Imperial College London
Department of Chemical Engineering

Predictive modelling of the solid—liquid solubility and chemical equilibrium of amino acids and oligopeptides: Exploring the influence of charges within a group-contribution framework

Ahmed Alyazidi

A thesis submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy and the Diploma of
Imperial College London, April 2025

Statement of Originality

I, Ahmed Alyazidi, declare that this thesis titled, "Predictive modelling of the solid–liquid solubility and chemical equilibrium of amino acids and oligopeptides: Exploring the influence of charges with a group-contribution framework", and the research on which it is based, are the product of my own work. I confirm that, where I have used the work of others, published or otherwise, it is fully acknowledged in accordance with the standard referencing practices of the discipline.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

The accurate prediction of thermodynamic properties of amino-acid and peptide solutions is critical in the pharmaceutical industry, where amino acids and oligopeptides are gaining increasing prominence as active pharmaceutical ingredients (APIs). A major challenge in this regard is the development of predictive models with broad applicability, parameter transferability, and minimal reliance on system-specific experimental data. In this thesis, this challenge is addressed by developing group-contribution (GC) thermodynamic models to predict the solubility of amino acids and oligopeptides across a wide range of solvent systems and thermodynamic conditions, with a particular focus on the complex effects of pH-dependent speciation and charge interactions.

The SAFT-γ Mie GC equation of state (EoS) is employed to describe solid—liquid equilibria in systems containing amino acids and oligopeptides. Parameter optimisation is carried out using experimental data from chemically related systems, demonstrating the high degree of transferability and reducing the dependency on direct solubility measurements. Chemical-equilibrium equations are coupled with phase-equilibrium equations to capture speciation effects accurately. Additionally, an exploratory molecular simulation study benchmarks the treatment of electrostatic interactions in ionic systems using the expanded-ensemble method, providing critical insights into modelling ionic chains within primitive-model expressions.

The resulting model delivers accurate solubility predictions for a wide variety of systems, underscoring its robustness and potential for application in drug development. However, the study also highlights the limitations imposed by the scarcity and inconsistency of experimental data, especially for melting properties and solid–liquid solubility. The molecular simulation study provides important insights into the treatment of electrostatic interactions in ionic-chain molecules, providing potential directions for future improvements to the primitive-model expressions used within the model.

Acknowledgements

In the long run, names and faces may be forgotten, but I hope that these acknowledgments serve as a reminder that it truly takes a village.

I would like to express my sincere gratitude to my thesis supervisors, Amparo Galindo and George Jackson, for their continued support throughout the entire duration of my PhD degree. Their valuable insights and words of encouragement helped push this project to the end. I want to thank Andrew Haslam for his consistent support and writing lessons that I will carry with me for a long time.

I thank my colleagues at the MSE group and the department for their friendship and support. Special thanks to Thomas, Shubhani, Malak, and Felipe for all the hours they contributed towards this thesis, from advice during meetings to conversations over coffee and beer. I would also like to thank Miriam for her warmth and friendship. I also extend my thanks to the other colleagues and friends at the office, in no particular order: Tanuj, Fabian, Gustavo, Stefanos, Lilly, Saher, Griffin, Lauren, Michele, Tom, Ilya, and Mathias. My thanks also go to the Sargent Centre members, with special thanks to Claire, Jana, and Ieke.

I'd also like to thank my friends and chosen family outside Imperial, whose love and support kept me alive and going. Big love and thanks to Arman for the lifelong friendship and for spending countless hours with me on Zoom as we co-worked, gamed, and chatted during the lockdown. Big love to Nabilah and Maram for their radiant loving warmth. Big love to Emil, Eray, Bayryam, Marah, Heidi, Shiza, Younis, Iman, Antonio, Byron, Nia, David, and Angelo for their grounding love, friendship, and support here in London, and to Erin, Razan, Moustafa, Lara, Mirna, Fatima, and Danish for their loving friendship and all the memories in Qatar. I send the most special thanks to my best friend in heaven, Moho, for his companionship, joy, and laughter. I will forever miss you, Moho!

Finally, I thank my family for their love, trust, and support. My dad, for all the long conversations about everything. My mom, for the endless love and tenderness. My siblings, for looking after me and lifting me up all these years. Big thanks to Elizabeth and Joe, for their love and generous support. Big love to Burtuqal, for being my shining light in the last 14 months.

Dedication

I dedicate this to my mom and dad, as I would be nothing without them. I also dedicate this to Moho. Moho, this is for both of us!

Contents

Abstract **v**

Acknowledgements **vii**

1 Introduction **1**

1.1 Motivation and objectives 1

1.2 Thesis overview 3

2 Thermodynamic models for amino acids and oligopeptides **5**

2.1 Introduction 5

2.2 Early theoretical developments 6

2.3 Local composition models 7

2.4 Molecular models 9

2.5 Modeling phase equilibria 13

2.5.1 Solid–liquid equilibrium 14

2.6 Conclusions 18

3 SAFT-γ Mie for amino-acid and oligopeptide solutions 20

3.1 Introduction 20

3.2 SAFT-γ Mie 22

3.2.1 The Helmholtz free energy for neutral systems 24

3.2.2 Electrostatic contributions to the Helmholtz free energy 26

3.3 Model development 33

3.4 SAFT-γ Mie models for amino acids and oligopeptides 35

3.4.1 NH_2−COOH 43

3.4.2 NH_2−CHOH 44

3.4.3 COOH−CHOH 45

3.4.4 NH_2−CH 47

3.4.5 The peptide (CONH) group 49

3.5 Conclusions 64

4 Prediction of solid–liquid solubility of amino acids and oligopeptides 66

4.1 Introduction 66

4.2 Sensitivity of solubility calculations to melting-property data 68

4.3 Solid–liquid solubility at the isoelecric point pI 69

4.3.1 Models 70

4.3.2 Solubility in pure solvents 71

4.3.3 Solubility in mixed-solvent media 77

4.4 Solid–liquid solubility under changing pH 85

4.4.1 SAFT-γ Mie models 85

4.4.2 Equations of chemical equilibria of speciation 87

4.4.3 Solubility prediction 91

4.5 Conclusions 94

5 Molecular simulations of charged chain molecules 97

5.1 Introduction 97

5.2 The expanded-ensemble method 100

5.3 Transition matrix Monte Carlo 103

5.4 Molecular models 107

5.5 Simulation details 109

5.6 Results and discussion 110

5.6.1 Effect of chain length 112

5.6.2 Effect of monomer size 113

5.6.3 Effect of charge position 117

5.7 Conclusions 123

6 Conclusions 126

6.1 Summary of thesis achievements 127

6.2 Future research 128

6.3 Publications 130

A Details of the electrostatic contributions to the the Helmholtz free-energy expression **131**

A.1 Electrostatic contributions to the SAFT-γ Mie Helmholtz free energy: Model M 131

A.2 Expressions for the chemical potential and pressure due to the ion contribution of Model G 133

A.3 Derivatives of the Helmholtz free energy with respect to screening length Γ . . . 135

Bibliography **138**

List of Tables

3.1 SAFT-γ Mie like interaction parameters of the groups considered in our current work (excluding association). The attractive range of the Mie potential is $\lambda_{kk}^{\mathrm{a}} = 6$ for all groups here. The asterisk indicates that the CONH group has been characterized in the current work. 39

3.2 Unlike group parameters (excluding association) for use with the SAFT-γ Mie approach. An asterisk indicates that the parameters have been characterised in the current work. CR indicates that combining rules are used [51]. 40

3.3 Group association parameters for use with the SAFT-γ Mie approach. An asterisk indicates that the parameters have been characterized in the current work. 41

3.4 Overview of the accuracy of SAFT-γ Mie, represented in %AAD and AAD, in the calculation of vapour pressure (P^{sat}), liquid pure-component and mixture densities (ρ), bubble temperature (T_{bub}), dew temperature (T_{dew}), bubble pressure (P_{bub}), dew pressure (P_{dew}), LLE temperature (T_{LLE}), and excess enthalpy (H^{E}). N^{D} is the number of experimental data points used to evaluate the accuracy of the model. 60

4.1 Melting properties of the amino acids and peptides used to calculate their solubility, reported with their experimental uncertainties; melting-point temperature (T_i^{fus}), enthalpy of fusion ($\Delta h_i^{\mathrm{fus}}$), and difference between the molar heat capacity of liquid and solid phases evaluated at T_i^{fus} ($\Delta c_{\mathrm{p,i}}^{\mathrm{SL}}$). 70

4.2 Overview of the accuracy of SAFT-γ Mie in the calculation of solubility for amino acids in water or alcohol, where N^{D} is the number of experimental data points used to calculate the average absolute deviation (AAD). 83

4.3 Overview of the accuracy of SAFT-γ Mie in the calculation of solubility for peptides in water, where N^{D} is the number of experimental data points used to calculate the average absolute deviation (AAD). 84

4.4 Acid–base equilibrium properties of glycine and alanine. 90

5.1 Force-field parameters of the LJ potential based on the model of Andreev et al. [245]. Parameters are presented in dimensionless units, as described in Equations 5.20 and 5.21. The same set of parameters is used for the MSA calculations, in both Model G and Model M. 108

5.2 EE simulation data of the ionic chemical potential ($\Delta\mu_{\mathrm{s}}^{\mathrm{Ion}}$) of the monomer, dimer, trimer, tetramer, and pentamer anion salts as a function of salt mole fraction (x_{s}). 115

5.3 EE simulation data of the ionic chemical potential ($\Delta\mu_{\mathrm{s}}^{\mathrm{Ion}}$) of systems containing dimer anions, which have different segment-to-ion diameter ratios ($\sigma_{\mathrm{seg-seg}}/\sigma_{\mathrm{ion-ion}}$), as a function of salt mole fraction (x_{s}). 117

5.4 EE simulation data of the ionic chemical potential ($\Delta\mu_{\mathrm{s}}^{\mathrm{Ion}}$) of systems containing trimer, tetramer, and pentamer anions, in which the charge is at the centre, as a function of salt mole fraction (x_{s}). 120

List of Figures

2.1 Thermodynamic cycle representing the transition from the solid state at T and P to the subcooled liquid state at T and P, adapted from Prausnitz et al. [10]. State **a** represents the solute in the solid phase at the operating temperature T and pressure P, state **b** represents the solute in the solid phase at the triple point (T_t and P_t), state **c** represents the solute in the liquid phase at the triple point (T_t and P_t), and state **d** represents the subcooled liquid at T and P. . . . 16

3.1 A cartoon representation of the models considered within the SAFT-γ Mie approach in the current work. The upper panel depicts an electrolyte comprising a chain-like cation (C_1) and a spherical anion (C_2) in water (C_3). The cation is modelled as a heteronuclear chain of groups including a charged group comprising two identical segments. The lower panels depict the effective models proposed for the evaluation of the ion and Born contributions of the Helmholtz free energy. Model M refers to “molecular" and Model G to “group". 28

3.2 SAFT-γ Mie representation of (a) glycine, (b) alanine, (c) serine, and (d) valine. A heteronuclear model with fused spherical segments is implemented in which short-range association sites are represented with smaller light-blue (sites of type H), red (type e_1), and brown (type e_2) circles. 36

3.3 SAFT-γ Mie representation the glycine cation, zwitterion and anion as modelled in the current work.A heteronuclear model with fused spherical segments is implemented in which short-range association sites are represented with smaller light-blue (sites of type H) and red (type e_1) circles. The reactions also involve water, the hydronium and hydroxide ions, as well as the acidic and basic counter ions, which are not shown here. 37

3.4 SAFT-γ Mie GC group interaction sub-matrix containing groups required to model systems in the current work. Blue cells indicate interactions previously published ([55, 57]), green cells indicate interactions developed in our current work, grey cells indicate interactions that are obtained using combining rules [51], and white cells indicate interactions that are not needed in the current work. . . 38

3.5 SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-COOH unlike interaction compared to experimental data. (a) Bubble pressure of water (1) + glycine (2) at 298 K [113]. The pressure is measured with a reported sensitivity of about 0.3 Pa. (b) Solid–liquid solubility of water (1) + glycine (2) at 1 bar; filled circles [111], squares [115], diamonds [112], and black triangles [30]. Empty symbols denote experimental data not used in the parameter optimisation: empty circles [116]; and green triangle [117]. 44

3.6 SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-COOH unlike interaction compared to experimental data (symbols)[114] of liquid densities of amino-acid mixtures at 1 bar. (a) Water (1) + 3-aminopropanoic acid (2). (b) Water (1) + 4-aminobutanoic acid. (c) Water (1) + 5-aminopentanoic acid. (d) Water (1) + 6-aminohexanoic acid. Different colours represent mixture densities at different temperature: 293.15 K (blue); 298.15 K (green); 303.15 K (yellow); and 308.15 K (orange). The estimated experimental uncertainty in the reported densities are estimated to be $\pm 0.05\,kg/m^3$ 45

3.7 SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-CHOH unlike interaction compared to experimental data (symbols) of liquid densities of mixtures of primary amines with secondary alcohols at 1 bar. (a) 2-butanol (1) + 1-propanamine (2) [122]. (b) 2-propanol (1) + 1-propanamine (2) [122]. (c) 2-butanol (1) + 1-butanamine (2) [123]. (d) 2-hexanol (1) + 1-butanamine (2) [124]. (e) 3-hexanol (1) + 1-butanamine (2) [124]. Different colours represent mixture densities at different temperature: 298.15 K (black); 303.15 K (purple); 308.15 K (blue); 313.15 K (light blue); 318.15 K (grey); 323.15 K (pink); and 328.15 K (green). The experimental uncertainties in the density measurements are reported to be 0.01 kg/m^3 [122] and 0.1 kg/m^3 [123, 124]. 46

3.8 SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-CHOH unlike interaction compared to experimental data (symbols) of VLE and excess enthalpies of mixtures of primary amines with secondary alcohols. (a) Vapour–liquid equilibrium isobar (Txy) [119] of 2-butanol (1) + 1-butanamine (2) at 1 bar. The maximum estimated uncertainty in liquid- and vapour-phase mole fractions is ± 0.0002. (b) Vapour–liquid equilibrium isotherm (Pxy) [118] of 2-butanol (1) + 1-butanamine (2) at 328.15 K. (c) Excess enthalpy of 2-butanol (1) + 1-butanamine (2) [118] at 298.15 K and 1 bar. (d) Excess enthalpies of mixtures of 2-propanol (1) with primary amines at 298.15 K and 1 bar: 1-butanamine (black circles) [121]; 2-butanamine (light blue triangles) [121]; and 2-propanamine (purple diamonds) [120]. 47

3.9 SAFT-γ Mie calculations (curves) used in the optimisation of the COOH-CHOH unlike interaction compared to experimental data (symbols) of VLE and liquid densities of mixtures of alkanoic acids with secondary alcohols. Filled symbols denote data that were used in the optimisation and empty symbols denote data that were not used. (a) Mixture liquid density of propanoic acid (1) + 2-butanol (2) at 1 bar, and at 293.15 K (purple) and 333.15 K (black) [127]. The uncertainty in the density measurements is reported to be 0.01 kg/m^3. (b) Temperature–composition isobar illustrating the vapour–liquid equilibrium of propanoic acid (1) + 2-butanol (2) at 1 bar [125]. (c) Temperature–composition isobar illustrating the vapour–liquid equilibrium of propanoic acid (1) + 2-propanol (2) at 1 bar [125]. (d) Pressure–composition isotherms illustrating the vapour–liquid equilibria of butanoic acid (1) + 2-propanol (2) at 333.15 K (light blue); 353.15 K (purple); and 373.15 K (black) [126]. The experimental uncertainty in measured pressures is reported to be less than 0.15 kPa. In (b), (c), and (d), triangles denote dew points and circles denote bubble points. 48

3.10 SAFT-γ Mie calculations (curves) of pure-component properties used in the optimisation of the NH_2-CH unlike interaction compared to experimental data (symbols). (a) Vapour-pressure of 2-propanamine (black), 2-butanamine (light blue), and 2-methyl-1-propanamine (purple). Symbols denote different experimental sources: black circles [128]; black triangle [139]; light blue circles [129]; purple circles [131]; and purple triangles [130]. SAFT-γ calculations for 2-butanamine and 2-methyl-1-propanamine are identical because the molecules are comprised of the same groups. (b) Isobaric liquid density of 2-propanamine (black) and 2-butanamine (light blue). The dashed curve denotes calculation of density at 1 bar whereas the solid curve denotes calculation of saturated density. Symbols denote different experimental sources: black circles [132]; blue circles [134]; and blue triangles [133] 49

3.11 SAFT-γ Mie calculations (curves) of mixture properties used in the optimisation of the NH_2-CH unlike interaction compared to experimental data (symbols). (a) Pressure–composition isotherms of hexane (1) + 2-propanamine (2) [136] at 283.15 K (black), 313.15 K (purple), and 333.15 K (light blue). (b) Pressure–composition isotherms of ethane (1) + 2-propanamine (2) [135] at 279.1 K (light blue), 328.3 K (purple), and 367.9 K (black). (c) Excess enthalpies of hexane (1) + 2-butanamine (2) at 298.15 K and 1 bar (black) [138] and heptane (1) + 2-butanamine (2) at 298.15 K and 1 bar (purple) [137]. 50

3.12 SAFT-γ Mie description of pure-component properties used in the optimisation of the CONH−CONH, CONH−CH_3 and CONH−CH_2 interactions. Empty symbols denote data not used in the parameter optimisation, filled symbols denote those used in the optimisation, and curves denote SAFT-γ Mie calculations. (a) Vapour pressures of *n*-methylacetamide (black) [162, 161], *n*-ethylacetamide (pink) [142, 143], *n*-propylacetamide (navy blue) [144, 143], *n*-butylacetamide (blue) [146, 145], *n*-pentylacetamide (light blue) [147], *n*-methylpropanamide (dark green) [148, 142], *n*-butylpropanamide (green) [144], and *n*-methylhexanamide (grey) [149]. *N*-ethylacetamide and *n*-methylpropanamide comprise the same groups and are therefore represented by the same calculation (pink curve); similarly for *n*-pentylacetamide, *n*-butylpropanamide, and *n*-methylhexanamide (green curve). (b) Isobaric liquid density of *n*-methylacetamide (black) [163, 164], *n*-ethylacetamide (pink) [150, 152, 151], *n*-methylpropanamide (dark green) [154, 153, 155, 156, 157], and *n*-methylbutanamide (purple) [158] at 1 bar. Solid curves represent saturation density calculations, whereas dashed curves represent isobaric density calculations at 1 bar. *N*-ethylacetamide and *n*-methylpropanamide comprise the same groups and are therefore represented by the same calculation (pink curve). 52

3.13 SAFT-γ Mie description of n-methylacetamide-mixture properties used in the optimisation of the CONH−CONH, CONH−CH_3, and CONH−CH_2 interactions. Empty symbols denote experimental data not used in the optimisation and curves represent SAFT calculations. (a) Pressure–composition isotherms illustrating the vapour–liquid–liquid equilibrium of octane (1) + n-methylacetamide (2) [142] at 363.15 K (black) and 383.15 K (purple). (b) Pressure–composition isotherm illustrating the vapour–liquid–liquid equilibrium of decane (1) + n-methylacetamide (2) at 413.49 K [142]. (c) Excess enthalpies of n-decane (1) + n-methylacetamide (2) at 413.15 K and 1.617 MPa (black) [159] and n-octane (1) + n-methylacetamide (2) at 398.15 K and 1.891 MPa (purple) [159]. 53

3.14 SAFT-γ Mie description of alkane–amide mixture properties used in the optimisation of the CONH−CONH, CONH−CH_3, and CONH−CH_2 interactions. Filled symbols denote experimental data used in the optimisation and curves represent SAFT calculations. (a) Pressure–composition isotherms illustrating the vapour–liquid equilibrium of octane (1) + n-methylpropanamide (2) [142] at 363.15 K (black) and 383.15 K (purple). (b) Pressure–composition isotherms illustrating the vapour–liquid–liquid equilibrium of decane (1) + n-ethylacetamide (2) at 363.15 K (black) and 383.15 K (purple) [142]. (c) Temperature–composition isobar illustrating the fluid-phase diagram of n-octane (1) + n-methylpropanamide (2) at 1 bar [142]. The SAFT calculations predict an upper critical solution temperature that lies below a VLE envelope characterized by a homogeneous azeotrope. Mixture data are only available for the liquid–liquid region, represented by circles [142]. The triangle symbol [165] denotes the normal boiling point of n-octane, which had been used to parameterise the CH_3 and CH_2 groups [51]. The diamond symbol [165] denotes the normal boiling points, which had not been used in parameterising any of the model interactions. 54

3.15 SAFT-γ Mie description of properties used in the optimisation of the CONH–H_2O interaction. (a) Temperature–composition isobar illustrating the vapour–liquid equilibrium of water (1) + *n*-methylacetamide (2) at 1 bar [161]. (b) Pressure–composition isotherms illustrating the vapour–liquid equilibrium of water (1) + *n*-methylacetamide (2) at 313 K (purple) [170], 333 K (blue) [171], 373 K (light blue) [171], and 413 K (green) [159]. (c) Excess enthalpies of water (1) + *n*-methylacetamide (2) [168] at 398.15 K and 17.54 bar (black); 363.15 K and 17.54 bar (purple); and 323.15 K and 18 bar (blue). (d) Excess enthalpies [169] at 308.15 K and 1 bar of water (1) + *n*-methylpropanamide (2) (circles) and water (1) + *n*-ethylacetamide (2) (triangles). Since *n*-methylpropanamide and *n*-ethylacetamide comprise the same SAFT-γ Mie groups, they are represented by the same calculation denoted by the black curve. 56

3.16 SAFT-γ Mie description of mixture properties used in the parameter optimisation and validation of the CONH–NH_2 interaction. Filled symbols denote data that were used in the parameter optimisation and empty symbols denote data used for validation only; curves denote SAFT-γ Mie predictions. (a) Excess enthalpy of 1-hexanamine (1) + *n*-methylacetamide (2) at 363.15 K and 12.03 bar [168]. (b) Liquid density of mixtures at 298.15 K and 1 bar of water (1) + methylurea (2) in dark blue and squares [177, 172, 174, 173], water (1) + ethylurea (2) in blue and triangles [174, 175], and water (1) + butylurea (2) in light blue and circles [174]. 58

3.17 SAFT-γ Mie description of mixture properties used in the parameter optimisation of the interaction CONH–COOH. Filled symbols denote experimental data that were used in the parameter optimisation, and curves denote SAFT-γ Mie calculations. (a) Isobaric mixture density of acetic acid (1) + n-methylacetamide (2) [178] at 298.15 K (black); 323.15 K (light blue); and 348.15 K (purple). Triangles [179] and diamonds [180] denote pure acetic acid densities. (b) Temperature–composition isobar illustrating the vapour–liquid equilibrium of acetic acid (1) + n-methylacetamide (2) at 1 bar [176]; diamonds and triangles denote dew and bubble temperature data, respectively. (c) Excess enthalpy of propanoic acid (1) + n-methylacetamide (2) at 363.15 K and 11.35 bar [168]. 59

4.1 Effect of uncertainty in melting temperature and heat of fusion on the SAFT-γ Mie description of the solid–liquid equilibrium of glycine (2) in water (1) at 1 bar. (a) The sensitivity of solubility calculations to the melting temperature for a fixed enthalpy of fusion value of 22 kJ/mol. The black curve represents calculations using the reported melting temperature (569 K), whereas blue and red curves denote calculations using the maximum (578 K) and minimum (560 K) of the uncertainty range, respectively. (b) A magnified version of (a) showing the region with available glycine solubility data. (c) The sensitivity of solubility calculations to the enthalpy of fusion for a fixed melting temperature of 569 K; the black curve represents calculations using the reported enthalpy of fusion (22 kJ mol^{-1}) whereas blue and red curves denote calculations using the maximum (25 kJ mol^{-1}) and minimum (19 kJ mol^{-1}) of the uncertainty range, respectively. (d) A magnified version of (c) showing the region with available glycine solubility data. Black filled symbols:circles [111]; triangles [30]; squares [115]; and diamonds [112] denote data used in optimising the interaction COOH-NH_2, and black empty symbols [116] denote data that have not been used in the optimisation. The green triangle [117] denotes the eutectic of the mixture. . . . 71

4.2 SAFT-γ Solid–liquid solubility prediction of alanine (2) in water (1) at 1 bar. (a) Full concentration range depicting the SLE up to the melting point, denoted by a black cross symbol. (b) Low alanine mole-fraction region illustrating the eutectic region of the mixture. The green cross denotes the melting point of water [193] and the green triangle denotes the eutectic point [194]. The empty black symbols correspond to solubility experimental data from various sources: circles [111]; squares [115]; and diamonds [195]. 73

4.3 SAFT-γ Solid–liquid solubility prediction of valine (2) in water (1) at 1 bar. (a) Full concentration range depicting the SLE up to the melting point, denoted by a black cross symbol. (b) Low valine mole-fraction region. The green cross denotes the melting point of water [193] and the green triangle [194] and green diamond [117] denote experimental measurements of the eutectic point. The empty black symbols correspond to solubility experimental data from different sources: circles [111]; and squares [115]. 73

4.4 SAFT-γ Solid–liquid solubility prediction of serine (2) in water (1) at 1 bar. (a) Full concentration range depicting the SLE up to the melting point, denoted by a black cross symbol. (b) Low serine mole-fraction region illustrating the eutectic region of the mixture. The green cross denotes the melting point of water [193] and the empty black circles [111] correspond to solubility experimental data. . . 74

4.5 Parity plot of the solubility of glycine, alanine, valine, and serine in water, ethanol, 1-propanol, 1-butanol, 2-methyl-1-propanol, 2-propanol, and 2-butanol. The solid diagonal denotes exact agreement between experiments and calculations, whereas each pair of dashed and dotted lines denotes a change in order of magnitude. Sources of the experimental data can be found in Table 4.2. 75

4.6 Parity plot of the solubility in water of di- and tri-peptides made up of residues of the amino acids glycine (Gly), alanine (Ala), leucine (Leu), and serine (Ser). The solid diagonal denotes exact agreement between experiments and calculations, whereas each pair of dashed lines denotes deviation of an increasing order of magnitude. Sources of the experimental data can be found in Table 4.3. 76

4.7 Parity plot of the solubility in water of di- and tri-peptides made up of residues of the amino acids glycine (Gly), alanine (Ala), leucine (Leu), and serine (Ser). Calculations do not include the term $\Delta c_{p,i}$; i.e., it is set to 0 in the solubility calculations. The solid diagonal denotes exact agreement between experiments and calculations, whereas each pair of dashed lines denotes deviation of an increasing order of magnitude. Sources of the experimental data can be found in Table 4.3. 77

4.8 SAFT-γ Mie description of the solid–liquid equilibria of glycine in mixed aqueous alcohol solvents expressed as the mole fraction of glycine as a function of solute-free alcohol mole fraction. (a) Solubility of glycine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (green); 298.15 K (blue); 323.15 K (purple); and 333.15 K (black). Different empty symbols denote different sources: circles [202]; triangles [205]; diamonds [206]; and squares [207]. "x" symbols denote pure-ethanol solubility [199, 30, 200, 17, 203]. (b) Solubility of glycine (2) in water (1) + 1-propanol (3) at 1 bar and: 283.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Symbols denote experiments from different sources: circles [202]; triangles [29]; and squares [207]. (c) Solubility of glycine (2) in water (1) + 1-butanol (3) at 1 bar and 298.15 K. Empty circles denote mixed-solvent data [208] and "x" symbols denote pure-1-butanol solubilities from different sources [196, 18]. The dotted section of the curve represents the region of liquid–liquid immiscibility of the water/1-butanol mixture. (d) Solubility of glycine (2) in water (1) + 2-propanol (3) at 1 bar and: 288.15 K (light blue); 298.15 K (purple); and 310 K (black). Empty symbols denote experiments from different sources: circles [202]; triangles [29]; diamonds [204]; and squares [209]. 80

4.9 SAFT-γ Mie description of the solid–liquid equilibria of alanine in mixed aqueous alcohol solvents expressed as the mole fraction of alanine as a function of solute-free alcohol mole fraction. (a) Solubility of alanine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Different empty symbols denote different sources: circles [202]; triangles [199]; and squares [206]. "x" symbols denote pure-ethanol alaninesolubility [199, 30, 17, 201, 202]. (b) Solubility of alanine (2) in water (1) + 1-propanol (3) at 1 bar and: 283.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Symbols denote experiments from different sources: circles [202]; triangles [195]; and squares [29]. (c) Solubility of alanine (2) in water (1) + 1-butanol (3) at 1 bar and 298.15 K. Empty circles denote mixed-solvent data [208]. The dotted section of the curve represents the region of liquid–liquid immiscibility of the water/1-butanol mixture. (d) Solubility of alanine (2) in water (1) + 2-propanol (3) at 1 bar and: 283.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Different symbols denote different sources of experimental data: circles [202]; triangles [195]; and squares [29]. 81

4.10 SAFT-γ Mie description of the solid–liquid equilibria of valine and alanine in mixed aqueous alcohol solvents expressed as the mole fraction of amino acid as a function of solute-free alcohol mole fraction. (a) Solubility of valine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (light blue); 298.15 K (purple); and 338.15 K (black). Different empty symbols denote different sources: circles [206]; and triangles [201]. (b) Solubility of serine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (light blue); 298.15 K (purple); and 338.15 K (black). Symbols denote experiments from different sources: circles [206]; and triangles [201]. (c) Solubility of serine (2) in water (1) + 1-propanol (3) at 1 bar and: 298.15 K (light blue); 313.15 K (purple); and 333.15 K (black). Empty circles denote mixed-solvent data [210] and "x" symbols denote pure-water serine solubility data [111]. (d) Solubility of serine (2) in water (1) + 1-butanol (3) at 1 bar and 298.15 K. Empty circles denote experimental mixed-solvent solubility data [208] and "x" symbols denote pure-water and 1-butanol solubility data [197, 111].The dotted section of the curve represents the region of liquid–liquid immiscibility of the water/1-butanol mixture. 82

4.11 Calculation of the temperature dependence of the pK_{A1}, pK_{A2}, and pI of glycine using the van't Hoff equation (Equation 4.11). Parameters used in the van't Hoff equation can be found in Table 4.4. 90

4.12 SAFT-γ Mie calculations of the acid-base titration curve for glycine at 298.15 K and 1 bar for an unsaturated glycine zwitterion molar composition of 0.00265. The OH^- equivalents are calculated as the proportion of OH^- molecules required to neutralise the glycine species; i.e. 50% of the cationic glycine is neutralised at pK_{A1} and 100% is neutralised at pI, at which glycine is predominantly in the zwitterionic form. At pK_{A2}, 50% of the zwitterionic species is ionised into the anionic form. Close to pH=12, glycine is predominantly in the anionic form. . . 91

4.13 (a) Relative concentration ($\xi_{l,i}$) of glycine zwitterion (continuous black curve), cation (long-dashed red curve), and anion (short-dashed blue curve) at 298.15 K and 1 bar, as a function of pH. The pK_{A} and pI values [222] are denoted by the vertical black dotted lines. (b) The solubility of glycine in water at 298.15 K (black); 318.15 K (blue); and 348.15 K (orange) and at 1 bar as a function of pH. The continuous curves represent SAFT-γ Mie calculations and symbols represent experimental solubility data; circles denote pH-dependent data by Needham et al. [215] and triangles solubility data at the isoelectric point (pI) by Dalton and Schmidt [111]. The filled circle represents the data point used in optimising the $NH_3^+ - COO^-$ interaction, while empty symbols denote data not used in the parameter estimation. 92

4.14 (a) The activity of water (a_{w}) calculated using SAFT-γ Mie as a function of pH for the system containing glycine (described in Figure 4.13) at 298.15 K (black), 318.15 K (blue), and 348.15 K (orange) and at 1 bar. (b) The solubility of glycine in water at 298.15 K (black), 318.15 K (blue), and 348.15 K (orange) and at 1 bar as a function of pH. Solid lines represent calculations in which the real value of a_{w} is used and dashed lines represent calculations in which a_{w} is taken to be unity. 94

4.15 (a)Relative concentration ($\xi_{l,i}$) of alanine zwitterion (continuous black curve), cation (long-dashed red curve), and anion (short-dashed blue curve) at 298.15 K and 1 bar, as a function of pH. The pK_{A} and pI values [222] are denoted by the vertical black dotted lines. (b) The solubility of alanine in water at 298.15 K (black); 318.15 K (blue); and 348.15 K (orange) and 1 bar, as a function of pH. The continuous curves represent SAFT-γ Mie predictions and the symbols the experimental solubility data; circles denote pH-dependent solubility data by Tseng et al. [216] and triangles denote solubility data at the isoelectric point (pI) by Dalton and Schmidt [111]. 95

5.1 A cartoon representation of the systems studied in this chapter. NVT simulation boxes contain solvent molecules (teal), cations (purple), and anions that consist of negatively charged segments (red) bonded to uncharged segments (grey). (a) NVT simulation box containing a trimer anion, in which the charged segment is at the head of the molecule, representing the systems studied in Section 5.6.1. (b) NVT simulation box containing a dimer anion with the uncharged segment being larger than the negatively charged segment, representing the systems studied in Section 5.6.2. (c) NVT simulation box containing a trimer anion, in which the charged segment is at the centre of the molecule, representing the systems studied in Section 5.6.3. 111

5.2 Schematic diagram representing the anionic species modeled to study the influence of chain length: (a) Monomeric, (b) Dimer, (c) Trimer, (d) Tetramer, and (e) Pentamer ion. 112

5.3 Coulombic contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Symbols denote EETM simulations of salts in which the anion is: a monomer (black); a dimer (purple); a trimer (blue); a tetramer (light blue); and a pentamer (green), where the charged bead is placed at the head of the ionic chain. The black solid curve denotes MSA calculations using Model G and dashed curves denote MSA calculations using Model M. 114

5.4 Coulomb contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Symbols denote EETM simulations of a salt in which the anion is a dimer consisting of the negatively charged ion segment bonded to a neutral monomer segment of different sizes. Different colours represent different sizes of the neutral monomer segment relative to the size of the ion ($\sigma_{seg}/\sigma_{ion}$): 0.25 (orange); 1 (purple); and 1.5 (dark red). The solid black curve denotes MSA calculations using Model G and dashed curves denote MSA calculations using Model M. 116

5.5 Schematic diagram representing the anionic species modeled to study the influence of charge shielding: (a) Monomer anion; (b) Dimer anion; (c) Trimer anion with the charge at the centre; (d) Tetramer anion with the charge at the centre; and (e) Pentamer anion with the charge at the centre. Note that this is a 2D representation; in the simulation models, the pentamer ion forms a tetrahedron. 118

5.6 Coulomb contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Symbols denote EETM simulations of salts in which the anion is: a monomer (black); a dimer (purple); a trimer (blue); a tetramer (light blue); and a pentamer (green). The anion's charged segment is always in the middle of the molecule. (a) EETM simulation data of the ionic chemical potential of the salt solutions considered. Solid lines represent calculations of the segment packing fraction, η_{seg}. (b) EETM simulation data alongside MSA calculations using Model G (black solid curve) and Model M (dashed curves). 119

5.7 A 2D representation of the volume around the central charge that is shielded from Coulombic interactions due to the presence of the uncharged segment. The shielded volume is the spherical conic section between the surface of the central charge at $r = \sigma_{ii}/2$ and the cut-off radius $r = r_c$ and is defined by the aperture angle ϕ. 122

5.8 Coulombic contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Empty symbols denote EETM simulations of salts in which the anion is a branched oligomer. Filled symbols are chemical potential simulation data of the monomeric anion scaled by φ_{free}. The reference monomer anion simulation data are denoted by empty black circles. The black curve denotes MSA calculations using Model G whereas the dashed curves represent calculations using Model M. (a) Chemical potential of the dimer-anion salt. (b) Chemical potential of the trimer-anion salt. (c) Chemical potential of the tetramer-anion salt. (d) Chemical potential of the pentamer-anion salt. 124

Chapter 1

Introduction

1.1 Motivation and objectives

Molecular thermodynamic models are at the heart of advancements in the engineering of our modern world. From the design and fabrication of materials used in everyday products to the design of large-scale manufacturing processes, molecular thermodynamics serves as the bridge between our most advanced understanding of the world at the quantum and atomic scales and the useful physico-chemical properties at the macroscale. This multiscale "bridge" is also indispensable in the understanding and development of modern pharmaceutical drugs; it is essential in the development of new chemical entities (NCEs), specifically in the synthesis and purification of active pharmaceutical ingredients (APIs).

In the world of drug design and development, computational thermodynamic models of all kinds have been increasingly used to aid in the screening, design, production, and purification of drug products. A key focus of such models is the prediction of solid–liquid solubility of the API molecule across a range of thermodynamic conditions and in a variety of solvent systems. According to Savjani et al. [1], approximately 40% of NCEs developed in the industry exhibit poor solubility. Furthermore, a huge amount of time, effort, and resources are spent on the development of a new drug. Thus, there is a constant need for computational models that can predict the solubility of APIs to improve the efficiency of the drug design process.

In recent decades, peptide therapeutics have gained increasing attention due to their specificity, potency, and low toxicity [2], with the market share of peptide therapeutics expected to increase significantly; the global market share of peptide-based therapeutics was estimated to be $21.5 billion in 2016 [3] and is projected to reach $68.7 billion by 2030 [4]. Peptides are gaining popularity as APIs for the treatment of illnesses such as diabetes [5], and some cancers [6, 7]. Of special interest are those that mimic natural hormones or can disrupt protein–protein interactions [8], while also exhibiting very low toxicity and good *in-vivo* stability [9]. As new peptide-based therapeutics are proposed, a good understanding of the thermodynamic properties of these molecules and of their mixtures is crucial for product development and manufacturing. In particular, the solubility of APIs in pure and mixed solvents is especially important, as it determines the bioavailability of the drug and the optimal solvent choice in the synthesis and purification stages. With the number of peptide-based APIs in discovery increasing year by year, it can be expensive and time-consuming to determine experimentally the solubility of each candidate under broad thermodynamic conditions and diverse solvent media. In this context, inexpensive computational tools that can be used to model and predict accurately the solid–liquid solubility of APIs in general, and peptides in particular, are becoming increasingly important in the development and design of pharmaceutical products.

Since peptides are oligomers made from a finite pool of amino acid residues, a good understanding of the physico-chemical properties of amino acid mixtures serves as a key to the modelling of peptide mixtures. Due to their zwitterionic nature, amino acids are generally soluble in water and insoluble in non-polar organic solvents, although the nature and size of side chains play an important role in the extent of their solubility. For example, amino acids with hydrophilic, i.e., charged and polar, side chains (e.g., lysine and serine) tend to have higher aqueous solubility than those with hydrophobic side chains (e.g., leucine and valine). Peptides exhibit varying solubility depending on their size, the nature of any side chains, the precise sequence of amino acid residues in their backbone, and the structure(s) they exhibit in the solid state. These factors contribute to other important phenomena, such as the formation of intramolecular hydrogen bonds and aggregate formation that further impact their phase-equilibrium behaviour.

In this thesis, the aim is to demonstrate the applicability of the SAFT-γ Mie GC framework in the prediction of solid–liquid solubility of amino acids and oligopeptides in pure and mixed solvents and under pH changes. To achieve that goal, the relevant SAFT-γ Mie parameters are optimised by adjusting to experimental thermodynamic data of a wide range of pure compounds and mixtures.

1.2 Thesis overview

In Chapter 2, a literature review of thermodynamic approaches that have been used in modelling the solubility of amino acids and oligopeptides is presented, starting with early theoretical work and concluding in recent advances in molecular thermodynamic models. This is done to highlight the common challenges faced in modelling the solubility of amino acids and peptides and some of the gaps that I try to address in this work.

In Chapter 3, the methodology followed for modelling the solid–liquid equilibria of the amino acids and oligopeptides in this work is presented. The chapter begins with the derivation of the solid–liquid equilibrium (SLE) equation to highlight the main assumptions made and the consequences they have on the rest of the work. Then, the SAFT-γ Mie group-contribution framework is introduced in detail by discussing each of the free-energy terms that are used to model the systems of interest. After that, a discussion of the parameter optimisation strategy and error quantification is presented. Then, the SAFT-γ Mie models for the amino-acid and oligopeptide solutions are presented to highlight the SAFT-γ Mie groups needed. Finally, each of the newly developed SAFT-γ Mie like and unlike interactions is discussed by highlighting the experimental data used for their optimisation.

In Chapter 4, a discussion of the solid–liquid solubility prediction is presented for various amino acids and oligopeptides in pure and mixed solvents at the isoelectric point (pI) and for the amino acids glycine and alanine under changing pH. This chapter demonstrates the applicability of the model developed to predict the solubility of many amino acids and oligopeptides under various thermodynamic conditions and in a variety of solvent media. Additionally, the results

discussed demonstrate the transferability of the SAFT-γ Mie parameters.

In Chapter 5, a molecular simulation study is conducted to benchmark mean-spherical-approximation (MSA) calculations of the ionic chemical potential for salt solutions containing ionic molecules. The method of expanded ensembles is used to investigate the influence that variables, such as the size of the ionic molecule and the position of the charge within the molecule, have on the ionic contribution to the chemical potential. Simulation data are compared with MSA calculations using two possible approaches for resolving the charge on an ionic chain.

In Chapter 6, key achievements of this thesis are summarised, and new avenues for future research work are proposed.

Chapter 2

Thermodynamic models of solid–liquid solubility for amino acids and oligopeptides

2.1 Introduction

Solid–liquid solubility is a measure of how much of a solid substance (the solute) can dissolve in a liquid medium (the solvent) at given thermodynamic conditions (T, P, $\mathbf{x}$, pH) to form a homogeneous mixture. Solubility depends strongly on the intermolecular forces between the solute and solvent molecules [10]. In thermodynamic models, this dependence on intermolecular forces is accounted for by the activity coefficient of the solute in the mixture at equilibrium. Additionally, solid–liquid solubility also depends strongly on the solid-state properties of the solute.

Solid–liquid solubility is an essential property that is of great importance in many fields, including predicting pollutant concentrations in bodies of water [11, 12], developing advanced materials like ionic liquids or eutectic solvents [13], and the design of crystallization processes for active pharmaceutical ingredients (API)[14] – which is the focus of this thesis.

This chapter presents an overview of thermodynamic modelling approaches that have been published to model the solid–liquid solubility of amino acids and oligopeptides. Section 2.2, a historical account of some of the earliest theoretical treatments of amino acid solutions, which have laid the groundwork for decades of research into the subject, is introduced. In 2.3, an overview of studies that have employed local-composition models to model the solubility of a variety of amino acids and oligopeptides, which have seen a surge in interest starting in the 1980s, is presented. In 2.4, an overview of more recent developments in the treatment of amino-acid and oligopeptides solubility, specifically those that have utilised molecular-based models, is presented.

2.2 Early theoretical developments

It has been long shown that amino acids exhibit properties of strong electrolytes when dissolved in solvents of high dielectric constant. Unlike strong electrolytes, however, amino acids do not increase the electrical conductivity of the solution in a significant manner. Adams [15] and Bjerrum [16] suggested that the acidic and basic groups of the amino acid ionize fully giving rise to a zwitterionic molecule with a zero net charge. In later years, the zwitterionic theory gained traction due to an accumulation of experimental evidence. Cohn [17] reported high values of the dielectric constant for amino-acid solutions, which could only be explained by very high dipole moments. Dipole moments of the magnitude that Cohn measured could only be produced by the amphoteric ionisation of the amino acids, yielding fully dissociated zwitterions. Building on the previous research, Wyman and McMeekin [18] used the Debye-Hückle [19] theory to study the effect of adding salts on the activity of zwitterions in amino-acid solutions and found that the logarithm of the activity coefficient did not contain a term proportional to the square root of the ionic strength of the salt. They concluded that a zwitterion behaves more like a highly dipolar non-electrolyte than like a real ion in salt solutions of low ionic strength. Later, Kirkwood [20] used a statistical-mechanical approach to model aqueous solutions of glycine, treating the amino acid as a sphere with discrete point charges and the solvent as a dielectric continuum to study the effect of the dielectric constant of the solvent on the solubility of amino acids.

However, Kirkwood neglected the dipole–dipole interactions between the amino acid molecules; thus, his theory applied strictly only to solutions at infinite dilution. In the same year, Cohn et al. [17] performed a systematic experimental investigation of the solubility behaviour of amino acids in water, alcohols, and alcohol–water mixtures, in which they found that amino acids behave similarly to strong electrolytes; i.e., they are soluble in water but highly insoluble in alcohols, and exhibit high solid densities, reflecting the charged nature of the molecules. Cohn et al. reported that "the activity coefficients of the larger amino acids deviate far more than those of glycine from any relation proportional to change in the mole fraction of alcohol or the dielectric constant of the solutions." This partly explains why Kirkwood's [20] theoretical treatment was accurate in the case of glycine only at very low concentrations.

2.3 Local composition models

In the 1980s and 1990s, there was a revived interest in studying amino-acid and oligopeptide systems through the adoption of semi-empirical and local composition models such as the Wilson equation [21], the non-random two-liquid (NRTL) equation [22], and the Flory-Huggins solution theory [23].

Local composition models are semi-empirical Gibbs free-energy models that were proposed to take into account the irregular structure of mixtures due to asymmetrical energetic interactions. Unlike the regular solution theory [24], which assumes that the mixture interactions are independent of each other, local composition models assume that the composition of a component B in the region surrounding a molecule of component A is markedly different from the bulk composition. This irregularity can be caused by specific interactions, such as hydrogen bonding and polarity. Wilson [21] was the first to hypothesise local composition; his model expresses the Gibbs free energy of a binary mixture as a function of the characteristic energy differences and the pure-component molar volumes. The Wilson equation was shown to successfully represent many systems of different chemical types, particularly those of solutions of associating or polar components in nonpolar solvents. However, a major disadvantage of the Wilson equation is

that it cannot represent partial miscibility. Thus, it is only useful for liquid solutions that are fully miscible.

The NRTL equation [22], developed based on the local composition theory of Wilson's, expresses the Gibbs free energy of a mixture as a function of the characteristic energy difference–like the Wilson equation. The NRTL equation differs from Wilson's equation in that it includes a non-randomness parameter instead of the pure-component molar volumes. This modification allows the NRTL equation to describe partially miscible as well as fully miscible mixtures. Later, the universal quasichemical activity coefficient model (UNIQUAC) [25] was developed with the goal of retaining some of the advantages of the Wilson equation without restriction to completely miscible mixtures. It expands on the quasichemical theory of Guggenheim [26] for nonrandom mixtures to solutions of components of different sizes by introducing the local area fraction as the primary concentration variable. The expression for the Gibbs free energy of the UNIQUAC model consists of a residual term that accounts for the intermolecular forces and depends on binary interaction parameters fitted to mixture data and a combinatorial term that accounts for the entropic contribution due to differences in the sizes and shapes of the mixture components and is dependent only on pure-component data.

In the context of modelling the solid–liquid solubility of amino acids and oligopeptides, Chen et al. [27] combined the NRTL equation with a Pitzer–Debye–Hückel term [28] that varies inversely with the solvent dielectric constant. Unlike Kirkwood's model, this model treats the coefficient of the electrostatic term as an adjustable parameter which allows it to capture the varying behaviour of different solvents. This model was used to correlate the solubility of glycine and alanine in aqueous ethanol solutions.

Orella and Kirwan [29] combined an excess-solubility approach with the three-suffix Margules, NRTL, and Wilson activity coefficient models to correlate solubility data of amino acids in mixed water + alcohol solvents. They used published solvent–solvent parameters and fitted solute–solvent parameters using solubility data. They reported that out of the three models, Wilson's yielded the best agreement with experiments, although, as concluded later by Ferreira et al. [30], the global quality of results in Orella and Kirwan's work seems to contradict this

conclusion.

Gude et al. [31] combined the excess-solubility approach with a simple activity-coefficient model comprising a combinatorial term based on the Flory–Huggins theory and a Margules residual term and reported good correlations of the solubility and partition coefficient of amino acids and small peptides in mixed water + alcohol solvents. Van Berlo et al. [32] extended the use of the model of Gude et al. to correlate the solubility of glycine in the ternary solvent water + ethanol + 1-butanol. They used phase behaviour data of the water + ethanol + 1-butanol solute-free system, along with single-solvent solubility data of glycine to develop a one-parameter excess Gibbs model to predict the solubility and partition coefficient of glycine in the ternary solvent system at the same temperature. Rudolph et al. [33] used the same model to study the solubility and partition coefficients of amoxicillin, ampicillin, and their precursors. They found, however, that they could not reproduce accurately the increasing relative solubility of the molecules in the aqueous phase observed for increasing 1-butanol concentration. To capture the correct aqueous solubility behaviour, they replaced the combinatorial term in Gude et al.'s model with that from the universal quasichemical activity-coefficient (UNIQUAC) model to account for the size and shape differences of the molecules. Neither of the two models was able to describe simultaneously the solubilities and partitioning accurately. The model of Ferreira et al. [30], in which the excess solubility approach is combined with the NRTL equation, showed an improvement in correlating the solubility of amino acids and small peptides in mixed solvents, as compared to the similar models of Orella and Kirwan [29] and Gude et al. [31], while using the same number of adjustable parameters.

2.4 Molecular models

In recent decades, molecular-based models, especially those based on the statistical associating fluid theory (SAFT)[34, 35, 36, 37], have been increasingly adopted to model the solid–liquid solubility of amino acids and oligopeptides. The term "molecular models" used here refers to models that are based on intermolecular potentials describing the forces between the molecules

in a system, as opposed to local-composition and excess Gibbs models, which are based on solution theory and represent the deviation from properties of the ideal solution. SAFT (EoS), in particular, are written in terms of the Helmholtz free energy of the mixture, from which thermodynamic properties can be derived.

Khoshkbarchi and Vera [38] developed a simplified perturbed hard-sphere model to correlate the activity coefficients and solubilities of amino acids in water, both at the isoelectric point (pI) and at varying pH. A Lennard-Jones (LJ) potential was used to model the dispersion forces and a Keesom term was introduced to account for the dipole–dipole interactions between amino acids. The solvent was treated as a dielectric continuum. The dipole moments of the amino acids were calculated using a quantum-mechanical approach whereas the LJ potential parameters were adjusted using activity-coefficient data. The enthalpy of fusion and melting temperature were adjusted using the amino acid solubility curves, and in order to model the pH-dependent solubility, experimental dissociation equilibrium constants were used.

In later works, the authors extended their treatment to study how the presence of salts in solution affects the solubility of amino acids [39, 40] and to model the solubility of a mixture of two amino acids in aqueous solutions [40]. These models marked a significant improvement in the description of the physical behaviour of amino acid solutions, especially in accounting for the large dipole moments of the amino acids, although the experimental solubility data required for adjusting the many model parameters are often scarce. It is also important to note that a key shortcoming of these approaches in modelling pH-dependent solubility is that they account for the speciation of the amino acids solely through the concentration of protons in the system. The activities of the amino acid cation and anion are accounted for only as mole fractions [38]. This assumption applies only to the system at or close to the isoelectric point and cannot be extrapolated to other pH values.

Fuchs et al. [41] have used the PC-SAFT [42] version of the statistical associating fluid theory (SAFT) [36, 37] to model the solubility of glycine, DL-alanine, and DL-methionine at their pI in pure water and alcohols and in mixed water + alcohol mixtures. Pure-component model parameters were adjusted using vapour-pressure and liquid-density data of the amino

acid aqueous solutions and an additional solute–solvent interaction parameter was adjusted using solubility data. Moreover, the enthalpy of fusion and melting temperature were also treated as adjustable parameters rather than using experimental values as had been the case in previous works; these solid-state properties were first calculated using the group-contribution method of Marrero and Gani [43] and then allowed to vary within the average deviation reported for the method (16%) to provide the best description of the experimental solubility data. The solubility in mixed-solvent systems was then predicted in fair agreement with experiments. The same model was also used to predict the solubility of the amino acids at variable pH, in good agreement with experimental data, although the influence of the different ionized forms of the amino acid was neglected (much as in [38]), with speciation accounted for solely through the concentration of protons. By adopting this strategy, the need to model the electrostatic interactions between amino acid species is circumvented.

Cameretti and Sadowski [44] also used the PC-SAFT equation of state (EoS) to correlate the density and vapour pressure of aqueous solutions of glycine, alanine, serine, proline, and valine, treating the enthalpy of fusion and melting temperature as adjustable parameters. The same model parameters were then used to model the properties of peptides with the additional adjustment of the segment-number parameter. Ferreira et al. [45] later employed a different strategy, treating amino acids as non-associating, and using only three component-specific PC-SAFT parameters; these were obtained by adjustment using liquid densities, activity and osmotic coefficients, vapour pressures, and water activities of unsaturated aqueous amino acid solutions. While neglecting association interactions led to a decrease in the number of model parameters, a binary solute–solvent parameter was needed to yield acceptable agreement with experiments. Additionally, the melting properties of the amino acids were adjusted using aqueous solubility data. The resulting prediction of solubility in pure alcohols showed poor agreement with experimental data. In both works the authors treated the amino acids and peptides without considering their speciation.

Grosse Daldrup et al. [46, 47] used the PC-SAFT EoS to model the mixed-solute solubility in water of amino acids of similar and of differing pIs, at variable pH, and Held et al. [48] modelled the density, vapour-pressure depression, activity coefficient, and solubility of aqueous solutions

of an extensive list of amino acids. The model parameters in these works were adjusted using experimental liquid-density and activity-coefficient data of aqueous solutions, with the melting properties treated as adjustable parameters. These studies capture the dissociation equilibria associated with pH change but do not incorporate a charge in any of the species, including the amino acid cation and anion. Charged PC-SAFT models have been used by Wysoczanska et al. [49], who studied the density, solubility, and partition coefficients of dinitrophenylated amino acids in aqueous two-phase systems, and by Aliyeva et al. [50] who have studied the impact of the addition of salts on the solubility of aromatic and dicarboxylic amino acids using the ePC-SAFT approach.

As highlighted in this literature review, several recurring limitations are evident in existing models for predicting the solubility of amino acids and oligopeptides. First, many models rely heavily on experimental data of amino acids and peptides for parameter fitting, which restricts their predictive capabilities–particularly problematic for peptides, given their vast diversity and the limited availability of solubility data. Consequently, these models often suffer from limited transferability and applicability beyond the specific systems they were developed for. Moreover, a number of these approaches involve a large set of adjustable parameters, which are typically optimised for only a narrow range of systems, further constraining their generalisability. In this work, I seek to address some of these limitations by developing models within the SAFT-γ Mie group-contribution (GC) framework [51, 52, 53, 54] that require minimal amino-acid and oligopeptide data and that are transferable across various solvent media and a broad range of conditions. Accurate modelling of phase-equilibrium properties is an essential part of developing thermodynamic models, as they provide fundamental insight into the interactions between molecules. Furthermore, the parameterisation of thermodynamic models relies on an accurate description of phase equilibrium behaviour. As such, the task of applying a thermodynamic model of any kind starts with formulating the phase-equilibrium relations needed to describe the properties of the system of interest.

