An Investigation into the Crystallisation Behaviour of Glycine Homopeptides

By

Mingxia Guo

A dissertation submitted to Imperial College London for the degree of

Doctor of Philosophy

Department of Chemical Engineering
Imperial College London
South Kensington
London SW7 2AZ

April 2022

Copyright Declaration

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

Declaration of Originality

I hereby declare that this thesis and the work presented herein is entirely my own work. Information presented and derived from the published and unpublished work of others has been acknowledged in the text and references as given.

Acknowledgments

The past four years in the Chemical Engineering Department of Imperial College London have provided me with invaluable research and life experiences. I am indebted to Prof. Jerry Heng for his insightful comments and ideas on my research, for encouraging and supporting me throughout my Ph.D. journey, and for his unshakable belief which made me have the chance to improve not only my research skills but also my soft skills. I learnt a lot from him and I can hardly imagine a better supervisor! I would like to express my gratitude to my Ph.D. mentor, Prof. Amparo Galindo, for her important advise and unwavering support throughout my research and life, most notably for her great guidance during the modelling work with SAFT. I want to offer my sincere thanks to the cooperators in my Ph.D. career: Prof. Qiuxiang Yin and Dr. Lingzhaou at Tianjin University in China, for their cooperation and support in my crystal structure research, which enabled me to continue working during the lockdown period caused by the covid-19 pandemic. I would like to thank Dr. Chin W. Yong and Dr. Ilian T. Todorov for their assistance with the molecular dynamic simulation. Dr. Emily Jeannette Guinn of Eli Lilly for her he informative discussion about peptide crystallisation.

I want to thank all the brilliant people I met in the Heng Research Group, for their friendliness and willingness to assist, and for their helpful discussion in every group meeting. I would like to express my thank to Dr. Huaiyu Yang and Dr. Wenqian Chen for their guidance and insightful advice on my research throughout my Ph.D. career. Dr. Vikram Karde and Dr. Vivek Verma for their instructive discussion on particle physics and crystallography. I would like to thank Dr. Ian Rosbottom for his encouragement and technical support on my research. I want to thank Ethan Errington for his infinite modeling wisdom, Enshu Liang and Hamish Mitchell for the communication and discussion regarding peptide crystallisation. I also want to thank Frederik Link for the helpful discussion about protein crystallisation, and Tsaone

Gosiamemang, Kebonyetsala Seanneng, Isha Bade, Angeliki Chalasti and Shunning Xiang for all the helpful discussion and support. Additionally, I'd want to express my gratitude to all of the visiting students and researchers from various universities for their fruitful sharing and collaboration. I'd also like to express my gratitude to Dr. Felipe Antonio Perdomo Hurtado of Prof. Galindo's group for always being so friendly and patient to teach me the modeling work about SAFT.

I also want to thank all the staff and technicians in the Chemical Engineering Department for their effort and support on my research. Especially Patricia Carry and Kaho Cheung for the training on the characterisation equipment, Dr. Umang Shah for the guidance when I did the graduate teaching assistant.

I want to give my special thanks to the most important people in my life-my parents-for their unwaving love and support all the time, making me can explore the world without fear and worries. I also want to thank my brother and sisters for the happiness and amazing moments they brought to my life, especially my younger sister Xiaoxia, who is a currently a Ph.D. candidate in the Chemistry department at UCL. We had a wonderful experience studying together in London. I also want to thank my childhood friend Yan Zhang who I knew from primary school. Many thanks for her companionship throughout my first year in the UK and for the enduring lifelong beautiful friendship.

Mingxia Guo

London, United Kingdom March 2022

Abstract

The combinations of amino acids into peptides and proteins, through peptide bonds, are the building blocks of life on earth. Their natural therapeutic properties has seen a significant increase in the application of these materials in the treatment of chronic diseases. The purest and most stable crystalline form provides structural information at the atomic level and is desirable for formulation into efficacious pharmaceutical products. Peptide crystallisation, as a good alternative to chromatographic purification, can also solve the shortcomings of traditional purification method, such as high cost, proteolytic degradation and physiochemical instability. However, peptide crystallisation still remains a major challenge due to highly flexible conformations especially in the case where water plays an integral role in the crystal structure. Glycine is the simplest amino acid and is known to play an important role in new biomimetic functional materials and biopharmaceutical research. Its hydrogen side chain makes the molecule an ideal candidate to study the effeect of chain length on the peptide solubility and crystallisation, without the effect of side chain.

The glycine homopeptides crystallisation research in this thesis includes three parts: thermodynamic properties, kinetic properties, and the relationship between peptides conformation and crystallisation. Firstly, the solubility of glycine homopeptides (glycine, diglycine, triglycine, tetraglycine, pentaglycine, and hexaglycine), amino acids with different side chains (aspartic acid, phenylalanine, histidine, and tyrosine) and their dipeptides (asp-phe, gly-asp, gly-phe, phe-phe, gly-gly, tyr-phe, gly-tyr, gly-his) in water from 278.15K to 313.15K were measured using the UV-Vis spectroscopy method and dynamic method. The modified Apelblat equation is used to correlate the relationship between solubility in water and temperature. Molecular dynamic (MD) simulation was further employed to investigate the solute-solvent interactions behind the dissolution behaviors. Moreover, the group-group

interaction matrix of the SAFT-γ Mie approach was extended for the prediction of the solubility of amino acids and peptides, exploring the application of SAFT-γ Mie to biomolecular thermodynamic properties. Secondly, the classical nucleation theory was applied to the short-chain glycine homopeptide crystallisation to explore the nucleation theory of macromolecules. The nucleation parameters (nucleation rate, growth rate, interfacial surface energy, and activation Gibbs energy) were calculated based on the classical nucleation theory to explore the chain length effect on the classical nucleation mechanism of peptides, providing kinetic data to the crystallisation conditions designed for industry and modeling tools, such as gPROMS. The evidence of the non-classical nucleation phenomenon was also observed and discussed. Finally, the interaction between water and peptide molecules which can stabilize the unfolded structure of peptides and proteins was revealed, the effect of temperature and salts on the transition between unfolded and folded structure was explored, giving an inspiration to the relationship between conformation and peptide crystallisation.

The research presented in this thesis investigates the thermodynamic and kinetic properties of glycine homopeptides, as well as the flexible conformation of peptides during crystallisation, thereby providing a comprehensive strategy for designing and optimising the crystallisation process. Additionally, the research establishes a fundamental understanding of peptide crystallisation, which is extremely beneficial for future macromolecular crystallisation research.

Publications

The papers published during the PhD career are listed as follows:

[1] **Mingxia Guo,** Ian Rosbottom, Lina Zhou, Chin W. Yong, Ling Zhou, Qiuxiang Yin, Ilian T. Todorov, Ethan Errington and Jerry Y.Y. Heng*. Triglycine (GGG) adopts a polyproline II (PPII) conformation in its hydrated crystal form: revealing the role of water in peptide crystallisation [J]. *The journal of physical chemistry Letters.* 2021, 12, 34, 8416–8422.

[2] **Mingxia Guo,** Wenqian Chen, Zhenhong Chang, Ling Zhou, Qiuxiang Yin and Jerry Y.Y. Heng*. The effect of chain length and side chains on the solubility of peptides in water from 278.15K to 313.15K: A case study in glycine homopeptides and dipeptides [J]. *The journal of molecule liquid.* 2022, 352, 118681.

[3] Vivek Verma, Hamish Mitchell, **Mingxia Guo**, Benjamin K. Hodnett, Jerry Y. Y. Heng. Studying the Impact of Pre-exponential Factor on Templated Nucleation [J]. *Faraday Discussions.* 2021.

[4] Wenqian Chen, Xiaoyu Li, **Mingxia Guo**, Frederik J. Link, Siti S. Ramli, Jinbo Ouyang, Ian Rosbottom and Jerry Y.Y.Heng*. Biopurification of monoclonal antibody (mAb) through crystallisation [J]. *Separation and Purification Technology.*2021, 263, 118358.

[5] Jinbo Ouyang*, Jian Chen, Wenqian Chen, Ian Rosbottom, **Mingxia Guo** and Jerry Y. Y. Heng*. Application of phenyl-functionalized porous silica for the selective crystallisation of carbamazepine metastable form II [J]. *Ind. Eng. Chem. Res.* 2021, 60, 2, 939–946

[6] Jinbo Ouyang*, Jian Chen, Wenqian Chen, Ian Rosbottom, **Mingxia Guo** and Jerry Y. Y. Heng*. Supersaturation and solvent dependent nucleation of carbamazepine polymorphs during rapid cooling crystallisation [J]. *CrystEngComm.*2021, 60, 2, 939–946

[7] Vikram Kardea, **Mingxia Guo** and Jerry Y.Y. Heng*. Influence of interparticle structuring on the surface energetics of a binary powder system [J]. *International Journal of Pharmaceutics*. 2020, 581, 119295.

In preparation:

[1] **Mingxia Guo,** Marie J F Jones, and Jerry Y. Y. Heng*. Nucleation of glycine homopeptides in water: the effect of the peptide bonds on the nucleation [J]. *Crystal Growth & Design* 2022,

In preparation.

[2] **Mingxia Guo**, Felipe A. Perdomo, Wenqian Chen, Amparo Galindo* and Jerry Y.Y. Heng*. SAFT-γ Mie prediction of peptides solubility based on amino acid residues. [J]. *Journal of molecular liquids.* 2022, In preparation.

[3] **Mingxia Guo** and Jerry Y.Y. Heng* The hydration and dehydrate process of triglycine hydrate to explore the stability of PPII conformation in peptide hydrates. [J].Angewandte Chemie International Edition 2022, In preparation.

[4] **Mingxia Guo** and Jerry Y.Y. Heng* Online spectrum measurement of the conformational change of glycine homopeptides with temperature to explore the balance of beta sheet and pPII conformation [J]. *The journal of physical chemistry Letters.* 2022 In preparation.

[5] **Mingxia Guo** and Jerry Y.Y. Heng* Review: The role of peptide conformation in the Peptide crystallisation [J]. *RSC advance* 2022, In preparation.

Conferences

2022 Early Career International Particle Technology Forum 2022, UK 202205

The relationship between unfolded conformation and crystallization conditions of glycine homopeptides. Poster presentation

2022 ChemEngDay UK 2022, UK 2022.04

The thermodynamic properties and nucleation of peptides: A case study of glycine homopeptides. Poster presentation

2020 Virtual AIChE Annual Meeting, USA 2020.11

The Role of Water in the Crystallisation and Nucleation of Glycine Homopeptides. Oral Presentation

2020 Virtual AIChE Annual Meeting, USA 2020.11

Solubility Measurement and SAFT-γ Mie Prediction of Peptides Based on Amino Acid Residues. Oral Presentation

50th Annual Conference of the British Association of Crystal Growth (BACG), UK 2019.07

Crystallisation of peptides: Case of glycine homopeptides. Poster presentation

3rd International Symposium on Continuous Manufacturing of Pharmaceuticals (ISCMP), UK 2018.10

Continuous crystallisation oscillatory flow platform for protein purification and bio-Separation. Poster presentation

The 4th Asian Crystallisation Technology Symposium (ACTS-2018), Singapore 2018.06

Peptide solubility and crystallisation. Oral Presentation

Table of Contents

Copyright Declaration 2
Declaration of Originality 3
Acknowledgments 4
Abstract 6
Publications 8
Conferences 10
Table of Contents 11
List of Tables 15
List of Figures 18
Chapter 1. Introduction 24
Chapter 2. Literature Review 30
2.1 Definition of peptides 30
2.1.1 The properties of amino acids 30
2.1.2 Bonds between amino acids 32
2.1.3 The secondary structure of folded and unfolded peptides 33
2.2 The thermodynamic properties of peptides 35
2.2.1 Solubility curves for small molecules and macromolecules 35
2.2.2 Solubility studies of peptides 37
2.3 Solubility prediction 39
2.3.1 Prediction of solubility based on classical thermodynamic equation of state 39
2.3.2 Prediction of solubility based on Statistical Associating Fluid Theory (SAFT) 40
2.3.3 Prediction of solubility based on amino acid composition 44
2.4 Nucleation during crystallisation 44
2.4.1 The classical nucleation theory 45
2.4.2 The non-classical nucleation theory 45
2.5 Peptide crystallisation 49
2.5.1 Crystallisation methods for small molecules 50
2.5.2 Crystallisation methods for macromolecules 50
2.5.3 Current crystallisation methods for peptides 53
Chapter 3. Methodology 57
3.1 Crystal structure analysis 57
3.1.1 Single crystal structure analysis 57
3.1.2 Powder X-ray diffraction (PXRD) 58
3.2 Thermal analysis 59

3.2.1 Differential scanning calorimetry (DSC) 59
3.2.2 Thermal gravimetric analysis (TGA) 60
3.3 Spectrum analysis 60
3.3.1 UV-Visible spectroscopy 60
3.3.2 Raman spectroscopy 62
3.3.3 Fourier transform infrared (FTIR) spectroscopy 64
3.4 Molecular dynamic (MD) modeling 64
Chapter 4. The effect of chain length and side chains on the solubility of peptides in water from 278.15K to 313.15K: A case study in glycine homopeptides and dipeptides 65
4.1 Abstract 65
4.2 Introduction 66
4.3 Materials and Methods 67
4.3.1 Materials 67
4.3.2 Powder X-ray diffraction analysis 70
4.3.3 Measurement of standard curves 70
4.3.4 Measurement of solubility 70
4.3.5 Thermodynamic equations 72
4.4 Results and Discussion 73
4.4.1 Material characterisation 73
4.4.2 The standard curves of the amino acids and peptides 76
4.4.3 Results and discussion of peptides hydrate 76
4.4.4 Effect of chain length on the solubility of glycine homopeptides 80
4.4.5 Effect of side chains on the solubility of dipeptides 82
4.4.6 Molecular dynamic simulations 85
4.5 Conclusion 89
Chapter 5 Solubility prediction of amino acids and peptides using SAFT- γ Mie 90
5.1 Abstract 90
5.2 Introduction 91
5.3 Theory and molecular models 94
5.3.1 The Group-Contribution Approach. 94
5.3.2 The interaction and parameters between two segments m and n 95
5.3.3 SAFT-γ Mie Equation of State 97
5.3.4 Solid-Liquid Phase Equilibrium 100
5.3.5 Parameter estimation $Fobj$ and the absolute deviation (%AAD) 100
5.4 Results and discussion 101
5.4.1 Structures and group parameters of studied amino acids and peptides 101
5.4.2 Parameter fitting process using pure component properties of glycine 107

5.4.3 Parameter fitting process incorporating properties of glycine-water mixture 109
5.4.4 Melting temperature and fusion enthalpy of amino acids and peptides 111
5.4.5 Analysis of the impact of fusion enthalpy 113
5.4.6 Only use bubble pressure and solubility of glycine-water to fit parameters 115
5.4.7 Only use bubble pressure of glycine-water, valine-water, alanine-water to fit parameters 120
5.4.8 Using the solubility and bubble pressure of glycine-water, valine-water, alanine-water to fit parameters 124
5.4.9 Other parameters in future work 126
5.5 Conclusion 130
Chapter 6. Nucleation of glycine homopeptides in water: the effect of the peptide bonds on the nucleation 131
6.1 Abstract 131
6.2 Introduction 132
6.3 Theory 133
6.3.1 Induction time t_d 133
6.3.2 Determination of nucleation rate J 134
6.3.3 Nucleation data derived from Classical Nucleation Theory 134
6.4 Experiments 136
6.4.1 Materials 136
6.4.2 Induction time measurement 137
6.5 Results and discussion 140
6.5.1 The characterisation of glycine homopeptides 140
6.5.2 Induction time measurement for glycine homopeptides (from mono-glycine to penta-glycine) 140
6.5.3 Nucleation parameters of glycine, diglycine and triglycine dihydrate 143
6.5.4 The evidence of non-classical nucleation phenomenon during the glycine homopeptides crystallisation 155
6.6 Conclusion 156
Chapter 7. The role of water in peptide crystallisation and the relationship between conformation and crystallisation conditions 158
7.1 Abstract 158
7.2 Introduction 159
7.3 Results and Discussion 161
7.3.1 The cooling crystallisation of mono-, di-and tri-glycine 161
7.3.2 The thermal and spectroscopic characterisation of triglycine anhydrate and dihydrate 162
7.3.3 The single crystal growth and structure determination process 166
7.3.4 Crystal packing of the glycine homopeptides 171
7.3.5 Role of water in stabilising the PPII conformation 176

7.3.6 The effect of temperature and salts on the conformation change in solution..... 185
7.4 Conclusion 188
Chapter 8 Conclusions and Future work..... 189
8.1 Overview..... 189
8.2 Key contributions..... 191
8.3 Future work..... 191
8.3.1 Solubility measurement and prediction..... 191
8.3.2 Solubility prediction using gSAFT 192
8.3.3 Nucleation of peptides 192
8.3.4 The transition between folded and unfolded conformation in peptide crystallisation 193
References..... 194
Appendix..... 208

List of Tables

Table 2.1 The dihedral angles for the main secondary structures 35

Table 2.2 The comparison between different SAFT equations* 44

Table 4.1 Properties of materials used in this chapter.....68

Table 4.2 Specific wavelengths and equations of the standard curves of amino acids and peptides 77

Table 4.4 Comparison of the solubility of triglycine anhydrate and dihydrate using the dynamic method and UV-Vis method 78

Table 4.5 Comparison of the solubility of Gly-Tyr anhydrate and dihydrate using the dynamic method and UV-Vis method. 79

Table 4.6 Comparison of the solubility of Gly-Tyr anhydrate and dihydrate using the dynamic method and UV-Vis method. 79

Table 4.8 Parameters and ARDs of the modified Apelblats equation about the solubility of glycine homopeptides in water (P = 0.1 MPa).....81

Table 4.10 Parameters and ARDs of the modified Apelblats equation about the solubility of amino acids and dipeptides in water (P = 0.1) MPa). 83

Table 4.11 Solvation free energy of glycine homopeptides in water at 298.15K.....88

Table 4.12 Solvation free energy of amino acids and dipeptides in water at 298.15K 88

Table 5.1 The like and unlike group-group interaction in glycine homopeptides.....104

Table 5.2 Like group-group parameters for studied amino acids in this work.....107

Table 5.3 Unlike group-group parameters for studied amino acids in this work 107

Table 5.4 Group-group parameters for NH_2-COOH using pure component properties of glycine.....108

Table 5.5 Group-group parameters of NH_2-COOH using pure properties of glycine and mixture properties of glycine-water 110

Table 5.6 The fusion enthalpy of glycine listed in literature.....112

Table 5.7 Fusion enthalpy of studied amino acids using different group contribution methods.....114

Table 5.8 Group-group parameters of NH_2-COOH using mixture properties of glycine-water 116

Table 5.9 The fusion enthalpy calculated by Gani method and melting points (in literatures) of glycine, valine

and alanine 121

Table 5.10 The group-group parameters of NH_2-COOH using bubble pressure of glycine-water, valine-water and alanine-water 121

Table 5.11 The adjusted fusion enthalpy and melting points (in literatures) of glycine, valine and alanine 123

Table 5.12 The adjusted fusion enthalpy and melting points (DSC) of glycine, valine and alanine.....124

Table 5.13 The new adjusted fusion enthalpy of glycine, valine and alanine 124

Table 5.14 The adjusted fusion enthalpy of glycine, valine and alanine using bubble pressure and solubility of glycine, valine and alanine as the input experimental data.....125

Table 5.15 Group-group parameters of NH_2-COOH using bubble pressure and solubilities of glycine-water, valine-water and alanine-water 125

Table 5.16 The AAD values for the predicted bubble pressure and solubility of glycine, valine and alanine 126

Table 5.17 Unlike group parameters for C=O-NH_2 and C=O-H_2O 127

Table 5.18 Unlike group parameters for C=O-COOH 128

Table 5.19 Unlike group parameters for C=O-NH.....129

Table 6.1 Properties of materials used in this work. 137

Table 6.2 The experimental conditions of the nucleation measurement of glycine, diglycine and triglycine 139

Table 6.3 The calculated nucleation kinetics for glycine homopeptides at 278.15 K and 283.15 K.....146

Table 6.4 Values obtained from linear fit and calculated nucleation parameters for glycine homopeptides at 283.15K 149

Table 6.5 Values obtained from linear fit and calculated nucleation kinetics for glycine homopeptides at 278.15K 150

Table 7.1 The crystallographic data of TGDH.....169

Table 7.2 Bond Lengths for TGDH 170

Table 7.3 Bond Angles for TGDH 170

Table 7.4 Torsion Angles for TGDH.....171

Table 7.5 The torsion of backbone and packing coefficient of glycine homopeptides and TGDH*.....174

Table 7.6 The comparison of dihedral angle of Gly(2) in triglycine dihydrate and other main secondary structures.....174
Table 7.7 Average count of hydrogen bond interactions observed for water-carboxylate, water-ammonium and ammonium-carboxylate functionalities in the simulated aqueous environment. 182
Table 7.8 Average count of hydrogen bond interactions observed for ammonium-carboxylate functionalities in the simulated aqueous environment.....182
Table 7.9 Average count of hydrogen bond interactions observed for water-carboxylate, functionalities in the simulated aqueous environment.....183
Table 7.10 Average count of hydrogen bond interactions observed for water-ammonium functionalities in the simulated aqueous environment.....183
Table 4.3 The calibration points and the standard deviation of studied amino acids and peptides.....208
Table 4.7 Experimental and calculated solubility of glycine homopeptides in water at different temperatures (0.1 MPa). [a] 217
Table 4.9 Experimental and calculated solubility of amino acids and dipeptides in water at different temperatures (0.1 MPa). [a] 219

List of Figures

Figure 1.1The research objectives of peptide crystallisation 28

Figure 2.1 The forming process of a dipeptide 30

Figure 2.2 Classification of 20 standard amino acids[17].....31

Figure 2.3 The three disulphide bonds in Insulin[29] 33

Figure 2.4 The four levels of peptides structure [32].....34

Figure 2.5 The solubility-supersaturation diagram of small molecules [45] 36

Figure 2.6 The solubility-supersolubility diagram of proteins (consisting of four zones representing different degrees of supersaturation）.[48] 37

Figure 2.7 Flow chart of the gravimetric and NMR methods used for determining solubility [66] 39

Figure 2.8 Continuous distribution of bond strengths, showing the span from simple van der Waals attractions to the formation of chemical bonds.[71] 42

Figure 2.9 The perturbation scheme and interactions for the formation of a molecule within the SAFT.....42

Figure 2.10 Sketch map of different nucleation theories/pathways and their free energy change during nucleation.[1] 47

Figure 2.11 Coexistence curves of LLPS in solutions of an incretin peptide (IP5) at pH 5.5.[97].....48

Figure 2.12 The liquid-liquid phase separation region and crystallisation region for macromolecules.[97]....48

Figure 2.13 Cryo-TEM images of the dynamic evolution process of Ag+ coordinated Fmoc-Ala(a) and Fmoc-His (b) self-assembling nanofibrils over time.[98].....49

Figure 2.14 The schematic diagrams of hanging-drop vapor-diffusion method (left) and sitting-drop vapor-diffusion method (right) [103] 51

Figure 2.15 Pipetting process for microbatch experiment [103,104].....52

Figure 2.16 Pipetting process for microdialysis crystallisation.[105] 52

Figure 2.17 Free-interface diffusion method [105] 53

Figure 2.18 Crystallisation of BmBKTx1 obtained by the racemic method [108].....54

Figure 2.19 (a) Unit cell of BmBKTx1, (b) The interface between L- and D-enantiomers [108].....54

Figure 2.20 [129]Xe-cryptophane biosensor (190) complexed with CAII [111] 55

Figure 2.21 Direct purification of L-Leu-L-Leu dipeptide through co-crystallisation[113] 55

Figure 3.1 A PXRD (Bruker D8 Advance) showing the goniometer and its drives inside the cabinet.[124] 58

Figure 3.2 General diagram for a differential scanning calorimeter.[127] 59

Figure 3.3 TGA equipment with different loading position. Left: top loading, middle: bottom loading, right: side loading.[130] 60

Figure 3.4 The Nanodrop UV-Vis equipment and the testing procedure 62

Figure 3.5. Change of the amide I, II, and III bands in Raman for different conformation of glycine homopeptides [135] 63

Figure 4.1 Chemical structures of the studied amino acids and peptides 69

Figure 4.2 The process of solubility measurement. 72

Figure 4.3 XRD patterns of amino acids and peptides before and after solubility measurement: A.Glycine (α form), B.Diglycine (α form), C.Triglycine (β form), D.Tetraglycine, E.Pentaglycine, F.Hexaglycine, G.Aspartic acid, H.Histidine, I. Phenylalanine, J.Tyrosine, K.Asp-Phe, L.Phe-Phe, M.Gly-Tyr, N.Gly-Asp, O.Tyr-Phe, P.Gly-Phe, Q.Gly-His 75

Figure 4.4 The TGA (solid line) and DSC (dash line) curves of Gly-Tyr (left) and Gly-Asp (right) 76

Figure 4.6 The solubilities of peptide anhydrate and hydrate. (black-anhydrate, red-hydrate, blue-UV-Vis measurement).....80

Figure 4.7 The solubility of glycine homopeptides in water from 278.15K to 313.15K 81

Figure 4.8 The chain length effect on the solubility of glycine homopeptides in water from 278.15K to 293.15K 81

Figure 4.9 Solubilities of amino acids and dipeptides in water from 278.15K to 313.15K. Solid lines represent the calculated solubilities. 83

Figure 4.10 The comparison of solubility between different amino acids and peptides 84

Figure 4.11 The amorphous cell of glycine homopeptides and water molecules for the solvation free energy calculation. 87

Figure 5.1 The different residue sequences of peptides with same segments 93

Figure 5.2 Example of the decomposition of glycine molecule into functional groups: one NH_2 group (shaded green), one CH_2 group (shaded gray), and one COOH group (shaded blue) 94

Figure 5.3 The four separate contributions of Helmholtz free energy in SAFT-γ Mie and the calculation of

chemical potential and pressure based on Helmholtz free energy 98
Figure 5.4 The different segments division strategies of peptides with and without side chains. 103
Figure 5.5 The like and unlike group-group parameters and interactions for glycine pure component and glycine-water mixture. The red boxes represent the unkwown parameters. 105
Figure 5.6 The amino acids molecules decomposed into functional groups and the like and unlike group-group parameters and interactions for the studied amino acids in water. 106
Figure 5.7 Comparison between experimental data and predicted data of glycine using pure component properties. The points mean the experimental data and the lines present the predicted data. 109
Figure 5.8 Comparison between experimental data and predicted data of glycine using pure and mixture component properties. 110
Figure 5.9 Left: the DSC curve (solid line) and TGA curve (dash line) for glycine. Right: the fusion enthalpy calculated by integrating the peak area in DSC curve. 111
Figure 5.10 DSC curves of glycine using different heating rate (10K/min, 50K/min, 100 K/min). 112
Figure 5.11 Heat flow rate curve of glycine using fast scanning calorimetry [55] 113
Figure 5.12 The comparison of predicted solubility for glycine with different fusion enthalpy 113
Figure 5.13 The comparison of experimental and predicted thermodynamic data of glycine and glycine-water. 117
Figure 5.14 Predicted solubility data of glycine, alanine, valine, Leucine, Lysine, Aspartic acid and Glutamic acid with fusion enthalpy calculated by Gani method (Left) and adjusted by solubility data (Right). 120
Figure 5.15 The comparison between experimental bubble pressure and predicted bubble pressure for glycine-water, alanine-water and valine-water 120
Figure 5.16 The experimental (points) and predicted(lines) bubble pressure and solubility of glycine, valine and alanine using fusion enthalpy estimated by Gani group contribution method 122
Figure 5.17 The experimental (points) and predicted (lines) solubility of glycine, valine and alanine using adjusted fusion enthalpy and melting temperature in literatures 123
Figure 5.18 The experimental (points) and predicted (lines) solubility of glycine, valine and alanine using adjusted fusion enthalpy and melting temperature by DSC 124
Figure 5.19 The experimental (points) and predicted (lines) solubility of glycine, valine and alanine using the new adjusted fusion enthalpy and melting temperature by DSC 124

Figure 5.20 The experimental (points) and predicted (lines) bubble pressure and solubilities of glycine, valine and alanine using bubble pressure and solubility as the input experimental data. 126

Figure 5.21 the parameters matrix and the structurally similar chemicals for glycine homopeptides parameters. 127

Figure 5.22 The experimental (points) and predicted (lines) bubble pressure and solubilities of acetamide in water 128

Figure 5.23 The experimental (points) and predicted (lines) vapour pressure and saturation liquid density of pyruvic acid in water 128

Figure 5.24 The experimental (points) and predicted (lines) vapour pressure and saturation liquid density of N-Methylacetamide in water 129

Figure 6.1 Chemical structures of Glycine homopeptides 137

Figure 6.2 Experimental set up for induction time measurement in the lab 139

Figure 6.3 Schematic representation of the nucleation experiment set up.....140

Figure 6.4 The XRD patterns of glycine homopeptides before and after induction time measurement.....141

Figure 6.5 The microscope images (a. glycine, b.diglycine, c.triglycine dihydrate, d.tetraglycine, e.pentaglycine, f.hexaglycine.) and SEM images (g. pentaglycine and h. hexaglycine) of glycine homopeptides.....142

Figure 6.6 Induction time of glycine homopeptides under different supersaturation levels at 278.15K. ...143

Figure 6.7 Induction time of glycine under different supersaturation levels at 278.15K and 283.15K. Solid lines represent the fit of the Poisson distribution.....144

Figure 6.8 Induction time of diglycine under different supersaturation levels at 278.15K and 283.15K. Solid lines represent the fit of the Poisson distribution.....144

Figure 6.9 Induction time of triglycine dihydrate under different supersaturation levels at 278.15K and 283.15K. Solid lines represent the fit of the Poisson distribution.....145

Figure 6.10 Comparison of Induction time of glycine homopeptides under different temperatures. Solid lines represent the fit of the Poisson distribution. 145

Figure 6.11 Correlation of nucleation rate with the supersaturation of glycine homopeptides under different temperatures.....147

Figure 6.12 The comparison of the relationships between nucleation rate and supersaturation level of glycine

homopeptides under different temperatures.....148

Figure 6.13 The relationship between nucleation rate J, supersaturation levels and the number of glycine residues. The darker colour are the peptides at 278.15K, whereas the lighter colour are peptides at 283.15K. 151

Figure 6.14 The relationship between interfacial energy γ, supersaturation levels and the number of glycine residues. The darker colour are the peptides at 278.15K, whereas the lighter colour are peptides at 283.15K. 153

Figure 6.15 The relationship between activation energy ΔGC , supersaturation levels and the number of glycine residues. The darker colour are the peptides at 278.15K, whereas the lighter colour are peptides at 283.15K. 154

Figure 6.16 The light and polarized microscope images of non-classical nucleation phenomenon during cooling crystallisation of hexaglycine.....155

Figure 6.17 Gelation phenomenon during the cooling crystallisation of triglycine dihydrate taken by light and polarized microscope 156

Figure 7.1 The chemical structures of Glycine homopeptides.....161

Figure 7.2 a The Optical microscope images and SEM images of triglycine and TGDH; b The XRD pattern of triglycine and TGDH.....162

Figure 7.3 The DSC and TGA images of triglycine (left) and TGDH (right).....163

Figure 7.4 The FT-IR (Top) and Raman (Bottom) spectrograms of triglycine anhydrate and TGDH.....164

Figure 7.5 The amide modes in FT-IR (Top) and Raman (Bottom) spectrograms of triglycine anhydrate and TGDH 166

Figure 7.6 a The asymmetric unit in TGDH; b-d The interaction between two water molecules and triglycine molecule; e a chain of hydrogen bonds formed by waters; f The bended triglycine backbone in TGDH ..168

Figure 7.7 The asymmetric unit in the crystal structure of TGDH. 171

Figure 7.8 The hydrogen bonding motifs of glycine, diglycine and triglycine anhydrate and TGDH in the single crystal structure. 173

Figure 7.9 The structure of glycine homopeptides with atom numbering 173

Figure 7.10 The interaction between water and triglycine molecules in the single crystal structure.....175

Figure 7.11 Left are the molecular structures and labelled H-bonding groups for alpha glycine, alpha

diglycine, beta triglycine and TGDH. Right are the H-bonding propensities tables, which show the coordination of H-bonds to these groups from other crystal structures in the CSD, with the number bolded corresponding to the coordination found in these glycine homopeptides crystal structure. 177

Figure 7.12 Ramachandran plot of the phi and psi angles for the central alpha carbon of triglycine molecules in aqueous solutions over 4 ns (top). 179

Figure 7.13 Radial distribution functions for the hydrogen the ammonium triglycine terminal (NH_3^+) and oxygen atoms of the carboxylate triglycine terminal (COO^-). Data is segmented into 1 ns sample sizes (e.g. 10 ns = 9.0 to 10.0 ns).....180

Figure 7.14 Radial distribution functions for the oxygen atoms of the carboxylate triglycine terminal (COO^-) and the hydrogen atom of water. Data is segmented into 1 ns sample sizes (e.g. 10 ns = 9.0 to 10.0 ns). .181

Figure 7.15 Radial distribution functions for the hydrogen atoms of the ammonium triglycine terminal (NH_3^+) and oxygen atoms of water. Data is segmented into 1 ns sample sizes (e.g. 10 ns = 9.0 to 10.0 ns).....181

Figure 7.16 The packing around a central diglycine molecule in the diglycine hydrate structure (above); the packing around a central diglycine molecule in the anhydrate form (below).....185

Figure 7.17 The online FTIR spectrum for diglycine (left) and triglycine (right) under different temperatures 186

Figure 7.18 The crystal structure of LiBr·Glyglygly[206] 187

Figure 7.19 The microscope images of final products when adding different salts under different temperature 187

Figure 4.5 Standard curves of the studied amino acids and peptides: A.Glycine (α form), B.Diglycine (α form), C.Triglycine(β form), D.Tetraglycine, E.Pentaglycine, F.Hexaglycine, G.Aspartic acid, H.Histidine, I. Phenylalanine, J.Tyrosine, K.Asp-Phe, L.Phe-Phe, M. Gly-Tyr, N.Gly-Asp, O.Tyr-Phe, P.Gly-Phe, Q.Gly-His.....217

Chapter 1. Introduction

Crystallisation is an environmentally friendly unit operation mainly used to separate or purify a wide variety of materials. It ranks as one of the oldest unit operation in the chemical engineering fields, having been improved significantly over the past more than 100 years by the development of diffraction.[2] In 1895, German scientist W. C. Roentgen discovered X-Ray, an electromagnetic radiation, which stands out as one of the notable landmarks in the progress of crystallisation.[3] This finding also earned him the first Nobel Prize in physics. Numerous scientists started to research the secret of the X-ray. Among them, M. von Laue discovered the diffraction of X-rays by crystals in 1912.[4] This discovery proves not only the wave property of X- rays, but also the existence of crystal lattice structure. Furthermore, electron diffraction was found in 1927 by C. J. Davisson and G. P. Thomson.[5] After that, the first electron microscope was established by E. Ruska in 1931.[6] These scientists were also awarded as the Nobel Laureates in physics in 1901, 1914, 1937 and 1986, respectively. These significant findings helped scientists to explore the inner structure of the objectives, and also pushed the development of crystallisation.

From medication design to materials research, there is an increasing need to determine the crystal structure of molecules in order to better understand their properties and to produce structures with unique attributes. Additionally, crystal size and morphology control is a critical stage in the development of industrial processes.[2] The natural therapeutic properties of peptides and proteins[7-9] has seen a significant increase in the application of these materials in the treatment of chronic and metabolic diseases (e.g. cancer, obesity and diabetes). [10-15] The most pure and stable crystalline form contains structural information at the atomic level, which is advantageous for formulation into effective medicines. The first macromolecule crystal was obtained by Lehman in 1853, which is the haemoglobin from blood. And then scientists started

to crystallise proteins in a very large range from egg white to plants. During the past decades, peptide crystallisation has also been reported, but compared with protein, peptides are still very difficult to crystallise. The main challenge about peptide crystallisation is just as same as protein crystallisation, which is identifying suitable solution environment to crystallise the products with ideal size, physical and chemical properties. Marraud summarized the reasons for the difficulty of crystallising linear peptides. Firstly, it is the flexible conformation of the peptides, which prevents the crystal nucleation. Secondly, the quantity of materials required for crystallisation is more than that used in biological experiments. Thirdly, the residual solvents from the former process, such as synthesis and lyophilisation, will decrease the purity of peptides and make crystals difficult to come out.

Glycine is the most basic amino acid and has been shown to be critical in the development of new biomimetic functional materials and biopharmaceutical research. Due to the absence of side chains, this amino acid is an excellent choice for studying the effect of chain length on peptide solubility and crystallisation. In this work, we aim to make glycine homopeptides as a model system to explain the crystallisation behaviour of peptides and analyze the relationship between peptides conformation and crystallisation conditions to optimize the peptide crystallisation process to get more regular, stable products.

The thermodynamic properties of peptides are significant in determining the crystallisation conditions.[16] To build a reasonable design of the conditions for peptide crystallisation, we strive to understand the influence of chain length on the solubility of glycine homopeptides and the effect of side chains on the solubility of eight dipeptides. The amino acid residues of dipeptides examined in this study were selected based on the distinct properties of their side chains. The solubility of glycine homopeptides (glycine, diglycine, triglycine, tetraglycine, pentaglycine and hexaglycine), amino acids with different side chains (aspartic acid, phenylalanine, histidine and tyrosine) and their dipeptides (asp-phe, gly-asp, gly-phe, phe-phe, gly-gly, tyr-phe, gly-tyr,

gly-his) in water from 278.15K to 313.15K were measured using the UV-Vis spectroscopy method and dynamic method. The modified Apelblat equation was used to correlate the relationship between solubility in water and temperature. Molecular dynamic (MD) simulation was further employed to investigate the solute-solvent interactions underlying the dissolution behaviours.

Based on these experimental data, gSAFT was used to predict the solubility of more peptides. The SAFT- γ Mie group contribution equation of state in which a heteronuclear model is implemented, with molecules fragmented into their chemical moieties (CH_3, CH_2, NH_2, COOH, etc…) was used to calculate the thermodynamic properties and phase equilibrium (including solid-liquid solubility) of fluids predictively. Peptides and amino acids of interest were modelled by dividing them into relevant groups. Known group-group interaction parameters were adopted from the literature or developed in this work using literature data as well as solubility data measured in this study. The NH_2-COOH parameters were determined based on the thermodynamic properties of glycine, alanine and valine. These parameters can be used to predict the properties of other amino acids (alanine, valine, aspartic acid, glutamic acid, isoleucine, leucine and lysine) and peptides.The group-group interaction matrix of the SAFT- γ Mie approach was extended for the prediction of the solubility of amino acids and peptides.

The nucleation mechanism for the crystallisation of glycine homopeptides has been investigated based on the thermodynamic properties measured in the previous work. The induction time of glycine, diglycine and triglycine in water were measured at different temperature and supersaturation levels. The nucleation parameters (nucleation rate, gowth rate, interfacial surface energy, activation Gibbs energy) were calculated based on the classical nucleation theory to explore the effect of chain length on the classical nucleation mechanism of peptides, providing kinetic data to the crystallisation conditions design for industry and

modeling tools, such as gPROMS.

The flexible structures of peptides in solution make the crystallisation process become more challenging. Polyproline II (pPII) is a left-handed 3_1-helix conformation, which has been observed to be the most abundant secondary structure in unfolded peptides and proteins. The pPII conformation has been shown to be preferable in proline-, alanine-, and glycine-rich regions of peptides and proteins as a substantial contributor to unfolded peptide structures. Water has been shown to stabilise the pPII conformation in unfolded portions of proteins, which are critical for a protein's function and dynamics. However, past research has failed to provide clear and consistent conclusions regarding how water interacts with peptides or with itself in order to stabilise the pPII conformation. Significant advances in spectroscopic and modelling approaches have increased the discovery of these peptide conformers in solutions, but the absence of high-resolution X-ray crystal structures has hampered efforts to comprehend peptide and protein conformations in the solid state. Crystal forms for up to five glycine residues in a peptide chain are reported, whereby all the anhydrous and hydrate forms show beta-sheet or fully extended conformations. In this thesis, the single crystal structure of the triglycine dihydrate was resolved for the first time, which is also the first resolved glycine homopeptide hydrate adopting pPII conformation in the solid state. The solution conformation and water structuring around triglycine was investigated using the Raman, FTIR and single crystal structure analysis to find the interaction between water and peptide backbones during the crystallisation process. This research demonstrates that even though the central glycine residue can contribute to the pPII conformation, the carboxylate-water and water-water interactions play an important role in stablizing the pPII conformation. Molecular dynamics simulation was used to obtain the Ramachandran plot to gain an in-depth understanding of the conformation distribution in solution to explore the solution to crystal pathway for peptides.

In summary, the glycine homopeptides crystallisation research includes three sections: thermodynamic properties, kinetic properties and the relationship between peptides conformation and crystallisation. The primary accomplishments include the following: To begin, we established a thermodynamic foundation for further development of the peptide crystallisation process and investigated the application of SAFT-γ Mie to biomolecular thermodynamic properties, providing a flatform for investigating the relationship between their physicochemical properties and biostructures. Second, we investigated the nucleation theory of macromolecules by applying the classical nucleation theory to the crystallisation of a short chain glycine homopeptide. Thirdly, discovered an interaction between water and peptide backbones that has the ability to stabilise the unfolded structure of peptides and proteins, providing insight into the relationship between conformation and peptide crystallisation. Figure 1.1 illustrates the link between the research objectives.

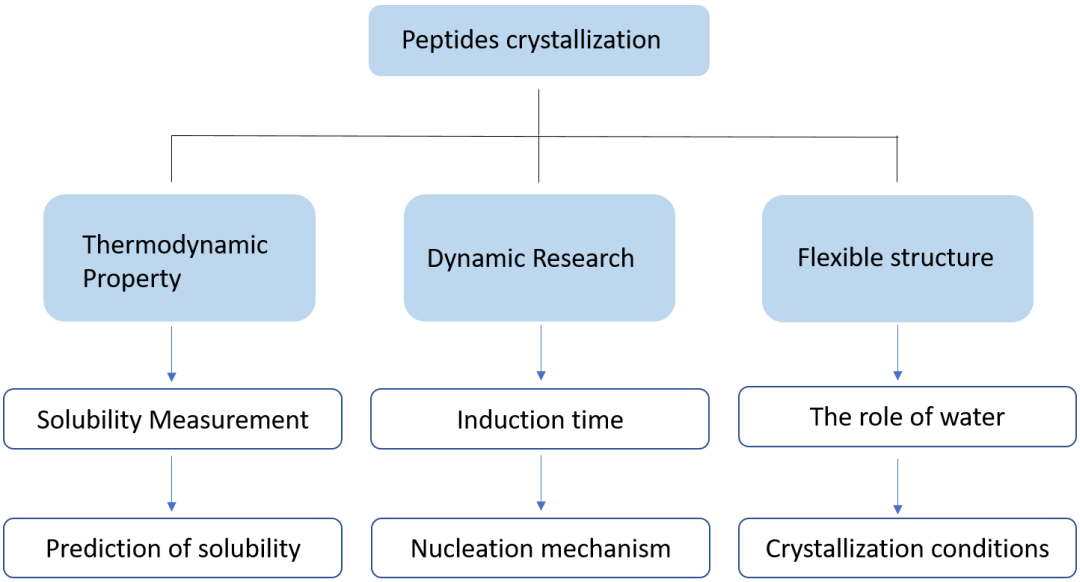


Figure 1.1The research objectives of peptide crystallisation

The layout of this thesis is as follows. In the first chapter, there is a brief introduction about the background and the objective of the PhD research. Chapter 2 presents a review of the literature, which includes a definition of peptides, current research progress on solubility

measurement, solubility prediction, the SAFT modelling tool, peptide conformation, and also approaches for peptide crystallisation. Chapter 3 introduces the research methodology, which include the characterisation techniques and modeling tools used in this work. In chapter 4, the solubility of glycine homopeptides and dipeptides in water from 278.15K to 313.15K were measured to explore the effect of chain length and side chains on peptides solubility. In chapter 5, the SAFT- γ Mie group contribution equation of state was expended to predict the solubility of amino acids and peptides. Based on the thermodynamic propeties, the nucleation mechanism of glycine homopeptides has been investigated in chapter 6. Chapter 7 discusses the role of water in the crystallisation of peptides hydrates and the relationship between the conformation and crystallisation conditions. In chapter 8, the conclution of the whole thesis and and future works based on the current research were gaven to provide the insight into peptide crystallisation.

Chapter 2. Literature Review

2.1 Definition of peptides

A peptide is a short chain protein composed of 2-50 amino acids connected by peptide bonds. Amino acids are the basic building bricks of peptides and proteins. Each amino acid has a side chain-R that confers unique chemical and physical properties on peptides. The amino acids are joined via a peptide bond (CO-NH) formed by the condensation process of the carboxyl group (COOH) and the amino group (NH_2). The amino acid fragment in the peptides is called the residue. The formation of a dipeptide is depicted in Figure 2.1.

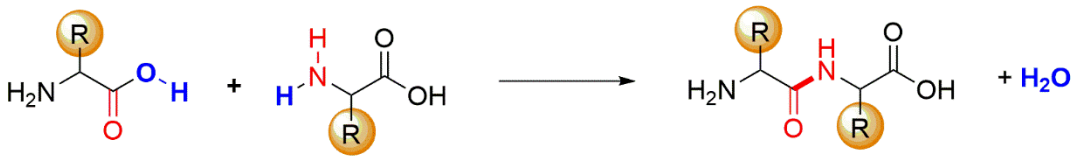


Figure 2.1 The forming process of a dipeptide

2.1.1 The properties of amino acids

There are 20 standard/coded amino acids, which includes 19 α-amino acids and the amino acid proline. Every amino acid has a one-letter and a three-letter abbreviation to help simplify the written sequence. Figure 2.2 shows the backbone structure and the different side chains of the 20 naturally occurring amino acids[17].

2.1.1.1 Acid/Base properties

Each amino acid contains α-amino and α-carboxylic acid group with pKa values of 9-10 and 2-2.5, respectively. However, these two groups do not dictate the acid/base properties of the 20 standard amino acids; rather, they are determined by the ionizable side chains of the amino acids. For example, aspartic acid contains a carboxylic acid group with a pKa value of

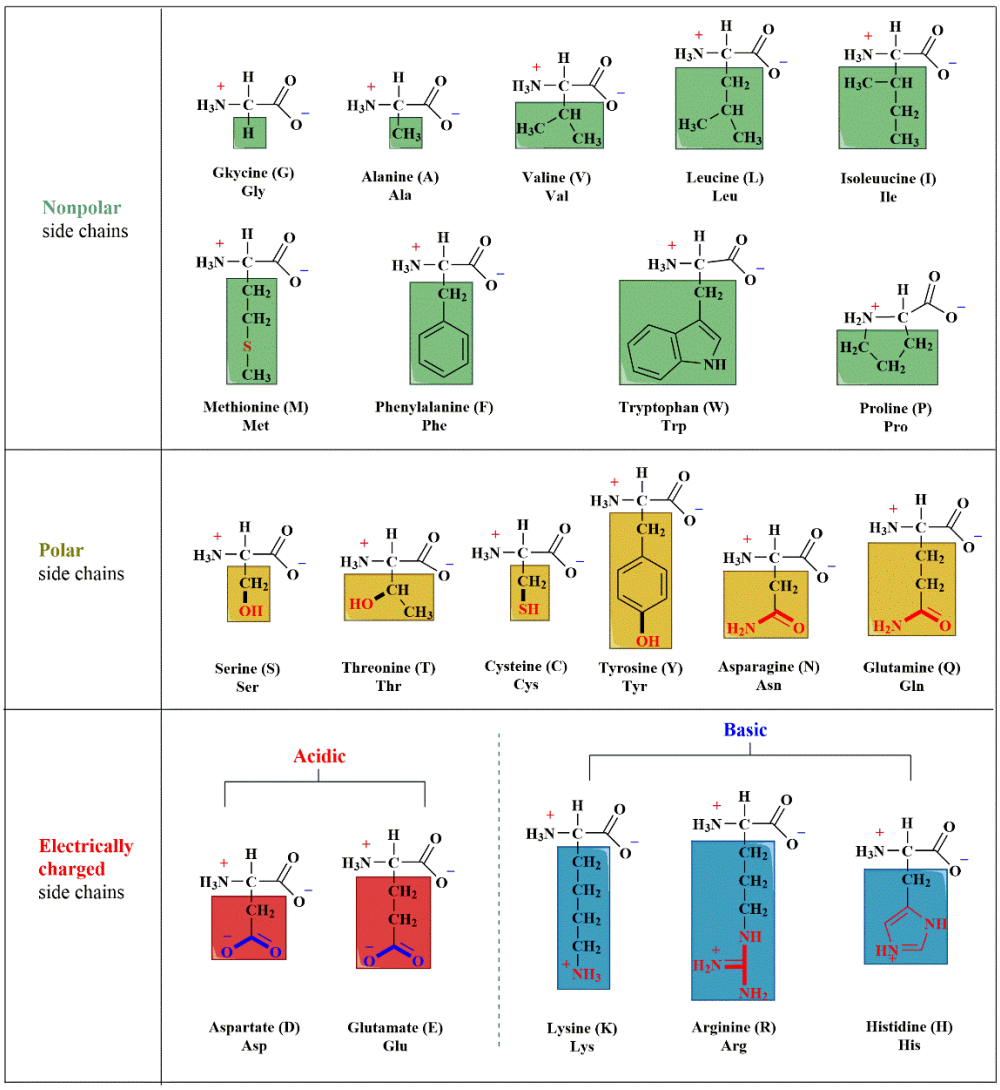


Figure 2.2 Classification of 20 standard amino acids[17]

3.5-4.5. Arginine contains a basic amino acid side chain with a pKa value of between 12 and 17.[18] Hence, in neutral solution, the acid residues in amino acids will carry negative charge and the basic one will be positively charged. Three amino acids have basic side chains in neutral pH solution, while two amino acids have acid side chains in neutral pH solution, as illustrated in figure 2.2. Each amino acid has its own particular pH value, also known as isoelectric point, at which the positive charge and negative charge are equal, and the total charge of the amino acid is zero.[19] Similarly, peptides and proteins also exhibit the isoelectric points.[20] At the isoelectric points, peptides or proteins have the lowest solubility and likely to crystallise out.[21]

2.1.1.2 Hydrophilic/Hydrophobic properties

Hydrophobic and hydrophilic amino acids are categorised according to the propensity of their side chains to interact with water. There are nine hydrophobic amino acids and six hydrophilic amino acids (figure 2.2). Hydrophobic amino acids include side chains that contain carbon and hydrogen, which have a natural tendency to reject water. Six hydrophilic amino acids have polar but uncharged side chains; these amino acids are frequently found on the surface of peptides.[22] Apart from acidic/basic and hydrophilic/hydrophobic, there are additional classification schemes for these twenty standard amino acids, including charged, polar, and nonpolar. For instance, Aftabuddin and Kundu classified them into hydrophobic, hydrophilic and charged.[23] Each classification standard is founded on the unique features of the side chains.

2.1.2 Bonds between amino acids

A peptide bond is a molecular bond which connects two amino acid monomers in peptide or protein chains [24]. When the carboxyl group of an amino acid A reacts with the amino group of an amino acid B, a peptide bond is formed with the release of a water molecule. The most important function of peptide bond is to connect amino acid residues together to form the primary structure of peptides and proteins[25]. Additionally, peptide bond has UV absorption property at 190-230 nm wavelength, so UV-Vis can be used to measure the concentration of peptides or proteins[26, 27]. Another significant link found in peptide structures is the disulfide bond, commonly known as the disulphide bridge, which is used to connect cysteine residues in peptides or proteins and thus plays a significant function in folding, such as in the formation of the secondary structure of peptides[28]. Disulfide bonds can occur between two peptide chains or just in one peptide chain. For instance, insulin contains three disulfide bonds: two inter-chain disulfide bonds and one intra-chain linkage (figure 2.3).[29]

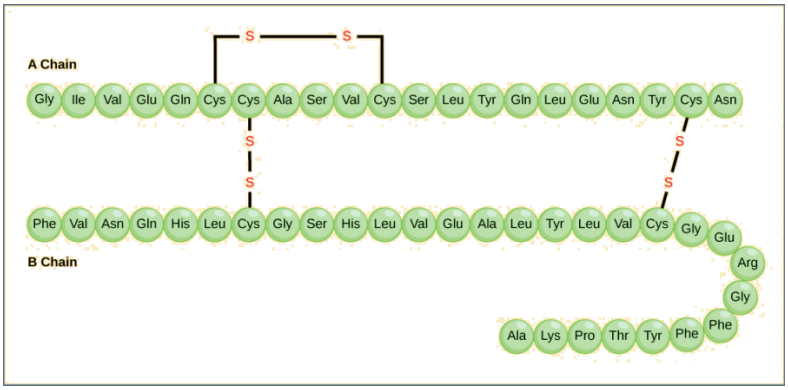


Figure 2.3 The three disulphide bonds in Insulin[29]

2.1.3 The secondary structure of folded and unfolded peptides

The structure of peptides can be classified into four levels: primary structure, secondary structure, tertiary structure and quaternary structure (figure 2.4).[30] The primary structure of a peptide is a linear sequence of amino acids starting at the amino-terminal (N) end and ending at the carboxyl-terminal (C) end. The secondary structure is the space conformation of these peptide linear chains by intramolecular hydrogen bonds.[31] Tertiary structure is a term that refers to the three-dimensional structure that a single peptide chain can form through salt bridges, disulfide bonds and the tight packing of side chains. The quaternary structure is a three-dimensional structure made up of two or more independent peptide chains stabilised through non-covalent interactions and disulfide bonds.

Among these, secondary structure serves as the foundation for tertiary and quaternary structures. And the alteration of the secondary structure is always employed to investigate the effect of additives on protein structures, protein denaturation, and protein stability. For the folded protein, secondary structure can be divided into three types, namely α helix, β sheet and coil. The former two structures are held in shape by hydrogen bonds, forming between the carbonyl (C=O) of one amino acid and the amino H (N-H) of another amino acid.

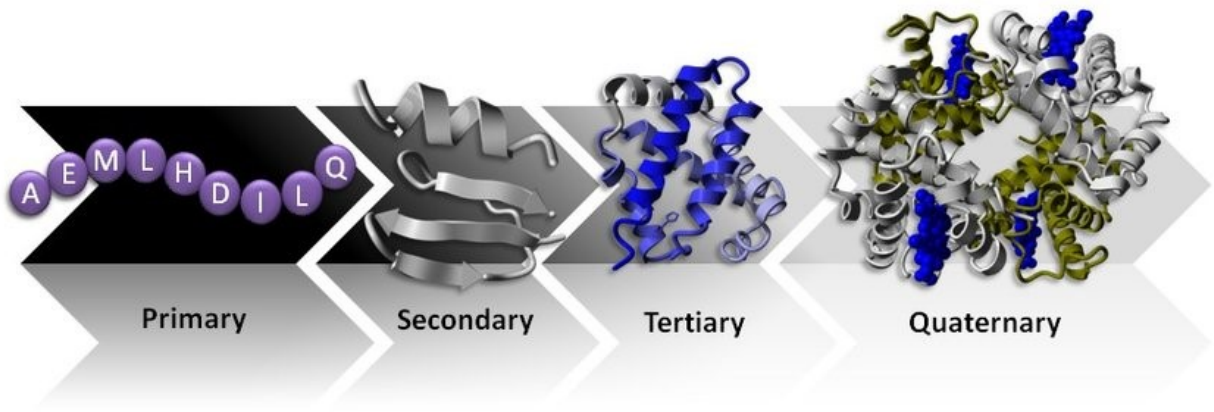


Figure 2.4 The four levels of peptides structure [32]

In α helix structures, the hydrogen bonds resemble the peptide chain to curl and helical structures, with each turn containing 3.6 amino acids. A β sheet structure is formed when two peptide chains are linked together by hydrogen bonds, which contains two types according to the direction of connection: anti parallel β sheet and parallel β sheet. The R groups of these two structures both stick outward, allowing them to interact with each other freely. Because of the plenty space left for side chains, the amino acids with ring structures in their R groups are frequently found in β sheets.

There is another structure class for unfolded proteins based on the new dihedral angle, which is called poly-L-proline type II (pPII or pGII). A peptide's dihedral angle is the internal angle of its backbone formed by the intersection of two adjacent planes. This structure was discovered in the last two decades and has since become the dominant conformation for unfolded peptides and proteins.[33-36] The CD spectra of pPII structure was first discovered by Tiffany and Krimm[37] and the X ray structure was first resolved by Bamford et al.[38], however, it has long been viewed as a random coil structure.[39] Following additional investigation into the secondary structure, this distinctive structure was discovered in an increasing number of peptides, including poly (Ala–Gly–Gly), poly(Pro–Ala–Gly), poly(Ala–Ala–Gly), and

poly(Pro–Ser–Gly).[33] This conformation also occurred frequently in the proline-, alanine- and glycine-rich regions in the connected and turning parts between the α-helix and β-structure in folded proteins. Besides, there are some other secondary structures (such as 3_{10}-helix, π -helix) reported that have unique dihedral angels but are not very common occurrence as the above secondary structures. The dihedral angles of different secondary structures can be seen in table 2.1.

Table 2.1 The dihedral angles for the main secondary structures

	α-helix	3_{10}-helix	Pi-helix	Parallel beta sheet	Anti- parallel beta sheet	Left-handed pPII helix
Frequency	Abundant	Infrequent	Rare	Abundant	Abundant	Infrequent
φ	-57	-49	-57	-119	-139	-75
ψ	-47	-26	-70	113	135	145

2.2 The thermodynamic properties of peptides

2.2.1 Solubility curves for small molecules and macromolecules

Solubility is defined as the ability of solids to dissolve into a given amount of solvent.[40] Investigating the solid-liquid equilibrium, solubility is a critical thermodynamic attribute of peptides in industrial crystallisation. It is helpful to design the phase diagram of the system and to optimize the crystallisation conditions. The fundamental theory of peptide crystallisation is similar to that governing the crystallisation of small molecules and proteins, namely the establishment of supersaturated solution systems. The difference in concentration between the actual concentration and the solubility at a particular temperature is called supersaturation. For example, when a solution that is saturated at a high temperature is cooled rapidly, the solution's concentration is greater than its solubility at the low temperature, and the supersaturation

provides a driving force for the materials to crystallise out until the solution's concentration reaches the solubility point at the low temperature, thereby achieving solid-liquid equilibrium.

For small molecules, the supersaturated state can be achieved by lowering the temperature (cooling crystallisation), increasing the concentration (evaporation crystallisation), and introducing anti-solvents (solvent-out crystallisation).[41-44] Figure 2.5 illustrates the phase change that occurs during the crystallisation process. There are primarily three zones depicted in this diagram: In undersaturated region, crystals will be dissolved without any nucleation occurring. In metastable region, the crystals will grow up if seeds are introduced. However, there is no nucleation happens automatically in this region if there are no seeds. The region in which solution can spontaneously nucleate is referred to as the labile region.[45, 46]

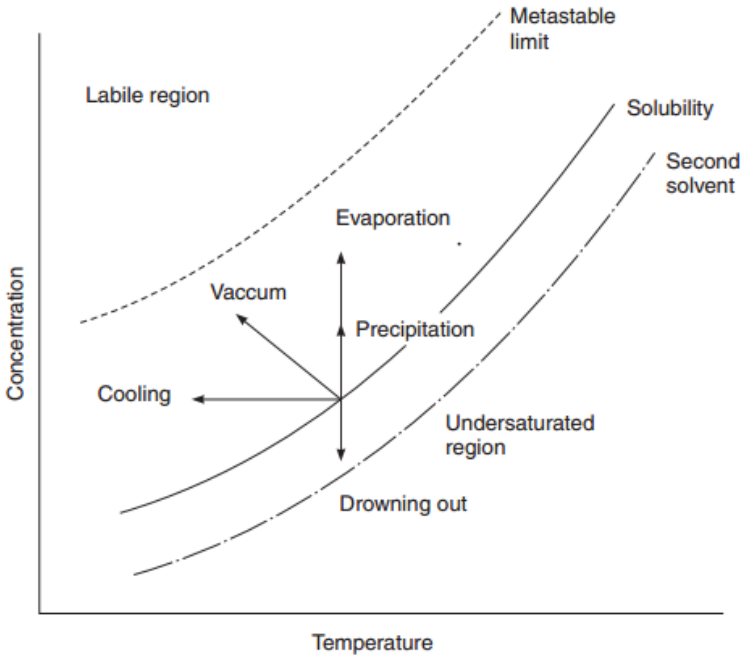


Figure 2.5 The solubility-supersaturation diagram of small molecules [45]

In the case of proteins, it is extremely difficult to quantify the change in solubility with temperature. Firstly, proteins may denature rapidly at high temperatures. Secondly, in order to stabilise proteins, precipitants, salts and buffers will be used during the crystallisation. The adjustable parameters such as the concentration of salts, the pH of buffers and the kinds of

precipitants will affect the solubility of peptides.[47] Figure 2.6 shows the phase diagram of protein crystallisation.[48]

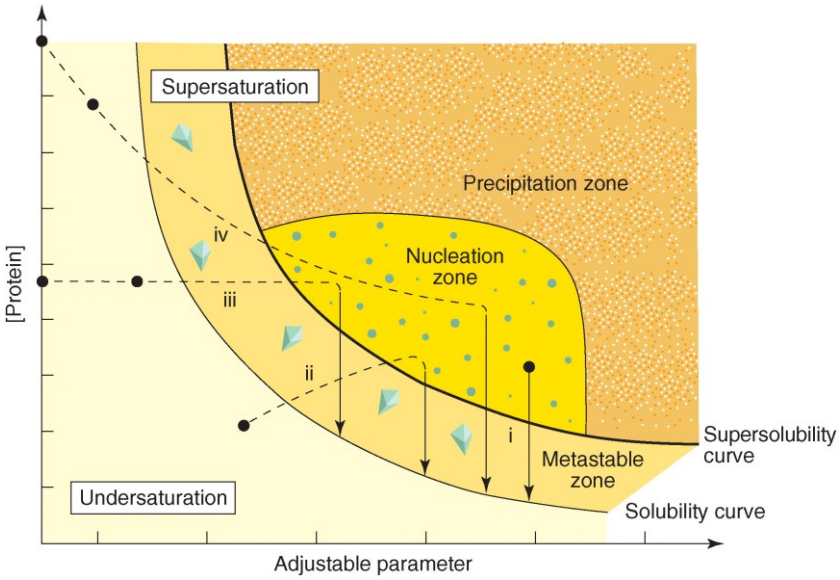


Figure 2.6 The solubility-supersolubility diagram of proteins (consisting of four zones representing different degrees of supersaturation: (i) Precipitation Zone: the protein will precipitate; (ii): Nucleation Zone: spontaneous nucleation will take place; (iii) Metastable Zone: crystals are stable and may grow but no further nucleation will take place; (iv) Undersaturation Zone: protein is fully dissolved and will never crystallise).[48]

2.2.2 Solubility studies of peptides

Accurate solubility data and associated thermodynamic parameters of peptides under different temperatures are used to prepare the supersaturated solutions, calculate the kinetic parameters and control the crystallisation conditions. However, few solubility data of peptides in organic solvent systems is available in the existing literature.[49] Compared with small molecules, the challenges of peptide crystallisation are their flexible conformation and biological activity, especially when the chain length is longer than five.

