribution approach. The unlike segment diameter $\sigma_{kl}$ is also obtained from combining rules [22] and all unlike attractive exponents $\lambda_{kl}^*$ = 6.0000. CR denotes that a combining rule is used [22].

<table>
<thead>
<tr>
<th>Group $k$</th>
<th>Group $l$</th>
<th>$(\varepsilon_{kl}/k_B)/K$</th>
<th>$\lambda_{kl}^*$</th>
<th>site a of group $k$</th>
<th>site b of group $l$</th>
<th>$(\varepsilon_{kl,ab}^{\text{HB}}/k_B)/K$</th>
<th>$K_{kl,ab}/\text{Å}^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_3$</td>
<td>CH$_3$</td>
<td>256.77</td>
<td>15.050</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH$_3$</td>
<td>CH$_2$OH</td>
<td>350.77</td>
<td>CR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH$_3$</td>
<td>H$_2$O</td>
<td>333.20</td>
<td>CR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH$_2$</td>
<td>CH$_2$</td>
<td>473.39</td>
<td>19.871</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH$_2$</td>
<td>CH$_2$OH</td>
<td>423.17</td>
<td>CR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH$_2$</td>
<td>H$_2$O</td>
<td>423.63</td>
<td>100.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CH$_2$OH</td>
<td>CH$_2$OH</td>
<td>407.22</td>
<td>22.699</td>
<td>e$_1$</td>
<td>H</td>
<td>2097.9</td>
<td>62.309</td>
</tr>
<tr>
<td>CH$_2$OH</td>
<td>H$_2$O</td>
<td>353.37</td>
<td>CR</td>
<td>e$_1$</td>
<td>H</td>
<td>2153.2</td>
<td>147.40</td>
</tr>
<tr>
<td>CH$_2$OH</td>
<td>H$_2$O</td>
<td>353.37</td>
<td>CR</td>
<td>H</td>
<td>e$_1$</td>
<td>621.68</td>
<td>425.00</td>
</tr>
</tbody>
</table>

3.3.1 The CH$_2$OH group and its binary interaction parameters CH$_2$OH - CH$_3$ and CH$_2$OH - CH$_2$

In general, the identification of groups in GC approaches is heuristic; the final choice is usually the combination of groups that results in the best representation of the experimental data and the trade-off of delivering transferability. In a group-contribution context, 1-alcohols can be modelled as having the alcohol functional group represented with either an OH or a CH$_2$OH group. In a recent publication on the modelling of 1-alcohols within the SAFT-γ SW approach [148], it was shown that the description of this chemical family using a CH$_2$OH group resulted in more accurate predictions of the fluid-phase behaviour of binary mixtures of 1-alcohols (including aqueous solutions), whilst retaining an excellent description of the pure-component VLE data employed in the development of the group parameters. Following the idea of including the first neighbouring methanediyl group in the definition of the alcohol group (as initially suggested by Wu and Sandler) [119], we propose to use this group here.

The CH$_2$OH group (cf. Fig. 3.1) is modelled with two identical segments ($\nu^*_{\text{CH}_2\text{OH}} = 2$) and two association site types ($N_{\text{ST,CH}_2\text{OH}} = 2$), two sites of type e$_1$ ($n_{\text{CH}_2\text{OH},e_1} = 2$) represent the
two electron lone-pairs on the oxygen atom and one site of type H \((n_{\text{CH}_2\text{OH},\text{H}} = 1)\) represents the hydrogen atom (a 3B association scheme according to Huang and Radosz [255]) and only sites of different type are allowed to interact (i.e., \(\epsilon_{\text{HB}^{\text{CH}_2\text{OH},\text{CH}_2\text{OH},c_1}} = \epsilon_{\text{HB}^{\text{CH}_2\text{OH},\text{CH}_2\text{OH},c_1\text{H}}} = 0\)). Of the remaining parameters, \(\lambda_{\text{CH}_2\text{OH},\text{CH}_2\text{OH}}\) is set to be equal to six, so that only the \(S_{\text{CH}_2\text{OH},\text{CH}_2\text{OH}}\), \(\lambda_{\text{CH}_2\text{OH},\text{CH}_2\text{OH}}\), \(\sigma_{\text{CH}_2\text{OH},\text{CH}_2\text{OH}}\), \(\varepsilon_{\text{CH}_2\text{OH},\text{CH}_2\text{OH}}\), \(\varepsilon_{\text{CH}_2\text{OH},\text{CH}_2\text{OH},c_1}\), \(K_{\text{CH}_2\text{OH},\text{CH}_2\text{OH},c_1\text{H}}\) and the unlike-interaction parameters: \(\varepsilon_{\text{CH}_2\text{OH},\text{CH}_3}\) and \(\varepsilon_{\text{CH}_2\text{OH},\text{CH}_2}\) need to be determined. These are characterised by comparison to experimental data. Vapour pressure \((N_{vap} = 214)\), saturated liquid density \((N_{sat} = 336)\) of the pure 1-alcohols from ethanol to 1-decanol [256] as well as some mixture data (molar excess enthalpy of mixing for 1-pentanol + n-heptane system [257] \((N_{h,E} = 25)\) and LLE data for the \(n\)-tetradecane + ethanol mixture [258]\((N_{x_{C14}} = 7\) and \(N_{y_{C14}} = 9)\)) are used. The objective function used in the parameter estimation is given by:

\[
\min_{\Omega} f_{\text{obj}} = w_1 \sum_{i=1}^{N_C} \sum_{q=1}^{N_{vap,i}} \left[ \frac{p_{\text{exp},i}(T_q) - p_{\text{calc},i}(T_q; \Omega)}{p_{\text{exp},i}(T_q)} \right]^2 \\
+ w_2 \sum_{i=1}^{N_C} \sum_{q=1}^{N_{sat,i}} \left[ \frac{\rho_{\text{exp},i}(T_q) - \rho_{\text{calc},i}(T_q; \Omega)}{\rho_{\text{exp},i}(T_q)} \right]^2 \\
+ w_3 \sum_{q=1}^{N_{h,E}} \left[ \frac{h_{E,\text{exp},q}(T_q, p_q, x_q) - h_{E,\text{calc},q}(T_q, p_q, x_q; \Omega)}{h_{E,\text{exp},q}(T_q, p_q, x_q)} \right]^2 \\
+ w_4 \sum_{q=1}^{N_{x_{C14}}} \left[ \frac{x_{\text{exp},q}(T_q, p_q) - x_{\text{calc},q}(T_q, p_q; \Omega)}{x_{\text{exp},q}(T_q, p_q)} \right]^2 \\
+ w_5 \sum_{q=1}^{N_{y_{C14}}} \left[ \frac{y_{\text{exp},q}(T_q, p_q) - y_{\text{calc},q}(T_q, p_q; \Omega)}{y_{\text{exp},q}(T_q, p_q)} \right]^2,
\]

where the first two sums are over the \(N_C\) pure components \(i\) included in the estimation over the number \(N_{vap}\) of experimental vapour pressure points \((N_{vap} = 214)\) or \(N_{sat}\) saturated liquid density points \((N_{sat} = 336)\), while the third term sums over the number \(N_{h,E}\) of experimental molar excess enthalpy points \((N_{h,E} = 25)\). The last two terms sum over LLE data points of the ethanol+tetradecane mixture: \(N_{x_{C14}}\) ethanol-rich equilibrium mole fraction data points \((N_{x_{C14}} = 7)\) and \(N_{y_{C14}}\) alkane-rich equilibrium mole fraction data points \((N_{y_{C14}} = 9)\). \(\Omega\) denotes the vector of parameters to be estimated. The desired level of accuracy for each calculated (calc.) property can be adjusted by means of weighting factors: \(w_1\) for \(N_{vap}\), \(w_2\) for \(N_{sat}\), \(w_3\) for \(N_{h,E}\), \(w_4\) for \(N_{x_j}\), and \(w_5\) for \(N_{y_j}\). In this case, \(w_1 = w_2 = 5, w_3 = 1, w_4 = w_5 = 10\) are employed. The estimations are performed using the commercial software package gPROMS [251].
Table 3.3: Percentage average absolute deviations (%AAD) for the vapour pressures $p_{\text{vap},i}(T)$ and the saturated liquid densities $\rho_{\text{sat},i}(T)$ obtained with the SAFT-$\gamma$ Mie group-contribution approach compared to experiment [256] (where $N_{p_{\text{vap},i}}$ and $N_{\rho_{\text{sat},i}}$ are the number of vapour pressure and saturated liquid density data points used for each of the 1-alcohols considered in the parameter estimation).

<table>
<thead>
<tr>
<th>Compound $i$</th>
<th>$T_{\text{range}}$/K</th>
<th>$N_{p_{\text{vap},i}}$</th>
<th>%AAD $p_{\text{vap},i}(T)$</th>
<th>$N_{\rho_{\text{sat},i}}$</th>
<th>%AAD $\rho_{\text{sat},i}(T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol</td>
<td>231-463</td>
<td>30</td>
<td>2.55</td>
<td>159-463</td>
<td>39</td>
</tr>
<tr>
<td>1-Propanol</td>
<td>280-483</td>
<td>25</td>
<td>5.85</td>
<td>169-483</td>
<td>38</td>
</tr>
<tr>
<td>1-Butanol</td>
<td>295-506</td>
<td>26</td>
<td>2.76</td>
<td>186-506</td>
<td>39</td>
</tr>
<tr>
<td>1-Pentanol</td>
<td>278-508</td>
<td>29</td>
<td>0.71</td>
<td>278-508</td>
<td>29</td>
</tr>
<tr>
<td>1-Hexanol</td>
<td>310-428</td>
<td>17</td>
<td>1.40</td>
<td>273-547</td>
<td>38</td>
</tr>
<tr>
<td>1-Heptanol</td>
<td>343-445</td>
<td>14</td>
<td>1.53</td>
<td>273-563</td>
<td>38</td>
</tr>
<tr>
<td>1-Octanol</td>
<td>296-549</td>
<td>31</td>
<td>3.59</td>
<td>263-583</td>
<td>39</td>
</tr>
<tr>
<td>1-Nonanol</td>
<td>366-481</td>
<td>15</td>
<td>1.08</td>
<td>293-596</td>
<td>38</td>
</tr>
<tr>
<td>1-Decanol</td>
<td>301-526</td>
<td>27</td>
<td>2.81</td>
<td>293-613</td>
<td>38</td>
</tr>
<tr>
<td>average</td>
<td>-</td>
<td>-</td>
<td>2.48</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The optimised parameters are presented in Tables 3.1 and 3.2. The deviation of the theoretical description in comparison to experimental (exp.) data is expressed by means of the percentage average absolute deviation (%AAD) as following:

$$%\text{AAD } R_j = \frac{1}{N_{R_j}} \sum_{i=1}^{N_{R_j}} \left| \frac{R_{j,i}^{\text{exp}} - R_{j,i}^{\text{calc}}}{R_{j,i}^{\text{exp}}} \right| \times 100,$$

(3.34)

where $N_{R_j}$ is the number of data points of a property $R_j$, $R_{j,i}^{\text{exp}}$ the experimental value, and $R_{j,i}^{\text{calc}}$ the calculated value for the same property, at the conditions of the $i^{th}$ experimental point.

The regression results in an average %AAD of 2.48% for the vapour pressures [256] and 1.82% for the saturated liquid densities [256] of the pure compounds (cf. Table 3.3). For the mixture data [258], the %AADs of the $n$-tetradecane composition in the ethanol-rich and alkane-rich phases are 5.13% and 1.19%, respectively. The average absolute error for the molar excess enthalpy of mixing for the 1-pentanol + $n$-heptane system [257] is 0.20 kJ/mol.

As mentioned earlier, it is possible also to use only pure-component data to characterise the functional group parameters within the SAFT-$\gamma$ Mie approach. Considering the same model
to treat the CH$_2$OH group, the optimised parameters estimated using pure component vapour-
liquid equilibrium data from ethanol to 1-decanol [256] only are: $\nu_{\text{CH}_2\text{OH}}^* = 1$, $S_{\text{CH}_2\text{OH}} = 0.92122$,
$\sigma_{\text{CH}_2\text{OH}} = 3.6569$ Å, $\lambda_{\text{CH}_2\text{OH}-\text{CH}_2\text{OH}} = 6.0000$, $\lambda_{\text{CH}_2\text{OH}-\text{CH}_2\text{OH}}^* = 11.548$, $\varepsilon_{\text{CH}_2\text{OH}-\text{CH}_2\text{OH}}/k_B = 415.91$ K, $\varepsilon_{\text{CH}_2\text{OH}-\text{CH}_2\text{OH},e_{1H}}/k_B = 2510.5$ K, and $K_{\text{CH}_2\text{OH}-\text{CH}_2\text{OH},e_{1H}} = 19.886$ Å$^3$. The unlike
dispersion energies with the other functional groups on the 1-alkanol are: $\varepsilon_{\text{CH}_2\text{OH}-\text{CH}_3}/k_B = 233.06$ K and $\varepsilon_{\text{CH}_2\text{OH}-\text{CH}_2}/k_B = 391.12$ K. This CH$_2$OH model is found to describe accurately
the pure-component data for the series of 1-alkanols considered, with %AADs for the entire
series of 1.56% for the vapour pressures and 1.31% for the saturated liquid densities, which are
smaller than those presented in Table 3.3. One is tempted to select these as the best performing.
However, when the two CH$_2$OH models are compared for the prediction of the phase behaviour
of binary mixtures containing alkanes and 1-alkanols a different conclusion emerges.

In Fig. 3.2, the fluid-phase behaviour of $n$-heptane + 1-pentanol [259] and $n$-undecane +
1-tetradecanol [260] mixtures is presented. As can be seen, both CH$_2$OH models lead to a good
description of the VLE of moderate and long alcohol mixtures with alkanes. The predicted
phase equilibrium of alkane + short alcohol (carbon number $\leq 4$) mixtures with the CH$_2$OH
model optimised solely from pure component data, however, exhibits a region of liquid-liquid
immiscibility at high pressure which is not observed experimentally. The predicted VLE of the
$n$-butane + ethanol [261] and $n$-heptane + 1-butanol [262] mixtures are shown in Fig. 3.3. The
two CH$_2$OH models are also used for the description of the LLE of the $n$-tetradecane + ethanol
mixture [258] (Fig. 3.4). Large deviations from the experimental data are observed for the LLE
predictions when the CH$_2$OH model optimised from pure component data alone is used, whereas
a significantly better description is achieved with the CH$_2$OH model developed from both pure
component and mixture data. The inability to capture accurately the properties of small polar
compounds and their mixtures is a common drawback of group-contribution approaches. The
polarisation caused by strongly polar groups undermines the underlying assumption of the GC
approach that each group is independent of others [173]. This proximity effect becomes more
pronounced in the prediction of mixtures involving small polar compounds. Although it is dif-
cult for the GC approach to model this type of compounds, techniques involving the use of
second-order group interactions [224] or the inclusion of mixture data to develop model param-
eters can be used to circumvent the problem. In this study, we show that with the inclusion of
mixture data (specifically, LLE of $n$-tetradecane + ethanol [258] and molar excess enthalpy of
mixing of 1-pentanol + $n$-heptane [257]) in the parameter estimation formulation the transfer-
ability of the CH$_2$OH model improves significantly. This is evidenced by the better description
of the fluid phase behaviour of alcohol and alkane mixtures, cf. Figs. 3.2 and 3.3. It is worth highlighting the accurate prediction obtained for the undecane + tetradecanol mixture [260] (Fig. 3.2 b), noting that these two compounds were not included in the regression of the group parameters.

Figure 3.2: Isothermal pressure–mole fraction \((P–x)\) phase diagrams of the vapour-liquid equilibria of \(n\)-alkanes and \(1\)-alcohols: (a) \(n\)-heptane + 1-pentanol mixture at temperatures of \(T = 348.15 \text{ K}\) (circles) [259] and \(T = 368.15 \text{ K}\) (triangles) [259], and (b) undecane + tetradecanol mixture at temperatures of \(T = 393.15 \text{ K}\) (circles) [260] and \(T = 413.15 \text{ K}\) (triangles) [260]. The symbols represent the experimental data, the dashed curves the SAFT-\(\gamma\) Mie predictions obtained with a CH\(_2\)OH model based on pure-component data alone, and the continuous curves the SAFT-\(\gamma\) Mie predictions obtained with a CH\(_2\)OH model developed from pure and mixture data (the parameters are reported in Tables 3.1 and 3.2).
Figure 3.3: Isothermal pressure–mole fraction ($P$–$x$) phase diagrams of the vapour-liquid equilibria of $n$-alkanes and short 1-alcohols: (a) $n$-butane + ethanol mixture at a temperatures of $T = 373.27$ K [261], and (b) $n$-heptane + 1-butanol mixture at a temperature of $T = 313.15$ K [262]. The symbols represent the experimental data and the curves the prediction with the SAFT-$\gamma$ Mie approach. The descriptions of the curves are given in Fig. 3.2.

Figure 3.4: Isobaric temperature–mole fraction ($T$–$x$) phase diagram of the liquid-liquid equilibria of $n$-tetradecane + ethanol at a pressure of $p = 0.101$ MPa. The circles [258] represent the experimental data and the curves the prediction with the SAFT-$\gamma$ Mie approach. The descriptions of the curves are given in Fig. 3.2.

3.3.2 CH$_3$ - H$_2$O and CH$_2$ - H$_2$O binary interaction parameters

In this section we determine the unlike interactions between CH$_3$/ CH$_2$ segments and H$_2$O by adjusting both $\varepsilon_{kl}$ and $\lambda_{kl}$ parameters to experimental mixture data. It is not possible to obtain these interactions using pure component data since H$_2$O is a molecular group. In a previous study [169], models for the CH$_3$ – H$_2$O and CH$_2$ – H$_2$O interactions were obtained by adjusting
only $\varepsilon_{kl}$ to achieve a good agreement of the experimental liquid-liquid equilibrium data for the n-heptane + water system. The rest of the parameters including the range of the repulsive exponent $\lambda_{kl}^r$ were calculated from combining rules. The previous models [169] provided a good description of the alkane solubilities in the water-rich phase only for the medium-chain-length alkanes. However, the solubilities predicted for mixtures with other alkanes were not accurate enough. Here in addition to considering both $\varepsilon_{kl}$ and $\lambda_{kl}^r$ for parameter estimation, we use three-phase (VLLIE) solubility data of n-pentane and n-octane aqueous mixtures [263, 264] in the temperature range of 280 – 400 K to optimise the unlike parameters. The coexisting liquid compositions at 298.00 K at the pressure of three-phase coexistence from n-hexane to n-decane aqueous mixtures [265–268] are also included in the parameter estimation ($N_{\text{total}} = 42$). The objective function used in the parameter estimation is given by:

$$
\min_{\theta} f_{\text{obj}} = w_1 \sum_{C_5, C_6} \sum_{q=1}^{N_{x_i}} \left[ \frac{x_i^{\text{exp}}(T_q) - x_i^{\text{calc}}(T_q, \Omega)}{x_i^{\text{exp}}(T_q)} \right]^2 \\
+ w_2 \sum_{C_5} \left[ \frac{x_i^{\text{exp}}(298) - x_i^{\text{calc}}(298, \Omega)}{x_i^{\text{exp}}(298)} \right]^2 \\
+ w_3 \sum_{C_5} \sum_{q=1}^{N_{y_j}} \left[ \frac{y_j^{\text{exp}}(T_q) - y_j^{\text{calc}}(T_q, \Omega)}{y_j^{\text{exp}}(T_q)} \right]^2,
$$

in which sums over the square of the relative residuals between the experimental (exp.) and calculated (calc.) equilibrium mole fraction of the alkane in the water-rich (liquid) phase $x_i(T)$ and $x_i(298)$ (water-rich phase compositions at $T = 298.00$ K) and alkane-rich (liquid) phase $y_j(T)$ of a given mixture at specified values of temperature (including $y_j(298)$) over all experimental points, where $i$ denotes n-alkane and $j$ denotes water. $N_{x_i} = 15$, $N_{y_j} = 21$, and weighting factors $w_1 = 5$, $w_2 = 10$ and $w_3 = 1$ are used. The minimisation is performed using the commercial software package gPROMS [251]. A multistart gradient-based optimisation algorithm (HELD algorithm) [252, 253] is used as input to local optimisations.

The optimal values for the unlike group interaction parameters between $\text{CH}_3 - \text{H}_2\text{O}$ and $\text{CH}_2 - \text{H}_2\text{O}$ groups are summarised in Table 3.2 and the SAFT-\(\gamma\) Mie descriptions of the three-phase equilibria (VLLIE) of the n-alkane + water mixtures are illustrated in Fig. 3.5. Though the optimal unlike dispersion energy parameters determined are relatively high ($\varepsilon_{\text{CH}_3-\text{H}_2\text{O}}/k_B = 358.18$ K and $\varepsilon_{\text{CH}_2-\text{H}_2\text{O}}/k_B = 423.63$ K), the two unlike repulsive exponents are also very large ($\lambda_{\text{CH}_3-\text{H}_2\text{O}}^r = 100.00$ and $\lambda_{\text{CH}_2-\text{H}_2\text{O}}^r = 100.00$) which conforms with the hydrophobic nature of the interactions between alkyl groups and water. Despite the use of a small number of mutual
solubility data points, the model can describe the solubilities of the shortest alkane (ethane) to the longest alkane considered (n-decane) accurately over a wide range of temperature, as shown in Fig. 3.5. Since the solubility of water in the alkane-rich phase is not a strong function of alkane length, only the solubilities of water in the n-pentane-rich and n-decane-rich phases are shown in Fig. 3.5 as representative of the accuracy of the model. While the description obtained from the solubility of water in the alkane-rich phase is not as accurate as the description of the solubility of the n-alkanes in the water-rich phase, the overall results are very satisfactory. The solubilities of ethane to n-decane in water span over six orders of magnitude in mole fraction over a range of temperatures. It is impressive that our group-contribution method can provide such an accurate description of this behaviour with only a few adjustable parameters.

**Figure 3.5:** Solubilities of water in the n-alkane-rich liquid phase (dashed curves) and alkanes (C_2-C_{10}) in the water-rich liquid phase (continuous curves) at conditions of three-phase equilibria as a function of temperature. The diamonds and squares denote experimental correlated data for solubilities of water in n-pentane-rich and n-decane-rich phases, respectively. The circles correspond to experimental and experimental correlated data for various alkane solubilities in water, ranging from ethane to n-decane. The filled symbols represent the mutual solubility data employed in the determination of the unlike CH_3 - H_2O and CH_2 - H_2O interaction parameters. The data for ethane + water, propane + water and n-butane + water are from Mokraoui et al [269]. The data for n-pentane + water to n-decane + water are from the IUPAC-NIST solubility data series [263–268].

The aqueous solubilities of various n-alkanes at 298.00 K are plotted separately in Fig. 3.6. The SAFT-γ Mic predictions are found to be in good agreement with experimental [263–270] and simulation data [196] with the exception of a number of data [271] for the longer alkanes.
(carbon number > 11). Measured data by Sutton et al. [271] exhibit a plateau in the \( n \)-alkanes solubilities for alkanes longer than \( n \)-dodecane, while in the more recent study of Tolls et al. [270] a decrease in the values of solubilities for all \( n \)-alkane studied (up to \( n \)-pentadecane) is presented. Tsonopoulos [191] had previously rationalised the existence of the solubility plateau as related to a “collapsed” conformation of the long alkanes that reduces the contact of the alkanes with water and results in a lower solubility decrease with carbon number. The measurement of solubility of highly hydrophobic compounds in water is notoriously difficult given the working solubility range of a part per billion or trillion; this could explain the uncertainty in the data. In this context theoretical studies can be useful in helping validate experimental data. Ferguson et al. [196] have performed molecular dynamics simulations of the solubility of \( n \)-alkanes in water. They observed no sharp break in the dependence of the solubility on increasing carbon number (cf. Fig. 3.6) in agreement with our SAFT-\( \gamma \) Mie calculations. The alkane solubility calculations using previously reported SAFT-\( \gamma \) Mie parameters [169] are also shown in Fig. 3.6 for reference. It is clear that the new model proposed here provides more accurate predictions.

Figure 3.6: Solubilities of \( n \)-alkanes in the water-rich liquid phase at conditions of three-phase equilibria at \( T = 298.00 \) K. The filled circles represent the experimental correlated data from IUPAC-NIST [263–268]; the circles [269], triangles [270] and diamonds [271] the experimental data; the crosses [196] the MD Simulation data; the continuous curve the SAFT-\( \gamma \) Mie calculations using parameters reported in Tables 3.1 and 3.2, and the dash curve the SAFT-\( \gamma \) Mie calculations using previously reported parameters [169].

The optimised CH\(_3\) – H\(_2\)O and CH\(_2\) – H\(_2\)O models are further assessed for transferability by predicting the fluid-phase behaviour of \( n \)-alkane + water binary mixtures over a wide range of conditions. Constant temperature pressure-composition slices for three mixtures, namely,
n-butane + water (at $T = 477.59$ K) [272], n-hexane + water (at $T = 473.15$ K) [273–275] and n-hexadecane + water (at $T = 523.15$ K) [274, 276] are presented in Figs. 3.7a to 3.7c), respectively. Overall, the predictions are very good in both the VLE and LLE regions for all systems examined. We emphasise that the simultaneous description of both the VLE and LLE of a system is considered a difficult challenge in the modelling of these systems. The ability of the method to describe accurately the high-temperature/pressure fluid-phase behaviour of these systems demonstrates the wide range of reliable applicability of SAFT-$\gamma$ Mie compared to other predictive approaches. It is especially noteworthy that no n-hexadecane data were included in the regression of the group parameters. As can be seen in Fig. 3.7c), the SAFT-$\gamma$ Mie method is shown to predict the fluid-phase behaviour of the n-hexadecane aqueous mixture [274, 276] with good accuracy. It is apparent that the newly estimated unlike group interaction parameters allow for an accurate description of all types of fluid phase equilibria considered (VLE, LLE and VLLE) over a wide range of thermodynamic conditions. In the following section, these parameters are transferred to study the more complex 1-alcohol + water mixtures.
Figure 3.7: Isothermal pressure–mole fraction (P–x) phase diagrams of (a) n-butane + water mixture at a temperature of $T = 477.59$ K [272], above the critical point of n-butane ($T_{c,C_4H_{10}} = 425.12$ K) [277]; (b) n-hexane + water mixture at a temperature of $T = 473.15$ K (VLE [273] and LLE) [274, 275]; (c) n-hexadecane + water mixture at a temperature of $T = 523.15$ K (VLE [276] and LLE) [274]. The symbols represent the experimental data, the continuous curves the prediction with the SAFT–γ Mie approach, the horizontal lines the calculated three phase line, the inset image in (a) corresponds to a magnified view of the water-rich phase, and the inset images in (b) and (c) correspond to a magnified view of the VLE region.

3.3.3 CH$_2$OH - H$_2$O binary interaction parameters

The addition of hydroxyl functional groups to the alkyl chain to form alcohols results in dramatic changes in the physical properties of the systems. Weak attractive van der Waals interactions dominate the phase behaviour of the very hydrophobic alkanes and water, while in the case of aqueous mixtures of alcohols strong attractive unlike interactions mediated by hydrogen bonding are also important. While all n-alkanes are markedly immiscible, a number of alcohols are fully miscible in water. In particular, aqueous mixtures of the shorter homologues of the 1-alcohol
series, i.e., methanol, ethanol, and 1-propanol [278] exhibit homogeneous liquid phases. Liquid-liquid immiscibility regions appear for aqueous solutions of longer chains, i.e., in 1-butanol and longer alcohol aqueous mixtures [278]. Modelling alcohol + water systems is challenging as the fluid-phase behaviour is determined by the relative magnitudes of the unlike dispersion energy and hydrogen bonding interactions. The simultaneous description of the VLE and LLE phase equilibria with a set of transferable parameters constitutes a stringent test of any model.