2.5 Modeling phase equilibria

When working with a free-energy formulation of an equation of state (EoS), such as the case in SAFT-γ Mie [51, 52, 53, 54], one can obtain all thermodynamic properties needed to describe the phase-equilibrium behaviour of a system from the derivatives of the Helmholtz free energy. The phase equilibrium between any number of phases is premised on the thermal, mechanical, and chemical equilibria between those phases [10]. These conditions of equilibrium can be expressed as

$$\begin{aligned} T^1 = T^2 &= ... = T^\pi, \\ P^1 = P^2 &= ... = P^\pi, \text{and} \\ \mu_i^1 = \mu_i^2 &= ... = \mu_i^\pi, \end{aligned} \tag{2.1}$$

where the superscript π denotes the number of phases and μ_i^π is the chemical potential of component i in phase π. For a system at temperature T with a total volume V and composition $\mathbf{x}$, the pressure is obtained from the Helmholtz free energy (A) as

$$P = -\left.\frac{\partial A(T,V,\mathbf{x})}{\partial V}\right|_{T,\mathrm{N}}. \tag{2.2}$$

The chemical potential at $(T,V,\mathbf{x})$ can be calculated as

$$\mu_i(T,V,\mathbf{x}) = -\left.\frac{\partial A(T,V,\mathbf{x})}{\partial N_i}\right|_{T,V,\mathrm{N_{j\neq i}}}. \tag{2.3}$$

The residual chemical potential, given by

$$\mu_i^{\mathrm{Res}}(T,P,\mathbf{x}) = -\left.\frac{\partial A^{\mathrm{Res}}(T,V,\mathbf{x})}{\partial N_i}\right|_{T,V,\mathrm{N_{j\neq i}}} - RT\ln Z(T,P,\mathbf{x}), \tag{2.4}$$

is used to calculate the fugacity coefficient of each component in the mixture, which is given by

$$\ln \hat{\varphi}_i(T, P, \mathbf{x}) = \frac{\mu_i^{\mathrm{Res}}(T, P, \mathbf{x})}{RT}, \tag{2.5}$$

where $A^{\mathrm{Res}} = A - A^{\mathrm{Ideal}}$ is the residual Helmholtz free energy and $Z = PV_P/NRT$ is the compressibility factor, with V_P is the total volume at the specified pressure.

In practice, the system is usually defined by the pressure rather than the volume, which requires a density root-finding routine to calculate the volume at the system pressure. The SAFT-γ Mie GC approach has been shown to produce accurate results for modelling various types of phase-equilibrium properties, including vapour–liquid equilibria (VLE), liquid–liquid equilibria (LLE), and vapour–liquid–liquid equilibria (VLLE) of a wide range of systems [51, 52, 56, 57, 58, 59, 60, 61, 62, 55].

2.5.1 Solid–liquid equilibrium

The solid–liquid solubility of a solute in solution is the amount of solute that is dissolved in the liquid solution in equilibrium with the solid phase. Because this is a phase-equilibrium problem, one can start with the equality of chemical potentials, at T and P, of the solute, component i, in the liquid and solid phases, as above, but considering now specifically the equilibrium between the solid phase and liquid phase, i.e.,

$$\mu_i^{\mathrm{S}}(T, P, \mathbf{x}^{\mathrm{S}}) = \mu_i^{\mathrm{L}}(T, P, \mathbf{x}^{\mathrm{L}}), \tag{2.6}$$

where the superscripts S and L denote the solid and liquid phases, respectively. This expression can be reformulated in terms of the fugacities of component i as

$$f_i^{\mathrm{S}}(T, P, \mathbf{x}^{\mathrm{S}}) = f_i^{\mathrm{L}}(T, P, \mathbf{x}^{\mathrm{L}}). \tag{2.7}$$

Assuming that upon equilibrium no solvent molecules have crossed from the liquid phase to the solid phase, i.e., the solid phase consists solely of the solute molecules i and the liquid

phase contains both solute and solvent molecules, Equation 2.7 becomes

$$f_i^{\mathrm{S},0}(T,P) = f_i^{\mathrm{L}}(T,P,\mathbf{x}^{\mathrm{L}}), \tag{2.8}$$

where the subscript 0 denotes pure-component fugacity.

Taking the pure solute i as a subcooled liquid at the system's temperature T and pressure P to be the reference state $f_i^{\mathrm{L},0}$, the fugacity of component i in the liquid phase can be expressed as

$$f_i^{\mathrm{L}}(T,P,\mathbf{x}^{\mathrm{L}}) = x_i^{\mathrm{L}}\gamma_i^{\mathrm{L}}(T,P,\mathbf{x}^{\mathrm{L}})f_i^{\mathrm{L},0}(T,P), \tag{2.9}$$

where x_i^{L} is the composition of the solute in solution, i.e., the solubility, and γ_i^{L} is the activity coefficient of i in the liquid phase at the system's temperature, pressure, and composition.

Thus, Equation 2.8 becomes

$$f_i^{\mathrm{S},0}(T,P) = x_i^{\mathrm{L}}\gamma_i^{\mathrm{L}}(T,P,\mathbf{x}^{\mathrm{L}})f_i^{\mathrm{L},0}(T,P). \tag{2.10}$$

Rearranging Equation 2.10 leads to

$$x_i^{\mathrm{L}}\gamma_i^{\mathrm{L}}(T,P,\mathbf{x}^{\mathrm{L}}) = \frac{f_i^{\mathrm{S},0}(T,P)}{f_i^{\mathrm{L},0}(T,P)}. \tag{2.11}$$

The right-hand side of Equation 2.11 depends only on the properties of the solute, regardless of the nature of the solvent, and it can be expressed in terms of the molar Gibbs free energy change from points a to d in the thermodynamic cycle in Figure 2.1.

The molar Gibbs free energy change associated with the transition from a to d is

$$\underset{a\to d}{\Delta g_i} = RT\ln\frac{f_i^{\mathrm{L},0}(T,P)}{f_i^{\mathrm{S},0}(T,P)}. \tag{2.12}$$

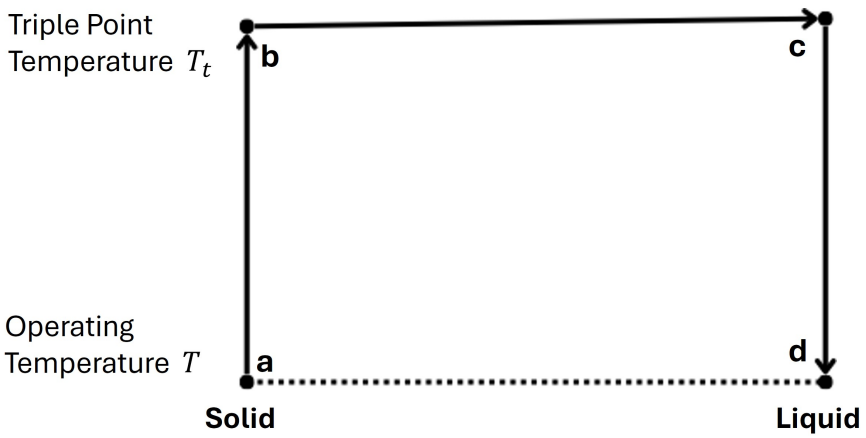


Figure 2.1: Thermodynamic cycle representing the transition from the solid state at T and P to the subcooled liquid state at T and P, adapted from Prausnitz et al. [10]. State **a** represents the solute in the solid phase at the operating temperature T and pressure P, state **b** represents the solute in the solid phase at the triple point (T_t and P_t), state **c** represents the solute in the liquid phase at the triple point (T_t and P_t), and state **d** represents the subcooled liquid at T and P.

The molar Gibbs free energy is defined as

$$\underset{a\to d}{\Delta g_i} = \underset{a\to d}{\Delta h_i} - T\underset{a\to d}{\Delta s_i}\,, \tag{2.13}$$

where $\underset{a\to d}{\Delta h_i}$ and $\underset{a\to d}{\Delta s_i}$ are the molar enthalpy and entropy changes associated with the transition $a \to d$. Because the molar Gibbs free energy is a state function, the transition $a \to d$ can be expressed as the sum of the transitions: $a \to b$, representing the "heating up" of the solid from the system's temperature T to the triple point T_t (assuming $T < T_t$); $b \to c$, representing the isothermal and isobaric phase change (fusion) from the solid to the liquid at T_t and P_t; and $c \to d$, representing the "cooling down" of the liquid from the triple point (T_t and P_t) to the system's T and P. For the molar enthalpy change, this leads to

$$\underset{a\to d}{\Delta h_i} = \underset{a\to b}{\Delta h_i} + \underset{b\to c}{\Delta h_i} + \underset{c\to d}{\Delta h_i}\,, \tag{2.14}$$

where $\underset{a\to b}{\Delta h_i}$ is associated with the molar enthalpy change in the solid phase ($\underset{a\to b}{\Delta h_i} = \int_T^{T_{\mathrm{t},i}} c^{\mathrm{S}}_{\mathrm{p},i} dT$), and similarly, $\underset{c\to d}{\Delta h_i}$ is associated with the molar enthalpy change in the liquid phase ($\underset{c\to d}{\Delta h_i} = \int_{T_{\mathrm{t},i}}^{T} c^{\mathrm{L}}_{\mathrm{p},i} dT$). The term $\underset{b\to c}{\Delta h_i}$ describes the latent enthalpy change associated with fusion at the

triple point ($\underset{b\to c}{\Delta h_i} = \Delta h_i^{\text{fus}}(T_{\text{t},i})$). Thus, Equation 2.14 becomes

$$\underset{a\to d}{\Delta h_i} = \Delta h_i^{\text{fus}}(T_{\text{t},i}) + \int_{T_{\text{t},i}}^{T} \Delta c_{\text{p},i}^{\text{SL}} dT \,, \tag{2.15}$$

where $\Delta c_{\text{p},i}^{\text{SL}} \equiv c_{\text{p},i}^{\text{L}} - c_{\text{p},i}^{\text{S}}$.

Similarly, for the entropy change from a to d,

$$\underset{a\to d}{\Delta s_i} = \underset{a\to b}{\Delta s_i} + \underset{b\to c}{\Delta s_i} + \underset{c\to d}{\Delta s_i} \,, \tag{2.16}$$

which, following a similar process to that of the enthalpy change, becomes

$$\underset{a\to d}{\Delta s_i} = \Delta s_i^{\text{fus}}(T_{\text{t},i}) + \int_{T_{\text{t},i}}^{T} \frac{\Delta c_{\text{p},i}^{\text{SL}}}{T} dT \,, \tag{2.17}$$

where $\Delta s_i^{\text{fus}}(T_{\text{t},i})$ is the entropy of fusion at the triple point which, at equilibrium, can be defined as

$$\Delta s_i^{\text{fus}} = \frac{\Delta h_i^{\text{fus}}}{T_{\text{t},i}} \,. \tag{2.18}$$

Substituting Equations 2.12, 2.15, 2.17, and 2.18 into Equation 2.12 and replacing the triple-point temperature $T_{\text{t},i}$ with the melting-point temperature T_i^{fus}–because the melting-point temperature is more abundantly available and for most compounds, the two are very similar–we get the well-known solubility equation [10]

$$\begin{aligned} \ln x_i(T,P,\mathbf{x}) = \; & -\frac{\Delta h_i^{\text{fus}}(T_i^{\text{fus}},P)}{R}\left(\frac{1}{T}-\frac{1}{T_i^{\text{fus}}}\right) \\ & -\frac{\Delta c_{p,i}^{\text{SL}}(T_i^{\text{fus}},P)}{R}\left(\ln\left(\frac{T_i^{\text{fus}}}{T}\right)-\frac{T_i^{\text{fus}}}{T}+1\right) \\ & -\ln\gamma_i(T,P,\mathbf{x}), \end{aligned} \tag{2.19}$$

where the activity coefficient of i γ_i is calculated from a thermodynamic model, in this work using SAFT-γ Mie, and the melting properties, Δh_i^{fus} and T_i^{fus}, are obtained from experiment.

The second term on the right-hand side of Equation (2.19) is often neglected, especially when the difference between T and T_i^{fus} is small [63]. In this thesis, this term is only neglected when the relevant $\Delta c_{p,i}^{\text{SL}}$ data are not available.

Furthermore, this framework can be extended to capture various kinds of solid–liquid equilibria. For example, the solubility of a solute in mixed solvents can easily be modelled by incorporating the properties of the mixed solvent in the activity coefficient of the solute. In principle, this is done with the same assumption employed in pure-solvent solubility, namely that the solid phase consists solely of the solute molecules. Only the interactions between the additional solvent with both the solute and the other solvents are needed to achieve this description.

Many systems relevant to the pharmaceutical industry form eutectic mixtures, in which the mixture exhibits a lower melting point than its components. In practice, a binary eutectic mixture is modelled by solving the SLE equation 2.19 twice: first by considering the solubility of component 1 in component 2 and then by considering the solubility of component 2 in component 1. As such, the melting properties of both compounds are required and the eutectic point is determined at the intersection of the curves produced in these two calculations.

The assumption that the solid phase is purely made up of solute molecules is not always valid. Using the current framework for systems that exhibit miscibility in the solid phase can yield erroneous results. Generally, it is not easy to know, a priori, whether a mixture will form a miscible solid phase. In practice, however, the general rule is that solid-phase miscibility is usually negligible when the two molecules are appreciably different in shape and size [10].

2.6 Conclusions

In this chapter, a review of classical thermodynamic approaches to modelling the solid–liquid solubility of amino acids and oligopeptides has been presented. Early historical developments, local composition models, and molecular-based equations of state have been discussed. Key limitations and challenges in the literature are identified to establish a foundation and motiva-

tion for the current work. The classical thermodynamic framework for the calculation of phase equilibrium is outlined, leading to a derivation of the solid—liquid equilibrium (SLE) equation from first principles. The application of the thermodynamic framework for solid–liquid equilibrium has been discussed, paving the way for the modelling approach developed in the subsequent chapters.

Chapter 3

SAFT-γ Mie theory and model development for amino-acid and oligopeptide solutions

3.1 Introduction

With the rise in interest in peptide therapeutics, there is now more than ever a salient need for computational tools that can provide property predictions, which are essential in the screening of potential APIs and the design of their manufacturing processes. The large number of possible peptide molecules that may have a potential application in the pharmaceutical industry renders the task of modelling their phase-equilibrium properties near impossible, without the availability of reliable and abundant experimental data or transferable parameters. In this context, a group-contribution (GC) approach can offer the needed utility, especially since peptide molecules and their constituent amino-acid residues share a finite set of functional groups. Moreover, amino acids and peptides exhibit complex inter- and intramolecular interactions, both charged and uncharged, that give rise to their properties and make them an interesting challenge for modelling approaches. The SAFT-γ Mie EoS provides an ideal tool for predicting the phase-equilibrium of amino-acid and peptide solutions, as it is a group-contribution equation of state

(EoS) [51, 52, 53], in which molecules are treated as heteronuclear chains of fused spherical segments that interact via the Mie potential, and which can account for hydrogen bonding and Coulombic interactions. It can also effectively account for intramolecular interactions with minimal modification to the model.

The SAFT-γ Mie GC approach has been used to model and study a wide variety of systems across many fields of applications. Sadeqzadeh et al [64] and Hutacharoen et al. [56] used this approach to study the phase equilibria of aqueous mixtures. Silva et al. [65] used the approach to study the phase-equilibrium behaviour of biodiesel systems. Papaioannou et al. [66] used SAFT-γ Mie to study fluids relevant to the oil and gas industry. Sabattie et al. [67] used SAFT-γ Mie to study oligomer-polymer mixtures. Samsatli et al. [68] used the model to describe the properties of hydrogen in the context of transport in pipelines. Bui et al. [69] used SAFT-γ Mie in a study of the chemisorption of CO_2 in amine solvents. Di Lecce et al. [59, 60] used the approach to model aqueous mixtures of choline geranate (CAGE), which is both an ionic liquid and a deep eutectic solvent. Morgado et al. [70] used the approach to model the fluid-phase equilibria of mixtures containing semifluorinated and perfluoroalkanes. Perdomo et al. [71] used SAFT-γ Mie to model the thermodynamic properties and fluid-phase behaviour of aqueous solutions of linear, branched, and cyclic amines. Wehbe et al. [62] used the approach to model the speciation and phase behaviour of binary and ternary mixtures of formaldehyde, water, and methanol. Bernet et al. used SAFT-γ Mie to model the thermodynamic properties of mixtures of lactones and a wide range of solvents [58] as well as cyclic alcohols, with a focus on applications to cyclohexanol and menthol [72]. Valsecchi et al. [73] used the approach to model the thermodynamic properties of mixtures of water and polyethylene glycol (PEG).

In the context of computer-aided molecular and process design (CAMPD), Gopinath et al. [74, 75] pioneered the use of SAFT-γ Mie to model solvent-based separation systems, Kazepidis et al. [76] and Papadopoulos et al. [77] used it to design post-combustion CO_2 capture based on phase-change solvents, Bowskill et al. [78] used the model for working-fluid design for organic Rankine cycles (ORC), and White and van Kleef et al. [79, 80, 81, 82] used the model in the thermoeconomic design, optimisation, and assessment of ORC with a focus on waste-heat recovery.

In the context of modelling solid–liquid equilibria for pharmaceutical industrial applications, Watson et al. [83] used the SAFT-γ Mie approach in the design of solvent blends for cooling and anti-solvent crystallisation of ibuprofen. Wehbe et al. [57] used the approach to model the solubility of ionizable acidic drugs, such as ibuprofen and ketoprofen. Febra et al. [61] used the approach to predict the solubility of benzoic acid, diphenylamine, and mefenamic acid, which is a nonsteroidal anti-inflammatory drug (NSAID). The current work, a part of which was published in Ref.[84], is a contribution to a large research effort to expand the applicability of the SAFT-γ Mie approach to model the solid–liquid solubility of APIs of relevance to the pharmaceutical industry.

In this chapter, the SAFT-γ Mie GC approach is presented by describing the free-energy terms it employs for describing the dispersion, association, and electrostatic interactions between the groups in molecules. The model is applied to the amino acids and oligopeptides of interest to this work, by describing the molecular groups they comprise and their like and unlike interaction parameters. The process of parameter optimisation is briefly discussed, followed by a presentation of the parameters developed in this work and their performance compared to thermodynamic pure-component and mixture data.

3.2 SAFT-γ Mie

SAFT-γ Mie [51, 52, 53, 54] is a group-contribution (GC) equation of state (EoS) based on the statistical associating fluid theory (SAFT) [35, 36, 37, 85]. In SAFT-γ Mie, molecules are treated as heteronuclear chains of fused spherical segments that are not necessarily of identical size. These segments, representing the functional groups comprising each molecule, interact with each other via Mie potentials of variable attractive and repulsive range, and hydrogen bonding between some of the groups is mediated through the interaction between association sites embedded in the segments.

In SAFT-γ Mie, a molecule is divided into its constituent groups, with the multiplicity of a group of type k in molecule i being denoted by $\nu_{k,i}$. Each group may consist of several (ν_k^*)

identical fused spherical segments. Two segments of type k and l that are at a distance r_{kl} from each other interact through Mie [86] potentials of variable attractive and repulsive ranges, given by

$$\Phi_{kl}^{\mathrm{Mie}}(r_{kl}) = C_{kl}\varepsilon_{kl}\left[\left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{\mathrm{r}}} - \left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{\mathrm{a}}}\right], \tag{3.1}$$

where σ_{kl} is the size parameter (which corresponds to the segment diameter when $k = l$), ε_{kl} is the depth of the potential well (alternatively the dispersion energy parameter), and $\lambda_{kl}^{\mathrm{r}}$ and $\lambda_{kl}^{\mathrm{a}}$ are, respectively, the repulsive and attractive exponents of the interaction between the two segments. The normalisation factor C_{kl} given by

$$C_{kl} = \frac{\lambda_{kl}^{\mathrm{r}}}{\lambda_{kl}^{\mathrm{r}} - \lambda_{kl}^{\mathrm{a}}}\left(\frac{\lambda_{kl}^{\mathrm{r}}}{\lambda_{kl}^{\mathrm{a}}}\right)^{\lambda_{kl}^{\mathrm{a}}/(\lambda_{kl}^{\mathrm{r}} - \lambda_{kl}^{\mathrm{a}})} \tag{3.2}$$

ensures that the minimum energy of interaction is $-\varepsilon_{kl}$.

The association interaction between a site a on on segment k and a site b on segment l is mediated by a square-well potential given by

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

where $r_{kl,ab}$ is the centre-to-centre distance between sites a and b (respectively positioned at $r_{kk,aa}^{d}$ and $r_{ll,bb}^{d}$ off-centre on their respective segments), $-\varepsilon_{kl,ab}^{\mathrm{HB}}$ is the association energy, and $r_{kl,ab}^{\mathrm{c}}$ is the cut-off distance of the square-well potential between sites a and b.

In the case of systems that contain charged groups, the Helmholtz free energy is extended following an unrestricted primitive model to account for the Coulombic interactions of the charged groups [87, 88]. Each ionic group is assigned a central point charge q and a parameter $\sigma_{kk}^{\mathrm{Born}}$ that represents the Born cavity diameter. The relative static permittivity ϵ_{r} of the dielectric medium proposed by Shreckenberg et al. [89] is used to replace the explicit representation of the polarity of the solvent.

The equation of state is obtained in terms of the Helmholtz free energy, which is written

as a sum of contributions [51, 52, 87]:

$$A = A^{\text{ideal}} + A^{\text{mono}} + A^{\text{chain}} + A^{\text{assoc}} + A^{\text{ion}} + A^{\text{Born}}, \tag{3.4}$$

where: A^{ideal} corresponds to the free energy of an ideal gas mixture; A^{mono} accounts for the interactions among monomer segments, described using Mie potentials; A^{chain} accounts for the free energy due to chain formation; A^{assoc} accounts for association mediated through short-range directional forces; A^{ion} accounts for the Coulombic ion–ion interactions; and A^{Born} accounts for ion– solvent electrostatic interactions. Only the first four terms are used when modeling neutral systems. The last two terms are included when charged groups are present in the system.

3.2.1 The Helmholtz free energy for neutral systems

The thermodynamic properties of neutral systems (i.e. molecular systems that contain no charges) are derived from the Helmholtz free-energy expression in Equation 3.4 by excluding the last two terms, namely the ion and Born terms. The first four terms account for the short-range repulsion, long-range dispersion, and association that account for the physico-chemical behaviour of said systems.

3.2.1.1 Ideal term

The Helmholtz free energy of an ideal mixture, which describes a system of non-interacting point particles, is given by [90]

$$\frac{A^{\text{ideal}}}{Nk_{\text{B}}T} = \left(\sum_{i=1}^{N_{\text{C}}} x_i \ln(\rho_i \Lambda_i^3)\right) - 1 \ , \tag{3.5}$$

where N is the total number of molecules in the system, k_{B} is the Boltzmann constant, T is the absolute temperature, N_{C} is the total number of components in the mixture, x_i is the mole fraction of component i, $\rho_i = N_i/V$ is the number density (V is the total volume of the system and N_i is the number of molecules of component i), and Λ_i^3 is the thermal de Broglie volume,

which implicitly accounts not only for the translational contribution to the kinetic energy but also those from rotations and vibrations of the molecules.

3.2.1.2 Monomer term

The monomer term accounts for repulsion and dispersion interactions between monomer segments and is expressed using a Barker–Henderson [91, 92] high-temperature perturbation expansion up to third order [93], i.e.,

$$\frac{A^{\mathrm{mono}}}{Nk_{\mathrm{B}}T} = \frac{A^{\mathrm{HS}}}{Nk_{\mathrm{B}}T} + \frac{A_1}{Nk_{\mathrm{B}}T} + \frac{A_2}{Nk_{\mathrm{B}}T} + \frac{A_3}{Nk_{\mathrm{B}}T} \ , \tag{3.6}$$

where A^{HS} is the hard-sphere reference free energy with temperature-dependent diameter d_{kk}, given by

$$\frac{A^{\mathrm{HS}}}{Nk_{\mathrm{B}}T} = \left(\sum_{i=1}^{N_{\mathrm{C}}} x_i \sum_{k=1}^{N_{\mathrm{G}}} \nu_{k,i} \nu_k^* S_k \right) a^{\mathrm{HS}}, \tag{3.7}$$

where N_{G} is the total number of groups in the mixture, $\nu_{k,i}$ is the multiplicity of group k in molecule i, ν_k^* is the number of identical spherical segments making up the group k and S_k is its shape factor which describes the contribution of group k to the overall Helmholtz free energy of the molecule in terms of a non-integer number of segments, and a^{HS} [94, 95] is the contribution to the hard-sphere free-energy per segment and is a function of the number density of the segments in the fluid. A_1, A_2, and A_3 are the first, second, and third-order terms of the expansion.

3.2.1.3 Chain term

The chain term corresponds to the change in free energy due to the connectivity of monomer segments forming the molecules of the system. This term is formulated using the TPT1 expression of Wertheim [96, 35] as

$$\frac{A^{\text{chain}}}{Nk_{\text{B}}T} = -\sum_{i=1}^{N_{\text{C}}} x_i \left(\sum_{k=1}^{N_{\text{G}}} \nu_{k,i} \nu_k^* S_k - 1 \right) \ln g_{ii}^{\text{Mie}}(\overline{\sigma}_{ii}; \zeta_x) \ , \tag{3.8}$$

where $g_{ii}^{\text{Mie}}(\overline{\sigma}_{ii}; \zeta_x)$ is the radial distribution function, calculated using a second-order expansion:

$$g_{ii}^{\text{Mie}}(\overline{\sigma}_{ii}; \zeta_x) = g_d^{\text{HS}}(\overline{\sigma}_{ii}; \zeta_x) exp \left[\beta \overline{\varepsilon}_{ii} g_1(\overline{\sigma}_{ii}) / g_d^{\text{HS}}(\overline{\sigma}_{ii}; \zeta_x) + (\beta \overline{\varepsilon}_{ii})^2 g_2(\overline{\sigma}_{ii} / g_d^{\text{HS}}(\overline{\sigma}_{ii}; \zeta_x) \right] , \tag{3.9}$$

evaluated at the average molecular segment diameter $\overline{\sigma}_{ii}$ of component i, in a hypothetical fluid of packing fraction ζ_x. The radial distribution function of a system of hard spheres g_d^{HS} is evaluated using the expression of Boublík [97]. g_1 and g_2, which appear in the first and second-order terms of the expansion, respectively, are evaluated at the contact distance $\overline{\sigma}_{ii} = \overline{d}_{ii}$. A more detailed description of this term can be found in the original SAFT-γ Mie paper [51]; note, however, that there is a typographical error (misplaced bracket) in the expression of the chain term (equation (46) in Ref. [51]), which is presented correctly in equation (3.8) here.

3.2.1.4 Association term

The association term accounts for the contribution to the free energy due to the association between molecules via short-range directional interactions and is expressed, based on the TPT1 perturbation theory of Wertheim [96], as

$$\frac{A^{\text{assoc}}}{Nk_{\text{B}}T} = \sum_{i=1}^{N_{\text{C}}} x_i \sum_{k=1}^{N_{\text{G}}} \nu_{k,i} \sum_{a=1}^{N_{\text{ST},k}} n_{k,a} \left(\ln X_{i,k,a} + \frac{1 - X_{i,k,a}}{2} \right) \ , \tag{3.10}$$

where $N_{\text{ST},k}$ is the total number of site types, $n_{k,a}$ the number of association sites of type a for group k, and $X_{i,k,a}$ the fraction of molecules of component i that are not bonded at site a on group k.

3.2.2 Electrostatic contributions to the Helmholtz free energy

In the case of mixtures containing ionic species, the electrostatic interactions between ions and those between ions and neutral solvent molecules are accounted for by the inclusion of the ion

and Born terms in the expression of the Helmholtz free energy of the mixture (Equation (3.4)).

3.2.2.1 Mapping the charge of an electrolytic chain molecule

In both the MSA [98, 99] and Born [100] approaches (as well as in the seminal Debye-Hückel model [19]), the ionic particles in the underlying molecular model are assumed to be spherical. The assumption of spherical charged particles in incorporating the ion and Born contributions to the Helmholtz free energy is commonly used in electrolyte equations of state, such as the eSAFT-VR Mie EoS [101] (in which the Debye–Hückel [19] term is used instead of the MSA), the electrolyte cubic plus association (eCPA) [102] (in which a Soave–Redlich–Kwong (SRK) [103] residual term, an association term [85], a Debye-Hückle term [19], and a Born term [100] are combined), and the ePC-SAFT EoS [104]. In the case of non-spherical ionic species, a decision must be made on how to reconcile the molecular (non-spherical) model of a charged species with the classical ionic expressions that assume a spherical charged particle.

Here, two possible mappings [87] for reconciling the heteronuclear SAFT-γ Mie molecular model with expressions accounting for spherical ionic interactions are discussed. These two routes are presented in Figure 3.1. As an example, consider an electrolyte containing a chain-like cation (component C_1), a spherical anion (C_2), and a solvent (C_3). The cation comprises a charged group k_1 formed by two identical segments $\nu^*_{k_1} = 2$, each of diameter $\sigma_{k_1k_1}$ and Born (solvated) diameter $\sigma^{\text{Born}}_{k_1k_1}$, which contribute to the overall free energy with a corresponding shape factor S_{k_1}. The group has a positive (central) charge Z_{k_1}. The rest of the cationic species is formed by a heteronuclear chain comprising three groups of type k_2 and one of type k_3. For simplicity, the anion is considered to be spherical here, but the expressions provided below are equally applicable to the case of chain-like anions. The neutral solvent molecules are represented with the standard SAFT representation of water in the figure.

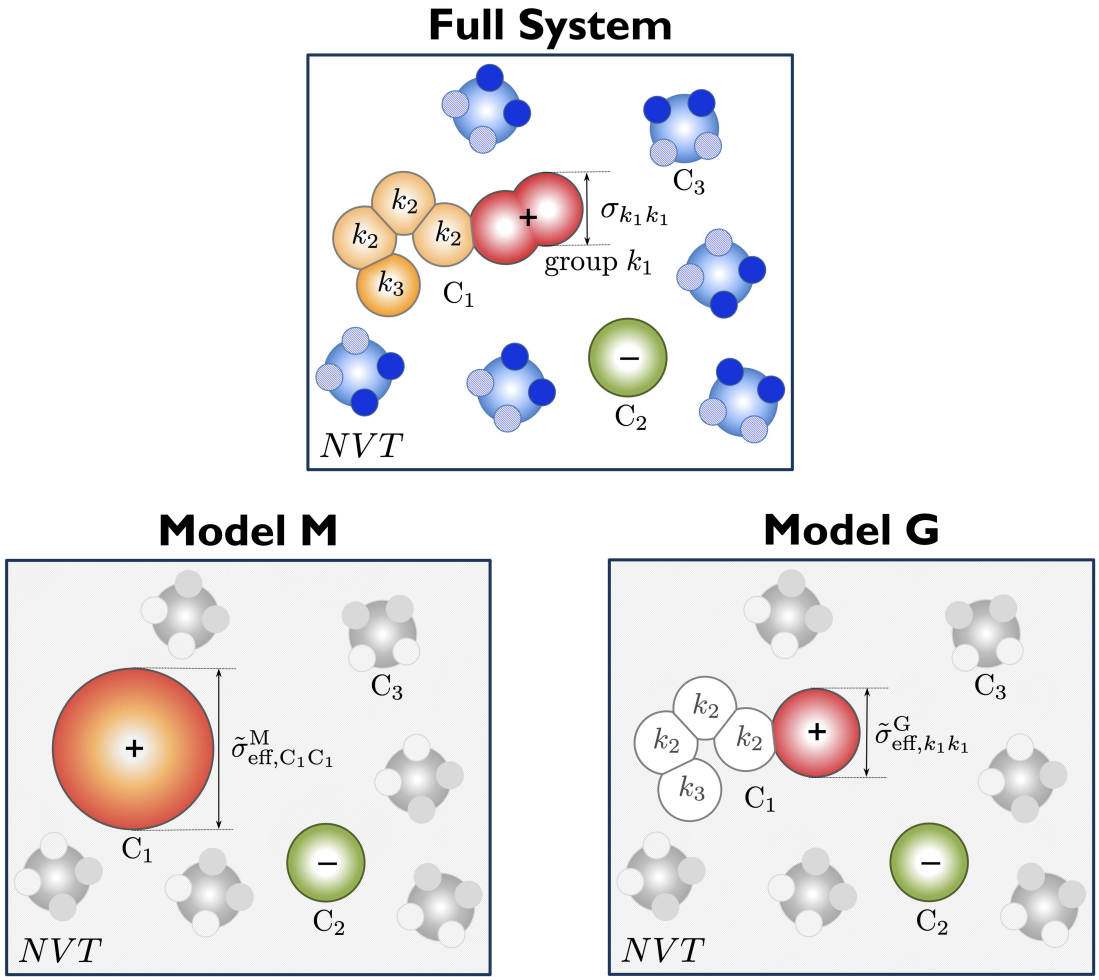


Figure 3.1: A cartoon representation of the models considered within the SAFT-γ Mie approach in the current work. The upper panel depicts an electrolyte comprising a chain-like cation (C_1) and a spherical anion (C_2) in water (C_3). The cation is modelled as a heteronuclear chain of groups including a charged group comprising two identical segments. The lower panels depict the effective models proposed for the evaluation of the ion and Born contributions of the Helmholtz free energy. Model M refers to “molecular" and Model G to “group".

3.2.2.1.1 Molecular Model (M)

In Model M (M for molecular), the entire ionic chain molecule i is mapped onto a single sphere of effective ionic diameter given as

$$\tilde{\sigma}_{\text{eff},\,ii}^{\text{MSA, M}} = \left(\sum_{k=1}^{N_{\text{G}}} \nu_{k,i} \nu_k^* S_k \sigma_{kk}^3 \right)^{1/3} , \tag{3.11}$$

which is defined to maintain the original molecular volume while mapping the chain onto a sphere obtained by the summation over all groups in the molecule (charged and neutral). An effective Born diameter, given by

$$\tilde{\sigma}_{\text{eff},\,ii}^{\text{Born, M}} = \left(\sum_{k=1;Z_k\neq 0}^{N_{\text{G}}} \nu_{k,i} \nu_k^* S_k (\sigma_{kk}^{\text{Born}})^3 + \sum_{k=1;Z_k=0}^{N_{\text{G}}} \nu_{k,i} \nu_k^* S_k \sigma_{kk}^3 \right)^{1/3} \tag{3.12}$$

is also defined. Thus, the two size parameters needed to describe a charged system are the group's Mie diameter σ_{kk} and an assigned Born diameter $\sigma_{kk}^{\text{Born}}$. Furthermore, the effective spherical ion carries a central point charge corresponding to the net charge of the original molecule; i.e.,

$$Z_{\text{eff},i}^{\text{M}} = \sum_{k=1}^{N_G} \nu_{k,i} Z_k \, , \tag{3.13}$$

where Z_k is the charge of group k, so that the overall molecular charge density of the ion remains unchanged through this mapping. Note that in the case in which the net charge of i is zero ($Z_{\text{eff},ii}^{\text{M}} = 0$), as is the case of zwitterions, the ion and Born terms A^{ion} and A^{Born} do not contribute to the free energy of the mixture (they are zero).

3.2.2.1.2 Group Model (G)

In Model G (G for group), only the charged group is mapped to a sphere of effective ionic and Born diameters $\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}$ and $\tilde{\sigma}_{\text{eff},kk}^{\text{Born,G}}$, given by

$$\tilde{\sigma}_{\text{eff},\,kk}^{\text{i,G}} = \left(\nu_k^* S_k \left(\sigma_{kk}^{\text{i}} \right)^3 \right)^{1/3} ; \text{i=MSA,Born} \tag{3.14}$$

such that only the charged group is mapped onto a spherical group of equivalent volume. The charge Z_k of the group remains unchanged so that the charge density of the group is preserved. The rest of the groups in the ion (e.g., groups k_2 and k_3 in the figure) remain unchanged and do not contribute to the ionic or Born terms, as they do not carry a charge. In the current thesis, Model G is adopted to model the solubility of amino acids, on the assumption that it represents a more physically accurate picture of the ionic molecule within a group-contribution framework, especially for the treatment of zwitterions. In Chapter 5, this assumption is tested against molecular simulations. The following presentation corresponds to Model G.

3.2.2.2 Ion term

The ion term A^{Ion}, which accounts for the contribution of Coulombic interactions between charged species (ion-ion) to the Helmholtz free energy is expressed following the classic expression of Blum for the solution of the mean spherical approximation (MSA) in an primitive unrestricted electrolyte model [98, 99] and is given by

$$\frac{A^{\text{ion}}}{Nk_{\text{B}}T} = \frac{U^{\text{MSA}}}{Nk_{\text{B}}T} + \frac{\Gamma^3}{3\pi\rho}, \tag{3.15}$$

where U^{MSA} is the MSA contribution to the internal energy and Γ is the screening length of the electrostatic forces. Implementing Model G in the SAFT-γ Mie formulation, U^{MSA} is given by

$$\frac{U^{\text{MSA}}}{Nk_{\text{B}}T} = -\frac{e^2}{(4\pi\epsilon_0)\,\rho k_{\text{B}}TD}\left[\Gamma\rho\sum_{i=1;Z_i\neq 0}^{N_{\text{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\text{G}}}\frac{x_i\nu_{k,i}Z_k^2}{\left(1+\Gamma\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}\right)} + \frac{\pi}{2\Delta}\Omega P_n^2\right], \tag{3.16}$$

where $e = 1.602\times10^{-19}$ C is the elementary charge, $\epsilon_0 = 8.854\times10^{-12}\text{C}^2\text{J}^{-1}\text{m}^{-1}$ is the static permittivity in vacuum, and D the relative static permittivity. Z_k is the charge of group k, while Z_i is the net charge of compound i, given by

$$Z_i = \sum_{k=1}^{N_G}\nu_{k,i}Z_k\,. \tag{3.17}$$

The constraint "$Z_i \neq 0$" indicates that the outer sum over components includes only ions with a net charge (i.e., it excludes uncharged molecules and zwitterions) and, similarly, the constraint "$Z_k \neq 0$" denotes that the sum over groups includes only those that are charged. Δ, given by

$$\Delta = 1 - \frac{\pi\rho}{6} \sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}} \sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}} x_i \nu_{k,i} \left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^3 , \tag{3.18}$$

describes the packing fraction of the ions as a function of the effective ionic diameter $\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}$ which is given by equation (3.14).

$$P_n = \frac{\rho}{\Omega} \sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}} \sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}} \frac{x_i \nu_{k,i} \tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}} Z_k}{1 + \Gamma \tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}} , \tag{3.19}$$

and

$$\Omega = 1 + \frac{\pi\rho}{2\Delta} \sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}} \sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}} \frac{x_i \nu_{k,i} \left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^3}{1 + \Gamma \tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}} , \tag{3.20}$$

are coupling parameters that are functions of the ionic parameters and the screening length of the ions. P_n couples the charges of the ions and Ω relates to the packing fractions of the ions. The screening length of the ions Γ is a function of the relative static permittivity, D, and the effective charge $Q_k(\Gamma)$ of the ions, leading to an implicit formulation through the electric charge of the individual ionic groups Q_k:

$$\Gamma^2 = \frac{\pi e^2 \rho}{(4\pi\epsilon_0)\, D k_{\mathrm{B}} T} \sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}} \sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}} x_i \nu_{k,i} Q_k^2, \tag{3.21}$$

where the effective charge is related to Q_k and the P_n coupling parameter:

$$Q_k = \frac{Z_k - \left(\tilde{\sigma}_{\mathrm{eff},\,kk}^{\mathrm{MSA,G}}\right)^2 P_n \left(\pi/\left(2\Delta\right)\right)}{1 + \Gamma \tilde{\sigma}_{\mathrm{eff},\,kk}^{\mathrm{MSA,G}}}. \tag{3.22}$$

3.2.2.3 Born term

The Born term [100] describes the contribution to the free energy arising from ion-solvent interactions. More specifically, this term accounts for the free energy associated with creating a cavity, called a Born cavity, to allow for the insertion of an ion from vaccuum into the solvent, at infinite dilution. The solvent is treated as a continuous dielectric medium, and the insertion of each ion is assumed to have no impact on the solvent properties near the other ions. The Born term, accounting for the implementation of Model G, is given as

$$\frac{A^{\text{Born}}}{Nk_{\text{B}}T} = -\frac{e^2}{4\pi\epsilon_0 k_{\text{B}}T}\left(1-\frac{1}{D}\right)\sum_{i=1;Z_i\neq 0}^{N_{\text{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\text{G}}}\frac{x_i\nu_{k,i}Z_k^2}{\tilde{\sigma}^{\text{Born,G}}_{\text{eff},\,kk}}\,, \tag{3.23}$$

where the effective Born diameter, $\tilde{\sigma}^{\text{Born}}_{\text{eff},\,kk}$, is calculated using equation (3.14) where the Born diameter of the spherical cavity created by each ionic group k in the dielectric medium is obtained independently of any other ionic group.

3.2.2.4 The relative permittivity

In the SAFT-γ Mie model, we follow the work of Shrekenberg et al. [105] to obtain the relative static permittivity, given by [89]

$$D = 1 + \rho_{\text{solv}}d, \tag{3.24}$$

where $\rho_{\text{solv}} = N_{\text{solv}}/V$ is the solvent number density of the system, with N_{solv} the number of molecules of the solvent. In the model proposed here only a species not containing charged groups (regardless of the net charge) is considered to be a solvent, meaning that any zwitterionic molecule is not a solvent. The parameter d is calculated as

$$d = \sum_{\substack{i=1;\,Z_i=0\\ i,\,j\,\text{are not zwitterions}}}^{N_{\text{C}}}\sum_{j=1;\,Z_j=0}^{N_{\text{C}}} x'_i x'_j d_{ij}\,, \tag{3.25}$$

where x'_i and x'_j are the salt-free mole fractions of solvents i and j. The summation is over solvent species only thanks to the constraints $Z_i = 0$ and $Z_j = 0$, with i and j not being a

zwitterion. The unlike d_{ij} term is obtained as:

$$d_{ij} = \frac{d_{ii} + d_{jj}}{2} \; ; i, j \text{ solvent} = 1, ..., N_{\mathrm{C}} \tag{3.26}$$

where d_{ii} is the temperature-dependent contribution to D from solvent i and is given by

$$d_{ii} = d_{i,V} \left(\frac{d_{i,T}}{T} - 1 \right) \; ; i \text{ solvent} = 1, ..., N_{\mathrm{C}}; \; , \tag{3.27}$$

where $d_{i,V}$ and $d_{i,T}$ are component-specific adjustable parameters. These have been provided for several solvents in previous work [89]. As can be gleaned from equations (3.24)–(3.26), the value of the dielectric constant of the medium is not affected by the choice of charge mapping (Model G or M).

3.3 Model development

The Helmholtz free energy expressions, presented in Sections 3.2.1 and 3.2.2, can only be evaluated once the set of like and unlike interaction parameters for the groups comprising the system have been obtained. A mixture with a composition x_i of component i is subdivided into its constituent groups, with $\nu_{k,i}$ counting the number of groups of type k in the molecule. Note that in cases where a functional group, or moiety, can be represented using different combinations of groups, the group containing the most atoms should be used [55]. For example, the CH_2OH group is used to represent primary alcohols instead of the combination of groups CH_2 and OH.

A group k is characterised by its shape factor (S_k), which determines the extent to which it contributes to the overall molecular properties, and the number of segments comprising it (ν_k^*). The dispersion interaction between two groups k and l - whether like ($k = l$) or unlike ($k \neq l$) - is fully determined by the following set of parameters: the size parameter (σ_{kl}); the dispersion energy parameter (ε_{kl}); the repulsive Mie exponent ($\lambda_{kl}^{\mathrm{r}}$); and the attractive Mie exponent ($\lambda_{kl}^{\mathrm{a}}$). Note that $\lambda_{kl}^{\mathrm{a}}$ is set equal to six to represent the London dispersion forces [106,

107].

For groups that are involved in association interactions, the following additional parameters are needed: the number of association sites, of different types, embedded in group k ($N_{\mathrm{ST},k}$); the number of sites of each given type ($n_{k,a}$, $n_{k,b}$,...,$n_{k,N_{\mathrm{ST},k}}$); the association energy between site a on group k and site b on group l ($\varepsilon^{\mathrm{HB}}_{kl,ab}$); and the association bonding-volume parameter ($K_{kl,ab}$). Note that the number and types of sites present in a group are decided on based on chemical intuition.

For charged systems, the charge of group k (Z_k) and its Born diameter of ($\sigma^{\mathrm{Born}}_{kk}$) are required. Additionally, the relative static permittivity of the dielectric medium, as presented by Shreckenberg [89] (ϵ_{r}) is used to account for the polarity of the solvent.

Once the system's group representation is decided on, the like and unlike parameters are obtained by minimising the objective function f_{obj}, which represents the weighted average error in model calculations relative to the available experimental data, and is calculated as

$$\min_{\Theta} f_{\mathrm{obj}} = \frac{1}{N^{\mathrm{exp}}} \sum_{s=1}^{N^{\mathrm{S}}} \sum_{p=1}^{N^{\mathrm{P}}_{s}} \sum_{i=1}^{N^{\mathrm{D}}_{s,p}} w_{s,p,i} \left(\frac{X^{\mathrm{exp}}_{s,p,i} - X^{\mathrm{calc}}_{s,p,i}(\Theta)}{X^{\mathrm{exp}}_{s,p,i}} \right)^2, \tag{3.28}$$

where Θ is the vector of model parameters, N^{exp} is the total number of experimental data points used in the parameter optimisation, N^{S} is the number of systems (pure compounds and mixtures), $N^{\mathrm{D}}_{s,p}$ is the number of property types for system s, $N^{\mathrm{D}}_{s,p}$ is the number of data points of property p of system s, and $w_{s,p,i}$ is a weight factor assigned to data point i of property p of system s. $X^{\mathrm{exp}}_{s,p,i}$ is the i^{th} experimental data point of property p of system s and $X^{\mathrm{calc}}_{s,p,i}(\Theta)$ is the i^{th} is the corresponding point calculated with the set of SAFT-γ Mie parameters Θ. The parameter optimisations in this work are conducted using Siemens' proprietary software, gProperties.

The performance of the model is evaluated by calculating the percent absolute average deviation (%AAD) of a property p of system s, given by

$$\%\mathrm{AAD}_{s,p} = \frac{1}{N^{\mathrm{D}}_{s,p}} \sum_{i=1}^{N^{\mathrm{D}}_{s,p}} \left| \frac{X^{\mathrm{exp}}_{s,p,i} - X^{\mathrm{calc}}_{s,p,i}}{X^{\mathrm{exp}}_{s,p,i}} \right| \times 100, \tag{3.29}$$

and the absolute average deviation (AAD), given by

$$\mathrm{AAD}_{s,p} = \frac{1}{N^{\mathrm{D}}_{s,p}} \sum_{i=1}^{N^{\mathrm{D}}_{s,p}} \left| X^{\mathrm{exp}}_{s,p,i} - X^{\mathrm{calc}}_{s,p,i} \right| . \tag{3.30}$$

3.4 SAFT-γ Mie models for amino acids and oligopeptides

In order to model the solid–liquid solubility of amino acids and oligopeptides using the SAFT-γ Mie GC approach, it is necessary to identify the groups that comprise the systems to be modeled and obtain their interaction parameters.

To model the solid–liquid solubility of amino acids and oligopeptides at the isoelectric point (pI), no charged groups are considered; i.e., not only is the amino acid (or peptide) neutral (as corresponds to a zwitterion), but each of the SAFT-γ Mie groups used to model the systems of interest is also a neutral (standard) group. For example, glycine is modelled as 1 $\times$ NH_2, 1 $\times$ CH_2, and 1 $\times$ COOH groups, alanine is modelled as 1 $\times$ NH_2, 1 $\times$ CH, 1 $\times$ CH_3, and 1 $\times$ COOH groups, valine is modelled as 1 $\times$ NH_2, 2 $\times$ CH, 2 $\times$ CH_3, and 1 $\times$ COOH groups, while serine contains 1 $\times$ NH_2, 1 $\times$ CH, 1 $\times$ CH_2OH, and 1 $\times$ COOH groups. A representation of these models is given in Figure 3.2.

In Section 4.3, the solid–liquid solubility of the amino acids glycine, alanine, serine, and valine in pure water, alcohols, and water-alcohol mixtures as well as that of di- and tri-peptides of glycine, alanine, leucine, and serine, are predicted. Since the corresponding solubility measurements are conducted at the pI, only the neutral models are used without treating the speciation of the zwitterions or any of the solvents. The similarity in structure of the amino acids and peptides, and the group-contribution nature of the SAFT-γ Mie approach, mean that a small number of groups is sufficient to model the properties of all the molecules of interest here. Specifically, the uncharged H_2O, COOH, NH_2, CH_3, CH_2, CH, CH_2OH, CHOH, and CONH

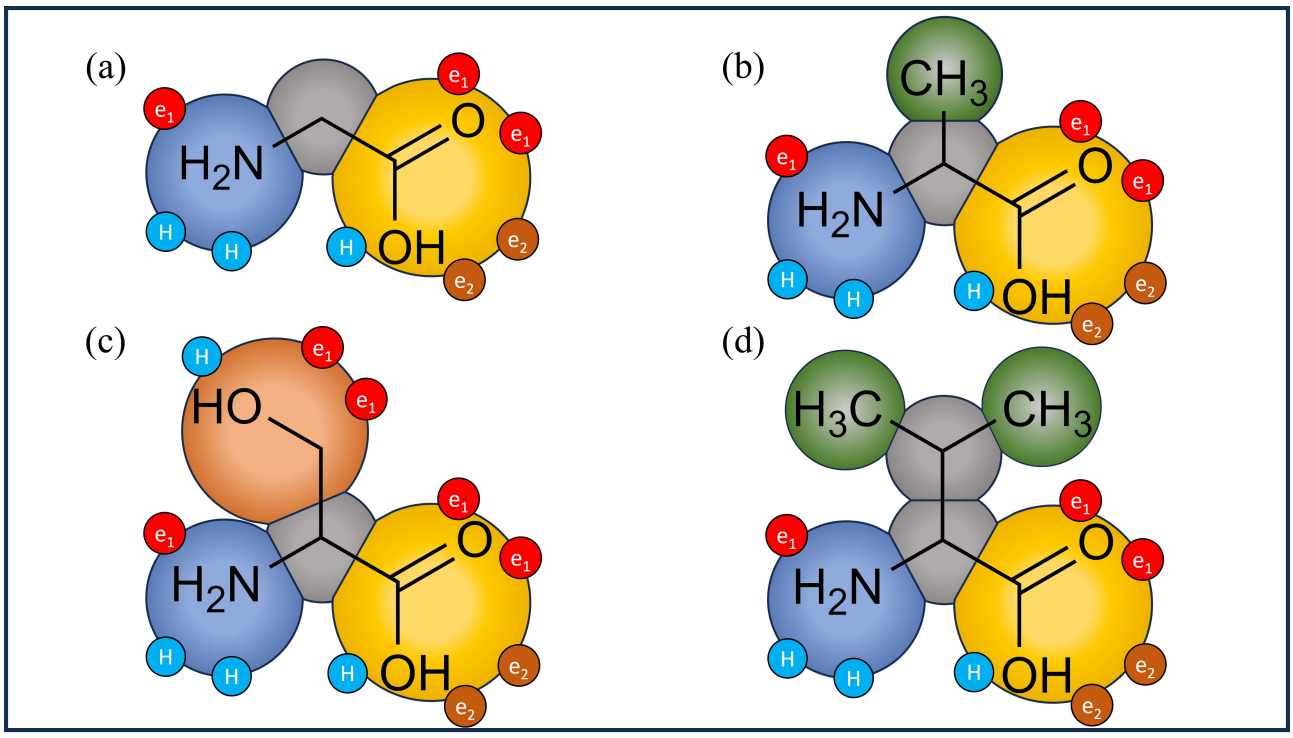


Figure 3.2: SAFT-γ Mie representation of (a) glycine, (b) alanine, (c) serine, and (d) valine. A heteronuclear model with fused spherical segments is implemented in which short-range association sites are represented with smaller light-blue (sites of type H), red (type e_1), and brown (type e_2) circles.

groups are used.

To model the solid–liquid solubility of amino acids and under changing pH, charged groups are considered. For example, zwitterionic glycine is modelled as 1 × NH_3^+, 1 × CH_2, and 1 × COO^- groups, cationic glycine is modelled as 1 × NH_3^+, 1 × CH_2, and 1 × COOH groups, and anionic glycine is modelled as 1 × NH_2, 1 × CH_2, and 1 × COO^- groups. A representation of the models used to SAFT-γ Mie models used to describe the speciation of glycine is given in Figure 3.3. Similarly, zwitterionic alanine is modelled as 1 × NH_3^+, 1 × CH, 1 × CH_3, and 1 × COO^- groups, cationic alanine is modelled as 1 × NH_3^+, 1 × CH, 1 × CH_3, and 1 × COOH groups, and anionic alanine is modelled as 1 × NH_2, 1 × CH, 1 × CH_3, and 1 × COO^- groups.

