A MOLECULAR THERMODYNAMIC APPROACH

TO PHASE PARTITIONING OF BIOMOLECULES AND PROTEIN FOLDING

by

YIZU ZHU

B.S. Chemical Engineering (1982), Dalian Institute of Technology
M.S. Chemical Engineering (1984), Dalian Institute of Technology
Dalian, China

SUBMITTED TO THE DEPARTMENT OF CHEMICAL ENGINEERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

CHEMICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

October, 1991

Signature of Author _____
Department of Chemical Engineering
October 30, 1991

Certified by _____
Professor Lawrence B. Evans
Thesis Supervisor, Department of Chemical Engineering

Accepted by _____
Professor William M. Deen
Chairman, Departmental Committee for Graduate Students

A MOLECULAR THERMODYNAMIC APPROACH TO PHASE PARTITIONING OF BIOMOLECULES AND PROTEIN FOLDING

by

YIZU ZHU

Submitted to the Department of Chemical Engineering, MIT, in partial fulfillment of the requirement for the Degree of Doctor of Philosophy in Chemical Engineering

Abstract

The phase partitioning of biomolecules and protein folding represent two important challenges facing today's biotechnology industry.

For over half a century, it has been the devoted efforts of many researchers to quantify the interactions that contribute to phase partitioning and stability of polypeptides and proteins. Much progress has been made. But, up to date, there is no comprehensive quantitative work done yet, and traditional modeling approaches are specific only for special properties or processes. Much progress is to be desired to deal with the complexities of the biomolecules.

An improved understanding of the nature of biological systems is instrumental to better represent the conformation and behavior of biomolecules in solutions. We have taken a practical approach using molecular thermodynamics to study the behavior of the biopolymers from the (amino acid) residue level as well as the molecular level. This methodology should endow more fundamental insights to the models for the behavior of the macrobiomolecules. This thesis is focused on the molecular thermodynamic modeling of phase partitioning of the biomolecules and the folding of polypeptides. It focuses on the development of Gibbs energy expression based on the nature of the interactions and the configurational entropy of polypeptide chains and solvents in solutions. The Gibbs energy expression was first developed for amino acids and small peptides. The expression is then extended for homopolypeptides and natural and synthetic polypeptides.

A molecular thermodynamic framework has been established for the representation of the solubilities of amino acids and small peptides. With this framework, satisfactory results have been obtained in representing and predicting the solubilities of amino acids and small peptides in aqueous solution as functions of temperature, ionic strength, dipolar species concentrations, solvent compositions, and pH.

The framework was then extended to describe the phase equilibrium behavior of β-lactam antibiotics (amino acid derivatives). It successfully correlates and represents the liquid-solid equilibrium behavior (solubilities) and the liquid-liquid equilibrium behavior

(phase partitioning coefficients) of β-lactam antibiotics as affected by temperature, ionic strength, solvent compositions, solute compositions, and pH.

A new molecular thermodynamic model is proposed in this work to represent the Gibbs energy of folding of homo-polypeptides in aqueous solutions. The enthalpy contribution is derived from a molecular segment-based local composition model which takes into consideration the residue-residue, residue-solvent, and solvent-solvent physical interactions and the distinct characteristics of amino acid residues in homo-polypeptides. The entropy contribution represents the entropy loss in folding a random-coiled polypeptide chain into a specific polypeptide conformation. The model interaction parameters between water and residues have been estimated with the UNIFAC method. A new hydrophobic scale has been developed for the twenty amino acid side chains based on the binary interaction parameters and the predicted Gibbs free energy of transferring amino acid side chains from ethanol to water. The model satisfactorily describes the folding of a number of homo-polypeptides into α-helices.

This thesis also presents a generalized molecular thermodynamic model for the free energy of folding of natural and synthetic polypeptides from coiled conformation into an α-helical conformation. The model attempts to account for the contributions resulting from both the local physical interactions and from the loss of configurational entropy upon folding the chain. The model is further generalized to a matrix form, it clearly differentiates the contributions to the enthalpy of folding due to solvation effect from that due to residue-residue interactions. The binary interaction parameters between the twenty amino acid residues were established using UNIFAC method. The model allows us to successfully study the formation of stable α-helices of several naturally occurring polypeptides in proteins including the C-peptide (residues 1-13) and S-peptide (residues 1-20) of RNase A (bovine pancreatic ribonuclease A), the Pα fragment in BPTI (Bovine Pancreatic Trypsin Inhibitor), and synthetic polypeptides (the copolymers of different amino acid residues) including the alanine-based peptides (16 or 17 residues long) in water. Mutation and temperature effects on the stability of polypeptide chain conformation were also examined with the model. The model prediction results suggest that the thermal unfolding of a polypeptide chain is a gradual conformational transition, instead of a two-state scheme.

This work has established essential foundations for further studies in predicting protein conformation and phase partitioning with the molecular thermodynamic approach.

Thesis Advisor: Dr. Lawrence B. Evans
Professor of Chemical Engineering

THESIS COMMITTEE

Chairman: Prof. Lawrence B. Evans (Chemical Engineering, MIT)

Members: Prof. Daniel Blankschtein (Chemical Engineering, MIT)

Dr. Chau-Chyun Chen (Aspen Technology, Inc.)

Prof. T. Alan Hatton (Chemical Engineering, MIT)

Prof. Peter S. Kim (Whitehead Institute and Biology, MIT)

Prof. Jonathan A. King (Biology, MIT)

Prof. David Soane (Chemical Engineering, UC Berkeley)

Prof. Daniel I.C. Wang (Chemical Engineering, MIT)

ACKNOWLEDGEMENTS

First of all, I wish to sincerely thank my thesis advisor, Professor Lawrence B. Evans, for the invaluable guidance and support he has given me in all of my efforts throughout the course of this work. His direction and encouragement have broadened my perspectives and provided me with many enriching experiences. I am certain that all my work in the future will still profit from his inspiring influence.

I owe many special thanks to my thesis committee members.

In particular, I am very grateful to Dr. Chau-Chyun Chen of Aspen Tech., Inc. for his valuable advice, constructive opinions and input to this work. His collaboration and many special contributions to this work are greatly appreciated. He has always been very helpful, enthusiastic, encouraging and challenging to my work over the years. His experience has shown me how one can build one's professional career through integrating engineering practice and academic research.

I would like to thank Professors Jonathan A. King and Peter S. Kim very much for providing valuable advice and comments about this project, especially, in polypeptide conformational studies. I would like to express my appreciation to Professors T. Alan Hatton and Daniel Blankschtein for their valuable comments on my work. Their group seminars 10.981/2 especially have been a great experience to me and provided me with a lot of input in colloid and polymer sciences. I am also grateful to all the students in Professors Hatton and Blankschtein's groups for many discussions. I also thank Professor David Soane of UC Berkeley very much for his many useful discussions with me on my thesis work. I would like to acknowledge Professor Daniel Wang very much for his enthusiasm, vision and valuable comments on this research.

I would like to express my sincere thanks to Professor James Wei, under whom I performed molecular modeling on diffusion and adsorption in zeolites, for his financial support and influential guidance. I have benefitted greatly from our many discussions. Working for him has taught me far more than science and engineering. He is always a good friend to talk with. His advice, encouragement and warmth have been most valuable along the way.

I owe special thanks to Professor John M. Prausnitz at UC Berkeley for his interest and

his discussions on my thesis research, and for his valuable comments on the homopolypeptide work of this thesis. I am very grateful to Professor John Edsall of Harvard University for his valuable input on amino acid studies of the thesis and Professor Harold A. Scheraga and his colleagues at Cornell for their comments on the homopolypeptide work.

I thank Shin-Goo Kang very much. His friendship, experience, interest and criticism have always been among the most valuable part of my life at M.I.T. Thanks also to Craig Zupke for helping me master HPLC and to UROP student Rachel Huggins who worked for me in the amino acid solubility study. My lab and office mates deserve special thanks for many discussions and help. Among them are Per Lindell, Dawn Orton, Gino Grampp, Demitrios Petridis and Mike Barrera. I thank all other BPEC colleague students for their help.

I would like to acknowledge financial support provided by the National Science Foundation's Biotechnology Process Engineering Center under Grant number CDR-88-03014. I also want to thank the people at Aspen Tech., Inc., Cambridge, MA, for their help. Special thanks to Dr. Suphat Watanasiri, Dr. Ian Gosling, Mr. Selim Anavi and Mr. Fred Ziegler for their assistance with my work.

My family on the other side of the Pacific is very special to me. I have been driven by their strong support, love and confidence from across the ocean. I cannot thank my family enough. My mother especially has given me invaluable support in the past. She has always been my example. It was my father who gave me the perceptual knowledge of efficiency and optimization by continuing to change the design of our home furnace, which inspired my interest and dream in engineering from the very beginning. It is to my Dad, who passed away ten years ago, and to my Mom that this thesis is dedicated. I am also indebted to my sisters who have been very supportive in many ways since I went to college.

Finally, I am deeply grateful to my wife, Jianping Chen and my son, Bo, for their love, encouragement and support along the road. Without them, I would not have made it. I owe them so much.

OUTLINE

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

The intellectual frontiers in biotechnology can be described on a continuum from microscale through mesocale to macroscale. At either end of this spectrum are highly interdisciplinary research topics that will require modeling tools currently used by chemical engineers in other context (National Research Council, 1988). The commercialization of the results of basic biological research requires the integration of the knowledge of both biologists and chemical engineers, and needs a strong knowledge base on which to design bioprocessing systems. Unfortunately, little attention has been given to date to this kind of research. Our knowledge of techniques for efficient large-scale recovery and purification of bioproducts from the complex mixtures of proteins is little more than rudimentary (Humphrey, 1989). Prompt and effective exploitation of biotechnology is dependent on the improvement of this disciplinary interface which is one of the most critical problems confronting bioengineering today.

This thesis tackles one of the challenging areas of today's biotechnology: the understanding of the protein behavior critical to the downstream processing of bioproducts. We try to explore the innovative synthesis of new concepts that combine engineering ideas with biological speculations, or the modeling of fundamental biological interactions using the methodologies of chemical engineering.

In general, biomolecules include proteins (amino acids, peptides, polypeptides and

proteins), nucleic acids (purine, pyrimidine, nucleotides, oligonucleotides, tRNA,DNA, and RNA), lipids, carbohydrates and steroids. But here in this thesis, by biomolecules, we mean amino acids, peptides, and proteins not only because they are extremely valuable compounds in both biological and therapeutic senses, but also their separation and purification present experimental and theoretical challenges for biotechnology and bioengineering. In particular, the manufacturing of protein products involve all the complexities any bioprocess can experience.

Most biomanufactureing systems require processes to produce, concentrate, and separate a target protein from a mixture of other proteins while maintaining its activity (or three dimensional structure). The key to separation and purification of bioproducts is the understanding of protein behavior (Humphrey, 1989). In addition, protein solution is generally aqueous, so the data on protein solubilities and a knowledge base are needed to predict the relation of solubility to protein structure and size, as well as to the physical parameters of the solution environment such as pH, ionic strength of the solution and the influence of organic solvents. In short, there is a compelling need to better understand the behavior of various protein molecules in different physical and chemical environments and to develop theories to explain the interactions of protein molecules with various surfaces, as well as with other proteins and with solute and solvent molecules. Protein solutions are highly complex, conformational, and nonideal systems due to the complicated physical and chemical interactions. Since protein phase partitioning always involves conformational problem, predicting the partitioning of a protein needs to

understand all the fundamental interactions occurring in protein solution which stabilize the native conformation of the protein and partition the protein into a desired phase. At present, the rules for protein folding are unknown yet and our understanding of protein folding phenomena is very limited, which has hindered the commercialization of novel techniques developed in laboratories, such as recombinant DNA technology to produce mammalian proteins in bacteria.

In fact, the forces which govern protein partitioning and stabilize protein structures, to a large extent, are essentially the same type of forces in nature which manipulate the physical and chemical properties of small organic molecules, typically the twenty amino acids that are the building blocks of polypeptides and proteins. For over half a century since the work of Cohn et al., (1930, 1933, 1941, 1943), it has been the endeavor of many researchers to quantify the solvation energy and the side-chain interactions that contribute to polypeptide and protein stability and phase equilibrium properties. But, up to date, there is no comprehensive work done yet. We have taken a rather practical approach to study the behavior of the biopolymers in solution from the amino acid residue level as well as the molecular level. The strategy taken in this research is to capture and represent the interactions governing the phase behavior of the small biomolecules (or, amino acids and their derivatives, antibiotics), then study the conformations of homopolypeptides (or, homopolymers of the twenty amino acid residues) and natural and synthetic polypeptides, based on the understanding of the forces in the smaller molecular systems. This methodology should endow more

fundamental insights to the behavior of the biopolymers.

To develop an improved understanding of the protein phase partitioning and folding phenomena in aqueous solutions, molecular thermodynamics can serve as a powerful tool to the biotechnology industry which needs the predictive models for the behavior of the biomolecules. Molecular thermodynamics has been well-established as a practical concept to study phase equilibria of both small and large molecular systems (Prausnitz, 1979), electrolytes (Chen and Evans, 1986) and surfactants and micelles (Blankschtein, et al., 1986; Chen, 1989). In such molecular systems, the Gibbs energies are dictated by the weak physical interactions that exist between various species and segments in the mixture. Molecular thermodynamics provides a semi-empirical theoretical framework to take proper account of such weak physical interactions.

This thesis is focused on the molecular thermodynamic modeling of phase partitioning of biomolecules and the folding of polypeptides. In other words, it is concentrated on the development of Gibbs free energy expression based on the nature of the physical interactions and the degree of flexibility of the amino acid residues and solvent molecules in polypeptide solutions. Such an expression should assist us in identifying the native protein conformation and in predicting protein partitioning between phases.

1.2 Thesis Objectives

The thesis objectives are:

1. To develop a comprehensive molecular thermodynamic framework which can

predict polypeptide conformation and to represent the solubility and the partitioning coefficient of the biomaterials (or, amino acids and their derivatives, synthetic and natural polypeptides and proteins) in solutions as a function of temperature, ionic strength, pH, solvent concentrations, and other factors.

2. To validate the proposed theoretical framework via a representative data base built on existing data.

3. To provide a molecular thermodynamic perspective of the complex mechanisms behind protein folding and the phase behavior of the biomolecules, which underlie all the non-bonding physical interactions both intra and inter molecules, either solute-solute type or solute-solvent type.

1.3 Summary of Thesis

Chapters 2 and 3 of the thesis provide background information, literature review and problem definitions. Chapter 2 presents an outline of the separation and purification techniques of biomolecules, the application of protein separation and purification techniques in blood plasma fractionation, the available data for the studies of phase partitioning of biomolecules, and the prior empirical models for phase partitioning of biomolecules.

Chapter 3 introduces protein conformational problem and gives a review on the empirical and theoretical methods in predicting protein conformation. Both the probabilistic methods and physico-chemical methods (including statistical mechanics methods and heuristic methods) are reviewed in this chapter.

Chapter 4 elucidates the molecular thermodynamic perspectives on the two challenging problems in today's biotechnology, the protein folding and the phase partitioning of biomolecules. This chapter points out that protein folding and phase partitioning are closely related intellectual and engineering problems confronting biologists, biochemists and biochemical engineers. The key to attack these problems is the sound understanding of the essential interactions occurring in protein solutions as a function of the environmental physico-chemical parameters. Molecular thermodynamics is a powerful tool to tackle the challenges.

Chapter 5 presents a molecular thermodynamic framework for the representation of the solubilities of amino acids and small peptides. The work extended the Electrolyte NRTL model (Chen and Evans, 1986) to aqueous solutions of zwitterionic species to predict the activity coefficients of amino acids and small peptides. With this framework, satisfactory results have been obtained in representing and predicting the solubilities of the small biomolecules in aqueous solution as functions of temperature, ionic strength, dipolar species concentrations, solvent compositions, and pH.

Based on the theoretical framework proposed in Chapter 5, Chapter 6 presents recent results in representing the liquid-solid equilibrium behavior (solubilities) and the liquid-liquid equilibrium behavior (phase partitioning) of β-lactam antibiotics, which are amino acid derivatives and important chemotherapeutic agents.

Chapter 7 and 8 established molecular thermodynamics models for the prediction of secondary structures of homo-polypeptides, synthetic peptides and protein fragments.

In Chapter 7, a semi-empirical molecular thermodynamic model is proposed to represent the Gibbs energy of folding of aqueous homo-polypeptide systems. The model takes into consideration both the entropy contribution and the enthalpy contribution of folding homo-polypeptide chains in aqueous solutions. The entropy contribution is derived from the Flory-Huggins expression for the entropy of mixing to account for the entropy loss in folding a random-coiled polypeptide chain into a specific polypeptide conformation. The enthalpy contribution is derived from a molecular segment-based local composition model which takes into consideration the residue-residue, residue-solvent and solvent-solvent binary physical interactions along with the local compositions of amino acid residues in aqueous homo-polypeptides. The UNIFAC group contribution method (Fredenslund et al., 1975; 1977), developed originally to estimate the excess Gibbs energy of solutions of small molecules, was used to estimate the binary interaction parameters.

The model yields a hydrophobicity scale for the twenty amino acid side chains which compares favorably with established scales (Nozaki and Tanford, 1971; Leodidis and Hatton, 1990). In addition, the model generates qualitatively correct thermodynamic constants and it accurately predicts thermodynamically favorable folding of a number of aqueous homo-polypeptides from coil states into α-helices. The model further facilitates estimation of the Zimm-Bragg helix growth parameter s and the nucleation parameter σ for amino acid residues (Zimm and Bragg, 1959). The calculated values of the two parameters fall into the ranges suggested by Zimm and Bragg.

In Chapter 8 we generalize the molecular thermodynamic model for aqueous homopolypeptides as a model for the free energy of folding of polypeptides from coiled conformation into an α-helical conformation. The generalization accounts for the contributions resulting from the loss of configurational entropy upon folding, the local physical interaction energies between solvent-residue pairs and the local physical interaction energies between residue-residue pairs. The solvent-residue interaction represents the intrinsic helix-forming potential of the residue in its homopolypeptide while the residue-residue interaction represents the cooperative potential of the twenty amino acid residues to form α-helices. The binary interaction parameters between the twenty amino acid residues were estimated using UNIFAC (universal functional activity coefficient) group contribution method (Fredenslund, et al., 1975, 1977).

The model has been used to investigate the formation of stable α-helices of several naturally occurring polypeptides including the Pα fragment in BPTI (Bovine Pancreatic Trypsin Inhibitor), the C-peptide (residues 1-13) and S-peptide (residues 1-20) of RNase A (bovine pancreatic ribonuclease A), the HBP-1 peptide in protein Heparinase and synthetic polypeptides including the alanine-based peptides (16 or 17 residues long) in water. Temperature and mutation effects on the stability of polypeptide chain conformation were also examined with the model.

Chapter 9 draws conclusions from this research and gives future directions to this on-going research effort.

Literature Cited

Blankschtein, D., G.M. Thurston, and G.B. Benedek, J. Chem. Phys., 85 (12), 7268-7288 (1986).

Chen, C.-C., "A Molecular Thermodynamic Model for the Gibbs Energy of Mixing of Micellar Solutions," paper presented at the Annual AIChE Meeting, San Francisco, 1989.

Chen, C.-C. and L.B. Evans, "A Local Composition Model for the Excess Gibbs Energy of Aqueous Electrolyte Systems," AIChE J., 32, 444 (1986).

Cohn, E.J., Berggren, R.E.L., and Hendry, J.L., J. Biol. Chem., LXXXVII, 81 (1930).

Cohn, E.J., and Berggren,R.E.L., J. Biol. Chem., 101, 45 (1933).

Cohn, E.J., Ferry, J.D., Livingood, J.J., and Blanchard, M.H., J. of Am. Chem. Sco., 63, 17-22 (1941).

Cohn, E.J. and J.T. Edsall, Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Reinhold Publishing Corporation, New York, 1943.

Evans, L.B., Bio/Technology Vol. 6, No. 2, February 1988.

Fredenslund, A., J. Gmehling, and P. Rasmussen, Vapor-Liquid Equilibrium Using UNIFAC, Elsevier Scientific Publishing Company, Amsterdam (1977).

Fredrickson, R.A., M.C. Chang, S.P. Powers, and H.A. Scheraga, Macromolecules, 14, 625-632 (1981).

Humphrey, A., in "Frontiers in Bioprocessing", Edited by Sikdar, S.K., M. Biet and P. Tood, CRC Press, Inc., Boca Raton, Florida, 1-13, 1989.

Leodidis, E.B., and T.A. Hatton, J. Phys. Chem., 94 (16), 6411-6420 (1990).

National Research Council, "Frontiers in Chemical Engineering," National Academy Press, Washington, D.C., 1988.

Nozaki, Y. and C. Tanford, J. Biol. Chem., 46 (7), 2211-2217 (1971).

Oas, T.G. and P.S. Kim, Nature, 336 (6194), 42-48 (1988).

Prausnitz, J.P., "Molecular Thermodynamics for Chemical Process Design," Science, 205, 24, 759-766 (1979).

Zimm, B.H. and J.K. Bragg, J. Chem. Phys., 31 (2), 526-535 (1959).

CHAPTER 2

THE PHASE PARTITIONING PROBLEM

Generally speaking, downstream processing is the most expensive part of the bio-production processes which requires efficient and extremely special separation and purification for fragile bioproducts from fermentation and cell culture media to ensure that adequate purity can be achieved without chemical change or loss of the configuration of the products. A key requirement to understand and to model such processes as fractional precipitation, crystallization, liquid-liquid extraction, and chromatography is to be able to predict the partitioning of biomolecules between liquid and solid, liquid and liquid, and liquid and immobilized media. Specifically, phase equilibria, especially liquid-liquid equilibria and solubilities in multicomponent solutions including electrolytes, activity coefficients in ionic solutions, conformational structures, pH dependence of various properties including phase behavior are among those of the most important properties to downstream bioprocessing (Olein, 1987).

2.1 The Outline of the Separation and Purification Techniques of Biomolecules

In certain protein mixtures, there are only a few kinds of proteins, and only one specific protein is predominant, but in some other mixtures (such as animal cells), there exist a great number of protein species with dilute concentrations. In order to separate a particular kind of protein species from a mixture, the non-protein species (such as sugar, lipids, nucleic acids and other organic compounds) should be separated first from the protein species by preliminary treatment, then different proteins are fractionated in

terms of the characteristics of proteins in the mixture (commonly, aqueous solution), such as solubility, hydrophobicity, molecular size, charge and affinity.

2.1.1 Methods Based on Solubility Difference --- Precipitation

Precipitation is the earliest experimental and practical methods of separating different types of proteins. The major factors which influence the solubility of proteins are pH, ionic strength, dielectric constant (solvent effect) and temperature

Under the same conditions, different proteins have different solubilities due to the different ratio of polar hydrophilic groups and nonpolar hydrophobic groups, and the different steric configurations of the protein molecules. By varying the above factors, we could control the solubility of a specific protein to reach separation and purification specifications.

Isoelectric precipitation and pH effect

Proteins are zwitterionic electrolytes. The ionization of the weakly acidic or basic amino acid side-chains of a protein is influenced strongly by pH, so the net charge of a protein varies with pH, and its solubility is a function of the net charge on these groups. At isoelectric point (pH = pI) of a protein, the molecule has zero net charge, where its solubility is reduced to minimum, aggregation and precipitation occurs due to the fact that there is no isoelectric repulsive force between protein molecules. When pH is away from the pI of a protein, the protein molecules carry net charges (positive for pH < pI, negative for pH > pI), and the molecules repel each other to avoid aggregation and precipitation, so the solubility is substantially increased. Different protein has different

pI. The minimum solubility behavior of proteins at their pI's makes it possible to separate and purify them from each other. When the pH of the solution is adjusted to the pI of the target protein, the most or almost of this protein is precipitated out from the solution, and the other kinds of protein species still remain in the mixture. Depending on the gentleness of the environment, the protein precipitates may still keep the natural configuration of the protein, and can be redissolved at appropriate pH.

Salting-in and salting-out

Neutral salts have significant influence on the solubility of globular proteins.

In dilute salt solution, adding neutral salts can increase the solubility of proteins, "salting-in" occurs since the adsorption of salt ions on the protein surface changes the electrostatic force field, makes protein molecules repel each other, and enhances the attractive interactions between protein and water molecules. From the dialysis (would be mentioned latter) point of view, the reason for globular proteins precipitate during dialysis process is that the removal of salt ions by semi-membrane increases protein-protein attraction and causes protein to aggregate and to precipitate.

As the ionic strength of the solution reaches certain value (or concentrated neutral salt solution), the solubility of protein begins to decrease. If the ionic strength is high enough, or in saturated or half saturated salt solution case, most of the protein can be precipitated out. This is called "salting-out", which is mainly because the addition of great amount of salt reduces the activity of water. The most part of free water molecules have become hydrated salt ion complex, which lowers the solubilization interaction between

water and the polar group of the protein. Salt precipitation is one of the most popular protein separation and purification techniques. Proteins via salt precipitation may remain their natural configuration as well, and can be redissolved. Ammonium sulphate is recommended as the best neutral salt for its high solubility in water (760 g/l H2O at 20°C), and the solubility is less temperature dependent.

Precipitation with organic solvent

Because of the high dielectric constant of water (79 at 20°C), the electrostatic interactions between charged polar groups of intra- or inter-protein molecules is reduced in aqueous solution, and the interaction between water and these polar groups favorites the solubilization of the protein. By adding some miscible organic solvent with lower dielectric constant (such as ethanol (26 at 20°C), acetone (21 at 20°C)), the dielectric constant of the solution is reduced. Some of the water molecules are repelled from protein surface by solvent molecules, which enhances the electrostatic interactions between protein molecules, or results in an increase in the attractive forces between the oppositely charged groups both intra and inter protein molecules, and proteins tend to aggregate and precipitate. The disadvantage of this method is that the operation has to be carried out at low temperature to avoid temperature denaturation.

Precipitation with synthetic polymers

PEG (polyethylene glycol) and dextran are most investigated biopolymers for protein precipitation.

PEG as a protein precipitant was first reported in 1956 (Stocking), but it did not receive serious attention until the work of Polston, et al (1964), who used PEG6000 to purify fibrinogen from human plasma.

Precipitation using PEG is the only simple, safe and commercialized purification technique for high purity Factor VIII, which is extremely valuable in therapeutic practice (the successful treatment of hemophilia). The advantages of PEG precipitation of proteins can be summarized as follows:

(1). Very low heat of solution of PEG in water. Therefore, large temperature change does not accompany the addition of the reagent, and precipitation can be carried out at room temperatures.

(2). Good performance (a mild environment for labile biomolecules) to minimize denaturation due to the solvency or low interfacial tension of the aqueous solution of PEG.

(3). Nontoxic (especially important for Factor VIII transfusion).

(4). It could be completely removed from clinical products by adsorption of the protein to either ion exchange or affinity supports since PEG is hydrophilic polymer without charge and affinity for such materials.

(5). Short processing time (only several minutes) required for the precipitation to reach equilibrium (compared with traditional precipitation methods (several hours)). This could greatly reduce problems of product contamination.

(6). The high resolution of the separation.

(7). This process can be easily integrated with the purification flowsheet (PEG is employed early and further purification involves one or more adsorption steps).

Furthermore, the low cooling requirement and high production yield, in addition to its relatively low cost of material make PEG precipitation of proteins a very attractive technique. Dextran are widely used as precipitant as well in laboratory study, but the cost of material is high.

The mechanisms for precipitation by biopolymers remain a big question, though researchers, Iverius and Laurent (1967), Ogston (1958, 1962) and their coworkers, interpreted the precipitation as polymer exclusion of protein molecules from the solution, and reduction of the effective amount of water available for their solvation. The phenomena are quite closely related to the formation of liquid-liquid extraction systems for protein partitioning developed by Albertsson (1958), Kula and coworkers (1977), and Baskir et al., (1987), and reviewed by Abbott and Hatton (1988).

Precipitation with polyelectrolytes

Acidic polysaccharides such as carboxymethylcellulose, alginate, pectate and carrangeenans have been used for food protein precipitation (Ledward, et al., 1978). A high molecular weight dextran sulphate has been used for the precipitation of serum proteins but not yet for fractionation of materials for clinical use. Anionic polymers, polyacrylic acid and polymethyacrylic acid, and the cationic polymers, polyethylene-imine and a polystyrene-based quaternary ammonium salt for the precipitation of whey proteins also have been studied by Hill and Zadow (1978). Salting-out and molecular exclusion

by polyelectrolytes are possible causes of this precipitation.

2.1.2 Methods Using Aqueous Two Phase Systems

Aqueous two-phase partition was introduced in late 1950s with applications for both cell particles and proteins by Albertsson (1958). Since then it has been applied to a large number of different materials, such as plant and animal cells, microorganisms, virus, chloroplasts, mitochondria, membrane vesicles, and biopolymers like proteins and nucleic acids.

The high potency of two phase system (or liquid-liquid extraction) in biotechnology is due to the fact that liquid-liquid extraction unit operation has been well established in chemical engineering both theoretically and practically, and the scale-up would be a lot easier than other processes.

Polymer/polymer/H_2O systems

Two-polymer phase system is a mixture of water with two different water-soluble polymers (not necessarily soluble in water in the entire range of concentration, such as PEG which itself could precipitate in pure water under certain temperature and concentration conditions). Both phases of the system are rich in water (around 90%w/w), and each phase is rich in one specific polymer. Phase separation occurs when water is mixed with different phase forming polymers due mainly to the mutual incompatibility (unfavorable energy of interactions) of the two polymers, interactions between the solvent (water) and either polymer play essentially no role (Brooks et al.,

1985).

Proteins or nucleic acids introduced in two phase system distribute unevenly between the phases. Systems used for partition of proteins include PEG-Dextran, Methylcellulose-Dextran, PEG-ficoll-Dextran, and PEG-Starch. Among these, PEG/Dextran/H_2O is the typical and most studied system. Because of the high cost of Dextran, researchers are trying to substitute crude dextran for high value dextran.

Polymer/salt/H_2O systems

Phase splitting can occur if a low molecular weight solute is added to a solution of a polymer in a good solvent such as PEG in H_2O, and the solute concentration reaches its critical concentration (for PEG/K_2SO_4/H_2O, 0.5M K_2SO_4). Salt-PEG system has a much higher ionic strength than Polymer-Polymer systems, since salts such as $(NH_4)_2SO_4$, $MgSO_4$, or K_3PO_4 have to be used in the concentration range of about 0.5-2M. These systems, however, can be used with advantage for purification of enzymes according to Kula, et al. (1982).

Reversed micelle systems

Reversed micelles are nanometer-scale surfactant aggregates of amphophilic molecules which may form spontaneously and reversely in nonpolar solvents capable of solubilizing significant amount of polar compounds, including proteins, in the bulk organic phase. They are aggregates in homogeneous solution, where the polar head groups of the surfactant molecules comprise the core of the aggregates and the hydrophobic tails

face the nonpolar solvent (Woll et al., 1987). The proteins can be induced to move from a bulk aqueous phase to the micelle-containing organic phase, and vice-versa by manipulating pH, ionic strength and surfactant concentration.

The partitioning of biomolecules between a bulk aqueous phase and a reversed micelle seems to be both the electrostatic interaction between the charged protein molecules and the surfactant headgroups, and the volume exclusion between protein and the micelle. It has been shown that protein mixtures can be readily resolved using this technique (Woll, et al., 1987). The investigation and the understanding of reversed micellar extraction of proteins is still in its infancy, compared with two-polymer systems (Abbott and Hatton, 1988).

2.1.3 Methods Based on Molecular Size Difference

The most evident characteristic of protein molecule is the large particle size, and different proteins have different sizes. Therefore, some simple techniques can be used to separate proteins with compounds of small size such as inorganic salts, sugar and water, and furthermore to differentiate different proteins.

Dialysis and ultrafiltration

The separation of proteins with small molecules is carried out by semi-membranes. Dialysis bag which contains protein solution is put into distillated water to allow small molecules diffuse through the membrane. In ultrafiltration method, small size solutes are forced to pass through the membrane by pressure and centrifugal

forces, and proteins are left on the membrane side.

Gel-filtration

Gel-filtration, or molecular-sieve chromatography, is one of the most efficient method for protein separation. In gravitational force field, protein species pass through gel media with different velocities , and are retarded to various degree depending on its sizes to give rise to separation in the outflow. Species with larger molecular size are washed out earlier, those with smaller size are trapped in the cross-linked gel network meshes, retarded and washed out latter.

2.1.4 Methods Based on Charge Characteristics

There are two kinds of separation methods based on the charge characteristics of proteins:

Electrophoresis (Ionphoresis)

The application of an electric field would make protein molecule move towards anode or cathode if pH is not at pI. The mobility of the molecule depends on its net charge (pH, ionic strength), size and shape, different proteins can be fractionated in terms of their property and the property of the solution by electrophoresis.

B. Ion-exchange chromatography

Ion-exchange chromatography media for protein separation consist of cellulose (such as carboxymethylcellulose and diethylaminoethylcellulose), which has a higher adsorptivity compared with ordinary ion-exchange resin. The interactions between

protein and the solid media relies on pH, which controls the degree of electrostatic attraction of the oppositely charged groups. Still, pH and salting effects play very important part here. pH determines the extent of dissociation of both protein and the ion-exchanger. The existence of salt could lower the electrostatic interaction between the protein and the ionic group of the exchanger. The protein species which has the weakest attraction force with the exchanger is washed out first by controlling pH and ionic strength.

2.1.5 Affinity Chromatography

Affinity chromatography is a very efficient separation method, and is especially used for further purification of proteins. The target protein is noncovalently bounded, via binding sites, to its ligand which is covalently linked to the column media. The binding force results from protein-ligand biospecific interaction. Other proteins which have no affinity to the ligand pass through the bed. The target protein then can be eluted from the chromatography media by free ligand solution.

Affinity chromatography will become of great importance in the future for the recovery of valuable active substances in blood plasma, but will not become an alternative to precipitation procedures when applied to the isolation of proteins which are present in high concentration.

In summary, generally, isoelectric precipitation, salt precipitation, and solvent precipitation are used as primary separation steps, especially used for the separation and

purification of large quantity of protein products. In some cases like the separation of antibodies, precipitation technique can give 95% or even higher purity. Two phase partitioning methods are attractive due to their relatively simple technique, but, at present, the phase forming polymers are still expensive to make liquid-liquid extraction industrialized due to the solvent recovery problem. Chromatographic methods are used only in purification steps commonly applied for small scale production because of the high resolution and the high cost.

2.2 The Application of Protein Separation and Purification Techniques in Blood Plasma Fractionation

Blood consists of only two phases: the cell phase and the plasma phase. Plasma is a complex medium containing a wide variety of proteins in aqueous solution. Proteins have been recovered from plasma for over forty years (Stryker, et al., 1985), primarily by fractional precipitation, so there is a wide body of experience and experimental data to evaluate the available theory for protein separation and to develop theory for better understanding and prediction of phase behavior of macromolecules. Human plasma fractionation system, therefore, can be chosen as the candidate for initial modeling on the partitioning behavior of biological substances.

A brief summary of available techniques for plasma fractionation is shown in the following Tables 2-1, 2-2 and 2-3:

Table 2-1. Methods Based on Interactions with Physical Parameters

Name	Main Product	Precipitant	Scale
Cryoprecipitation	AHF(FactorVIII)	EtOH	Large
Thermal denaturation	NSA	EtOH/PEG	
Electrodialysis	NSA		
Electrophoresis	NSA(95% pure)		

Table 2-2. Methods Based on Interactions with Chemical Precipitant

Name	Main Product	Precipitant	Scale
Cohn method 6	NSA	EtOH	Large
Cohn method 9	ISG	EtOH	Large
Gerlough(method 6G)	NSA	EtOH	
Kistler & Nitschmann	NSA, ISG	EtOH	Large
Deutsch method	ISG	EtOH	Large
Kink method	PPF(90%NSA,9%ISG)	EtOH	Large
Hao method	NSA,ISG	EtOH	
Ammonium sulfate	NSA, ISG,	$(NH_4)_2SO_4$	Large
Caprylic acid	IgG6,IgA,Ceroloplasmin	$CH_3(CH_2)_6COOH$	
Rivanol	IgG	Lactates	
PEG	High Purity AHF	PEG	Large
	NSA,AT-III	PEG	Large
Zinc	Immune globulin	Zinc ion/EtOH	
Amino acids	Fibrinogen,AHF	Glycine-alanine	

Table 2-3. Methods Based on Interactions with Solid Phases

Name	Main Product	Precipitant	Scale
Silicon dioxide	ISG,NSA	Porous glass beads	
Calcium phosphate	FactorIX complex	$(Ca)_3(PO_4)_2$	
Ionexchange chromatog.	NSA Sephadex	G-25	Pilot
	FactorIX complex	DEAE-sephadex	
Solid phase polyions	FactorIX complex	Copolymers	
Affinity chromatog.	High purity products	Cross-linked polysacchorides	
	AHF	Agarose/dextran	
	AT-III	Insolublized heparin	
	Plasminogen	Lysine-agarose	
	FactorIX	DEAE-sephadex	

2.3 Data Available

There already exist certain amount of data in the vast publications of scientific literature. Among these, solubility and phase partitioning data are the most abundant. Table 2-4 is only a partial list of them existing in the literature. Active research on the measurement of solubility and phase partitioning data is still going on. For protein phase partitioning data, please refer to Albertsson (1985) and Brooks, et al., (1985).

Table 2-4. Data Available in Literature for Amino Acids, Peptides and Protein

Amino Acids					
	s	γ^a	$\Delta G_{Soln}{}^b$	Titration Data	References
Glycine	+	+	+	+	Dalton, Dunn
L-Alanine	+	+	+		Dalton,Hade
L-Valine	+	+	+		Hade
L-Leucine	+		+		Dolton
L-Isoleucine	+		+		Hade
L-Phenylalanine	+		+		Dalton,Dunn
L-Tyrosine	+		+	+	Dalton,Dunn
L-Tryptophan	+		+		Dalton
L-Serine	+	+	+		Hade
L-Threonine		+			Smith
L-Cysteine					
L-Methionine	+	+	+		Hade
L-Aspartic acid	+		+		Dalton,Dunn
L-Asparagine	+*		+		Dalton
L-Glutamic acid	+		+		Hade
L-Glutamine			+		Fasman
L-Lysine	+**				Hade
L-Arginine	+***	+***	+***		Hade
L-Histidine					
L-Proline	+	+	+		Tomiyama
Betaine		+			Smith
L-Cystine	+		+		Dalton
Sarcosine		+		+	Smith

Peptides

	s	γ [a]	ΔG_{Soln} [b]	Titration Data	References
Alanylalanine		+			Smith
Alanylglycine		+			Smith
Glycylglycine		+			Ellerton
Glycylalanine		+			Smith
Tri-glycine		+			Smith

Proteins

	s	γ [a]	P_{Osm}	Titration Data	References
Albumin	+				
Carboxy-hemoglobin	+				Green
Casein	+	+		+	Pertzoff
Egg albumin	+	+			Green
Fibrinogen	+				Florkin
Foetal-serum albumin			+		McCarthy
Gelatin	+				Cohn
Globulin	+			+	Palmer, Mellanby
Growth hormone			+		Li
Haemoglobin	+		+		Green
Insulin	+				Cohn
Oxyhemoglobin	+	+			Green
Paracasein	+				Pertzoff
Psedoglobulin	+				Cohn
Sera	+				Greenwald
Serum-globulin	+				Cohn
Tuberin	+				Cohn
Zein	+			+	Cohn

*	Solubility of L-Asparagine.H2O
**	Solubility of L-Lysine.HCl
***	Solubility of L-Arginine.HCl
a	Activity coefficient (no indication of L- or D- form)
b	Free energy of solution is the free energy change for transporting one mole of the solute from the saturated solution to a hypothetical aqueous solution at an activity of 1 molal.
c	In various salts
s	Solubility
γ	Activity coefficient
P_{osm}	Osmotic pressure
ΔG_{Soln}	Gibbs free energy of solution

2.4 Prior Models for Phase Partitioning of Biomolecules

As mentioned in previous section, precipitation is one of the most popular and important techniques for both the laboratory and industrial scale separation and purification of proteins from fermentation broth, enzymes, and human and animal blood plasma. Modeling and simulation of precipitation processes involve almost all the physical interactions occurring in downstream processing. Also with its high potency, the two phase extraction technique draws great attention of biotechnology industry. In the following, therefore, we review modeling methods related to precipitation and two phase partitioning only.

2.4.1 Models for Precipitation

Salting-out effect

Cohn and Edsall (1943) proposed that the logarithm of the protein solubility is linearly proportional to the salt ionic strength.

$$\log S = \beta - K' I \qquad (2\text{-}1)$$

Where S: Solubility of the protein at ionic strength I

I: Ionic strength

β, K': Constants

β is essentially salt-independent, but varies significantly with protein, and is a strong function of pH and temperature, usually passing through a minimum at the isoelectric point. K', the slope of the salting-out curve, is independent of pH, and temperature, but varies with the salt and protein involved.

Melander and Horvath (1977) treated salt precipitation as a balance between a salting-in process due to electrostatic effects of the salt and a salting-out process due to hydrophobic effects. Their expression is as follows:

$$\text{Log} (S/S_0) = - K' I + \beta' \qquad (2\text{-}2)$$

Where S_0 is the solubility at zero ionic strength, and K' is correlated with the protein contact area, induced dipole, and repulsive interactions.

$$K' = \Omega \sigma - \lambda \qquad (2\text{-}3)$$

Here Ω: Contact area between the protein molecules

σ: Induced dipoles

λ: Repulsive interaction between molecules with like charges

Organic solvent effect

Bell, et al. (1983) proposed that the relative change in solubility of a protein due to the reduction of solvent dielectric constant can be empirically described by the expression:

$$\log S/S_0 = K''/D_s^2 \qquad (2\text{-}4)$$

where D_s: Dielectric constant of the solution

K': Constant embracing the dielectric constant of the original aqueous medium

S_0: extrapolated solubility

Non-ionic polymer solvent effect

Juckes (1971) proposed a correlation similar to the one given by Ogston (1958):

$$\text{Log } S = K - \beta\, \omega \qquad (2\text{-}5)$$

Where S: Solubility of protein

ω: Concentration of polymer(PEG)

β, K: Constants

The slope in equation (2-5) increases with the size of the protein molecule and is otherwise unaffected by conditions. This indicated that the technique of PEG precipitation of proteins is treated quantitatively the same way as the well-known salting-out technique we discussed earlier. The author showed that the precipitation of the proteins carboxyhemoglobin, ovalbumin and bovine serum albumin, as well as brome grass mosaic virus were found to conform with the above equation.