To complete the development of a model to treat alcohol + water mixtures, unlike-interaction parameters between the CH$_2$OH and H$_2$O groups are developed building on the CH$_3$ – H$_2$O and CH$_2$ – H$_2$O parameters from the previous section. Both CH$_2$OH and H$_2$O are associating groups, modelled with 3B and 4C association schemes, respectively, in the notation of Huang and Rapadoz [255]. The unlike association interactions between the two groups are assumed to be asymmetric, i.e., $\varepsilon^{HB}_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1}\neq \varepsilon^{HB}_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1}$ and $K_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1} \neq K_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1}$ (cf. Fig. 3.8). Here we allow the unlike dispersion energy ($\varepsilon_{\text{CH}_2\text{OH} - \text{H}_2\text{O}}$), the unlike association energies ($\varepsilon^{HB}_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1}$ and $\varepsilon^{HB}_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1}$) and the unlike bonding volumes ($K_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1}$ and $K_{\text{CH}_2\text{OH} - \text{H}_2\text{O},e_1}$) to be adjusted by comparison to 1-octanol + water LLE data ($N_{\text{x\text{oct}}} = 8$ and $N_{\text{y\text{oct}}} = 8$) at $p = 0.101$ MPa [90]. The unlike segment diameter ($\sigma_{\text{CH}_2\text{OH} - \text{H}_2\text{O}}$), the unlike repulsive range ($\lambda^r_{\text{CH}_2\text{OH} - \text{H}_2\text{O}}$) and the unlike attractive range ($\lambda^a_{\text{CH}_2\text{OH} - \text{H}_2\text{O}}$) are obtained through combining rules (Eqs. 3.20 and 3.23). The corresponding objective function is given by:

$$\min_{\Omega} f_{\text{obj}} = w_1 \sum_{q=1}^{N_{x_1}} \left[ \frac{x_i^{\text{exp}}(T_q,p_q) - x_i^{\text{calc}}(T_q,p_q;\Omega)}{x_i^{\text{exp}}(T_q,p_q)} \right]^2 + w_2 \sum_{q=1}^{N_{y_1}} \left[ \frac{y_i^{\text{exp}}(T_q,p_q) - y_i^{\text{calc}}(T_q,p_q;\Omega)}{y_i^{\text{exp}}(T_q,p_q)} \right]^2,$$

where the first sum is over the square of the relative residuals between the experimental (exp.) and calculated (calc.) values of the equilibrium mole fractions of the water-rich (liquid) phase $x_i(T,p)$ and the second over the equilibrium mole fractions of the octanol-rich (liquid) phase $y_i(T,p)$ and where $i$ denotes 1-octanol. Here $N_{x_1} = 8$, $N_{y_1} = 8$, $w_1 = 1$ and $w_2 = 20$. 
Figure 3.8: Asymmetric association scheme between CH$_2$OH and H$_2$O groups

The optimal values for the unlike-interaction parameters are summarised in Table 3.2 and the resulting description of the liquid-liquid equilibrium of the binary mixture of 1-octanol + water is displayed in Fig. 3.9a). As can be seen in the figure the unlike-interaction parameters determined result in an excellent description of the data [90, 142]. In addition to the fluid-phase behaviour of the 1-octanol + water binary mixture, the SAFT-γ Mie predictions of the densities of the two coexisting liquid phases are also found to be in excellent agreement with the experimental data [90] (Fig. 3.9b).

Figure 3.9: (a) Isobaric temperature–mole fraction (T–x) phase diagram of the vapour-liquid-liquid equilibria of 1-octanol + water at a pressure of $p = 0.101$ MPa; (b) saturated densities of the water-rich (right) and octanol-rich phase (left). The circles [90] and triangles [142] represent the experimental data, the continuous curves the description with the SAFT-γ Mie approach, and the horizontal line the calculated three phase line. The inset image in figure (a) corresponds to a magnified view of the water-rich boundary.
In Fig. 3.10, two examples of the predictive capabilities of the model are given based on 1-hexanol + water and 1-butanol + water. The SAFT-\( \gamma \) Mie predictions of both binary mixtures are shown to be in good agreement with the experimental fluid phase behaviour of these systems [279–282]. Both the VLE and the LLE regions of the phase envelope are predicted well, together with the location of the three-phase line. The level of accuracy for the alcohol solubilities in the water-rich phase (the magnified regions in Figs. 3.9 and 3.10) is exceptional considering the predictive nature of the calculations. While at ambient conditions the 1-butanol + water exhibits liquid-liquid demixing, the shorter alcohols are completely miscible with water [278]. The complete miscibility of the ethanol + water and 1-propanol + water mixtures is also predicted with our model. This suggests that the right balance of hydrophobic and hydrophilic interactions is captured in the model. We emphasise again that this is achieved while using a unique set of transferable interaction parameters for all systems. The transferability of the model is to a large extent due to the accuracy of the alkyl-water interactions, as well as to the use of the asymmetric association parameters in the description of the \( \text{CH}_2\text{OH} - \text{H}_2\text{O} \) interactions.

**Figure 3.10:** Isobaric temperature–mole fraction (\( T-x \)) phase diagrams of the vapour-liquid-liquid equilibria at \( p = 0.101 \) MPa of (a) 1-hexanol + water (triangles [279] and circles [280]) and (b) 1-butanol + water (triangles [281] and circles [282]). The symbols represent the experimental data, the continuous curves the predictions with the SAFT-\( \gamma \) Mie approach and the horizontal lines the predicted three phase lines. The inset image in figure (a) corresponds to a magnified view of the water-rich boundary.
3.4 Prediction of solvation properties

An important application of the models developed here is the study of the infinite dilution (solvation) properties. To the best of our knowledge, only a few reports are found in the literature which utilise group-contribution approaches for the prediction of different types of fluid phase equilibria (VLE, LLE, and VLLE) over the entire range of composition including the infinite dilution region. Pereda et al. [157–159, 161] used the GCA-EoS model to predict the vapour-liquid, liquid-liquid equilibria and infinite dilution activity coefficients of mixtures containing water, alcohols, and hydrocarbons, although different sets of alkyl-water unlike parameters (H2O-CH3∞ and H2O-CH2∞) were required to deliver accurate predictions of the mutual solubility of water and hydrocarbons. Possani et al. [221] have used the F-SAC model to represent the mutual solubilities as well as the infinite dilution activity coefficients of hydrocarbon + water mixtures, and find good agreement with the experimental data, although it should be noted that experimental $\gamma_{ij}^\infty$ data were used in the calibration of the model parameters. In this chapter, a unique set of SAFT-γ Mie group parameters (cf. Tables 3.1 and 3.2) is used to predict the phase equilibria as well as the infinite dilution properties for the hydrocarbon and alcohol aqueous systems without the need to include the experimental infinite dilution data to calibrate the model parameters or further adjust any parameters.

Several experimental studies [269, 283–285] report the distribution of a solute between water and a gaseous phase in terms of Henry's law constants $K_{H_{i,j}}$. In Fig. 3.11 experimental data of the $K_{H_{i,j}}$ for ethane, $n$-butane, $n$-hexane, and $n$-octane in water over a temperature range are compared with the corresponding SAFT-γ Mie predictions (Sec. 3.2.3). As can be seen in the figure our predictive approach is found to represent correctly the experimental measurements for the different hydrocarbons considered. In addition, in Fig. 3.12 solvation Gibbs free energy $\Delta G_{i,j}^{sol}$ data, which are directly related to the Henry's law constants (cf. Eq. (3.31)), are presented for a series of $n$-alkanes and 1-alcohols in an aqueous environment at $T = 298.15$ K and $p = 0.100$ MPa. The calculations are performed as single-phase calculations, i.e., through Eq. (3.30) with specified $T$, $p$, and $x_i = 10^{-10}$. Calculating $\Delta G_{i,j}^{sol}$ from $K_{H_{i,j}}$ through the phase equilibrium at $T = 298.15$ K and $p = 0.100$ MPa leads to the same values as when single-phase calculations are carried out in the case of the solutions of $n$-alkanes and 1-alcohols with carbon numbers $\geq 7$ since the coexisting composition of the solutes in these mixtures are effectively of the order considered as the dilute limit. In the case of the shorter alcohols, however, the coexisting composition of the alcohol in the aqueous environment is far from the dilute limit at the specified $T$ and $p$ and therefore the single-phase calculation route in which the composition of the solute is specified.
explicitly is used. As expected from the hydrophobic nature of the \( n \)-alkanes, large positive values are seen for the \( \Delta G^\text{sol}_{i,j} \) of these molecules (Fig. 3.12a), and we note that the values increase as the chain length of the hydrocarbon increases. On the other hand, the hydration of alcohols is favourable, with corresponding negative values of \( \Delta G^\text{sol}_{i,j} \) (Fig. 3.12b). The predicted \( \Delta G^\text{sol}_{i,j} \) of the \( n \)-alkanes (Fig 3.12a) are in excellent agreement with the experimental data [286], especially for ethane to \( n \)-undecane, with slightly larger deviations observed for the longer \( n \)-alkanes. The level of agreement seen here is consistent with the solubility predictions of \( n \)-alkanes in water shown earlier in Fig. 3.6. The Gibbs free energy of hydration is obtained from experimental aqueous solubility data (or Henry’s law constants or activity coefficients of a solute in aqueous solution at infinite dilution) [286]. The larger uncertainties reported in \( \Delta G^\text{sol}_{i,j} \) data [286] for the longer alkanes are due to the higher uncertainty in the solubility measurements of these highly hydrophobic compounds. While our solvation Gibbs free energy predictions for the \( n \)-alkanes are overall in good agreement with the experimental data [286] (Fig. 3.12a), a larger deviation is observed for the case of the shorter alcohols (Fig. 3.12b). In these small polar molecules, due to proximity effects, the assumption of the transferability of group parameters is less applicable.

![Figure 3.11](image_url): Henry’s law constants \( K_{H_{i,j}} \) for hydrocarbons in water as a function of temperature at the corresponding vapour pressure of the mixture. The symbols represent the experimental data and the continuous curves the predictions with the SAFT-\( \gamma \) Mie approach: ethane (triangles [283] and diamonds [284]), \( n \)-butane (circles) [269], \( n \)-hexane (squares) [269] and \( n \)-octane (pentagons) [285].
Figure 3.12: Gibbs free energies of solvation $\Delta G_{i,j}^{\text{sol}}$ for (a) $n$-alkanes and (b) 1-alcohols in water at $T = 298.15$ K and $p = 0.100$ MPa. The circles [286] represent the experimental data, the error bars the standard deviations, and the continuous curves the predictions with the SAFT-$\gamma$ Mie approach.

The infinite dilution activity coefficient $\gamma_{i,j}^{\infty}$ provides a measure of solvation with reference to the fugacity coefficient of the pure solute (Eq. (3.32)). It describes the behaviour of a solute $i$ entirely surrounded by solvent $j$ and reflects the maximum deviation from the ideal solution behaviour. Reliable information on $\gamma_{i,j}^{\infty}$ is of importance in predicting the fate of chemicals in the environment [287]. In Figs. 3.13 and 3.14, the $\gamma_{i,j}^{\infty}$ of $n$-alkanes and 1-alcohols, respectively, in water predicted with the SAFT-$\gamma$ Mie model are seen to be in good agreement with experimental data [270, 288]. Some deterioration of the agreement is seen for the longer $n$-alkanes (carbon number $> 11$ [271]), though we note there are discrepancies among different sets of experimental data [271] and simulation results [196]. As in the case of the measurement of $\Delta G_{i,j}^{\text{sol}}$, the $\gamma_{i,j}^{\infty}$ data for long $n$-alkanes are commonly obtained from aqueous solubility data [270, 271] (cf. Fig. 3.6), which can be extremely difficult to measure. Therefore, the same argument used to explain the disagreement between the experimental and theoretical values of the aqueous solubility and $\Delta G_{i,j}^{\text{sol}}$ of the longer $n$-alkanes can be applied here as well. The quantitative agreement with most of the experimental data and with the simulation data suggests that our theoretical approach may be used to validate the experimental measurements for the longer $n$-alkanes. In Fig. 3.14 predicted values of $\gamma_{i,j}^{\infty}$ of $n$-alkanes and 1-alcohols at different temperatures are compared with experimental data. As can be seen very good agreement with the experimental data is observed for the entire temperature range available for alkanes and alcohols larger than 1-butanol. We have only considered 1-alcohols larger than 1-butanol since, as was pointed out earlier, our group-contribution method is not as accurate for the shorter alcohols. The level of agreement
observed for the longer alcohols suggests that our SAFT-\(\gamma\) Mie model should also lead to accurate predictions of \(K_{i,OW}\) given the relation between the two thermodynamic quantities (Eq. (2.4)). The \(K_{i,OW}\) prediction of alkanes, alcohols, and other families of organic compounds is the main focus for Chapter 5.

**Figure 3.13:** Infinite dilution activity coefficients \(\gamma_{i,j}^{\infty}\) of \(n\)-alkanes (top line) and 1-alcohols (bottom line) in aqueous solution at \(T = 298.15\) K and \(p = 0.101\) MPa (except for ethane-butane, where the calculations are done at 5.00 MPa). The circles [288], triangles [270], diamonds [271] and squares [288] represent the experimental data, the crosses [196] the MD simulation data, and the continuous curves the predictions with the SAFT-\(\gamma\) Mie approach.

**Figure 3.14:** Infinite dilution activity coefficients \(\gamma_{i,j}^{\infty}\) of \(n\)-alkanes (hexane and pentane) [269] and 1-alcohols (octanol [289], heptanol–butanol [290]) in aqueous solution at \(p = 0.101\) MPa (except for \(n\)-pentane and \(n\)-hexane, where the calculations were carried out at 7.00 MPa). The symbols represent the experimental data and the continuous curves the predictions with the SAFT-\(\gamma\) Mie approach.
3.5 Conclusions

Modelling the phase equilibria of alkane and alcohol aqueous mixtures using theoretically sound models especially in a group-contribution (GC) framework remains a challenge due to the non-idealality and complexity of interactions that the mixtures exhibit. For design purposes, a thermodynamic model should ideally be able to predict quantitatively the phase behaviour of the systems of interest over a range of thermodynamic conditions and the entire composition range, including the infinite dilution region, with a single set of parameters. In this chapter, the SAFT-$\gamma$ Mie group-contribution approach is used to predict vapour-liquid and liquid-liquid equilibria of mixtures containing $n$-alkanes, 1-alcohols, and water using a single set of group-group model parameters. New interaction parameters for the description of the family of 1-alcohols by means of a CH$_2$OH functional group are obtained, and revised unlike parameters between the alkyl CH$_3$ and CH$_2$ groups and the H$_2$O group used for water are reported. The predictive ability of the model is confirmed in the comparison of calculations of the fluid phase equilibria of $n$-alkane + 1-alcohol and $n$-alkane + water mixtures with experimental data. Alkane solubilities in the water-rich phase at conditions of three-phase coexistence are predicted for long $n$-alkanes in good agreement with simulation data for concentrations as low as $10^{-14}$ in mole fraction of the alkane for the larger chains (C$_{22}$). This is an interesting result that highlights the difficulty in measuring these extremely low concentrations. New model parameters characterising the CH$_2$OH - H$_2$O unlike interaction have also been determined and used to predict the phase behaviour of a number of alcohol + water binary mixtures that are not used in the development of the model. Our SAFT-$\gamma$ Mie models provide good predictions for several $n$-alkane and 1-alcohol aqueous mixtures over a wide range of thermodynamic conditions, including VLE, LLE, and VLLE. The robustness of the group parameters is further demonstrated in the prediction of several infinite dilution properties. The Henry’s law constants, solvation Gibbs free energies and infinite dilution activity coefficients of $n$-alkanes and 1-alcohols in water are predicted and found to be in good agreement with the experimental data. The findings in this study serve as a good basis for the development of more complex functional groups which will ultimately lead to the modelling of complex molecules of relevance to pharmaceutical systems.
Chapter 4

Development of the SAFT-$\gamma$ Mie group interaction parameters characterising pharmaceutical molecules

4.1 Introduction

In the work presented thus far, the performance of the SAFT-$\gamma$ Mie approach in the accurate description of the thermodynamic properties and fluid-phase equilibria of complex mixtures containing alkane, alcohol, and water has been discussed. In this chapter a development of further group interaction parameters that are commonly found in pharmaceutical molecules is demonstrated.

The study of the properties of active pharmaceutical ingredients (APIs) using SAFT-$\gamma$ Mie requires the determination of the relevant group interaction parameters as it is a group-contribution (GC) method. The GC element of the method allows a very broad predictive capability, but this can only be harnessed if the parameters describing each group and group-group interactions are carefully obtained in a physically appropriate manner. The first step in determining these interactions is to decompose the target molecules into functional groups. There is no unique way to define groups. The main consideration in decomposing molecules into groups is the number of atoms per groups: fewer atoms per groups will lead to more general groups, i.e., greater transferability while more atoms per groups will take into account more
effects due to local polarity or conformation of the molecules. This leads to a balance between the transferability and accuracy of the functional groups. An example of the decomposition of ibuprofen molecule into distinct functional groups is given in Fig. 4.1. In the SAFT-\(\gamma\) Mie approach, functional groups are developed in a successive manner. This means that new groups are developed based on previous groups and thus any change in the model cascades to other interactions and leads to the possible re-estimation of several interactions. We refer to this as the “jenga effect” in reference to the tower building game in which the integrity of the tower is compromised when the based building block is pull off. Given that parameter estimation plays a key role in the application of GC approaches, this procedure needs to be developed with care. In the early stages of group development [22], parameters were estimated using almost exclusively vapour-liquid equilibrium (VLE) data, i.e., saturated-liquid densities (\(\rho_{\text{sat}}\)) and vapour pressures (\(p_{\text{vap}}\)) of pure components and coexistence compositions of binary mixtures. As experienced is gained by looking at different groups and properties our parameter estimation procedure is enhanced to include different types of experimental data such as liquid-liquid equilibrium (LLE), vapour-liquid-liquid equilibrium (VLLE), and excess properties of mixing. A summary of functional groups available in the SAFT-\(\gamma\) Mie framework at the time of writing this thesis for the modelling of APIs is presented in Table 4.1

In this chapter, the development for group-group interaction parameters is carried out to complete the ibuprofen + octanol + water interactions based on the pre-determined like and unlike group parameters indicated in Table 4.2. In particular, the unlike group interaction parameters between the branched-alkyl/ aromatic/ COOH groups and the \(\text{H}_2\text{O}\) group, as well as between the branched-alkyl/ aromatic/ COOH groups and the \(\text{CH}_2\text{OH}\) group are obtained. Developing the unlike interaction with \(\text{H}_2\text{O}\) group involves the modelling of aqueous solutions of hydrocarbons which are of great interest in many chemical engineering applications ranging from the petrochemical to biological industries [186, 187, 292]. The thermodynamic modelling
Table 4.1: Groups developed for use within the SAFT-\(\gamma\) Mie approach\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>CH(_3)</th>
<th>CH(_2)</th>
<th>CH</th>
<th>C</th>
<th>aCH</th>
<th>aCCH(_2)</th>
<th>aCCH(_3)</th>
<th>CH=</th>
<th>C=</th>
<th>cCH(_2)</th>
<th>COOH</th>
<th>COO</th>
<th>H(_2)O</th>
<th>OH</th>
<th>CH(_2)OH</th>
<th>aCCOaC</th>
<th>CH(_3)CO(_2)H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)The published parameters and the combining rules can be found in [22, 169, 239, 240, 291].

of branched alkanes and aromatic aqueous mixtures, similar to the alkane and alcohol aqueous systems discussed in Chapter 3, is particularly challenging due to the highly non-ideal behaviour that the mixtures exhibit over a wide range of thermodynamic conditions. Characteristics of the fluid phase behaviour of these systems include heterogeneous azeotropes, bounding regions of vapour–liquid and liquid–liquid equilibria, highly asymmetric mutual solubilities, and a minimum solubility in the water-rich phase around ambient temperature. These extreme natures of the fluid phase behaviour is a direct consequence of the strong hydrogen bonding interactions between the water molecules. Therefore, thermodynamic tools successful in the description of aqueous solutions are mainly the more sophisticated methods that explicitly account for the effects of association (\textit{e.g.}, variants of SAFT [78, 79] and CPA EoSs [107]). For instance, Dong \textit{et al.} have used the GC polar perturbed-chain SAFT EoS to describe the VLE and LLE of alkyl benzenes + water binary mixtures [165] as well as the more complex mixtures of alkyl phenols + water [293]. The CPA [162, 212] and GCA (group contribution with association) EoSs [157, 160, 161] have also been used to model the aqueous mixtures of branched alkanes and aromatic compounds. Note that different sets of binary interaction parameters between the alkyl group - water (CH\(_3\)\(^\infty\)-H\(_2\)O and CH\(_2\)\(^\infty\)-H\(_2\)O) and aromatic group - water (\textit{e.g.}, aCH\(^\infty\)-H\(_2\)O)
[157, 161] were used in the GCA model to obtain an accurate description of the mutual solubility between water and hydrocarbons. The authors explained that the environment of the alkyl groups in a paraffin + water mixture is completely different from that, for instance, in an alcohol + water solution. A recent review on the application of molecular-based equations of state for water and aqueous solutions focusing on the application of SAFT, CPA, and GCA EoSs can be found in work of Vega and Llorell [204].

The modelling of alkanolic acids + water mixtures is also of great theoretical and industrial interest as they serve as important commodities with an increasingly wide range of industrial applications. Most theoretical studies of these systems also use EoSs with an explicit treatment of association. The CPA EoS has been used to study the VLE of short acids (up to propanoic acid) + water mixture [294, 295] and to correlate the LLE of the longer chains (pentanoic to decanoic acids) + water [296]. The GCA EoS has also been used for the modelling of VLE for short-chain acids with water and alcohols [297]. Applications of the SAFT and CPA methods to organic acids mixtures have been summarised in the work of Breil et al. [295]. To our knowledge, no group-contribution approach has been employed to study both VLE and LLE of carboxylic acid + water mixtures simultaneously.

In developing the CH$_2$OH unlike interactions, a number of alcohol mixtures of aromatic compounds and alkanolic acids are involved. Pure component data for compounds pertaining the CH$_2$OH group and the group of interest, e.g. CH and C groups, can also be used to obtain the interaction parameters taking advantage of the heteronuclear formulation of the SAFT-$\gamma$ theory. Modelling alcohol mixtures can be challenging as they can exhibit very interesting non-ideal phase behaviour because of the strong and highly directional nature of their hydrogen bonding interaction. Especially with the alcohol + alkanolic acid mixtures, the interaction can be complicated given the presence of the two strongly associating functional groups in the mixtures. McCabe et al. have used the GC-SAFT-VR approach [125] (a heteronuclear GC SAFT based on a square-well potential) to treat mixtures of 1-butanol + butanoic acid [298] and several alcohols + alkyl benzenes [299] with good accuracy. The alcohol + alkyl benzene mixtures have also been studied using a GCA EoS [157] and the alcohol + propanoic acid mixtures using a CPA EoS [294].
**Table 4.2:** Like group parameters for use in the SAFT-\(\gamma\) Mie group-contribution approach: \(n_k\) is the number of segments constituting group \(k\), \(S_k\) the shape factor, \(\lambda_k^1\) the attractive exponent, \(\lambda_k^2\) the repulsive exponent, \(\sigma_{kk}\) the segment diameter of group \(k\), and \(\varepsilon_{kk}\) the dispersion energy of the Mie potential characterising the interaction of two \(k\) groups; \(N_{ST,k}\) represents the number of association site types on group \(k\), with \(n_{k,H}\), \(n_{k,e_1}\), and \(n_{k,e_2}\) denoting the number of association sites of type H, \(e_1\), and \(e_2\) respectively. Unpublished\(\dagger\) parameters can be found in Dufal and Papaioannou personal communications.

<table>
<thead>
<tr>
<th>(k)</th>
<th>Group (k)</th>
<th>(n_k)</th>
<th>(S_k)</th>
<th>(\lambda_k^1)</th>
<th>(\lambda_k^2)</th>
<th>(\sigma_{kk}/\AA)</th>
<th>(\varepsilon_{kk}/k_B)/K</th>
<th>(N_{ST,k})</th>
<th>(n_{k,H})</th>
<th>(n_{k,e_1})</th>
<th>(n_{k,e_2})</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CH(_3)</td>
<td>1</td>
<td>0.57255</td>
<td>15.050</td>
<td>6.0000</td>
<td>4.0773</td>
<td>256.77</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>[22]</td>
</tr>
<tr>
<td>2</td>
<td>CH(_2)</td>
<td>1</td>
<td>0.22932</td>
<td>19.871</td>
<td>6.0000</td>
<td>4.8801</td>
<td>473.39</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>[22]</td>
</tr>
<tr>
<td>3</td>
<td>CH</td>
<td>1</td>
<td>0.07210</td>
<td>8.0000</td>
<td>6.0000</td>
<td>5.2950</td>
<td>95.621</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>[169]</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>1</td>
<td>0.04072</td>
<td>8.0000</td>
<td>6.0000</td>
<td>5.6571</td>
<td>50.020</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>[169]</td>
</tr>
<tr>
<td>5</td>
<td>aCH</td>
<td>1</td>
<td>0.32184</td>
<td>14.756</td>
<td>6.0000</td>
<td>4.0578</td>
<td>371.53</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>[169]</td>
</tr>
<tr>
<td>6</td>
<td>aCCH(_2)</td>
<td>1</td>
<td>0.20859</td>
<td>8.5433</td>
<td>6.0000</td>
<td>5.2648</td>
<td>591.56</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>[169]</td>
</tr>
<tr>
<td>7</td>
<td>aCCH(_3)</td>
<td>1</td>
<td>0.31655</td>
<td>23.627</td>
<td>6.0000</td>
<td>5.4874</td>
<td>651.41</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>[291]</td>
</tr>
<tr>
<td>8</td>
<td>aCH</td>
<td>1</td>
<td>0.20650</td>
<td>8.0000</td>
<td>6.0000</td>
<td>4.3128</td>
<td>61.325</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>[169]</td>
</tr>
<tr>
<td>9</td>
<td>CH(_2=)</td>
<td>1</td>
<td>0.4489</td>
<td>20.271</td>
<td>6.0000</td>
<td>4.3175</td>
<td>300.90</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>[169]</td>
</tr>
<tr>
<td>10</td>
<td>CH(=)</td>
<td>1</td>
<td>0.2004</td>
<td>15.974</td>
<td>6.0000</td>
<td>4.7488</td>
<td>952.54</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>[169]</td>
</tr>
<tr>
<td>11</td>
<td>C=</td>
<td>1</td>
<td>0.1533</td>
<td>8.0000</td>
<td>6.0000</td>
<td>4.0330</td>
<td>1500.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>unpublished(\dagger)</td>
</tr>
<tr>
<td>12</td>
<td>cCH(_2)</td>
<td>1</td>
<td>0.2475</td>
<td>20.386</td>
<td>6.0000</td>
<td>4.7852</td>
<td>477.36</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>[169]</td>
</tr>
<tr>
<td>13</td>
<td>COOH</td>
<td>1</td>
<td>0.5593</td>
<td>8.0000</td>
<td>6.0000</td>
<td>4.3331</td>
<td>405.78</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>[240]</td>
</tr>
<tr>
<td>14</td>
<td>COO</td>
<td>1</td>
<td>0.6526</td>
<td>31.189</td>
<td>6.0000</td>
<td>3.9939</td>
<td>868.92</td>
<td>1</td>
<td>–</td>
<td>2</td>
<td>–</td>
<td>[169]</td>
</tr>
<tr>
<td>15</td>
<td>H(_2)O</td>
<td>1</td>
<td>1.0000</td>
<td>17.020</td>
<td>6.0000</td>
<td>3.0063</td>
<td>266.68</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>[239]</td>
</tr>
<tr>
<td>16</td>
<td>OH</td>
<td>1</td>
<td>0.9634</td>
<td>20.702</td>
<td>6.0000</td>
<td>2.7990</td>
<td>410.31</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>unpublished(\dagger)</td>
</tr>
<tr>
<td>17</td>
<td>CH(_2)OH</td>
<td>2</td>
<td>0.58538</td>
<td>22.699</td>
<td>6.0000</td>
<td>3.4054</td>
<td>407.22</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>this work</td>
</tr>
<tr>
<td>18</td>
<td>aCCO(_2)C</td>
<td>3</td>
<td>0.18086</td>
<td>9.8317</td>
<td>6.0000</td>
<td>4.0670</td>
<td>656.71</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>unpublished(\dagger)</td>
</tr>
<tr>
<td>19</td>
<td>CH(_2)COCH(_3)</td>
<td>3</td>
<td>0.72135</td>
<td>17.433</td>
<td>6.0000</td>
<td>3.5981</td>
<td>286.02</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>[240]</td>
</tr>
</tbody>
</table>

### 4.2 Parameter estimation of functional groups

In this chapter, the unlike group-group interaction parameters needed to model the systems of interest are estimated using experimental data for the fluid-phase equilibria of appropriate pure components and mixtures. The parameter estimation follows the procedure described in Chapter 3 which involves a minimisation of a least-squares objective function consisting of an appropriate sum of individual residuals. The minimisation is performed using a multistart gradient-based optimisation algorithm (HELD algorithm) [252, 253] or the numerical solvers available in the
gPROMS software package [251]. The general form of the objective function \( f_{\text{obj}} \) for mixture data is given by:

\[
\min_{\Omega} f_{\text{obj}} = w_1 \sum_{q=1}^{N_x} \left[ \frac{\left( x_i^{\text{exp}}(T_q, p_q) - x_i^{\text{calc}}(T_q, p_q; \Omega) \right)}{x_i^{\text{exp}}(T_q, p_q)} \right]^2 + w_2 \sum_{u=1}^{N_y} \left[ \frac{\left( y_i^{\text{exp}}(T_u, p_u) - y_i^{\text{calc}}(T_u, p_u; \Omega) \right)}{y_i^{\text{exp}}(T_u, p_u)} \right]^2,
\]

(4.1)

where the first sum is over the square of the relative residuals between the experimental (exp.) and calculated (calc.) values of the equilibrium mole fractions of the first phase \( x_i(T, p) \) over the \( N_x \) mole fraction data points with the weighting factors \( w_1 \) and the second over the equilibrium mole fractions of the second phase \( y_i(T, p) \) over the \( N_y \) mole fraction data points with the weighting factors \( w_2 \). Note that when the three-phase (VLE) solubility data are used, the coexisting liquid compositions are a function of temperature only, i.e., \( x_i(T) \) and \( y_i(T) \); see Eq. (3.35). The objective function for pure-component data is given by:

\[
\min_{\Omega} f_{\text{obj}} = w_1 \sum_{i=1}^{N_C} \sum_{q=1}^{N_{\text{vap},i}} \left[ \frac{\left( \rho_{\text{vap},i}(T_q) - \rho_{\text{vap},i}^{\text{calc}}(T_q; \Omega) \right)}{\rho_{\text{vap},i}(T_q)} \right]^2 + w_2 \sum_{i=1}^{N_C} \sum_{q=1}^{N_{\text{sat},i}} \left[ \frac{\left( \rho_{\text{sat},i}(T_q) - \rho_{\text{sat},i}^{\text{calc}}(T_q; \Omega) \right)}{\rho_{\text{sat},i}(T_q)} \right]^2,
\]

(4.2)

where the two sums are over the \( N_C \) pure components included in the estimation and over the number \( N_{\text{vap},i} \) of experimental vapour pressure points for compound \( i \) with the weighting factor \( w_1 \) and \( N_{\text{sat},i} \) saturated liquid density points for compound \( i \) with the weighting factor \( w_2 \).