Most of the current research about the peptides solubility focus on short chain peptides. Nozaki investigated the solubility of amino acids, diglycine, and triglycine in water, as well as the effects of ethanol and dioxane.[50] Narita et al. found that the solubility and the potential of

beta-sheet conformation of Octapeptides which includes Gln-Val-Gly, Asn-Ala-IIe and Phe—Leu are not strongly dependent on their amino acid sequence but rather on their amino acid residues composition.[51] Ching reported the solubility of glycine homopeptides in water at various pH values under room temperature, as well as the effect of ethanol, NaCl and PEG 6000.[52] The investigation discovered a cross pH point at which the solubility has the lowest value, which corresponds to the isoelectric point mentioned in Section 2.1.1.1. Additionally, it has been discovered that Ethanol and PEG 6000 can lower the solubility and can be employed as anti-solvents during the crystallisation process. They also found NaCl can increase the solubilities of diglycine and triglycine, however, for glycine, the solubility increases initially and then decreases as the NaCl concentration increases. Castronuovo and Breil et al. measured the solubilities of the cyclic dipeptides Gly-Gly, Ala-Ala, and Leu-Gly in water and Gly-Gly and Gly-Ala in aqueous electrolyte solutions at 298.15 K.[53, 54] Held et al. researched the melting properties and solubilities of dipeptides composed of glycine or alanine.[55] Their group enlarged the data set to include di-and tripeptides composed of glycine, alanine, leucine, proline, and serine.[56] However, all of these studies only measured the solubility at room temperature and did not explore the temperature dependence. Besides, the effect of side chains, conformation and the morphology on the peptides solubility has not been studied systematically.

There are two common ways widely described in literature for solubility measurement: the static method and dynamic method.[57-59] For the static approach, the extra solids are added into the solvents and stirred for a long time until the system research to the solid-liquid equilibrium, then the concentration of the filtrated liquid supernatant is the solubility.[60, 61] There are also several methods for determining the concentration, including gravimetric analysis, spectrum analysis and NMR.[62-65] Garcia and co-workers explained the different flow chart for the gravimetric and NMR methods (figure 2.7).[66] For gravimetric analysis, the weight of the supernatant and the weight of the dissolved solids remaining after drying were measured using

a balance. The solubility can be calculated as the mass of the solids divided by the mass of the solvent. For spectrum analysis techniques such as UV-Vis, FTIR, HPLC, and NMR, the calibration curve of the solute is measured in order to determine the linear relationship between the concentration and absorbance of the specific lights.[60] The solubility can be calculated by the absorbance of the specific wavelength based on the calibration curve. Compared with gravimetric method, the spectrum methods give a higher accuracy for the low solubility and can be applied for a very small amount of sample. For the dynamic method, the solute is added continuously into the solvent until it reaches thermodynamic equilibrium under the specific temperature. The solubility is calculated from the amount of the solute and the mass of the solvent.

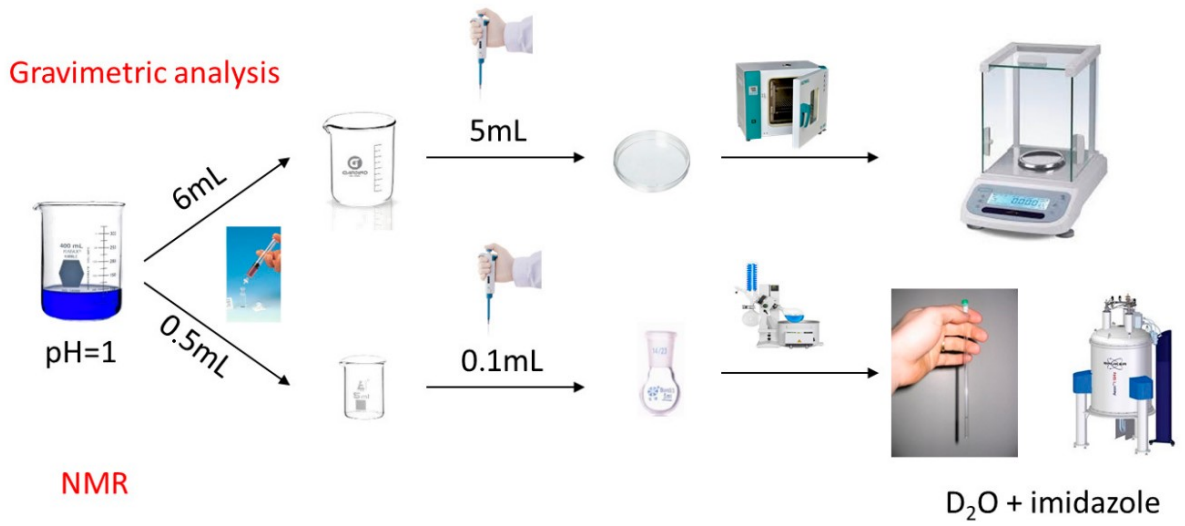


Figure 2.7 Flow chart of the gravimetric and NMR methods used for determining solubility [66]

2.3 Solubility prediction

2.3.1 Prediction of solubility based on classical thermodynamic equation of state

Due to the lack of the available experimental solubility data for most peptides and proteins, many researchers have focused on the solubility prediction. For the prediction, there are two

different pathways: one is to predict the solubility in different solvents under different temperatures using the classical thermodynamic models with the correlated parameters. Another one is predicting the solubility of macromolecules based on their structures, like predicting the solubility of the proteins using the amino acid composition and sequence.

Several classical thermodynamic equations are commonly applied to correlate the solubility of small molecules in a given solvent system in literatures.[67-70] For example, the modified Apelblat equation (2.1) is used to correlate the relationship between solubility in pure solvent and temperatures. While, CNIBS/R–K model (2.2) illustrates the relationship between two solvent components and isothermal solubility.

$$\ln x_1 = A + \frac{B}{T} + C\ln T \tag{2.1}$$

where x_1 and T is the mole fraction solubility of solute and temperature, respectively. A, B, and C are all empirical constants.

$$\ln x_1 = x_2^0 \ln(x_1)_2 + x_3^0 \ln(x_1)_3 + x_2^0 x_3^0 \sum_{i=0}^{n} s_i (x_2^0 - x_3^0)^i \tag{2.2}$$

where x_1 is the mole fraction solubility of a material, $(x_1)_i$ is the mole fraction solubility of the material in pure solvent I; x_2^0 and x_3^0 express the initial mole fraction of two different solvent compositions in the solvent system. s_i is the model parameter; n represents the number of "curve-fit" parameters.

2.3.2 Prediction of solubility based on Statistical Associating Fluid Theory (SAFT)

The most classical thermodynamic equations of state are variations on the van der Waals equation with the hard sphere as the reference system. The repulsive and dispersion interactions and other long range forces are considered.[71] However, the hard-sphere reference is not suitable for most of the fluids which are non-spherical and associating. To extend prediction beyond

simple fluids with classical thermodynamic equations of state to more complicated fluids with associating interactions, the statistical mechanics concept was applied to the equations of state, resulting in the Statistical Associating Fluid Theory (SAFT).[71] Figure 2.8 shows the bond energy for the different classifications of fluids, which covers a wide range with several orders of magnitude. SAFT is an approach which combines Wertheim's thermodynamic perturbation theory (TPT) with equations of state to calculate the thermodynamic properties and phase equilibrium in the associated fluid systems where the hydrogen-bonded and donor-acceptor clusters are exist.[72, 73] For simple molecules, repulsion and dispersion are the most important interactions, as well as the weak electrostatic forces. In the SAFT equation of state, the studied molecules are regarded as a chain which composed by different spherical segments (figure 2.9). The total intermolecular potential includes three major contributions: the repulsive interactions between hard-sphere, chain formation of several hard spheres, and association. From figure 2.9, short chains are formed using covalent bonds to produce tetramers and the hydrogen bonds at the terminal sites of tetramers connect the chains together to form chain oligomers. Furthermore, there are also dispersion forces that are to be considered. For SAFT, there are many variations developed currently based on the varying potentials between the hard-spheres. The three main variations are PC-SAFT, SAFT-VR and SAFT- γ Mie.

PC-SAFT is based on the perturbed-chain statistical association theory.[74, 75] The advantage of PC-SAFT is the good prediction for the co-polymers with the interacting parameters derived from the homopolymers, which means the properties of a macromolecule solution can be predicted with the known pure component parameters of the specific amino acids.[76] Cameretti modeled the vapor pressures, liquid densities and solubilities of glycine, alanine, serine, proline, and valine using PC-SAFT. The parameters of these amino acids were then used to predict those properties of glycine homopeptides and Ala $(Gly)_{n=1-4}$.[76] Fuchs and co-workers predicted

the solubility of glycine and alanine in water and ethanol mixtures by PC-SAFT.[49] The vapor pressures and densities of glycine and alanine in water were employed to fit the pure component parameters, which can be used to predict the solubility of glycine and alanine in water-ethanol mixtures without fitting any additional parameters. Do et al. presented the solubility prediction

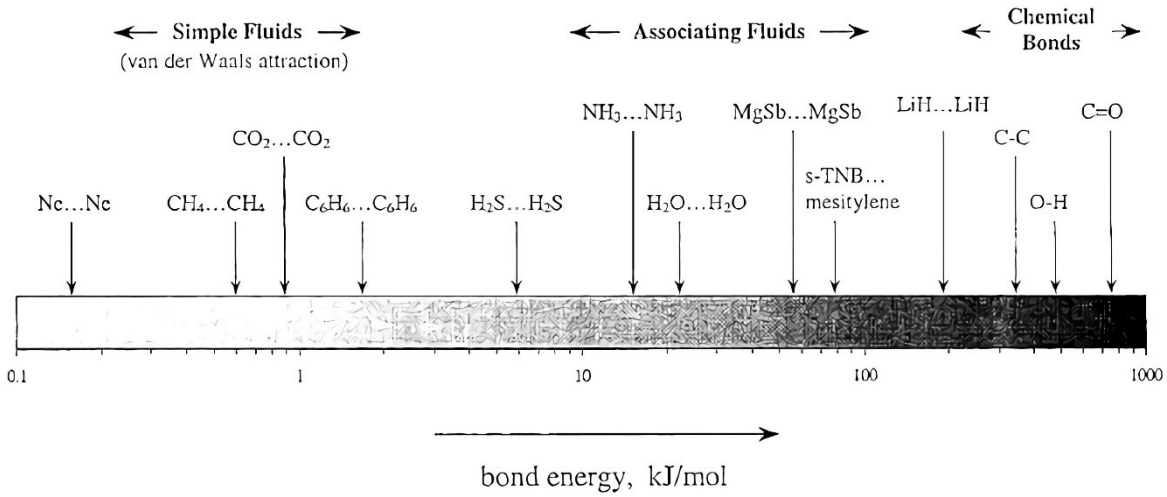


Figure 2.8 The perturbation scheme and interactions for the formation of a molecule within the SAFT

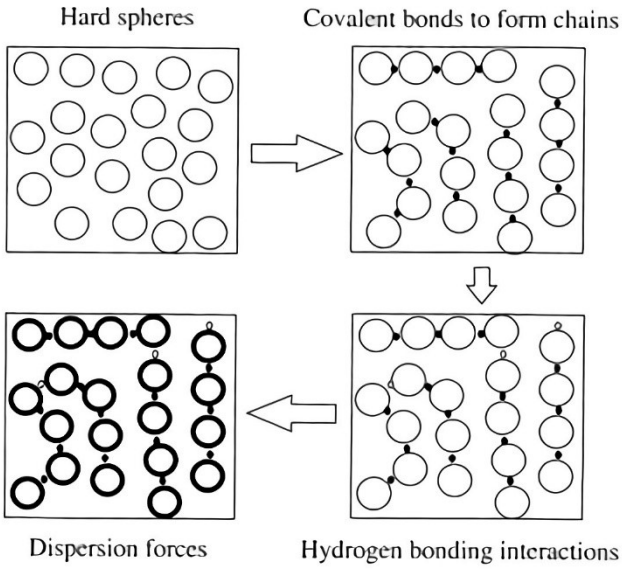


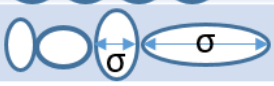
Figure 2.9 Continuous distribution of bond strengths, showing the span from simple van der Waals attractions to the formation of chemical bonds.[71]

of four dipeptides and one cyclo-peptide composed of glycine and alanine using PC-SAFT, which distinguished the different solubility properties of isomers.[55]

SAFT-VR was established by Gil-Villegas and coworkers for chain molecules with variable attractive potentials.[77] They predicted the vapour-liquid properties of pure alkanes and perfluoroalkanes using SAFT-VR, improving the chain contribution and the description for the dispersion force of SAFT. Anvari and colleagues used SAFT-VR to determine the vapour pressure and activity of water in amino acid solutions containing different salts and then verified the accuracy of the model.[78] Peng et al. integrated the GC and SAFT approaches to develop a GC-SAFT-VR model capable of distinguishing the structural isomers with the specified chain connectivity.[79]

Compared with these SAFT Eos, SAFT-γ Mie discussed in this work is based on a Mie potential of variable repulsive and attractive range and combines group-contribution (GC) theory, which means it can predict different structurally similar peptides without any further parameters from the new peptides. The details about SAFT-γ Mie theory can be found in chapter 5. Dufal et al. examined the performance of the SAFT-γ Mie in the description of solid−liquid equilibria of long-chain linear alkanes in n-butyl acetate, obtaining a good prediction result and expanding the database of the SAFT-γ Mie.[80] Lecce and co-workers used SAFT-γ Mie to predict the thermophysical properties of aqueous mixtures of choline and geranate (CAGE) and got a good agreement with the experimental results.[81] There is no literature reported the prediction of properties of amino acids and peptides using SAFT-γ Mie so far. The comparison of these variations of SAFT can be seen in table 2.2.

Table 2.2 The comparison between different SAFT equations*

	λ^r	λ^a	σ
PC-SAFT	12	6	
SAFT VR	variable	6	
SAFT- γ Mie	variable	variable	

* λ^r, λ^a, σ represent the repulsive exponent, attractive exponent, and the segment diameter, respectively.

2.3.3 Prediction of solubility based on amino acid composition

Under the given experimental conditions, theoretical predictions of solubility from the sequence of proteins have been reported currently. Frishman established PROSO, a machine-learning approach, which can assess the solubility of a protein in Escherichia coli through the amino acid composition.[82] Roger used statistical analysis to predict the solubility of recombinant proteins using only the protein's amino acid composition.[83] There is also a web server called Protein-Sol, which can predict the solubility of proteins from sequence. This web server uses available data for Escherichia coli protein solubility in a cell-free expression system, the calculated results can represents 35 sequence-based properties.[84]

2.4 Nucleation during crystallisation

As the first step in the process of crystallisation, nucleation plays a decisive role in the control of the particle size distribution, polymorphism, perfection, and other characteristics of crystals. Therefore, the design of controlled process that ensures desired properties of a compound highly relies on the understanding of the fundamentals of nucleation. Generally, there are two mainly theories of nucleation summarized from current crystal nucleation research.

2.4.1 The classical nucleation theory

The classic nucleation theory, proposed by Gibbs in the 19th century remains the most common theoretical model for the understanding of nucleation. It prefers that the formation of clusters are caused by concomitant density and order fluctuations.[85, 86] The clusters then proceed to consolidate to create the crystalline nucleus, from which crystals can be formed. Homogeneous and heterogeneous nucleation both can be explained by classical nucleation. From the kinetic view of classic nucleation theory, the state-state rate of nucleation (J) is the number of particles formed from a supersaturated solution per unit time per unit volume, which is expressed in the equation (1): [87]

$$J = A\exp(-\frac{B}{\ln^2 S}) \tag{2.3}$$

where A is the pre-exponential factor, B is the thermodynamic factor and S is the supersaturation ratio; and the exponent $B/\ln^2 S = W/kT$ also represents the dimensionless nucleation work.[88]

2.4.2 The non-classical nucleation theory

Due to the apparent discrepancy between theoretical predictions and practical outcomes for a large number of systems, non-classical nucleation theories have been proposed to describe the complex nucleation behaviour. Lutsko et al. calculated the droplet free energies using DFT and discovered they differ from the calculated value from CNT. Additionally, he discovered the free energy during the crystal growth stage exhibits a shallow minimum, indicating that a metastable form exists at the start and nucleation requires a higher free energy in the subsequent step.[89] Jacobson demonstrated a multistep clathrate crystallisation process using MD simulations: solution –blob–amorphous clathrate–crystalline clathrate.[90]

Numerous non-classical nucleation hypotheses have been developed over the last few

decades. Among them, the two-step nucleation theory, in which the solution to crystal transformation proceeds through a dense, disordered liquid phase and then into crystalline order was suggested by simulations and described experimentally later for protein system.[91-93] Besides, there are some other nucleation mechanisms proposed in recent years (figure 2.10), including the prenucleation clusters (PNCs) theory and crystallisation by particle attachment (CPA). [1, 94, 95]

Before macromolecules nucleated, an essential event known as liquid–liquid phase separation occurred, in which a solute-rich phase coexisted with a solute-poor phase. The exchange of macromolecule/water interactions for macromolecule/macromolecule and water/water interactions in the liquid-liquid phase separation caused the nucleation of the crystallisation.[96] Wang and co-workers reported the first observation of liquid−liquid phase separation (LLPS) in peptide solutions (figure 2.11) and they found the temperature of the liquid−liquid binary coexistence curve is lower than the solubility line, which can be seen in figure 2.12.[97] Yuan et al. reported the liquid–liquid phase separation (figure 2.13) is a fundamental step leading to the nucleation of supramolecular nanofibrils and the form of the liquid droplets is an entropy driven process.[98]

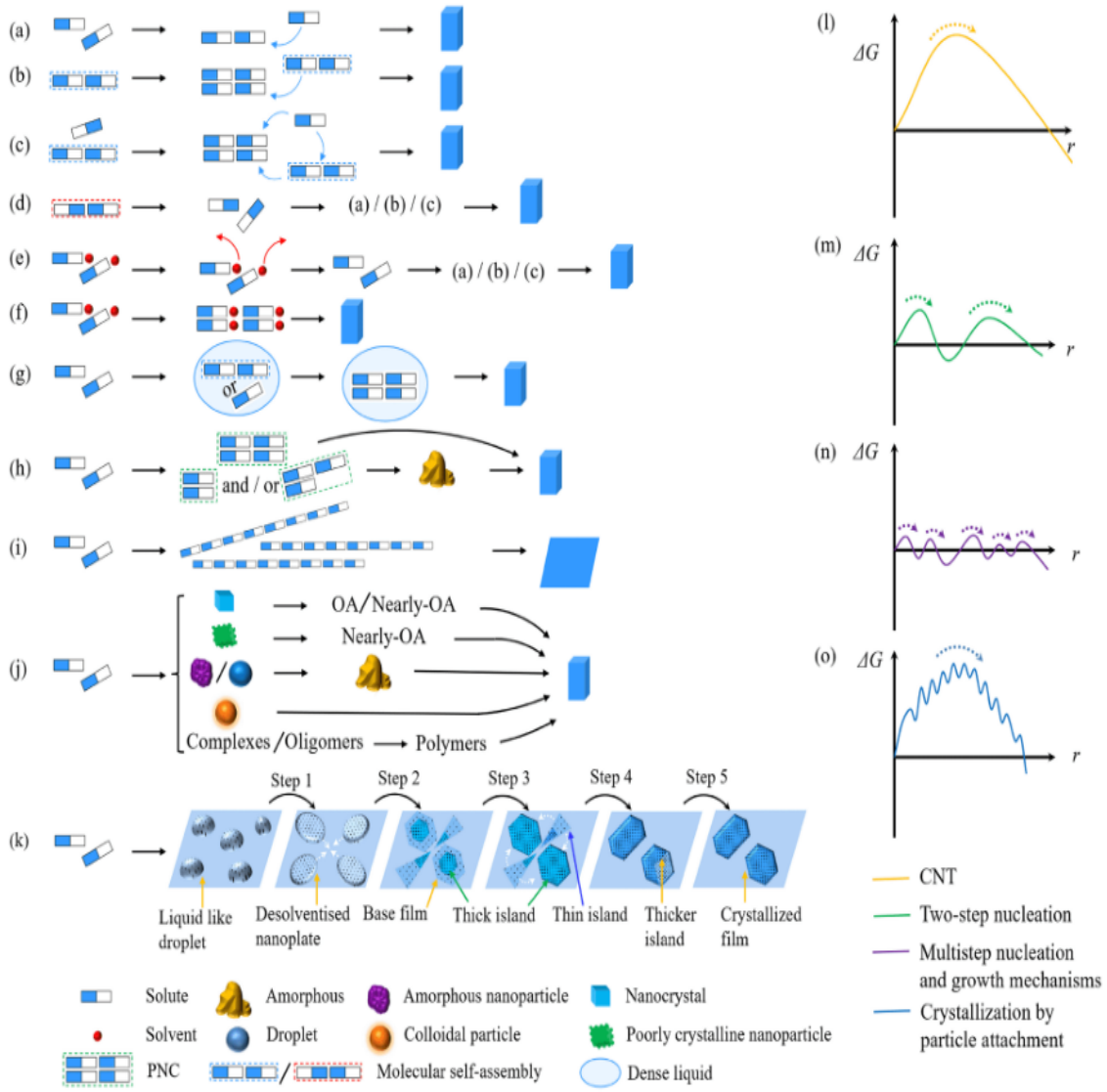


Figure 2.10 Sketch map of different nucleation theories/pathways and their free energy change during nucleation. (a) CNT; (b) nucleation by solute molecular self-assemblies; (c) nucleation by solute monomers and solute molecular self-assemblies; (d) nucleation after molecular rearrangement; (e) nucleation by solvated solute molecules after desolvation; (f) nucleation by solvated solute molecules; (g) two-step theory; (h) PNCs theory; (i) building one row at a time; (j) CPA; (k) multistep nucleation and growth mechanisms; (l) the free energy change of CNT; (m) the free energy change of two-step nucleation mechanism; (n) the free energy change of multistep nucleation and growth mechanisms; (o) the free energy change of CPA (color online).[1]

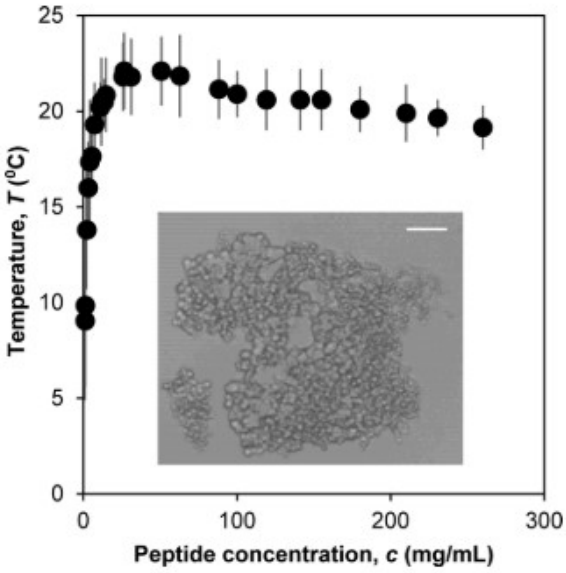


Figure 2.11 Coexistence curves of LLPS in solutions of an incretin peptide (IP5) at pH 5.5. Vertical bars indicate the temperature hysteresis between the clouding and clearing of the solutions. Inset shows a bright-field microscope image of the peptide-rich liquid condensed phase formed in a 30 mg/mL peptide solution at 21 °C. The white scale bar in the upper right corner represents 10 μm.[97]

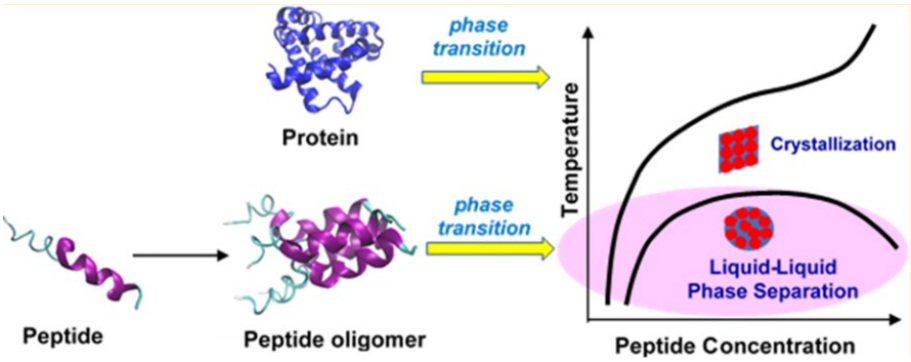


Figure 2.12 The liquid-liquid phase separation region and crystallisation region for macromolecules.[97]

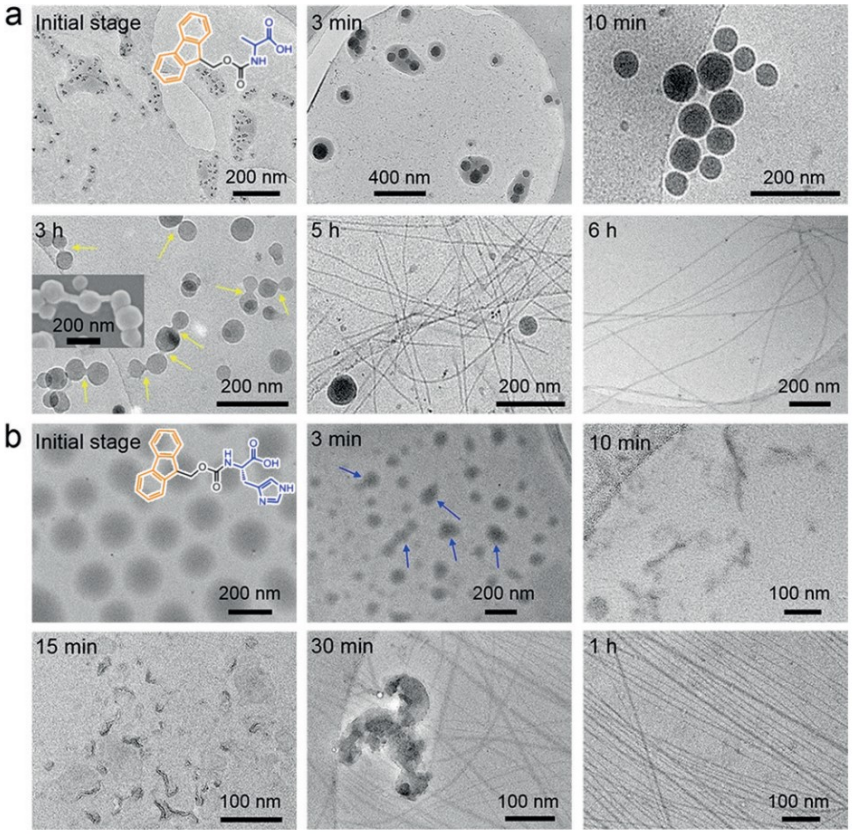


Figure 2.13 Cryo-TEM images of the dynamic evolution process of Ag^+ coordinated Fmoc-Ala (a) and Fmoc-His (b) self-assembling nanofibrils over time. These images demonstrate the structural evolution from the amino acid dense liquid droplets to the metastable solidified droplets and, finally, to thermodynamically favorable nanofibrils through the liquid–liquid phase separation, nucleation, and subsequent nucleation–elongation processes. Yellow arrows denote the assembly intermediates formed through the elongation of the nuclei from the solidified droplet surfaces, which is further confirmed by the inserted SEM image (see inset). Blue arrows represent the coalescence of Fmoc-His liquid droplets.[98]

2.5 Peptide crystallisation

When the number of amino acids residue is less than five, some peptides can crystallise using the same crystallisation methods of small molecules. However, with the increase of the residue numbers, the crystal is very difficult to crash out and will have to adopt the crystallisation methods of macromolecules, like proteins.

2.5.1 Crystallisation methods for small molecules

According to the crystallisation mechanism outlined in Section2.2.3, the crystallisation methods used in industry can be classified as follows: cooling crystallisation, anti-solvent crystallisation, evaporation crystallisation, and reactive crystallisation.

Cooling crystallisation normally is defined as the formation of crystals with desirable and consistent properties through decreasing temperature from solution, where the solute solubility sharply increases with temperature.[99] Cooling crystallisation is the most common way for batch crystallisation in pharmaceutical, petrochemical and semiconductor industry as the temperature profile can be easily controlled to get perfect control of supersaturation profile. Anti-solvent crystallisation is a process in which a second solvent (anti-solvent) is added to the solution for creating supersaturation.[100] Normally the solute is more insoluble in the anti-solvent than in the primary solvent which causes a solubility decrease in the miscible binary solvent mixture. This technique is used for the substance whose solubility has weak function of temperature or which is unstable in high temperature. Evaporative crystallisation is a technique of choice for substances whose solubility is weakly dependent on temperature.[44] Supersaturation is achieved in this procedure by evaporating the solvent. Normally evaporative crystallisation is operated at constant temperature and reduced pressure in practice. Reactive crystallisation is operated by creating supersaturation of a compound by chemical reaction.[101] Examples of industrial relevance include the liquid-phase oxidation of para-xylene to terephthalic acid, and the absorption of ammonia in aqueous sulfuric acid to form ammonium sulfate. In this method, reactions normally can be very fast which result in high nucleation rate and small size crystals. Sometimes it also could be called precipitation.

2.5.2 Crystallisation methods for macromolecules

Currently, macromolecules are crystallised in primarily two ways: vapour diffusion and

microbatch. Additionally, there are other techniques available, including dialysis, free-interface diffusion, and microfluidic chip technology.

Vapor diffusion is a commonly used method of macromolecules crystallisation, which includes hanging-drop and sitting-drop [102]. A droplet containing purified protein, buffer, and precipitant will equilibrate with a reservoir containing a greater concentration of the same buffer and precipitant in a sealing environment. The theory is that water diffuses from a low concentration droplet to a higher concentration reservoir, resulting in the crystal emerging with a gradual reduction in the droplet's concentration. The difference between hanging-drop and sitting-drop is the location of the droplet, which is either on an inverted cover slip or on a pedestal, respectively [103]. The details can be seen in figure 2.14.

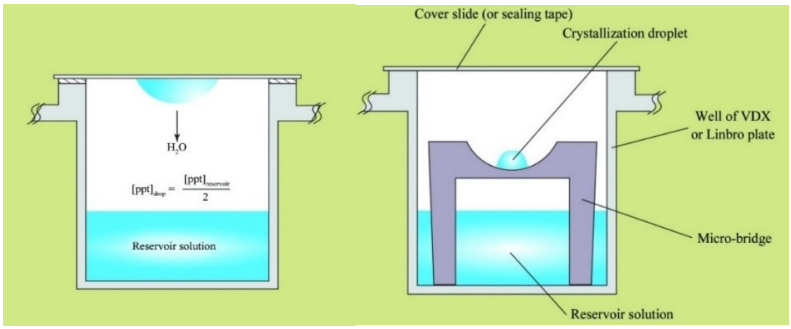


Figure 2.14 The schematic diagrams of hanging-drop vapor-diffusion method (left) and sitting-drop vapor-diffusion method (right) [103]

Microbatch is a technique in which small amounts of protein (1 μl) are combined with stock solutions and immersed in oil to prevent evaporation. The advantage of this method is the small amount of the sample required and the whole experiment can be protected from exposure to the air. Figure 2.15 illustrates the pipetting procedure in detail.[104]

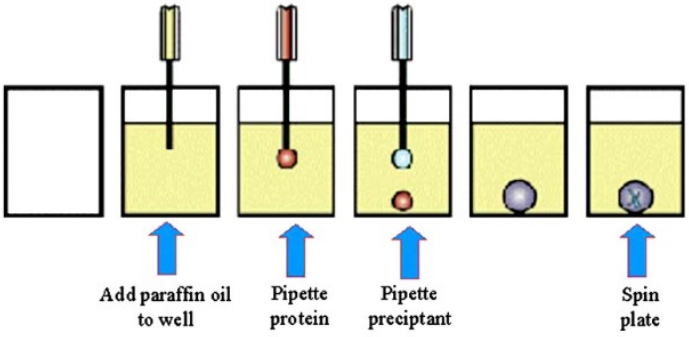


Figure 2.15 Pipetting process for microbatch experiment [103,104]

Besides vapour diffusion and microbatch methods, microdialysis is also an effective method using the theory of salting out and salting in. A semi-permeable membrane, which only allow small molecule and ions to pass across is employed to separate the high concentrations of salt and protein solutions. When the system reach equilibrium, the protein solution become supersaturated, and the crystals will come out, which can be seen in figure 2.16.[105]

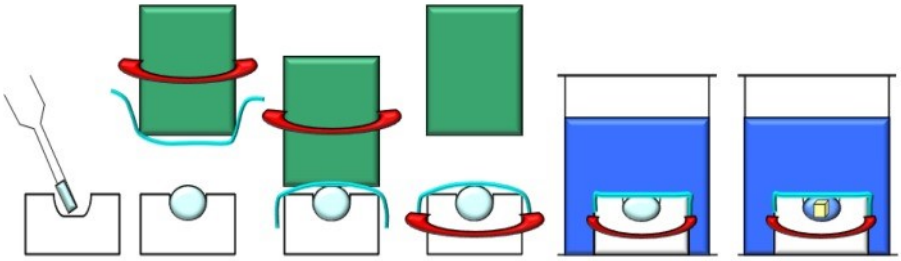


Figure 2.16 Pipetting process for microdialysis crystallisation. The droplet of protein solution is placed in a small button, which is covered with a dialysis membrane and closed with a thin elastic band. The button is immersed in the precipitant solution and the equilibration of precipitant molecules occurs through the membrane.[105]

Another macromolecule crystallisation method is free-interface diffusion (figure 2.17). The process of this method is putting the protein and precipitant solution together but without mixing, and then sealling them using wax in a capillary. Then, the system slowly reaches equilibrium with the slowly diffusion of the protein and the precipitant. This process gives the protein solution a supersaturation so that the crystals can come out and grow.[105]

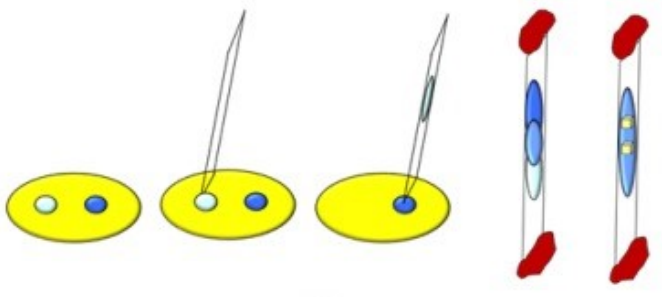


Figure 2.17 Free-interface diffusion method [105]

2.5.3 Current crystallisation methods for peptides

Although peptides are difficult to crystallise, some of the new methods have been reported to crystallise peptides.

A. Racemic peptide crystallisation

Structural biologists use racemic crystallography to generate protein crystals by mixing an equimolar quantity of naturally chiral L-protein molecule with its D-protein mirror copy. L-protein molecules contain 'left-handed' L-amino acids and glycine, while D-protein molecules are composed of Right-handed D-amino acids and glycine. In 1993, Zawadzke and Berg crystallised a centrosymmetric crystal of rubredoxin, which is made up of 45 amino acids. The crystal grown from the racemic mixture of rubredoxin. The unit cell of the crystal structure contains two molecules – the enantiomers.[106] In 1995, Wukovitz and Yeates predicted that proteins would crystallise much easier from racemic mixtures. When using this mixture, an racemic crystal symmetries can be obtained.[107] Mandal and co-workers crystallised BmBKTx1, which contains 31 amino acids using racemic crystallisation in 2009. This peptide has six cysteine residues and three disulphide bonds. The crystals of BmBKTx1 in the tetragonal centrosymmetric space group $I4_1/a$ was obtained finally. The racemic crystallisation process of BmBKTx1 is depicted in figure 2.18. The structure of BmBKTx1.95 is seen in Figure 2.19.[108] Todd O. Yeates and Stephen B.H. Kent tested racemic crystallography on a wide range of

protein molecules in 2012. Their experimental data verified that racemic protein mixtures which contains the enantiomeric forms can crystallise ideal crystals under suitable conditions.[109]

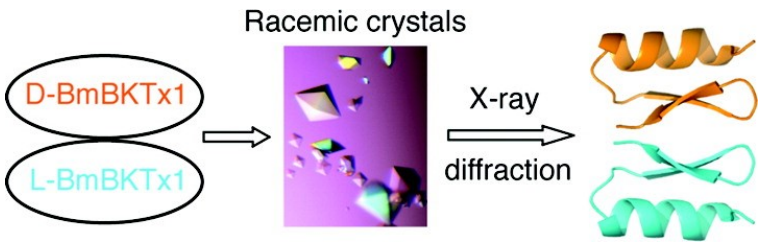


Figure 2.18 Crystallisation of BmBKTx1 obtained by the racemic method [108]

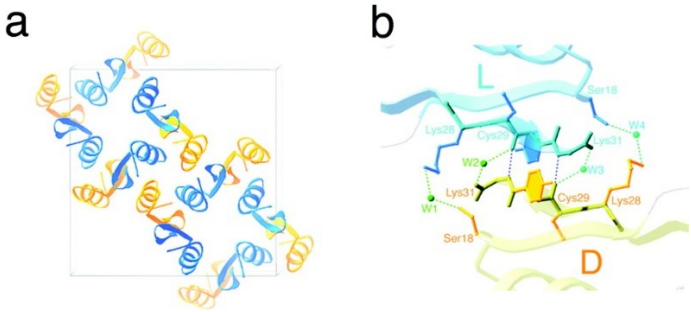


Figure 2.19 (a) Unit cell of BmBKTx1, (b) The interface between L- and D-enantiomers [108]

B. Co-crystallisation method for peptide crystallisation

Co-crystallisation approach gives peptides a new opportunity to crystallise with large protein. Some research focus on the crystal structure of ligand-bound protein complexes, in which the ligand bounded to the active site of the large protein can be peptides. Garai and coworkers crystallised the wild-type protein (ERK2 WT) with six different peptides.[110] In 2008, Christianson crystallised a ^{129}Xe-cryptophane biosensor which complexed with human carbonic anhydrase II (CAII) (figure 2.20).[111, 112] Using this approach to crystallisation, a small peptide could be linked to a small molecule inhibitor of a readily crystallisable enzyme and co-crystallise with enzyme. The disadvantage of this method is that many receptor proteins for

natural peptides are not identified.

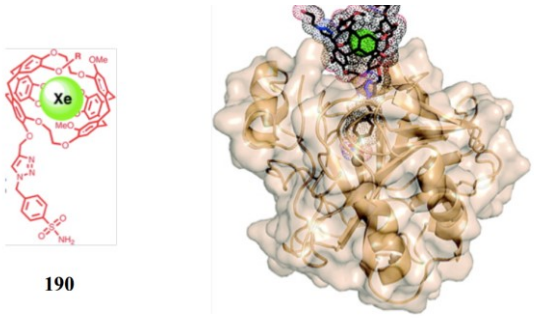


Figure 2.20 ^{129}Xe-cryptophane biosensor (190) complexed with CAII [111]

In order to purify the dipeptide from trifluoroacetate salt which is necessary in the peptide synthesis, Nicholas and coworkers obtained TFA-free crystals of a dipeptide using co crystallisation method with different solid coformers (figure 2.21). Thirty distinct coformers were utilised for a comprehensive screening. These molecules contain a broad variety of structural characteristics (e.g., linear, aromatic, heterocyclic), chemical functions (i.e., hydrogen bond donors and acceptors), and chemical properties (i.e., acids and bases). This work introduces new concepts to peptide crystallisation and proves that crystallisation is an effective method to purify combined materials.[113]

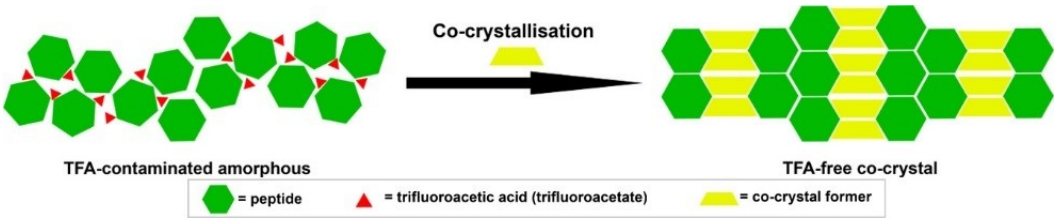


Figure 2.21 Direct purification of L-Leu-L-Leu dipeptide through co-crystallisation[113]

C. New approaches

Besides, numerous approaches are developed for peptide crystallisation currently. For instance, the templates approach, which involves incorporating templates into crystallisation

solutions and allowing crystals to grow on them or accelerating crystal growth with the templates. This method has been applied to a large number of protein crystallisation.[114, 115] For example, Link used amino acids as soft-templates to modulate nucleation of insulin to enhance the crystallisation rate.[116] Chen demonstrated that mesoporous silica with a high protein loading is a suitable nucleant for lysozyme and thaumatin due to its favourable interaction with these proteins during batch crystallisation.[117]

Another approach is to crystallise peptides under special environments, like electric field, magnetic field or ultrasonic sounds. Richard Lakerveld and coworkers found optimal electric-field conditions (the nonuniform electric field of 10 Vpp and 1 MHz induced by a specific pattern) can enhance the crystallisation of lysozyme and insulin.[118] Wakayama reported the protein crystal quality can be increased under intense magnetic fields. The research includes magnetic orientation of protein crystals, magnetic regulation of effective gravity on earth, creation of protein crystals in diverse effective gravity, magnetic improvement in crystal quality, magnetic increase in viscosity of protein aqueous solutions, and Lorentz force dampening of natural convection. [119] Ultrasonic sounds has been found can control the crystal qualities, such as particle size and particle size distribution and it can also promote the liquid-phase epitaxy growth.[120, 121]

Chapter 3. Methodology

3.1 Crystal structure analysis

3.1.1 Single crystal structure analysis

In determining structure of molecules, there are numerous powerful techniques such as X-ray crystallography, EM crystallography, neutron diffraction and NMR. Both X-ray crystallography and NMR are excellent and well-established complementary techniques. Adelinda et al. suggested that NMR spectrum and optimized crystallisation can be used in parallel for the screening of all small proteins' structure.[122] However, the success rates of this is still relatively low with only 21 targets (8%) where both NMR and X-ray crystallography methods were successful in the determination of the molecular structure. It has also been suggested that NMR may not also reveal accurate or exact conformation due to complications of the technique. In instance, the assumption that all NMR signals originate from the same main conformer may not hold true for flexible molecules such as peptides. In such a circumstance, traditional NMR structure refinement approaches are insufficient for reliable experimental data interpretation.[123] Currently, ~90% of the total protein structures deposited in the PDB (Protein Data Bank https://www.rcsb.org/) are resolved by X-ray crystallography. It is evident that X-ray crystallography is advantageous in being able to provide a full atomic position for the whole macromolecular structure.

The Single Crystal XRD (SCXRD) data obtained in this thesis was collected on a Rigaku Saturn 70 abstract diffractometer with an Mo K\ α radiation source (λ = 0.71073 Å). Using Olex2[122], the structure was solved with the ShelXT[125] structure solution program using Intrinsic Phasing and refined with the ShelXL refinement package[125] using Least Squares minimisation.

3.1.2 Powder X-ray diffraction (PXRD)

Powder X-ray diffraction (PXRD) is normally used to obtain the diffraction pattern for powder samples to get their crystal structure information. The advantage of PXRD is that powder samples (i.e. no large single crystal is required) makes it become more convenient and easier to distinguish the different morphology and solvates. Figure 3.1 shows an example of the Bruker D8 Advance PXRD equipment.[126] The transition between polymorphs under different temperatures and humidity can also be examined by PXRD with the variable temperature stages and humidity accessory to observe the diffraction pattern change under different environment.[127] The XRD patterns are also used to study the crystalline content, calculate the purity, identify the crystalline phases and the spacing between the lattices.[126]

Figure 3.1 A PXRD (Bruker D8 Advance) showing the goniometer and its drives inside the cabinet.[126]

3.2 Thermal analysis

3.2.1 Differential scanning calorimetry (DSC)

DSC is a powerful technique to measure the thermodynamic properties of samples, such as melting point, fusion enthalpy, and decomposed points.[128] Figure 3.2 shows the basic components of a DSC equipment.[129] With the increasing temperature, the heat flow detected by the sensor can be used to monitor the morphology change and the phase transition for small molecules and also peptides and proteins.[128] However, for peptides, the melting process is accompanied by decomposition.

In this thesis, DSC of samples were measured with a Discovery DSC Q2000 (TA Instruments, New Castle, DE, USA) from 298.15 K to 850 K at a scan rate of 10 $K \cdot min^{-1}$. The measurements were carried out using 5.00 ± 0.10 mg samples under a helium atmosphere (50 $mL \cdot min^{-1}$).

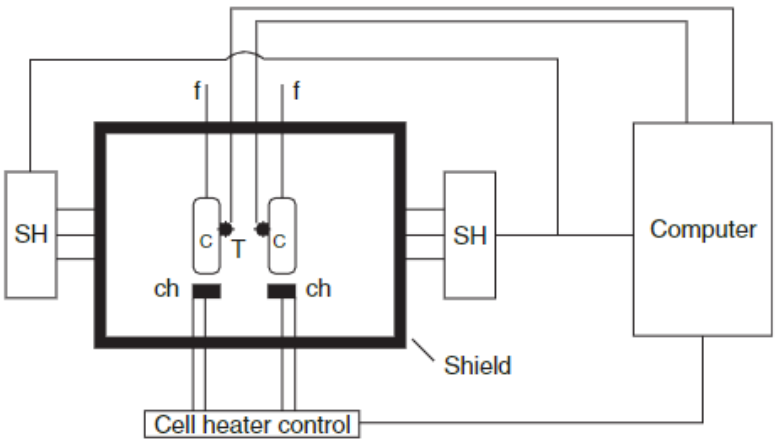


Figure 3.2 General diagram for a differential scanning calorimeter. Cells (c) are located within a shield, which is in contact with shield heaters (SH). Individual cell heaters (ch) control the temperature of the sample and reference cells. Temperature sensors (T) are located on the cell surfaces, which determine if there is a temperature difference between the two cells, and through computer control apply appropriate compensating heat to the cells to keep temperature difference near zero. The compensating energy per unit time is recorded as the calorimetric signal.[129]

3.2.2 Thermal gravimetric analysis (TGA)

Thermal gravimetric analysis (TGA) can monitor the weight loss of the samples with temperature to characterize the phase change during the heating process.[131] The content of the solvents in the sample can be calculated based on the weight loss ratio. Figure 3.5 shows the different loading directions in the equipment of TGA.[132] DSC, gas chromatography, mass spectrometry can be coupled with TGA together to measure the different physical and chemical properties.

In this thesis, the weight loss of samples was determined by thermogravimetric analysis using a TGA Q500 (TA Instruments, New Castle, DE, USA). The TGA measurements were performed at a heating rate of 10 K/min. A constant nitrogen flow of 60 mL/min was used to prevent thermal oxidation processes.

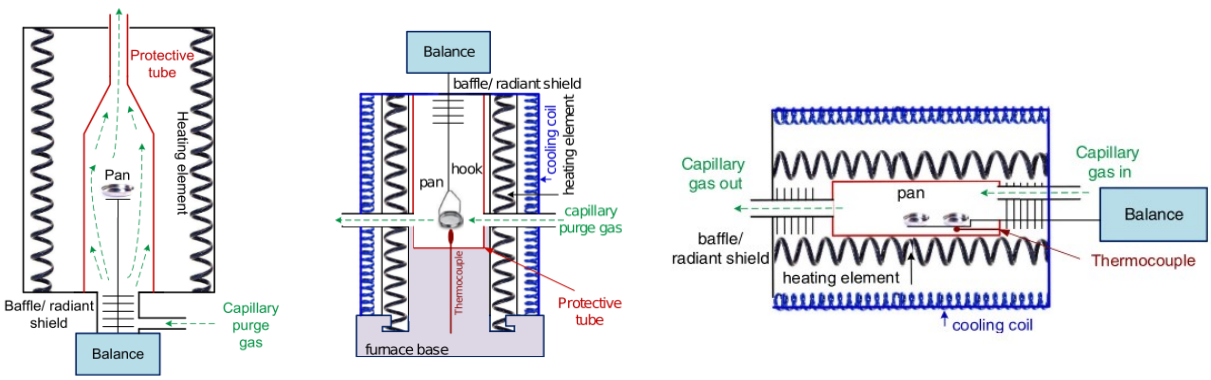


Figure 3.3 TGA equipment with different loading position. Left: top loading, middle: bottom loading, right: side loading.[132]

3.3 Spectrum analysis

3.3.1 UV-Visible spectroscopy

The UV-visible light (200-900 nm) can be absorbed by samples in their specific

wavelength, which is commonly used in the concentration measurement during the crystallisation. A plot of the linear relationship between concentration and light absorbance can be obtained based on the Beer–Lambert law, the equation is represented as following:

$$C = \frac{A}{\varepsilon \times b} \tag{3.1}$$

Where C is the concentration in molarity, A represents the UV absorbance in absorbance units, ε is the extinction coefficient, b means the pathlength in cm.

This law is only relevant at low concentrations, which means that the solution must be diluted prior to measurement. In order to measure the concentration of unknown materials, the first step is to prepare solutions with varying concentrations and dilute them to extremely low concentrations to measure their absorbance A. After that, a standard curve can be established, which is a straight line equation that links concentration and absorbance. The parameter A and the extinction coefficient ε can be determined from this standard curve. Finally, after measuring a new sample, the absorbance value can be used to compute the concentration using the standard curve.

Nanodrop-UV is an approach which is commonly used to measure the ultraviolet absorption property of proteins and other materials. The advantage of this method is the tiny amounts of samples and the short testing time. The equipment and the testing procedure can be seen in figure 3.4. The sample volume is only 2 μL, and the test takes only 5 seconds to complete. It is not suited for solvents with a high evaporation rate, as the concentration of the sample will increase as a result of evaporation.

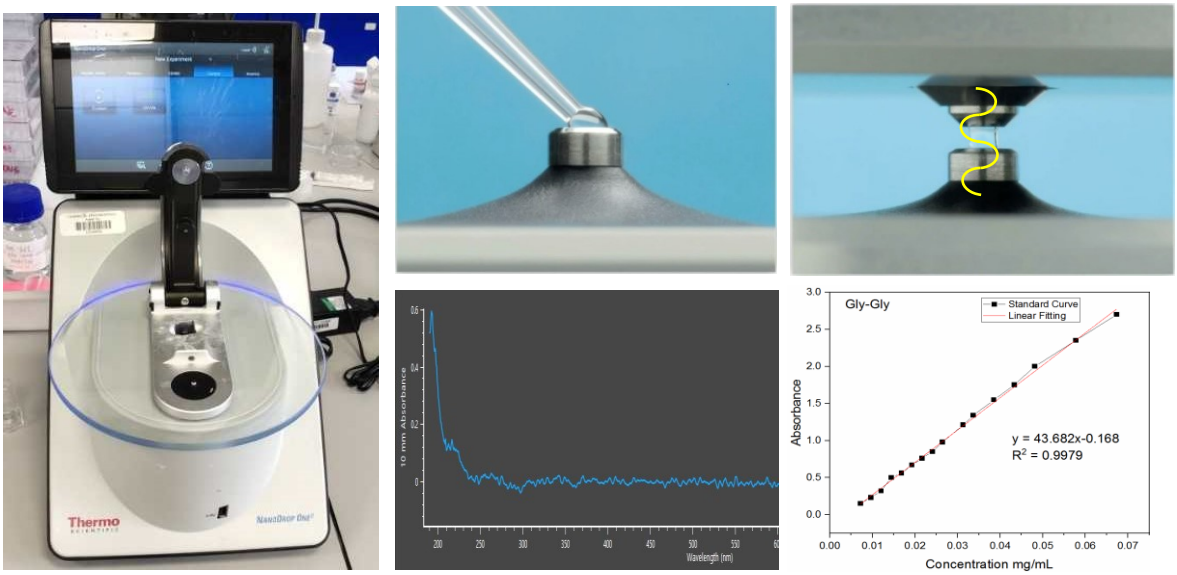


Figure 3.4 The Nanodrop UV-Vis equipment and the testing procedure

3.3.2 Raman spectroscopy

Due to the absence of chiral atoms in glycine, the circular dichroism (CD) technique that is typically employed to determine the secondary structure of peptides and proteins cannot be applied in this thesis. For glycine homopeptides research, Raman and FTIR become the best methods which can monitor the conformation change under different conditions. The theory of Raman spectroscopy is energy exchange between the molecule and the incident electromagnetic field of the scattered light, which results in a difference in the vibrational frequency of the incident and the scattered light.[133] For a vibrational transition to be Raman active, the molecule's polarizability must change during the vibration. Polarizability refers to the ease with which electrons can be shifted from their original position. With increasing electron density, increasing bond strength, and decreasing bond length, a molecule's polarizability decreases. For peptides and proteins, there are particular bands in Raman where the different peptides bonds vibration present different conformation information.[134, 135] There are mainly three bands, which are located at around 1650 cm^{-1} (Amide I, 80% C=O stretching); 1550 cm^{-1} (Amide II, 60% NH bending and 40% C-N stretching), and 1200-1340 cm^{-1} (Amide

II|, 40% C-N stretching and 30% NH bending). For the main secondary structures, there are specific patterns in these bands to show the conformation. Figure 3.5 illustrates how the three major amide bands shift in response to the conformation of glycine homopeptides. [136]

Bykov et al. investigated the conformational preferences of glycine homopeptides in water using Raman, as well as the effect of Li^+ on the conformation change.[136] They later employed Raman to differentiate the contribution of internal and penultimate peptide bonds.[137] In this thesis, solid Raman was used to measure the conformation change of triglycine dihydrate in the crystalline form and in solution.

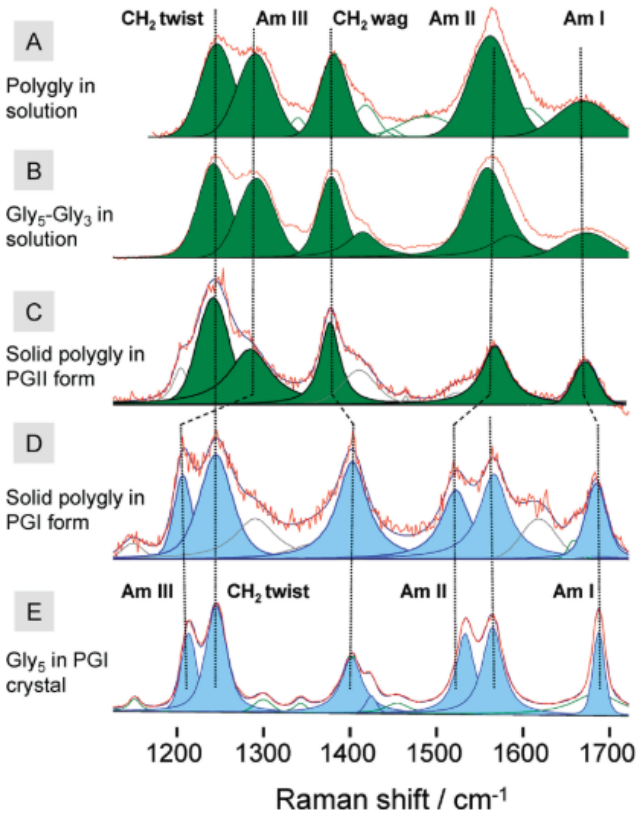


Figure 3.5. Change of the amide I, II, and III bands in Raman for different conformation of glycine homopeptides [137]

3.3.3 Fourier transform infrared (FTIR) spectroscopy

Compared with Raman, the theory of FTIR is the measurement of the infrared light absorbed by the sample, instead of the light scattering. The transmission of infrared light can also be used to describe the absorption of infrared light. The IR absorption of carbonyl group of peptides and protein makes the FTIR as a good technique to monitor the conformation change. The vibrations of different groups correspond to different infrared absorption frequencies. The wavenumber for the specific vibration from different conformation in FTIR is similar with those in Raman.[138] However, FTIR is inactive to symmetric group vibrations. To be IR active, a molecule's dipole moment must change as a result of the vibration caused by IR light absorption. Thus, Raman and FTIR are always combined together to research the conformation in solution and structure information for peptides.[139-141]

3.4 Molecular dynamic (MD) modeling

Molecular dynamic simulation is a common modeling method to research the dynamic change of energy and structure of the samples in atomic level and explore the pathways to the secondary structures of peptides.[142-144] However, the result of MD depends on the force fields strongly. There are several commonly used force fields, which are AMBER, CHARMM, GROMOS, and OPLS Force Fields.[145, 146] The suitable force fields of molecular dynamic simulation were chosen to estimate the real conformation of peptides and proteins in solution based on the experimental data obtained by Raman, FTIR, CD and NMR.[147]

In this thesis, MD simulation was used to calculate the solvation free energy of glycine homopeptides and dipeptides to explore the effect on the solubility, and also applied to calculate the conformation of triglycine in solution to explore the transition between folded and unfolded secondary structure under different temperatures.

Chapter 4. The effect of chain length and side chains on the solubility of peptides in water from 278.15K to 313.15K: A case study in glycine homopeptides and dipeptides

4.1 Abstract

The thermodynamic properties of peptides are significant in terms of the crystallisation conditions of biomaterials. In this chapter, I seek to understand and explain the effect of side chains and chain length on the solubility of peptides. The amino acid residues of dipeptides investigated here were chosen based on their difference in the side chain properties. The modified Apelblat equation was used to correlate the relationship between the solubility in water and temperature. In order to explore solute-solvent interactions, the solvation free energies of these peptides were calculated by Molecular Dynamic simulations. This work gives an indication of the effects of side chains and chain length on the solubility of amino acids and peptides in water under different temperatures, which not only provides the thermodynamic data for peptides, but is also critical in the prediction of peptide solubility using Statistical Associating Fluid Theory (SAFT).

4.2 Introduction

Peptides are short chain biomolecules containing 50 or fewer amino acids, which are known to have important therapeutic properties and crucial potential applications in the treatment of chronic diseases.[12, 13, 148-154] Peptide crystallisation, as a promising alternative to chromatographic purification, can circumvent the shortcomings of the traditional purification methods.[155] Accurate solubility data is a crucial thermodynamic property in peptide crystallisation, as it corresponds to the stable solid-liquid phase boundary in the phase diagram, which is needed to characterize supersaturation and calculate nucleation parameters in the prediction and design of crystallisation conditions.[61, 62, 68, 156] Currently, only limited peptide solubility data in water is available in literature.

Compared with small molecules, the solubility measurement of longer chain peptides is challenging due to their flexible structure, especially in the case of peptides composed of more than five amino acid residues. Glycine is the simplest amino acid, of which the non-heavy atom side chain makes the molecule an ideal candidate to study the chain length effect on the peptide solubility, without the effect of side chains. The effect of chain length on the thermodynamic properties of glycine homopeptides was studied to establish a rational design of the conditions for the homopeptide crystallisation. Apart from glycine, side chains of different amino acid residues in the peptides determine the final thermodynamic properties. In the standard genetic code, there are 20 different amino acids which can be divided into four categories based on the chemical properties of their side chain: polar, nonpolar, acidic, and basic side chains.[155] In this work, four amino acids were chosen from each category to form eight dipeptides to research the effect of side chains on peptides solubility. The solubilities of glycine homopeptides (glycine, diglycine, triglycine, tetraglycine, pentaglycine and hexaglycine), four amino acids with different side chains (aspartic acid, phenylalanine, histidine and tyrosine) and their dipeptides (asp-phe, gly-asp, gly-phe, phe-phe, gly-gly, tyr-phe, gly-tyr, gly-his) in water from

278.15K to 313.15K were measured using UV-Vis spectrophotometry.

Several classical thermodynamic equations are commonly applied to correlate the solubility of small molecules in a given solvent system.[61, 156, 158] In this chapter, the modified Apelblat equation (Equation 2) was used to correlate the relationship between solubility in pure solvent and temperatures.[70, 159, 160] The solvation free energies of peptides were calculated using molecular dynamic simulation to investigate the solute-solvent interactions during the dissolution process.[161]

4.3 Materials and Methods

4.3.1 Materials

Glycine homopeptides (mono-(α form), di-(α form), tri-(β form), tetra-, penta-, hexa-glycine), four other amino acids (aspartic acid, phenylalanine, tyrosine, histidine) and dipeptides (gly-gly, gly-phe, phe-phe, tyr-phe, gly-tyr, asp-phe, gly-asp, gly-his) were purchased from Sigma-Aldrich Company Ltd. (Figure 4.1). Deionized water was produced in the laboratory. All chemicals were used without further purification. The detailed information about the chemicals used in this work has been listed in Table 4.1.