On the basis of studies with purified proteins, Middaugh, et al. (1979, 1980) reported the similar exponential dependence of solubility of proteins on PEG concentration. As shown in equation (2-6):

$$\text{Log } S = \text{Log } S_0 - K\,[\text{PEG}] \qquad (2\text{-}6)$$

Semilog plots of solubility data are usually linear, analogous to the salting-out phenomenon (Cohn, 1943), where S and S_0 are the solubilities in the presence and absence of PEG, respectively.

The above authors' equations for the relationship between PEG and protein concentration could not precisely fit the S-shaped experimental curve because the equation is valid only if applied close to the isoelectric point of proteins since far from the isoelectric point the protein can not be completely precipitated by PEG. Hasko, et al. (1982) suggested another kind of formula for PEG precipitation:

$$\text{Ln}\,(1 - C) = -k^* + b\,[\,\text{Ln}\,(S - S_v)\,] \qquad (2\text{-}7)$$

where k^*, b: Constants

C, S: PEG and protein concentration respectively

S_v: The final protein concentration (that does not change with the addition of further PEG)

Foster, et al. (1973) showed that PEG precipitation of yeast intracellular proteins can be expressed by a simplified equation:

$$\text{Ln}\,S + f\,S = X - a\,C \qquad (2\text{-}8)$$

$$\text{Here: } X = (\mu_i - \mu_0)/RT = \text{Ln}\,m_i + f_{ii} * m_i + f_{ij}\,m_j \qquad (2\text{-}9)$$

μ_i: Chemical potential of component i

μ_{i0}: Standard chemical potential of component i

m_i, m_j: Molality of component i, j respectively

f_{ij}: Coefficient for the interaction between i and j

S: Solubility of the protein

C: PEG concentration

f, a: Constant

The similar relationship as equation (2-8) was obtained by Atha and Ingham

(1981).

More recently, based on the excluded volume interpretation, Haire, et al. (1984) presented an expression:

$$\text{Ln } C_{Hb} + \text{Ln } \gamma'_{Hb} = \text{Ln } a_{Hb0} + A' C_{PEG} \quad (2\text{-}10)$$

Where C_{Hb}, C_{PEG}: Concentration of hemoglobin and PEG

γ'_{Hb}: Activity coefficient for hemoglobin in PEG free solution

a_{Hb0}: Activity that remains constant as long as the uncharged solid phase is present.

The linear relationship could hold until the solubility reaches 150 g/l, but at concentrations higher than 150g/l, deviations increase rapidly.

2.4.2 Models for Two Phase Extraction

A model predicting the partitioning of biomolecules should be able to successfully predict the splitting of the phase system at the first place.

Phase splitting

The key to be able to describe phase separation behavior in two phase system is to obtain an expression for the free energy associated with the formation of a polymer solution from pure components, the free energy of mixing ΔG_m:

$$\Delta G_m = \Delta H_m - T \Delta S_m \qquad (2\text{-}11)$$

Where ΔH_m: Enthalpy of mixing

ΔS_m: Entropy of mixing

T: Absolute temperature

Brooks, et al. (1985) applied Flory-Huggins lattice model for polymer solution to two phase system, and gave an expression of free energy of mixing of two polymers in a single solvent:

$$\begin{aligned} Gm &= kT [n_1 \operatorname{Ln} \phi_1 + n_2 \operatorname{Ln} \phi_2 + n_3 \operatorname{Ln} \phi_3 \\ &+ (n_1 + n_2 P_2 + n_3 P_3) (\phi_1 \phi_2 \chi_{12} + \phi_1 \phi_3 \chi_{13} + \phi_2 \phi_3 \chi_{23})] \end{aligned} \qquad (2\text{-}12)$$

Where component 1 refers to solvent and components 2 and 3 to two polymer species characterized by molecular weight parameters P_2 and P_3 (or their relative molecular volume), respectively.

ϕ_i $= n_i P_i/(n_1 + n_2 P_2 + n_3 P_3)$, the fraction of lattice sites occupied by polymer segments.

n_i: the number of molecules of component i on the lattice, $i = 1,2,3$

χ_{ij} $= z \Delta \omega_{ij}/kT$, the Flory interaction parameter.

$\Delta\omega_{ij} = (\omega_{ii} - \omega_{jj})/2$

ω_{ij}: the energy associated with a contact between component i and j.

Using stability criteria, they also derived the critical conditions for miscibility and the incipient appearance of two phases as:

$$\phi_{2c} = \phi_{3c} = (1 - \phi_{1c})/2 \quad (2\text{-}13)$$

$$\chi_{23c} = 1/P_2\phi_{2c} \quad (2\text{-}14)$$

It is very clear, from equations (2-13) and (2-14), that the larger the molecular weight of the polymer (or P_2), the easier the phase splits. The interactions between solvent and polymers (χ_{12} and χ_{13}) are not important because only χ_{23} (polymer-polymer interaction parameters) appears in phase separation criteria.

Kang and Sandler (1987) obtained Flory-Huggins interaction parameter χ_{ij} from osmotic pressure and plait point data of PEG/Dextran/H2O system, and also applied Flory-Huggins model to predict the phase separation. They concluded that the binodal is very sensitive to the interaction between unlike polymers. Their prediction reasonably fits the experimental binodals.

Partitioning of biomolecules

Flory-Huggins theory based models

With the assumptions of random coiling homopolymer chain and zero volume

change during mixing, Brooks, et al. (1985) also derived expressions for the partition effects of both neutral and charged molecular species (component i in the following expressions).

For uncharged molecular species:

$$K_i = C_i^T/C_i^B = f_{iB}/f_{iT} \exp(-\Delta\mu_{i0}/RT) \tag{2-15}$$

For charged species:

$$K_i = C_i^T/C_i^B = f_{iB}/f_{iT} \exp[-(\Delta\mu_{i0} + z_i F \Delta\psi /RT) \tag{2-16}$$

where K_i: Partition coefficient

C_i^T, C_i^B: Mole concentration of component i in top and bottom phases, respectively

f_{iT}, f_{iB}: Activity coefficient of i in top and bottom phases

μ_{i0T}, μ_{i0B} : Standard chemical potential of i in top and bottom phases.

$-\Delta\mu_{i0}$ $= -(\mu_{i0T} - \mu_{i0B})$

$$= P_i [(\phi_{1T} - \phi_{1B})(1 - \chi_{1i}) + (\phi_{2T} - \phi_{2B})(1/P_2 - \chi_{2i}) + (\phi_{3T} - \phi_{3B})(1/P_3 - \chi_{3i})] \tag{2-17}$$

z_i: Net charge of species i

F $= NA\, e$, Faraday constant

NA: Avogadro's number

e: Electron charge

$\Delta\psi$ $= \psi^T - \psi^B$

ψ^T, ψ^B: Electrostatic potential in the top and bottom phases

The above expressions show a number of features which are consistent with experimental observations:

(1). The partition coefficient is very sensitive to relevant properties (exponential dependence).

(2). The one-sided partition coefficient K_i for larger biomolecule i being distributed (or larger P_i makes larger K_i).

(3). The difference in polymer concentrations between the two phases ($\phi_{2T} - \phi_{2B}$) and ($\phi_{3T} - \phi_{3B}$) favors the one sided partition.

(4). The reduction of the molecular weight of one phase polymer will favor the partition of the biomolecule into that phase.

but, the ideal solution behavior (zero volume change of mixing) and the random coiling structure assumptions still remain questionable.

Baskir, et al., (1987) modified the liquid lattice theory of Scheutjens and Fleer (1980) for the adsorption of polymers from dilute solution to infinite flat surfaces, applied the theory to globular proteins, with the "hard sphere" assumption, and gave a Gibbs free energy model as follows:

$$g/kT = L_1 (u_{1|s}/kT - \chi_s \phi_{21}) + \Sigma_i L_i [\phi_{1,i} Ln(\phi_{1,i}/\phi_{1,*}) + \phi_{2,i} Ln P_i - (\phi_{1,i} - \phi_{1,*}) - (\phi_{2,i} - \phi_{2,*})/r] + \chi \Sigma_i L_i [\phi_{1,i} (<\phi_{2,i}> - \phi_{2,*}) - \phi_{1,*} (\phi_{2,i} - \phi_{2,*})] \quad (2\text{-}18)$$

where g: Gibbs free energy of the molecule/particle

k: Boltzmann's constant

L_i: Number of lattice size in layer i

$u_1|_s$: Adsorption energy for a solvent molecule

X_s : "Segment-protein" interaction energy

$\phi_{j,i}$: Volume fraction of j in layer i

$\phi_{1,*}$: Solvent volume fraction in bulk phase

$\phi_{2,*}$: Polymer volume fraction in bulk phase

$<\phi_{2,*}>$: Site average volume fraction of chain segments in layer i

χ: Flory-Huggins interaction parameter between polymer and solvent

r: Length of polymer chain in monomer units (i.e., the degree of polymerization)

The partition coefficient was defined by:

$$K_p = \exp[(g_B - g_A)/kT] \quad (2\text{-}19)$$

where g_A and g_B are particle Gibbs energies in phases A and B.

The model takes into account the biomaterial size and shape as a factor determining its interaction with the surrounding polymer solution. The model shows that the partitioning is extremely sensitive to the polymer-protein interaction energy. Their calculation agrees with the interpretation and experimental observations of other groups (Albertsson, et al, 1985; Higuchi, et al, 1983). In their model, protein molecule were considered as neutral species and the effects of both pH and ionic strength (salting effect) on partitioning were not investigated yet.

Diamond and Hsu (1988) found an approximate linear relationship between the natural logarithm of the partitioning coefficient of dipeptides and proteins and the concentration of one phase forming polymer. They also interpreted their results in terms of the Flory-Huggins model by introducing several assumptions. The physical meaning of their assumptions remains questionable.

Virial expansion based model

The Virial expansion of chemical potential of each component in terms of the molality of all the components in a multicomponent mixture was first suggested by Ogston (1962). King and Prausnitz (1986) extended Ogston's work to the partition of biomolecules in polymer-polymer-H2O system by adding a term considering the interactions between the biomolecule and each of the phase forming components.

With the assumption of dilute protein solution, the expression for the partition coefficient of protein molecule is:

$$K_p = m_{p,1}/m_{p,2} = \exp[\, a_{2,p} (m_{2,1} - m_{2,2}) + a_{3,p} (m_{3,1} - m_{3,2})] \qquad (2\text{-}20)$$

where $m_{p,1}$, $m_{p,2}$: Molality of a protein in phase 1 and 2

$a_{i,p}$: Interaction coefficient of one molecule of phase polymer i with one protein molecule, i=2 (1st phase polymer), i=3 (2nd phase polymer)

$m_{i,j}$: Molality of component i in phase j (i=2, 3; j=1, 2)

For the neutral salt effect, King, et al. (1987) added another term (electrostatic potential difference between two phases) to equation (2-20):

$$Kp = \exp[\, a_{2,p} (m_{2,1} - m_{2,2}) + a_{3,p} (m_{3,1} - m_{3,2}) + z_p F \Delta\psi/RT \,] \qquad (2\text{-}21)$$

It is clear that the above equation shows the one-sided feature of Kp with phase polymer concentration, which is similar to the prediction of Flory-Huggins based models. The model prediction agrees with experimental measurement for a number of Protein/PEG/Dextran/Salt systems, but still, most of the models assumed that protein molecule is uncharged, and the pH influence is also not accounted for by these models.

Blackburn and his coworkers (1982) also developed Virial expansion type of approach for the energetics of pairwise interactions of peptides in aqueous solution. Their method consists of two steps: first, the excess specific Gibbs free energy, G^{ex}, for

a binary solute system containing species A and B can be written as an empirical series expansion:

$$G^{ex} = g_{AA}\, m_{A2} + g_{BB}\, m_{B2} + 2\, g_{AB}\, m_A\, m_B \qquad (2\text{-}22)$$

Here, g_{AB} represents the free energy of pairwise interaction between species A and B, m_A and m_A are the molality of species A (CH_3-CONH-CH(CH_3)-$CONH_2$) and B (CH_3-CONH-CH_2-$CONH_2$), respectively, and A_2 and B_2 stand for dimers of A and B. Secondly, they treat polypeptides as two simple groups, -CONH-, the peptide group, and -CH_2-, the methylene group, using the simplification that a methyl group is equivalent to 1.5 methylene groups and a -CH- group to half a methylene group. This results in the following expression for free energy coefficient g_{AB} :

$$\begin{aligned} g_{AB} &= n_{CH2(A)}\, n_{CH2(B)}\, G_{CH2\text{-}CH2} + n_{Pep(A)}\, n_{Pep(B)}\, G_{Pep\text{-}Pep} \\ &\quad + \left(n_{CH2(A)}\, n_{Pep(B)} + n_{Pep(A)}\, n_{CH2(B)} \right) G_{CH2\text{-}Pep} \end{aligned} \qquad (2\text{-}23)$$

Where $G_{CH2\text{-}CH2}$, $G_{CH2\text{-}Pep}$ and $G_{Pep\text{-}Pep}$ are terms representing the free energies of interactions between two methylene groups, methylene group and amide group, and two amide groups respectively.

Several critiques should be placed on Blackburn's empirical expansion approach:

(1). Solute-solvent (or, functional group-solvent) interactions are not considered.

So, this approach cannot describe the hydration effect which plays a very important role in stabilizing the conformation of polypeptides and proteins and in partitioning these molecules between phases.

(2). Blackburn's approach works well only for dipeptides, but does not for some of the tripeptides they studied. It has been found that the addition of a methyl group changes the hydrophobicity greatly (Cohn and Edsall, 1943), but the author's approach could not predict this properly.

(3). The empirical approach suffers from being unable to represent amino acid sequence, the pH, charge, solvent and dipolar ion effects on the behavior of biomolecules in solution.

(4). Functional groups in peptide chain are oversimplified, the definition of functional groups and group counts are arbitrary, and steric effects are completely ignored due to the assumption that each amide group or methylene group is able to interact freely with any of these groups on another molecule. The distinct groups are so few (only two, -CONH- and -CH_2-) as to neglect significant effects in molecular structure.

In summary, there are models based on empirical correlations, but previous work has been largely empirical and applicable only to specific systems and specific phenomena and the coefficients or parameters of these equations lack clear physical meaning, thus are unable to make predictions. To date, no comprehensive theoretical models are available to satisfactorily predict the phase behavior of biomolecules (amino acids,

peptides, especially proteins) in solution. The charged, long-chain molecules present a real challenge to traditional modeling approaches.

Notation

C	=	Concentration
D	=	Dielectric constant
F	=	NA * e , Faraday constant (NA = Avogadro's number)
G	=	Gibbs free energy
G	=	Free energy of interaction in equation (2-23)
I	=	Ionic strength in molality scale
K	=	Thermodynamic chemical equilibrium constant, partitioning coefficient, or empirical constant
K'	=	Empirical constant in equation (2-1) and (2-2)
K"	=	Empirical constant in equation (2-4)
L	=	Number of lattice size in a layer of lattice model
P	=	Molecular weight parameter of polymer species
R	=	Gas constant
S	=	Solubility of proteins
S_v	=	The final protein concentration (that does not change with the addition of further PEG)
T	=	Temperature, K
X_s	=	"Segment-protein" interaction energy
χ_{ij}	=	Flory-Huggins interaction parameter.
a	=	Activity
$a_{i,p}$	=	Interaction coefficient of one molecule of phase 2 polymer i with one protein molecule, i=2 (1st phase polymer), i=3 (2nd phase polymer)
b	=	Constant
e	=	Electric charge, 1.602189*10-19C
f	=	Interaction coefficient
g	=	Gibbs free energy of the molecule/particle
g_{AB}	=	Free energy of pairwise interaction between species A and B, and m_A is the molality of species A.
g^{ex}	=	Molar excess Gibbs energy
k	=	Boltzmann constant, 1.380662*10-23J.K-1
k*	=	Constant
m	=	Molality, g-mole/kg of solvent
$m_{p,1}$, $m_{p,2}$	=	Molality of protein in phase 1 and 2

$m_{i,j}$	=	Molality of component i in phase j (i=2, 3; j=1, 2)
n_i	=	Number of molecules of component i on lattice, i=1,2,3
r	=	Length of polymer chain in monomer units
s	=	Solid
$u_1\|_s$	=	Adsorption energy for a solvent molecule
z_i	=	Net charge of species i

Greek Letters

Ω	=	Contact area between the protein molecules
β	=	Empirical constant
γ	=	Activity coefficient, mole fraction scale
λ	=	Repulsive interactions between molecules with like charges
σ	=	Induced dipoles
ϖ	=	Concentration of polymer
μ	=	Chemical potential
ϕ	=	Fraction of lattice sites occupied by polymer segments.
$\phi_{j,i}$	=	Volume fraction of j in layer i
$\phi_{1,*}$	=	Solvent volume fraction in bulk phase
$\phi_{2,*}$	=	Polymer volume fraction in bulk phase
$<\phi_{2,*}>$	=	Site average volume fraction of chain segments in layer i
ψ	=	Electrostatic potential in top or bottom phases
ϖ_{ij}	=	Energy associated with a contact between component i and j

Superscripts

B	=	Bottom phase
T	=	Top phase
ex	=	Excess property
0	=	Standard state

Subscripts

A,B	=	Phases
Hb	=	Hemoglobin
Osm	=	Osmotic
Soln	=	Solution
i,j	=	Any species
0	=	Zero ionic strength or zero PEG concentration
1	=	Solvent
2,3	=	Polymer species

Literature Cited

Abbott, N.L. and T.A. Hatton, "Liquid-Liquid Extraction for Protein Separation," Chemical Engineering Progress, 31-41, August 1988.

Albertsson, P.A., "Particle Fractionation in Liquid Two Phase Systems," Biochim. Biophys. Acta, 27, 378-395 (1958).

Albertsson, P.A., Partition of Cell Particles and Macromolecules, Wiley, New York, 1985.

Atha, D.H., and K.C. Ingham, "Mechanism of Precipitation of Proteins by Polyethylene Glycols," J. Biol. Chem. 256 (23), 12108-12117 (1981).

Baskir, J.N., T.A. Hatton, and U.W. Suter, "Thermodynamics of the Partition of Biomaterials in Two-Phase Aqueous Polymer Systems: Effect of the Phase Forming Polymers," Macromolecules, 20, 1300-11 (1987).

Bell, D.J., M. Hoare, and P. Dunnill, "The Formation of Protein Precipitates and Their Centrifugal Recovery," Advances in Biochemical Engineering/Biotechnology, 26, 1-72 (1983).

Blackburn, G.M., T.H. Lilley, and E. Walmsley, "Aqueous Solutions Containing Amino Acids and Peptides," J. Chem. Soc., Faraday Trans. 1 (78), 1641-1665 (1982).

Brooks, D.E., K.A. Sharp, and D. Fisher, "Theoretical Aspects of Partitioning," in Partitioning in Aqueous Two-Phase Systems, Theory, Methods, Uses, and Applications to Biotechnology, Academic Press, Inc., Orlando, 1985.

Cohn, E.J., Berggren, R.E.L., and Hendry, J.L., J. Biol. Chem., LXXXVII, 81 (1930).

Cohn, E.J., and Berggren,R.E.L., J. Biol. Chem., 101, 45 (1933).

Cohn, E.J., Ferry, J.D., Livingood, J.J., and Blanchard, M.H., J. of Am. Chem. Sco., 63, 17-22 (1941).

Cohn, E.J. and J.T. Edsall, Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Reinhold Publishing Corporation, New York, 1943.

Dalton, J.B., P.L. Kirk, and C.L.A. Schmidt, "The Apparent Dissociation Constants of Diiodotyrosine, Its Heat of Solution, and Its Apparent Heat of Ionization," J. Biol. Chem., 88, 589 (1930).

Dalton, J.B. and C.L.A. Schmidt, J. Biol. Chem., 103, 549 (1933).

Dalton, J.B. and C.L.A. Schmidt, J. Biol. Chem., 109, 241 (1935).

Dalton, J.B. and C.L.A. Schmidt, J. Gen. Physiol., ,19, 767 (1936).

Diamond, A.D. and J.T. Hsu, "Fundamental Studies of Biomolecule Partitioning in Aqueous Two-Phase Systems," Paper presented at the AIChE Annual Meeting, 1988.

Dunn, M.S., F.J. Ross, and L.S. Read, J. Biol. Chem., 103, 579 (1933).

Ellerton, H.D., G. Reinfield, D.E. Mulcathy, and P.J. Dunlop, J. Phys. Chem., 68, 398 (1964).

Florkin, M., J. Biol. Chem., XCIII, 631 (1931).

Foster, P.R., P. Dunnill, and M.D. Lilly, Biochim. Biophys. Acta 317, 505-516 (1973).

Green, A.A., J. Biol. Chem., XCIII, 551 (1931).

Greenwald, I., J. Biol. Chem., XCIII, 551 (1931).

Hade, Thesis, University of Chicago, December 1962.

Haire, R.N., W.A. Tisel, J.G. White and A. Rosenberg, Biopolymers, Vol. 23, 2761-2779 (1984).

Hasko, F. and R. Vaszileva, Biotech. & Bioeng., Vol XXIV, 1931-1939 (1982).

Higuchi, A., D. Rigby, and R.F.T. Stepto, in Symposium on Adsorption from Solution, Ottewill, R.H., C.H. Rochester, and A.L. Smith (eds.), Academic Press, New York, 273-286 (1983).

Hill, R.D. and J.G. Zadow, New Zealand J. Dair Sci. Technol. 13, 61 (1978).

Iverius, P.H., and T.C. Laurent, Biophys. Acta 133, 371 (1967).

Juckes, I.R.M., Biochim. Biophys. Acta 229, 535-546 (1971).

Kang, C.H., and S.I. Sandler, Fluid Phase Equilibria, 38, 245-272 (1987).

King, R., H.W. Blanch, and J.M. Prausnitz, Thermodynamics of Aqueous Polymer-Polymer Two Phase Systems," Presented at ACS National Convention in Anaheim, California. September 10, 1986.

King, R., H.W. Blanch, and J.M. Prausnitz, "Molecular Thermodynamics of Aqueous Two-Phase systems for biological separations," Presented at AIChE National Meeting in New York City, New York. November, 1987.

Kula, M.R., A. Buckmann, H. Hustedt, K.H. Kroner, and M. Korr, "Aqueous Two Phase Systems for the Large-Scale Purification of Enzymes," Enzyme Eng., 4, 47-53 (1977).

Kula, M.R., K.H. Kroner, and H. Hustedt, "Purification of Enzymes by Liquid-Liquid Extraction," Adv. Biochem. Eng., 24, 73-118 (1982).

Leward, D.A., "Protein-Polysaccharide Interactions," in Polysaccharides in Food, Blanchard, J.M.V., and J.R. Mitchell (eds.), London: Butterworths, 1978.

Li, C.H., Lyons, W.R., and Hormones, H., J. Gen. Physiol., 303 (1941).

McCarthy, E.F., J. Physiol., 104, 443-448 (1946).

Melander, W. and C. Horvath, "Salt Effects on Hydrophobic Interactions in Precipitation

and Chromatography of Proteins: An Interpretation of the Lyotropic Series," Arch. of Biochem. and Biophy., 183, 200-215 (1977).

Mellanby, J., J. of Physiol., 36, 289

Middaugh, C.R., W.A. Tisel, R.N. Haire, and A. Rosenberg, J. Biol. Chem. 254, 367-370 (1979).

Middaugh, C.R., E.O.Lawson, G.W. Litman, W.A. Tisel, D.A. Mood,and A. Rosenberg, J. Biol. Chem. 255, 6532-6534 (1980).

Ogston, A.G., Trans. Faraday Soc. 54, 1754-1761 (1958).

Ogston, A.G., Arch. Biochem. Biophys. 1, 39-51 (suppl.) (1962).

Olein, N.A., Chemical Engineering Progress, 83 (10), 45-48 (1987).

Palmer, A.H., G. Biol. Chem., 104, 359 (1934).

Pertzoff, V., J. Gen. Physiol., 961 (1927).

Polson, A., G.M. Potgieter, J.F. Largier, G.E.F. Mears, and F.G. Jourbet, "The Fractionation of Protein Mixtures by Linear Polymers of High Molecular Weight," Biochim Biophys Acta 82:463, 1964.

Smith, P.K. and E.R.B. Smith, J. Biol. Chem., 121, 607 (1937).

Smith, E.R.B. and P.K. Smith, J. Biol. Chem., 135, 273 (1940).

Stocking, C.R., "Precipitation of Enzymes During Isolation of Chloroplasts in Carbowax," Science 123:1032, 1956.

Stryker, M.H., M.J. Bertolini, and Y.-L. Hao, "Blood Fractionation: Proteins", in Advances in Biotechnological Processes (Mizrahi, A. and A.L. van Wezel, ed.), Vol. 4, 275-336 (1985).

Tomiyama and Schmidt, J. Gen. Physiol., 19, 379 (1935).

Woll, J.M., A.S. Dillon, R.S. Rahaman, and T.A. Hatton, in Protein Purification: Micro to Macro (R. Burgess, ed.), Alan R. Liss, New York, 1987.

Zhu, Y., L.B. Evans, and C.C. Chen, "Representation of Phase Equilibrium Behavior of Antibiotics," Biotechnology Progress, 266 (1990).

CHAPTER 3

THE PROTEIN FOLDING PROBLEM

The prediction of the conformations of proteins has developed from an intellectual exercise into a serious practical endeavor that has great promise to yield new stable enzymes, products of pharmacological significance (Fasman, 1989).

It has been realized that both the sequence and the complex interactions of the amino acid residues are responsible for the conformation and the phase behavior of a protein. By employing our knowledge about amino acid sequence and non-covalent forces operating within the system, it should be promising to predict the most stable conformation for an given polypeptide chain under the solution conditions (solvent, temperature, pH, buffer, and so on). However, to make such prediction requires a through command of all the interactions occurring in the polypeptide solution system, between side chains, side chains and solvents, and between solvent molecules, both intra- and inter-molecular type, and it also requires an algorithm of searching among all possible structural arrangements of the chain and the solvent molecules.

3.1 Protein Conformation

Protein molecules have several levels of structure. Primary conformation is the amino acid sequence (and the location of disulfide, Stryer, 1988), or the covalent-bond connection of amino acid residues of a protein. After a protein molecule is folded or partially folded, the molecule possess higher ranking structures. Secondary structure refers to the spatial arrangement of amino acid residues that are near one another in the

linear sequence, such as α-helices and β-sheets. Tertiary conformation refers to the spatial arrangement of amino acid residues that are far apart in the linear sequence, or the correct packing of the secondary structural units of the chain. The helix, sheet and other secondary structures are packed into a particular shape to make all the residues in the polypeptide chain feel comfortable under particular environmental conditions. Quaternary conformation refers to the spatial arrangement of subunits (such as a single polypeptide chain structure of hemoglobin).

For example, Figures 3-1a and 3-1b show the amino acid sequence, the secondary structural units, and the folded native conformation of BPTI (Bovine Pancreatic Trypsin Inhibitor), a single polypeptide chain protein. It has a molecular weight 6,520 with 58 residues. The main portion of the BPTI molecule consists of a right-handed α-helix and an antiparallel β-sheet, they both have regularly repeating patterns. As I have mentioned earlier, BPTI is one of the very simplest proteins, but it is already very complicated as you can imagine. All kind of forces come into play when you fold or partition this protein.

After folding, as we can see from Figure 3-2, the hydrophilic residues will be exposed to the surface of the protein, and the hydrophobic residues will be buried in the protein core region. People only partially know how and why protein molecules fold into this particular conformation. There are two aspects regarding the folding of a protein into its final configuration, the thermodynamics and the kinetics. In terms of the work of Anfinsen (1961), the native conformation of a protein should be at the lowest free

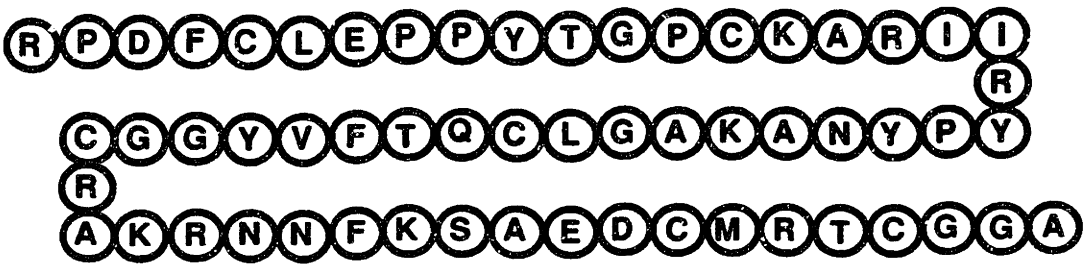
energy state of the protein system. People also find that there are more than one folding pathways which are manipulated by folding kinetics. Our current work is mainly focused on the thermodynamics part of the problem since for helix and sheet formation, the kinetics might play less role.

Figure 3-1(a). Conformational States of BPTI

Bovine Pancreatic Trypsin Inhibitor (BPTI) MW: 6,520

Residues: 58

Primary structure



Secondary structure

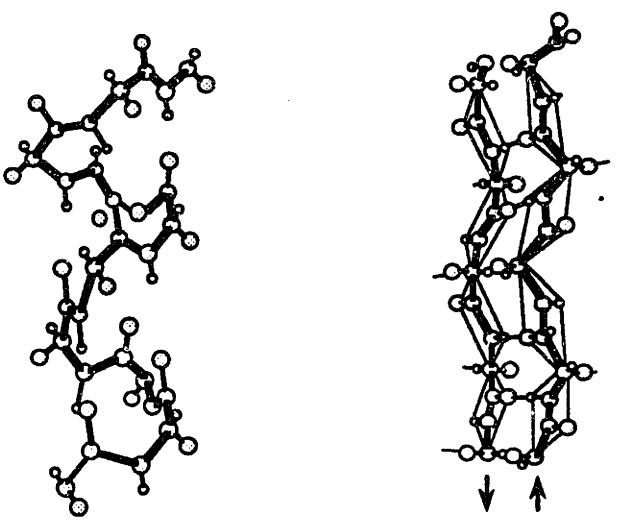
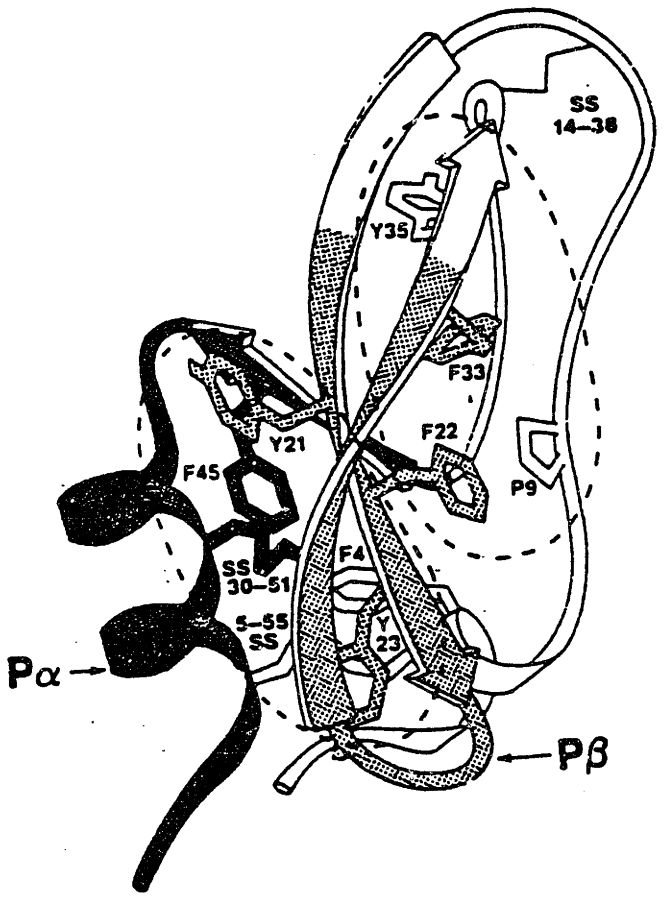
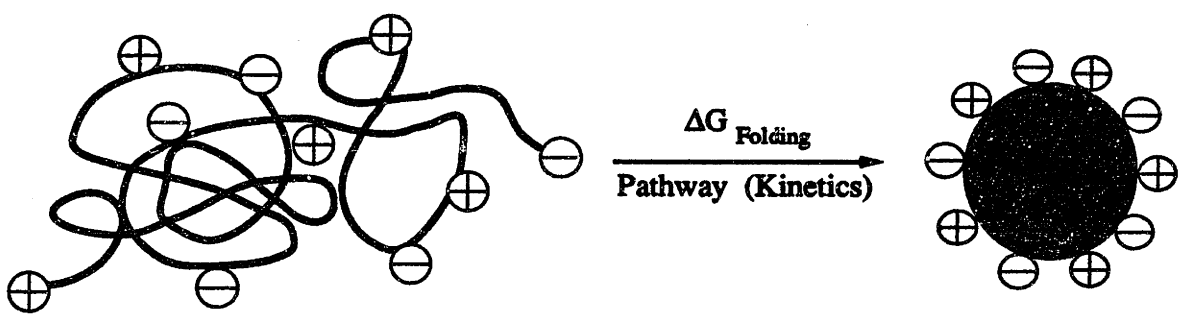


Figure 3-1(b). Tertiary Structure of BPTI



(Oas and Kim, 1988)

Figure 3-2. Schematic Diagram of Protein Folding



Coil (unfolded)

Native (folded)

3.2 Prediction of Protein Conformation

The biological function, activity and the stability of proteins are intercorrelated, and are closely reliant on the conformation of proteins. It has been a long cherished dream to predict the tertiary structure of proteins from the amino acid sequence since the experimental work of Anfinsen et al. (1961) on bovine pancreatic ribonuclease, who found that the incorrect secondary and tertiary structure results in inactivation of enzymes, and the chain will be driven by thermodynamic forces (through noncovalent interactions) to rearrange itself toward the most probable form, native protein. They concluded that the information for the native secondary and tertiary structures of proteins is contained in the amino acid sequence itself.

The prediction of the secondary structure of proteins from the amino acid sequence should be the first trial in protein folding studies since according to statistics, on the average, fifty percent of the residues in a protein is in alpha-helices (35%) or beta-sheets (15%), and 25% in reverse turns, another 25% in coiled form (Schulz and Schirmer, 1979). For example, in protein BPTI, the $P\alpha P\beta$ domain alone contains 30 residues out of the 58 residues. Another reason to carry on the study of protein secondary structure prediction is that secondary structure might be the early formed subunits upon which the higher order structures are built in the folding process. The experimental results of Oas and Kim (1988) implies that native-like structure can form early in protein folding.

Practically, the existence of certain secondary structures which can remain their

"native" conformation in proteins at low temperatures makes the prediction of the "native" secondary structure of proteins feasible. These secondary structural blocks are called aqueous, monomeric "autonomous folding units" (AFUs) (Goodman and Kim, 1989). A few examples of the AFUs are C-peptide (residues 1-13 of ribonuclease A) showing ~30% α-helical structure at 0° C in aqueous solution (Shoemaker, et al., 1987), S-peptide (residues 1-20 of ribonuclease A), and Pα (residues 43-58) and Pβ (residues 20-33) in BPTI (Oas and Kim, 1988). Oas and Kim have designed and synthesized this 30 residue synthetic analogue designated as [30-51], which is a disulfide-bonded peptide pair. The PαPβ is >90% folded in aqueous solution at pH 6 and temperature 4°C. They also showed that PαPβ contains much of the secondary and tertiary structure present in the corresponding region of native BPTI.

There has been a great amount of work done in the area of prediction of secondary structure of proteins since the early work of Blout et al. (1960), Davies (1964), Kotelchuck, Dygert, and Scheraga (1969), Lim (1974), Schulz, et al. (1974a,b), Chou and Fasman (1974), Garnier, et al. (1978), Fasman (1989), Chan and Dill (1989), Tirado-Rives and Jorgensen (1991), and others. In particular, helix-coil transition and the prediction of α-helix structure has been subject to intensive studies. However, the beta pleated sheets were not treated until the work of Ptitsyn and Finkelstein (1970), though they predicted only one of the six beta sheets of Ribonuclease and two of the six beta regions of papain.

Schulz and Schirmer (1979) gave a comprehensive review on the prediction of

secondary structure from the amino acid sequence. They placed the "prediction methods" into two categories: probabilistic and physico-chemical. The former extracts rules and parameters using purely statistical analysis of the protein data base, or relies on the known correlations between sequence and structure available in the data base. While the latter applies structural information (both experimental and theoretical) from outside the data base.

The original probabilistic method (Dirkx, 1972) determined the singlet propensities of each of the 20 residues to occur in α-, β-and reverse turn conformation by considering each residue separately without attention to near or far neighbors. Another rigorous probabilistic method was proposed by Periti (1974) using doublet propensities which account for binary residue-residue interactions. Periti limited the extent of interaction within a distance of up to 6 residues in each direction from a central residue. This model contains no adjustable parameters, and predicts α- and β- structures simultaneously. Wu and Kabat (1971) and Kabat and Wu (1972, 1973) predicted both secondary structure and complete chain folding of cytochromes C and immunoglobulins on the basis of triplet information of the (ϕ, Ψ)-angles of the central residue of all the triplets considered.

Schulz and Schirmer (1979) divided the physico-chemical prediction methods again into two classes: the methods based on statistical mechanics which formulates the problem in terms of partition function, and the methods based on stereochemical considerations which treat the problem according to heuristic rules on the relationship

between the geometrical location of the residues and their polarity/hydrophobicity and size.

Zimm and Bragg theory (1959) was among the earliest statistical mechanics treatment of the phase transition between helix and random coil in polypeptide chains. Zimm and Bragg constructed the partition function and the probability of hydrogen bonding. While the method by Kotelchuck and Scheraga (1968, 1969) can be considered as a representative of the statistical methods. They classified residues as helix making and helix breaking, proposed that four or more helix making residues in a raw will make a helix with two helix breaking ones terminating the helix, and introduced nucleation and unidirectional growth in their algorithm, assumed (implicitly) nearest neighbor interactions, and finally established statistical weights from empirical data for left-handed α-helix and extended sheet. Here, due to the difficulties in calculating the free energy of the whole system (including both chain and solvent), they fit energy function using observed propensities of a given residue from "Ramachandran Map". They correctly predicted 61% of the helices and 78% of the total residues in four proteins as helix or coil.

A detailed review of Chou-Fasman method (Chou and Fasman, 1974) is given by Prevelige and Fasman (Fasman, 1989). In brief, Chou and Fasman investigated 15 sample proteins (with 2473 residues in all) whose conformation had been determined by x-ray crystallography, and assigned empirical values for their tendencies to form α-helix $P\alpha = f\alpha / <f\alpha>$ and sheet $P\beta = f\beta / <f\beta>$ for the 20 naturally occurring amino acid

residues. Here $f\alpha$ and $f\beta$ are the fractions of each type of residue in helix and sheet, respectively. $P\alpha$, $P\beta$ and Pc are normalized conformational parameters (or potentials) for the helix, beta sheet, and the coil conformations. They classified all the 20 residues as formers, breakers, and indifferent to α-helical and β-sheet regions. Based on the facts that residues with the highest helical potential (with high $P\alpha$) reside mostly at the helix center while strong helix breakers (with low $P\alpha$) cluster just beyond the helix ends, Chou and Fasman proposed that helix nucleation could start at the center and propagate in both directions. They also applied these empirical parameters and postulates to predict the secondary structure of proteins such as BPTI. Their algorithm predicted that BPTI has 29% α-helix and 33% β-sheet, which was 87% correct for helices and 95 % correct for the β-sheets (see Table 3-1). And the overall accuracy of prediction of the secondary structure of globular proteins is 80%. Their computed $P\alpha$ values are within 10% of the experimental Zimm-Bragg helix growth parameters, s, evaluated from poly (α-amino acids). They also compared the frequencies at the helical ends to the experimental Zimm-Bragg helix initiation parameter, σ (derived from synthetic polymers), the correlation coefficient between the logarithms is +0.75.

A comparison of the prediction accuracy ("percentage predicted correct") for three well-known methods is given in Table 3-1.

Table 3-1. Some Model Prediction Results for α-Helices and β-Sheets

	BPTI		RNase	Adenylate kinase		Proteins	
	α-helix	β-sheet	β-sheet	α-helix	β-sheet	α-helix	β-sheet
Chou & Fasman (1974)	87%	95%		9/10	5/5	80%	86%
Ptitsyn & Finkelstein(1970)			1/6	9/10	3/5		
Kotelchuck & Scheraga(1969)						61%	78%

The GOR method (Garnier, Osgusthorpe, and Robson, 1978; Garnier and Robson, 1989) is another example of statistical procedure. The method is based on information theory. Its basic conformational parameters are very similar to those used by Chou and Fasman (1974). It assigned four conformational states of a residue, α-helix, extended chain, reverse turn and coil. Four information parameters were obtained from statistical analysis of 25 proteins of known sequence and conformation. The whole method is based on empirical correlations between the conformation of a singlet residue and the identities of the residue itself and of the neighboring residues at different locations surrounding it up to 8 residues at each side along the chain. The method correctly predicted 49% of the residue states in a sample of 26 proteins.