The accuracy of the SAFT-\( \gamma \) Mie methodology for the description of the pure component properties of the molecules of interest here is evaluated using a percentage average absolute deviation (\( \%\text{AAD} \)) defined as:

\[
\%\text{AAD} \, R_i = \frac{1}{N_{R_i}} \sum_{j=1}^{N_{R_i}} \left| \frac{R_{i,j}^{\text{exp}} - R_{i,j}^{\text{calc}}}{R_{i,j}^{\text{exp}}} \right| \times 100, \quad R_1 = \rho_{\text{vap}}, \quad R_2 = \rho_{\text{sat}},
\]

(4.3)

where \( N_{R_i} \) is the number of data points of property \( R_i \), \( R_{i,j}^{\text{exp}} \) the experimental value of \( R_i \) at the \( j^{\text{th}} \) point, and \( R_{i,j}^{\text{calc}} \) the calculated value for the same property at the conditions of the \( j^{\text{th}} \) experimental point. The average absolute error (AAE), over the \( N_p \) data points, is used to evaluate the infinite dilution activity coefficient \( (\gamma_i^\infty) \) prediction which is defined as:
Chapter 4: Group parameter development

\[ AAE = \frac{1}{N_p} \sum_{i=1}^{N_p} |\gamma_{i,\text{exp}} - \gamma_{i,\text{calc}}| \]  \hspace{1cm} (4.4)

4.3 Obtaining H\textsubscript{2}O unlike group interaction parameters

The unlike interaction parameters between the H\textsubscript{2}O group and the branched-alkyl groups (C and CH), aromatic groups (aCCH, aCCH\textsubscript{2}, and aCCH\textsubscript{3}), and the COOH group are obtained in this section based on the CH\textsubscript{3}/CCH\textsubscript{2} - H\textsubscript{2}O water interactions from the previous chapter. Here the unlike interaction parameters that are regressed to the experimental fluid-phase equilibrium data include the dispersion energy \( \varepsilon_{kl} \), as well as the hydrogen bonding energy \( \varepsilon_{\text{HB}}^{kl,ab} \) and the bonding volume \( K_{kl,ab} \) in the case of two associating groups. The unlike segment diameter \( \sigma_{kl} \), the unlike repulsive range \( \lambda_{kl}^r \) are obtained through combining rules [22] (Eqs. 3.20 and 3.23). The unlike attractive range \( \lambda_{kl}^a \) is kept equal to six. The optimised parameters obtained in this work and those involved in the modelling of APIs are summarised in Tables 4.4 - 4.5.

As water is modelled as a single functional group, obtaining the unlike interactions between water and other functional groups must therefore be done based on available experimental data for the appropriate mixtures. Isoalkanes + water mutual solubilities at three-phase equilibrium conditions are used to obtain the CH - H\textsubscript{2}O dispersion energy \( \varepsilon_{\text{CH-H}_2\text{O}} \). The main mutual solubility data employed for the parameter estimation is the isopentane (2-methylbutane) + water mixture at various temperature [263]. Additionally, the mutual solubility of 3-methylpentane [265], 3-methylhexane [266], 2,4-dimethylpentane [266], and 2,3,4-trimethylpentane [264] with water at \( T = 298.15 \) K are used to ensure the accuracy of the parameter for property prediction at ambient temperature. This approach is similar to that adopted to derive the H\textsubscript{2}O - CH\textsubscript{3} and H\textsubscript{2}O - CH\textsubscript{2} interactions. The description of the isopentane + water mutual solubilities with the SAFT-\( \gamma \) Mie approach is shown in Fig. 4.2a) and the deviations from the fitted solubility values at 298.15 K of every branched alkane + water mixtures used are reported in Table 4.3.

After the CH - H\textsubscript{2}O parameter is established, it is transferred to be used as a basis for the C - H\textsubscript{2}O parameter estimation. The C - H\textsubscript{2}O dispersion energy \( \varepsilon_{\text{C-H}_2\text{O}} \) is regressed to the mutual solubilities at three-phase equilibria of isooctane (2,2,4-trimethylpentane) + water [264] at \( T = 280.15 - 303.15 \) K, and 2,2-dimethylbutane + water [265] and 2,2,5-trimethylhexane + water [267] at \( T = 298.15 \) K. As the aqueous mixtures of 2,2,4-trimethylpentane and 2,2,5-trimethylhexane, which contain both CH and C functional groups, are used for the parameter estimation, the unlike interaction parameter between the CH and C groups is obtained simultaneously along with the C - H\textsubscript{2}O parameter (cf. Table 4.4). The description of the isooctane + water mutual
solubilities and the prediction at higher temperature are shown in Fig. 4.2b) and the deviations from the fitted solubility values at 298.15 K are reported in Table 4.3.

![Graphs showing solubility vs temperature](image)

### Figure 4.2: Mutual solubilities of (a) isopentane (2-methylbutane) [263] and (b) isoctane (2,2,4-trimethylpentane) [264] in the water-rich liquid phase (continuous curves) and water in the alkane-rich phase liquid (dotted curves) at conditions of three-phase equilibria as a function of temperature. The squares are experimental data for the mutual solubilities of water in the alkane-rich phase and the circles are the mutual solubilities of alkane in the water-rich phase. The dotted and continuous curves represent the description with the SAFT-γ Mie approach.

### Table 4.3: Absolute deviation (AD = |R^{exp} - R^{calc}|, R = x_i, y_{water}) and percentage relative deviation (%RD = \frac{|R^{exp} - R^{calc}|}{R^{exp}} \times 100) for solubilities of branched alkanes (i) in water-rich (liquid) phase x_i and water in alkane-rich (liquid) phase y_{water} at three-phase equilibria at 298.15 K used for CH - H_2O and C - H_2O parameter estimation.

<table>
<thead>
<tr>
<th>Mixture: compound i + water</th>
<th>AD x_i</th>
<th>%RD x_i</th>
<th>AD y_{water}</th>
<th>%RD y_{water}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Methylbutane + water</td>
<td>8.3100×10^{-7}</td>
<td>6.7676</td>
<td>3.6016×10^{-4}</td>
<td>54.224</td>
</tr>
<tr>
<td>3-Methylpentane + water</td>
<td>4.7736×10^{-7}</td>
<td>15.112</td>
<td>2.8286×10^{-4}</td>
<td>46.523</td>
</tr>
<tr>
<td>3-Methylhexane + water</td>
<td>1.3717×10^{-7}</td>
<td>20.001</td>
<td>2.5983×10^{-4}</td>
<td>42.748</td>
</tr>
<tr>
<td>2,4-Dimethylpentane + water</td>
<td>7.4938×10^{-8}</td>
<td>10.395</td>
<td>3.0463×10^{-4}</td>
<td>45.913</td>
</tr>
<tr>
<td>2,3,4-Trimethylpentane + water</td>
<td>3.0928×10^{-8}</td>
<td>11.074</td>
<td>1.9910×10^{-4}</td>
<td>33.484</td>
</tr>
<tr>
<td>2,2-dimethylbutane + water</td>
<td>1.1113×10^{-7}</td>
<td>2.7104</td>
<td>2.9846×10^{-4}</td>
<td>42.637</td>
</tr>
<tr>
<td>2,2,4-Trimethylpentane + water</td>
<td>7.6407×10^{-9}</td>
<td>3.0441</td>
<td>2.2405×10^{-4}</td>
<td>33.143</td>
</tr>
<tr>
<td>2,2,5-Trimethylhexane + water</td>
<td>1.3026×10^{-9}</td>
<td>2.2852</td>
<td>1.9431×10^{-4}</td>
<td>29.088</td>
</tr>
</tbody>
</table>
Next, the unlike interaction between aromatic functional groups and water is investigated based on the established group parameters listed in Table 4.2. The aCH - H₂O interactions are taken from the literature [169] which were obtained from regressing the $\varepsilon_{a\text{CH} - \text{H}_2\text{O}}$ and $\lambda_{a\text{CH} - \text{H}_2\text{O}}$ to the mutual solubilities of benzene and water [300]. Although the aromatic hydrocarbons themselves are non self-associating, it is known that they are able to cross associate with water [301, 302]. An $e_1$ site is therefore placed on the aCH group to mediate strong interaction with the H₂O group. A similar association scheme is applied to the modelling of other aromatic groups: aCCH, aCCH₂, and aCCH₃ with water. Given that the association between the aromatic and H₂O groups stems from the $\pi$ system of the phenyl ring, which is composed of the aCH groups, rather than the alkylated part of the group, the hydrogen bonding energies and bonding volumes of aCCH, aCCH₂, and aCCH₃ group with H₂O group are assumed to be equal to the $\varepsilon_{a\text{CH} - \text{H}_2\text{O},e_1\text{H}}^{\text{HB}}$ and $K_{a\text{CH} - \text{H}_2\text{O},e_1\text{H}}$ (Table 4.5). This approximation reduces the complexity of the aromatic - H₂O parameter estimation by leaving only the $\varepsilon_{kl}$ to be regressed with the experimental data. The mutual solubilities at three-phase equilibria of toluene + water and 1,2,4-trimethylbenzene + water are used to obtain $\varepsilon_{a\text{CCH}_3 - \text{H}_2\text{O}}$, those of ethylbenzene + water and butylbenzene + water to obtain $\varepsilon_{a\text{CCH}_2 - \text{H}_2\text{O}}$, and those of cumene + water to obtain $\varepsilon_{a\text{CCH} - \text{H}_2\text{O}}$. The SAFT-$\gamma$ Mie descriptions of the mutual solubilities are illustrated in Figs. 4.3a) - 4.3c). The mutual solubilities of different trimethylbenzene isomers are also shown in Fig. 4.3a). The experimental data [303] indicate a slight difference in the aqueous phase solubilities of 1,2,3-trimethylbenzene, 1,2,4-trimethylbenzene, and 1,3,5-trimethylbenzene while the SAFT-$\gamma$ Mie method results in the same description for all isomers. Overall, the model provides a good description of the mutual solubilities of water and all alkyl benzenes studied in both phases.
Figure 4.3: Mutual solubilities of water and alkyl benzenes [303] at conditions of three-phase equilibria as a function of temperature. The alkyl benzenes include (a) methylbenzene or toluene (squares), 1,2,3-trimethylbenzene (circles, not used in fitting), 1,2,4-trimethylbenzene (triangles), and 1,3,5-trimethylbenzene (diamonds, not used in fitting); (b) ethylbenzene (squares) and butylbenzene (circles); (c) (1-methylethyl)benzene or cumene (circles). The continuous curves represent the mutual solubilities of alkyl benzenes in the water-rich liquid phase and the dotted curves the mutual solubilities of water in the hydrocarbon-rich liquid phase.

The carboxyl (COOH) functional group is the most commonly occurring functional group conferring acidity to drug molecules including ibuprofen. In spite of its importance in the agrochemical and pharmaceutical industries, the experimental data for the mutual solubilities of carboxylic acid + water mixtures are scarce. Theoretical studies of phase equilibria of the mixtures are also limited. Most studies focus on the acetic acid mixture due to its industrial importance as both a final product and an intermediate in the production of many other chemicals [295]. To our knowledge, no group-contribution approach has been employed to study both VLE and LLE of carboxylic acid + water mixtures simultaneously. Therefore the ambitious task of being able to model the VLE of short acids + water and LLE of longer chains + water
with a single GC model is one of the targets of our acid-water model.

In the SAFT-\(\gamma\) Mie model for carboxylic acids, an H site on the COOH functional group is used to mediate the dimerisation with another COOH group via its H site, cf. Fig. 4.4. For carboxylic acid + water mixtures, two of each additional electronic site type \(e_1\) and \(e_2\) on the COOH group are switched on to mediate the hydrogen bonding interaction with water. The unlike association interactions between COOH and \(\text{H}_2\text{O}\) segments are asymmetric, e.g., \(\varepsilon_{\text{COOH-H}_2\text{OH},e_1\text{H}} \neq \varepsilon_{\text{COOH-H}_2\text{OH},e_2\text{H}}\) (Fig. 4.4). The optimising unlike interaction parameters between COOH - \(\text{H}_2\text{O}\) include one dispersion energy (\(\varepsilon\)), three hydrogen bonding association energies (\(\varepsilon_{\text{He}_1}, \varepsilon_{\text{e}_1\text{H}}, \text{ and } \varepsilon_{\text{e}_2\text{H}}\)), and three bonding volume parameters (\(K_{\text{He}_1}^{\text{HB}}, K_{\text{e}_1\text{H}}^{\text{HB}}, \text{ and } K_{\text{e}_2\text{H}}^{\text{HB}}\)). It is noted that these COOH - \(\text{H}_2\text{O}\) unlike interaction parameters have been reported in the literature [240] based on a previous alkane-water model [169]. In this chapter, we revise these parameters based on the recently developed alkane-water model shown in Chapter 3. Butanoic acid + water VLE, hexanoic acid + water, and octanoic acid + water LLE data are used to regress these parameters; the optimised values are reported in Tables 4.4 and 4.5. The description of fluid-phase behaviour of the \(n\)-alkanoic acids + water mixtures is shown in Fig. 4.5. As seen in the figure, the SAFT-\(\gamma\) Mie method is able to describe reasonably both VLE and LLE of the different mixtures which is a very challenging task for any predictive methods, despite a significant deviation from the LLE experimental data at higher temperatures. The deviation suggests that the optimised interaction between acid and water is too weak at high temperature. To resolve this problem, temperature-dependent interaction parameters can be used to account for a great difference in water-acid interactions at different temperatures. This \(n\)-alkanoic acid-water model is especially accurate in the description of hexanoic acid and octanoic acid solubilities in the water-rich phase as seen in the magnifications of Figs. 4.5b) and 4.5c). This quantitative description is inherited from the accurate alkane-water model which the alkanoic acid-water model is built upon.
Figure 4.4: Association schemes for COOH - COOH (dimerisation) and COOH - H₂O groups (asymmetric association)

Figure 4.5: Fluid-phase equilibria of (a) n-butanoic acid + water binary mixture (VLE) [304] at pressures of $p = 0.013$ MPa (circles) and $p = 0.101$ MPa (squares); (b) $n$-hexanoic acid + water binary mixture at ambient and elevated pressure, the symbols: crosses [305], circles [306], triangles [296], and squares [307] represent the experimental data; (c) $n$-octanoic acid + water binary mixture at conditions of three-phase existence (circles) [306] and pressures of 5.00 - 15.0 MPa (triangles) [308]. The red curves represent the VLE description in (a) and the description at conditions of three-phase equilibria in (b) and (c). The inserts in (b) and (c) magnify $n$-hexanoic acid and $n$-octanoic acid solubilities in the water-rich phase.
4.4 Obtaining CH$_2$OH unlike group interaction parameters

This section reveals how the unlike-interaction parameters between the branched-alkyl/ aromatic/ COOH groups and the CH$_2$OH group are obtained based on the predetermined CH$_3$, CH$_2$, and CH$_2$OH interaction parameters. One of the main features of the SAFT-\(\gamma\) Mie approach lies in its ability to predict fluid phase behaviour and thermodynamic properties of mixtures based solely on pure component data, as long as the necessary information pertaining to the groups present in the mixture can be gleaned from the pure components. Taking advantage of this, the unlike interaction parameters for CH$_2$OH - CH and CH$_2$OH - C groups are obtained using pure component data for the branched alcohols [309–313]. The CH - CH$_2$OH unlike dispersion energy \(\varepsilon_{\text{CH-CH}_2\text{OH}}\) is estimated from the vapour pressure \(p_{\text{vap}}(T)\) [309, 310] and the saturated liquid density \(\rho_{\text{sat}}^L(T)\) [311] of pure isobutanol (2-methyl-1-propanol) which contains both CH and CH$_2$OH groups within the molecule. The %AADs between the calculated and the experimental values are 4.35 for the \(p_{\text{vap}}(T)\) and 2.20 for the \(\rho_{\text{sat}}^L(T)\). The C - CH$_2$OH unlike dispersion energy \(\varepsilon_{\text{C-CH}_2\text{OH}}\) is estimated from \(p_{\text{vap}}(T)\) [312] and \(\rho_{\text{sat}}^L(T)\) [313] of pure neopentyl alcohol (2,2-dimethyl-1-propanol). The %AAD \(p_{\text{vap}}(T)\) and %AAD \(\rho_{\text{sat}}^L(T)\) of the fitted data are 14.09 and 6.58, respectively. The seemingly high %AAD of both sets of the neopentyl alcohol data are largely due to the %AAD of pure neopentane (15.44 % for \(p_{\text{vap}}(T)\) and 5.80 % for \(\rho_{\text{sat}}^L(T)\) [169]), which is the basis for the neopentyl alcohol molecule. The optimised \(\varepsilon_{\text{CH-CH}_2\text{OH}}\) and \(\varepsilon_{\text{C-CH}_2\text{OH}}\) values are reported in Table 4.4.

Although the CH$_2$OH functional group is modelled as an associating group in pure alcohols and in aqueous mixtures, its interactions with the aromatic groups are governed solely by the dispersion energy. In other words, no associating site is used to mediate the interaction between aromatic and alcohol compounds. The unlike interaction is obtained by regressing the \(\varepsilon_{kl}\) to the VLE data of relevant mixtures. The VLE data for 1-hexanol + benzene [314] and 1-butanol + benzene [315] at the ambient pressure are used to obtain \(\varepsilon_{\text{CH-CH}_2\text{OH}}\). The accuracy of the description of these binary mixtures is illustrated in Fig. 4.6a). The aCH - CH$_2$OH model is then transferred to obtain \(\varepsilon_{kl}\) between CH$_2$OH and other aromatic groups. The SAFT-\(\gamma\) Mie descriptions for the corresponding binary mixtures used to obtain the \(\varepsilon_{kl}\) for aCH$_3$ - CH$_2$OH, aCH$_2$ - CH$_2$OH, and aCH - CH$_2$OH interactions are shown in Figs. 4.6b), 4.6c), and 4.6d), respectively. The theory provides a good description of the vapour-liquid phase behaviour of the mixtures studied, most of which are characterised by a minimum boiling azeotrope. It is interesting to observe that the shift in the azeotropic composition to higher compositions of 1-butanol with increasing size of benzyl derivatives can be captured accurately by the theory,
cf. Figs. 4.6a) - 4.6d). Moreover, the change in the azeotrope of the ethylbenzene mixtures (Fig. 4.6c) with different alcohols, ethanol and 1-butanol, is also accurately reproduced by the theory.

Figure 4.6: Isobaric temperature–mole fraction (T – x) phase diagrams of the vapour-liquid equilibria of: (a) 1-hexanol + benzene (squares) [314] and 1-butanol + benzene (circles) [315] at pressure \( p = 0.101 \) MPa; (b) 1-butanol + 1,3-dimethylbenzene (triangles) [316], 1-butanol + 1,4-dimethylbenzene (squares) [316], and 1-butanol + methylbenzene (circles) [317] at \( p = 0.101 \) MPa; (c) 1-butanol + ethylbenzene (squares) [316] at \( p = 0.101 \) MPa and ethanol + ethylbenzene (circles) [318] at \( p = 0.096 \) MPa; (d) 1-butanol + (1-methylethyl)benzene at \( p = 0.101 \) MPa [319]. The symbols represent experimental data and the continuous and dotted curves the description with the SAFT-\( \gamma \) Mie approach (the dotted curves represent 1-butanol mixtures).

In modelling alkanic acid + alcohol mixtures, a full (asymmetric) association scheme is used to describe the COOH - CH\(_2\)OH interactions. Thus, a total of seven parameters \( \varepsilon_{\text{CH}_2\text{OH} - \text{COOH}} \), \( \varepsilon^\text{HB}_{\text{CH}_2\text{OH} - \text{COOH},e} \), \( \varepsilon^\text{HB}_{\text{CH}_2\text{OH} - \text{COOH},e} \), \( \varepsilon^\text{HB}_{\text{CH}_2\text{OH} - \text{COOH},e} \), \( K^\text{HB}_{\text{CH}_2\text{OH} - \text{COOH},e} \), \( K^\text{HB}_{\text{CH}_2\text{OH} - \text{COOH},e} \), and \( K^\text{HB}_{\text{CH}_2\text{OH} - \text{COOH},e} \) are regressed to the VLE data of butanoic acid + butanol mixtures.
The optimised parameters result in a quantitatively accurate description of the experimental data as shown in Fig. 4.7, which is comparable with the prediction from the similar group-contribution (GC-SAFT-VR) approach [298].

Figure 4.7: Isobaric temperature–mole fraction \((T - x)\) phase diagram of the vapour-liquid equilibria of \(n\)-butanoic acid + 1-butanol at pressures of \(p = 0.053\) MPa (squares) and 0.027 MPa (circles). The symbols represent the experimental data [320], the continuous curves the description with the SAFT-\(\gamma\) Mie approach.

Table 4.4: Group dispersion interaction energies \(\varepsilon_{kl}\) and repulsive exponent \(\lambda_{kl}^r\) for use in the SAFT-\(\gamma\) Mie group-contribution approach. The unlike segment diameter \(\sigma_{kl}\) is obtained from the arithmetic combining rule and all unlike attractive exponents \(\lambda_{kl}^a = 6.0000\); these are not shown in the table. CR indicates that the unlike repulsive exponent \(\lambda_{kl}^r\) is obtained from a combining rule [22]. * indicates parameters obtained in this work and † unpublished parameters (Dufal and Papaioannou personal communications).