In Section 4.4.3, the aqueous solubility of glycine and alanine is predicted under changing pH, in which case the speciation of the zwitterion and water is essential. There, both the uncharged and charged groups are considered to model the phase and chemical equilibria of the system of interest. Thus, the groups needed to model both of these systems are: NH_2, NH_3^+,

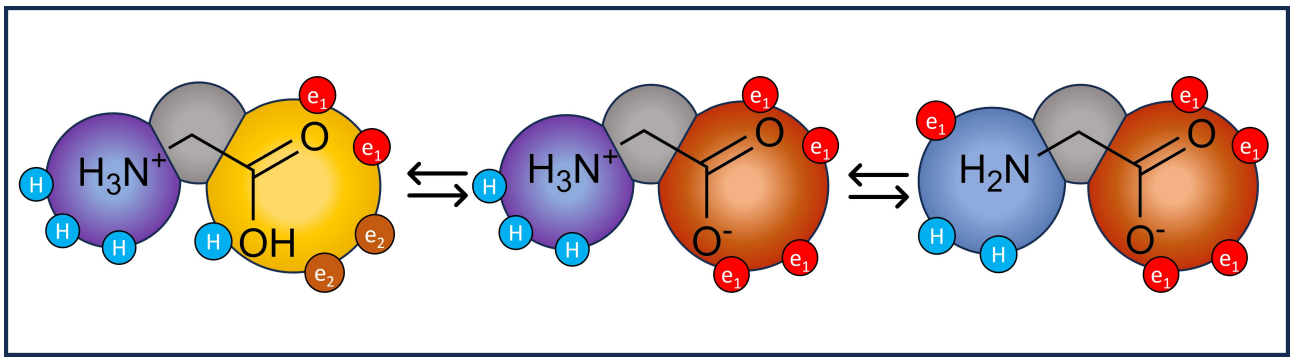


Figure 3.3: SAFT-γ Mie representation the glycine cation, zwitterion and anion as modelled in the current work.A heteronuclear model with fused spherical segments is implemented in which short-range association sites are represented with smaller light-blue (sites of type H) and red (type e_1) circles. The reactions also involve water, the hydronium and hydroxide ions, as well as the acidic and basic counter ions, which are not shown here.

COOH, COO^-, CH_3, CH_2, CH, H_2O, H_3O^+, OH^-, Na^+, and Cl^-.

The parameter sub-matrix in Figure 3.4 contains all the groups needed for the solid–liquid solubility predictions in this work. Most of these groups and their like and unlike interactions have been characterised in previous works, with one new group and 11 unlike interactions needing to be determined as part of the current thesis.

The interaction parameters of the groups, shown in Tables 3.1, 3.2, and 3.3, are optimised by adjustment using experimental thermodynamic-property data of pure-component and binary systems that contain the groups of interest.

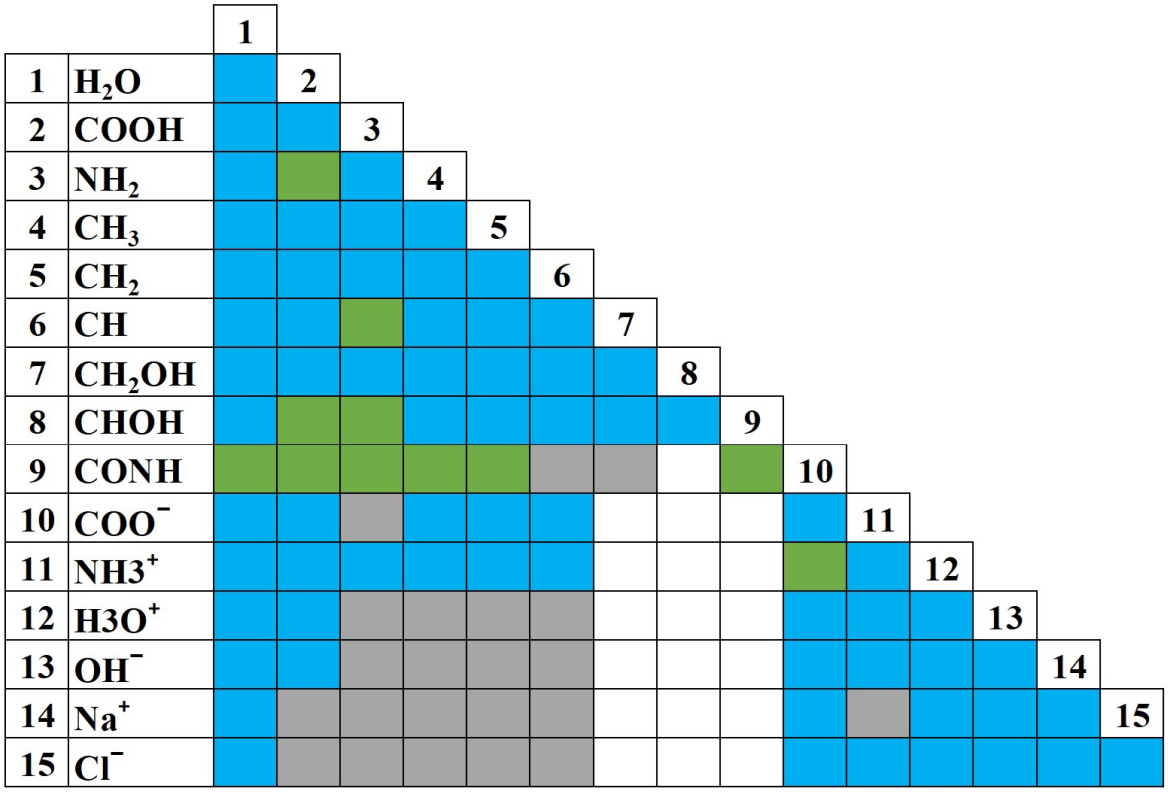


Figure 3.4: SAFT-γ Mie GC group interaction sub-matrix containing groups required to model systems in the current work. Blue cells indicate interactions previously published ([55, 57]), green cells indicate interactions developed in our current work, grey cells indicate interactions that are obtained using combining rules [51], and white cells indicate interactions that are not needed in the current work.

Table 3.1: SAFT-γ Mie like interaction parameters of the groups considered in our current work (excluding association). The attractive range of the Mie potential is $\lambda^{a}_{kk} = 6$ for all groups here. The asterisk indicates that the CONH group has been characterized in the current work.

Group k	ν^{*}_{k}	S_k	σ_{kk}/Å	$\sigma^{\mathrm{Born}}_{kk}$Å	λ^{r}_{kk}	$(\varepsilon_{kk}/k_B)/K$	$N_{ST,k}$	$n_{k,H}$	n_{k,e_1}	n_{k,e_2}	Ref.
H_2O	1	1.0000	3.0063	—	17.020	266.68	2	2	2	—	[53, 54]
COOH	1	0.55593	4.3331	—	8.0000	405.78	3	1	2	2	[52]
NH_2	1	0.79675	3.2477	—	10.254	284.78	2	2	1	—	[71]
CH_3	1	0.57255	4.0772	—	15.050	256.77	—	—	—	—	[51]
CH_2	1	0.22932	4.8801	—	19.871	473.39	—	—	—	—	[51]
CH	1	0.072100	5.2950	—	8.0000	95.621	—	—	—	—	[52]
CH_2OH	2	0.58538	3.4054	—	22.699	407.22	2	1	2	—	[56]
CHOH	2	0.18963	4.5381	—	18.185	599.66	2	1	2	—	[55]
CONH	2	0.73764	2.9482	—	29.839	156.12	2	1	2	—	*
COO^-	1	0.55593	4.3331	4.6364	8.0000	21.264	1	—	4	—	[60]
NH_3^+	1	0.79675	3.2477	3.4750	10.254	48.300	1	3	—	—	[57]
H_3O^+	1	1.0000	3.0063	3.0063	17.020	68.190	1	3	—	—	[88]
OH^-	1	1.0000	2.4600	3.0063	17.020	170.24	1	—	3	—	[88]
Na^+	1	1.0000	2.3200	3.3600	12.000	31.711	—	—	—	—	[88]
Cl^-	1	1.0000	3.3400	3.8740	12.000	113.77	—	—	—	—	[88]

Table 3.2: Unlike group parameters (excluding association) for use with the SAFT-γ Mie approach. An asterisk indicates that the parameters have been characterised in the current work. CR indicates that combining rules are used [51].

Group k	Group l	$(\varepsilon_{kl}/k_{\mathrm{B}})$	$\lambda^{\mathrm{r}}_{kl}$	Ref.	Group k	Group l	$(\varepsilon_{kl}/k_{\mathrm{B}})$	$\lambda^{\mathrm{r}}_{kl}$	Ref.
H_2O	COOH	289.76	CR	[56]	CH_3	COO^-	255.99	CR	[59, 60]
H_2O	NH_2	358.55	CR	[71]	CH_3	NH_3^+	244.15	CR	[57]
H_2O	CH_3	358.18	100.00	[56]	CH_3	H_3O^+	CR	CR	–
H_2O	CH_2	423.63	100.00	[56]	CH_3	Na^+	CR	CR	–
H_2O	CH	275.75	CR	[108]	CH_3	Cl^-	CR	CR	–
H_2O	CH_2OH	353.37	CR	[56]	CH_3	OH^-	CR	CR	–
H_2O	CHOH	479.16	CR	[55]	CH_3	CONH	430.60	CR	*
H_2O	COO^-	171.61	CR	[59, 60]	CH_2	CH	506.21	CR	[52]
H_2O	NH_3^+	450.21	CR	[57]	CH_2	CH_2OH	423.17	CR	[56]
H_2O	H_3O^+	391.04	CR	[88]	CH_2	CHOH	517.64	CR	[55]
H_2O	OH^-	134.41	CR	[88]	CH_2	COO^-	413.74	CR	[59, 60]
H_2O	CONH	379.59	CR	*	CH_2	NH_3^+	348.39	CR	[57]
H_2O	Na^+	539.68	CR	[88]	CH_2	H_3O^+	CR	CR	–
H_2O	Cl^-	95.406	CR	[88]	CH_2	OH^-	CR	CR	–
COOH	NH_2	285.00	CR	*	CH_2	CONH	315	CR	*
COOH	CH_3	255.99	CR	[52]	CH_2	Na^+	CR	CR	–
COOH	CH_2	413.74	CR	[52]	CH_2	Cl^-	CR	CR	–
COOH	CH	504.99	CR	[52]	CH	CH_2OH	329.22	CR	[108]
COOH	CH_2OH	488.18	CR	[109]	CH	CHOH	0	CR	[55]
COOH	CHOH	1154.3	50.000	*	CH	COO^-	504.99	CR	[57]
COOH	COO^-	405.78	8.0000	[59, 60]	CH	NH_3^+	151.01	CR	[57]
COOH	NH_3^+	388.58	CR	[57]	CH	H_3O^+	CR	CR	–
COOH	H_3O^+	CR	CR	[57]	CH	OH^-	CR	CR	–
COOH	OH^-	CR	CR	[57]	CH	CONH	CR	CR	–
COOH	CONH	670.04	CR	*	CH	Na^+	CR	CR	–
COOH	Na^+	CR	CR	–	CH	Cl^-	CR	CR	–
COOH	Cl^-	CR	CR	–	CH_2OH	CHOH	389.23	CR	[55]
NH_2	CH_3	244.15	CR	[71]	CH_2OH	CONH	CR	CR	–
NH_2	CH_2	348.39	CR	[71]	COO^-	NH_3^+	26.330	CR	[57]
NH_2	CH	278.26	CR	*	COO^-	H_3O^+	27.74	CR	[57]
NH_2	CH_2OH	528.21	52.305	[110]	COO^-	OH^-	44.520	CR	[57]
NH_2	CHOH	415.54	10.643	*	COO^-	Na^+	9.9125	CR	[59, 60]
NH_2	COO^-	CR	CR	–	COO^-	Cl^-	21.265	CR	[59, 60]
NH_2	NH_3^+	284.78	CR	[57]	NH_3^+	H_3O^+	56.958	CR	[57]
NH_2	H_3O^+	CR	CR	–	NH_3^+	OH^-	62.238	CR	[57]
NH_2	OH^-	CR	CR	–	NH_3^+	Na^+	CR	CR	–
NH_2	CONH	150.77	CR	*	NH_3^+	Cl^-	65.257	CR	[57]
NH_2	Na^+	CR	CR	–	H_3O^+	OH^-	66.439	CR	[88]
NH_2	Cl^-	CR	CR	–	H_3O^+	Na^+	37.48	CR	[57]
CH_3	CH_2	350.77	CR	[51]	H_3O^+	Cl^-	70.552	CR	[88]
CH_3	CH	387.48	CR	[52]	OH^-	Na^+	27.898	CR	[88]
CH_3	CH_2OH	333.20	CR	[56]	OH^-	Cl^-	123.21	CR	[57]
CH_3	CHOH	479.38	CR	[55]	Na^+	Cl^-	27.938	CR	[88]

Table 3.3: Group association parameters for use with the SAFT-γ Mie approach. An asterisk indicates that the parameters have been characterized in the current work.

group k	site a	group l	site b	$(\varepsilon^{HB}_{kl,ab}/k_B)$ / K	$K^{HB}_{kl,ab}$ / Å^3	Ref.
H_2O	H	H_2O	e_1	1985.4	101.69	[53, 54]
H_2O	e_1	COOH	H	2567.7	270.09	[56]
H_2O	H	COOH	e_1	1451.8	280.89	[56]
H_2O	H	COOH	e_2	1252.6	150.98	[56]
H_2O	H	NH_2	e_1	1460.0	179.60	[71]
H_2O	e_1	NH_2	H	1988.3	55.824	[71]
H_2O	H	CH_2OH	e_1	2153.2	147.40	[56]
H_2O	e_1	CH_2OH	H	621.68	425.00	[56]
H_2O	H	CHOH	e_1	2140.9	19.478	[55]
H_2O	e_1	CHOH	H	2289.1	63.813	[55]
H_2O	H	COO^-	e_1	802.21	52.555	[59, 60]
H_2O	e_1	NH_3^+	H	2016.6	49.397	[57]
H_2O	e_1	H_3O^+	H	1985.4	101.69	[88]
H_2O	H	OH^-	e_1	1492.0	76.411	[88]
H_2O	H	CONH	e_1	1986.2	236.59	*
H_2O	e_1	CONH	H	3061.5	130.81	*
COOH	H	COOH	H	6427.9	0.80620	[52]
COOH	H	NH_2	e_1	4000.0	100.00	*
COOH	e_1	NH_2	H	1446.6	100.00	*
COOH	e_2	NH_2	H	1220.1	100.00	*
COOH	H	CH_2OH	e_1	3238.4	36.050	[109]
COOH	e_1	CH_2OH	H	1062.1	210.67	[109]
COOH	e_2	CH_2OH	H	997.89	227.07	[109]
COOH	H	CHOH	e_1	4000.0	10.834	*
COOH	e_1	CHOH	H	305.02	1.0000	*
COOH	e_2	CHOH	H	319.87	0.10000	*

group k	site a	group l	site b	$(\varepsilon^{\mathrm{HB}}_{kl,ab}/k_{\mathrm{B}})$ / K	$K^{\mathrm{HB}}_{kl,ab}$ / Å^3	Ref.
COOH	e_1	NH_3^+	H	334.08	13.500	[57]
COOH	e_2	NH_3^+	H	366.54	9990.0	[57]
COOH	e_1	H_3O^+	H	1451.8	280.89	[57]
COOH	e_2	H_3O^+	H	1252.6	150.98	[57]
COOH	H	OH^-	e_1	2036.0	214.16	[57]
COOH	H	CONH	e_1	2487.2	242.73	$*$
COOH	e_1	CONH	H	1723.5	461.80	$*$
COOH	e_2	CONH	H	1723.5	461.80	$*$
NH_2	e_1	NH_2	H	1070.8	95.225	[71]
NH_2	H	CH_2OH	e_1	629.88	346.08	[110]
NH_2	e_1	CH_2OH	H	2403.8	26.192	[110]
NH_2	H	CHOH	e_1	1524.9	103.22	$*$
NH_2	e_1	CHOH	H	1470.3	303.47	$*$
NH_2	H	COO^-	e_1	1220.1	100.00	$*$
NH_2	e_1	NH_3^+	H	1070.8	95.225	$*$
NH_2	e_1	H_3O^+	H	1460.0	179.60	$*$
NH_2	H	OH^-	e_1	1511.4	42.436	$*$
NH_2	e_1	CONH	H	2807.1	122.83	$*$
NH_2	H	CONH	e_1	1687.6	122.83	$*$
CH_2OH	H	CH_2OH	e_1	2097.9	62.309	[56]
CH_2OH	H	CHOH	e_1	2500.0	10.444	[55]
CH_2OH	e_1	CHOH	H	1464.1	591.55	[55]
CH_2OH	H	CONH	e_1	2583.6	101.84	$*$
CH_2OH	e_1	CONH	H	2583.6	101.84	$*$
CHOH	H	CHOH	e_1	2480.6	8.4740	[55]
COO^-	e_1	NH_3^+	H	1767.0	13.500	$*$, [57]
COO^-	e_1	H_3O^+	H	802.21	52.555	[57]
NH_3^+	H	OH^-	e_1	1988.3	55.824	[57]

group k	site a	group l	site b	$(\varepsilon^{\mathrm{HB}}_{kl,ab}/k_{\mathrm{B}})$ / K	$K^{\mathrm{HB}}_{kl,ab}$ / Å^3	Ref.
H_3O^+	H	OH^-	e_1	1492.0	76.411	[88]
CONH	H	CONH	e_1	3181.7	155.31	$*$

The parameter optmisation process introduced in Section 3.3 is followed for each like and unlike interaction and the corresponding SAFT-γ Mie calculation is presented alongside the experimental data used. The NH_2 [71], COOH [52], CHOH [55], and CH [52] groups have been developed in previous works, but the NH_2–COOH, NH_2–CH, COOH–CHOH, and NH_2–CHOH unlike interactions needed to be characterised. Experimental data of simple mixtures containing these groups were used in the parameter estimation with as few as possible amino-acid data, as the goal is to develop a SAFT-γ Mie framework that is predictive for the properties of amino acid and peptide solutions. Additionally, for the modelling of oligopeptide solutions, the peptide (amide) CONH group was characterised with all its like and unlike parameters. Finally, the NH_3^+–COO^- interaction was slightly manually modified when modelling the solubility of amino acids under changing pH.

3.4.1 NH_2−COOH

The NH_2−COOH interaction is key to modeling amino acids and peptides, however, its optimisation represents a major challenge due to the lack of experimental data for mixtures of primary amines (R−NH_2) and alkanoic acids (R−COOH) as a result of the reactive nature of these mixtures. Instead, experimental aqueous-mixture data of amino acids belonging to the glycine homologous series $H_2N-(CH_2)_n-COOH$ with n=1–5 was used. These amino acids were chosen to minimise the influence of groups, other than NH_2 and COOH, on the resulting interaction parameters [58]. Specifically the SLE [111, 30, 112] and VLE [113] of glycine in water, and the liquid density of aqueous solutions [114] of glycine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-aminobutanoic acid, and 6-aminohexanoic acid were used. Calculations using the optimised parameters, illustrated in Figures 3.5 and 3.6, exhibit satisfactory agreement with the available experimental data–given the constraints of the model and the

scarcity of relevant VLE data. Optimisation strategies incorporating other combinations of thermodynamic-property data, but excluding solubility, led to non-physical parameter values and larger deviations between calculated and experimental solubilities. The AAD and %AAD values quantifying the model's agreement with experiments for these systems can be found in Table 3.4

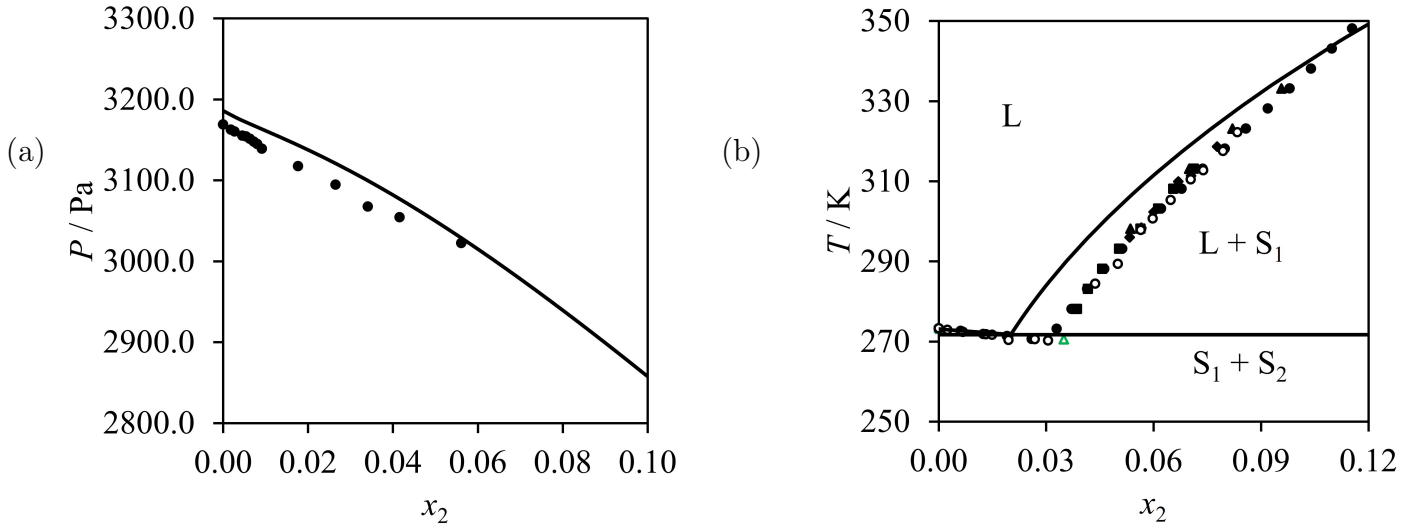


Figure 3.5: SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-COOH unlike interaction compared to experimental data. (a) Bubble pressure of water (1) + glycine (2) at 298 K [113]. The pressure is measured with a reported sensitivity of about 0.3 Pa. (b) Solid–liquid solubility of water (1) + glycine (2) at 1 bar; filled circles [111], squares [115], diamonds [112], and black triangles [30]. Empty symbols denote experimental data not used in the parameter optimisation: empty circles [116]; and green triangle [117].

3.4.2 NH_2–CHOH

The NH_2−CHOH interaction is optimised by adjustment using experimental data of secondary amine + primary amine mixtures. Specifically: VLE data of 2-butanol + 1-butanamine [118, 119]; excess-enthalpy data of 2-propanol + 2-propanamine [120], 2-propanol + 1-butanamine [121], 2-propanol + 2-butanamine [121], and 2-butanol + 1-butanamine [118]; and density data of 2-propanol + 1-propanamine [122], 2-butanol + 1-propanamine [122], 2-butanol + 1-butanamine [123], 2-hexanol + 1-butanamine [124], and 3-hexanol + 1-butanamine [124] are considered. Figures 3.7 and 3.8 illustrate the agreement between SAFT-γ Mie calculations and experimental data for the aforementioned systems; the AAD and %AAD values corresponding

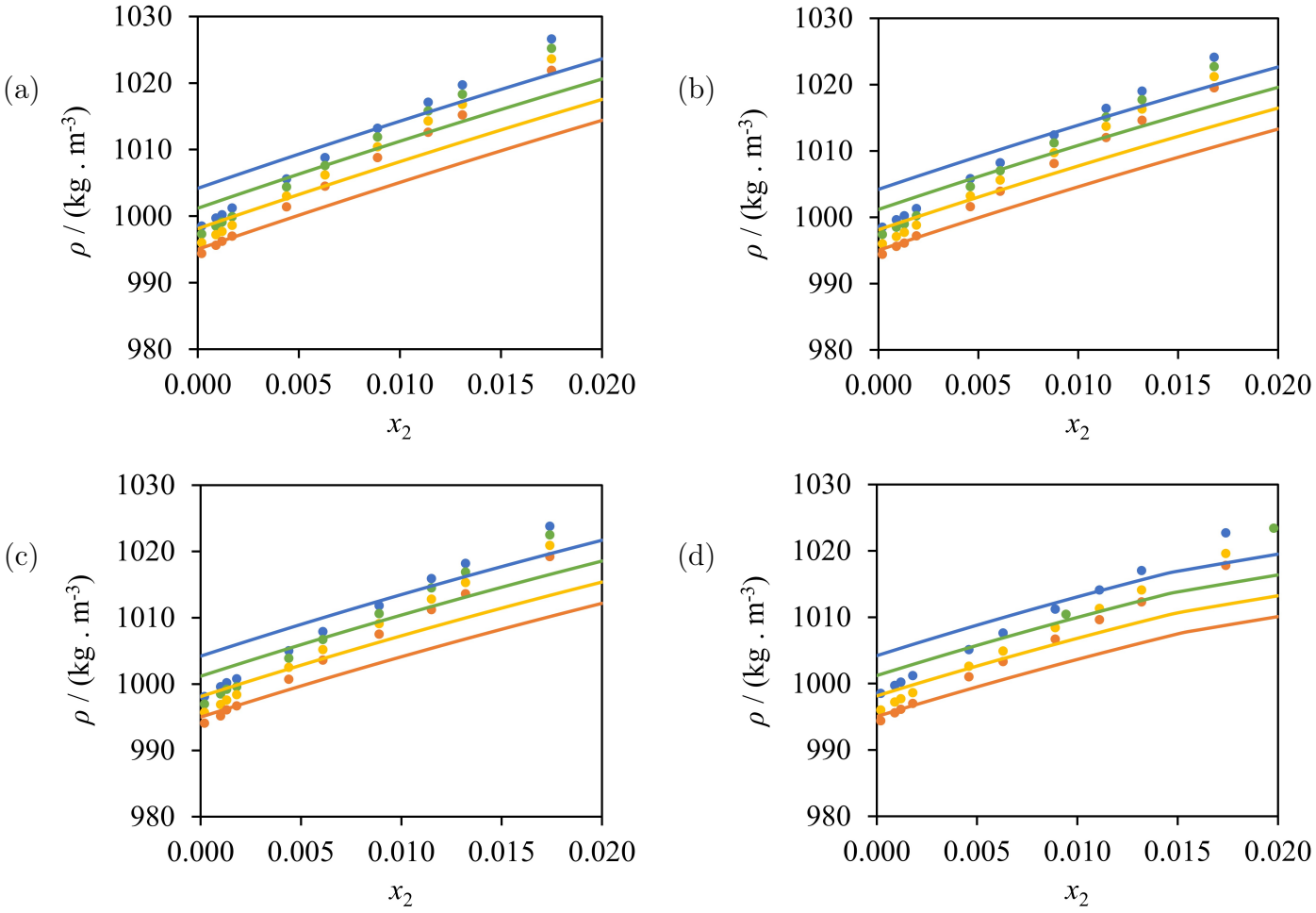


Figure 3.6: SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-COOH unlike interaction compared to experimental data (symbols)[114] of liquid densities of amino-acid mixtures at 1 bar. (a) Water (1) + 3-aminopropanoic acid (2). (b) Water (1) + 4-aminobutanoic acid. (c) Water (1) + 5-aminopentanoic acid. (d) Water (1) + 6-aminohexanoic acid. Different colours represent mixture densities at different temperature: 293.15 K (blue); 298.15 K (green); 303.15 K (yellow); and 308.15 K (orange). The estimated experimental uncertainty in the reported densities are estimated to be ±0.05 kg/m^3

to these calculations can be found in Table 3.4.

3.4.3 COOH−CHOH

The COOH−CHOH interaction is optimised using VLE data of the binary mixtures propanoic acid + 2-propanol [125] and butanoic acid + 2-propanol [126], and VLE [125] and liquid-density [127] data of the mixture propanoic acid + 2-butanol. The SAFT-γ Mie calculations show satisfactory agreement with experiments, as illustrated by the plots in 3.9 and %AAD

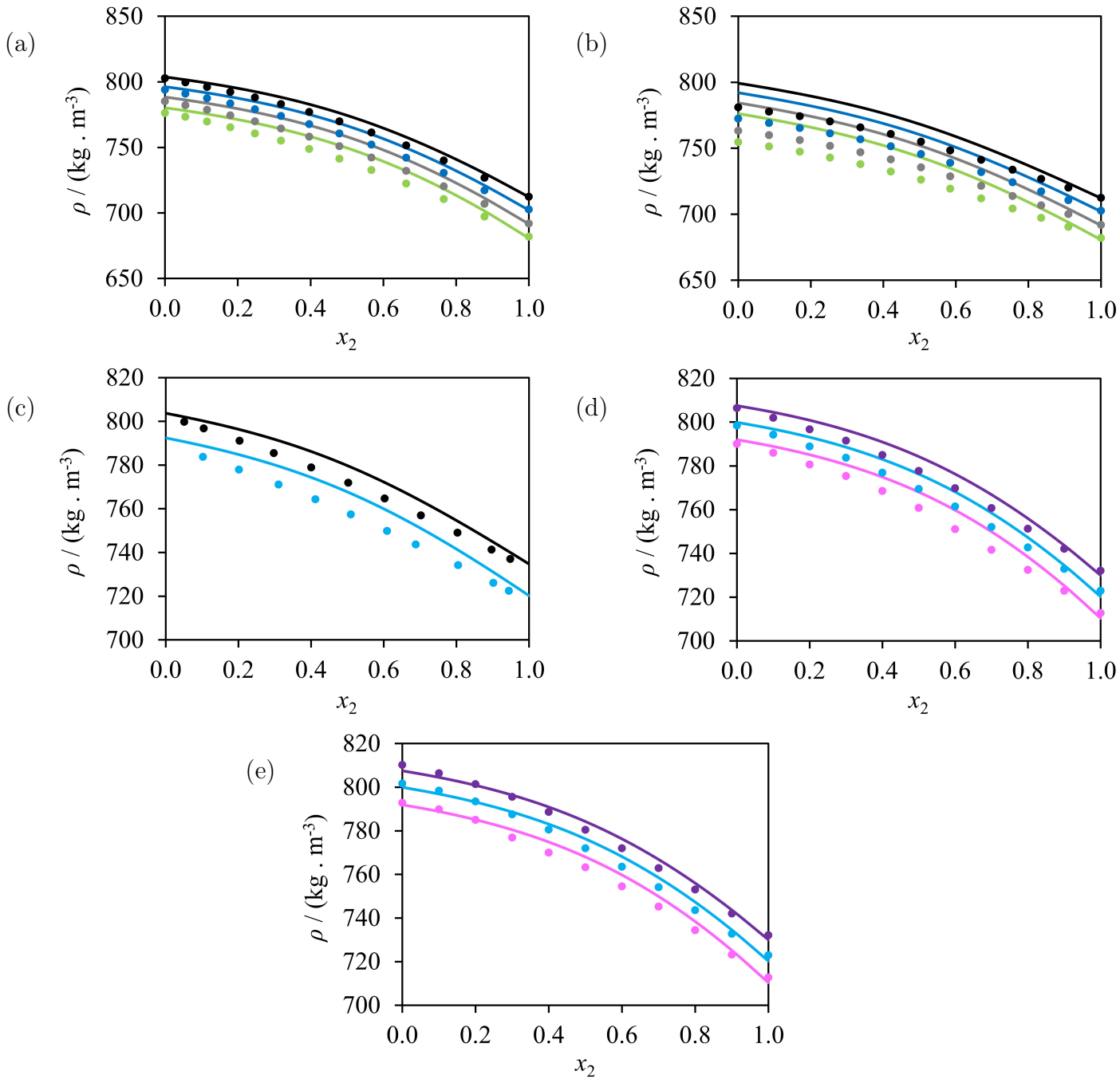


Figure 3.7: SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-CHOH unlike interaction compared to experimental data (symbols) of liquid densities of mixtures of primary amines with secondary alcohols at 1 bar. (a) 2-butanol (1) + 1-propanamine (2) [122]. (b) 2-propanol (1) + 1-propanamine (2) [122]. (c) 2-butanol (1) + 1-butanamine (2) [123]. (d) 2-hexanol (1) + 1-butanamine (2) [124]. (e) 3-hexanol (1) + 1-butanamine (2) [124]. Different colours represent mixture densities at different temperature: 298.15 K (black); 303.15 K (purple); 308.15 K (blue); 313.15 K (light blue); 318.15 K (grey); 323.15 K (pink); and 328.15 K (green). The experimental uncertainties in the density measurements are reported to be 0.01 kg/m^3 [122] and 0.1 kg/m^3 [123, 124].

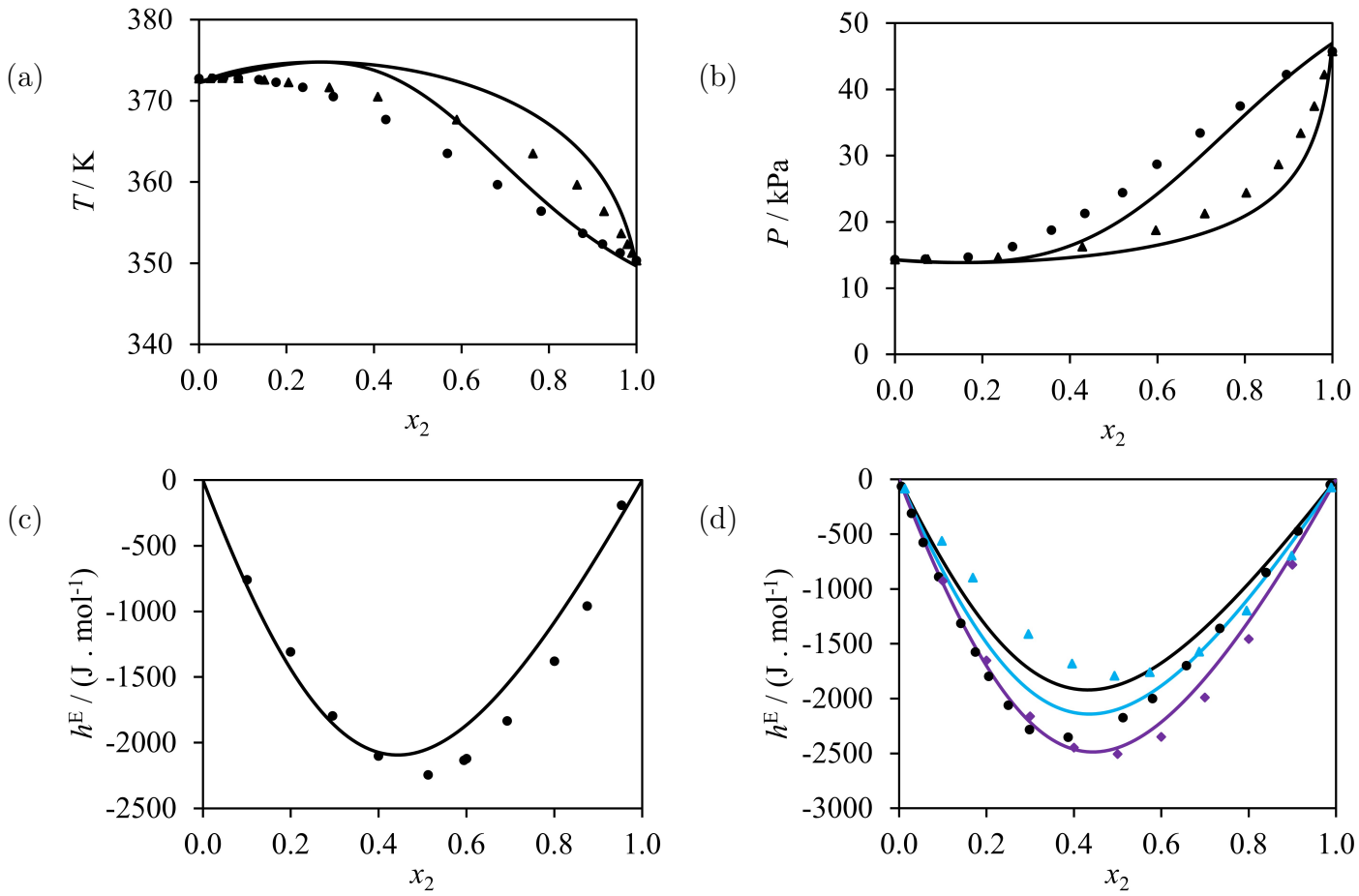


Figure 3.8: SAFT-γ Mie calculations (curves) used in the optimisation of the NH_2-CHOH unlike interaction compared to experimental data (symbols) of VLE and excess enthalpies of mixtures of primary amines with secondary alcohols. (a) Vapour–liquid equilibrium isobar (Txy) [119] of 2-butanol (1) + 1-butanamine (2) at 1 bar. The maximum estimated uncertainty in liquid- and vapour-phase mole fractions is ±0.0002. (b) Vapour–liquid equilibrium isotherm (Pxy) [118] of 2-butanol (1) + 1-butanamine (2) at 328.15 K. (c) Excess enthalpy of 2-butanol (1) + 1-butanamine (2) [118] at 298.15 K and 1 bar. (d) Excess enthalpies of mixtures of 2-propanol (1) with primary amines at 298.15 K and 1 bar: 1-butanamine (black circles) [121]; 2-butanamine (light blue triangles) [121]; and 2-propanamine (purple diamonds) [120].

values in Table 3.4.

3.4.4 NH_2–CH

The NH_2–CH interaction is optimised by adjustment using experimental vapour-pressure data of 2-propanamine [128], 2-butanamine [129], and 2-methyl-1-propanamine [130, 131], liquid-density data of 2-propanamine [132] and 2-butanamine [133, 134], VLE data of ethane + 2-propanamine [135], and hexane + 2-propanamine [136] mixtures, and excess-enthalpy data of

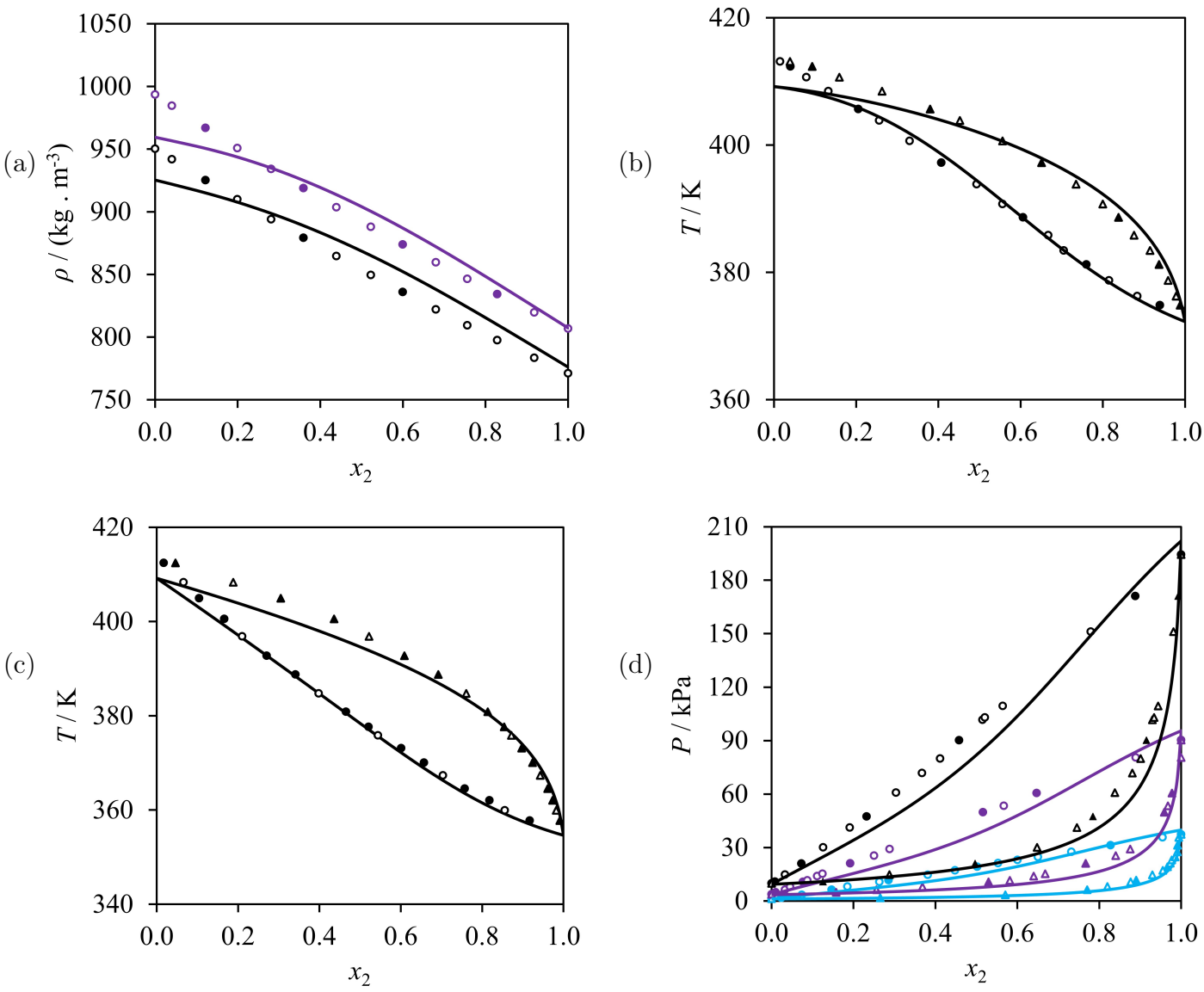


Figure 3.9: SAFT-γ Mie calculations (curves) used in the optimisation of the COOH-CHOH unlike interaction compared to experimental data (symbols) of VLE and liquid densities of mixtures of alkanoic acids with secondary alcohols. Filled symbols denote data that were used in the optimisation and empty symbols denote data that were not used. (a) Mixture liquid density of propanoic acid (1) + 2-butanol (2) at 1 bar, and at 293.15 K (purple) and 333.15 K (black) [127]. The uncertainty in the density measurements is reported to be 0.01 kg/m^3. (b) Temperature–composition isobar illustrating the vapour–liquid equilibrium of propanoic acid (1) + 2-butanol (2) at 1 bar [125]. (c) Temperature–composition isobar illustrating the vapour–liquid equilibrium of propanoic acid (1) + 2-propanol (2) at 1 bar [125]. (d) Pressure–composition isotherms illustrating the vapour–liquid equilibria of butanoic acid (1) + 2-propanol (2) at 333.15 K (light blue); 353.15 K (purple); and 373.15 K (black) [126]. The experimental uncertainty in measured pressures is reported to be less than 0.15 kPa. In (b), (c), and (d), triangles denote dew points and circles denote bubble points.

hexane + 2-butanamine [137] and heptane + 2-butanamine [138] mixtures. The theory yields very good agreement with the experimental data for the pure-compound properties, as shown

in Figure 3.10. In the case of mixture properties, the model yields satisfactory agreement with experimental data as can be seen in Figure 3.11. Corresponding percentage average absolute deviations (%AAD) and average absolute deviations (AAD) are presented in Table 3.4.

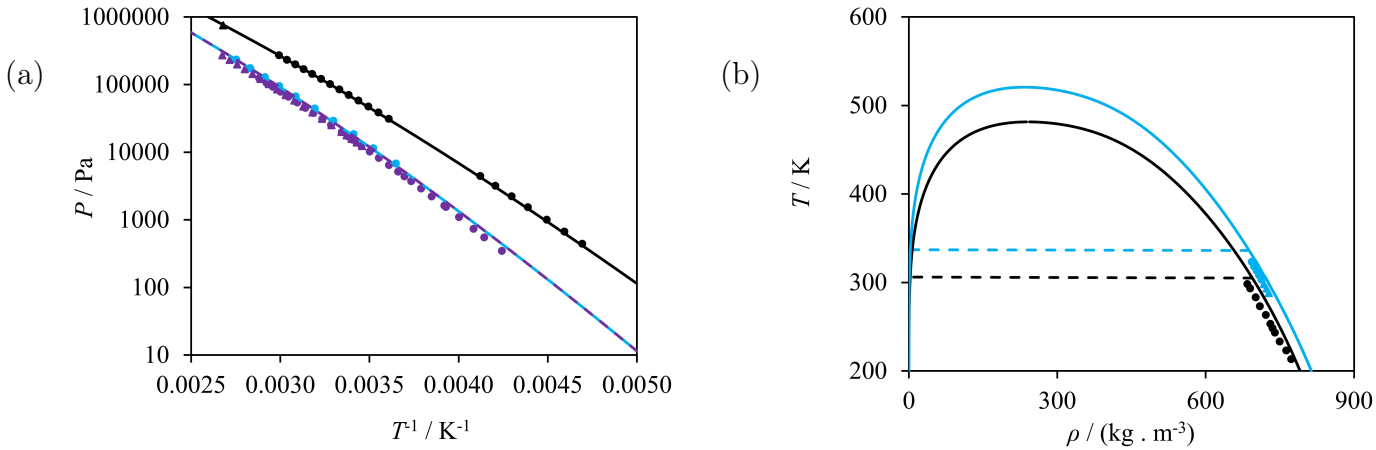


Figure 3.10: SAFT-γ Mie calculations (curves) of pure-component properties used in the optimisation of the NH_2-CH unlike interaction compared to experimental data (symbols). (a) Vapour-pressure of 2-propanamine (black), 2-butanamine (light blue), and 2-methyl-1-propanamine (purple). Symbols denote different experimental sources: black circles [128]; black triangle [139]; light blue circles [129]; purple circles [131]; and purple triangles [130]. SAFT-γ calculations for 2-butanamine and 2-methyl-1-propanamine are identical because the molecules are comprised of the same groups. (b) Isobaric liquid density of 2-propanamine (black) and 2-butanamine (light blue). The dashed curve denotes calculation of density at 1 bar whereas the solid curve denotes calculation of saturated density. Symbols denote different experimental sources: black circles [132]; blue circles [134]; and blue triangles [133]

3.4.5 The peptide (CONH) group

We have considered the possibility of modelling the amide functional group as separate C=O and NH groups, which have been parametrised in previous work [140, 71], but we find that such a model does not lead to an accurate representation of the properties of aqueous dipeptides. This is most likely because the C=O and NH groups were parametrised using experimental data of 2-ketones and secondary amines, respectively. In these families the two groups are not adjacent, and large polarisation effects that arise when the two groups are adjacent, as is the case in peptide molecules are therefore neglected. Hence, to model mixtures containing peptides, a new group CONH is introduced and parametrised using experimental data of amides (R−CONH−R′), which are structurally more similar to peptides. The CONH group is modelled

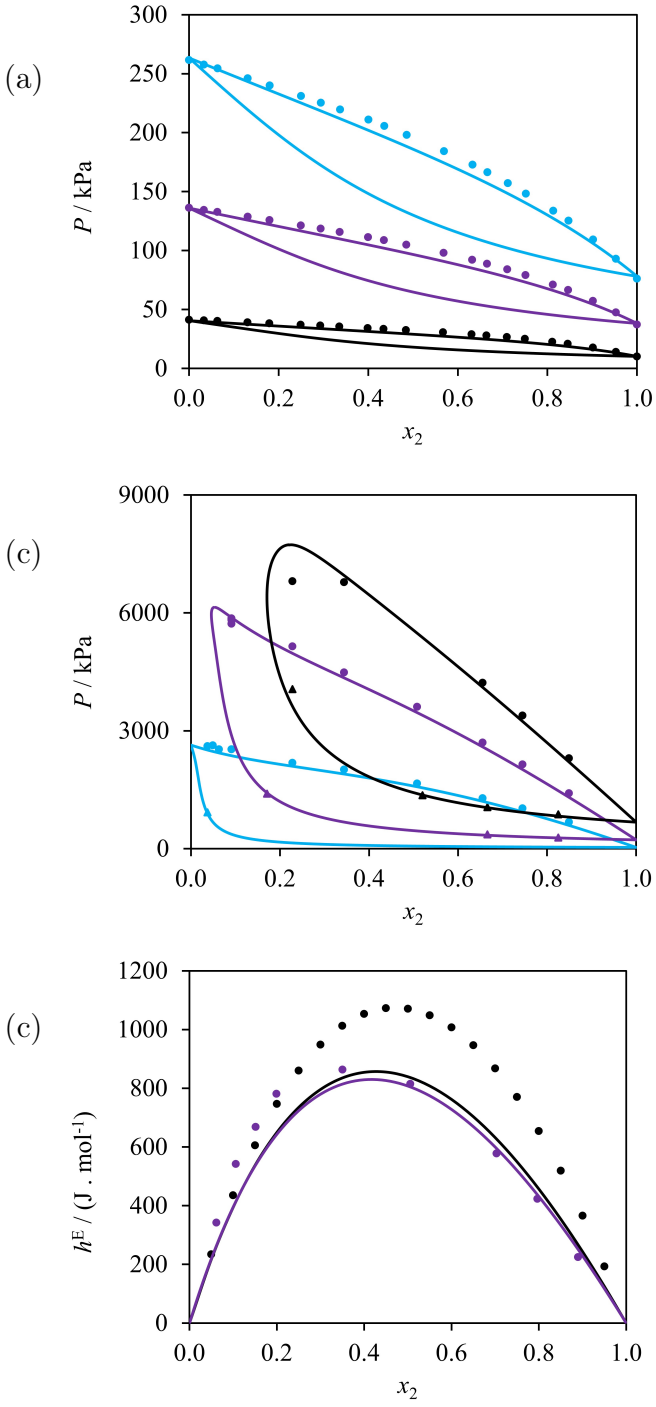


Figure 3.11: SAFT-γ Mie calculations (curves) of mixture properties used in the optimisation of the NH_2-CH unlike interaction compared to experimental data (symbols). (a) Pressure–composition isotherms of hexane (1) + 2-propanamine (2) [136] at 283.15 K (black), 313.15 K (purple), and 333.15 K (light blue). (b) Pressure–composition isotherms of ethane (1) + 2-propanamine (2) [135] at 279.1 K (light blue), 328.3 K (purple), and 367.9 K (black). (c) Excess enthalpies of hexane (1) + 2-butanamine (2) at 298.15 K and 1 bar (black) [138] and heptane (1) + 2-butanamine (2) at 298.15 K and 1 bar (purple) [137].

with three associating sites, two of type e (e_1 in Tables 3.1-3.3), representing the lone pairs of the oxygen atom, and one of type H corresponding to the hydrogen. Contrasting our treatment of amine groups, we find that it is not necessary to add a further e-type site to account for the electron pair of the nitrogen atom. This may reflect that the ground state of an amide is stabilised by the delocalisation of the nitrogen lone-pair electrons through orbital overlap with the carbonyl group of the amide [141]; this delocalisation is the principal reason that amides are non-basic in nature, whereas amines, in which the nitrogen lone pair is not delocalised, are quite strong bases.

The CONH−CONH, CONH−CH_3, and CONH−CH_2 interaction parameters are adjusted simultaneously using pure n-alkylamide and n-alkylamide + n-alkane mixture data. Specifically, pure compound vapour-pressure data of n-ethyl-acetamide [142, 143], n-propyl-acetamide [143, 144], n-butylacetamide [145, 146], n-pentyl-acetamide [147], n-methyl-propanamide [148], n-methyl-hexanamide [149], and n-butyl-propanamide [144], and liquid-density data of n-ethyl-acetamide [150, 151, 152], n-methyl-propanamide [153, 154, 155, 156, 157], and n-methyl-butanamide [158] are used. Mixture bubble-pressure and excess-enthalpy data of n-decane + n-methyl-acetamide [159], n-octane + n-methyl-acetamide [160, 142], bubble-pressure of n-decane + n-ethyl-acetamide [142] and bubble-pressure and liquid–liquid-equilibrium (LLE) data of n-octane + n-methyl-propanamide [142] are also used. Selected systems used in this parametrisation are represented in Figures 3.12 – 3.14. As can be seen in the figures, the SAFT-γ Mie calculations yield satisfactory agreement with experimental data of the pure and mixed systems, especially considering the stringent test of delivering the vapour–liquid–liquid equilibria observed in the mixtures of n-octane + n-methylacetamide, n-decane + n-methylacetamide, and n-decane + n-ethylacetamide and the vapour–liquid and liquid–liquid equilibria observed in n-octane + n-methylpropanamide. The extrapolative suitability of the group parameters is validated by comparison to a small number of vapour pressure and liquid density of n-methylacetamide [161] data points not included in the parameter estimation.

The interaction CONH−H_2O, which is crucial in modelling the aqueous solubility of peptides, is optimised using vapour–liquid equilibrium [161, 159], density [166, 167] and excess-enthalpy [168] data of water + n-methylacetamide mixtures and density [166], and excess-

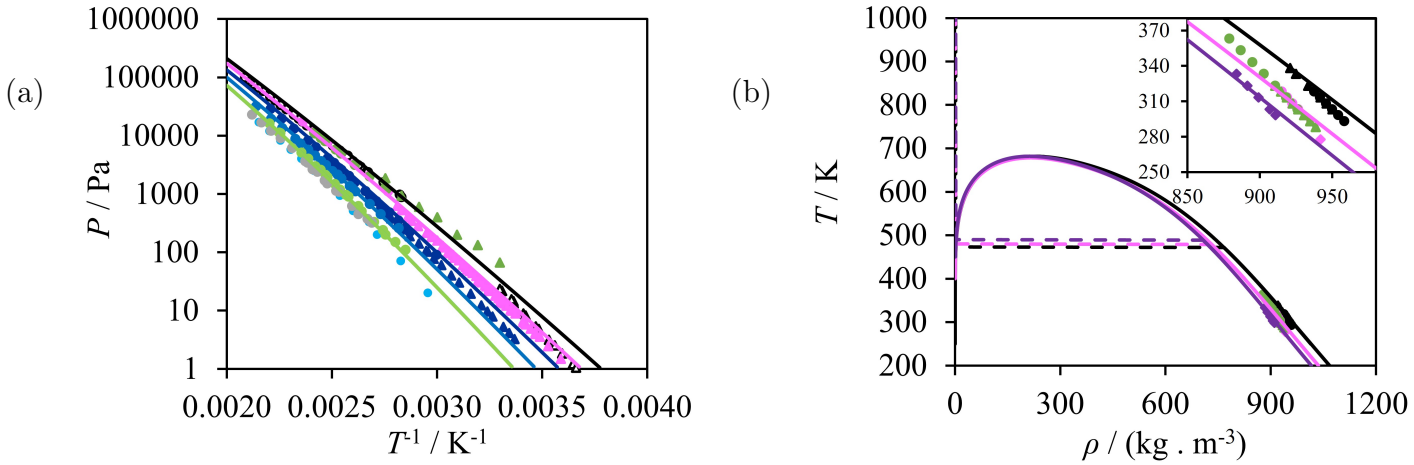


Figure 3.12: SAFT-γ Mie description of pure-component properties used in the optimisation of the CONH−CONH, CONH−CH_3 and CONH−CH_2 interactions. Empty symbols denote data not used in the parameter optimisation, filled symbols denote those used in the optimisation, and curves denote SAFT-γ Mie calculations. (a) Vapour pressures of *n*-methylacetamide (black) [162, 161], *n*-ethylacetamide (pink) [142, 143], *n*-propylacetamide (navy blue) [144, 143], *n*-butylacetamide (blue) [146, 145], *n*-pentylacetamide (light blue) [147], *n*-methylpropanamide (dark green) [148, 142], *n*-butylpropanamide (green) [144], and *n*-methylhexanamide (grey) [149]. *N*-ethylacetamide and *n*-methylpropanamide comprise the same groups and are therefore represented by the same calculation (pink curve); similarly for *n*-pentylacetamide, *n*-butylpropanamide, and *n*-methylhexanamide (green curve). (b) Isobaric liquid density of *n*-methylacetamide (black) [163, 164], *n*-ethylacetamide (pink) [150, 152, 151], *n*-methylpropanamide (dark green) [154, 153, 155, 156, 157], and *n*-methylbutanamide (purple) [158] at 1 bar. Solid curves represent saturation density calculations, whereas dashed curves represent isobaric density calculations at 1 bar. *N*-ethylacetamide and *n*-methylpropanamide comprise the same groups and are therefore represented by the same calculation (pink curve).