Based on the stereochemistry of the Pauling-Corey α-helix and β-structure, Lim's theory (1974a, b) classified residues according to their hydrophobicity and size, established a system rules to locate α-helical, and β-structural regions, respectively. In the case of α-helices, he emphasizes the intertwining of hydrophobic triplets at positions such as i, i+1, i+4, or i, i+3, i+4. He pointed out that large polar residues can stabilize adjacent nonpolar clusters by shielding them from the solvent water. In the case of β-

structure, the subunits of internal, semi-surface, and surface sheet strands are defined and considered individually. Lim theory gives a satisfactory agreement with experiment (80% accuracy for a-helices and 85% accuracy for b-sheets) for the 25 proteins tested by the author himself. However, it does not completely predict the experimentally observed α-helical regions and β-structural regions, and also predicts "excess" β-regions.

Using data on 62 proteins of known structure with more than 10,000 residues, Kabsch and Sander (1983) tested the three most widely used methods, Chou and Fasman (1974), Lim (1974), and Garnier, et al., (1978), for the prediction of protein secondary structure from the amino acid sequence. Their comparison showed that for the three states helix, sheet, and loop/turn, a success rate can be of about 50% with Chou and Fasman's method, and of 55-56% with either GOR or Lim's method. The typical results of their comparison of the methods are given in Table 3-2 for each protein and each method as the percentage of residues predicted correctly in a three state description of secondary structure.

Table 3-2 Comparison on Some Secondary Structure Prediction Methods

Proteins	BPTI	RNaseS	CytC	Lysozyme	Insulin	Adenylate Kinase	Proteins
No. Residues	58	124	103	129	51	194	
$\mathcal{P}$ Structure	$\alpha+\beta$	$\alpha+\beta$	α	$\alpha+\beta$	α	$\alpha+\beta$	
Methods	Fraction Correct (%) Kabsch and Sander (1983)						
Chou&Fasman(1974)	71	57	43	58	40	52	50
Lim (1974)	72	66	55	64	71	65	55-56
GOR(Garnier,1978)	74	66	54	63	44	73	55-56

Recently, Chan and Dill (1989) suggested that the driving force for formation of secondary structures in proteins may be nonspecific steric interactions rather than hydrogen-bonding or other specific interactions.

In summary, the vital role of solvent water in solvation and in stabilizing the native conformation of protein molecules has been recognized by both experimentalists and theoreticians, but the mechanism for the function of the solvent is still poorly understood. Therefore, almost all the prior prediction methods of secondary structure of proteins neglect the role of solvent in the formation of secondary structure and in the complete chain folding of the proteins, or the previous methods could only deal with an isolated secondary structure or chain in vacuum. This is one key reason that the previous probabilistic, statistical mechanics, and stereochemical heuristic methods have resulted in a big prediction error from the native conformation of the proteins concerned. According to Kabsch and Sander (1983), the prediction success rate is, at most, around 50~56% for the secondary structures of proteins, or less than 56% of the residues are predicted correctly.

It is the attempt of this work to use molecular thermodynamics to understand the rules governing the formation of secondary structures of polypeptides, to correlate peptide chain conformation with amino acid sequence and all the interactions occurring in the system, and to give a better understanding of protein folding phenomena.

Notation

D	=	Dielectric constant
I	=	Ionic strength in molality scale
K	=	Thermodynamic chemical equilibrium constant
Kp	=	Two phase partitioning coefficient
K'	=	Empirical constant in Equation 1
K"	=	Empirical constant in Equation 4
M_s	=	Solvent molecular weight, kg/kmol
R	=	Gas constant
S	=	Solubility
T	=	Temperature, K
X	=	Concentration variable, see Chen and Evans (1986)
k	=	Boltzmann constant, $1.380662*10^{-23}$J.K-1
m	=	Molality, g-mole/kg of solvent
s	=	Solid

Greek Letters

Ω	=	Contact area between the protein molecules
β	=	Empirical constant in equation 1
γ	=	Activity coefficient, mole fraction scale
λ	=	Repulsive interactions between molecules with like charges
σ	=	Induced dipoles

Subscripts

i,j,k	=	Any species
m	=	Molecular species
s	=	Solvent species

Literature Cited

Anfinsen, C.B., E. Haber, M. Sela, and F.H. White, Jr., Proc. Nat. Acad. Sci. U.S., 47, 1309, 1961.

Blout, E.R., C. de Loze, S.M. Bloom, and G.D. Fasman, J. Amer. Chem. Soc. 82, 3787-3789, 1960.

Chan, H.S. and K.A. Dill, Macromolecules, 22, 4559-4573 (1989)

Chou, P.Y., and G.D. Fasman, Biochemistry, 13, 222-245, 1974.

Davies, D.R., J. Mol. Biol. 9, 605-609, 1964.

Dirkx, J., Arch. Int. Physiol. Biochim. 80, 185-187 (1972).

Fasman, G.D., in "Prediction of Protein Structure and the Principles of Protein Conformation," 193-316, Edited by G.D. Fasman, Plenum Press, New York, 1989.

Garnier, J., D.J. Osgusthorpe, and B. Robson, J. Mol. Biol., 120, 97-120, 1978.

Garnier, J. and B. Robson, in "Prediction of Protein Structure and the Principles of Protein Conformation", Edited by Fasman, G.D., Plenum Press, New York, 417-467 (1989).

Goodman, E.M., and P.S. Kim, Biochemistry, 28, 4343-4347 (1989).

Kabat, E.A., and T.T. Wu, Proc. Nat. Acad. Sci. USA 69, 960-964 (1972).

Kabat, E.A., and T.T. Wu, J. Mol. Biol., 75, 13-31 (1973).

Kabsch, W., and C. Sander, FEBS Letters, 155 (2), 179-182 (1983).

Kotelchuck, D., M. Dygert, and H.A. Scheraga, Proc. Nat. Acad. Sci. U.S., 63, 615 (1969).

Kotelchuck, D., and H.A. Scheraga, Proc. Nat. Acad. Sci. U.S.A., 61, 1163-1170 (1968).

Kotelchuck, D., and H.A. Scheraga, Proc. Nat. Acad. Sci. U.S.A., 62, 14-21 (1969).

Lim, V.I., J. Mol. Biol., 88, 857-872 (1974a).

Lim, V.I., J. Mol. Biol., 88, 873-894 (1974b).

Oas, T.G., and P.S. Kim, Nature, 33 (6194), 42-48 (1988).

Periti, P.F., Boll. Chim. Farm. 113, 187-218 (1974).

Ptitsyn, O.B., and A.V. Finkelstein, Biofizika 15, 757 (1970).

Scheutjens, J.M.H.M., and G.J. Fleer, J. Phys. Chem. 83, 1619-1635 (1979).

Scheutjens, J.M.H.M., and G.J. Fleer, J. Phys. Chem. 84, 178-190 (1980).

Schulz, G.E., M. Elzinga, and R.H. Schirmer, Three Dimentional Structure of Adenyl Kinase, Nature 250, 120-123, (1974a).

Schulz, G.E., C.D. Barry, J. Friedman, P.Y. Chou, G.D. Fasman, A.V. Finkelstein, V.I. Lim, O.B. Ptitsyn, E.A. Kabat, T.T. Wu, M. Levitt, B. Robson, and K. Nagano, Nature 250, 140-142 (1974b).

Schulz, G.E., and R.H. Schirmer, "Principles of Protein Structure", Springer-Verlag, New York, 1979.

Shoemaker, K.R., R. Fairman, P.S. Kim, E.J. York, J.M. Stewart, and R.L. Baldwin, Cold Spring Harbor Symposia on Quantitative Biology, Volume LII., 391, 1987.

Stryer, L., Biochemistry, W.H. Freeman and Company, New York, 1988.

Tirado-Rives, J. and W.J. Jorgensen, Biochemistry, 30, 3864-3871 (1991).

Wu, T.T., and E.A. Kabat, Proc. Nat. Acad. Sci. USA 68, 1501-1506 (1971).

Zimm, B.H., and J.K. Bragg, J. Chem. Phys., 31 (2), 526-535 (1959).

CHAPTER 4

MOLECULAR THERMODYNAMICS PERSPECTIVES OF PROTEIN PHASE PARTITIONING AND FOLDING

Protein phase partitioning and protein folding are two important challenges facing today's biotechnology industry. The conformation of a protein is very important and closely related to its phase partitioning behavior since the immediate neighbors of the residues in a protein solution change greatly with the conformation of the protein. Likewise, the physical properties and the phase behavior of the protein depend strongly on its molecular conformation (the folded or unfolded state).

It is the same physical interactions working in the biological solution systems that govern both the conformation of a protein and its phase partitioning. There exist great opportunities in better understanding and predicting the phase behavior of the charged, long-chain biomolecules and in building the bridges between the solution properties of the building blocks (amino acids) and the biopolymers (proteins), and between the protein folding and its partitioning.

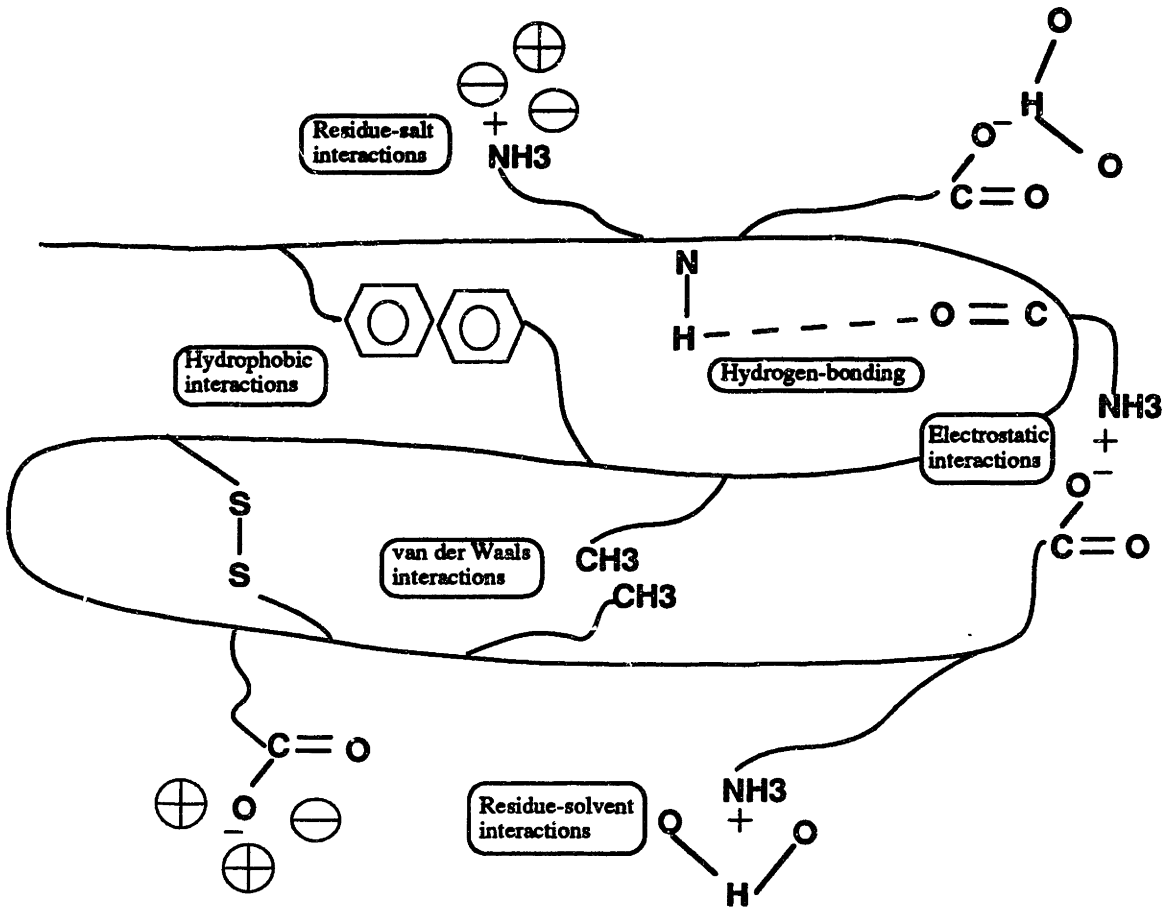
4.1 Essential Interactions in Protein Folding and Phase Partitioning

Proteins are conformational polyelectrolytes with zwitterionic features. The side chains of the twenty residues differ in size, shape, charge, hydrophobicity, and/or hydrogen-bonding capacity. The diversity of the 20 amino acid side chains and their integrated mutual interactions endow protein molecules with properties far complicated than small organic molecules and much different from the simple summation of the

residues in the polypeptide chain. In their isoelectric solutions, these biological molecules share the characteristic that they exist as dipolar ions, or zwitterions with distinct dual charges. The dual charges contribute to the ionic strength of the solution and make the biomolecules possess both the characteristics of nonelectrolyte and those of electrolytes.

The interactions affecting the state of proteins in a polar aqueous environment involving various agents are very complicated. As we can see from Figure 4-1, all kinds of forces come into play in protein solutions. These forces include hydrophobic interactions, hydrogen-bonding, van der Waals interactions, salt bridges, interactions with solvent molecules and with salts, etc. These forces might interplay, then the problem will be much more difficult. How do we quantify these complicated forces? These manifold non-bonding interactions are individually energetically weak but collectively of great importance with respect to the phase behavior and the conformational stability of proteins.

Figure 4-1. Interactions Present in Protein Folding and Phase Partitioning
(Modified after Anfinsen, C.B., 1959)



4.2 Molecular Thermodynamics Perspectives

As mentioned earlier, molecular thermodynamics can facilitate as a useful tool for us to develop an improved understanding of the protein phase partitioning and folding phenomena.

Chemical engineers have been very successful in treating conventional chemical systems using molecular thermodynamics. These successful examples include athermal solutions in most of the polymer systems described by Lattice Model (or, Flory-Huggins model)(Flory, 1941; Huggins, 1941), non-electrolyte regular solutions in petrochemical systems described by NRTL (non-random two liquid) model (Renon and prausnitz, 1968), conventional aqueous electrolyte systems (Pitzer, 1973) represented by Electrolyte NRTL model (Chen, et al, 1982, 1986), and surfactant systems (Blankschtein, 1986; Chen, 1989). Most of these models have been serving chemical and petroleum systems very well.

From molecular thermodynamic view point, biological solutions are profoundly different from conventional solution systems. A comparison between biomolecular systems and conventional chemical systems is given in Table 4-1.

Table 4-1. Complexities of Biomolecular Systems

Biomolecules	Similar Chemical Systems
Zwitterions	Dipolar ionic surfactant (low molecular weight)
Strong interactions	Electrolytes (small molecules)
Chain structure	Polymers (random coil, nearly athermal)
Conformation	Micelles (no tertiary, quaternary structure)
Multicharges	Polyelectrolytes (no high rank conformations)

As seen above, biomolecules have all the complexities existing in conventional chemical systems. As a result, they are more complicated than any conventional chemicals or polymers. The important requirement of the thermodynamics model for biomolecular solution is to be able to represent simultaneously the characteristics of the zwitterions, strong nonidealities, high molecular weight, chain conformation and multicharges. The existing molecular thermodynamic models available for electrolyte systems (or, systems with strong interactions), can only deal with small molecules of similar size. Given the large heat of solution and the non-Gaussian steric distribution of the polypeptide chain, the lattice model of Flory-Huggins is not appropriate for the molecules with helical, pleated sheet, and spherical shapes, or ones that form "aggregates" (polypeptides, proteins, and DNA are typical examples), or ones that are very compact. Energetically, the lattice model is also completely inappropriate for intramolecular hydrogen-bonding systems, it can, at most, represent only the polymeric character of a fully unfolded polypeptide chain. The existing NRTL types of models also can not represent the polymeric and conformational characteristics of biopolymers. In fact, most of the conventional chemical compounds do not have conformational problems. The only system which possesses any similarity to the conformation problem of the biopolymers is the surfactant system, where solute molecules have preferred orientation at the surfactant concentrations above CMC (we might call these "orientations" as secondary conformations of micellar systems).

In other words, considering the differences of biological solutions from ordinary

chemical solutions, the development of a molecular thermodynamic model for biomolecular systems needs to take into account not only the size effect (or, the non-ideal entropy of mixing), but also the short range interactions (or, the heats of interactions due to hydrophobicity, hydrogen-bonding, Van der Waals forces) and the long range interactions (electrostatic forces between charged groups). More importantly, four additional distinct characteristics of biological systems should also be considered. First, all biomolecules are zwitterionic species. Second, polypeptides and proteins are conformational polyelectrolytes with secondary (including supersecondary), tertiary (including domains) and quaternary structures. Third, solvent plays a vital role in affecting the conformation and the partition of biomolecules. Finally, residue-residue (or side chain-side chain) interactions in a peptide chain are critical to achieve and maintain the specific conformational state of the molecule.

Aiming at the investigation of the solution behavior of protein systems, we began with a systematic study of the properties of systems containing smaller molecules which incorporate functional groups present in proteins. As the primary building units of polypeptides and proteins, the twenty amino acid residues are deemed as the appropriate definition of "segment" units to be able to combine both biological relevance and structural simplicity, and the definition of group counts are accurate. The interactions occurring in solutions of amino acid, peptide and their derivatives contain the fundamental features of the interactions in protein solutions. The studies for these interactions would give insight into the factors which affect the behavior of proteins.

According to this approach, the challenges are to express the properties of polypeptides and proteins, in terms of the properties of twenty amino acid residues and the forces governing their mutual interactions and their interactions with solvent, salts, additives and other species present in solutions.

Literature Cited

Anfinsen, C.B., The Molecular Basis of Evolution, John Wiley & Sons, Inc., N.Y., 1959.

Chen, C.-C., "A Molecular Thermodynamic Model for the Gibbs Energy of Mixing of Micellar Solutions," paper presented at the Annual AIChE Meeting, San Francisco, 1989.

Chen, C.-C. and L.B. Evans, "A Local Composition Model for the Excess Gibbs Energy of Aqueous Electrolyte Systems," AIChE J., 32, 444 (1986).

Flory, P.J., Chem. Phys., 9, 660 (1941).

Huggins, M.L., J. Chem. Phys., 9, 440 (1941).

Olein, N.A., Chemical Engineering Progress, 83 (10), 45-48 (1987).

Pitzer, K.S., J. Phys. Chem., 77, 268 (1973).

Prausnitz, J.M., Fluid Phase Equilibria, 53, 439-451 (1989).

Renon, H. and J.M. Prausnitz, "Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures," AIChE J., 14, 135 (1968).

CHAPTER 5*

PHASE PARTITIONING OF BIOMOLECULES
(Amino Acid Solubility Studies)

* Chapter 5 has been published in *Biotechnology Progress,* 5 (3), 111-118 (1989), with Chau-Chyun Chen and Lawrence B. Evans

Introduction

The development of biotechnology has resulted in an increased interest in the development of efficient methods for the separation, concentration, and purification of amino acids, peptides, proteins and other biological products from fermentation and cell culture media (Albertsson et al., 1985; Abbott and Hatton, 1988).

Fractional precipitation and crystallization are widely used techniques for the separation of proteins from fermentation broth, human and animal blood plasma, etc. Therefore, the study of the solubilities of amino acids, peptides, and proteins in aqueous solution is of fundamental importance to the design and optimization of downstream separations of protein products (Bell et al., 1983).

The investigation of the solubilities of amino acids and small peptides in aqueous solutions is the first step toward the representation of the complex mechanisms behind the solubility behavior of peptides and proteins. These biological molecules share the characteristic that they exist as dipolar ions, or zwitterions, in the liquid phase. Hitchcock (1924), Dalton, et al. (1930), Cohn and Edsall (1943), and Greenberg (1951), among others, have made important pioneering research in representing the amino acid

solubilities. However, to date, no comprehensive theoretical models are available to represent the solubility behavior of amino acids, peptides, and proteins in aqueous solutions.

The objective of this study is to develop a generalized molecular thermodynamic framework to represent the solubility behavior of the biomolecules. Specifically, this study is aimed at developing a framework that can represent the solubility behavior of amino acids and small peptides as affected by temperature, ionic strength, dipolar species concentrations, solvent compositions, pH, etc.

It should be noted that, although this work applied the proposed framework only to the solubility behavior of the biomolecules, the framework is equally applicable to other phase partitioning phenomena.

5.1 Prior Modeling Work

Several empirical correlations have been proposed in the literature to describe the solubility behavior of amino acids, peptides, and proteins (Bell, et al., 1983). Cohn and Edsall (1943) proposed that the logarithm of the protein solubility in the aqueous phase is linearly proportional to the salt ionic strength.

$$\log S = \beta - K' I \tag{5-1}$$

Melander and Horvath (1977) attempted to correlate the K' constant in Equation (5-1) with the protein contact area, induced dipole, and repulsive interactions.

$$K' = \Omega \sigma - \lambda \tag{5-2}$$

Bell, et al. (1983) proposed that the relative change in the protein solubility at its isoelectric point with a change in the dielectric constant of the solvent mixture can be described by the expression:

$$\log \Delta S = K''/D^2 \tag{5-3}$$

Recently, Nass (1988) proposed to represent the amino acid solubilities with a molecular thermodynamic framework. The chemical theory was used to describe the amino acid chemical reaction relationships among the true ionic species in the aqueous phase; and the Wilson excess Gibbs energy equation (Wilson, 1964) was used to describe the physical interactions between the two apparent components of the amino acid solutions, i.e., water and amino acid. Nass was able to successfully represent the solubilities of amino acids in pure water as functions of temperature and pH. However, due to the lack of a comprehensive excess Gibbs energy expression to fully describe the physical interactions among the true species of the biomolecular systems, Nass' work was limited to aqueous single amino acid systems. The effects of ionic strength, dipolar species concentrations, and solvent compositions on the amino acid solubilities were not investigated.

In brief, previous works have largely been empirical and applicable only to limited systems and limited phenomena. To date, no comprehensive thermodynamic framework is available to describe the solubility behavior of amino acids and small peptides in aqueous phase.

5.2 Proposed Thermodynamic Framework for Solubilities of Amino Acids

Amino acids, peptides, and proteins are ampholytes which exist in solutions partly as neutral dipolar species and partly as cations and anions. The neutral dipolar species are the predominant species in isoelectric solutions of amino acids and proteins. Being dipolar ions or zwitterions, these neutral dipolar species carry dual electric charges even in isoelectric solutions. The ampholyte solution chemistry and the dipolar ionic structure are the key factors characterizing the solubility behavior of the biomolecules.

The following four chemical reactions take place with the dissolution of amino acids into aqueous phase:

$$A_{(s)} \quad <--> \quad A^{+-} \qquad (5\text{-}4)$$

$$A^{+} \quad <--> \quad A^{+-} \quad + \quad H^{+} \qquad (5\text{-}5)$$

$$A^{+-} \quad <--> \quad A^{-} \quad + \quad H^{+} \qquad (5\text{-}6)$$

$$H_2O \quad <--> \quad OH^{-} \quad + \quad H^{+} \qquad (5\text{-}7)$$

The first reaction is the formation of the dipolar species in the liquid phase. The dipolar species then participates in acid-base reactions to form amino acid cations

(Equation 5-5) and amino acid anions (Equation 5-6).

The chemical equilibrium relationships for the four reactions above can be written as follows:

$$K_s = a_{A+-} \quad (5\text{-}8)$$

$$K_{A+} = a_{A+-}\ a_{H+}\ / a_{A+} \quad (5\text{-}9)$$

$$K_{A-} = a_{A-}\ a_{H+}\ / a_{A+-} \quad (5\text{-}10)$$

$$K_w = a_{H+}\ a_{OH-}\ / a_{H2O} \quad (5\text{-}11)$$

where

$a_i = x_i \gamma_i$ for solvent species (5-12a)

$a_i = x_i \gamma_i^*$ for solute species (5-12b)

The thermodynamic chemical equilibrium constants of Equations (5-9) and (5-10) are available in the literature (Cohn and Edsall, 1943; Greenstein and Winitz, 1961).

As shown in Equations (5-8) to (5-12), the solubilities of amino acids are determined by the thermodynamic chemical equilibrium constants of the reactions, and the concentrations as well as the activity coefficients of the true species in the liquid phase. In representing the solubilities of amino acids, it is essential that the solution chemistry and the activity coefficients of the true species be properly accounted for. These quantities are affected by variations in system temperature, ionic strength, dipolar

species concentrations, solvent compositions, pH, etc.

In the proposed molecular thermodynamic framework, we incorporate explicitly the solution chemistry of the biomolecules and we apply the Electrolyte NRTL model (Chen and Evans, 1986; Scauflaire, et al., 1989) to compute the activity coefficients for the true species. The Electrolyte NRTL model provides proper accounts of the physical interactions that exist in the biomolecular solutions. These physical interactions involve binary molecule-molecule interactions (including molecular species and neutral dipolar species), binary molecule-ion interactions, and binary ion-ion interactions.

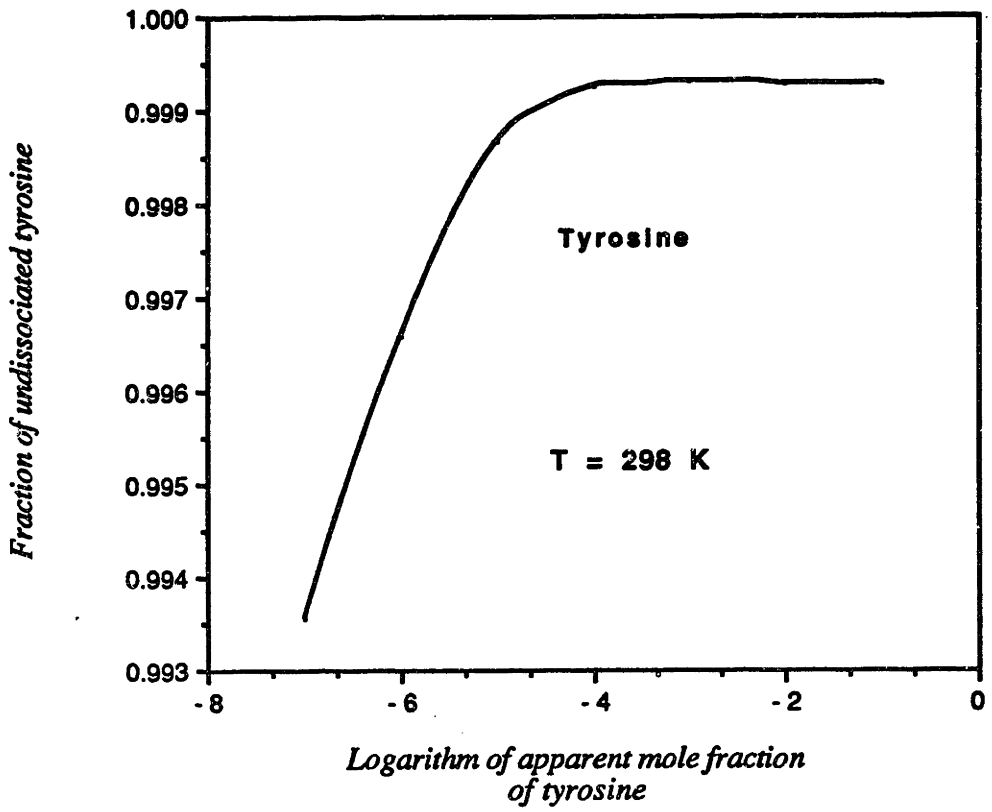
As stated previously, the neutral dipolar species are the predominant species in the isoelectric solution of amino acids. The amino acid cation formation reaction and the amino acid anion formation reaction are insignificant for amino acids in pure water.

For example, in pure water, 99.9% of tyrosine remains as neutral dipolar species. Figure 5-1 shows the computed percentage of tyrosine existing as dipolar species as a function of tyrosine concentration in water. Table 5-1 summarizes the values of the chemical equilibrium constants used in the solution chemistry calculations. Note that in addition to the three species described above, tyrosine generates a double-charged anion by releasing a hydrogen ion from the single-charged tyrosine anion:

$$A^{-} \quad <--> \quad A^{=} \quad + \quad H^{+} \tag{5-13}$$

The corresponding thermodynamic relationship is given below.

Figure 5-1. Dissociation of Tyrosine in Water



$$K_{A-} = a_{A-} \quad a_{H+} \quad / \, a_{A-} \tag{5-14}$$

Table 5-1. Chemical Equilibrium Constants* for Tyrosine in Water (Greenstein,1961)

$\ln K_{A+}$	= -5.066
$\ln K_{A-}$	= -20.98
$\ln K_{A-}$	= -23.18

* molality scale

5.3 The Electrolyte NRTL Model

The Electrolyte NRTL model was originally proposed by Chen, et al. (Chen, et al., 1982) and Chen and Evans (1986) as an excess Gibbs energy expression for aqueous electrolytes. It was later extended for mixed-solvent electrolytes (Scauflaire, et al., 1989; Mock et al., 1986). The model has proved to be very successful in representing thermodynamic properties of various aqueous and mixed-solvent electrolyte systems. The study represents the first attempt to apply the model to systems containing zwitterions.

The model assumes that there are two contributions for the excess Gibbs energy of electrolyte systems. The first contribution is the result of the long-range ion-ion interactions and the second contribution is the result of the local interactions between ion and ion, between molecule and molecule, and between ion and molecule. The long-range interaction contribution is accounted for by the combination of the Pitzer-Debye-Huckel

equation (Pitzer, 1980) and the Born equation (Lewis and Randall, 1961). The local interaction contribution is represented by a modified form of the Non-random Two Liquid (NRTL) equation (Renon and Prausnitz, 1968). Note that the Born term is associated with the effect of transferring charged species from pure water to a mixed solvent. In the absence of nonaqueous solvents, the Born term reduces to zero.

$$g^{ex^*}/RT = g^{ex^*,NRTL}/RT + g^{ex^*,PDH}/RT + g^{Born}/RT \tag{5-15}$$

$$\frac{g^{ex*,NRTL}}{RT} = \sum_m X_m \frac{\sum_j X_j G_{jm} \tau_{jm}}{\sum_k X_k G_{km}} + \sum_c X_c \sum_{a'} \frac{X_{a'}}{\sum_{a''} X_{a''}} \frac{\sum_j X_j G_{jc,a'c} \tau_{jc,a'c}}{\sum_k X_k G_{kc,a'c}} + \sum_a X_a \sum_{c'} \frac{X_{c'}}{\sum_{c''} X_{c''}} \frac{\sum_j X_j G_{ja,c'a} \tau_{ja,c'a}}{\sum_k X_k G_{ka,c'a}} \tag{5-16}$$

$$\frac{g^{ex*,PDH}}{RT} = -\left(\sum_k x_k\right)\left(\frac{1000}{M_s}\right)^{1/2}\frac{\left(4A_\phi I_x\right)}{\rho}\ln\left(1+\rho I_x^{1/2}\right) \tag{5-17}$$

$$\frac{g^{Born}}{RT} = \frac{e^2}{2kT}\left(\frac{1}{D} - \frac{1}{D_w}\right)\left(\sum_i \frac{x_i Z_i^2}{r_i}\right) 10^{-2} \qquad (5\text{-}18)$$

The reference state for solvent molecules is the pure liquid at system temperature and pressure. The reference state for ionic species is the infinite dilution state in aqueous phase. In this work, the reference state for neutral dipolar species, or zwitterions, has also been chosen to be the infinite dilution state in aqueous phase. These reference states are used regardless of whether the systems are aqueous systems or mixed-solvent systems.

The mole fraction scale unsymmetric activity coefficient expression for ionic species can be derived from Equation (5-15).

$$\ln \gamma_i^* = \ln \gamma_i^{*,NRTL} + \ln \gamma_i^{*,Born} + \ln \gamma_i^{*,PDH} \qquad (5\text{-}19)$$

The specific expressions for the three terms in Equation (5-19) have been given in the literature (Chen and Evans, 1986; Scauflaire, et al., 1989).

Similarly, the mole fraction scale unsymmetric activity coefficient expression for neutral dipolar species can be derived.

$$\ln \gamma_{A+-}^* = \ln \gamma_{A+-}^{*,NRTL} + \ln \gamma_{A+-}^{*,Born} \qquad (5\text{-}20)$$

Note that the Pitzer-Debye-Huckel term computes solution ionic strength from the net charge of the species while the Born term is computed with the total distinct charges of the species. Therefore, the Pitzer-Debye-Huckel term drops out since the dipolar species is neutral with a net charge of zero. However, the Born term remains since the neutral dipolar species exist in the liquid phase with distinct dual charges (one positive charge and one negative charge) residing on the species.

The mole fraction scale activity coefficient expression for molecular solvent species can also be derived.

$$\ln \gamma_m = \ln \gamma_m^{NRTL} \tag{5-21}$$

Note that the molality scale activity coefficients can be computed from the mole fraction scale activity coefficients.

There are two types of binary interaction parameters with the Electrolyte NRTL model: nonrandomness factor and energy parameters. These binary parameters are associated with binary molecule-molecule pairs, binary molecule-electrolyte pairs, and binary electrolyte-electrolyte pairs. Being neutral, molecular solvent species and dipolar species are considered as "molecules" while cation-anion pairs are treated as "electrolytes," in the context of this Chapter.

The nonrandomness factor parameters are symmetric. Following the convention

of Chen and Evans (1986), in this work, the nonrandomness factor was fixed at 0.2 for binary electrolyte-molecule pairs and binary electrolyte-electrolyte pairs. It was set at 0.3 for binary molecule-molecule pairs. The energy parameters were used as adjustable parameters in correlating literature data.

The energy parameters are asymmetric. In other words, there are two energy parameters associated with each pair, including molecule-molecule pairs, molecule-electrolyte pairs, and electrolyte-electrolyte pairs. For the biomolecular systems, concentrations of the biomolecules are often relatively small in comparison to that of solvents. In such cases, one energy parameter per pair alone is found to be sufficient in representing the literature data. Nevertheless, as a generalized model, the Electrolyte NRTL model provides two energy parameters per binary pair as adjustable parameters in correlating the solubility data.

The proposed molecular thermodynamic framework is general and it offers important advantages over previous empirical methods in representing the solubility behavior of the biomolecules. The framework requires only binary parameters in computing activity coefficients of the true species in the biomolecular systems. The binary parameters identified for one binary biomolecular system are applicable to multicomponent biomolecular systems containing the same binary biomolecular subsystem. Furthermore, many of the binary parameters are already available in the literature for molecule-molecule pairs, molecule-electrolyte pairs, and electrolyte-electrolyte pairs. These parameters can be readily applied in the investigation

of the biomolecular systems.

5.4 Activity Coefficients of Amino Acids

Fasman (1976) has compiled the molality scale unsymmetric activity coefficient data of amino acids in pure water. Bonner (1982) also has obtained additional activity coefficient data for lysine and arginine from isopiestic equilibrations. These data have been successfully correlated in this study.

As discussed earlier, the dissociation extents of amino acids in unbuffered, pure water are relatively insignificant. In other words, neutral dipolar species are the predominant amino acid species while the concentrations of ionic amino acid species are negligible. Therefore, in representing the activity coefficient data, it is not necessary to explicitly address the solution chemistry. In other words, the trace ionic amino acid species need not be taken into account. These trace species have little impact upon the activity coefficients of neutral, dipolar amino acid species.

In the absence of nonaqueous solvent, the Born term is to be dropped from Equation (5-20) for the activity coefficient of dipolar species.

$$\ln \gamma_{A\pm}^{*} = \ln \gamma_{A\pm}^{*,NRTL} \tag{5-22}$$

Figure 5-2 shows the representation of the activity coefficient data with the model. The data were correlated satisfactorily with only one or two adjustable energy parameters for water-amino acid (zwitterion) pairs. It is interesting to note that the amino acid

activity coefficients are mostly linear functions of amino acid concentrations, similar to that of nonelectrolytes.

Figure 5-2. Activity Coefficient of Amino Acids in Water

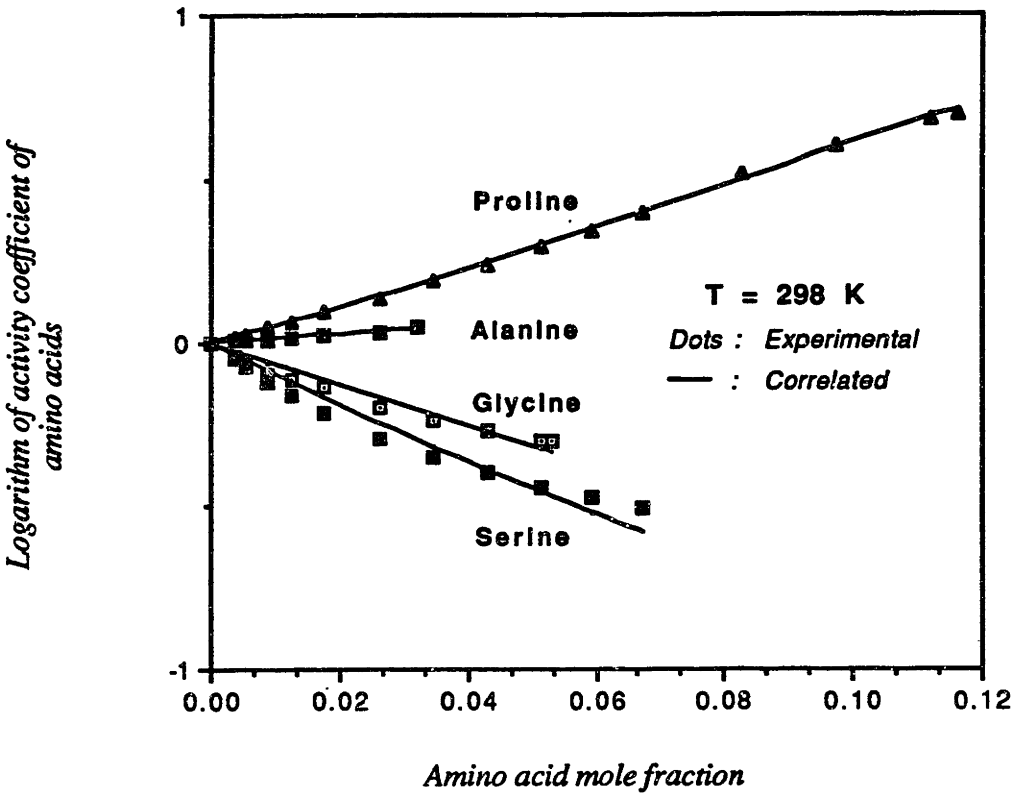


Table 5-2. summarizes the values of the energy parameters.

Table 5-2. Electrolyte NRTL Model Parameters for Water(1)-Amino Acid(2) Pairs

Amino Acid	m_{max}	τ_{12}	τ_{21}	r.m.s. error
alanine	1.5	-2.352	0*	0.0004
α-aminobutyric acid	2.0	4.627	-3.117	0.0032
arginine	1.6	5.806	-3.239	0.0056
glycine	3.0	1.719	0*	0.0207
l-hydroxyproline	2.0	-1.108	0*	0.0036
lysine	5.5	6.883	-3.979	0.0748
methionine	0.3	3.755	0*	0.0086
proline	4.0	4.138	-2.942	0.0130
serine	4.0	2.404	0*	0.0280
threonine	2.0	0.424	0*	0.0070
valine	0.5	-3.347	0*	0.0447
alanylalanine	1.0	-1.161	0*	0.0204
alanylglycine	1.0	2.517	0*	0.0292
diglycine	1.5	3.278	0*	0.0325
glycylalanine	1.0	2.451	0*	0.0244
triglycine	0.3	4.880	0*	0.0067

* Parameters fixed

5.5 Solubilities of Amino Acids in Water

The solubilities of amino acids in pure water generally increase with temperature. The solubility data can be represented by regressing the chemical equilibrium constant, or thermodynamic solubility constant, of Equation (5-8). In pure water, the formations of amino acid cations and amino acid anions have little impact on the amino acid solubilities and they can be ignored.

In this work, the solubility data of many amino acids from Fasman (1976) have been correlated with the solubility constant coefficients of Equation (5-23) adjusted. The data cover temperature ranges from 273.15 K to 373.15 K. Table 5-3 summarizes the values of the thermodynamic solubility constants. It was assumed that the solid forms of the amino acids were anhydrous. For the amino acids that are tabulated in Table 5-2, the amino acid activity coefficients in pure water were computed with the Electrolyte NRTL model energy parameters reported in Table 5-2. For the amino acids that are not tabulated in Table 5-2, the energy parameters were set to zero.

$$\ln K_s = a + b/T + c \ln T \tag{5-23}$$

Figures 5-3a and 5-3b show the correlation of the solubility data with the framework.