<table>
<thead>
<tr>
<th>(k)</th>
<th>(l)</th>
<th>Gr. (k)</th>
<th>Gr. (l)</th>
<th>(\varepsilon_{kl}/k_B/K)</th>
<th>(\lambda_{kl}^r)</th>
<th>Ref.</th>
<th>(k)</th>
<th>(l)</th>
<th>Gr. (k)</th>
<th>Gr. (l)</th>
<th>(\varepsilon_{kl}/k_B/K)</th>
<th>(\lambda_{kl}^r)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>CH(_1)</td>
<td>CH(_3)</td>
<td>256.77</td>
<td>15.050</td>
<td>[22]</td>
<td>4</td>
<td>15</td>
<td>C</td>
<td>H(_2)O</td>
<td>420.82</td>
<td>CR</td>
<td>*</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>CH(_3)</td>
<td>CH(_2)</td>
<td>350.77</td>
<td>CR</td>
<td>[22]</td>
<td>4</td>
<td>16</td>
<td>C</td>
<td>OH</td>
<td>0.0000</td>
<td>CR</td>
<td>†</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>CH(_3)</td>
<td>CH</td>
<td>387.48</td>
<td>CR</td>
<td>[169]</td>
<td>4</td>
<td>17</td>
<td>C</td>
<td>CH(_2)OH</td>
<td>0.0000</td>
<td>CR</td>
<td>*</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>CH(_3)</td>
<td>C</td>
<td>339.91</td>
<td>CR</td>
<td>[169]</td>
<td>5</td>
<td>5</td>
<td>aCH</td>
<td>aCH</td>
<td>371.53</td>
<td>14.756</td>
<td>[169]</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>CH(_3)</td>
<td>aCH</td>
<td>305.81</td>
<td>CR</td>
<td>[169]</td>
<td>5</td>
<td>6</td>
<td>aCH</td>
<td>aCH(_2)</td>
<td>416.69</td>
<td>CR</td>
<td>[169]</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>CH(_3)</td>
<td>aCH(_2)</td>
<td>386.91</td>
<td>CR</td>
<td>[169]</td>
<td>5</td>
<td>7</td>
<td>aCH</td>
<td>aCH(_3)</td>
<td>471.23</td>
<td>CR</td>
<td>[291]</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>CH(_3)</td>
<td>aCH(_3)</td>
<td>358.58</td>
<td>CR</td>
<td>[291]</td>
<td>5</td>
<td>8</td>
<td>aCH</td>
<td>aCH(_4)</td>
<td>429.16</td>
<td>CR</td>
<td>[169]</td>
</tr>
</tbody>
</table>
Table 4.4: Continued.

| k   | t     | Gr. k | Gr. t | $(\epsilon_k/\kappa_T)/K$ | $\lambda_{ij}$ | Ref. | k   | t     | Gr. k | Gr. t | $(\epsilon_k/\kappa_T)/K$ | $\lambda_{ij}$ | Ref. |
|-----|-------|-------|-------|-----------------------------|-----------------|------|-----|-------|-------|-----------------------------|-----------------|------|     |
| 1   | 8     | CH₃   | aCCH  | 455.85                      | CR [169]        |      | 5   | 13    | aCH   | COOH | 331.61                      | 9.0687          | [169]|
| 1   | 9     | CH₃   | Cl₂   | 333.48                      | CR [169]        |      | 5   | 15    | aCH   | H₂O  | 357.78                      | 38.640          | [169]|
| 1   | 10    | CH₃   | Cl=   | 252.41                      | CR [169]        |      | 5   | 17    | aCH   | Cl₂OH| 386.05                      | CR *            |      |
| 1   | 11    | CH₃   | C=    | 281.40                      | CR t            |      | 5   | 19    | aCH   | CH₂COCH₃| 333.11                    | CR [169]        |      |
| 1   | 12    | CH₃   | <CH₂  | 355.95                      | CR [169]        | 6    | 6    | aCCH₃| aCCH₂| 591.56                      | 8.5433          | [169]|
| 1   | 13    | CH₃   | COOH  | 255.99                      | CR [240]        | 6    | 8    | aCCH₂| aCCH₁| 462.04                      | CR [169]        |      |
| 1   | 14    | CH₃   | COO   | 402.75                      | CR [169]        | 6    | 13   | aCCH₂| COOH | 471.66                      | CR [169]        |      |
| 1   | 15    | CH₃   | H₂O   | 358.18                      | 100.00          * | 6    | 15   | aCCH₂| H₂O  | 329.03                      | CR *            |      |
| 1   | 16    | CH₃   | OH    | 314.67                      | CR t            | 6    | 17   | aCCH₂| CH₂OH| 434.37                      | CR *            |      |
| 1   | 17    | CH₃   | CH₂   | 345.20                      | CR *            | 6    | 19   | aCCH₂| CH₃COCH₃| 394.83                    | CR [169]        |      |
| 1   | 18    | CH₃   | CH₅COCH₃| 233.48                    | 14.449          | 7    | 7    | aCCH₃| aCCH₁| 654.41                      | 23.637          | [293]|
| 2   | 2     | CH₂   | CH₂   | 473.38                      | 19.871          [22] | 7    | 15   | aCCH₃| H₂O  | 360.70                      | CR *            |      |
| 2   | 3     | CH₂   | CH    | 500.21                      | CR [169]        | 7    | 17   | aCCH₃| CH₂OH| 486.62                      | CR *            |      |
| 2   | 4     | CH₂   | C     | 300.07                      | CR [169]        | 8    | 8    | aCCH  | aCCH  | 61.325                      | 8.0000          | [169]|
| 2   | 5     | CH₂   | aCH   | 415.64                      | CR [169]        | 8    | 13   | aCCH  | COOH | 599.28                      | CR [169]        |      |
| 2   | 6     | CH₂   | aCCH₁| 454.16                      | CR [169]        | 8    | 15   | aCCH | H₂O  | 314.03                      | CR *            |      |
| 2   | 7     | CH₂   | aCCH₂| 569.18                      | CR [240]        | 8    | 17   | aCCH | CH₂OH| 430.14                      | CR *            |      |
| 2   | 8     | CH₂   | aCCH  | 345.80                      | CR [169]        | 8    | 19   | aCCH | CH₃COCH₃| 459.22                    | CR [169]        |      |
| 2   | 9     | CH₂   | CH₅= | 386.80                      | CR t            | 9    | 9    | CH₅=| CH₅= | 300.90                      | 20.271          | [169]|
| 2   | 10    | CH₂   | CH=  | 459.40                      | CR t            | 9    | 10   | CH₂=| CH=  | 275.75                      | CR [169]        |      |
| 2   | 11    | CH₂   | C=   | 286.58                      | CR t            | 10   | 10   | CH= | CH= | 952.54                      | 15.974          | [169]|
| 2   | 12    | CH₂   | <CH₂  | 469.67                      | CR t            | 10   | 11   | CH= | C=  | 1195.3                      | CR t            |      |
| 2   | 13    | CH₂   | COOH  | 413.74                      | CR [240]        | 10   | 12   | CH= | <CH₂ | 398.35                      | CR [169]        |      |
| 2   | 14    | CH₂   | COO   | 498.86                      | CR [169]        | 10   | 14   | CH= | COO | 818.79                      | CR t            |      |
| 2   | 15    | CH₂   | H₂O   | 423.63                      | 100.00          * | 10   | 16   | CH= | OH  | 625.17                      | CR t            |      |
| 2   | 16    | CH₂   | OH    | 386.27                      | CR t            | 11   | 11   | C= | C=  | 1500.0                      | 8.0000          | t      |
| 2   | 17    | CH₂   | CH₂OH | 423.17                      | CR t            | 11   | 12   | C= | <CH₂ | 846.19                      | CR t            |      |
| 2   | 18    | CH₂   | CH₅COCH₃| 289.48                    | 11.594          | 11   | 14   | C= | COO | 868.11                      | CR t            |      |
| 3   | 3     | CH     | CH    | 95.621                      | 8.0000          [240] | 11   | 16   | C= | OH  | 784.51                      | CR t            |      |
| 3   | 4     | CH     | C     | 2.0000                      | CR *            | 12   | 12   | <CH₂ | <CH₂ | 477.36                      | 20.386          | [169]|
| 3   | 5     | CH     | aCH   | 441.43                      | CR [169]        | 12   | 14   | <CH₂ | COO | 498.60                      | CR [169]        |      |
| 3   | 6     | CH     | aCCH₁| 65.410                      | CR [169]        | 12   | 16   | <CH₂ | OH  | 376.57                      | CR t            |      |
| 3   | 7     | CH     | aCCH  | 67.510                      | CR [169]        | 13   | 13   | COOH | COOH | 405.78                      | 8.0000          | [240]|
| 3   | 9     | CH     | CH₅= | 426.76                      | CR [169]        | 13   | 15   | COOH | H₂O  | 289.76                      | CR *            |      |
| 3   | 10    | CH     | CH=  | 502.99                      | CR [169]        | 13   | 17   | COOH | CH₂OH| 656.80                      | CR *            |      |
| 3   | 11    | CH     | C=   | 378.72                      | CR t            | 13   | 19   | COOH | CH₃COCH₃| 393.71                    | CR [169]        |      |
| 3   | 12    | CH     | <CH₂  | 570.45                      | CR [169]        | 14   | 14   | COO  | COO | 868.92                      | 31.189          | [169]|
| 3   | 13    | CH     | COOH  | 504.99                      | CR [169]        | 14   | 16   | COO  | OH  | 490.95                      | CR t            |      |
| 3   | 14    | CH     | COO   | 353.65                      | CR t            | 15   | 15   | H₂O  | H₂O | 266.68                      | 17.020          | [239]|


Table 4.4: Continued.

<table>
<thead>
<tr>
<th>k</th>
<th>l</th>
<th>Gr. k</th>
<th>Gr. l</th>
<th>( (\varepsilon_{kl}/k_B) )/K</th>
<th>( \lambda_{kl} )</th>
<th>Ref.</th>
<th>k</th>
<th>l</th>
<th>Gr. k</th>
<th>Gr. l</th>
<th>( (\varepsilon_{kl}/k_B) )/K</th>
<th>( \lambda_{kl} )</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>15</td>
<td>OH</td>
<td>H(_2)O</td>
<td>275.75</td>
<td>CR</td>
<td>*</td>
<td>15</td>
<td>17</td>
<td>H(_2)O</td>
<td>CH(_3)OH</td>
<td>353.37</td>
<td>CR</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>OH</td>
<td>OH</td>
<td>198.08</td>
<td>CR</td>
<td>†</td>
<td>15</td>
<td>19</td>
<td>H(_2)O</td>
<td>CH(_3)COCH(_3)</td>
<td>287.26</td>
<td>CR</td>
<td>[240]</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>OH</td>
<td>CH(_2)OH</td>
<td>329.22</td>
<td>CR</td>
<td>*</td>
<td>16</td>
<td>16</td>
<td>OH</td>
<td>OH</td>
<td>410.31</td>
<td>20.702</td>
<td>†</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>OH</td>
<td>CH(_3)COCH(_3)</td>
<td>637.29</td>
<td>CR</td>
<td>[240]</td>
<td>17</td>
<td>17</td>
<td>CH(_2)OH</td>
<td>CH(_2)OH</td>
<td>407.22</td>
<td>22.699</td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>C</td>
<td>C</td>
<td>50.020</td>
<td>8.0000</td>
<td>[169]</td>
<td>18</td>
<td>18</td>
<td>aCCOaC</td>
<td>aCCOaC</td>
<td>656.71</td>
<td>9.8317</td>
<td>†</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>C</td>
<td>–CH(_2)</td>
<td>0.0000</td>
<td>CR</td>
<td>†</td>
<td>19</td>
<td>19</td>
<td>CH(_2)COCH(_3)</td>
<td>CH(_2)COCH(_3)</td>
<td>286.02</td>
<td>17.433</td>
<td>[240]</td>
</tr>
</tbody>
</table>

Table 4.5: Group association energies \( \varepsilon_{kl,ab}^{HB} \) and bonding volume parameters \( K_{kl,ab} \) for use within the SAFT-\( \gamma \) Mie group-contribution approach.

<table>
<thead>
<tr>
<th>k</th>
<th>l</th>
<th>Group k</th>
<th>Site a of group k</th>
<th>Group l</th>
<th>Site b of group l</th>
<th>( (\varepsilon_{kl,ab}^{HB}/k_B) )/K</th>
<th>( K_{kl,ab}/\text{Å}^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>11</td>
<td>aCH</td>
<td>e(_1)</td>
<td>H(_2)O</td>
<td>H</td>
<td>563.56</td>
<td>339.61</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>aCCH(_2)</td>
<td>e(_1)</td>
<td>H(_2)O</td>
<td>H</td>
<td>563.56</td>
<td>339.61</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>aCCH(_3)</td>
<td>e(_1)</td>
<td>H(_2)O</td>
<td>H</td>
<td>563.56</td>
<td>339.61</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>aCCH</td>
<td>e(_1)</td>
<td>H(_2)O</td>
<td>H</td>
<td>563.56</td>
<td>339.61</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>COOH</td>
<td>H</td>
<td>COOH</td>
<td>H</td>
<td>6427.9</td>
<td>0.8062</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>COOH</td>
<td>e(_1)</td>
<td>H(_2)O</td>
<td>H</td>
<td>1451.8</td>
<td>280.89</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>COOH</td>
<td>e(_2)</td>
<td>H(_2)O</td>
<td>H</td>
<td>1252.6</td>
<td>150.98</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>COOH</td>
<td>H</td>
<td>H(_2)O</td>
<td>e(_1)</td>
<td>2567.7</td>
<td>270.09</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>COOH</td>
<td>e(_1)</td>
<td>CH(_2)OH</td>
<td>H</td>
<td>1015.5</td>
<td>21.827</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>COOH</td>
<td>e(_2)</td>
<td>CH(_2)OH</td>
<td>H</td>
<td>547.42</td>
<td>53.150</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>COOH</td>
<td>H</td>
<td>CH(_2)OH</td>
<td>e(_1)</td>
<td>524.04</td>
<td>14.017</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>COO</td>
<td>e(_1)</td>
<td>OH</td>
<td>H</td>
<td>1920.08</td>
<td>114.85</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>H(_2)O</td>
<td>e(_1)</td>
<td>H(_2)O</td>
<td>H</td>
<td>1985.4</td>
<td>101.69</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>H(_2)O</td>
<td>e(_1)</td>
<td>CH(_2)OH</td>
<td>H</td>
<td>621.68</td>
<td>425.00</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>H(_2)O</td>
<td>H</td>
<td>CH(_2)OH</td>
<td>e(_1)</td>
<td>2153.2</td>
<td>147.40</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>CH(_2)OH</td>
<td>e(_1)</td>
<td>CH(_2)OH</td>
<td>H</td>
<td>2097.9</td>
<td>62.309</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>CH(_3)COCH(_3)</td>
<td>e(_1)</td>
<td>CH(_3)COCH(_3)</td>
<td>H</td>
<td>980.20</td>
<td>2865.2</td>
</tr>
<tr>
<td>19</td>
<td>15</td>
<td>CH(_3)COCH(_3)</td>
<td>e(_1)</td>
<td>H(_2)O</td>
<td>H</td>
<td>1588.7</td>
<td>772.77</td>
</tr>
<tr>
<td>19</td>
<td>15</td>
<td>CH(_3)COCH(_3)</td>
<td>e(_2)</td>
<td>H(_2)O</td>
<td>H</td>
<td>417.24</td>
<td>1304.3</td>
</tr>
<tr>
<td>19</td>
<td>15</td>
<td>CH(_3)COCH(_3)</td>
<td>H</td>
<td>H(_2)O</td>
<td>e(_1)</td>
<td>1386.8</td>
<td>188.83</td>
</tr>
</tbody>
</table>
4.5 Thermodynamic prediction with the SAFT-\(\gamma\) Mie approach

One of the main features of the GC approaches such as the SAFT-\(\gamma\) Mie method lies in its ability to predict the thermodynamic properties of compounds which have not been included in the model development, \(i.e.,\) the parameter estimation procedure. In this section, the models developed in previous sections are examined for their transferability by comparing the theoretical predictions with the corresponding experimental data for binary mixtures that contain the relevant functional groups.

The mutual solubilities in the aqueous binary mixtures of 2-methylpropane and 2,3-dimethylbutane are examined for the transferability of the CH - H\(_2\)O model. The two aqueous isoalkane mixtures are considered to be limiting cases to examine the CH - H\(_2\)O interaction as they are the two smallest isoalkanes, for which the CH - H\(_2\)O interaction thus has the largest impact upon the description of the properties. As can be seen in Fig. 4.8a), a good agreement between the predicted mutual solubilities and the experimental data of both mixtures, particularly the solubilities of the isoalkanes in the water-rich phase, confirms the reliability and transferability of the CH - H\(_2\)O parameters. A similar observation can be made for the transferability of the aromatic - H\(_2\)O parameters. The mutual solubilities in the aqueous binary mixtures of 1,2-dimethylbenzene, 1,3-dimethylbenzene, and 1,4-dimethylbenzene are displayed in Fig. 4.8b). The three dimethylbenzene isomers show similar mutual solubilities in both phases which are predicted accurately by our model. Note that the three isomers are represented by the same SAFT-\(\gamma\) Mie functional groups so their predictions are identical. The quantitative mutual solubility predictions for (1-methylpropyl)benzene and especially the di-substituted derivatives: 1,3-diethylbenzene, 1,4-diisopropylbenzene and water substantiate the predictability of our aromatic - H\(_2\)O model as evidenced in Figs. 4.8c) and 4.8d).
Figure 4.8: Mutual solubilities of water and hydrocarbons at conditions of three-phase equilibria as a function of temperature of: (a) water + isobutane (squares [321] and circles [269]) and water + 2,3-dimethylbutane (pentagons and triangles [263]); (b) water + 1,2-dimethylbenzene (crosses), water + 1,3-dimethylbenzene (squares), water + 1,4-dimethylbenzene (circles) [303]; (c) water + 1,3-diethylbenzene [303]; (d) water + (1-methylpropyl)benzene (squares) and water + 1,4-diisopropylbenzene (triangle) [303]. The continuous curves represent the mutual solubilities of hydrocarbons in the water-rich liquid phase and the dotted curves the mutual solubilities of water in the hydrocarbon-rich liquid phase.

As discussed in Sec. 4.3, though not perfect, the COOH - H₂O model can be used to describe simultaneously the VLE and LLE of different alkanoic acid + water mixtures. The COOH group was designed originally to model medium-length alkanoic acids as such the pure-component data of butanoic, heptanoic, and decanoic acid and the excess enthalpies of mixing of the binary mixtures of n-heptane + hexanoic acid and n-heptane + butanoic acid were considered in the development for its like-interaction parameters [240]. We follow this by including only aqueous mixtures of butanoic and longer acids for the COOH - H₂O parameter estimation. The main reason for excluding the shorter homologues is due to the strong polarity of the COOH group,
which greatly polarises other constituting groups, making these molecules unsuitable to be described by GC approaches. Particularly for the short acids such as ethanoic (acetic) acid, the molecular group model is recommended as has been used by the CPA [295] or the simplified SAFT approach [322] given its industrial importance and complex phase behaviour. Despite excluding the short acid homologues in the development of the COOH group, the complete miscibility of the ethanoic acid + water and propanoic acid + water mixtures at ambient conditions is correctly predicted by the model. In fact, the model gives a reasonable prediction of the VLE of propanoic acid + water mixture as shown in Fig. 4.9a).

A more relevant property to the pharmaceutical industry is the infinite dilution activity coefficient ($\gamma_i^\infty$) which describes the behaviour of a single solute molecule completely surrounded by the solvent in the absence of solute-solute interaction. This property provides a direct link to the octanol-water partition coefficient, particularly $\gamma_i^\infty$ in water (approximately $\gamma_i^{\infty,WR}$) because in the octanol + water mixture the octanol solubility in the water-rich phase is minimal (the aqueous phase in essentially water). Here, the COOH - H$_2$O model is employed for the prediction of $\gamma_i^\infty$ of $n$-alkanoic acid solutes ($i$) in water at $T = 298.15$ K and $p = 0.101$ MPa as shown in Fig. 4.9b). The predictions obtained from a previous model [240] are also shown as a reference. As can be seen in the figure, the SAFT-$\gamma$ Mie predictions based on the current model agree very well with the experimental data; the results are consistent with the accurate description of the solubilities of hexanoic and octanoic acids in the aqueous phase as seen in Figs. 4.5b) and 4.5c). An average absolute error (AAE) of 0.83 is obtained from the current model compared to 2.99 from the previous model. The marked improvement with the current model is due to a better basis for the alkane-water interactions and the additional use of LLE data to obtain the COOH - H$_2$O interaction parameters. The SLE prediction of APIs containing the COOH functional group, i.e., azelaic acid, ibuprofen and ketoprofen, in an aqueous environment will be demonstrated in the next chapters to confirm the validity of the COOH - H$_2$O model.
The transferability of the aromatic - CH$_2$OH model is also investigated for the prediction of the VLE of the binary mixtures that have not been used to develop the model. In Fig. 4.10a), the theoretical descriptions of the VLE of benzene and short alcohols: ethanol and 1-propanol are compared with the experimental data. As seen from the figure, the SAFT-$\gamma$ Mie predictions are in excellent agreement with the experimental data [326, 327], capturing accurately the minimum boiling azeotrope that the mixtures exhibit. The VLE of benzene and a longer alcohol, 1-octanol, is also investigated and presented in Fig. 4.10b). The reliable prediction for the fluid-phase behaviour of benzene and different alcohols indicates the accuracy and transferability of our aromatic - CH$_2$OH model. The model is further assessed for the prediction of excess thermodynamic property of benzene + alcohol mixture; such information has not been used to develop the model. Excess properties of mixing are very sensitive to both the like and unlike intermolecular interactions [328, 329] and constitute a stringent test for any model. Again, an excellent agreement with the experimental data [330] is observed for the SAFT-$\gamma$ Mie prediction for the excess molar enthalpy of mixing of the benzene + ethanol mixture (Fig. 4.10c).

The performance of the SAFT-$\gamma$ Mie approach is further examined for the fluid-phase behaviour of a highly associating mixture between alkanoic acid and alcohol. A satisfactory prediction is achieved for the n-propanoic acid + ethanol VLE (Fig. 4.10d) despite the exclusion
of propanoic acid when obtaining the COOH group interaction parameters (discussed earlier). These results are particularly encouraging considering that both \( n \)-propanoic acid and ethanol are short polar compounds from different chemical families, so that their properties are difficult to capture by any GC approaches due to the relative dominance of the proximity effect caused by the polar groups.

**Figure 4.10:** Isobaric temperature–mole fraction \( (T - x) \) phase diagram of the vapour-liquid equilibria of (a) 1-propanol + benzene (green triangles) [326] and ethanol + benzene (blue diamonds) [327] at pressure \( p = 0.101 \) MPa; (b) 1-octanol + benzene [314] at \( p = 0.100 \) MPa; (d) \( n \)-propanoic acid + ethanol at \( p = 0.101 \) MPa [331]. (c) Isothermal–isobaric excess molar enthalpy of mixing of ethanol + benzene at \( T = 303.15 \) K and \( p = 0.101 \) MPa [330]. The symbols represent the experimental data and the dashed curves the prediction with the SAFT-\( \gamma \) Mie approach.
4.6 Conclusions

In this chapter, the unlike interaction parameters between water/alcohol and the chemical groups characterising ibuprofen have been established. The parameter estimation procedure involves the use of different types of the fluid-phase equilibrium data including saturated vapour pressures and saturated liquid densities of pure components, VLE, LLE, and mutual solubilities along the three-phase coexistence of several binary aqueous mixtures. In most cases, the SAFT-\(\gamma\) Mie approach is found to describe the phase behaviour of the systems studied in very good agreement with the experimental data and accurately capture the effects of change in structure and molecular composition on phase behaviour particularly the mutual solubility in the aqueous phase. The same group parameters are further assessed for their transferability by predicting thermodynamic properties of mixtures that are not included in the regression set for the determination of the group-group parameters. An accurate prediction of the mutual solubilities and VLE of several binary mixtures for a wide range of thermodynamic conditions is obtained with no further adjustment of the parameters. The physical robustness of the group parameters is demonstrated in the simultaneous description of different types of the phase behaviour (VLE and LLE) where the agreement of the predictions with the experimental data is quite reasonable considering the predictive nature of the GC approach and the highly non-ideal phase behaviour that the mixtures exhibit. One of the key features of the approach lies in the description not only of the phase equilibria, but also of the single-phase thermodynamic properties, such as excess properties of mixing and activity coefficients at infinite dilution, which are predicted accurately by the model developed. The results obtained in this chapter encourage the extension of this work towards more complex systems such as the prediction of partition coefficients and solubilities of multifunctional compounds, e.g., APIs.
Chapter 5

Predicting the partition coefficients and solubilities of organic compounds

5.1 Introduction

In this chapter, the SAFT-$\gamma$ Mie approach is applied for the prediction of partition coefficients and solubilities of a number of organic compounds using the functional groups studied in Chapters 3 and 4. Details of the definition, use, and measurement of partition coefficients, specifically the octanol-water partition coefficient $K_{i,OW}$, have been discussed extensively in Chapter 2. In brief, they are an infinite dilution property describing the distribution of a solute $i$ across two phases which are fairly different in polarity, i.e., organic and aqueous phases. In this study, the partition coefficient of solutes in hexane-water ($K_{i,C6W}$), hexadecane-water ($K_{i,C16W}$), and octanol-water systems ($K_{i,OW}$) are considered. Partition coefficients are very useful properties in product and process design applications, as they are often used to indicate the lipophilicity of compounds [28]. A key example is the application of $K_{i,OW}$ in predicting the pharmacokinetic properties and toxicity of organic chemicals, especially for drug molecules [1, 194]. The prediction of $K_{i,OW}$ using thermodynamic approaches is an ongoing subject of research [66, 93, 96, 100, 129].

Solid phase solubility is a fundamental parameter in the design of crystallisation processes commonly used in the pharmaceutical and agrochemical industries. The development and optimisation of these processes require quantitative knowledge of the solubility of the compound of interest (the active pharmaceutical ingredient, API), traditionally achieved by screening a large
number of solvent systems for which the solubility of the compound has to be measured as a function of temperature and solvent composition. Solvent screening can be very time-consuming and cumbersome to carry out experimentally, especially in the case of solvent mixtures, given the wide range of solvents and compositions of the mixture that may have to be examined. The number of experiments that can be conducted is often limited by the small quantities and tremendous cost of APIs in the early stages of drug discovery and development. For the specific case of APIs, the importance of the solubility data extends much beyond the scope of production. The solubility of a solid drug in aqueous media can dictate its bioavailability and indicate the appropriate drug formulation. In particular, the solubility advantage of the transformation of poorly soluble crystalline APIs to the amorphous state for achieving adequate solubility can be properly evaluated [332, 333]. Given the difficulty of the experimental measurement of solubility, thermodynamic tools that can predict the solubility of APIs in an array of solvents become an interesting alternative for solvent screening. The accurate modelling of the solubility can be very useful in the initial phases of the solvent selection process, and can help manipulating the solubility in order to achieve an optimal production process. The application of such methodologies has been highlighted as a key element in the progress of the development of efficient manufacturing in the pharmaceutical industry [6].

A variety of modelling tools have been applied to describe the solubility of complex organic molecules including APIs. A review of tools and their use in solvent-selection procedures can be found in Refs. [334, 335]. Unlike the $K_{i,OW}$ prediction that is typically done based only on the chemical structure of the molecule, solubility prediction usually requires some experimental data, e.g., the molar enthalpy of fusion $\Delta H_f$ and the melting temperature $T_m$ of the compound of interest. The availability of compounds at the stage of prediction allows the application of correlative methods such as NRTL, UNIQUAC, and Wilson equations for solubility modelling [336] as these methods require some solubility data of the target compound for the model calibration. One of the most widely applied thermodynamic tools for estimating solubilities of pharmaceuticals is the NRTL-SAC model [76, 77, 108, 109, 116, 117]. This semiempirical activity coefficient model requires regression to extensive experimental data for solid-liquid equilibria of the target compound in a wide range of solvents [76]. In a similar manner, non-group-contribution equations of state have been applied to the study of solid-liquid equilibria, e.g., the work with SAFT-VR on the study of the solubility of aromatic compounds and alkanoic acid in organic solvents [337]. The PC-SAFT EoS [105] has been recently applied to the prediction of solubility of paracetamol [114], ibuprofen, lovastatin [338], and some heterocyclic compounds such as
hydrochlorothiazide, indomethacin, and itraconazole [333]. In these studies, the PC-SAFT parameters describing the API were determined from their solubility data (most commonly aqueous solubility data). This is mainly because these compounds tend to be in the solid state at normal conditions thus experimental vapour pressures and saturated liquid densities are rarely available. The solubility data were also used to determine the unlike interactions between the APIs and solvents. A similar EoS which is also based on Wertheim’s theory for fluids, the CPA EoS [107], has also been applied for the solubility modelling of different APIs such as acetanilide, acetylsalicylic acid, ibuprofen, paracetamol, etc. in aqueous [339] and organic solvents [340]. The pure component parameters for these APIs were obtained from fitting to their corresponding vapor pressure and liquid density data. Although the pure predictions were inaccurate, the results improved when a temperature independent binary interaction parameter $k_{ij}$ estimated from experimental solubility data was used.

A common denominator in the methodologies mentioned thus far is that solubility data of the target compound are typically required for the model development. The thermodynamic tools that are able to predict solubility without resorting to any solubility data of the compound of interest are a priori models such as COSMO-SAC [96] and group-contribution approaches such as UNIFAC [17] and SAFT-γ Mie [22]. Although the compound must be already available at the stage of prediction since pure-component data such as $\Delta H_f$ and $T_m$ of the API are required, the predictive methods are still very useful because these data are much easier to obtain in comparison to the solubility data, especially at the early stage development where the amount of API is very limited. A number of studies [110, 111, 116, 341, 342] have used the COSMO-SAC approach to predict solubility of APIs including ibuprofen, paracetamol, and lovastatin. However in order to improve the accuracy for solubility predictions, experimental solubility data of the drug was used to refine the model’s parameter [111, 341, 342]. Comparison studies for the solubility prediction by COSMO-SAC with other modelling approaches such as NRTL-SAC and UNIFAC are also available [109, 113, 116, 181]; these studies suggest that COSMO-SAC is useful in estimating solubility as it is a priori model but the semi-predictive NRTL-SAC and the predictive UNIFAC methods offer superior performance. As one of the most popular predictive approaches in chemical engineering applications, UNIFAC has been employed for the solubility prediction of several pharmaceuticals [108, 113, 343, 344]. In order to maximise the performance of the method, a recent incarnation named Pharma Mod. UNIFAC has been presented [112], where the fragmentation of groups was geared specifically to the study of APIs (e.g., the definition of the sulfonic acid and amine groups) and several solubility data were
included in the regression of the parameters. The Pharma Mod. UNIFAC performs significantly better compared to the results obtained with the original UNIFAC model. However, the method was developed by neglecting several group-group interactions and should not be applied to concentrations of APIs that are greater than 10% mol/mol. Although this covers the range of concentrations encountered in most crystallisation purposes, the limited composition range can be considered a significant limitation of the Pharma Mod. UNIFAC, especially for highly soluble compounds.

After a success in accurately modelling a variety of thermodynamic properties including phase equilibria, derivative properties, and excess properties in previous chapters, in this chapter, the SAFT-\(\gamma\) Mie approach is applied for the prediction of partition coefficients and solubilities using the group parameters developed thus far. The target compounds in this chapter range from “simple” organic molecules to APIs which do not involve intramolecular association. Details for the calculation of partition coefficients and solubility (SLE) are discussed in the next section. Results for the \(K_{i,C_6W}\), \(K_{i,C_{16}W}\), and \(K_{i,OW}\) predictions of several families of organic compounds and the solubility of APIs are presented afterwards.