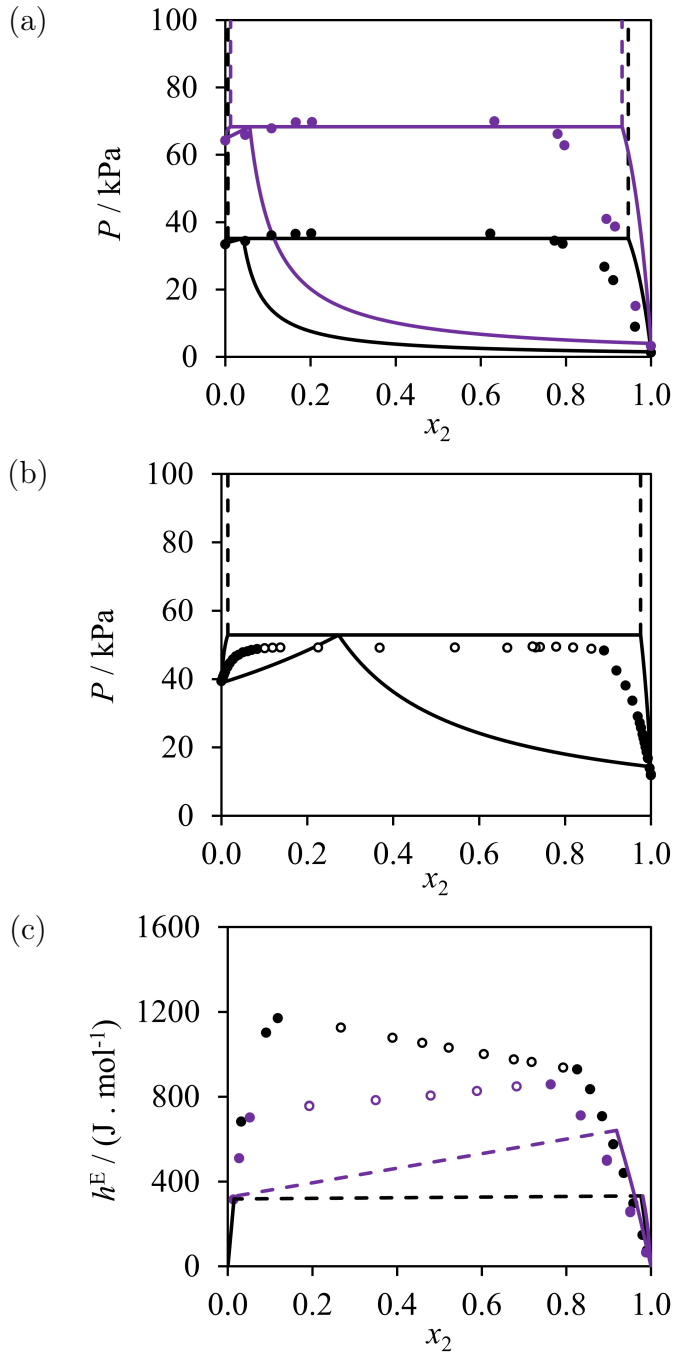


Figure 3.13: SAFT-γ Mie description of n-methylacetamide-mixture properties used in the optimisation of the CONH−CONH, CONH−CH_3, and CONH−CH_2 interactions. Empty symbols denote experimental data not used in the optimisation and curves represent SAFT calculations. (a) Pressure–composition isotherms illustrating the vapour–liquid–liquid equilibrium of octane (1) + *n*-methylacetamide (2) [142] at 363.15 K (black) and 383.15 K (purple). (b) Pressure–composition isotherm illustrating the vapour–liquid–liquid equilibrium of decane (1) + *n*-methylacetamide (2) at 413.49 K [142]. (c) Excess enthalpies of *n*-decane (1) + *n*-methylacetamide (2) at 413.15 K and 1.617 MPa (black) [159] and *n*-octane (1) + *n*-methylacetamide (2) at 398.15 K and 1.891 MPa (purple) [159].

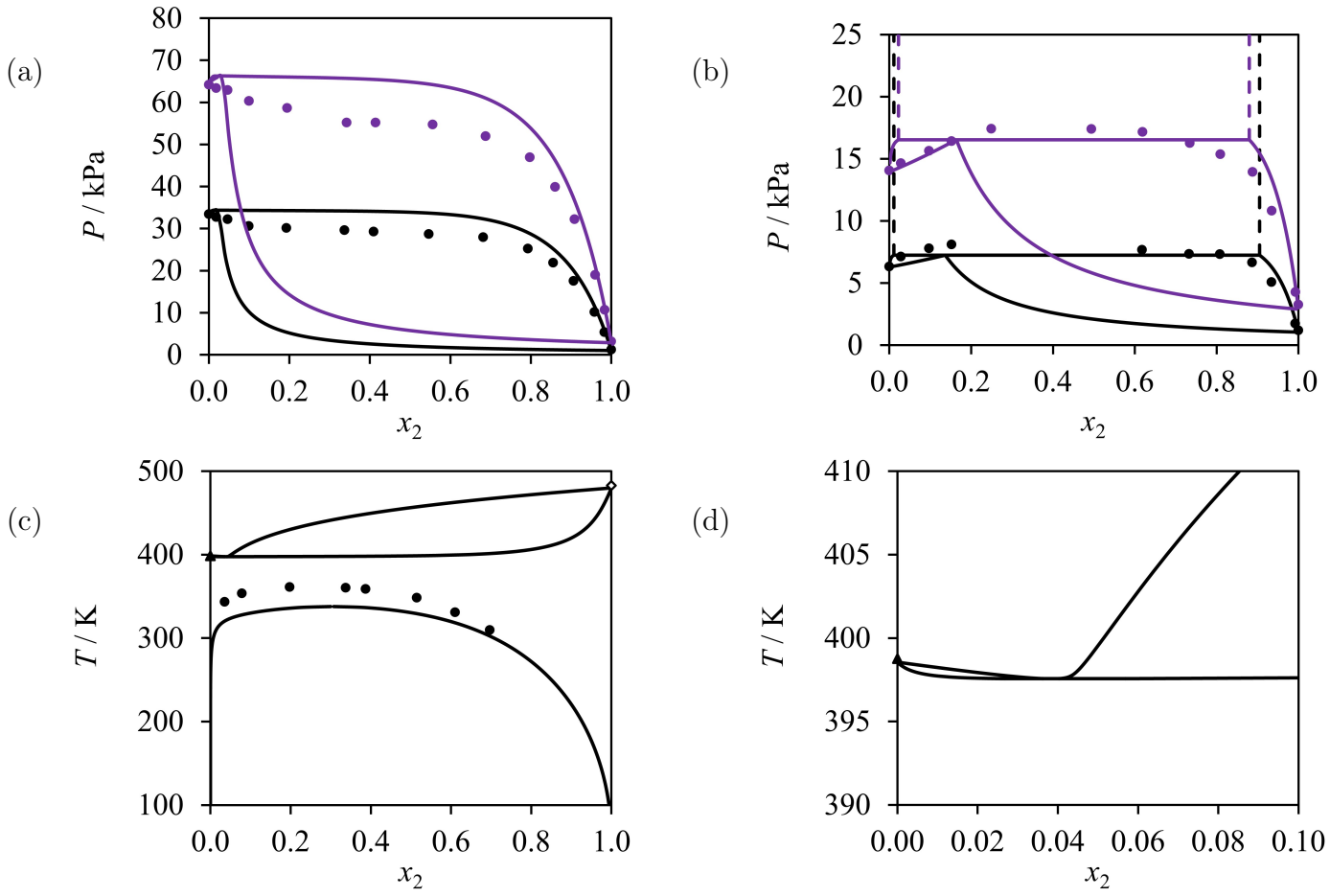


Figure 3.14: SAFT-γ Mie description of alkane–amide mixture properties used in the optimisation of the CONH−CONH, CONH−CH_3, and CONH−CH_2 interactions. Filled symbols denote experimental data used in the optimisation and curves represent SAFT calculations. (a) Pressure–composition isotherms illustrating the vapour–liquid equilibrium of octane (1) + *n*-methylpropanamide (2) [142] at 363.15 K (black) and 383.15 K (purple). (b) Pressure–composition isotherms illustrating the vapour–liquid–liquid equilibrium of decane (1) + *n*-ethylacetamide (2) at 363.15 K (black) and 383.15 K (purple) [142]. (c) Temperature–composition isobar illustrating the fluid-phase diagram of *n*-octane (1) + *n*-methylpropanamide (2) at 1 bar [142]. The SAFT calculations predict an upper critical solution temperature that lies below a VLE envelope characterized by a homogeneous azeotrope. Mixture data are only available for the liquid–liquid region, represented by circles [142]. The triangle symbol [165] denotes the normal boiling point of *n*-octane, which had been used to parameterise the CH_3 and CH_2 groups [51]. The diamond symbol [165] denotes the normal boiling points, which had not been used in parameterising any of the model interactions.

enthalpy data [169] of water + n-ethylacetamide and water + n-methylpropanamide mixtures. The optimised parameters lead to calculations in good agreement with VLE experimental data and satisfactory agreement with excess enthalpy data (cf. Table 3.4). Selected phase diagrams comparing our SAFT-γ Mie calculations and the experimental data used in the parameter development are presented in Figure 3.15. As can be seen in the figure, an accurate description of the bubble and dew pressures is achieved. The excess enthalphies of the water + n-methylacetamide and water + n-ethylacetamide mixtures are also described in close agreement with the available data, while in the case of the water + n-methylpropanamide mixture, the calculations present slightly larger deviations from the data. In a SAFT-γ Mie model using first-order groups, such as those developed in the current work, n-ethylacetamide and n-methylpronanamide are modelled with identical groups, and as such have identical calculated properties. Experimentally, however, the two molecules have noticeably different values of the excess enthalpy (cf. Figure 3.15(c)).

The unlike interactions between CONH, NH_2, and COOH, also need to be characterised in order to model peptides in our approach. Due to the lack of experimental data of mixtures containing the CONH and NH_2 groups, the cross-interaction is optimised using only one set of excess-enthalpy data of n-methylacetamide + 1-hexanamine mixtures [168]. The parameters obtained are validated by predicting the liquid density of aqueous mixtures of alkylureas (methyl- [172, 173, 174], ethyl- [174, 175], and butylurea [174], are used). As can be seen in Figure 3.16, the optimised parameters yield predictive calculations of the density of these solutions in satisfactory agreement with the experimental data available. The CONH−COOH interaction is optimised using experimental data of alkanoic acid + n-methylacetamide mixtures, although as with the previous groups, few experimental data are found for mixtures including these two groups alone. The optimisation is carried out using excess-enthalpy data of propanoic acid + n-methylacetamide [168], and isobaric VLE and density data of acetic acid + n-methylacetamide [176]. The resulting calculations are compared to experimental data in Figure 3.17. As can be seen, the excess enthalpy of the propanoic acid + n-methylacetamide mixtures is described in reasonably good agreement with the data (note the small units of J mol^{-1}), but in the case of the acetic acid + n-methylacetamide vapour–liquid phase equilibrium,

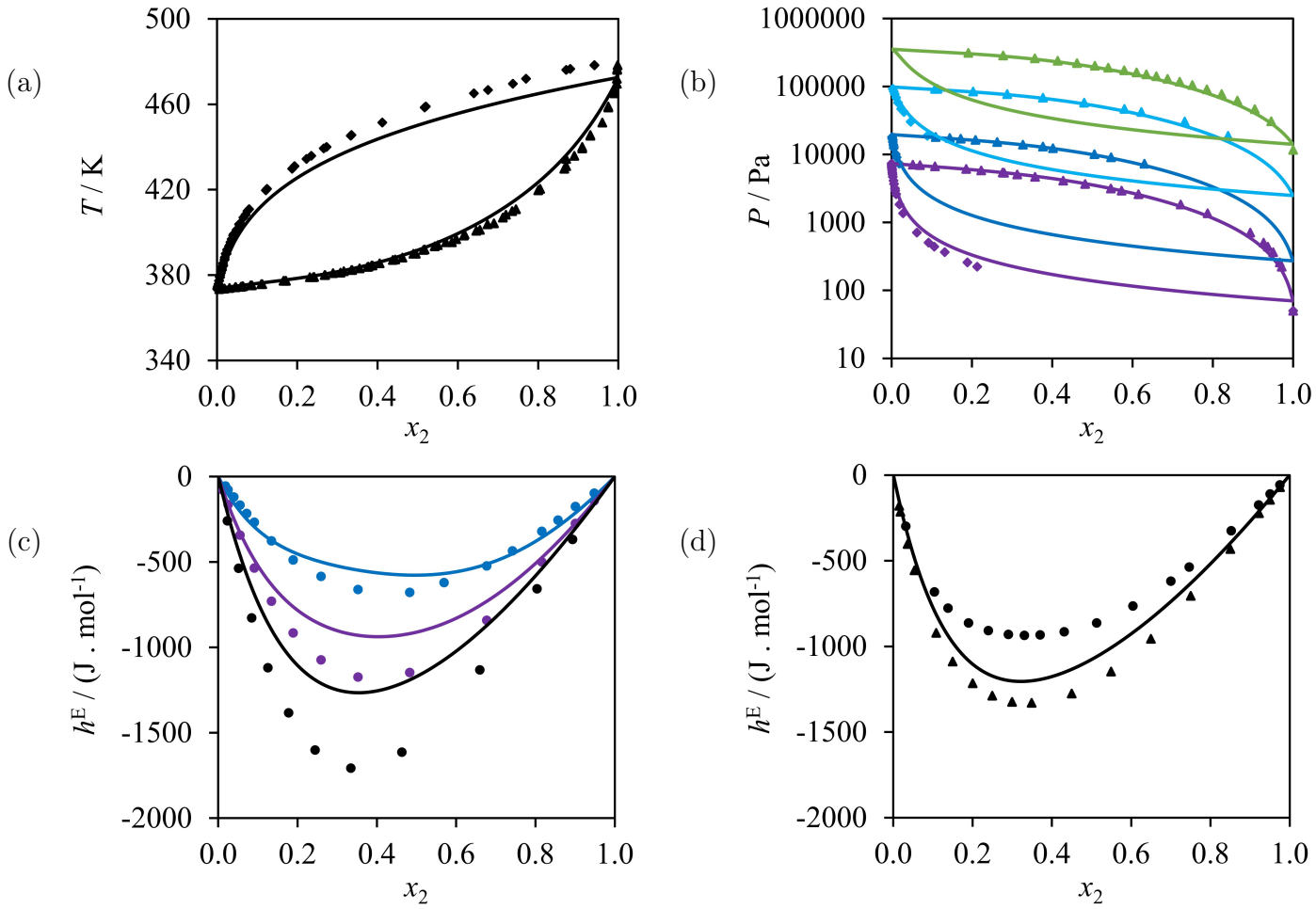


Figure 3.15: SAFT-γ Mie description of properties used in the optimisation of the CONH–H_2O interaction. (a) Temperature–composition isobar illustrating the vapour–liquid equilibrium of water (1) + *n*-methylacetamide (2) at 1 bar [161]. (b) Pressure–composition isotherms illustrating the vapour–liquid equilibrium of water (1) + *n*-methylacetamide (2) at 313 K (purple) [170], 333 K (blue) [171], 373 K (light blue) [171], and 413 K (green) [159]. (c) Excess enthalpies of water (1) + *n*-methylacetamide (2) [168] at 398.15 K and 17.54 bar (black); 363.15 K and 17.54 bar (purple); and 323.15 K and 18 bar (blue). (d) Excess enthalpies [169] at 308.15 K and 1 bar of water (1) + *n*-methylpropanamide (2) (circles) and water (1) + *n*-ethylacetamide (2) (triangles). Since *n*-methylpropanamide and *n*-ethylacetamide comprise the same SAFT-γ Mie groups, they are represented by the same calculation denoted by the black curve.

larger deviations are seen. The underestimation of the saturation temperature of pure acetic acid is especially noticeable. A degree of deviation between the SAFT-γ Mie calculations and experiment for acetic acid mixtures is expected, as no acetic acid data were used in optimising the previously characterised COOH−COOH like and COOH−CH_3 unlike interactions [52], added to the fact that in such a small molecule the group-contribution proposition is likely to break-down [55]. The decision to include acetic acid to characterise first-order groups here, is based on the scarcity of other alkanoic acid + amide mixture data. Overall, the model for CONH and its unlike interactions lead to acceptable agreement with experiments as evidenced by the AAD and %AAD values presented in Table 3.4.

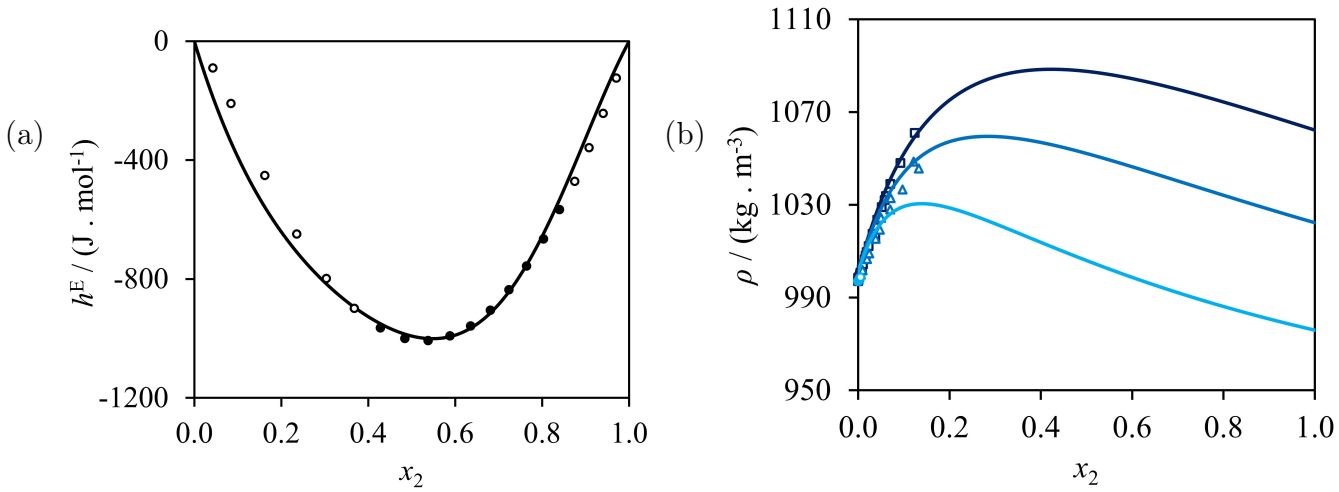


Figure 3.16: SAFT-γ Mie description of mixture properties used in the parameter optimisation and validation of the CONH–NH_2 interaction. Filled symbols denote data that were used in the parameter optimisation and empty symbols denote data used for validation only; curves denote SAFT-γ Mie predictions. (a) Excess enthalpy of 1-hexanamine (1) + *n*-methylacetamide (2) at 363.15 K and 12.03 bar [168]. (b) Liquid density of mixtures at 298.15 K and 1 bar of water (1) + methylurea (2) in dark blue and squares [177, 172, 174, 173], water (1) + ethylurea (2) in blue and triangles [174, 175], and water (1) + butylurea (2) in light blue and circles [174].

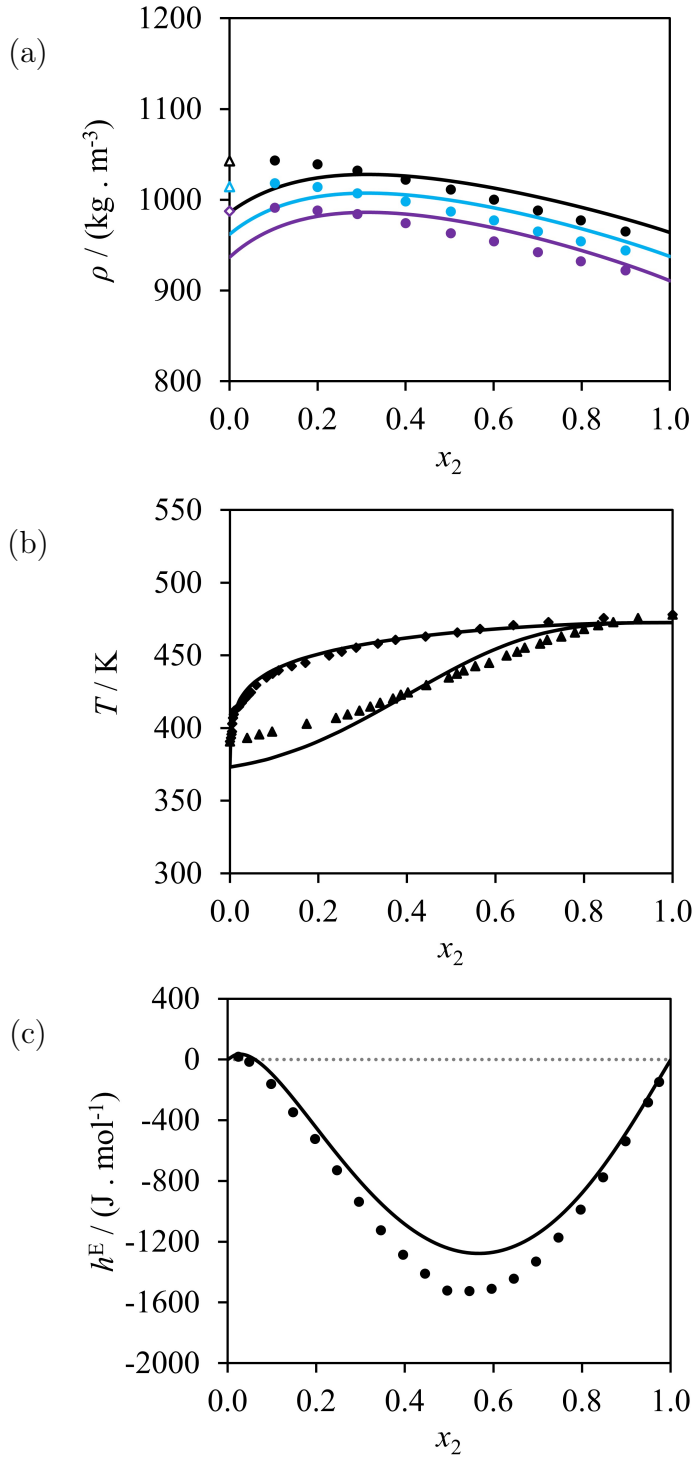


Figure 3.17: SAFT-γ Mie description of mixture properties used in the parameter optimisation of the interaction CONH–COOH. Filled symbols denote experimental data that were used in the parameter optimisation, and curves denote SAFT-γ Mie calculations. (a) Isobaric mixture density of acetic acid (1) + *n*-methylacetamide (2) [178] at 298.15 K (black); 323.15 K (light blue); and 348.15 K (purple). Triangles [179] and diamonds [180] denote pure acetic acid densities. (b) Temperature–composition isobar illustrating the vapour–liquid equilibrium of acetic acid (1) + *n*-methylacetamide (2) at 1 bar [176]; diamonds and triangles denote dew and bubble temperature data, respectively. (c) Excess enthalpy of propanoic acid (1) + *n*-methylacetamide (2) at 363.15 K and 11.35 bar [168].

Table 3.4: Overview of the accuracy of SAFT-γ Mie, represented in %AAD and AAD, in the calculation of vapour pressure (P^{sat}), liquid pure-component and mixture densities (ρ), bubble temperature (T_{bub}), dew temperature (T_{dew}), bubble pressure (P_{bub}), dew pressure (P_{dew}), LLE temperature (T_{LLE}), and excess enthalpy (H^{E}). N^{D} is the number of experimental data points used to evaluate the accuracy of the model.

	T range	P range			%AAD	AAD		
System	/ K	/ kPa	x_2 range	N^{D}	P^{sat}	P^{sat} / kPa	Figures	Ref.
2-propanamine	213 – 373	—	—	21	2.3	1.1	3.10(a)	[128, 139]
2-butanamine	274 – 363	—	—	10	3.5	1.3	3.10(a)	[129]
2-methyl-1-propanamine	236 – 374	—	—	47	17	8.2	3.10(a)	[130, 131]
n-ethylacetamide	279 – 423	—	—	59	21	0.096	3.12(a)	[142, 143]
n-propylacetamide	297 – 453	—	—	39	31	0.41	3.12(a)	[143, 144]
n-butylacetamide	355 – 502	—	—	19	6.7	0.96	3.12(a)	[146, 145]
n-pentylacetamide	374 – 384	—	—	14	34	1.3	3.12(a)	[147]
n-methylpropanamide	303 – 685	—	—	30	23	0.25	3.12(a)	[148, 142]
n-methylhexanamide	372 – 472	—	—	18	20.	1.2	3.12(a)	[149]
n-butylpropanamide	351 – 454	—	—	20	4.2	0.045	3.12(a)	[144]
	T range	P range			%AAD	AAD		
System	/ K	/ kPa	x_2 range	N^{D}	ρ	ρ / (kg m^{-3})	Figures	Ref.
2-propanamine	213 – 298	101.33	—	11	2.1	15	3.10(b)	[132]
2-butanamine	288 – 323	101.33	—	8	1.4	10.	3.10(b)	[133, 134]
n-ethylacetamide	278 – 318	101.33	—	8	0.32	3.0	3.12(b)	[150, 152, 151]
n-methylpropanamide	288 – 363	101.33	—	34	0.34	3.4	3.12(b)	[154, 153, 155, 156, 157]
n-methylbutanamide	298 – 333	101.33	—	5	0.33	3.0	3.12(b)	[158]
System	T range	P range			%AAD	AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	ρ	ρ / (kg m^{-3})	Figures	Ref.
water + 3-aminopropanoic acid	293 – 308	101.33	0.0002 – 0.002	40	0.33	3.3	3.5(c)	[114]
water + 4-aminobutanoic acid	293 – 308	101.33	0.0002 – 0.002	40	0.32	3.2	—	[114]
water + 5-aminopentanoic acid	293 – 308	101.33	0.0002 – 0.002	40	0.32	3.3	—	[114]
water + 6-aminohexanoic acid	293 – 308	101.33	0.0002 – 0.002	35	0.45	4.6	—	[114]
2-propanol + 1-propanamine	298 – 328	81.50	0.00 – 1.00	52	1.8	13	—	[122]

2-butanol + 1-propanamine	298 – 328	81.50	0.00 – 1.00	52	0.75	5.7	—	[122]
2-butanol + 1-butanamine	298 – 313	101.33	0.05 – 0.95	21	0.82	6.3	—	[123]
2-hexanol + 1-butanamine	303 – 323	101.33	0.00 – 1.00	33	0.59	4.5	—	[124]
3-hexanol + 1-butanamine	303 – 323	101.33	0.00 – 1.00	33	0.35	2.7	—	[124]
2-butanol + propanoic acid	293 – 333	101.33	0.00 – 1.00	84	1.4	12	—	[127]
water + *n*-methylacetamide	298 – 323	101.33	0.00 – 1.00	107	1.3	13	—	[166, 167]
water + *n*-ethylacetamide	298	101.33	0.00 – 1.00	19	1.9	19	—	[166]
water + *n*-methylpropanamide	293 – 313	101.33	0.00 – 1.00	20	1.1	10.	—	[166]
water + methyl urea	274 – 333	101.33	0.00 – 0.12	256	0.76	7.7	3.16(b)	[177, 172, 174, 173]
water + ethyl urea	288 – 323	101.33	0.00 – 0.13	76	0.75	7.6	3.16(b)	[174, 175]
water + butyl urea	288 – 308	101.33	0.00 – 0.005	30	0.54	5.4	3.16(b)	[174]

System	T range	P range			%AAD	AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	T_{bub}	T_{bub} / K	Figures	Ref.
2-propanol + propanoic acid	358 – 413	101.33	0.00 – 1.00	18	0.28	1.1	—	[125]
water + *n*-methylacetamide	373 – 478	101.33	0.00 – 1.00	82	0.47	2.0	3.15(a)	[161]
acetic acid + *n*-methylacetamide	391 – 478	101.33	0.00 – 1.00	31	1.6	6.9	3.17(b)	[176]

System	T range	P range			%AAD	AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	T_{dew}	T_{dew} / K	Figures	Ref.
2-propanol + propanoic acid	358 – 413	101.33	0.00 – 1.00	18	0.51	2.0	—	[125]
water + *n*-methylacetamide	373 – 478	101.33	0.00 – 1.00	82	0.78	3.3	3.15(a)	[161]
acetic acid + *n*-methylacetamide	373 – 473	101.33	0.00 – 1.00	31	0.69	3.0	3.17(b)	[176]

System	T range	P range			%AAD	AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	P_{bub}	P_{bub} / kPa	Figures	Ref.
water + glycine	298	3.0 – 3.2	0.00 – 0.06	14	0.66	0.021	3.5(a)	[113]
2-butanol + 1-butanamine	328	14 – 46	0.00 – 1.00	12	9.0	2.2	—	[118]
2-propanol + butanoic acid	333 – 373	1 – 194	0.00 – 1.00	50	16	5.2	—	[126]
ethane + 2-propanamine	279 – 372	676 – 6850	0.04 – 0.85	100	4.3	140	3.10(c)	[135]
hexane + 2-propanamine	283 – 333	10 – 262	0.00 –1.00	63	5.3	4.2	3.10(d)	[136]
n-octane + *n*-methylacetamide	363 – 398	2 – 98	0.00 – 1.00	45	27	9.5	—	[142, 160]
n-decane + *n*-methylacetamide	414	12 – 50	0.00 – 1.00	43	30.	8.7	—	[160]

n-decane + n-ethylacetamide	363 – 383	1 – 17	0.00 – 1.00	24	8.2	0.71	—	[142]
n-octane + n-methylpropanamide	363 – 383	1 – 64	0.00 – 1.00	30	12	4.1	—	[142]
water + n-methylacetamide	313 – 413	0.05 – 313	0.00 – 1.00	67	6.0	2.6	3.15(b)	[170, 171, 159]

System	T range	P range			%AAD	AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	P_{dew}	P_{dew} / kPa	Figures	Ref.
2-butanol + 1-butanamine	328	14 – 46	0.00 – 1.00	12	9.2	2.5	—	[118]
2-propanol + butanoic acid	333 – 373	1 – 194	0.00 – 1.00	50	16	6.1	—	[126]
ethane + 2-propanamine	279 – 369	208 – 4860	0.04 – 0.82	73	6.0	150	3.10(c)	[135]
water + n-methylacetamide	313 – 373	0.05 – 69	0.00 – 1.00	47	16	1.8	3.15(b)	[170, 171]

System	T range	P range			%AAD	AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	T_{LLE}	T_{LLE} / K	Figures	Ref.
n-octane + n-methylpropanamide	310 – 361	101.33	0.04 – 0.70	8	5.6	20.	3.14(c)	[142]

System	T range	P range			%AAD	AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	H^{E}	H^{E} / (J mol^{-1})	Figures	Ref.
2-propanol + 2-propanamine	298	101.33	0.10 – 0.90	9	5.4	85	—	[120]
2-propanol + 1-butanamine	298	101.33	0.00 – 0.90	137	19	290	—	[121]
2-propanol + 2-butanamine	298	101.33	0.01 – 0.99	11	24	220	—	[121]
2-butanol + 1-butanamine	298	101.33	0.1 – 0.95	11	14	170	—	[118]
n-hexane + 2-butanamine	298	101.33	0.06 – 0.89	9	12	66	3.10(e)	[137]
n-heptane + 2-butanamine	298	101.33	0.05 – 0.95	19	22	160	3.10(f)	[138]
n-octane + n-methylacetamide	398	1891	0.00 – 1.00	5	51	102	3.14(c)	[159]
n-decane + n-methylacetamide	413	1617	0.00 – 1.00	2	91	101	3.14(c)	[159]
water + n-methylacetamide	323 – 398	101.33	0.01 – 0.95	43	15	104	3.15(c)	[168]
water + n-ethylacetamide	308	101.33	0.02 – 0.98	18	14	111	3.15(c)	[169]
water + n-methylpropanamide	308	101.33	0.03 – 0.98	17	16	110	3.15(c)	[169]
1-hexanamine + n-methylacetamide	363	1203	0.03 – 0.96	20	15	38	3.16(a)	[168]
propanpoic acid + n-methylacetamide	363	1135	0.03 – 0.98	21	31	125	3.17(a)	[168]

System	T range	P range			AAD		
(1 + 2)	/ K	/ kPa	x_2 range	N^{D}	x_2^{SLE}	Figures	Ref.
2-propanol + octanoic acid	272 – 283	101.33	0.50 – 0.80	2	0.04	—	[181]

2-propanol + nonanoic acid	273 – 283	101.33	0.60 – 0.90	2	0.037	—	[181]
2-propanol + decanoic acid	273 – 303	101.33	0.20 – 0.95	4	0.023	—	[181]
2-propanol + undecanoic acid	273 – 293	101.33	0.20 – 0.60	3	0.029	—	[181]
2-propanol + dodecanoic acid	273 – 318	101.33	0.05 – 1.00	31	0.013	—	[181, 182, 183]
2-propanol + tridecanoic acid	273 – 313	101.33	0.05 – 0.95	5	0.020	—	[181]
2-propanol + tetradecanoic acid	273 – 323	101.33	0.00 – 0.80	6	0.017	—	[181]
2-propanol + pentadecanoic acid	273 – 323	101.33	0.02 – 0.80	6	0.014	—	[181]
2-propanol + hexadecanoic acid	273 – 336	101.33	0.00 – 1.00	28	0.044	—	[181, 184, 183]
2-propanol + heptadecanoic acid	273 – 333	101.33	0.00 – 0.94	7	0.018	—	[181]
2-propanol + octadecanoic acid	283 – 344	101.33	0.00 – 1.00	42	0.042	—	[181, 185, 186, 183]

3.5 Conclusions

This chapter presents a detailed development of the SAFT-γ Mie group-contribution (GC) approach, with emphasis on the Helmholtz free-energy expressions used to model both neutral and charged molecular systems. Particular attention is given to the treatment of electrostatic interactions, comparing two approaches used for representing the charges on ionic chain molecules. Following the theoretical development, the SAFT-γ Mie GC framework is applied to amino acids and oligopeptides. The relevant functional groups required for modelling these systems are identified, and the parameter optimisation process is outlined. A comprehensive set of like and unlike interaction parameters is presented, including those for the CONH (peptide) group. The accuracy of the approach is demonstrated through comparisons with fluid-phase experimental data for a wide range of thermodynamic properties. These include saturated and liquid densities, vapour pressures, and various phase equilibria (VLE, VLLE, LLE, and SLE), as well as excess enthalpies for binary mixtures involving diverse compound classes such as water, alcohols, amines, amides, acids, and ureas. In total, 11 new interactions are introduced and validated against data from 11 pure compounds and 43 binary mixtures, demonstrating the transferability and broad applicability of the SAFT-γ Mie GC framework. Overall, the models lead to satisfactory agreement with experimental data of the different systems considered. The deviations reported are to be expected when using a group-contribution model especially when using the same set of parameters to describe different pure-component and mixture properties of various systems. In predicting the solid–liquid equilibria of amino acids and oligopeptides, other factors such as the reliability and availability of melting-property data, the omission of electrostatic contribution (specifically for calculations at the isoelectric point), and the omission of intramolecular hydrogen-bonding interactions are expected to outweigh the deviations obtained during the parameter optimisation process.

In the following chapter, the SAFT-γ Mie GC framework is extended to model the solid–liquid solubility of amino acids under changing pH. This extension involves the modelling of chemical equilibria, in addition to solid–liquid equilibrium, which produce charged species in

solution. This requires the treatment of the like and unlike Coulombic interactions of charged species.

Chapter 4

Prediction of solid–liquid solubility of amino acids and oligopeptides

4.1 Introduction

The goal of this chapter is to assess the performance of the SAFT-γ Mie GC approach in predicting the solid–liquid solubility of amino acids and oligopeptides in various solvent media and across a range of thermodynamic conditions. To achieve this goal, one needs to solve the solid–liquid equilibrium (SLE) equation, detailed in Section 2.5.1. Equation 2.19 requires as input the activity coefficient of the solute in solution $\gamma_i(T, P, \mathbf{x})$ as well as the melting properties of the pure solute, namely the enthalpy of fusion Δh_i^{fus}, the melting temperature T_i^{fus}, and the difference in the molar, constant-pressure heat capacity between the solid and liquid phases $\Delta c_{p,i}^{\text{SL}}$. The calculation of activity coefficients depends on the availability of SAFT-γ Mie parameters describing the intermolecular interactions between groups making up the molecules of the system; these parameters have been obtained by adjusting to experimental thermodynamic data of pure compounds and mixtures, as detailed in the previous chapter. Provided the melting properties of the pure solute (the amino acids and oligopeptides here) are known, predictive solubility calculations can be carried out if the groups are available. A major challenge in predicting the solubility of amino acids and oligopeptides is the scarcity

of reliable melting property data. Amino acids and oligopeptides undergo decomposition well below their melting temperatures [187], hence specialised techniques are needed to measure their melting properties. A more detailed discussion of this challenge and the sensitivity of solubility calculations to the melting properties is presented in Section 4.2. In terms of assessing model performance, there is a scarcity in the solubility data of amino acids and oligopeptides that are essential for validating solubility predictions. Thus, the availability of experimental data on the melting properties and solubility were the limiting factors taken into consideration when choosing the systems to be modeled in this work. Indeed, other factors have also contributed to this consideration, such as the availability of experimental thermodynamic data necessary for optimising interaction parameters.

In the first part of this chapter, predictions of the solid–liquid solubilities of amino acids and oligopeptides at the pI of the solution are presented and discussed. In these systems, the amino acids and oligopeptides exist solely as zwitterions; they do not ionize into their cationic or anionic counterparts due to the absence of pH-modifying agents (i.e. acids and bases). Thus, they are modeled with neutral groups for simplicity; the NH_2 and COOH groups are used, as shown in Figure 3.2 in the previous chapter. In this first part, the solubilities of the amino acids glycine, alanine, valine, and serine are predicted in pure and binary mixtures of water and alcohol across a wide temperature range. Additionally, the aqueous solubilities of 17 di- and tri-peptides comprising the amino acids glycine, alanine, leucine, and serine are also predicted.

In the second part of this chapter, the solid–liquid solubilities of the amino acids glycine and alanine are predicted as a function of pH. These multicomponent systems are composed of: the amino acid zwitterions along with their ionized cationic and anionic species; water along with the hydronium and hydroxide ions, which are produced in water dissociation reactions; and counterions produced from the full dissociation of the pH-modifying strong acid and strong base. Glycine and alanine are chosen as example amino acids due to the availability of their solubility data at variable pH. In principle, the same modeling framework can be extended to other amino acids, pending the availability of the necessary interaction parameters.

The chapter begins with a discussion of the sensitivity of solid–liquid solubility calculations

to the melting properties in Section 4.2. In Section 4.3, predictions of the solubility of amino acids and oligopeptides at their pI are discussed and compared with experimental data. In Section 4.4, the speciation of amino acids is discussed and the predictions of pH-dependent solubility are presented and compared to experimental data. Finally, the chapter ends with conclusions in Section 4.5.

4.2 Sensitivity of solubility calculations to melting-property data

Before presenting the predictions of solid–liquid solubility of amino acids and oligopeptides, it is important to consider the dependence of the solubility equation (Eq. 2.19) on the melting properties of the solute, namely the enthalpy of fusion (Δh_i^{fus}) and the melting temperature (T_i^{fus}). More specifically, it is important to understand how these measurements were conducted and how their reported uncertainties influence the solubility calculations.

Experimentalists face unique challenges when measuring the melting properties of amino acids and oligopeptides. Amino acids have high melting points, ranging from about 500 K to 600 K, likely due to the strong ionic interactions between the zwitterions in the solid phase. Oligopeptides have even higher melting temperatures due to more complex intermolecular interactions and higher molecular weights. Unfortunately, amino acids and oligopeptides undergo decomposition before melting when subjected to the heating rates of traditional methods, such as differential scanning calorimetry (DSC). The challenge faced by experimentalists is demonstrable in the huge variability in reported melting properties in the literature. For example, the melting point of glycine is reported to be as low as 531 K [188] and as high as 569 K [189]. The same variability is observed in the enthalpy of fusion of glycine, which is reported to be as low as 21 kJ/mol [189] and as high as 43 kJ/mol [188].

In the current thesis, the melting properties reported by Held and coworkers [190, 191, 192, 189] are used, and can be found in Table 4.1. This choice was made for two main reasons: first,

the group of authors used fast-scanning calorimetry (FSC) to circumvent the problem of decomposition; and second, they reported the melting properties of a large number of amino acids and oligopeptides along with their associated experimental uncertainties. Using experimental melting-property data from the same source greatly simplifies the task of solubility modeling, although it does not completely eliminate the inherent uncertainty in the measurements. For example, glycine is reported to have a melting temperature of 569 K with an uncertainty of ± 9 K and an enthalpy of fusion of 22 kJ/mol with an uncertainty of ± 3 kJ/mol. Figure 4.1 illustrates the sensitivity of solubility calculations to the enthalpy of fusion and melting temperature. As can be seen in Figures 4.1, the uncertainty in the fusion enthalpy has a much greater impact on the calculated solubility, especially in dilute concentrations below the boiling point of water (373 K), than the melting temperature. At 298 K, the uncertainty in the melting temperature results in an uncertainty in the solubility of glycine in water of about 0.0033 (in mole fraction), while the uncertainty in the enthalpy of fusion results in an uncertainty of about 0.026, which is higher by an order of magnitude. It is important to note that in this work the difference between the molar heat capacity of solid and liquid phases ($\Delta c_{\mathrm{p,i}}^{\mathrm{SL}}$) is used in modelling systems for which it is experimentally available, which is only for oligopeptides; i.e, $\Delta c_{\mathrm{p,i}}^{\mathrm{SL}}$ is not used.

4.3 Solid–liquid solubility at the isoelecric point pI

In this section, the solid–liquid solubility is predicted for the amino acids glycine, alanine, valine, and serine in pure solvents, namely water and primary and secondary alcohols, and in mixed binary water + alcohol solvents. Moreover, aqueous solubility is predicted for 17 di- and tri-peptides made up of the amino acids glycine, alanine, leucine, and serine. Because the amino acids and oligopeptides do not ionize in these systems, their constituent groups are treated as fully neutral to simplify the task of predicting their solubility.

Table 4.1: Melting properties of the amino acids and peptides used to calculate their solubility, reported with their experimental uncertainties; melting-point temperature (T_i^{fus}), enthalpy of fusion (Δh_i^{fus}), and difference between the molar heat capacity of liquid and solid phases evaluated at T_i^{fus} ($\Delta c_{\text{p,i}}^{\text{SL}}$).

Solute	T_i^{fus}/K	Δh_i^{fus}/kJmol^{-1}	$\Delta c_{\text{p,i}}^{\text{SL}}$/Jmol^{-1}K^{-1}	Ref.
glycine	569 ± 9	22 ± 3	—	[189]
alanine	608 ± 9	23 ± 3	—	[189]
valine	529 ± 7	44 ± 6	—	[191]
leucine	518 ± 8	43 ± 5	—	[191]
serine	519 ± 7	28 ± 3	—	[191]
gly-gly	593 ± 7	40 ± 6	51 ± 6	[190]
gly-gly-gly	594 ± 7	54 ± 7	57 ± 15	[192]
ala-ala	606 ± 7	54 ± 7	62 ± 18	[190]
ala-ala-ala	606 ± 7	72 ± 9	124 ± 8	[192]
gly-ala	551 ± 7	41 ± 5	55 ± 6	[190]
ala-gly	611 ± 7	52 ± 7	57 ± 3	[190]
gly-gly-ala	592 ± 10	70 ± 8	66 ± 7	[192]
gly-ala-gly	623 ± 7	61 ± 7	78 ± 11	[192]
ala-gly-ala	557 ± 8	58 ± 7	98 ± 5	[192]
leu-gly-gly	530 ± 7	74 ± 8	111 ± 11	[192]
gly-leu-gly	545 ± 7	60 ± 7	139 ± 11	[192]
gly-gly-leu	521 ± 7	55 ± 7	160 ± 7	[192]
gly-ala-leu	578 ± 7	77 ± 9	112 ± 6	[192]
gly-ser	530 ± 8	49 ± 6	67 ± 6	[192]
ser-gly	553 ± 7	62 ± 7	61 ± 9	[192]
ala-ser	556 ± 7	43 ± 5	48 ± 6	[192]
ser-ala	609 ± 7	73 ± 8	55 ± 3	[192]

4.3.1 Models

As discussed in Section 3.4, to model the amino acid and oligopeptide solutions, a number of SAFT-γ Mie like and unlike interaction parameters are needed. The amino acids glycine, alanine, valine, and serine are made up of the groups NH_2, COOH, CH, CH_2, CH_3, and CH_2OH, as shown in Figure 3.2. When modelling the oligopeptide solutions, the inclusion of the peptide (CONH) group is required, in addition to the amino-acid groups. Water is modelled with the group H_2O and the alcohol solvents are modelled with the groups CH_3, CH_2, CH, CH_2OH, and CHOH. The SAFT-γ Mie interaction parameters for these groups can be found in Tables 3.1, 3.2, and 3.3.

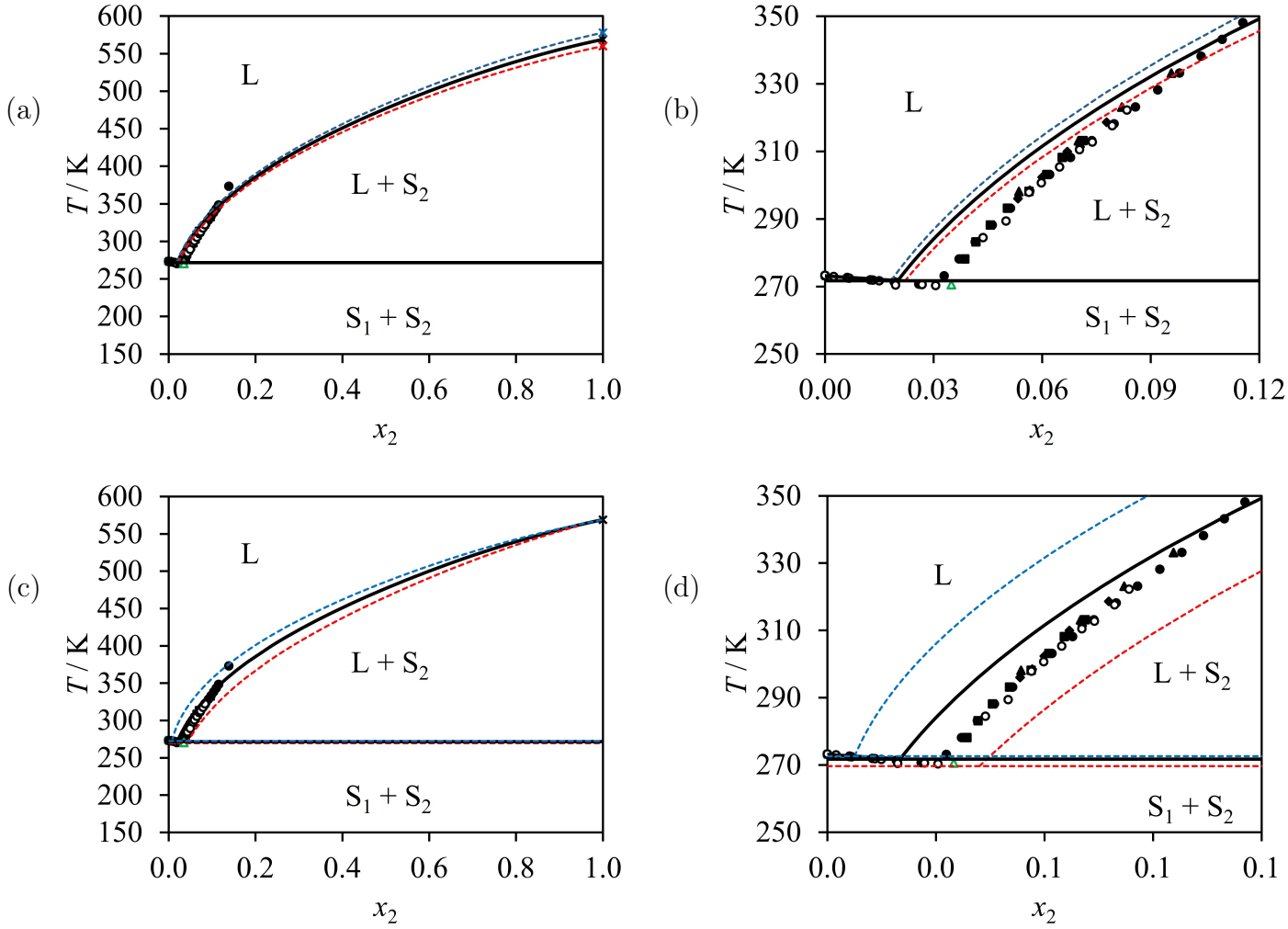


Figure 4.1: Effect of uncertainty in melting temperature and heat of fusion on the SAFT-γ Mie description of the solid–liquid equilibrium of glycine (2) in water (1) at 1 bar. (a) The sensitivity of solubility calculations to the melting temperature for a fixed enthalpy of fusion value of 22 kJ/mol. The black curve represents calculations using the reported melting temperature (569 K), whereas blue and red curves denote calculations using the maximum (578 K) and minimum (560 K) of the uncertainty range, respectively. (b) A magnified version of (a) showing the region with available glycine solubility data. (c) The sensitivity of solubility calculations to the enthalpy of fusion for a fixed melting temperature of 569 K; the black curve represents calculations using the reported enthalpy of fusion (22 kJ mol^{-1}) whereas blue and red curves denote calculations using the maximum (25 kJ mol^{-1}) and minimum (19 kJ mol^{-1}) of the uncertainty range, respectively. (d) A magnified version of (c) showing the region with available glycine solubility data. Black filled symbols:circles [111]; triangles [30]; squares [115]; and diamonds [112] denote data used in optimising the interaction COOH-NH$_2$, and black empty symbols [116] denote data that have not been used in the optimisation. The green triangle [117] denotes the eutectic of the mixture.

4.3.2 Solubility in pure solvents

The SAFT-γ Mie approach is used here to predict the solubility of the amino acids glycine, alanine, valine, and serine in water and alcohols, and of the di- and tripeptides — comprising

these four amino acids — in water.

The SLE diagrams of the aqueous mixtures of glycine, alanine, serine, and valine can be seen in Figures 4.1 - 4.4. Except for glycine (cf. Section 3.4.1), the activity coefficients are entirely predicted; the solubility of glycine in water was used in characterising the interaction NH_2-COOH. The solubility of alanine in water is predicted to be in overall good agreement with experiments over the entire temperature range measured, as illustrated in Figure 4.2, but the eutectic point is underpredicted compared to the sole experimental data point available. The prediction of serine solubility in water is accurate at lower temperatures but the deviation gets gradually larger as the temperature increases. This is likely due to the presence of the hydrogen-bonding group CH_2OH which strongly interacts with the water group H_2O leading to an overprediction of solubility at higher temperatures. Valine, shown in Figure 4.3, contains more hydrophobic groups than the other amino acids considered in this work, which leads to a vast underprediction in solubility as the hydrophobic groups interact more strongly with each other than they do with the water group. It is also important to note that the different experimental sources report significantly different values of both the solubility of valine and its eutectic point. The model, however, underpredicts both by orders of magnitude and predicts an extremely narrow eutectic region that is not visible in the phase diagram. The temperature dependence of the predicted solubility of valine in water is in satisfactory agreement with the experimental data, but the predicted values are visibly lower than those measured (e.g., $x_{\text{valine}} = 0.0001$ is predicted at 298 K, while the measured value is 0.0108).

Although the accuracy of the model could be improved by treating the melting properties as adjustable parameters (as in the works [38, 41, 44, 45, 46, 47, 48]), or by using some of these solubility data to refine the group parameters, we consider the current results satisfactory, and use the models presented to predict the solubility in other solvents and to study di, and tripeptides. We are interested in considering a fully predictive approach at this point, and in assessing the merits of standard, neutral, groups to treat these solutions, neglecting in addition any speciation. Moreover, the uncertainty inherent in the measurement of the melting properties of amino acids and peptides (cf. Section 4.2), means that using solubility data to estimate molecular model parameters may lead to unexpected biasing of the molecular model

developed.

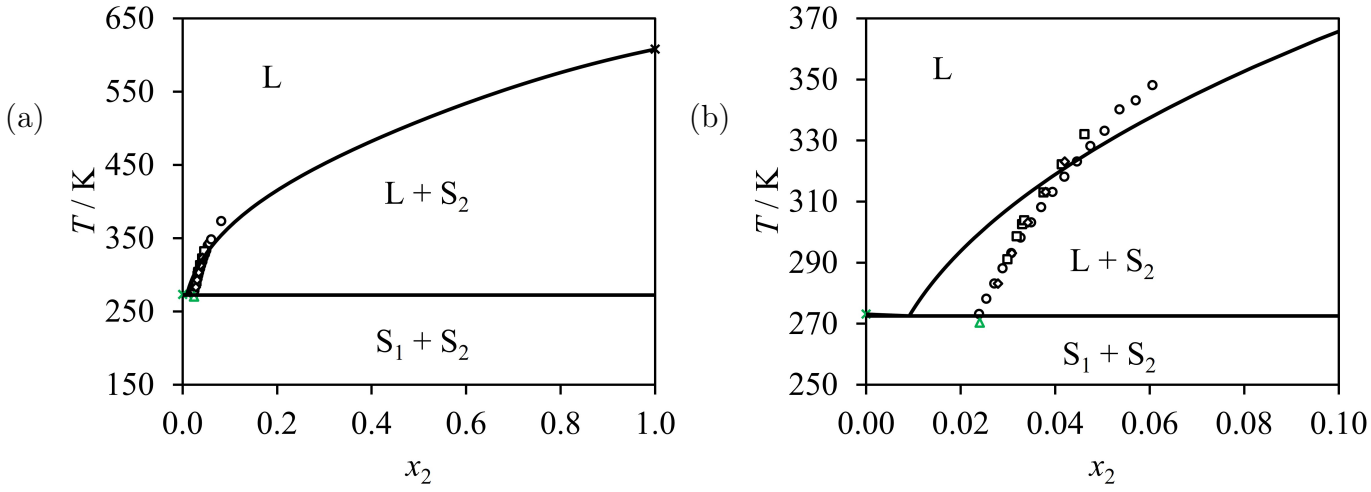


Figure 4.2: SAFT-γ Solid–liquid solubility prediction of alanine (2) in water (1) at 1 bar. (a) Full concentration range depicting the SLE up to the melting point, denoted by a black cross symbol. (b) Low alanine mole-fraction region illustrating the eutectic region of the mixture. The green cross denotes the melting point of water [193] and the green triangle denotes the eutectic point [194]. The empty black symbols correspond to solubility experimental data from various sources: circles [111]; squares [115]; and diamonds [195].