Figure 5-3a. Amino Acid Solubillity in Water

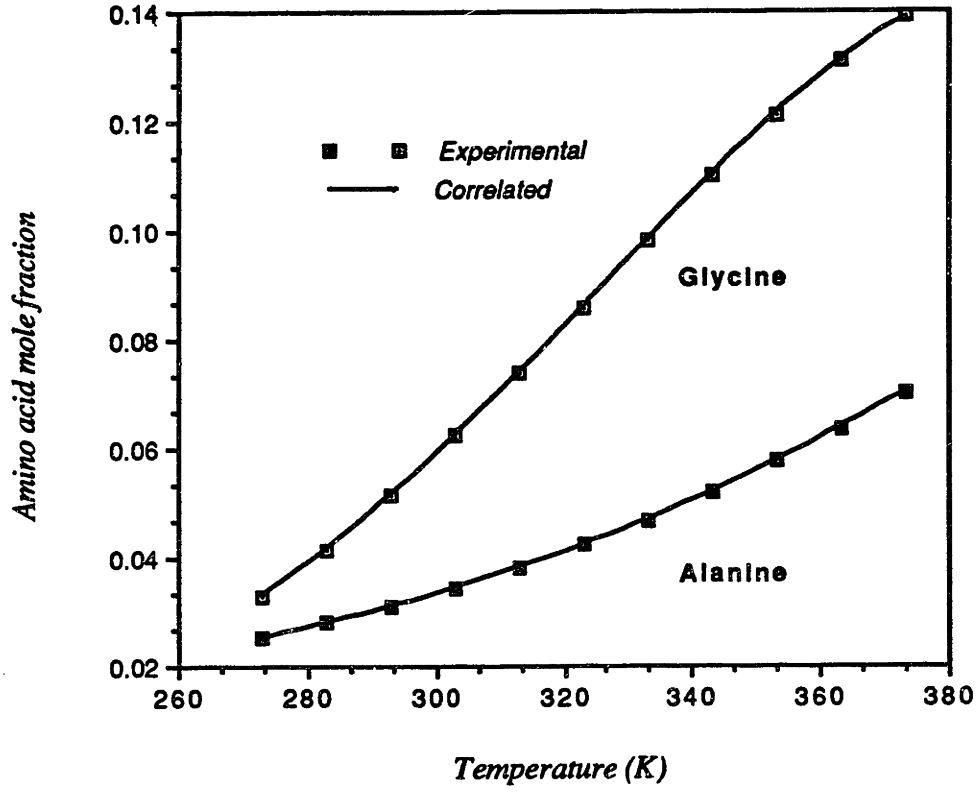


Figure 5-3b. Amino Acid Solubility in Water

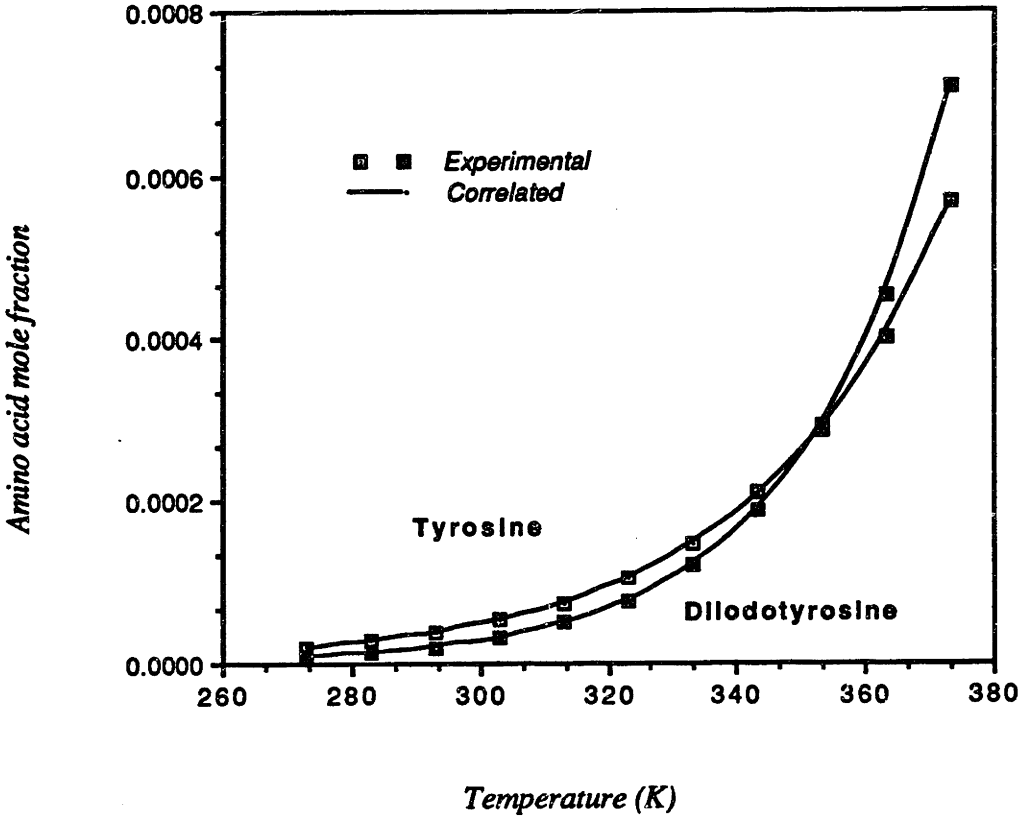


Table 5-3. Solubility Constants of Amino Acids in Water

	a	b	c	r.m.s. rel. error
l-alanine*	-137.8478	4827.688	20.8295	0.0067
dl-alanine*	-65.7263	1763.498	9.91053	0.0019
l-aspartic acid**	-142.2452	3324.175	21.72465	0.0550
dl-aspartic acid**	43.5974	-5438.725	-5.65442	0.0298
l-cystine**	-141.9555	3453.390	20.82660	0.0051
l-diiodotyrosine**	-199.7653	5136.637	30.18313	0.0155
glycine*	81.2501	-4796.998	-11.99306	0.0014
l-glutamic acid**	-88.5340	884.618	13.81428	0.0117
dl-glutamic acid**	-139.9315	3317.609	21.55557	0.0013
l-hydroxyproline*	-41.2903	1012.144	6.12857	0.0007
l-isoleucine**	-163.5717	6671.286	23.83987	0.0107
dl-isoleucine**	-194.0435	7637.427	28.54204	0.0120
l-methionine*	-19.9356	-375.778	2.81495	0.0011
dl-methionine*	28.4148	-3239.851	-4.06103	0.0202
dl-norleucine**	-199.2678	7532.739	29.40484	0.0125
l-phenylalanine**	-70.1121	1657.000	10.32303	0.0015
dl-phenylalanine**	-195.2954	7253.531	28.86780	0.0124
l-proline*	-50.6762	942.310	8.30864	0.0014
l-serine*	430.2341	-20413.49	-64.07024	0.0508
dl-serine*	101.4502	-6963.363	-14.55624	0.0285
taurine**	207.0671	-12087.37	-29.96454	0.0340
tyrosine**	-155.2931	3967.347	23.16516	0.0129
dl-tyrosine**	-173.2559	4828.088	25.78009	0.0974
l-tryptophan**	-185.0812	6860.580	27.23113	0.0116
l-valine*	-116.45586	4554.057	16.93207	0.0050

* See Table 5-2 for Electrolyte NRTL model energy parameters

** Electrolyte NRTL model energy parameters set to zero

5.6 Solubilities of Amino Acids in the Presence of Salts

The presence of salts in the liquid phase could either have a salting-in effect or a salting-out effect on the solubilities of amino acids. Generally, salts dissociate completely to ions in the liquid phase. It is the ion-molecule (dipolar amino acid species) physical interactions that result in the salting effects.

Cohn and Edsall's data (1943) on the solubilities of glycine and dl-α-aminocaproic acid in the presence of lithium chloride and the solubilities of l-asparagine in the presence of lithium chloride and sodium chloride have been represented satisfactorily with the framework. The salt-amino acid energy parameters of the Electrolyte NRTL model were adjusted to represent the ion-molecule physical interactions in the liquid phase.

Figure 5-4 shows the comparison of the literature data and the correlation results with the model. The l-asparagine (anhydrous) data show a strong salting-in effect of salts. The glycine data and dl-α-aminocaproic acid data demonstrate relatively weak salting effects. Table 5-4 summarizes the regressed values of the energy parameters determined in this work. For strong salting-in effects, the absolute values of the salt-amino acid energy parameters were found to be relatively large. Note that the water-amino acid pair energy parameters have been set to zero for l-asparagine and dl-α-aminocaproic acid. The water-glycine pair energy parameters of Table 5-2 have been used for water-glycine interactions. The energy parameters for electrolyte (salt)-water pairs were obtained from Chen and Evans (1986).

Figure 5-4. Salting-In Effect of Sodium Chloride and Lithium Chloride on Amino Acid Solubility

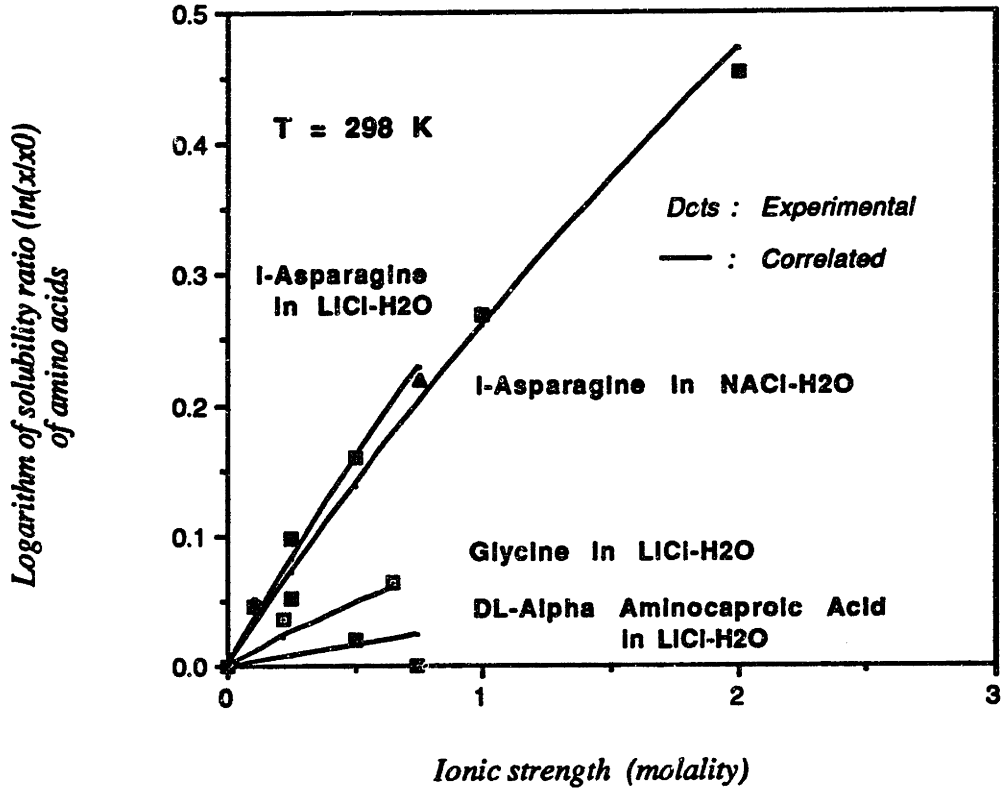


Table 5-4. Electrolyte NRTL Model Parameters for Salt(1)-Amino Acid(2) Pairs

Salt	Amino Acid	τ_{12}	τ_{21}	r.m.s. rel. error
LiCl	dl-α-aminocaproic acid	-4.500	8.045*	0.0254
NaCl	l-asparagine	-5.728	8.045*	0.0150
LiCl	l-asparagine	-5.930	8.045*	0.0078
LiCl	glycine	-3.032	8.045*	0.0036

K_s (dl-α-aminocaproic acid) = -6.489045
K_s (l-asparagine) = -5.695815
K_s (glycine) = -3.175186

* Parameters fixed

5.7 Solubilities of Amino Acids in the Presence of Dipolar Amino Acid Species

Similar to the salting effects, the presence of other dipolar amino acid species in the liquid phase may either increase or decrease the solubilities of amino acids. It is the molecule-molecule physical interactions between the dipolar amino acid species that result in the change in the solubilities.

The data of Cohn and Edsall (1943) for the solubilities of l-asparagine and cystine in the presence of glycine, diglycine, alanine, and α-aminobutyric acid were represented successfully with the framework. The energy parameters for the amino acid-amino acid pairs were adjusted to account for the physical interactions. Table 5-5 summarizes the values of the energy parameters determined in this work. For strong dipolar species effects, the absolute values of the amino acid-amino acid energy parameters were found to be relatively large. Figures 5-5a and 5-5b show the comparison of the literature data

and the correlation results with the model.

Table 5-5. Electrolyte NRTL Model Parameters for Amino Acid-Amino Acid Pairs

Amino Acid(1)	Amino Acid(2)	τ_{12}	τ_{21}	r.m.s. rel. error
alanine	l-asparagine	-2.651	0*	0.0118
α-aminobutyric acid	l-asparagine	0*	1.123	0.0049
diglycine	l-asparagine	-2.233	0*	0.0134
glycine	l-asparagine	-2.620	0*	0.0165
alanine	cystine	-3.783	0*	0.0270
α-aminobutyric acid	cystine	0*	0.698	0.0258
diglycine	cystine	-4.356	0*	0.0554
glycine	cystine	-3.979	0*	0.0644

K_s (l-asparagine) = -5.695815
K_s (cystine) = -11.71138

* Parameters fixed

Figure 5-5a. Dipolar Ion Effect on Cystine Solubility

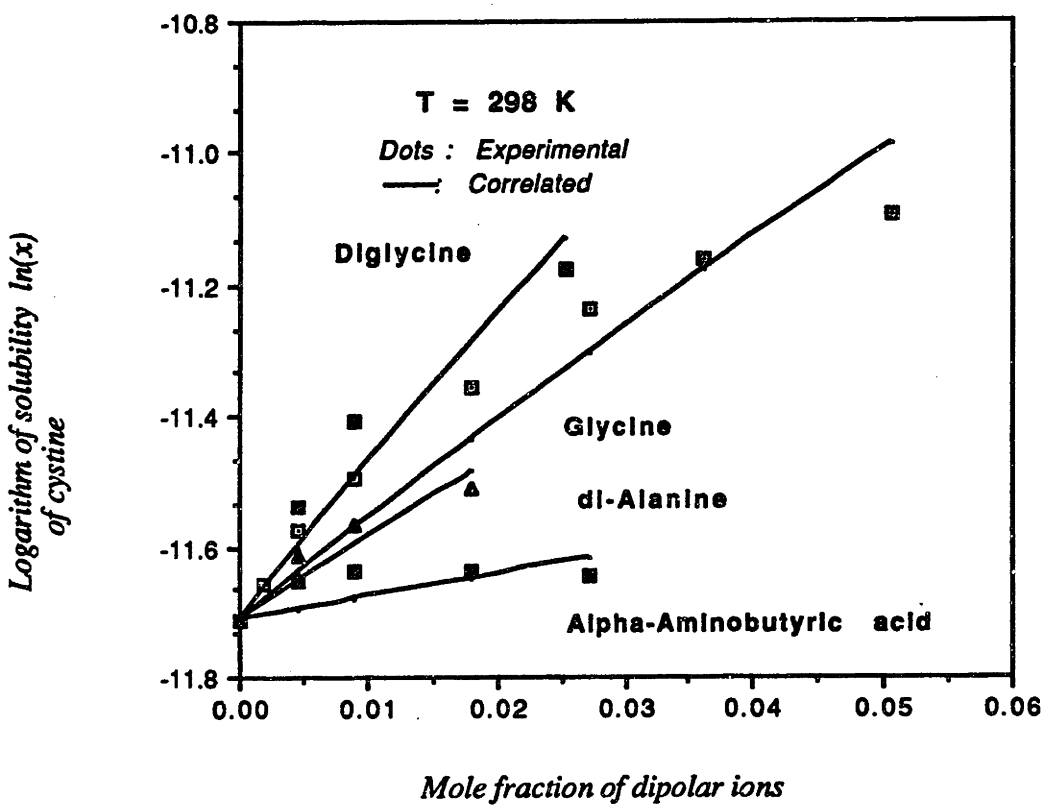
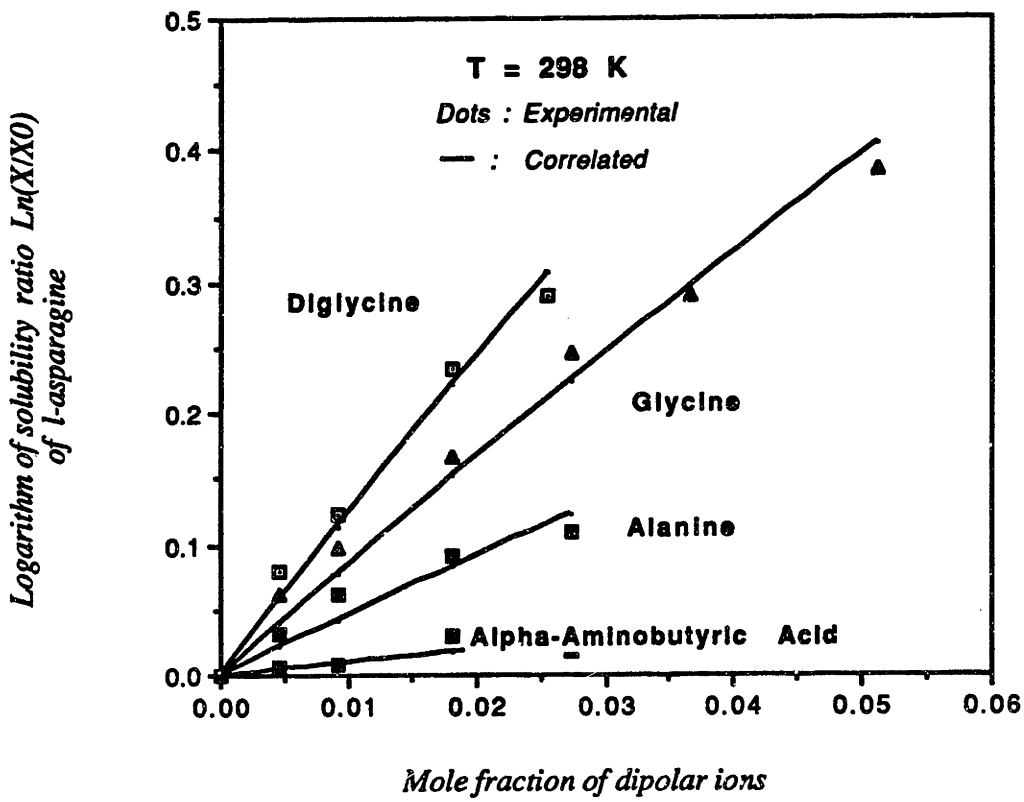


Figure 5-5b. Dipolar Ion Effect on l-Asparagine Solubility



5.8 Solubilities of Amino Acids in Alcohol-Water Mixtures

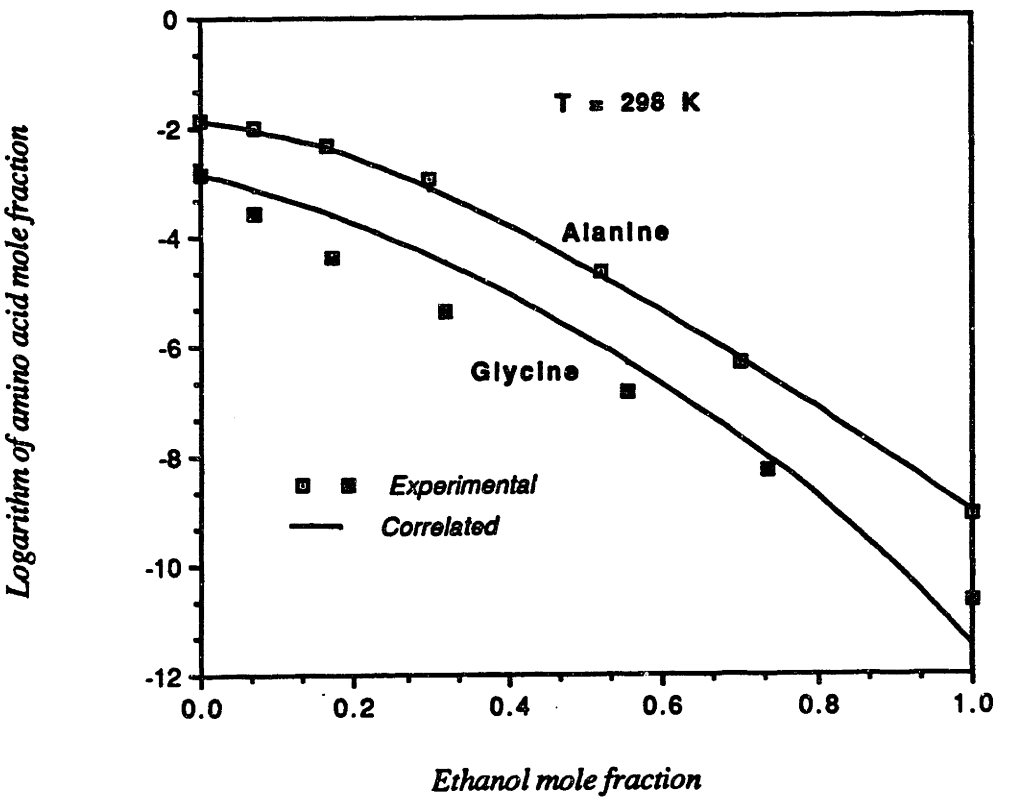
The change in solvent compositions has a profound effect on the solubilities of amino acids. This effect can be studied by examining the Born term of Equation (5-18), which accounts for the energy of transfer of charge-carrying species from water to the mixed solvent. When the alcohol content of the mixed-solvent increases, the dielectric constant of the mixed-solvent decreases. As a result, the Born term makes a significant, positive contribution to the logarithm of activity coefficient of dipolar amino acid species. The sharp increase causes a steep drop in the solubilities of amino acids in alcohol-water mixtures.

The data of Greenstein and Winitz (1961) for the solubilities of glycine and β-alanine in ethanol-water mixtures have been represented satisfactorily in this work. The energy parameters for the ethanol-water pair were taken from Gmehling and Onken (1977). The dielectric constant of ethanol was set to 24.55 at 298.15 K. Born term radii of the dipolar amino acid species were adjusted to fit the data. Figure 5-6 shows the comparison of the literature data and the correlation results with the model. Table 5-6 summarizes the regressed values of the Born term radii.

Table 5-6. Born Term Radii of Amino Acids

	Born Term Radii	r.m.s. rel. error
β-alanine	0.2179 nm	0.0797
glycine	0.1564 nm	0.8141

Figure 5-6. Solvent Effect of Ethanol on Amino Acid Solubility



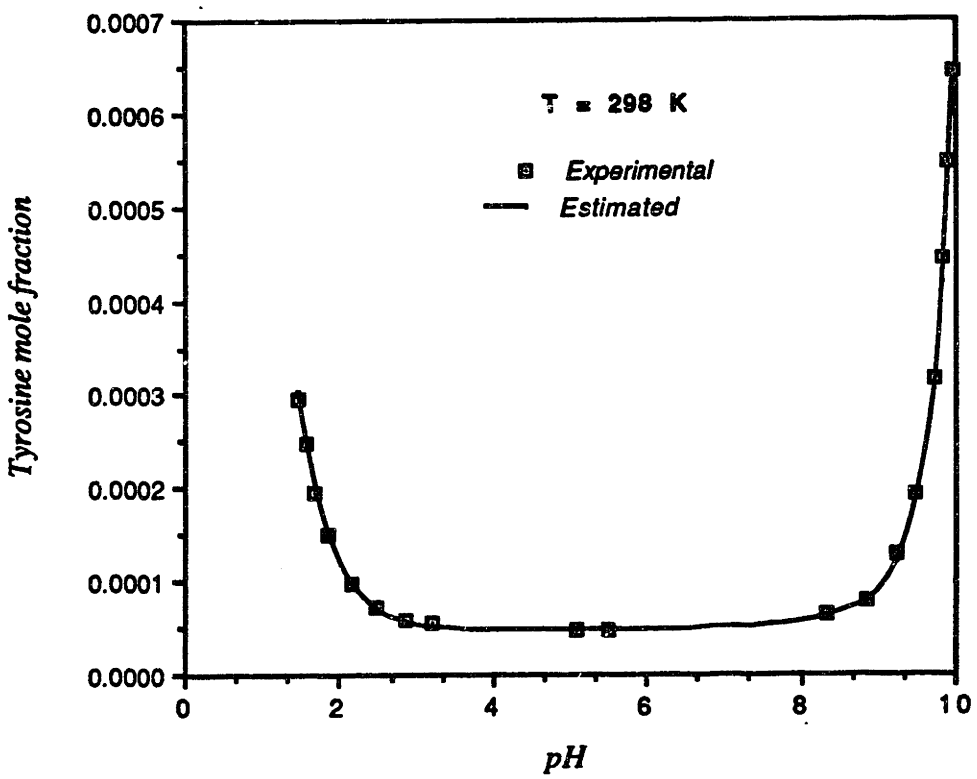
Note that the regressed Born term radii are small in comparison to radii of the amino acids. Each amino acid is carrying two charges. The Born term radius is the radius of each of the two charge-carrying nucleus.

5.9 Solubilities of Amino Acids as a Function of pH

The effect of pH on the solubilities of amino acids is best explained by the solution chemistry of amino acids. The change in pH makes dramatic shifts in the distribution of amino acid species in the liquid phase. In isoelectric solutions, i.e., pH is equal to or near pI, the neutral dipolar amino acid species are the predominant species. At pH much greater than pI, anionic amino acid species become predominant, while cationic amino acid species become predominant at pH much less than pI. Because the thermodynamic solubility constant is only related to the activity of the neutral dipolar amino acid species (see Equation 5-8), the solubilities of amino acids are the lowest at pI and the solubilities increase when pH is moved away from pI.

Figure 5-7 shows the prediction of the tyrosine solubility as a function of pH versus the experimental data of Hitchcock (1924). By explicitly accounting for the solution chemistry, the tyrosine solubility can be predicted satisfactorily without any parameters adjusted.

Figure 5-7. pH Effect on Tyrosine Solubility



Notation

A	=	Amino acid or small peptide
A^{+-}	=	Dipolar ion or zwitterion
A^{+}	=	Amino acid cation
A^{-}	=	Amino acid anion
A_{ϕ}	=	Debye-Huckel constant for the osmotic coefficient
D	=	Solvent dielectric constant
G	=	Boltzmann factor in Equation (5-16)
I	=	Ionic strength in molality scale
I_x	=	Ionic strength in mole fraction scale
K	=	Thermodynamic chemical equilibrium constant
K_s	=	Thermodynamic solubility constant
K'	=	Empirical correlation constant in Equation (5-1)
K"	=	Empirical correlation constant in Equation (5-3)
M_s	=	Solvent molecular weight, kg/kmol
R	=	Gas constant
S	=	Solubility
T	=	Temperature, K
X	=	Concentration variable in Equation (5-16)
Z	=	Absolute value of ionic charge
a	=	Activity
e	=	Electric charge, 1.602189*10-19C
g^{ex}	=	Molar excess Gibbs energy
k	=	Boltzmann constant, 1.380662*10-23J.K-1
m	=	Molality, g-mole/kg of solvent
r	=	Born term ionic radius
s	=	Solid
x	=	True liquid phase mole fraction based on all species: molecular and ionic

Greek Letters

Ω	=	Contact area between the protein molecules
β	=	Empirical constant in Equation (5-1)

γ	=	Activity coefficient, mole fraction scale
λ	=	Repulsive interactions between molecules with like charges
ρ	=	The closest approach parameter of the Pitzer-Debye-Huckel equation
σ	=	Induced dipoles
τ	=	NRTL binary interaction energy parameter

Superscripts

*	=	Unsymmetric convention
Born	=	Long range contribution, represented by the Born equation
PDH	=	Long range contribution, represented by the Pitzer-Debye-Huckel equation
NRTL	=	Local contribution, represented by the Non-Random Two Liquid equation

Subscripts

a, a', a"	=	Anion
c, c', c"	=	Cation
i	=	Ionic species
j, k	=	Any species
m	=	Molecular species
s	=	Solvent species
w	=	Water

Literature Cited

Albertsson, P.A., Partition of Cell Particles and Macromolecules, Wiley, New York (1985).

Abbott , N.L. and T.A. Hatton, "Liquid-Liquid Extraction for Protein Separation," Chemical Engineering Progress, 31-41, August 1988.

Bell, D.J., M. Hoare, and P. Dunnill, "The Formation of Protein Precipitates and Their Centrifugal Recovery," Advances in Biochemical Engineering/Biotechnology, 26, 1-72 (1983).

Bonner, O.D., "Osmotic and Activity Coefficients of Some Amino Acids and Their Hydrochloride Salts at 298.15 K," J. Chem. Eng. Data, 27, 422-423 (1982).

Chen, C.-C., H.I. Britt, J.F. Boston, and L.B. Evans, "Local Composition Model for Excess Gibbs Energy of Electrolyte Systems," AIChE J., 25, 820 (1982).

Chen, C.-C. and L.B. Evans, "A Local Composition Model for the Excess Gibbs Energy of Aqueous Electrolyte Systems," AIChE J., 32, 444 (1986).

Chen, C.-C., Y. Zhu, and L.B. Evans, Biotechnology Progress, 5, (3), 111-118 (1989).

Cohn, E.J. and J.T. Edsall, Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Reinhold Publishing Corporation, New York, 1943.

Dalton, J.B., P.L. Kirk, and C.L.A. Schmidt, "The Apparent Dissociation Constants of Diiodotyrosine, Its Heat of Solution, and Its Apparent Heat of Ionization," J. Biol. Chem., 88, 589 (1930).

Fasman, G.D., ed., Handbook of Biochemistry and Molecular Biology, 3rd ed., Vol. I, Page 115,120, CRC Press, Cleveland, OH (1976).

Gmehling, J. and U. Onken, Vapor-Liquid Equilibrium Data Collection, Aqueous-Organic Systems; Chemistry Data Series, Vol. 1, Part 1, DECHEMA,

Frankfurt/Main, 1977.

Greenberg, D.M., Amino Acids and Proteins, Theory, Methods, Applications., Charles C. Thomas Publisher, Springfield, Illinois, 1951.

Greenstein, J.P. and M. Winitz, Chemistry of the Amino Acids, Vol. 1, Pages 486-491, 547, John Wiley & Sons, 1961.

Hitchcock, D.I., "The Solubility of Tyrosine in Acid and in Alkali," J. Gen. Physiol., 6, 747 (1924).

Lewis, G.N. and M. Randall, Thermodynamics, revised by K.S. Pitzer and L. Brewer, 2nd ed., McGraw-Hill, New York (1961).

Melander, W. and C. Horvath, "Salt Effects on Hydrophobic Interactions in Precipitation and Chromatography of Proteins: An Interpretation of the Lyotropic Series," Arch. of Biochem. and Biophy., 183, 200-215(1977).

Mock, B., L.B. Evans, and C.-C. Chen, "Thermodynamic Representation of Phase Equilibria in Multiple Solvent Electrolyte Systems," AIChE J., 32, 1655 (1986).

Nass, K.K.,"Representation of Solubility Behavior of Amino Acids in Water," AIChE J., 34, 1257 (1988).

Pitzer, K.S., "Electrolytes: From Dilute Solutions to Fused Salts," J. of Am. Chem. Soc., 102:9, 2902 (1980).

Renon, H. and J.M. Prausnitz, "Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures," AIChE J., 14, 135 (1968).

Scauflaire, P., D. Richards, and C.-C. Chen, "Ionic Activity Coefficients of Mixed-Solvent Electrolyte Systems," paper submitted for publication in AIChE J. (1989).

Wilson, G., "Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Gibbs Energy of Mixing," J. Am. Chem. Soc., 86, 127 (1964).

CHAPTER 6*

PHASE EQUILIBRIUM BEHAVIOR OF ANTIBIOTICS

* Chapter 6 has been published in *Biotechnology Progress,* 6, 266-272 (1990), with Lawrence B. Evans and Chau-Chyun Chen.

Introduction

Antibiotics are antimicrobial compounds produced by living organisms. A number of antibiotics have achieved immense commercial success for the pharmaceutical industry, with penicillin being the most widely known example. This commercial success, and the well established markets for antibiotics in human and veterinary medicine, the food industry, in agriculture, and in raising livestock has prompted a great deal of laboratory work. This work has lead to dramatic growth in the number of antibiotics identified (Berdy, 1978; Smith, 1988). However, only a small fraction of those identified have been commercialized primarily due to prohibitive production costs. One major stumbling block is in the area of recovery and separation of these compounds. The investigation of phase equilibrium behavior of antibiotics is important both in understanding the partition mechanisms and in the design and optimization of downstream recovery processes (Strong, 1986; Evans, 1988). Although there are empirical equations used for correlating experimental data (Tsuji, et al., 1977; Tsuji, et al., 1978; Tsuji, et al., 1979; and Bogardus and Palepu, 1979), comprehensive and predictive models for antibiotic solutions that can be used to scale up and optimize manufacturing processes do not exist. Development of

a predictive model for representing the phase equilibrium behavior of antibiotics has been difficult due to the lack of experimental data and (even more importantly) the lack of a sound theoretical approach. This chapter presents recent results in applying and extending the molecular thermodynamic framework in Chapter 5 for amino acids to represent the liquid-solid equilibrium behavior (solubilities) and the liquid-liquid equilibrium behavior (partitioning coefficients) of antibiotics as functions of temperature, ionic strength, solvent compositions, and pH.

β-lactam antibiotics are amino acid derivatives. They were chosen as the model compounds in this work since this class of antibiotics includes the most important chemotherapeutic agents. The almost complete lack of toxicity and wide range of antibacterial activity, together with excellent pharmaceutical properties, make these natural and semisynthetic antibiotics "wonder drugs" (Tomlinson and Regosz, 1985). Furthermore, phase equilibrium data for this class of antibiotics have been accumulating ever since Fleming's discovery of penicillin in 1928. These data are essential for the application of the molecular thermodynamic framework.

6.1 Structures of β-Lactam Antibiotics

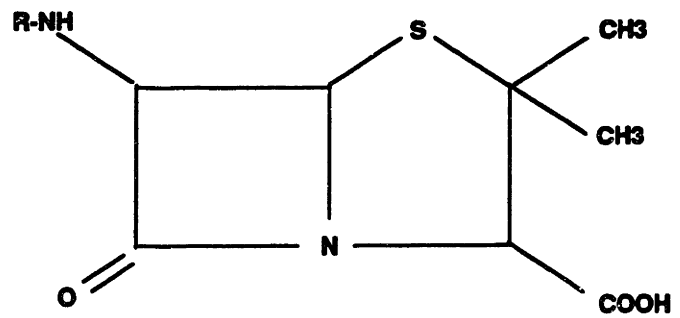
The β-lactam antibiotics comprise two major groups of therapeutic agents of considerable clinical importance - the penicillins and the cephalosporins. This category of substances contain a 4-membered lactam as their name implies. The β-lactam is fused through the nitrogen atom and the adjacent tetrahedral carbon to a second heterocycle - a 5-membered thiazolidine for the penicillins, or a 6-membered dihydrothiazine for the

cephalosporins. These antibiotics carry a variety of substituents which contribute to their different biological activities and physico-chemical properties. There are two structural features common to virtually all of the β-lactam antibiotics. The first is the carboxyl group on the carbon adjacent to the lactam nitrogen, and the second is the functionized amino group on the carbon atom opposite the nitrogen of the β-lactam (Hoover and Nash, 1978).

As shown in Figures 6-1a and 6-1b, the penicillins possess a 6-aminopenicillanic acid nucleus, and the cephalosporins have a 7-aminocephalosporanic acid nucleus. The two amino acids, 6-aminopenicillianic acid and 7-aminocephalosporanic acid, are zwitterions (i.e., dipolar ions). A large number of antibiotics derived from these two amino acids are also zwitterions. Some typical examples of these compounds are shown in Figures 6-1a and 6-1b. These amphoteric antibiotics can exist in solution as molecules carrying nuclei with both positive and negative charges, i.e., as zwitterions.

Antibiotic zwitterions share the same characteristics as amino acids in aqueous solutions. They carry charges like ions, and side chains and the two ring frame like other organic compounds. At their pI's, they do not conduct electric current. However, internally ionized, they are charged species with the net apparent charge being zero. The two dissociation constants K_1 and K_2 correspond mainly to the 2-carbon free carboxyl group and the 9-carbon (for penicillins) or 10-carbon (for cephalosporins) free amino group, respectively.

Figure 6-1a. The Structure and Examples of Penicillins



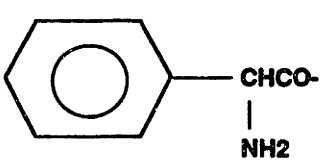
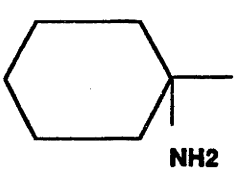
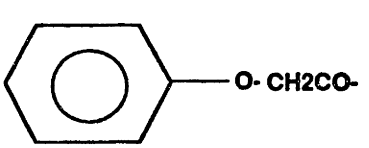
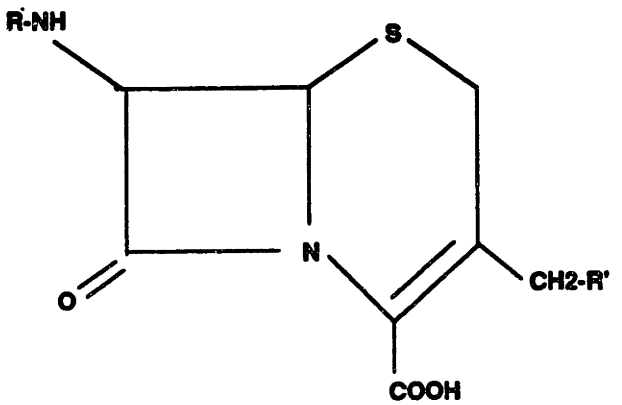
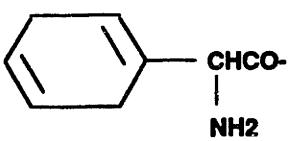
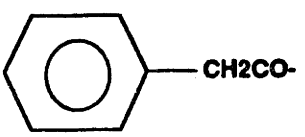
Penicillins	Side Chain R
6-Aminopenicillanic Acid	H -
Ampicillin	
Cyclacillin	
Penicillin V	

Figure 6-1b. The Structure and Examples of Cephalosporins



Cephalosporins	Side Chain R	Side Chain R'
7-Amino Cephalosporanic Acid	H-	-OCOCH3
Cephradine		-H
7-phenylacetamidodeacetoxy--cephalosporanic Acid		-H
7-aminodeacetoxycephalosporanic Acid	H-	-H

It should be noted that not all of the β-lactam antibiotics possess the feature of zwitterions. Certain β-lactam imino acid antibiotics such as penicillins G and V, and 7-phenylacetamideacetoxycephaloaporanic acid (7-PDA) do not possess a free amino group on their substituted side chain R. The basicity of the imino groups in these molecules is relatively weak compared with the acidity of their carboxyl groups. As such, these antibiotics behave more like "acids" than "zwitterions."

It becomes obvious that the phase equilibrium behavior of β-lactam antibiotics can be treated the same way as we did for amino acids in Chapter 5. In other words, the key factors characterizing the phase equilibrium behavior of the β-lactam antibiotics are the ampholyte solution chemistry and the nonideality of the true species, including zwitterions, in the solution.

6.2 Previous Correlation Methods

Using an empirical equation developed from the ampholyte solution chemistry, Tsuji, et al. (1977) correlated their apparent partitioning coefficient data of several β-lactam antibiotics between octanol and water:

$$P_{app}\ (a_{H+}/K_a + 1) = P_u\ a_{H+}/K_a + P_i \qquad (6\text{-}1)$$

where P_{app}, P_u and P_i are the partitioning coefficients for the apparent, undissociated, and ionized forms of the antibiotics, respectively. a_{H+} is the hydrogen-ion activity, and K_a is the dissociation constant. Using equation (6-1), they correlated the intrinsic partitioning

coefficients, P_u and P_i, for the true species present in the solutions. They also reported empirical equations correlating the intrinsic partitioning coefficients.

Tsuji, et al. (1978) established U-shaped curves for the pH dependency of the solubilities of ampicillin and other penicillins at a constant ionic strength (I_m=0.5) at 310.15 K. They also represented the solubility-pH curves using an expression derived from the solution chemistry of the penicillins:

$$C_T = C_0 (a_{H+}/K_1 + 1 + K_2/a_{H+}) \qquad (6\text{-}2)$$

where C_T is the total solubility, C_0 is the intrinsic solubility of amphoteric cephalosporins with the electrically neutral zwitterions, a_{H+} is the hydrogen-ion activity of the solution, and K_1 and K_2 are the dissociation constants for the carboxylic acid and the conjugated acid of the α-amino acid group, respectively. Tsuji, et al. (1979) also used equation (6-2) to predict the solubilities of several aminocephalosporins as a function of pH. Their calculated values agree well with their experimental data with the exception of some aminocephalosporins at lower pH.

Bogardus and Palepu (1979) gave a similar formula for the pH influence upon the total solubility of β-lactam antibiotics. An additional term was incorporated to address the diprotonation of CP-38,371 in the solution chemistry.

In summary, there exist empirical relationships based primarily on the ampholyte solution chemistry to represent the pH effect upon the solubility and phase partitioning

of β-lactam antibiotics. However, no prior studies have ever taken into considerations the liquid phase nonideality in representing the phase equilibrium behavior of antibiotics.

6.3 Theoretical Framework

In Chapter 5, we proposed a molecular thermodynamic framework which successfully represents the phase equilibrium, including liquid-solid equilibrium of amino acids and small peptides as functions of temperature, ionic strength, solvent compositions, and pH.

Following the framework, we incorporated explicitly the solution chemistry of the β-lactam antibiotics and the Electrolyte NRTL activity coefficient model to fully account for both the chemical interactions and the physical interactions of the antibiotic solutions. The thermodynamic framework is general and it offers important advantages over previous empirical methods in representing the phase equilibrium behavior of the antibiotics. The framework requires only binary interaction parameters in computing activity coefficients of the true species in the systems. The binary interaction parameters identified for one antibiotic system are applicable to multicomponent systems containing the same antibiotic subsystem.

The Electrolyte NRTL model was originally proposed by Chen, et al. (1982) and Chen and Evans (1986) as an excess Gibbs energy expression for aqueous electrolytes and mixed-solvent electrolytes. The model has proved to be very successful in representing thermodynamic properties of various aqueous and mixed-solvent electrolyte systems, and systems containing zwitterions (see Chapter 5).