Table 4.1 Properties of materials used in this chapter

Material	Molar Mass ($g \cdot mol^{-1}$)	CAS registry no.	Mass Fraction Purity	Source
Glycine	75.07	56-40-6	≥0.990	Sigma-Aldrich Company Ltd.
Diglycine	132.12	556-50-3	≥0.990	Sigma-Aldrich Company Ltd.
Triglycine	189.17	556-33-2	≥0.990	Sigma-Aldrich Company Ltd.
Tetraglycine	246.22	637-84-3	≥0.980	Sigma-Aldrich Company Ltd.
Pentaglycine	303.27	7093-67-6	≥0.980	Sigma-Aldrich Company Ltd.
Hexaglycine	360.32	3887-13-6	≥0.980	Sigma-Aldrich Company Ltd.
Aspartic acid	133.10	56-84-8	≥0.980	Sigma-Aldrich Company Ltd.
Phenylalanine	165.19	63-91-2	≥0.990	Sigma-Aldrich Company Ltd.
Histidine	155.15	71-00-1	≥0.990	Sigma-Aldrich Company Ltd.
Tyrosine	181.19	60-18-4	≥0.990	Sigma-Aldrich Company Ltd.
Asp-Phe	280.28	13433-09-5	≥0.970	Sigma-Aldrich Company Ltd.
Gly-Phe	222.24	3321-03-7	≥0.970	Sigma-Aldrich Company Ltd.
Gly-Asp	190.15	4685-12-5	≥0.990	Sigma-Aldrich Company Ltd.
Gly-Tyr	238.24	658-79-7	≥0.980	Sigma-Aldrich Company Ltd.
Gly-His	212.21	2489-13-6	≥0.980	Sigma-Aldrich Company Ltd.
Tyr-Phe	328.36	17355-11-2	≥0.980	Sigma-Aldrich Company Ltd.
Phe-Phe	312.36	2577-40-4	≥0.980	Sigma-Aldrich Company Ltd.

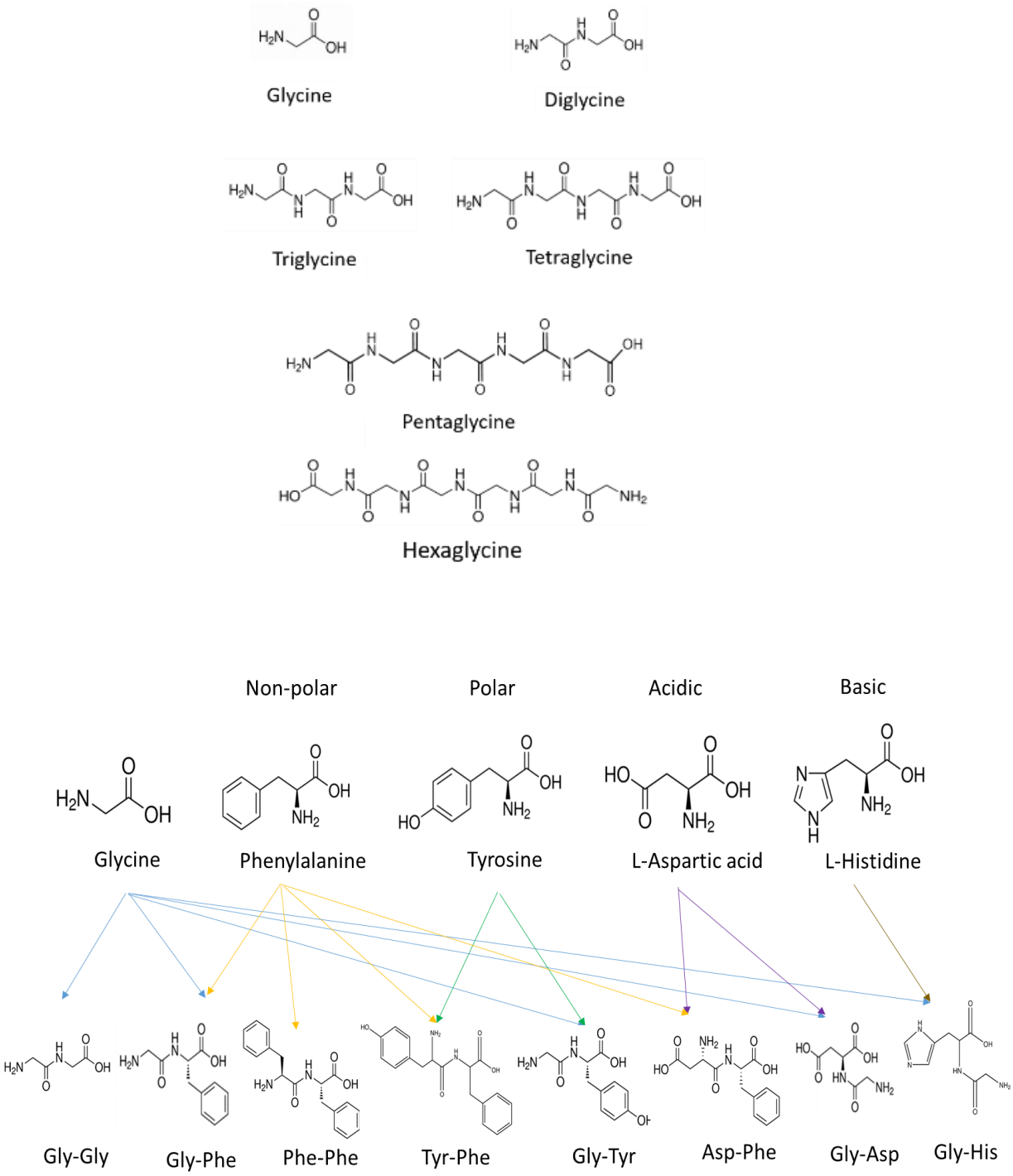


Figure 4.1 Chemical structures of the studied amino acids and peptides

4.3.2 Powder X-ray diffraction analysis

Powder X-ray diffraction (PXRD) patterns of amino acids and peptides during solubility measurement were collected by a PANalytical X'Pert PRO X-ray diffractometer. All of the samples were ground to fine powders before being placed flush against the sample container with a glass slide, so that the surface of the sample will be flat. The samples were carefully placed in the proper XRD slot after extra powder was removed from the sample holder's edges. The X-ray diffraction measurement was performed using Cu Kα radiation (1.5405 Å) at 40 kV and 40 mA. All samples were conducted at a scanning rate of 1 step/s over a diffraction angle range from 2 to 50°.

4.3.3 Measurement of standard curves

The NanoDrop Micro-UV/Vis Spectrophotometer (NanoDrop One C, Thermo Fisher Scientific) was used to measure the solubility of all amino acids and peptides in water from 278.15K to 313.15K. The standard calibration curves were obtained by measuring a series of known concentration solutions (the details can be seen in Table 4.3), which were diluted when necessary until they followed the Beer–Lambert law.[160] Spectra of the prepared samples were acquired with multiple scans (3–5) from 190 to 280 nm to find the characteristic wavelength and the concentration scope where the absorbance is linearly associated with concentration.

The linear equation of the standard curve is as follows:

$$Y = a\,x + C \tag{4.1}$$

where Y is the UV absorbance at the characteristic wavelength, and x represents the concentration of the peptide in mg (peptide)/mL (water).

4.3.4 Measurement of solubility

4.3.4.1 UV–Vis spectrometer method

The solubility data of tested amino acids and peptides were measured by the UV–Vis spectrometer, which is commonly used to measure the concentration of solution, especially when the concentration is low[163] (Figure 4.2a). 10 mL glass tubes were filled with deionized water, into which excess amounts of amino acids and peptides were added. These tubes were placed in a 1 L cylindrical double-jacketed glass vessel connected to a water bath (GP200, Grant Instruments (Cambridge) Ltd.). The temperature of the vessel can be controlled within ±0.1 K. The solid–liquid mixtures were continuously stirred by magnetic stir bar for 48 h until the concentration of the solution measured by UV-Vis were no longer changed afterwards, which was deemed long enough to reach solid-liquid phase equilibrium. Then the solution was kept in a quiescent state without stirring for 12 hours until all the undissolved particles had settled to the bottom of the vials. The supernatant was the saturated solution under this specific temperature, which was filtered with a filter membrane (0.22μm). Then the filtered solution was diluted to the specific concentration scope with deionized water according to the standard curve. The diluted samples were taken to measure the concentration by NanoDrop Micro-UV/Vis Spectrophotometers. Each sample ran in triplicate and each measurement was repeated at least three times. The average value was used to calculate the mole fraction solubility (x) of peptides in water by their standard calibration curves (Equation (4.1)), respectively.

4.3.4.2 Dynamic method

The dynamic method has been used to explore the solubility of metastable crystal forms, which would transform to the most stable form during the long stirring times required by the UV-Vis method (figure 4.2b). 2g water was added into an empty tube and stirred until the temperature reached the set point, which was controlled by the same water bath described in Section 2.4.1. Small amounts (10% of solubility) of peptides were repeatedly added to the tube and stirred to dissolve until solution started to become cloudy. Since the saturation point was exceeded easily at this point, more water was added again to make the solution became clearer.

However, if the solution became absolutely clear, more peptides were added again until there were only a few particles existed on the bottom which cannot be dissolved anymore. By continuing to add peptide and water in this cyclical fashion until the calculated change in concentration was less than 1%, the solutions were therefore considered saturated. The whole cycling time cannot be longer than the transition time in which the metastable form can transfer to the most stable form. The different crystal habits for the anhydrates (plate) and dihydrates (needle) can help to ensure whether the metastable form was converted to the stable form during measurement.

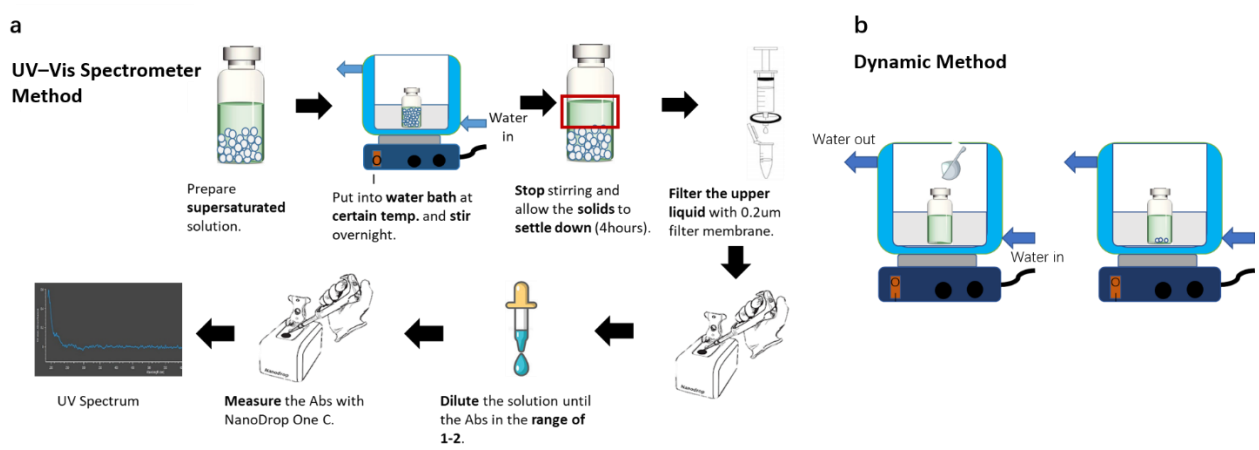


Figure 4.2 The process of solubility measurement. a. UV-Vis spectrometer methhod; b. Dynamic method.

4.3.5 Thermodynamic equations

Many thermodynamics equations can be used to correlate solubility with temperature. The melting point and fusion enthalpy for peptides are very difficult to obtain via DSC, due to the decompositon of samples during the melting process. As a result, the only thermodynamic equation which can be used is the following modified Apelblat equation to correlate solubility data in pure solvent.

There are three empirical constants in this equation, A and B reflect the non-ideality of the

real solution, C represents the effect of temperature on fusion enthalpy:

$$\ln x_1 = A + \frac{B}{T} + C\ln T \tag{4.2}$$

where x_1 and T represent the mole fraction solubility of peptide and the experimental temperature, respectively.

The average relative deviation percentage (*ARD*) is an index which indicates the differences between experimental and calculated data and also can be used in this work to test whether the thermodynamic model was adequately fitted to the experimental results. The equation of *ARD* is as following:

$$ARD\% = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{x_{1,i} - x_{1,i}^{cal}}{x_{1,i}}\right| \tag{4.3}$$

where $x_{1,i}$ and $x_{1,i}^{cal}$ represent the experimental and the calculated solubility, respectively. N is the number of the experimental points.

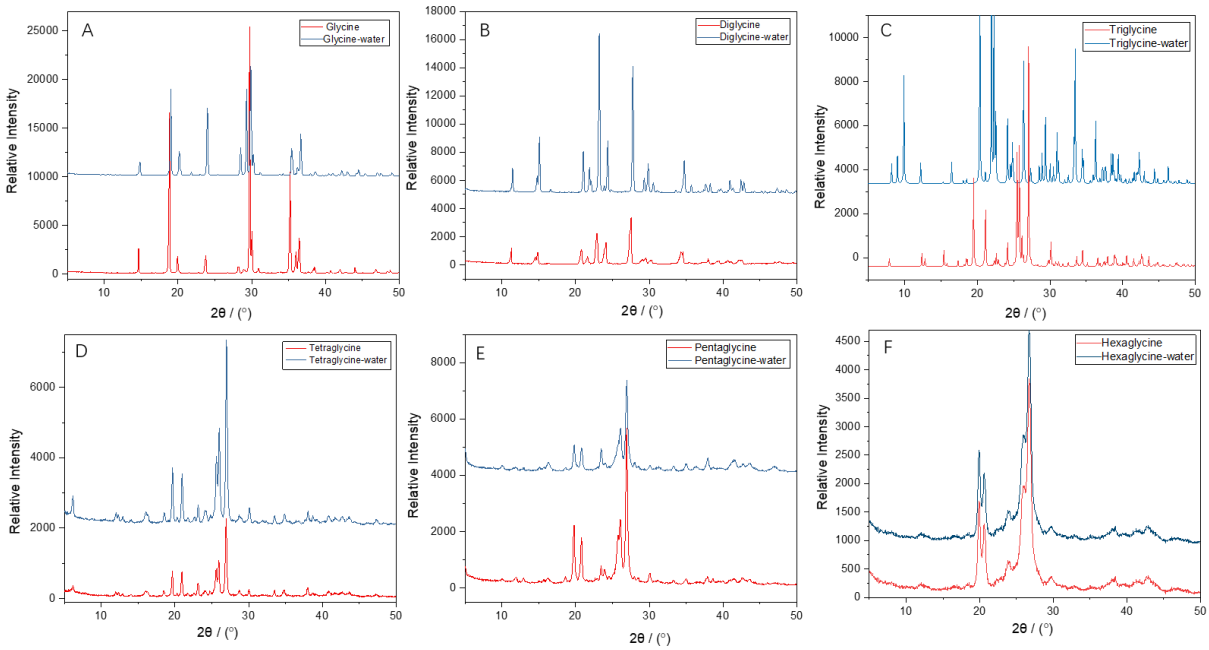
4.4 Results and Discussion

4.4.1 Material characterisation

The PXRD patterns of amino acids and peptides used in this work can be seen in Figure 4.3. All samples were tested before and after the solubility measurement process to ensure that there was no phase transition. For glycine, diglycine and triglycine, there are different morphologies, such as α, β and γ. Here, α form of glycine and diglycine, β form of triglycine were used, separately. As shown in Figure 4.3, all the samples kept the same XRD pattern, except triglycine, gly-asp and gly-tyr.

From our previous research on triglycine, it is known that triglycine can form a dihydrate at 298.15 K.[164] TGA and DSC (Figure 4.4) of new phases of gly-asp and gly-tyr were measured to see whether they can also form hydrates. From Figure 4.4, there are two steps for both gly-

try and gly-asp, which verifies the existence of hydrate forms. Since UV-Vis can only measure the most stable solubility data in water at specific temperatures, the dynamic method was used to measure the metastable form for these three peptides.



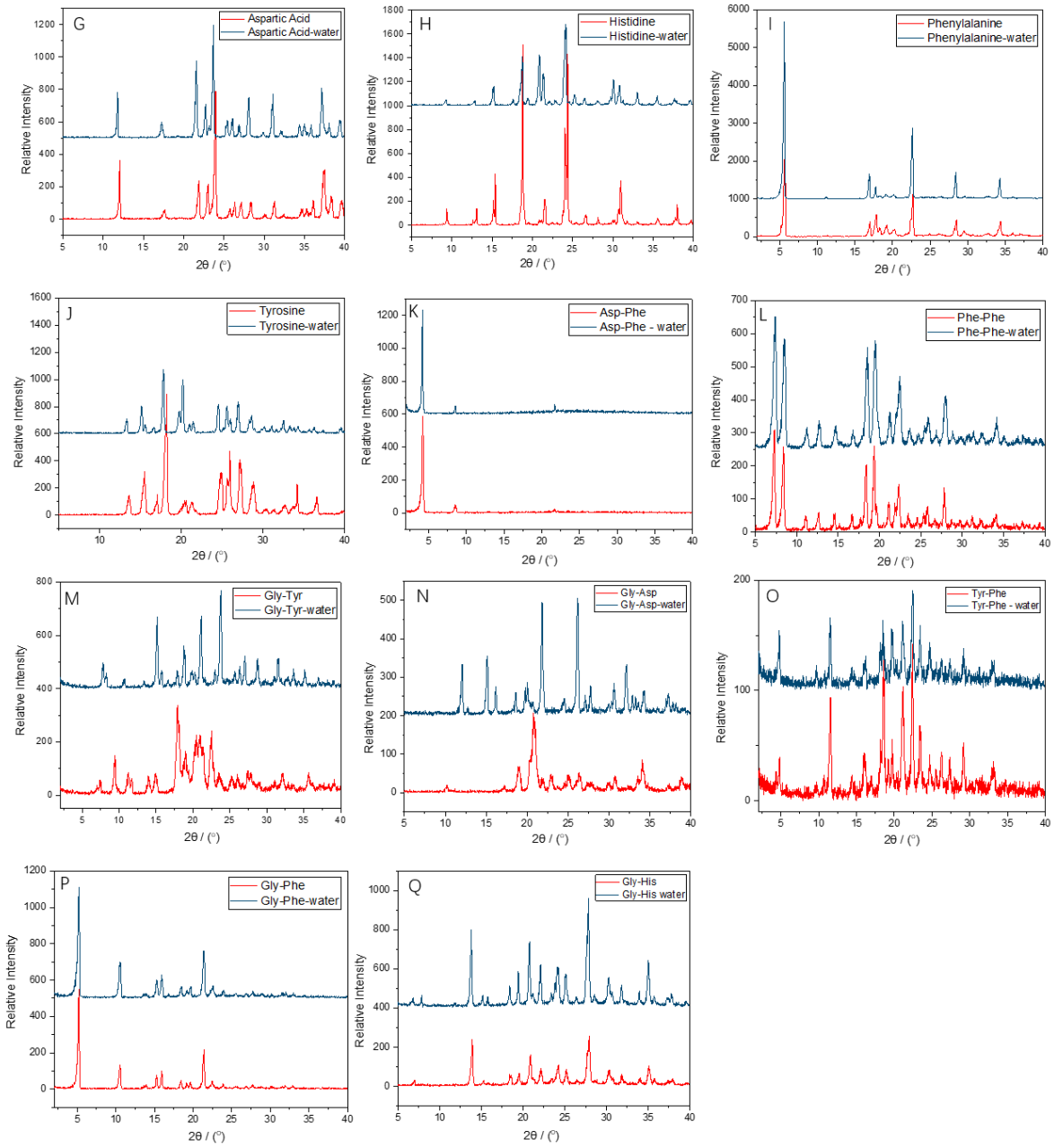


Figure 4.3 XRD patterns of amino acids and peptides before and after solubility measurement: A.Glycine (α form), B.Diglycine (α form), C.Triglycine (β form), D.Tetraglycine, E.Pentaglycine, F.Hexaglycine, G.Aspartic acid, H.Histidine, I. Phenylalanine, J.Tyrosine, K.Asp-Phe, L.Phe-Phe, M.Gly-Tyr, N.Gly-Asp, O.Tyr-Phe, P.Gly-Phe, Q.Gly-His

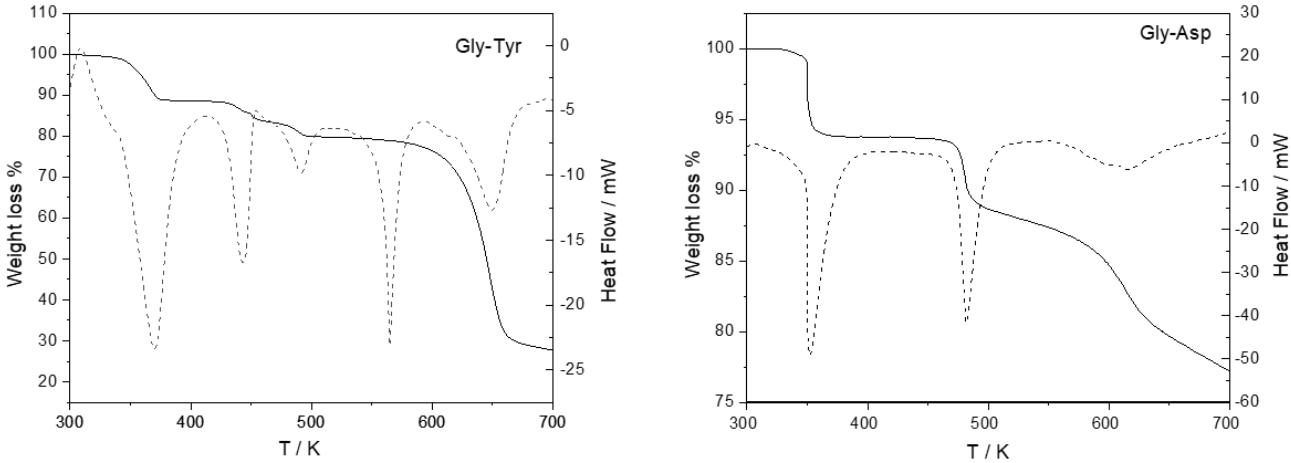


Figure 4.4 The TGA (solid line) and DSC (dash line) curves of Gly-Tyr (left) and Gly-Asp (right)

4.4.2 The standard curves of the amino acids and peptides

The maximum absorption wavelengths and equations of standard curves of amino acids and peptides can be seen in Figure 4.5 (Appendix) and Table 4.2. Each calibration point and the standard deviation are shown in Table 4.3 (Appendix).

4.4.3 Results and discussion of peptides hydrate

For triglycine, gly-asp and gly-tyr, the difference between anhydrate and hydrate crystalline forms makes the solubility measurement process more difficult. In section 4.3.4, the solubility measured using UV-Vis is only for the most stable form, however, the most stable form will be changed under different temperatures (for enantiotropic systems). In order to get solubility data for the metastable forms, the dynamic method was used over the whole range of temperatures. Tables 4.4-4.6 and Figure 4.6 show the experimental data and the plots of solubility trends for all three peptides.

Table 4.2 Specific wavelengths and equations of the standard curves of amino acids and peptides

Materials	λ_{max}	Standard curve	R^2
Glycine	190	A=4.3081C+0.2242	0.993
Diglycine	192	A= 43.682C-0.168	0.998
Triglycine	192	A = 45.438C+0.4279	0.998
Tetraglycine	192	A =57.265 C+0.4549	0.994
Pentaglycine	190	A =76.085 C+0.0592	0.997
Hexaglycine	192	A= 46.602C+0.1486	0.9977
Aspartic acid	190	A=7.667C-0.1993	0.999
Phenylalanine	260	A =1.1291C-0.3303	0.999
Histidine	207	A =34.293C-0.37466	0.999
Tyrosine	193	A =142.192C+0.23489	0.995
Asp-Phe	260	A =0.65877C+0.0493	0.998
Gly-Phe	190	A =106.38C+0.4027	0.995
Gly-Asp	190	A=37.252C+0.1386	0.995
Gly-Tyr	193	A=194.35C+0.0676	0.998
Gly-His	193	A=34.6478C+0.1713	0.988
Tyr-Phe	193	A = 123.493C+0.0656	0.995
Phe-Phe	190	A =250.75C-0.0902	0.996

From the results, it can be seen that for all three peptides, the solubility of the anhydrous and the hydrated forms intersect, signifying a critical temperature for anhydrate stability. For triglycine, when the temperature is lower than 303 K, the dihydrate form is the most stable form, while when the temperature is higher than 303 K, only the anhydrate can be obtained in water. The results also shown the comparison between experimental data measured by dynamic and UV–Vis method. It can be seen from figure 4.6 that the results of UV–Vis measurement always get similar values to the solubility of the most stable form under that specific

temperature, which measured by the dynamic method.

Table 4.3 Comparison of the solubility of triglycine anhydrate and dihydrate using the dynamic method and UV-Vis method

T/K	Triglycine Anhydrate (mol/mol)	Standard Deviation	Triglycine Dihydrate (mol/mol)	Standard Deviation	UV-Vis (mol/mol)	Standard Deviation
278.15	0.004485013	5.688E-5	0.002447487	1.432E-5	0.002647487	2.543E-5
283.15	0.004788502	3.213E-5	0.003026769	3.526E-5	0.003026769	1.654E-5
288.15	0.005152794	2.061E-5	0.003621552	3.281E-5	0.003521552	1.721E-5
293.15	0.005674934	5.724E-5	0.004279686	6.003E-5	0.004379686	1.690E-5
298.15	0.006195222	1.623E-5	0.005254334	5.687E-5	0.005354334	4.326E-5
303.15	0.006855379	1.523E-5	0.006701441	7.438E-5	0.006617318	3.598E-5
308.15	0.008088579	2.132E-5	0.008793356	4.621E-5	0.007983911	1.291E-5
313.15	0.009173054	1.873E-5	0.011191544	3.542E-5	0.009173054	2.540E-5

Table 4.4 Comparison of the solubility of Gly-Tyr anhydrate and dihydrate using the dynamic method and UV-Vis method.

T/K	Gly-Tyr Anhydrate(mol/mol)	Standard Deviation	Gly-Tyr Dihydrate(mol/mol)	Standard Deviation	UV-Vis(mol/mol)	Standard Deviation
278.15	0.001832398	2.654E-5	0.001287580	1.548E-5	0.001377219	2.368E-5
283.15	0.001964406	3.425E-5	0.001579693	1.632E-5	0.001608128	1.578E-5
288.15	0.002183512	1.679E-5	0.001816501	1.677E-5	0.001782055	3.254E-5
293.15	0.002388188	2.164E-5	0.002112099	2.432E-5	0.002119995	1.640E-5
298.15	0.002824558	1.561E-5	0.002594589	5.543E-5	0.002544049	2.398E-5
303.15	0.003406030	1.901E-5	0.003406505	4.678E-5	0.003512000	2.020E-5
308.15	0.004340120	2.210E-5	0.004703471	1.708E-5	0.004342000	4.392E-5
313.15	0.005293637	2.433E-5	0.005907235	1.321E-5	0.005218000	3.125E-5

Table 4.5 Comparison of the solubility of Gly-Tyr anhydrate and dihydrate using the dynamic method and UV-Vis method.

T/K	Gly-Asp Anhydrate (mol/mol)	Standard Deviation	Gly-Asp Dihydrate (mol/mol)	Standard Deviation	UV-Vis (mol/mol)	Standard Deviation
278.15	0.002850184	2.241E-5	0.001571146	1.211E-5	0.001775700	2.736E-5
283.15	0.00312385	1.675E-5	0.001830643	1.453E-5	0.001920555	1.688E-5
288.15	0.003468704	1.651E-5	0.002308203	2.698E-5	0.002108428	3.890E-5
293.15	0.003786484	1.551E-5	0.002706168	4.431E-5	0.002553973	2.031E-5
298.15	0.004164186	3.680E-5	0.003353977	5.698E-5	0.003194400	3.324E-5
303.15	0.004981260	2.700E-5	0.004749883	3.612E-5	0.004399777	1.579E-5
308.15	0.005909760	1.702E-5	0.006712112	1.002E-5	0.005632038	1.628E-5
313.15	0.006903708	2.092E-5	0.008829273	1.543E-5	0.006848977	2.397E-5

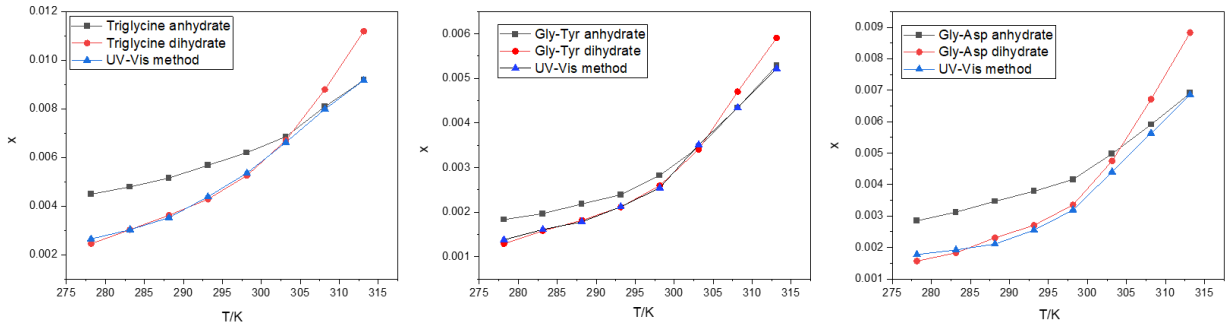


Figure 4.6 The solubilities of peptide anhydrate and hydrate. (black-anhydrate, red-hydrate, blue-UV-Vis measurement)

4.4.4 Effect of chain length on the solubility of glycine homopeptides

The experimental solubility of glycine homopeptides in water from 278.15K to 313.15K is shown in Table 4.7 (Appendix) and plotted against temperature in Figure 4.7. From these results, the solubility of glycine, diglycine, triglycine, tetraglycine, pentaglycine and hexaglycine monotonically decreased with an increasing number of glycine residues. Among them, pentaglycine and hexaglycine have very low solubilities in water. Moreover, the solubilities of these six glycine homopeptides exhibit a positive temperature dependence in water. The modified Apelblat equation was used to correlate solubility data. The results of this, alongside the ARD, can be seen in Table 4.8. For glycine homopeptides, all ARD values for the correlations are less than 5%, which means this modified Apelblat equation can be used to correlate the solubility of glycine homopeptides under different temperatures.

Even though the mole fraction solubility of diglycine is around half of the solubility of glycine, there is no linear relationship between the solubility and chain length in the overall picture (Figure 4.8). When the chain length is longer than two, the solubility is further reduced by one order of magnitude. Figure 4.8 shows the trend of decrease in solubility with the chain length.

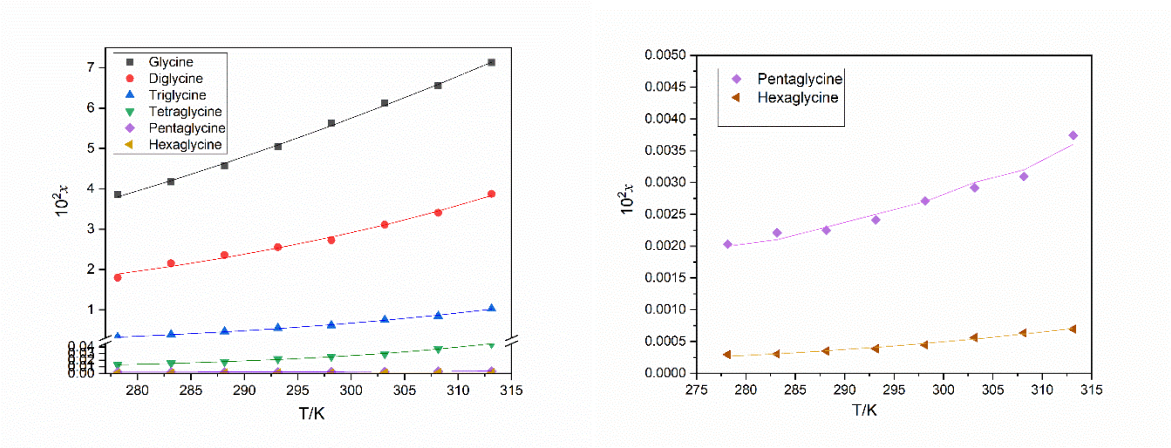


Figure 4.7 The solubility of glycine homopeptides in water from 278.15K to 313.15K

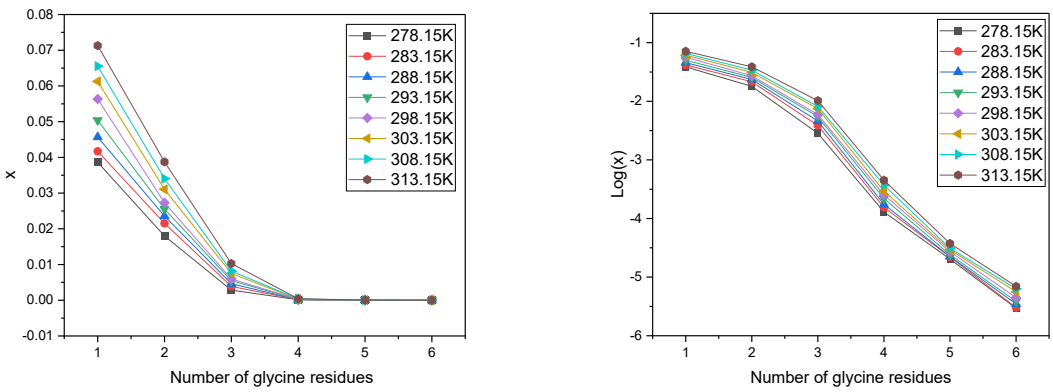


Figure 4.8 The chain length effect on the solubility of glycine homopeptides in water from 278.15K to 293.15K

Table 4.6 Parameters and ARDs of the modified Apelblats equation about the solubility of glycine homopeptides in water (P = 0.1 MPa).

Parameters	Glycine	Di glycine	Tri glycine	Tetra glycine	Penta glycine	Hexa glycine
A	14.4	-118.2	-4190.0	-343.0	-109.9	-66.4
B	-2101.7	3555.4	16737	12136	3109.8	322.6
C	-1.8	18	63	52	15.6	9.3
10^2ARD	0.9809	2.1429	0.8434	1.9268	3.2924	4.0389

4.4.5 Effect of side chains on the solubility of dipeptides

The experimental solubility of four amino acids (aspartic acid, phenylalanine, tyrosine, histidine,) and eight dipeptides (gly-gly, gly-phe, phe-phe, tyr-phe, gly-tyr, asp-phe, gly-asp, gly-his) in water from 278.15K to 313.15K are shown in Table 4.9 (Appendix) and plotted against temperature in Figure 4.9. The solubility of all amino acids and peptides in this study exhibit a temperature dependence in water, meaning that more sample can be dissolved at higher temperatures. The mole fraction solubility order of the amino acids and peptides studied is as following: Gly > Gly-Gly > Gly-His > Histidine > Gly-Asp > Gly-Tyr > Phenylalanine > Gly-Phe > Aspartic Acid > Asp-Phe > Tyrosine > Tyr-Phe > Phe-Phe. Interestingly, tyrosine has a very low solubility despite the presence of a polar side chain. The order is the same throughout the whole temperature range studied, except for gly-tyr which obeys this order when the temperature is from 295K to 303K. The modified Apelblat equation was used to correlate the solubility data, and the parameters and ARD of these correlations can be seen in Table 4.10. From these results, the capability of the modified Apelblat equation to correlate the solubility of amino acids and peptides with the small ARD values was verified.

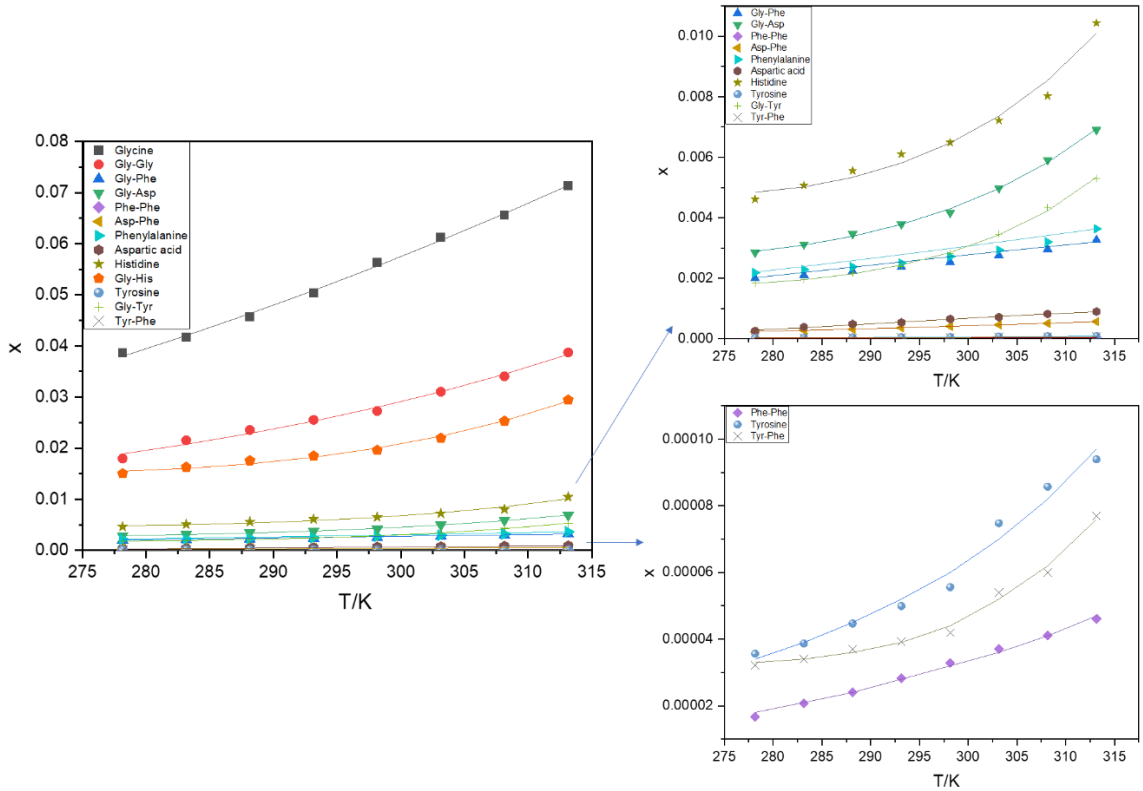


Figure 4.9 Solubilities of amino acids and dipeptides in water from 278.15K to 313.15K. Solid lines represent the calculated solubilities.

Table 4.7 Parameters and ARDs of the modified Apelblats equation about the solubility of amino acids and dipeptides in water (P = 0.1) MPa).

	Aspartic Acid	Phenylalanine	Glycine	Asp-Phe	Gly-Asp	Gly-Phe	Phe-Phe	Histidine	Tyrosine	Gly-Tyr	Tyr-Phe	Gly-His
A	528	-373	14.4	80.3	-503	-212	-70.5	-59.4	-212.3	-777	-826	-491
B	-26009	15135	-2101	-5692	2008	8094	641	2443	6714	31765	34222	21041
C	-79	55	-1.8	-12.1	75	31	10.17	8.9	31.6	117	123	74
10^2ARD	3.8643	0.6644	0.9809	0.789	1.242	0.342	2.598	3.3605	3.8437	1.221	2.759	1.838

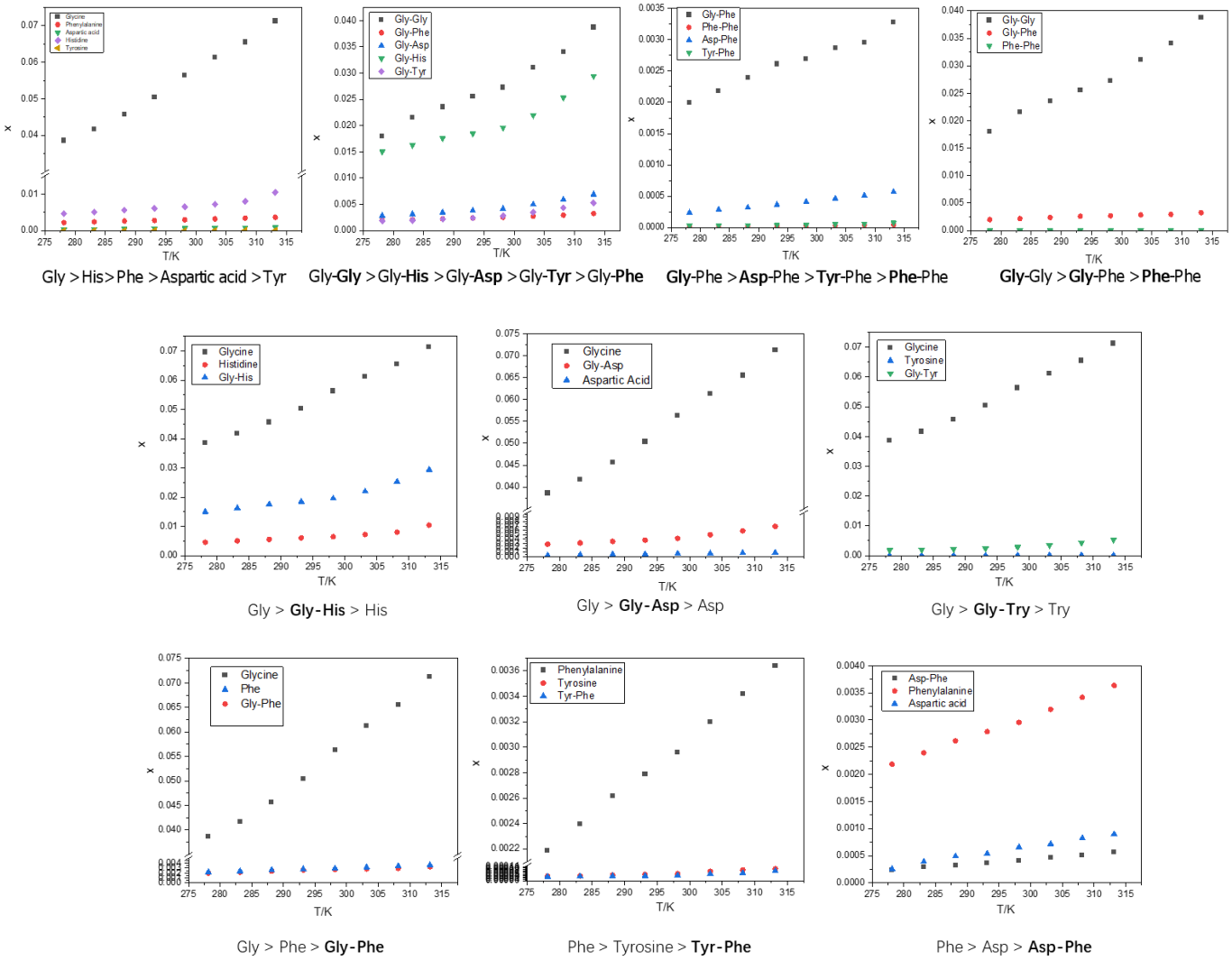


Figure 4.10 The comparison of solubility between different amino acids and peptides

Figure 4.10 compares the solubilities of dipeptides with different side chains. For the four amino acids, the solubility order of phenylalanine (non-polar), tyrosine (polar), aspartic acid (acidic), histidine (basic) is as following: Glycine > Histidine > Phenylalanine > Aspartic acid > Tyrosine. From Figure 4.10, we can see the solubility of dipeptides with glycine residue follows the order of: Gly-Gly > Gly-His > Gly-Asp > Gly-Tyr > Gly-Phe. For glycine, histidine, aspartic acid and tyrosine, adding glycine residue does not affect the solubility order. Interestingly, Gly-Phe has the lowest solubility of the dipeptides considered, even though

phenylalanine has the third highest solubility of the amino acids studied. In order to explore the effect of phenylalanine, the solubilities of Gly-Phe, Asp-Phe, Tyr-Phe and Phe-Phe were compared. The order of their solubilities is Gly-Phe > Asp-Phe > Tyr-Phe > Phe-Phe. It can be noted that this series obeys a similar order to the dipeptides containing glycine residues, but phenylalanine residues reduce the dipeptides' solubilities significantly. As well as this, the solubilities of Gly-Gly, Gly-Phe and Phe-Phe are compared in Figure 4.10. The ratio of glycine residues appears to increase the solubility of the dipeptides, while phenylalanine reduces the solubility. The solubility of a dipeptide and its two constituent amino acid residues have also been compared. Interestingly for Gly-His, Gly-Tyr and Gly-Asp, the solubility of the dipeptide lies between those of the two constituent amino acids. However, dipeptides containing phenylalanine have a lower solubility than their constituent amino acids, with phe-phe having the lowest solubility of all dipeptides studied.

4.4.6 Molecular dynamic simulations

In order to investigate the role of peptide-water interactions on the solid-liquid phase equilibrium behaviour, the solvation free energy of the peptides studied were calculated with molecular dynamic simulations.

The solvation free energy is the free energy requirement to transfer peptides from their vapor phase to a solvent, which is a key factor in determining the solubility in one solvent or the relative solubility in solvent mixtures.[161] The Accelrys Materials Studio Forcite program (Accelrys Software Inc., US) was employed in this work to perform molecular dynamic simulation.[24] The Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies (COMPASS) was chosen from the force field list.

The total solvation free energy includes three separate free energies, which are the ideal free energy (removing the charges in vacuum), the van der Waals free energy (adding a

neutralized molecule in the solvent), and the electrostatic free energy (adding the charges on the solute).

First, geometry optimization of peptide and water molecule was performed using charged groups until the total energy of all structures were minimized. Then an amorphous cell was constructed with 1 peptide molecule and 500 water molecules. Figure 4.11 shows an example of the amorphous cells of glycine homopeptides and water molecules. Before starting the calculation, this initial structure was equilibrated to remove the excess heat from the system, which was performed by using the NVT ensemble and NHL thermostat for 500 ps at 298.15 K. The solvation free energy calculation was launched using the final structure of the equilibration run, which was divided into the three contributors: ideal free energy, electrostatic free energy and van der Waals free energy. Each contribution was run with 50000 steps for equilibration and 100000 steps for production.

The calculated solvation free energies of amino acids and peptides in water at 298.15K are listed in Table 4.11. The larger the absolute value of solvation free energy, the stronger the interactions between peptide and water molecules. To compare the solvation free energy of the glycine homopeptides and dipeptides, the number of water molecule were all fixed as 500, the solute molecule was set as 1. However, because of the low solubility properties of long chain length glycine homopeptides and some dipeptides, their real dissolved molar ratio is much lower than 1:500. To investigate this, calculations were performed with varying numbers of solvent molecules, which shows that the solvation free energy is proportional to the solute molecules. As a result, one sample was chosen as a reference, and the relative solvation free energy was calculated based on the relative solubilities. The equation of the relative solvation free energy is as follow:

$$Y = Y_0 \cdot \frac{c}{c_r} \tag{4.4}$$

where Y is the relative solvation free energy, Y_0 represents the solvation free energy, c and c_r

are the solubility and reference's solubility, respectively.

Tables 4.11 and 4.12 show the solvation free energy and the relative solvation free energy of the amino acids and peptides. The order of the relative solvation free energy has a positive correlation with that of the solubility values, which means that the solvation free energy can provide a useful indication of solubility. It has to be stated that since these results are the free energy required to transfer peptides from their gas phase to a solvent, they neglect any effect of lattice energy on solubility.

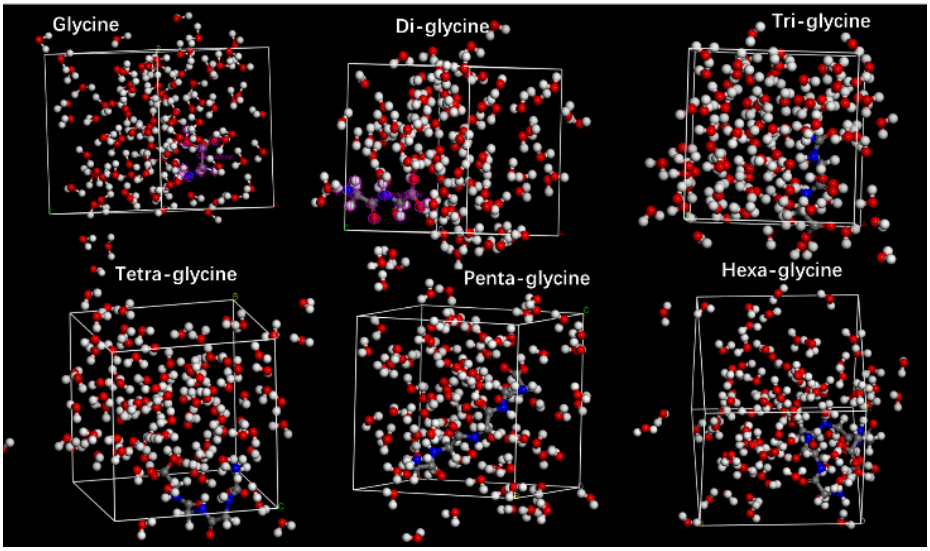


Figure 4.5 The amorphous cell of glycine homopeptides and water molecules for the solvation free energy calculation.

Table 4.8 Solvation free energy of glycine homopeptides in water at 298.15K

Glycine homopeptides	Solubility $10^2 x$ (mol/mol)	Number of solute molecules	Number of solvent molecules	T/K	Ideal Free Energy (kJ/mol)	Electrostatic free energy (kJ/mol)	Van der Waals free energy (kJ/mol)	Solvation free energy (kJ/mol)	Relative solvation free energy (kJ/mol)
Glycine	5.635	28.174	500	298	-35.137	25.288	2.726	-7.123	-7.123
Diglycine	2.727	13.636	500	298	-46.478	32.153	2.588	-11.736	-5.680
Triglycine	0.614	3.070	500	298	-60.301	42.179	2.637	-15.486	-1.687
Tetraglycine	0.025	0.127	500	298	-73.36	51.787	2.201	-19.372	-0.087
Pentaglycine	0.0027	0.014	500	298	-85.202	64.511	1.732	-22.201	-0.011
Hexaglycine	0.00045	0.002	500	298	-85.524	61.205	1.212	-23.107	-0.002

Table 4.9 Solvation free energy of amino acids and dipeptides in water at 298.15K

Amino acids and dipeptides	Solubility $10^2 x$ (mol/mol)	Number of Solute molecules	Number of Solvent molecules	T/K	Ideal Free Energy(kJ/mol)	Electrostatic Free Energy(kJ/mol)	Van der Waals free energy(kJ/mol)	Solvation free energy(kJ/mol) (kJ/mol)	Relative solvation free energy (kJ/mol)
Gly-His	1.9578383	9.000	500	298	-43.169	23.243	2.826	-17.100	-51.603
Histidine	0.6487760	3.000	500	298	-27.175	11.924	1.937	-13.313	-13.313
Gly-Asp	0.4164186	1.500	500	298	156.963	-178.854	1.847	-20.044	-12.866
Gly-Tyr	0.2824558	2.000	500	298	-12.243	-6.919	2.977	-16.186	-7.047
Phenylalanine	0.2718000	1.500	500	298	-12.95	-1.960	3.050	-11.860	-4.969
Gly-phe	0.2530000	1.347	500	298	-45.615	29.622	3.494	-12.500	-4.875
Aspartic acid	0.0657350	0.329	500	298	-46.348	32.331	2.407	-11.609	-1.176
Asp-Phe	0.0414390	0.200	500	298	3.004	-24.662	3.328	-18.330	-1.171
Tyrosine	0.0055600	0.028	500	298	-17.153	1.423	2.919	-12.811	-0.110
Tyr-Phe	0.0041890	0.021	500	298	-26.228	5.971	3.976	-16.283	-0.105
Phe-phe	0.0032844	0.016	500	298	-40.049	23.727	4.006	-12.316	-0.062

4.5 Conclusion

In this work, we determined the solubility of glycine homopeptides (glycine to hexaglycine), four amino acids with different side chains, and eight dipeptides in water with experimental and computational methods. Results show that the solubility of glycine homopeptides decreases (from $3.8604 \cdot 10^{-2}$ to $0.29 \cdot 10^{-5}$ mol/mol) with increasing chain length (from one to six residues). The solubility of diglycine is around fifty percent of that of glycine, while when the chain length is longer than two, the solubility is further reduced by one order of magnitude. For glycine, aspartic acid and histidine, adding glycine residues and phenylalanine residues have no effect on the solubility order. However, adding phenylalanine residue causes a large decrease in the solubility value of the dipeptide. Compared to the solubility of dipeptides and their two constituent amino acids, glycine residues resulted in the solubility of dipeptides falling between those of its constituent amino acids, while phenylalanine residues reduce the dipeptides solubility below those of both constituent amino acids. The more phenylalanine residues in the peptides, the lower the solubility of the peptides. These results presented that compared with basic, acidic, or polar side chains, the non-polar (hydrophobic) side chain causes the most apparent effect on the solubility. The experimental solubility data in water for all the amino acids and peptides are well correlated using the modified Apelblat model. The vast majority of average relative deviations (ARD) for all the correlations are less than 5% and only the ARD for hexaglycine is larger than 4%. Furthermore, the solvation free energy of the peptides in water calculated by molecular dynamic simulations suggests that the peptide-water interaction is influenced by chain length and side chains of peptides, with stronger interactions leading to higher solubilities. This thermodynamic research of amino acids and peptides helps to understand the effect of chain length and side chains on the solubility of peptides, which provides a basic database not only for solubility prediction research, but also for the design and optimization of peptide crystallisation.

Chapter 5 Solubility prediction of amino acids and peptides using SAFT- γ Mie

5.1 Abstract

The SAFT-Mie group contribution equation of state was used to predict the thermodynamic characteristics and phase equilibria (including solid-liquid solubility) of fluids using a heteronuclear model with molecules divided into their chemical moieties (CH_3, CH_2, NH_2, COOH, etc...).[71] Known group-group interaction parameters are adopted from the database or developed in this work using literature data as well as solubility data measured in Chapter 4. The NH_2-COOH parameters are determined based on the thermodynamic properties of glycine, alanine and valine. These parameters can be used to predict the properties of other structurally similar amino acids (aspartic acid, glutamic acid, isoleucine, leucine and lysine) and expend the parameter data bank for the longer chain peptides.

5.2 Introduction

Thermodynamic tools play an important role in the design and optimization of crystallisation process, especially in the prediction of thermodynamic properties.[165, 166] The accuracy of the description of the thermodynamic properties using predictive approaches can significantly affect process-design decisions. Classical equations of state (EoSs) which have been discussed in Chapter 4 (such as vdW, RK, SRK and PR) can correlate solubility with temperature and solvent component, giving guidance in solid-liquid phase equilibrium during various crystallisation methods in industry. However, these cubic equations are based on the hard-sphere fluid, which have poor performance for macromolecules that are highly non-spherical and associating molecules, such as hydrogen-bonding molecules. There is another class of molecular based equations of state called Statistical Associating Fluid Theory (SAFT), which has been well established as an approach for the description of the thermodynamics and phase equilibria of a wide variety of fluid systems.[71]

The SAFT EoS developed since the early 1950s, is based on Wertheim's thermodynamic perturbation theory (TPT1) and expressed in terms of Helmholtz energy, which has been explained in section 2.3.2. This EoS helps scientists to predict the thermodynamic properties of non-spherical and associating molecules.[71] During the past decades, SAFT has developed many variations based on their different systems and the intermolecular potentials and even form a family of SAFT of equations of state that can predict thermodynamic properties of complicated associating fluid systems, such as PC-SAFT (perturbed-chain), LJ-SAFT (Lennard-Jones spheres served as a reference for the chain formation), and VR-SAFT (the attractive potentials are allowed to show variable widths). Compared with these SAFT EoS, SAFT-γ Mie discussed in this chapter is based on a Mie potential (generalised Lennard-Jones) of variable repulsive and attractive range and combines a specific class of predictive theory that has drawn much attention over the past 40 years, which is group-contribution (GC)

method.[81] In this SAFT-γ Mie, the parameters between different molecular segments (such as functional groups) can be used to determine the final properties of a molecule or mixture. By accounting for the appropriate contributions of the segments, the interaction parameters between them were fitted using thermodynamic properties, such as vapor pressure, bubble pressure and solubility. The group–group interaction parameters can be used in the other systems which have the same segments. SAFT-γ Mie has already been verified to have the capacity to predict not only the vapor–liquid equilibrium (VLE) of many compounds for different chemical families, but also the liquid–liquid equilibrium (LLE) and solid–liquid equilibrium (SLE) for binary mixtures.[80] However, it has never been used to predict the peptides system, which is formed of amino acids and has repetitive peptide bonds and segments. This special structure of peptides makes them potential candidates to be researched based on SAFT-γ Mie group contribution method.

For glycine homopeptides, even though the residues are all glycine, there is still slight difference because of their positions. The one on the left end has NH_2 group but no carboxylate group, while the right ending one has carboxyl group without NH_2 group, and the residues in the middle only have peptide bonds, which means every amino acids can be divided into three different segments. One segment contains NH_2, the second segment consists of peptide bonds, while the third segment is composed of carboxylic acid group. And then if we want to express a peptide, we can simply select and combine different segments together. For example, there is a dipeptide called AB as shown in Figure 5.1. That is, we must first locate the left section of A and then join it to the right segment of B. Following that, we can move on to tri-and tetrapeptides. But the question is, can we predict the solubility based on the sequence of the amino acid residues? If we can predict the solubility, does the ABAB has a similar solubility with AABB if they have the same left and right segment? The only difference is the order of the two middle residues. So based on these questions, we need to measure the solubility first,

fit the parameters and then try to predict them

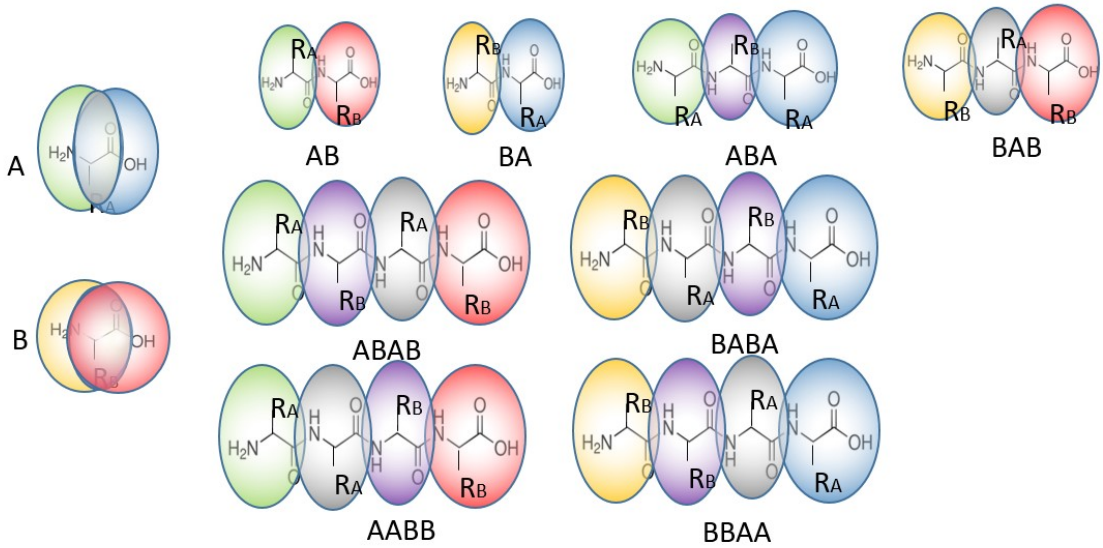


Figure 5.1 The different residue sequences of peptides with same segments

In this chapter, the SAFT-γ Mie group contribution equation of state in which a heteronuclear model is implemented, with molecules split into their chemical moieties (CH_3, CH_2, NH_2, COOH, etc…) was used to calculate the thermodynamic properties and phase equilibrium (including solid-liquid solubility) of fluids predictively. The SAFT-γ Mie theoretical formalisms for amino acids and peptides are first described, followed by group-group interaction parameters fitting process using literature data as well as solubility data measured in Chapter 4. On the basis of the thermodynamic characteristics of glycine, alanine, and valine, the NH_2-COOH parameters were determined. Other amino acids (aspartic acid, glutamic acid, isoleucine, leucine, and lysine) and peptides can be predicted using these parameters. The group-group interaction matrix of the SAFT-γ Mie approach was extended for the prediction of the solubility of amino acids and peptides.

5.3 Theory and molecular models

5.3.1 The Group-Contribution Approach.

In the SAFT-γ Mie approach, amino acids and peptides molecules can be divided into various chemical moieties which are represented as fused united-atom heteronuclear spherical segments based on the group contribution theory.[80] An example of the decomposition of a glycine molecule can be seen in figure 5.2. The association interactions between every segments are considered in this model as well, which are represented by the association sites to the corresponding chemical segments. There are two unique types of association sites: "e" and "H" sites, which exhibit electronegativity and electropositivity, respectively. The overall thermodynamic properties of the molecule in a system (pure component or mixture) are composed by the separate contributions of each of the divided segments with a shape factor. All the groups can be transferred across different molecules, which means the group-group interaction parameters in an unknown molecule can be addressed the same way they were in a studied molecule. For example, a CH_2 group in glycine is treated same as a CH_2 group in alanine.

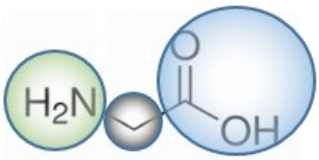


Figure 5.2 Example of the decomposition of glycine molecule into functional groups: one NH_2 group (shaded green), one CH_2 group (shaded gray), and one COOH group (shaded blue)

5.3.2 The interaction and parameters between two segments *m* and *n*

The interaction between two segments *m* and *n* can be presented as equation 5.1, in which they interact via a variable-range Mie potential:

$$\phi_{\mathrm{mn}}^{\mathrm{Mie}}(r_{mn}) = \mathrm{C}_{mn}\epsilon_{mn}\left[\left(\frac{\sigma_{mn}}{r_{mn}}\right)^{\lambda_{mn}^{r}} - \left(\frac{\sigma_{mn}}{r_{mn}}\right)^{\lambda_{mn}^{\alpha}}\right] \tag{5.1}$$

where r_{mn} expresses the distance between the segment m and segment n, ϵ_{mn} is the depth of the potential well, σ_{mn} is the segment diameter, and λ_{mn}^{r} and λ_{mn}^{α} represent the repulsive and attractive exponents, respectively. The item C_{mn} is a function of λ_{mn}^{r} and λ_{mn}^{α} to guarantee the minimum value of the potential is$-\epsilon_{mn}$.