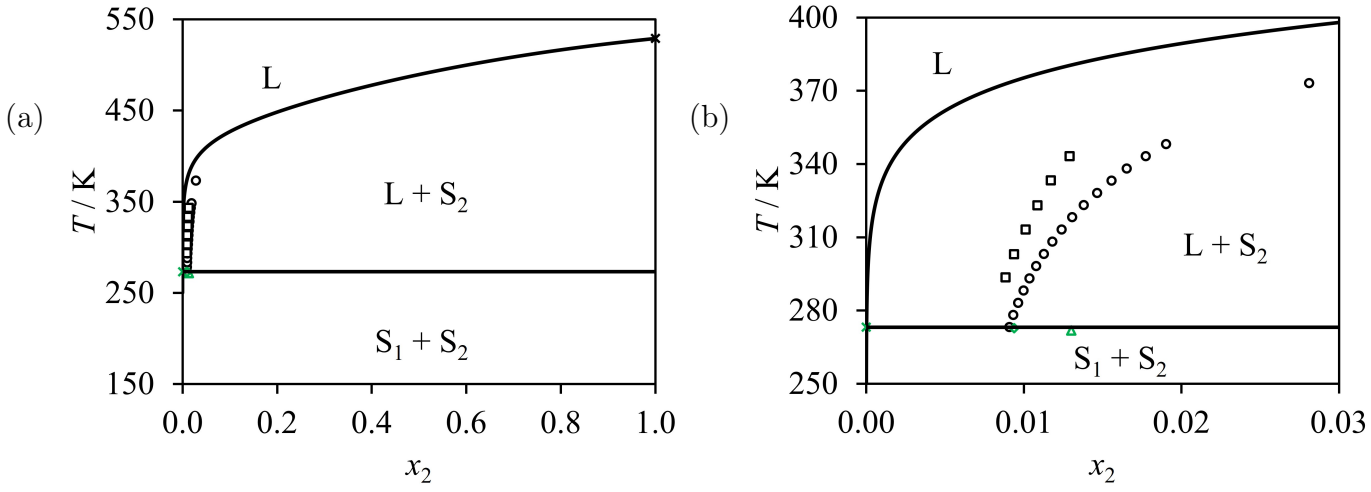


Figure 4.3: SAFT-γ Solid–liquid solubility prediction of valine (2) in water (1) at 1 bar. (a) Full concentration range depicting the SLE up to the melting point, denoted by a black cross symbol. (b) Low valine mole-fraction region. The green cross denotes the melting point of water [193] and the green triangle [194] and green diamond [117] denote experimental measurements of the eutectic point. The empty black symbols correspond to solubility experimental data from different sources: circles [111]; and squares [115].

The predictive capability of the model is now assessed by calculating the solubilities of glycine, alanine, valine, and serine, in various primary (ethanol [196, 199, 30, 200, 17, 201, 202, 203, 195, 197, 198], 1-propanol [202, 196, 30, 29, 195], 1-butanol [196, 31, 18, 195, 197], and

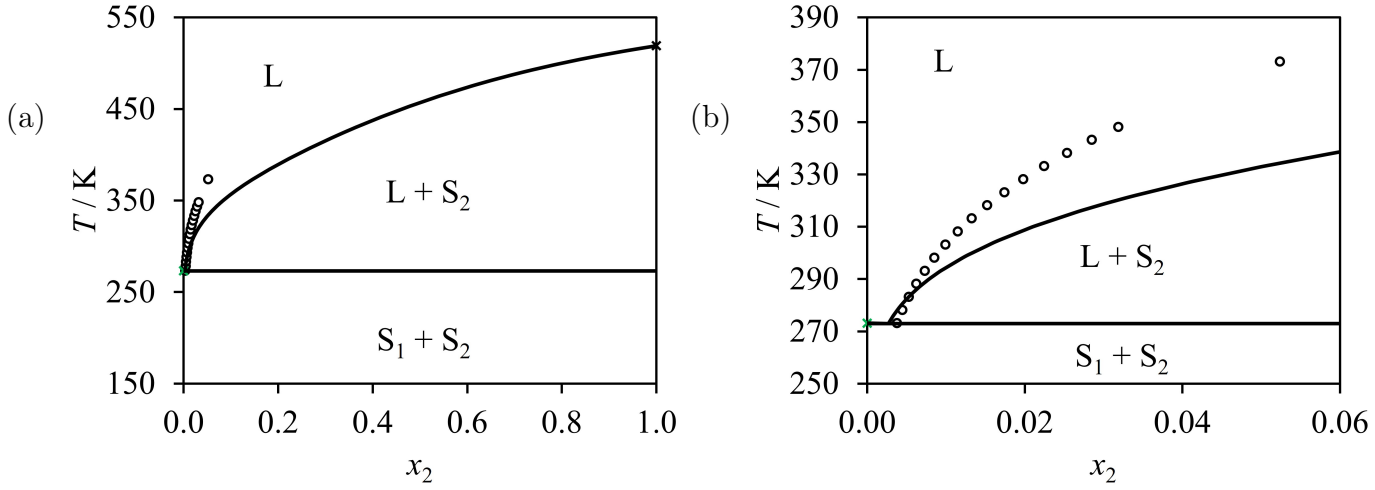


Figure 4.4: SAFT-γ Solid–liquid solubility prediction of serine (2) in water (1) at 1 bar. (a) Full concentration range depicting the SLE up to the melting point, denoted by a black cross symbol. (b) Low serine mole-fraction region illustrating the eutectic region of the mixture. The green cross denotes the melting point of water [193] and the empty black circles [111] correspond to solubility experimental data.

2-methyl-1-propanol [195]) and secondary alcohols (2-propanol [202, 196, 204, 30, 29, 197, 198] and 2-butanol [195]). The predictions are presented in Figure 4.5 as a parity plot against the experimental data available. The aqueous solubility calculations and data shown in Figures 4.1 – 4.4 are also included for completeness, and AADs for each of the systems considered are presented in Table 4.2. Given the very low solubility values of some of the systems, these results confirm the predictive capability of the method and validate the use of neutral models as proposed here. It can be seen that the model performs best for the prediction of solubility in water, and that deviations increase as the order of magnitude of solubility becomes smaller, as is the case in alcohols. It is, however, worth noting that the solubility measurements of amino acids in alcohols reported in the literature vary significantly depending on the source. One clear example is the solubility of glycine in 2-propanol. The prediction is in very good agreement with the data reported by Bouchard et al. [204], but off by two orders of magnitude when compared to the data of Abraham et al. [196], even though in both works, the solubilities were measured using the gravimetric method. A similar observation can be made for glycine in ethanol, alanine in ethanol, and alanine in 1-butanol.

Having assessed the performance of the SAFT-γ Mie model for the prediction of solubility

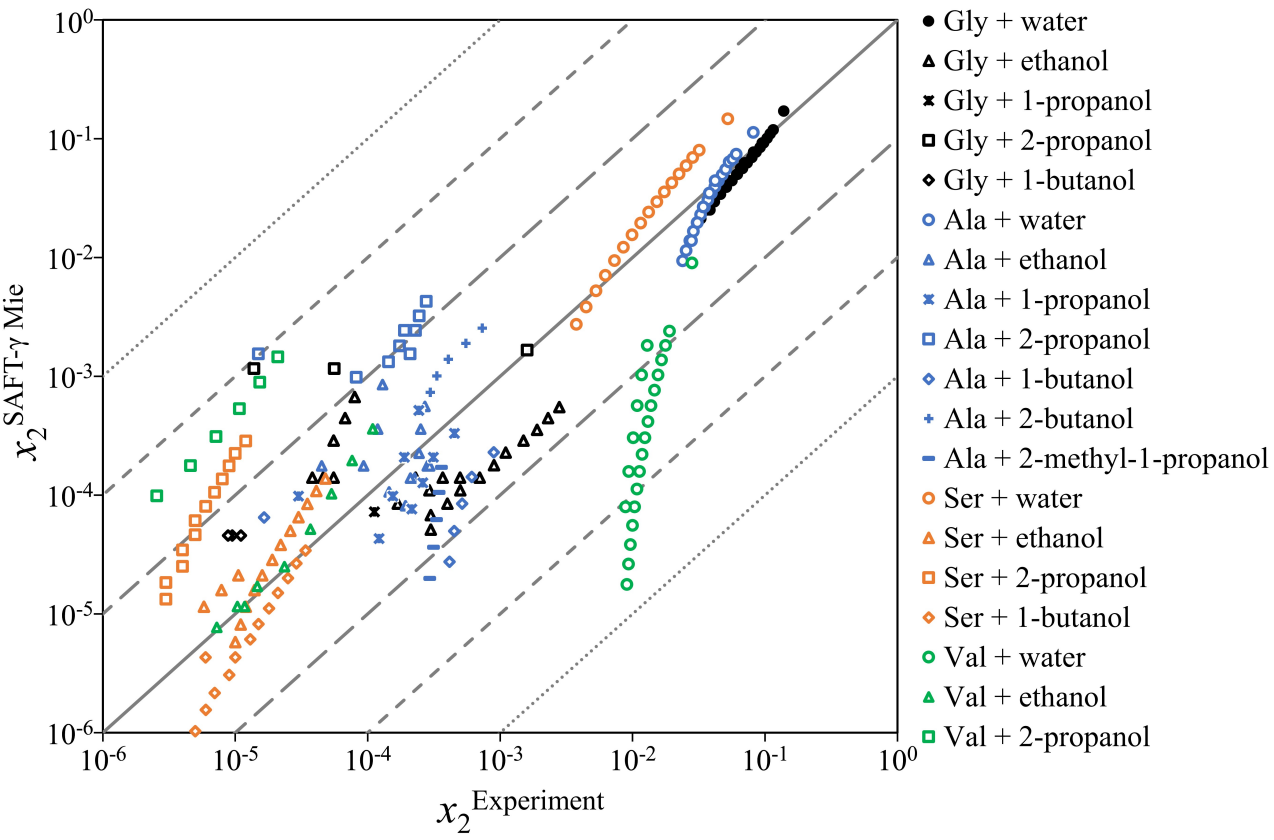


Figure 4.5: Parity plot of the solubility of glycine, alanine, valine, and serine in water, ethanol, 1-propanol, 1-butanol, 2-methyl-1-propanol, 2-propanol, and 2-butanol. The solid diagonal denotes exact agreement between experiments and calculations, whereas each pair of dashed and dotted lines denotes a change in order of magnitude. Sources of the experimental data can be found in Table 4.2.

of amino acids when treated as uncharged unspeciated mixtures, we consider now the prediction of solubility in peptide solutions. The predicted aqueous solubility of 17 di- and tri-peptides containing glycine (Gly), alanine (Ala), leucine (Leu), and serine (Ser), is compared to experimental data in a parity plot in Figure 4.6. The corresponding melting temperatures and enthalpies of fusion of each of the peptides are reported in Table 4.1. We note that, as in the case of amino acid solubility, the solubility calculations for peptides are highly sensitive to uncertainty in the enthalpy of fusion, which unfortunately are reported with larger uncertainty ranges than those of the amino acids. We use the actual values reported in Do et al. [190, 192] in all of our calculations. We find the most accurate predictions are those for glycine homopeptides (Gly-Gly and Gly-Gly-Gly), and we find that with few exceptions, the predicted values are within an order of magnitude of the experimental values which, given the very low solubility of these larger compounds, is a promising result. Furthermore, it is interesting to note

that our model leads either to very good agreement with experiments, or to an overprediction; in none of the cases are the solubilities underpredicted.

Moreover, the solubility predictions of the 17 oligopeptides use the difference in molar heat capacity between the liquid and solid phases ($\Delta c_{p,i}$) as reported in Table 4.1. There is an argument against using the heat-capacity terms in the solubility equation due to the fact that the term was not used in optimising the interaction parameters of COOH-NH_2, which is present in all the oligopeptides. The impact of not using the heat-capacity terms leads to the solubility predictions in Figure 4.7. In this case, the model still leads to satisfactory agreement with experiments, albeit exhibiting both under- and overpredictions of solubility.

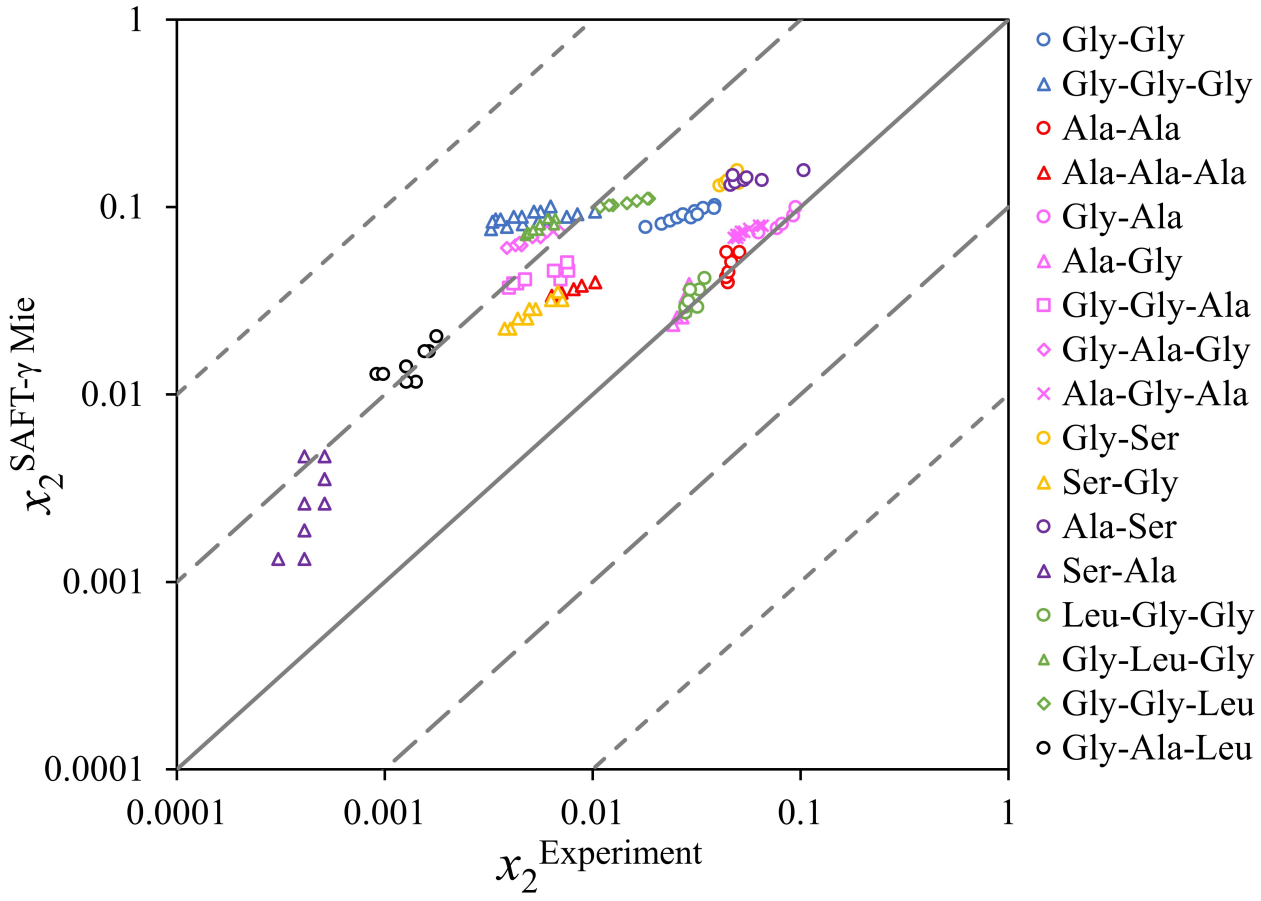


Figure 4.6: Parity plot of the solubility in water of di- and tri-peptides made up of residues of the amino acids glycine (Gly), alanine (Ala), leucine (Leu), and serine (Ser). The solid diagonal denotes exact agreement between experiments and calculations, whereas each pair of dashed lines denotes deviation of an increasing order of magnitude. Sources of the experimental data can be found in Table 4.3.

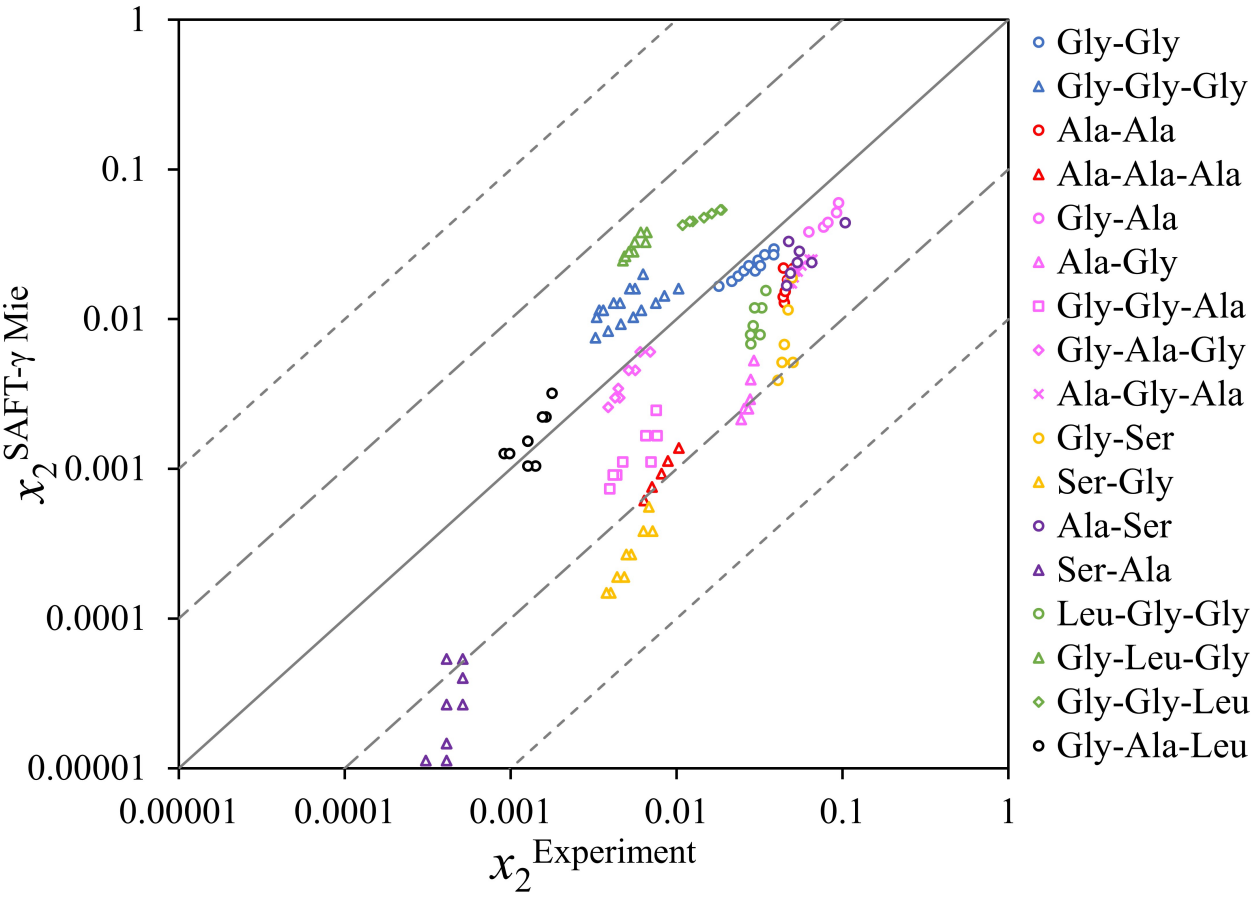


Figure 4.7: Parity plot of the solubility in water of di- and tri-peptides made up of residues of the amino acids glycine (Gly), alanine (Ala), leucine (Leu), and serine (Ser). Calculations do not include the term $\Delta c_{p,i}$; i.e., it is set to 0 in the solubility calculations. The solid diagonal denotes exact agreement between experiments and calculations, whereas each pair of dashed lines denotes deviation of an increasing order of magnitude. Sources of the experimental data can be found in Table 4.3.

4.3.3 Solubility in mixed-solvent media

The modelling framework is extended to predict the solubility of amino acids in binary water + alcohol solvents. In principle, the same equation (Eq 2.19) used to predict the solubility in pure solvents is used. The additional solvent is taken into account through the activity coefficient of the amino acid, which is a function of the temperature, pressure, and the mixture composition, and through the material balance equation. Here, we predict the solubilities of glycine, alanine, valine, and serine in aqueous solutions of ethanol, 1-propanol, 1-butanol, 2-propanol, and 2-butanol. Those mixtures are chosen based on the availability of experimental data.

Figures 4.8 - 4.10 illustrate the solubility of the four amino acids in different aqueous alcohol mixtures. The solubility of the amino acid is represented in the y-axis in mole fraction x_2 of the ternary mixture while the composition of the solvent is represented in the x-axis in solute-free mole fraction $x_3^{'}$ of the alcohol solvent. The latter is calculated as

$$x_3^{'} = \frac{x_3}{x_1 + x_3}, \tag{4.1}$$

where x_1 and x_3 are the mole fractions of the solvents.

The solubility of glycine in water + ethanol mixtures is predicted with good agreement with experiments at low ethanol concentrations and high temperatures. As the concentration of the ethanol solvent increases, the measured solubility of glycine drops significantly and while the model captures this steep decrease in solubility, it overpredicts the solubility near the pure-ethanol region. A similar observation can be made for glycine solubility in water + 1-propanol (Figure 4.8(b)), water + 1-butanol (Figure 4.8(c)), and water + 2-propanol (Figure 4.8(d)).

The model also predicts a slight increase in solubility at low ethanol, 1-propanol, and 1-butanol concentrations before the steep decrease. This effect is intensified by increasing the temperature, as illustrated in Figure 4.8(b); the down-facing curvature increases as the temperature increases from 283.15 K (light blue) to 333.15 K (black). This effect is also intensified as the alcohol molecule increases in size; at a fixed temperature of 298.15 K, the curvature goes from almost non-existent in water + ethanol (blue curve in Figure 4.8(a)) to slightly more pronounced in water + 1-propanol (purple curve in Figure 4.8(b)) to even more pronounced in water + 1-butanol (curve in Figure 4.8(c)). This maximum solubility is not observed anywhere in the data and is most likely due to the model parameters describing the unlike interactions between the water and alcohol. At higher temperatures, the association strength is reduced due to the greater thermal energy disrupting hydrogen bonds and, in longer-chain alcohol mixtures, the presence of additional hydrophobic groups "dilutes" the association interactions in the mixture. This reduction in association between water molecules reduces the energetic cost of creating solute-sized voids, leading to enhanced solubilities. As the concentration of the alcohol increases, the solubility drops due to the dominant interaction between the solute and alcohol

molecules, which is inherently less favorable than those between the solute and water molecules. This is supported by the fact that alanine exhibits a more pronounced maximum solubility peak than glycine as illustrated in Figure 4.9, since alanine contains one more hydrophobic group (CH_3) than glycine. This is also supported by the much larger maximum solubility that valine exhibits in water + ethanol, as illustrated in Figure 4.10(a), owing to the fact that valine has an isopropyl side group, which is larger and renders the solute more hydrophobic. Following this logic, the predicted solubilities of serine, which contains a CH_2OH group, in water + alcohol mixtures should be expected to show little, if any, maximum-solubility effects. This is indeed what the model predicts, as illustrated in Figures 4.10(b-d).

All amino acids considered have solubilities that are orders of magnitude lower in pure alcohol solvents than in pure water. This difference in solubility is well predicted by the model. The model yields qualitative predictions of solubility in mixed aqueous-alcohol solvents for all the amino acids considered and could be improved further by introducing mixture-dependent binary interaction parameters.

Overall, the results presented suggest that treating amino acids and oligopeptides as neutral species can lead to satisfactory predictions of solubility in pure water and alcohols and in mixed aqueous alcohol solvents. The assumption of neutrality is useful when modelling systems at the isoelectric point, without the need to account for the zwitterionic nature of the amino acid or the peptide and their speciation, thus reducing the number of species (and equations) that need to be accounted for in the model. However, to model the behaviour of amino acids or peptides at pH values different from the pI, it becomes essential to account explicitly for their zwitterionic nature. We consider these models in the next chapter.

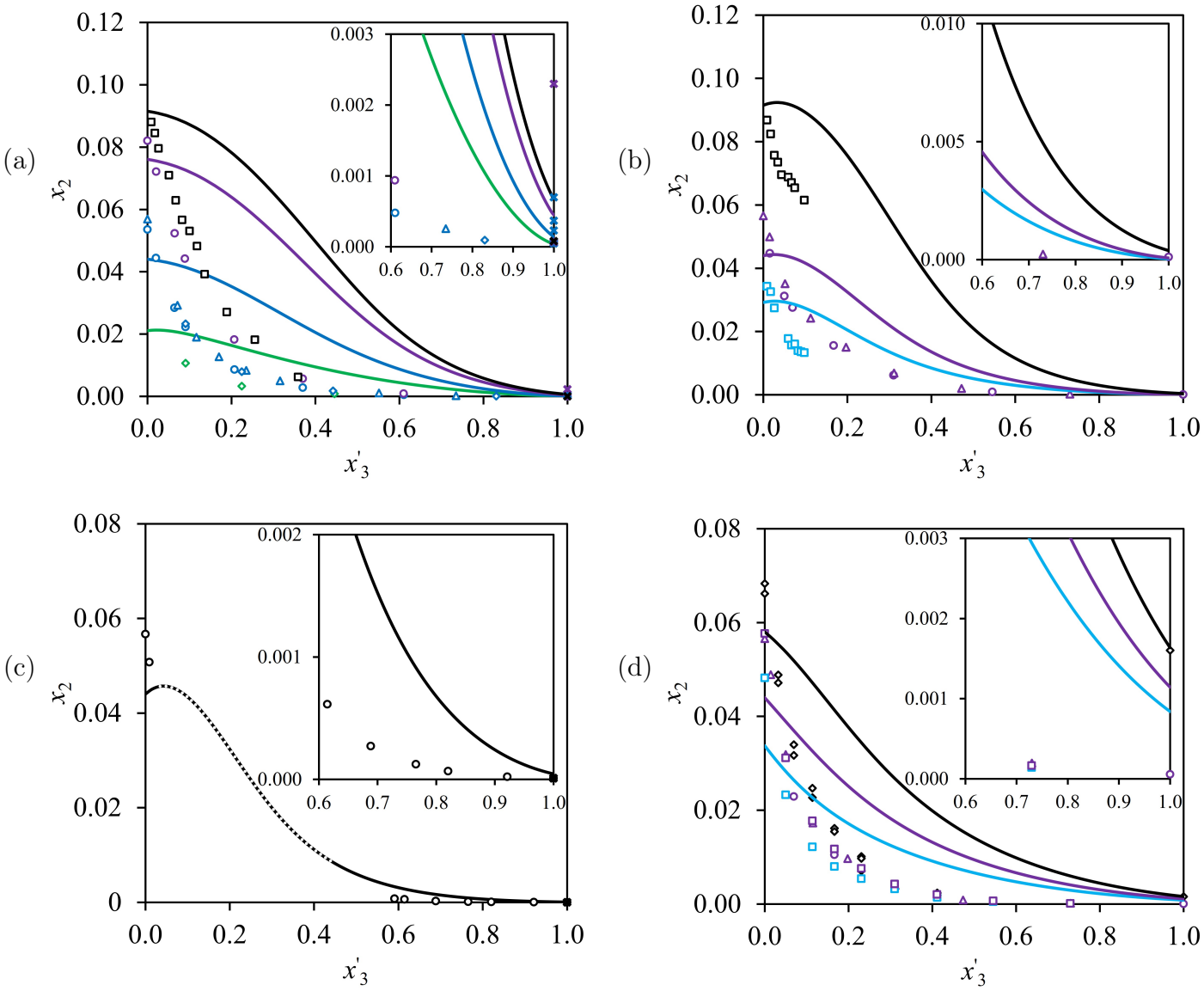


Figure 4.8: SAFT-γ Mie description of the solid–liquid equilibria of glycine in mixed aqueous alcohol solvents expressed as the mole fraction of glycine as a function of solute-free alcohol mole fraction. (a) Solubility of glycine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (green); 298.15 K (blue); 323.15 K (purple); and 333.15 K (black). Different empty symbols denote different sources: circles [202]; triangles [205]; diamonds [206]; and squares [207]. "x" symbols denote pure-ethanol solubility [199, 30, 200, 17, 203]. (b) Solubility of glycine (2) in water (1) + 1-propanol (3) at 1 bar and: 283.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Symbols denote experiments from different sources: circles [202]; triangles [29]; and squares [207]. (c) Solubility of glycine (2) in water (1) + 1-butanol (3) at 1 bar and 298.15 K. Empty circles denote mixed-solvent data [208] and "x" symbols denote pure-1-butanol solubilities from different sources [196, 18]. The dotted section of the curve represents the region of liquid–liquid immiscibility of the water/1-butanol mixture. (d) Solubility of glycine (2) in water (1) + 2-propanol (3) at 1 bar and: 288.15 K (light blue); 298.15 K (purple); and 310 K (black). Empty symbols denote experiments from different sources: circles [202]; triangles [29]; diamonds [204]; and squares [209].

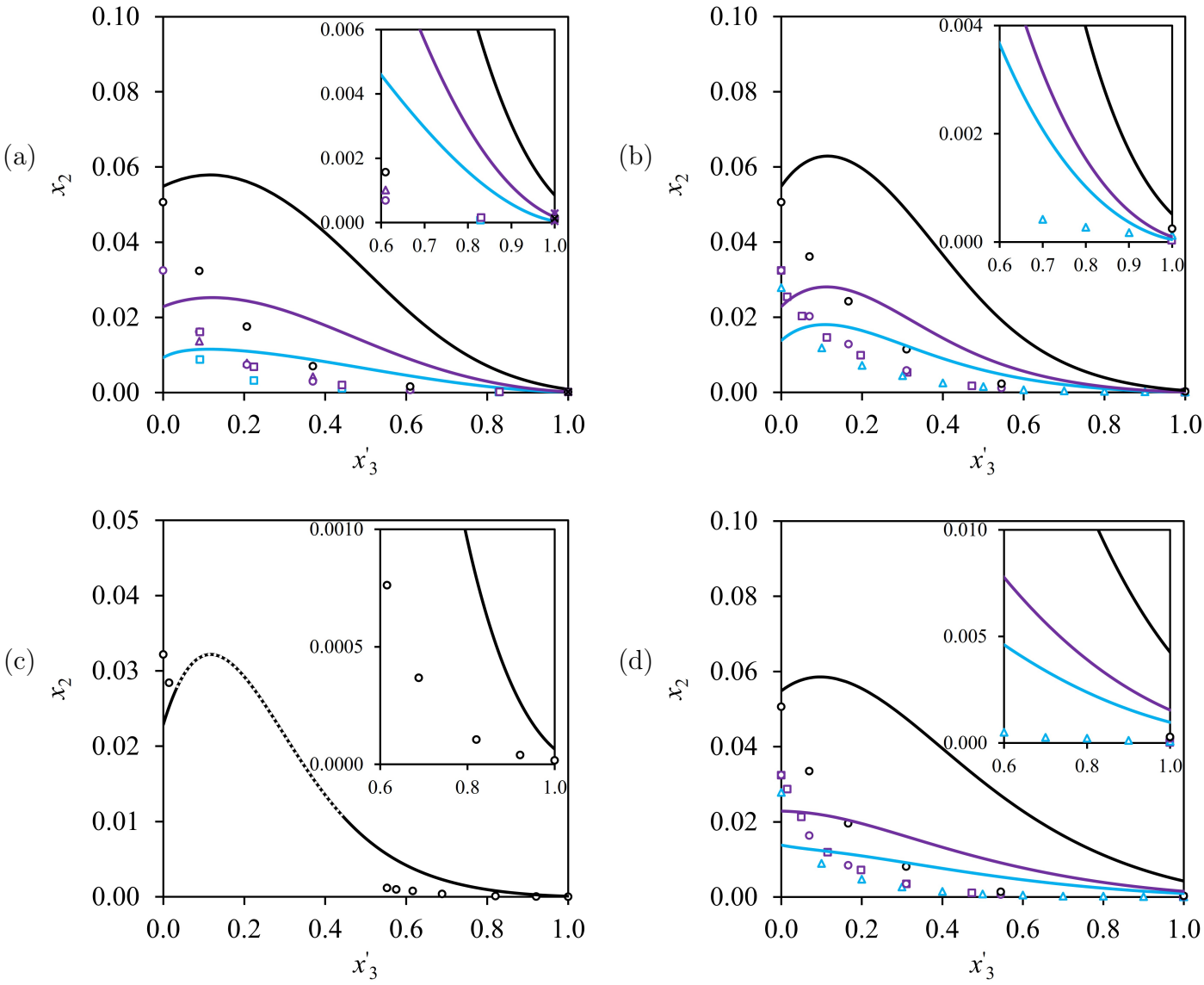


Figure 4.9: SAFT-γ Mie description of the solid–liquid equilibria of alanine in mixed aqueous alcohol solvents expressed as the mole fraction of alanine as a function of solute-free alcohol mole fraction. (a) Solubility of alanine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Different empty symbols denote different sources: circles [202]; triangles [199]; and squares [206]. "x" symbols denote pure-ethanol alaninesolubility [199, 30, 17, 201, 202]. (b) Solubility of alanine (2) in water (1) + 1-propanol (3) at 1 bar and: 283.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Symbols denote experiments from different sources: circles [202]; triangles [195]; and squares [29]. (c) Solubility of alanine (2) in water (1) + 1-butanol (3) at 1 bar and 298.15 K. Empty circles denote mixed-solvent data [208]. The dotted section of the curve represents the region of liquid–liquid immiscibility of the water/1-butanol mixture. (d) Solubility of alanine (2) in water (1) + 2-propanol (3) at 1 bar and: 283.15 K (light blue); 298.15 K (purple); and 333.15 K (black). Different symbols denote different sources of experimental data: circles [202]; triangles [195]; and squares [29].

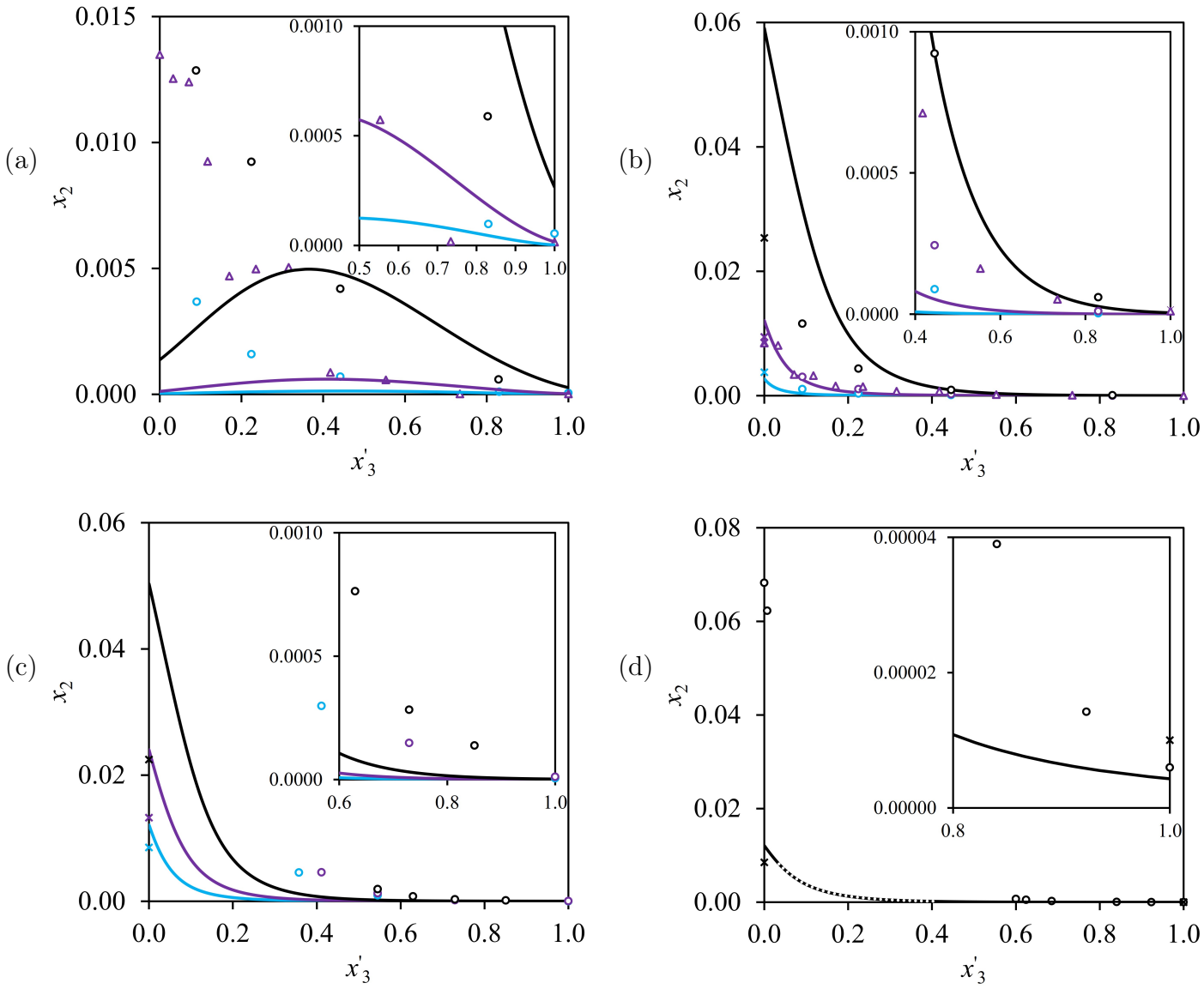


Figure 4.10: SAFT-γ Mie description of the solid–liquid equilibria of valine and alanine in mixed aqueous alcohol solvents expressed as the mole fraction of amino acid as a function of solute-free alcohol mole fraction. (a) Solubility of valine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (light blue); 298.15 K (purple); and 338.15 K (black). Different empty symbols denote different sources: circles [206]; and triangles [201]. (b) Solubility of serine (2) in water (1) + ethanol (3) at 1 bar and: 273.15 K (light blue); 298.15 K (purple); and 338.15 K (black). Symbols denote experiments from different sources: circles [206]; and triangles [201]. (c) Solubility of serine (2) in water (1) + 1-propanol (3) at 1 bar and: 298.15 K (light blue); 313.15 K (purple); and 333.15 K (black). Empty circles denote mixed-solvent data [210] and "x" symbols denote pure-water serine solubility data [111]. (d) Solubility of serine (2) in water (1) + 1-butanol (3) at 1 bar and 298.15 K. Empty circles denote experimental mixed-solvent solubility data [208] and "x" symbols denote pure-water and 1-butanol solubility data [197, 111].The dotted section of the curve represents the region of liquid–liquid immiscibility of the water/1-butanol mixture.

Table 4.2: Overview of the accuracy of SAFT-γ Mie in the calculation of solubility for amino acids in water or alcohol, where N^{D} is the number of experimental data points used to calculate the average absolute deviation (AAD).

System	T / K	N^{D}	AAD (x_1^{sat})	Figure	Ref.
glycine (1) + water (2)	273 – 373	36	1.01×10^{-2}	4.1,4.5	[111, 112, 211, 30]
glycine (1) + ethanol (2)	278 – 333	26	5.35×10^{-4}	4.5	[196, 199, 30, 200, 17, 201, 202, 203]
glycine (1) + 1-propanol (2)	298	2	4.00×10^{-5}	4.5	[202, 196, 30]
glycine (1) + 2-propanol (2)	298 – 310	4	8.49×10^{-4}	4.5	[202, 196, 204, 30]
glycine (1) + 1-butanol (2)	298	3	3.59×10^{-5}	4.5	[196, 31, 18]
alanine (1) + water (2)	273 – 373	29	8.02×10^{-3}	4.2,4.5	[111, 195]
alanine (1) + ethanol (2)	283 – 333	16	1.98×10^{-4}	4.5	[199, 195, 202, 201, 17, 30]
alanine (1) + 1-propanol (2)	283 – 333	12	1.12×10^{-4}	4.5	[29, 195, 30, 202]
alanine (1) + 2-propanol (2)	283 – 333	12	2.12×10^{-3}	4.5	[29, 195, 202, 30]
alanine (1) + 1-butanol (2)	283 – 323	6	4.00×10^{-4}	4.5	[31, 195]
alanine (1) + 2-butanol (2)	283 – 323	5	1.04×10^{-3}	4.5	[195]
alanine (1) + 2-methyl-1-propanol (2)	283 – 323	5	2.50×10^{-4}	4.5	[195]
serine (1) + water (2)	273 – 373	17	1.93×10^{-2}	4.4,4.5	[111]
serine (1) + ethanol (2)	278 – 333	15	2.20×10^{-5}	4.5	[201, 197]
serine (1) + 2-propanol (2)	278 – 333	12	9.40×10^{-5}	4.5	[197]
serine (1) + 1-butanol (2)	278 – 333	13	4.67×10^{-6}	4.5	[31, 197]
valine (1) + water (2)	273 – 343	23	1.26×10^{-2}	4.3,4.5	[111]
valine (1) + ethanol (2)	288 – 343	9	4.89×10^{-5}	4.5	[201]
valine (1) + 2-propanol (2)	293 – 343	6	5.67×10^{-4}	4.5	[198]

Table 4.3: Overview of the accuracy of SAFT-γ Mie in the calculation of solubility for peptides in water, where N^{D} is the number of experimental data points used to calculate the average absolute deviation (AAD).

Solute	T / K	N^{D}	AAD (x_1^{sat})	Ref.
gly-gly	278 – 313	11	0.0061	[192, 115]
gly-gly-gly	278 – 313	16	0.045	[192, 115]
ala-ala	293 – 323	6	0.039	[192]
ala-ala-ala	288 – 308	5	0.0068	[192]
gly-ala	293 – 323	5	0.045	[192]
ala-gly	293 – 323	6	0.019	[192]
gly-gly-ala	293 – 323	9	0.012	[192]
gly-ala-gly	293 – 323	8	0.020	[192]
ala-gly-ala	288 – 308	9	0.026	[192]
leu-gly-gly	293 – 323	7	0.022	[192]
gly-leu-gly	293 – 323	9	0.016	[192]
gly-gly-leu	288 – 308	9	0.016	[192]
gly-ala-leu	293 – 323	9	0.00079	[192]
gly-ser	293 – 323	6	0.047	[192]
ser-gly	288 – 308	9	0.0033	[192]
ala-ser	293 – 323	7	0.036	[192]
ser-ala	288 – 308	10	0.00022	[192]

4.4 Solid–liquid solubility under changing pH

The treatment of amino acids and oligopeptides with fully neutral groups limits the ability of the model to describe pH-dependent solubility, which is an important property in pharmaceutical and biological applications. The bioavailability of a pharmaceutical product depends on its solubility in the human body, within which there are significant variations in pH [212]. Furthermore, the ability to model the speciation of amino acids and oliogopeptides is also essential to understanding their behaviour in salt solutions, which plays a substantial role in screening solvent conditions [213, 214].

Amino acids are ampholytes, meaning that they possess a dual acid–base nature conferred by the presence of the NH_2 amine and COOH carboxyl groups, which ionize to NH_3^+ and COO^-, respectively. Additionally, some amino acids contain ionizable side groups, making them polyprotic ampholytes. In this work, only diprotic amino acids (i.e. those containing non-ionizable side groups) are considered. Specifically, the pH-dependent solubility of glycine and alanine is modeled here due to the availability of reliable experimental data [215, 216] for these two amino acids. The SAFT-γ Mie approach is implemented, in which the amino acids are treated as zwitterions; i.e., molecules that carry an overall charge of zero, but which lead to the formation of cationic and anionic species in solution as the pH changes. Chemical equilibrium equations, which govern the speciation of amino acids in solution, are incorporated with the solid–liquid equilibrium equation to describe the solubiliy of amino acids under variable pH and temperature.

4.4.1 SAFT-γ Mie models

In this work, to describe the speciation behaviour and pH-dependent solubility of an amino acid (AA) in water, the following chemical species are modeled: the zwitterionic ($AA^{\pm}$), cationic(AA^+), and anionic (AA^-) species of the amino acid; water (H_2O) along its dissociation products, namely the hydronium (H_3O^+) –which is essential for computing the pH of

solution– and hydroxide (OH^-) ions; and the counterions, namely Cl^- and Na^+, that are produced in the dissolution of the strong acid (HCl) and strong base (NaOH), respectively. The proposed SAFT-γ Mie model of glycine as a zwitterion, with its corresponding cation and anion, is shown in Figure 3.3. Modelling these species in solution requires characterising the like and unlike interactions of the COO^-, NH_3^+, H_3O^+, OH^-, Na^+, and Cl^- groups, in addition to the interactions of the relevant neutral groups, which have been discussed at length in Sections 3.4 and 4.3.1.

As shown in the parameter sub-matrix of Figure 3.4, most of the interactions of charged groups have been presented in previous works. The hard-sphere diameter σ_{kk} and shape factor S_k of charged groups are based on those of the corresponding uncharged group (they are assigned the same value). Additionally, a Born diameter is estimated by increasing the bare diameter of the ion by 7% ($\sigma_{kk}^{\text{Born}} = 1.07\sigma_{kk}$) to correct for the non-sphericity of the solvation ion cavity following the proposition of Rashin and Honig [217]. The charge is assigned according to the type of the group: $Z_k = +1$ for a cationic group and $Z_k = -1$ for an anionic one. To model the amino acid zwitterion, the NH_3^+ and COO^- groups are used, but an overall charge of zero is assigned for the molecule ($Z_i = 0$), such that no Coulombic or Born contribution arises in the calculation of the free energy contribution of this species (so that these contributions equal zero for species with $Z_i = 0$). The number of association sites of the charged groups is also different from that of the neutral groups, reflecting the loss or gain of a proton and the different tendency to form hydrogen bonds. The dispersion energy of the charged groups is different from that of the corresponding neutral group, as can be expected, and is either calculated by the use of combining rules or adjusted to experimental data, if available. A number of the unlike interactions involving the charged groups are obtained using combining rules; this was shown to be reliable in predicting the solubility of ibuprofen [57] and, accordingly, the same strategy is followed here.

In the case of the unlike COO^-–NH_3^+ interaction, however, it was found that the association energy parameter ($\varepsilon_{kl,ab}^{\text{HB}}$) between the e_1 site on the COO^- group and the H site on the NH_3^+ needs further refining to capture accurately the reported solubility of glycine at the isoelectric point. This change, while significantly improving the model's description of the pH-

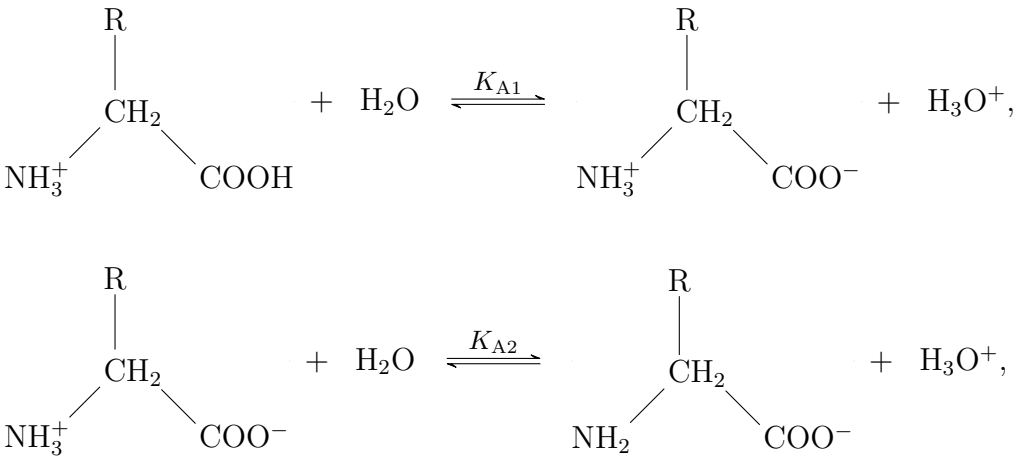
dependent solubility of glycine, had almost no impact on calculations of ibuprofen solubility [57]. This is because the ibuprofen molecule is much larger, and contains more association interactions than the amino acids studied. In other words, the relative contribution of the unlike COO^-–NH_3^+ interaction to the properties of ibuprofen is much smaller than its contribution to the properties of amino acids.

The optimised and calculated like and unlike parameters of the SAFT-γ Mie groups relevant to the aqueous solutions of glycine, alanine, water, and the related ions that result from their speciation, are presented in Tables 3.1, 3.2, and 3.3.

4.4.2 Equations of chemical equilibria of speciation

To model the solid–liquid solubility of diprotic amino acids, the solid–liquid equilibrium equation (2.19) is solved for the amino acid zwitterion, taking into account the speciation (chemical equilibrium) equations of the amino acid and the ionisation of water in the liquid phase. The concentrations of each of these species are determined as a function of changing pH, at given T and P. Thus, the liquid phase is a mixture containing the neutral zwitterion, the amino acid cation and anion, the counterions of the acid and base, and the species related to water, in coexistence with a solid phase containing only the neutral amino acid.

For a diprotic amino acid (NH_3^+−RCH_2−COO^-), where R is a non-ionisable side group in aqueous solution, the acid–base chemical equilibria between the different species of the amino acid can be written as:



$$2H_2O \quad \overset{K_W}{\rightleftharpoons} \quad H_3O^+ \; + \; OH^-,$$

where, K_{A1} and K_{A2} are the (true) equilibrium constants [218] associated with the speciation of the zwitterion ($AA^{\pm}$) into the cation (AA^+) and anion (AA^-) amino acids, respectively, and K_W is the dissociation constant of water. These are given as,

$$K_{\mathrm{A1}} = \frac{a_{\mathrm{AA}^{\pm}} \; a_{\mathrm{H_3O^+}}}{a_{\mathrm{AA}^{+}} \; a_{\mathrm{H_2O}}}, \tag{4.2}$$

$$K_{\mathrm{A2}} = \frac{a_{\mathrm{AA}^{-}} \; a_{\mathrm{H_3O^+}}}{a_{\mathrm{AA}^{\pm}} \; a_{\mathrm{H_2O}}}, \tag{4.3}$$

and

$$K_{\mathrm{W}} = \frac{a_{\mathrm{H_3O^+}} \; a_{\mathrm{OH^-}}}{\left(a_{\mathrm{H_2O}}\right)^2}, \tag{4.4}$$

where a_i is the activity of species i. The equilibrium constants are related to the corresponding $\mathrm{p}K_i$ by

$$\mathrm{p}K_i = -\log_{10}(K_i); \; i = \mathrm{A1}, \mathrm{A2}, \mathrm{W}, \tag{4.5}$$

and we calculate pH as [219]

$$\mathrm{p}H = -\log_{10}(a_{\mathrm{H_3O^+}}). \tag{4.6}$$

The activity a_i is calculated following the asymmetric convention, i.e.,

$$a_i = \frac{m_i}{m_0}\tilde{\gamma}_{m,i}, \tag{4.7}$$

at each pressure, temperature, and composition for all species except water. Here m_i is the molality of i, $m_0 = 1$ mol kg^{-1} is the reference molality and $\tilde{\gamma}_{m,i}$ is the asymmetric molality-based activity coefficient, calculated as

$$\tilde{\gamma}_{m,i} = x_j \frac{\varphi_i(T, P, \mathbf{x})}{\varphi_i(T, P, \mathbf{x}^\infty)}, \tag{4.8}$$

where x_j is the mole fraction of the solvent (water), φ_i the fugacity coefficient of i, calculated using the SAFT-γ Mie approach, $\mathbf{x}$ the composition vector of the mixture, and $\mathbf{x}^\infty$ the composition vector of the reference which is an infinitely dilute mixture. In the calculations of this work, a mole fraction of 1×10^{-15} is used for the infinite-dilution fugacity coefficient. The water activity (a_w) is calculated according to the symmetric convention as:

$$a_\mathrm{w} = x_\mathrm{w}\gamma_\mathrm{w}, \tag{4.9}$$

where x_w is the mole fraction of water and γ_w is the symmetric mole-fraction-based activity coefficient, calculated as:

$$\gamma_\mathrm{w} = \frac{\varphi_\mathrm{w}(T, P, \mathbf{x})}{\varphi_\mathrm{w}(T, P, x_\mathrm{w} = 1)}. \tag{4.10}$$

To account for the effect of temperature on the equilibrium constants, the van't Hoff equation is used:

$$K_r(T) = K_r(T_0) \exp\left(-\frac{\Delta h_{\mathrm{Protonation},r}}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right); r = \mathrm{A1}, \mathrm{A2}, \mathrm{W}, \tag{4.11}$$

where $K_r(T)$ is the equilibrium constant of reaction r at the system temperature, $K_r(T_0)$ is the equilibrium constant at a known reference temperature (here $T_0 = 298.15$ K), and $\Delta h_{\mathrm{Protonation},r}$ is the enthalpy of protonation of reaction r. The values of the equilibrium constant for reactions involving glycine and alanine, and the corresponding enthalpies of protonation can be found in Table 4.4. The temperature dependence of the water dissociation constant K_W is calculated using equation (4.11) with $K_\mathrm{W}(T_0) = 1.0077 \times10^{-14}$ [220] and $\Delta h_{\mathrm{Protonation,W}} = 56.149$ kJ/mol [221]. The base-10 logarithm values of the acid dissociation constants, pK_A1 and pK_A2, as well as the isoelectric point, pI, change significantly as a function of temperature. As illustrated in Figure 4.11, the pK_A and pI values span up to two pH points in the temperature range 280 K

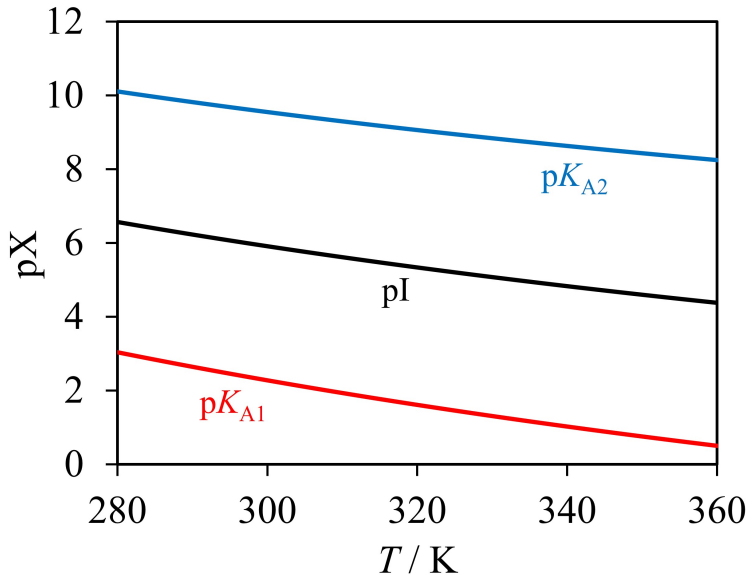


Figure 4.11: Calculation of the temperature dependence of the pK_{A1}, pK_{A2}, and pI of glycine using the van't Hoff equation (Equation 4.11). Parameters used in the van't Hoff equation can be found in Table 4.4.