The model assumes that there are two contributions to the excess Gibbs energy of electrolyte systems. The first contribution accounts for the local interactions between ion and ion, between molecule and molecule, and between ion and molecule. The second contribution accounts for the long-range ion-ion interactions. The local interaction contribution is represented by a modified form of the Non-random Two Liquid (NRTL) equation. The long-range interaction contribution is represented by the combination of the Pitzer-Debye-Huckel equation and the Born equation.

$$g^{ex^*}/RT = g^{ex^*,NRTL}/RT + g^{ex^*,PDH}/RT + g^{Born}/RT \tag{6-3}$$

$$\frac{g^{ex^*,NRTL}}{RT} = \sum_m X_m \frac{\sum_j X_j G_{jm} \tau_{jm}}{\sum_k X_k G_{km}} + \sum_c X_c \sum_{a'} \frac{X_{a'}}{\sum_{a''} X_{a''}} \frac{\sum_j X_j G_{jc,a'c} \tau_{jc,a'c}}{\sum_k X_k G_{kc,a'c}} + \sum_a X_a \sum_{c'} \frac{X_{c'}}{\sum_{c''} X_{c''}} \frac{\sum_j X_j G_{ja,c'a} \tau_{ja,c'a}}{\sum_k X_k G_{ka,c'a}} \tag{6-3a}$$

$$\frac{g^{ex^*,PDH}}{RT} = -\left(\sum_k x_k\right)\left(\frac{1000}{M_s}\right)^{1/2} \frac{\left(4A_\phi I_x\right)}{\rho} \ln\left(1+\rho I_x^{1/2}\right) \tag{6-3b}$$

$$\frac{g^{Born}}{RT} = \frac{e^2}{2kT}\left(\frac{1}{D} - \frac{1}{D_w}\right)\left(\sum_i \frac{x_i Z_i^2}{r_i}\right) 10^{-2} \qquad (6\text{-}3c)$$

Note that the Born term is used to account for the Gibbs energy of transferring charged species (ions and zwitterions) from the infinite dilution state in pure water to the infinite dilution state in a mixed solvent. In the absence of nonaqueous solvents, the Born term reduces to zero.

The main adjustable parameters with the Electrolyte NRTL model are the binary interaction energy parameters, τ, associated with binary molecule-molecule pairs, binary molecule-electrolyte pairs, and binary electrolyte-electrolyte pairs. There are two binary interaction energy parameters per binary pair since the binary parameters are asymmetric. Being neutral, molecular solvent species and zwitterionic species are considered as "molecules" while cation-anion pairs are treated as "electrolytes," in the context of the Electrolyte NRTL model.

6.4 Liquid-Solid Equilibrium (Solubility)

Starting from solid antibiotics, the dissolution of the antibiotics and the dissociation of the zwitterions, the cations, and the anions can be represented by the following chemical equilibrium reactions:

$$A_{(s)} \quad <--> \quad A^{+-} \qquad (6\text{-}4)$$

$$A^{+} \quad <--> \quad A^{+-} + H^{+} \tag{6-5}$$

$$A^{+-} \quad <--> \quad A^{-} + H^{+} \tag{6-6}$$

$$H_2O \quad <--> \quad OH^{-} + H^{+} \tag{6-7}$$

The first reaction is the formation of the zwitterions in the liquid phase. The zwitterions then participate in acid-base reactions to form antibiotic cations (acidic form) and anions (basic form).

The chemical equilibrium relationships for the four reactions can be written as follows:

$$K_s = a_{A+-} \tag{6-8}$$

$$K_{A+} = a_{A+-}\, a_{H+} / a_{A+} \tag{6-9}$$

$$K_{A+-} = a_{A-}\, a_{H+} / a_{A+-} \tag{6-10}$$

$$K_w = a_{H+}\, a_{OH-} / a_{H2O} \tag{6-11}$$

where

$a_i = x_i \gamma_i$ for solvent species (6-12a)

$a_i = x_i \gamma^*_i$ for solute species (6-12b)

The thermodynamic chemical equilibrium constants of equations (6-9) and (6-10) are available in the literature (Yamana and Tsuji, 1976; Tsuji, et al., 1978; Hoover and Nash, 1978; and Tomlinson and Regosz, 1985).

The apparent solubilities of β-lactam antibiotics are the sum of the concentrations of the true species A^{+-}, A^{+}, and A^{-}.

$$x_{app} = x_{A+-} + x_{A+} + x_{A-} \tag{6-13}$$

As shown in equations (6-8) to (6-12), the solubilities are determined by the thermodynamic chemical equilibrium constants of the reactions as well as the activity coefficients of the true species in the liquid phase. The activity coefficients are functions of system temperature, solute compositions including ionic strength, solvent compositions, pH, etc. In representing the solubilities of antibiotics, it is essential that both the solution chemistry and the activity coefficients of the true species be properly accounted for.

6.5 Liquid-Liquid Equilibrium (Phase Partitioning)

The thermodynamic representation of liquid-liquid equilibrium of antibiotic solutions must take into account the following chemical reactions:

$$A^{+} \quad <--> \quad A^{+-} + H^{+} \tag{6-5}$$

$$A^{+-} \quad <--> \quad A^{-} + H^{+} \tag{6-6}$$

$$H_2O \quad <--> \quad OH^{-} + H^{+} \tag{6-7}$$

The chemical reactions exist in both the aqueous phase and the organic phase. The chemical equilibrium relationships for the above three reactions are given in equations (6-9) to (6-11).

In addition, the activity of each true species in the aqueous phase must be equal to the activity of the species in the organic phase.

$$(x_i\gamma_i)_{aq} = (x_i\gamma_i)_{org} \quad \text{for solvent species} \tag{6-14a}$$

$$(x_i\gamma^*_i)_{aq} = (x_i\gamma^*_i)_{org} \quad \text{for solute species} \tag{6-14b}$$

The apparent partition coefficient of the antibiotics is the ratio of the sum of the concentrations of the true species A^{+-}, A^+, and A^- in the organic phase to that in the aqueous phase.

$$K_{app} = (x_{A^{+-}} + x_{A^+} + x_{A^-})_{org} / (x_{A^{+-}} + x_{A^+} + x_{A^-})_{aq} \tag{6-15}$$

The apparent partition coefficients are therefore determined by the solution chemistry and the activity coefficients of the true species in the liquid phases. They are functions of system temperature, ionic strength, solvent compositions, pH, etc.

For β-lactam antibiotic "acids" such as Penicillin V, their basicity characteristic is weak. Either the anionic form or the neutral form of these compounds can become predominant in solutions, depending on pH of the solution. For these antibiotic "acids",

equation (6-5) should be dropped and the zwitterion, A^{+-}, in equation (6-6) should be replaced by a non-charged antibiotic, A.

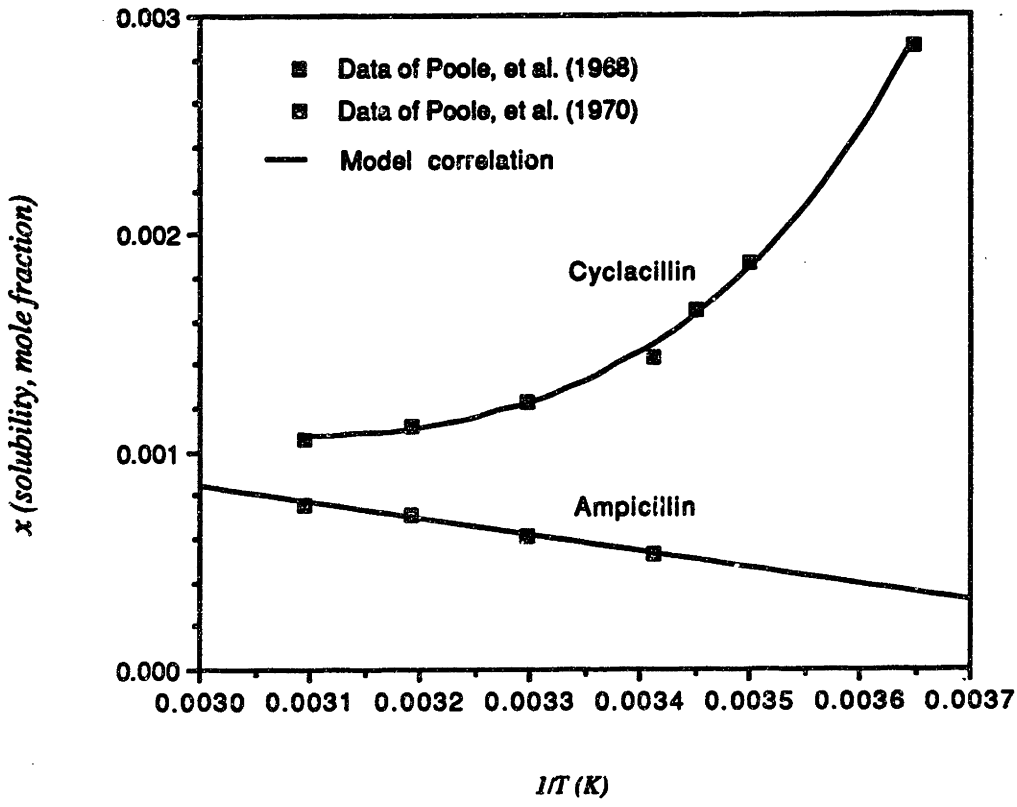
Results and Discussions

The theoretical framework has been used to correlate and predict the liquid-solid equilibrium behavior (solubilities) and the liquid-liquid equilibrium behavior (phase partitioning) of β-lactam antibiotics.

6.6 Temperature Effect on the Solubilities of β-lactam Antibiotics in Water

Like most chemical compounds, the solubility of antibiotics varies with temperature. Figure 6-2 gives the correlation results for the solubility data (Poole and Bahal, 1968; 1970) of ampicillin anhydrate and cyclacillin anhydrate in pure water over the temperature range of 293.15 K to 323.15 K. Due to the lack of data on activity coefficients, the activity coefficients were set to unity in regressing the solubility product constants. The plot of ampicillin anhydrate solubility vs. the reciprocal of absolute temperature gives straight line relationship, a typical van't Hoff plot. However, the plot for cyclacillin anhydrate is parabolic. This deviation from linearity is probably due to the degradation of the antibiotic at the higher temperatures (Tomlinson and Regosz, 1985).

Figure 6-2. Solubilities of Ampicillin Anhydrate and Cyclacillin Anhydrate vs. Temperature



The solubility product constant coefficients in equation (6-16) are adjustable parameters here.

$$\ln K_s = a + b/T + c \ln t \qquad (6\text{-}16)$$

The values of the coefficients in equation (6-16) are listed in Table 6-1.

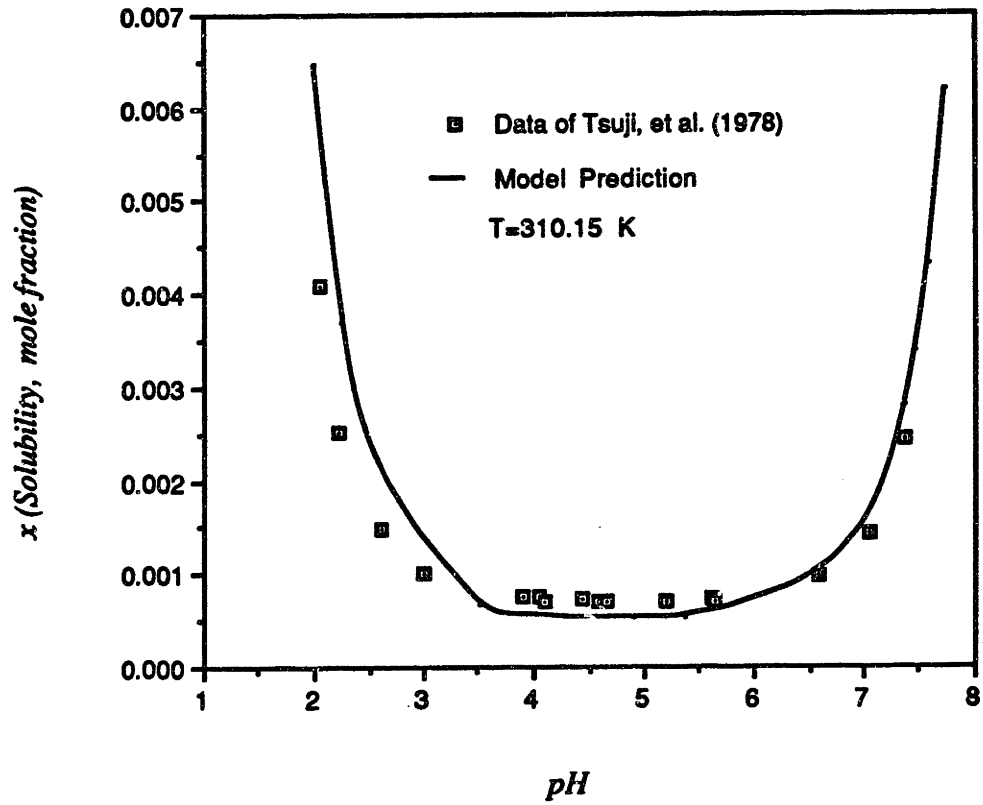
Table 6-1. Solubility Product Constant Coefficients of β-Lactam Antibiotics in Water

Antibiotics	a	b	c	r.m.s. rel. dev.
Ampicillin	217.3042	-11297.87	-32.79578	.0052
Cyclacillin	-541.4134	25370.73	78.93511	.0172

6.7 pH Effect on the Solubilities of β-lactam Antibiotics

Solution chemistry plays the key role with respect to the pH effect. With the activity coefficients set to unity, the chemical equilibrium calculations for the antibiotic solution chemistry give Figure 6-3 which shows the predicted pH effect upon the solubility of ampicillin anhydrate in water at 310.15 K. Values of pK_{A+} (=2.53) and pK_{A+-} (=7.24) were taken from Hoover and Nash (1978). The prediction results match well the experimental data of Tsuji, et al. (1978). The U-shaped solubility curve, with the minimum solubility at the pH near the isoelectric point of ampicillin anhydrate, is a typical feature shared by all the zwitterionic β-lactam antibiotics.

Figure 6-3. Solubility of Ampicillin Anhydrate vs. pH



As mentioned earlier, zwitterions and their acidic and basic forms are in chemical equilibrium in aqueous solutions. The nature of this equilibrium depends on the acid and base strength of the ionizing groups involved. Figure 6-4 shows that over the range of pH around the pI of the antibiotic (3.5 to 6 in this case), the zwitterion is the predominant species in the solution. As the pH drops to 2.5 or lower, the cationic form of the antibiotic, A^+, becomes dominant in the solution. As the pH increases to a value higher than 7, the anionic form of the antibiotic, A^-, becomes the dominant species.

6.8 Salt Effect on the Solubilities of β-Lactam Antibiotics

The salting-in and salting-out phenomena of antibiotics are widely applied in the precipitation of the biomolecules. In particular, ammonium salts are very popular reagents for precipitation processes because the solubilities of these inexpensive salts in water are high and relatively temperature-independent. The addition of salts in the solution introduces ion-molecule interactions in the aqueous phase which alternate the activity coefficients of the true species in solution.

Figure 6-5 shows the data of Nys, et al. (1979) for the salting effect of ammonium chloride, NH_4Cl, upon the solubility of 7-aminodeacetoxycephalosporanic acid (7-ADA) in water at 298.15 K and at pH=pI. At this condition, 7-ADA exists in water mainly as a zwitterion. To sufficiently describe the physical interactions in the system, the following binary interaction parameters need to be identified: τ 's for the water-ammonium chloride pair, τ 's for the water-7-ADA pair, and τ 's for the 7-ADA-ammonium chloride pair. τ 's for pairs involving electrolytes of the antibiotic cations or anions are insignificant since

Figure 6-4. Predicted Ampicillin Anhydrate True Species Concentrations in Solution

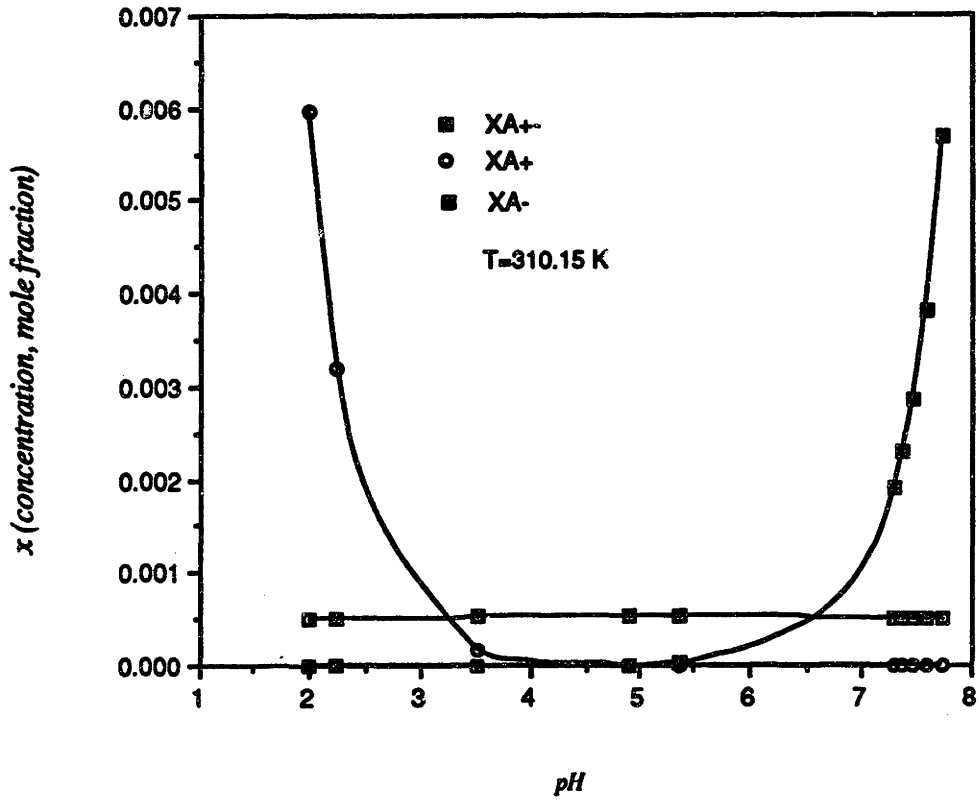
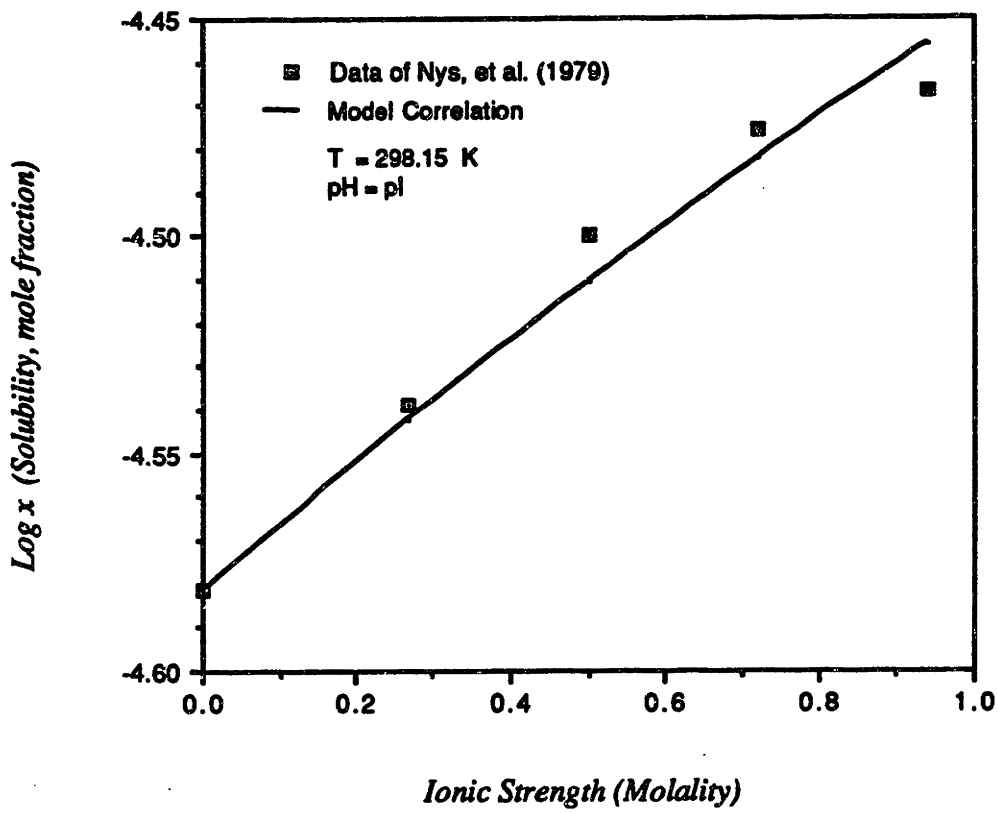


Figure 6-5. Effect of Ammonium Chloride upon the Solubility of 7-ADA



the dominant species in the system are water, ammonium ion, chloride ion, and the antibiotic zwitterion.

In applying the thermodynamic framework, τ 's for the water-ammonium chloride pair (τ_{12}=7.860, τ_{21}=-4.012) were taken from Chen and Evans (1986); τ 's for the water-7-ADA pair were set to zero ($\tau_{12}=\tau_{21}=0$); and τ 's for the 7-ADA-ammonium chloride pair were used as the correlation parameters to regress the salting effect upon the ampicillin anhydrate solubility. The determined values for the correlation parameters are given in Table 6-2.

Table 6-2. Electrolyte NRTL Model Parameters for Ammonium Chloride (1) - 7-aminodeacetoxycephalosporanic Acid (2) Pair

Salt	Antibiotics	τ_{12}	τ_{21}	r.m.s. rel. dev.
NH_4Cl	7-ADA	-5.697	8.045*	0.0182

$\ell n\ K_s$ = -10.54975
* parameter fixed

6.9 Solvent Effect on the Solubilities of β-lactam Antibiotics

Zwitterions in general are much less soluble in organic solvents than in water due to their charged groups (-COO^- and NH_3^+ groups). The addition of ethanol (dielectric constant 24.55 at 298.15 K) in the aqueous solution lowers the dielectric constant of the water-based solvent and raises the activity coefficients of true species ions and zwitterions. As a result, the solubilities drop.

Figure 6-6 shows the regression results for the data of Hou and Poole (1969) which illustrate the combined effects of solvent ethanol and potassium chloride salt upon the solubility of ampicillin anhydrate in water at 298.15 K. To sufficiently describe the physical interactions in the system, the following binary interaction parameters need to be identified: τ 's for the water-potassium chloride pair, τ 's for the water-ampicillin pair, τ 's for the ampicillin-potassium chloride pair, τ 's for the water-ethanol pair, τ 's for the ethanol-potassium chloride pair, and τ 's for the ethanol-ampicillin pair. Again, τ 's for pairs involving electrolytes of the antibiotic cations or anions are insignificant since none of these ions are the dominant species in the system.

In applying the thermodynamic framework, τ 's for the water-potassium chloride pair (τ_{12}=8.134, τ_{21}=-4.134) were taken from Chen and Evans (1986). τ 's for the water-ampicillin pair were set to zero ($\tau_{12}=\tau_{21}=0$). τ 's for the ampicillin-potassium chloride pair were treated as the adjustable parameters in fitting the ampicillin solubility data in the presence of potassium chloride. τ 's for the water-ethanol pair (τ_{12}=1.822, τ_{21}=-0.134, α=0.3) were taken from Mock et al. (1986). τ 's for the ethanol-potassium chloride pair were set to be the same as τ 's for the water-potassium chloride pair (τ_{12}=8.134, τ_{21}=-4.134). τ 's for the ethanol-ampicillin pair were set to zero ($\tau_{12}=\tau_{21}=0$). The Born radius for ampicillin was adjusted to regress the ampicillin solubility data in the presence of both potassium chloride and ethanol.

The regressed τ 's for the ampicillin-potassium chloride pair and the Born radius for ampicillin are given in Table 6-3.

Figure 6-6. Effects of Ethanol and Potassium Chloride Upon the Solubility of Ampicillin Anhydrate

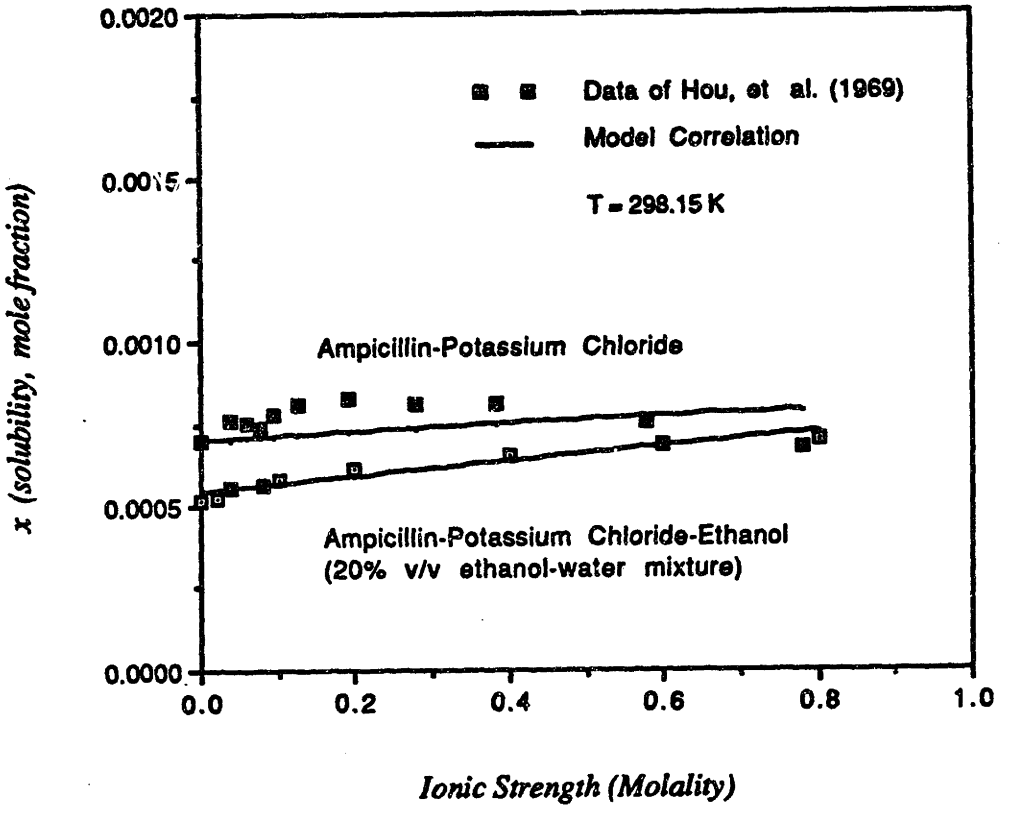


Table 6-3. Born Radii and Electrolyte NRTL Model Parameters for Potassium Chloride (1) - Ampicillin (2) - Ethanol - Water System

Salt	Antibiotics	τ_{12}	τ_{21}	r.m.s. rel. dev.
KCl	Ampicillin	-5.061	8.045*	0.0938

Antibiotics	Born Radius	r.m.s. rel. dev.
Ampicillin	0.302 nm	0.0593

$\ell n\ K_s$ = -7.26964
* parameter fixed

6.10 Solute Effect on the Solubilities of β-lactam Antibiotics

The influence of 7-phenylacetamideacetoxycephaloaporanic acid (7-PDA) on the solubility of 7-aminodeacetoxycephaloaporanic acid (7-ADA) in water (at 298.15 K and at pH=pI of 7-ADA) is shown in Figure 6-7. To represent this influence, the following binary interaction parameters need to be identified: τ 's for the water-7-ADA pair, τ 's for the water-7-PDA pair, and τ 's for the 7-ADA-7-PDA pair. Again, τ 's for pairs involving electrolytes of the antibiotic cations or anions are negligible since none of these ions are the dominant species in the system.

In applying the thermodynamic framework, τ 's for the water-7-ADA pair and the water-7-PDA pair were set to zero ($\tau_{12}=\tau_{21}=0$). τ 's for the 7-ADA-7-PDA were treated as adjustable parameters. It was assumed that, at pH=pI of 7-ADA, 7-ADA exists mainly as a zwitterion and 7-PDA exists mainly as an acid. The results of regression on the data

of Nys, et al. (1979) are summarized in Table 6-4 and Figure 6-7.

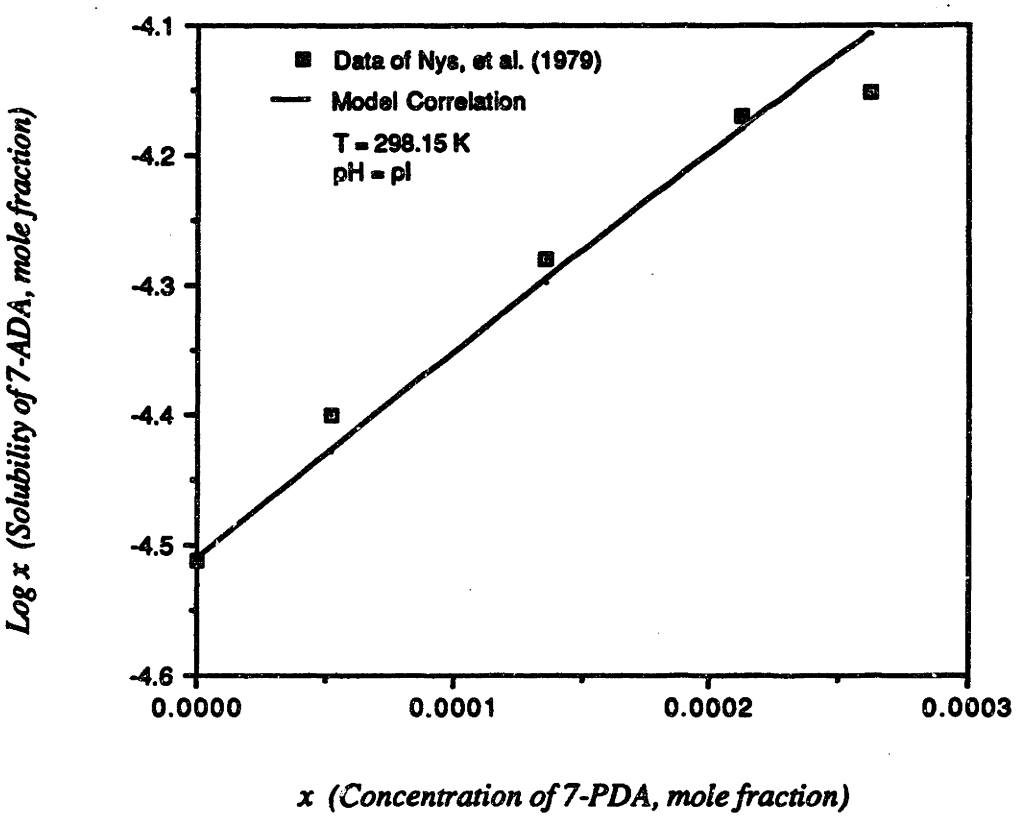
Table 6-4. Electrolyte NRTL Model Parameters for 7-Phenylacetamideacetoxycephaloaporanic Acid (1) - 7-Aminodeacetoxycephaloaporanic Acid (2) Pair

antibiotics(1)	(2)	τ_{12}	τ_{21}	r.m.s. rel. dev.
7-PDA	7-ADA	-17.838	0*	0.0431

$\ell n\ K_s$ = -10.39092

* parameter fixed

Figure 6-7. Effect of 7-PDA upon the Solubility of 7-ADA



6.11 pH Effect on the Phase Partitioning of β-lactam Antibiotics

The mechanism behind the pH effect upon the phase partitioning of antibiotics is similar to that of the solubility behavior. There are two major factors which determine the phase partitioning behavior of β-lactam antibiotics in two phase system. The first factor corresponds to the electric charges that the true species ions and zwitterions carry with them. The activity coefficients of the true species ions and zwitterions in the aqueous phase (high dielectric constant) are lower than those in the organic phase (low dielectric constant) due to the Born term of equation (6-3). As a result, the charge effect would favor dissolution of β-lactam antibiotics in the aqueous phase. However, the attractive physical interactions between the hydrophobic groups on the side chains of β-lactam antibiotics and the organic solvents would lower the activity coefficients of the true species and, therefore, favor dissolution of β-lactam antibiotics in the organic phase.

The data of Tsuji, et al. (1977) on the apparent partition coefficients of penicillin V in octanol-water system at 310.15 K as a function of pH were investigated in this work. (See equation (6-15) for the definition of the apparent partition coefficient.) Tsuji, et al. also reported the dissociation constants of the penicillin V solution chemistry.

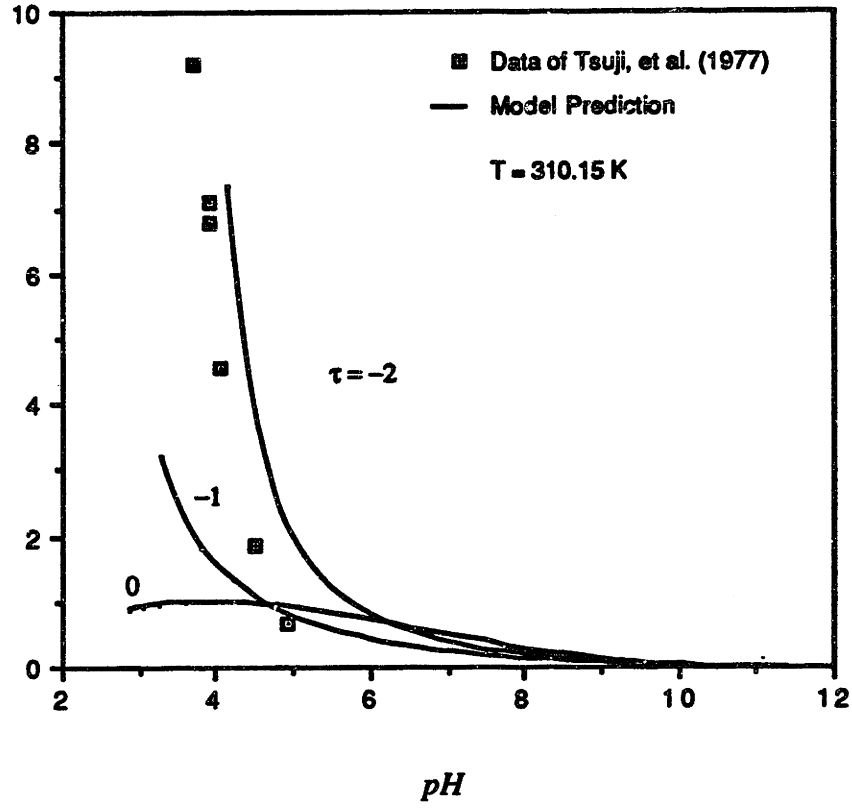
To fully describe the physical interactions in the system, the following binary interaction parameters need to be identified: τ's for the water-octanol pair, τ's for the water-penicillin V pair, τ's for the octanol-penicillin V pair, τ's for the water-electrolyte pairs, τ's for the octanol-electrolyte pairs, and τ's for the penicillin V-electrolyte pairs.

The τ's for the water-octanol pair have been reported by Sorensen and Arlt (1979)

and used directly in this work ($\tau_{12}=2682.7/T$, $\tau_{21}=12.0/T$, $\alpha=0.2$). These τ's faithfully describe the water-octanol liquid-liquid equilibrium. The water-penicillin V pair was assumed to be ideal and the τ's for the pair were set to zero ($\tau_{12}=\tau_{21}=0$). The octanol-penicillin V pair were varied with the simplifying assumption that the τ's are symmetric ($\tau_{12}=\tau_{21}$). The τ's for water-electrolyte pair were assumed ($\tau_{12}=8$, $\tau_{21}=-4$) along with the τ's for organics (including molecular penicillin V)-electrolyte pairs ($\tau_{12}=12$, $\tau_{21}=0$). The dielectric constant for octanol (10.34 at 298.15 K) was obtained from Dean (1985).

As discussed previously, the organic solvent dielectric constant and the solvent-antibiotic physical interactions are the two major factors controlling the phase partitioning behavior. For the water-octanol-penicillin V system, the influence of the solvent-antibiotic physical interactions were examined in this work. Figure 6-8 shows the computed apparent partition coefficients of penicillin V versus pH with the τ's for the octanol-penicillin V pair adjusted. As expected, the τ's for the octanol-penicillin V pair have a dramatic influence on the apparent partition coefficients. The more attractive the physical interactions between octanol and penicillin V (the more negative for the τ's) are, the greater the apparent partition coefficients are and the greater the concentrations of the antibiotics in the octanol phase become. The apparent partition coefficients approach zero at higher pH because penicillin V exists primarily as ions at higher pH and the higher solvent dielectric constant of water favors the existence of the ions in the aqueous phase over the organic phase.

Figure 6-8. Apparent Partition Coefficients of Penicillin V vs. pH



Notation

A	=	Antibiotic
A^{+-}	=	Zwitterion or dipolar ion
A^{+}	=	Antibiotic cation
A^{-}	=	Antibiotic anion
A_{ϕ}	=	Debye-H ckel constant for the osmotic coefficient
D	=	Dielectric constant
I_m	=	Ionic strength in molality
I_x	=	Ionic strength in mole fraction scale
K	=	Thermodynamic chemical equilibrium constant
K_s	=	Thermodynamic solubility constant
K_{app}	=	Apparent partition coefficient
M_s	=	Solvent molecular weight, kg/kmol
R	=	Gas constant
T	=	Temperature, K
X	=	Charge-weighted concentration
Z	=	Absolute value of ionic charge
a	=	Activity
e	=	Electric charge, $1.602189*10^{-19}$C
g	=	Molar Gibbs energy
g^{ex}	=	Molar excess Gibbs energy
k	=	Boltzmann constant, $1.380662*10^{-23}$J.K^{-1}
r	=	Born term ionic radius
x	=	Liquid phase mole fraction based on all true species: molecular and ionic
x_{app}	=	Apparent solubility

Greek Letters

α	=	Nonrandomness factor in the NRTL equation
γ	=	Activity coefficient, mole fraction scale
ρ	=	the closest approach parameter of the Pitzer-Debye-H ckel equation
τ	=	NRTL binary interaction energy parameters

Superscripts

*	=	Unsymmetric convention
Born	=	Long-range interaction contribution, represented by the Born equation
PDH	=	Long-range interaction contribution, represented by the Pitzer-Debye-Huckel equation
NRTL	=	Local interaction contribution, represented by the Non-Random Two Liquid equation

Subscripts

aq	=	aqueous phase
a,a',a"	=	anion
c,c',c"	=	cation
i	=	ionic species (including zwitterions)
j, k	=	any species
m	=	molecular species
org	=	organic phase
s	=	solid
w	=	water

Literature Cited

Berdy, Janos, CRC Handbook of Antibiotic Compounds, CRC Press, Inc., Boca Raton, Florida (1978).

Bogardus, J.B. and N.R. Palepu, International Journal of Pharmaceutics, 4, 159-170 (1979).

Chen, C.-C., H.I. Britt, J.F. Boston, and L.B. Evans, AIChE J., 25, 820 (1982).

Chen, C.-C. and L.B. Evans, AIChE J., 32, 444 (1986).

Chen, C.-C., Y. Zhu, and L.B. Evans, Biotechnology Progress, 5, (3), 111-118 (1989).

Dean, J.A. (editor), Lange's Handbook of Chemistry, 13th ed., P.10:113, McGraw-Hill, New York (1985).

Evans, L.B., "Bioprocess Simulation: A New Tool for Process Development," Bio/Technology, 6, 200-203, (1988).

Hoover, J.R.E., and C.H. Nash, Antibiotics (β-Lactams) in Kirk-Othmer, Encyclopedia of Chemical Technology, 3rd Edition, Volume 2, John Wiley & Sons, New York (1978).

Hou, J.P. and J.W. Poole, J. Pharm. Sci., 58, 1510-1515 (1969).

Mock, B., L.B. Evans, and C.-C. Chen, AIChE J., 32, 1655 (1986).

Nys, P.S., L.M., Elizarovskaya, N.N., Shellenberg and E.M., Savitskaya, Antibiotiki (Moscow), 24, 895-902 (1979).

Poole, J.W., and C.K. Bahal, J. Pharm. Sci., 57, 1945 (1968).

Poole, J.W., and C.K. Bahal, J. Pharm. Sci., 59, 1265 (1970).

Smith, J. E., Biotechnology, 2nd Ed., Edward Arnold, London (1988).

Sorensen, J.M., and W. Arlt, Liquid-Liquid Equilibrium Data Collection; Binary Systems, Chemistry Data Series, Vol. V, Part 1, P. 501, DECHEMA, Frankfurt/Main, Germany (1979).

Strong, J.E., An Example of Computer Simulation and Optimization of a Biochemical Process: Penicillin Recovery, M.S. thesis, Dept. of Chem. Eng., Mass. Inst. of Tech. (1986).

Tomlinson, E., and A. Regosz, Antibiotics: I. β-Lactam Antibiotics, in Solubility Data Series, Vol. 16/17, Pergamon Press, New York (1985).

Tsuji, A., O. Kubo, E. Miyamoto, and T. Yamana, J. Pharm. Sci., 66, 1675 (1977).

Tsuji, A., E. Nakashima, S. Hamano, and T. Yamana, J. Pharm. Sci., 67, 1059 (1978).

Tsuji, A., E. Nakashima, and T. Yamana, J. Pharm. Sci., 68, 308-11 (1979).

Yamana, T., and A. Tsuji, J. Pharm. Sci., 65, 1563 (1976).

Zhu, Y., L.B. Evans, and C.C. Chen, Biotechnology Progress, 266 (1990).

CHAPTER 7*

A MOLECULAR THERMODYNAMIC APPROACH TO PREDICT THE SECONDARY STRUCTURE OF HOMO-POLYPEPTIDES IN AQUEOUS SYSTEMS

* Chapter 7 has been submitted for publication in *Biopolymers,* with Chau-Chyun Chen, Jonathan A. King and Lawrence B. Evans

Introduction

The property of polypeptide chains of folding into discrete compact stable structures distinguishes them from many other organic polymers. The presence of twenty different side chains along the polypeptide backbone gives them much higher structural complexity than the macromolecules which underlie many industrial materials. Since the unfolded forms of proteins can refold in aqueous phase back into their mature structure, it has long been believed that the native state represents one of the lowest Gibbs energy (Anfinsen, 1973). In these cases the conformation of the native state must then be determined solely by the interactions among amino acids, the amino acids and the solvent, and among solvent molecules (Kim and Baldwin, 1990).

A considerable body of experimentation has revealed some general features of the native state: charged residues are located preferentially at the protein surface, where they can interact with water; residues in the interior are close-packed, with the solvent essentially excluded and no voids; and burying hydrophobic groups and surfaces is a

major source of stabilization (King, 1989). Each residue exhibits unique physical interactions with its neighboring residues. Along with steric hindrance, the physical interaction characteristics of these residues in a polypeptide chain determine the most favorable local environment around each residue.