### 5.2 Property calculations

#### Partition Coefficient Calculation

The partition coefficient is defined as the equilibrium ratio of the molar concentrations of the infinitely diluted solute in two mutually saturated liquid phases, usually of large difference in polarity (e.g., a hydrocarbon-rich and a water-rich phase). In this study, the partition coefficient of the solute in hexane-water \((K_{i,C_6W})\), hexadecane-water \((K_{i,C_{16}W})\), and octanol-water systems \((K_{i,OW})\) are considered. The partition coefficient can be calculated as the ratio of infinite dilution activity coefficients of the solute in the water-rich (WR) and hydrocarbon-rich (HR) phases with molar volumes of the two phases usually at 298.15 K and 0.101 MPa as \([135]\):

\[
K_{HC-W_i}(T,p) = \lim_{x_i \to 0} \frac{C_{i,HR}}{C_{i,WR}} = \frac{v_{WR} \gamma_{i,WR}^\infty(T,p)}{v_{HR} \gamma_{i,HR}^\infty(T,p)}.
\]  

(5.1)

The molar volumes of the water-rich \(v_{WR}\) and the hydrocarbon-rich \(v_{HR}\) phases are obtained using the SAFT-\(\gamma\) Mie equation of state at \(T = 295.15\) K and \(p = 0.101\) MPa. Infinite dilution activity coefficients of the solute in the two phases \(\gamma_{i,WR}^\infty\) and \(\gamma_{i,HR}^\infty\) are calculated at the same \(T\) and \(p\) at infinite dilution (here taken to be \(x_i = 10^{-8}\) in the calculations) and with the solvent compositions for the water-rich and hydrocarbon-rich phases taken to be those corresponding
to the LLE of the water-hydrocarbon mixtures (the third component $i$ is assumed to have no effect on the hydrocarbon-water LLE).

**Solubility Calculation**

The solubility of a compound (*e.g.*, API) in a solvent or solvent mixture is defined as the maximum quantity of the compound that can be dissolved in a given amount of a solvent to form a homogeneous solution. This can be determined according to the solid-liquid equilibrium (SLE) between the crystalline API solid and the liquid API in solution using the following expression [345]:

$$
\ln x_{API}(T,p) = \frac{\Delta H_f}{R} \left( \frac{1}{T_m} - \frac{1}{T} \right) - \frac{\Delta c_p(T_m)}{R} \left[ \ln \frac{T_m}{T} - \frac{T_m}{T} + 1 \right] - \ln \gamma_{API}(T,p,x), \quad (5.2)
$$

where $x_{API}(T,p)$ is the solubility of the API solute in the liquid phase and $\Delta H_f$ and $T_m$ are the molar enthalpy of fusion and the melting temperature of the solute, respectively. In this expression, the difference between the heat capacities of the solid and the liquid form of the solute, $\Delta c_p$, is taken explicitly into account but is assumed to be temperature-independent at the $T_m$. For compounds that do not exhibit melting points far from the temperature of interest ($T_m \approx T$), the $\Delta c_p(T_m)$ term has insignificant effect on the solubility and is therefore neglected. The above expression is then reduced to:

$$
\ln x_{API}(T,p) = \frac{\Delta H_f}{R} \left( \frac{1}{T_m} - \frac{1}{T} \right) - \ln \gamma_{API}(T,p,x). \quad (5.3)
$$

The values for $\Delta H_f$, $T_m$, and $\Delta c_p(T_m)$ are taken from experimental data. The SAFT-\(\gamma\) Mie EoS is employed for the calculation of the activity coefficient of the API in the liquid phase, $\gamma_{API}(T,p,x)$. The SLE calculation of APIs is carried out in a predictive fashion where neither SLE data nor other physical property data relating to APIs are used to obtain interaction parameters.
Chapter 5: Partition coefficients and solubility

5.3 Results

5.3.1 Partition coefficients

We present the predicted partition coefficients of aliphatic and aromatic compounds in \( n \)-hexane-water \( (K_{i,C_6W}) \), \( n \)-hexadecane-water \( (K_{i,C_{16}W}) \), and 1-octanol-water \( (K_{i,OW}) \) systems. The ability to predict the partition coefficients in different solvent mixtures is considered a major advantage for group-contribution approaches. The SAFT-\( \gamma \) Mie models for the alkane-water and alcohol-water systems have been established and described in full detail in Chapter 3. In brief, the alkane-water model is able to predict accurately a wide range of alkane aqueous solubilities (spanning approximately ten orders of magnitude) as well as other types of fluid-phase equilibria, and results to highly accurate hexane-water and hexadecane-water models (Sec. 3.3.2). The octanol-water model provides an excellent description of the mutual solubility data and saturated liquid densities, and is transferable to other alcohol + water mixtures (Sec. 3.3.3).

For the partition coefficient prediction with SAFT-\( \gamma \) Mie, the models are used to calculate both the activity coefficients and saturated volumes (Eq. (5.1)), it is worth noting that other thermodynamic approaches \[93, 96, 100, 129\] often use the experimental saturated volumes for their prediction instead of predicted or calculated ones. The hexane-water, hexadecane-water, and octanol-water partition coefficients of \( n \)-alkanes, 1-alcohols, and \( n \)-alkanoic acids are shown in Fig. 5.1. As can be seen the SAFT-\( \gamma \) Mie predictions are overall in excellent agreement with the experimental values \[104, 143, 346, 347\]. Though the partition coefficients are calculated from the ratio of the activity coefficients of solutes in the water-rich \( \gamma_i^{\infty}_{i,WR} \) and the hydrocarbon-rich phases \( \gamma_i^{\infty}_{i,HR} \) (Eq. (5.1)), the variation of \( \gamma_i^{\infty}_{i,WR} \) is much greater than that of \( \gamma_i^{\infty}_{i,HR} \). Saying this, most solutes behave more non-ideally in the water-rich phase \[191, 292\] and we find that an accurate description of \( \gamma_i^{\infty}_{i,WR} \) plays a vital role in the reliability of the partition coefficient prediction. It is also useful to note that the water-rich phase is essentially pure water given the very small amount of hexane, hexadecane, or octanol present and hence the \( \gamma_i^{\infty} \) of a solute in pure water (see Chapters 3 and 4) can provide a direct link to the partition coefficient calculation.
Figure 5.1: Partition coefficients of \( n \)-alkanes, 1-alcohols, and \( n \)-alkanoic acids in (a) hexane-water, (b) hexadecane-water, and (c) octanol-water systems at \( T = 298.15 \) K and \( p = 0.101 \) MPa. The symbols [104, 143, 346, 347] represent the experimental data, the error bars the associated uncertainty [104], and the lines the predictions with the SAFT-\( \gamma \) Mie approach.

In addition to the linear aliphatic compounds, partition coefficients of the branched aliphatic and aromatic compounds are also predicted and reported in Tables 5.1 - 5.3. The average absolute errors (AAE = \( \frac{1}{N_p} \sum_{i=1}^{N_p} |\log K_{i, \text{exp}} - \log K_{i, \text{calc}}| \) for \( \log K_{i, \text{C}_8\text{W}} \), \( \log K_{i, \text{C}_{16}\text{W}} \), and \( \log K_{i, \text{OW}} \) are 0.19 (\( N_p = 28 \)), 0.16 (\( N_p = 60 \)), and 0.38 (\( N_p = 78 \)), respectively, for alkane, alcohol, carboxylic acid, and aromatic solutes. The overall comparison between experimental and calculated partition coefficients in the three systems are shown in Fig. 5.2. The excellent agreement with the experimental data suggests that our SAFT-\( \gamma \) Mie models are accurate in modelling a variety of compounds in different solvent systems and also transferable since all compounds are based on the same fundamental functional groups. These results verify the applicability of our group-contribution method for the partition coefficient prediction and allow us to investigate more complex solutes which may undergo intramolecular association (Chapter 6).
Figure 5.2: Comparison between experimental and predicted (a) $K_{i,C_6W}$, (b) $K_{i,C_{16}W}$, and (c) $K_{i,OW}$ at $T = 298.15$ K and $p = 0.101$ MPa. The experimental data can be found in Refs. [104, 143, 346, 347]. The blue circles represent the $n$-alkanes, the purple circles the branched alkanes, the red squares the 1-alcohols, the orange squares the branched alcohols, the dark-green diamonds the $n$-alkanoic acids, the light-green diamonds the branched acids, the brown diamonds the dicarboxylic acids, the light-blue triangles the alkylbenzenes, and the yellow pentagons the phenylalkanoic acids (also see Chapter 6).
Table 5.1: Hexane-water partition coefficients $K_{i,C_6W}$ of alkanes, alcohols, alkanoic acids, and alkylbenzenes (at $T = 298.15$ K and $p = 0.101$ MPa). The deviation is defined as $|\log K_{i,C_6W}^{\text{exp}} - \log K_{i,C_6W}^{\text{calc}}|$. The experimental data are taken from Ref. [347].

<table>
<thead>
<tr>
<th>Formula</th>
<th>Compound</th>
<th>Exp. $\log K_{i,C_6W}$</th>
<th>Predicted $\log K_{i,C_6W}$</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₂H₆</td>
<td>ethane</td>
<td>2.05</td>
<td>2.06</td>
<td>0.01</td>
</tr>
<tr>
<td>C₃H₈</td>
<td>propane</td>
<td>2.73</td>
<td>2.76</td>
<td>0.03</td>
</tr>
<tr>
<td>C₄H₁₀</td>
<td>$n$-butane</td>
<td>3.38</td>
<td>3.46</td>
<td>0.08</td>
</tr>
<tr>
<td>C₅H₁₂</td>
<td>$n$-pentane</td>
<td>4.06</td>
<td>4.15</td>
<td>0.09</td>
</tr>
<tr>
<td>C₆H₁₄</td>
<td>$n$-hexane</td>
<td>4.58</td>
<td>4.85</td>
<td>0.27</td>
</tr>
<tr>
<td>C₇H₈O</td>
<td>ethanol</td>
<td>-2.08</td>
<td>-2.17</td>
<td>0.09</td>
</tr>
<tr>
<td>C₃H₄O</td>
<td>1-propanol</td>
<td>-1.39</td>
<td>-1.47</td>
<td>0.08</td>
</tr>
<tr>
<td>C₄H₁₀O</td>
<td>1-butanol</td>
<td>-0.82</td>
<td>-0.77</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>2-methyl-1-propanol</td>
<td>-0.60</td>
<td>-0.92</td>
<td>0.32</td>
</tr>
<tr>
<td>C₅H₁₂O</td>
<td>1-pentanol</td>
<td>-0.28</td>
<td>-0.07</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>3-methyl-1-butanol</td>
<td>-0.30</td>
<td>-0.22</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>2,2-dimethyl-1-propanol</td>
<td>-0.35</td>
<td>-0.49</td>
<td>0.14</td>
</tr>
<tr>
<td>C₆H₁₄O</td>
<td>1-hexanol</td>
<td>0.48</td>
<td>0.63</td>
<td>0.15</td>
</tr>
<tr>
<td>C₇H₁₆O</td>
<td>1-heptanol</td>
<td>1.04</td>
<td>1.32</td>
<td>0.28</td>
</tr>
<tr>
<td>C₈H₁₈O</td>
<td>1-octanol</td>
<td>1.56</td>
<td>2.02</td>
<td>0.46</td>
</tr>
<tr>
<td>C₂H₄O₂</td>
<td>ethanoic acid</td>
<td>-3.06</td>
<td>-3.47</td>
<td>0.41</td>
</tr>
<tr>
<td>C₃H₆O₂</td>
<td>propanoic acid</td>
<td>-2.23</td>
<td>-2.75</td>
<td>0.52</td>
</tr>
<tr>
<td>C₄H₈O₂</td>
<td>butanoic acid</td>
<td>-1.70</td>
<td>-2.04</td>
<td>0.34</td>
</tr>
<tr>
<td>C₅H₁₀O₂</td>
<td>pentanoic acid</td>
<td>-0.92</td>
<td>-1.33</td>
<td>0.41</td>
</tr>
<tr>
<td>C₆H₁₂O₂</td>
<td>hexanoic acid</td>
<td>-0.49</td>
<td>-0.63</td>
<td>0.14</td>
</tr>
<tr>
<td>C₇H₁₄O₂</td>
<td>heptanoic acid</td>
<td>-0.13</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>C₁₀H₂₀O₂</td>
<td>decanoic acid</td>
<td>1.87</td>
<td>2.18</td>
<td>0.31</td>
</tr>
<tr>
<td>C₆H₆</td>
<td>benzene</td>
<td>2.30</td>
<td>2.43</td>
<td>0.13</td>
</tr>
<tr>
<td>C₇H₈</td>
<td>methylbenzene</td>
<td>2.85</td>
<td>2.94</td>
<td>0.09</td>
</tr>
<tr>
<td>C₈H₁₀</td>
<td>ethylbenzene</td>
<td>3.34</td>
<td>3.33</td>
<td>0.01</td>
</tr>
<tr>
<td>C₆H₁₂</td>
<td>1,2-dimethylbenzene</td>
<td>3.26</td>
<td>3.45</td>
<td>0.19</td>
</tr>
<tr>
<td>C₇H₁₄</td>
<td>1,3-dimethylbenzene</td>
<td>3.24</td>
<td>3.45</td>
<td>0.21</td>
</tr>
<tr>
<td>C₈H₁₆</td>
<td>1,4-dimethylbenzene</td>
<td>3.44</td>
<td>3.45</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Average deviation 0.19
Table 5.2: Hexadecane-water partition coefficients $K_{i,C_{16}W}$ of alkanes, alcohols, alkanolic acids, and alkylbenzenes (at $T = 298.15$ K and $p = 0.101$ MPa). The deviation is defined as $|\log K_{i,C_{16}W}^{\text{exp}} - \log K_{i,C_{16}W}^{\text{calc}}|$. The experimental data are taken from Ref. [346, 347].

<table>
<thead>
<tr>
<th>Formula</th>
<th>Compound</th>
<th>Exp. $\log K_{i,C_{16}W}$</th>
<th>Predicted $\log K_{i,C_{16}W}$</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₂H₆</td>
<td>ethane</td>
<td>1.83</td>
<td>1.84</td>
<td>0.01</td>
</tr>
<tr>
<td>C₃H₈</td>
<td>propane</td>
<td>2.49</td>
<td>2.52</td>
<td>0.03</td>
</tr>
<tr>
<td>C₄H₁₀</td>
<td>$n$-butane</td>
<td>3.13</td>
<td>3.21</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>2-methylpropane</td>
<td>3.11</td>
<td>3.03</td>
<td>0.08</td>
</tr>
<tr>
<td>C₅H₁₂</td>
<td>$n$-pentane</td>
<td>3.87</td>
<td>3.89</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>2-methylbutane</td>
<td>3.76</td>
<td>3.71</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>2,2-dimethylpropane</td>
<td>3.66</td>
<td>3.38</td>
<td>0.28</td>
</tr>
<tr>
<td>C₆H₁₄</td>
<td>$n$-hexane</td>
<td>4.49</td>
<td>4.57</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>2-methylpentane</td>
<td>4.34</td>
<td>4.40</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>3-methylpentane</td>
<td>4.42</td>
<td>4.40</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>2,2-dimethylbutane</td>
<td>4.19</td>
<td>4.07</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>2,3-dimethylbutane</td>
<td>4.22</td>
<td>4.22</td>
<td>0.00</td>
</tr>
<tr>
<td>C₇H₁₆</td>
<td>$n$-heptane</td>
<td>5.14</td>
<td>5.25</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>2-methylhexane</td>
<td>5.15</td>
<td>5.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>3-methylhexane</td>
<td>5.03</td>
<td>5.08</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>2,2-dimethylpentane</td>
<td>4.91</td>
<td>4.75</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>2,3-dimethylpentane</td>
<td>4.87</td>
<td>4.90</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>2,4-dimethylpentane</td>
<td>4.89</td>
<td>4.90</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>3,3-dimethylpentane</td>
<td>4.83</td>
<td>4.75</td>
<td>0.08</td>
</tr>
<tr>
<td>C₈H₁₈</td>
<td>$n$-octane</td>
<td>5.79</td>
<td>5.93</td>
<td>0.14</td>
</tr>
</tbody>
</table>
### Table 5.2: Continued.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Compound</th>
<th>Exp. log$K_{i,C_{18}W}$</th>
<th>Predicted log$K_{i,C_{18}W}$</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_3H_{20}$</td>
<td>n-nonane</td>
<td>5.69</td>
<td>5.76</td>
<td>0.07</td>
</tr>
<tr>
<td>$C_3H_{20}$</td>
<td>n-decane</td>
<td>6.49</td>
<td>6.61</td>
<td>0.12</td>
</tr>
<tr>
<td>$C_4H_{10}$</td>
<td>1-propanol</td>
<td>5.28</td>
<td>5.26</td>
<td>0.02</td>
</tr>
<tr>
<td>$C_4H_{10}$</td>
<td>1-butanol</td>
<td>5.36</td>
<td>5.41</td>
<td>0.05</td>
</tr>
<tr>
<td>$C_6H_{12}$</td>
<td>1-pentanol</td>
<td>6.45</td>
<td>6.11</td>
<td>0.66</td>
</tr>
<tr>
<td>$C_6H_{12}$</td>
<td>1-hexanol</td>
<td>7.01</td>
<td>7.29</td>
<td>0.28</td>
</tr>
<tr>
<td>$C_6H_{14}$</td>
<td>1-hexanol</td>
<td>-2.19</td>
<td>-2.33</td>
<td>0.14</td>
</tr>
<tr>
<td>$C_7H_{16}$</td>
<td>1-heptanol</td>
<td>-1.53</td>
<td>-1.65</td>
<td>0.12</td>
</tr>
<tr>
<td>$C_8H_{18}$</td>
<td>1-octanol</td>
<td>-0.86</td>
<td>-0.97</td>
<td>0.11</td>
</tr>
<tr>
<td>$C_9H_{20}$</td>
<td>1-nonanol</td>
<td>-0.89</td>
<td>-1.16</td>
<td>0.27</td>
</tr>
<tr>
<td>$C_{10}H_{22}$</td>
<td>1-decanol</td>
<td>-0.24</td>
<td>-0.29</td>
<td>0.05</td>
</tr>
<tr>
<td>$C_{10}H_{22}$</td>
<td>1-decanol</td>
<td>-0.23</td>
<td>-0.47</td>
<td>0.24</td>
</tr>
<tr>
<td>$C_{12}H_{26}$</td>
<td>1-decanol</td>
<td>2.96</td>
<td>3.12</td>
<td>0.16</td>
</tr>
<tr>
<td>$C_{12}H_{26}$</td>
<td>ethanoic acid</td>
<td>-3.16</td>
<td>-3.55</td>
<td>0.39</td>
</tr>
<tr>
<td>$C_{12}H_{26}$</td>
<td>propanoic acid</td>
<td>-2.45</td>
<td>-2.85</td>
<td>0.40</td>
</tr>
<tr>
<td>$C_{12}H_{26}$</td>
<td>butanoic acid</td>
<td>-1.83</td>
<td>-2.15</td>
<td>0.32</td>
</tr>
<tr>
<td>$C_{12}H_{26}$</td>
<td>pentanoic acid</td>
<td>-1.14</td>
<td>-1.46</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Table 5.2: Continued.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Compound</th>
<th>Exp. log(K_{i,C_{16}W})</th>
<th>Predicted log(K_{i,C_{16}W})</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{C}<em>6\text{H}</em>{12}\text{O}_2)</td>
<td>3-methylbutanoic acid</td>
<td>-1.33</td>
<td>-1.67</td>
<td>0.34</td>
</tr>
<tr>
<td>(\text{C}<em>7\text{H}</em>{14}\text{O}_2)</td>
<td>hexanoic acid</td>
<td>-0.64</td>
<td>-0.77</td>
<td>0.13</td>
</tr>
<tr>
<td>(\text{C}<em>8\text{H}</em>{16}\text{O}_2)</td>
<td>heptanoic acid</td>
<td>-0.06</td>
<td>-0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>(\text{C}<em>8\text{H}</em>{16}\text{O}_2)</td>
<td>octanoic acid</td>
<td>0.56</td>
<td>0.61</td>
<td>0.05</td>
</tr>
<tr>
<td>(\text{C}_6\text{H}_6)</td>
<td>benzene</td>
<td>2.15</td>
<td>2.28</td>
<td>0.13</td>
</tr>
<tr>
<td>(\text{C}_7\text{H}_8)</td>
<td>methylbenzene</td>
<td>2.68</td>
<td>2.91</td>
<td>0.23</td>
</tr>
<tr>
<td>(\text{C}<em>8\text{H}</em>{10})</td>
<td>ethylbenzene</td>
<td>3.20</td>
<td>3.00</td>
<td>0.20</td>
</tr>
<tr>
<td>(\text{C}<em>9\text{H}</em>{12})</td>
<td>1,2-dimethylbenzene</td>
<td>3.28</td>
<td>3.53</td>
<td>0.25</td>
</tr>
<tr>
<td>(\text{C}<em>9\text{H}</em>{12})</td>
<td>1,3-dimethylbenzene</td>
<td>3.23</td>
<td>3.53</td>
<td>0.30</td>
</tr>
<tr>
<td>(\text{C}<em>9\text{H}</em>{12})</td>
<td>1,4-dimethylbenzene</td>
<td>3.25</td>
<td>3.53</td>
<td>0.28</td>
</tr>
<tr>
<td>(\text{C}<em>{10}\text{H}</em>{14})</td>
<td>isopropylbenzene</td>
<td>3.86</td>
<td>3.44</td>
<td>0.42</td>
</tr>
<tr>
<td>(\text{C}<em>{11}\text{H}</em>{16})</td>
<td>1,2,3-trimethylbenzene</td>
<td>3.68</td>
<td>4.16</td>
<td>0.48</td>
</tr>
<tr>
<td>(\text{C}<em>{11}\text{H}</em>{16})</td>
<td>1,2,4-trimethylbenzene</td>
<td>3.81</td>
<td>4.16</td>
<td>0.35</td>
</tr>
<tr>
<td>(\text{C}<em>{11}\text{H}</em>{16})</td>
<td>1,3,5-trimethylbenzene</td>
<td>3.68</td>
<td>4.16</td>
<td>0.48</td>
</tr>
<tr>
<td>(\text{C}<em>{11}\text{H}</em>{16})</td>
<td>1-methyl-2-ethylbenzene</td>
<td>3.59</td>
<td>3.63</td>
<td>0.04</td>
</tr>
<tr>
<td>(\text{C}<em>{11}\text{H}</em>{16})</td>
<td>1-methyl-4-ethylbenzene</td>
<td>3.59</td>
<td>3.63</td>
<td>0.04</td>
</tr>
<tr>
<td>(\text{C}<em>{11}\text{H}</em>{16})</td>
<td>(n)-pentylnbenzene</td>
<td>5.06</td>
<td>5.05</td>
<td>0.01</td>
</tr>
<tr>
<td>(\text{C}<em>{12}\text{H}</em>{18})</td>
<td>(n)-hexylbenzene</td>
<td>5.69</td>
<td>5.74</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Average deviation: 0.16
**Table 5.3:** Octanol-water partition coefficients ($K_{i,OW}$) of alkanes, alcohols, alkanoic acids, alkylbenzenes, and phenylalkanoic acids (at $T = 298.15$ K and $p = 0.101$ MPa). The deviation is defined as $|\log K_{i,OW}^{exp} - \log K_{i,OW}^{calc}|$. The experimental data are taken from Ref. [104] unless stated otherwise.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Compound</th>
<th>Exp. log$K_{i,OW}$</th>
<th>Uncertainty</th>
<th>Predicted log$K_{i,OW}$</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C$_2$H$_6$</td>
<td>ethane</td>
<td>1.81[143]</td>
<td>n/a</td>
<td>1.70</td>
<td>0.11</td>
</tr>
<tr>
<td>C$_3$H$_8$</td>
<td>propane</td>
<td>2.36[143]</td>
<td>n/a</td>
<td>2.34</td>
<td>0.02</td>
</tr>
<tr>
<td>C$<em>4$H$</em>{10}$</td>
<td>n-butane</td>
<td>2.89[143]</td>
<td>n/a</td>
<td>2.98</td>
<td>0.08</td>
</tr>
<tr>
<td>2-methylpropane</td>
<td></td>
<td>2.76[143]</td>
<td>n/a</td>
<td>2.77</td>
<td>0.01</td>
</tr>
<tr>
<td>C$<em>5$H$</em>{12}$</td>
<td>n-pentane</td>
<td>3.45</td>
<td>0.20</td>
<td>3.62</td>
<td>0.17</td>
</tr>
<tr>
<td>2,2-dimethylpropane</td>
<td></td>
<td>3.11</td>
<td>0.20</td>
<td>3.02</td>
<td>0.09</td>
</tr>
<tr>
<td>C$<em>6$H$</em>{14}$</td>
<td>n-hexane</td>
<td>4.00</td>
<td>0.25</td>
<td>4.26</td>
<td>0.26</td>
</tr>
<tr>
<td>3-Methylpentane</td>
<td></td>
<td>3.60</td>
<td>0.20</td>
<td>4.05</td>
<td>0.45</td>
</tr>
<tr>
<td>2,3-Dimethylbutane</td>
<td></td>
<td>3.85</td>
<td>0.40</td>
<td>3.83</td>
<td>0.02</td>
</tr>
<tr>
<td>2,2-Dimethylbutane</td>
<td></td>
<td>3.82</td>
<td>0.40</td>
<td>3.66</td>
<td>0.16</td>
</tr>
<tr>
<td>C$<em>7$H$</em>{16}$</td>
<td>n-heptane</td>
<td>4.50</td>
<td>0.25</td>
<td>4.90</td>
<td>0.40</td>
</tr>
<tr>
<td>C$<em>8$H$</em>{18}$</td>
<td>n-octane</td>
<td>5.15</td>
<td>0.45</td>
<td>5.54</td>
<td>0.39</td>
</tr>
<tr>
<td>C$<em>9$H$</em>{20}$</td>
<td>n-nonane</td>
<td>5.65</td>
<td>0.60</td>
<td>6.18</td>
<td>0.53</td>
</tr>
<tr>
<td>C$<em>{10}$H$</em>{22}$</td>
<td>n-decane</td>
<td>6.25</td>
<td>0.70</td>
<td>6.82</td>
<td>0.57</td>
</tr>
<tr>
<td>C$<em>{11}$H$</em>{24}$</td>
<td>n-dodecane</td>
<td>6.80</td>
<td>1.00</td>
<td>8.10</td>
<td>1.30</td>
</tr>
<tr>
<td>C$<em>{12}$H$</em>{26}$</td>
<td>n-tetradecane</td>
<td>8.00</td>
<td>1.00</td>
<td>9.38</td>
<td>1.38</td>
</tr>
<tr>
<td>C$_2$H$_6$O</td>
<td>ethanol</td>
<td>-0.30</td>
<td>0.30</td>
<td>-0.56</td>
<td>0.26</td>
</tr>
<tr>
<td>C$_3$H$_4$O</td>
<td>1-propanol</td>
<td>0.25</td>
<td>0.05</td>
<td>0.07</td>
<td>0.18</td>
</tr>
<tr>
<td>C$<em>4$H$</em>{10}$O</td>
<td>1-butanol</td>
<td>0.84</td>
<td>0.04</td>
<td>0.71</td>
<td>0.13</td>
</tr>
<tr>
<td>2-Methyl-1-propanol</td>
<td></td>
<td>0.76</td>
<td>0.07</td>
<td>0.50</td>
<td>0.26</td>
</tr>
<tr>
<td>C$<em>5$H$</em>{12}$O</td>
<td>1-pentanol</td>
<td>1.15</td>
<td>0.05</td>
<td>1.35</td>
<td>0.20</td>
</tr>
<tr>
<td>Formula</td>
<td>Compound</td>
<td>Exp. log(K_{i,OW})</td>
<td>Uncertainty</td>
<td>Predicted log(K_{i,OW})</td>
<td>Deviation</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------</td>
<td>-----------------------</td>
<td>-------------</td>
<td>----------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>C₆H₁₂O₂</td>
<td>ethanol</td>
<td>-0.17</td>
<td>0.20</td>
<td>-0.62</td>
<td>0.45</td>
</tr>
<tr>
<td>C₄H₆O₂</td>
<td>propanoic acid</td>
<td>0.33</td>
<td>0.10</td>
<td>0.04</td>
<td>0.29</td>
</tr>
<tr>
<td>C₆H₁₂O₂</td>
<td>butanoic acid</td>
<td>0.79</td>
<td>0.10</td>
<td>0.69</td>
<td>0.10</td>
</tr>
<tr>
<td>C₅H₁₀O₂</td>
<td>pentanoic acid</td>
<td>1.39</td>
<td>0.15</td>
<td>1.34</td>
<td>0.05</td>
</tr>
<tr>
<td>C₆H₁₂O₂</td>
<td>hexanoic acid</td>
<td>1.92</td>
<td>0.15</td>
<td>1.99</td>
<td>0.07</td>
</tr>
<tr>
<td>C₄H₁₀O₂</td>
<td>2-methylpentanoic acid</td>
<td>1.80</td>
<td>0.15</td>
<td>1.75</td>
<td>0.05</td>
</tr>
<tr>
<td>C₄H₁₀O₂</td>
<td>2-ethylhexanoic acid</td>
<td>1.68</td>
<td>0.20</td>
<td>1.75</td>
<td>0.07</td>
</tr>
<tr>
<td>C₆H₁₂O₂</td>
<td>1,6-hexanedioic acid</td>
<td>0.08[348]</td>
<td>n/a</td>
<td>-0.24</td>
<td>0.32</td>
</tr>
<tr>
<td>C₅H₁₂O₂</td>
<td>1,7-heptanedioic acid</td>
<td>0.61[349]</td>
<td>n/a</td>
<td>0.40</td>
<td>0.21</td>
</tr>
<tr>
<td>C₆H₁₆O₂</td>
<td>octanoic acid</td>
<td>3.05</td>
<td>0.20</td>
<td>3.27</td>
<td>0.22</td>
</tr>
<tr>
<td>C₅H₁₀O₂</td>
<td>2-propylpentanoic acid</td>
<td>2.75</td>
<td>0.25</td>
<td>3.04</td>
<td>0.29</td>
</tr>
<tr>
<td>C₄H₁₀O₂</td>
<td>2-ethylhexanoic acid</td>
<td>2.64</td>
<td>0.25</td>
<td>3.04</td>
<td>0.40</td>
</tr>
<tr>
<td>C₆H₁₂O₂</td>
<td>2-propylhexanoic acid</td>
<td>3.01</td>
<td>0.30</td>
<td>3.69</td>
<td>0.68</td>
</tr>
<tr>
<td>C₅H₁₂O₂</td>
<td>1,9-nonanedioic acid</td>
<td>1.57[348]</td>
<td>n/a</td>
<td>1.69</td>
<td>0.12</td>
</tr>
<tr>
<td>C₁₀H₂₀O₂</td>
<td>decanoic acid</td>
<td>4.09</td>
<td>0.30</td>
<td>4.56</td>
<td>0.47</td>
</tr>
<tr>
<td>C₁₂H₂₆O₂</td>
<td>2-butyllhexanoic acid</td>
<td>3.55</td>
<td>0.40</td>
<td>4.33</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Table 5.3: Continued.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Compound</th>
<th>Exp. $\log K_{i,OW}$</th>
<th>Uncertainty</th>
<th>Predicted $\log K_{i,OW}$</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{10}H_{18}O_4$</td>
<td>1,10-decanedioic acid</td>
<td>2.20[350]</td>
<td>n/a</td>
<td>2.33</td>
<td>0.13</td>
</tr>
<tr>
<td>$C_{12}H_{24}O_2$</td>
<td>decaenoic acid</td>
<td>4.60</td>
<td>0.50</td>
<td>5.84</td>
<td>1.24</td>
</tr>
<tr>
<td>$C_{14}H_{26}O_2$</td>
<td>tetradecanoic acid</td>
<td>6.10</td>
<td>0.60</td>
<td>7.13</td>
<td>1.03</td>
</tr>
<tr>
<td>$C_{16}H_{32}O_2$</td>
<td>hexadecanoic acid</td>
<td>7.17</td>
<td>0.70</td>
<td>8.41</td>
<td>1.24</td>
</tr>
<tr>
<td>$C_{18}H_{36}O_2$</td>
<td>octadecanoic acid</td>
<td>8.23</td>
<td>0.82</td>
<td>9.69</td>
<td>1.46</td>
</tr>
<tr>
<td>$C_{20}H_{40}O_2$</td>
<td>eicosanoic acid</td>
<td>9.29</td>
<td>0.93</td>
<td>10.97</td>
<td>1.68</td>
</tr>
<tr>
<td>$C_6H_6$</td>
<td>benzene</td>
<td>2.13</td>
<td>0.10</td>
<td>2.29</td>
<td>0.07</td>
</tr>
<tr>
<td>$C_7H_8$</td>
<td>methylbenzene</td>
<td>2.73</td>
<td>0.10</td>
<td>2.78</td>
<td>0.05</td>
</tr>
<tr>
<td>$C_8H_{10}$</td>
<td>1,2-dimethylbenzene</td>
<td>3.12</td>
<td>0.20</td>
<td>3.37</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>1,3-dimethylbenzene</td>
<td>3.20</td>
<td>0.20</td>
<td>3.37</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>1,4-dimethylbenzene</td>
<td>3.15</td>
<td>0.20</td>
<td>3.37</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>ethylbenzene</td>
<td>3.15</td>
<td>0.20</td>
<td>2.91</td>
<td>0.24</td>
</tr>
<tr>
<td>$C_9H_{12}$</td>
<td>isopropylbenzene</td>
<td>3.66</td>
<td>0.20</td>
<td>3.35</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>1,2,3-trimethylbenzene</td>
<td>3.60</td>
<td>0.20</td>
<td>3.95</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>1,2,4-trimethylbenzene</td>
<td>3.63</td>
<td>0.15</td>
<td>3.95</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>1,3,5-trimethylbenzene</td>
<td>3.42</td>
<td>0.30</td>
<td>3.95</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>1-methyl-2-ethylbenzene</td>
<td>3.53</td>
<td>0.30</td>
<td>3.49</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>1-methyl-4-ethylbenzene</td>
<td>3.63</td>
<td>0.40</td>
<td>3.49</td>
<td>0.14</td>
</tr>
<tr>
<td>$C_{10}H_{14}$</td>
<td>4-isopropyltoluene</td>
<td>4.10</td>
<td>0.10</td>
<td>3.93</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>1,2,3,4-tetramethylbenzene</td>
<td>4.00</td>
<td>0.20</td>
<td>4.54</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>1,2,3,5-tetramethylbenzene</td>
<td>4.10</td>
<td>0.20</td>
<td>4.54</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>1,2,4,5-tetramethylbenzene</td>
<td>4.10</td>
<td>0.20</td>
<td>4.54</td>
<td>0.44</td>
</tr>
<tr>
<td>$C_{11}H_{16}$</td>
<td>$n$-pentylbenzene</td>
<td>4.90</td>
<td>0.30</td>
<td>4.94</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Table 5.3: Continued.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Compound</th>
<th>Exp. log$K_{i,OW}$</th>
<th>Uncertainty</th>
<th>Predicted log$K_{i,OW}$</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C$<em>{12}$H$</em>{18}$</td>
<td>$n$-hexylbenzene</td>
<td>5.52</td>
<td>0.30</td>
<td>5.47</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>hexamethylbenzene</td>
<td>4.75</td>
<td>0.25</td>
<td>5.71</td>
<td>0.96</td>
</tr>
<tr>
<td>C$<em>{14}$H$</em>{14}$</td>
<td>bibenzyl</td>
<td>4.70</td>
<td>0.20</td>
<td>4.13</td>
<td>0.57</td>
</tr>
<tr>
<td>C$<em>{14}$H$</em>{22}$</td>
<td>$n$-octylbenzene</td>
<td>6.30</td>
<td>0.40</td>
<td>6.75</td>
<td>0.45</td>
</tr>
<tr>
<td>C$<em>{16}$H$</em>{26}$</td>
<td>$n$-decylbenzene</td>
<td>7.35</td>
<td>0.40</td>
<td>8.04</td>
<td>0.69</td>
</tr>
<tr>
<td>C$<em>{12}$H$</em>{16}$O$_2$</td>
<td>6-phenylhexanoic acid</td>
<td>3.30[351]</td>
<td>n/a</td>
<td>3.22</td>
<td>0.08</td>
</tr>
<tr>
<td>C$<em>{14}$H$</em>{18}$O$_2$</td>
<td>7-phenylheptanoic acid</td>
<td>3.65[351]</td>
<td>0.15</td>
<td>3.86</td>
<td>0.21</td>
</tr>
<tr>
<td>C$<em>{14}$H$</em>{20}$O$_2$</td>
<td>8-phenyloctanoic acid</td>
<td>4.00[351]</td>
<td>0.10</td>
<td>4.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Average deviation