Table 4.4: Acid–base equilibrium properties of glycine and alanine.

Compound	K_{A1}(298.15 K)	Δh_{A1} (kJ/mol)	K_{A2}(298.15 K)	Δh_{A2} (kJ/mol)	Ref.
Glycine	4.525×10^{-3}	61.101	2.54×10^{-10}	44.780	[222]
Alanine	4.535×10^{-3}	61.101	2.06×10^{-10}	47.919	[222]

- 360 K.

In Figure 4.12, the amphoteric speciation behaviour of a diprotic amino acid, in this case glycine, at fixed $T = 298.15$ K, $P = 1$ bar, and zwitterion molar composition $x_{\mathrm{AA}^{\pm},\mathrm{glycine}} = 0.00265$, is illustrated. The zwitterion composition is fixed to a value known to be below the solid–liquid solubility to model an unsaturated solution of the amino acid fully in the liquid phase. The OH^- equivalents are calculated from the net charge of the amino acid species of i:

$$\mathrm{OH}^-_{\mathrm{equiv},i} = 1 - \sum_l \xi_{l,i} Z_l;\ l = \mathrm{AA}^{\pm}, \mathrm{AA}^{+}, \mathrm{AA}^{-}, \tag{4.12}$$

where $\xi_{l,i}$ are the relative concentrations of the amino-acid species l, given by

$$\xi_{l,i} = \frac{x_{l,i}}{x_{\mathrm{AA}^{\pm},i} + x_{\mathrm{AA}^{+},i} + x_{\mathrm{AA}^{-},i}}. \tag{4.13}$$

This is in agreement with the definition of OH^- equivalents as the number of moles of OH^- ions

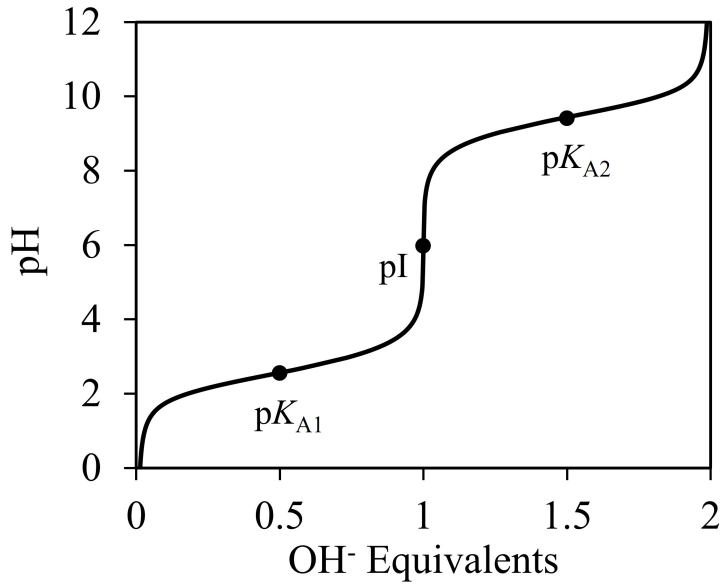


Figure 4.12: SAFT-γ Mie calculations of the acid-base titration curve for glycine at 298.15 K and 1 bar for an unsaturated glycine zwitterion molar composition of 0.00265. The OH^- equivalents are calculated as the proportion of OH^- molecules required to neutralise the glycine species; i.e. 50% of the cationic glycine is neutralised at pK_{A1} and 100% is neutralised at pI, at which glycine is predominantly in the zwitterionic form. At pK_{A2}, 50% of the zwitterionic species is ionised into the anionic form. Close to pH=12, glycine is predominantly in the anionic form.

required to convert 50% of the glycine cations to zwitterions (at the 0.5 equivalence point), to convert 100% of the glycine cations to zwitterions (at the 1.0 equivalence point), and to convert 50% of the glycine zwitterions to anions (at the 1.5 equivalence point). The glycine zwitterions are fully converted to anions at the 2.0 equivalence point.

4.4.3 Solubility prediction

The solubility equation (Eq. 2.19) is solved to obtain the mole fraction of the zwitterionic species of the amino acid in equilibrium with the solid phase and in chemical equilibria with the cationic and anionic species.

In Figure 4.13 the calculated concentrations of the glycine zwitterion, anion, and cation, and the overall solid–liquid solubility, are presented as a function of pH at 298.15 K and 1 bar. In Figure 4.13(a) the relative concentrations of the glycine zwitterion $AA^{\pm}$, cation AA^+, and anion AA^-, given by equation (4.13) can be seen. The calculated solubility is compared to

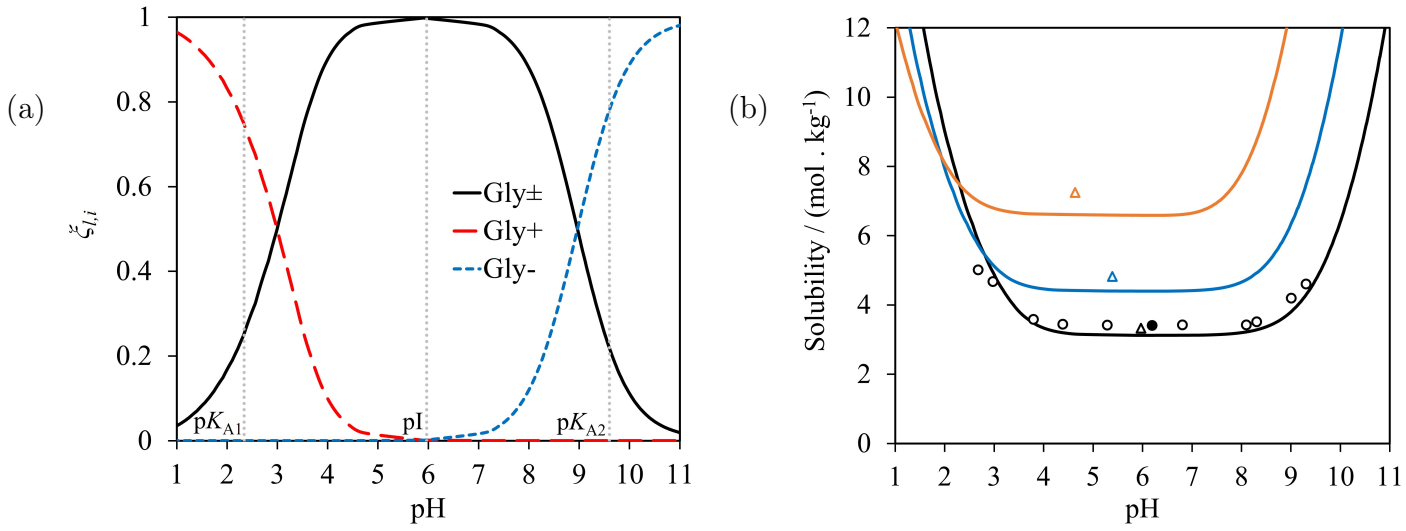


Figure 4.13: (a) Relative concentration ($\xi_{l,i}$) of glycine zwitterion (continuous black curve), cation (long-dashed red curve), and anion (short-dashed blue curve) at 298.15 K and 1 bar, as a function of pH. The pK_A and pI values [222] are denoted by the vertical black dotted lines. (b) The solubility of glycine in water at 298.15 K (black); 318.15 K (blue); and 348.15 K (orange) and at 1 bar as a function of pH. The continuous curves represent SAFT-γ Mie calculations and symbols represent experimental solubility data; circles denote pH-dependent data by Needham et al. [215] and triangles solubility data at the isoelectric point (pI) by Dalton and Schmidt [111]. The filled circle represents the data point used in optimising the $NH_3^+-COO^-$ interaction, while empty symbols denote data not used in the parameter estimation.

the experimental data available [216, 111] in Figure 4.13(b). The equilibrium constants K_{A1} and K_{A2} are calculated using equation (4.11) with the parameters in Table 4.4, and K_W as described earlier.

As can be seen in Figure 4.13, for a significant range of pH close to the isoelectric point, the prevalent species in solution is the zwitterion, which has a lower water solubility than the cation and anion, and hence leads to a solubility minimum (cf. Figure 4.13(b)). At low values of pH, the equilibrium tends towards the left-hand side of equation (4.2) and the glycine cation is the prevalent species in solution. This accumulation of positively charged ions is balanced by the presence of negatively charged counterions (Cl^- from HCl). At high values of pH, the equilibrium tends to the right-hand side of equation (4.3) resulting in the glycine anion becoming the prevalent species, which is now electrostatically balanced by a similar concentration of positively charged counter-ions (Na^+ from NaOH). The cationic and anionic forms of glycine are highly soluble in water due to favourable solvation interactions between these ions and water molecules, leading to the higher solubility seen at both ends of the pH scale

in the figure. The calculated solubility is in good agreement with the experimental data (Figure 4.13(b)), with only a small deviation noticeable at pH = 2.7. The influence of temperature on the pH-dependent solubility can also be seen in Figure4.13 (b). At 348.15 K, the pK_{A1} and pK_{A2} values calculated using equations (4.11) and (4.5) (0.807 and 8.47, respectively) are significantly lower than those at 298.15 K (2.34 and 9.60, respectively). This results in a shift in the pH-solubility profile to the left, centered around the pI value, which is calculated as the arithmetic mean of the pK_{A1} and pK_{A2} values. Additionally, at higher temperatures the solubility of glycine is higher, in accordance with equation (2.19), which leads to an upwards shift in the solubility minimum. This upwards and leftwards shift causes the two curves to cross in the low-pH region.

It is common practice in modelling chemical equilibria in aqueous solutions to assume that the water activity (a_w) is equal to 1, especially in dilute solutions, and to neglect this contribution in the equilibrium constant equations. While this can be a valid approximation near the isoelectric point (pI), at the high and low ends of the pH range, a_w is, however, no longer close to 1, and values below 0.5 can be found, as can be seen in Figure 4.14(a). At the pH extremes, the concentrations of the speciated amino acid and the counterions are high, reflecting their high solubility, and the solution cannot be considered to be close to the reference dilute molality of 1 mol kg^{-}1 for these species. The impact of incorporating a_w in the chemical equilibrium equations on the solubility calculations, can be seen in Figure 4.14(b). As expected, including a_w in equations (4.2), (4.3) and (4.4) leads to a negligible change in the calculated solubility at the pI. However, in the low and high pH range, a more pronounced deviation between calculations that include a_w and those that take it to be unity can be seen. The calculations shown in Figure 4.13(b) are obtained with the inclusion of water activities throughout the range of thermodynamic conditions presented.

In order to assess the transferability of the $NH_3^+-COO^-$ interaction, which is optimised using the solubility data of glycine as a function of pH, similar SAFT-γ Mie calculations for alanine in water (at 298.15 and 348.15 K and 1 bar) as a function of pH, for which no data have been used during parameter estimation, is used. A comparison of the calculations with the available experimental data is shown in Figure 4.15. As in the case of glycine, the alanine

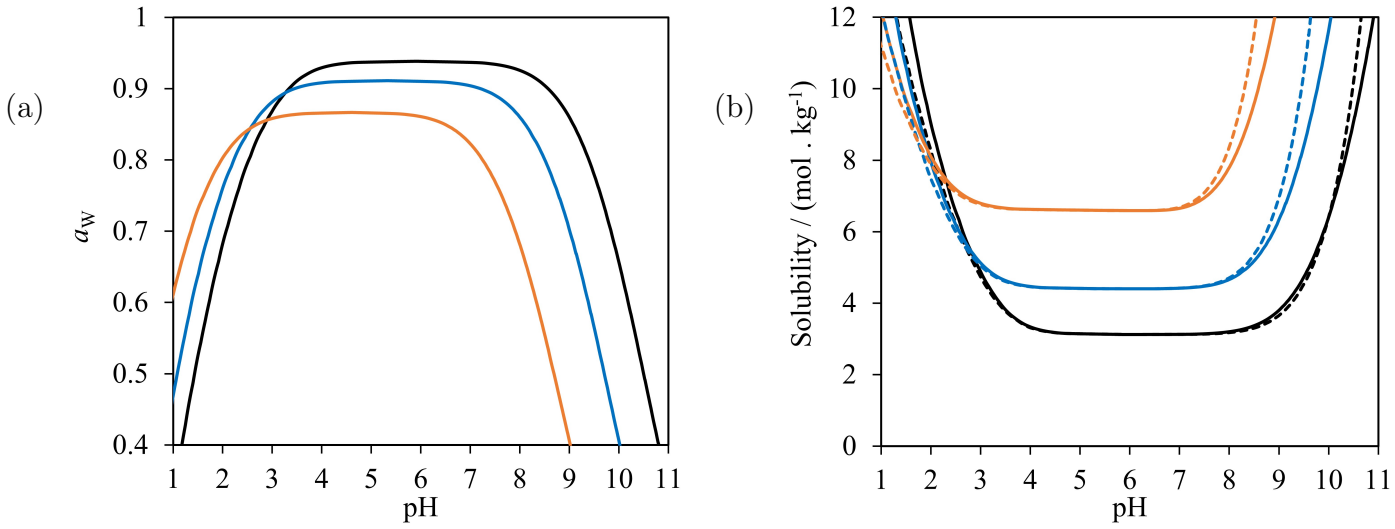


Figure 4.14: (a) The activity of water (a_w) calculated using SAFT-γ Mie as a function of pH for the system containing glycine (described in Figure 4.13) at 298.15 K (black), 318.15 K (blue), and 348.15 K (orange) and at 1 bar. (b) The solubility of glycine in water at 298.15 K (black), 318.15 K (blue), and 348.15 K (orange) and at 1 bar as a function of pH. Solid lines represent calculations in which the real value of a_w is used and dashed lines represent calculations in which a_w is taken to be unity.

zwitterion dominates over a wide range of pH around the isoelectric point, resulting in the minimum in the solubility shown in Figure 4.15(b). Note that these calculations are fully predictive, and that they show a good agreement with the experimental data of Tseng et al. [216] and Dalton and Schmidt [111]. The calculations exhibit a small under-prediction of the solubility across the pH range at 298 K, but accurately capture the qualitative trend in the data as pH changes. In these calculations, as discussed in the previous figure, the activity of water is included at each state (i.e., it is not assumed to be 1). The level of agreement with the experimental data available confirms the predictive ability of this modelling approach and the robustness of the SAFT-γ Mie model presented.

4.5 Conclusions

In this chapter, the SAFT-γ Mie GC framework is used to predict the solubility of amino acids and oligopeptides in various solvent media and across a wide range of thermodynamic conditions. The sensitivity of solubility calculations to changes in the fusion enthalpy and

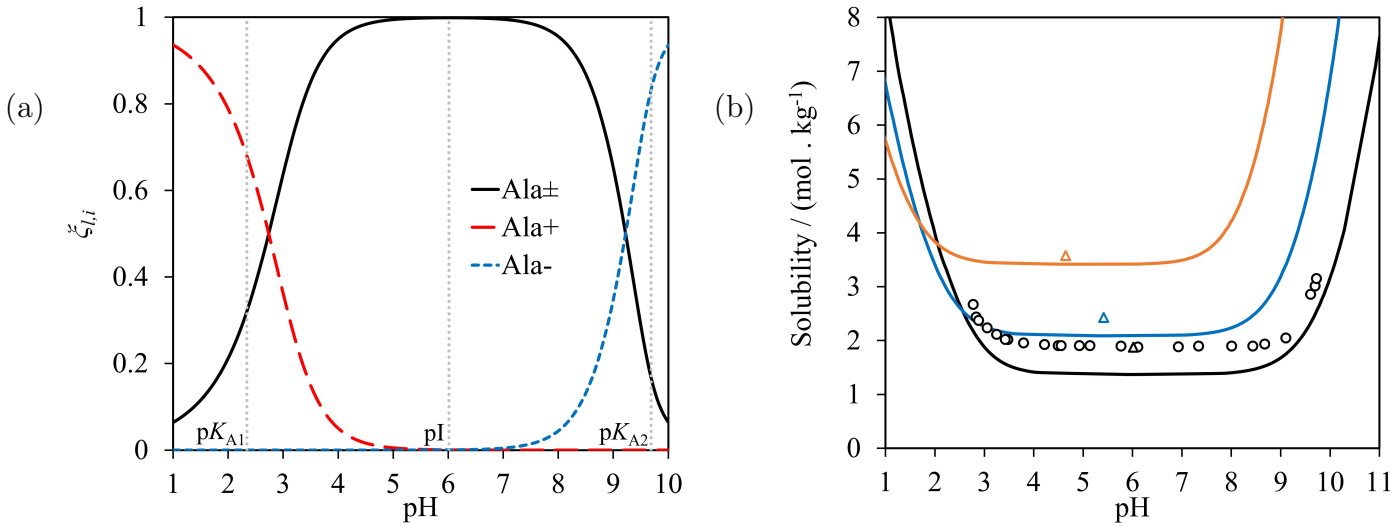


Figure 4.15: (a)Relative concentration ($\xi_{l,i}$) of alanine zwitterion (continuous black curve), cation (long-dashed red curve), and anion (short-dashed blue curve) at 298.15 K and 1 bar, as a function of pH. The pK_A and pI values [222] are denoted by the vertical black dotted lines. (b) The solubility of alanine in water at 298.15 K (black); 318.15 K (blue); and 348.15 K (orange) and 1 bar, as a function of pH. The continuous curves represent SAFT-γ Mie predictions and the symbols the experimental solubility data; circles denote pH-dependent solubility data by Tseng et al. [216] and triangles denote solubility data at the isoelectric point (pI) by Dalton and Schmidt [111].

melting temperature is discussed. Given the large uncertainty in reported experimental melting-property data, the solubility predictions are shown to vary significantly, especially with respect to the uncertainty in fusion enthalpy data. The uncertainty in melting temperature, on the other hand, is relatively smaller and leads to less significant variability in solubility predictions.

The SAFT-γ Mie GC framework is first used to predict the solid–liquid solubility of the amino acids glycine, alanine, valine, and serine in pure water and alcohols and in mixed water + alcohol solvents at the isoelectric point (pI) of the solutions. Despite the variability in the reported experimental solubility data, the model yields fairly accurate results. The model is then extended to predict the solid–liquid solubility of 17 di- and tri-peptides comprising the amino acids glycine, alanine, leucine, and serine, yielding good agreement with experimental data.

To predict the pH-dependent solid–liquid solubility of amino acids, the model is extended to treat the chemical equilibria associated with the ionisation of the amino acid zwitterion and the dissociation of water molecules under changing pH. Solving the solid–liquid equilibrium

equation alongside the chemical equilibrium equations, the model yields very good agreement with experimental data of pH-dependent solubility at different temperatures.

Overall, the SAFT-γ Mie GC approach is successfully shown to produce reliable solubility predictions for amino acids and oligopeptides in water, alcohol, and water + alcohol solvents, both at the isoelectric point and under changing pH.

Chapter 5

Molecular simulations of charged chain molecules: comparison with SAFT-γ Mie molecular and group models

5.1 Introduction

A good understanding of the thermodynamics of organic electrolytes is essential in applications ranging from renewable energy to drug design and development. A particularly prominent use of organic electrolytes today is in the design and manufacturing of electrochemical batteries, such as Lithium-ion (Li-ion) [223], Sodium-ion (Na-ion) [224], and Potassium-ion (K-ion) [225] batteries. Specifically, organic solutions of alkali salts, such as Lithium hexafluorophosphate (in the case of secondary commercial Li-ion batteries), are used as the electrolytic component of the battery; they act as the charge carriers between the battery's electrodes.

Modelling electrolyte solutions is also important in the context of drug development, as many active pharmaceutical ingredients (APIs) are ionisable and thus produce electrolytic solutions upon dissociation. Examples of such APIs include ibuprofen and procaine [57]. Similarly, amino acids and peptides speciate into their charged species in acidic, basic, and salt solutions.

This significantly alters their thermodynamic properties, such as the solid–liquid solubility, as discussed in previous chapters.

Accurate estimation of chemical potentials is crucial for predicting the thermodynamic behavior of molecular systems. A common method to estimate chemical potentials is the Widom insertion method[226], which calculates the average free-energy change associated with placing an additional molecule into an already equilibrated system. This straightforward method is especially effective for simple, non-polar systems. However, in complex systems such as aqueous solutions, inserting a molecule often encounters energetically unfavorable configurations, leading to insufficient sampling and unreliable results.

To overcome these limitations, advanced sampling techniques are employed, which utilize bias potentials to thoroughly explore the energy landscape. A bias potential is an artificial modification of the energy landscape that encourages the system to visit less favorable but important states more frequently.

One example of an advanced biased sampling method is the Bennett acceptance ratio (BAR) [227]. BAR minimises variance in free-energy calculations by balancing the sampling between pairs of closely related states. This approach has been successfully used, together with methods involving gradual molecule insertions, to accurately determine properties like mean ionic activity coefficients (MIAC) and solubilities of alkali halide solutions [228, 229]. Another powerful technique is metadynamics [230], in which exploration is accelerated by applying a bias potential associated with the values of collective variables (CVs). CVs are carefully chosen parameters, such as distances, angles, or structural features, that describe the important characteristics of a molecular state transition. Metadynamics has proven effective in tasks such as crystal structure prediction (CSP), notably in predicting high-pressure phases of CO_2 [231], and in estimating the solid–liquid interface free energies of Lennard-Jones systems [232].

The Expanded Ensemble method, proposed by Lyubartsev et al. [233], is another valuable biased sampling technique. It involves creating multiple subensembles, each representing closely related system states. Monte Carlo simulations then allow the system to move between these subensembles, thoroughly sampling the energy landscape and providing accurate estimates of

the free energy difference between them. The expanded ensemble method has been used in a variety of applications. Hurley et al. [234] use it to predict the binding free energies of small molecules against water-soluble pillar[6]ene, which has shown potential in drug delivery and bio-diagnostics. Chodera and Shirts [235] use a replica-exchange expanded ensemble method as a Gibbs sampling technique. They apply their method to compute the free energy of transfer of methane from gas to water and the potential of mean force (PMF) of terminally-blocked alanine dipeptide in implicit solvent. Zhang et al. [236] use the expanded ensemble method to predict the relative binding free energies between proteins and ligands. Paluch et al. [237] develop an automated expanded ensemble method to calculate the residual chemical potential and solvation free energies of amino acid analogs in water. By employing these advanced methods—BAR, metadynamics, and expanded ensembles—researchers can reliably estimate chemical potentials in complex systems, addressing the shortcomings associated with simpler approaches like the Widom insertion method.

In this thesis, the ion contribution to chemical potential is calculated using the EE transition matrix (EETM) method [238, 239] which combines the EE method [233], for sampling across multiple thermodynamic states with the transition matrix method [239] that is based on transition matrix Monte Carlo (MC) ideas developed by Fitzgerald et al. [240], for recording the attempted transitions between the different thermodynamic sub-ensembles. Specifically, the EETM method used here combines the principles of replica exchange MD (REMD) [241] with those of the transition matric MC (TMMC) method [242]. In the REMD method, several copies (i.e., replicas) of the system are simulated in parallel at different values of the collective variable (CV). Periodically, replicas exchange configurations based on a criterion, allowing the algorithm to explore different regions in the CV space. In the TMMC method, the probabilities associated with transitions between different states along the CV are collected in a matrix that is used to calculate the density of states (DOS) and free energy as a function of the chosen CV.

Previously, Kournopoulos et al. [243] benchmarked the ionic free-energy contributions of the DH and MSA theories by conducting EETM simulations of solutions of NaCl. In this work, ionic chain molecules are considered. Specifically, EETM simulations are conducted for various systems of model Lennard-Jones ionic chain molecules to benchmark the theoretical MSA

calculations of the ionic chemical potential. The goal is to gain insight into the performance of theoretical calculations based on Models M and G that were presented in Section 3.2.2 by using simulation simulation data of model systems.

In Section 5.2, the method of expanded ensembles is introduced in the context of estimating the ionic chemical potential of salt solutions, highlighting the sampling algorithms followed to achieve this goal. Then, the molecular models and force field parameters used to represent the atomic species in the simulations are presented. In Section 5.6, MSA calculations of the ionic chemical potential using Models G and M are assessed by comparison to EE simulation data. The effects of chain length, size of the uncharged segments, and charge shielding are discussed. This is followed by a discussion of the role of free volume in explaining some of the simulation data. In Section 5.7, some concluding remarks are given.

5.2 The expanded-ensemble method

The expanded ensemble technique [233] is a method for free-energy calculation in a single Monte Carlo (MC) step, in which an NVT ensemble is expanded and modified. Expansion implies the replication of the ensemble into a set of sub-ensembles. Modification implies the introduction of a weight factor for each sub-ensemble. Additionally, each sub-ensemble is characterised by different values of a chosen thermodynamic variable, which is referred to as the collective variable (CV). A Monte Carlo (MC) simulation of an ensemble does not allow the determination of the full corresponding partition function. However, for any pair of sub-ensembles, a standard MC simulation is enough to determine the ratio of their partition functions. The partition function of a sub-ensemble is proportional to the probability of the system to occur in this sub-ensemble [233]. With that, one can compute the free-energy difference associated with the change from one sub-ensemble to the other. The goal is to estimate the free-energy difference between two states; in this thesis, since the goal is to evaluate the validity of the theoretical expressions used in the ion term of the free energy in SAFT-γ Mie, where chain-like ions are considered, the reference state is one in which the electrostatic interactions are "switched off",

and the other state is one in which they are "switched on". Thus, the CV is chosen to be a factor that scales the charge of an ion pair from 0 to 1. In practice, this factor is incremented within the given range, and the free energy of the interaction is estimated by averaging the free energy differences associated with those incremental changes. What follows is a detailed introduction to the method of expanded ensembles.

In the canonical ensemble of constant number of particles (N), volume (V), and temperature (T), the configurational integral of the N-particle system is given by

$$Z(N,V,T) = \int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N)], \tag{5.1}$$

where U is the total potential energy of the system, $\mathbf{r}^N$ is the position vector of all N particles in the system, and $\beta = 1/k_{\mathrm{B}}T$ is the reciprocal of thermodynamic temperature T. In the canonical ensemble, the Helmholtz free energy is given by

$$A = -\frac{1}{\beta} \ln Z(N,V,T)\,. \tag{5.2}$$

To calculate the free energy associated with the Coulombic interaction of an ion pair, the free energy associated with the charging and discharging of a randomly selected pair of ions is determined. The ion charges of the selected pair are incremented linearly, taking on discrete values between the uncharged state ($\lambda_i = 0$) and the fully charged state ($\lambda_i = 1$). Each value of λ_i corresponds to the collective variable (CV) of a sub-ensemble, for which the configurational integral is given by

$$Z_i = \int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N; \lambda_i)]. \tag{5.3}$$

Following the derivations in the work of Lyubartsev et al. [233], the configurational integral of the expanded ensemble (EE) can be defined as

$$Z_{\mathrm{EE}} = \sum_{i=1}^{N_{\mathrm{SE}}} Z_i \exp(-\eta_i), \tag{5.4}$$

where N_{SE} is the number of sub-ensembles in the EE and η_i is a weight assigned to each sub-ensemble, ensuring an equiprobable sampling of the phase space.

Considering two states of the (N, V, T) sub-ensemble, corresponding to λ_0 and λ_1 and described by the configurational integrals: $Z_0 = \int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N; \lambda_0)]$ and $Z_1 = \int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N; \lambda_1)]$, one can write the difference in the Helmholtz free energy between them as:

$$\Delta A = -\beta \ln \frac{Z_1}{Z_0} = -\beta \langle \exp\left[-\beta \Delta U(\mathbf{r}^N; (\lambda_0 \rightarrow \lambda_1))\right] \rangle_{\lambda_0}, \tag{5.5}$$

where $\langle ... \rangle_{\lambda_0}$ denotes an ensemble average at the reference state of λ_0. First, the system is simulated at the reference state $(N, V, T; \lambda_0)$ then it is perturbed to $(N, V, T; \lambda_1)$ to obtain the potential energy change $\Delta U(\mathbf{r}^N; (\lambda_0 \rightarrow \lambda_1))$. After that, the reference state is restored, and the simulation is continued. In order to obtain an accurate estimate of the free-energy difference ΔA, the target state (λ_1) must be visited frequently, and a large number of states must be explored and ensemble-averaged.

The relationship between the difference in Helmholtz free energy and the weights η_i, in Equation 5.4, is shown in the work of Kournopoulos et al. [243], and is repeated here in more detail. Following a Boltzmann distribution, the probability of visiting a state i in the expanded ensemble can be written as

$$p_i = \frac{\exp(\eta_i) Z_i}{Z_{\mathrm{EE}}}. \tag{5.6}$$

Assuming that the weights η_i have been chosen to impose an equiprobable sampling of the phase space in the expanded ensemble implies

$$p_i = p_j; i, j = 1, 2, ..., N_{\mathrm{SE}}, \tag{5.7}$$

where i and j denote two sub-ensembles. Substituting Equation 5.6 into 5.7 and rearranging, we get

$$\frac{Z_i}{Z_j} = \frac{\exp(\eta_j)}{\exp(\eta_i)}\,. \tag{5.8}$$

Then, following the definition of the Helmholtz free energy in Equation 5.2, the difference in the Helmholtz free energy between the states of the two sub-ensembles i and j can be written as:

$$A_j - A_i = -\frac{1}{\beta}\ln(\frac{Z_i}{Z_j})\,. \tag{5.9}$$

Finally, substituting Equation 5.8 into Equation 5.9, and simplifying, we get:

$$\beta(A_j - A_i) = \eta_j - \eta_i\,. \tag{5.10}$$

Equation 5.10 implies that an equiprobable sampling of the phase space of the expanded ensemble relies on a choice of weights η_i that correspond to the free-energy difference between the sub-ensembles. Therefore, calculating the free energy is equivalent to finding the set of weights that flatten the histogram of visited sub-ensemble states. In practice, the total Helmholtz free-energy difference is calculated by adding the free-energy differences of all the intermediate sub-ensemble transitions.

5.3 Transition matrix Monte Carlo

The transition matrix Monte Carlo (TMMC) method [238, 239, 233] is used to efficiently compute the Helmholtz free-energy difference between the uncharged and fully charged states by propagating along a Markov chain. Accordingly, the method samples the transition probabilities between different sub-ensembles of the EE. The unbiased macroscopic transition probabilities between adjacent sub-ensembles ($\Pi_{i\to i+1}$) are calculated by averaging the corresponding unbiased microscopic transition probabilities ($\varpi_{i\to i+1}$).

The detailed balance of the macroscopic probability of the transition between adjacent sub-ensembles, which is given by

$$P_i \Pi_{i \to i+1} = P_{i+1} \Pi_{i+1 \to i}, \tag{5.11}$$

can be rewritten in terms of the configurational integrals as

$$\frac{Z_i}{Z_{i+1}} = \frac{\Pi_{i+1 \to i}}{\Pi_{i \to i+1}}. \tag{5.12}$$

Thus, the free-energy difference between the two unbiased states ($\Delta A_{i \to i+1}$) can be written in terms of the configurational integrals (cf. Equation 5.9, and hence in terms of the macroscopic probabilities, as

$$\frac{\Delta A_{i \to i+1}}{k_\mathrm{B} T} = -\ln\left(\frac{Z_{i+1}}{Z_i}\right) = -\ln\left(\frac{\Pi_{i \to i+1}}{\Pi_{i+1 \to i}}\right). \tag{5.13}$$

As mentioned previously, the unbiased macroscopic probabilities ($\Pi_{i \to i+1}$) are calculated by averaging the microscopic unbiased probabilities ($\varpi_{i \to i+1}$) associated with sub-ensemble transitions. Each time a sub-ensemble transition is attempted, the microscopic unbiased probability of the transition is updated and recorded in a collection matrix C according to

$$C_{ii} \rightsquigarrow C_{ii} + (1 - \varpi_{i \to j}), \tag{5.14}$$

and

$$C_{ij} \rightsquigarrow C_{ij} + \varpi_{i \to j}\,, \tag{5.15}$$

where the curly arrow denotes an update of the value of the transition probability and the row i and column j represent the origin and target sub-ensembles, respectively. When a sub-ensemble change is attempted, both the element of the collection matrix associated with the transition and its complementary element are updated since the inclusion of the latter in the averaging speeds up the calculation significantly. The complementary probability of leaving a

sub-ensemble (Equation 5.14) is that of staying in the original sub-ensemble (Equation 5.15), and since they are complementary, they should always add up to one. It is important to note that all sub-ensemble transition attempts are recorded and used in the averaging, even those that are not accepted.

In this work, the goal is to calculate the free-energy change associated with charging and discharging the Coulombic interaction of an ion pair. As such, the collective variable chosen is λ, which ranges between 0 and 1 and scales the ionic charges according to λq_i. Any transition attempt that would change the value of the collective variable λ out of its range is rejected and penalised. The TMMC sub-ensemble transition step is presented in Algorithm 1.

Algorithm 1 Transition matrix Monte Carlo step

1: Select a sub-ensemble transition $i \to j$.
2: Calculate the unbiased microscopic transition probability:

$$\varpi_{i\to j} = \exp(-\beta \Delta U_{ij})$$

3: Accept the sub-ensemble transition $i \to j$ with biased probability:

$$\xi_{i\to j} = \exp(-\beta \Delta U_{ij}) \cdot \exp(\eta_j - \eta_i)$$

4: Update the collection matrix:

$$C_{ij} = C_{ij} + \varpi_{i\to j}$$
$$C_{ii} = C_{ii} + (1 - \varpi_{i\to j})$$

The macroscopic probabilities can then be calculated from the collection matrix by

$$\Pi_{i\to j} = \frac{C_{i,i\to j}}{\sum_k C_{i,i\to k}}, \tag{5.16}$$

where the denominator represents the sum over all transitions from and to the sub-ensemble$i$.

At the start of the EE simulation, the weights of each of the sub-ensembles are yet to be determined. Therefore, a Wang-Landau (WL)[244] algorithm is used to obtain initial estimates for the chemical potential. The Wang-Landau method is a Monte Carlo method designed to estimate the density of states (DoS) of a system by performing a random walk in energy space. In the context of this work, the WL method is used to quickly explore all of the sub-ensembles.

This is achieved by updating the weight given to a sub-ensemble transition (η_i) with a small penalty γ such that the probability of revisiting the same sub-ensemble i is reduced:

$$\eta_i \rightsquigarrow (\eta_i - \gamma)\,. \tag{5.17}$$

The WL scheme ensures a quick exploration of all sub-ensembles by starting with a large value of γ, and it is kept running until the histogram of the number of visits to each sub-ensemble is uniform. In practice, the flattening of the histogram is considered achieved when its slope is less than a prescribed value s_c, which is commonly chosen to be close to 1. Once that is achieved, a new WL iteration begins – this time with a new scaled γ given by

$$\gamma \rightsquigarrow \alpha\gamma, \tag{5.18}$$

where α is a prescribed factor, which is set equal to 0.5 in this work. The WL iteration described is represented in Algorithm 2

Algorithm 2 Wang-Landau iteration

1: Select a sub-ensemble transition $i \to j$.
2: Calculate the unbiased microscopic transition probability:

$$\varpi_{i\to j} = \exp(-\beta\Delta U_{ij})$$

3: Accept the sub-ensemble transition $i \to j$ with biased probability:

$$\xi_{i\to j} = \exp(-\beta\Delta U_{ij}) \cdot \exp(\eta_j - \eta_i)$$

4: **if** Transition $i \to j$ is accepted **then**
5: Update the histogram of visited states with a visit to state j.
6: Penalise the weight of the transition: $\eta_j = \eta_j - \gamma$.
7: **else**
8: Update the histogram of visited states with a visit to state i.
9: Penalise the weight of the transition: $\eta_i = \eta_i - \gamma$.
10: **end if**
11: **if** The slope of the histogram of visited states reaches s_c **then**
12: Reset the histogram.
13: Update the penalty factor: $\gamma = \alpha\gamma$.
14: **end if**

5.4 Molecular models

The model proposed by Andreev et al. [245] to model aqueous solutions of monovalent salts was used as the basis for studying aqueous NaCl solutions in the work of Kournopoulos et al. [243] and is the basis for the models considered here. In the original model by Andreev et al. [245], all species (cation, anion, and solvent) are represented by single spherical Lennard-Jones (LJ) beads. The LJ size parameter, σ, is taken to be equal for all single spherical species for simplicity. The energy parameter, ϵ, is fixed for all like interactions (cation-cation, anion-anion, and solvent-solvent). The cation and anion are distinguished from each other by having different charges – $+q$ and $-q$, respectively – and different unlike interaction energies, ϵ, with the solvent.

In this thesis, the model is extended to treat aqueous solutions of monovalent salts in which the anion is made up of multiple spherical segments by adding uncharged segments in different configurations. Specifically, the uncharged segments are either bonded sequentially to the same end of the ionic molecule, forming linear chains in which the charged segment is bonded only to one of the uncharged segments, or they are all bonded directly to the charged segment, forming trigonal-planar and tetrahedral ionic molecules.

The spherical beads used to represent the ions, solvent, and uncharged segments interact with each other through the LJ potential, i.e.,

$$u_{ij}^{\mathrm{LJ}}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{5.19}$$

where r_{ij} is the centre-to-centre distance between particles i and j, σ_{ij} is the separation at which the potential is zero, and ϵ_{ij} is the minimum of the potential well. The LJ parametersused in this thesis are summarised in Table 5.1 and are given in reduced, dimensionless units defined as

$$\sigma_{ij}^{*} = \frac{\sigma_{ij}}{\sigma_{\mathrm{H_2O-H_2O}}} \tag{5.20}$$

Table 5.1: Force-field parameters of the LJ potential based on the model of Andreev et al. [245]. Parameters are presented in dimensionless units, as described in Equations 5.20 and 5.21. The same set of parameters is used for the MSA calculations, in both Model G and Model M.

Particle i	Particle j	σ^*_{ij}	ϵ^*_{ij}
Solvent	Solvent	1.00	1.00
Cation	Cation	1.00	1.00
Anion	Anion	1.00	1.00
Uncharged segment	Uncharged segment	1.00	1.00
Solvent	Cation	1.00	1.25
Solvent	Anion	1.00	1.00
Solvent	Uncharged segment	1.00	1.00
Cation	Uncharged segment	1.00	1.00
Anion	Uncharged segment	1.00	1.00

and

$$\epsilon^*_{ij} = \frac{\epsilon_{ij}}{\epsilon_{\mathrm{H_2O-H_2O}}}\,. \tag{5.21}$$

The charges on the cation and anion interact via the Coulomb potential, given by

$$u^{\mathrm{Coul}}_{ij}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}, \tag{5.22}$$

where u^{Coul}_{ij} is the Coulomb potential energy between the ions i and j as a function of their centre-to-centre distance (r_{ij}), q_i is the charge of ion i, ε_0 is the permittivity of vacuum, and ε_r is the relative permittivity of the dielectric medium. To reduce the Coulombic energy in dimensionless units, a fixed Bjerrum length is used. The Bjerrum length is the separation between charges at which the Coulombic energy (u^{Coul}_{ij}) is comparable with the thermal fluctuations of the system ($k_{\mathrm{B}}T$). In this work, following the methodology of Kournopoulos et al.[243], the Bjerrum length is reduced using the size parameter of water:

$$l^*_{\mathrm{B}} = \frac{q^2}{4\pi\varepsilon_0\varepsilon_r k_{\mathrm{B}} T \sigma_{\mathrm{H_2O-H_2O}}}\,. \tag{5.23}$$

Here, the relative permittivity of the dielectric medium is set equal to 1, and the charges of the ions are adjusted to match the desired Bjerrum length at a given temperature.

5.5 Simulation details

All of the simulations presented in this chapter are run with 1000 particles in a cubic periodic simulation box. The size of the simulation box is calculated such that $\rho^* = \rho\sigma^3_{\mathrm{H_2O-H_2O}} = 0.8$. The simulations are run at $T = 420$ K such that $T^* = k_{\mathrm{B}}T/\epsilon_{\mathrm{H_2O-H_2O}} = 0.75$. This mapping of the temperature is introduced by Andreev et al. [245] to reproduce the critical temperature and pressure of water. Additionally, the value of the charges is set $q = 0.1364$ for the cation and anion such that $l^*_{\mathrm{B}} = 1.85$. Simulations are conducted for a range of salt mole fraction between 0 and 0.20. It is important to note that since the Bjerrum length does not depend on the ion concentration, fixing the value of l^*_{B} is equivalent to neglecting the dependence of the dielectric constant on ion concentration.

The simulation procedure follows three stages, all of which are MD simulations conducted using the Large-scale Aromic/Molecular Massively Parallel Simulator (LAMMPS) [246]. First, the initial configuration of the system is created by inserting the desired number of cations and anions into a crystal made of the solvent particles. Random swaps are carried out to minimise the likelihood of ion clusters forming, as this would lead to unnecessarily long equilibration times. Then, once the initial configuration is set up, an NVT equilibration simulation is run to melt the crystal for 1 ns (with a time step of 2 fs). Finally, once the system has equilibrated, the EE simulations are carried out for 4 ns, with a subensemble change attempted every 40 fs.

At the early stages of the EE simulation, the Wang-Landau (WL) [244] scheme is used to provide an initial estimate of the chemical potential with the weights initialised at zero. With every sub-ensemble change, the weight η_i is updated with the penalty (γ) to reduce the probability of visiting the same sub-ensemble again.

Initially, the value of γ is set equal to 1 to quickly force the simulation to explore all the sub-ensembles of the system. This is ensured by flattening the histogram containing the number of visits to each sub-ensemble. Once that is achieved–when the slope of the histogram is less than a prescribed factor (s_{c}=0.95)–the visits histogram is reset, and the penalty factor is halved ($\alpha = 0.5$). Once γ reaches a value less than 10^{-4}, the WL scheme is terminated, and the guesses

from the TMMC are used to refine the free-energy profile.

5.6 Results and discussion

Expanded-ensemble simulations are carried out for a range of systems to benchmark the ionic chemical potential calculations of SAFT-γ Mie, which rely on the expression of the mean-spherical approximation (MSA) [98, 99], calculated according to Model G and Model M (cf. Section 3.2.2). MSA calculations of the ionic contribution to the chemical potential of the salt ($\Delta\mu_{\mathrm{s}}^{\mathrm{Ion}}$) are compared with the EETM simulation results for model ionic chain molecules of varying chain length and charge placement. First, the impact of increasing the chain length of the ionic chain molecule is studied by conducting simulations of salt solutions containing linear anionic molecules of increasing length as exemplified in Figure 5.1(a); this is done by adding uncharged monomeric segments to the anionic particle. Then, simulations and calculations are conducted to explore the impact of the size of the uncharged monomeric segment relative to the charged segment that it is bonded to; such systems are exemplified in Figure 5.1(b). Then, simulations are conducted to explore the effect of charge screening by bonding uncharged monomeric segments to the charged segment in a "branched" fashion, as exemplified in Figure 5.1(c).

The LJ forcefield selected to represent all atomic species in the system, including solvent particles, results in the exclusion of the solvation contribution (i.e., ion–solvent interactions) from the system, such that the ion contribution to the free energy is isolated. As a result, the electrostatic chemical potential estimated in the EETM simulations represents only ion–ion interactions, with zero Born free energy. In terms of the expressions presented in Chapter 3, MSA calculation of the free energy and chemical potential requires the input of ion diameters, which are set equal to those used in the simulation.

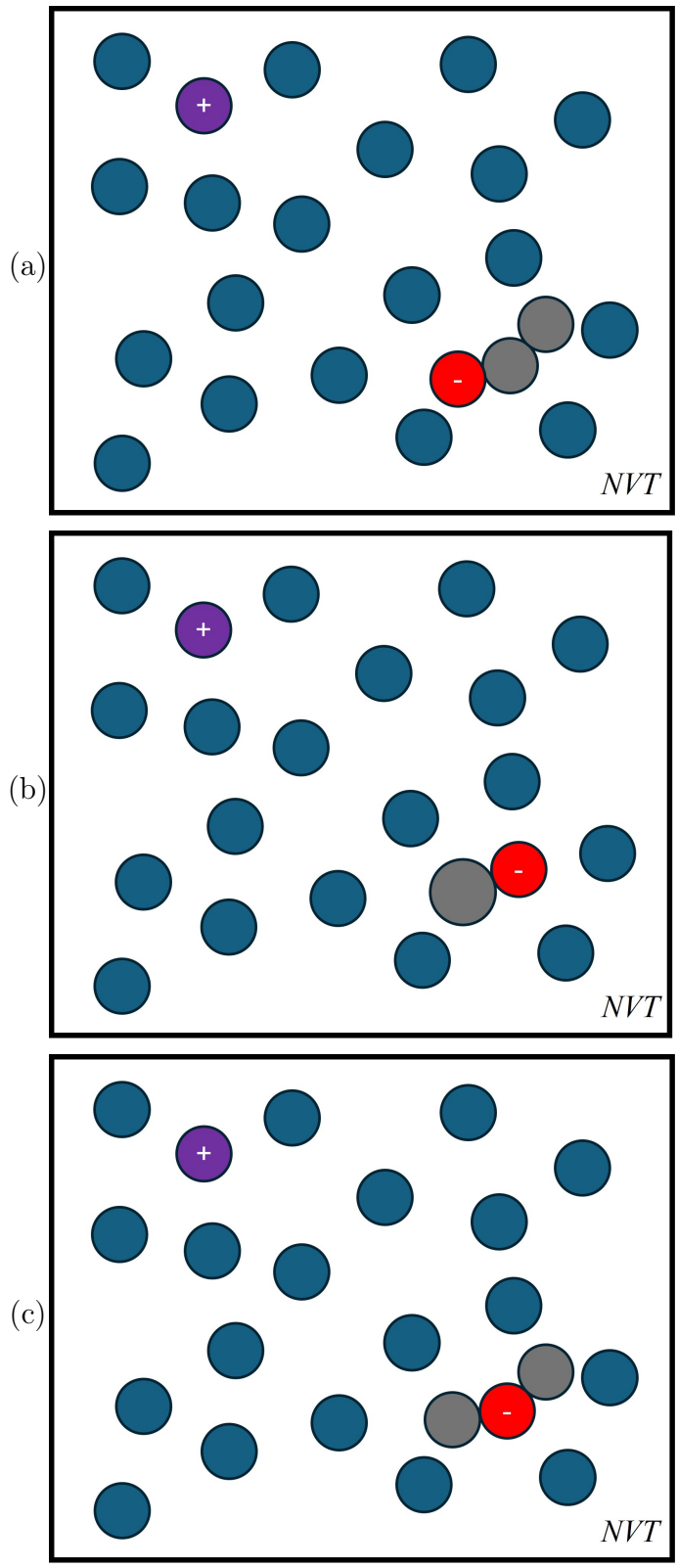


Figure 5.1: A cartoon representation of the systems studied in this chapter. NVT simulation boxes contain solvent molecules (teal), cations (purple), and anions that consist of negatively charged segments (red) bonded to uncharged segments (grey). (a) NVT simulation box containing a trimer anion, in which the charged segment is at the head of the molecule, representing the systems studied in Section 5.6.1. (b) NVT simulation box containing a dimer anion with the uncharged segment being larger than the negatively charged segment, representing the systems studied in Section 5.6.2. (c) NVT simulation box containing a trimer anion, in which the charged segment is at the centre of the molecule, representing the systems studied in Section 5.6.3.

5.6.1 Effect of chain length

The influence of the anion's chain length on the ionic chemical potential is investigated by conducting a series of EE simulations for systems containing spherical cations with anions of increasing chain length – specifically, monomeric, dimeric, trimeric, tetrameric, and pentameric anionic species. The chain length is increased by bonding uncharged segments to the negatively charged particle, linearly and via flexible bonds, as illustrated in Figure 5.2. For each system, an independent EE simulation is performed over a range of salt mole fractions (x_{s}) of up to 0.2 at $T = 420$ K and $l_{\mathrm{B}}^{*} = 1.85$. Note that the salt mole fraction is defined as

$$x_s = \frac{x_{\mathrm{cat}} + x_{\mathrm{an}}}{\sum_{i=1}^{N_{\mathrm{C}}} x_i}, \tag{5.24}$$

where x_{cat} is the mole fraction of the cation, x_{an} is the mole fraction of the anion, and the sum in the denominator is calculated over all components in the system (cation, anion, and solvent).

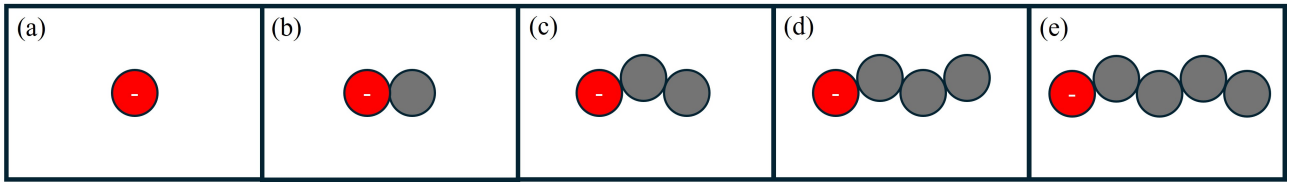


Figure 5.2: Schematic diagram representing the anionic species modeled to study the influence of chain length: (a) Monomeric, (b) Dimer, (c) Trimer, (d) Tetramer, and (e) Pentamer ion.

In Figure 5.3, simulation data (Table 5.2) of the salt solutions are illustrated along with MSA calculations of the ionic chemical potential as a function of salt mole fraction. The simulation data show that the chemical potential, for all salt solutions, increases (in absolute value) with increasing salt concentration. MSA calculations, both using Model G and Model M, successfully capture this qualitative trend. The observed change in the absolute value of the chemical potential between the monomer and dimer anions is quite significant, and it results from the reduction in Coulombic interaction between the ion pair (in the case of dimer anions) due to screening by the uncharged segment. More interestingly, the simulation data show no significant difference in the chemical potentials of the dimer, trimer, tetramer, and pentamer salts. This can be explained by the fact that the neutral segments, which act as a screen for

Coulombic interactions, are added to the same tail end of the anion particle. Thus, beyond the addition of the first neutral segment, the length of the chain has no impact on the Coulombic interaction between the ion pair. Note that no simulations were conducted for the pure solvent (i.e., $x_{\text{salt}} = 0$). For such system, the value of $\Delta\mu_{\text{s}}^{\text{Ion}}$ is equal to zero since there are no charges present.

According to Model G, the free energy depends only on the diameter of the charged segment. As a result of that, the chemical potential calculations are identical for all of the salts studied, regardless of the number of uncharged segments bonded to the ionic particle, if any. Surprisingly, however, the Model G calculations are in excellent agreement with the simulation data of the oligomeric salts, but not with the monomeric salt.

Model M calculations, on the other hand, show a less significant decreasing chemical potential as the chain length of the anion increases. This is because in Model M, the ionic diameter scales with the size of the entire ionic molecule (cf. Section 3.2.2). As the size of the ionic molecule increases, the effective ion diameter ($\tilde{\sigma}_{\text{eff},\,ii}^{\text{MSA, M}}$) increases, which in turn leads to a decrease in the charge density (charge per volume of molecule) and thus a decrease in the ionic chemical potential. Note that as the chain length of the anion increases, the difference in the calculated chemical potential diminishes. This is to be expected as the size of the neutral segment becomes smaller relative to the size of the ionic molecule.

5.6.2 Effect of monomer size

The influence of the size of the neutral segment on the ionic chemical potential is investigated by conducting three EE simulations for salt solutions of dimer anions, in which the diameter of the neutral segment is varied. The simulations are run for salt concentrations of up to 0.2 in salt mole fraction at 420 K and a reduced Bjerrum length of 1.85. Note that only the size parameters of the uncharged segment are varied. Specifically, the like σ_{ii} value for the uncharged segment is varied, and its unlike σ_{ij} values are calculated as the arithmetic mean of σ_{ii} and σ_{jj}. The simulation data of the systems considered are presented in Figure 5.4 and Table 5.3.

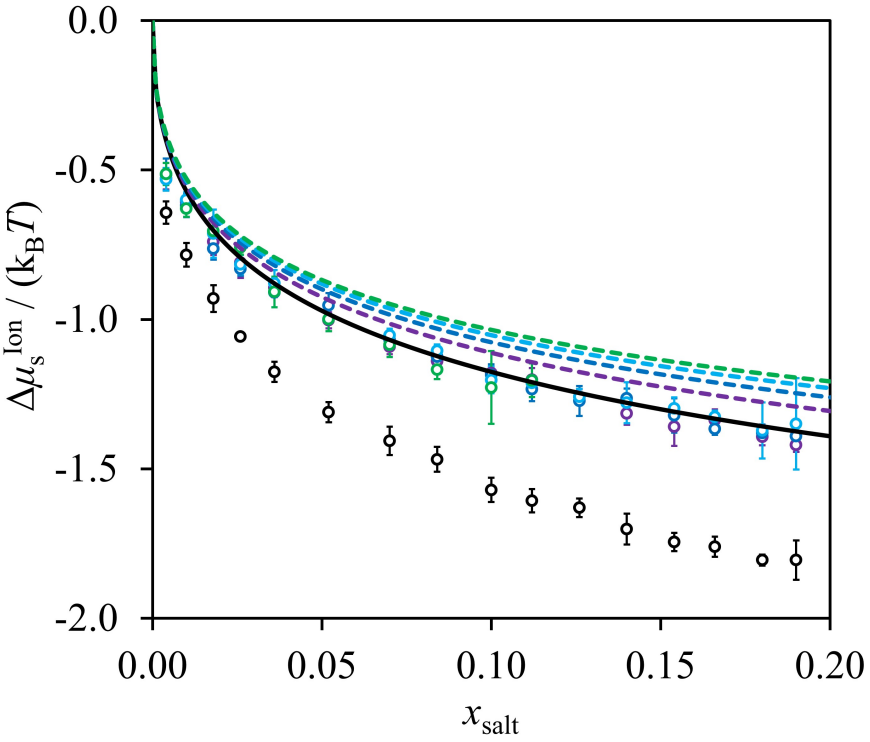


Figure 5.3: Coulombic contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Symbols denote EETM simulations of salts in which the anion is: a monomer (black); a dimer (purple); a trimer (blue); a tetramer (light blue); and a pentamer (green), where the charged bead is placed at the head of the ionic chain. The black solid curve denotes MSA calculations using Model G and dashed curves denote MSA calculations using Model M.