7.1 Prediction of Secondary Structure and Molecular Thermodynamics

The problem of quantitatively describing these folding processes has remained elusive. There is a large literature in the prediction of secondary structure of proteins and polypeptides since the early works of Blout et al. (1960), Davies (1964), Kotelchuck et al. (1969), Lim (1974), Schulz et al. (1974a,b), Chou and Fasman (1974a,b; 1978), Garnier et al. (1978), and others. In particular, helix-coil transition and the prediction of α-helix structure have been subject to intensive experimental and theoretical studies (Schellman, 1955; Zimm and Bragg, 1959; Poland and Scheraga, 1970; Scheraga, 1978; Dill, 1985; Chan and Dill, 1989; Ben-Naim, 1990).

Schulz and Schirmer (1979) gave a comprehensive review on the prediction of secondary structure from amino acid sequences. They placed the "prediction methods" into two categories: probabilistic and physico-chemical. The former extracts rules and parameters using purely statistical analysis of the protein data base, or relies on the known correlations between sequence and structure available in the data base. The latter applies, either exclusively or in addition to the data base correlations, structural information (both experimental and theoretical) from outside the data base. So far these prediction methods have had only limited success (von Heijne, 1987).

Molecular thermodynamics has been well-established as a useful, semi-empirical approach to represent the system excess Gibbs energies and to study fluid phase equilibria of both small and large molecular systems (Prausnitz, 1979). Examples of these molecular systems include nonelectrolytes, electrolytes (Chen and Evans, 1986), amino acids (Chapter 5; Chen et al., 1989), antibiotics (Chapter 6; Zhu, et al., 1990), and surfactants and micelles (Blankschtein et al., 1986; Chen, 1989). In these systems, the system Gibbs energies and the phase equilibrium behaviors are often dictated by the configurational entropy change and the weak physical interactions that exist between various species in the solution. These physical interactions are conveniently accounted for via binary interaction parameters. These binary interaction parameters characterize the potential energies of the binary physical interactions in both binary systems and multicomponent systems. A great deal of research has been carried out in the last twenty years to determine values of the interaction parameters for systems involving small organic and inorganic molecules in solution by regressing experimental phase equilibrium data. Methods have also been developed to predict the phase equilibrium behavior from functional group contribution techniques and to regress the interaction parameters from the predicted phase equilibrium data (Fredenslund et al., 1975; 1977).

In this study, we investigated, from the molecular thermodynamic viewpoint, the rules governing the formation of secondary structures of aqueous polypeptides. We developed a method to determine the thermodynamically favorable peptide chain secondary structure based on the knowledge of amino acid sequence and the physical

interactions between two amino acid residues and between residues and solvents. Specifically, we approximate the physical interactions in aqueous polypeptide systems with the binary interaction parameters of the Non-random Two Liquid (NRTL) excess Gibbs energy model for nonelectrolyte solutions (Renon and Prausnitz, 1968). The model has been later extended by Chen et al. (1982) and Chen and Evans (1986) as an excess Gibbs energy expression for aqueous electrolytes, mixed-solvent electrolytes (Mock et al., 1986), and systems containing zwitterions (Chapter 5 and Chapter 6).

7.2 Local Compositions of Aqueous Polypeptide Solutions

Local compositions of aqueous polypeptide solutions are different from those of typical small molecular fluid systems in several ways. In small molecular fluid systems, each component is fully dispersed in the fluid and the local compositions of a species depend upon the local physical interactions and the bulk compositions of the fluid. For aqueous polypeptide solutions, the residues are confined in a local lattice structure because they are covalently bonded together with a specific amino acid sequence. Within this local lattice structure, the local compositions of the residues are determined by the weak physical interactions between the residues and between the residues and the solvent molecules, the amount of solvent molecules that are solvated by the polypeptide chain, and the amino acid sequence.

We envision that patterned conformations of aqueous polypeptides, folded or unfolded, reflect favored local compositions of aqueous polypeptides. For completely unfolded and fully extended polypeptides, the favored nearest neighboring species (except

the covalently-linked neighboring residues) around a given residue approximate totally solvent water. For folded polypeptides, the nearest neighboring species around each residue of the folded structure reflect a state of minimum Gibbs energy and it is subject to the constraints of steric hindrance and intramolecular covalent bonding (including disulfide bonds).

A good way to represent polypeptide conformations in aqueous solutions is the contact map. When a residue i is at the nearest neighborhood of residue j, the element ij of the contact map matrix becomes nonzero. In these maps, α-helices show up as a broadening of the diagonal (down from top left to bottom right), antiparallel β-sheets appear as a band perpendicular to the diagonal, and parallel β-sheets appear as a band running parallel to the diagonal. The native conformation of bovine pancreatic trypsin inhibitor (BPTI) is shown in Figure 7-1. A contact map of BPTI is given in Figure 7-2.

7.3 A Molecular Thermodynamic Model for Aqueous Homo-polypeptides

The energetics of the folding of residues can be represented as the following two individual contributions: the entropy contribution and the enthalpy contribution.

$$\Delta G = \Delta H - T\Delta S \tag{7-1}$$

Figure 7-1. Model Polypeptide Mimics Folding in BPTI
(Oas and Kim, Nature, 336, 6194, 42-48, 1988)

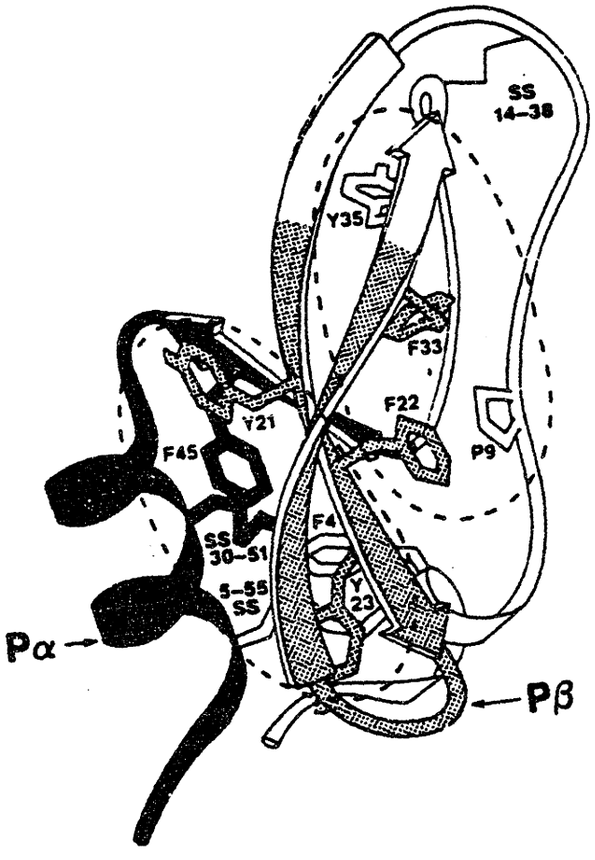
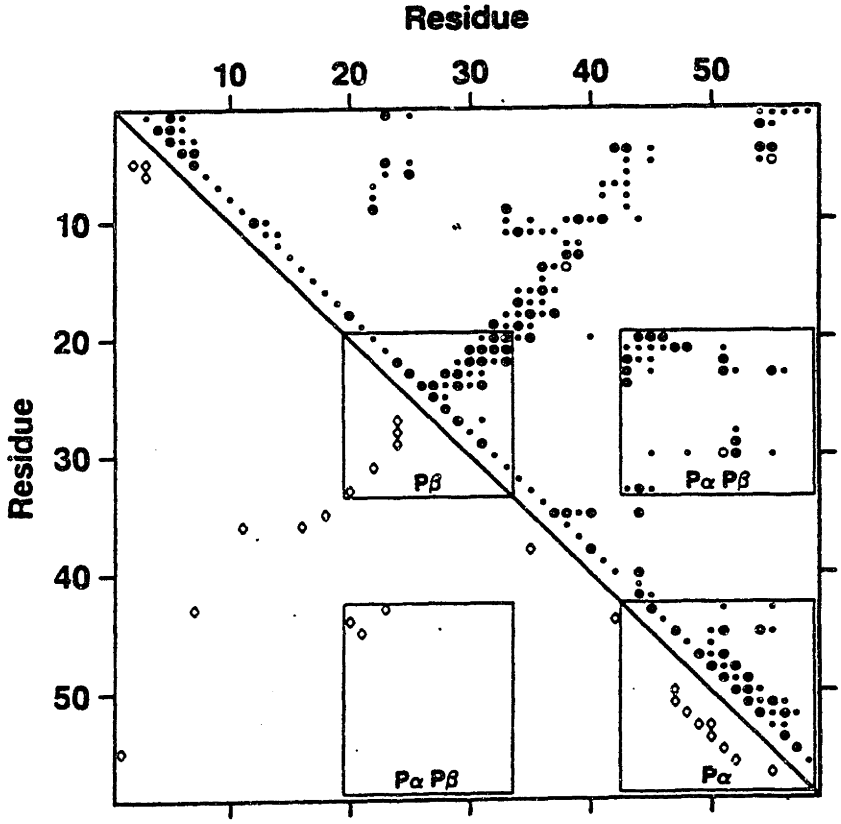


Figure 7-2. Contact Map of Bovine Pancratic Trypsin Inhibitor
(Oas and Kim, Nature, 336, 6194, 42-48, 1988)



The Gibbs energies of folding residues in a polypeptide chain from a coiled state into a secondary structure (either an α-helix or a β-sheet state) is the Gibbs energy difference between state α (or β) and the random-coiled state c.

$$\Delta G^{c\to\alpha} = \Delta H^{c\to\alpha} - T\Delta S^{c\to\alpha} \tag{7-2}$$

$$\Delta G^{c\to\beta} = \Delta H^{c\to\beta} - T\Delta S^{c\to\beta} \tag{7-3}$$

The Gibbs energy of folding is subject to the entropy change in folding residues into a specific polypeptide conformation and the physical interactions among amino acid residues, solvents, and all other species present in the local lattice structure at a set of given conditions (temperature, pH, ionic strength, solvent, etc.). For convenience, we chose the system to be one single polypeptide chain with the amount of solvent molecules necessary to "solvate" all the amino acid residues on the random-coiled chain. In other words, the system consists of the polypeptide chain and the nearest neighboring solvent molecules that constitute the hydration shell. A typical coordination number of six is assumed in this treatment (Prausnitz et al., 1986). Based on the experience with molecular thermodynamic models, we believe the choice on the coordination number does not significantly impact the results.

The reference states are chosen to be the pure liquid for water and a hypothetical amino acid residue aggregate state for residues, or the residue in pure homo-polypeptides. In this aggregate state for residues, the residues are surrounded by residues

of the same amino acid. The reference state for water is characterized by the water-water physical interactions. The reference state for residues is characterized by the residue-residue physical interactions.

7.4 The Entropy Contribution

Folded polypeptides are tightly packed like molecular crystals and the configurational entropy of mixing for aqueous polypeptide systems can be considered as partially lost or completely lost depending on the folded conformations.

The lattice model of Flory (1941, 1942) and Huggins (1941, 1942) gives a simple representation for the entropy of mixing of a polymer solution. The configurational entropy of mixing the disoriented polymer and the solvent to form a random coil polymer solution is :

$$\Delta S_{mix} = - R \{N_1 * \ln [N_1/(N_1 + N_2 * x)] + N_2 * \ln [N_2 * x/(N_1 + N_2 * x)]\} \qquad (7\text{-}4)$$

where, N_1: number of solvent molecules.

N_2: number of polymer chain molecules of x segments.

x: number of segments of a chain, or the ratio of the molar volume of a chain over the molar volume of the solvent.

We apply the Flory-Huggins expression to the system of one polypeptide chain with a shell of water molecules solvating the random-coiled polypeptide molecule with

the assumption that a residue occupies a single lattice site as that of a solvent water molecule. Then,

$$N_1 = 4x+2 \tag{7-5}$$

$$N_2 = 1 \tag{7-6}$$

$$x = n \tag{7-7}$$

Here, the coordination number of six is used.

It should be noted that x, the molar volume ratio of polymer chain over that of water, is equal to the x cells in Flory's lattice that are occupied by the segments and need not be equal to the number of amino acid residues in the polypeptide molecule. However, in our treatment, the x of Flory-Huggins theory is considered to be the degree of polymerization, or the number of residues in the chain. This treatment is consistent with the understanding that, in the folded secondary structure (helix or sheet), the side chains of the residues retain essentially the same degrees of freedom compared to the side chains in the coil state (McGregor et al., 1987).

The α-helix conformation of a polypeptide chain is very condensed while the β-pleated sheet is almost fully extended. As a first approximation, we assume that a residue should lose its entire configurational entropy by folding into a helical conformational state, and lose part of its entropy by folding into a β-sheet since the sheet may still have motions in two directions perpendicular to the plane. β-sheets in globular

proteins, generally packed against each other or against helices, are quite stiff. However, we are modeling the first organized state or form of a β-sheet, which has not yet reached the highly organized sandwich state in folded proteins. The entropy loss upon folding a residue from a random coil state into an α-helical state is taken as the entropy of mixing of a polypeptide molecule of chain length n-1 with its solvated water molecules minus that of a polypeptide molecule of chain length n:

$$\Delta s^{c\to\alpha} = (\Delta S^{c,FH}|_{n-1} - \Delta S^{c,FH}|_{n}) \tag{7-8}$$

or

$$\Delta S^{c\to\alpha} = n(\Delta S^{c,FH}|_{n-1} - \Delta S^{c,FH}|_{n}) \tag{7-9}$$

The entropy loss upon folding a residue from a random-coiled conformation in a polypeptide chain of n residues into an antiparallel sheet conformation is considered as the entropy of mixing of a polypeptide molecule of chain length n/2 with its solvated water molecules minus that of a polypeptide molecule of chain length n, then averaged over n:

$$\Delta s^{c\to\beta} = (\Delta S^{c,FH}|_{n/2} - \Delta S^{c,FH}|_{n})/n \tag{7-10}$$

or

$$\Delta S^{c\to\beta} = (\Delta S^{c,FH}|_{n/2} - \Delta S^{c,FH}|_{n}) \tag{7-11}$$

7.5 The Enthalpy Contribution

The enthalpy contribution must account for the unique residue-residue and residue-solvent interactions which exist in the polypeptide solution. For each conformational state of a polypeptide chain, the residual interaction energies are to be summed up from the local interactions as expressed in the contact map. With a coordination number of six for residues in secondary structures, the neighboring species around an i th residue include the i-1 th residue, the i+1 th residue, the i-3 th residue, the i+3 th residue, and two solvent molecules in the case of α-helices. However, for β-pleated sheets such as antiparallel sheets, the neighboring species around an i th residue include the i-1 th residue and the i+1 th residue from the same chain, the j th residue from the antiparallel chain, and three solvent molecules (see Figures 7-3a and 7-3b).

It should be noted that by a "residue," here we mean a peptide unit, -CH(R_i)-C(=O)-NH-, rather than -NH-CH(R_i)-C(=O)-. In a polypeptide chain the former notation is more representative for showing the interactions involving amino acid residues than the later.

We define the interaction energies per residue in the states random coil, α-helix, and β-sheets as:

$$g_R^c = (2\, g_{RR} + 4\, g_{WR})/6 \tag{7-12}$$

$$g_R^\alpha = (4\, g_{RR} + 2\, g_{WR})/6 \tag{7-13}$$

$$g_R^\beta = (3\, g_{RR} + 3\, g_{WR})/6 \tag{7-14}$$

Figure 7-3a. Interactions in an α-helical Conformation

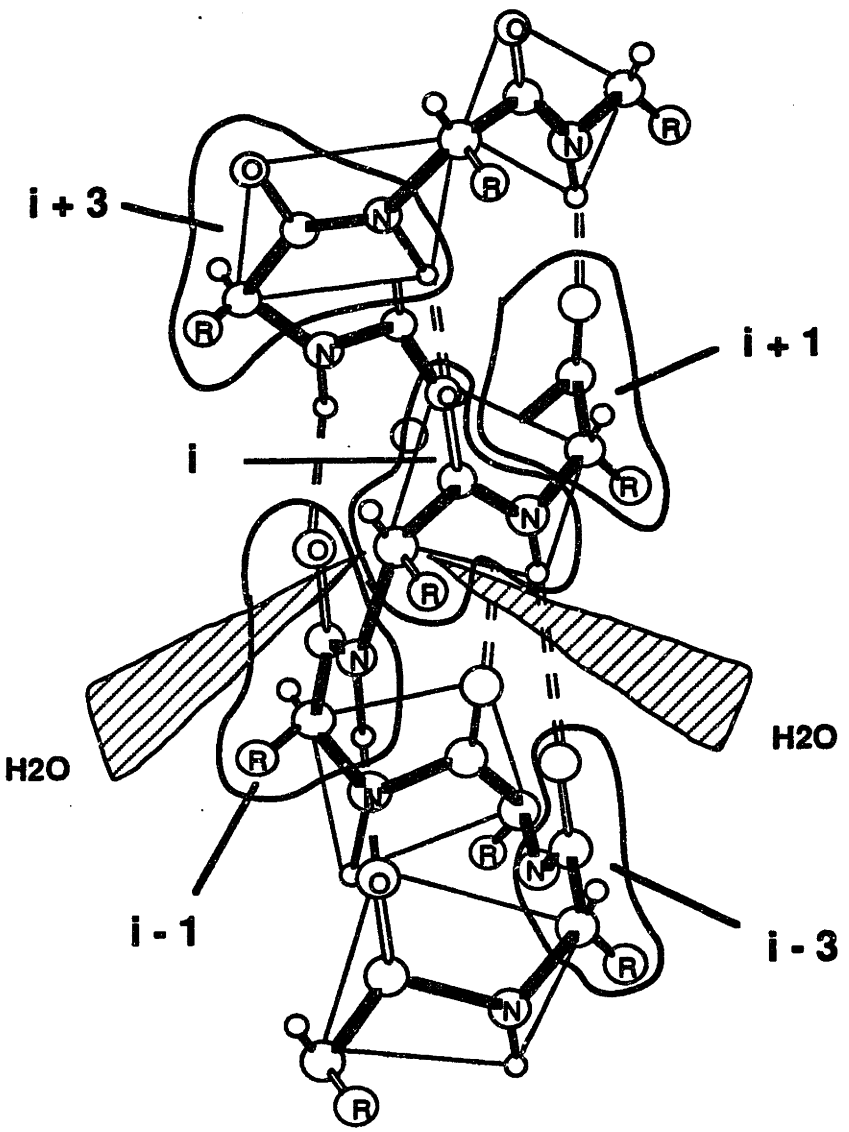
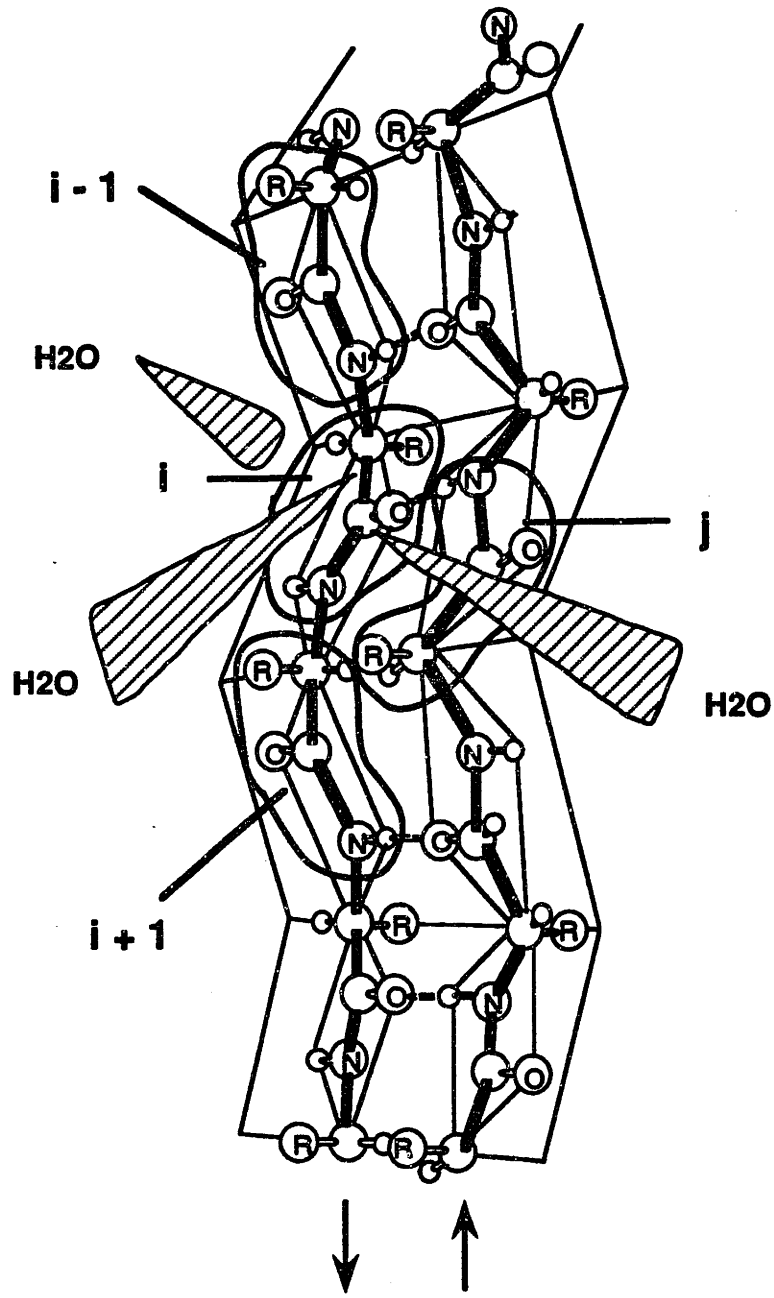


Figure 7-3b. Interactions in an Antiparallel β-sheet Conformation



The quantity g_{ij} is the potential energy of interaction between i-j pair of species, and are inherently symmetric ($g_{ij}=g_{ji}$).

The interaction energies per water molecule in a solvated state (hydrogen-bonded to amide groups of the polypeptide chain) and a pure water state (free water beyond the hydration shell of the chain) are:

$$g_w^s = (g_{RW} + 5\, g_{WW})/6 \tag{7-15}$$

$$g_w^w = g_{WW} \tag{7-16}$$

where, g_{RR}, g_{WR} and g_{WW} are the potential energies of interaction between two amino acid residues, between a water molecule and an amino acid residue, and between two water molecules, respectively. In this formulation, only first neighbor contacts are considered.

Therefore, the interaction energy changes caused by folding a residue from its coil state to a helical state and a sheet state are:

$$\begin{aligned}\Delta g_R^{c\to\alpha} &= g_R^{\alpha} - g_R^{c} \\ &= 2\,(g_{RR} - g_{WR})/6 \\ &= -\tau_{WR} * RT/3\end{aligned} \tag{7-17}$$

$$\Delta g_R^{c\to\beta} = g_R^{\beta} - g_R^{c}$$

$$= (g_{RR} - g_{WR})/6$$

$$= - \tau_{WR} * RT/6 \qquad (7\text{-}18)$$

The interaction energy change of a solvent molecule from its solvated state to its pure water state is:

$$\Delta g_W^{s \to w} = g_W^s - g_W^w$$

$$= (g_{WW} - g_{RW})/6$$

$$= - \tau_{RW} * RT/6 \qquad (7\text{-}19)$$

here, τ_{ji}'s are interaction parameters between species j and i defined as the following:

$$\tau_{ji} = (g_{ji} - g_{ii})/RT \qquad (7\text{-}20)$$

For each residue, the formation of an α-helix releases two water molecules from the solvated state. The formation of a β-sheet releases one water molecule. Therefore, the change of the system enthalpy upon folding a particular residue in a polypeptide chain from coil conformation to α-helix or β-sheet structure can be formulated as:

α-helix formation:

$$\Delta h^{c \to \alpha} = \Delta g_R^{c \to \alpha} + 2\Delta g_W^{s \to w}$$

$$= -2(\tau_{WR} + \tau_{RW}) * RT/6$$

$$= -(\tau_{WR} + \tau_{RW}) * RT/3 \qquad (7\text{-}21)$$

β-sheet formation (antiparallel):

$$\Delta h^{c\to\beta} = \Delta g_R^{c\to\beta} + \Delta g_W^{s\to w}$$

$$= -(\tau_{WR} + \tau_{RW}) * RT/6 \qquad (7\text{-}22)$$

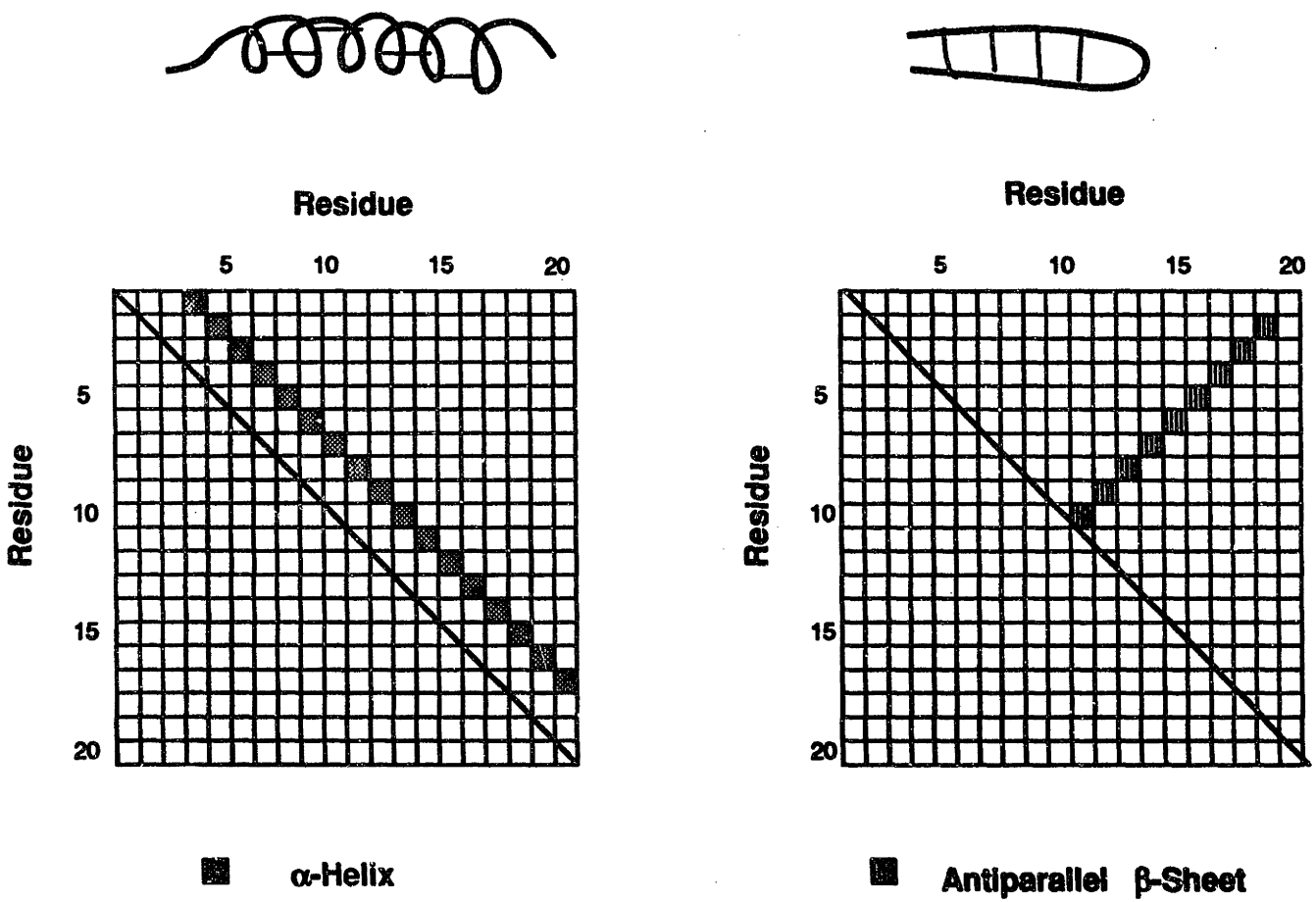
The contact maps in Figure 7-4 illustrate the residue-residue interactions occurring in the α-helix and β-sheet conformations of homo-polypeptides. The contacts due to the covalent bond linkage along the chain are not shown on the maps. The broadening of the diagonal band represents an α-helix. Antiparallel β-sheet conformation is shown as a band perpendicular to the diagonal. Each filled square represents 1/3 of the interaction energy between two neighboring residues, g_{RR}.

The thermodynamic quantities of the aqueous polypeptide system are the summation of the contributions resulting from the change of the thermodynamic quantities of each individual residue. For a homo-polypeptide molecule with chain length of n, with the chain ends effect ignored, the enthalpy of mixing can be written as follows:

$$\Delta H^{c\to\alpha} = \Sigma_{i=1}^{n} \Delta h^{c\to\alpha} \qquad (7\text{-}23)$$

$$\Delta H^{c\to\beta} = \Sigma_{i=1}^{n} \Delta h^{c\to\beta} \qquad (7\text{-}24)$$

Figure 7-4. Contact Maps of α-helix and β-sheet Conformations of Polypeptides



7.6 Effect of Chain Ends

According to Pauling et al. (1951), at the N-terminus of an α-helix, there are four residues whose amino groups -NH- do not form hydrogen bonds (contacts) with the carbonyl groups -CO- of the chain. At the C-terminus of the α-helix, there are also four residues whose carbonyl groups -CO- are not hydrogen-bonded with the amino groups of the chain. For β-sheets, there are always two residues per chain whose amino groups -NH- do not form hydrogen bonds (contacts) with the carbonyl groups -CO- of the chain, and two residues whose carbonyl groups -CO- are not hydrogen bonded with the amino groups of the chain. Therefore, the change of the system enthalpy in folding a coil into an α-helix or a β-sheet should be adjusted for the chain ends:

$$\Delta H^{c\to\alpha} = -(n-4)(\tau_{WR} + \tau_{RW}) * RT/3 \qquad (7\text{-}25)$$

$$\Delta H^{c\to\beta} = -(n-2)(\tau_{WR} + \tau_{RW}) * RT/6 \qquad (7\text{-}26)$$

The chain end effect also needs to be addressed in the entropy calculation in the case of β-sheet. Two end residues per chain (each at one end) always stay free from folding in the β-sheet case. In other words, these end residues retain their degrees of freedom even after the folding. Therefore, equation 7-11 should take the following form:

$$\Delta S^{c\to\beta} = (n-2)(\Delta S^{c,FH}|_{n/2} - \Delta S^{c,FH}|_{n})/n \qquad (7\text{-}27)$$

7.7 Estimation of Binary Interaction Parameters

The enthalpy contribution requires two binary interaction parameters for each residue-solvent (water) pair. The definition of these binary parameters is given in equation 7-20. To obtain an estimate of the interaction parameters, we applied the UNIFAC (Universal Functional Activity Coefficient) group contribution method of Fredenslund et al. (1975, 1977).

The UNIFAC method was developed for the purpose of predicting activity coefficients for organic mixtures where little or no experimental data are available (Prausnitz, 1977). In this method, each molecule is treated as the sums of the functional groups which constitute that molecule. The UNIFAC parameters for the contribution of each functional group were developed by regressing phase equilibrium data of a wide range of systems composed of small molecules with various functional groups. We used ASPEN PLUS™ (Aspen Technology, 1988) software system to identify the set of UNIFAC functional groups for each residue and to calculate the contribution of each functional group to the interaction parameters.

The estimation technique for the binary interaction parameters is carried out in two steps. First, infinite dilution activity coefficients are estimated for a given binary residue-water mixture with the UNIFAC method. Then, with the infinite dilution activity coefficients, the Non-Random-Two-Liquid (NRTL) equation (Renon and Prausnitz, 1968) is solved for the NRTL binary interaction parameters with the nonrandomness factor set to 0.3. The NRTL equation is chosen here because the equation has been successfully

applied to biomolecular systems such as amino acids, antibiotics, and surfactants (Chapter 5, Chapter 6, Chen, 1989).

Due to the complexity of the twenty amino acid side chain structure, the limited number of the UNIFAC functional groups available in the literature, and the lack of the phase equilibrium data to evaluate the interactions for new functional groups, several approximations were made in estimating the binary interaction parameters. The peptide backbone >CH-C(=O)-NH- for a residue was represented as >CH-C(=O)-O-. The side chain of methionine was treated as $-CH_2-CH_2-O-CH_3$, asparagine as $-CH_2-CH(OH)-NH_2$, glutamine as $-CH_2-CH_2-CH(OH)-NH_2$, and arginine as $-(CH2)_3-NH-CH(OH)-NH_2$. In addition, the double bond >C=CH-NH- in the five-membered ring of tryptophan was treated as two single bonds, or, >C< and >CH-NH-. The double bond -N=CH- in the ring of histidine side chain was represented as two single bonds $-NH-CH_2-$. The guanidine group $-NH-C(=NH)-NH_2$ in arginine was approximated as $-NH-CH(-OH)-NH_2$.

The representation of the twenty amino acid residues as functional groups and the definition of the functional groups are given in Tables 7-1 and 7-2. Tryptophan, for instance, was represented as 1005 1/3300 1/1010 1/1000 1/1105 6/1700 1 in the calculation. It means that its backbone consists of one >CH- group (1005) and one -C(=O)-NH- group (treated as -C(=O)-O-, 3300); its side chain has one $-CH_2-$ group (1010), one >C= group (approximated as >C<, 1000), six aromatic carbon -CH= groups (1105), and one =CH-NH- group (approximated as >CH-NH-, 1700).

The estimated binary interaction parameters for the pairs of the amino acid residues and water are listed in Table 7-3.

Amino acids may be considered to constitute four classes: amino acids with a nonpolar side chain (alanine, valine, leucine, isoleucine, methionine, phenylalanine, proline, and tryptophan), amino acids with a polar side chain (glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine), amino acids with a charged cationic side chain (lysine, arginine and histidine), and amino acids with a charged anionic side chain (aspartate and glutamate). The amino acids with a polar side chain or a charged side chain are hydrophilic amino acids. Hydrogen bonds may be formed between these hydrophilic amino acid residues and the solvated water molecules. The amino acids with a nonpolar side chain are hydrophobic amino acids. These amino acid residues are known to form solvent-excluding protein interiors due to attractive residue-residue van der Waals contacts.

Table 7-3 shows that both τ_{WR}'s and τ_{RW}'s are positive for the hydrophobic amino acid residues (except methionine, which has a -S-CH_3 group). Also, as the nonpolar side chain becomes bulkier, the values of the interaction parameters become larger; the interactions between amino acid residues and water becomes more repulsive; and the residues tend to dislike water to a greater degree.

The values of the interaction parameters for the amino acid residues with a neutral polar side chain show that they have relatively attractive interactions with water compared with the nonpolar side chain amino acid residues. Tyrosine, which has a

hydrophobic benzene ring, is an exception. Threonine is representative of the second class. Being neutral polar, a threonine residue has a less repulsive interaction with water compared with a valine residue or a leucine residue. However, a threonine residue also has a longer side chain with more -CH2- groups than that of an alanine residue. Therefore, it has a more repulsive interaction with water than an alanine residue does.

Table 7-1. The Representation of Amino Acid Residues Using UNIFAC Functional Groups

Residue	Alias	Backbone Group	Side Chain Group
Alanine	Ala	1005 1/3300 1	1015 1
Valine	Val	1005 1/3300 1	1005 1/1015 2
Leucine	Leu	1005 1/3300 1	1005 1/1010 1/1015 2
Isoleucine	Ile	1005 1/3300 1	1005 1/1010 1/1015 2
Proline	Pro	1005 1/3300 1	1010 3
Methionine	Met	1005 1/3300 1	1010 2/1615 1
Phenylalanine	Phe	1005 1/3300 1	1105 5/1155 1
Tryptophan	Trp	1005 1/3300 1	1010 1/1000 1/1105 6/1700 1
Glycine	Gly	1010 1/3300 1	
Serine	Ser	1005 1/3300 1	1010 1/1200 1
Threonine	Thr	1005 1/3300 1	1005 1/1015 1/1200 1
Cysteine	Cys	1005 1/3300 1	2400 1
Tyrosine	Tyr	1005 1/3300 1	1105 4/1155 1/1350 1
Asparagine	Asn	1005 1/3300 1	1010 1/1200 1/1650 1
Glutamine	Gln	1005 1/3300 1	1010 2/1200 1/1650 1
Aspartic acid	Asp	1005 1/3300 1	1010 1/1955 1
Glutamic acid	Glu	1005 1/3300 1	1010 2/1955 1
Lysine	Lys	1005 1/3300 1	1010 3/1655 1
Arginine	Arg	1005 1/3300 1	1010 2/1200 1/1650 1/1705 1
Histidine	His	1005 1/3300 1	1155 1/1700 1/1705 1

Table 7-2. Definition of the Functional Groups

Group Code	Group
1000	>C<
1005	-CH<
1010	$-CH_2$
1015	$-CH_3$
1100	(AC) (AC: Aromatic carbon)
1105	(AC)-H
1150	(AC)-CH<
1155	(AC)-CH_2-
1200	-OH
1350	(AC)-OH
1615	$-OCH_3$
1650	>CH-NH_2
1655	$-CH_2-NH_2$
1700	>CH-NH-
1705	$-CH_2$-NH-
1955	-COOH
2400	$-CH_2$-SH
3300	-(C=O)-O-

Table 7-3. Estimated Interaction Parameters, τ's, for the Pairs of Water and Amino Acid Residues (at 25° C)

No.	Solvent(W)	Residue(R)	τ_{WR}	τ_{RW}
1	water	ala	2.221	0.461
2	water	val	3.988	0.978
3	water	leu	4.920	1.272
4	water	ile	4.920	1.272
5	water	pro	3.833	0.934
6	water	met	5.548	-1.489
7	water	phe	7.704	2.505
8	water	trp	7.825	0.211
9	water	gly	1.409	0.269
10	water	ser	2.882	-1.288
11	water	thr	3.280	-1.114
12	water	cys	3.115	-0.676
13	water	tyr	5.487	-0.533
14	water	asn	3.954	-2.138
15	water	gln	4.276	-2.010
16	water	asp	3.222	-1.065
17	water	glu	3.689	-0.940
18	water	lys	4.380	-1.459
19	water	arg	0.006	-1.317
20	water	his	3.030	-1.712

The third class of amino acid residues consists of those with acidic side chains. Both aspartic acid residues and glutamic acid residues have a carboxyl group on their side chains. They are more hydrophilic, especially when the acidic side chains lose the proton at a pH greater than the pK_a's of the side chains. This class of amino acid residues should have a more attractive interaction with water compared to the nonpolar side chain residues. As shown in Table 7-3, aspartic acid residues and glutamic acid residues show a more attractive interaction with water than valine residues do.

The fourth class of amino acids consists of those with basic side chains. The UNIFAC predictions for these three residues show the same trend as the third class of amino acid residues. Their interaction parameters show that these residues are very hydrophilic.

It should be noted that the side chains of amino acids of the third and fourth classes may carry positive or negative charges depending on the pH of the solution. In general, the interaction parameters associated with the binary pairs of water and charged species show strong attractive interaction between water and the charged species (Chen and Evans, 1986), i.e., strong hydrophilicity. It is known that the UNIFAC method can only predict reliable interaction parameters for nonelectrolyte systems. We believe the UNIFAC predictions under-estimate the hydrophilicity of the charged amino acid residues.

7.8 A Hydrophobicity Scale

Hydrophobicity of an amino acid (or its residue or side chain) is a relative measure of its likeness for nonpolar organic solvents. Among the hydrophobicity scales in the literature, the hydrophobic scale of Nozaki and Tanford (1971) is the most popular reference for the relative hydrophobicities of amino acids. Nozaki and Tanford measured the solubilities of amino acids both in pure water and in ethanol-water mixtures. From the solubility data, they obtained the Gibbs energy of transferring an amino acid residue from ethanol to water, which is then used as a scale of hydrophobicity.

We have generated a hydrophobicity scale from the binary interaction parameters estimated with the UNIFAC method. The comparison between our computed Gibbs energies of transferring amino acid residues from organic solvent to water and the known hydrophobicity scales in the literature serves the purpose to examine the reliability of the UNIFAC predictions for the binary interaction parameters.

The hydrophobicity scale in this work is formulated according to the definition of the interaction parameters:

$$\tau_{WR} = (g_{WR}-g_{RR})/RT \tag{7-28}$$

$$\tau_{RW} = (g_{RW}-g_{WW})/RT \tag{7-29}$$

$$\tau_{ER} = (g_{ER}-g_{RR})/RT \tag{7-30}$$

$$\tau_{RE} = (g_{RE}-g_{EE})/RT \tag{7-31}$$

where W, R, and E stand for water, amino acid residue, and ethanol, respectively. From these definitions, we obtain the Gibbs energy of transferring an amino acid residue from ethanol to water:

$$g_{WR} - g_{ER} = (\tau_{WR} - \tau_{ER})RT \tag{7-32}$$

Likewise, we obtain the Gibbs energy of transferring glycine residue from ethanol to water:

$$g_{WG} - g_{EG} = (\tau_{WG} - \tau_{EG})RT \tag{7-33}$$

The Gibbs energy of transferring an amino acid side chain from ethanol to water is further defined by subtracting equation 7-33 from equation 7-32:

$$\Delta g_{tr} = [(\tau_{WR} - \tau_{ER}) - (\tau_{WG} - \tau_{EG})]RT \tag{7-34}$$

Table 7-4 lists the values of the binary interaction parameters for the pairs of ethanol and the twenty amino acid residues (τ_{ER}'s and τ_{RE}'s). The values for τ_{WR}'s and τ_{RW}'s are given in Table 7-3. Table 7-5 lists the values of the Gibbs energy of transfer from various hydrophobicity scales.