0.29

0.38

5.3.2 Solubility

After successfully predicting the partition coefficients of solutes in different solvent mixtures, the SAFT-$\gamma$ Mie method is employed for a solubility prediction of a simple API. In previous works [169, 335], the solubility prediction for several organic compounds has confirmed the ability of the SAFT-$\gamma$ Mie method to provide quantitative agreement with experimental data. In this section, the solubility of azelaic acid, a topical anti-inflammatory drug comprising only a linear alkyl chain and two carboxylic acid groups (structure shown in Fig. 5.3), is investigated using Eq. (5.3) where the $\Delta c_p(T_m)$ term is assumed to have negligible effect on the solubility. The $T_m$ and $\Delta H_f$ values are taken from Ref. [352]. In Fig. 5.3a), the solubility prediction of azelaic acid in stearic acid and stearic acid in azelaic acid together with the eutectic point for the mixture compared to the experimental data [353] is illustrated. The experimental data for $T_m$ and $\Delta H_f$ for the stearic acid are taken from Ref. [354]. A good agreement between the predictions and the experimental data for the solubilities and the eutectic point emphasises the applicability of the SAFT-$\gamma$ Mie method. In Fig. 5.3b), the azelaic acid solubility prediction in water is shown in
comparison to three sets of experimental data [355–357]. As can be seen in the figure, there are inconsistencies between the sets of experimental data in the literature. Examining the figure, one may propose to use the theoretical SAFT-γ Mie approach to discriminate between the different sets. Our predictions find remarkable agreement with the data of Chen et al [356]. These authors attribute in their paper the deviation between the different sets of experimental data principally to the different measurement techniques used in each of the studies and to standard experimental error. Though it is worth mentioning that the differences observed appear larger than might be expected as arising only from the use of different techniques. Overall, the quantitative agreement between the SAFT-γ Mie predictions and the experimental data demonstrates the applicability of the method to predict the solubility of this API in different solvents.

The solubilities of more complex APIs, such as ibuprofen, ketoprofen, lovastatin and simvastatin, in different solvents as predicted using the SAFT-γ Mie approach will be presented in the next chapter. Thermodynamic properties of these multifunctional compounds may also involve the competition between inter- and intramolecular hydrogen bonding which will be discussed in detail there.

Figure 5.3: Solubility predictions (solid-liquid equilibrium) of (a) azelaic acid (structure shown in the figure) and stearic acid at ambient pressure [353] and (b) azelaic acid in water (triangles [355], diamonds [356] and circles [357]) at $p = 0.101$ MPa. The symbols represent experimental data, the continuous curves the solubility predictions, and the dashed curve the solidus line of the mixture predicted by the SAFT-γ Mie approach.
5.4 Conclusions

In this chapter, the robustness of the group parameters obtained from Chapters 3 and 4 is demonstrated in view of partition coefficient and solubility predictions. The SAFT-\(\gamma\) Mie approach is shown to provide an accurate prediction for the partition coefficients of several families of organic compounds in both alkane-water and alcohol-water mixtures, i.e., the \(K_{i,C_6W}\), \(K_{i,C_{18}W}\), and \(K_{i,OW}\). The average absolute errors are 0.19 \((N_p = 28)\), 0.16 \((N_p = 60)\), and 0.38 \((N_p = 78)\), for \(\log K_{i,C_6W}\), \(\log K_{i,C_{18}W}\), and \(\log K_{i,OW}\), respectively, for alkane, alcohol, carboxylic acid, and aromatic solutes. Obtaining accurate predictions of partition coefficients of these chemical families is very much reliant on an accurate description of the alkane solubility in the aqueous phase; this in turn relates to the group interaction parameters between the alkyl groups, which are a fundamental building block for many compounds, and water. The first example of the predictive capability of the SAFT-\(\gamma\) Mie approach for the solubility of active pharmaceutical ingredients is also presented. The solubility predictions of azelaic acid in stearic acid and in water are shown to be in a very good agreement with the corresponding experimental data. The SAFT-\(\gamma\) Mie approach captures a few order-of-magnitude difference in solubility of the API in these two solvents accurately. These results are especially important considering no data related to the APIs was used to obtain the model parameters. The findings in this chapter emphasise the applicability and generality of SAFT-\(\gamma\) Mie theory and confirm it as a promising tool in modelling complex molecules of relevance to pharmaceutical systems. In the next chapter, thermodynamic properties of more complex solutes, which may undergo intramolecular hydrogen bonding, will be investigated based on the same model used in this chapter.
Chapter 6

The competition between inter- and intramolecular association in modelling partition coefficients and solubilities of organic molecules

6.1 Introduction

After having successfully applied the SAFT-\(\gamma\) Mie model for the partition coefficient and solubility prediction for several families of organic compounds, in this chapter, the model is applied for the study of more complex molecules with properties that may be governed by a competition between inter- and intramolecular hydrogen bonding.

The properties and molecular structure of organic compounds are determined from the interplay of several molecular interactions such as van der Waals' forces, electrostatic interactions, and hydrogen bonds. For example, the large difference in boiling points of ethane and ethanol is mainly due to the different type and degree of intermolecular interactions that stabilise them. Ethane molecules interact weakly by dispersion forces with a gas phase being stable at ambient conditions, while ethanol molecules are held together closely by a network of hydrogen bonds and forming a liquid at ambient conditions. The formation of hydrogen bonds plays a fundamental role in determining the properties of organic- and biomolecules; it is central to water's life-providing properties [358] and contributes to the architecture of DNA and functional form of proteins in their secondary and tertiary structures [359].
Hydrogen bonds do not only occur between molecules but also occur intramolecularly. The intramolecular hydrogen bond (IMHB) is an attractive interaction in which an electropositive hydrogen atom interacts with an electronegative fragment of the same molecule and stabilises the molecule in a gauche or syn conformation [360]. The hydrogen bond is strong enough to restrict the rotation of the fragment by forming, most commonly, five- to eight-membered rings [27]. In addition to electronegative atoms, e.g., O and N, the electronegative fragment can be a π-group in the molecule. As examples, the n-propylbenzene and phenylacetic acid shown in Fig. 6.1 contain a phenyl ring, which acts as the π-group. A number of experimental [361, 362] and ab initio [363, 364] studies have confirmed the existence of the gauche conformer of alkylbenzenes such as n-propylbenzene, n-butylbenzene, and isobutylbenzene. This nonconventional XH/π (X = O, N, C, etc.) hydrogen bond is prevalent in organic compounds [364]. The presence of intramolecular hydrogen bonding has been shown to significantly alter the molecular properties due to the formation of various conformers, which in turn influence thermodynamic properties such as solubility, partition coefficient, and biochemical phenomena [365–367]. For instance, the 2-hydroxybenzoic acid (salicylic acid), commonly known to form IMHB, has a distinct solubility profile from other isomers which do not form IMHB [366]. For molecules with IMHB, the hydrogen bond donor and acceptor are used intramolecularly, shielding the polarity from the environment, and therefore resulting in an increase in lipophilicity and higher membrane permeability [368]. In drug design, the IMHB stretches the Lipinski’s ‘rule-of-five’ [2] (i.e., hydrogen bond donor and acceptor counts) allowing a more diverse drug design beyond the ‘rule-of-five’ chemical space [367, 369].

![Diagram](image)

**Figure 6.1**: Intramolecular association via a CH/π hydrogen bond of (I) n-propylbenzene and an OH/π hydrogen bond of (II) phenylacetic acid
Chapter 6: Considering intramolecular association

Intramolecular hydrogen bonds can be investigated directly using various spectroscopic techniques [370–374]. However, given that the IMHB affects the thermodynamic properties of molecules, we can use a reverse approach to envisage the existence and impact of the bond from thermodynamic data. For instance, the partitioning equilibrium of solutes between water and immiscible organic solvents, which has long been used primarily to model physicochemical and transport properties [28], can be employed to investigate the intramolecular and intermolecular interactions between the solute and each solvent [375, 376]. The partition coefficients are rich in structural information [377] and solvatochromic parameters for linear solvation energy relationships (LSER), e.g., the solute dipolarity/polarisability, H-bond donor acidity, and H-bond acceptor basicity [378, 379]. Among the pairs of solvents used, the octanol-water mixture is the most commonly used system, possibly because it serves as an ideal surrogate in modelling biological activity [29]. As a result, the octanol-water partition coefficient $K_{i,OW}$ is considered a key physicochemical parameter in medicinal chemistry [194]. The partition coefficients in other systems, e.g., hexane-water [347, 380], hexadecane-water [346, 347], and cyclohexane-water [376] also provide useful information for the LSER. In fact, the difference between $\log K_{i,OW}$ and $\log K_{i,Alkane-W}$ values ($\Delta \log K$ or $\Delta \log P$) has been used as a measure of the hydrogen bonding potential of a molecule (including IMHB) and is identified as a target for predictive modelling [347, 375].

Despite the very pronounced effect of IMHB on molecular structure and physical properties [27], they are often under-recognised and seldom predicted. This is because the IMHB investigation from the $\Delta \log K$ requires the prediction of the partition coefficient in both octanol-water and alkane-water systems which is usually not feasible from a single empirical model. In addition, the thermodynamic equilibrium of gauche versus open conformation, which is governed by a competition between inter- and intramolecular interactions, depends on a number of complex factors including the type of solvent and geometry. These interrelated factors make the IMHB difficult to examine based on a 2D structure. Given the considerable impact and ongoing challenge of IMHB, the main objective of this chapter is to model thermodynamic properties of complex molecules which are capable of forming IMHB.

Recently, Shalaeva et al. [381] have used COSMO-RS (conductor-like screening model for real solvents) [175, 382] to describe compounds’ propensity to form IMHB from $\Delta \log K_{i,oct-tol}$ (difference between $\log K_{i,OW}$ and $\log K_{i,toluene-W}$). The COSMO-RS method was chosen because it takes the 3D structure of the molecule into account (inherited from the QM nature of the method) and also allows for the calculation of $\log K$ values in non-octanol/water systems. Given
that no general guidelines on the interpretation of $\Delta \log K$ in relation to the presence and the strength of IMHB are reported in the literature, the propensity of a compound to form IMHB is then indicated by comparing the $\Delta \log K$ of the target compound to the control value ($\Delta \log K$ of similar lettered compounds that are not capable of forming IMHB). This a priori approach is especially attractive for prospective design of IMHB, however, the categorisation becomes uncertain when the $\Delta \log K$ values of sample and control are close to each other. Moreover, the choice of the control compound can be subjective to user’s experience.

In addition to QM-based approaches, e.g., the COSMO-RS [175] and COSMO-SAC [96] (COSMO-segment activity coefficient) for the modelling of compounds with IMHB, molecular dynamic simulations have been used to estimate partition coefficients with an account for intramolecular bonding energy [66]. On the other hand, the property prediction for multifunctional molecules with IMHB is a real challenge for group-contribution (GC) approaches [16] despite their success in estimating many physical and thermodynamic properties of pure substances and mixtures. GC methods dissect a molecule into distinct chemical groups, which are assumed to be independent of each other, and the overall properties of the molecule are obtained by summing up the contributions from all groups. This assumption leads to one of the common drawbacks of the GC methods that occurs in a molecule with two or more functional groups strongly interacting with each other, in this context via IMHB, such that the groups are no longer independent. Although a number of molecular-based GC approaches including UNIFAC (universal quasichemical functional-group activity coefficients model) [20, 112] and GC-PPC-SAFT (GC polar perturbed-chain statistical association fluid theory) [167, 168] have been used to calculate partition coefficients and solubility of organic compounds including APIs, none of these approaches has taken the IMHB into consideration for their predictions. Lin and Sandler [100, 172] have developed a group contribution solvation (GCS) model, which was parameterised based on the result from previous QM calculations, and used it to predict $K_{1,OW}$. Similar to other GC approaches, the model encountered some limitations in the case of multifunctional compounds due to proximity effects (which could also arise from IMHB). A correction method [173] was later employed to correct for the structure and proximity effects through QM calculations for the isolated molecule. Another common solution to address this issue is to use empirical approaches, such as including additional functional groups and introducing second- and higher-order functional groups to account for the intramolecular interactions between functional groups [383].
Chapter 6: Considering intramolecular association

To the best of our knowledge, none of the GC methods has been used to successfully model compounds that pose IMHB without the use of higher-order functional groups or the assistance from QM calculations, which can be computationally expensive. In this chapter, a heteronuclear group-contribution equation of state, SAFT-γ Mie, is used to predict the partition coefficient and solubility of various organic compounds including APIs, some of which contain IMHB. In the previous chapters, the SAFT-γ Mie method has been shown to deliver an accurate description of phase equilibria over a broad range of compositions of aqueous mixtures, including the infinite dilution regime, as well as a quantitative prediction of the partition coefficients and solubility. These developments make the method promising for the thermodynamic modelling of multifunctional molecules in a pharmaceutical context. Moreover, the SAFT-γ Mie approach takes into account the hydrogen bonding energy explicitly which enables the association sites to be switched on and off, i.e., the unlike induced association-site models [240], when the associating group is present in different environments. This highly versatile association-site scheme leads to a novel approach for the effective treatment of the IMHB within the SAFT-type equation of state.

A brief outline of the definition of groups and the GC model for APIs are described in the next section. The effective treatment of the IMHB within the SAFT-γ Mie framework is elaborated next. In the following section, the partition coefficients of complex molecules which are capable of forming IMHB are investigated. The effective treatment for the IMHB is then applied to the prediction of the octanol-water partition coefficient and solubility of APIs without the assistance from QM calculations or resorting to higher-order functional groups.

6.2 Definition of groups

The definition of functional groups in group-contribution approaches is often arbitrary; the final choice is usually the combination of groups that results in the best representation of the experimental data. In our work, the SAFT-γ Mie functional groups are developed by considering the fine balance between transferability and accuracy of the model to best represent the fluid-phase equilibrium data, e.g., vapour-liquid equilibrium and liquid-liquid equilibrium data [169, 384]. The functional groups are chosen to be consistent with a physical representation of the molecules/chemical groups, e.g., the number of associating sites is determined with respect to the number of lone pairs of electrons and hydrogen atoms. The models have been designed to be versatile for a wide range of thermodynamic conditions. No partition coefficients or solid-liquid equilibrium data are used to develop the model. All functional group parameters used in this
Chapter 6: Considering intramolecular association

chapter are taken from our previous publications [169, 291, 384] and Chapters 3 and 4 of this thesis (summarised in Table 4.2); the details on how each of the group parameters is developed can be found there.

The main solvents in this chapter include water, acetone, alkanes, alcohols, acetates, and their mixtures. In the SAFT-\(\gamma\) Mie framework, water and acetone are modelled as molecular groups with association sites. The H\(_2\)O group for water has two \(e_1\) sites and two H sites [239] and the CH\(_3\)COCH\(_3\) group for acetone has one each of \(e_1\), \(e_2\), and H sites [240]. Alkanes are composed of the CH\(_3\) and CH\(_2\) groups. Alcohols and acids contain the CH\(_2\)OH (or OH) and COOH groups, respectively, in addition to the alkyl groups to represent their functionalities. These groups are also modelled with association sites, the CH\(_2\)OH and OH groups contain two \(e_1\) sites and one H site (cf. Fig. 3.1), and the COOH group contains two \(e_1\) sites, two \(e_2\) sites and one H site (cf. Fig. 4.1). The COO group is used to represent the functional group of acetate; its two \(e_1\) sites are activated when modelled with the OH group. The SAFT-\(\gamma\) Mie models for aliphatic solutes are equivalent to the solvents. The CH and C functional groups are used to represent the branched compounds. For instance, 2,2,4-trimethylpentane (isooctane) is composed of 5\(\times\) CH\(_3\), 1\(\times\) CH\(_2\), 1\(\times\) CH, and 1\(\times\) C groups. The cCH\(_2\) and CH=/C= are used to model the cyclic and unsaturated hydrocarbons, respectively. The aromatic hydrocarbons are modelled with aCH, aCCH, aCCH\(_2\) and aCCH\(_3\) groups. Each of the aromatic functional groups also carries one \(e_1\) site to mediate the interaction with the H\(_2\)O group. The decomposition of the APIs studied in this chapter is shown Table 6.1.
Table 6.1: Active pharmaceutical ingredients mentioned in this chapter and their decomposition into SAFT-γ Mie groups.

<table>
<thead>
<tr>
<th>API</th>
<th>Group</th>
<th>Occurrence</th>
<th>API</th>
<th>Group</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Succinic acid</td>
<td>CH₂</td>
<td>2</td>
<td>Valproic acid</td>
<td>CH₃</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>COOH</td>
<td>2</td>
<td></td>
<td>CH₂</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CH</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>COOH</td>
<td>1</td>
</tr>
<tr>
<td>Azelaic acid</td>
<td>CH₂</td>
<td>7</td>
<td>Phenylacetic acid</td>
<td>aCH</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>COOH</td>
<td>2</td>
<td></td>
<td>aCCH₂</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>COOH</td>
<td>1</td>
</tr>
<tr>
<td>Hydratropic acid</td>
<td>CH₃</td>
<td>1</td>
<td>Hydrocinnamic acid</td>
<td>CH₂</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>aCH</td>
<td>5</td>
<td></td>
<td>aCH</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>aCCH</td>
<td>1</td>
<td></td>
<td>aCCH₂</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>COOH</td>
<td>1</td>
<td></td>
<td>COOH</td>
<td>1</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>CH₃</td>
<td>3</td>
<td>Ketoprofen</td>
<td>CH₃</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>COOH</td>
<td>1</td>
<td></td>
<td>COOH</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>aCH</td>
<td>4</td>
<td></td>
<td>aCH</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>aCCH</td>
<td>1</td>
<td></td>
<td>aCCH</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>aCCH₂</td>
<td>1</td>
<td></td>
<td>aCCOaC</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CH</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lovastatin</td>
<td>CH₃</td>
<td>4</td>
<td>Simvastatin</td>
<td>CH₃</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>CH₂</td>
<td>3</td>
<td></td>
<td>CH₂</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>CH</td>
<td>8</td>
<td></td>
<td>CH</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>cCH₂</td>
<td>3</td>
<td></td>
<td>cCH₂</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>COO</td>
<td>2</td>
<td></td>
<td>COO</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CH=</td>
<td>3</td>
<td></td>
<td>CH=</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>C=</td>
<td>1</td>
<td></td>
<td>C=</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>OH</td>
<td>1</td>
<td></td>
<td>OH</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td>1</td>
</tr>
</tbody>
</table>
6.3 Effective treatment for intramolecular association and gauche conformation

In the case of molecules with IMHB forming a gauche conformation, an effective treatment is applied to the SAFT-\(\gamma\) Mie model by deactivating the association sites that are involved in the intramolecular association or that are inaccessible by other associating groups in the puckered conformation of the molecule. The SAFT-\(\gamma\) Mie models for \(n\)-propylbenzene and phenylacetic acid are composed of aCH, aCCH\(_2\), CH\(_2\), CH\(_3\) (in case of \(n\)-propylbenzene) and COOH (in case of phenylacetic acid) groups. The aCCH\(_2\) and COOH groups in the SAFT-\(\gamma\) Mie approach are modelled with association site(s) in certain mixtures. In the alcohol-water system, the e\(_1\) site on aCCH\(_2\) group is able to associate with the H site on H\(_2\)O group and the e\(_1\), e\(_2\), and H site on COOH group is able to associate with H\(_2\)O and CH\(_2\)OH groups [384]. In addition, the COOH group is able to dimerise with another COOH group via its H site [240]. Within the syn conformations of \(n\)-propylbenzene and phenylacetic acid (Fig. 6.1), the terminal CH\(_3\) and the (CO)OH groups are gauche to the phenyl ring due to the CH/\(\pi\) and OH/\(\pi\) intramolecular hydrogen bonds, respectively. This conformation shields the e\(_1\) site on the aCCH\(_2\) group of the phenyl ring to associate with the H\(_2\)O group and hinders the H site on the COOH group to associate with H\(_2\)O, CH\(_2\)OH, and another COOH group. The straightforward effect of the IMHB is the reduction of availability of the association site for intermolecular association, therefore we “mimic” this effect by switching off the site involved in the IMHB so that they are no longer available to bond with the solvents or other associating groups. Examples for the effective treatment for molecules existing in a syn or a gauche conformation are illustrated in Fig. 6.2.
Figure 6.2: The effective treatment for gauche conformations of (a) \( n \)-propylbenzene, (b) phenylacetic acid, (c) propanedioic acid, and (d) ethylene glycol within the SAFT-\( \gamma \) Mie framework.