Table 5.2: EE simulation data of the ionic chemical potential ($\Delta\mu_s^{Ion}$) of the monomer, dimer, trimer, tetramer, and pentamer anion salts as a function of salt mole fraction (x_s).

Monomer			Dimer		
x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)	x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)
0.004	-0.643	0.038	0.004	-0.529	0.036
0.010	-0.784	0.040	0.010	-0.608	0.031
0.018	-0.930	0.045	0.018	-0.740	0.044
0.026	-1.06	0.013	0.026	-0.810	0.052
0.036	-1.18	0.034	0.036	-0.894	0.032
0.052	-1.31	0.034	0.052	-1.00	0.027
0.070	-1.41	0.048	0.070	-1.09	0.025
0.084	-1.47	0.041	0.084	-1.14	0.012
0.100	-1.57	0.041	0.100	-1.18	0.021
0.112	-1.61	0.039	0.112	-1.20	0.038
0.126	-1.63	0.031	0.126	-1.26	0.027
0.140	-1.70	0.052	0.140	-1.32	0.038
0.154	-1.75	0.031	0.154	-1.36	0.065
0.166	-1.76	0.033	0.166	-1.34	0.036
0.180	-1.81	0.018	0.180	-1.39	0.029
0.190	-1.81	0.066	0.190	-1.42	0.023
Trimer			Tetramer		
x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)	x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)
0.004	-0.515	0.053	0.004	-0.531	0.038
0.010	-0.619	0.038	0.010	-0.600	0.025
0.018	-0.763	0.037	0.018	-0.713	0.081
0.026	-0.831	0.023	0.026	-0.816	0.032
0.036	-0.910	0.015	0.036	-0.879	0.044
0.052	-0.954	0.042			
0.070	-1.07	0.023	0.070	-1.05	0.022
0.084	-1.13	0.036	0.084	-1.10	0.020
0.100	-1.19	0.036	0.100	-1.20	0.045
0.112	-1.23	0.040	0.112	-1.21	0.022
0.126	-1.27	0.050	0.126	-1.26	0.026
0.140	-1.26	0.032	0.140	-1.28	0.069
0.154	-1.32	0.058	0.154	-1.30	0.038
0.166	-1.37	0.021	0.166	-1.33	0.021
0.180	-1.38	0.028	0.180	-1.37	0.094
0.190	-1.39	0.040	0.190	-1.35	0.15
Pentamer					
x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)			
0.004	-0.514	0.037			
0.010	-0.630	0.028			
0.018	-0.705	0.043			
0.026	-0.759	0.024			
0.036	-0.908	0.051			
0.052	-1.00	0.040			
0.070	-1.09	0.042			
0.084	-1.17	0.033			
0.100	-1.23	0.12			
0.112	-1.20	0.056			

In Figure 5.4, EE simulation data are compared with MSA calculations using Models G and M. As can be seen in the figure, the chemical potential is the most negative for the smallest size ratio =($\sigma_{\mathrm{seg-seg}}/\sigma_{\mathrm{ion-ion}} = 0.25$) and the largest for the largest size ratio ($\sigma_{\mathrm{seg-seg}}/\sigma_{\mathrm{ion-ion}} = 1.5$). This is explained by the fact that the smaller the neutral segment bonded to the ion, the smaller the surface area of the ion that is shielded from Coulombic forces and hence the largest Coulombic contribution to the chemical potential. The difference between the simulation data of the system with $\sigma_{\mathrm{seg-seg}}/\sigma_{\mathrm{ion-ion}} = 1$ and that with $\sigma_{\mathrm{seg-seg}}/\sigma_{\mathrm{ion-ion}} = 1.5$ is barely noticeable because the size difference between the neutral segments in those cases is not significant enough. Larger values of the ratio $\sigma_{\mathrm{seg-seg}}/\sigma_{\mathrm{ion-ion}}$ should result in less negative values of chemical potential. However, simulations that include uncharged segments with $\sigma_{\mathrm{seg-seg}}$ greater than 6 (i.e. $\sigma_{\mathrm{seg-seg}}/\sigma_{\mathrm{ion-ion}} > 1.5$) exhibit precipitation and thus lead to unreliable calculation of the chemical potential.

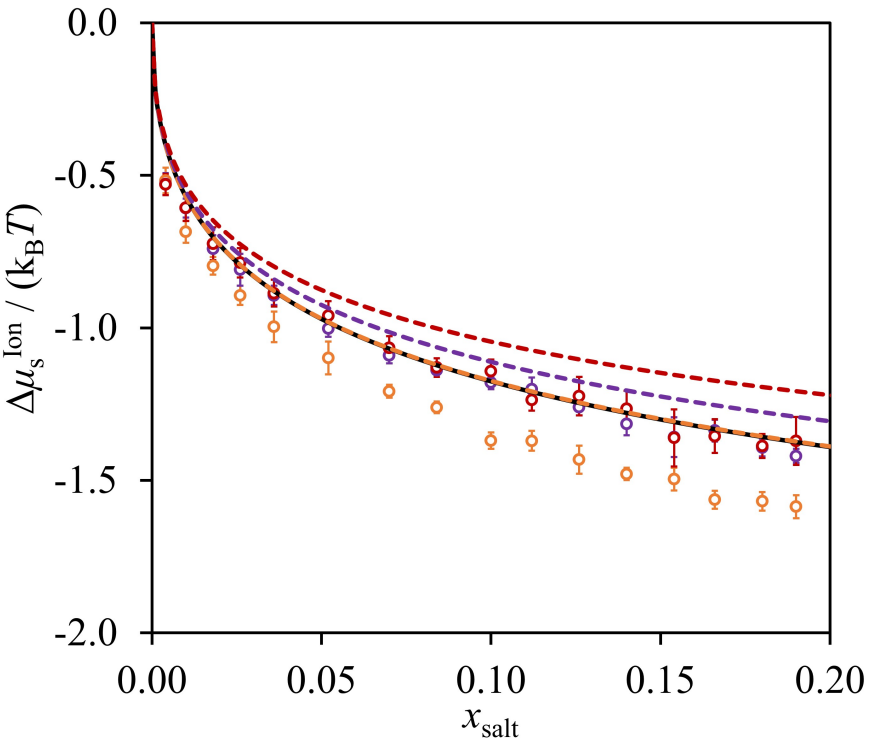


Figure 5.4: Coulomb contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Symbols denote EETM simulations of a salt in which the anion is a dimer consisting of the negatively charged ion segment bonded to a neutral monomer segment of different sizes. Different colours represent different sizes of the neutral monomer segment relative to the size of the ion ($\sigma_{\mathrm{seg}}/\sigma_{\mathrm{ion}}$): 0.25 (orange); 1 (purple); and 1.5 (dark red). The solid black curve denotes MSA calculations using Model G and dashed curves denote MSA calculations using Model M.

Table 5.3: EE simulation data of the ionic chemical potential ($\Delta\mu_s^{Ion}$) of systems containing dimer anions, which have different segment-to-ion diameter ratios ($\sigma_{seg-seg}/\sigma_{ion-ion}$), as a function of salt mole fraction (x_s).

$\sigma_{seg-seg}/\sigma_{ion-ion} = 0.25$			$\sigma_{seg-seg}/\sigma_{ion-ion} = 1.5$		
x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)	x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)
0.004	-0.517	0.042	0.004	-0.530	0.035
0.010	-0.685	0.036	0.010	-0.606	0.044
0.018	-0.797	0.029	0.018	-0.725	0.054
0.026	-0.894	0.031	0.026	-0.787	0.048
0.036	-0.997	0.050	0.036	-0.886	0.044
0.052	-1.10	0.054	0.052	-0.960	0.048
0.070	-1.21	0.021	0.070	-1.07	0.038
0.084	-1.26	0.019	0.084	-1.13	0.031
0.100	-1.37	0.027	0.100	-1.14	0.038
0.112	-1.37	0.033	0.112	-1.24	0.035
0.126	-1.43	0.046	0.126	-1.22	0.063
0.140	-1.48	0.020	0.140	-1.27	0.059
0.154	-1.50	0.037	0.154	-1.36	0.093
0.166	-1.56	0.030	0.166	-1.36	0.055
0.180	-1.57	0.031	0.180	-1.39	0.040
0.190	-1.59	0.037	0.190	-1.37	0.079

5.6.3 Effect of charge position

The simulation results in the previous sections (5.6.1 and 5.6.2) showcase the influence of two variables on the ionic chemical potential: the chain length and the size of uncharged segments. Both reflect the impact of charge shielding due to the presence of the neutral segments that are bonded to the charged segment. This effect can be observed in the fact that increasing the chain length, as shown in Figure 5.2, has no discernable impact on the ionic chemical potential of the salt (cf. Figure 5.3), since the charge shielding is the same in all of the salts shown. Charge shielding also explains the less negative values in the ionic chemical potential when the size of the uncharged segment increases relative to the ionic particle (cf. Figure 5.4). In this section, the concept of charge shielding is further explored by conducting three additional independent EE simulations of ionic molecules in which the charge shielding is increased. Specifically, simulations are conducted for ions in which all the uncharged segments are bonded to the ionic particle, as represented in Figure 5.5. The monomer and dimer anions used here are identical to the ones presented in previous sections. The trimer anion is formed by bonding

the charged particle to two uncharged segments– one on each side. The tetramer anion is formed by bonding the charged particle to three charged segments–forming a trigonal planar ionic molecule with the charge at the centre. The pentamer is formed by bonding the charged particle to four uncharged segments–forming a tetrahedral ionic molecule with the charge at the centre. Similar to the simulations in the previous two sections, the simulations here are run for salt concentrations of up to 0.2 in salt mole fraction at 420 K and a reduced Bjerrum length of 1.85.

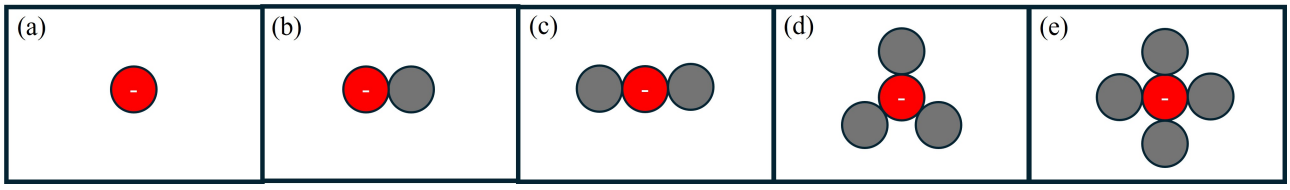


Figure 5.5: Schematic diagram representing the anionic species modeled to study the influence of charge shielding: (a) Monomer anion; (b) Dimer anion; (c) Trimer anion with the charge at the centre; (d) Tetramer anion with the charge at the centre; and (e) Pentamer anion with the charge at the centre. Note that this is a 2D representation; in the simulation models, the pentamer ion forms a tetrahedron.

The EE simulation data of the ionic chemical potential for the trimer, tetramer, and pentamer anions–all with the charges at the centre–are shown alongside simulation data of the monomer and dimer anions in Figure 5.6. As can be seen in Figure 5.6(a), the value of the chemical potential gets increasingly less negative when going from the dimer anion to the pentamer anion. The secondary axis in Figure 5.6(a) represents the segment packing fraction in the system, given by

$$\eta_{\text{seg}} = \frac{N_{\text{seg}} V_{\text{seg}}}{V_{\text{box}}}, \tag{5.25}$$

where N_{seg} is the total number of segments in the system (ions, uncharged segments, and solvent particles), $V_{\text{seg}} = 4\pi\sigma_{\text{seg}}^3/3$ is the volume of each segment, and V_{box} is the total volume of the simulation box. The simulation data is presented for salt mole fractions of up to 0.125 for the tetramer anion and up to 0.1 for the pentamer anion. Beyond these concentrations, the tetramer and pentamer salt solutions exhibit precipitation as evidenced by calculations of the mean squared displacement (MSD) and supported by calculations of the packing fraction, as

shown in Figure 5.6(a).

MSA calculations of the ionic chemical potential using Models G and M are presented in Figure 5.6(b), alongside the same simulation data. As shown in previous sections, chemical potential calculations for all the salt systems are identical using Model G. Model M calculations here are identical to the ones done for the linear oligomeric anions (cf. Figure 5.3).

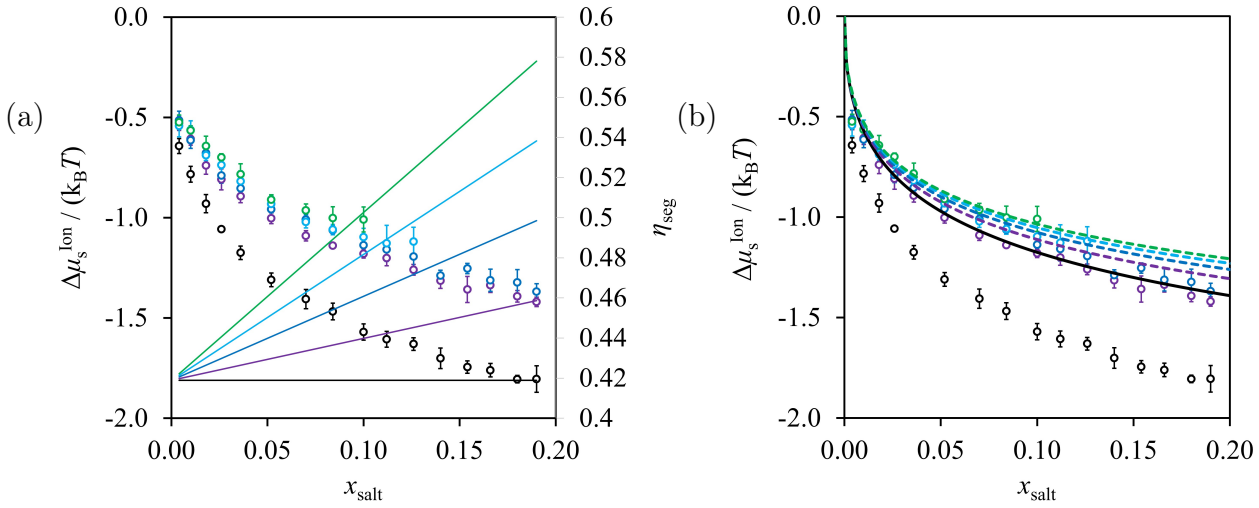


Figure 5.6: Coulomb contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Symbols denote EETM simulations of salts in which the anion is: a monomer (black); a dimer (purple); a trimer (blue); a tetramer (light blue); and a pentamer (green). The anion's charged segment is always in the middle of the molecule. (a) EETM simulation data of the ionic chemical potential of the salt solutions considered. Solid lines represent calculations of the segment packing fraction, η_{seg}. (b) EETM simulation data alongside MSA calculations using Model G (black solid curve) and Model M (dashed curves).

5.6.3.1 Scaling using the free-volume fraction

Even though calculations of the ionic chemical potential using Model M capture qualitatively the differentiation between the different branched oligomeric anions, they do so as a result of the averaging over the size of the molecule (cf. Equation 3.11). This same averaging is also the reason Model M yields different values of the chemical potential for the linear oligomeric anions; a result that is not supported by simulation data. More specifically, the averaging over the size of the molecule in Model M does not include a distinction between linear and branched molecules, as it calculates the effective diameter of the ionic molecule simply based on the number of segments present in it and their corresponding segment diameters. Inversely, Model

Table 5.4: EE simulation data of the ionic chemical potential ($\Delta\mu_s^{Ion}$) of systems containing trimer, tetramer, and pentamer anions, in which the charge is at the centre, as a function of salt mole fraction (x_s).

Trimer			Tetramer		
x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)	x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)
0.004	-0.515	0.046	0.004	-0.543	0.054
0.010	-0.615	0.040	0.010	-0.563	0.046
0.018	-0.678	0.025	0.018	-0.688	0.028
0.026	-0.790	0.038	0.026	-0.738	0.039
0.036	-0.854	0.027	0.036	-0.818	0.051
0.052	-0.959	0.040	0.052	-0.932	0.042
0.070	-1.01	0.026	0.070	-1.02	0.031
0.084	-1.06	0.023	0.084	-1.06	0.025
0.100	-1.14	0.033	0.100	-1.10	0.044
0.112	-1.16	0.032	0.112	-1.13	0.090
0.126	-1.19	0.069	0.126	-1.12	0.071
0.140	-1.29	0.027			
0.154	-1.25	0.025			
0.166	-1.31	0.055			
0.180	-1.32	0.065			
0.190	-1.37	0.039			

Pentamer		
x_s	$\Delta\mu_s^{Ion}$ / (k_BT)	Uncertainty / (k_BT)
0.004	-0.524	0.036
0.010	-0.565	0.024
0.018	-0.643	0.049
0.026	-0.700	0.018
0.036	-0.784	0.052
0.052	-0.910	0.024
0.070	-0.964	0.032
0.084	-1.00	0.057
0.100	-1.01	0.063

G calculations show no differentiation between any of the salts simulated in this work; this is because, according to Model G, the ionic free energy depends solely on the size parameter of the ionic particle (i.e., charged segment). While the calculations of Model G capture the near-identical simulation data for the different linear oligomeric anions, it fails to capture the impact of charge shielding present in the branched ionic molecules.

Here, an attempt is made to investigate the charge shielding effect observed in the simulation data of the branched chain anions (cf. Figure 5.6) by scaling the molecular simulation data of the monomer (reference) using a factor corresponding to the free volume around the ionic charge. The assumption is that every uncharged segment that is bonded to the ionic segment shields it from interacting via the Coulomb potential within a spherical cone that is proportional to the size of the uncharged segment, as shown in Figure 5.7. This assumption is justified by the fact that the presence of an uncharged segment bonded to the charged particle sterically hinders the access of a counterion within the volume occupied by the uncharged segment. As a consequence, the effective Coulombic force between the bonded charged segment and a counterion in solution is dampened. The spherical cone, with its vertex at the centre of the ionic particle and its base at the cut-off radius of the potential, represents the volume around the ionic particle that is shielded from interacting with the counter ion. Consequently, the volume available for interaction is calculated as the volume of the sphere with a radius equal to the potential cut-off radius minus the volume of the cones shielded by the bonded uncharged segments. The fraction of free (available) volume can thus be calculated for each n-mer ionic molecule as:

$$\varphi_{\text{free,n}} = \frac{V_{\text{cutoff}} - nV_{\text{sh}}}{V_{\text{cutoff}}}, \tag{5.26}$$

where V_{cutoff} is the volume of the sphere concentric with the ion and with a radius equal to the cut-off radius (r_{c}) of the Coulomb potential: $V_{\text{cutoff}} = 4\pi r_{\text{c}}^3/3$, V_{sh} is the volume of the spherical conic section shielded from Coulombic interaction due to the bonded uncharged segment and is calculated as the difference between the larger ($r = r_{\text{c}}$) and smaller ($r = \sigma_{ii}/2$) cones, i.e., $V_{\text{sh}} = 2\pi(r_{\text{c}}^3 - (\sigma_{ii}/2)^3)[1 - \cos(\phi)]/3$. n is the number of uncharged segments bonded to the

ion, where $\phi = 45^\circ$ is the cone aperture angle for the tangent segment, i.e., the angle between the axis of the cone and its rim. A schematic diagram illustrating the calculation of free volume is shown in Figure 5.7. The fraction of free volume (φ_{free}) is calculated to be approximately: 0.85 for the dimer (n=1) anion; 0.71 for the trimer (n=2) anion; 0.56 for the tetramer (n=3) anion; and 0.41 for the pentamer (n=4) anion.

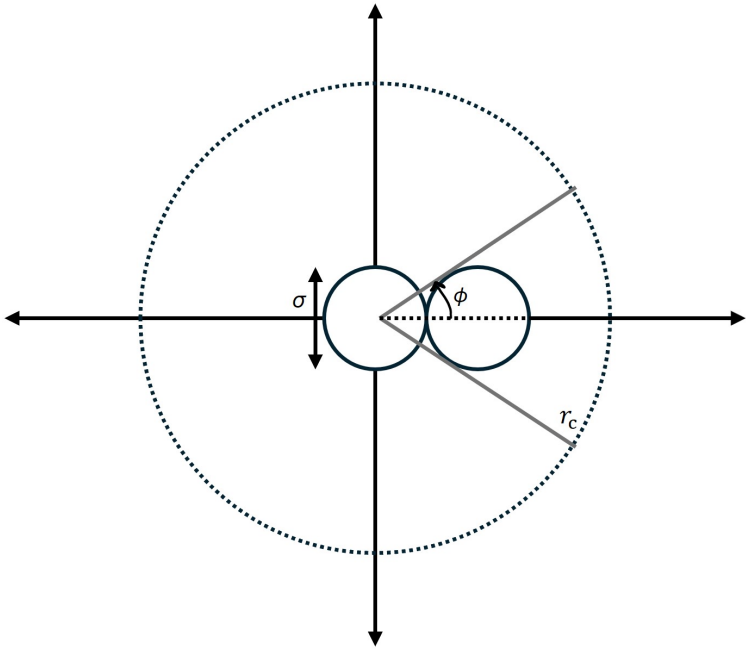


Figure 5.7: A 2D representation of the volume around the central charge that is shielded from Coulombic interactions due to the presence of the uncharged segment. The shielded volume is the spherical conic section between the surface of the central charge at $r = \sigma_{ii}/2$ and the cut-off radius $r = r_c$ and is defined by the aperture angle ϕ.

The ionic chemical potential ($\Delta\mu_s^{\text{Ion}}$) of a salt of stoichiometry MX (e.g., NaCl) is calculated as the sum of the ionic chemical potentials of the cation ($\Delta\mu_{\text{cat}}^{\text{Ion}}$) and anion ($\Delta\mu_{\text{an}}^{\text{Ion}}$) such that

$$\Delta\mu_s^{\text{Ion}} = \Delta\mu_{\text{cat}}^{\text{Ion}} + \Delta\mu_s^{\text{Ion}}\,. \tag{5.27}$$

The fraction of free volume (φ_{free}) is used to calculate a scaled ionic chemical potential of the n-mer anions by taking the monomer anion as the reference, such that:

$$\Delta\mu_{\text{s,n-mer}}^{\text{Ion}} = \Delta\mu_{\text{cat}}^{\text{Ion}} + \varphi_{\text{free,n}}\Delta\mu_{\text{an,mono}}^{\text{Ion}}, \tag{5.28}$$

where $\Delta\mu^{\mathrm{Ion}}_{\mathrm{s,n-mer}}$ is the ionic chemical potential of the n-mer salt and $\Delta\mu^{\mathrm{Ion}}_{\mathrm{an,mono}}$ is the ionic chemical potential of the reference spherical monomer anion. Note that the scaling factor ($\varphi_{\mathrm{free,n}}$) is only applied to the anion and not the cation, since only the former is bonded to uncharged segments and thus experiences the charge shielding effect.

In Figure 5.8, the simulation data of branched oligomeric anions is illustrated alongside the scaled data and MSA calculations, using both Model G and Model M. The scaled chemical potential points are in excellent agreement with data of the pentameric anion (Figure 5.8(d)) and increasingly deviate from the data as the shielding decreases; i.e., as one goes from the pentameric anion to the dimeric anion. This is likely because at the limit of minimum shielding, i.e., in the case of the dimeric anion, other counter ions in the solution may contribute to and increase the total Coulombic energy. Whereas in the limit of maximum shielding, i.e., in the case of the pentameric anion, the dominant contribution to the Coulombic potential comes from pairwise interactions. The latter is due to the steric hindrance imposed by the four uncharged segments bonded to the ion, which makes it harder for other counter ions in the solution to interact with the ion at the centre of the pentamer.

5.7 Conclusions

In this chapter, mean spherical approximation (MSA) calculations of the ionic chemical potential are assessed against expanded-ensemble (EE) simulation data of various solutions of salts made up of ionic chain molecules, with a particular focus on benchmarking the performance of Models G and M. The study employs a coarse-grained (CG) representation of water that excludes explicit polar interactions, allowing for an isolated evaluation of the ion–ion interaction.

The systems under investigation comprise both linear and branched ionic oligomers, enabling a systematic exploration of the effects of molecular size on ion—ion interactions. Key variables such as chain length, the size of uncharged segments, and electrostatic shielding are examined to understand their influence on the ionic chemical potential. The results reveal notable differences between the two models in capturing these effects.

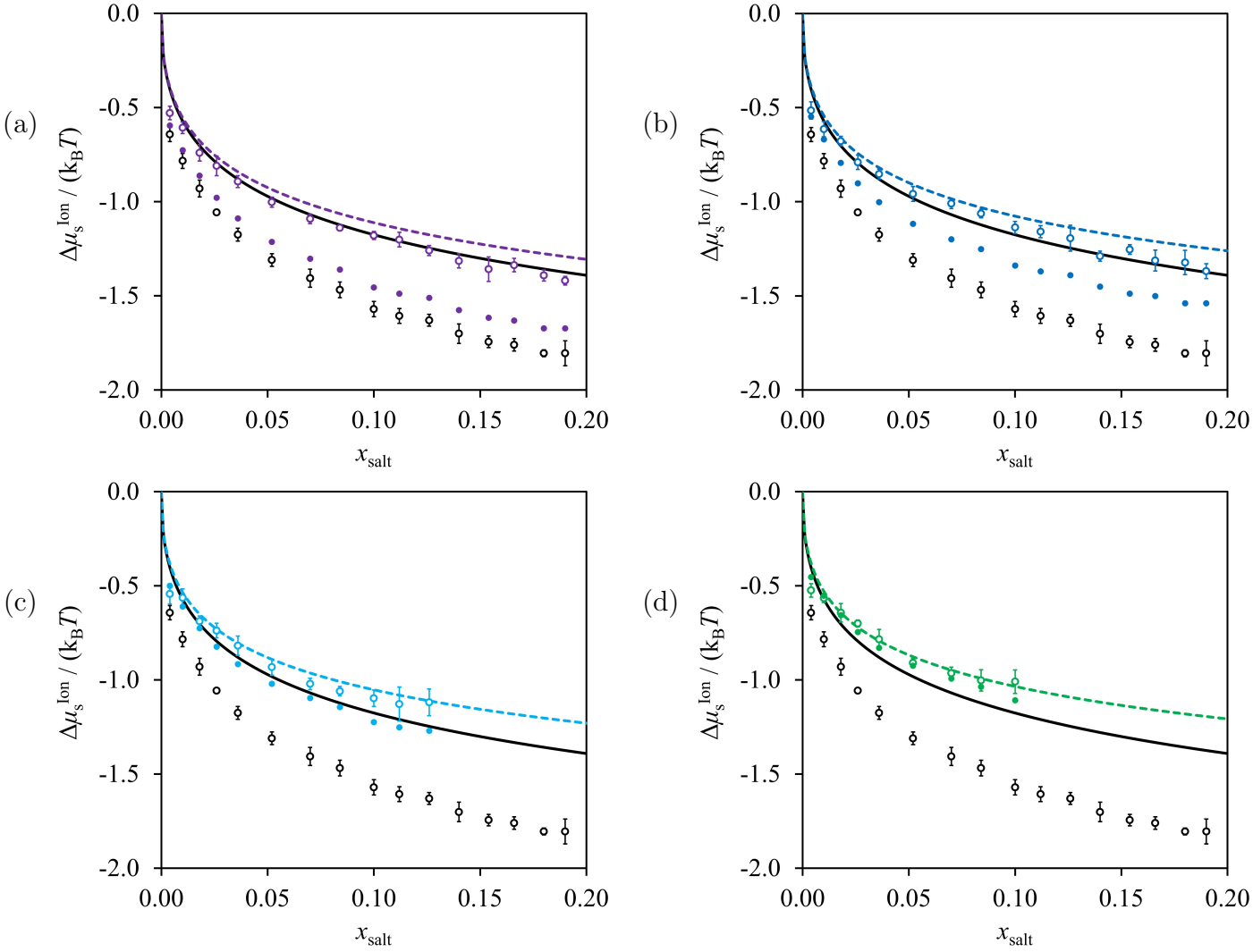


Figure 5.8: Coulombic contribution to the chemical potential as a function of salt mole fraction at $T^* = 0.75$, $\rho^* = 0.8$, and $l_B^* = 1.85$. Empty symbols denote EETM simulations of salts in which the anion is a branched oligomer. Filled symbols are chemical potential simulation data of the monomeric anion scaled by φ_{free}. The reference monomer anion simulation data are denoted by empty black circles. The black curve denotes MSA calculations using Model G whereas the dashed curves represent calculations using Model M. (a) Chemical potential of the dimer-anion salt. (b) Chemical potential of the trimer-anion salt. (c) Chemical potential of the tetramer-anion salt. (d) Chemical potential of the pentamer-anion salt.

Model G, in which the calculation of the ionic free energy is based solely on the size of the ion particle, yields identical results for all the salt solutions considered. This is in agreement with simulation data for linear oligomeric anions, in which the ion particle experiences the same amount of shielding regardless of the length of the uncharged chain bonded to it. On the other hand, Model G fails to capture the influence of the size of the uncharged segment within the ionic molecule due to its dependence on the size of the ionic particle only. For the same reason, Model G fails to differentiate between different branched oligomeric ions, which exhibit different levels of charge shielding.

Model M, in which the calculation of the ionic free energy takes into account the average size of the ionic molecule, captures the influence of the size of the uncharged chain bonded to the ion, but incorrectly differentiates between the linear oligomeric ions. Even though Model M qualitatively captures the change in chemical potential between the different branched oligomeric anions, it does not take into account the underlying cause, namely charge shielding.

Finally, in an attempt to further explore the effect of charge shielding, a simple yet effective scaling approach based on the concept of free volume is proposed. The scaled points, based on the reference monomeric-ion data, provide useful insights into the underlying principle of charge shielding and a potential direction in future research involving the treatment of the electrostatic interactions of ionic molecules within primitive-model expressions.

Chapter 6

Conclusions

Amino acids and oligopeptides are biomolecules of great importance in the pharmaceutical industry, specifically for the development of novel active pharmaceutical ingredients (APIs). Therefore, there is a high level of demand for thermodynamic models that can produce accurate solid–liquid solubility predictions for amino acids and oligopeptides across various solvent media and wide ranges of conditions. This is particularly important due to the scarcity of experimental data and the sheer number of potential oligopeptides that can be synthesised.

In this thesis, the SAFT-γ Mie group-contribution (GC) approach was used to predict the solid–liquid solubility of amino acids and oligopeptides in pure and mixed solvents at the isoelectric point (pI) and under changing pH and across wide temperature ranges. To achieve that, like and unlike interaction parameters were optimised by adjusting to experimental data of chemically related systems, demonstrating a high degree of transferability. Chemical equilibrium equations were solved alongside the solid–liquid equilibrium equation to predict the speciation of amino acids in multicomponent systems as a function of pH. Despite scatter in the reported experimental data of solubility, especially in some alcohols and mixed alcohol + water solvents, the model overall yields satisfactory predictions of solubility. To assess the treatment of charge interactions in ionic chain molecules using the SAFT-γ Mie model, a molecular simulations study was carried out to benchmark the Coulombic interactions between ions in systems which contain ionic chains.

6.1 Summary of thesis achievements

1. The Helmholtz free-energy equations used to calculate the thermodynamic properties of both neutral and charged systems in the SAFT-γ Mie GC approach are provided. A detailed discussion of the electrostatic contributions to the Helmholtz free energy is presented, comparing the two possible approaches, Model G and Model M, for resolving the charge on ionic chain molecules.

2. The SAFT-γ Mie GC framework is applied to the amino acids and oligpeptides of interest. The groups needed to model the amino-acid and oligopeptide systems are highlighted, and the needed like and unlike interaction parameters are presented along with calculations and experimental data of the pure compounds and mixtures included in the optimisation.

3. 11 new interactions are optimised by adjusting to experimental data, including the interactions of the peptide (CONH) group, which is newly added to the SAFT-γ Mie database. In total, around 11 pure compounds and 43 mixtures are described with great accuracy.

4. The SAFT-γ Mie GC framework is used to predict the solid–liquid solubility of a number of amino acids and oligopeptides in pure and mixed solvents, across a range of temperatures and under changing pH.

5. The solid–liquid solubility is predicted, at the isoelectric point, for the amino acids glycine, alanine, valine, and serine in pure water and primary and secondary alcohols. Despite the variability in the reported experimental solubility data, the model yields fairly accurate results. The models are then extended to predict, with good accuracy, the solubility of the aforementioned amino acids in binary water-alcohol solvents without the use of mixture-dependent interaction parameters. The newly optimised peptide (CONH) group parameters are used to predict the aqueous solubility of 17 di- and tri-peptides made up of the amino acids glycine, alanine, leucine, and serine.

6. The solubilities of glycine and alanine are predicted as a function of pH at different temperatures. The chemical-equilibrium equations governing the speciation of the amino

acids are presented in detail, highlighting the impact of temperature change on pH-dependent solubility. A system of equations, which consists of the SLE equation, chemical-equilibrium equations, and material and charge balances, is solved to model the 7-component system and produce accurate predictions of pH-dependent solubility at different temperatures. The impact of including the activity of water, a_{W}, is discussed by presenting calculations of a_{W} as a function of pH and temperature.

7. Expanded-ensemble transition matrix (EETM) simulations are conducted to benchmark the theoretical calculations of the ionic chemical potential ($\Delta\mu_{\mathrm{s}}^{\mathrm{Ion}}$) for various model systems containing ionic chain molecules and zwitterions. The ionic contribution to the chemical potential of ionic molecules of variable chain length, relative monomeric segment size, and charge placement is studied. The study focuses on the impact of charge screening in ionic chain molecules on the chemical potential.

8. A comparison of the two approaches used to resolve charge interactions in ionic molecules (Model G and Model M), which are discussed in Chapter 3, is presented in comparison with the EETM simulation results.

9. A simple empirical correlation is proposed in an attempt to explain the effect of charge screening observed in the simulation results.

6.2 Future research

In the context of obtaining the SAFT-γ Mie model parameters of the relevant groups, mixture densities were used. To better represent the impact of newly obtained parameters on calculations of mixture volumetric properties, the use of partial molar volume should be considered. Such a choice is more suitable because it represents the influence of the cross interaction between groups while removing the influence of like-group parameters. In the context of predicting the solid–liquid solubility of amino acids and oligopeptides at the isoelectric point (pI), future studies should explore the use of charged groups to model the zwitterions. In this work, a decision was made to treat the groups comprising the amino-acid and oligopeptide zwitterions as

fully neutral. This simplifying assumption, while widely used in literature, does not accurately encapsulate the physical intermolecular behaviour of zwitterions. The incorporation of explicit charge interactions might, for example, be the key to capturing the large drop in solubility that is observed when moving from pure water to pure alcohol solvents. Moreover, to overcome the challenge of obtaining reliable experimental melting-property data, future efforts should be invested in estimating the fusion enthalpy and melting temperature from molecular simulations or density functional theory (DFT) methods. In the case of modelling the thermodynamic properties of oligopeptide solutions, future studies should investigate the potential presence of intramolecular hydrogen bonding (IMHB) and its impact on the solid–liquid solubility. Future studies should extend the SAFT-γ Mie parameters to expand the list of amino acids, oligopeptides, and solvents that can be modelled. For example, an interesting area to investigate is the speciation of polyprotic amino acids and oligopeptides – those containing ionisable side groups – and the influence those ionisable groups have on the solid–liquid solubility. Other adjacent areas of research that can be explored include modelling the solubility in multi-solute amino acid mixtures and the solubility of amino acids and oligopeptides in salt solutions; these are particularly important for protein drug formulation and stability as well as in the solubility optimisation of peptide therapeutics.

More broadly, future work should investigate the relative performance of the current model compared to other predictive and group-contribution methods, such as UNIFAC and COSMO-RS. Due to the scarcity of reliable data, the use of average absolute deviations may not be enough to benchmark the performance of such models. Therefore, future effort should be expended in finding reliable metrics to judge the accuracy and reliability of such methods in predicting the solubility of amino acids and oligopeptides.

Given that the current study in Chapter 5 only focuses on the ionic contribution to the Helmholtz free energy, future research should extend the scope of this work to include other electrostatic contributions that are important in electrolyte thermodynamics. Future research should benchmark the theoretical calculations of the Born term, accounting for ion–solvent interactions (solvation), by incorporating a polar model for water in the molecular simulations.

6.3 Publications

1. **Alyazidi, A.**, Paliwal, S., Perdomo, F.A., Mead, A., Guo, M., Heng, J.Y., Bernet, T., Haslam, A.J., Adjiman, C.S., Jackson, G. and Galindo, A. Predicting the Solubility of Amino Acids and Peptides with the SAFT-γ Mie Approach: Neutral and Charged Models. *Industrial & Engineering Chemistry Research*, **63**(46), 20397-20423 (2024).

2. Bernet, T., Paliwal, S., **Alyazidi, A.**, Standish, R., Haslam, A.J., Adjiman, C.S., Jackson, G. and Galindo, A. Modeling the thermodynamic properties of cyclic alcohols with the SAFT-γ Mie approach: application to cyclohexanol and menthol systems. *Molecular Physics*, **122**(21-22), e2394132 (2024).

3. Paliwal, S., Perdomo, F.A., Bernet, T., **Alyazidi, A.**, Haslam, A.J., Jackson, G. and Galindo, A. Extending the SAFT-γ Mie group-contribution approach to model cyclic ketones: application to monoterpenoids. (In press, Molecular Physics) (2025).

Appendix A

Details of the electrostatic contributions to the the Helmholtz free-energy expression

A.1 Electrostatic contributions to the SAFT-γ Mie Helmholtz free energy: Model M

We describe here the treatment of ionic molecules when adopting the Molecular Model (M), as described in Figure 3.1 of the thesis. The derivation, which is largely analogous to that relating to Model G, follows a similar path beginning with the ion term from the SAFT-γ Mie free energy (equation 10, reproduced here as equation (A.1)):

$$\frac{A^{\text{ion}}}{Nk_{\text{B}}T} = \frac{U^{\text{MSA}}}{Nk_{\text{B}}T} + \frac{\Gamma^3}{3\pi\rho}, \tag{A.1}$$

where U^{MSA} is the MSA contribution to the internal energy U, $\rho = N/V$ is the number density of molecules in the system, V is the system volume, and Γ is the screening length of the

electrostatic forces. Considering spherical ions, U^{MSA} is given by

$$\frac{U^{\mathrm{MSA}}}{Nk_{\mathrm{B}}T} = -\frac{e^2}{(4\pi\epsilon_0)\rho k_{\mathrm{B}}TD}\left[\Gamma\rho\sum_{i\in I}\left(\frac{x_i Z_i^2}{1+\Gamma\sigma_{ii}}\right)+\frac{\pi}{2\Delta}\Omega P_n^2\right], \tag{A.2}$$

where I is the set of ionic species in the mixture and

$$\Delta = 1-\frac{\pi\rho}{6}\sum_{i\in I}x_i\sigma_{ii}^3, \tag{A.3}$$

$$P_n = \frac{\rho}{\Omega}\sum_{i\in I}\frac{x_i\sigma_{ii}Z_i}{1+\Gamma\sigma_{ii}}, \tag{A.4}$$

and

$$\Omega = 1+\frac{\pi\rho}{2\Delta}\sum_{i\in I}\frac{x_i\sigma_{ii}^3}{1+\Gamma\sigma_{ii}}. \tag{A.5}$$

Δ represents the packing fraction of the ions as a function of the diameters σ_{ii} ($i \in I$). P_n and Ω are coupling parameters. The coupling parameter P_n relates to the charge of the ions, whereas Ω couples to the packing fraction of the ions. Both parameters are functions of the ionic diameters, as well as the screening length, Γ:

$$\Gamma^2 = \frac{\pi e^2\rho}{(4\pi\epsilon_0)\,Dk_{\mathrm{B}}T}\sum_{i\in I}x_i Q_i^2, \tag{A.6}$$

where the effective charge $Q_i(\Gamma)$ is given by

$$Q_i = \frac{Z_i-\sigma_{ii}^2 P_n\left(\pi/\left(2\Delta\right)\right)}{1+\Gamma\sigma_{ii}}. \tag{A.7}$$

In adopting Model M one each non-spherical ion onto a single, large sphere of diameter

$$\tilde{\sigma}_{\mathrm{eff},\,ii}^{\mathrm{MSA,M}} = \left(\sum_{k=1}^{N_{\mathrm{G}}}\nu_{k,i}\nu_k^* S_k\sigma_{kk}^3\right)^{1/3},\ i\in I \tag{A.8}$$

and σ_{ii} is replaced by this effective $\tilde{\sigma}_{\mathrm{eff},\,ii}^{\mathrm{MSA,M}}$ in equations (A.2) to (A.7).

Adopting the Born model, one assumes that a spherical cavity of diameter $\sigma_{ii}^{\mathrm{Born}}$ is created for each ion, i, (independently of any others) in the dielectric medium. This leads to a

contribution to the Helmholtz free energy given by:

$$A^{\text{Born}} = -\frac{e^2}{4\pi\epsilon_0}\left(1-\frac{1}{D}\right)\sum_{i\in I}\frac{N_i Z_i^2}{\sigma_{ii}^{\text{Born}}}. \tag{A.9}$$

The Born diameter $\tilde{\sigma}_{\text{eff},\,ii}^{\text{Born, M}}$ of the single large sphere used to map each non-spherical ion in model M is given by:

$$\tilde{\sigma}_{\text{eff},\,ii}^{\text{Born, M}} = \left(\sum_{k=1}^{N_{\text{G},i}} \nu_{k,i}\nu_k^* S_k \left(\sigma_{kk}^{\text{Born}}\right)^3\right)^{1/3}, \; i \in I$$

where $N_{\text{G},i}$ represents the total number of groups making up the ionic molecule of species i, and $\nu_{k,i}$ is the number of groups of type k within the ionic molecule.

This larger diameter is then incorporated into the Born free-energy expression:

$$A^{\text{Born, M}} = -\frac{e^2}{4\pi\epsilon_0}\left(1-\frac{1}{D}\right)\sum_{i\in I}\frac{N_i Z_i^2}{\tilde{\sigma}_{\text{eff},\,ii}^{\text{Born, M}}}, \tag{A.10}$$

or, equivalently (in terms of mole fractions, and dividing through in customary fashion by $k_{\text{B}}T$ to give a dimensionless free energy),

$$\frac{A^{\text{Born, M}}}{Nk_{\text{B}}T} = -\frac{e^2}{4\pi\epsilon_0 k_{\text{B}}T}\left(1-\frac{1}{D}\right)\sum_{i\in I}\frac{x_i Z_i^2}{\tilde{\sigma}_{\text{eff},\,ii}^{\text{Born, M}}}\,. \tag{A.11}$$

The Born contribution is thereby reduced by a factor of $\sim \frac{\tilde{\sigma}_{\text{eff},\,ii}^{\text{Born, M}}}{\sigma_{ii}^{\text{Born}}}$.

A.2 Expressions for the chemical potential and pressure due to the ion contribution of Model G

The algebra involved in taking the derivatives of A^{ion} to obtain the chemical potential, $\mu_i^{\text{MSA, G}}$, and pressure, $p^{\text{MSA, G}}$, is cumbersome, comprising multiple applications of the chain rule of calculus, although the procedure is simplified a little by recognising that the derivatives of

A^{ion} with respect to Γ are zero (see "Derivatives of the Helmholtz free energy with respect to screening length Γ"). The chemical potential of component i is thus given by

$$\begin{aligned}\frac{\mu_i^{\mathrm{ion,\,G}}}{k_{\mathrm{B}}T} = \frac{1}{k_{\mathrm{B}}T}\left(\frac{\partial A^{\mathrm{ion,\,G}}}{\partial N_i}\right)_{V,T,N_{j\neq i}} = & -\frac{e^2V}{4\pi\epsilon_0 D k_{\mathrm{B}}T}\left\{\frac{\Gamma}{V}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{\nu_{k,i}Z_k^2}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},\,kk}^{\mathrm{MSA,G}}\right)}\right. \\ & + \left(\frac{\partial A^{\mathrm{MSA,\,G}}}{\partial \Delta}\right)_{V,T,N_{j\neq i}}\left(\frac{\partial \Delta}{\partial N_i}\right)_{V,T,N_{j\neq i}} \\ & + \left(\frac{\partial A^{\mathrm{MSA,\,G}}}{\partial \Omega}\right)_{V,T,N_{j\neq i}}\left(\frac{\partial \Omega}{\partial N_i}\right)_{V,T,N_{j\neq i}} \\ & \left. + \left(\frac{\partial A^{\mathrm{MSA,\,G}}}{\partial P_n}\right)_{V,T,N_{j\neq i}}\left(\frac{\partial P_n}{\partial N_i}\right)_{V,T,N_{j\neq i}}\right\} \\ & - \frac{U^{\mathrm{MSA,G}}N}{Dk_{\mathrm{B}}T}\left(\frac{\partial D}{\partial N_i}\right)_{V,T,N_{j\neq i}}. \end{aligned} \tag{A.12}$$

The individual partial derivatives appearing in equation (A.12) are obtained as follows:

$$\left(\frac{\partial A^{\mathrm{MSA,\,G}}}{\partial \Delta}\right)_{V,T,N_{j\neq i}} = -\frac{\pi\Omega}{2}\left(\frac{P_n}{\Delta}\right)^2; \tag{A.13}$$

$$\left(\frac{\partial \Delta}{\partial N_i}\right)_{V,T,N_{j\neq i}} = \frac{-\pi}{6V}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\nu_{k,i}\left(\tilde{\sigma}_{\mathrm{eff},\,kk}^{\mathrm{MSA,\,G}}\right)^3; \tag{A.14}$$

$$\left(\frac{\partial A^{\mathrm{MSA,\,G}}}{\partial \Omega}\right)_{V,T,N_{j\neq i}} = \frac{\pi P_n^2}{2\Delta}; \tag{A.15}$$

$$\left(\frac{\partial \Omega}{\partial N_i}\right)_{V,T,N_{j\neq i}} = \frac{\pi}{2\Delta V}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{\nu_{k,i}Z_k^2}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},\,kk}^{\mathrm{MSA,G}}\right)} + \frac{(1-\Omega)}{\Delta}\left(\frac{\partial \Delta}{\partial N_i}\right)_{V,T,N_{j\neq i}}; \tag{A.16}$$

$$\left(\frac{\partial A^{\mathrm{MSA,\,G}}}{\partial P_n}\right)_{V,T,N_{j\neq i}} = \frac{\pi\Omega P_n}{\Delta}; \tag{A.17}$$

$$\left(\frac{\partial P_n}{\partial N_i}\right)_{V,T,N_{j\neq i}} = \frac{1}{\Omega V}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{\nu_{k,i}\tilde{\sigma}_{\mathrm{eff},\,kk}^{\mathrm{MSA,\,G}}Z_k}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},\,kk}^{\mathrm{MSA,G}}\right)} - \frac{P_n}{\Omega}\left(\frac{\partial \Omega}{\partial N_i}\right)_{V,T,N_{j\neq i}}. \tag{A.18}$$

The pressure is derived from equations 3.15 and 3.16 of the thesis, once again involving multiple applications of the chain rule. The volume derivatives of Δ, Ω and P_n are first required:

$$\left(\frac{\partial \Delta}{\partial V}\right)_{T,N} = \frac{\pi}{6V^2}\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}x_i\nu_{k,i}\left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^3 = \frac{1-\Delta}{V}; \tag{A.19}$$

$$\left(\frac{\partial\Omega}{\partial V}\right)_{T,N} = -\left(\frac{\pi}{2V^2\Delta}+\frac{\pi}{2V\Delta^2}\left(\frac{\partial\Delta}{\partial V}\right)\right)\sum_{i=1;Z_i\neq 0}^{N_\mathrm{C}}\sum_{k=1;Z_k\neq 0}^{N_\mathrm{G}}\frac{x_i\nu_{k,i}\left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^3}{1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}}$$
$$= \frac{1-\Omega}{V}+\left(\frac{1-\Omega}{\Delta}\right)\left(\frac{1-\Delta}{V}\right)=\frac{1-\Omega}{V\Delta}\,; \tag{A.20}$$

$$\left(\frac{\partial P_n}{\partial V}\right)_{T,N} = -\left(\frac{1}{V^2\Omega}+\frac{1}{V\Omega^2}\left(\frac{\partial\Omega}{\partial V}\right)\right)\sum_{i=1;Z_i\neq 0}^{N_\mathrm{C}}\sum_{k=1;Z_k\neq 0}^{N_\mathrm{G}}\frac{x_i\nu_{k,i}\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}Z_k}{1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}}$$
$$= -\frac{P_n}{V}-\frac{P_n}{\Omega}\left(\frac{1-\Omega}{V\Delta}\right). \tag{A.21}$$

For the purposes of the volume derivative, the final term in equation 3.16 of the thesis can be thought of as a constant $\times\, V\Omega P_n^2/\Delta$; the relevant derivative is thus obtained via

$$\frac{\partial}{\partial V}\left(\frac{V\Omega P_n^2}{\Delta}\right)_{T,N} = \frac{\Omega P_n^2}{\Delta}+V\left\{\Omega\frac{2P_n}{\Delta}\left(\frac{\partial P_n}{\partial V}\right)-\frac{P_n^2}{\Delta^2}\left(\frac{\partial\Delta}{\partial V}\right)+\frac{P_n^2}{\Delta}\left(\frac{\partial\Omega}{\partial V}\right)\right\}$$
$$= \ldots = \left(\frac{P_n}{\Delta}\right)^2(\Omega\Delta-2\Omega\Delta-2(1-\Omega)-\Omega(1-\Delta)+(1-\Omega))$$
$$= -\left(\frac{P_n}{\Delta}\right)^2. \tag{A.22}$$

One finally obtains

$$p^{\mathrm{ion,\,G}} = \left(-\frac{\partial A^{\mathrm{ion,\,G}}}{\partial V}\right)_{N,T}$$
$$= -\frac{\Gamma^3 k_\mathrm{B}T}{3\pi}-\frac{e^2}{8\epsilon_0 D}\left(\frac{P_n}{\Delta}\right)^2+\frac{U^{\mathrm{MSA,\,G}}}{D}\left(\frac{\partial D}{\partial V}\right)_{T,N}. \tag{A.23}$$

A.3 Derivatives of the Helmholtz free energy with respect to screening length Γ

(This demonstration is adapted from the PhD thesis of Schreckenberg [105].)

Derived thermodynamic properties in the case of the MSA (in the primitive model) are obtained by taking the derivative of the Helmholtz free energy with respect to the appropriate state

variable, such as volume, number of components or temperature. Applying the chain rule, the first derivative with respect to any state variable Ψ is obtained as

$$\begin{aligned}\frac{\partial A}{\partial \Psi} &= \left(\frac{\partial A}{\partial \Psi}\right)_{\Gamma,\Delta,\Omega,P_n} + \left(\frac{\partial A}{\partial \Gamma}\right)_{\Psi,\Delta,\Omega,P_n}\left(\frac{\partial \Gamma}{\partial \Psi}\right) + \left(\frac{\partial A}{\partial \Delta}\right)_{\Psi,\Gamma}\left(\frac{\partial \Delta}{\partial \Psi}\right) \\ &+ \left(\frac{\partial A}{\partial \Omega}\right)_{\Psi,\Gamma}\left(\frac{\partial \Omega}{\partial \Psi}\right) + \left(\frac{\partial A}{\partial P_n}\right)_{\Psi,\Gamma}\left(\frac{\partial P_n}{\partial \Psi}\right). \end{aligned} \tag{A.24}$$

We provide first the demonstration for the simpler case of Model M. Taking the derivative with respect to Γ in the second term on the right-hand side of equation (A.24), as applied to A^{ion} given by equations (A.1) and (A.2) for Model M leads to

$$\frac{\partial A^{\text{ion}}}{\partial \Gamma} = -\frac{Ve^2}{4\pi D\epsilon_0}\left[\frac{1}{V}\sum_{i\in I}\frac{N_i Z_i^2}{1+\Gamma\sigma_{ii}}\left(1-\frac{\Gamma\sigma_{ii}}{1+\Gamma\sigma_{ii}}\right)+\frac{\pi P_n}{2\Delta}\left(P_n\frac{\partial\Omega}{\partial\Gamma}+2\Omega\frac{\partial P_n}{\partial\Gamma}\right)\right]+\frac{VkT\Gamma^2}{\pi}. \tag{A.25}$$

From equation (A.4) we have

$$\begin{aligned}\frac{\partial P_n}{\partial \Gamma} &= \frac{-1}{\Omega^2 V}\frac{\partial\Omega}{\partial\Gamma}\sum_{i\in I}\frac{N_i\sigma_{ii}Z_i}{1+\Gamma\sigma_{ii}} - \frac{1}{\Omega V}\sum_{i\in I}\frac{N_i\sigma_{ii}^2 Z_i}{(1+\Gamma\sigma_{ii})^2} \\ &= \frac{-P_n}{\Omega}\frac{\partial\Omega}{\partial\Gamma} - \frac{1}{\Omega V}\sum_{i\in I}\frac{N_i\sigma_{ii}^2 Z_i}{(1+\Gamma\sigma_{ii})^2}, \end{aligned} \tag{A.26}$$

whence

$$\begin{aligned}\frac{\partial A^{\text{ion}}}{\partial \Gamma} &= -\frac{Ve^2}{4\pi D\epsilon_0}\left[\frac{1}{V}\sum_{i\in I}\frac{N_i Z_i^2}{(1+\Gamma\sigma_{ii})^2}+\frac{\pi P_n}{2\Delta}\left(P_n\frac{\partial\Omega}{\partial\Gamma}-\frac{2}{V}\sum_{i\in I}\frac{N_i\sigma_{ii}^2 Z_i}{(1+\Gamma\sigma_{ii})^2}-2P_n\frac{\partial\Omega}{\partial\Gamma}\right)\right] \\ &+ \frac{VkT\Gamma^2}{\pi}. \end{aligned} \tag{A.27}$$

Equation (A.5) provides

$$\frac{\partial\Omega}{\partial\Gamma} = \frac{-\pi}{2\Omega V}\sum_{i\in I}\frac{N_i\sigma_{ii}^4}{(1+\Gamma\sigma_{ii})^2}, \tag{A.28}$$

allowing cancellation of V from the numerator of the prefactor in equation (A.33) and leading

to

$$\frac{\partial A^{\text{ion}}}{\partial \Gamma} = -\frac{e^2}{4\pi D\epsilon_0}\left[\sum_{i\in I}\frac{N_i Z_i^2}{(1+\Gamma\sigma_{ii})^2} - \frac{\pi P_n}{\Delta}\sum_{i\in I}\frac{N_i\sigma_{ii}^2 Z_i}{(1+\Gamma\sigma_{ii})^2} + \frac{\pi^2 P_n^2}{4\Delta^2}\sum_{i\in I}\frac{N_i\sigma_{ii}^4}{(1+\Gamma\sigma_{ii})^2}\right] + \frac{VkT\Gamma^2}{\pi}\,. \tag{A.29}$$

Recalling equation (A.7) for Q_i, the terms in square brackets in equation (A.35) can be recognised as $\sum_{i\in I} N_i Q_i^2$, whereby this expression reduces to

$$\frac{\partial A^{\text{ion}}}{\partial \Gamma} = -\frac{e^2}{4\pi D\epsilon_0}\sum_{i\in I} N_i Q_i^2 + \frac{VkT\Gamma^2}{\pi} = 0\,. \tag{A.30}$$

The final equality follows upon substitution for Γ^2 using equation (A.6).