The function of C_{mn} can be seen from the following equation:

$$\mathrm{C}_{mn} = \frac{\lambda_{mn}^{r}}{\lambda_{mn}^{r} - \lambda_{mn}^{\alpha}}\left(\frac{\lambda_{mn}^{r}}{\lambda_{mn}^{\alpha}}\right)^{\frac{\lambda_{mn}^{\alpha}}{\lambda_{mn}^{r}-\lambda_{mn}^{\alpha}}} \tag{5.2}$$

where λ_{mn}^{r} and λ_{mn}^{α} represent the repulsive and attractive exponents, respectively.

Besides, the association sites of segments *m* and *n* can be multiple, so the interaction between associating site *a* of segment *m* and associating site *b* of segment *n* was considered, which is given by:

$$\phi_{mn,ab}^{\mathrm{HB}}(r_{mn,ab}) = \begin{cases} -\epsilon_{mn,ab}^{HB} & \text{if } r_{mn,ab} \le r_{mn,ab}^{c} \\ 0 & \text{if } r_{mn,ab} > r_{mn,ab}^{c} \end{cases} \tag{5.3}$$

where $r_{mn,ab}$ is the distance between centers of site *a* of segment *m* and site *b* of segment *n*, $-\epsilon_{mn,ab}^{HB}$ represents the association energy, $r_{mn,ab}^{c}$ is the cutoff range of the interaction between two association sites, which can also be described as the term of bonding volume $K_{mn,ab}$.

The unlike group parameters are calculated using combining rules (CR) by combining the two individual groups *m* and *n* in the SAFT EoS.[80] The unlike segment diameter σ_{mn} and the unlike reference hard-sphere diameter d_{mn} are obtained using the Lorentz-like arithmetic mean, the unlike dispersion energy ε_{mn} is calculated using augmented-geometric mean (Berthelot-like rule)[167], the repulsive λ_{mn}^{r} and attractive λ_{mn}^{a} exponents are obtained by Berthelot-like combing rules, association energy $\varepsilon_{mn,ab}^{HB}$ is calculated based on a simple geometric mean, all the equations can be seen in the following:

$$\sigma_{mn} = \frac{\sigma_{mm}+\sigma_{nn}}{2} \tag{5.4}$$

$$d_{mn} = \frac{d_{mm}+d_{nn}}{2} \tag{5.5}$$

$$\varepsilon_{mn} = \frac{\sqrt{\sigma_{mm}^{3}\sigma_{nn}^{3}}}{\sigma_{mn}^{3}}\sqrt{\varepsilon_{mm}\varepsilon_{nn}} \tag{5.6}$$

$$\lambda_{mn} = 3 + \sqrt{(\lambda_{mm}-3)(\lambda_{nn}-3)} \tag{5.7}$$

$$\varepsilon_{mn,ab}^{HB} = \sqrt{\varepsilon_{mm,aa}^{HB}\varepsilon_{nn,bb}^{HB}} \tag{5.8}$$

The bonding volume of the unlike group is calculated using equation (5.9)

$$K_{mn,ab} = \left(\frac{\sqrt[3]{K_{mm,aa}}+\sqrt[3]{K_{nn,bb}}}{2}\right)^{3} \tag{5.9}$$

The combining rules have a cheap computing cost and are frequently utilised in complex systems where there is insufficient experimental data. However, it is unable to give good prediction results, the unlike group parameters have to be fitted by the experimental data again.

In summary, one chemical segment *m* can be characterised within the SAFT-γ Mie by the following parameters: the number of identical spherical segments in the molecule (ν_m^*) , the

shape factor (S_k). For the like segments, there are the diameter of each segment (σ_{mm}), the dispersion energy (ε_{mm}), and the repulsive (λ_{mm}^{r}) and attractive (λ_{mm}^{α}) exponents of the Mie potential. For the unlike segments, there are dispersion energy (ε_{mn}), and the repulsive (λ_{mn}^{r}) and attractive exponents (λ_{mn}^{a}) of the Mie potential. Furthermore, if segment m has different association sites, the association energy $\varepsilon_{mn,ab}^{HB}$ between the site α of segment m and site b of segment n, and the bonding volume $K_{mn,ab}$ should be considered.

5.3.3 SAFT-γ Mie Equation of State

In order to get the above parameters to predict the thermodynamic properties, the Helmholtz free energy was used to link the molecular model and the macroscopic thermodynamic properties (chemical potential and pressure) within the canonical ensemble based on the statistical−mechanical treatment[80], which expressed as the following SAFT-γ Mie EoS with four separate contributions (equation 5.10). The contributions from the four terms A^{ideal}, A^{mono}, A^{chain} and A^{assoc} can be expressed in figure 5.3. Other properties of the system, such as chemical potential and pressure, can be derived by standard thermodynamic relations.

$$\frac{\mathrm{A}}{Nk_BT} = \frac{A^{ideal}}{Nk_BT} + \frac{A^{mono}}{Nk_BT} + \frac{A^{chain}}{Nk_BT} + \frac{A^{assoc}}{Nk_BT} \tag{5.10}$$

where N, k_B and T represent the total number of molecules, Boltzmann's constant and Kelvin temperature, respectively.

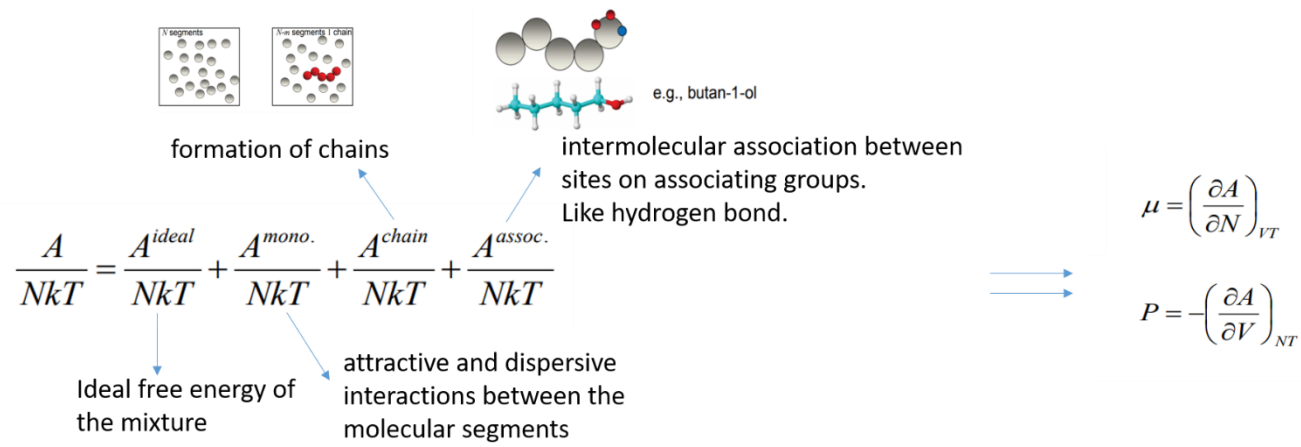


Figure 5.3 The four separate contributions of Helmholtz free energy in SAFT-γ Mie and the calculation of chemical potential and pressure based on Helmholtz free energy

5.3.3.1 Ideal term

The ideal term represents the ideal free energy contribution as shown in Equation 5.11:[80]

$$\frac{A^{ideal}}{Nk_BT} = \left(\sum_{i=1}^{N_C} x_i \, ln(\rho_i \Lambda_i^3)\right) - 1 \tag{5.11}$$

where N_i is the number of molecules of the component i in the system, x_i is the mole fraction of component i, ρ_i is the number density of component i, which calculated by $\rho_i = N_i/V$, and Λ_i^3 is the thermal de Broglie volume.

5.3.3.2 Monomer term

The monomer term which corresponds the repulsive and attractive interactions can be seen in equation 5.12.

$$\frac{A^{mono}}{Nk_BT} = \frac{A^{HS}}{Nk_BT} + \frac{A_1}{Nk_BT} + \frac{A_2}{Nk_BT} + \frac{A_3}{Nk_BT} \tag{5.12}$$

Where term A^{HS} is the free energy contributed by a hard-sphere reference system, A_1, A_2 and A_3 represent the correction terms, which are produced by performing comparable summations over the free-energy contributions for each segment, each with their associated power of inverse temperature, so that the mean attractive energy, the energy fluctuation, and the third-

order terms are given by

$$\frac{A_q}{Nk_BT} = \left(\frac{1}{k_BT}\right)^q \left(\sum_{i=1}^{N_C} x_i \sum_{k=1}^{N_C} \nu_{k,i}\, \upsilon_k^* S_k\right) a_q \qquad q = 1,2,3 \qquad (5.13)$$

where a_q shows the the free-energy contributions per segment.

5.3.3.3 Chain term

The chain term which represent the contribution due to chain formation can be seen in the following:

$$\frac{A^{chain}}{Nk_BT} = -\sum_{i=1}^{N_C} x_i \left(\sum_{m=1}^{N_G} \upsilon_{m,i}\, \upsilon_m^* S_m - 1\right) \ln\left(g_{ii}^{Mie}\right) \qquad (5.14)$$

where N_G is the number of types of segments, $\upsilon_{m,i}$ is the number of occurrences of segment m in component i, x_i is the mole fraction of component i, υ_m^* presents the number of identical spherical segments m, S_m is the shape factor of the segment m, and g_{ii}^{Mie} is the radial distribution function (RDF).

The segment diameter ($\bar{\sigma}_{mm}$), reference hard-sphere diameter ($\bar{d}_{mm}$), dispersion energy ($\bar{\varepsilon}_{mm}$) and Mie exponents ($\bar{\lambda}_{mm}$) of component m are also used in the chain term. The details about the equations of these terms can be found in literature.[71,80]

5.3.3.4 Association term

The association term denotes the interaction of two association sites, as illustrated in equation 5.15.

$$\frac{A^{assoc}}{Nk_BT} = -\sum_{i=1}^{N_C} x_i \sum_{m=1}^{N_G} \upsilon_{m,i} \sum_{a=1}^{N_{ST,m}} n_{m,a} \left(lnX_{i,m,a} + \frac{1-X_{i,m,a}}{2}\right) \qquad (5.15)$$

where $N_{ST,m}$ is the number of types of association sites on group m, $n_{m,a}$ is the number of sites of type a on group m, x_i is the mole fraction of component i, and $X_{i,m,a}$ is the fraction of molecules of component i that are non-bonded at a site of type a on group m. $X_{i,m,a}$ is a function

of both the depth of the association-energy well ($\varepsilon_{mn,ab}^{HB}$), which characterises the strength of the interaction, and the bonding volume of site a on group *m* and site b on group *n* ($K_{mn,ab}$).

5.3.4 Solid-Liquid Phase Equilibrium

A solid-liquid equilibrium (SLE) occurs when there is no more solids can be dissolved in the solution. The relationship between the potential of solid phase and liquid phase are shown in the following equation:

$$\mu_i^s(T) = \mu_i^l(T) + \Delta H^{l\to s}(T_m)\left(1 - \frac{T}{T_m}\right) + \int_{T_m}^{T} \Delta C_p^{l\to s}(T')dT' - T\int_{T_m}^{T} \frac{\Delta C_p^{l\to s}(T)}{T'}dT' \quad (5.16)$$

Where $\mu_i^s(T)$ and $\mu_i^l(T)$ are the potential of solid and liquid phase at specific temperature T, respectively; $\Delta H^{l\to s}$ and $\Delta C_p^{l\to s}$are the melting enthalpy and melting heat capacity when the solid melt to liquid, respectively; T_m is the melting temperature.

The mole fraction solubility of peptides can be expressed in the following equation (5.17).

$$\ln x_j = \left(\frac{T}{T_{fus,j}} - 1\right)\frac{\Delta H_{fus,j}}{Nk_BT} - ln\gamma_j(T,P,x_j) \quad (5.17)$$

Where x_j is the mole fraction solubility of component *j*, $T_{fus,j}$ is the fusion temperature of component *j*, $\Delta H_{fus,j}$ is the fusion enthalpy of component *j*, $\gamma_j(T,P,x_j)$ is the activity coefficient of component *j* at the specified temperature T, pressure P, and composition x_j.

The activity coefficient of the solute $\gamma_j(T,P,x_j)$ is derived from the Helmholtz free energy using standard thermodynamic relations. The fusion temperature $T_{fus,j}$and fusion enthalpy $\Delta H_{fus,j}$ as the input parameters need to obtained from experimental data.

5.3.5 Parameter estimation F_{obj} and the absolute deviation (%AAD)

Multiple thermodynamic properties can be used as input experimental data during the fitting process, and the optimal solution of the parameters is determined by minimising the following objective function F_{obj}, which is a summation of the relative deviations between the

experimental and predicted data. This function takes into account a weighing factor W_Z for each property.

$$\min F_{obj} = \sum_{k=1}^{N_z} \frac{W_Z}{N_{p,k}} \sum_{l=1}^{N_{p,k}} \left[\left(\frac{Z_{k,l}^{exp} - Z_{k,l}^{calc}}{Z_{k,l}^{exp}} \right)^2 \right] \tag{5.18}$$

where N_z represents the number of properties used for fitting the parameters; $N_{p,k}$ means the number of data points used for each type of properties; W_Z is the weighting factor given to each type of experimental data.

The percentage average absolute deviation (%AAD) was used to compare the experimental data and calculated data. A generic property R which contains n_R separate data points is shown as follow:

$$\%\text{AADR} = \frac{1}{n_R} \sum_{i=1}^{n_R} \left| \frac{R_i^{exp} - R_i^{calc}}{R_i^{exp}} \right| \times 100 \tag{5.19}$$

5.4 Results and discussion

5.4.1 Structures and group parameters of studied amino acids and peptides

Glycine homopeptides were selected as model peptides because glycine is the simplest amino acid without a side chain, which can exclude the interaction between side chains and glycine and reduce the number of unknown parameters. Based on the group contribution approach, there are different division strategies: glycine homopeptides can be divided into NH_2, CH_2, NH, CO, and COOH segments, or be divided into NH_2CH_2CO, $NHCH_2CO$ and $NHCH_2COOH$ segments, or be divided into NH_2, CH_2, CONH, COOH segments. Figure 5.4 illustrates the various segment division strategies. For peptides that contain side chains, the side chains can be further split into one or more segments. The more segments seperated, the more parameters required for modelling. However, because there are few experimental data available on peptides, the parameters can be determined using glycine and other structurally

similar chemicals which are composed of the same segments. The matrix in table 5.2 shows the group-group interaction of glycine homopeptide-water system using the first strategy, the dark blue ones mean the parameters that have already been fitted and deposited in the databank from previous work, while the grey values indicate parameters that must be developed using experimental data in this chapter.

The group-group parameter matrix for pure glycine and the mixture system of glycine and water is shown in Figure 5.5. The only unresolved interaction parameters are those between NH_2 and COOH, which are the specific groups existing in every amino acid and peptide. Thus, the experimental data from the pure component, such as the vapour pressure, as well as the thermodynamic properties of the mixture system like solubility and bubble pressure can be utilised to fit the parameters between NH_2 and COOH. These parameters can be further used to predict more solubility of other amino acids with the same functional groups.

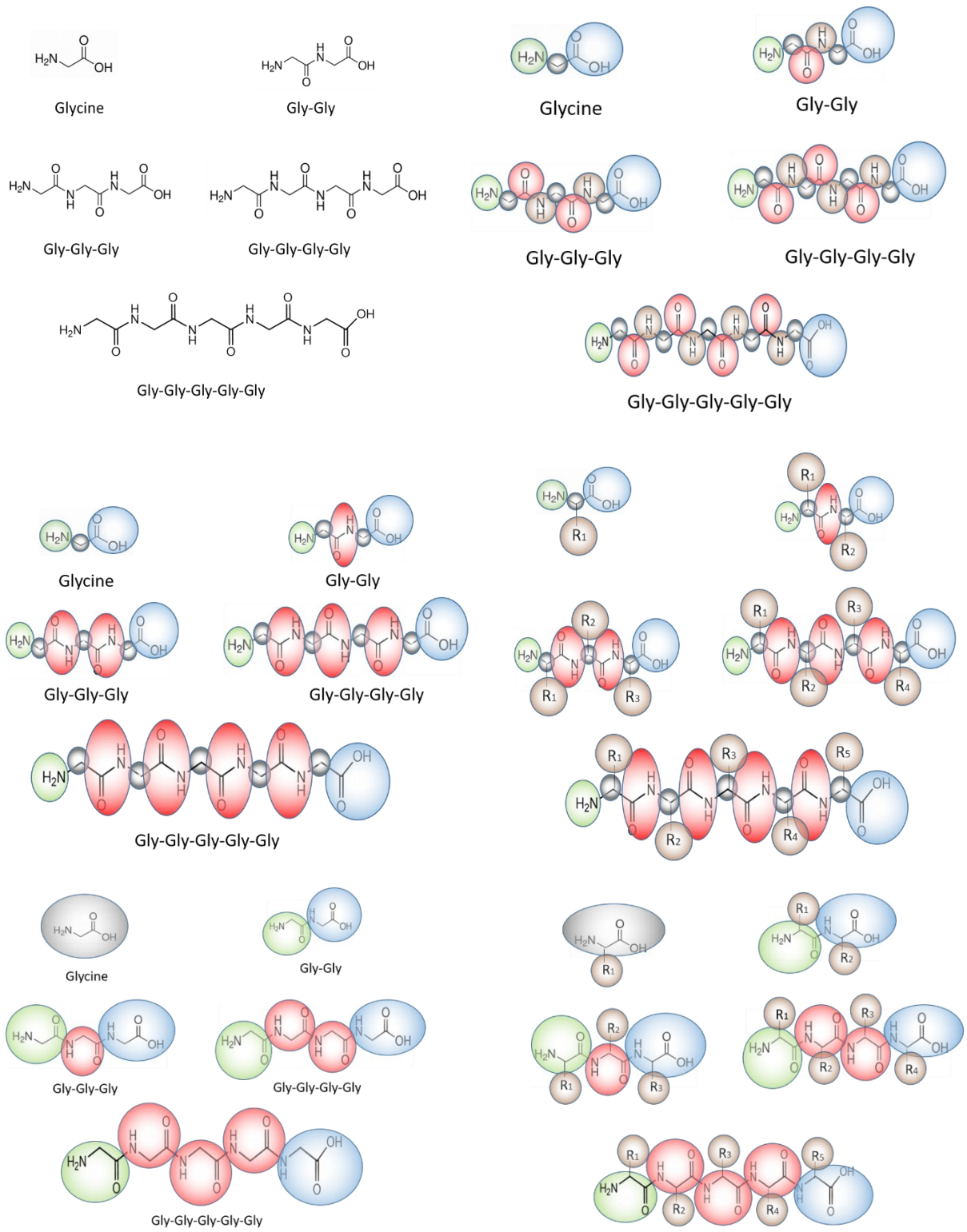


Figure 5.4 The different segments division strategies of peptides with and without side chains.

Table 5.1 The like and unlike group-group interaction in glycine homopeptides

	NH_2	CH_2	COOH	C=O	NH	H_2O
NH_2	■					
CH_2	■	■				
COOH	□	■	■			
C=O	□	■	□	■		
NH	□	■	□	□	■	
H_2O	■	■	■	□	■	■

For NH_2 and COOH which have hydrogen acceptors and donors, the association interaction should be considered in SAFT-γ Mie approach. For NH_2, there are three association sites: one electron "e" site located on the nitrogen atom and two hydrogen "H" sites located on hydrogen atom. For COOH group, there are two "e1" sites located on the carbonyl oxygen, two "e2" sites located on the hydroxyl oxygen and one "H" site located on the hydrogen atom. For water, there are four association sites containing two "e1" sites on the oxygen atom and two "H" sites located on the hydrogen atoms. The interaction between these association sites can be seen in figure 5.5.

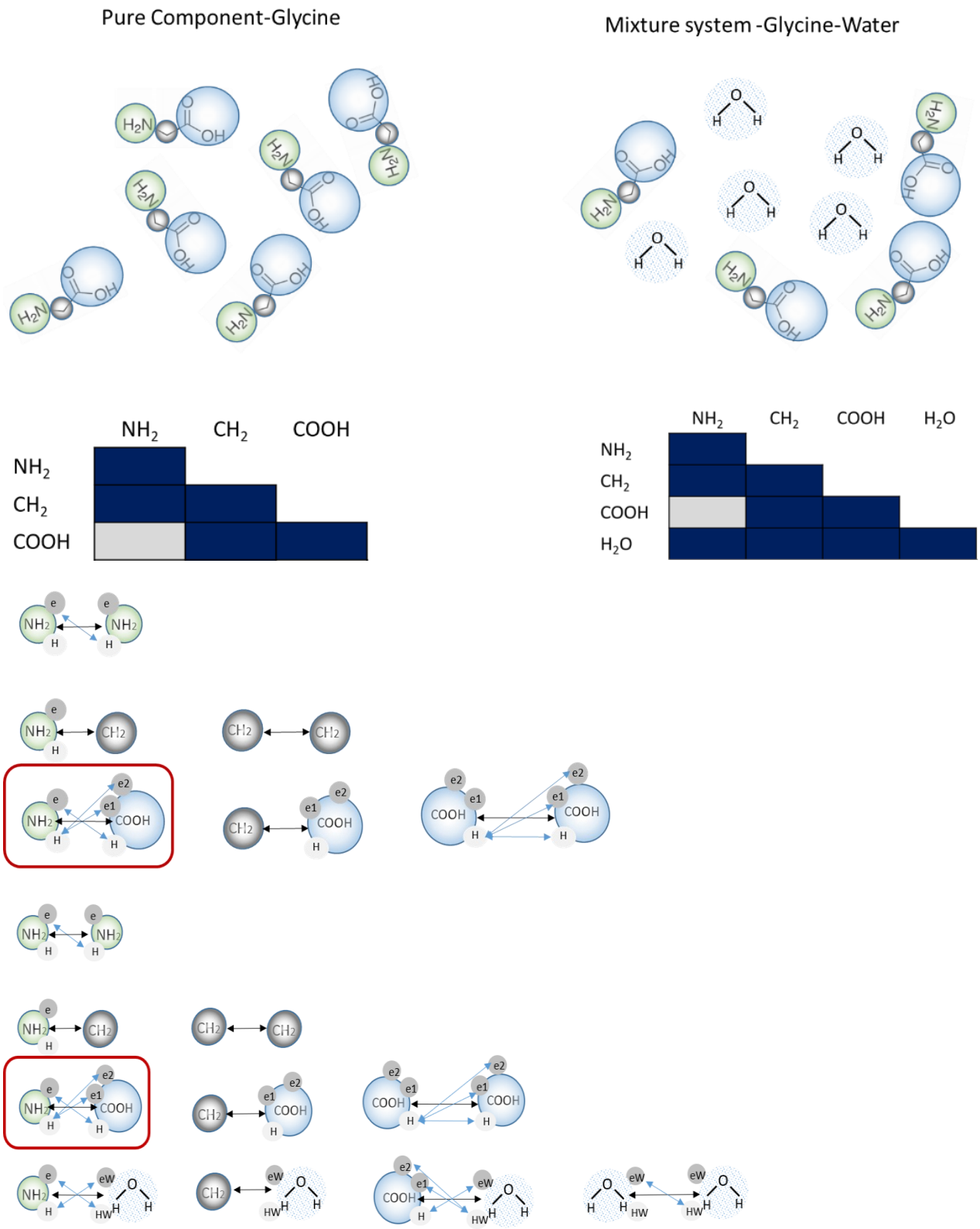


Figure 5.5 The like and unlike group-group parameters and interactions for glycine pure component and glycine-water mixture. The red boxes represent the unknown parameters.

In order to find more structurally similar chemicals, the following amino acids which have extra CH_3 and CH groups compared to glycine were considered to be the model peptides too (Figure 5.6). The matrix shows the group-group interaction of these amino acids, the dark blue one means the parameters fitted and deposited in the databank from previous work, and the grey parameters are those need to be developed in this work using experimental data. Combing rules are used for the parameters between CH and NH_2 groups.

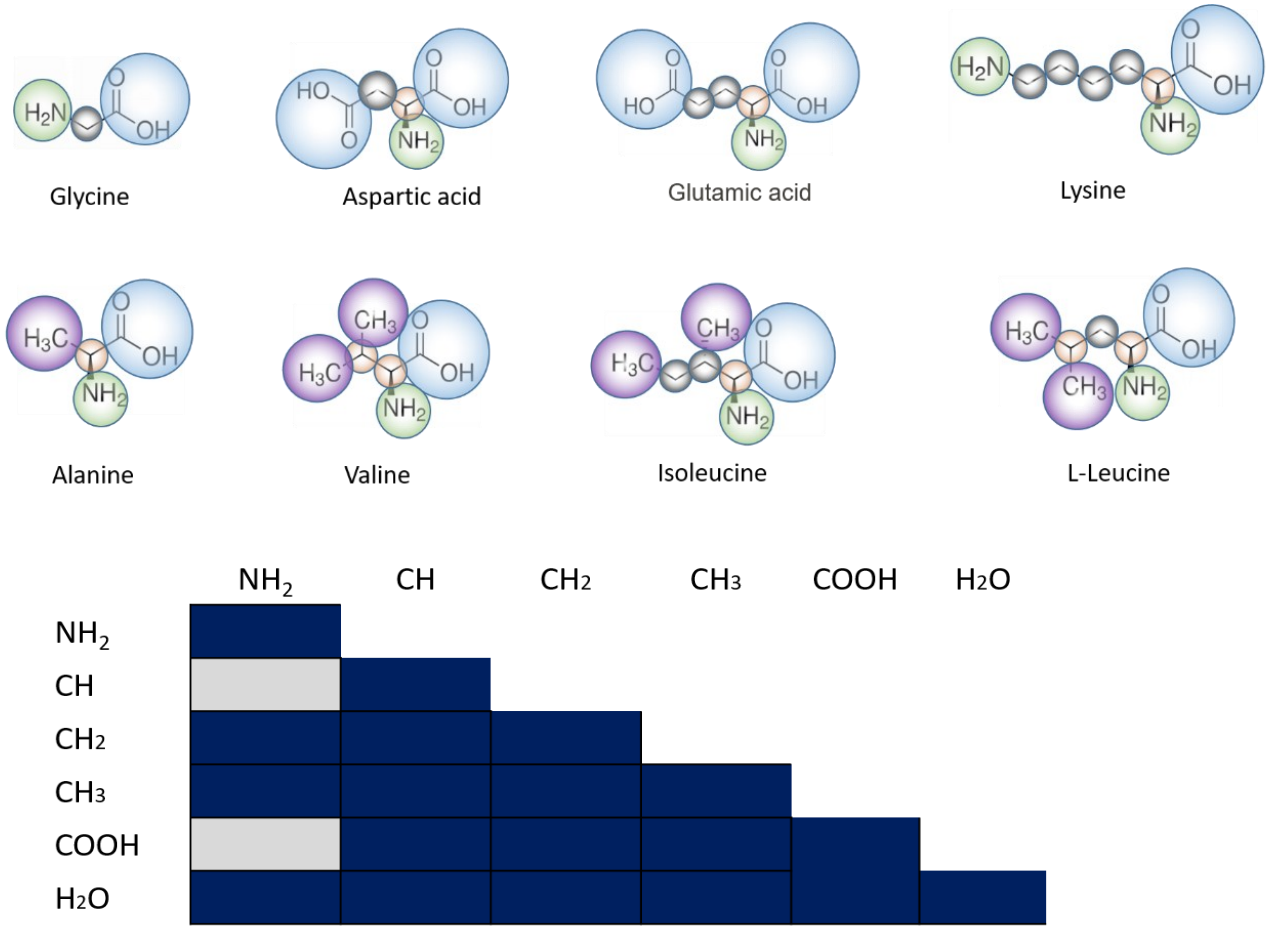


Figure 5.6 The amino acids molecules decomposed into functional groups and the like and unlike group-group parameters and interactions for the studied amino acids in water.

In summary, the following table lists the parameters that exist in database:

Table 5.2 Like group-group parameters for studied amino acids in this work

m	group m	ν_m^*	S_k	λ_{mm}^r	λ_{mm}^α	σ_{mm} /Å	$(\varepsilon_{mm}/k_B)/K$	$N_{ST,m}$	$n_{m,H}$	n_{m,e_1}	n_{m,e_2}
1	NH_2	1	0.7968	10.254	6.0000	3.2477	284.78	2	1	2	-
2	CH	1	0.07210	8.0000	6.0000	5.2950	95.621	-	-	-	-
3	CH_2	1	0.22932	19.871	6.0000	4.8801	473.39	-	-	-	-
4	CH_3	1	0.57255	15.050	6.0000	4.0773	256.77	-	-	-	-
5	COOH	1	0.55593	8.0000	6.0000	4.3331	405.78	3	1	2	2
6	H_2O	1	1.00000	17.020	6.0000	3.0063	266.68	2	2	2	0

Table 5.3 Unlike group-group parameters for studied amino acids in this work

m	n	group m	group n	$(\varepsilon_{mn}/k_B)/K$	λ_{mn}^r	m	n	group m	group n	$(\varepsilon_{mn}/k_B)/K$	λ_{mn}^r
1	2	NH_2	CH	CR	CR	2	6	CH	H_2O	101.89	CR
1	3	NH_2	CH_2	348.39	″	3	4	CH_2	CH_3	350.7	17.258
1	4	NH_2	CH_3	244.15	″	3	5	CH_2	COOH	413.74	″
1	5	NH_2	COOH	In this work	″	3	6	CH_2	H_2O	284.53	″
1	6	NH_2	H_2O	358.55	″	4	5	CH_3	COOH	255.99	″
2	3	CH	CH_2	506.21	″	4	6	CH_3	H_2O	274.80	″
2	4	CH	CH_3	387.48	″	5	6	COOH	H_2O	399.00	″
2	5	CH	COOH	504.99	″						

5.4.2 Parameter fitting process using pure component properties of glycine

The interactions between NH_2 and COOH can be seen in figure 5.5, which corresponds to the following parameters: the dispersion energy ε_{mn}, association energy $\varepsilon_{mn,ab}^{HB}$ and the bonding volume $K_{mn,ab}$ for the three pairs of association sites. The thermodynamic properties of glycine pure component were checked in Detherm (https://dechema.psds.ac.uk/), a databank

that collected all published thermodynamic data from literature. Vapour pressure, saturation liquid density, vaporization enthalpy, heat capacity of glycine were used as the input experimental data in SAFT. The parameters in table 5.2 and 5.3 were used to fit the new parameters. The results of the parameters between NH_2 and COOH can be seen in Table 5.4. Figure 5.7 shows the comparison between experimental data and calculated data. For vapour pressure, saturation liquid density, and heat of vaporization, it has a good prediction with small AAD values (all lower than 0.8). However, for the properties in glycine-water mixture, such as heat capacity and bubble pressure, the predicted results cannot be fitted with the experimental data very well, especially for bubble pressure. The parameters can not be used to predict the solubility at the same time.

Table 5.4 Group-group parameters for NH_2-COOH using pure component properties of glycine

Objective Function	Epsilon	Cross Epsilon	Bonding Volume	Cross Epsilon	Bonding Volume	Cross Epsilon	Bonding Volume
	COOH – NH_2	COOH (e1) – NH_2 (H)	COOH (e1) – NH_2 (H)	COOH (e2) – NH_2 (H)	COOH (e2) – NH_2 (H)	COOH (H) – NH_2 (e)	COOH (H) – NH_2 (e)
81.49334501	336.224544	8279.099968	1.57E-35	3410.468813	1.06E-29	4106.947791	9.01E-28

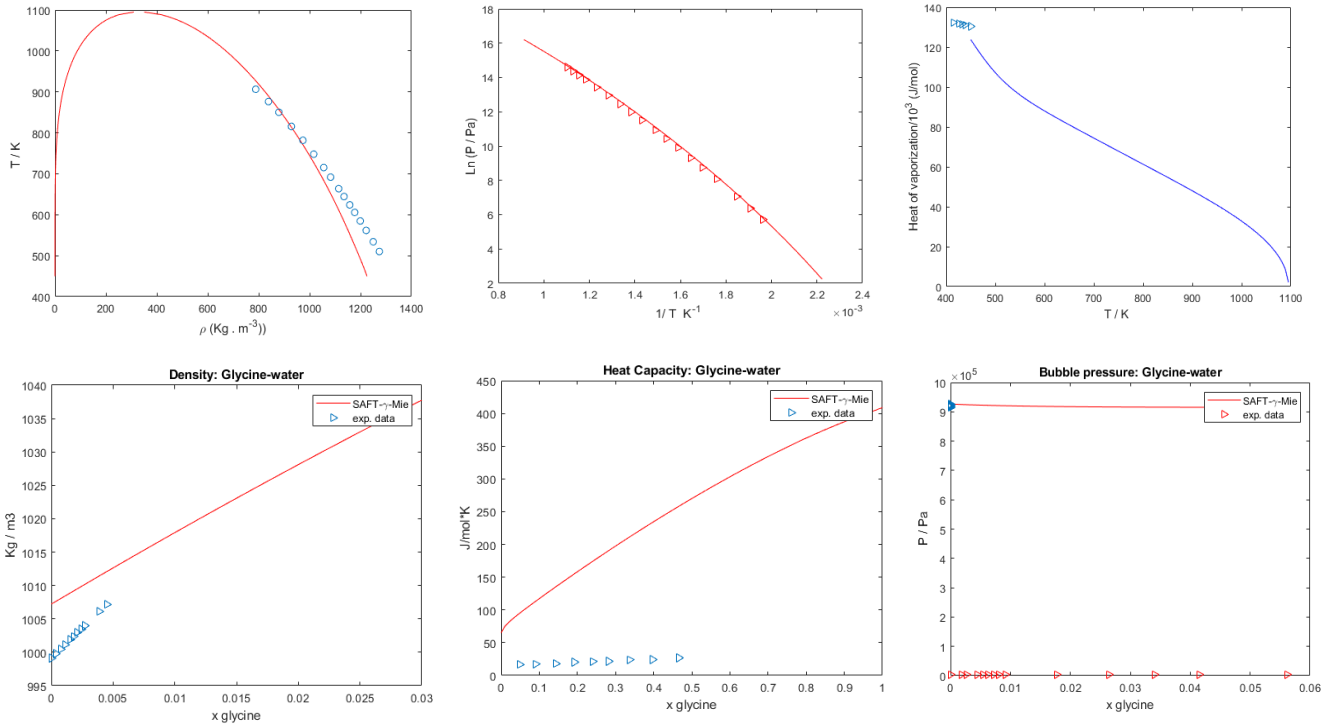


Figure 5.7 Comparison between experimental data and predicted data of glycine using pure component properties. The points mean the experimental data and the lines present the predicted data.

5.4.3 Parameter fitting process incorporating properties of glycine-water mixture

In order to get better predicted results, the experimental data of solubility and bubble pressure of glycine-water system were added to fit the parameters between COOH and NH_2. From the results in figure 5.8, all the properties have a good predicted results with a low AAD value except heat capacity which still has a large difference compared with the experimental data, the parameters can be seen in table 5.5. However, the value of bonding volume which describes the volume accessible to bonding between COOH (e1)-NH2 (H) is too small to have its physical meaning, which is at least larger than 1E-30.

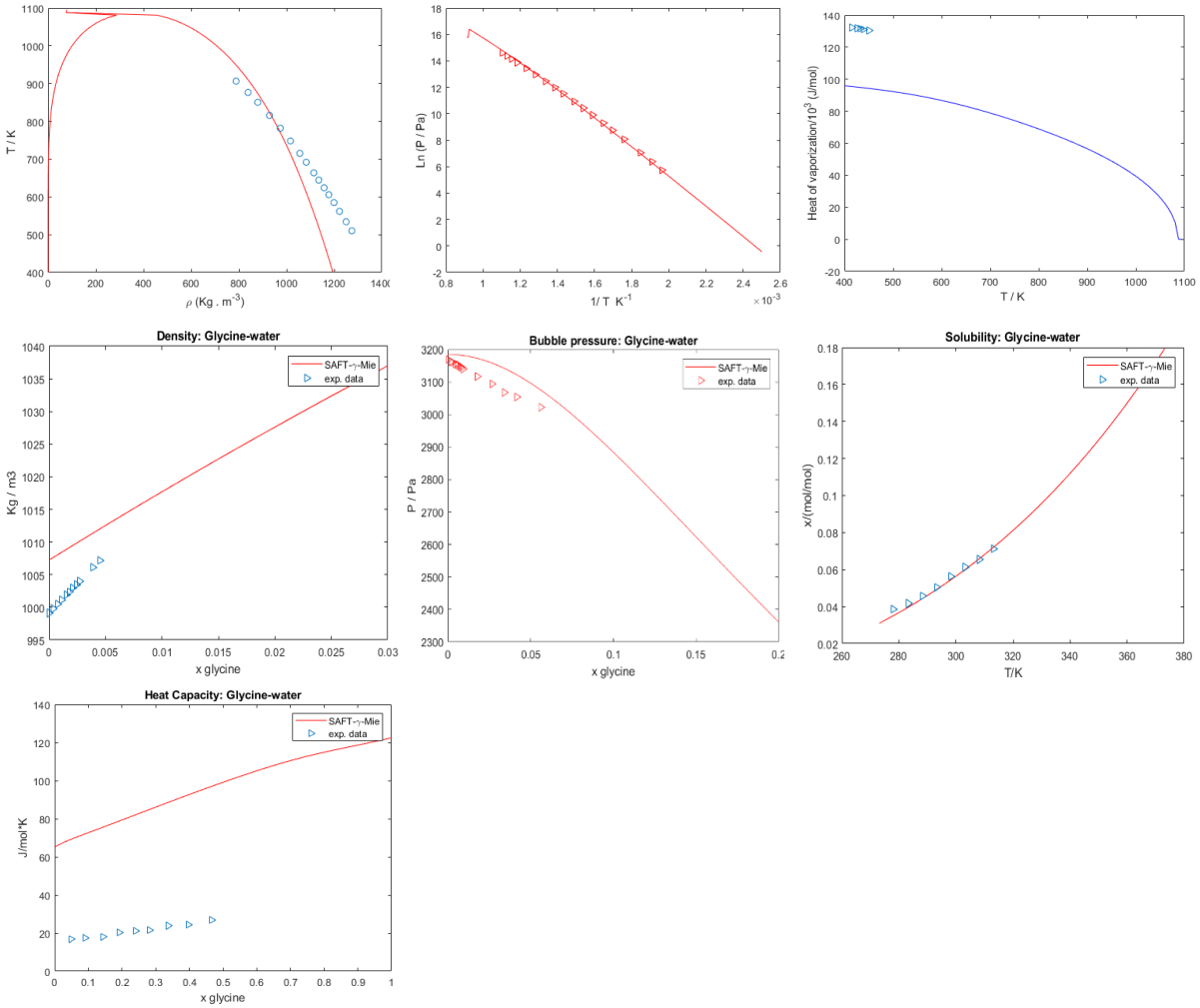


Figure 5.8 Comparison between experimental data and predicted data of glycine using pure and mixture component properties. The points mean the experimental data and the lines mean the predicted data.

Table 5.5 Group-group parameters of NH_2-COOH using pure properties of glycine and mixture properties of glycine-water

Objective Function	Epsilon	Assoc Epsilon	Bonding Volume	Assoc Epsilon	Bonding Volume	Assoc Epsilon	Bonding Volume
	COOH - NH_2	COOH (e1) -NH_2 (H)	COOH (e1) - NH_2 (H)	COOH (e2) - NH_2 (H)	COOH (e2) - NH_2 (H)	COOH (H) - NH_2 (e)	COOH (H) -NH_2 (e)
209.672	225.12268	5071.7463	1.01E-37	1000.044	5.52E-28	6867.044	2.53E-29

5.4.4 Melting temperature and fusion enthalpy of amino acids and peptides

The melting temperature and fusion enthalpy of glycine were determined using a differential scanning calorimetry (DSC) pattern from 298.15 K to 600 K at a scan rate of 10 K/min. The thermogravimetric analysis (TGA) of glycine was carried out from 298.15 K to 600 K at a heating rate of 10 K/min. From the results in figure 5.9, the decomposition process was tested during the melting process at the same time, so the heat calculated by the integral of the endothermic peak in the DSC pattern includes not only the heat from melting but also the heat from decomposition process, implying that the fusion enthalpy cannot be measured using DSC.

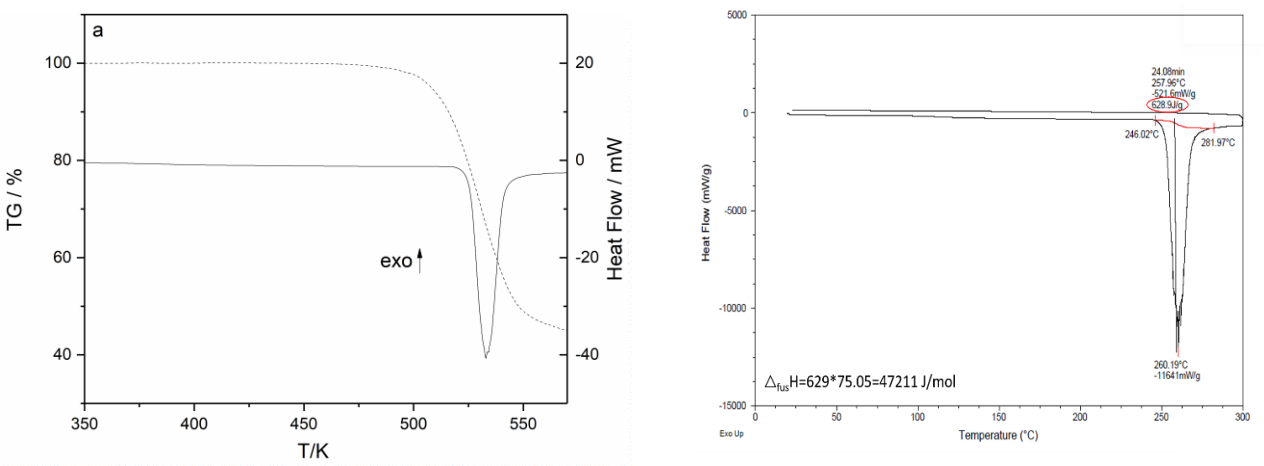
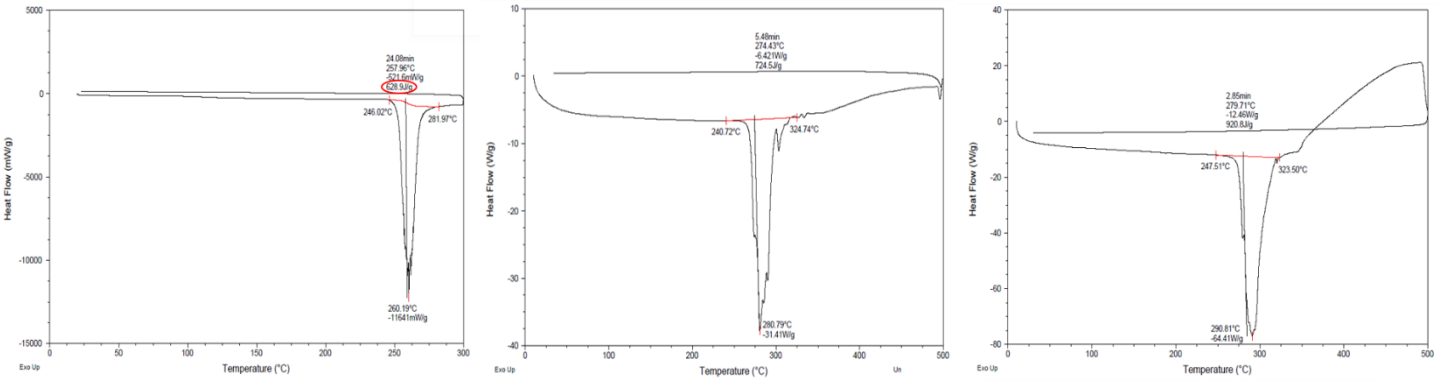


Figure 5.9 Left: the DSC curve (solid line) and TGA curve (dash line) for glycine. Right: the fusion enthalpy calculated by integrating the peak area in DSC curve.

This decomposition process occurred for all amino acids and peptides. Different heating rate (10 K/min, 50 K/min, 100 K/min) were set in DSC to attempt to separate the melting process and the decompose processes, but none of them have two separate peaks (figure 5.10). Christoph and co-workers used the fast scanning calorimetry to separate the two peaks successfully, in which the heating rate can be as high as 1000 K/s (figure 5.11).[55, 168] The fusion

enthalpy can be calculated based on the separated melting peak. However, the degree of uncertainty is still 20%. Other fusion enthalpy in the literature were summarized in table 5.6. For the fusion enthalpy obtained using the group contribution method, the value is similar with those obtained using the fast scanning calorimetry.



10K/min $\triangle_{fus}H$=629*75.05=47211 J/mol 50K/min $\triangle_{fus}H$=724.5*75.05=54373 J/mol 100K/min $\triangle_{fus}H$=920.8*75.05=69106 J/mol

Figure 5.10 DSC curves of glycine using different heating rate (10K/min, 50K/min, 100 K/min).

Table 5.6 The fusion enthalpy of glycine listed in the literature

Fusion enthalpy (J/mol)	Method	Reference
42790	Van't Hoff equation $\Delta H^{\Theta}_{sol} = -R^{*}\left[\frac{\partial \ln x_A}{\partial(1/T - 1/T_{hm})}\right]$	169
24114	Adjusted and fitted to experimental solubility data	76
21000 ± 4000	Fast scanning calorimetry	168
24100	Group contribution + modelling adjust	170
47211	DSC in this work	-

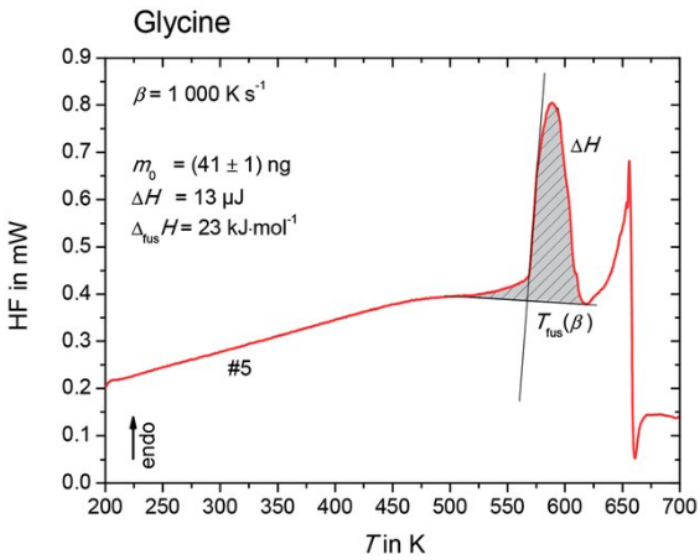


Figure 5.11 Heat flow rate curve of glycine using fast scanning calorimetry [168]

5.4.5 Analysis of the impact of fusion enthalpy

Since the fusion enthalpy cannot be obtained by DSC, group contribution methods were used to predict the values. The sensitivity of fusion enthalpy on the final prediction results was also examined in this section.

In order to test the impact of fusion enthalpy, the solubility of glycine was predicted using different fusion enthalpies listed in table 5.6. The results can be seen in figure 5.12. There is a 30% deviation for the solubility even though the fusion enthalpy only has a 10% difference.

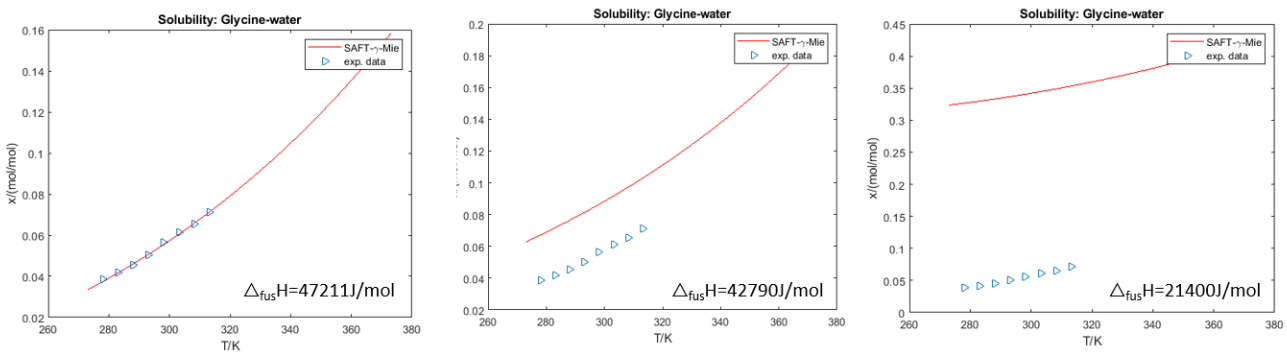


Figure 5.11 The comparison of predicted solubility for glycine with different fusion enthalpy (The points are the experimental data and the lines are the predicted values)

Different group contribution methods were used to predict the fusion enthalpy of the studied amino acids, which is shown in table 5.7. For Joback method, an extension of the Lydersen method, there is no interactions considered between the groups and only a small number of necessary parameters required to calculate thermophysical and transport properties.[171] Gani's method is performed at three levels, taking into account poly-functional compounds and structural groups based on conjugation theory.[172] This method broadens the range of applicability and improves the accuracy of group contribution methods. In order to get reasonable fusion enthalpy in this work, the fusion enthalpy was first estimated using Gani's method and then adjusted the value based on the final results within its deviation range which is around 16%. Another way is to fit the parameters without solubility data, and estimating the value based on the predicted solubility results. The fusion enthalpy of amino acids (glycine, alanine, valine, aspartic acid, glutamic acid, isoleucine, leucine and lysine) were estimated using the first method (details can be seen in section 5.4.5).

Table 5.7 Fusion enthalpy of studied amino acids using different group contribution methods

	Fusion enthalpy			
Amino acid	Joback Method (J/mol)	Gani Method (J/mol)	Adjust through the results (J/mol)	Deviation*
Glycine	16276	28400	28400	20.42%
Alanine	15340	22861	27061	-18.37%
Valine	17000	25051	28051	-15.97%
Aspartic acid	28080	56194	70194	-24.91%
Glutamic acid	30670	29722	81722	-174.95%
Isoleucine	19590	27294	24294	10.99%
Leucine	19590	27690	25690	7.22%
Lysine	28310	48266	22566	53.25%

*The deviation values mean the deviation between the adjusted values and the values calculated by Gani's method.

5.4.6 Only use bubble pressure and solubility of glycine-water to fit parameters

The DSC measurement in section 5.4.4 revealed that the saturated liquid density experimental data utilised in sections 5.4.2 and 5.4.3 exceeded their physical meaning bounds, of which the temperature range is from 510 K to 906 K, while the initial melting point measured in DSC is 533.15 K, implying that there is no liquid state above this temperature, let alone saturated liquid density. As a result, the saturated liquid density data must be discarded during the subsequent fitting process.

Since the saturated liquid density of glycine is not suitable to fit the parameters and predict the properties in glycine –water mixture system, the new fitting processes based on other properties were attempted. According to the results, when simply vapour pressure or other pure component qualities are utilised in the fitting process, the solubility data cannot always be predicted, as demonstrated in Section 5.4.2. Another strategy which includes vapor pressure, mixture liquid density, bubble pressure, heat capacity and solubility were also considered, but still cannot predict solubility and bubble pressure, however the value of heat capacity improved significantly. After analysing the experimental data, it is possible that the low vapour pressure (down to 4 Pa) is causing the problem; the relative mean deviations are fairly large due to the extremely low absolute vapour pressures. Thus, after validating all experimental data, the mixture properties (the bubble pressure and solubility) were determined to provide the optimum findings.

The obtained parameters between NH_2 and COOH are shown in the following table 5.8. The fusion enthalpy of glycine is calculated by Gani group contribution method (28400 J/mol). This set of parameters were used to predict other pure component properties of glycine to verify the accuracy, which can be seen in figure 5.13. At the same time, the parameters were used to predict the solubility of other amino acids, and the fusion enthalpy of the predicted amino acids were adjusted based on the solubility experimental data, the new fitting results can be seen in

the right side of figure 5.14. The parameters were also used to predict the bubble pressure of alanine and valine, which shows a good prediction (figure 5.15).

Table 5.8 Group-group parameters of NH_2-COOH using mixture properties of glycine-water

Objective Function	Epsilon	Assoc Epsilon	Bonding Volume	Assoc Epsilon	Bonding Volume	Assoc Epsilon	Bonding Volume
	COOH – NH_2	COOH (e1) – NH_2 (H)	COOH (e1) – NH_2 (H)	COOH (e2) – NH_2 (H)	COOH (e2) – NH_2 (H)	COOH (H) – NH_2 (e)	COOH (H) – NH_2 (e)
3.094777	115.2255	1306.751	5.35E-28	1137.188	8.91E-28	1361.993	4.47E-28

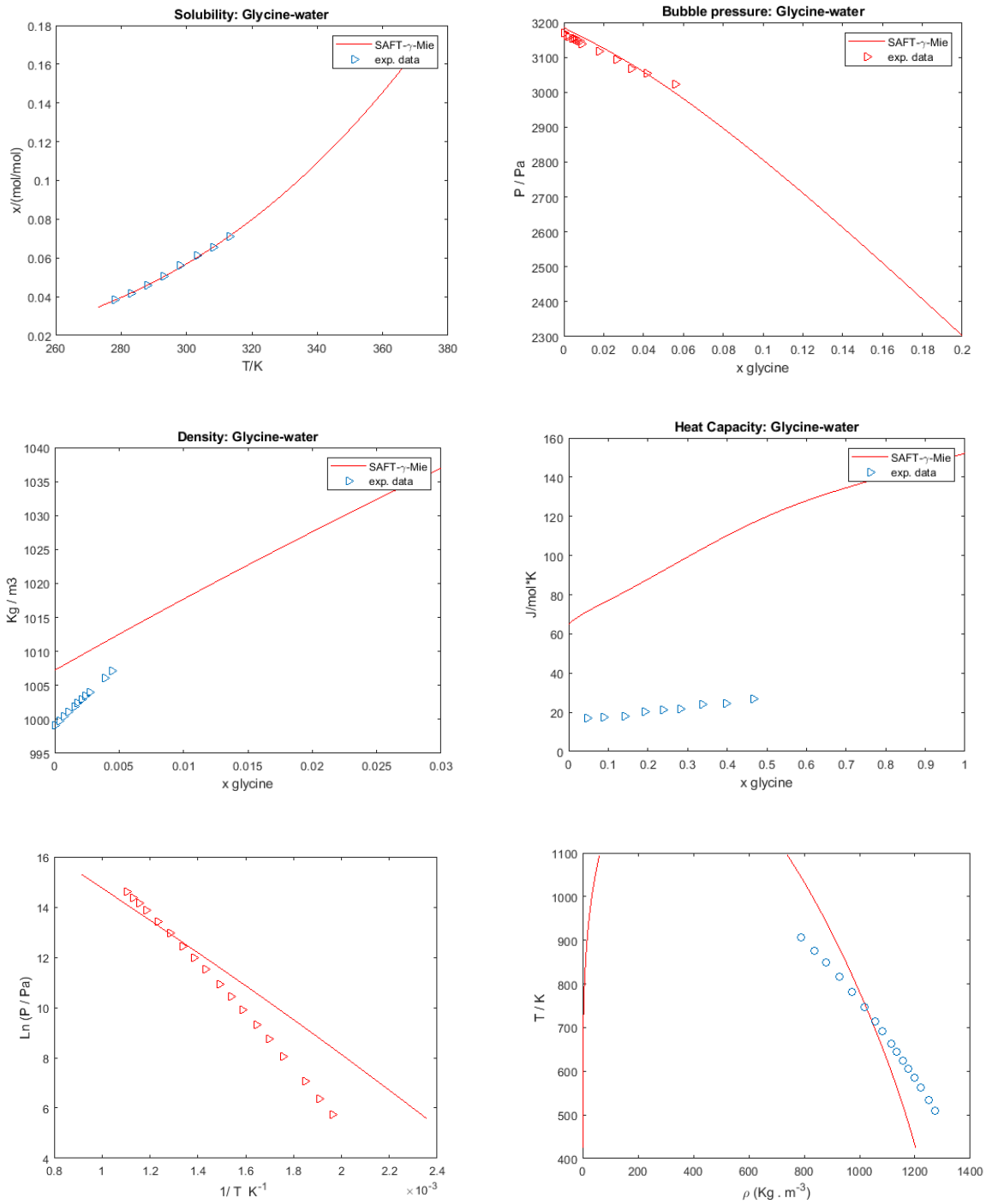
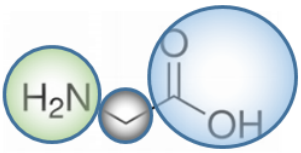
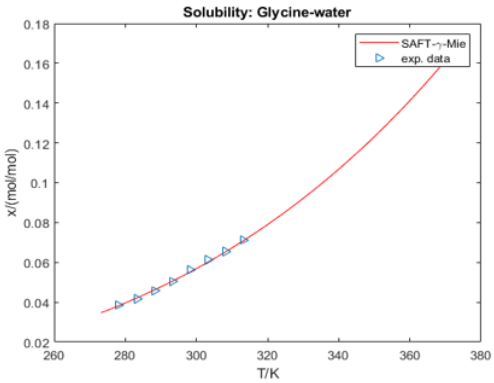


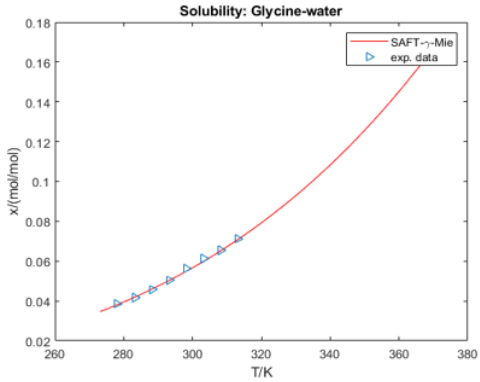
Figure 5.12 The comparison of experimental and predicted thermodynamic data of glycine and glycine-water.