In modelling \( n \)-propylbenzene (Fig. 6.2a) only the \( e_1 \) site on the aCCH\(_2\) group is switched off (the CH\(_3\) group has no association site) inhibiting the association with H\(_2\)O while accentuating the influence of the unlike dispersion interaction between aCCH\(_2\) and CH\(_3\) groups. For the phenylacetic acid (Fig. 6.2b), the \( e_1 \) site on the aCCH\(_2\) group as well as the H site on the COOH group are deactivated. It is worth noting that the COOH group can still associate with the H\(_2\)O and CH\(_2\)OH groups via its \( e_1 \) and \( e_2 \) sites since only its H site is switched off. This treatment is applied when the terminal group is able to interact with its phenyl ring to form five- to six-membered ring for alkylbenzene (up to \( n \)-butylbenzene) and five- to eight-membered ring for phenylacetic acid (up to 5-phenylpentanoic acid), as well as to APIs that present the same characteristic, i.e., ibuprofen and ketoprofen.

A similar idea is applied to the effective treatment for the gauche conformation of non-aromatic compounds, e.g., malonic acid (propanedioic acid) and ethylene glycol (1,2-ethanediol) as shown in Figs. 6.2c) and 6.2d). For some dicarboxylic acids, e.g., propanedioic acid, in which the IMHB between two COOH groups favours the gauche conformation, the sites involved in the intramolecular association (one \( e_1 \) site on one COOH group and the H site on the other COOH group) are switched off. The criteria for this effective treatment is similar to the phenylacetic acid in which the intramolecular rings of five- to eight-member are allowed, i.e., from ethanedioic
acid up to pentanedioic acid. For ethylene glycol, which is composed of two CH₂OH groups, one H site on the first CH₂OH and both e₁ sites on the other CH₂OH are switched off since these sites are shielded in the gauche conformation which is the predominant conformation of ethylene glycol \[385\].

6.4 Results

6.4.1 Partition coefficients of alkylbenzenes, phenylalkanoic acids, and dicarboxylic acids

Having fully characterised and described the SAFT-\(\gamma\) Mie models, they are then used to investigate molecules which their thermodynamic properties may be governed by the competition between inter- and intramolecular hydrogen bond. In the literature, experimental evidence [364] and \textit{ab initio} calculations [363] suggest the predominance of the \textit{syn} conformation for \(n\)-propylbenzene, \(n\)-butylbenzene, and isobutylbenzene. We apply our SAFT-\(\gamma\) Mie model to “mimic” the CH/\(\pi\) interaction (Fig. 6.1) that stabilises the \textit{syn} conformation of these compounds as described in the previous section. The alkane-water and octanol-water partition coefficients predicted by the theory (via Eq. (5.1)) with the effective IMHB treatment generally have better agreement with the experimental data in comparison to the model without the treatment (Tables 6.2 - 6.4). The average absolute errors, AAE = \(\frac{1}{N_p} \sum_{i=1}^{N_p} |\log K_i^{exp} - \log K_i^{calc}|\), have improved slightly for all cases of the partition coefficients as summarised in Table 6.5. The change in the partition coefficient predictions of alkylbenzenes after the treatment is found to be minor, around 3 % for all values, because only the e₁ site on the aCCH₂ group is switched off and this site only interacts with H₂O (cf. Fig. 6.2a). The treatment impacts the results much more significantly when applied to treating molecules with more and/or stronger associating groups. This can be observed in the case of the \(K_{i,OW}\) prediction of phenylalkanoic acids (Table 6.4) for which the intramolecular treatment results up to 75 % change in the \(\log K_{i,OW}\) value of phenylacetic acid. Consequently, the AAE value improves more significantly compared to the prediction for less associating solutes, alkylbenzenes, as seen in Table 6.5.
**Table 6.2:** Hexane-water partition coefficient $K_{i,C_6W}$ of alkylbenzene (at $T = 298.15$ K and $p = 0.101$ MPa) with and without intramolecular association treatment. The deviation is defined as $|\log K_{i,C_6W}^{\text{exp}} - \log K_{i,C_6W}^{\text{calc}}|$. \\

<table>
<thead>
<tr>
<th>Compound</th>
<th>Exp. $\log K_{i,C_6W}$[347]</th>
<th>With IMHB treatment</th>
<th>Deviation</th>
<th>Without IMHB</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$-propylbenzene</td>
<td>4.11</td>
<td>4.15</td>
<td>0.04</td>
<td>4.03</td>
<td>0.08</td>
</tr>
</tbody>
</table>

**Table 6.3:** Hexadecane-water partition coefficient $K_{i,C_{16}W}$ of alkylbenzenes (at $T = 298.15$ K and $p = 0.101$ MPa) with and without intramolecular association treatment. The deviation is defined as $|\log K_{i,C_{16}W}^{\text{exp}} - \log K_{i,C_{16}W}^{\text{calc}}|$. \\

<table>
<thead>
<tr>
<th>Compound</th>
<th>Exp. $\log K_{i,C_{16}W}$[347]</th>
<th>With IMHB treatment</th>
<th>Deviation</th>
<th>Without IMHB</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$-propylbenzene</td>
<td>3.84</td>
<td>3.81</td>
<td>0.03</td>
<td>3.69</td>
<td>0.15</td>
</tr>
<tr>
<td>$n$-butylbenzene</td>
<td>4.44</td>
<td>4.49</td>
<td>0.05</td>
<td>4.37</td>
<td>0.07</td>
</tr>
<tr>
<td>Isobutylbenzene</td>
<td>4.62</td>
<td>4.31</td>
<td>0.31</td>
<td>4.19</td>
<td>0.43</td>
</tr>
<tr>
<td>Average deviation</td>
<td></td>
<td></td>
<td>0.13</td>
<td>0.22</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.4: Octanol-water partition coefficients \((K_{i,OW})\) of dicarboxylic acids, alkylbenzenes, and phenylalkanoic acids (at \(T = 298.15\) K and \(p = 0.101\) MPa) with and without intramolecular association treatment. The deviation is defined as \(|\log K_{i,OW}^{\text{exp}} - \log K_{i,OW}^{\text{calc}}|\).

<table>
<thead>
<tr>
<th>Compound</th>
<th>Exp. (\log K_{i,OW}[104])</th>
<th>Exp. uncertainty</th>
<th>With IMHB treatment</th>
<th>Deviation</th>
<th>Without IMHB Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>propanedioic acid</td>
<td>-0.81[349]</td>
<td>n/a</td>
<td>-1.03</td>
<td>0.22</td>
<td>-2.17</td>
</tr>
<tr>
<td>1,4-butanedioic acid</td>
<td>-0.59[348]</td>
<td>n/a</td>
<td>-0.39</td>
<td>0.20</td>
<td>-1.53</td>
</tr>
<tr>
<td>1,5-pentanedioic acid</td>
<td>-0.29[349]</td>
<td>n/a</td>
<td>0.25</td>
<td>0.54</td>
<td>-0.88</td>
</tr>
<tr>
<td>(n)-propylbenzene</td>
<td>3.69</td>
<td>0.15</td>
<td>3.66</td>
<td>0.03</td>
<td>3.55</td>
</tr>
<tr>
<td>(n)-butylbenzene</td>
<td>4.26</td>
<td>0.15</td>
<td>4.30</td>
<td>0.04</td>
<td>4.19</td>
</tr>
<tr>
<td>isobutylbenzene</td>
<td>4.01[386]</td>
<td>n/a</td>
<td>4.08</td>
<td>0.07</td>
<td>3.97</td>
</tr>
<tr>
<td>phenylacetic acid</td>
<td>1.41</td>
<td>0.15</td>
<td>1.12</td>
<td>0.29</td>
<td>0.64</td>
</tr>
<tr>
<td>2-phenylpropanoic acid</td>
<td>1.80</td>
<td>0.20</td>
<td>1.55</td>
<td>0.25</td>
<td>1.07</td>
</tr>
<tr>
<td>3-phenylpropanoic acid</td>
<td>1.84</td>
<td>0.15</td>
<td>1.77</td>
<td>0.07</td>
<td>1.29</td>
</tr>
<tr>
<td>3-methylphenylacetic acid</td>
<td>1.95</td>
<td>0.20</td>
<td>1.72</td>
<td>0.23</td>
<td>1.24</td>
</tr>
<tr>
<td>4-methylphenylacetic acid</td>
<td>1.86</td>
<td>0.20</td>
<td>1.72</td>
<td>0.14</td>
<td>1.24</td>
</tr>
<tr>
<td>4-phenylbutanoic acid</td>
<td>2.42[387]</td>
<td>n/a</td>
<td>2.41</td>
<td>0.01</td>
<td>1.93</td>
</tr>
<tr>
<td>5-phenylpentanoic acid</td>
<td>2.92[388]</td>
<td>n/a</td>
<td>3.05</td>
<td>0.13</td>
<td>2.58</td>
</tr>
<tr>
<td>Average deviation</td>
<td></td>
<td></td>
<td></td>
<td>0.17</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 6.5: Average absolute error (AAE) for alkane-water \((\log K_{i,C_{6}W} \text{ and } \log K_{i,C_{15}W})\) and octanol-water partition coefficients \((\log K_{i,OW})\) of alkylbenzenes, phenylalkanoic, and dicarboxylic acids with and without intramolecular association treatment.

<table>
<thead>
<tr>
<th>Compound</th>
<th>(\log K_{i,C_{6}W}) with IMHB</th>
<th>(\log K_{i,C_{15}W}) with IMHB</th>
<th>(\log K_{i,OW}) without IMHB</th>
<th>(\log K_{i,OW}) with IMHB</th>
</tr>
</thead>
<tbody>
<tr>
<td>alkylbenzenes</td>
<td>0.08</td>
<td>0.04</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>(N_p = 1)</td>
<td>(N_p = 3)</td>
<td>(N_p = 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>phenylalkanoic acids</td>
<td>-</td>
<td>-</td>
<td>0.60</td>
<td>0.16</td>
</tr>
<tr>
<td>(N_p = 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dicarboxylic acids</td>
<td>-</td>
<td>-</td>
<td>0.96</td>
<td>0.32</td>
</tr>
<tr>
<td>(N_p = 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the effective treatment for phenylalkanoic acids, the e1 site on the aCCH2 group and the H site on the COOH group are assumed to participate in the intramolecular hydrogen bonding.
and are therefore switched off (Fig. 6.2b). For the branched phenylalkanoic acids, e.g., 2-phenylpropanoic acid, which is composed of a acCCH group instead of a acCCH$_2$ group, the $e_1$ site on the acCCH is switched off. In our SAFT-$\gamma$ Mie model, the H site on the COOH group is allowed to associate with the $e_1$ sites of both the H$_2$O and CH$_2$OH groups in our octanol-water model [384], as well as to dimerise with another COOH group [240]. To the best of our knowledge, there is no study regarding the intramolecular hydrogen bond for this chemical family. Due to the greater partial positive charge $\delta^+$ on the H atom of the COOH group (i.e., greater hydrogen-bond donor strength [389]) in comparison to the H atom on the CH$_3$ group of alkylbenzene, we postulate that the intramolecular OH/π interaction in phenylalkanoic acids should be stronger and, therefore, able to gauche longer molecules. The $K_{i,OW}$ prediction by the two SAFT-$\gamma$ Mie models, with and without the intramolecular treatment, of n-phenylalkanoic acids is shown in Fig. 6.3, from which it is clear that the intramolecular treatment is suitable for short molecules. The results also suggest that the effective treatment should be applied up to the 5-phenylpentanoic acid in order to obtain the best representation of the experimental data [104, 351, 387, 388]. For the compounds longer than the 5-phenylpentanoic acid (IMHB forms greater than eight-membered ring), it would be entropically unfavourable for the terminal COOH group to gauche back to associate with the phenyl ring. Therefore, these compounds should be represented by the fully associated model, i.e., the model with only intermolecular association and no sites switched off. After the IMHB assessment, i.e., when the appropriate model is used, the AAE of $K_{i,OW}$ for this compound family reduces from 0.50 to 0.19 (including the branched compounds, see Tables 5.3 and 6.4).
Figure 6.3: The octanol-water partition coefficient ($K_{i,OW}$) of phenyl-$n$-alkanoic acids at $T = 298.15$ K and $p = 0.101$ MPa. The symbols represent experimental data [104, 351, 387, 388], the error bars the experimental uncertainty, the dashed curve the SAFT-γ Mie prediction based on the fully associated model, and the continuous curve the prediction after the IMHB assessment (when the appropriate model is used).

In a similar fashion to the treatment of intramolecular hydrogen bonds of the aromatic compounds, we apply the effective treatment to model $K_{i,OW}$ of short dicarboxylic acids which are able to form IMHB from the two COOH ends of the molecule. The presence of IMHB in malonic acid (propanedioic acid) has been investigated using NMR [390]. The conformation and IMHB of the short homologues of the dicarboxylic acid series (propanedioic acid to hexanedioic acid) have also been investigated using density functional methods (DFT) [391], however, the study suggested that the gauche conformation with IMHB is not the most stable form in aqueous solution. The IMHB treatment within the SAFT-γ Mie method works by deactivating one $e_1$ association site on one COOH group and the H site on the other COOH group, cf. Fig. 6.2c), as these sites are presumed to be involved in IMHB. The $K_{i,OW}$ prediction of the dicarboxylic acid series is shown in Fig. 6.4. The results are in line with the $K_{i,OW}$ prediction of phenylalkanoic acids showing that the treatment significantly improves the $K_{i,OW}$ prediction of the short molecules. Based on these results, the IMHB treatment should be applied to diacids up to 1,5-pentanedioic, which means that a maximum of eight-membered ring is allowed to form from the association between the two COOH ends (same ring size with the 5-phenylpentanoic acid). The AAE of the short diacids: propanedioic acid, 1,4-butanedioic acid, and 1,5-pentanedioic acid has improved
drastically from 0.96 (model without IMHB treatment) to 0.32 as shown in Table 6.5. For molecules longer than the 1,5-pentanedioic acid, the intramolecular association is too unfavourable therefore the fully associated model is used to describe the \(K_{i,OW}\) of these compounds as shown in Fig. 6.4. The overall AAE of this chemical family is 0.25 after the IMHB assessment, i.e., when the appropriate model is used.

Based on the results of partition coefficient predictions of alkylbenzenes, phenylalkanoic, and dicarboxylic acids, it is evident that the effective treatment of the SAFT-\(\gamma\) Mie model has significantly improved the partition coefficient predictions of molecules that present IMHB. We assume that the intramolecular bonding energy, which is not accounted for in the current theory, plays a small contribution to the overall free energy of the molecule compared to the intermolecular hydrogen bonding energy. It is noted that, in our theory, molecules are modelled as existing in a single form while in the real system they coexist in both syn and anti forms [364]. This is a limitation for molecules which have no dominant forms, perhaps, such as the 1,5-pentanedioic acid whose the experimental \(\log K_{i,OW}\) data lies between the two modes of prediction, cf. Fig. 6.4. In the following section, the same concept is applied to the \(K_{i,OW}\) prediction of APIs that present structural similarity to the compounds with IMHB.

![Figure 6.4](image)

**Figure 6.4**: Octanol-water partition coefficients (\(K_{i,OW}\)) of dicarboxylic acids at \(T = 298.15\) K and \(p = 0.101\) MPa. The symbols represent experimental data [348–350], the dashed curve the SAFT-\(\gamma\) Mie prediction based on the fully associated model, and the continuous curve the prediction after the IMHB assessment (when the appropriate model is used).
6.4.2 Octanol-water partition coefficients of APIs

Having successfully predicted partition coefficients of several organic compounds, we follow how the IMHB treatment is applied within the SAFT-γ framework as well as the types of molecules on which the treatment should be implemented. This concept is applied to the $K_{i,OW}$ prediction of APIs and a toxicant, the molecules chosen are multifunctional. The SAFT-γ Mie functional groups used for these compounds are shown in Table 6.1. The $K_{i,OW}$ values of ethylene glycol (toxicant), succinic acid (food additive and dietary supplement), valproic acid (anti-epileptic drug), azelaic acid (topical anti-inflammatory drug), phenylactic acid (drug precursor), hydratropic acid (human metabolite/anti-inflammatory drug), hydrocinnamic acid (food additive) and ibuprofen/ketoprofen (anti-inflammatory drugs) are reported in Table 6.6. In the table, the predicted log$K_{i,OW}$ values before and after the IMHB assessment are reported. The values before the assessment are referred to the values without the IMHB treatment and the values after the assessment can be from either with or without the treatment, such that only the compounds that present IMHB are subjected to the IMHB treatment. For instance, the compounds that contain no characteristic for IMHB, such as valproic acid and azelaic acid, the IMHB treatment is not applied so the $K_{i,OW}$ value after the assessment is the same value from the fully associated model. The IMHB treatment for succinic acid (dicarboxylic acid), phenylactic acid, hydratropic acid (2-phenylpropanoic acid), and hydrocinnamic acid (phenylalkanoic acids) has been discussed in detail in the previous section. Ibuprofen and ketoprofen partly carry the same functional groups (aCCH and COOH groups) as hydratropic acid, so the same intramolecular treatment is applied to treat these APIs. In particular, the $e_1$ site on the aCCH group and the $H$ site on the COOH group are switched off as they are presumed to be involved in the OII/π intramolecular hydrogen bond. As can be seen in Table 6.6, the $K_{i,OW}$ predictions have improved significantly with the treatment; the deviations from the experimental data [392] are reduced to 0.18 for ibuprofen and to 0.06 for ketoprofen. In the case of ethylene glycol, the conformational analysis by NMR spectroscopy indicates that the molecule exists preferentially in the gauche conformation in water and chloroform [385]. Although the NMR spectroscopy suggests that intramolecular hydrogen bonding between the hydroxyl groups is unlikely to be a significant factor in the preferred conformation, the effective treatment can still be applied by deactivating association sites that are inaccessible by the solvent due to the gauche conformation. The treatment for ethylene glycol is done by deactivating two $e_1$ sites on one CH$_2$OH group and one $H$ site on the other CH$_2$OH group as illustrated in Fig. 6.2d). The predicted $K_{i,OW}$ value with the treatment has a very good agreement with experimental data (Table 6.6).
Table 6.6: Octanol-water partition coefficients ($K_{i,OW}$) of pharmaceutical-related molecules (at $T = 298.15$ K and $p = 0.101$ MPa). The deviation is defined as $|\log K_{i,OW}^{exp} - \log K_{i,OW}^{calc}|$.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Experiment</th>
<th>Uncertainty</th>
<th>Before IMHB assessment</th>
<th>Deviation</th>
<th>After IMHB assessment</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethylene glycol</td>
<td>-1.34[393]</td>
<td>n/a</td>
<td>-2.82</td>
<td>1.48</td>
<td>-1.42</td>
<td>0.08</td>
</tr>
<tr>
<td>Succinic acid</td>
<td>-0.59[348]</td>
<td>n/a</td>
<td>-1.53</td>
<td>0.94</td>
<td>-0.37</td>
<td>0.22</td>
</tr>
<tr>
<td>Valproic acid</td>
<td>2.75[104]</td>
<td>0.25</td>
<td>3.04</td>
<td>0.29</td>
<td>3.04</td>
<td>0.29</td>
</tr>
<tr>
<td>Azelaic acid</td>
<td>1.57[348]</td>
<td>n/a</td>
<td>1.69</td>
<td>0.12</td>
<td>1.69</td>
<td>0.12</td>
</tr>
<tr>
<td>Phenylacetic acid</td>
<td>1.41[104]</td>
<td>0.15</td>
<td>0.64</td>
<td>0.77</td>
<td>1.12</td>
<td>0.29</td>
</tr>
<tr>
<td>Hydratropic acid</td>
<td>1.80[104]</td>
<td>0.20</td>
<td>1.07</td>
<td>0.73</td>
<td>1.55</td>
<td>0.25</td>
</tr>
<tr>
<td>Hydrocinnamic acid</td>
<td>1.84[104]</td>
<td>0.15</td>
<td>1.29</td>
<td>0.55</td>
<td>1.77</td>
<td>0.07</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>3.50[392]</td>
<td>0.50</td>
<td>2.84</td>
<td>0.66</td>
<td>3.32</td>
<td>0.18</td>
</tr>
<tr>
<td>Ketoprofen</td>
<td>3.12[392]</td>
<td>n/a</td>
<td>2.58</td>
<td>0.54</td>
<td>3.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>

As some association sites of the solute are deactivated in the effective IMHB treatment, the solute is able to associate with water to a lesser extent and preferentially migrates to the organic phase. The treatment therefore leads to an overall increase in the partition coefficient values. This trend is in line with the experimental observation which indicates that the IMHB results in less polar molecular conformations, increase in lipophilicity, and higher passive membrane permeability [368]. In other words, our modelling of the IMHB reflects the shielding of the hydrogen bond donor and acceptor atoms from water and the decrease in the drug polarity which have been noted experimentally, thereby reducing the energetic penalty of desolvation required in moving from an aqueous environment through a phospholipid bilayer [394].

Despite the presence of a larger number of functional groups and complex interactions in the APIs considered, the $K_{i,OW}$ predictions in this study agree very well with the experimental data. The quantitative predictions of the behaviour of solutes at infinite dilution indicate an appropriate balance within the set of group interaction parameters that have been determine, especially in terms of their like and unlike dispersion and hydrogen bonding energies. In addition, the results confirm the application of the effective IMHB approach in modelling complex interactions such as the competition between inter- and intramolecular association. We emphasise that the prediction of the thermodynamic properties of compounds containing multiple functional groups with multicomponent systems is a very challenging task for a group-contribution approach. Next, we extend the predictive framework of the SAFT-$\gamma$ model to solubility prediction, where neither SLE data nor any data related to the API has been used in the development of the model.
6.4.3 Solubility prediction of APIs

Aromatic compounds

In this section, the solid-liquid-equilibrium solubilities of aromatic APIs are calculated using Eq. (5.3). Phenylacetic acid (PAA) is considered first. Its experimental data for $\Delta H_f$ and $T_m$ are taken from Ref. [395]. Despite relatively simple chemical composition (in terms of our groups the molecule is modelled with $aCH$, $aCCH_2$, and COOH groups, cf. Table 6.1), the $K_{i, OW}$ prediction from the previous section suggests that intramolecular OH/$\pi$ hydrogen bonds, which affect the bulk thermodynamic properties, may also be present. The same effective treatment for IMHB used to model the $K_{i, OW}$ prediction of PAA (Fig. 6.2b) is applied to model the solubility of PAA in ethanol and acetone. In these solvents, the $e_1$ site on the $aCCH_2$ group of PAA is primitively inactive. All of the sites on the COOH group ($H$, $e_1$, $e_2$) are active in ethanol (with the $CH_2OH$ group) but inactive in acetone (Table 4.5). The IMHB treatment for PPA in these systems then works by deactivating only the $H$ site on the COOH group. The comparison of the two models with and without the treatment is shown in Fig. 6.5a). The two models provide similar solubility values because the degree of association of the switched-off sites with other associating groups is very small. The impact comes directly from switching off the $H$ site on the COOH group. Although the COOH group does not associate with the acetone group, the slight difference of the PAA solubility in acetone is seen between the two models because the $H$ site on the COOH group can be used to dimerise with another COOH group (in the fully associated model, without IMHB).

While PPA is a relatively simple molecule, most APIs tend to be more complex and contain many types of functional groups. For example ibuprofen and ketoprofen are multifunctional compounds with at least one aromatic ring, alkyl chains and carboxylic acid groups. The solubility predictions of these APIs in 1-butanol, water, and acetone are displayed in Figs. 6.5b) - 6.5d). The experimental data for $\Delta H_f$ and $T_m$ for ibuprofen and ketoprofen are taken from Refs. [395] and [396], respectively. These predictions are calculated with and without taking into account IMHB (see Sec. 6.4.2). As can be seen from the figures, only a small change in the predicted solubilities is observed when IMHB is considered in the case where the functional groups of the drug (COOH and $aCCH$ groups) weakly associate or do not at all associate with the solvents (1-butanol and acetone), while a more significant change is observed where the functional groups associate strongly with the solvent ($H_2O$); this is the case for both ibuprofen and ketoprofen. In the case of the solubility of ibuprofen in water, despite being one of the most commonly used drugs, a range of values is found in the literatures [397–402] as seen in Fig.
6.5c). The lack of consistency in the data may be due to the use of different buffer solutions, the purity of the API, and the very low solubility values. The model with IMHB treatment gives an excellent agreement with the more recent data [400, 401] and has a similar solubility trend with Ref. [402]. In the case of ketoprofen, the unlike interactions between the aCCOaC group, which characterises the connectivity between two phenyl rings, with other functional groups are not obtained from any experimental data but from combining rules [22]. Despite the approximated interactions, the overall prediction of the solubility of this API is in a very good agreement with experimental data [403–405]. The solubilities of both ibuprofen and ketoprofen vary greatly in different solvents; the solubilities in water of these APIs are four orders of magnitude smaller than those in 1-butanol or acetone and, as can be seen, this difference in solubility is captured accurately by our model. These results validate the model parameters as well as the proposed technique for the treatment of intramolecular association and demonstrate the applicability of the SAFT-\(\gamma\) Mie method in predicting the solubility of a selection of aromatic APIs in different solvents.
Figure 6.5: Prediction of the solid-liquid equilibria in binary systems of APIs in different solvents as a function of temperature at $p = 0.101$ MPa. The symbols represent the experimental data and the continuous and dashed curves the prediction with the SAFT-$\gamma$ Mie approach with and without effective intramolecular treatment, respectively: (a) phenylacetic acid in ethanol (circle [395] and purple curves) and in acetone (triangle [395] and orange curves); (b) ibuprofen (triangles [406] and green curves) and ketoprofen (squares [403] and blue curves) in 1-butanol; (c) ibuprofen (diamonds [397], star [398], plus [399], circle [400], squares [401], triangles [402], and green curves) and ketoprofen (cross [404], pentagons [405], and blue curves) in water; (d) ibuprofen (triangles [406] and green curves) and ketoprofen (circles [405] and blue curves) in acetone. Chemical structures for phenylacetic acid, ibuprofen, and ketoprofen are shown in figure (a), (b), and (c), respectively.

Statins

Statins provide very effective treatment for the management of hypercholesterolaemia by lowering the plasma level of low density lipoprotein (LDL or bad) cholesterol with an exceptional safety record [407]. The drugs’ mode of action is through the inhibition of the (3S-hydroxy)-3-methylglutaryl coenzyme A (HMG-CoA) reductase which results in an upregulation of the LDL receptor. In this manner, the statins reduce human plasma cholesterol by increasing the uptake
of LDL via the LDL receptor [408]. The global market for statins generated over $27 billions in revenues in 2009 [409]. Despite being in clinical use for more than 20 years, there are still considerable efforts in developing novel formulations [410] and understanding structure-energetics relationships of the drugs [411] in order to achieve a better stability and a higher bioavailability after oral administration.