In the case of Model G, taking the derivative with respect to Γ in the second term on the right-hand side of equation (A.24), as applied to A^{ion} given by equations 3.15 and 3.16 of the thesis leads to

$$\frac{\partial A^{\text{ion}}}{\partial \Gamma} = -\frac{Ve^2}{4\pi D\epsilon_0}\left[\frac{1}{V}\sum_{i=1;Z_i\neq 0}^{N_{\text{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\text{G}}}\frac{N_i\nu_{k,i}Z_k^2}{1+\Gamma\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}}\left(1-\frac{\Gamma\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}}{1+\Gamma\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}}\right) + \frac{\pi P_n}{2\Delta}\left(P_n\frac{\partial\Omega}{\partial\Gamma}+2\Omega\frac{\partial P_n}{\partial\Gamma}\right)\right] + \frac{VkT\Gamma^2}{\pi}\,. \tag{A.31}$$

From equation 3.19 of the thesis we have

$$\begin{aligned}\frac{\partial P_n}{\partial\Gamma} &= \frac{-1}{\Omega^2 V}\frac{\partial\Omega}{\partial\Gamma}\sum_{i=1;Z_i\neq 0}^{N_{\text{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\text{G}}}\frac{N_i\nu_{k,i}\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}Z_k}{1+\Gamma\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}} \\ &- \frac{1}{\Omega V}\sum_{i=1;Z_i\neq 0}^{N_{\text{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\text{G}}}\frac{N_i\nu_{k,i}\left(\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}\right)^2 Z_k}{\left(1+\Gamma\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}\right)^2} \\ &= \frac{-P_n}{\Omega}\frac{\partial\Omega}{\partial\Gamma} - \frac{1}{\Omega V}\sum_{i=1;Z_i\neq 0}^{N_{\text{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\text{G}}}\frac{N_i\nu_{k,i}\left(\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}\right)^2 Z_k}{\left(1+\Gamma\tilde{\sigma}_{\text{eff},kk}^{\text{MSA,G}}\right)^2}\,,\end{aligned} \tag{A.32}$$

whence

$$\begin{aligned}\frac{\partial A^{\mathrm{ion}}}{\partial \Gamma} &= -\frac{Ve^2}{4\pi D\epsilon_0}\left[\frac{1}{V}\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{N_i\nu_{k,i}Z_k^2}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2}\right.\\ &+ \left.\frac{\pi P_n}{2\Delta}\left(P_n\frac{\partial\Omega}{\partial\Gamma}-\frac{2}{V}\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{N_i\nu_{k,i}\left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2 Z_k}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2}-2P_n\frac{\partial\Omega}{\partial\Gamma}\right)\right]\\ &+ \frac{VkT\Gamma^2}{\pi}\,. \end{aligned} \tag{A.33}$$

Equation 3.20 of the thesis provides

$$\frac{\partial\Omega}{\partial\Gamma}=\frac{-\pi}{2\Omega V}\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{N_i\nu_{k,i}\left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^4}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2}\,, \tag{A.34}$$

allowing cancellation of V from the numerator of the prefactor in equation (A.33) and leading to

$$\begin{aligned}\frac{\partial A^{\mathrm{ion}}}{\partial \Gamma} &= -\frac{e^2}{4\pi D\epsilon_0}\left[\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{N_i\nu_{k,i}Z_k^2}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2}\right.\\ &\quad -\frac{\pi P_n}{\Delta}\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{N_i\nu_{k,i}\left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2 Z_k}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2}\\ &\quad \left.+\frac{\pi^2P_n^2}{4\Delta^2}\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}\frac{N_i\nu_{k,i}\left(\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^4}{\left(1+\Gamma\tilde{\sigma}_{\mathrm{eff},kk}^{\mathrm{MSA,G}}\right)^2}\right]\\ &\quad +\frac{VkT\Gamma^2}{\pi}\,. \end{aligned} \tag{A.35}$$

Recalling equation 3.22 of the thesis for Q_k, the terms in square brackets in equation (A.35) can be recognised as $\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}N_i\nu_{k,i}Q_k^2$, whereby this expression reduces to

$$\frac{\partial A^{\mathrm{ion}}}{\partial\Gamma}=-\frac{e^2}{4\pi D\epsilon_0}\sum_{i=1;Z_i\neq 0}^{N_{\mathrm{C}}}\sum_{k=1;Z_k\neq 0}^{N_{\mathrm{G}}}N_i\nu_{k,i}Q_k^2+\frac{VkT\Gamma^2}{\pi}=0\,. \tag{A.36}$$

The final equality follows upon substitution for Γ^2 using equation 3.21 of the thesis.

Bibliography

[1] K. T. Savjani, A. K. Gajjar, and J. K. Savjani. Drug solubility: importance and enhancement techniques. International Scholarly Research Notices 2012 (1) (2012), 195727.

[2] B. J. Bruno, G. D. Miller, and C. S. Lim. Basics and recent advances in peptide and protein drug delivery. Therapeutic Delivery 4 (11) (2013), 1443–1467.

[3] K. Fosgerau and T. Hoffmann. Peptide therapeutics: current status and future directions. Drug Discovery Today 20 (1) (2015), 122–128.

[4] K. Sharma et al. Peptide-based drug discovery: Current status and recent advances. Drug Discovery Today 28 (2) (2023), 103464.

[5] F. G. Banting et al. Pancreatic extracts in the treatment of diabetes mellitus. Canadian Medical Association Journal 12 (3) (1922), 141–6.

[6] O. W. Press et al. Phase II trial of 131I-B1 (anti-CD20) antibody therapy with autologous stem cell transplantation for relapsed B cell lymphomas. The Lancet 346 (8971) (1995), 336–340.

[7] D. M. Goldenberg et al. Use of radiolabeled antibodies to carcinoembryonic antigen for the detection and localization of diverse cancers by external photoscanning. New England Journal of Medicine 298 (25) (1978), 1384–1388.

[8] A. C.-L. Lee et al. A comprehensive review on current advances in peptide drug development and design. International Journal of Molecular Sciences 20 (10) (2019), 2383.

[9] P. Matsson et al. Cell permeability beyond the rule of 5. Advanced Drug Delivery Reviews 101 (2016), 42–61.

[10] J. M. Prausnitz, R. N. Lichtenthaler, and E. G. De Azevedo. Molecular Thermodynamics of Fluid-Phase Equilibria. Pearson Education, 1998.

[11] E. Ruckenstein and I. Shulgin. Solubility of hydrophobic organic pollutants in binary and multicomponent aqueous solvents. Environmental science & technology 39 (6) (2005), 1623–1631.

[12] B. L. Rivas, B. F. Urbano, and J. Sánchez. Water-soluble and insoluble polymers, nanoparticles, nanocomposites and hybrids with ability to remove hazardous inorganic pollutants in water. Frontiers in chemistry 6 (2018), 320.

[13] L. Xu et al. Completely inorganic deep eutectic solvents for efficient and recyclable liquid–liquid interface catalysis. Advanced Materials 36 (29) (2024), 2313853.

[14] P. Crafts. "The role of solubility modeling and crystallization in the design of active pharmaceutical ingredients". *Computer Aided Chemical Engineering*. Vol. 23. Elsevier, 2007, 23–85.

[15] E. Q. Adams. Relations between the constants of dibasic acids and of amphoteric electrolytes. J. Am. Chem. Soc 38 (1916), 1503–1510.

[16] N. Bjerrum. Die Konstitution der Ampholyte, besonders der Aminosäuren, und ihre Dissoziationskonstanten. Zeitschrift für Physikalische Chemie 104 (1) (1923), 147–173.

[17] E. J. Cohn et al. Studies in the physical chemistry of amino acids, peptides and related substances. II. The solubility of α-amino acids in water and in alcohol–water mixtures. Journal of the American Chemical Society 56 (11) (1934), 2270–2282.

[18] T. L. McMeekin, E. J. Cohn, and J. H. Weare. Studies in the physical chemistry of amino acids, peptides and related substances. VII. A comparison of the solubility of amino acids, peptides and their derivatives. Journal of the American Chemical Society 58 (11) (1936), 2173–2181.

[19] V. P. Debye. Zur theorie der electrolyte. Phyikalishce Zeitschrift (1923), 185–206.

[20] J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. The Journal of Chemical Physics 2 (7) (1934), 351–361.

[21] G. M. Wilson. Vapor-liquid equilibrium. XI. A new expression for the excess free energy of mixing. Journal of the American Chemical Society 86 (2) (1964), 127–130.

[22] H. Renon and J. M. Prausnitz. Local compositions in thermodynamic excess functions for liquid mixtures. AIChE Journal 14 (1) (1968), 135–144.

[23] P. J. Flory. Principles of Polymer Chemistry. Cornell University Press, 1953.

[24] J. S. Rowlinson. The Theory of regular solutions. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 214 (1117) (1952), 192–206.

[25] D. S. Abrams and J. M. Prausnitz. Statistical thermodynamics of liquid mixtures: a new expression for the excess Gibbs energy of partly or completely miscible systems. AIChE Journal 21 (1) (1975), 116–128.

[26] E. Guggenheim. General theory—Introduction to general theory. Discussions of the Faraday Society 15 (1953), 24–27.

[27] C.-C. Chen, Y. Zhu, and L. B. Evans. Phase partitioning of biomolecules: solubilities of amino acids. Biotechnology Progress 5 (3) (1989), 111–118.

[28] K. S. Pitzer. Thermodynamics of electrolytes. I. Theoretical basis and general equations. The Journal of Physical Chemistry 77 (2) (1973), 268–277.

[29] C. J. Orella and D. J. Kirwan. Correlation of amino acid solubilities in aqueous aliphatic alcohol solutions. Industrial & Engineering Chemistry Research 30 (5) (1991), 1040–1045.

[30] L. A. Ferreira, E. A. Macedo, and S. P. Pinho. Solubility of amino acids and diglycine in aqueous–alkanol solutions. Chemical Engineering Science 59 (15) (2004), 3117–3124.

[31] M. T. Gude, L. A. M. van der Wielen, and K. C. A. M. Luyben. Phase behavior of α-amino acids in multicomponent aqueous alkanol solutions. Fluid Phase Equilibria 116 (1-2) (1996), 110–117.

[32] M. van Berlo et al. Partition coefficients and solubilities of glycine in the ternary solvent system 1-butanol+ ethanol+ water. Industrial & Engineering Chemistry Research 36 (6) (1997), 2474–2482.

[33] E. S. J. Rudolph et al. Solubilities and partition coefficients of semi-synthetic antibiotics in water+ 1-butanol systems. Industrial & Engineering Chemistry Research 40 (1) (2001), 398–406.

[34] G. Jackson, W. G. Chapman, and K. E. Gubbins. Phase equilibria of associating fluids: Spherical molecules with multiple bonding sites. Molecular Physics 65 (1) (1988), 1–31.

[35] W. G. Chapman, G. Jackson, and K. E. Gubbins. Phase equilibria of associating fluids: chain molecules with multiple bonding sites. Molecular Physics 65 (5) (1988), 1057–1079.

[36] W. G. Chapman et al. SAFT: Equation of state solution model for associating fluids. Fluid Phase Equilibria 52 (1989), 31–38.

[37] W. G. Chapman et al. New reference equation of state for associating liquids. Industrial & Engineering Chemistry Research 29 (8) (1990), 1709–1721.

[38] M. K. Khoshkbarchi and J. H. Vera. A simplified perturbed hard-sphere model for the activity coefficients of amino acids and peptides in aqueous solutions. Industrial & Engineering Chemistry Research 35 (11) (1996), 4319–4327.

[39] M. K. Khoshkbarchi and J. H. Vera. Effect of NaCl and KCl on the solubility of amino acids in aqueous solutions at 298.2 K: measurements and modeling. Industrial & Engineering Chemistry Research 36 (6) (1997), 2445–2451.

[40] A. Soto et al. Effect of the cation and the anion of an electrolyte on the solubility of DL-aminobutyric acid in aqueous solutions: measurement and modelling. Biophysical Chemistry 73 (1-2) (1998), 77–83.

[41] D. Fuchs et al. Solubility of amino acids: Influence of the pH value and the addition of alcoholic cosolvents on aqueous solubility. Industrial & Engineering Chemistry Research 45 (19) (2006), 6578–6584.

[42] J. Gross and G. Sadowski. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. Industrial & Engineering Chemistry Research 40 (4) (2001), 1244–1260.

[43] J. Marrero and R. Gani. Group-contribution based estimation of pure component properties. Fluid Phase Equilibria 183 (2001), 183–208.

[44] L. F. Cameretti and G. Sadowski. Modeling of aqueous amino acid and polypeptide solutions with PC-SAFT. Chemical Engineering and Processing: Process Intensification 47 (6) (2008), 1018–1025.

[45] L. A. Ferreira et al. Thermodynamic modeling of several aqueous alkanol solutions containing amino acids with the perturbed-chain statistical associated fluid theory equation of state. Industrial & Engineering Chemistry Research 48 (11) (2009), 5498–5505.

[46] J.-B. Grosse Daldrup et al. Measurement and modeling solubility of aqueous multisolute amino-acid solutions. Industrial & Engineering Chemistry Research 49 (3) (2010), 1395–1401.

[47] J.-B. Grosse Daldrup et al. Modeling pH and solubilities in aqueous multisolute amino acid solutions. Industrial & Engineering Chemistry Research 50 (6) (2011), 3503–3509.

[48] C. Held, L. F. Cameretti, and G. Sadowski. Measuring and modeling activity coefficients in aqueous amino-acid solutions. Industrial & Engineering Chemistry Research 50 (1) (2011), 131–141.

[49] K. Wysoczanska et al. Solubility of DNP-amino acids and their partitioning in biodegradable ATPS: Experimental and ePC-SAFT modeling. Fluid Phase Equilibria 527 (2021), 112830.

[50] M. Aliyeva et al. Salt effects on the solubility of aromatic and dicarboxylic amino acids in water. The Journal of Chemical Thermodynamics 177 (2023), 106929.

[51] V. Papaioannou et al. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. The Journal of Chemical Physics 140 (5) (2014), 054107.

[52] S. Dufal et al. Prediction of thermodynamic properties and phase behavior of fluids and mixtures with the SAFT-γ Mie group-contribution equation of state. Journal of Chemical & Engineering Data 59 (10) (2014), 3272–3288.

[53] S. Dufal et al. The A in SAFT: developing the contribution of association to the Helmholtz free energy within a Wertheim TPT1 treatment of generic Mie fluids. Molecular Physics 113 (9-10) (2015), 948–984.

[54] G. Jackson et al. Corrigendum: the A in SAFT: developing the contribution of association to the Helmholtz free energy within a Wertheim TPT1 treatment of generic Mie fluids. Molecular Physics 116 (2) (2018), 283–285.

[55] A. J. Haslam et al. Expanding the Applications of the SAFT-γ Mie Group-Contribution Equation of State: Prediction of Thermodynamic Properties and Phase Behavior of Mixtures. Journal of Chemical & Engineering Data 65 (12) (2020), 5862–5890.

[56] P. Hutacharoen et al. Predicting the solvation of organic compounds in aqueous environments: from alkanes and alcohols to pharmaceuticals. Industrial & Engineering Chemistry Research 56 (38) (2017), 10856–10876.

[57] M. Wehbe et al. Phase behaviour and pH-solubility profile prediction of aqueous buffered solutions of ibuprofen and ketoprofen. Fluid Phase Equilibria 560 (2022), 113504.

[58] T. Bernet et al. Modeling the Thermodynamic Properties of Saturated Lactones in Nonideal Mixtures with the SAFT-γ Mie Approach. Journal of Chemical & Engineering Data 69 (2) (2023), 650–678.

[59] S. Di Lecce et al. Modelling and prediction of the thermophysical properties of aqueous mixtures of choline geranate and geranic acid (CAGE) using SAFT-γ Mie. RSC Advances 9 (65) (2019), 38017–38031.

[60] S. Di Lecce et al. Correction: Modelling and prediction of the thermophysical properties of aqueous mixtures of choline geranate and geranic acid (CAGE) using SAFT-γ Mie. RSC Advances 10 (33) (2020), 19463–19465.

[61] S. A. Febra et al. Extending the SAFT-γ Mie approach to model benzoic acid, diphenylamine, and mefenamic acid: Solubility prediction and experimental measurement. Fluid Phase Equilibria 540 (2021), 113002.

[62] M. Wehbe et al. Thermodynamic modelling of the nature of speciation and phase behaviour of binary and ternary mixtures of formaldehyde, water and methanol. Molecular Physics 121 (19-20) (2023), e2197712.

[63] S. A. Febra. Ring formation in a statistical associating fluid theory framework. PhD thesis. Imperial College London, 2018.

[64] M. Sadeqzadeh et al. The development of unlike induced association-site models to study the phase behaviour of aqueous mixtures comprising acetone, alkanes and alkyl carboxylic acids with the SAFT-γ Mie group contribution methodology. Fluid Phase Equilibria 407 (2016), 39–57.

[65] C. Silva et al. Phase equilibria of triolein to biodiesel reactor systems. Fluid Phase Equilibria 409 (2016), 171–192.

[66] V. Papaioannou et al. Application of the SAFT-γ Mie group contribution equation of state to fluids of relevance to the oil and gas industry. Fluid Phase Equilibria 416 (2016), 104–119.

[67] E. F. Sabattié et al. Predicting oligomer/polymer compatibility and the impact on nanoscale segregation in thin films. Soft Matter 13 (19) (2017), 3580–3591.

[68] S. Samsatli, I. Staffell, and N. J. Samsatli. Optimal design and operation of integrated wind-hydrogen-electricity networks for decarbonising the domestic transport sector in Great Britain. International Journal of Hydrogen Energy 41 (1) (2016), 447–475.

[69] M. Bui et al. Dynamic operation and modelling of amine-based CO2 capture at pilot scale. International Journal of Greenhouse Gas Control 79 (2018), 134–153.

[70] P. Morgado et al. Modeling the fluid-phase equilibria of semifluorinated alkanes and mixtures of (n-Alkanes+ n-Perfluoroalkanes) with the SAFT-γ Mie group-contribution approach. Journal of Chemical & Engineering Data 65 (12) (2020), 5909–5919.

[71] F. A. Perdomo et al. Description of the thermodynamic properties and fluid-phase behavior of aqueous solutions of linear, branched, and cyclic amines. AIChE Journal 67 (3) (2021), e17194.

[72] T. Bernet et al. Modeling the thermodynamic properties of cyclic alcohols with the SAFT-γ Mie approach: application to cyclohexanol and menthol systems. Molecular Physics 122 (21-22) (2024), e2394132.

[73] M. Valsecchi, A. Galindo, and G. Jackson. Modelling the thermodynamic properties of the mixture of water and polyethylene glycol (PEG) with the SAFT-γ Mie group-contribution approach. Fluid Phase Equilibria 577 (2024), 113952.

[74] S. Gopinath et al. Outer approximation algorithm with physical domain reduction for computer-aided molecular and separation process design. AIChE Journal 62 (9) (2016), 3484–3504.

[75] S. Gopinath et al. "A feasibility-based algorithm for Computer Aided Molecular and Process Design of solvent-based separation systems". *Computer Aided Chemical Engineering*. Vol. 38. Elsevier, 2016, 73–78.

[76] P. Kazepidis et al. Optimal design of post combustion CO_2 capture processes based on phase-change solvents. Vol. 46. Elsevier, 2019, 463–468.

[77] A. I. Papadopoulos et al. An approach for simultaneous computer-aided molecular design with holistic sustainability assessment: Application to phase-change CO_2 capture solvents. Computers & Chemical Engineering 135 (2020), 106769.

[78] D. H. Bowskill et al. Beyond a heuristic analysis: integration of process and working-fluid design for organic Rankine cycles. Molecular Systems Design & Engineering 5 (2) (2020), 493–510.

[79] M. T. White et al. Industrial waste-heat recovery through integrated computer-aided working-fluid and ORC system optimisation using SAFT-γ Mie. Energy Conversion and Management 150 (2017), 851–869.

[80] M. White et al. Integrated computer-aided working-fluid design and thermoeconomic ORC system optimisation. Energy Procedia 129 (2017), 152–159.

[81] M. White et al. Computer-aided working-fluid design, thermodynamic optimisation and thermoeconomic assessment of ORC systems for waste-heat recovery. Energy 161 (2018), 1181–1198.

[82] L. M. van Kleef, O. A. Oyewunmi, and C. N. Markides. Multi-objective thermo-economic optimization of organic Rankine cycle (ORC) power systems in waste-heat recovery applications using computer-aided molecular design techniques. Applied Energy 251 (2019), 112513.

[83] O. L. Watson et al. Computer-aided Design of Solvent Blends for the Cooling and Anti-solvent Crystallisation of Ibuprofen. Vol. 46. Elsevier, 2019, 949–954.

[84] A. Alyazidi et al. Predicting the Solubility of Amino Acids and Peptides with the SAFT-γ Mie Approach: Neutral and Charged Models. Industrial & Engineering Chemistry Research 63 (46) (2024), 20397–20423.

[85] S. H. Huang and M. Radosz. Equation of state for small, large, polydisperse, and associating molecules. Industrial & Engineering Chemistry Research 29 (11) (1990), 2284–2294.

[86] G. Mie. Zur kinetischen Theorie der einatomigen Körper. Annalen der Physik 316 (8) (1903), 657–697.

[87] G. Lazarou. Development of the SAFT-γ Mie equation of state for predicting the thermodynamic behaviour of strong and weak electrolyte solutions. PhD Thesis. Imperial College London, 2017.

[88] D. K. Eriksen et al. Development of intermolecular potential models for electrolyte solutions using an electrolyte SAFT-VR Mie equation of state. Molecular Physics 114 (18) (2016), 2724–2749.

[89] J. M. A. Schreckenberg et al. Modelling of the thermodynamic and solvation properties of electrolyte solutions with the statistical associating fluid theory for potentials of variable range. Molecular Physics 112 (17) (2014), 2339–2364.

[90] L. L. Lee. Molecular Thermodynamics of Nonideal Fluids. Butterworth-Heinemann, 2016.

[91] J. A. Barker and D. Henderson. What is “liquid"? Understanding the states of matter. Reviews of Modern Physics 48 (4) (1976), 587.

[92] J. A. Barker and D. Henderson. Perturbation theory and equation of state for fluids. II. A successful theory of liquids. The Journal of Chemical Physics 47 (11) (1967), 4714–4721.

[93] T. Lafitte et al. Accurate statistical associating fluid theory for chain molecules formed from Mie segments. The Journal of Chemical Physics 139 (15) (2013), 154504.

[94] T. Boublık. Hard-sphere equation of state. The Journal of chemical physics 53 (1) (1970), 471–472.

[95] G. Mansoori et al. Equilibrium thermodynamic properties of the mixture of hard spheres. The Journal of Chemical Physics 54 (4) (1971), 1523–1525.

[96] M. S. Wertheim. Thermodynamic perturbation theory of polymerization. The Journal of Chemical Physics 87 (12) (1987), 7323–7331.

[97] T. Boublık. Background correlation functions in the hard sphere systems. Molecular Physics 59 (4) (1986), 775–793.

[98] L. Blum. Mean spherical model for asymmetric electrolytes: I. Method of solution. Molecular Physics 30 (5) (1975), 1529–1535.

[99] L. Blum and J. S. Hoye. Mean spherical model for asymmetric electrolytes. 2. Thermodynamic properties and the pair correlation function. The Journal of Physical Chemistry 81 (13) (1977), 1311–1316.

[100] M. Born. Volumen und hydratationswärme der ionen. Zeitschrift für physik 1 (1) (1920), 45–48.

[101] M. A. Selam, I. G. Economou, and M. Castier. A thermodynamic model for strong aqueous electrolytes based on the eSAFT-VR Mie equation of state. Fluid Phase Equilibria 464 (2018), 47–63.

[102] B. Maribo-Mogensen, K. Thomsen, and G. M. Kontogeorgis. An electrolyte CPA equation of state for mixed solvent electrolytes. AIChE Journal 61 (9) (2015), 2933–2950.

[103] G. Soave. Equibrium constants from a modified Redlich-Kwong equation of state. Chemical Engineering Science 27 (1972), 1198–1203.

[104] C. Held et al. Measuring and modeling aqueous electrolyte/amino-acid solutions with ePC-SAFT. The Journal of Chemical Thermodynamics 68 (2014), 1–12.

[105] J. M. A. Schreckenberg. Modelling of the Thermodynamic and Solvation Properties of Electrolyte Solutions with the Statistical Associating Fluid Theory for Potentials of Variable Range. PhD thesis. Imperial College London, UK, 2011.

[106] A. J. Haslam, A. Galindo, and G. Jackson. Prediction of binary intermolecular potential parameters for use in modelling fluid mixtures. Fluid Phase Equilibria 266 (1-2) (2008), 105–128.

[107] F. London. The general theory of molecular forces. Transactions of the Faraday Society 33 (1937), 8b–26.

[108] P. Hutacharoen. Prediction of partition coefficients and solubilities of active pharmaceutical ingredients with the SAFT-γ Mie group-contribution approach. PhD Thesis. Imperial College London, 2017.

[109] N. R. Nur Jazlan. Modelling the free energy of solvation: from data-driven to statistical mechanical approaches. PhD Thesis. Imperial College London, 2020.

[110] F. A. Perdomo et al. A predictive group-contribution framework for the thermodynamic modelling of CO_2 absorption in cyclic amines, alkyl polyamines, alkanolamines and phase-change amines: New data and SAFT-γ Mie parameters. Fluid Phase Equilibria 566 (2023), 113635.

[111] J. B. Dalton and C. L. A. Schmidt. The solubilities of certain amino acids in water, the densities of their solutions at twenty-five degrees, and the calculated heats of solution and partial molal volumes. Journal of Biological Chemistry 103 (2) (1933), 549–578.

[112] Y. Yi et al. Development of a small-scale automated solubility measurement apparatus. Industrial & Engineering Chemistry Research 44 (15) (2005), 5427–5433.

[113] H. Kuramochi et al. Measurements of vapor pressures of aqueous amino acid solutions and determination of activity coefficients of amino acids. Journal of Chemical & Engineering Data 42 (3) (1997), 470–474.

[114] C. M. Romero and J. C. Cadena. Effect of temperature on the volumetric properties of α, ω-amino acids in dilute aqueous solutions. Journal of Solution Chemistry 39 (2010), 1474–1483.

[115] M. Guo et al. The effect of chain length and side chains on the solubility of peptides in water from 278.15 K to 313.15 K: A case study in glycine homopeptides and dipeptides. Journal of Molecular Liquids 352 (2022), 118681.

[116] R. Neerup et al. Solid–Liquid Phase Boundaries in the System: Glycine–NaOH–NaHCO3–H2O. Journal of Chemical & Engineering Data 67 (6) (2022), 1550–1564.

[117] D. A. Amilova, B. S. Zakirov, and B. M. Beglov. POLYTHERM FOR THE SOLUBILITY OF THE AMMONIUM SULFATE-METHIONINE-WATER SYSTEM. Zhurnal Obshchei Khimii 56 (1) (1986), 25–27.

[118] S. D. Pradhan. "The chain length and isomeric effect of alcohol on the excess properties of amine-alcohol systems: Excess free energy of mixing, enthalpy of mixing and volume change on mixing". *Proceedings of the Indian Academy of Sciences–Chemical Sciences*. Vol. 90. Springer. 1981, 261–273.

[119] M. Domınguez et al. Isobaric Vapor–Liquid Equilibrium for the Binary Mixtures (2-Butanol+ n-Hexane) and (2-Butanol+ 1-Butylamine) and for the Ternary System (2-Butanol+ n-Hexane+ 1-Butylamine) at 101.3 kPa. Journal of Chemical & Engineering Data 47 (3) (2002), 405–410.

[120] R. Thacker and J. S. Rowlinson. The physical properties of some polar solutions. Part 1.—Volumes and heats of mixing. Transactions of the Faraday Society 50 (1954), 1036–1042.

[121] T. Kimura et al. Excess enthalpies of binary mixtures of butylamines + propanols at 298.15 K. Journal of Thermal Analysis and Calorimetry 113 (2013), 1467–1474.

[122] H. Iloukhani and M. Soleimani. Measurement and modeling the excess molar volumes and refractive index deviations of binary mixtures of 2-Propanol, 2-butanol and 2-pentanol with N-propylamine. Journal of Solution Chemistry 46 (2017), 2135–2158.

[123] M. Dominguez et al. Speed of sound and isentropic compressibility of the ternary mixture (2-butanol + n-hexane + 1-butylamine) and the constituent binary mixtures at 298.15 K and 313.15 K. Journal of Molecular Liquids 88 (2-3) (2000), 243–258.

[124] W. L. Weng and J. T. Chen. Density and viscosity measurement of n-butylamine with hexyl alcohol isomer binary systems. Journal of Chemical & Engineering Data 49 (6) (2004), 1748–1751.

[125] S. Amer Amezaga. Vapor–liquid equilibrium at 760 mm of binary systems formed by methyl, ethyl, propyl, isopropyl, butyl, isobutyl, sec-butyl, and tert-butyl alcohols with propionic acid. Anales de Quimica 71 (1975), 117–126.

[126] S. A. Iwarere et al. Vapour–liquid equilibrium of carboxylic acid–alcohol binary systems: 2-Propanol + butyric acid, 2-butanol + butyric acid and 2-methyl-1-propanol + butyric acid. Fluid Phase Equilibria 380 (2014), 18–27.

[127] M. Behroozi and H. Zarei. Volumetric properties of highly nonideal binary mixtures containing ethanoic acid and propanoic acid with butan-2-ol, methyl-2-propanol, and 2-methyl-2-butanol at different temperatures. Journal of Chemical & Engineering Data 57 (4) (2012), 1089–1094.

[128] A. G. Osborn and D. R. Douslin. Vapor pressure relations of 13 nitrogen compounds related to petroleum. Journal of Chemical & Engineering Data 13 (4) (1968), 534–537.

[129] N. C.-B. Ahmed et al. Experimental determination of the isothermal (vapour + liquid) equilibria of binary aqueous solutions of sec-butylamine and cyclohexylamine at several temperatures. The Journal of Chemical Thermodynamics 44 (1) (2012), 116–120.

[130] A. G. Osborn and D. W. Scott. Vapor pressures of 17 miscellaneous organic compounds. The Journal of Chemical Thermodynamics 12 (5) (1980), 429–438.

[131] A. Simon and J. Huter. Zur Kenntnis der Dampfdruckkurven, Schmelzpunkte und der chemischen Konstanten von Dimethyl-, Trimethyl-und Isobutyl-Amin. Zeitschrift für Elektrochemie und angewandte physikalische Chemie 41 (1) (1935), 28–33.

[132] M. Shirai. Dielectric Polarization of Some Aliphatic Amines in the Liquid State. Bulletin of the Chemical Society of Japan 29 (1956), 518–521.

[133] R. B. Tôrres and H. E. Hoga. Volumetric properties of binary mixtures of dichloromethane and amines at several temperatures and p = 0.1 MPa. Journal of Molecular Liquids 143 (1) (2008), 17–22.

[134] M. A. Saleh, S. Akhtar, and A. R. Khan. Excess molar volumes of aqueous solutions of butylamine isomers. Physics and Chemistry of Liquids 38 (1) (2000), 137–149.

[135] T. W. De Loos, H. R. Tijsseling, and J. De Swaan Arons. Vapor–liquid equilibria of the system ethane + 2-aminopropane. Journal of Chemical and Engineering Data 32 (3) (1987), 374–377.

[136] H. Wolff and A. Shadiakhy. The vapour-pressure behaviour and the association of isomeric propylamines and n-deuteriopropylamines in mixtures with n-hexane. Fluid Phase Equilibria 11 (3) (1983), 267–287.

[137] S. D. Pradhan and H. B. Mathur. Thermodynamic study of binary mixtures of isomeric butylamines with n-hexane: Enthalpy of hydrogen bonding. Proceedings of the Indian Academy of Sciences-Section A, Chemical Sciences 87 (1978), 23–29.

[138] E. Matteoli, L. Lepori, and A. Spanedda. Thermodynamic study of heptane+ amine mixtures: I. Excess and solvation enthalpies at 298.15 K. Fluid Phase Equilibria 212 (1-2) (2003), 41–52.

[139] E. Brunner. Löslichkeit von Wasserstoff in Aminen. Berichte der Bunsengesellschaft für Physikalische Chemie 82 (8) (1978), 798–805.

[140] K. Payne. Modelling the Carbonyl Group in 2-Ketones using the SAFT-γ Mie Methodology. MSc Thesis. Imperial College London, 2017.

[141] J. McMurry. Organic Chemistry. 2nd. (See section 25.4). Pacific Grove, California, USA: Brooks/Cole Publishing Company, 1988.

[142] J. Schmelzer and J. Pusch. Phase equilibria in binary systems containing N-monosubstituted amides and hydrocarbons. Fluid Phase Equilibria 110 (1-2) (1995), 183–196.

[143] K. V. Zaitseva, M. A. Varfolomeev, and S. P. Verevkin. Vapour pressures and enthalpies of vaporisation of N-alkyl acetamides. Journal of Molecular Liquids 293 (2019), 111453.

[144] J. Hahn and S. Stoeck. Determination of the vapor pressure of N-butylpropionamide and N-propylacetamide. Leuna-Protokoll (1983), 7271.

[145] S. I. Gertler. Screening tests of some N-substituted acetamides as insecticides and acarides. ARS (Series)(United States. Agricultural Research Service); 33-14 (1955).

[146] H. Eshghi and P. Shafieyoon. P_2O_5/SiO_2 as a mild and efficient reagent for acylation of alcohols, phenols and amines under solvent-free conditions. Journal of Chemical Research 2004 (12) (2004), 802–805.

[147] G. Mueller and K. Moerke. Determination of the vapor–liquid equilibrium in the system trichloroethene – N-pentylacetamide. Leuna-Protokoll (1988), 4131.

[148] R. Gopal and S. A. Rizvi. Vapour pressures of some mono-and di-alkyl substituted aliphatic amides at different temperatures. Journal of the Indian Chemical Society 45 (1) (1968), 13.

[149] J. Hahn and K. Moerke. Vapor pressures of some caprolactam impurities. Leuna-Protokoll (1984), 4121.

[150] J. Gmehling et al. Pure compound data from DDB. Dortmund Data Bank 2014 (1983).

[151] B. Mukesh et al. Thermodynamic, DFT and molecular dynamics studies of intermolecular interactions between 2-methoxyaniline and N-substituted amide mixtures. Chemical Data Collections 22 (2019), 100241.

[152] B. Jović, A. Nikolić, and B. Kordić. Densitometric and spectroscopic investigation of interactions of selected N-substituted amides and acetonitrile. Journal of Molecular Liquids 191 (2014), 10–15.

[153] R. Gopal and S. A. Rizvi. Physical properties of some mono-and dialkyl-substituted amides at different temperatures. Journal of the Indian Chemical Society 43 (1966), 179–182.

[154] F. J. Millero. Relative viscosity and apparent molal volume of N-methylpropionamide solutions at various temperatures. The Journal of Physical Chemistry 72 (9) (1968), 3209–3214.

[155] T. B. Hoover. Conductance of potassium chloride in highly purified N-methylpropionamide from 20 to 40°. The Journal of Physical Chemistry 68 (4) (1964), 876–879.

[156] T. B. Hoover. The N-Methylpropionamide–water system. Densities and dielectric constants at 20–40°. The Journal of Physical Chemistry 73 (1) (1969), 57–61.

[157] J. E. Van Evercooren, G. V. Merken, and H. P. Thun. The Conductivity of Hydrochloric Acid in N-Methylpropionamide at Temperatures from 15 to 50 °C. Bulletin des Sociétés Chimiques Belges 84 (6) (1975), 533–539.

[158] L. R. Dawson, R. H. Graves, and P. G. Sears. Solvents Having High Dielectric Constants. III. Solutions of Sodium and Potassium Halides in N-Methylpropionamide and in N-Methylbutyramide from 30 to 60°. Journal of the American Chemical Society 79 (2) (1957), 298–300.

[159] A. de Haan et al. Vapor–liquid equilibria and enthalpies of mixing for binary mixtures of N-methylacetamide with aniline, decane, ethylene glycol, naphthalene, phenol, and water. Journal of Chemical & Engineering Data 42 (5) (1997), 875–881.

[160] A. B. de Haan et al. Vapor–Liquid Equilibria and Excess Enthalpies for Octane + N-Methylacetamide, Cyclooctane + N-Methylacetamide, and Octane + Acetic Anhydride at 125 °C. Journal of Chemical & Engineering Data 40 (6) (1995), 1228–1232.

[161] G. Kortüm and H. Biedersee. Dampf/Flüssigkeit-Gleichgewichte (Siedediagramme) binärer Systeme hoher relativer Flüchtigkeit. Wasser/N-Methylacetamid, Wasser/N-Methylformamid und N-Methylformamid/N-Methylacetamid. Chemie Ingenieur Technik 42 (8) (1970), 552–560.

[162] V. Štejfa et al. Thermodynamic study of acetamides. Journal of Molecular Liquids 319 (2020), 114019.

[163] A. K. Nain. Densities and volumetric properties of (acetonitrile + an amide) binary mixtures at temperatures between 293.15 K and 318.15 K. The Journal of Chemical Thermodynamics 38 (11) (2006), 1362–1370.

[164] P. Pacak. Refractivity and density of some organic solvents. Chemical Papers 45 (227232) (1991), 29.

[165] K. Simmrock, R. Janowsky, and A. Ohnsorge. Critical Data of Pure Substances, Chemical Data Ser. Vol. II, Part I, DECHEMA (1986), 614.

[166] P. Assarsson and F. R. Eirich. Properties of amides in aqueous solution. I. Viscosity and density changes of amide–water systems. An analysis of volume deficiencies of mixtures based on molecular size differences (mixing of hard spheres). The Journal of Physical Chemistry 72 (8) (1968), 2710–2719.

[167] J. F. Casteel and E. S. Amis. Conductance of sodium perchlorate in water–N-methylacetamide (NMA) solvent system. Journal of Chemical & Engineering Data 19 (2) (1974), 121–128.

[168] A. B. de Haan and J. Gmehling. Excess enthalpies for various binary mixtures with N-methylacetamide or acetic anhydride. Journal of Chemical & Engineering Data 41 (3) (1996), 474–478.

[169] A. M. Zaichikov. Enthalpies of mixing of water with secondary amides of carboxylic acids. Russian Journal of General Chemistry 67 (9) (1997), 1355–1360.

[170] J. Zielkiewicz. Vapour + liquid equilibrium measurements and correlation of the ternary mixture (N-methylacetamide + ethanol + water) at the temperature 313.15 K. The Journal of Chemical Thermodynamics 32 (1) (2000), 55–62.

[171] J. Manczinger and G. Kortüm. Thermodynamische Mischungseffekte im System Wasser (1)/N-Methylacetamid (2). (1975).

[172] G. I. Egorov and D. M. Makarov. Densities and Volumetric Properties of Aqueous Solutions of {water (1) + N-methylurea (2)} Mixtures at Temperatures of 274.15–333.15 K and at Pressures up to 100 MPa. Journal of Chemical & Engineering Data 62 (12) (2017), 4383–4394.

[173] J. M. Saleh and L. H. Al-Azzawi. Determination of the transfer energies of hydrobromic acid in N-methylurea + water mixtures at different temperatures. Iraqi Journal of Science 21 (1980), 507–525.

[174] J. Krakowiak, J. Wawer, and A. Panuszko. Densimetric and ultrasonic characterization of urea and its derivatives in water. The Journal of Chemical Thermodynamics 58 (2013), 211–220.

[175] S. Lapanje et al. Effect of temperature on the apparent molal volume of ethylurea in aqueous solutions. Journal of Chemical and Engineering Data 30 (1) (1985), 29–32.

[176] W. Chang et al. Isobaric vapor–liquid equilibria for water + acetic acid +(N-methyl pyrrolidone or N-methyl acetamide). Fluid Phase Equilibria 242 (2) (2006), 204–209.

[177] M. Singh. Determination of Densities of Amino Compounds for Molar Volumes in Aqueous Solutions with Magnetic Float Densimeter at Various Temperatures. Biological Sciences-PJSIR 49 (3) (2006), 160–169.

[178] Y. Y. Fialkov and A. Suprunenko. STRENGTH OF CARBONIC-ACIDS IN N-METHYLACETAMIDE. UKRAINSKII KHIMICHESKII ZHURNAL 41 (11) (1975), 1214–1215.

[179] W.-T. Vong and F.-N. Tsai. Densities, molar volumes, thermal expansion coefficients, and isothermal compressibilities of organic acids from 293.15 K to 323.15 K and at pressures up to 25 MPa. Journal of Chemical & Engineering Data 42 (6) (1997), 1116–1120.

[180] H. L. Ritter and J. H. Simons. The molecular state of acetic acid vapor. Journal of the American Chemical Society 67 (5) (1945), 757–762.

[181] C. W. Hoerr and A. W. Balston. THE SOLUBILITIES OF THE NORMAL SATURATED FATTY ACIDS. II. The Journal of Organic Chemistry 09 (4) (1944), 329–337.

[182] E. A. Cepeda, R. Bravo, and B. Calvo. Solubilities of Lauric Acid in n-Hexane, Acetone, Propanol, 2-Propanol, 1-Bromopropane, and Trichloroethylene from (279.0 to 315.3) K. Journal of Chemical & Engineering Data 54 (4) (2009), 1371–1374.

[183] A. B. Gonçalves Bonassoli et al. Solubility Measurement of Lauric, Palmitic, and Stearic Acids in Ethanol, n-Propanol, and 2-Propanol Using Differential Scanning Calorimetry. Journal of Chemical & Engineering Data 64 (5) (2019), 2084–2092.

[184] B. Calvo, I. Collado, and E. A. Cepeda. Solubilities of Palmitic Acid in Pure Solvents and Its Mixtures. Journal of Chemical & Engineering Data 54 (1) (2009), 64–68.

[185] U. Domańska. Solid-liquid phase relations of some normal long-chain fatty acids in selected organic one- and two-component solvents. Industrial & Engineering Chemistry Research 26 (6) (1987), 1153–1162.

[186] B. Calvo and E. A. Cepeda. Solubilities of Stearic Acid in Organic Solvents and in Azeotropic Solvent Mixtures. Journal of Chemical & Engineering Data 53 (3) (2008), 628–633.

[187] V. Yablokov, I. Smel'tsova, and V. Faerman. Thermal stability of amino acids. Russian Journal of General Chemistry 83 (2013), 476–480.

[188] H. Sun, L. Wang, and B. Liu. Solubility of α-glycine in water with additives at a temperature range of (293.15–343.15) K: Experimental data and results of thermodynamic modeling. Fluid Phase Equilibria 434 (2017), 167–175.

[189] Y. Z. Chua et al. New experimental melting properties as access for predicting amino-acid solubility. RSC Advances 8 (12) (2018), 6365–6372.

[190] H. T. Do et al. Melting properties of peptides and their solubility in water. Part 1: dipeptides based on glycine or alanine. RSC Advances 9 (56) (2019), 32722–32734.

[191] H. T. Do et al. Melting properties of amino acids and their solubility in water. RSC Advances 10 (72) (2020), 44205–44215.

[192] H. T. Do et al. Melting properties of peptides and their solubility in water. Part 2: Di-and tripeptides based on glycine, alanine, leucine, proline, and serine. Industrial & Engineering Chemistry Research 60 (12) (2021), 4693–4704.

[193] G. Wolf, H.-G. Schmidt, and K. Bohmhammel. Differential scanning calorimetry. A reliable method of enthalpy calibration. Thermochimica acta 235 (1) (1994), 23–29.

[194] J. Huang and L. Yu. Effect of molecular chirality on racemate stability: α-amino acids with nonpolar R groups. Journal of the American Chemical Society 128 (6) (2006), 1873–1878.

[195] M. An et al. Measurement and Correlation for Solubility of L-Alanine in Pure and Binary Solvents at Temperatures from 283.15 to 323.15 K. Journal of Chemical & Engineering Data 65 (2) (2020), 549–560.

[196] M. H. Abraham and P. L. Grellier. Substitution at saturated carbon. Part XIX. The effect of alcohols and water on the free energy of solutes and on the free energy of transition states in SN and SE reactions. Journal of the Chemical Society, Perkin Transactions 2 (15) (1975), 1856–1863.

[197] Y. Fan et al. The Research and Measurement about the Solubility of L-Serine in Eight Common Pure Solvents and Four Binary Mixed Solvents for T=(278.15–333.15) K. Journal of Chemical & Engineering Data 64 (10) (2019), 4398–4411.

[198] C. Zhang et al. Measurement and Correlation of Solubility of L-Valine in Water + (Ethanol, N, N-Dimethylformamide, Acetone, Isopropyl Alcohol) from 293.15 K to 343.15 K. Journal of Chemical & Engineering Data 59 (9) (2014), 2732–2740.

[199] N. A. Bowden, J. P. M. Sanders, and M. E. Bruins. Solubility of the proteinogenic α-amino acids in water, ethanol, and ethanol–water mixtures. Journal of Chemical & Engineering Data 63 (3) (2018), 488–497.

[200] Z. Cao et al. Solubility of glycine in binary system of ethanol+ water solvent mixtures: Experimental data and thermodynamic modeling. Fluid Phase Equilibria 360 (2013), 156–160.

[201] B. P. Dey and S. C. Lahiri. Solubilities of Amino Acids in Ethanol + Water Mixture at Different Temperatures. Journal of the Indian Chemical Society 69 (1992), 552–557.

[202] B.-W. Long, L.-S. Wang, and J.-S. Wu. Solubilities of 1, 3-benzenedicarboxylic acid in water + acetic acid solutions. Journal of Chemical & Engineering Data 50 (1) (2005), 136–137.

[203] G. Pucher and W. M. Dehn. Solubilities in mixtures of two solvents. Journal of the American Chemical Society 43 (8) (1921), 1753–1758.

[204] A. Bouchard, G. W. Hofland, and G.-J. Witkamp. Solubility of glycine polymorphs and recrystallization of β-glycine. Journal of Chemical & Engineering Data 52 (5) (2007), 1626–1629.

[205] Y. Nozaki and C. Tanford. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions: establishment of a hydrophobicity scale. Journal of Biological Chemistry 246 (7) (1971), 2211–2217.

[206] M. S. Dunn and F. J. Ross. Quantitative Investigations of Amino Acids and Peptides: IV. The Solubilities of the Amino Acids in Water-Ethyl Alcohol Mixtures. Journal of Biological Chemistry 125 (1) (1938), 309–332.

[207] Z. H. Ansari and Z. Li. Solid–Liquid Equilibria for the Glycine–Alcohol–NaCl–H2O System. Journal of Chemical & Engineering Data 62 (10) (2017), 3551–3560.

[208] M. T. Gude et al. Partition coefficients and solubilities of α-amino acids in aqueous 1-butanol solutions. Industrial & engineering chemistry research 35 (12) (1996), 4700–4712.

[209] C. J. Orella. The crystallization of amino acids from mixtures of water and alcohol. PhD thesis. University of Virginia, 1990.

[210] L. A. Ferreira, S. P. Pinho, and E. A. Macedo. Solubility of L-serine, L-threonine and L-isoleucine in aqueous aliphatic alcohol solutions. Fluid phase equilibria 270 (1-2) (2008), 1–9.

[211] H. Talukdar, S. Rudra, and K. K. Kundu. Thermodynamics of transfer of glycine, diglycine, and triglycine from water to aqueous solutions of urea, glycerol, and sodium nitrate. Canadian Journal of Chemistry 66 (3) (1988), 461–468.

[212] W. J. Waddell and R. G. Bates. Intracellular pH. Physiological Reviews 49 (2) (1969), 285–329.

[213] D. T. Manallack. The acid–base profile of a contemporary set of drugs: implications for drug discovery. SAR and QSAR in Environmental Research 20 (7-8) (2009), 611–655.

[214] D. T. Manallack et al. The significance of acid/base properties in drug discovery. Chemical Society Reviews 42 (2) (2013), 485–496.

[215] T. E. Needham Jr, A. N. Paruta, and R. J. Gerraughty. Solubility of amino acids in mixed solvent systems. Journal of Pharmaceutical Sciences 60 (2) (1971), 258–260.

[216] H.-C. Tseng et al. Solubilities of amino acids in water at various pH values under 298.15 K. Fluid Phase Equilibria 285 (1–2) (2009), 90–95.

[217] A. A. Rashin and B. Honig. Reevaluation of the Born model of ion hydration. The journal of physical chemistry 89 (26) (1985), 5588–5593.

[218] S. I. Sandler. Chemical, Biochemical, and Engineering Thermodynamics. 4^{th}. Wiley, 2005.

[219] D. Schick et al. Predicting CO_2 solubility in aqueous and organic electrolyte solutions with ePC-SAFT advanced. Fluid Phase Equilibria 567 (2023), 113714.

[220] H. S. Harned and B. B. Owen. The Physical Chemistry of Electrolyte Solutions. Reinhold, 1958, 634–643.

[221] A. K. Covington, M. I. A. Ferra, and R. A. Robinson. Ionic product and enthalpy of ionization of water from electromotive force measurements. Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases 73 (1977), 1721–1730.

[222] G. E. K. Branch and S. Miyamoto. Dissociation constants and heats of ionization of some simple amino acids and peptides. Journal of the American Chemical Society 52 (3) (1930), 863–868.

[223] M. Li et al. 30 years of lithium-ion batteries. Advanced materials 30 (33) (2018), 1800561.

[224] J.-Y. Hwang, S.-T. Myung, and Y.-K. Sun. Sodium-ion batteries: present and future. Chemical Society Reviews 46 (12) (2017), 3529–3614.

[225] R. Rajagopalan et al. Advancements and challenges in potassium ion batteries: a comprehensive review. Advanced Functional Materials 30 (12) (2020), 1909486.

[226] B. Widom. Some topics in the theory of fluids. The Journal of Chemical Physics 39 (11) (1963), 2808–2812.

[227] C. H. Bennett. Efficient estimation of free energy differences from Monte Carlo data. Journal of Computational Physics 22 (2) (1976), 245–268.

[228] Z. Mester and A. Z. Panagiotopoulos. Temperature-dependent solubilities and mean ionic activity coefficients of alkali halides in water from molecular dynamics simulations. The Journal of chemical physics 143 (4) (2015).

[229] Z. Mester and A. Z. Panagiotopoulos. Mean ionic activity coefficients in aqueous NaCl solutions from molecular dynamics simulations. The Journal of chemical physics 142 (4) (2015).

[230] A. Barducci, M. Bonomi, and M. Parrinello. Metadynamics. Wiley Interdisciplinary Reviews: Computational Molecular Science 1 (5) (2011), 826–843.

[231] J. Sun et al. High-pressure polymeric phases of carbon dioxide. Proceedings of the National Academy of Sciences 106 (15) (2009), 6077–6081.

[232] S. Angioletti-Uberti et al. Solid-liquid interface free energy through metadynamics simulations. Physical Review B—Condensed Matter and Materials Physics 81 (12) (2010), 125416.

[233] A. Lyubartsev et al. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. The Journal of chemical physics 96 (3) (1992), 1776–1783.

[234] M. F. Hurley et al. Expanded ensemble predictions of absolute binding free energies in the SAMPL9 host–guest challenge. Physical Chemistry Chemical Physics 25 (47) (2023), 32393–32406.

[235] J. D. Chodera and M. R. Shirts. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. The Journal of chemical physics 135 (19) (2011).

[236] S. Zhang et al. Expanded ensemble methods can be used to accurately predict protein-ligand relative binding free energies. Journal of chemical theory and computation 17 (10) (2021), 6536–6547.

[237] A. S. Paluch, J. K. Shah, and E. J. Maginn. Efficient solvation free energy calculations of amino acid analogs by expanded ensemble molecular simulation. Journal of Chemical Theory and Computation 7 (5) (2011), 1394–1403.

[238] A. S. Paluch et al. A method for computing the solubility limit of solids: Application to sodium chloride in water and alcohols. The Journal of chemical physics 133 (12) (2010).

[239] J. R. Errington. Direct calculation of liquid–vapor phase equilibria from transition matrix Monte Carlo simulation. The Journal of chemical physics 118 (22) (2003), 9915–9925.

[240] M. Fitzgerald, R. Picard, and R. Silver. Canonical transition probabilities for adaptive Metropolis simulation. Europhysics Letters 46 (3) (1999), 282.

[241] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. Chemical physics letters 314 (1-2) (1999), 141–151.

[242] J.-S. Wang and R. H. Swendsen. Transition matrix monte carlo method. Journal of statistical physics 106 (2002), 245–285.

[243] S. Kournopoulos et al. The contribution of the ion–ion and ion–solvent interactions in a molecular thermodynamic treatment of electrolyte solutions. The Journal of Physical Chemistry B 126 (47) (2022), 9821–9839.

[244] F. Wang and D. P. Landau. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. Physical Review E 64 (5) (2001), 056101.

[245] M. Andreev et al. Coarse-grained model of the dynamics of electrolyte solutions. The Journal of Physical Chemistry B 121 (34) (2017), 8195–8202.

[246] A. P. Thompson et al. LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. Computer physics communications 271 (2022), 108171.