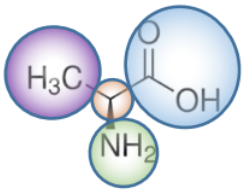
Glycine



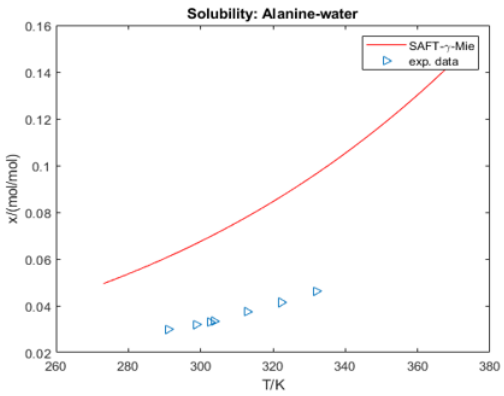
H=28400 J/ mol



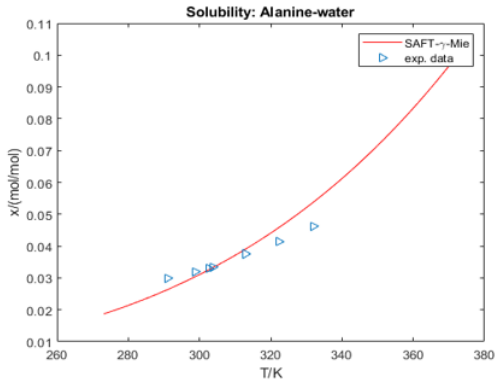
H=28400 J/ mol



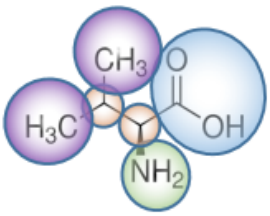
Alanine



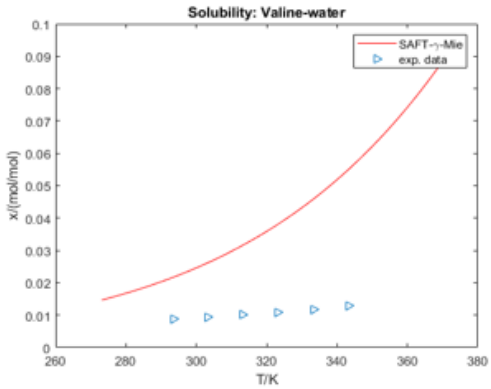
H=22861 J/ mol



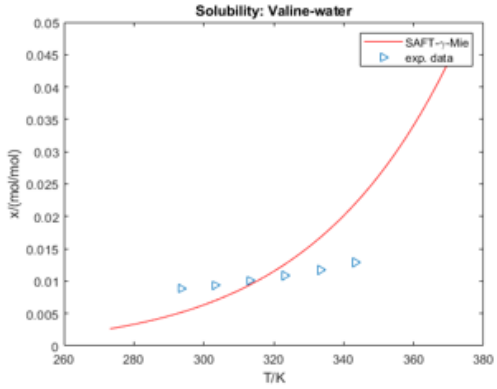
H=27061 J/ mol



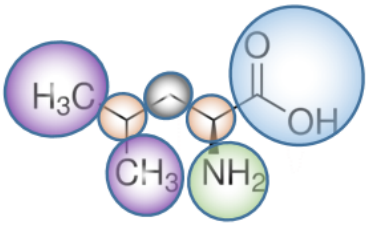
Valine



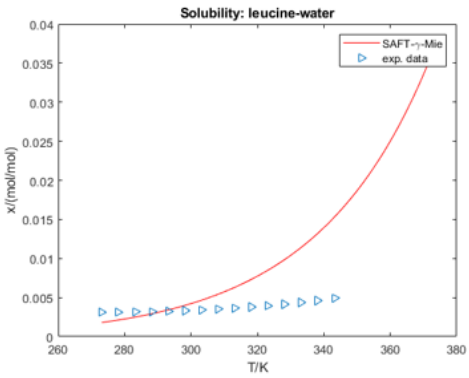
H=25051 J/ mol



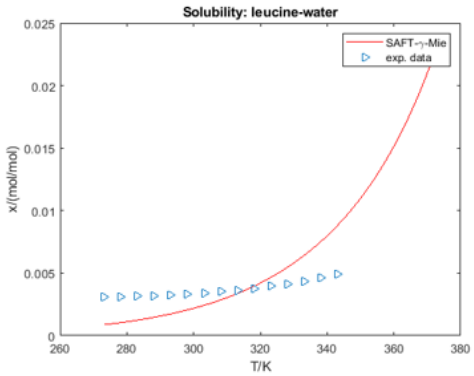
H=28051 J/ mol



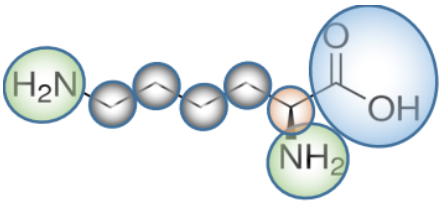
L-Leucine



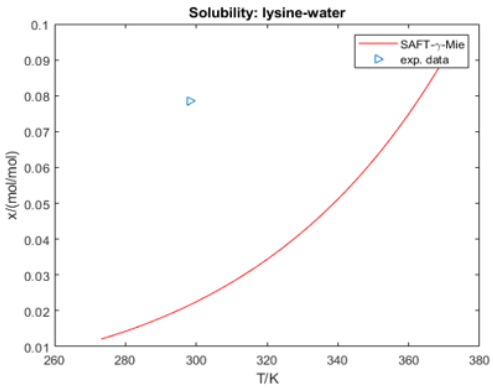
H=27690 J/ mol



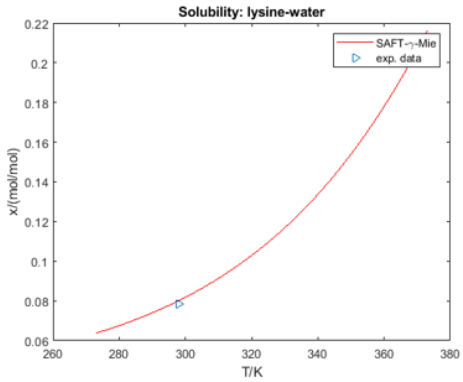
H=25690 J/ mol



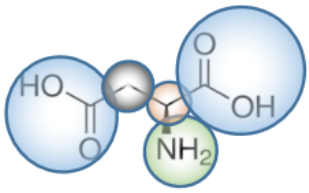
Lysine



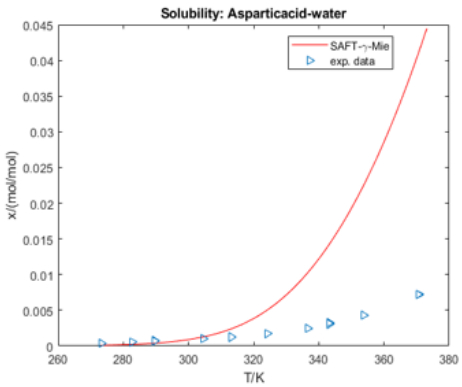
H=48266 J /mol



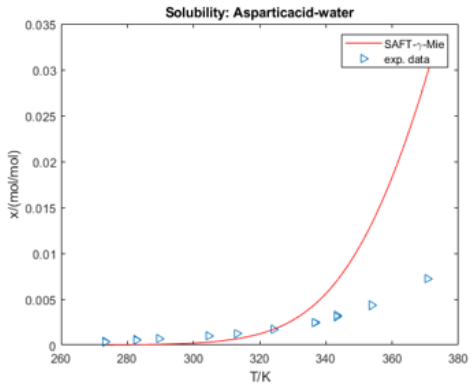
H=22566 J / mol



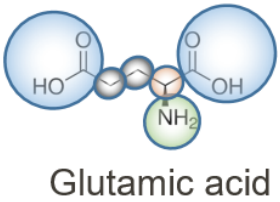
Aspartic acid



H=22861 J/ mol



H=70194 J/ mol



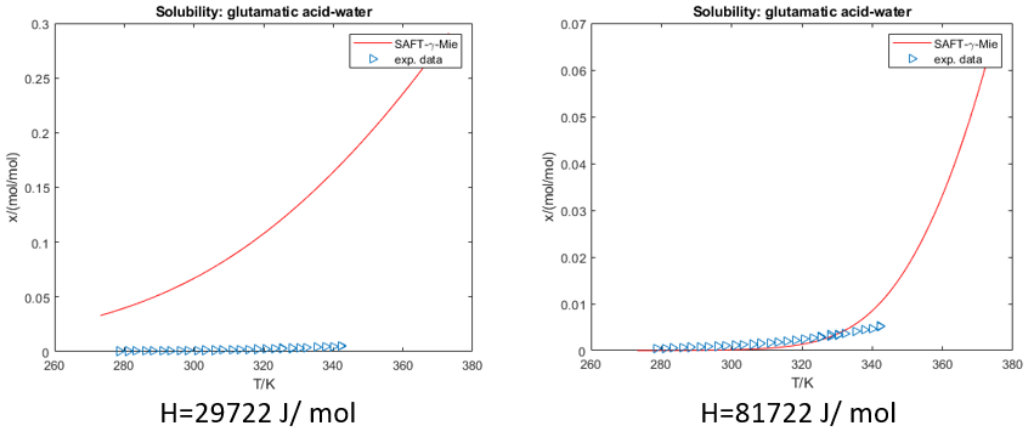


Figure 5.13 Predicted solubility data of glycine, alanine, valine, Leucine, Lysine, Aspartic acid and Glutamic acid with fusion enthalpy calculated by Gani method (Left) and adjusted by solubility data (Right).

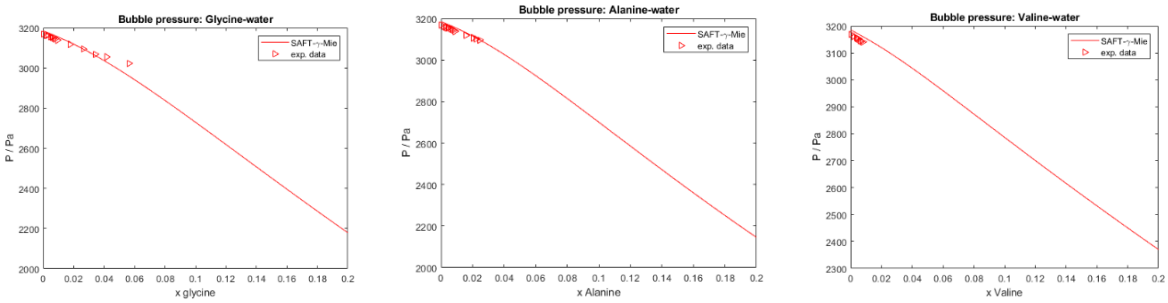


Figure 5.14 The comparison between experimental bubble pressure and predicted bubble pressure for glycine-water, alanine-water and valine-water

5.4.7 Only use bubble pressure of glycine-water, valine-water, alanine-water to fit parameters

In order to avoid the effect of fusion enthalpy and melting temperature during the fitting process, the parameters were fitted using only the bubble pressures of glycine-water, valine-water, and alanine-water. For the solubility prediction, the initial fusion enthalpy of glycine, valine and alanine were estimated by Gani group contribution method which listed in table 5.9, as well as the melting points reported in literature. The fitting parameters and the predicted

results can be seen in table 5.10 and figure 5.16, respectively. From figure 5.16, the bubble pressure has a good correlation but the solubility has a large error when compared with the experimental values. As a result, the fusion enthalpy was modified to new set values in order to reduce the inaccuracy.

Table 5.9 The fusion enthalpy calculated by Gani method and melting points (in literatures) of glycine, valine and alanine

	Glycine	Valine	Alanine
Fusion Enthalpy (J/mol)	28400	25051.0	22861
Melting point(K)	535.15	571.15	608

Table 5.10 The group-group parameters of NH_2-COOH using bubble pressure of glycine-water, valine-water and alanine-water

Objective Function	Epsilon	Cross Epsilon	Bonding Volume	Cross Epsilon	Bonding Volume	Cross Epsilon	Bonding Volume
	COOH – NH_2	COOH (e1) – NH_2 (H)	COOH (e1) – NH_2 (H)	COOH (e2) – NH_2 (H)	COOH (e2) – NH_2 (H)	COOH (H) – NH_2 (e)	COOH (H) – NH_2 (e)
1.294888	284.9384	969.1467	4.47E-28	1340.408	2.57E-28	1941.213	3.33E-28

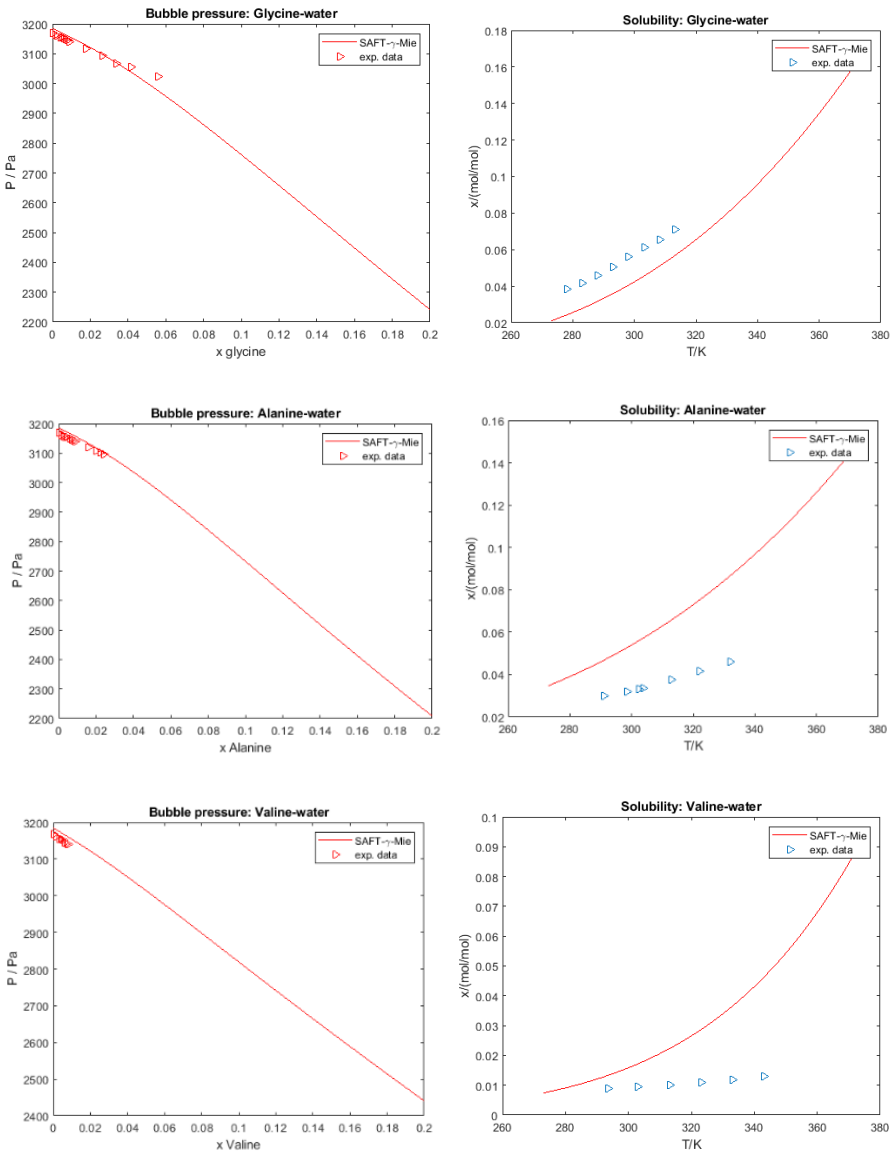


Figure 5.15 The experimental (points) and predicted(lines) bubble pressure and solubility of glycine, valine and alanine using fusion enthalpy estimated by Gani group contribution method

The adjusted fusion enthalpy were listed in table 5.11, and the new corresponding predicted solubility results can be seen in figure 5.17. The order of the goodness of the prediction is as following: glycine > alanine > valine, which corresponds to the order of the number of CH_3 group in the side chain. For glycine, there is no side chain but only a single hydrogen atom filling in for the beta carbon. For alanine, the side chain is one CH_3 group, while for valine, there are two CH_3 groups. This means more complicated the side chains, more difficult to get the accurate results.

Table 5.11 The adjusted fusion enthalpy and melting points (in literatures) of glycine, valine and alanine

	Glycine	Valine	Alanine
Fusion Enthalpy (J/mol)	25600	29051.0	27061
Melting point(K)	535.15	571.15	608

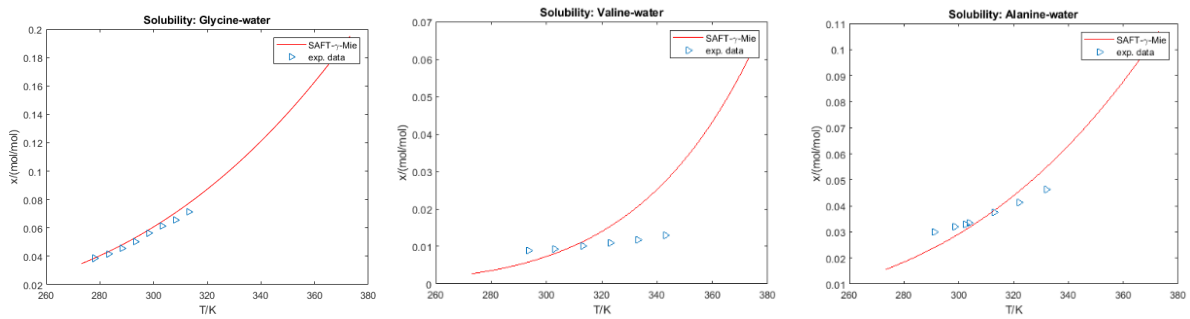


Figure 5.17 The experimental (points) and predicted (lines) solubility of glycine, valine and alanine using adjusted fusion enthalpy and melting temperature in literatures

The melting temperature measure by DSC in our lab were also used to fit parameters, combined the adjusted fusion enthalpy values (listed in table 5.12). Figure 5.18 illustrates the prediction results, indicating that there is still some discrepancy between the experimental and anticipated solubility. Thus, the fusion enthalpy of glycine, valine and alanine were adjusted based on the solubility again. The updated modified fusion enthalpy and the final anticipated outcomes are shown in Table 5.13 and Figure 5.19. In comparison to the values in Table 5.12, the fusion enthalpy increases as the melting temperature decreases.

Table 5.12 The adjusted fusion enthalpy and melting points (DSC) of glycine, valine and alanine

	Glycine	Valine	Alanine
Fusion Enthalpy (J/mol)	25600	29051.0	27061
Melting point(K)	525.7	513.15	526.22

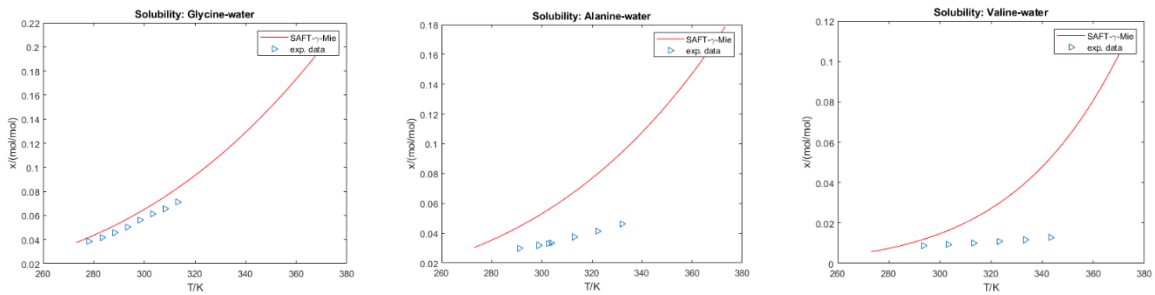


Figure 5.16 The experimental (points) and predicted (lines) solubility of glycine, valine and alanine using adjusted fusion enthalpy and melting temperature by DSC

Table 5.13 The new adjusted fusion enthalpy of glycine, valine and alanine

	Glycine	Valine	Alanine
Fusion Enthalpy (J/mol)	26600	32051	31061
Melting point(K)	525.7	513.15	526.22

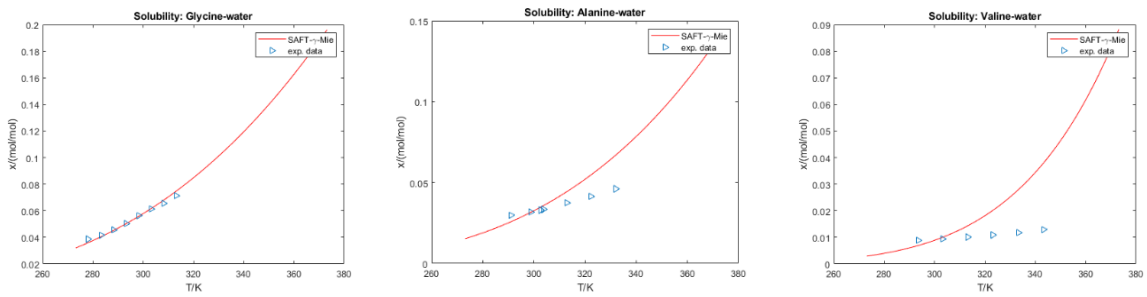


Figure 5.19 The experimental (points) and predicted (lines) solubility of glycine, valine and alanine using the new adjusted fusion enthalpy and melting temperature by DSC

5.4.8 Using the solubility and bubble pressure of glycine-water, valine-water, alanine-water to fit parameters

Based on the fusion enthalpy obtained from section 5.4.6, the solubility data of glycine,

valine and alanine were combined with the bubble pressure together to fit the parameters. The adjusted fusion enthalpy are listed in table 5.14. The parameters of NH_2-COOH and the predicted results can be seen in table 5.15 and figure 5.20. The AAD values of the predicted results can be seen in table 5.16. The low AAD values indicate that the parameters and fusion enthalpy are suitable for accurate prediction.

Table 5.14 The adjusted fusion enthalpy of glycine, valine and alanine using bubble pressure and solubility of glycine, valine and alanine as the input experimental data

	Glycine	Valine	Alanine
Fusion Enthalpy (J/mol)	22600	29051	27061
Melting point(K)	525.7	513.15	526.22

Table 5.15 Group-group parameters of NH_2-COOH using bubble pressure and solubilities of glycine-water, valine-water and alanine-water

Objective Function	Epsilon	Cross Epsilon	Bonding Volume	Cross Epsilon	Bonding Volume	Cross Epsilon	Bonding Volume
	COOH - NH2	COOH (e1) - NH2 (H)	COOH (e1) – NH2 (H)	COOH (e2) -NH2 (H)	COOH (e2) – NH2 (H)	COOH (H) -NH2 (e)	COOH (H)- NH2 (e)
119.1631	388.577	334.0828	1.35E-29	366.5358	9.99E-27	2054.621	9.97E-27

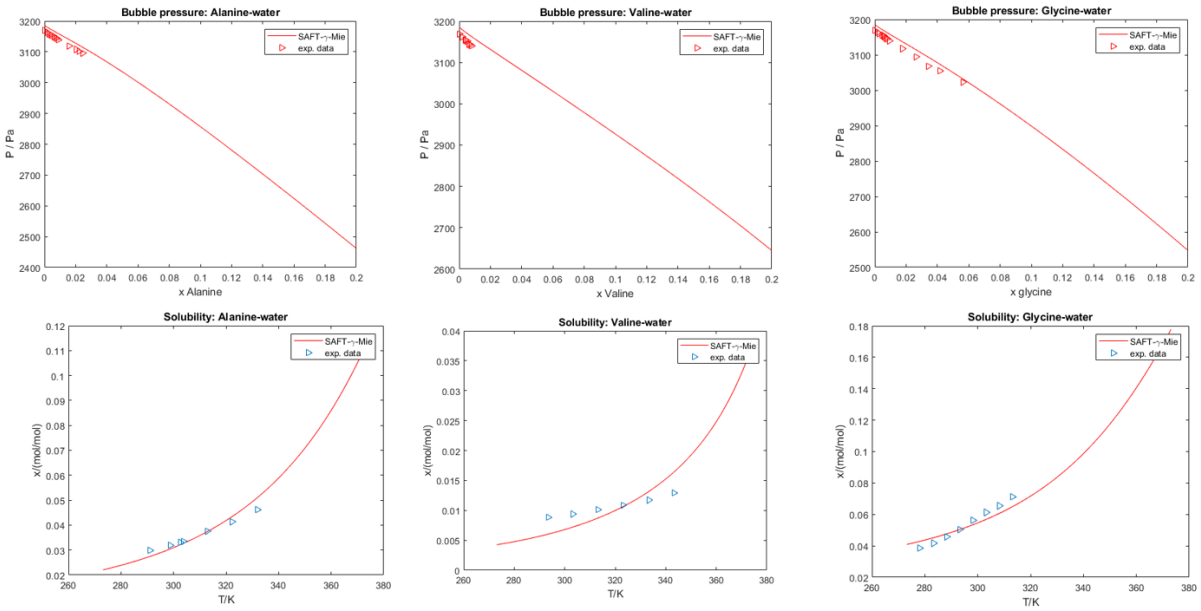


Figure 5.20 The experimental (points) and predicted (lines) bubble pressure and solubilities of glycine, valine and alanine using bubble pressure and solubility as the input experimental data.

Table 5.16 The AAD values for the predicted bubble pressure and solubility of glycine, valine and alanine

	Bubble Pressure			Solubility		
	Alanine-water	Glycine-water	Valine-water	Alanine-water	Glycine-water	Valine-water
AAD	0.511020548	0.491955652	0.605973249	2.928910876	11.04231869	14.4099816

5.4.9 Other parameters in future work

As shown in table 5.1, besides the parameters between NH_2 and COOH, there are other parameters should be fitted for the glycine homopeptides-water system. Because of the few experimental data of dipeptides and short chain peptides, the parameters between these groups have to be fitted by other experimental data from other structurally similar chemicals. Acetamide has C=O and NH_2 groups, which makes it as a potential chemical to fit the two sets parameters for C=O - NH_2 and C=O - H_2O. The Pyruvic acid has C=O and COOH groups, which can be used to fit the parameters between C=O and COOH groups. N-Methylacetamide has C=O and NH groups, which can be used to fit the parameters between C=O and NH. Figure

5.21 shows the parameters matrix and the structurally similar chemicals for glycine homopeptides parameters.

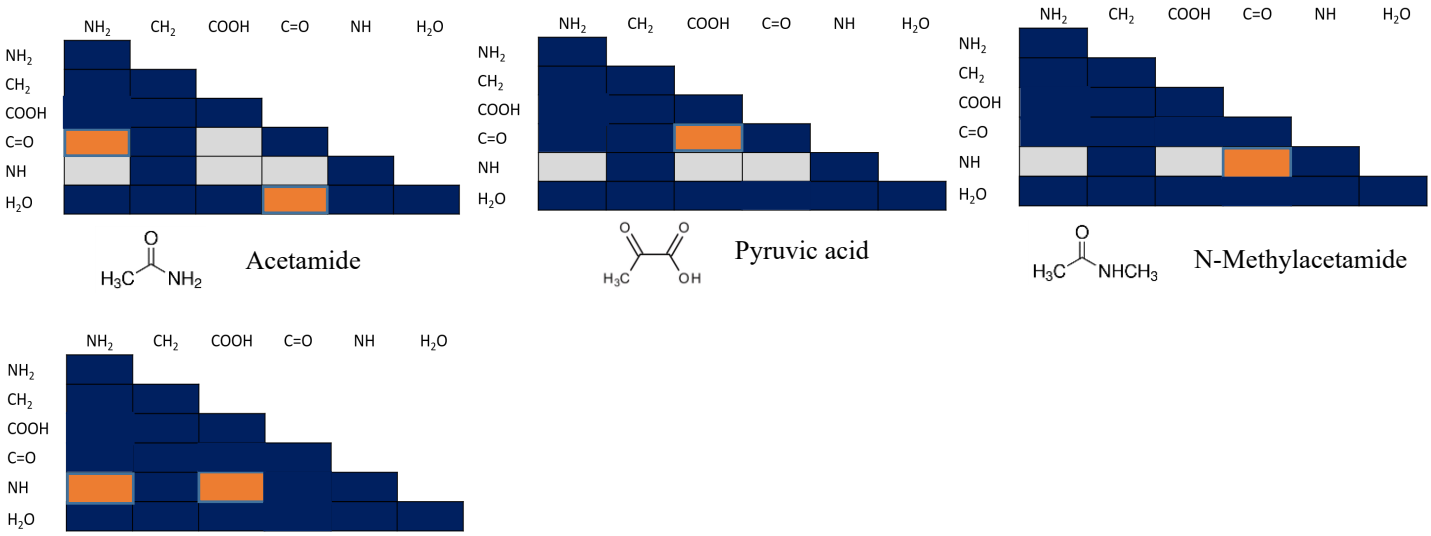


Figure 5.21 the parameters matrix and the structurally similar chemicals for glycine homopeptides parameters.

A) Parameters for C=O - NH_2 and C=O - H_2O

The pure component properties (heat capacity, vapour pressure, saturation liquid density) of Acetamide and the density of acetamide-water mixture were used as the input experimental data. The fitting parameters can be seen in table 5.17. The parameters were used to predict the bubble pressure and solubility of acetamide in water, which shown a good agreement with the experimental data (figure 5.22).

Table 5.17 Unlike group parameters for C=O-NH_2 and C=O-H_2O

Objective Function	Epsilon	Cross Epsilon	Bonding Volume	Epsilon	Cross Epsilon	Bonding Volume
	C=O –NH_2	C=O (e1) –NH_2 (H)	C=O (e1) –NH_2 (H)	C=O- H_2O	C=O(e1)-H_2O (H)	C=O (e1) –H_2O (H)
1.996399	100.878	2010.076	4.34E-29	213.7331	2023.083	9.18E-29

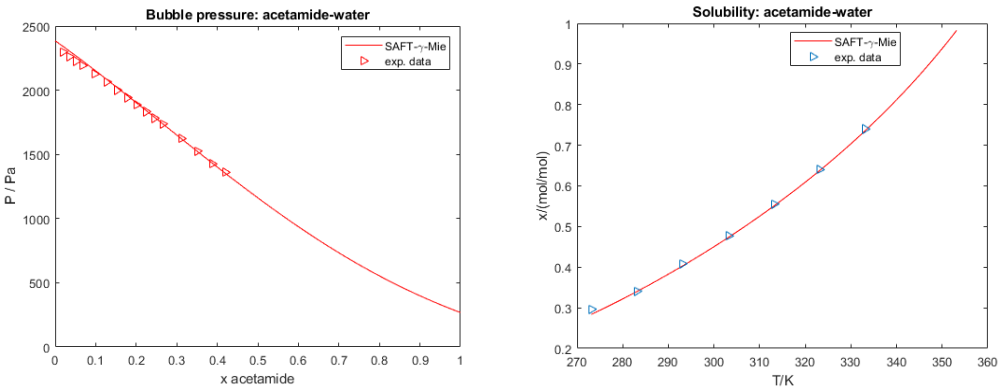


Figure 5.22 The experimental (points) and predicted (lines) bubble pressure and solubilities of acetamide in water

B) Parameters for C=O - COOH

The vapour pressure and saturation liquid density of pyruvic acid were used as the input experimental data. The fitting parameters and the predicted vapour pressure and saturation liquid density can be seen in table 5.18 and figure 5.23. The vapour pressure has a good prediction, but the saturation liquid density has a deviation compared with experimental data.

Table 5.18 Unlike group parameters for C=O-COOH

Objective Function	Epsilon	Lambda Repulsive	Cross Epsilon	Bonding Volume
	C=O - COOH	C=O - COOH	C=O (e1) - COOH (H)	C=O (e1) - COOH (H)
10.35796	63.57487	17.51772	2673.647	1.04E-28

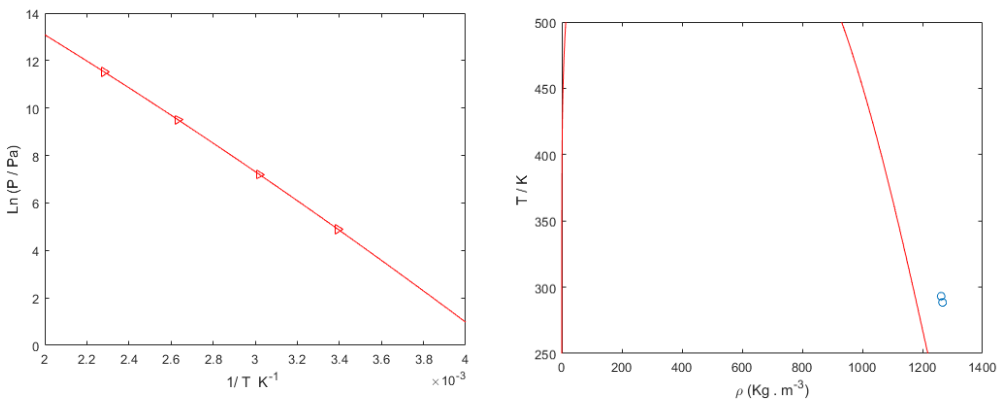


Figure 5.23 The experimental (points) and predicted (lines) vapour pressure and saturation liquid density of pyruvic acid in water

C) The parameters for C=O - NH.

The vapour pressure and saturation liquid density of N-Methylacetamide were used as the input experimental data. The fitting parameters and the predicted vapour pressure and saturation liquid density can be seen in table 5.19 and figure 5.24.

Table 5.19 Unlike group parameters for C=O-NH

Objective Function	Epsilon	Lambda Repulsive	Cross Epsilon	Bonding Volume
	C=O - NH	C=O - NH	C=O (e1) -NH (H)	C=O (e1) -NH (H)
16.7698	40	28.61472	2548.725	2.62E-28

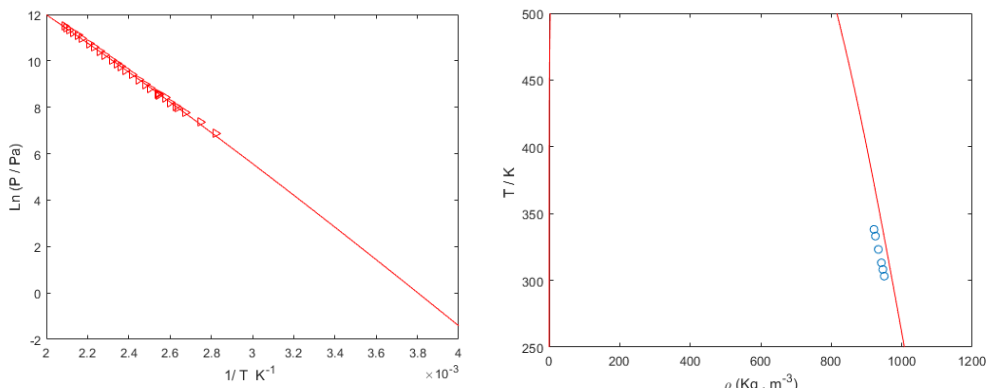


Figure 5.24 The experimental (points) and predicted (lines) vapour pressure and saturation liquid density of N-Methylacetamide in water

In summary, the above parameters fitted by the structurally similar chemicals can give some pure component properties (such as vapour pressure and saturation liquid density) a good prediction, but because of the lack of binary phase properties (such as solubility), the parameters in this section needs a further verification with more experimental data.

5.5 Conclusion

In this section, the group-group interaction matrix of the SAFT-γ Mie approach is extended for the prediction of the solubility of amino acids and peptides. The NH_2-COOH parameters are determined based on the thermodynamic properties of glycine, alanine and valine. These parameters can be used to predict the properties of other structurally similar amino acids (alanine, valine, aspartic acid, glutamic acid, isoleucine, leucine and lysine). Because of the decomposion during the melting process for peptides, the fusion enthalpy is difficult to be measured using DSC. The analysis of the impact of fusion enthalpy on the solubility prediction is also provided since the group contribution methods has to be applied to estimate the fusion enthalpy. For the longer chain peptides, more parameters still need to be fitted in the future. This research explores the application of SAFT-γ Mie to biomolecular thermodynamic properties, providing a platform to explore the relationship between their physicochemical properties and biostructures.

Chapter 6. Nucleation of glycine homopeptides in water: the effect of the peptide bonds on the nucleation

6.1 Abstract

In this PhD research work, triglycine was found to form a dihydrate with unfolded conformation (pPII) at low temperature (Chapter 7). In order to explore the peptide bonds and conformation effect on the nucleation of peptides, the primary nucleation induction time of glycine homopeptides in pure water at different supersaturation levels (1.4, 1.6, 1.8, 1.9, 1.95, 2.0, 2.2, 2.4) under different temperatures have been determined. At each condition, 100 experiments have been conducted in 2 mL scale to capture the statistics of the nucleation process. The data were analysed using the probability distribution function of the induction time within the framework of classical nucleation theory. Overall, the nucleation parameters (nucleation rate J, growth time t_g, interfacial energy γ, critical radius: r_C, number of molecules in critical nucleus n_C and activation Gibbs energy ΔG_C) of these glycine homopeptides have been calculated and compared. From the results, longer chain length will prolong the induction time, especially for the chain length which is longer than three, the nucleation will happen in several days. On the other hand, nuceation increased with an increase in the supersaturation for all homopeptides. Lower temperature makes the nucleation become more difficult and the induction time increases. However, for triglycine, the dihydrate form was produced during the nucleation process. The interfacial energy and activation Gibbs energy at low temperature are both lower than those at high temperature, while the induction time is longer, indicating the classical nucleation theory is not suitable to explain the nucleation phenomenon of triglycine

dihydrate. Moreover, the gelation and liquid-liquid separation of longer chain glycine homopeptides were observed, which provided insight into how the nucleation process evolves with increasing chain length, thereby offering a fundamental understanding of the complexed nucleation process for peptides.

6.2 Introduction

Nucleation is the first and essential step in the process of crystallisation, where a new thermodynamic phase forms with a lower free energy.[173] The kinetics of this phase transformation process is determined by the surface free energy of the emerging phase boundary.[174] Nucleation plays a decisive role in the control of the particle size distribution, polymorphism, perfection, and other characteristics of crystals. Therefore, the design of controlled process that ensures desired properties of a compound highly relies on the understanding of the fundamentals of nucleation. Generally, there are classical nucleation theory and non-classical nucleation theory summarized from current crystal nucleation research. Among them, the classical nucleation theory, remains the most common theoretical model for the understanding of nucleation, which prefers that concomitant density and order fluctuations cause the formation of clusters.[85,86] Then the clusters begin to aggregate to form the nucleus. The difference between these two nucleation theories was discussed in section 2.4.

Compared with small molecules and proteins, for short chain peptides, the nucleation mechanism is still not clear because of the flexible conformation. The solubility of glycine homopeptides (glycine, diglycine, triglycine, tetraglycine, pentaglycine and hexaglycine) were measured using UV-Vis at 190-260nm wavelength in Chapter 4. The result shows that solubility will decrease with the increase of the glycine residues and increase with the increasing temperature. In the work in Chapter 7, triglycine was found to form a dihydrate

under low temperature. Based on these thermodynamic and morphology research, the crystallisation process (cooling crystallisation) at different supersaturation levels and temperatures were researched to explore the peptide bonds, temperature, and hydration effect on the nucleation of peptide crystallisation.

6.3 Theory

6.3.1 Induction time t_d

Before a nucleus can be detected by instruments or human vision, it must have reached a certain size or be present in a large number, which means that the time t_n for a single nucleus to be formed cannot be accurately recorded. The period of time between the moment of a constant supersaturation created and the formation of crystals which can be detected is defined as "induction time" (t_d).[175] The nuclei have to grow to a detectable size in order to obtain measurements since it is impossible to detect the real induction time (t_n) when the critical nuclei form. Due to this, evaluating "induction time" is the only way. Induction time is larger than t_n, it cannot be considered a fundamental property of the system as its values depends on the technique employed to detect formation of the new phase. However, analysis of the values for induction time can help comprehend the mechanisms of new phase formation and growth from critical nuclei into crystals[176].

The induction time can be expressed as the sum of three terms as following:

$$t_d = t_{tr} + t_n + t_g \qquad (6.2)$$

where t_d is the measured induction time, t_{tr} is the time needed for reaching steady-state nucleation, t_n is the nucleation time, and t_g is the growth time required for critical nucleus to grow to a larger detected size.

By finding out the induction time, the nucleation kinetics can be calculated and used to

find the thermodynamic and kinetic parameters (such as interfacial energy, critical nucleation free energy, pre-exponential factor, and critical radius and number of molecules in the nucleus) required for efficient crystallisation design which will be beneficial for future works in crystallisation.

6.3.2 Determination of nucleation rate J

Due to the stochastic nature of crystallisation, the measured induction times of glycine homopeptides can be approximated to a cumulative probability distribution $P(t)$. For N isolated experiments, the probability $P(t)$ of observing an induction time between time zero and t is defined as following:

$$P(t) = \frac{n(t)}{N} \qquad (6.3)$$

where $n(t)$ is the number of trials for which crystals were detected at time *t*.

The experimentally determined cumulative probability distribution for the induction times was found by Jiang and ter Horst to resemble the Poisson Distribution[177]. They have indeed expressed the probability of finding at least one nucleus at time t_n as the Poisson distribution:

$$P(t_n) = 1 - \exp\ (-JVt_n) \qquad (6.4)$$

where *J* is the nucleation rate and *V* is the volume of solution.

Besides, the probability distribution for the detection time can thus be rewritten, on the model of the Poisson distribution, as:

$$P(t_d) = 1 - \exp\left(-JV\left(t_d - t_g\right)\right) \qquad (6.5)$$

6.3.3 Nucleation data derived from Classical Nucleation Theory

The Classical Nucleation Theory was used to determine the thermodynamic factor B from which nucleation data can be derived and compared for each glycine homopeptide.

The supersaturation S can be expressed as

$$S = \frac{c}{c^*} \tag{6.6}$$

where c is the actual concentration of the solution and c^* is the solubility at the specific temperature, the unit of all data is mole fraction.

The relation between the nucleation rate J, the induction time obtained experimentally, and supersaturation S could be expressed as the following equations based on different limiting steps: [178-180]

$$\ln J = lnA - \frac{B}{ln^2 S} \tag{6.7}$$

$$\ln(\frac{J}{S}) = lnA - \frac{B}{ln^2 S} \tag{6.8}$$

$$\ln(\frac{J}{Sln(S)}) = lnA - \frac{B}{ln^2 S} \tag{6.9}$$

$$\ln J + 3\ln(\ln S) = lnA - \frac{B}{ln^2 S} \tag{6.10}$$

where A and B are usually considered to be constants, and the exponent $B/\ln^2 S = W/kT$ is the dimensionless nucleation energy barrier for nucleation.

Equation 6.7 and 6.8 is for interface-transfer control, equation 6.9 is for volume-diffusion control, and equation 6.10 is another interface-transfer control expression according to the assumption that nucleus growth takes place by a surface nucleation mechanism. It also be verified in the following section 6.5.3 to show the best fitting performance for glycine homopeptides. Equation 6.10 was plotted for each poly-glycine in this work, and the slope of this linear regression provided an estimation of factor B, defined as:

$$B = \frac{16\pi v^2 \gamma^3}{3k^3 T^3} \tag{6.11}$$

where v is the volume of one molecule, γ represents the interfacial energy, k is the Boltzmann constant, and T is the absolute temperature in Kelvin.

Interfacial energy is the work required to generate a new interface between the

supersaturated solution and the solid phase contacting it.[181] Critical radius r_c is the critical radius corresponds to the minimum size at which a particle can survive in a solution without being redissolved.[182] Number of molecules in critical nucleus n_C is the number of molecules that must be included in the initial nucleus. Activation Gibbs energy ΔG_C is the energy barrier that the cluster needs to conquer to form a nucleus during homogeneous nucleation process.[171]

From the thermodynamic factor B, the interfacial energy γ and the following nucleation parameters can be calculated. [180]

$$\text{Interfacial energy } \gamma = \left(\frac{3k^3T^3B}{16\pi v^2}\right)^{1/3} \quad (6.12)$$

$$\text{Critical radius: } r_C = \frac{2v\gamma}{RTlnS} \quad (6.13)$$

$$\text{Number of molecules in critical nucleus } n_C = \frac{4\pi r_C^3}{3v} \quad (6.14)$$

$$\text{Activation Gibbs energy } \Delta G_C = \frac{16\pi\gamma^3 v^2}{3k^2T^2(\ln S)^2} \quad (6.15)$$

where v is the volume of one molecule, S is the supersaturation, γ represents the interfacial energy, k is the Boltzmann constant, and T is the absolute temperature in Kelvin.

6.4 Experiments

6.4.1 Materials

Glycine homopeptides (glycine, diglycine, triglycine, tetraglycine and pentaglycine) were supplied by Sigma-Aldrich Company Ltd. (Figure 6.1). Deionized water was produced in laboratory. All chemicals were used without further purification. The detailed information about the chemicals used in this work can be seen in Table 6.1.

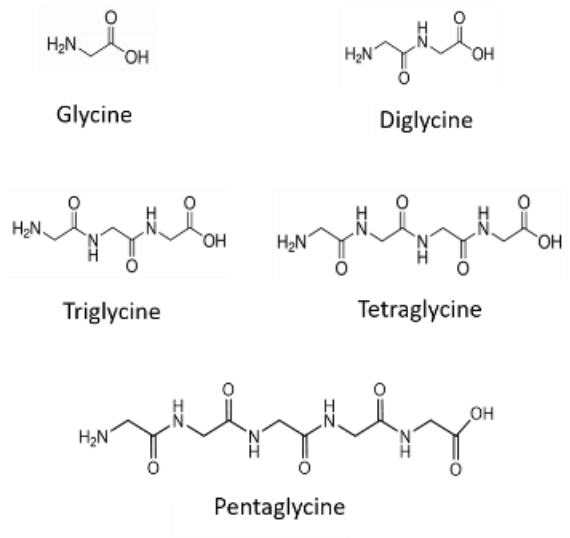


Figure 6.1 Chemical structures of Glycine homopeptides

Table 6.1 Properties of materials used in this work.

Material	Molar Mass ($g \cdot mol^{-1}$)	Mass Fraction Purity	Source
Glycine	75.07	≥0.990	Sigma-Aldrich Company Ltd.
Diglycine	132.12	≥0.990	Sigma-Aldrich Company Ltd.
Triglycine	189.17	≥0.990	Sigma-Aldrich Company Ltd.
Tetraglycine	246.22	≥0.990	Sigma-Aldrich Company Ltd.
Pentaglycine	303.27	≥0.990	Sigma-Aldrich Company Ltd.

6.4.2 Induction time measurement

The induction times of glycine homopeptides crystallisation were determined in deionized water at 278.15K and 283.15 K. Solutions with different supersaturation levels (listed in Table 6.2) were prepared for each glycine homopeptide in test tubes. For glycine and diglycine, 2 mL solutions were made with 2 g of deionized water, whereas a volume of 1.5 mL was preferred for triglycine due to the increased feasibility of nucleation. The tubes were equipped with a small magnetic stir bar and meticulously sealed with rubber lids wrapped with parafilm both

inside and outside the cap. The samples were placed into the hot thermostat bath maintained at 333.15 K, well above the supersaturation temperature of glycine homopeptides. The solutions were stirred at 500 rpm in the water bath until all the glycine homopeptides had fully dissolved (Figure 6.2). The clear tubes were then immersed in a cold thermostat bath held at the low constant nucleation temperature (278.15K and 283.15K). During the experiment, the solutions of different supersaturations for each peptide were tested in parallel and continuously magnetically stirred at 250 rpm. The induction time was recorded as the time of the first observation of the solution becoming cloudy. Once the solutions had nucleated, the tubes were transferred back to the water bath held at a temperature above the supersaturation temperature, to dissolve before repeating the nucleation experiment cycle displayed in Figure 6.3. 100 induction time data were obtained for each glycine homopeptide and each supersaturation level to capture the stochastic nature of nucleation.

After all the induction time measurements were performed, the solutions were filtrated, and the powders were tested with PXRD to determine the morphology of the crystallised glycine homopeptides to make sure there is no polymorphism transformation during the nucleation process.

Table 6.2 The experimental conditions of the nucleation measurement of glycine, diglycine and triglycine.

S	GLYCINE		DIGLYCINE		TRIGLYCINE	
	5°C	10°C	5°C	10°C	5°C	10°C
1.4						
1.6						
1.8						
1.9						
1.95						
2.0						
2.2						
2.3						

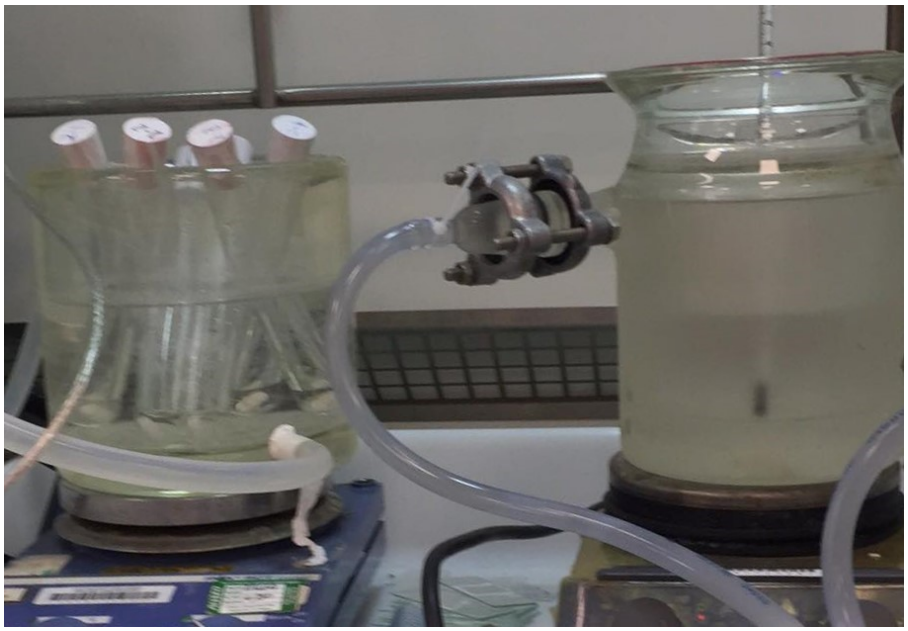


Figure 6.2 Experimental set up for induction time measurement in the lab

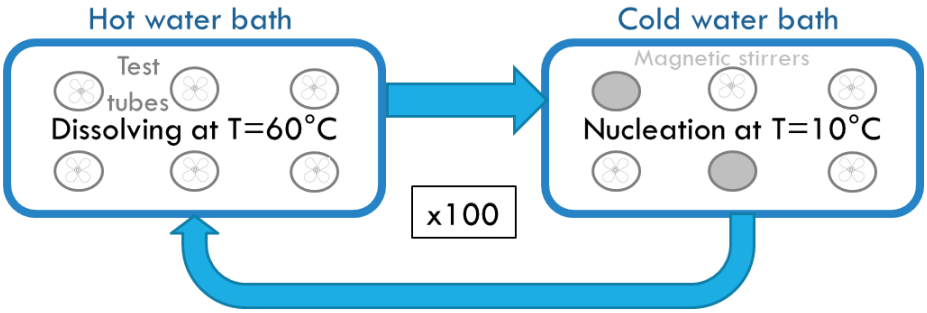


Figure 6.3 Schematic representation of the nucleation experiment set up

6.5 Results and discussion

6.5.1 The characterisation of glycine homopeptides

The PXRD patterns of glycine homopeptides used in this work have been tested before and after the induction time measurement to make sure that there was no phase transition. As the examples shown in figure 6.4, all the samples kept the same XRD pattern, except triglycine. The new XRD patterns of triglycine comes from the dihydrate morphology based on the further characterisation (Chapter 7). From the research, the dihydrate form of triglycine can be obtained when the temperature is lower than 298.15 K.[164]

6.5.2 Induction time measurement for glycine homopeptides (from mono-glycine to penta-glycine)

The induction time of glycine, diglycine, triglycine, tetraglycine and pentaglycine were measured under 278.15K at different supersaturation level S (1.5, 2, 3) in 2 mL solutions. The graph of the crystal products captured by microscope and SEM can be seen in Figure 6.5. From the graph, the crystals of glycine homopeptides are all have regular shapes – rodlike for glycine, needle for triglycine dehydrate, plate for di, tetra-, penta- and hexa-glycine. The solution of pentaglycine and hexaglycine became cloudy first until there were small crystals came out, the induction time were longer than one week.

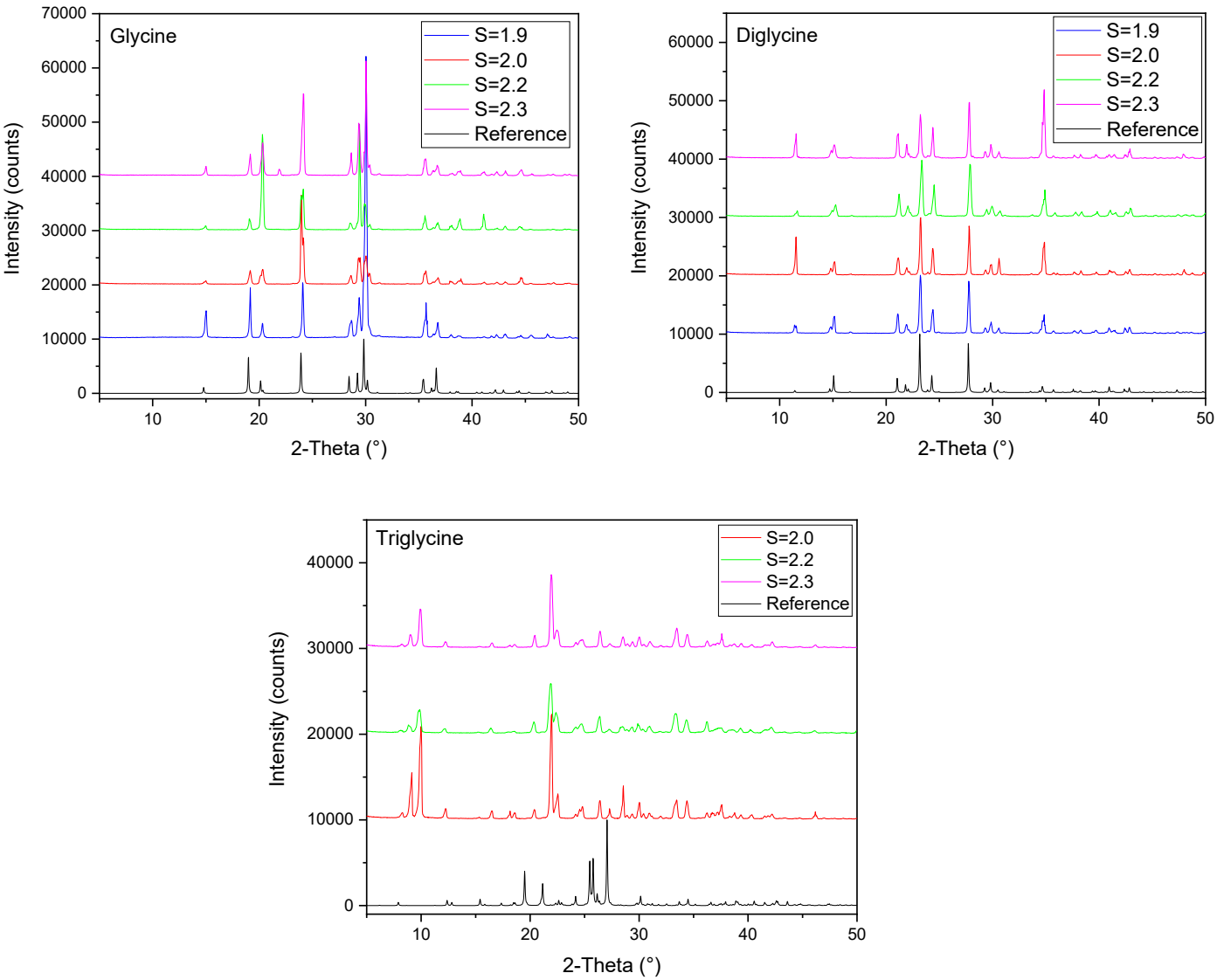


Figure 6.4 The XRD patterns of glycine homopeptides before and after induction time measurement

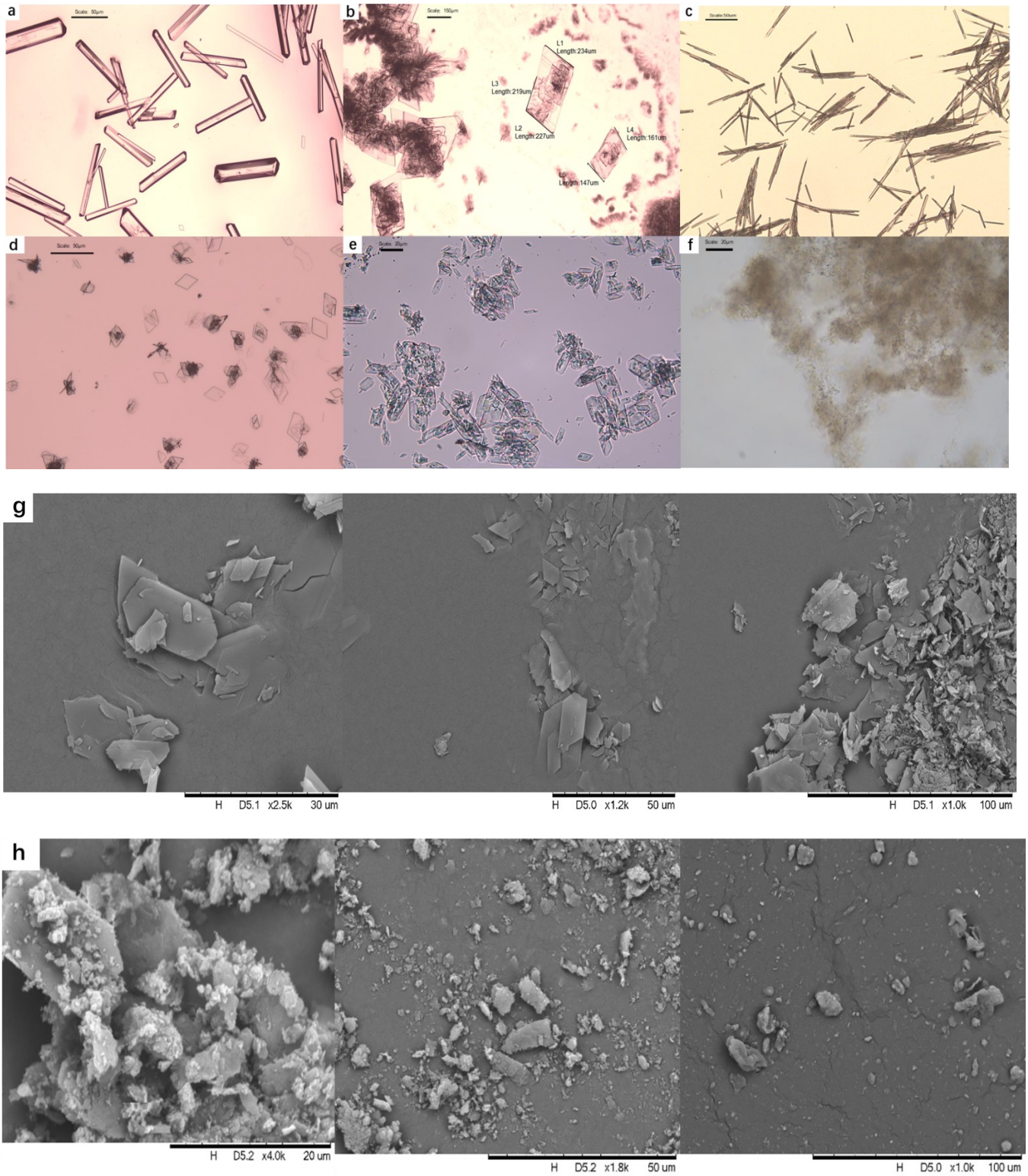


Figure 6.5 The microscope images (a. glycine, b.diglycine, c.triglycine dihydrate, d.tetraglycine, e.pentaglycine, f.hexaglycine.) and SEM images (g. pentaglycine and h. hexaglycine) of glycine homopeptides

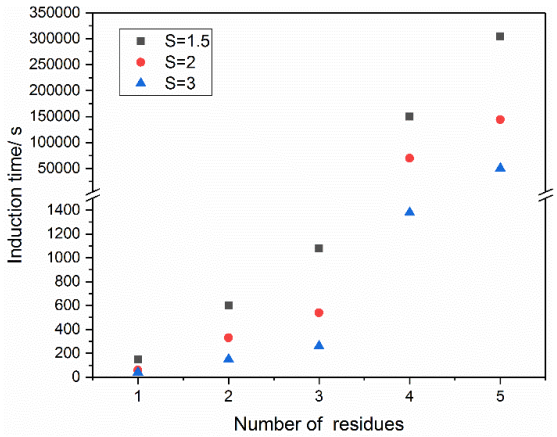


Figure 6.6 Induction time of glycine homopeptides under different supersaturation levels at 278.15K.

The experiment results shown that the induction time increased with the increasing number of the amino acid residues (Figure 6.6). Glycine has the shortest induction time and pentaglycine has the longest induction time compared with others. Moreover, when the number of amino acids in glycine homopeptides exceeds three, the induction time increases by an order of magnitude, which also indicates that when the chain length is long enough, there is no linear relationship between number of peptide bonds and nucleation rate. Moreover, the induction time decreased with the increase of the supersaturation level S.

6.5.3 Nucleation parameters of glycine, diglycine and triglycine dihydrate

In order to explore the chain length and the hydration effect on the nucleation process. The nucleation parameters (nucleation rate J, growth time t_g, interfacial energy γ, critical radius: r_C, number of molecules in critical nucleus n_C and activation Gibbs energy ΔG_C) of glycine, diglycine and triglycine dihydrate were calculated based on the stochastic nature of nucleation in small scale volumes and compared with each other. The distribution of induction time is independent of the repetitions of 100 measurements for glycine homopeptides, which are shown in figure 6.7-figure 6.10. The exponential-based Poisson function proposed by Jiang and ter Horst (equation 6.4) was used to correlate the experimental data. The results can be

seen in Table 6.3. According to section 6.3, the relationship between the nucleation rate and the supersaturation can be described by equations 6.7-6.10 for different rate-limiting steps in the nucleation process. In order to determine which one fit the experimental data best, the fitting goodness of these four functions were compared, which can be seen in figure 6.11. Equation 6.10 provides the best fitting results (the lowest R^2 value), which indicates that the nucleation rate for glycine homopeptides in water is controlled by the interface-transfer condition, and the nucleus growth takes place by a surface nucleation mechanism.

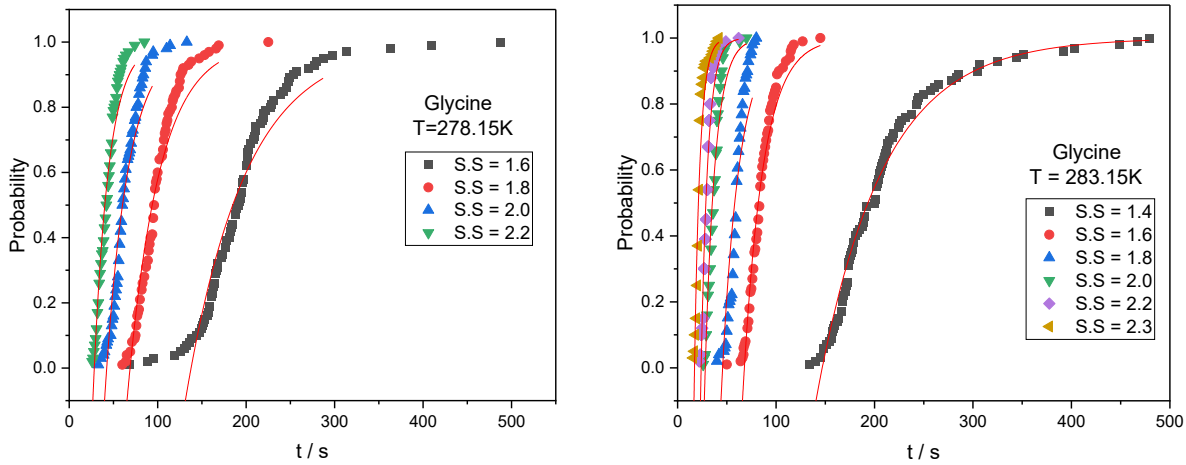


Figure 6.7 Induction time of glycine under different supersaturation levels at 278.15K and 283.15K. Solid lines represent the fit of the Poisson distribution.

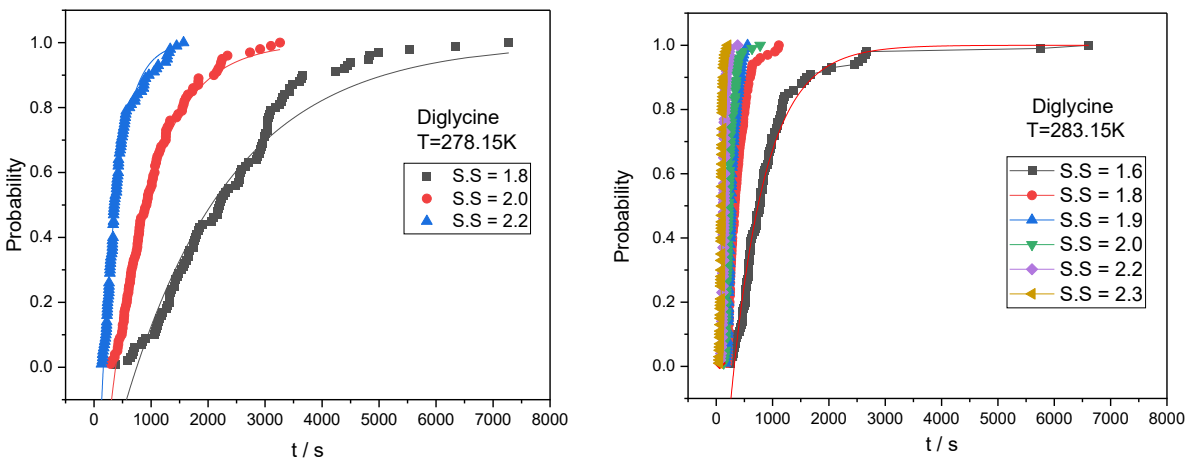


Figure 6.8 Induction time of diglycine under different supersaturation levels at 278.15K and 283.15K. Solid lines represent the fit of the Poisson distribution.