Here we use the SAFT-γ Mie approach to predict the solubility of lovastatin and simvastatin in a number of alcohols and acetates. The functional group parameters characterising the statins have been developed by MSE members at Imperial College and are reported in Tables 4.2 - 4.5. The group-contribution model for lovastatin within the SAFT-γ Mie framework is represented by 4× CH₃, 3× CH₂, 8× CH, 3× cCH₂, 2× COO, 3× CH=, 1× C=, and 1× OH as listed in Table 6.1 and illustrated in Fig. 6.6. The simvastatin molecule is composed of similar building blocks but with 5× CH₃, an additional C group, and only 2× CH₂ groups (cf., Table 6.1). The solvents in this study are 1-alcohols, composed of CH₃, CH₂, and OH groups, and acetates, composed of CH₃, CH₂, CH, C and COO groups. Previously in this thesis, the 1-alcohols were modelled by the CH₂OH functional group but the OH group is used in this case for both the alcohols and the statin molecules. The association scheme for the OH group is the same as the CH₂OH group (2× e₁ and 1× H sites). The COO group is modelled with two e₁ sites which are allowed to associate with other H sites such as the one from the OH group. The IMHB treatment is applied to treat both lovastatin and simvastatin by deactivating the H site on the OH functional group and one e₁ on the COO group in the lactone ring as illustrated in Fig. 6.6 as these two sites are presumed to participate in the IMHB formation. This IMHB within the lactone ring of statins causes a boat-like conformation as depicted in Fig. 6.6 (right). An experimental investigation on this hypothesis is currently underway at Pfizer and the MSE group at Imperial. As a final note it is important to mention that with this effective treatment, the IMHB can also be attributed from the association between the OH group and another COO group outside the lactone ring since the group connectivity is not explicitly specified in this GC approach.
Figure 6.6: The SAFT-γ Mie group-contribution model for lovastatin highlighting the IMHB treatment and the plausible conformational equilibrium of the lactone ring showing the IMHB (right)

As the solubility (solid-liquid equilibrium) prediction relies heavily on the experimental data of the pure component of solutes, via Eq. (5.2), the input data must be chosen very carefully. The required data: the enthalpy of fusion $\Delta H_f$, melting point temperature $T_m$, and the difference between the heat capacity of the solid and the liquid at the melting point $\Delta c_p(T_m)$ available in the literature are summarised in Table 6.7. Given a significant difference between the $T_m$ of both statins and the saturation temperature range, the assumption of negligible $\Delta c_p(T_m)$ is no longer valid as demonstrated by a number of studies [412–414]. As can be seen in the table, the experimental data vary quite significantly between different studies especially for the $\Delta c_p(T_m)$ and $\Delta H_f$. Note that the lovastatin pure-component data from Refs. [412] (Int. J. Pharm.) and [415] (PhD thesis) and the simvastatin pure-component data from Refs. [413] (Fluid Phase Equilib.) and [415] (PhD thesis) are from the same group of authors, but the reported $\Delta c_p(T_m)$ and $\Delta H_f$ values are different (while the $T_m$ and the solubility data are reported the same). In the case of simvastatin, the reported $\Delta c_p(T_m)$ values are markedly different between different studies and no evidence of polymorphism, which can lead to different physical properties, is found between 293 K and the fusion temperature [411]. The pure-component data for both statins from Ref. [415] are chosen since this reference provides extensive solubility data (the same data reported in Refs. [412, 413]) which can be used to compare with our predictions. The pure-component data also agree well with the in-depth study from Ref. [411] in the case of simvastatin as shown in Table 6.7.

The SAFT-γ Mie calculations for the solubility of lovastatin and simvastatin in different alcohols are performed with and without the IMHB effective treatment. As can be seen in Fig.
Table 6.7: Literature data on lovastatin and simvastatin pure component and solubility in alcohols and alkyl acetates

<table>
<thead>
<tr>
<th>Reference</th>
<th>Solute</th>
<th>$\Delta H_{1}$/ J mol$^{-1}$</th>
<th>$T_{m}$/ K</th>
<th>$\Delta c_p(T_{m})$/ J mol$^{-1}$K$^{-1}$</th>
<th>Solvent</th>
<th>$T$ range / K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun et al. [416]</td>
<td>lovastatin</td>
<td>n/a</td>
<td>452.15 ± 0.5</td>
<td>n/a</td>
<td>ethanol/acetates</td>
<td>278.3 - 323.7</td>
</tr>
<tr>
<td>Nui-Gyabaah et al. [412]</td>
<td>lovastatin</td>
<td>431.36</td>
<td>445.5 ± 0.5</td>
<td>255</td>
<td>alcohols</td>
<td>285.7 - 312.6</td>
</tr>
<tr>
<td>Nui-Gyabaah et al. [417]</td>
<td>lovastatin</td>
<td>431.36</td>
<td>445.5 ± 0.5</td>
<td>255</td>
<td>acetates</td>
<td>285.1 - 312.2</td>
</tr>
<tr>
<td>Nui-Gyabaah et al. [415]</td>
<td>simvastatin</td>
<td>32170 ± 500</td>
<td>412.6 ± 0.5</td>
<td>1149 ± 5</td>
<td>alcohols/acetates</td>
<td>279.1 - 315.5</td>
</tr>
<tr>
<td>Nui-Gyabaah et al. [413]</td>
<td>simvastatin</td>
<td>32169 ± 500</td>
<td>412.65 ± 0.5</td>
<td>230 ± 5</td>
<td>alcohols/acetates</td>
<td>279.1 - 313.5</td>
</tr>
<tr>
<td>Aceves-Hernández et al. [414]</td>
<td>simvastatin</td>
<td>24464</td>
<td>410.92 ± 0.5</td>
<td>268.61</td>
<td>alcohols</td>
<td>285.7 - 312.0</td>
</tr>
<tr>
<td>Yan et al. [418]</td>
<td>simvastatin</td>
<td>30130</td>
<td>415.1</td>
<td>n/a</td>
<td>acetates</td>
<td>278.2 - 318.2</td>
</tr>
<tr>
<td>Simões et al. [411]</td>
<td>simvastatin</td>
<td>30400 ± 200</td>
<td>412.2 ± 0.2</td>
<td>142 ± 13</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

6.7a), it is evident that the SAFT-\(\gamma\) Mie prediction with the IMHB treatment provides an exceptional agreement with two independent sets of lovastatin solubility data in ethanol [412, 416]. The solubility predictions with the consideration of IMHB for lovastatin in different alcohols are shown in Fig. 6.7b) to demonstrate the overall trend. The comparison to the experimental data [412] by means of the log$_{10}$ squared error (log$_{10}$ SE = $\frac{1}{N_p}$ $\sum_{n=1}^{N_p}$ $[\log_{10}(x_{n}^{\text{exp}}) - \log_{10}(x_{n}^{\text{calc}})]^2$) is reported in Table 6.8. While the experimental solubility data for lovastatin in ethanol are consistent between the two studies [412, 416], the solubility data for simvastatin in ethanol (and other alcohols) reported by Refs. [413] and [414] are quite different for $T > 300$ K as illustrated in Fig. 6.7c). Nevertheless, the SAFT-\(\gamma\) Mie prediction for simvastatin + ethanol mixture (with IMHB treatment) lies within the two experimental data sets [413, 414] with a good agreement with Ref. [413] at low temperature and Ref. [414] at $T > 300$ K. The log$_{10}$ SE values are calculated based on the data from Ref. [413] for consistency. From the log$_{10}$ SE reported in Table 6.8, in most cases, the solubility predictions for both statins with the IMHB approach provide better agreement to the experimental data compared to the fully associated model (without IMHB treatment). The level of agreement is remarkable and the log$_{10}$ SE values are found to be lower than 0.09 for both statins in every alcohol studied (ethanol to 1-octanol). The marked improvement for the IMHB treatment is seen in the solubility prediction in ethanol which results in 2-3 times lower in the statins solubilities. A possible rationalisation may be because the contribution of the associating OH group in ethanol is very large, thus the probability of this group in finding other associating groups (being the OH and COO groups in the statin molecule in this case) is proportionally large, therefore the significant effect of the deactivation of association sites in the IMHB treatment is observed. There are also some cases where the two modes of calculation present a similar log$_{10}$ SE or the model without the IMHB yields slightly
lower log$_{10}$ SE values. The reason for this may lie in the current model only being able to take into account one form of solute, either with or without IMHB, while in the real systems, the two forms coexist in equilibrium.

![Graphical representations](image)

**Figure 6.7:** Solubilities of lovastatin in (a) ethanol and (b) 1-alcohols and (c) simvastatin in ethanol at pressure $p = 0.101$ MPa as a function of temperature. The circles [416], squares [412], triangles [413], diamonds [414] represent experimental data, the continuous curves the SAFT-$\gamma$ Mie predictions with effective IMHB treatment, and the dashed curves the predictions without the treatment.

The solubilities of lovastatin and simvastatin in a number of linear and branched alkyl acetates are also investigated. A graphical comparison between the SAFT-$\gamma$ Mie predictions and the experimental data for these statins in ethyl acetate is provided in Fig. 6.8. Similar to the solubility in alcohols, the experimental data for lovastatin are in good agreement between the two independent studies [416, 417] while the data for simvastatin differ especially at higher temperature [413, 418] as seen in Figs. 6.8a) and 6.8b), respectively. Despite the better agreement between the SAFT-$\gamma$ Mie prediction and the data from Ref. [418], where the trend with increasing temperature is captured, the log$_{10}$ SE values in Table 6.8 are calculated based on the data
from Ref. [413] for consistency. The average log_{10} SE values for the two statins in seven different acetates, both linear and branched compounds, are quite low either with or without the IMHB approach. These results substantiate the transferability and accuracy of the GC model. The largest error observed for the two statins is associated with the solubilities in tert-butyl acetate which is modelled by the C functional group. This group provides the most general method for the modelling of compounds with quaternary carbon, i.e., the most transferable model, but it may not be the most accurate model [169]. The accuracy of the C group parameters in the SAFT-\gamma Mie approach is only moderate given the relatively high %AAD for the vapour pressure (4.08 - 15.88 %) and the saturated liquid density (5.23 - 7.90 %) for branched alkanes comprising the quaternary carbon [169]. Although the C group accounts for a very small contribution of the whole molecule in terms of its shape factor ($S_k = 0.04072$), its effect becomes significant for the tert-butyl acetate + statins systems given a very large fraction of the tert-butyl acetate in the mixture ($x_{acetate} > 0.96$). The C group is also used to represent the quaternary carbon in simvastatin (Fig. 6.7c); its solubility predictions in acetates and alcohols are, still, in excellent agreement with the data because the mole fraction of the simvastatin is much smaller than the solvent, hence the overall contribution of the C group remains small.

The solubility predictions for lovastatin and simvastatin with and without the IMHB approach are similar as seen in Fig. 6.8. The impact of the IMHB approach on the solubility in acetates is less than what is observed in alcohols (cf. Fig. 6.7). This can be explained because the COO functional group in the acetate solvent only associates with the OH group of the statins, while the OH group in the alkanol solvent associates with both the OII and the COO groups, consequently switching off the association site on these groups (i.e., the IMHB treatment) has a greater impact in the case of the alkanol solvents. In other words, this means that the statins do not associate to a great extent with acetate solvents, but they do with alkanol solvents.

Regarding the solvent ranking for all acetates studied, Ref. [413] reports the simvastatin solubilities in decreasing order: $n$-butyl acetate $>$ ethyl acetate $>$ $n$-propyl acetate $>$ sec-butyl acetate $>$ iso-butyl acetate $>$ iso-propyl acetate $>$ tert-butyl acetate while Ref. [418] reports ethyl acetate $>$ $n$-propyl acetate $>$ $n$-butyl acetate $>$ iso-butyl acetate. The SAFT-\gamma Mie prediction follows the ranking from Ref. [418]. The decrease in solubility observed in going from linear to branched isomers is, however, reported the same by both studies [413, 418] and this ranking is predicted correctly by our model for the simvastatin solubility in both $n$-propyl/isopropyl and $n$-butyl/isobutyl/tert-butyl acetate isomers. The same pattern of agreement is also observed for
the solubility of lovastatin in linear and branched acetates [417]. Note that the solubility predictions in isobutyl acetate and sec-butyl acetate are identical since these solvents are represented by the same GC model.

**Figure 6.8:** Solubilities of (a) lovastatin and (b) simvastatin in ethyl acetate at pressure $p = 0.101$ MPa as a function of temperature. The squares [417], circles [416], triangles [413], diamonds [418] represent experimental data, the continuous curves the SAFT-$\gamma$ Mie prediction with effective IMHB treatment, and the dashed curves the prediction without the treatment.
Table 6.8: \( \log_{10} \) squared error between the reported experimental values \([412, 413, 417]\), \( N_p \) being the total number of experimental points, and the SAFT-\( \gamma \) Mic predictions for the solubility of lovastatin and simvastatin in alcohols and alkyl acetates, with and without the effective treatment of the IMHB

<table>
<thead>
<tr>
<th>Solute</th>
<th>Solvent</th>
<th>( T / \text{K} )</th>
<th>( N_p )</th>
<th>( \log_{10} ) SE without IMHB treatment</th>
<th>( \log_{10} ) SE with IMHB treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>lovastatin</td>
<td>ethanol</td>
<td>286.2 - 310.6</td>
<td>9</td>
<td>0.1034</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>1-propanol</td>
<td>286.2 - 310.6</td>
<td>9</td>
<td>0.0027</td>
<td>0.0324</td>
</tr>
<tr>
<td></td>
<td>1-butanol</td>
<td>285.7 - 310.1</td>
<td>10</td>
<td>0.0159</td>
<td>0.0550</td>
</tr>
<tr>
<td></td>
<td>1-pentanol</td>
<td>289.2 - 310.6</td>
<td>9</td>
<td>0.0031</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>1-hexanol</td>
<td>293.9 - 312.0</td>
<td>7</td>
<td>0.0167</td>
<td>0.0144</td>
</tr>
<tr>
<td></td>
<td>1-octanol</td>
<td>285.7 - 310.1</td>
<td>10</td>
<td>0.0496</td>
<td>0.0163</td>
</tr>
<tr>
<td></td>
<td>(average)</td>
<td></td>
<td></td>
<td>0.0319</td>
<td>0.0214</td>
</tr>
<tr>
<td></td>
<td>ethyl acetate</td>
<td>285.1 - 312.2</td>
<td>10</td>
<td>0.1129</td>
<td>0.0743</td>
</tr>
<tr>
<td></td>
<td>( n )-propyl acetate</td>
<td>285.1 - 312.2</td>
<td>10</td>
<td>0.0600</td>
<td>0.0608</td>
</tr>
<tr>
<td></td>
<td>isopropyl acetate</td>
<td>285.1 - 312.2</td>
<td>10</td>
<td>0.0120</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>( n )-butyl acetate</td>
<td>285.1 - 312.2</td>
<td>10</td>
<td>0.0073</td>
<td>0.0198</td>
</tr>
<tr>
<td></td>
<td>isobutyl acetate</td>
<td>285.1 - 312.2</td>
<td>10</td>
<td>0.0492</td>
<td>0.0255</td>
</tr>
<tr>
<td></td>
<td>sec-butyl acetate</td>
<td>285.1 - 312.2</td>
<td>10</td>
<td>0.0615</td>
<td>0.0345</td>
</tr>
<tr>
<td></td>
<td>tert-butyl acetate</td>
<td>285.1 - 312.2</td>
<td>10</td>
<td>0.4444</td>
<td>0.3187</td>
</tr>
<tr>
<td></td>
<td>(average)</td>
<td></td>
<td></td>
<td>0.1068</td>
<td>0.0777</td>
</tr>
<tr>
<td>simvastatin</td>
<td>ethanol</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.2836</td>
<td>0.0482</td>
</tr>
<tr>
<td></td>
<td>1-propanol</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0798</td>
<td>0.0197</td>
</tr>
<tr>
<td></td>
<td>1-butanol</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0248</td>
<td>0.0226</td>
</tr>
<tr>
<td></td>
<td>1-pentanol</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0159</td>
<td>0.0376</td>
</tr>
<tr>
<td></td>
<td>1-hexanol</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0673</td>
<td>0.0812</td>
</tr>
<tr>
<td></td>
<td>1-octanol</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.1279</td>
<td>0.0876</td>
</tr>
<tr>
<td></td>
<td>(average)</td>
<td></td>
<td></td>
<td>0.0999</td>
<td>0.0495</td>
</tr>
<tr>
<td></td>
<td>ethyl acetate</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.2314</td>
<td>0.1706</td>
</tr>
<tr>
<td></td>
<td>( n )-propyl acetate</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.1607</td>
<td>0.1407</td>
</tr>
<tr>
<td></td>
<td>isopropyl acetate</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0208</td>
<td>0.0143</td>
</tr>
<tr>
<td></td>
<td>( n )-butyl acetate</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0402</td>
<td>0.0482</td>
</tr>
<tr>
<td></td>
<td>isobutyl acetate</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0064</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>sec-butyl acetate</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.0098</td>
<td>0.0083</td>
</tr>
<tr>
<td></td>
<td>tert-butyl acetate</td>
<td>279.1 - 315.5</td>
<td>10</td>
<td>0.2943</td>
<td>0.2211</td>
</tr>
<tr>
<td></td>
<td>(average)</td>
<td></td>
<td></td>
<td>0.1091</td>
<td>0.0870</td>
</tr>
</tbody>
</table>
6.5 Conclusions

Intramolecular hydrogen bonds (IMHB) can greatly influence molecular properties and pharmacokinetic properties of the drug. They can be observed by various evidence such as thermodynamic data, spectroscopic data, or ab initio calculations. Modelling multifunctional molecules with the consideration of IMHB is challenging, especially in the context of group-contribution approaches which do not consider molecular conformation and group connectivity. In this chapter, an effective IMHB-treatment method has been developed within the SAFT-\(\gamma\) Mie framework for the modelling of compounds that exhibit IMHB. In the formation of IMHB, the association sites (the hydrogen bond donor and acceptor) are used intramolecularly, therefore reducing the availability of the sites for intermolecular association. We “mimic” this phenomenon by precluding the association sites in a given solute (API) that would have participated in the IMHB, preventing them from interacting with other associating groups (mainly the associating solvent). This method has been proven to be a simple but powerful technique to predict the partition coefficients of organic compounds including several APIs, for which the values are governed by the competition between inter- and intramolecular association. The treatment should be applied to molecules with IMHB forming five- to eight-membered ring which conforms with experimental observations. This effective treatment results in an increase in the partition coefficients, i.e., increase in the hydrophobicity of molecules, which is also observed experimentally for molecules with IMHB. The same approach is successfully applied on solubility predictions of complex APIs including ibuprofen, ketoprofen, simvastatin and lovastatin in different solvents. The level of agreement with the corresponding experimental data is exceptional given the simple GC model and the effective IMHB-treatment approach used for the modelling. The impact of the treatment is most significant when it is applied to solubility prediction of highly associated solutes, whose functional groups participate in IMHB, in polar solvents, e.g., water and ethanol. These results highlight the ability of the theory to treat hydrogen bonding systems from a molecular standpoint and the versatility of the method for the description of complex bonding aggregates without the need for additional parameters regression. We emphasise that no QM calculations or higher-order functional groups are used to improve the thermodynamic properties prediction of SAFT-\(\gamma\) Mie. The work presented in this chapter advocates the significant impact of the IMHB on the thermodynamic properties as well as the need to incorporate such interactions in the model used for drug design, and also demonstrates the fidelity of the SAFT-\(\gamma\) Mie approach in pharmaceutical applications.
Chapter 7

Conclusions and future work

A reliable predictive, rather than correlative, method to obtain physicochemical properties of APIs is a powerful tool to reduce drug’s attrition at early stages and improve R&D productivity for pharmaceutical industries. Of many predictive approaches, group-contribution (GC) methods are excellent candidates given their ability to provide an accurate description of the thermodynamic properties of fluids and fluid mixtures and their extensive application in the general area of fluid formulation and computer-aided molecular design (CAMD). Opportunities for GC methods for the prediction of octanol-water partition coefficients ($K_{OW}$), one of the most widely used physicochemical properties to indicate drug’s lipophilicity, over other predictive and correlative methods were discussed in Chapter 2. UNIFAC [17], despite being the most widely employed GC method in industrial applications due to its extensive parameter table, has exhausted its applicability to account for highly asymmetric systems, especially aqueous solutions with complex multifunctional compounds [20]. In this thesis, I have demonstrated that the challenging task of modelling multifunctional molecules (in both aqueous and organic environments) can be accomplished using an advanced GC method, the SAFT-$\gamma$ Mie equation of state [22]. In SAFT-$\gamma$ Mie, functional groups are modelled as fused spherical segments that interact via Mie (generalized Lennard-Jones) potentials of variable repulsive and attractive range. The method also takes into account association energy explicitly which makes it specially compatible with associating species and asymmetric systems such as hydrocarbon + water mixtures. The modelling of alkanes and alcohols in aqueous environment using the SAFT-$\gamma$ Mie approach was demonstrated in Chapter 3, where the GC model developed was shown to lead to accurate predictions of phase equilibria, including vapour-liquid and liquid-liquid equilibria, free energies of solvation and other infinite dilution properties. The estimation of functional group parameters presented in this work was done with a consideration of a fine balance between transferability
and accuracy. Different types of phase equilibria as well as single-phase properties of a number of compounds characterised by the GC model were shown in Chapter 4. The applicability of the model stretches much beyond the thermodynamic conditions of interest for the pharmaceutical industry; nevertheless, it is useful to use pharmaceutically relevant properties as a stringent test for the model’s fidelity as these properties are very sensitive to the model parameters. Accurate predictions for $K_{i,OW}$ and solubilities of several complex organic solutes including APIs were demonstrated in Chapters 5 and 6. The GC model within the SAFT-\(\gamma\) Mie platform was shown to overcome a general difficulty in modelling complex multifunctional compounds in highly non-ideal systems. In addition to the great transferability of the model parameters, the success in modelling APIs is attributed to the unlike induced association-site models [240], which lead to a simple but effective treatment for intramolecular hydrogen bonds (IMHB). The IMHB occur prevalently in organic compounds and can significantly alter their molecular conformations and thermodynamic properties [364]. Despite playing a critical role in medicinal chemistry, IMHB are hardly recognised by many predictive approaches especially within the GC framework. Even though the 3D structure of the compound is not fully considered in our predictive method, the IMHB treatment approach proposed in this thesis is found to be very effective in accounting for changes in thermodynamic properties exhibited from the IMHB. The works presented thus far demonstrate the applicability of the SAFT-\(\gamma\) Mie approach as a potential tool to improve R&D productivity for pharmaceutical industries. This molecular-based approach could aid drug design beyond the Lipinski’s ‘rule-of-five’ chemical space [2], generating truly innovative and cost-effective new medicines, and could also be incorporated into process models for the design of optimised production precesses.

\section{7.1 Summary of the key contribution of this thesis}

The contributions of the work presented in this thesis can be summarised as follows:

- Highlighted a significant impact of including mixture data in the model development. The predictability of two models for alcohols— one developed also with mixture data— on the phase behaviour of different alcohol mixtures was assessed. The inclusion of mixture data to develop model parameters can be used to circumvent the proximity effect, which is a common issue for GC approaches, especially in the description of fluid-phase behaviour of mixtures involving small polar compounds.
Extended the SAFT-$\gamma$ Mie method to the study of aqueous solution of hydrocarbons. In this case, the ability of the theory to successfully describe different types of phase behaviour of highly non-ideal mixtures was demonstrated. It was shown that the SAFT-$\gamma$ Mie method provides an accurate simultaneous description of the vapour-liquid and liquid-liquid equilibria that mixtures exhibit. The model was accurate and highly transferable even for the prediction of properties that are known to be challenging, such as mutual solubilities of hydrocarbon and water which span over ten orders of magnitude, and solvation properties of hydrocarbons in an aqueous environment.

Developed group parameters for use within the SAFT-$\gamma$ Mie approach. Several group parameters, especially unlike-interaction parameters, that characterise pharmaceutical compounds have been established. The functional group parameters were developed mostly with phase equilibria data; no $K_{i,OW}$ or SLE data was used. The accuracy and transferability of the parameters obtained were examined extensively on different mixtures, which have not been included in the parameter estimation, for their fluid-phase equilibrium data (VLE, LLE, and VLLE) and single-phase thermodynamic properties, such as excess properties of mixing and activity coefficients at infinite dilution.

Demonstrated the predictive capability of SAFT-$\gamma$ Mie in the accurate description of solubility and partition coefficient of several families of organic compounds in different solvent mixtures, i.e., hexane-water, hexadecane-water and octanol-water using a consistent GC model.

Proposed a novel effective method to treat intramolecular hydrogen bonds. The treatment was proven to account for the change in thermodynamic properties that might be exhibited by the formation of IMHB in some organic compounds. We also validated the application of the SAFT-$\gamma$ Mie approach further for the prediction of $K_{i,OW}$ and solubilities of several APIs, including valproic acid, ibuprofen, ketoprofen, lovastatin, and simvastatin, some of which contain IMHB.

7.2 Directions for future work

Expanding SAFT-$\gamma$ Mie group parameters

In common with any group-contribution approach, the predictive capability of the SAFT-$\gamma$ Mie approach depends primarily on the extent of the group parameters available. Thus, in order to
enhance its predictive power, an extensive group parameter development should be performed based on the parameters presented in this work. Functional groups such as amines and different aromatic substituents, *e.g.*, aC, aCOH, aCCOOH, aCNHCO, aN, and aCF, should be given a priority for the next phase of development given their general appearance in pharmaceutical molecules. Note that the development of group parameters is done in a successive manner; therefore, it has to be carried out very carefully and examined thoroughly to avoid the “jenga effect”.

**Solubility prediction**

The difference between the heat capacities of the solid and the liquid form of the solute, $\Delta c_p(T_m)$, is often needed in order to calculate solubility; however, this value is very challenging to obtain experimentally. Experimentalists have also tried to evaluate the $\Delta c_p(T_m)$ from solubility data, but the value varies widely by 50%, and in some cases up to 300% [412]. Without a reliable $\Delta c_p(T_m)$ data, it is extremely difficult to obtain activity coefficients from the solubility data for thermodynamic analysis. Given the ability to predict activity coefficients accurately, the SAFT-$\gamma$ Mie approach can be used to estimate $\Delta c_p(T_m)$ from regression to the solubility data. It is worthy of note that $\Delta c_p(T_m)$ is a property of a pure crystalline, and therefore, the value should remain constant regardless of the dissolving solvent.

**Improving property prediction**

More than 60% of pharmaceutical compounds listed in the *World Drug Index* contain ionisable groups [419]. In order to model the lipophilicity profile of a drug, expressed in terms of a distribution coefficient log$D_{i,OW}$ value as a function of pH, the theory must take ionisation into account. The future predictive method should also be able to consider the IMHB or structural changes of the molecule intrinsically, based on the free energy of the system. Although incorporating these two extra features together with the GC method could be particularly challenging, its predictive capability would make a significant impact on many industrial applications.

**Extended applications**

One of the ultimate aims of developing the GC approach, in addition to its use in property prediction as demonstrated in this work, is its extensive application in computer-aided molecular design or the integrated design of solvents and processes, where the molecular characteristics of solvents are determined as part of the optimisation process. For pharmaceutical applications,
several constrains including drug-like properties (e.g., \( K_{i,OW} \)) and solvent-selection criteria could be used to formulate an optimisation problem to determine novel drug candidates systematically.
Bibliography


[248] \textit{gPROMS} v. 3.7.1; PSE Ltd.: London, 2013; \url{http://www.pselectorprise.com}.


[251] \textit{gPROMS} v. 3.7.1; PSE Ltd.: London, 2013; \url{http://www.pselectorprise.com}.


Bibliography

[276] R. G. Sultanov, V. G. Skripka, and A. Y. Namiot. Phase equilibria in the system water-

[277] D. Ambrose and C. Tsonopoulos. Vapor-liquid critical properties of elements and com-

[278] M. Góral, B. Wiśniewska-Gocłowska, and A. Maczyński. Recommended liquid-liquid equi-
2006.

[279] S. P. Tunik, T. M. Lesteva, and V. I. Chernaya. Phase equilibria in the water-alcohols-
formaldehyde systems. I. The liquid-vapor equilibria in the systems water-C₆ and C₇ al-


High-precision determination of Henry’s constants for methane and ethane in liquid water

[284] R. Crovetto, R. Fernandez-Prini, and M. L. Japas. The solubility of ethane in water up

constants and air/water partition coefficients of n-octane and two halogenated octanes.

[286] V. P. Andrey and L. S. Everett. Thermodynamic functions of hydration of hydrocarbons

coefficients at infinite dilution using gas-liquid chromatography. 6. Results for systems