In comparing various hydrophobicity scales, one should recognize the fact that they either use different organic solvents in transferring amino acids or apply different criteria for the calculations. For example, Leodidis and Hatton (1990) transferred amino acids from AOT (Aerosol OT(bis(2-ethylhexyl) sodium sulfosuccinate) interface to water. von Heijne and Blomerg (1979) considered the Gibbs energy of transfer from three individual contributions, i.e. the covering of hydrophobic surface area, the breakage of hydrogen bonds, and the charge neutralization. Eisenberg et al. (1984) normalized five different hydrophobicity scales: Nozaki and Tanford (1971), Chothia (1976), Janin (1979), von Heijne and Blomerg (1979) and Wolfenden et al. (1981). However, one thing in common among various hydrophobicity scales is that nonpolar side chains, without hydrogen bonding capacities such as phenylalanine, leucine, and alanine, require a certain amount of energy to transfer to water while polar side chains like histidine, glutamate, and arginine show less positive values or negative values in their Gibbs energy of transfer. Table 7-5 shows that, in general, the profiles of the twenty side chains on various hydrophobicity scales follow similar order. The hydrophobicity profile reported in this study is in good agreement with the other measures of the Gibbs energy of transfer, except for amino acid residues with either acidic or basic side chains for which the estimates were based on uncharged side chains.

Figures 7-5 and 7-6 demonstrate that the predicted values for the Gibbs energy of transferring amino acid side chains correlate well with the values of Nozaki and Tanford (1971) and Eisenberg et al. (1986).

Figure 7-5. Computed Gibbs Energy of Transfer of Amino Acid Side Chains in Comparison with Data of Norzaki and Tanford (1971)

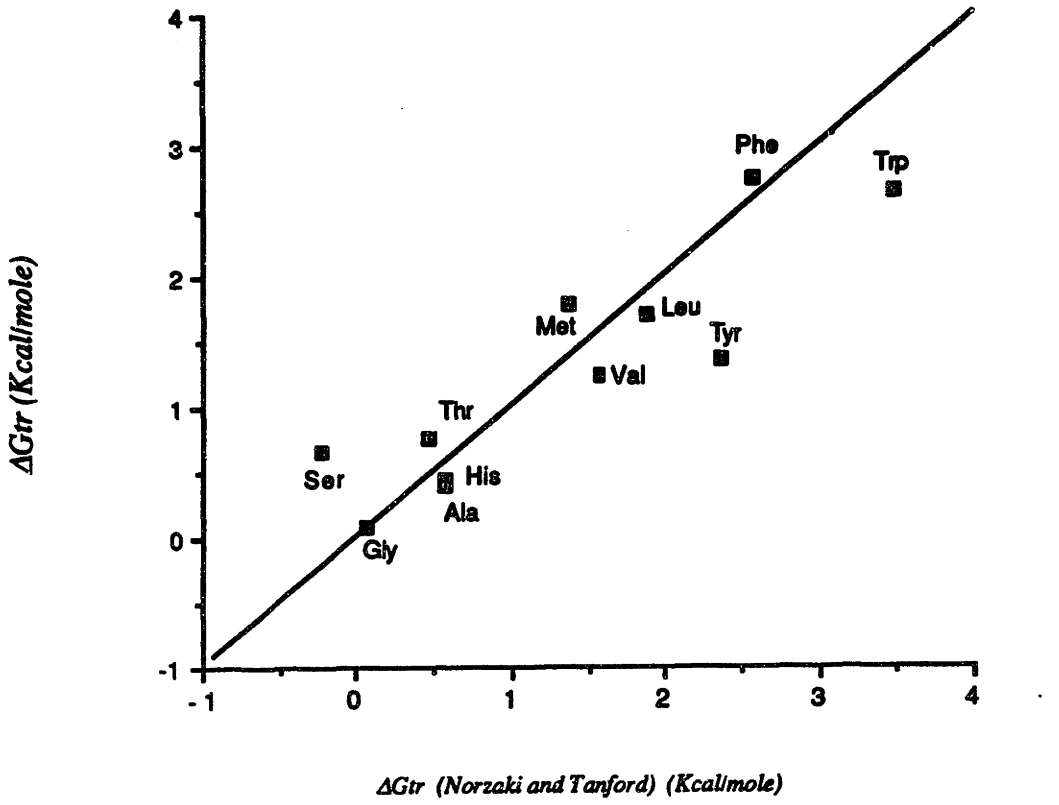


Figure 7-6. Computered Gibbs Energy of Transfer of Amino Acid Side Chains in Comparison with Values of Eisenberg et al. (1986)

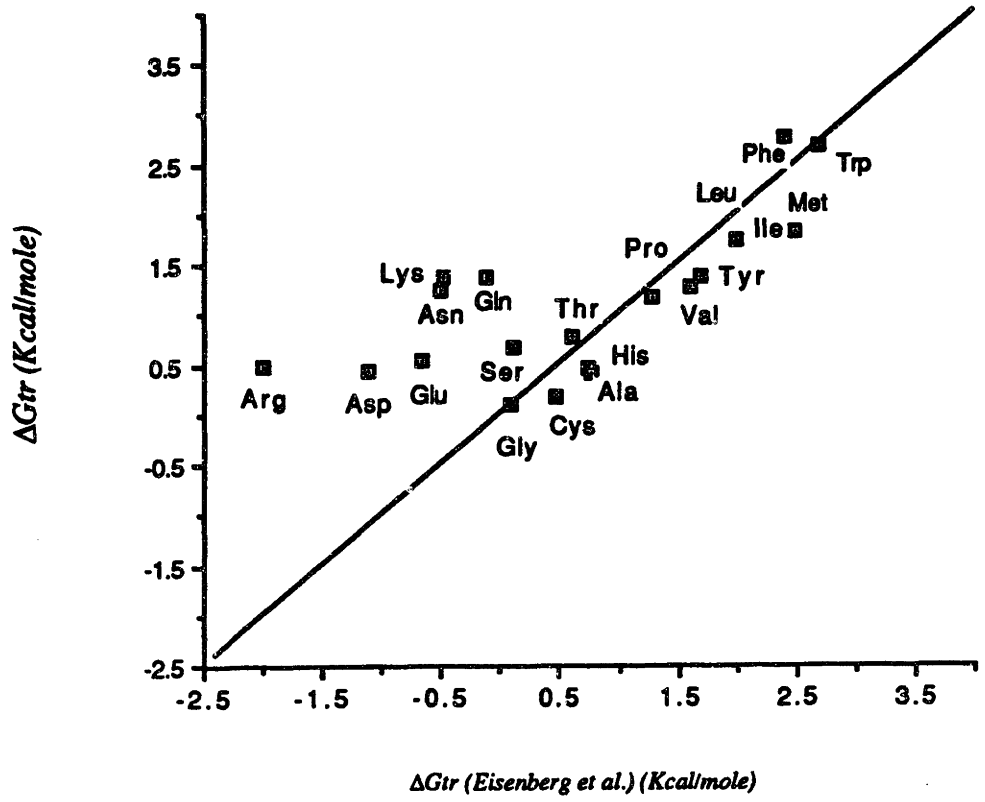


Table 7-4. Estimated Interaction Parameters, τ's, for the Pairs of Ethanol and Amino Acid Residues (at 25° C)

No.	Solvent(E)	Residue(R)	τ_{ER}	τ_{RE}
1	Ethanol	Ala	0.432	0.591
2	Ethanol	Val	0.256	0.927
3	Ethanol	Leu	0.213	1.085
4	Ethanol	Ile	0.213	1.085
5	Ethanol	Pro	0.316	0.945
6	Ethanol	Met	-0.006	1.574
7	Ethanol	Phe	0.092	2.137
8	Ethanol	Trp	-0.441	2.417
9	Ethanol	Gly	0.816	0.319
10	Ethanol	Ser	-0.087	0.814
11	Ethanol	Thr	-0.458	1.035
12	Ethanol	Cys	0.746	1.886
13	Ethanol	Tyr	-1.481	2.253
14	Ethanol	Asn	-0.407	0.938
15	Ethanol	Gln	-0.654	1.034
16	Ethanol	Asp	-0.282	1.589
17	ethanol	Glu	-0.694	-1.872
18	Ethanol	Lys	-0.727	1.145
19	Ethanol	Arg	2.695	-1.716
20	Ethanol	His	-0.707	1.349

Table 7-5. Computed Values for the Gibbs Energy of Transfer of Amino Acid Side Chains and Other Hydrophobicity Scales

Amino Acid	This Work Δg_{tr}[a]	Nozaki & Tanford Δg_{tr}[b] (1971)	Leodidis & Hatton Δg_{tr}[c] (1990)	Fauchere & Pliska Δg_{tr}[d] (1983)	Eisenberg et al. Δg_{tr}[e] (1986)	von Heijne & Blomberg Δg_{tr}[f] (1979)	(Eisenberg et al.) Δg_{tr}[g] (1984)
Phenylalanine	2.653	2.5	2.66	2.44	2.30	5.2	1.19
Tryptophan	2.558	3.4	3.35	3.07	2.60	3.9	0.81
Methionine	1.709	1.3	1.75	1.68	2.40	2.1	0.64
Leucine	1.627	1.8	1.93	2.32	1.90	4.4	1.06
Isoleucine	1.627		1.68	2.46	1.90	4.2	1.38
Glutamine	1.275			-0.30	-0.22	-0.52	-0.85
Tyrosine	1.271	2.3	1.32	1.31	1.60	3.6	0.26
Lysine	1.271			-1.35	-0.57	-2.3	-1.50
Valine	1.168	1.5	0.90	1.66	1.50	3.9	1.08
Asparagine	1.142			-0.82	-0.60	-1.0	-0.78
Proline	1.066		0.46	0.98	1.20	1.1	0.12
Threonine	0.684	0.4		0.35	0.52	1.2	-0.05
Serine	0.580	-0.3		-0.05	0.01	0.36	-0.18
Glutamic acid	0.431			-0.87	-0.76	-4.0	-0.74
Arginine	0.375			-1.37	-2.10	-9.4	-2.53
Histidine	0.350	0.5		0.18	0.64	-1.5	-0.40
Aspartic acid	0.322			-1.05	-1.20	-5.6	-0.90
Alanine	0.320	0.5		0.42	0.67	2.9	0.62
Cystine	0.083			1.34	0.38	-0.08	0.29
Glycine	0.000	0.0		0.00	0.00	1.90	0.48

a: Calculated Gibbs energy of transfer of an amino acid side chain from ethanol to water at 25°C (Kcal/mole).
b: Gibbs energy of transfer of an amino acid side chain from organic solvent (here refers to ethanol) to water (Kcal/mole).
c: Gibbs energy of transfer of amino acids from AOT interface to water at 25°C (Kcal/mole).
d: Gibbs energy of transfer for N-acetyl-amino acid amides from 1-octanol to water.
e: Gibbs energy of transfer, calculated from the data of Fauchere and Pliska (1983).
f: Theoretical Gibbs energy of transfer of the amino acid side chains from a hydrophobic phase to a hydrophilic phase (Kcal/mole).
g: The averaged values of five other hydrophobicity scales.

7.9 Predictions of Stable Conformations of Homo-polypeptides

Table 7-6 summarizes the thermodynamic parameters of the homo-polypeptides as predicted by the model. They have been favorably compared with the "experimental" values based on the "host-guest" technique of Scheraga (1978). It should be noted that Scheraga's values are not raw experimental data. Rather, they were calculated from the slope and the intercept of the Zimm-Bragg helix growth parameter, s, versus temperature plot. The parameter s was obtained by fitting helical fraction data using a statistical model of Lifson (1963) and Allegra (1967) for copolymers. Several assumptions and simplifications were introduced in their technique, such as: 1) the statistical weight of a residue being in a helical or coil conformation is independent of the chemical nature of the neighbors, 2) the helix nucleation parameter is independent of temperature.

A comparison of the model predictions with other available literature data for the thermodynamic parameters of the folding processes are listed in Table 7-7. The predictions are qualitatively consistent with the literature data.

When $\Delta g^{c\to\alpha}$ is negative, α-helix formation is favored over random coil. Likewise, when $\Delta g^{c\to\beta}$ is negative, β-sheet formation is favored. Based on the molecular thermodynamic model, our predictions for the secondary structure of homo-polypeptides match well with the experimental observations reported by previous researchers using infrared dichroism, ultraviolet adsorption, titration, and optical rotary techniques. It is found that polypeptide chains of hydrophobic residues favor folding of the chain while

polypeptide chains of hydrophilic residues favor the unfolded conformation. Furthermore, α-helix represents a more favorable conformational state than β-sheet does.

The documented conformations of homo-polypeptides in aqueous solutions along with the model predictions are listed in Table 7-8.

Alanine has long been recognized as a strong helix former (Alter et al., 1972). With no steric hindrance, its aliphatic side chain is ideally suited to the helical structure. Our result is consistent with the work of Richardson and Richardson (1989).

Valine has a nonpolar side chain. The hydrophobic interactions between valine side chains in poly-valine are stronger than those in the case of alanine. The model predicts an α-helical structure for poly-valine. According to the copolymer work of Go and Scheraga (1984), poly-valine does form helical structure at higher temperature.

Leucine is also a strong helix maker. It has a very large Zimm-Bragg helix growth parameter, s, among amino acid residues (Chou and Fasman, 1972). Our model predicts a stable helix conformation of poly-leucine in aqueous solution which was reported by Blout et al. (1960).

Although isoleucine has a branched short side chain, the model predicts that poly-isoleucine forms α-helix at 25°C, which is consistent with the copolymer experimental work of Go and Scheraga (1984), but contrary to the earlier work of Blout and Shechter (1963).

Though methionine residue is mostly found in β-sheets (Chou and Fasman, 1972, 1974a), its homo-polypeptide forms an α-helix as observed by Blout et al. (1960). Our model also predicts the helical structure for poly-methionine.

The lack of hydrophobicity makes the homo-polypeptide chain of serine stay as a random coil. Our model predicts a random coil structure for poly-serine as experimentally observed by Blout et al. (1960).

Our calculation indicates that aspartic acid residue is hydrophilic. As a result, poly-aspartic acid forms random coils in aqueous solutions, which is in agreement with the conclusion of Olander and Holtzer (1968).

Like poly-leucine, poly-glutamic acid forms a very stable helix. The prediction of the model agrees well with the data of Olander and Holtzer (1968) and Bychkova et al. (1971).

The model predicts a helical conformation for uncharged poly-lysine, which is consistent with the study of Davidson and Fasman (1967). But more study is needed on the transition between the helical and sheet structures of poly-lysine as reported intensively by Hermans (1966), and Scheraga's group.

7.10 Estimated Zimm-Bragg Parameters

There are two parameters associated with the Zimm-Bragg model (Zimm and Bragg, 1959): the helix growth parameter, s, and the nucleation parameter σ. The model

provides a facility to estimate the Zimm-Bragg helix growth parameter s and the nucleation parameter σ for amino acid residues.

The s parameter represents a measure for the intrinsic α-helicity of an amino acid residue, defined originally as the equilibrium constant for the formation of a hydrogen bond in α-helices. Zimm and Bragg assumed that hydrogen bonding between amide groups was the sole driving force of α-helix formation. It is now recognized that although hydrogen bonding plays important roles in the formation and the stabilization of the polypeptide conformation, other types of forces, such as the side chain interactions, and the interactions of the chain with solvent and buffer molecules, also make significant contributions. Therefore, the experimentally determined s parameter is not simply the equilibrium constant for the formation of a hydrogen bond, but the equilibrium constant for the conversion of an amino acid residue from its coiled state to the helical state in a specific solvent environment.

As an equilibrium constant, the Zimm-Bragg helix growth parameter is related to the thermodynamic quantity $\Delta g^{c\to\alpha}$, the Gibbs energy of conversion of a coil residue to a helical one, via equation 7-35 (Sueki et al., 1984):

$$\begin{aligned} s &= \exp(-\Delta g^{c\to\alpha}/RT) \\ &= \exp(-\Delta h^{c\to\alpha}/RT + \Delta s^{c\to\alpha}/R) \end{aligned} \tag{7-35}$$

The σ parameter measures the difficulty in initiating a helix. There exists a large entropy loss upon bending a coil residue to form the first helical "element." This entropy loss must be adequately compensated by favorable side-chain contacts, an enthalpy contribution, when the helical turns are formed. Whether this enthalpy contribution to the Gibbs energy can sufficiently overcome the entropy loss of these residues depends on the nature of the residue-residue interactions.

According to Zimm and Bragg (1959), a helical sequence with j helical residues is characterized by a statistical weight of σs^j. For the same helical sequence, the Gibbs energy change for the conversion of the j coil residues to a helical sequence is:

$$\Delta G^{c\to\alpha}(j) = (j-4)\Delta h^{c\to\alpha} - jT\Delta s^{c\to\alpha} \qquad (7\text{-}36)$$

Here, the number 4 in (j-4) represents the deficit in the helical hydrogen bonds and other side-chain interactions at the two ends of the j helical sequence. According to von Dreele et al. (1971), the relationship between the Zimm-Bragg parameters and the Gibbs energy of formation of a helical sequence of j residues from the coil state is:

$$\Delta G^{c\to\alpha}(j)/RT = -\ln(\sigma s^j) \qquad (7\text{-}37)$$

Applying equations 7-35 and 7-36 to equation 7-37, we can obtain the nucleation parameter σ:

$$\sigma = \exp(4\Delta h^{c\to\alpha}/RT) \tag{7-38}$$

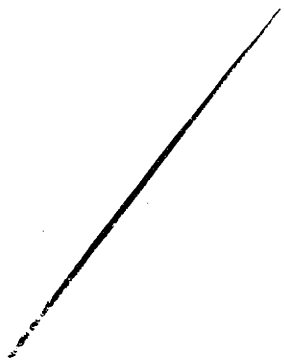


Table 7-6. Predicted Thermodynamic Parameters for the Secondary Structures of Homo-polypeptides per Residue

Peptide	This work (25°C)					Scheraga et al.'s work (20°C)						This work (25°C)		
	$\Delta h^{c\to\alpha}/RT$	$\Delta s^{c\to\alpha}/R$	$\Delta g^{c\to\alpha}/RT$	s	$\sigma x10^4$	$\Delta h^{c\to\alpha}/RT$	$\Delta s^{c\to\alpha}/R$	$\Delta g^{c\to\alpha}/RT$	s	$\sigma x10^4$	ref.	$\Delta h^{c\to\beta}/RT$	$\Delta s^{c\to\beta}/R$	$\Delta g^{c\to\beta}/RT$
Poly-ala	-0.890	-0.798	-0.092	1.10	284.	-0.415	-0.354	-0.069	1.07	8.0	a	-0.446	-0.420	-0.026
Poly-val	-1.649	-0.798	-0.851	2.34	14.	1.098	1.032	0.069	0.95	1.0@	b	-0.826	-0.420	-0.406
Poly-leu	-2.056	-0.798	-1.258	3.52	2.7	-0.256	-0.134	-0.130	1.14	33.	c	-1.030	-0.420	-0.610
Poly-ile	-2.056	-0.798	-1.258	3.52	2.7	-0.980	-0.846	-0.130	1.14	55.	d	-1.030	-0.420	-0.610
Poly-pro	-1.583	-0.798	-0.785	2.19	17.	3.392	1.661	1.732	0.27	0.1@	e	-0.793	-0.420	-0.373
Poly-met	-1.348	-0.798	-0.550	1.73	46.	-1.401	-1.208	-0.182	1.20	54.	f	-0.675	-0.420	-0.255
Poly-phe	-3.390	-0.798	-2.591	13.35	0.01	-0.292	-0.232	-0.076	1.09	18.	g	-1.698	-0.420	-1.278
Poly-trp	-2.668	-0.798	-1.870	6.49	0.23	-0.776	-0.674	-0.101	1.11	77.	h	-1.337	-0.420	-0.917
Poly-gly	-0.557	-0.798	0.241	0.79		1.073	0.503	0.561	0.59	0.1@	i	-0.279	-0.420	0.141
Poly-ser	-0.529	-0.798	0.269	0.76		-0.173	-0.453	0.271	0.76	0.1@	j	-0.265	-0.420	0.155
Poly-thr	-0.719	-0.798	0.079	0.92		0.774	0.574	0.203	0.82	0.1@	k	-0.360	-0.420	0.060
Poly-cys	-0.810	-0.798	-0.012	1.01	392.	-0.362	-0.393	0.033	0.99	1.0@	l	-0.406	-0.420	0.014
Poly-tyr	-1.645	-0.798	-0.847	2.33	14.	-1.597	-1.560	-0.024	1.02	66.	m	-0.824	-0.420	-0.404
Poly-asn	-0.603	-0.798	0.195	0.82		0.687	0.443	0.244	0.78	0.1@	n	-0.302	-0.420	0.118
Poly-gln	-0.753	-0.798	0.046	0.96		-0.846	-0.856	0.022	0.98	33.	o	-0.377	-0.420	0.043
Poly-asp	-0.716	-0.798	0.082	0.92		-1.760	-2.003	0.242*	0.78	210.	p	-0.359	-0.420	0.061
Poly-glu	-0.913	-0.798	-0.115	1.12	260.	-1.837	-1.560	-0.299*	1.35	100	q	-0.457	-0.420	-0.037
Poly-lys	-0.970	-0.798	-0.172	1.19	207.	0.567	0.503	0.062*	0.94&	1.0@,&	r	-0.486	-0.420	-0.066
Poly-arg	0.435	-0.798	1.233	0.29		-0.280	-0.252	-0.033*	1.03&	0.1@,&	s	0.218	-0.420	0.638
Poly-his	-0.438	-0.798	0.361	0.70		-2.850	-3.020	0.161*	0.85	210	t	-0.219	-0.420	0.201

@: Arbitrarily chosen (fixed) as the best σ at which both s and the least square deviation in θ becomes independent of σ.

*: Poly-asp, at pH 1.5 (asp is uncharged)
Poly-glu, at pH 2.3 (glu is uncharged)
Poly-lys, at neutral pH (lys is positively charged)
Poly-arg, at neutral pH (arg is positively charged)
Poly-his, at pH 9.0 (his is uncharged)

&: Charged side chains

a: Platzer et al. (1972).
b: Alter et al. (1973).
c: Alter et al. (1972).
d: Fredrickson et al. (1961).
e: Altmann et al. (1990).
f: Hill et al. (1977).
g: van Wart et al. (1973).
h: Nagy et al. (1980).
i: Ananthanarayanan et al. (1971).
j: Hughes et al. (1972).
k: Hecht et al. (1978).
l: Wojcik et al. (1990).
m: Scheule et al. (1976).
n: Matheson et al. (1977).
o: Denton et al. (1982).
p: Kobayashi et al. (1977).
q: Maxfield et al. (1975).
r: Dygert et al. (1976).
s: Konishi et al. (1977).
t: Sueki et al. (1984).

Table 7-7. Comparison between Predicted and Observed Thermodynamic Parameters for the Secondary Structures of Homo-polypeptides

Peptide	$\Delta h^{c\to\alpha}/RT$		$\Delta s^{c\to\alpha}/R$		$\Delta g^{c\to\alpha}/RT$		Ref
	Model*	Data	Model	Data	Model	Data	
Poly-alanine	-0.890	-0.32	-0.798	-0.28	-0.092	-0.04	a
Poly-glutamic acid	-0.963	-1.65	-0.798	-1.34	-0.115	-0.31	b
Poly-lysine	-0.970	-1.49	-0.798	-1.36	-0.172	-0.13	c

*: Model prediction is carried out at 25 °C.
a: Ingwall et al. (1968)
b: For uncharged side chain of glutamic acid at 25 °C, Olander and Holtzer (1968)
c: For uncharged side chain of lysine at 25 °C, Hermans (1966)

Table 7-8. Comparison between Predicted and Observed Conformations of Homo-polypeptides in Aqueous Solutions

Peptide	Conformation		Reference
	Model*	Literature	
Poly-alanine	α	α	a
Poly-valine	α	α	b,c
Poly-leucine	α	α	c,c
Poly-isoleucine	α	α	b
Poly-methionine	α	α	c
Poly-serine	c	c	c
Poly-aspartic acid**	c	c	d
Poly-glutamic acid**	α	α	d, e
Poly-lysine**	α	α	f, g

*: Model prediction is carried out at 25 °C.
**: Both the data and the calculations are for uncharged side chains of aspartic acid,

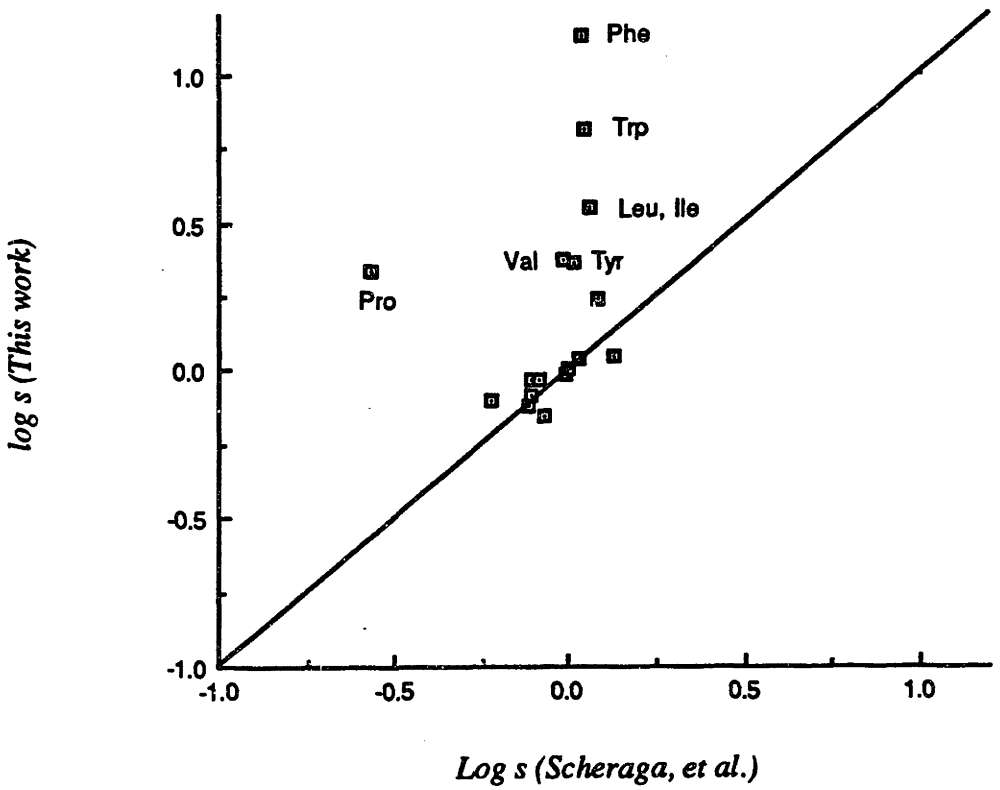
glutamic acid and lysine.

a: Richardson and Richardson (1989)
b: 30 °C, Go and Scheraga (1984).
c: Blout et al. (1960)
d: Blout and Shechter (1963)
e: 25 °C, Olander and Holtzer (1968)
f: 25 °C, Bychkova et al. (1971)
g: 25 °C, Hermans (1966)
h: Davidson and Fasman (1967)

Using their data on random copolymers of hydroxy-propyl-glutamine with one of the twenty amino acids to avoid solubility problems of the amino acid homopolymer, Scheraga's group has calculated the s values for the twenty residues in water in a series of studies as summarized by Sueki et al. (1984). Table 7-6 lists the s and σ parameters as predicted by the molecular thermodynamic model and those of Scheraga et al. Figure 7-7 shows the comparison of the values for the helix growth parameter. Note that, for those residues with the s value less than unity, the predicted values for σ are not listed here since they are not favored for helical conformation.

It is encouraging that, given the uncertainty nature of the s and σ parameters, our predicted values for s and σ fall in the same order of magnitude with the quantities suggested by Zimm and Bragg.

Figure 7-7. Predicted Helix Growth Parameter vs. the Values of Scheraga et al. (1984)



7.11 Effect of Chain Length

Polypeptides may assume different conformations at different chain lengths (Schellman, 1955). As the chain length increases, both the entropy and the enthalpy terms scaled on a per residue basis change. After the chain length becomes fairly long (say, 100 residues), these thermodynamic parameters would remain constant. In other words, the "end effect" diminishes as the chain length increases.

Figures 7-8 and 7-9 show the model-predicted values for the thermodynamic parameters in folding poly-alanine into α-helices and β-sheets as a function of chain length x. It is clear that the stability of helical structures will increase with the chain length. In other words, the Δg of folding becomes more negative as the chain length increases. This effect diminishes as the chain length reaches a value of about 100. The potential per residue to form a sheet structure may increase or decrease with the chain length depending on the interaction strength of the amino acid residues with solvent water. In the case of poly-alanine, the potential of forming a sheet may decrease with the chain length since the entropy contribution to the Gibbs energy of folding may increase faster than the enthalpy contribution.

Figure 7-8 also shows that the model predicts a critical chain length for the helix-coil transition as suggested by Schellman (1955). If the chain length is below the critical length, the attractive interaction between the residues may not be sufficient to offset the entropy loss upon folding the residues to a helix. The more hydrophobic the amino acid side chain is, the shorter the critical chain length becomes. For example, according to our calculation, the critical length for poly-alanine is 15 residues long. In other words, a poly-alanine molecule with the degree of polymerization (DP) less than 15 will not fold

into α-helical conformation. This result is consistent with the experimental work of Ingwall et al. (1968).

Figure 7-8. Predicted Thermodynamic Parameters of α-helix Formation of Poly-alanine as a Function of Chain Length

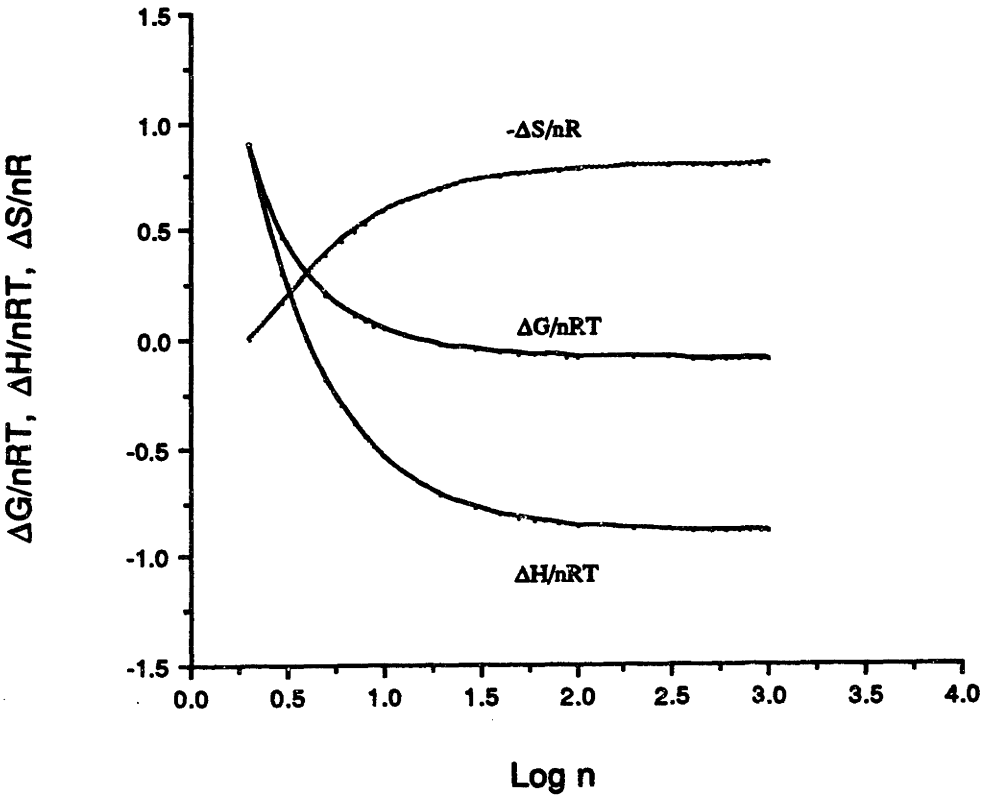
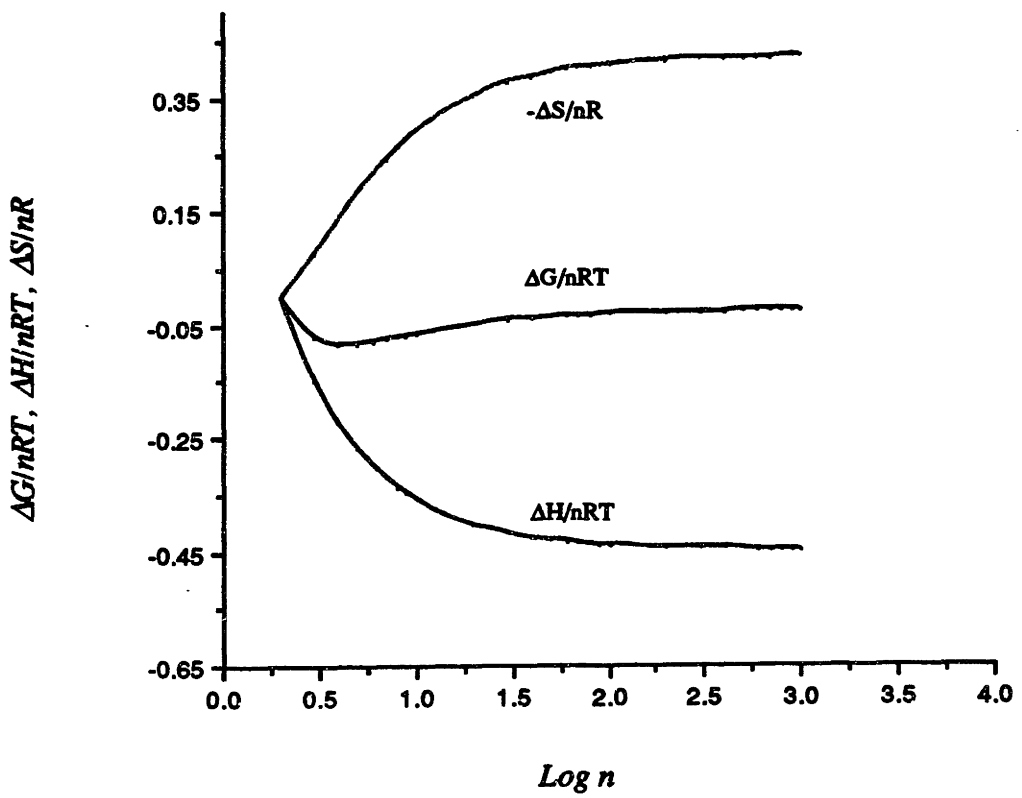


Figure 7-9. Predicted Thermodynamic Parameters of β-sheet Formation of Poly-alanine as a Function of Chain Length



Notation

ΔG	=	Gibbs energy
Δg	=	molal Gibbs energy
ΔH	=	enthalpy
Δh	=	molal enthalpy
n	=	number of residues
N_1	=	number of solvent molecules.
N_2	=	number of polymer chain molecules of x segments.
R	=	gas constant, residues
s	=	Zimm-Bragg helix growth parameter
ΔS	=	entropy
Δs	=	molal entropy
T	=	temperature, K
u	=	interaction energy
x	=	number of segments per polymeric chain

Greek Letters

σ	=	Zimm-Bragg nucleation parameter
τ	=	NRTL binary interaction energy parameter

Superscripts

α	=	α-helix
β	=	β-sheet
c	=	coil
FH	=	Flory-Huggins term
s	=	solvated state
w	=	pure water state

Subscripts

E	=	ethanol
G	=	glycine
i,j	=	any species
R	=	peptide unit, or "residue"
S	=	side chain
tr	=	transfer
W	=	water

Literature Cited

Allegra, G., J. Poly. Sci., Part C., 16, 2815 (1967).

Alter, J.E., T. Taylor, and H.A. Scheraga, Macromolecules, 5(6), 739-746 (1972).

Alter, J.E., R.H. Andereatta, T. Taylor, and H.A. Scheraga, Macromolecules, 6 (4), 564-570 (1973).

Altmann, K.-H., J. Wojcik, M. Vasquez, and H.A. Scheraga, Biopolymers, 30, 107-120 (1990).

Ananthanarayanan, V.S., R.H. Andreatta, D. Poland, and H.A. Scheraga, Macromolecules, 4 (4), 417-424 (1971).

Anfinsen, C.B., Science, 181, 223-230 (1973).

Aspen Technology, ASPEN PLUS User Guide (1988).

Ben-Naim, A., Biopolymers, 29, 567-596 (1990).

Blankschtein, D., G.M. Thurston, and G.B. Benedek, J. Chem. Phys., 85(12), 7268-7288 (1985).

Blout, E.R., C. de Loze, S.M. Bloom, and G.D. Fasman, J. Amer. Chem. Soc., 82, 3787-3789 (1960).

Blout, E.R. and E. Shechter, Biopolymers, 1, 568 (1963).

Bychkova, V.E., O.B. Ptitsyn, and T.V. Barskaya, Biopolymers, 10, 2161-2179 (1971).

Chan, H.S. and K.A. Dill, Macromolecules, 22, 4559-4573 (1989).

Chen, C.-C., J.F. Boston, H.I. Britt, and L.B. Evans, AIChE J., 23 (4), 588-596 (1982).

Chen, C.-C. and L.B. Evans, AIChE J., 32 (3), 444-454 (1986).

Chen, C.-C., Y. Zhu, and L.B. Evans, Biotechnology Progress, 5 (3), 111-118 (1989).

Chen, C.-C., "A Molecular Thermodynamic Model for the Gibbs Energy of Mixing of Micellar Solutions," paper presented at the Annual AIChE meeting, San Francisco, 1989.

Chothia, C., J. Mol. Biol., 105 (1), 1-12 (1976).

Chou, P.Y. and G.D. Fasman, Biochemistry, 11, 3028-3043 (1972).

Chou, P.Y. and G.D. Fasman, Biochemistry, 13, 211-222 (1974a).

Chou, P.Y. and G.D. Fasman, Biochemistry, 13, 222-245 (1974b).

Chou, P.Y. and G.D. Fasman, Adv. Enzymol., 47, 45-148 (1978).

Davidson, B. and G.D. Fasman, Biochemistry, 6, 1616 (1967).

Davies, D.R., J. Mol. Biol., 9, 605-609 (1964).

Denton, J.B., S.P. Powers, B.O. Zweifel, and H.A. Scheraga, Biopolymers, 21, 51-77 (1982).

Dill, K.A., Biochemistry, 24, 1501-1509 (1985).

Dygert, M.K., G.T. Taylor, F. Cardinaux, and H.A. Scheraga, Macromolecules, 9, 794 (1976).

Eisenberg, D., E. Schwarz, M. Komaromy, and W. Wilcox, J. Mol. Biol., 179 (1), 125-142 (1984).

Eisenberg, D., W. Wilcox, and A.D. McLachlan, J. Cell Biochem., 31 (1), 11-17 (1986).

Fauchere, J.L. and V. Pliska, Eur. J. Med. Chem.-Chim. Ther., 18 (4), 369-375 (1983).

Flory, P.J., J. Chem. Phys., 9, 660 (1941); 10, 51 (1942).

Fredenslund, A., R.L. Jones, and J.M. Prausnitz, AIChE J., 21, 1086 (1975).

Fredenslund, A., J. Gmehling, and P. Rasmussen, Vapor-Liquid Equilibrium Using UNIFAC, Elsevier Scientific Publishing Company, Amsterdam (1977).

Fredrickson, R.A., M.C. Chang, S.P. Powers, and H.A. Scheraga, Macromolecules, 14, 625-632 (1981).

Garnier, J., D.J. Osgusthorpe, and B. Robson, J. Mol. Biol., 120, 97-120 (1978).

Go, M., and H.A. Scheraga, Biopolymers, 23, 1961-1977 (1984).

Hecht, M.H., B.O. Zweifel, and H.A. Scheraga, Macromolecules, 11 (3), 545-551 (1978).

Hermans, J. Jr., J. Phys. Chem., 70, 510 (1966).

Hill, D.J.T., F. Cardinaux, and H.A. Scheraga, Biopolymers, 16, 2447-2467 (1977).

Huggins, M.L., J. Phys. Chem., 9, 440 (1941).

Huggins, M.L., Ann. N.Y. Acad. Sci., 43, 1 (1942).

Hughes, L.J., R.H. Andreatta, and H.A. Scheraga, Macromolecules, 5 (2), 187-197 (1972).

Ingwall, R.T., H.A. Scheraga, N. Lotan, A. Berger, and E. Katchalski, Biopolymers, 6, 331 (1968).

Janin, J., Nature, 277 (5696), 491-492 (1979).

Kim, P.S. and R.L. Baldwin, Annu. Rev. Biochem., 59, 631-60 (1990).

King, J., Chemical & Engineering News, 67 (15), 32-54 (1989).

Kobayashi, Y., F. Cardinaux, B.O. Zweifel, and H.A. Scheraga, Macromolecules, 10 (6), 1271-1283 (1977).

Konishi, Y., J.W. Van Nispen, G.D. Davenport, and H.A. Scheraga, Macromolecules, 10 (6), 1264-1271 (1977).

Kotelchuck, D., M. Dygert, and H.A. Scheraga, Proc. Nat. Acad. Sci. USA, 63, 615 (1969).

Leodidis, E.B., and T.A. Hatton, J. Phys. Chem., 94 (16), 6411-6420 (1990).

Lifson, S., Biopolymers, 1, 25 (1963).

Lim, V.I., J. Mol. Biol., 88, 857-872 (1974).

Maxfield, F.R., J.E. Alter, G.T. Taylor, and H.A. Scheraga, Macromolecules, 8 (4), 479-491 (1975).

Matheson, R.R. Jr., R.A. Nemenoff, F. Cardinaux, and H.A. Scheraga, Biopolymers, 16, 1567-1586 (1977).

McGregor, M.J., S.A. Islam and M.J.E. Sternberg, J. Mol. Biol., 198, 295-310 (1987).

Mock, B., L.B. Evans, and C.-C. Chen, AIChE J., 32 (10), 1655-1664 (1986).

Nagy, J.A., S.P. Powers, B.O. Zweifel, and H.A. Scheraga, Macromolecules, 13 (6), 1428-1440 (1980).