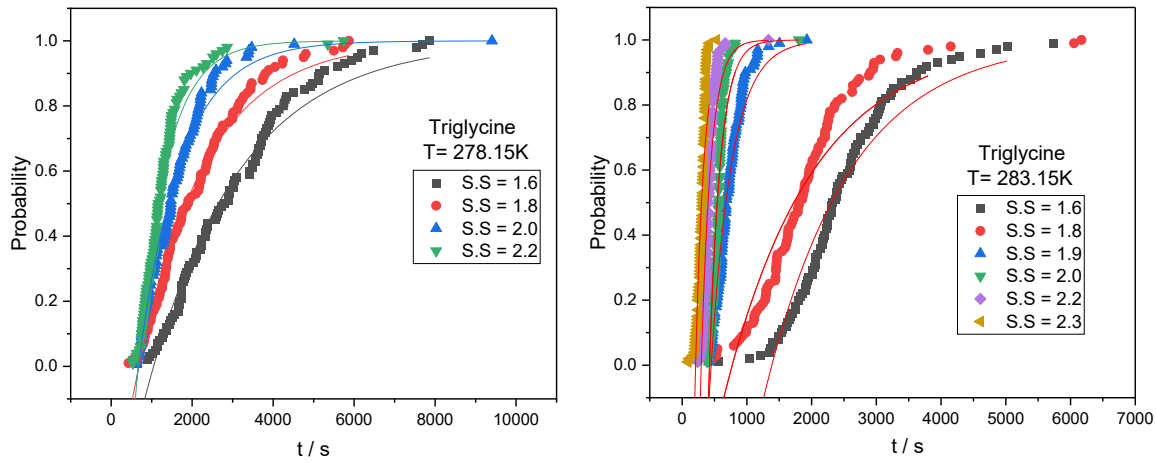


Figure 6.9 Induction time of triglycine dihydrate under different supersaturation levels at 278.15K and 283.15K. Solid lines represent the fit of the Poisson distribution.

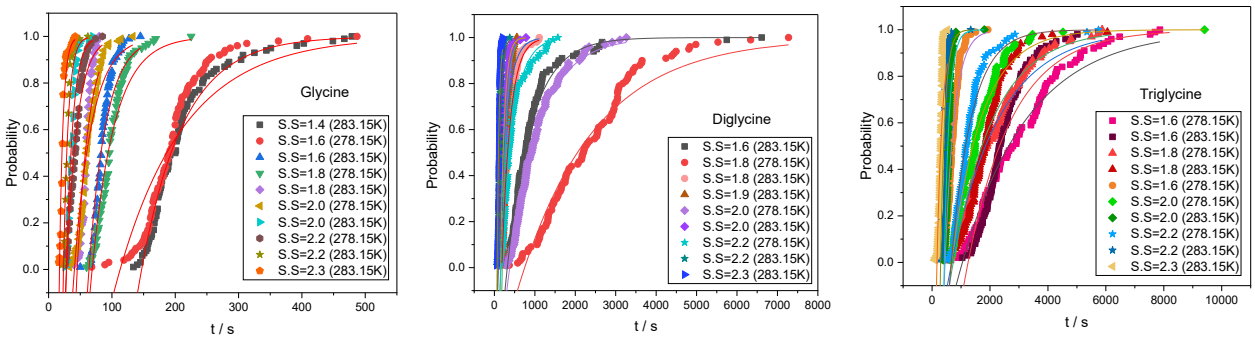


Figure 6.10 Comparison of Induction time of glycine homopeptides under different temperatures. Solid lines represent the fit of the Poisson distribution.

Table 6.3 The calculated nucleation kinetics for glycine homopeptides at 278.15 K and 283.15 K

T/°C	T/K	V/cm³	S	R²	J /m⁻³ s⁻¹	t_g (s)	1/(lnS)²	lnJ	ln(J/S)	ln(J/SlnS)	lnJ + 3ln(lnS)
						Glycine					
10	283.15	2	1.4	0.9778	74700	146.8992	8.8329	8.9187	8.5822	9.6714	5.6509
10	283.15	2	1.6	0.9621	24600	67.8182	4.5269	10.1105	9.6405	10.3955	7.8455
10	283.15	2	1.8	0.8238	29690	46.4636	2.8944	10.2986	9.7108	10.2422	8.7044
10	283.15	2	2.0	0.9399	48670	28.1022	2.0814	10.7928	10.0997	10.4662	9.6933
10	283.15	2	2.2	0.9334	72760	23.9071	1.6086	11.1949	10.4065	10.6441	10.4819
10	283.15	2	2.3	0.8599	88130	17.1760	1.4415	11.3866	10.5537	10.7365	10.8381
5	278.15	2	1.6	0.9275	70150	135.3543	4.5269	8.8558	8.3858	9.1408	6.5908
5	278.15	2	1.8	0.9344	13845	68.7710	2.8944	9.5358	8.9479	9.4793	7.9415
5	278.15	2	2	0.9222	19720	42.4799	2.0814	9.8894	9.1962	9.5628	8.7898
5	278.15	2	2.2	0.9564	29225	28.3442	1.6086	10.2828	9.4943	9.7320	9.5697
						Diglycine					
10	283.15	2	1.6	0.9795	8200	319.1015	4.5269	6.7093	6.2393	6.9943	4.4443
10	283.15	2	1.8	0.9972	2655	196.0230	2.8944	7.8842	7.2964	7.8278	6.2900
10	283.15	2	1.9	0.9928	5215	196.9189	2.4273	8.5593	7.9174	8.3608	7.2291
10	283.15	2	2.0	0.9765	5385	151.3666	2.0814	8.5914	7.8982	8.2647	7.4918
10	283.15	2	2.2	0.9153	9990	95.8312	1.6086	9.2093	8.4209	8.6586	8.4963
10	283.15	2	2.3	0.89660	12795	72.8940	1.4415	9.4568	8.6239	8.8067	8.9083
5	278.15	2	1.8	0.95857	265.3485	748.1490	2.8944	5.5810	4.9936	5.5246	3.9869
5	278.15	2	2.0	0.99076	670	378.2064	2.0814	6.5073	5.8141	6.1806	5.4077
5	278.15	2	2.2	0.98397	1680	162.4202	1.6086	7.4265	6.6381	6.8758	6.7135
						Triglycine dihydrate					
10	283.15	1.5	1.6	0.8862	425.8047	1243.1838	4.5269	6.0540	5.5840	6.3390	3.7890
10	283.15	1.5	1.8	0.8953	480.2700	758.4529	2.8944	6.1743	5.5866	6.1180	4.5802
10	283.15	1.5	1.9	0.9672	2046.6667	448.3031	2.4273	7.6240	6.9821	7.4255	6.2938
10	283.15	1.5	2.0	0.8964	3893.3333	423.2787	2.0814	8.2670	7.5739	7.9404	7.1675
10	283.15	1.5	2.2	0.8897	4426.6667	295.5602	1.6086	8.3954	7.6069	7.8446	7.6824
10	283.15	1.5	2.3	0.8221	5106.6667	208.3258	1.4415	8.5383	7.7054	7.8882	7.9898
5	278.15	1.5	1.6	0.9770	289.4140	1047.4838	4.5269	5.6679	5.1979	5.9529	3.4028
5	278.15	1.5	1.8	0.9843	398.7593	686.0846	2.8944	5.9884	5.4057	5.9320	4.3942
5	278.15	1.5	2.0	0.9860	634.7140	674.9353	2.08141	6.4532	5.7600	6.1265	5.3536
5	278.15	1.5	2.2	0.9669	946.6667	665.6701	1.6086	6.8529	6.0645	6.3022	6.1399

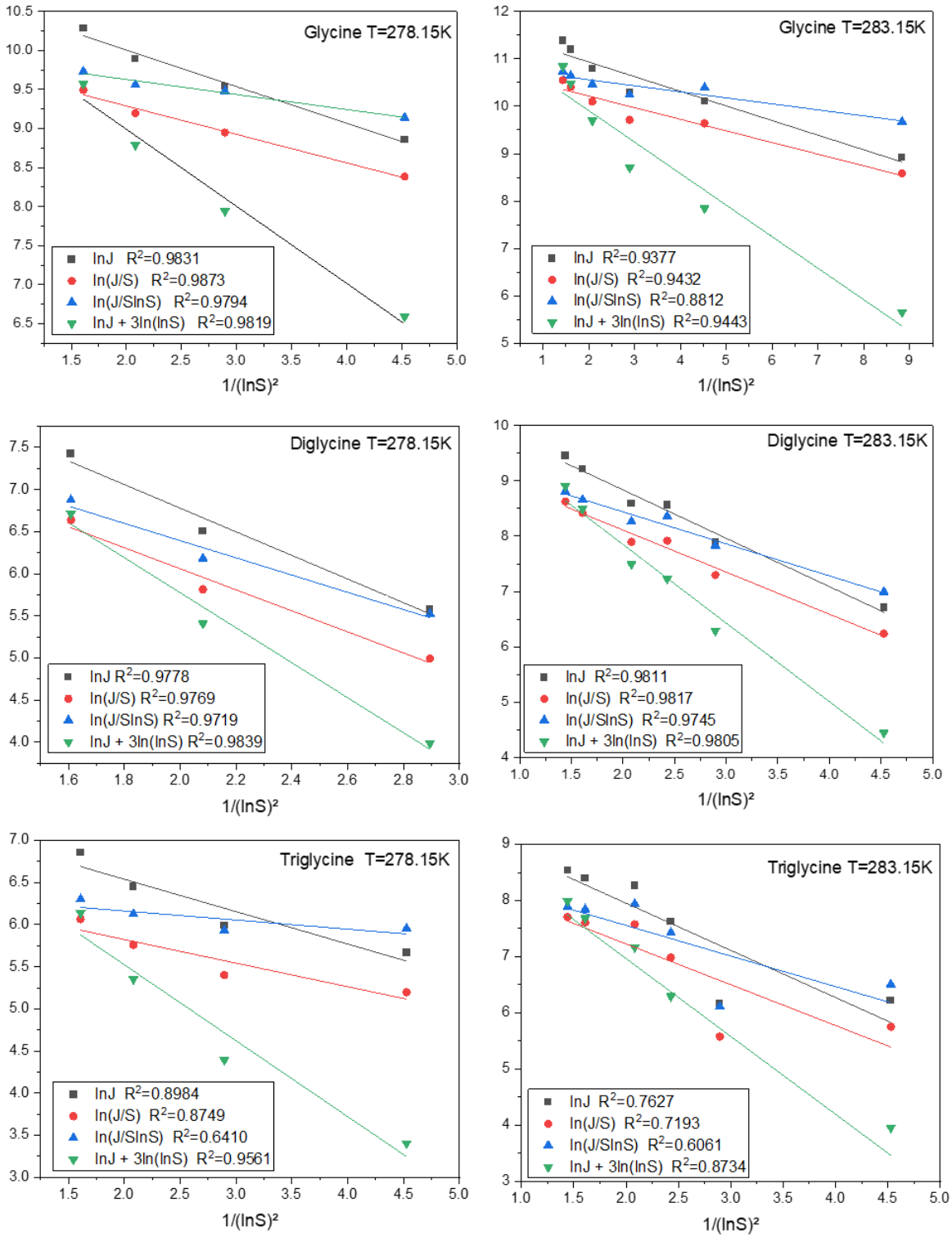


Figure 6.11 Correlation of nucleation rate with the supersaturation of glycine homopeptides under different temperatures.

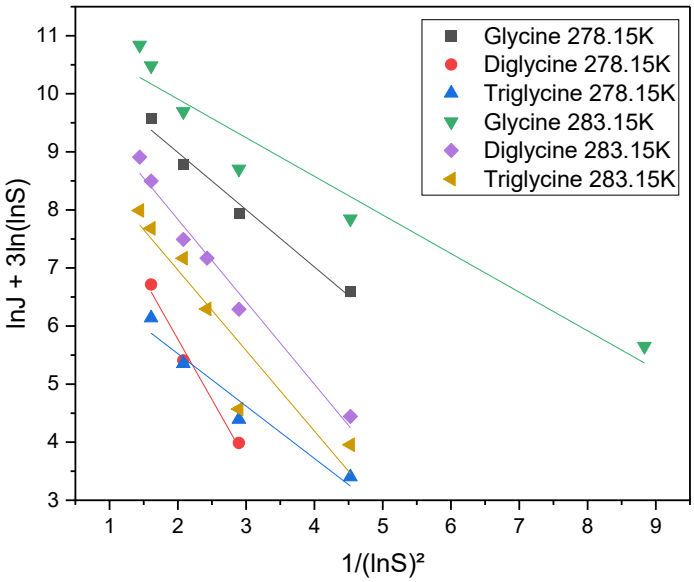


Figure 6.12 The comparison of the relationships between nucleation rate and supersaturation level of glycine homopeptides under different temperatures

The relationship between $1/(\ln S)^2$ and $\ln J + 3\ln(\ln S)$ of glycine homopeptides in equation (6.10) were shown in Figure 6.12. The linear fits in Figure 6.12, all have a high R^2 value which already shown in Figure 6.11, indicate high accuracy of results. From the intercept ln A and the slope B of each linear line, the pre-exponential factor and the interfacial energy can be determined, respectively. Under same temperature, the steeper the slope, the larger the B, the higher value of $\upsilon^2\gamma^3$ and the activation Gibbs energy become higher, then the induction time should be longer. From the interfacial energy, the free-energy barrier to nucleation ΔG_C, the size of the critical nucleus r_c, and the number of molecules n_c in the critical nucleus can be determined by equation 6.11-6.13. The values were listed in Table 6.4 and Table 6.5.

Table 6.4 Values obtained from linear fit and calculated nucleation parameters for glycine homopeptides at 283.15K

Molecular weight (g/mol)	Density (g/m^3)	v one molecule	S	A	B	γ /$mJ.m^{-2}$	r_c /nm	n_c	△Gc /$kJmol^{-1}$
				Glycine					
75.07	1509550	8.26E-29	1.4	76114.952	0.665	7.031	0.883	34.934	13.823
			1.6				0.632	12.817	7.084
			1.8				0.506	6.553	4.530
			2.0				0.429	3.996	3.257
			2.2				0.377	2.715	2.517
			2.3				0.357	2.303	2.256
				Diglycine					
132.12	1477200	1.48571E-28	1.6	43914.507	1.421	6.122	0.990	27.366	15.126
			1.8				0.792	13.991	9.671
			1.9				0.725	10.745	8.110
			2.0				0.671	8.532	6.955
			2.2				0.590	5.797	5.375
			2.3				0.559	4.917	4.816
				Triglycine dihydrate					
225.17	1518350	2.46344E-28	1.6	18839.293	1.4392	4.389	1.177	27.726	15.325
			1.8				0.941	14.176	9.798
			1.9				0.862	10.887	8.217
			2.0				0.798	8.644	7.046
			2.2				0.702	5.873	5.446
			2.3				0.664	4.982	4.880

Table 6.5 Values obtained from linear fit and calculated nucleation kinetics for glycine homopeptides at 278.15K

Molecular weight (g/mol)	Density (g/m^3)	v one molecule	S	A	B	$\gamma/mJ.m^{-2}$	r_c /nm	n_c	△Gc /$kJmol^{-1}$
				Glycine					
75.07	1509550	8.26081E-29	1.6	57699.28	0.986	7.876	0.721	18.999	10.316
			1.8				0.576	9.714	6.596
			2.0				0.489	5.923	4.743
			2.2				0.430	4.024	3.666
				Diglycine					
132.12	1477200	1.48571E-28	1.8	20625.84	2.080	6.952	0.899	20.486	14.161
			2.0				0.762	12.492	10.183
			2.2				0.670	8.488	7.870
				Triglycine dihydrate					
225.17	1518350	2.46E-28	1.6	1523.552	0.902	3.690	1.007	17.381	9.437
			1.8				0.805	8.886	6.0340
			2.0				0.683	5.419	4.3390
			2.2				0.600	3.682	3.3534

All in all, according to the results obtained in tables 6.4 and 6.5, it can be seen that as the supersaturation level increases, the Gibbs free energy, critical radius and number of molecules in one nucleus decreases. This relationship was seen across the different results of the same peptide in both temperatures of different chain length. This is potentially due to the amount of energy required to form the earliest nucleus of critical size is lesser at higher supersaturation levels hence, induction time would be shorter.

The comparison of nucleation parameters between different peptides is shown in Figure 6.13-6.15. When crystals form quickly, the nucleation rate should be higher. From figure 6.13, crystals nucleate more rapidly with a shorter chain length; thus, glycine, which contains only one amino acid residue, nucleates the quickest and has the highest J, while triglycine dihydrate,

which has three amino acid residues, nucleates the slowest and has the lowest J. Nucleation rate at 283.15K is greater than it is at 278.15K, which could be because the molecules are more active at a higher temperature, resulting in a faster rate of surface integration and hence a faster growth rate.

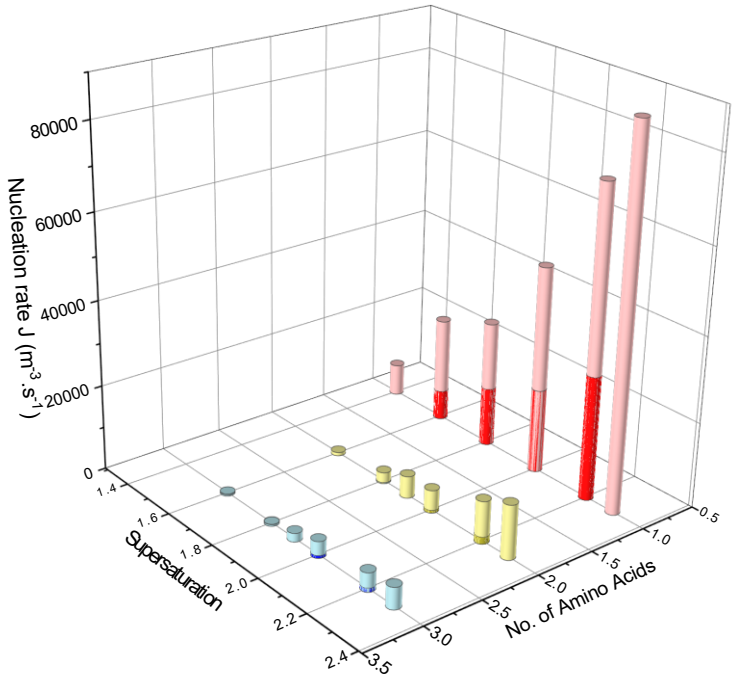


Figure 6.13 The relationship between nucleation rate J, supersaturation levels and the number of glycine residues. The darker colour are the peptides at 278.15K, whereas the lighter colour are peptides at 283.15K.

According to literature, a greater difficulty in nucleation equals to a higher solid-liquid interfacial energy.[181] The order of the interfacial energy of glycine homopeptides in water under the same temperature is as following: glycine > diglycine > triglycine dihydrate, triglycine dihydrate which is supposed to have the highest γ as an increase in the number of amino acid residues resulting in an increase in the difficulty of nucleation however results do not align. From the definition of interfacial energy, it is proportional to thermodynamic factor B and

inversely proportional to the volume of one peptide molecule. Since the order of volume is glycine < diglycine < triglycine dihydrate, the order of B is diglycine > glycine > triglycine dihydrate at 278.15K, and diglycine > triglycine dihydrate > glycine at 283.15 K, so triglycine can adopt the lowest interfacial energy compared with glycine and diglycine. The interfacial energy of various peptides at the same temperature is incomparable. The influence of temperature on the interfacial energy, on the other hand, can be investigated. When the temperature is increased, all interfacial energy become lower, except for triglycine dihydrate, which shows a higher interfacial energy with temperature increasing during the nucleation process. This means interfacial energy of triglycine dihydrate cannot be used as the only evidence whether the nucleation rate is high or not. Triglycine adopts a different conformation which shows an unfolded state in crystals and solutions and will transfer to the folded conformation when the temperature become higher. The details will be discussed in Chapter 7. In summary, the variable conformation under different temperature made the classical nucleation theory become unsuitable for triglycine dihydrate.

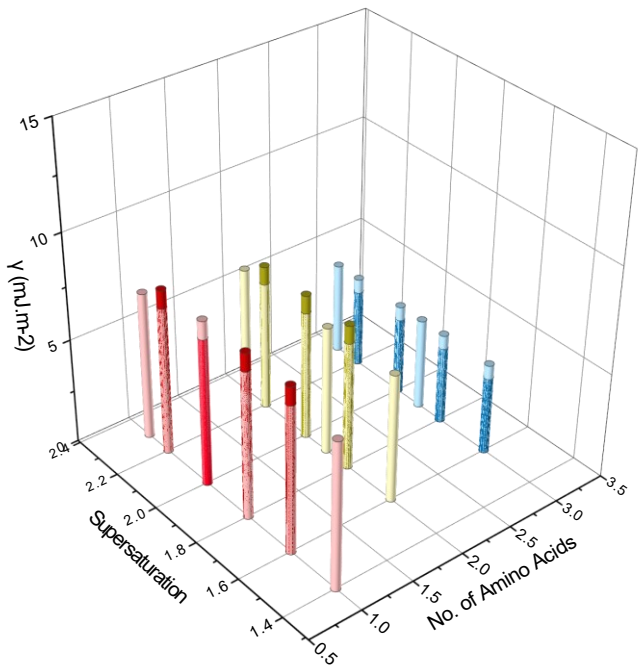


Figure 6.14 The relationship between interfacial energy γ, supersaturation levels and the number of glycine residues. The darker colour are the peptides at 278.15K, whereas the lighter colour are peptides at 283.15K.

From figure 6.15, for all peptides and temperatures, as supersaturation increases, ΔG_C decreases, indicating that higher driving force accelerated nucleation. When investigating the effect of chain length on activation energy, the energy at the same supersaturation level and temperature were compared. When the temperature is 283.15K, triglycine dihydrate has the largest ΔG_C and longest induction time, which aligns with the trend of activation energy. When the temperature is lower as 278.15K, triglycine has the lowest activation energy compared with mono- and diglycine, which is contradictory with its longest induction time.

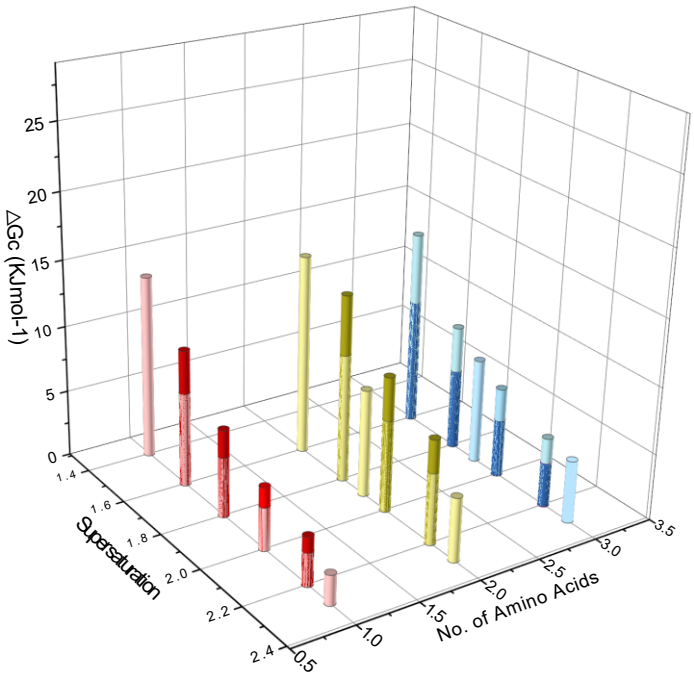


Figure 6.15 The relationship between activation energy ΔG_C , supersaturation levels and the number of glycine residues. The darker colour are the peptides at 278.15K, whereas the lighter colour are peptides at 283.15K.

As it has been already mentioned before, under same temperature, the steeper the slope, the larger the B, the higher value of $\upsilon^2\gamma^3$ and the activation Gibbs energy become higher, then the induction time should be longer. For mono- and diglycine, ΔG_C decreases when the temperature is higher, while, for triglycine dihydrate, the ΔG_C increases when the temperature is higher. For triglycine dihydrate, there are hydrogen bonds formed during the dihydrate crystallisation process. The new conformation of triglycine dihydrate is more stable under lower temperature compared with the anhydrate conformation (details can be seen in chapter 7), which can also be the reason why the calculated ΔG_C is higher at higher temperature.

6.5.4 The evidence of non-classical nucleation phenomenon during the glycine homopeptides crystallisation

Apart from the classical nucleation theory, other non-classical nucleation models have been created over the last two decades, as discussed in Section 2.4.2. For the glycine homopeptides, the liquid -liquid separation was also found in hexaglycine before the nucleation, which can be seen in figure 6.16. The solution become blurry and there are some gel-like particles won't settle for a long time (figure 6.16), similar to the liquid-liquid separation process mentioned in the literature.[97] For triglycine, gelation appeared during the cooling crystallisation of triglycine dihydrate (figure 6.17). This discovery raises a question that whether there is any relationship between gelation and liquid-liquid separation during the peptide crystallisation. However, because the solubility of hexaglycine is very low, the gelation is difficult to observe. This observation provided to us a new direction of future research into the nucleation mechanism for longer chain peptides.

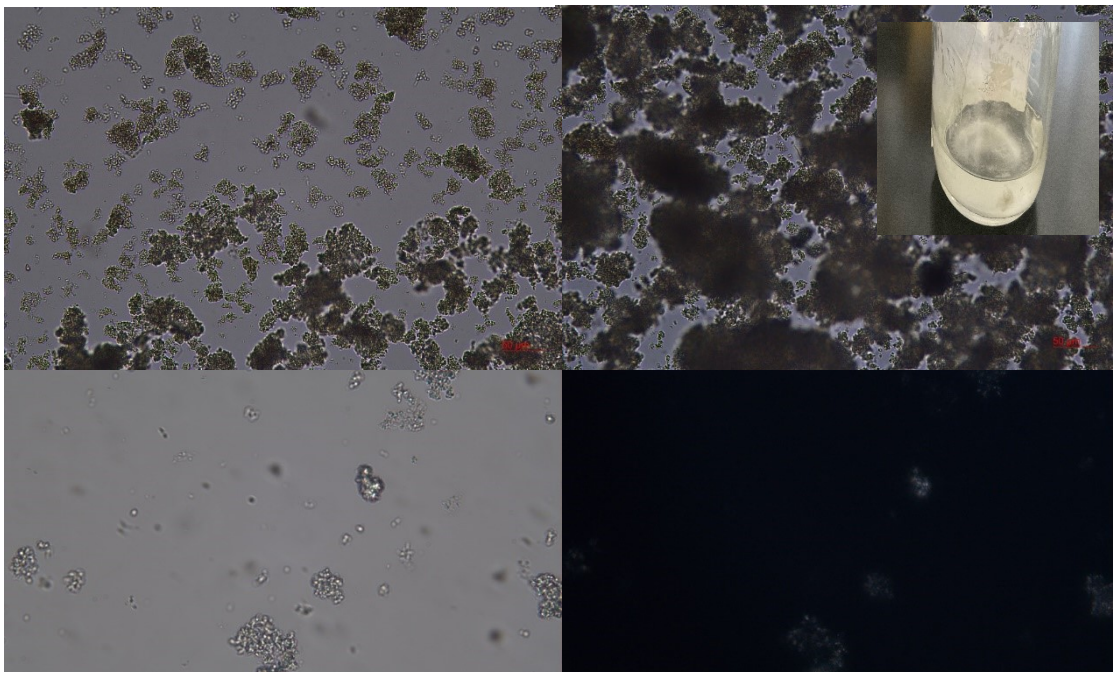


Figure 6.16 The light and polarized microscope images of non-classical nucleation phenomenon during cooling crystallisation of hexaglycine

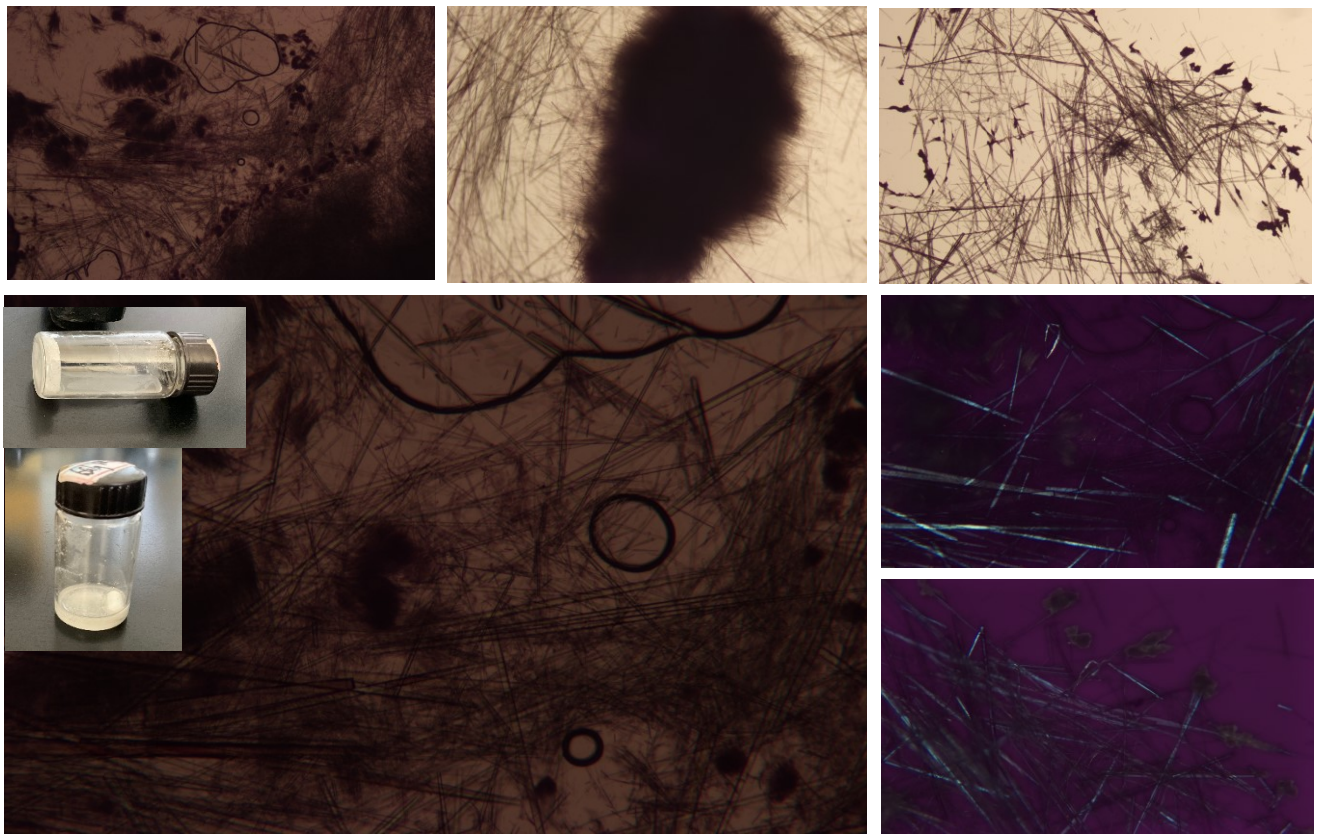


Figure 6.17 Gelation phenomenon during the cooling crystallisation of triglycine dihydrate taken by light and polarized microscope

6.6 Conclusion

In summary, the nucleation of glycine homopeptides was investigated in this chapter to explore the effect of chain length, supersaturation and temperature on the nucleation of short chain homopeptides. The induction time increases with the increasing number of peptide bonds, especially when the number of glycine residues is larger than three, the induction time increases exponentially. For the glycine, diglycine and triglycine, the nucleation parameters (nucleation rate J, growth time t_g, interfacial energy γ, critical radius r_C, number of molecules in critical nucleus n_C and activation Gibbs energy ΔG_C) were calculated. With increasing supersaturation, activation energy, critical radius, number of molecules and growth time decreases, while the nucleation rate increase for all homopeptides. Lower temperature makes the nucleation become

more difficult and the induction time become longer. When the temperature is lower, the interfacial energy γ and Gibbs energy $\Delta G_C$are higher for glycine and diglycine. However, for triglycine, the dihydrate was formed during the nucleation process, the values of interfacial energy and Gibbs energy at low temperature are both lower than those at high temperature, which means the classical nucleation theory is not suitable to explain the nucleation phenomenon of triglycine dihydrate. Moreover, the gelation phenomenon of triglycine dihydrate was found during the cooling crystallisation, as well as the liquid-liquid separation of hexaglycine. For glycine homopeptides, non-classical nucleation theory shows a better explanation when the chain length is longer than three. This work gives a better understanding on the nucleation mechanism for different chain length peptides.

Chapter 7. The role of water in peptide crystallisation and the relationship between conformation and crystallisation conditions

7.1 Abstract

Polyproline II (pPII) is a left-handed 3_1-helix conformation, which has been observed to be the most abundant secondary structure in unfolded peptides and proteins compared to α-helix and β-sheet. Although pPII has been reported as the most stable conformation for several unfolded short chain peptides in aqueous solution, it is rarely observed in their solid-state. Here, we show for the first time a glycine homopeptide (gly-gly-gly) adopting the pPII conformation in its crystalline dihydrate structure. The single crystal X-ray structure with molecular dynamic simulation suggest that a network of water and the charged carboxylate group is critical in stabilising the pPII conformation in solid state, offering an insight into the structures of unfolded regions of proteins and the role of water in peptide crystallisation. Besides, Li_2SO_4, Na_2SO_4 and K_2SO_4 were found to have a negative effect on the formation of dihydrate during cooling crystallisation, which provides a guidance about the effect of salts on the transition between the unfolded and folded conformation in the solid state.

7.2 Introduction

Intrinsically disordered proteins (IDP) make up more than 30% of human genome, and these unstructured regions are often key to the function.[184, 185] pPII conformation, as a significant contributor in unfolded peptide structures, has been observed to be preferred in proline-, alanine- and glycine-rich regions in peptides and proteins.[181-188] Significant improvements in spectroscopic and simulation techniques have improved the detection of these peptide conformers in solutions, while the lack of high resolution X-ray crystal structures has frustrated efforts to understand peptide and protein conformations in the solid-state.[36, 141, 193-198]

The presence of water has been proved to be able to stabilize the pPII conformation in unfolded regions of proteins, which are important in the function and dynamics of a protein.[197-199] However, previous research has not given us a clear and consistent results about how water interacts with peptides, or how water interacts with itself to help stabilize the pPII conformation. Density functional theory (DFT) calculations and molecular dynamics (MD) simulations of alanine-based peptide have been reported to show the hydration of the backbone is the main contributor to the stabilization of pPII.[202, 203] There are also some research emphasize the solvent accessibility area of side chains associated with transitions between pPII and β-sheet like conformations.[204] Ilawe et al., reported not only the indirect solvation is sufficient to support the pPII preference, but also the solvent self-energies.[205] From the current experimental and calculation results, it is challenging to observe the exact coordination of the water molecules to the peptides or proteins for pPII conformation.[140, 205]

Previous studies observed that short chain peptides that are rich in glycine and alanine form the pPII conformation in aqueous solutions, but so far the pPII conformation has been only rarely observed in the solid-state.[202, 206-208] Studies of short chain peptides can be useful to infer information about the regions of proteins that are rich in the amino acids within these

short chain peptides. For example, studies of GAG[200], GGG[200] and GGAGG[36] have revealed that they all have a high proportion of pPII conformation in water. Bykov et al., found high concentration Li^+ cations made the pPII conformation of polyglycine become more favorable in water based on Raman experiment and speculated the reason was Li^+ polarized the peptide bond carbonyls.[209] Meulemans et al., reported a lithium bromide salt of GGG crystallising with the pPII conformation, where the Li^+ ions were attached on the carboxylate groups and one C=O group and they can even link different carboxylate groups together to form a ring. Without the interaction between carboxylate group, the N-terminal group N-H formed hydrogen bond with the left C=O in backbone of triglycine.[210] This crystal structure therefore can support what was observed by Bykov. However, none of the above-mentioned materials were reported to adopt the pPII conformation in the anhydrate or hydrate solid-state. The only other short homopeptides reported in literature which has been found to exhibit a pPII conformation in crystalline form is a hexaproline anhydrous structure.[16] For longer peptides, Crick's research shows the polyglycine can adopt pPII conformation in its anhydrate solid form, however, it still cannot provide us any idea about how water interact with peptide molecules and stabilize the pPII conformation.[39]

Crystal forms for up to five glycine residues in a peptide chain are reported, whereby all the anhydrous and hydrate forms show beta-sheet or fully extended conformations.[211-213] Here, we report the single crystal structure of the triglycine dihydrate (TGDH) which adopts pPII conformation in solid state and discuss its crystallisability and structure, in terms of the conformation and water interactions of glycine rich peptides. The solution conformation and water structuring around triglycine are investigated using molecular dynamics simulations to gain a fuller understanding of the solution to crystal pathway. [143, 214-217]

7.3 Results and Discussion

7.3.1 The cooling crystallisation of mono-, di-and tri-glycine

The cooling crystallisation of mono-, di-and tri-glycine (Figure 7.1) were launched in this work. All the samples were purchased from Sigma Aldrich (all >99% purity) and crystallised from aqueous solution at variable supersaturations and cooling rates.

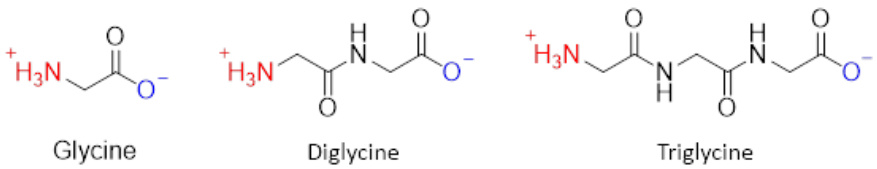


Figure 7.1 The chemical structures of Glycine homopeptides

5 mL triglycine solutions with different supersaturation levels (S=1.4, 1.6, 1.8, 1.9, 2.0, 2.1, 2.2 and 2.3) were prepared in 10 mL crystallisers at 333.15K and then cooled down to 283.15K at different cooling rate (0.1K/min, 0.3K/min, 0.5K/min, 1K/min, 2K/min). The solutions were stirred at 500 rpm during the cooling process. TGDH was collected after filtration at room temperature. Suspension crystallisation was also launched to confirm dihydrate was the most stable form in water when the temperature was lower than 283.15 K. For mono- and di-glycine, they both formed anhydrate,while for triglycine, the PXRD patterns were inconsistent with the commonly crystallised β-triglycine. Figure 7.2 shows the rodshaped crystals from the experiments differed substantially from the known plate morphology of the anhydrous triglycine. It is interesting to observe that the TGDH is easy to crystallise from cooling crystallisation of aqueous solutions, but the glycine and diglycine hydrated structures were not observed.

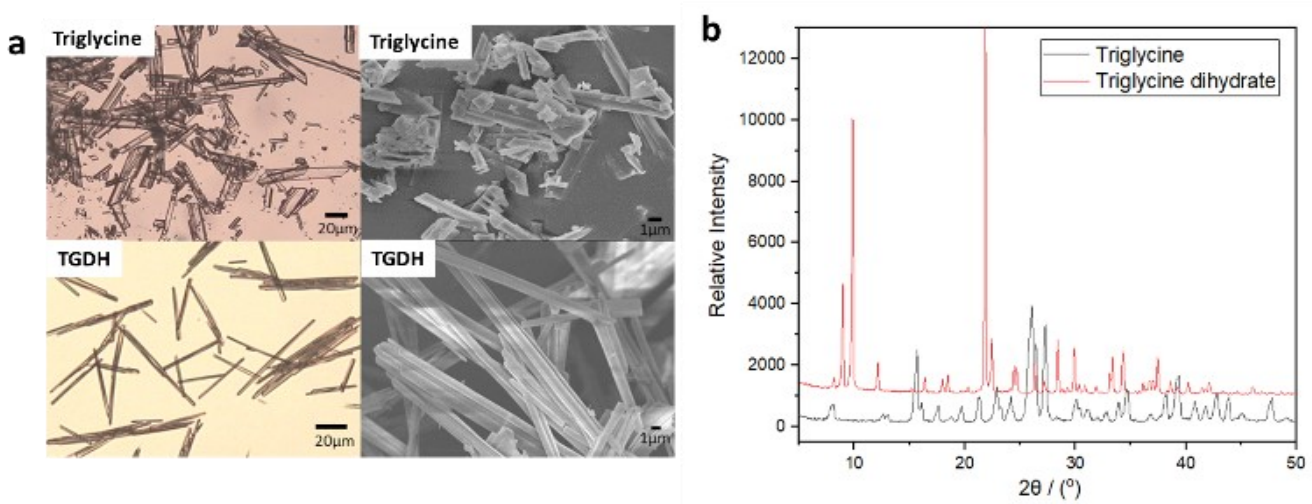


Figure 7.2 a The Optical microscope images and SEM images of triglycine and TGDH; b The XRD pattern of triglycine and TGDH

7.3.2 The thermal and spectroscopic characterisation of triglycine anhydrate and dihydrate

In order to identify the solid form and structure of TGDH, the powder products of triglycine and TGDH were characterised by Optical microscopy, Scanning electron microscopy (SEM), Powder X-ray Diffraction (PXRD), Thermogravimetric analysis (TGA), Differential scanning calorimetry (DSC), Fourier transform infrared spectroscopy (FTIR) and Raman microscope.

The SEM pictures of Triglycine and TGDH were taken by a JEOL-JSM-6700F scanning electron microscope. The samples were coated with a thin layer of gold. Figure 1b shows that the triglycine is plate shape, however the dihydrate looks like thin rod, and very easily to aggregate together.

Powder XRD patterns of triglycine and TGDH were obtained on a PANalytical X' Pert PRO X-ray diffractometer at 40 kV and 40 mA with Cu Kα radiation (λ = 1.5406 Å). The samples were conducted over a 2-theta range from 2° to 50° at a scanning rate of 1 step/s. The XRD pattern (Figure 7.2 b) of the rodlike crystal has several new diffraction peaks (at 2θ =9.03, 9.93, 12.24 and 21.93 degree), which suggests a new polymorph for triglycine.

The water loss of triglycine and TGDH were determined by TGA, which were performed from 298.15 K to 850 K at a heating rate of 10 K/min. A constant nitrogen flow of 60 mL/min was used to prevent thermal oxidation processes. Sample weight was approximately 6 mg. The weight loss of TGDH at 353.68K was 15.99 %, corresponding to two water molecules loss in TGDH (Figure 7.3).

DSC of Triglycine and TGDH were measured with a Discovery DSC Q2000. The measurements were carried out using 5.00 ± 0.10 mg samples under a helium atmosphere (50 $mL \cdot min^{-1}$). From DSC pattern, there is a new endothermic peak before 400K in TGDH, which corresponds to heat absorption from water loss in the crystal structure (Figure 7.3).

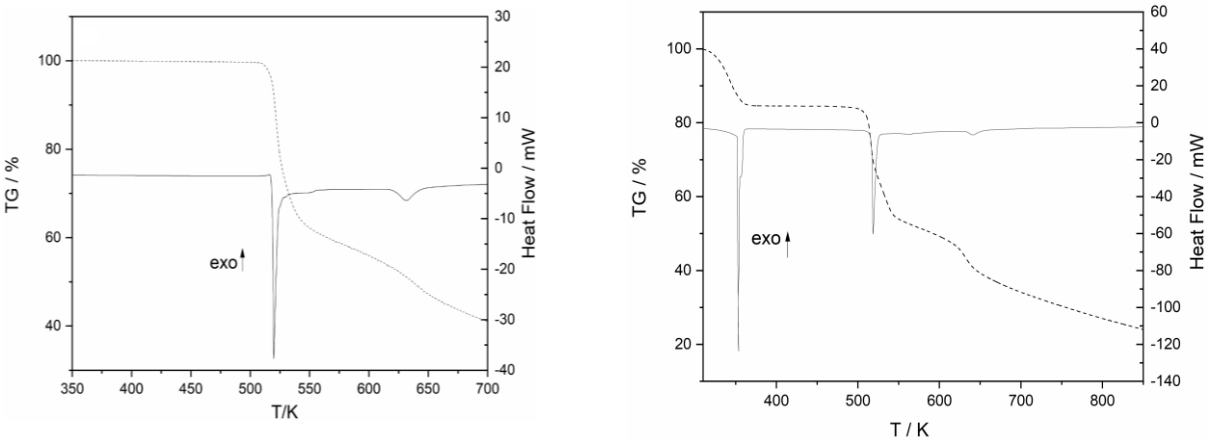


Figure 7.3 The DSC and TGA images of triglycine (left) and TGDH (right)

The interaction of water molecule and triglycine molecule can be verified using the FT-IR and Raman spectroscopy, which were shown in Figure 7.4-7.5. FT-IR spectroscopy was recorded on Cary 630 FT-IR spectrometer. The range of the wavenumber was 4000−650 cm^{-1} with resolution 2 cm^{-1}. Raman spectra was obtained using the Raman microscope Bruker SenterraII. The integration time was 4 s with 1 co-addition for each measurement spot. The power of the laser was 12.5 mW with a wavelength of 532 nm.

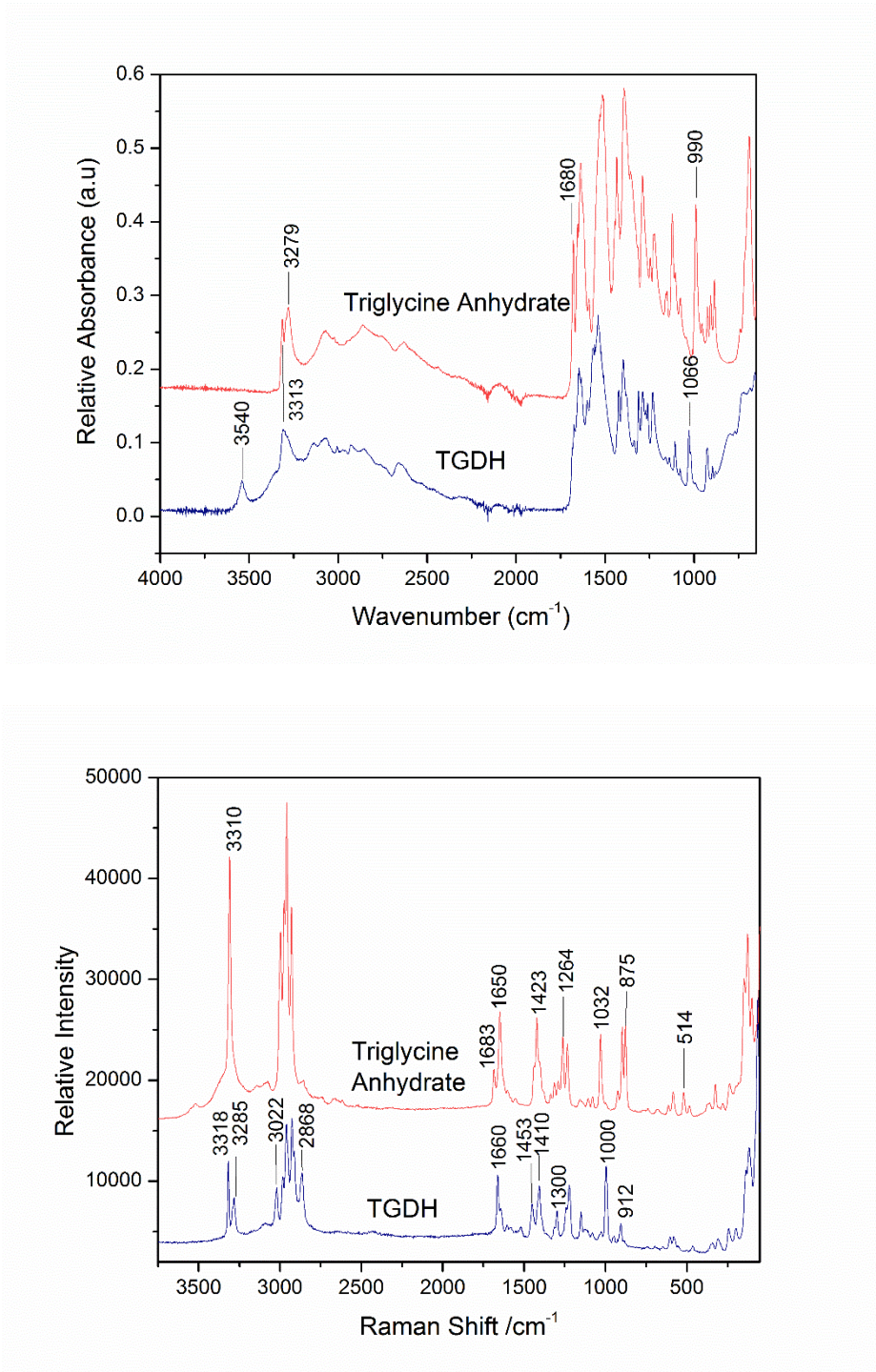


Figure 7.4 The FT-IR (Top) and Raman (Bottom) spectrograms of triglycine anhydrate and TGDH

As shown in FT-IR spectra, the broad band of O-H stretching vibration at frequency 2500-3300 cm^{-1} can be seen in both anhydrate and dihydrate, which are caused by the

interaction between COO^- and NH^{3+} groups of adjacent triglycine molecules. However, there is a new peak at 3540 cm^{-1} corresponds to the hydrogen-bonded O-H group, which comes from the two water molecules in TGDH crystal structure. Besides, the bands for anhydrate at 3313cm^{-1} and 3279 cm^{-1} in FT-IR, corresponding to the stretching vibration of N-H bond and N-H_2 bond, merge to one wider band in TGDH, which indicates the existence of hydrogen bond between N-H and water. This change also leads to the band at 990 cm^{-1} attributed to C-N stretch vibration has a slight shift to 1066 cm^{-1}. The new hydrogen bonding between N-H and water can also been verified from significant change of bands at 3310 cm^{-1} and 2850-3000 cm^{-1} which contributed to bonded symmetric N-H_2 stretching and bonded C-H stretching, respectively in Raman spectra.

Furthermore, it can be seen in figure 7.5 that the amide I band shifts from 1680 cm^{-1} for β-sheet in anhydrate to around 1660 cm^{-1} for pPII in dihydrate in both FT-IR and Raman. The amide II band upshifts from 1518 cm^{-1} for β-sheet to 1540 cm^{-1} and 1568 cm^{-1} for pPII in FT-IR. The peaks in amide III band (around 1300 cm^{-1}) upshift to the higher wavenumbers which correspond the transformation from β-sheet to pPII conformation. Besides, some vibration changes of other groups can be seen from Raman: the bands at 1423 cm^{-1}, 1264 cm^{-1}, 1032 cm^{-1} and 875cm^{-1} corresponding to stretching vibration of C-H_2, C-H, C-C and C-O, respectively, all slightly shift because of the new conformation of the dihydrate.

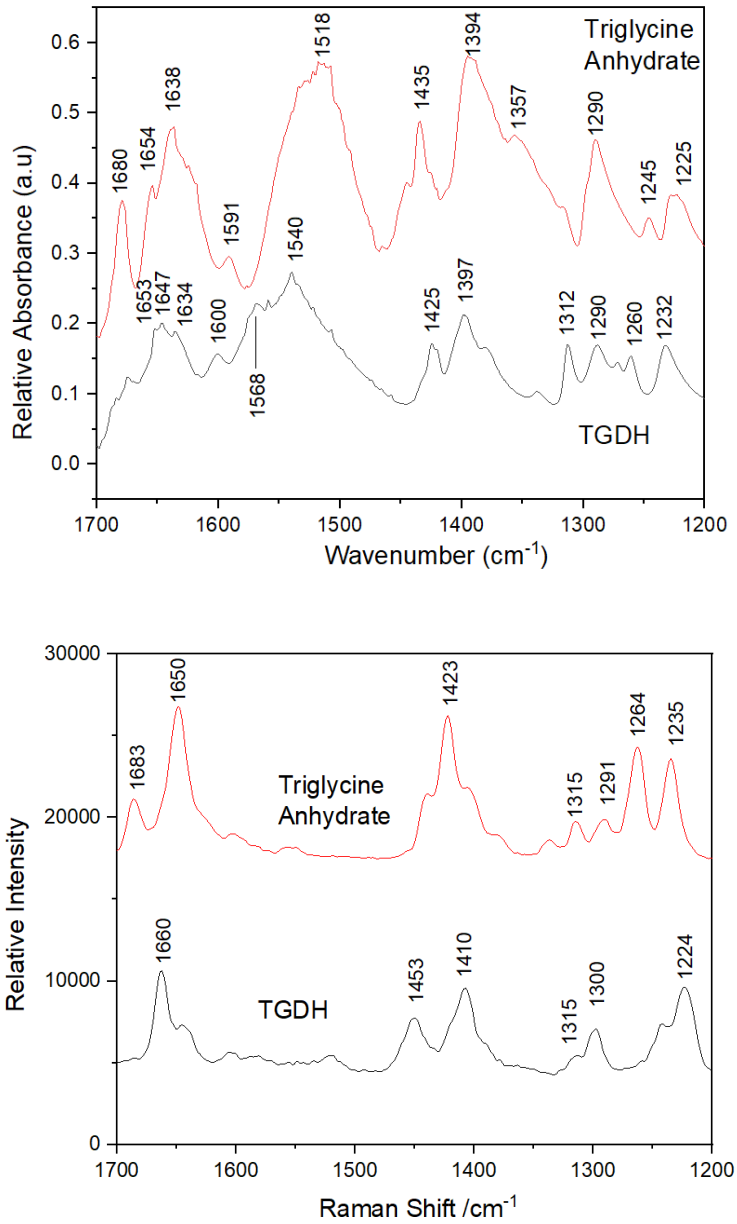


Figure 7.5 The amide modes in FT-IR (Top) and Raman (Bottom) spectrograms of triglycine anhydrate and TGDH

7.3.3 The single crystal growth and structure determination process

Large single crystals of TGDH for single crystal X-ray diffraction experiments were grown by slow solvent evaporation from an aqueous solution at 278.15 K. Triglycine aqueous solutions were prepared in different concentrations (20 mg/mL, 25 mg/mL, 30 mg/mL, 35

mg/mL, 40 mg/mL, 45 mg/mL, 50 mg/mL) at room temperature and filtered with filter membranes (0.22 μm) , then the clear liquid were moved into 35 small sealed glass vessels with different hole numbers (1, 2, 3, 4, 5) on the caps and kept at 278.15 K. The single crystal of TGDH was cultured by allowing the solvent to slowly evaporate from the aqueous solution of triglycine at 278.15 K. After two weeks, the single crystal appeared in the 50 mg/mL solution and the other single crystals appeared in the following days in the 45 mg/mL and 40 mg/mL solutions. The crystals found in the bulk solution during solvent evaporation were all identified as the TGDH form.

The SCXRD data was collected on an Rigaku Saturn 70 abstract diffractometer with a Mo K\ α radiation source (λ = 0.71073 Å). The single crystals with suitable sizes were placed on a fiber needle on the goniometer of the X-ray diffractometer and then purged with a cooled nitrogen gas stream at 113.15 K throughout the data collection. Using Olex2[124], the structure was solved with the ShelXT[125] structure solution program using Intrinsic Phasing and refined with the ShelXL refinement package[125] using Least Squares minimisation. The crystal data was already sent to CCDC with a deposition Number 2049957.

Single crystal X-ray diffraction revealed that the TGDH crystallises in an Orthorhombic Pca21 space group, with 4 triglycine molecules and 8 water molecules in the unit cell. The details of the single crystal structure information can be seen in Table 7.1-7.4. The interaction of TGDH molecules and water molecules can be seen from Figure 7.6. The asymmetric unit in TGDH structure contains one triglycine and two water molecules (Figure 7.7). Each triglycine molecule is surrounded by four water molecules via one N-H···O and three O···H-O hydrogen

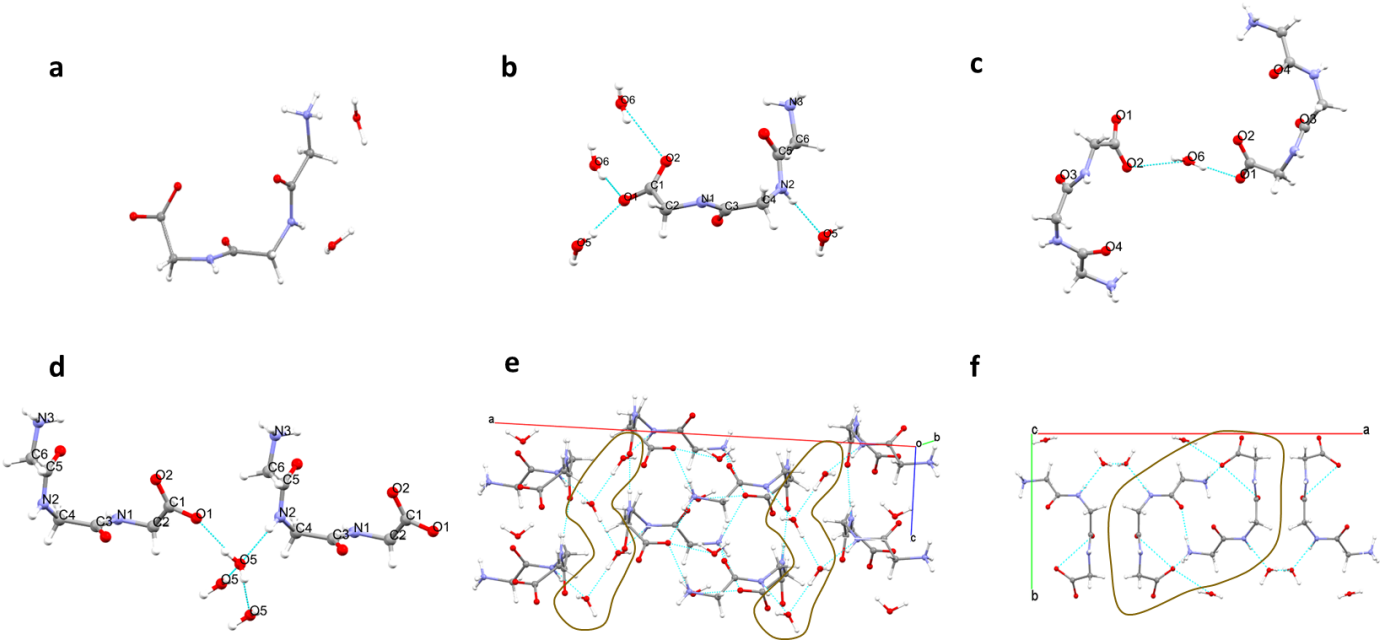


Figure 7.6 a The asymmetric unit in TGDH; b-d The interaction between two water molecules and triglycine molecule; e a chain of hydrogen bonds formed by waters; f The bended triglycine backbone in TGDH

bonds. For the two water molecules in the unit cell, one water molecule (O6) is surrounded by two triglycine molecules via two O-H···O hydrogen bonds, in which the two hydrogen atoms from the water molecule are hydrogen bond donors and two oxygen atoms of carboxyl group from two triglycine molecules are hydrogen bond acceptors. While, the second water molecule (O5) is surrounded by other two water molecules(O5) (via one O-H···O and one O···H-O hydrogen bond) and two triglycine molecules (via one O···H-N and O-H···O hydrogen bond). In this case, three hydrogen atoms from two water molecules and one hydrogen atom from NH group of one triglycine molecule are hydrogen bond donors, while two oxygen atoms from two water molecules and one oxygen atom of carboxyl group from another triglycine molecule are hydrogen bond acceptors.

Table 7.1 The crystallographic data of TGDH

Parameters	Triglycine Dihydrate (TGDH)
Formula	$C_6H_{11}N_3O_4$▪ $2(H_2O)$
Formula weight (g/mol)	225.21
Temperature (K)	113.15
Crystal system, space group	Orthorhombic, Pca21
a (Å)	21.5256(8)
b (Å)	9.7968(4)
c (Å)	4.7685(2)
α (°)	90
β (°)	90
γ (°)	90
V ($Å^3$)	1005.59(7)
Z	4
Density (calculated) (mg/m^3)	1.488
F (000)	480
Crystal colour / morphology	Colourless needles
Crystal size (mm^3)	0.38 x 0.10x 0.08
μ/mm^{-1}	0.132
Radiation	MoKα (λ = 0.71073)
2Θ range for data collection (°)	3.784 to 57.394
Index ranges	$-29 \leq h \leq 29$, $-13 \leq k \leq 13$, $-6 \leq l \leq 6$
Reflections collected	12303
Independent reflections	2602 [Rint = 0.0521, Rsigma = 0.0335]

Data/restraints/parameters	2602/1/172
Goodness-of-fit on F2	1.041
Final R indexes [I>=2σ (I)]	R1 = 0.0466, wR2 = 0.1277
Final R indexes [all data]	R1 = 0.0492, wR2 = 0.1305
Largest diff. peak/hole / e $Å^{-3}$	0.43/-0.29
Flack parameter	0.5(6)

Table 7.2 Bond Lengths for TGDH

Atom	Atom	Length(Å)	Atom	Atom	Length(Å)
O1	C1	1.256(3)	N2	C4	1.457(3)
O2	C1	1.262(3)	N2	C5	1.335(3)
O3	C3	1.232(3)	N3	C6	1.482(3)
O4	C5	1.239(3)	C1	C2	1.538(4)
N1	C2	1.452(3)	C3	C4	1.524(3)
N1	C3	1.343(3)	C5	C6	1.517(3)

Table 7.3 Bond Angles for TGDH

Atom	Atom	Atom	Angle(Å)	Atom	Atom	Atom	Angle(Å)
C3	N1	C2	119.2(2)	O3	C3	C4	120.9(2)
C5	N2	C4	120.8(2)	N1	C3	C4	116.4(2)
O1	C1	O2	124.8(2)	N2	C4	C3	110.7(2)
O1	C1	C2	116.3(2)	O4	C5	N2	123.9(2)
O2	C1	C2	118.9(2)	O4	C5	C6	120.6(2)
N1	C2	C1	113.42(19)	N2	C5	C6	115.4(2)

O3	C3	N1	122.7(2)	N3	C6	C5	109.86(19)

Table 7.4 Torsion Angles for TGDH

A	B	C	D	Angle(Å)	A	B	C	D	Angle(Å)
O1	C1	C2	N1	159.1(2)	C2	N1	C3	O3	3.1(3)
O2	C1	C2	N1	-21.1(3)	C2	N1	C3	C4	-176.54(19)
O3	C3	C4	N2	41.9(3)	C3	N1	C2	C1	-62.6(3)
O4	C5	C6	N3	-3.0(3)	C4	N2	C5	O4	7.0(4)
N1	C3	C4	N2	-138.4(2)	C4	N2	C5	C6	-173.3(2)
N2	C5	C6	N3	177.3(2)	C5	N2	C4	C3	79.9(3)

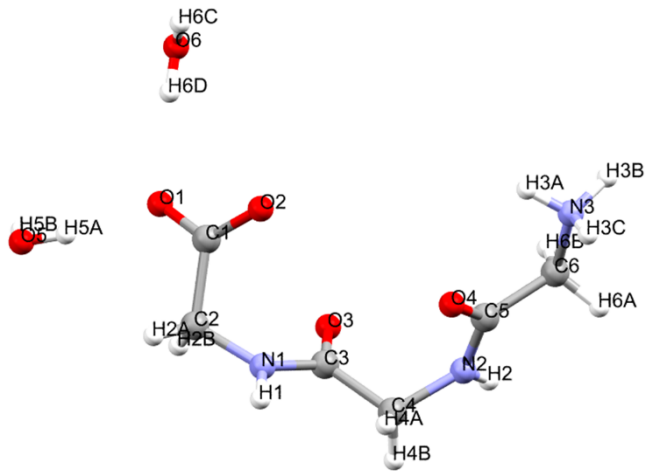


Figure 7.7 The asymmetric unit in the crystal structure of TGDH.

7.3.4 Crystal packing of the glycine homopeptides

The packing coefficient of the new TGDH form was calculated to compare to other structures to investigate whether to crystallisability of the hydrated and anhydrous forms are affected by their packing efficiencies. The glycine molecules are the most tightly packed in the crystal structure of the three peptides examined. However, the packing coefficient (Table 7.5) does not decrease with increasing size of the peptide, with the α-diglycine structure having a

lower packing coefficient than the β-triglycine structure. This suggests that it is not simply the packing efficiency driving the crystallisability of these oligomer polymorphs.

The crystal packing of the new TGDH form was compared to the packing of the other anhydrous forms crystallised in this study, shown in Figure 7.8. The molecules in diglycine and triglycine crystal structure adopt an antiparallel β-sheet and almost fully extended conformation, respectively (Table 7.5), facilitating strong H-bonds between the oligomers. Previous studies have shown that the stable forms of crystalline tetraglycine and pentaglycine also display fully extended conformations.[201] However, the triglycine molecule in the newly discovered TGDH structure displays pPII conformation (Table 7.6), with a left-handed 3_1-helix-like dihedral angles (φ=79.85, ψ=-138.46) for the second glycine residue Gly(2), and the terminal glycine residues have different dihedral angle(ψ=177.30 for Gly(1) and φ=-62.54 for Gly(3)).

Examination of the carboxylate group in the TGDH structure shows that it forms three H-bonding interactions with water molecules, and two with the ammonium groups. The water molecules in the solid state can also form some extended network which includes water-water interactions. This suggests that the stabilization of the pPII conformation in the solid state may also be stabilized by surrounding water molecules. It is interesting to note that Figure 7.10 reveals that the water molecules not only form favorable interactions with the negatively charged carboxylate group, but also link negatively charged COO groups, which would otherwise repel each other at such distances. Furthermore, we speculate that the extended network of water not only stabilizes the pPII conformer but plays a role in reducing the charge-charge repulsions that would arise when packing zwitterions in the condensed solid phase.

Interestingly, there is no interaction between water and the N-terminus in the TGDH crystal structure, but N-terminus can form a hydrogen bond with the unsaturated C=O in the backbone of the adjacent peptide molecule. Also, one water molecule is attached to N-H in the backbone, which acts to impede the inter molecule hydrogen bond between N-H and C=O that normally appears to support the parallel β-sheet structure.

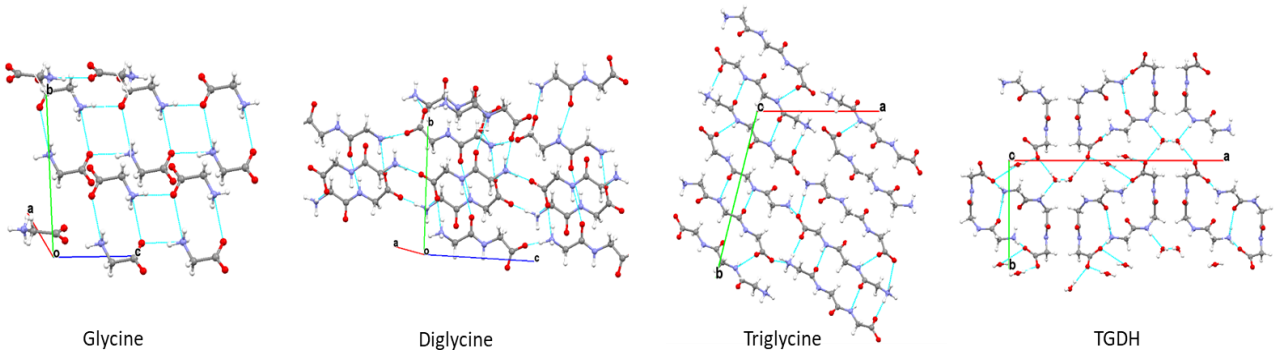


Figure 7.8 The hydrogen bonding motifs of glycine, diglycine and triglycine anhydrate and TGDH in the single crystal structure.

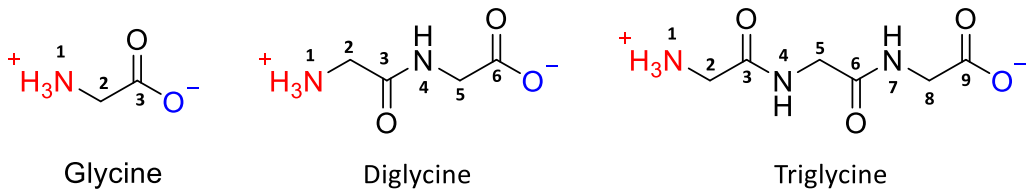


Figure 7.9 The structure of glycine homopeptides with atom numbering

Table 7.5 The torsion of backbone and packing coefficient of glycine homopeptides and TGDH*

Peptides	Dihedral Angles (°)				Packing coefficient
	(1,2,3,4) ψ	(3,4,5,6)φ	(4,5,6,7)ψ	(6,7,8,9)φ	
Alpha Glycine	-	-	-	-	0.78
Alpha Diglycine	152.33	154.61	-	-	0.74
Triglycine moleA	-149.88	177.63	-171.50	172.84	0.76
Triglycine moleB	-161.83	-165.91	174.50	172.79	0.76
TGDH	177.30	79.85	-138.46	-62.54	0.73

* Dihedral angles named using the atom numbering in Figure 7.9.

Table 7.6 The comparison of dihedral angle of Gly(2) in triglycine dihydrate and other main secondary structures

	α-helix	3_{10}-helix	Pi-helix	Parallel beta sheet	Anti- parallel beta sheet	Left-handed pPII helix	Triglycine dihydrate Gly(2)
Frequency	Abundant	Infrequent	Rare	Abundant	Abundant	Infrequent	In this work
φ	-57	-49	-57	-119	-139	-75	79.85
ψ	-47	-26	-70	113	135	145	-138.46

* Glycine has a symmetric Cα, so the dihedral angle (φ, ψ) is equal to (-φ, -ψ). The second glycine residue Gly(2) in the structure adopted similar dihedral angle with left-handed pPII helix.

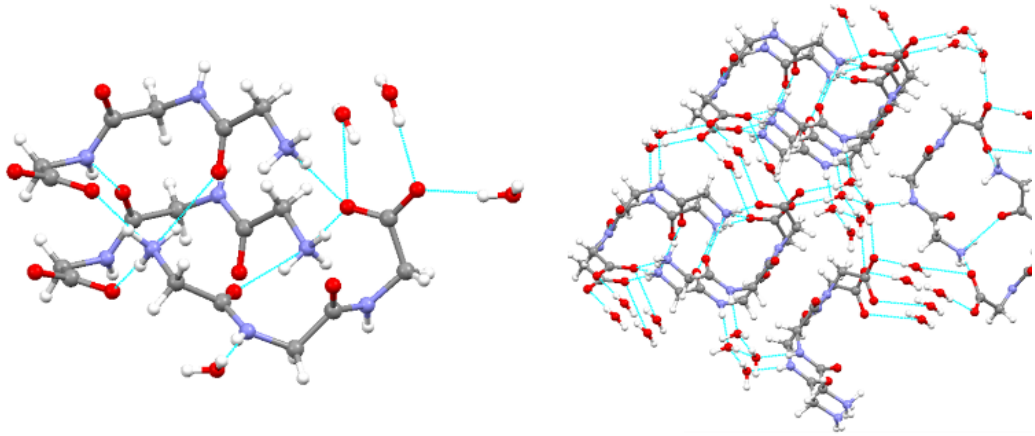
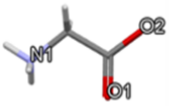


Figure 7.10 The interaction between water and triglycine molecules in the single crystal structure.

The stable packing of the TGDH structure is further confirmed by hydrogen bonding propensity calculations (Figure 7.11), which compare the hydrogen bonds formed by a functional group in a crystal structure to other similar functional groups in other crystal structures.[218] The hydrogen bond propensity calculations reveal how well coordinated hydrogen bonding groups of a molecule are, as compared to similar functional groups in different molecules in the Cambridge Structural Database, where under coordinated hydrogen-bonding interactions can indicate a metastable crystalline form.[219] Calculations show that the stable solid forms of glycine and diglycine are either forming the most common amount of H-bonds for this functional group, or a greater number. Such stable and hyper coordinated H-bonding patterns indicate that these are very stable solid-state structures. The anhydrous triglycine structure was found to have two oxygen atoms which are under H-bond (O5 and O8) coordinated in comparison to oxygen atoms in similar environments in other crystal structures. The O8 is one of the carboxylate oxygens. However, the TGDH structure has optimal or above H-bonding in its structure, whereby the water molecules complete the favoured coordination shell for the carboxylate group.

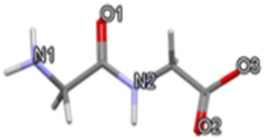
7.3.5 Role of water in stabilising the PPII conformation

Stable conformations in solution have been found to play a role in the crystallisation of specific solid forms.[220, 221] Theoretical studies have indicated that the most stable conformation of triglycine is a compact conformation, rather than the fully extended conformer seen in the anhydrous forms.[222, 223] Coupled with the experimental observations of the pPII conformation in solution, it is suggested that this is a stable solution conformer. [209] Graf et al. explored the conformation distribution of cationic triglycine in water by MD simulations, they found a rugged free energy landscape with a large number of minima distributed all over the Ramachandran space.[147] However, Andrew et al., recently reported the conformation distribution of triglycine in water at room temperature using MD simulation, and show that pPII is a mesostate which enables the functional backbone groups of the central residue to form the most hydrogen bonds with water.[202]



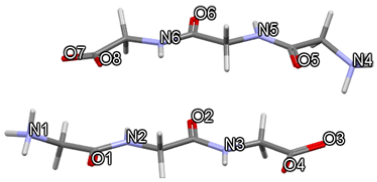
Alpha glycine

Atom (D/A)	= 0	= 1	= 2	= 3	= 4	= 5	>= 6
N1 of acyclic_T4C_nh3 (d)	0.000222169	0.00228635	0.00181489	0.425452	**0.384529**	0.142083	0.0436126
O1 of carboxylate (a)	0.0147985	**0.274169**	0.611357	0.0859672	0.0137088	0	0
O2 of carboxylate (a)	0.0116566	0.199474	0.661438	**0.112368**	0.0150633	0	0



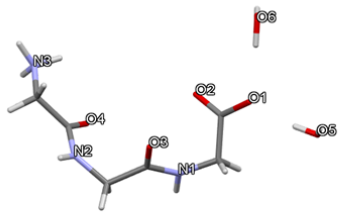
Alpha di-glycine

Atom (D/A)	= 0	= 1	= 2	= 3	= 4	= 5	>= 6
N1 of acyclic_T4C_nh3 (d)	0.000225077	0.00163915	0.00389396	0.473299	**0.371274**	0.109792	0.0398768
N2 of acyclic_amide (d)	0.0529412	**0.929939**	0.0171199	0	0	0	0
O1 of acyclic_amide (a)	0.105109	**0.832179**	0.0598389	0.00287246	0	0	0
O2 of carboxylate (a)	0.0154385	0.282182	**0.609378**	0.0828543	0.0101465	0	0
O3 of carboxylate (a)	0.0127447	0.222374	**0.654147**	0.100458	0.0102755	0	0



Triglycine

Atom (D/A)	= 0	= 1	= 2	= 3	= 4	= 5	>= 6
N1 of acyclic_T4C_nh3 (d)	0.000183852	0.000912672	0.00467055	**0.494947**	0.355698	0.0979732	0.045615
N2 of acyclic_amide (d)	0.0328259	**0.950129**	0.017045	0	0	0	0
N3 of acyclic_amide (d)	0.0721432	**0.913688**	0.0141691	0	0	0	0
N4 of acyclic_T4C_nh3 (d)	0.000272025	0.00109748	0.00602817	0.511659	**0.353421**	0.0882684	0.0392533
N5 of acyclic_amide (d)	0.0577438	**0.926778**	0.0154779	0	0	0	0
N6 of acyclic_amide (d)	0.0456387	**0.93902**	0.0153408	0	0	0	0
O1 of acyclic_amide (a)	0.0952948	**0.820605**	0.0808868	0.00321296	0	0	0
O2 of acyclic_amide (a)	0.0834248	**0.821325**	0.091268	0.00398208	0	0	0
O3 of carboxylate (a)	0.0124426	0.208087	0.654468	**0.113418**	0.0115836	0	0
O4 of carboxylate (a)	0.0160801	0.326274	**0.585178**	0.0671825	0.00528572	0	0
O5 of acyclic_amide (a)	**0.103467**	0.821071	0.0727918	0.00267051	0	0	0
O6 of acyclic_amide (a)	0.0920262	**0.82402**	0.0807502	0.0032034	0	0	0
O7 of carboxylate (a)	0.0119408	0.194273	**0.660564**	0.120764	0.0124582	0	0
O8 of carboxylate (a)	0.0158519	**0.31984**	0.590108	0.0687775	0.00542244	0	0



TGDH

Atom (D/A)	= 0	= 1	= 2	= 3	= 4	= 5	>= 6
N1 of acyclic_amide (d)	0.0119678	**0.968118**	0.0199146	0	0	0	0
N2 of acyclic_amide (d)	0.0148798	**0.966104**	0.019016	0	0	0	0
N3 of acyclic_T4C_nh3 (d)	0.000810314	0.000810314	0.008936	**0.516216**	0.348496	0.08907	0.0356615
O5 of water (d)	0.0176761	0.0798905	**0.778133**	0.110896	0.0134036	0	0
O6 of water (d)	0.0176761	0.0798905	0.778133	**0.110896**	0.0134036	0	0
O1 of carboxylate (a)	0.00421729	0.0933393	**0.637013**	0.234137	0.0312937	0	0
O2 of carboxylate (a)	0.0073327	0.245075	0.627765	**0.107889**	0.0119386	0	0
O3 of acyclic_amide (a)	0.0539267	**0.798634**	0.144538	0.00290162	0	0	0
O4 of acyclic_amide (a)	0.0746331	**0.825891**	0.0980415	0.00143443	0	0	0
O5 of water (a)	0.102736	0.625291	**0.250058**	0.0138867	0.00802926	0	0
O6 of water (a)	0.102736	**0.625291**	0.250058	0.0138867	0.00802926	0	0

Figure 7.11 Left are the molecular structures and labelled H-bonding groups for alpha glycine, alpha diglycine, beta triglycine and TGDH. Right are the H-bonding propensities tables, which show the coordination of H-bonds to these groups from other crystal structures in the CSD, with the number bolded corresponding to the coordination found in these glycine homopeptides crystal structure. If the label is green, it indicates that the coordination found in the crystal structure corresponds with the majority of other structures found in the database, if it is red then this coordination is more unusual.

In this work, MD simulation of triglycine in water at 278 K has led to the prediction of both the extended conformer from the anhydrate structure and the kinked conformer from the TDGH structure. The molecular dynamics simulations were carried out using the DL_POLY 4 [125, 224], where DL_FIELD[225] was used to create DL_POLY 4 input files. The OPLS_AA/M forcefield[6] was utilised with electrostatic potential[226] fractional charges, which were calculated at the density functional theory (B3LYP/6-31G*) level[227], and the TIP3_P water model[228] was also utilised. A single molecule of triglycine, extracted from the anhydrous crystal structure, was placed in the centre of a 64Å X 64Å X 64Å simulation box, which was then filled with water molecules.

The simulations were gently heated to 278 K using an NVE ensemble and 12 Å cut-off radius for computation of interactions, whereby if the temperature remained constant without any scaling, it was considered equilibrated. The NPT ensemble, using the Nose-Hoover thermostat and barostat[229, 230], with relaxation times of 4 and 5 ps respectively, was applied and the simulation ran for 10 ns with a timestep of 0.002 picoseconds. Further analysis, such as creating trajectory files that omitted the water molecules to extract conformations, utilised DL_ANALYSER[231]. Further details on the development of these methods can be found in the following publication.[232]

The results of the MD simulation are evidenced in the Ramachandran plot shown in Figure 7.12, which shows that both the pPII and β-sheet conformations show a high preference compared with other possible conformations over the 4ns time period considered (from 6 ns to 10 ns). To visualise the probability of observing conformers in the Ramachandran space, phi and psi angles of the central alpha carbon in triglycine were taken from simulation trajectories in 2 ns samples. The symmetry in phi and psi angles which arises from the chirality of triglycine, effectively doubling the dataset size from 580,000 to 1,160,000. A 2D histogram was then created in the phi, psi space with a total of 250,000 bins – corresponding to 500 bins in each

dimension.

Frequent observation of the β-sheet conformer in Figure 7.12 suggests that there is minimal energetic barrier between these two conformations of triglycine. Liang reported a pentapeptide GGAGG has a high proportion of pPII conformation in water at low temperature and shifted to beta-strand structure at high temperature to show the temperature dependent propensity of the alanine residue in this peptide.[36] This is especially so, given that the initial conformation of trigycine molecules used to generate the simulation box were specified in their extended form. It is therefore interesting that results do not suggest the presence of a preference for the pPII conformation in solution, though the conformation is still of a significant proportion.

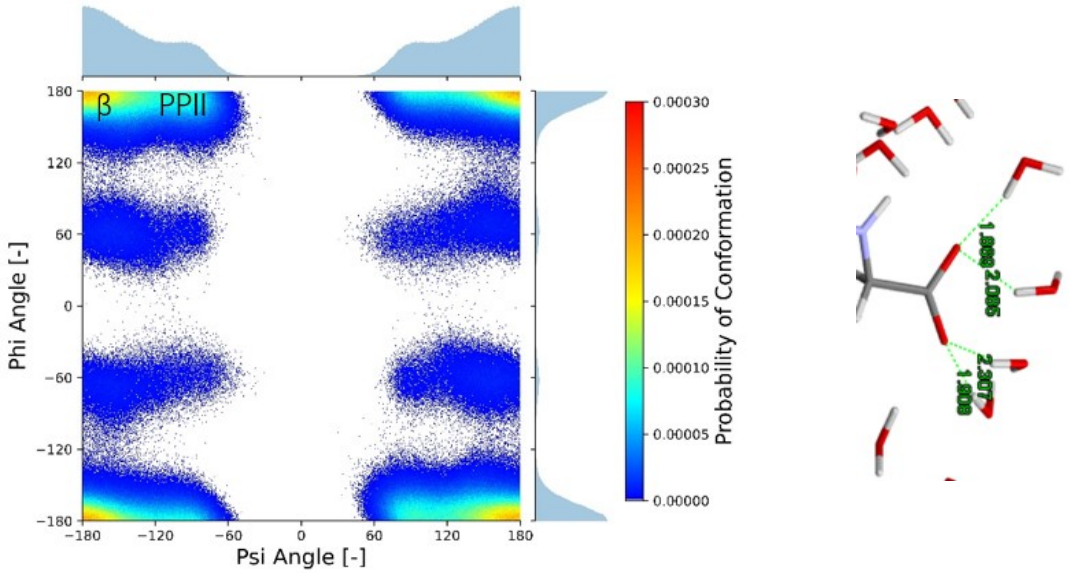


Figure 7.12 Ramachandran plot of the phi and psi angles for the central alpha carbon of triglycine molecules in aqueous solutions over 4 ns (top). Since the central carbon is achiral, the positive and negative regions of the Ramachandran plot can be treated as the same. Below is a snapshot demonstrating the structuring of water structuring around the carboxylate – both from MD simulation.

Consequently, further research is required to investigate potential sources of model uncertainty including the effect of force field and molecule specification - as recently done for

similar glycine peptide systems by Andrews et al.[202] and Zhang et al.[233]. Moreover, while the unchanging structure of water at terminals solution (see radial density functions in Figure 7.13-7.15) suggest that the simulation is converged, it may be fruitful for dedicated simulation works to test the consistency of results for simulations of large sizes and over longer time periods.

Radial density functions have been provided as an alternate means of understanding the structure and potential orientation of water around triglycine in solution in Figure 7.13, Figure 7.14 and Figure 7.15. The radial distribution function, $G(r)_{A-B}$ indicates the density of an atom, B, a given distance, r, from a reference atom, A relative to its bulk density within a system; this is proportional to the probability of finding atom B at the distance r from atom A.

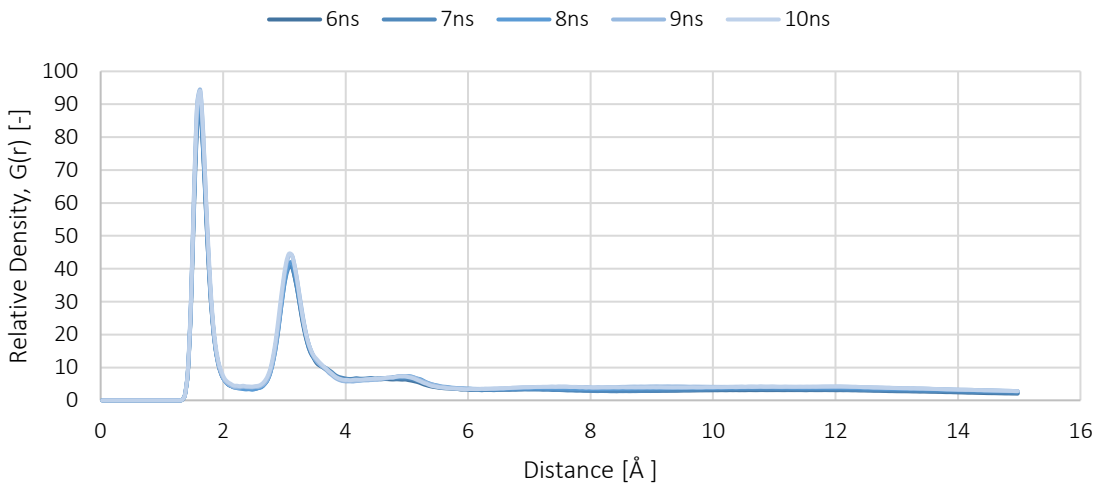


Figure 7.13 Radial distribution functions for the hydrogen the ammonium triglycine terminal (NH_3^+) and oxygen atoms of the carboxylate triglycine terminal (COO^-). Data is segmented into 1 ns sample sizes (e.g. 10 ns = 9.0 to 10.0 ns)

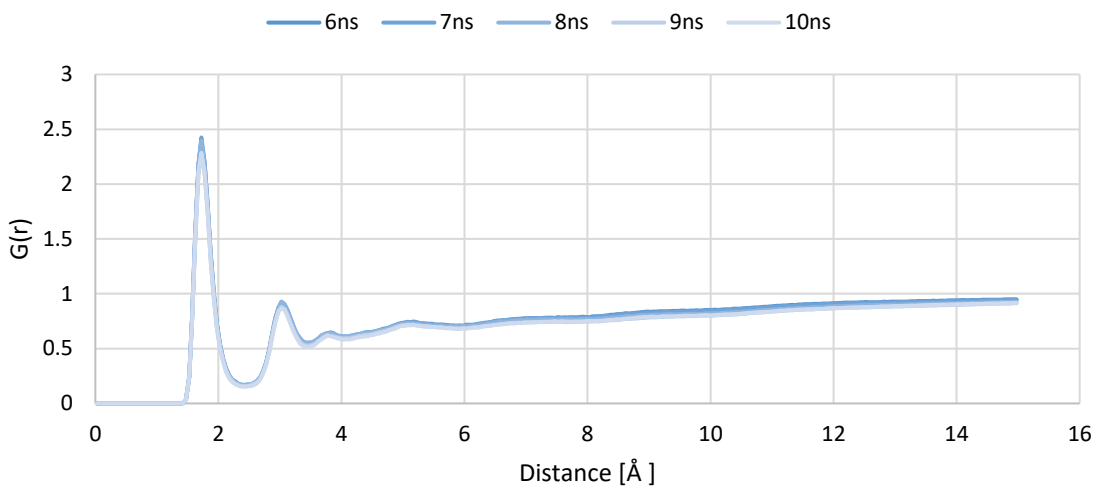


Figure 7.14 Radial distribution functions for the oxygen atoms of the carboxylate triglycine terminal (COO^-) and the hydrogen atom of water. Data is segmented into 1 ns sample sizes (e.g. 10 ns = 9.0 to 10.0 ns).

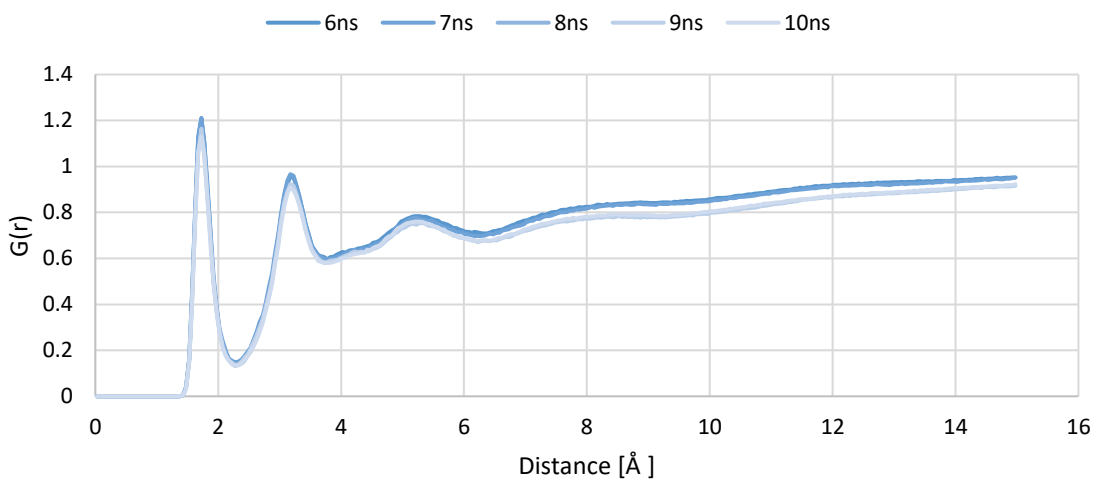


Figure 7.15 Radial distribution functions for the hydrogen atoms of the ammonium triglycine terminal (NH_3^+) and oxygen atoms of water. Data is segmented into 1 ns sample sizes (e.g. 10 ns = 9.0 to 10.0 ns).

A snapshot of the molecular dynamics simulation in Figure 7.12 shows the afore mentioned four molecule hydration shell around the carboxylate group, consistent with the hydration seen in the solid-form structure and previous studies. Analysis of the water interactions with the two

molecules of triglycine is shown in Table 7.7, whereby each carboxylate group is coordinated to between 3 and 4 molecules of water and each ammonium is coordinated to between 1 and 2 molecules of water. This is consistent with waters preference to interact with the carboxylate group and these stable hydrogen bonds remain consistent between the solid and solution state. Nonetheless, future investigations may benefit from extending current understanding of interactions by considering the importance of backbone solvation and specific orientation of water in stabilizing conformed under the conditions considered.[202]

A summary of the average, standard deviation and mean squared error of interactions counted between triglycine terminals and water are shown as a function of time in Table 7.7, 7.8 and 7.9. Averages are taken over a 4ns time interval. H-bonds are defined as within 2.0Å and a donor acceptor angle of at least 120°.

Table 7.7 Average count of hydrogen bond interactions observed for water-carboxylate, water-ammonium and ammonium-carboxylate functionalities in the simulated aqueous environment.

Interactions	Water-Carboxyl	Water-Ammonium	Ammonium-Carboxyl
Average count	3.478 ± 0.020	1.341 ± 0.007	1.310 ± 0.003

Table 7.8 Average count of hydrogen bond interactions observed for ammonium-carboxylate functionalities in the simulated aqueous environment.

Time [ns]	Average	Standard Deviation	Mean Squared Error
2-4	0.932	9.860	0.029
4-6	1.160	5.577	0.009
6-8	1.273	3.366	0.003
8-10	1.339	2.326	0.002

Table 7.9 Average count of hydrogen bond interactions observed for water-carboxylate, functionalities in the simulated aqueous environment.

Time [ns]	Average	Standard Deviation	Mean Squared Error
2-4	4.060	8.307	0.021
4-6	3.667	9.543	0.027
6-8	3.573	7.399	0.016
8-10	3.411	6.093	0.011

Table 7.10 Average count of hydrogen bond interactions observed for water-ammonium functionalities in the simulated aqueous environment.

Time [ns]	Average	Standard Deviation	Mean Squared Error
2-4	1.696	5.648	0.009
4-6	1.428	5.491	0.009
6-8	1.392	4.322	0.006
8-10	1.305	3.911	0.005

DFT was also used to understand about the pPII conformation in solution and vacuo in the previous research. Ilawe used DFT to calculate the energetics of four representative GxG peptides, where x = alanine (A), leucine (L), valine (V), and isoleucine (I) in explicit water and in vacuo.[203] They found that the large energetic contributions to the stabilization of pPII result mainly from peptide–water, water–water interactions, and changes of the solvent self-energy. Compared with our research, peptide GGG which has no heavy atom in the side chain, actually can be used to analyze these three types of interactions as well. Our crystal structure result shows the water can form hydrogen bonds with carboxylate group and these water molecules can also form a water network which increase the distance between C terminus and N terminus of adjacent peptide molecules. This means that N terminus can form a hydrogen bond with the unsaturated C=O in the backbone of the adjacent peptide molecule. Besides, for the central glycine residue, there is one water interacts with the NH group to form a hydrogen bond (Figure 7.10). This structure is therefore a powerful proof for what was discussed in this modeling

literature about the peptide–water, water–water interactions, and also give the current evidence which states that the hydration of the backbone is the main contributor to the stabilization of pPII a new sight to consider about the carboxylate-water and water-water interactions.

The hydrated forms of glycine and diglycine have been reported to be much more challenging to crystallise than TGDH and were not observed from the crystallisation experiments in this study. The single crystal of mono-glycine hydrate was difficult to obtain due to the coexistence of glycine and ice at low temperature conditions (208K). Its crystal structure was obtained using computational predictions[234] and is consistent with NMR measurements[235]. For diglycine, the reported hydrate form has anti-parallel beta-sheet conformation but does not remain stable when exposed to ambient conditions.[212] In contrast, TGDH remains stable at room temperature for up to a year from our current observations.

Figure 7.16 shows that the hydrate structure of diglycine is closely related to its anhydrate. The lack of flexibility of the diglycine molecule is likely to be a contributing factor to the similarity in packing of the hydrated and anhydrous forms. In contrast, the increased flexibility of the triglycine molecule can generate an alternate and stable packing arrangement with a strong peptide-water and water-water interactions. We speculate that the isostructural nature of the diglycine hydrate and its anhydrate facilitate the conversion between the forms. Whereas the very different packing of TGDH, compared to its anhydrous form, provides an energetic barrier to conversion and stabilizes it upon isolation.

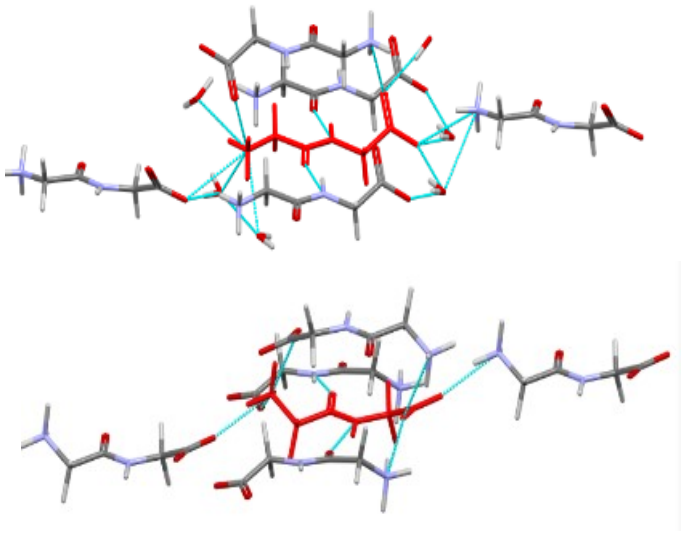


Figure 7.16 The packing around a central diglycine molecule in the diglycine hydrate structure (above); the packing around a central diglycine molecule in the anhydrate form (below)

7.3.6 The effect of temperature and salts on the conformation change in solution

Online FTIR was used to measure the conformation change under different temperatures and salts to explore the factors that can affect the interaction of water and peptides molecules. A certain amount of diglycine and triglycine were added into crystalliser at low temperature as 278.15K first and kept stirring to measure the FTIR spectrum until it reached to the equilibrium, which means the absorbance of FTIR will not increase any more. Then, the temperature was increased to 283.15K to get the spectrum at that temperature. The process was repeated until the temperature went to 333.15K, and then the solution was cooled down again to 278.15 K.

The results can be seen in figure 7.17. From the results, there is no band shifting for diglycine under different temperatures, only an increase in absorbance with increasing concentration. It means there is no conformation change for diglycine. While for triglycine, when the temperature increase, the amide I band at 1650 cm^{-1} shifted to left and the peaks at 1580 also has a right shifting, which corresponds to the beta-sheet conformation. This spectrum change shows that triglycine adopts the unfolded PPII conformation at low temperature and

changes to the folded beta-sheet conformation at higher temperature. The conformation of peptides in solution is identical to the conformation adopted in the crystalline form at the corresponding temperature, implying that the conformation of peptides in solution is critical for the final crystalline form.

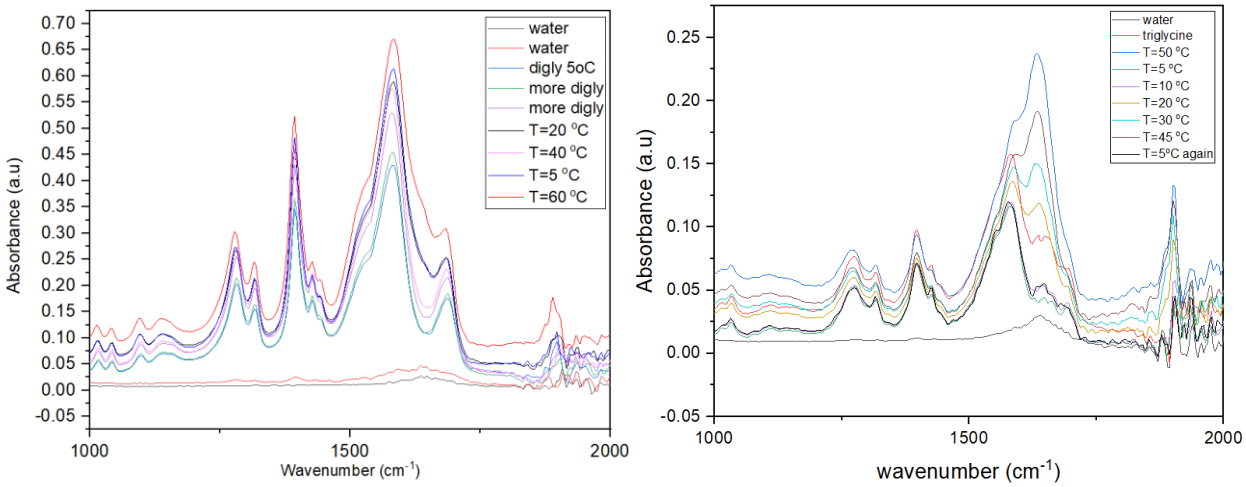


Figure 7.17 The online FTIR spectrum for diglycine (left) and triglycine (right) under different temperatures

The effect of ions on the PPII conformation in the final crystal structure and in the solution has been researched. Meulemans et al., reported a lithium bromide salt of triglycine crystallised with the pPII conformation (figure 7.18), demonstrating that the interaction between Li^+ and triglycine exists on the carboxylate groups and one C=O group.[208] Without the carboxylate group, the N-terminus formed hydrogen bond with the remaining C=O in the triglycine backbone.[210] The interaction between the polar and carboxylate group is analogous to that found in triglycine dihydrate. To investigate the effect of cations on the conformation and crystal structure, we added three different sulphate salts (Li_2SO_4, Na_2SO_4, and K_2SO_4) to the solution during the cooling crystallisation process. The microscope images of the products can be seen in figure 7.19. In water, triglycine can form anhydrate when the temperature is higher than 303.15K and dihydrate when the temperature is lower than that. From figure 7.19, the triglycine anhydrate (plate form) appeared at 293.15K and 298.15K, which means the cations actually

prohibit the formation of the dihydrate. The transition of the morphology was also verified by XRD measurement.

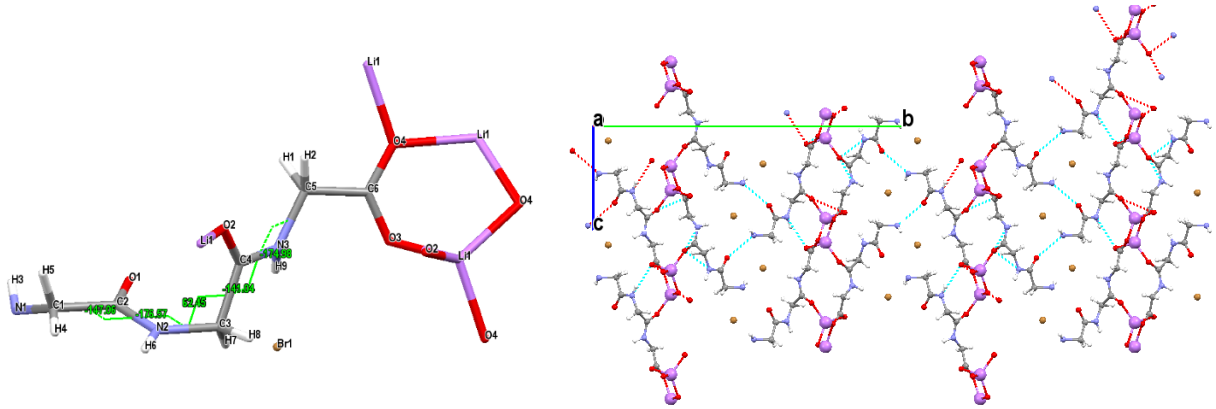


Figure 7.18 The crystal structure of LiBr·Glyglygly[206]

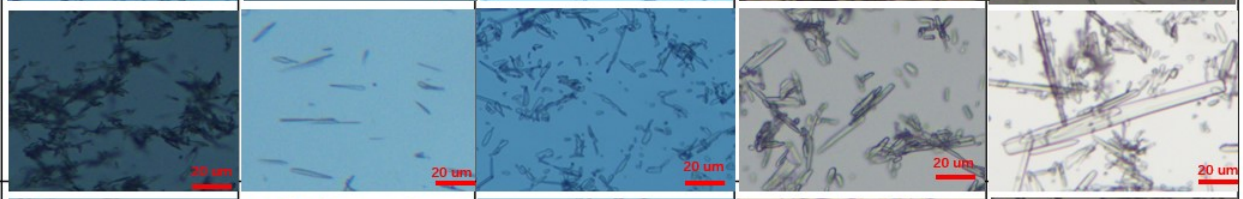
Salts	T/K				
	278	283	288	293	298
Li_2SO_4 0.355M					
Na_2SO_4 0.4730M					
K_2SO_4 0.3147M					

Figure 7.19 The microscope images of final crystals when adding different salts under different temperature

7.4 Conclusion

In summary, the stability of the pPII conformer, the ease of accessing kinked and straight chain conformers in solution and the extended hydration of the carboxylate group from solution to solid-state, likely drives the crystallisation of the triglycine dihydrate. Furthermore, the significant rearrangement needed to convert to the anhydrous form can create an energy barrier to solid-state conversion, hence stabilising the hydrated form at ambient conditions. The understanding of the role of water and conformation in proteinogenic amino acids is crucial to a more mature understanding of unfolded peptides and proteins crystallisation. Our research find that even though the central glycine residue can contribute to the pPII conformation, the carboxylate-water and water-water interactions play a more important role in stablizing the pPII conformation. Li_2SO_4, Na_2SO_4 and K_2SO_4 were found to have a negative effect on the PPII conformation in the crystalline form, providing insight into the effect of salts on the structure stability of peptides. Once the determinants of pPII formation are known in more detail, it will become possible to apply them to the understanding of protein unfolded states. These data will also allow for modeling of the unfolded state ensembles of specific proteins with a level of realism that has not been previously anticipated.

Chapter 8 Conclusions and Future work

8.1 Overview

The crystallisation behaviour of glycine homopeptides was examined in this thesis to gain a better understanding of peptide crystallisation. The research includes the solubility measurement, solubility prediction, nucleation mechanism and the relationship between conformation and crystallisation conditions.

The solubilities of glycine homopeptides (glycine to hexaglycine), four amino acids with different side chains, eight dipeptides, three peptide hydrates were studied in water with experimental and computational methods. The results indicate that the solubility of glycine homopeptides diminishes as the chain length increases, and can be lowered by an order of magnitude when the chain length exceeds two. Distinct side chains have a different effect on the solubility of peptides. For glycine, aspartic acid and histidine, the addition of glycine and phenylalanine residues has no effect on the order of their solubility, but the addition of phenylalanine residue can result in a significant fall in solubility value. Glycine and phenylalanine also perform differently when compared to the solubility of dipeptides and their two constituent amino acids. Dipeptides have a solubility between that of their constituent amino acids, however phenylalanine residues reduce the solubility of dipeptides below that of both constituent amino acids. The greater the number of phenylalanine residues in the peptides, the less soluble the peptides are. These findings demonstrate that, in comparison to basic, acidic, or polar side chains, the non-polar (hydrophobic) side chain has the most noticeable effect on solubility. Furthermore, the solvation free energy of the peptides in water calculated by molecular dynamic simulations suggests that the peptide-water interaction is influenced by chain length and side chains of peptides, with stronger interactions leading to higher solubilities.

Based on the experimental data, SAFT-γ Mie approach was extended for the prediction of the solubility of amino acids and peptides. The NH_2-COOH parameters were determined based on the thermodynamic properties of glycine, alanine and valine. These parameters can be used to predict the properties of other structurally similar amino acids (alanine, valine, aspartic acid, glutamic acid, isoleucine, leucine and lysine).

The nucleation of glycine homopeptides was investigated in chapter 6 to determine the nucleation mechanism for short chain homopeptides as well as the critical parameters affecting nucleation. The research found the induction time increase with the increasing number of peptide bonds, especially when the number of glycine residues is larger than three, the induction time increase an order of magnitude. Supersaturation is still a key factor which has positive effect on the nucleation of all peptides. Although lower temperatures make nucleation more difficult, for triglycine, the dihydrate form was formed during the nucleation process, the values of interfacial energy and Gibbs energy are both lower at low temperatures than at high temperatures, indicating that the classical nucleation theory does not adequately explain the nucleation phenomenon of triglycine dihydrate. Additionally, the gelation of triglycine dihydrate during cooling crystallisation was observed, as was the liquid-liquid separation of hexaglycine. Moreover, the gelation phenomenon of triglycine dihydrate was found during the cooling crystallisation, as well as the liquid-liquid separation of hexaglycine. When the chain length of glycine homopeptides exceeds three, non-classical nucleation theory appears to provide a better explanation.

The research in chapter 7 revealed the role of water in the peptide crystallisation through the triglycine dihydrate. My research found that even though the central glycine residue can contribute to the pPII conformation, the carboxylate-water and water-water interactions play a more important role in stablizing the pPII conformation. Li_2SO_4, Na_2SO_4 and K_2SO_4 were also found to have a negative effect on the PPII conformation in the crystalline form.

8.2 Key contributions

1. The thermodynamic analysis of amino acids and peptides enables the study of the effect of chain length and side chains on the solubility of peptides, thereby providing a fundamental database for not only solubility prediction research, but also for the design and optimization of peptide crystallisation.
2. The SAFT-γ Mie was used to predict the thermodynamic properties of amino acids and peptides, providing a flat form for investigating the link between their physicochemical properties and biostructures.
3. The nucleation mechanism of glycine homopeptides with varying chain lengths was investigated to determine how the nucleation process changes with increasing chain length.
4. The unfolded secondary structure in the crystalline form of peptides was researched in this thesis. The single crystal structure of triglycine dihydrate was resolved, which reveals the role of water in stabilising the PPII conformation. The relationship between conformation and crystallisation conditions was also explored, which is crucial to understanding of unfolded peptides and proteins crystallisation. These experimental data will enable a new level of realism in modelling the unfolded state ensembles of individual proteins.

8.3 Future work

8.3.1 Solubility measurement and prediction

In chapter 4, the solubilities of six glycine homopeptides, four amino acids with varying side chains, eight dipeptides, three peptide hydrates in water under different temperature were already measured to research the chain length and side chain effect. There are several work should be done for the thermodynamic properties of peptides in the future:

1. Additional data from other homopeptides and peptides with various side chains in various solvents should be collected in order to develop a peptides solubility database.
2. The effect of salts, pH and other additives on the peptides solubility also need to be considered, which will give a guidance for designing the crystallisation procedure.

8.3.2 Solubility prediction using gSAFT

Because of the few experimental data of peptides, so the work in this thesis only fitted the parameters between NH2 and COOH groups, the following work can be done to expand the database to predict solubility of more amino acids and peptides.

1. The decomposition happens along the melting process for peptides, so the fusion enthalpy is still difficult to be measured using DSC. The accurate fusion enthalpy needs to be measured experimentally.
2. For the longer chain peptides, more parameters still need to be fitted in the future, which can be seen in section 5.4.8.

8.3.3 Nucleation of peptides

Gel phenomena and liquid-liquid phase separation were seen in my current work during the cooling crystallisation of glycine homopeptides, however the subtleties were not explained. Thus, the following studies can serve as a good starting point for learning more about the mechanism of peptide nucleation:

1. The induction time of other short chain peptides should be measured to verify the effect of chain length on the homopeptides nucleation.
2. To capture the liquid-liquid phase separation, dynamic light scattering (DLS) study and online detection characterisation regarding the pre-nucleation period should be conducted.
3. Investigate the influence of gelation on the transition between triglycine anhydrate and dihydrate.

8.3.4 The transition between folded and unfolded conformation in peptide crystallisation

In Chapter 7, triglycine dihydrate was shown to adopt the peculiar PPII structure that is observed in unfolded peptides and proteins and has distinct biochemical properties. Additional research should be conducted to determine the transition from folded to unfolded structures and the mechanisms by which unfolded structures can be stabilised in solution and solid states.

1. The dehydration of triglycine dihydrate should be investigated. Some preliminary work has been conducted, and evidence for the shift from dihydrate to anhydrate at elevated temperatures has already been observed. Different temperature settings and heating times must be used to conduct a thorough investigation of dehydration.
2. In-situ FTIR and AFM should be explored to monitor the transition of peptides between folded and unfolded conformation in different solvents, as well as the effect of salts and pH on the conformation in solution and in final crystalline state.
3. The effect of water activity on the stabilising ability need to be explored further, as this will aid in developing a thorough understanding of the role of water in peptide crystallisation.

References

1 Li, X.; Wang, J.; Wang, T.; Wang, N.; Zong, S.; Huang, X.; Hao, H., Molecular Mechanism of Crystal Nucleation from Solution. *Science China Chemistry* **2021,** 64, 1460–1481.

2 Gao, Z.; Rohani, S.; Gong, J.; Wang, J., Recent Developments in the Crystallization Process: Toward the Pharmaceutical Industry. *Engineering* **2017,** *3* (3), 343-353.

3 Robotti, N., The Discovery of X-Ray Diffraction. *Rendiconti Lincei* **2013,** *24* (1), 7-18.

4 Eckert, M., Disputed Discovery: The Beginnings of X-Ray Diffraction in Crystals in 1912 and Its Repercussions. *Zeitschrift für Kristallographie* **2012,** *227* (1), 27-35.

5 Davisson, C.; Germer, L. H., Diffraction of Electrons by a Crystal of Nickel. *Physical Review* **1927,** *30* (6), 705-740.

6 Freundlich, M. M., Origin of the Electron Microscope. *Science* **1963,** *142* (3589), 185-188.

7 Fosgerau, K.; Hoffmann, T., Peptide Therapeutics: Current Status and Future Directions. *Drug Discov Today* **2015,** *20* (1), 122-128.

8 Chaparro-Aguirre, E.; Segura-Ramirez, P. J.; Alves, F. L.; Riske, K. A.; Miranda, A.; Silva Junior, P. I., Antimicrobial Activity and Mechanism of Action of a Novel Peptide Present in the Ecdysis Process of Centipede Scolopendra Subspinipes Subspinipes. *Scientific Reports* **2019,** *9* (1), 13631.

9 Jindal, H. M.; Le, C. F.; Mohd Yusof, M. Y.; Velayuthan, R. D.; Lee, V. S.; Zain, S. M.; Isa, D. M.; Sekaran, S. D., Antimicrobial Activity of Novel Synthetic Peptides Derived from Indolicidin and Ranalexin against Streptococcus Pneumoniae. *PLoS One* **2015,** *10* (6), e0128532.

10 Cicero, A. F. G.; Fogacci, F.; Colletti, A., Potential Role of Bioactive Peptides in Prevention and Treatment of Chronic Diseases: A Narrative Review. *British Journal of Pharmacology* **2017,** *174* (11), 1378-1394.

11 Marqus, S.; Pirogova, E.; Piva, T. J., Evaluation of the Use of Therapeutic Peptides for Cancer Treatment. *Journal of Biomedical Science* **2017,** *24* (1), 21.

12 Pittala, S.; Levy, I.; De, S.; Kumar Pandey, S.; Melnikov, N.; Hyman, T.; Shoshan-Barmatz, V., The Vdac1-Based R-Tf-D-Lp4 Peptide as a Potential Treatment for Diabetes Mellitus. *Cells* **2020,** *9* (2), 481.

13 Tomaro-Duchesneau, C.; LeValley, S. L.; Roeth, D.; Sun, L.; Horrigan, F. T.; Kalkum, M.; Hyser, J. M.; Britton, R. A., Discovery of a Bacterial Peptide as a Modulator of Glp-1 and Metabolic Disease. *Scientific Reports* **2020,** *10* (1), 4922.

14 Jeong, W. J.; Bu, J.; Han, Y.; Drelich, A. J.; Nair, A.; Kral, P.; Hong, S., Nanoparticle Conjugation Stabilizes and Multimerizes Beta-Hairpin Peptides to Effectively Target Pd-1/Pd-L1 Beta-Sheet-Rich Interfaces. *Journal of the American Chemical Society* **2020,** *142* (4), 1832-1837.

15 Sun, H.; Choi, W.; Zang, N.; Battistella, C.; Thompson, M. P.; Cao, W.; Zhou, X.; Forman, C.; Gianneschi, N. C., Bioactive Peptide Brush Polymers Via Photoinduced Reversible-Deactivation Radical Polymerization. *Angewandte Chemie International Edition* **2019,** *58* (48), 17359-17364.

16 Spencer, R. K.; Nowick, J. S., A Newcomer's Guide to Peptide Crystallography. *Israel Journal of Chemistry* **2015,** *55* (6-7), 698-710.

17 Myerson, A. S., *Handbook of Industrial Crystallization.* 2nd ed. **2002,** Butterworth-Heinemann: Wobern, MA. 267-285.

18 Langel, U.; Cravatt, B. F.; Graslund, A.; Heijne, N. G. H. V.; Zorko, M.; Land, Tiit.; Niessen, Sherry., *Introduction to Peptides and Proteins*. 1st ed. **2010,** CRC Press, Boca Raton.

19 Liu, H. X.; Zhang, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T., Prediction of the Isoelectric Point of an Amino Acid Based on GA-PLS and SVMs. *Journal of Chemical Information and Modeling* **2004***;44* (1):161-167.

20 Audain, E.; Ramos, Y.; Hermjakob, H.; Flower, D. R.; Perez-Riverol, Y., Accurate Estimation of Isoelectric Point of Protein and Peptide Based on Amino Acid Sequences. *Bioinformatics* **2016,** *32* (6), 821-827.

21 Kirkwood, J.; Hargreaves, D.; O'Keefe, S.; Wilson, J., Using Isoelectric Point to Determine the Ph for Initial Protein Crystallization Trials. *Bioinformatics* **2015,** *31* (9), 1444-51.

22 Mant, C. T.; Kovacs, J. M.; Kim, H. M.; Pollock, D. D.; Hodges, R. S., Intrinsic Amino Acid Side-Chain Hydrophilicity/Hydrophobicity Coefficients Determined by Reversed-Phase High-Performance Liquid Chromatography of Model Peptides: Comparison with Other Hydrophilicity/Hydrophobicity Scales. *Biopolymers* **2009,** *92*(6):573-595.

23 Aftabuddin, M.; Kundu, S., Hydrophobic, Hydrophilic, and Charged Amino Acid Networks within Protein. *Biophysical Journal* **2007,** *93* (1), 225-31.

24 Gunter, F., Peptidyl-Prolyl Cis/Trans Isomerases and Their Effectors. *Angewandte Chemie International Edition in English* **1994,** *33* (14), 1415-1436.

25 Yawen, B.; S., M. J.; Leland, M.; Walter, E. S., Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins: Structure, Function, and Bioinformatics* **1993,** *17* (1), 75-86.

26 Anthis,N. J.; Clore, G. M., Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Science* **2013,** *22* (6), 851-858.

27 Nielsen, E. B.; Schellman, J. A., The Absorption Spectra of Simple Amides and Peptides. *The Journal of Physical Chemistry* **1967,** *71* (7), 2297-2304.

28 Sevier, C. S.; Kaiser, C. A., Formation and Transfer of Disulphide Bonds in Living Cells. *Nature Reviews Molecular Cell Biology* **2002,** *3*, 836.

29 Chen, X.; Hu, C.; Huang, Y.; Chen, Y., Role of Disulfide Bonds in Activity and Stability of Tigerinin-1r. *International Journal of Molecular Sciences* **2018,** *19* (2).288.

30 Peter D. Sun; Christine E. Foster; Boyington, J. C., Overview of Protein Structural Andfunctional Fold. *Current Protocols in Protein Science* **2004**, 17.1.1-17.1.189.

31 Pauling L.; Corey R. B.; Branson H. R., The Structure of Proteins Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proceedings of the National Academy of Sciences of the United States of America* **1951,** 37(4), 205-211.

32 Joosten, R., X-Ray Structure Re-Refinement. Combining Old Data with New Methods for Better Structural Bioinformatics. *Electronic Notes in Theoretical Computer Science - ENTCS* **2010**.

33 Adzhubei, A. A.; Sternberg, M. J.; Makarov, A. A., Polyproline-II Helix in Proteins: Structure and Function. *Journal of Molecular Biology* **2013,** *425* (12), 2100-2132.

34 Warkentin, E.; Weidenweber, S.; Schuhle, K.; Demmer, U.; Heider, J.; Ermler, U., A Rare Polyglycine Type II-Like Helix Motif in Naturally Occurring Proteins. *Proteins* **2017,** *85* (11), 2017-2023.

35 Wilhelm, P.; Lewandowski, B.; Trapp, N.; Wennemers, H., A Crystal Structure of an Oligoproline PpII-Helix, at Last. *Journal of the American Chemical Society* **2014,** *136* (45), 15829-15832.

36 Ding, L.; Chen, K.; Santini, P. A.; Shi, Z.; Kallenbach, N. R., The Pentapeptide Ggagg Has PII Conformation. *Journal of the American Chemical Society* **2003,** *125*, 8092-8093.

37 Rucker, A. L.; Creamer, T. P., Polyproline II Helical Structure in Protein Unfolded States: Lysine Peptides Revisited. *Protein Science* **2002,** *11* (4), 980-5.

38 Ramachandran, G.N.; Sasisekharan, V.; Ramakrishnan, C., Molecular Structure of Polyglycine II. *Biochimica et Biophysica Acta (BBA) - Biophysics including Photosynthesis* **1966,** *112*(1), 168-170.

39 Crick, F.H.; Rich, A. Structure of Polyglycine II. *Nature* **1955**, *176*, 780–781.

40 Black, S. N., Crystallization in the Pharmaceutical Industry. *Handbook of Industrial Crystallization* **2019,** Butterworth-Heinemann: Wobern, MA. 380-413.

41 Zhou, L.; Wang, Z.; Zhang, M.; Guo, M.; Xu, S.; Yin, Q., Determination of Metastable Zone and Induction Time of Analgin for Cooling Crystallization. *Chinese Journal of Chemical Engineering* **2017,** *25* (3), 313-318.

42 Ouyang, J.; Chen, J.; Rosbottom, I.; Chen, W.; Guo, M.; Heng, J. Y. Y., Supersaturation and Solvent Dependent Nucleation of Carbamazepine Polymorphs During Rapid Cooling Crystallization. *CrystEngComm* **2021,** *23* (4), 813-823.

43 Zhou, F.; Shemchuk, O.; Charpentier, M. D.; Matheys, C.; Collard, L.; Ter Horst, J. H.; Leyssens, T., Simultaneous Chiral Resolution of Two Racemic Compounds by Preferential Cocrystallization*. *Angewandte Chemie International Edition* **2021,** *60* (37), 20264-20268.

44 McBride, S. A.; Dash, S.; Khan, S.; Varanasi, K. K., Evaporative Crystallization of Spirals. *Langmuir* **2019,** *35* (32), 10484-10490.

45 Jones, A. G., 3 - Crystallization Principles and Techniques. *Crystallization Process Systems*. Butterworth-Heinemann: Oxford, **2002**; 58-79.

46 Jones, A. G., 4 - Solid-Liquid Separation Processes. *Crystallization Process Systems*, Jones, A. G., Ed. Butterworth-Heinemann: Oxford, **2002**; 80-122.

47 Asherie, N., Protein Crystallization and Phase Diagrams. *Methods* **2004,** *34* (3), 266-272.

48 Chayen, N. E., Turning Protein Crystallisation from an Art into a Science. *Current Opinion in Structural Biology* **2004,** *14* (5), 577-583.

49 Fuchs, D.; Fischer J.; Tumakaka F.; Sadowski, Gabriele., Solubility of Amino Acids: Influence of the pH value and the Addition of Alcoholic Cosolvents on Aqueous Solubility. *Industrial & Engineering Chemistry Research* **2006,** *45*, 6578-6584.

50 Nozaki, Y.; Tanford, C. The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions. *The journal of Biological Chemistry* **1971,** *246* (7), 2211-2217.

51 Ou-uchi, S.; Yang, J.; Lee, J.; Murakawa, Y.; Narita, M., Only Weak Dependence of the Protected Peptides Solubility in Organic Solvents on Their Amino Acid Sequence. *Polymer Journal 1***996,** *28*, 1033-1038.

52 Lu, J.; Wang, X.; Yang, X.; Ching, C., Solubilities of Glycine and Its Oligopeptides in Aqueous Solutions. *Journal of Chemical & Engineering Data* **2006**, *51,* 1593-1596.

53 Castronuovo, G.; Elia, V.; Niccoli, M.; Velleca, F., Simultaneous Determination of Solubility, Dissolution and Dilution Enthalpies of a Substance from a Single Calorimetric Experiment. *Thermochimica Acta* **1998,** *320,* 13-22.

54 Breil, M. P.; Mollerup, J. M.; Rudolph, E. S. J.; Ottens, M.; van der Wielen, L. A. M., Densities and Solubilities of Glycylglycine and Glycyl-L-Alanine in Aqueous Electrolyte Solutions. *Fluid Phase Equilibria* **2004,** *215* (2), 221-225.

55 Do, H. T.; Chua, Y. Z.; Habicht, J.; Klinksiek, M.; Hallermann, M.; Zaitsau, D.; Schick, C.; Held, C., Melting Properties of Peptides and Their Solubility in Water. Part 1: Dipeptides Based on Glycine or Alanine. *RSC Advances* **2019,** *9* (56), 32722-32734.

56 Do, H. T.; Chua, Y. Z.; Habicht, J.; Klinksiek, M.; Volpert, S.; Hallermann, M.; Thome, M.; Pabsch, D.; Zaitsau, D.; Schick, C.; Held, C., 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* **2021,** *60* (12), 4693-4704.

57 Yang, X.; Wang, X.; Ching, C. B., Solubility of Form A and Form Γ of Glycine in Aqueous Solutions. *Journal of Chemical & Engineering Data* **2008,** *53* (5), 1133-1137.

58 Yin, Y.; Bao, Y.; Gao, Z.; Wang, Z.; Liu, D.; Hao, H.; Wang, Y., *Solubility of Cefotaxime Sodium in Ethanol + Water Mixtures under Acetic Acid Conditions*. 2014; Vol. 59, p 1865–1871.

59 Marrero, J.; Gani, R., Group-Contribution Based Estimation of Pure Component Properties. *Fluid Phase Equilibria* **2001,** *183-184*, 183-208.

60 Guo, M.; Chang, Z. H.; Liang, E.; Mitchell, H.; Zhou, L.; Yin, Q.; Guinn, E. J.; Heng, J. Y. Y., 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* **2022,** *352*.

61 Guo, M.; Yin, Q.; Wang, C.; Huang, Y.; Li, Y.; Zhang, Z.; Zhang, X.; Wang, Z.; Zhang, M.; Zhou, L., Thermodynamic Properties of Metamizol Monohydrate in Pure and Binary Solvents at Temperatures from (283.15 to 313.15) K. *Chinese Journal of Chemical Engineering* **2017,** *25* (10), 1481-1491.

62 Zhang, X.; Yin, Q.; Li, X.; Zhang, M.; Huang, J.; Wang, C.; Zhang, Z.; Huang, Y.; Guo, M.; Li, Y., Solubility and Mixing Thermodynamics Properties of Erythromycin Ethylsuccinate in Different Organic Solvents. *Journal of Molecular Liquids* **2017,** *237*, 46-53.

63 Sakthivel, S.; Velusamy, S.; Gardas, R. L.; Sangwai, J. S., Experimental Investigation on the Effect of Aliphatic Ionic Liquids on the Solubility of Heavy Crude Oil Using Uv–Visible, Fourier Transform-Infrared, and 13c Nmr Spectroscopy. *Energy & Fuels* **2014,** *28* (9), 6151-6162.

64 Behera, S., Uv-Visible Spectrophotometric Method Development and Validation of Assay of Paracetamol Tablet Formulation. *Journal of Analytical & Bioanalytical Techniques* **2012,** *03* (06). 1000151

65 Lin, M.; Tesconi, M.; Tischler, M., Use of (1)H Nmr to Facilitate Solubility Measurement for Drug Discovery Compounds. *International Journal of Pharmaceutics* **2009,** *369* (1-2), 47-52.

66 Garcia, S. N.; Yang, X.; Bereczki, L.; Konya, D., Aqueous Solubility of Organic Compounds for Flow Battery Applications: Symmetry and Counter Ion Design to Avoid Low-Solubility Polymorphs. *Molecules* **2021,** *26* (5).

67 Li, Y.; Yin, Q.; Huang, Q.; Guo, M.; Hao, H.; Lou, Y.; Su, N.; He, M.; Wang, Y., Estimation and Confirmation of the Thermodynamic Stability Relationships of the Enantiotropic Polymorphs of Glycolide. *The Journal of Chemical Thermodynamics* **2018,** *118*, 26-33.

68 Huang, J.; Yin, Q.; Hao, H.; Zhang, M.; Zhang, Z.; Guo, M.; Wang, J.; Bao, Y.; Hou, B.; Xie, C., Solid-Liquid Equilibrium of Erythromycin Thiocyanate Dihydrate in Four Mono-Solvents and Three Binary Solvent Mixtures. *Journal of Molecular Liquids* **2017,** *234*, 408-416.

69 Zhou, X.; Fan, j.; Li, N.; Du, Z.; Ying, H.; Wu, J.; Xiong, J.; Bai, J., Solubility of L-Phenylalanine in Water and Different Binary Mixtures from 288.15 to 318.15k. *Fluid Phase Equilibria* **2012,** *316(25),* 26-33.

70 Chen, X.; Zhou, Z.; Chen, J.; Chu, C.; Zheng, J.; Wang, S.; Jia, W.; Zhao, J.; Li, R.; Han, D., Solubility Determination and Thermodynamic Modeling of Buprofezin in Different Solvents and Mixing Properties of Solutions. *Journal of Chemical & Engineering Data* **2019,** *64* (3), 1177-1186.

71 Mu1ller, E. A.; Gubbins, K. E., Molecular-Based Equations of State for Associating Fluids: A Review of Saft and Related Approaches. *Industrial & Engineering Chemistry Research* **2001,** *40*, 2193-2211.

72 Müller, E. A.; Gubbins, K. E., Molecular-Based Equations of State for Associating Fluids: A Review of Saft and Related Approaches. *Industrial & Engineering Chemistry Research* **2001,** *40* (10), 2193-2211.

73 Economou, I. G., Statistical Associating Fluid Theory: A Successful Model for the Calculation of Thermodynamic and Phase Equilibrium Properties of Complex Fluid Mixtures. *Industrial & Engineering Chemistry Research* **2002,** *41* (5), 953-962.

74 Solms, N.; Michelsen, M. L.; Kontogeorgis, G. M., Computational and Physical Performance of a Modified PC-SAFT Equation of State for Highly Asymmetric and Associating Mixtures. *Industrial & Engineering Chemistry Research.* **2003,** *42*, 1098-1105.

75 Gross, J.; Sadowski, G.; Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Industrial & Engineering Chemistry Research.* **2001***, 40,* 1244-1260.

76 Cameretti, L. F.; Sadowski, G., Modeling of Aqueous Amino Acid and Polypeptide Solutions with Pc-Saft. *Chemical Engineering and Processing: Process Intensification* **2008,** *47* (6), 1018-1025.

77 Gil-Villegas, A.; Galindo, A.; Whitehead, P. J.; Mills, S. J.; Jackson, G.; Burgess, A. N., Statistical Associating Fluid Theory for Chain Molecules with Attractive Potentials of Variable Range. *The Journal of Chemical Physics* **1997,** *106* (10), 4168-4186.

78 Anvari, M. H.; Pazuki, G.; Kakhki, S. S.; Bonakdarpour, B., Application of the Saft-Vr Equation of State in Estimation of Physiochemical Properties of Amino Acid Solutions. *Journal of Molecular Liquids* **2013,** *184*, 24-32.

79 Peng, Y.; Goff, K. D.; dos Ramos, M. C.; McCabe, C., Developing a Predictive Group-Contribution-Based Saft-Vr Equation of State. *Fluid Phase Equilibria* **2009,** *277* (2), 131-144.

80 Dufal, S.; Papaioannou, V.; Sadeqzadeh, M.; Pogiatzis, T.; Chremos, A.; Adjiman, C. S.; Jackson, G.; Galindo, A., 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* **2014,** *59* (10), 3272-3288.

81 Di Lecce, S.; Lazarou, G.; Khalit, S. H.; Adjiman, C. S.; Jackson, G.; Galindo, A.; McQueen, L., Modelling and Prediction of the Thermophysical Properties of Aqueous Mixtures of Choline Geranate and Geranic Acid (Cage) Using Saft-Γ Mie. *RSC Advances* **2019,** *9* (65), 38017-38031.

82 Pawel, S.; Martin-Galiano, A. J.; Mikolajka, A.; Girschick, T.; Holak, T.; Frishman, D., *Protein Solubility: Sequence Based Prediction and Experimental Verification*. **2007,** *23*, 2536-42.

83 Wilkinson, D. L.; Harrison, R. G., Predicting the Solubility of Recombinant Proteins in Escherichia Coli. *Bio/Technology* **1991,** *9*, 443-448.

84 Hebditch, M.; Carballo-Amador, M. A.; Charonis, S.; Curtis, R.; Warwicker, J., Protein–Sol: A Web Tool for Predicting Protein Solubility from Sequence. *Bioinformatics* **2017,** *33* (19), 3098-3100.

85 Gibbs, J. W., On the Equilibrium of Heterogeneous Substances. *American Journal of Science* **1878,** 441–458.

86 Weissbuch, I.; Lahav, M.; Leiserowitz, L., Toward Stereochemical Control, Monitoring, and Understanding of Crystal Nucleation. *Crystal Growth & Design* **2003,** *3* (2), 125-150.

87 Lacmann, R., Crystallization, Third Edition. J. W. Mullin, Butterworth-Heinemann, Oxford 1997, 527 Seiten, Zahlr. Abb. Und Isbn 0-7506-3759-5. *Chemie Ingenieur Technik* **2004,** *70* (11), 1468-1468.

88 Kashchiev, D.; Rosmalen G. M. v, Review: Nucleation in Solutions Revisited. *Crystal Research and Technology* **2003,** *38* (7-8), 555-574.

89 Lutsko, J. F., How Crystals Form a Theory of Nucleation Pathways. *Science Advances* **2019,** *5*, eaav7399.

90 Jacobson, L. C.; Hujo W.; Molinero, Valeria., Amorphous Precursors in the Nucleation of Clathrate Hydrates. *Journal of the American Chemical Society* **2010,** *132,* 11806–11811.

91 Vekilov, P. G., Nucleation. *Crystal Growth & Design* **2010,** *10* (12), 5007-5019.

92 Wolde, P. R. t.; Frenkel, D., Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science* **1997,** *277* (5334), 1975-1978.

93 Erdemir, D.; Lee, A. Y.; Myerson, A. S., Nucleation of Crystals from Solution: Classical and Two-Step Models. *Accounts of Chemical Research* **2009,** *42* (5), 621-629.

94 Vekilov, P. G., Dense Liquid Precursor for the Nucleation of Ordered Solid Phases from Solution. *Crystal Growth & Design* **2004,** *4* (4), 671-685.

95 Mirabello, G.; Ianiro, A.; Bomans, P. H. H.; Yoda, T.; Arakaki, A.; Friedrich, H.; de With, G.; Sommerdijk, N., Crystallization by Particle Attachment Is a Colloidal Assembly Process. *Nature Materials* **2020,** *19* (4), 391-396.

96 Alberti, S.; Gladfelter, A.; Mittag, T., Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019,** *176* (3), 419-434.

97 Wang, Y.; Lomakin, A.; Kanai, S.; Alex, R.; Benedek, G. B., Liquid-Liquid Phase Separation in Oligomeric Peptide Solutions. *Langmuir* **2017,** *33* (31), 7715-7721.

98 Yuan, C.; Levin, A.; Chen, W.; Xing, R.; Zou, Q.; Herling, T. W.; Challa, P. K.; Knowles, T. P. J.; Yan, X., Nucleation and Growth of Amino Acid and Peptide Supramolecular Polymers through Liquid-Liquid Phase Separation. *Angewandte Chemie International Edition* **2019,** *58* (50), 18116-18123.

99 Myerson, A. S., Crystallization in the Pharmaceutical and Bioprocessing Industries. *Handbook of Industrial Crystallization.* 2nd ed. **2002,** Butterworth-Heinemann: Wobern, MA. 267-285.

100 Mostafa Nowee, S.; Abbas, A.; Romagnoli, J. A., Antisolvent Crystallization: Model Identification, Experimental Validation and Dynamic Simulation. *Chemical Engineering Science* **2008,** *63* (22), 5457-5467.

101 Wang, H.-Y.; Ward, J. D., Seeding and Optimization of Batch Reactive Crystallization. *Industrial & Engineering Chemistry Research* **2015,** *54* (38), 9360-9368.

102 Dessau, M.; Modis, Y., Protein Crystallization for X-Ray Crystallography. *Journal of Visualized Experiments* **2011,** 47, e2285

103 Alexander, M.; Gavira, J. A., Introduction to Protein Crystallization. *Acta Crystallographica Section F* **2014,** *70* (1), 2-20.

104 D'Arcy, A.; Sweeney, A. M.; Haber, A., Practical Aspects of Using the Microbatch Method in Screening Conditions for Protein Crystallization. *Methods* **2004,** *34* (3), 323-328.

105 Russo Krauss, I.; Merlino, A.; Vergara, A.; Sica, F., An Overview of Biological Macromolecule Crystallization. *International Journal of Molecular Sciences* **2013,** *14* (6), 11643-11691.

106 Zawadzke, L. E; Berg, J. M., The Structure of a Centrosymmetric Protein Crystal. *Proteins: Structure, Function, and Bioinformatics* **1993,** *16* (3), 301-305.

107 Wukovitz, S. W.; Yeates, T. O, Why Protein Crystals Favor Some Space-Groups over Others. *Nature Structural & Molecular Biology* **1995,** *2*, 1062-1067.

108 Mandal, K.; Pentelute, B. L.; Tereshko, V.; Kossiakoff, A. A.; Kent, S. B. H., X-ray Structure of Native Scorpion Toxin BmBKTx1 by Racemic Protein Crystallography Using Direct Methods. *Journal of the American Chemical Society* **2009,** *131*, 4, 1362–1363.

109 Yeates, T. O.; Kent, S. B. H., Racemic Protein Crystallography. *Annual Review of Biophysics* **2012,** *41*(1), 41-61.

110 Gógl, G.; Törő, I.; Reményi, A., Protein-Peptide Complex Crystallization: A Case Study on the Erk2 Mitogen-Activated Protein Kinase. *Acta crystallographica Section D, Biological crystallography* **2013,** *69* (Pt 3), 486-489.

111 Gobeaux, F.; Tarabout, C.; Fay, N.; Meriadec, C.; Ligeti, M.; Buisson, D. A.; Cintrat, J. C.; Artzner, F.; Paternostre, M., Directing Peptide Crystallization through Curvature Control of Nanotubes. *Journal of Peptide Science* **2014,** *20* (7), 508-516.

112 Aaron, J., *Structure, Function and Applications of Metal-Requiring Enzymes: Carbonic Anhydrase and Epi-Isozizaene Synthase*. **2010**.

113 Lucaioli, P.; Nauha, E.; Singh, I.; Blagden, N., First Steps for the Direct Purification of L-Leu-L-Leu Dipeptide through Co-Crystallization. *Crystal Growth & Design* **2017,** 18, 1062-1069.

114 Pechkova, E.; Nicolini, C., Accelerated Protein Crystal Growth by Protein Thin Film Template. *Journal of Crystal Growth* **2001,** *231* (4), 599-602.

115 Artusio, F.; Pisano, R., Surface-Induced Crystallization of Pharmaceuticals and Biopharmaceuticals: A Review. *International Journal of Pharmaceutics* **2018,** *547* (1-2), 190-208.

116. Link, F. J. & Heng, J. Y. Y. Enhancing the crystallisation of insulin using amino acids as soft-templates to control nucleation. *CrystEngComm* **23**, 3951–3960 (2021).

117. Chen, W. *et al.* High Protein-Loading Silica Template for Heterogeneous Protein Crystallization. *Crystal Growth & Design* **20**, 866–873 (2019).

118 Li, F.; Lakerveld, R., Influence of Alternating Electric Fields on Protein Crystallization in Microfluidic Devices with Patterned Electrodes in a Parallel-Plate Configuration. *Crystal Growth & Design* **2017,** *17* (6), 3062-3070.

119 Wakayama, N. I., Effects of a Strong Magnetic Field on Protein Crystal Growth. *Crystal Growth & Design* **2003,** *3*(1), 17–24.

120 Jordens, J.; Canini, E.; Gielen, B.; Van Gerven, T.; Braeken, L., Ultrasound Assisted Particle Size Control by Continuous Seed Generation and Batch Growth. *Crystals* **2017,** *7* (7), 195.

121 Kozhemyakin, G.; V. Zolkina, L.; Inatomi, Y., *Influence of Ultrasound on Crystal Growth from Solution and Related Flow Visualization*. **2006**; Vol. 6.

122 Yee, A. A.; Savchenko, A.; Ignachenko, A.; Lukin, J.; Xu, X.; Skarina, T.; Evdokimova, E.; Liu, C. S.; Semesi, A.; Guido, V.; Edwards, A. M.; Arrowsmith, C. H., NMR and X-Ray Crystallography, Complementary Tools in Structural Proteomics of Small Proteins. *Journal of the American Chemical Society* **2005,** *127*, 16512-16517.

123 Glattli, A.; van Gunsteren, W. F., Are NMR-Derived Model Structures for Beta-Peptides Representative for the Ensemble of Structures Adopted in Solution? *Angewandte Chemie International Edition* **2004,** *43* (46), 6312-6.

124 Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H., Olex2: A Complete Structure Solution, Refinement and Analysis Program. *Journal of Applied Crystallography* **2009,** *42* (2), 339-341.

125 Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T., Dl_Poly_3: New Dimensions in Molecular Dynamics Simulations Via Massive Parallelism. *Journal of Materials Chemistry* **2006,** *16* (20), 1911–1918.

126 Chauhan, A., Powder XRD Technique and Its Applications in Science and Technology. *Journal of Analytical & Bioanalytical Techniques* **2014,** *5* (6). 1000212

127 Borghetti, G. S.; Carini, J. P.; Honorato, S. B.; Ayala, A. P.; Moreira, J. C. F.; Bassani, V. L., Physicochemical Properties and Thermal Stability of Quercetin Hydrates in the Solid State. *Thermochimica Acta* **2012,** *539*, 109-114.

128 Lopez, M. M.; Makhatadze, G. I., Differential Scanning Calorimetry. In *Calcium-Binding Protein Protocols: Volume 2: Methods and Techniques*, Springer: **2002,** 113-119.

129 Spink, C. H., Differential Scanning Calorimetry. In *Biophysical Tools for Biologists, Volume One: In Vitro Techniques*, **2008;** 115-141.

130 Baltes, L.; Costiuc, L.; Patachia, S.; Tierean, M., Differential Scanning Calorimetry—a Powerful Tool for the Determination of Morphological Features of the Recycled Polypropylene. *Journal of Thermal Analysis and Calorimetry* **2019,** *138* (4), 2399-2408.

131 Lothenbach, B.; Durdziński, P.; Weerdt, K. D., Thermogravimetric analysis. *A Practical Guide to Microstructural Analysis of Cementitious Materials* 1st ed. **2002,** Boca Raton 267-285.

132 Saadatkhah, N.; Carillo Garcia, A.; Ackermann, S.; Leclerc, P.; Latifi, M.; Samih, S.; Patience, G. S.; Chaouki, J., Experimental Methods in Chemical Engineering: Thermogravimetric Analysis—Tga. *The Canadian Journal of Chemical Engineering* **2019,** *98* (1), 34-43.

133 Jones, R. R.; Hooper, D. C.; Zhang, L.; Wolverson, D.; Valev, V. K., Raman Techniques: Fundamentals and Frontiers. *Nanoscale Research Letters* **2019,** *14* (1), 231.

134 Huang, M. C.; Chen, W. H.; Huang, C. W.; Huang, K. Y.; Horng, J. C.; Hayashi, M.; Chen, I. C., Investigation of the Cis–Trans Structures and Isomerization of Oligoprolines by Using Raman Spectroscopy and Density Functional Theory Calculations: Solute–Solvent Interactions and Effects of Terminal Positively Charged Amino Acid Residues. *RSC Advances* **2020,** *10* (57), 34493-34500.

135 Schweitzer, S.; Reinhard; Eker; Fatma; Griebenow; Kai; Cao; Xiaolin; Nafie; Laurence, A., The Conformation of Tetraalanine in Water Determined by Polarized Raman, FTIR, and VCD Spectroscopy. *Journal of the American Chemical Society* **2004**, *126*, 9, 2768–2776

136 Bykov, S.; Asher, S., Raman Studies of Solution Polyglycine Conformations. *The Journal of Physical Chemistry B* **2010,** *114*, 6636–6641.

137 Bykov, S. V.; Asher, S. A., Uv Resonance Raman Elucidation of the Terminal and Internal Peptide Bond Conformations of Crystalline and Solution Oligoglycines. *The Journal of Physical Chemistry Letters* **2010,** *1* (1), 269-271.

138 Zhang, S.; Schweitzer-Stenner, R.; Urbanc, B., Do Molecular Dynamics Force Fields Capture Conformational Dynamics of Alanine in Water? *Journal of Chemical Theory and Computation* **2020,** *16* (1), 510-527.

139 Schweitzer-Stenner, R.; Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A., The Conformation of Tetraalanine in Water Determined by Polarized Raman, Ft-Ir, and Vcd Spectroscopy. *Journal of the American Chemical Society* **2004,** *126*, 2768-2776.

140 Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K., Dihedral Angles of Trialanine in D2O Determined by Combining FTIR and Polarized Visible Raman Spectroscopy. *Journal of the American Chemical Society* **2001,** *123*, 39, 9628–9633

141 Schweitzer-Stenner, R.; Measey, T.; Kakalis, L.; Jordan, F.; Pizzanelli, S.; Forte, C., Conformations of Alanine-Based Peptides in Water Probed by FTIR, Raman, Vibrational Circular Dichroism, Electronic Circular Dichroism, and NMR Spectroscopy. *Biochemistry* **2007**, *46, 6*, 1587–1596

142 Mu, Y.; Stock, G., Conformational Dynamics of Trialanine in Water: A Molecular Dynamics Study The Journal of Physical Chemistry B J. Phys. Chem. B **2002**, 106, 5294-5301

143 Meral, D.; Toal, S.; Schweitzer-Stenner, R.; Urbanc, B., Water-Centered Interpretation of Intrinsic Ppii Propensities of Amino Acid Residues: In Vitro-Driven Molecular Dynamics Study. *The Journal of Physical Chemistry B* **2015,** *119* (42), 13237-13251.

144 Salvi, N.; Abyzov, A.; Blackledge, M., Analytical Description of Nmr Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. *Angewandte Chemie International Edition* **2017,** *56* (45), 14020-14024.

145 Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L., Improved Peptide and Protein Torsional Energetics with the Oplsaa Force Field. *Journal of Chemical Theory and Computation* **2015,** *11* (7), 3499-3509.

146 Sun, H., COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase Applicationss-Overview with Details on Alkane and Benzene Compounds. *The Journal of Physical Chemistry B* **1998,** *102*, 7338-7364.

147 Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H., Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. *Journal of the American Chemical Society* **2007**, 129, 5, 1179–1189.

148 Marqus, S.; Pirogova, E.; Piva, T. J., Evaluation of the Use of Therapeutic Peptides for Cancer Treatment. *Journal of Biomedical Science* **2017,** *24* (1), 21.

149 Todd, J. F.; Bloom, S. R., Incretins and Other Peptides in the Treatment of Diabetes. *Diabetic Medicine* **2007,** *24* (3), 223-232.

150 Cicero, A. F. G.; Fogacci, F.; Colletti, A., Potential Role of Bioactive Peptides in Prevention and Treatment of Chronic Diseases: A Narrative Review. *British Journal of Pharmacology* **2017,** *174* (11), 1378-1394.

151 Zhou, M.; Qian, Y.; Xie, J.; Zhang, W.; Jiang, W.; Xiao, X.; Chen, S.; Dai, C.; Cong, Z.; Ji, Z.; Shao, N.; Liu, L.; Wu, Y.; Liu, R., Poly(2-Oxazoline)-Based Functional Peptide Mimics: Eradicating Mrsa Infections and Persisters While Alleviating Antimicrobial Resistance. *Angewandte Chemie International Edition* **2020,** *59* (16), 6412-6419.

152 Barry, D. M.; Liu, X. T.; Liu, B.; Liu, X. Y.; Gao, F.; Zeng, X.; Liu, J.; Yang, Q.; Wilhelm, S.; Yin, J.; Tao, A.; Chen, Z. F., Exploration of Sensory and Spinal Neurons Expressing Gastrin-Releasing Peptide in Itch and Pain Related Behaviors. *Nature Communication* **2020,** *11* (1), 1397.

153 Carlini, A. S.; Gaetani, R.; Braden, R. L.; Luo, C.; Christman, K. L.; Gianneschi, N. C., Enzyme-Responsive Progelator Cyclic Peptides for Minimally Invasive Delivery to the Heart Post-Myocardial Infarction. *Nature Communication* **2019,** *10* (1), 1735.

154 Otvos, L., Jr.; Wade, J. D., Current Challenges in Peptide-Based Drug Discovery. *Frontiers in Chemistry* **2014,** *2*, 62.

155 Giffard, M.; Ferte, N.; Ragot, F.; El Hajji, M.; Castro, B.; Bonnete, F., Urate Oxidase Purification by Salting-in Crystallization: Towards an Alternative to Chromatography. *PLoS One* **2011,** *6* (5), e19013.

156 Han, D.; Liu, Y.; Li, X.; Zhang, T.; Li, Z.; Du, S.; Guo, M.; Xu, S.; Yu, B.; Hou, B.; Gong, J., Effect of B-Alanine and the Solvent Composition on the Solubility of Solvate of Calcium D-Pantothenate Containing Four Molecules of Methanol and One Molecule of Water (D-Pc·4meoh·1h2o). *The Journal of Chemical Thermodynamics* **2017,** *106*, 36-46.

157 Tessari, P.; Lante, A.; Mosca, G., Essential Amino Acids: Master Regulators of Nutrition and Environmental Footprint? *Scientific Report* **2016,** *6*, 26074.

158 Hazaveie, S. M.; Sodeifian, G.; Sajadian, S. A., Measurement and Thermodynamic Modeling of Solubility of Tamsulosin Drug (Anti Cancer and Anti-Prostatic Tumor Activity) in Supercritical Carbon Dioxide. *The Journal of Supercritical Fluids* **2020,** *163*, 104875

159 Zhao, S.; Ma, Y.; Tang, W., Thermodynamic Analysis and Molecular Dynamic Simulation of Solid-Liquid Phase Equilibrium of Griseofulvin in Three Binary Solvent Systems. *Journal of Molecular Liquids* **2019,** *294*.

160 Hussain, A.; Altamimi, M. A.; Alshehri, S.; Imam, S. S., Assessment of Solubility and Hansen Solubility Parameters of Rifampicin in Various Permeation Enhancers: Experimental and Computational Approach. *Journal of Molecular Liquids* **2021,** *328*.

161 Shi, P.; Ma, Y.; Han, D.; Du, S.; Zhang, T.; Li, Z., Uncovering the Solubility Behavior of Vitamin B6 Hydrochloride in Three Aqueous Binary Solvents by Thermodynamic Analysis and Molecular Dynamic Simulation. *Journal of Molecular Liquids* **2019,** *283*, 584-595.

162 Mamouei, M.; Budidha, K.; Baishya, N.; Qassem, M.; Kyriacou, P. A., An Empirical Investigation of Deviations from the Beer-Lambert Law in Optical Estimation of Lactate. *Sci Rep* **2021,** *11* (1), 13734.

163 Park, G.; Kim, T.; Lee, Y.-W., A Method for Measuring the Solubility of Disperse Red 60 in Supercritical Carbon Dioxide Using Variable-Volume View Cell with in-Situ Uv–Vis Spectrometer. *The Journal of Supercritical Fluids* **2021,** *176*.

164 Guo, M.; Rosbottom, I.; Zhou, L.; Yong, C. W.; Zhou, L.; Yin, Q.; Todorov, I. T.; Errington, E.; Heng, J. Y. Y., Triglycine (Ggg) Adopts a Polyproline II (pPII) Conformation in Its Hydrated Crystal Form: Revealing the Role of Water in Peptide Crystallization. *The Journal of Physical Chemistry Letters* **2021**, 8416-8422.

165 Kontogeorgis, G. M.; Dohrn, R.; Economou, I. G.; de Hemptinne, J. C.; Ten Kate, A.; Kuitunen, S.; Mooijer, M.; Zilnik, L. F.; Vesovic, V., Industrial Requirements for Thermodynamic and Transport Properties: 2020. *Industrial & Engineering Chemistry Research* **2021,** *60* (13), 4987-5013.

166 Wei, S. Y.; Sadus, R. J., Equations of State for the Calculation of Fluid-Phase Equilibria. *Thermodynamics* **2000**, *46*(1), 169-196.

167 Haslam, A. J.; Galindo, A.; Jackson, G., Prediction of Binary Intermolecular Potential Parameters for Use in Modelling Fluid Mixtures. *Fluid Phase Equilibria* **2008,** *266* (1-2), 105-128.

168 Chua, Y. Z.; Do, H. T.; Schick, C.; Zaitsau, D.; Held, C., New Experimental Melting Properties as Access for Predicting Amino-Acid Solubility. *RSC Advances* **2018,** *8* (12), 6365-6372.

169 Sun, H.; Wang, L.; Liu, B., Solubility of A-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* **2017,** *434*, 167-175.

170 Fuchs D.; Fischer, Jan.; Tumakaka, F.; Sadowski, G., Solubility of Amino Acids Influence of the pH Value and the Addition of Alcoholic Cosolvents on Aqueous Solubility. *Industrial & Engineering Chemistry Research* **2006,** *45*(19), 6578–6584.

171 Joback, K. G.; Reid, R. C., Estimation of Pure-Component Properties from Group-Contributions. *Chemical Engineering Communications* **2007,** *57* (1-6), 233-243.

172 Marreroa, J.; Gani, Rafiqul., Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria* **2001,** *183–184* (1), 183-208.

173 Karthika, S.; Radhakrishnan, T. K.; Kalaichelvi, P., A Review of Classical and Nonclassical Nucleation Theories. *Crystal Growth & Design* **2016,** *16* (11), 6663-6681.

174 Vekilov, P. G., Nucleation. *Crystal Growth & Design* **2010,** *10* (12), 5007-5019.

175 Kulkarni, S. A.; Kadam, S. S.; Meekes, H.; Stankiewicz, A. I.; ter Horst, J. H., Crystal Nucleation Kinetics from Induction Times and Metastable Zone Widths. *Crystal Growth & Design* **2013,** *13* (6), 2435-2440.

176 Andre´ Bernardo, C. E. C., and Everson Alves Miranda, Induction Time as an Instrument to Enhance Comprehension of Protein Crystallization. *Crystal Growth & Design* **2004,** *4,* 4, 799–805.

177 Jiang, S.; ter Horst, J. H., Crystal Nucleation Rates from Probability Distributions of Induction Times. *Crystal Growth & Design* **2010,** *11* (1), 256-261.

178 Davey, R. J.; Schroeder, S. L.; ter Horst, J. H., Nucleation of Organic Crystals--a Molecular Perspective. *Angewandte Chemie International Edition in English* **2013,** *52* (8), 2166-2179.

179 Dunning, W.; Notley, N., Kinetics of Crystallization. III. *Zeitschrift für Elektrochemie, Berichte der Bunsengesellschaft für physikalische Chemie* **2010,** *61*, 55-59.

180 Yang, Y.; Zhou, L.; Zhang, X.; Yang, W.; Zhang, S.; Xiong, L.; Wei, Y.; Zhang, M.; Hou, B.; Yin, Q., Influence of Solvent Properties and Intermolecular Interaction between Solute and Solvent on Nucleation Kinetics of Hmbtad. *Journal of Crystal Growth* **2018,** *498*, 77-84.

181 Omar, W.; Mohnicke, M.; Ulrich, J., Determination of the Solid Liquid Interfacial Energy and Thereby the Critical Nucleus Size of Paracetamol in Different Solvents. *Crystal Research and Technology* **2006,** *41* (4), 337-343.

182 Thanh, N. T. K.; Maclean, N.; Mahiddine, S., Mechanisms of Nucleation and Growth of Nanoparticles in Solution. *Chemical Reviews*, **2014,** *114* (15), 7610–7630.

183 Waid Omar, M. M., and Joachim Ulrich, Determination of the Solid Liquid Interfacial Energy and Thereby the Critical Nucleus Size of Paracetamol in Different Solvents. *Crystal Research & Technology*, **2006,** *41*(4):337 – 343.

184 Wei, L.; Zhao, Y.; Hu, X.; Tang, L., Redox-Responsive Polycondensate Neoepitope for Enhanced Personalized Cancer Vaccine. *ACS Central Science* **2020,** *6* (3), 404-412.

185 Sicinski, K. M.; Montanari, V.; Raman, V. S.; Doyle, J. R.; Harwood, B. N.; Song, Y. C.; Fagan, M. P.; Rios, M.; Haines, D. R.; Kopin, A. S.; Beinborn, M.; Kumar, K., A Non-Perturbative Molecular Grafting Strategy for Stable and Potent Therapeutic Peptide Ligands. *ACS Central Science* **2021,** *7* (3), 454-466.

186 Jappe, E. C.; Garde, C.; Ramarathinam, S. H.; Passantino, E.; Illing, P. T.; Mifsud, N. A.; Trolle, T.; Kringelum, J. V.; Croft, N. P.; Purcell, A. W., Thermostability Profiling of Mhc-Bound Peptides: A New Dimension in Immunopeptidomics and Aid for Immunotherapy Design. *Nature Communications* **2020,** *11* (1), 6305.

187 Zhou, P.; Wang, J.; Wang, M.; Hou, J.; Lu, J. R.; Xu, H., Amino Acid Conformations Control the Morphological and Chiral Features of the Self-Assembled Peptide Nanostructures: Young Investigators Perspective. *Journal of Colloid and Interface Science* **2019,** *548*, 244-254.

188 Na, J. H.; Lee, W. K.; Yu, Y. G., How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. *International Journal of Molecular Sciences* **2018,** *19* (2), 381.

189 Chaudhary, H.; Fernandes, R. M. F.; Gowda, V.; Claessens, M.; Furo, I.; Lendel, C., Intrinsically Disordered Protein as Carbon Nanotube Dispersant: How Dynamic Interactions Lead to Excellent Colloidal Stability. *Journal of Colloid and Interface Science* **2019,** *556*, 172-179.

190 Alexei, A. A.; Michael, J. E. S.; Alexander, A. M., Polyproline- II Helix in Proteins: Structure and Function. *Journal of Molecular Biology* **2013**.

191 Drozdov, A. N.; Grossfield, A.; Pappu, R. V., Role of Solvent in Determining Conformational Preferences of Alanine Dipeptide in Water. *Journal of the American Chemical Society* **2004,** *126*, 2574-2581.

192 Hamburger, J. B.; Ferreon, J. C.; Whitten, S. T.; Hilser, V. J., Thermodynamic Mechanism and Consequences of the Polyproline II (PII) Structural Bias in the Denatured States of Proteins. *Biochemistry* **2004,** *43*, 9790-9799.

193 Measey, T.; Schweitzer-Stenner, R., Simulation of Amide I′ Band Profiles of Trans Polyproline Based on an Excitonic Coupling Model. *Chemical Physics Letters* **2005,** *408*, 123-127.

194 Gorbunov, R. D.; Nguyen, P. H.; Kobus, M.; Stock, G., Quantum-Classical Description of the Amide I Vibrational Spectrum of Trialanine. *The Journal of Chemical Physics* **2007,** *126*, 054509

195 Siobhan, T.; Derya, M.; Daniel, V.; Brigita, U.; Reinhard, S.-S., Ph-Independence of Trialanine and the Effects of Termini Blocking in Short Peptides: A Combined Vibrational, NMR, UVCD, and Molecular Dynamics Study. *The Journal of Physical Chemistry B* **2013,** *117*, 3689−3706.

196 Mu, Y.; Kosov, D. S.; Stock, G., Conformational Dynamics of Trialanine in Water. 2. Comparison of Amber, Charmm, Gromos, and Opls Force Fields to NMR and Infrared Experiments. *The Journal of Physical Chemistry B* **2003,** *107*, 5064-5073.

197 Mikhonin; Aleksandr, V.; Myshakina; Nataliya, S.; Bykov; Sergei, V.; Asher; Sanford, A., Uv Resonance Raman Determination of Polyproline Ii, Extended 2.5(1)-Helix, and Beta-Sheet Psi Angle Energy Landscape in Poly-L-Lysine and Poly-L-Glutamic Acid. *Journal of the American Chemical Society* **2005,** *127*, 21, 7712–7720.

198 Ilawe, N. V.; Schweitzer-Stenner, R.; DiGuiseppi, D.; Wong, B. M., Is a Cross-Beta-Sheet Structure of Low Molecular Weight Peptides Necessary for the Formation of Fibrils and Peptide Hydrogels? *Physical Chemistry Chemical Physics* **2018,** *20* (27), 18158-18168.

199 Levy, Y.; Onuchic, J. N., Water and Proteins: A Love-Hate Relationship. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (10), 3325-3326.

200 Brian, A.; Shuting, Z.; Reinhard, S.-S.; Brigita, U., Glycine in Water Favors the Polyproline II State. *Biomolecules* **2020,** *10*, 1121-1141.

201 Angel, E. G., Characterization of Non-Alpha Helical Conformations in Ala Peptides. *Polymer* **2004**, *45* (2), 669-676.

202 Brian, A.; Shuting, Z.; Reinhard, S.-S.; Brigita, U., Glycine in Water Favors the Polyproline II State. *Biomolecules* **2020,** *10*(8), 1121

203 Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D., Polyproline II Helix Is the Preferred Conformation for Unfolded Polyalanine in Water. *Proteins* **2004,** *55* (3), 502-507.

204 Fleming, P. J.; Fitzkee, N. C.; Mezei, M.; Srinivasan, R.; Rose, G. D., A Novel Method Reveals That Solvent Water Favors Polyproline Ii over Beta-Strand Conformation in Peptides and Unfolded Proteins: Conditional Hydrophobic Accessible Surface Area (Chasa). *Protein Science* **2005,** *14* (1), 111-8.

205 Ilawe, N. V.; Raeber, A. E.; Schweitzer-Stenner, R.; Toal, S. E.; Wong, B. M., Assessing Backbone Solvation Effects in the Conformational Propensities of Amino Acid Residues in Unfolded Peptides. *Physical Chemistry Chemical Physics* **2015,** *17* (38), 24917-24.

206 Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R., Intrinsic Propensities of Amino Acid Residues in GxG. *Journal of the American Chemical Society* **2010,** *132*, 540–551.

207 Siegrist, K.; Bucher, C. R.; Mandelbaum, I.; Walker, A. R. H.; Balu, R.; Gregurick, S. K.; Plusquellic, D. F., High-Resolution Terahertz Spectroscopy of Crystalline Trialanine: Extreme Sensitivity to Beta-Sheet Structure and Cocrystallized Water. *Journal of the American Chemical Society* **2006,** *128*, 5764-5775.

208 Meulemans, P. R.; Meerssche, P. P. E. M. V., Structure Cristalline De Derives D'acides Amines. Iv. Le Compose D'addition Libr. Glycylglycylglycine. *Acta Crystallographica* **1971,** (27), 1187-1190.

209 Bykov, S.; Asher, S., Raman Studies of Solution Polyglycine Conformations. *The Journal of Physical Chemistry B* **2010,** *114*, 6636–6641.

210 Meulemans, P. R.; Meerssche, P. P. E. M. V., Structure Cristalline De Derives D'acides Amines. Iv. Le Compose D'addition Libr. Glycylglycylglycine. *Acta Crystallographica* **1971,** (27), 1187-1190.

211 Kameda, T.; Takeda, N.; Ando, S.; Ando, I.; Hashizume, D.; Ohashi, Y., Structure of Glygly Peptide in the Crystalline State as Studied by X-Ray Diffraction and Solid State 13c-NMR Methods. *Biopolymers* **1998,** *45*, 333–339.

212 Drebushchak, T. N.; Kolesnik, E. N.; Boldyreva, E. V., Variable Temperature (100—295 K) Single-Crystal X-Ray Diffraction Study of the A-Polymorph of Glycylglycine and a Glycylglycine Hydrate. *Zeitschrift für Kristallographie – Crystalline Materials* **2006,** *221* (2), 128-138.

213 Smith, A. J.; Ali, F. I.; Soldatov, D. V., Glycine Homopeptides: The Effect of the Chain Length on the Crystal Structure and Solid State Reactivity. *CrystEngComm* **2014,** *16* (31), 7196-7208.

214 Gunsteren, W. F. v.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; Vegt, N. F. A. V. D.; Yu, H. B., Biomolecular Modeling: Goals, Problems, Perspectives. *Angewandte Chemie International Edition* **2006,** *45* (25), 4064-4092.

215 Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardere, L.; Schnieders, M. J.; Piquemal, J. P.; Ren, P. Y.; Ponder, J. W., Tinker 8: Software Tools for Molecular Design. *Journal of Chemical Theory and Computation* **2018,** *14* (10), 5273-5289.

216 Klitou, P.; Rosbottom, I.; Simone, E., Synthonic Modeling of Quercetin and Its Hydrates: Explaining Crystallization Behavior in Terms of Molecular Conformation and Crystal Packing. *Crystal Growth and Design* **2019,** *19* (8), 4774-4783.

217 Mirkin, N. G.; Krimm, S., Water Interaction Differences Determine the Relative Energetic Stability of the Polyproline II Conformation of the Alanine Dipeptide in Aqueous Environments. *Biopolymers* **2012,** *97* (10), 789-794.

218 Thompson, H. P. G.; Day, G. M., Which Conformations Make Stable Crystal Structures? Mapping Crystalline Molecular Geometries to the Conformational Energy Landscape. *Chemical Science* **2014,** *5* (8), 3173-3182.

219 Galek, P. T.; Fabian, L.; Motherwell, W. D.; Allen, F. H.; Feeder, N., Knowledge-Based Model of Hydrogen-Bonding Propensity in Organic Crystals. *Acta Crystallographica Section B* **2007,** *63* (Pt 5), 768-782.

220 Cruz-Cabeza, A. J.; Bernstein, J., Conformational Polymorphism. *Chemical Reviews* **2014,** *114* (4), 2170-91.

221 Chakraborty, D.; Sengupta, N.; Wales, D. J., Conformational Energy Landscape of the Ritonavir Molecule. *Journal of Physical Chemistry B* **2016,** *120* (19), 4331-40.

222 Chaudhuri, P.; Canuto, S., Conformational Behavior of Different Possible Ways of Oligoglycine Formation in a Solvent-Free Environment. *Journal of Molecular Structure: THEOCHEM* **2008,** *849*, 25-32.

223 Yu, W.; Wu, Z.; Chen, H.; Liu, X.; MacKerell, A. D., Jr.; Lin, Z., Comprehensive Conformational Studies of Five Tripeptides and a Deduced Method for Efficient Determinations of Peptide Structures. *Journal of Physical Chemistry B* **2012,** *116* (7), 2269-2283.

224 Bush, I. J.; Todorov, I. T.; Smith, W., A Daft Dl_Poly Distributed Memory Adaptation of the Smoothed Particle Mesh Ewald Method. *Computer Physics Communications* **2006,** *175* (5), 323-329.

225 Yong, C. W., Descriptions and Implementations of Dl_F Notation: A Natural Chemical Expression System of Atom Types for Molecular Simulations. *Journal of Chemical Information and Modeling* **2016,** *56* (8), 1405-1409.

226 Besler, B. H., Atomic Charges Derived from Semiempirical Methods. *Journal of Computational Chemistry* **1989,** *11* (4), 431-439.

227 Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *The Journal of Chemical Physics* **1993,** *98* (7), 5648-5652.

228 Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B: Condensed Matter and Materials Physics* **1988,** *37* (2), 785-789.

229 Posch, H. A.; Hoover, W. G.; Vesely, F. J., Canonical Dynamics of the Nose Oscillator: Stability, Order, and Chaos. *Physical review A : General physics* **1986,** *33* (6), 4253-4265.

230 Hoover, W. G.; Holian, B. L., Kinetic Moments Method for the Canonical Ensemble Distribution. *Physics Letters A* **1996,** *211*, 253-257.

231 Yong, C. W.; Todorov, I. T., Dl_Analyser Notation for Atomic Interactions (Danai): A Natural Annotation System for Molecular Interactions, Using Ethanoic Acid Liquid as a Test Case. *Molecules* **2017,** *23* (1), 36-50.

232 Rosbottom, I.; Yong, C. W.; Geatches, D. L.; Hammond, R. B.; Todorov, I. T.; Roberts, K. J., The Integrated Dl_Poly/Dl_Field/Dl_Analyser Software Platform for Molecular Dynamics Simulations for Exploration of the Synthonic Interactions in Saturated Benzoic Acid/Hexane Solutions. *Molecular Simulation* **2019**, 1-16.

233 Zhang; Jiajing; King; Michael; Suggs; Laura; Ren; Pengyu, Molecular Modeling of Conformational Properties of Oligodepsipeptides. *Biomacromolecules* **2007,** 8, 10, 3015–3024.

234 Xu, W.; Zhu, Q.; Hu, C. T., The Structure of Glycine Dihydrate: Implications for the Crystallization of Glycine from Solution and Its Structure in Outer Space. *Angewandte Chemie International Edition* **2017,** *56* (8), 2030-2034.

235 Cerreia Vioglio, P.; Mollica, G.; Juramy, M.; Hughes, C. E.; Williams, P. A.; Ziarelli, F.; Viel, S.; Thureau, P.; Harris, K. D. M., Insights into the Crystallization and Structural Evolution of Glycine Dihydrate by in Situ Solid-State NMR Spectroscopy. *Angewandte Chemie International Edition* **2018,** *57* (22), 6619-6623.

Appendix

Table 4.1 The calibration points and the standard deviation of studied amino acids and peptides

Concentration (mg/mL)	Average Absorbance	Standard Deviation
	Glycine	
0.09760	0.61	0.023
0.12200	0.73	0.054
0.14640	0.80	0.038
0.17080	0.97	0.022
0.19520	1.04	0.012
0.21960	1.23	0.019
0.24400	1.29	0.027
0.29280	1.56	0.013
0.34160	1.77	0.020
0.43920	2.10	0.040
0.53680	2.47	0.022
	Diglycine	
0.00722	0.15	0.037
0.00963	0.23	0.021
0.01204	0.32	0.029
0.01444	0.50	0.040
0.01685	0.56	0.014
0.01926	0.67	0.046
0.02167	0.76	0.028
0.02407	0.85	0.031
0.02648	0.98	0.014

0.03129	1.21	0.028
0.03370	1.34	0.041
0.03851	1.55	0.021
0.04333	1.75	0.045
0.04814	2.00	0.015
0.05777	2.35	0.021
0.06740	2.70	0.021
0.08665	3.00	0.038
0.09628	3.35	0.011
0.10592	3.70	0.043
0.11554	4.00	0.024
0.13479	4.52	0.016
0.15405	4.88	0.017
	Triglycine	
0.02259	1.43	0.028
0.02485	1.55	0.031
0.02937	1.77	0.019
0.03163	1.90	0.020
0.03389	1.95	0.031
0.03615	2.08	0.011
0.03840	2.16	0.012
0.04066	2.22	0.023
0.04292	2.38	0.030
0.05196	2.87	0.011
0.06100	3.22	0.022
0.07003	3.63	0.024
0.07907	3.97	0.028

	Tetraglycine	
0.01197	0.97	0.030
0.01317	1.10	0.012
0.01437	1.25	0.013
0.01557	1.30	0.032
0.01916	1.59	0.031
0.02155	1.80	0.041
0.02395	1.92	0.013
0.02395	1.93	0.023
0.02874	2.10	0.021
0.03113	2.29	0.011
0.03592	2.40	0.013
0.04550	3.16	0.034
0.05029	3.36	0.026
0.05508	3.75	0.031
0.05987	3.85	0.016
0.06466	4.16	0.011
0.06945	4.40	0.016
0.07424	4.57	0.018
	Pentaglycine	
0.00967	0.80	0.021
0.01544	1.22	0.031
0.01985	1.60	0.014
0.02480	1.90	0.010
0.02914	2.30	0.041
	Hexaglycine	
0.01000	0.66	0.025

0.01500	0.80	0.031
0.02000	1.06	0.034
0.02500	1.32	0.022
0.03000	1.52	0.029
0.03500	1.80	0.015
0.04000	2.04	0.013
0.04500	2.28	0.024
0.05000	2.50	0.019
0.05500	2.74	0.018
0.06000	2.88	0.012
	Aspartic Acid	
0.11682	0.70	0.029
0.08524	0.45	0.015
0.06616	0.30	0.016
0.05360	0.22	0.032
	Histidine	
0.05000	1.33	0.012
0.10000	3.10	0.023
0.20000	6.48	0.018
0.03630	0.83	0.024
0.04000	1.02	0.013
0.08000	2.38	0.011
0.14000	4.40	0.026
	Phenylalanine	
8.21294	8.89	0.017
5.47530	5.98	0.034
4.10722	4.30	0.015

3.28578	3.30	0.018
2.05361	2.00	0.014
3.00000	3.06	0.028
5.00000	5.32	0.016
7.00000	7.57	0.012
	Tyrosine	
0.01000	1.68	0.029
0.00800	1.33	0.021
0.00574	1.05	0.022
0.00503	0.94	0.011
0.00446	0.89	0.012
0.00600	1.09	0.013
0.00700	1.23	0.015
0.00900	1.52	0.042
	Aps-Phe	
2.49350	1.68	0.022
2.45435	1.66	0.016
1.22717	0.90	0.013
0.81812	0.60	0.026
0.61358	0.46	0.018
0.49087	0.33	0.014
1.00000	0.71	0.024
0.70000	0.51	0.035
2.00000	1.37	0.019
	Phe-Phe	
0.00788	1.90	0.012
0.00656	1.60	0.013

0.00563	1.30	0.011
0.00492	1.10	0.021
0.00438	1.00	0.025
0.00394	0.90	0.019
0.00358	0.80	0.032
0.00328	0.70	0.013
0.00303	0.65	0.027
0.00281	0.60	0.029
0.00263	0.55	0.024
0.00246	0.55	0.017
0.00219	0.50	0.016
0.00197	0.44	0.023
	Gly-Tyr	
0.00333	0.58	0.025
0.00443	0.80	0.031
0.00411	0.73	0.014
0.00285	0.49	0.018
0.00374	0.65	0.029
	Gly-Asp	
0.00334	0.27	0.013
0.00501	0.32	0.021
0.00835	0.45	0.054
0.01169	0.55	0.037
0.01502	0.69	0.026
0.01836	0.85	0.018
0.02170	1.00	0.021
0.02504	1.05	0.013

0.02838	1.15	0.019
0.03172	1.30	0.017
0.03506	1.50	0.026
0.03840	1.55	0.024
	Tyr-Phe	
0.03303	4.03	0.031
0.02202	2.85	0.016
0.01651	2.18	0.011
0.01321	1.80	0.019
0.01101	1.45	0.010
0.00944	1.28	0.012
0.00826	1.14	0.025
0.00734	0.98	0.033
0.00661	0.80	0.024
0.00600	0.78	0.016
0.00550	0.68	0.02
0.00508	0.60	0.011
	Gly-Phe	
0.00550	1.05	0.021
0.01179	1.60	0.012
0.01336	1.78	0.016
0.02358	2.95	0.014
0.01000	1.47	0.029
0.00700	1.14	0.031
0.01700	2.21	0.024
0.02000	2.53	0.017
	Gly-His	

0.02673	1.10	0.035
0.02005	0.90	0.022
0.01604	0.68	0.014
0.01336	0.62	0.018
0.01145	0.56	0.019
0.01002	0.52	0.015
0.00891	0.50	0.013
0.00802	0.46	0.012

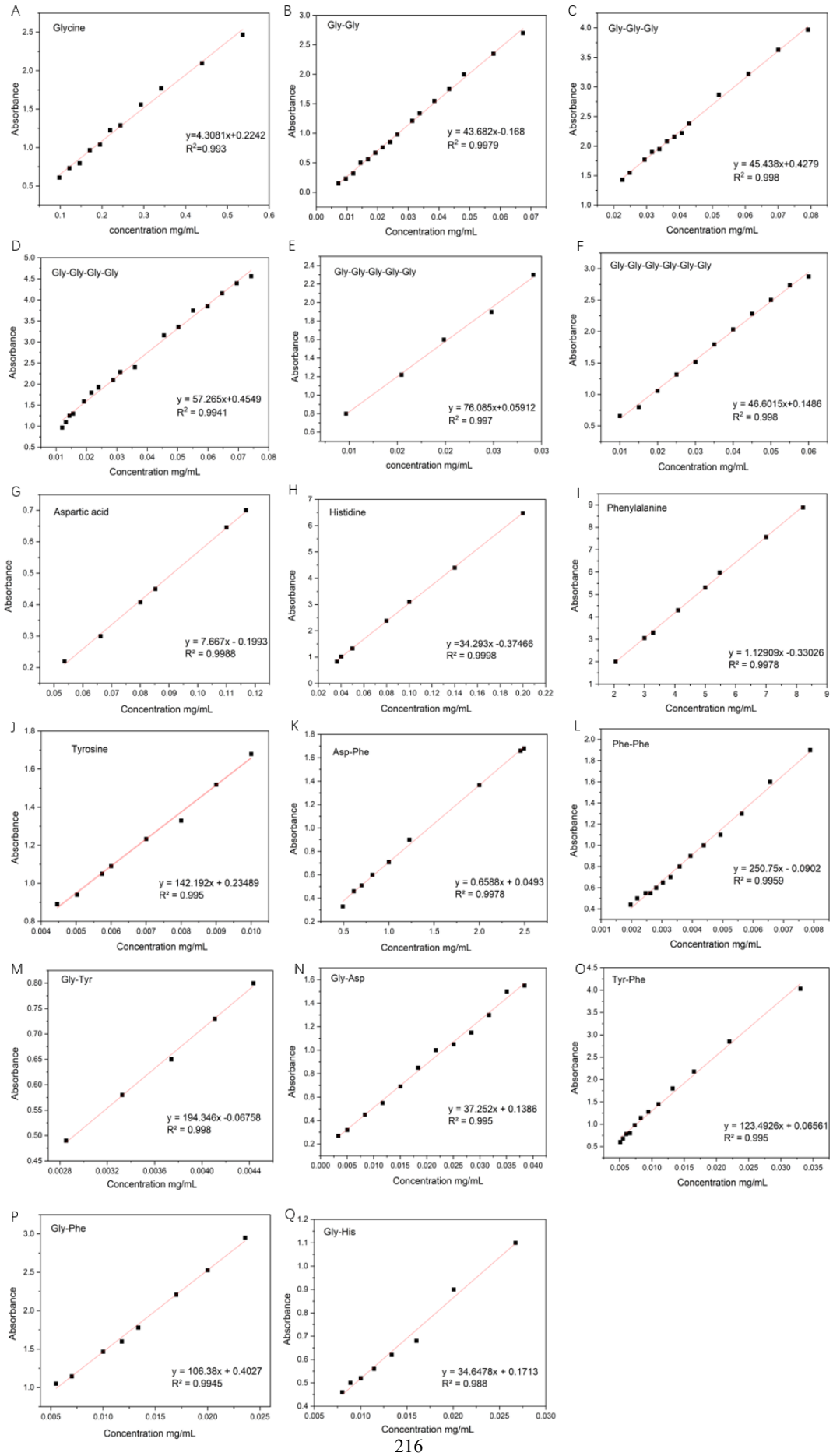

A
Glycine
y=4.3081x+0.2242
R²=0.993
Absorbance
concentration mg/mL
B
Gly-Gly
y = 43.682x-0.168
R² = 0.9979
Absorbance
Concentration mg/mL
C
Gly-Gly-Gly
y = 45.438x+0.4279
R² = 0.998
Absorbance
Concentration mg/mL
D
Gly-Gly-Gly-Gly
y = 57.265x+0.4549
R² = 0.9941
Absorbance
Concentration mg/mL
E
Gly-Gly-Gly-Gly-Gly
y = 76.085x+0.05912
R² = 0.997
Absorbance
concentration mg/mL
F
Gly-Gly-Gly-Gly-Gly-Gly
y = 46.6015x+0.1486
R² = 0.998
Absorbance
Concentration mg/mL
G
Aspartic acid
y = 7.667x - 0.1993
R² = 0.9988
Absorbance
Concentration mg/mL
H
Histidine
y =34.293x -0.37466
R² = 0.9998
Absorbance
Concentration mg/mL
I
Phenylalanine
y = 1.12909x -0.33026
R² = 0.9978
Absorbance
Concentration mg/mL
J
Tyrosine
y = 142.192x + 0.23489
R² = 0.995
Absorbance
Concentration mg/mL
K
Asp-Phe
y = 0.6588x + 0.0493
R² = 0.9978
Absorbance
Concentration mg/mL
L
Phe-Phe
y = 250.75x - 0.0902
R² = 0.9959
Absorbance
Concentration mg/mL
M
Gly-Tyr
y = 194.346x -0.06758
R² = 0.998
Absorbance
Concentration mg/mL
N
Gly-Asp
y = 37.252x + 0.1386
R² = 0.995
Absorbance
Concentration mg/mL
O
Tyr-Phe
y = 123.4926x + 0.06561
R² = 0.995
Absorbance
Concentration mg/mL
P
Gly-Phe
y = 106.38x + 0.4027
R² = 0.9945
Absorbance
Concentration mg/mL
Q
Gly-His
y = 34.6478x + 0.1713
R² = 0.988
Absorbance
Concentration mg/mL

Figure 4.5 Standard curves of the studied amino acids and peptides: A.Glycine (α form), B.Diglycine (α form), C.Triglycine(β form), D.Tetraglycine, E.Pentaglycine, F.Hexaglycine, G.Aspartic acid, H.Histidine, I. Phenylalanine, J.Tyrosine, K.Asp-Phe, L.Phe-Phe, M. Gly-Tyr, N.Gly-Asp, O.Tyr-Phe, P.Gly-Phe, Q.Gly-His

Table 4.2 Experimental and calculated solubility of glycine homopeptides in water under different temperature (0.1 MPa). [a]

T/K	$10^2 x_1^{exp}$	$10^2 x_1^{cal}$(equation(2))
	Glycine	
278.15	3.8604	3.7983
283.15	4.1690	4.2034
288.15	4.5671	4.6327
293.15	5.0386	5.0863
298.15	5.6347	5.5639
303.15	6.1295	6.0654
308.15	6.5524	6.5904
313.15	7.1279	7.1386
	Diglycine	
278.15	1.7990	1.8878
283.15	2.1527	2.0770
288.15	2.3570	2.2902
293.15	2.5532	2.5304
298.15	2.7271	2.8009
303.15	3.1050	3.1053
308.15	3.4060	3.4480
313.15	3.8731	3.8335
	Triglycine	

278.15	0.44850	0.4516
283.15	0.47885	0.4773
288.15	0.5153	0.5132
293.15	0.56750	0.5608
298.15	0.61952	0.6220
303.15	0.6855	0.6997
308.15	0.8089	0.7973
313.15	0.9173	0.9195
	Tetraglycine	
278.15	0.0129	0.0132
283.15	0.0155	0.0152
288.15	0.0174	0.0179
293.15	0.0219	0.0211
298.15	0.0254	0.0252
303.15	0.0292	0.0303
308.15	0.0373	0.0368
313.15	0.0450	0.0450
	Pentaglycine	
278.15	0.00203	0.0020
283.15	0.00221	0.0021
288.15	0.00225	0.0023
293.15	0.00241	0.0025
298.15	0.00271	0.0027
303.15	0.00292	0.0030
308.15	0.00309	0.0032

313.15	0.00374	0.0036
	Hexaglycine	
278.15	0.00029	0.00027
283.15	0.00030	0.00031
288.15	0.00035	0.00035
293.15	0.00039	0.00041
298.15	0.00045	0.00047
303.15	0.00056	0.00054
308.15	0.00063	0.00062
313.15	0.00069	0.00070

[a] x_1^{exp} represents the experimental solubility calculated by equation (2).

Table 4.3 Experimental and calculated solubility of amino acids and dipeptides in water at different temperatures (0.1 MPa). [a]

T/K	$10^2 x_1^{exp}$	$10^2 x_1^{cal}$(equation(5.2))
	Aspartic Acid	
278.15	0.0257	0.0289
283.15	0.0389	0.0370
288.15	0.0490	0.0460
293.15	0.0541	0.0554
298.15	0.0657	0.0649
303.15	0.0717	0.0740
308.15	0.0823	0.0822
313.15	0.0899	0.0891
	Phenylalanine	
278.15	0.2186	0.2206
283.15	0.2289	0.2268

288.15	0.2378	0.2369
293.15	0.2504	0.2511
298.15	0.2718	0.2700
303.15	0.2936	0.2940
308.15	0.3201	0.3241
313.15	0.3638	0.3613
	Tyrosine	
278.15	0.003562	0.0034
283.15	0.003869	0.0039
288.15	0.004468	0.0045
293.15	0.004991	0.0052
298.15	0.00556	0.006
303.15	0.007480	0.007
308.15	0.008574	0.0082
313.15	0.009394	0.0097
	Histidine	
278.15	0.4610	0.4845
283.15	0.5071	0.5018
288.15	0.5556	0.5335
293.15	0.6102	0.581
298.15	0.6488	0.6473
303.15	0.7218	0.7364
308.15	0.8183	0.8545
313.15	1.0437	1.0098
	Gly-Gly	
278.15	1.7990	1.8878

283.15	2.1527	2.0770
288.15	2.3570	2.2902
293.15	2.5532	2.5304
298.15	2.7271	2.8009
303.15	3.1050	3.1053
308.15	3.4060	3.4480
313.15	3.8731	3.8335
	Gly-Phe	
278.15	0.1997	0.2003
283.15	0.2105	0.2102
288.15	0.2232	0.2223
293.15	0.2369	0.2368
298.15	0.2530	0.2541
303.15	0.2756	0.2743
308.15	0.2960	0.2979
313.15	0.3263	0.3254
	Phe-Phe	
278.15	0.0017	0.0018
283.15	0.0021	0.0021
288.15	0.0024	0.0024
293.15	0.0028	0.0028
298.15	0.0033	0.0032
303.15	0.0037	0.0036
308.15	0.0041	0.0041
313.15	0.0046	0.0047
	Tyr-Phe	

278.15	0.003212	0.0033
283.15	0.003407	0.0034
288.15	0.003697	0.0036
293.15	0.00392	0.0039
298.15	0.004189	0.0044
303.15	0.005399	0.0052
308.15	0.005994	0.0062
313.15	0.007687	0.0076
	Gly-Tyr	
278.15	0.1832	0.1833
283.15	0.1964	0.1952
288.15	0.2184	0.2149
293.15	0.2388	0.2442
298.15	0.2825	0.2856
303.15	0.3406	0.3433
308.15	0.4340	0.4231
313.15	0.5294	0.5340
	Asp-Phe	
313.15	0.0239	0.0243
283.15	0.0288	0.0281
288.15	0.0319	0.0322
293.15	0.0366	0.0366
298.15	0.0414	0.0413
303.15	0.0463	0.0463
308.15	0.0512	0.0515
313.15	0.0570	0.0569

	Gly-Asp	
278.15	0.28502	0.2888
283.15	0.3124	0.3097
288.15	0.3468	0.339
293.15	0.3786	0.3784
298.15	0.4164	0.4299
303.15	0.4981	0.4966
308.15	0.5909	0.5827
313.15	0.6904	0.6937
	Gly-His	
278.15	1.5031	1.5515
283.15	1.6288	1.6035
288.15	1.7581	1.6932
293.15	1.8483	1.8239
298.15	1.9578	2.0014
303.15	2.1963	2.2346
308.15	2.5305	2.5355
313.15	2.9430	2.9206

[a] x_1^{exp} represents the experimental solubility calculated by equation (2).