Nozaki, Y. and C. Tanford, J. Biol. Chem., 46 (7), 2211-2217 (1971).

Olander, D.S. and A. Holtzer, J. American Chem. Soc., 90 (17), 4549-4560 (1968).

Pauling, L., R.B. Corey, and H.R. Branson, Proc. Natl. Acad. Sci. USA, 37, 205 (1951).

Platzer, K.E.B., V.S. Ananthanarayanan, R.H. Andreatta, and H.A. Scheraga, Macromolecules, 5 (2), 177-187 (1972).

Poland D. and H.A. Scheraga, Theory of Helix-Coil Transitions in Biopolymers, Academic Press, New York (1970).

Prausnitz, J.M., Preface in "Vapor-Liquid Equilibrium Using UNIFAC" by Fredenslund, A., J. Gmehling, and P. Rasmussen, Elsevier Scientific Publishing Company, Amsterdam, (1977).

Prausnitz, J.P., Science, 205 (24), 759-766 (1979).

Prausnitz, J.M., R.N. Lichtenthaler and E. G. de Azevedo, "Molecular Thermodynamics of Fluid-Phase Equilibria," p. 294, Prentice-Hall, Inc., Englewood Cliffs, NJ (1986).

Renon, H. and J.M. Prausnitz, AIChE J., 14 (1), 135 (1968).

Richardson, J.S. and D.C. Richardson, "Prediction of Protein Structure and the Principles of Protein Conformation", pp1-98, Edited by Fasman, G.D., Plenum Press, New York (1989).

Schellman, J.A., Compt. Rent. Trav. Lab. Carlsburg, Ser. Chim., 29 (14-15), 230-259 (1955).

Scheraga, H.A., Pure & Appl. Chem., 50, 313-324 (1978).

Scheule, R.K., F. Gardinaux, G.T. Taylor, and H.A. Scheraga, Macromolecules, 9 (1), 23-33 (1976).

Schulz, G.E., M. Elzinga, and R.H. Schirmer, Three Dimentional Structure of Adenyl Kinase, Nature, 250, 120-123 (1974a).

Schulz, G.E., C.D. Barry, J. Friedman, P.Y. Chou, G.D. Fasman, A.V. Finkelstein, V.I. Lim, O.B. Ptitsyn, E.A. Kabat, T.T. Wu, M. Levitt, B. Robson, and K. Nagano, Nature, 250, 140-142 (1974b).

Schulz, G.E., and R.H. Schirmer, Principles of Protein Structure, Springer-Verlag, New York (1979).

Sueki, M., S. Lee, S.P. Powers, J.B. Denton, Y. Konishi, and H.A. Scheraga, Macromolecules, 17, 148-155 (1984).

van Wart, R.E., Taylor, G.T., and H.A. Scheraga, Macromolecules, 6 (2), 266-273 (1973).

von Dreele, P.H., D. Poland and H.A. Scheraga, Macromolecules, 4 (4), 390-407 (1971).

von Heijne, G. and Blomberg, C., Eur. J. Biochem., 97, 175-81 (1979).

von Heijne, G., Sequence Analysis in Molecular Biology, Academic Press, Inc., San Diego (1987).

Wojcik, J., K.-H. Altmann, and H.A. Scheraga, Biopolymers, 30, 121-134 (1990).

Wolfenden, R., L. Anderson, P.M. Culis, and C.C.B. Southgate, Biochemistry, 20, 849-855 (1981).

Zhu, Y., L.B. Evans, and C.C. Chen, Biotechnology Progress, 6, 266-272 (1990).

Zimm, B.H. and J.K. Bragg, J. Chem. Phys., 31 (2), 526-535 (1959).

CHAPTER 8

MOLECULAR THERMODYNAMIC MODELING OF POLYPEPTIDE CHAIN FOLDING

Introduction

The α-helix is the most abundant unit of secondary structure in proteins. α-helix forms spontaneously under favorable environmental conditions and it can be unfolded by increasing temperature, GuHCl, or urea (Kim and Baldwin, 1984). Though significant progress has been made in understanding the factors stabilizing homopolypeptide α-helices (Scheraga, 1978; Sueki, et al., 1984; Davidson and Fasman, 1967), not much work was done in the studies on synthetic peptide (copolymers of the twenty amino acid residues) and natural peptide conformations until the recent efforts of Kim and Baldwin (1984), Marqusee, et al. (1987, 1989), Padmanabhan, et al. (1990), and others.

Molecular thermodynamics has been well-established as a practical method to derive semi-empirical expressions for the excess Gibbs energy of mixtures of small molecules. In Chapter 7, we have successfully applied the molecular thermodynamic approach and developed a model for the Gibbs energy of folding of aqueous homopolypeptides. The resulting model has shown to generate critical results that are consistent with the critical observations reported in the literature for the folding of the aqueous homopolypeptides. In this Chapter we generalize the molecular thermodynamic model for aqueous homopolypeptides as a molecular thermodynamic model for the free

energy of folding of polypeptides from coiled conformation into an α-helical conformation.

8.1 The Molecular Thermodynamic Model

As stated in Chapter 7, the energetics of the folding of residues can be represented as the following two individual contributions: the enthalpy contribution and the entropy contribution.

$$\Delta G = \Delta H - T\Delta S \tag{8-1}$$

The Gibbs energies of folding residues in a polypeptide chain from a coiled state into an α-helix is the Gibbs energy difference between state α and the random-coiled state c.

$$\Delta G^{c\rightarrow\alpha} = \Delta H^{c\rightarrow\alpha} - T\Delta S^{c\rightarrow\alpha} \tag{8-2}$$

The Gibbs energy of folding is subject to the entropy change in folding residues into a specific polypeptide conformation and the physical interactions among amino acid residues, solvents, and all other species present in the local lattice structure at a set of given conditions (temperature, pH, ionic strength, solvent, etc.). For convenience, we chose the system to be one single polypeptide chain with the amount of solvent molecules necessary to "solvate" all the amino acid residues on the random-coiled chain. In other

words, the system consists of the polypeptide chain and the nearest neighboring solvent molecules that constitute the hydration shell. A typical coordination number of six is assumed in this treatment (Prausnitz et al., 1986). Based on the experience with molecular thermodynamic models, we believe the choice on the coordination number does not significantly impact the results.

The reference states are chosen to be the pure liquid for water and a hypothetical amino acid residue aggregate state for residues, or the residue in pure homo-polypeptides. In this aggregate state for residues, the residues are surrounded by residues of the same amino acid. The reference state for water is characterized by the water-water physical interactions. The reference state for residues is characterized by the residue-residue physical interactions.

It should be noted here that to be more representative for the interactions involving peptides, the segment along the chain is treated as a peptide unit -CH(R_i)-C(O)-NH-, instead of -NH-CH(R_i)-C(O)-. In the context of this work, therefore, a "residue" stands for a peptide unit.

8.2 The Enthalpy Contribution

We generalize the definition for the interaction energies for a residue i in the states random coil and α-helix as:

$$g_{R(i)}^{c} = [\, g_{R(i-1),R(i)} + g_{R(i+1),R(i)} + 4\, g_{W,R(i)}]/z \qquad (8\text{-}3)$$

$$g_{R(i)}^{\alpha} = [\, g_{R(i-3),R(i)} + g_{R(i-1),R(i)} + g_{R(i+1),R(i)} + g_{R(i+3),R(i)} + 2\, g_{W,R(i)}]/z \qquad (8\text{-}4)$$

Here, $g_{R(j),R(i)}$'s represent the interaction energy between jth and ith residues along the peptide chain, $g_{W,R(i)}$ is the interaction energy between a water molecule and the ith amino acid residue, and z is the coordination number. This concept of the contact energy for a central helical residue in the helix (or, a residue located in the middle portion of an α-helix, labeled as i= 4 through N_h-3) represented in equation (8-4) is illustrated in Figure 8-1.

It is important, according to the illustration in Figure 8-2, to take into consideration the end effect of the helix portion due to the deficit of hydrogen bonds and other types of interactions there. For instance, there always exist three residues at the N terminus of a helix where their amino groups are not hydrogen-bonded, and three residues at the C-terminus of the helix in deficit of hydrogen-bonding with their carboxyl groups.

For a residue in positions 2 and 3 at the N-terminal end of a helical sequence in a polypeptide chain:

$$g_{R(i)}^{\alpha}\,|_{\text{N-terminal}} = [\, g_{R(i-1),R(i)} + g_{R(i+1),R(i)} + g_{R(i+3),R(i)} + 3\, g_{W,R(i)}]/z \qquad (8\text{-}5)$$

Figure 8-1. First Neighbor Interactions for a Residue in the Middle Portion of an α-helical Sequence

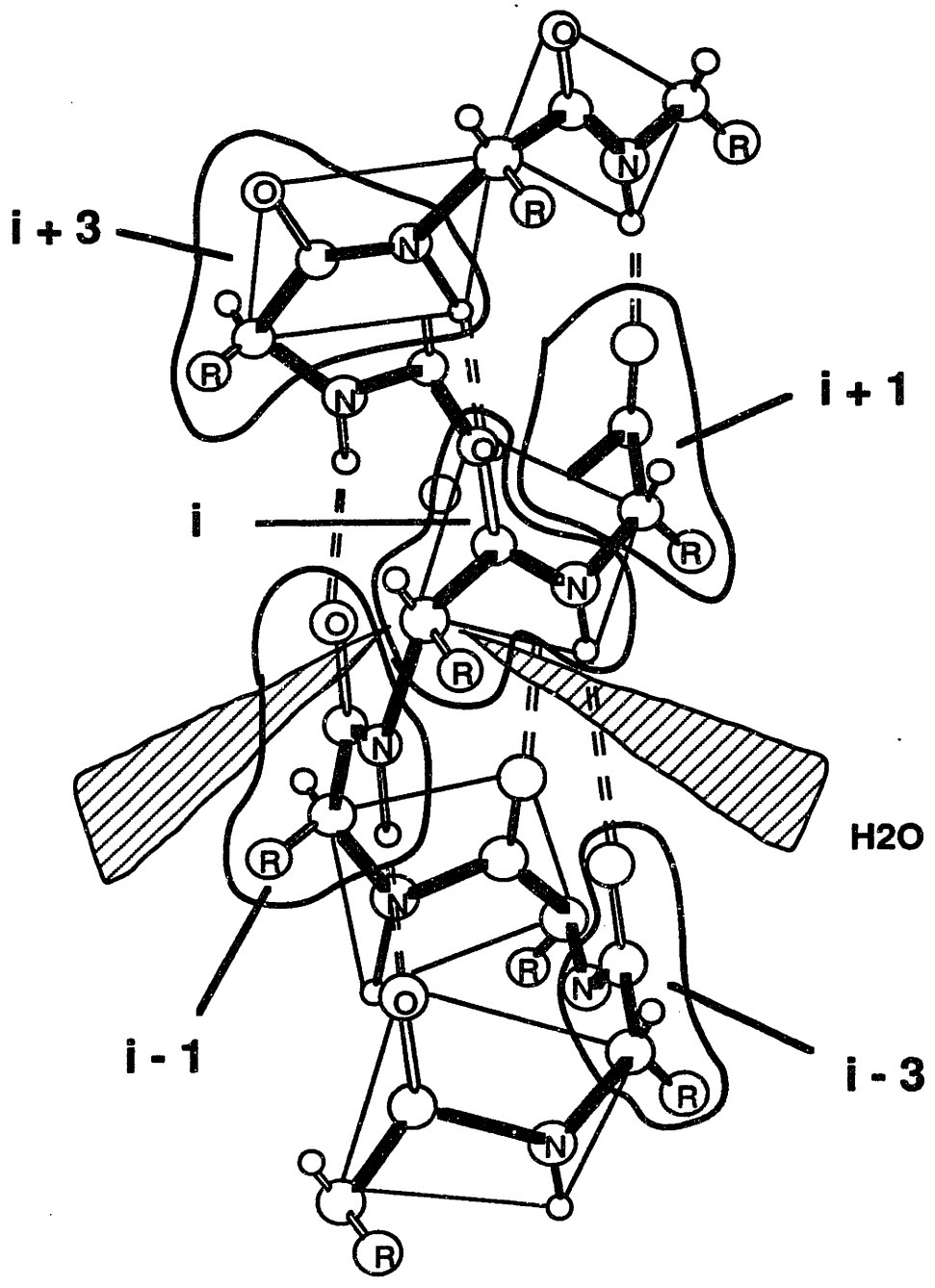
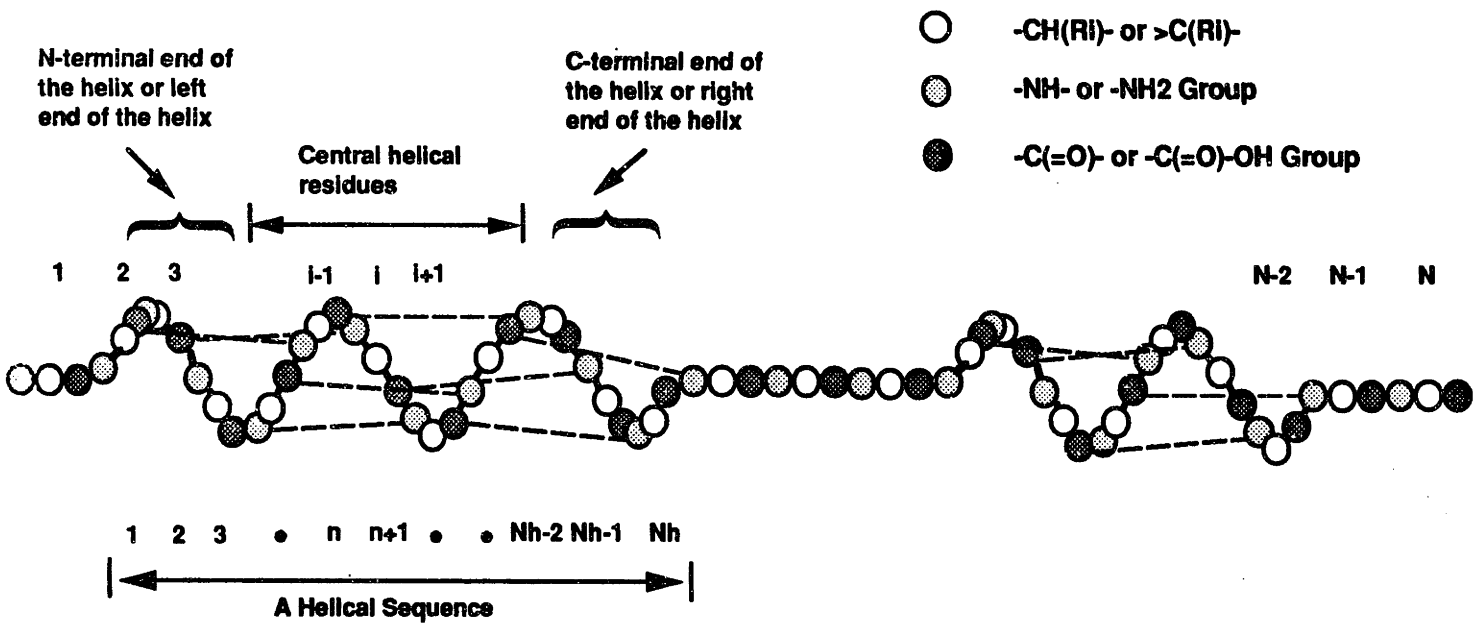


Figure 8-2. Definition of the α-helix Conformation of a Polypeptide Chain



and a residue in positions N_h-2 and N_h-1 at the C-terminal end of a helical sequence:

$$g_{R(i)}^{\alpha}\,|_{\text{C-terminal}} = [\, g_{R(i-3),R(i)} + g_{R(i-1),R(i)} + g_{R(i+1),R(i)} + 3\, g_{W,R(i)}]/z \qquad (8\text{-}6)$$

Set $g_{R(i-1),R(i)}$ term equal to zero and add another $g_{W,R(i)}$ term, equations (8-3) and (8-5) will hold for helical residue number 1 at the N-terminus of the helical sequence. While substitute a term $g_{W,R(i)}$ for $g_{R(i+1),R(i)}$, equations (8-3) and (8-6) will be valid for residue number N_h at the C-terminus of the sequence.

The interaction energies per water molecule in a solvated state (hydrogen-bonded to amide groups or the side chain groups of the polypeptide molecule) and a free state (free water beyond the hydration shell of the chain) are:

$$g_{R(i),w}^{s} = [\, g_{R(i),w} + 5\, g_{W,W}]/z \qquad (8\text{-}7)$$

$$g_{W}^{f} = g_{W,W} \qquad (8\text{-}8)$$

respectively. Where, $g_{R(i),w}$ and $g_{W,W}$ are interaction energies of contacts between amino acid residues i and a water molecule, and between two water molecules, respectively. In all the formulations above, only first neighbor contacts are considered.

Therefore, the interaction energy change caused by folding a residue from its coil state to helical state is given in terms of its location in the helical sequence as follows:

If the folded residue is a central helical residue, then

$$\Delta g_{R(i)}^{c\to\alpha}\,|_{\text{Central}} = g_{R(i)}^{\alpha} - g_{R(i)}^{c}$$

$$= [\, g_{R(i-3),R(i)} + g_{R(i+3),R(i)} - 2\, g_{W,R(i)}]/z$$

$$= [\, \tau_{R(i-3),R(i)} + \tau_{R(i+3),R(i)} - 2\, \tau_{W,R(i)}\,] * RT/z \qquad (8\text{-}9)$$

If the folded residues are located at the two ends of the helical sequence, then

$$\Delta g_{R(i)}^{c\to\alpha}\,|_{\text{N-terminal}} = [\, g_{R(i+3),R(i)} - g_{W,R(i)}]/z$$

$$= [\, \tau_{R(i+3),R(i)} - \tau_{W,R(i)}\,] * RT/z \qquad (8\text{-}10)$$

$$\Delta g_{R(i)}^{c\to\alpha}\,|_{\text{C-terminal}} = [\, g_{R(i-3),R(i)} - g_{W,R(i)}]/z$$

$$= [\, \tau_{R(i-3),R(i)} - \tau_{W,R(i)}\,] * RT/z \qquad (8\text{-}11)$$

The interaction energy change of a solvent molecule from its solvated state to its free state is:

$$\Delta g_{R(i),w}^{s\to f} = g_w^s - g_{R(i),w}^f$$

$$= (g_{w,w} - g_{R(i),w})/z$$

$$= -\tau_{R(i),W} * RT/z \qquad (8\text{-}12)$$

Defined as the following, $\tau_{R(i),R(j)}$'s are interaction parameters between residues i and j, and $\tau_{R(i),W}$'s between residue i and a solvent (water) molecule:

$$\tau_{R(i),R(j)} = [g_{R(i),R(j)} - g_{R(j),R(j)}]/RT \qquad (8\text{-}13)$$

$$\tau_{R(i),W} = [g_{R(i),W} - g_{w,w}]/RT \qquad (8\text{-}14)$$

For each central helical residue, the formation of an α-helix releases two water molecules bonded to the residue. And for each residue at the two ends of the helical sequence, or residues 1, 2, 3 and N_h-2, N_h-1 and N_h, folding into α-helical conformation frees only one solvent molecule. Therefore, the change of the system enthalpy for folding a residue R(i) in the polypeptide chain into α-helix formation from a coiled state can be formulated as:

$$\Delta h^{c\to\alpha} = \Delta g_{R(i)}^{c\to\alpha} + 2\frac{\Delta g_{W,R(i)}^{c\to\alpha}}{m_i} \qquad (8\text{-}15)$$

Here, $\Delta g_{R(i)}^{c\to\alpha} = \Delta g_{R(i)}^{c\to\alpha}|_{\text{N-terminal}}$ (for N-terminal end helical residues)

$= \Delta g_{R(i)}^{c\to\alpha}|_{\text{Central}}$ (for central helical residues)

$= \Delta g_{R(i)}^{c\to\alpha}|_{\text{C-terminal}}$ (for C-terminal end helical residues)

and $m_i = 1$ (for central helical residues)

$= 2$ (for terminal helical residues)

m_i is an index differentiating central helical residues from terminal helical residues and the conformational states of residue i.

Therefore, equation (8-15) can be rewritten as:

$$\frac{\Delta h^{c\to\alpha}}{RT} = -\frac{\Delta \tau_i}{z} \qquad (8\text{-}16)$$

And here,

$\Delta \tau_i = [(\tau_{W,R(i)} + \tau_{R(i),W}) - \tau_{R(i+3),R(i)}]$ (for N-terminal end helical residues)

$= [2(\tau_{W,R(i)} + \tau_{R(i),W}) - (\tau_{R(i-3),R(i)} + \tau_{R(i+3),R(i)})]$ (for central helical residues)

$= [(\tau_{W,R(i)} + \tau_{R(i),W}) - \tau_{R(i-3),R(i)}]$ (for C-terminal end helical residues) (8-17)

8.3 The Entropy Contribution

The configurational entropy of mixing a disoriented polymer and the solvated solvent to form a random coiled polymer solution is given by Flory and Huggins (1941, 1942):

$$\Delta S_{mix} = - R \{N_1 \ln [N_1/(N_1 + N_2 x)] + N_2 \ln [N_2 x/(N_1 + N_2 x)]\} \quad (8\text{-}18)$$

Here, N_1 is the number of solvent molecules, N_2 is the number of polymer chain molecules of x segment long, and x is the number of the segments in a polymer chain.

The backbone of the α-helical conformation of a chain is taken to be one with no randomness at all. Therefore, we estimate the loss of entropy by setting the configurational entropy of the helical conformation to zero and calculating the entropy loss for the chain based on the number of residues in the coil state, n.

$$\Delta S^{c \to \alpha} = \Delta S^{c,FH}|_{x=n-1} - \Delta S^{c,FH}|_{x=n} \quad (8\text{-}19)$$

8.4 The Gibbs Energy of Folding

The Gibbs energy of folding of a polypeptide chain from coiled conformation to α-helical conformation is summarized in matrix form as follows:

Here, the elements of the matrices $\Delta \tau_i$'s take the same form as represented by equation

$$\frac{\Delta G^{c\to\alpha}}{RT} = -e^{T}\left[\frac{\Delta\tau_i}{z} - \frac{\Delta s_i^{c\to\alpha}}{R}\right]_{NxN}\left[s_i\right]_{Nx1} \qquad (8\text{-}18)$$

(8-17), and s_i is the conformational index of the polypeptide sequence. e^T is the transpose of unit vector e with N elements. These parameters take the following values:

m_i: maintain the same meaning as previously defined (see equation (8-17))

s_i = 0 for residues in coil conformation

= 1 for residues in helical conformation

As an example, the matrix representation of folding a polypeptide from a coiled conformation into the lowest Gibbs free energy single helical sequence conformation is given as follows (refer to Figures 8-1 and 8-2 for related symbols):

$$
{}_{\Delta}G^{c\rightarrow\alpha}/RT = - \; e^{\left[- \begin{vmatrix} \left(\tau_{W,R(1)} + \tau_{R(1),W}\right) & & & & & & & & & \\ & \left(\tau_{W,R(2)} + \tau_{R(2),W}\right) & & & & & & & & \\ & & \left(\tau_{W,R(3)} + \tau_{R(3),W}\right) & & & & & & & \\ & & & 2\left(\tau_{W,R(4)} + \tau_{R(4)W}\right) & & & & & & \\ & & & & \cdots\cdots & & & & & \\ & & & & & 2\left(\tau_{W,R(I)} + \tau_{R(I),W}\right) & & & & \\ & & & & & & \cdots\cdots & & & \\ & & & & & & & 2\left(\tau_{W,R(Nh-3)} + \tau_{R(Nh-3),W}\right) & & \\ & & & & & & & & \left(\tau_{W,R(Nh-2)} + \tau_{R(Nh-2),W}\right) & \\ & & & & & & & & & \left(\tau_{W,R(Nh-1)} + \tau_{R(Nh-1),W}\right) \\ & & & & & & & & & & \left(\tau_{W,R(Nh)} + \tau_{R(Nh),W}\right) \end{vmatrix} \right.}
$$

$$
\left. - \begin{vmatrix} \tau_{R(4),R(1)} & & & & & & & & & \\ & \tau_{R(5),R(2)} & & & & & & & & \\ & & \tau_{R(6),R(3)} & & & & & & & \\ & & & \cdots\cdots & & & & & & \\ & & & & \tau_{R(I+3),R(i)} + \tau_{R(I-3),R(I)} & & & & & \\ & & & & & \cdots\cdots & & & & \\ & & & & & & \tau_{R(Nh-5),R(Nh-2)} & & & \\ & & & & & & & \tau_{R(Nh-4),R(Nh-1)} & & \\ & & & & & & & & \tau_{R(Nh-3),R(Nh)} & \end{vmatrix} \right\| \left\| \; 1/z - \frac{\left(\Delta S^{c,FH}\right|_{I-1} - \Delta S^{c,FH}\left|_{I}\right)}{R} \right\|^{I'} \left\| \begin{matrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \\ \cdot \\ \cdot \\ 1 \\ 1 \end{matrix} \right\|
$$

(8-19)

Here, e'^{T} = (1 1 ... 1 1) is the transpose of unit vector with N_h elements, and I' is N_hxN_h unit matrix.

It is clear, from equations (8-17) and (8-19), that there are two contributions to the enthalpy of folding, the contribution due to intrinsic helix-forming potential of the residue in its homopolypeptide represented by the term ($\tau_{W,R(i)} + \tau_{R(i),W}$), and the contribution due to the cooperative potential of the amino acid residues to form α-helices (or, the residue-residue interactions) by $\tau_{R(i-3),R(i)}$ and $\tau_{R(i+3),R(i)}$. In other words, the model suggests that the helical conformation of a peptide is determined by both the intrinsic conformational preferences of the individual residues in a specific solvent and their cooperative interactions with neighboring residues in the chain.

The formulation can be applied to multiple helix formation in a polypeptide chain. The definition of an α-helical conformation of a polypeptide chain is shown in Figure 8-2. Where N is the total number of residues in the polypeptide, i is the sequence index of the peptide. A helical sequence is defined as a series of consecutive residues in α-helical conformation.

N_h is the number of helical residues in a specific helical sequence (if multiple helices occur in a chain, there will be more than one N_h's), and n is the helical sequence index for a specific helical sequence (say, N_h) in the polypeptide chain. A central helical residue is meant to be a residue located at positions rather than the left-end (or N-terminal end) and the right-end (or C-terminal end) of a helical sequence as illustrated in Figure 8-2.

8.5 Interaction Energy Parameters

The key parameters with the model are the binary interaction energy parameters. In this work, we estimated the binary interaction parameters, τ's, from the UNIFAC (universal functional activity coefficient) method using ASPEN PLUS™ software (Aspen Technology, 1988). For the convenience and the representativity of the characteristics of the peptide behavior, we define the interaction parameters on the basis of peptide unit (-CH(R_i)-CO-NH-) when dealing with the conformations of polypeptides and proteins. For more detailed information on the estimate techniques for the interaction parameters, refer to Chapter 7.

In general, residue-water interactions are stronger than residue-residue interactions (see the 21x21 matrix of the interaction parameters shown in Tables 8-1 and 8-2). Consequently, the hydrophobic interactions play dominant role in peptide folding resulting in the partial burying of nonpolar side chain residues. In addition, the residue-residue interactions can also contribute significantly toward the stabilization or destabilization of the folded structure by enhancing or weakening the helix-forming potential of the residue under consideration. In certain cases, residue-solvent interactions are even weaker than residue-residue interactions (see Tables 8-1 and 8-2). The energy of interaction of two groups can easily be as great as thermal energy on per residue basis. Some of the interactions are intensive, with the τ's in the range of ~2.5 (the forces are especially strong for some cases with tyrosine).

Table 8-1. Residue-residue and Residue-water Interaction Parameter τ_{ij}'s at 25 °C

	j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
i		A	V	L	I	P	M	F	W	G	S	T	C	Y	N	Q	D	E	K	R	H	H2O	
1	A	0.00	0.61	0.85	0.85	0.50	-1.23	-0.03	-0.10	-0.12	1.15	1.13	-0.98	-3.03	1.78	1.67	2.22	2.35	-1.62	2.08	1.42	0.46	A
2	V	-0.47	0.00	0.43	0.43	0.34	0.26	0.70	0.89	-0.72	1.00	1.06	0.03	-2.32	1.87	1.85	1.27	1.50	0.58	2.33	0.95	0.98	V
3	L	-0.59	-0.38	0.00	0.43	0.16	0.11	0.47	0.69	-0.81	0.95	1.00	-0.04	-1.96	1.86	1.84	1.06	1.20	0.55	2.35	0.74	1.27	L
4	I	-0.59	-0.38	-0.38	0.00	0.16	0.11	0.47	0.69	-0.81	0.95	1.00	-0.04	-1.96	1.86	1.84	1.06	1.20	0.55	2.35	0.74	1.27	I
5	P	-0.40	-0.30	-0.13	-0.13	0.00	0.36	0.79	0.91	-0.65	1.04	1.10	0.08	-2.57	1.87	1.85	1.47	1.70	-0.82	2.32	1.09	0.93	P
6	M	1.63	-0.07	0.27	0.27	-0.23	0.00	0.39	0.57	2.23	0.48	0.55	-0.23	-2.54	1.26	1.26	1.48	1.81	-0.91	1.68	0.62	-1.49	M
7	F	-0.08	-0.52	-0.26	-0.26	-0.67	-0.25	0.00	-0.46	-1.00	0.68	0.69	-0.55	-1.27	1.44	1.44	1.33	1.43	-0.33	1.94	0.15	2.50	F
8	W	-0.05	-0.58	-0.31	-0.31	-0.72	-0.32	0.51	0.00	-1.13	-0.07	-0.07	-0.63	-1.10	0.76	0.75	0.22	0.26	-0.11	1.21	0.00	0.21	W
9	G	0.18	1.27	1.59	1.59	1.07	-1.55	1.34	1.45	0.00	1.06	0.97	-1.73	-3.33	1.34	0.03	2.88	2.89	-1.75	-0.15	1.53	0.27	G
10	S	-0.43	0.38	0.80	0.80	0.21	0.17	0.82	1.39	-0.67	0.00	0.39	0.08	-0.07	-1.32	-1.25	-1.58	-1.57	0.84	-1.99	0.53	-1.29	S
11	T	-0.62	-0.06	0.27	0.27	-0.17	-0.05	0.45	0.97	-0.72	-0.30	0.00	-0.01	-0.16	-1.13	-1.28	-1.42	-1.54	0.53	-1.98	0.40	-1.11	T
12	C	1.23	0.47	0.86	0.86	0.24	0.22	1.09	1.31	2.61	0.79	0.86	0.00	0.15	1.58	1.63	0.00	0.19	0.90	2.04	0.84	-0.67	C
13	Y	6.69	4.28	3.42	3.42	4.98	4.80	1.75	1.43	8.07	-1.30	-1.29	-0.74	0.00	-0.10	-0.17	0.10	0.24	-0.86	-2.03	-0.42	-0.53	Y
14	N	-0.86	-0.13	0.27	0.27	-0.29	-0.33	0.34	0.95	-1.00	1.77	1.47	-0.30	-0.65	0.00	0.37	-0.11	0.32	0.27	0.51	0.66	-2.14	N
15	Q	-0.97	-0.42	-0.09	-0.09	-0.54	-0.46	0.06	0.63	-0.20	1.65	1.70	-0.37	-0.67	-0.30	0.00	0.05	-0.03	0.00	-0.44	0.68	-2.01	Q
16	D	-0.91	0.22	0.77	0.77	-0.08	-0.83	0.43	0.90	-1.33	2.35	2.08	-0.57	0.48	-0.19	-0.25	0.00	0.40	1.43	-1.69	0.50	-1.06	D
17	E	-1.16	-0.36	0.09	0.09	-0.57	-1.12	-0.01	0.47	-1.43	2.33	2.25	-0.74	0.16	-0.53	-0.24	-0.30	0.00	1.66	-1.71	0.99	-0.94	E
18	K	2.39	-0.50	-0.32	-0.32	0.99	1.12	0.41	0.54	2.68	-0.68	-0.67	-0.50	0.54	0.18	0.21	-0.94	-1.25	0.00	0.88	0.47	-1.46	K
19	R	-1.18	-0.60	-0.27	-0.27	-0.73	-0.67	-0.27	0.21	-0.14	3.22	3.18	-0.64	3.26	-0.44	0.49	2.55	2.56	-0.28	0.00	0.26	-1.31	R
20	H	-0.88	-0.19	0.20	0.20	-0.46	-0.26	0.11	0.12	-1.08	-0.64	-0.62	-0.64	0.08	-0.26	-0.36	-0.41	-0.83	0.07	-0.07	0.00	-1.71	H
21	w	2.22	3.99	4.92	4.92	3.83	5.55	7.70	7.82	1.41	2.88	3.28	3.11	5.49	3.95	4.28	3.22	3.69	4.38	0.01	3.03	0.00	w

Table 8-2. Residue-residue and Residue-water Interaction Parameter τ_{ij}'s at 0 °C

	j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
i		A	V	L	I	P	M	F	W	G	S	T	C	Y	N	Q	D	E	K	R	H	H2O	
1	A	0.00	0.67	0.93	0.93	0.55	-1.34	-0.03	-0.11	-0.13	1.25	1.23	-1.07	-3.31	1.94	1.82	2.43	2.57	-1.77	2.27	1.55	0.50	A
2	V	-0.51	0.00	0.47	0.47	0.37	0.29	0.76	0.97	-0.78	1.09	1.16	0.03	-2.53	2.04	2.02	1.39	1.63	0.64	2.54	1.04	1.07	V
3	L	-0.65	-0.41	0.00	0.47	0.17	0.11	0.51	0.75	-0.88	1.04	1.09	-0.04	-2.14	2.04	2.01	1.16	1.31	0.60	2.57	0.81	1.39	L
4	I	-0.65	-0.41	-0.42	0.00	0.17	0.11	0.51	0.75	-0.88	1.04	1.09	-0.04	-2.14	2.04	2.01	1.16	1.31	0.60	2.57	0.81	1.39	I
5	P	-0.44	-0.33	-0.14	-0.14	0.00	0.39	0.86	0.99	-0.71	1.14	1.20	0.09	-2.80	2.04	2.02	1.60	1.86	-0.90	2.53	1.19	1.02	P
6	M	1.78	-0.07	0.29	0.29	-0.25	0.00	0.42	0.62	2.44	0.52	0.60	-0.25	-2.77	1.37	1.38	1.62	1.97	-0.99	1.83	0.67	-1.62	M
7	F	-0.09	-0.57	-0.29	-0.29	-0.73	-0.27	0.00	-0.51	-1.09	0.74	0.75	-0.60	-1.39	1.57	1.57	1.45	1.56	-0.35	2.11	0.16	2.73	F
8	W	0.06	-0.63	-0.34	-0.34	-0.79	-0.35	0.55	0.00	-1.24	-0.08	-0.08	-0.69	-1.20	0.83	0.82	0.24	0.29	-0.12	1.32	0.00	0.23	W
9	G	0.19	1.39	1.74	1.74	1.17	-1.69	1.46	1.58	0.00	1.15	1.05	-1.89	-3.64	1.47	0.04	3.14	3.16	-1.91	-0.16	1.67	0.29	G
10	S	0.47	0.41	0.87	0.87	0.23	0.18	0.89	1.52	-0.74	0.00	0.42	0.08	-0.08	-1.44	-1.36	-1.73	-1.72	0.92	-2.17	0.57	-1.41	S
11	T	-0.68	-0.07	0.29	0.29	-0.19	-0.06	0.49	1.06	-0.78	-0.33	0.00	-0.01	-0.17	-1.23	-1.39	-1.54	-1.68	0.57	-2.16	0.44	-1.22	T
12	C	1.34	0.52	0.94	0.94	0.26	0.24	1.19	1.44	2.84	0.86	0.93	0.00	0.17	1.73	1.78	0.00	0.21	0.98	2.23	0.91	-0.74	C
13	Y	7.31	4.68	3.73	3.73	5.44	5.24	1.91	1.56	8.81	-1.42	-1.41	-0.81	0.00	-0.11	-0.19	0.10	0.26	-0.94	-2.21	-0.46	-0.58	Y
14	N	-0.93	-0.14	0.30	0.30	-0.32	-0.35	0.37	1.03	-1.09	1.93	1.60	-0.33	-0.71	0.00	0.40	-0.12	0.35	0.29	0.56	0.72	-2.33	N
15	Q	-1.05	-0.46	-0.10	-0.10	-0.59	-0.51	0.06	0.69	-0.22	1.80	1.86	-0.41	-0.74	-0.33	0.00	0.05	-0.03	0.00	-0.48	0.74	-2.19	Q
16	D	-1.00	0.24	0.85	0.85	-0.09	-0.90	0.47	0.98	-1.45	2.56	2.27	-0.62	0.53	-0.21	-0.28	0.00	0.44	1.56	-1.85	0.54	-1.16	D
17	E	-1.27	-0.39	0.10	0.10	-0.63	-1.23	-0.01	0.51	-1.56	2.55	2.45	-0.81	0.17	-0.57	-0.26	-0.33	0.00	1.81	-1.87	1.08	-1.03	E
18	K	2.61	-0.54	-0.35	-0.35	1,08	1.22	0.45	0.59	2.92	-0.74	-0.73	-0.54	0.59	0.19	0.23	-1.03	-1.37	0.00	0.96	0.52	-1.59	K
19	R	-1.29	-0.66	-0.29	-0.29	-0.80	-0.73	-0.30	0.23	-0.15	3.51	3.47	-0.70	3.55	-0.48	0.54	2.78	2.79	-0.31	0.00	0.28	-1.44	R
20	H	-0.96	-0.21	0.22	0.22	-0.50	-0.29	0.12	0.13	-1.18	-0.70	-0.67	-0.70	0.08	-0.28	-0.40	-0.45	-0.91	0.08	-0.08	0.00	-1.87	H
21	w	2.42	4.35	5.37	5.37	4.18	6.06	8.41	8.54	1.54	3.14	3.58	3.40	5.99	4.32	4.67	3.51	4.03	4.78	0.01	3.31	0.00	w

Note that the UNIFAC method is not capable of estimating the contribution of ion-ion interactions and ion-molecule interactions.

According to prior homopolypeptide work in Chapter 7, His, Asp, Ser and Thr behave as helix breakers intrinsically. It can also be seen from the 21x21 matrices (Tables 8-1 and 8-2) that there exist unfavorable interactions with theses residues (or, $\tau_{R(j),R(i)}$'s are positive for most of the cases when these residues are under consideration (as the ith residue)). As an example, Table 8-3 shows that if we assume that all the 20 residues in S-peptide fold into α-helical conformation, the net contribution from each of the above mentioned residues His, Ser and Thr (except Asp14) is unfavorable. These kinds of residues (from residues 12 to 18) line up in a row along the peptide chain, they are the sources of local instabilities to helical conformations, consequently, this portion of the chain remains as coil structure.

Table 8-3. Individual Contributions of the Twenty Residues in S-peptide to the Gibbs Energy of Folding and the Components of the Gibbs Energy of Folding of Each Residue in the Chain (at 0°C)

			Intrinsic	Cooperative		Chain End Effect
i	Residue	$\Delta G_i^{c\to\alpha}/RT$	$\Delta G_i^{c\to\alpha *}/RT$	$\tau_{R(i-3),R(i)}/z$	$\tau_{R(i+3),R(i)}/z$	$(\tau_{R(i),W}+\tau_{W,R(i)})/z$
1	K	-0.027	-0.261		-0.295	0.529
2	E	0.727	-0.199		0.428	0.498
3	T	0.610	0.012		0.205	0.393
4	A	0.695	-0.175	0.435	0.435	
5	A	-0.401	-0.175	-0.211	-0.015	
6	A	-0.499	-0.175	-0.113	-0.211	
7	K	-0.608	-0.262	-0.295	-0.051	
8	F	-2.901	-2.905	-0.006	0.010	
9	E	0.077	-0.199	0.428	-0.151	
10	R	1.739	1.273	0.160	0.306	
11	Q	0.191	-0.024	0.262	-0.046	
12	H	0.596	0.320	0.181	0.096	
13	M	-0.765	-0.674	-0.121	0.031	
14	D	-0.233	0.016	0.009	-0.257	
15	S	0.104	0.220	-0.116	0.000	
16	S	0.516	0.220	0.087	0.209	
17	T	0.595	0.012	0.378	0.205	
18	S	0.509	0.219	0.000		0.290
19	A	0.233	-0.174	-0.079		0.486
20	A	0.199	-0.174	-0.113		0.486

Here, i is the chain length counter, $\Delta G_i^{c\to\alpha}/RT$ is the free energy of folding of each residue alone, $\Delta G_i^{c\to\alpha *}/RT$ the "intrinsic" helix-forming tendency of a specific residue, $\tau_{R(i-3),R(i)}/z$ and $\tau_{R(i+3),R(i)}/z$ the "cooperative" contributions from the i and i-3th residue interaction and the i and i+3th residue interaction, respectively, $(\tau_{R(i),W}+\tau_{W,R(i)})/z$ the chain end effect due to the deficit of hydrogen-bonding and other types of interactions

at the two ends of the helix, and z, the coordination number of the lattice structure of the peptide solution.

8.6 Approach to Search for the Stable Conformation of Polypeptides in Aqueous Solutions

With the above interaction parameters and the proposed model in the earlier section for the Gibbs energy of folding, an exhaustive search procedure has been developed to find the most stable conformation of the polypeptide chains in water. Assume that the helix length for a specific helical sequence is N_h ($\leq$ N, the chain length of the peptide chain), and N_h could take any value from 4 to N since the first hydrogen bond starts to form with at least 4 residues apart along the chain. Also assume that the helix could possibly start from any residue of the chain. The search was done according to the following steps:

Step 1: Choose a starting residue N_s. For an exhaustive search, the possible starting residue could be any residue from residues 1 to N-4 of the peptide chain.

Step 2: Choose a helical length N_h. Let the helix start from the lowest helical length $N_h=4$.

Step 3: Calculate $\Delta G^{c\rightarrow\alpha}$, the free energy of folding the peptide chain into a helical conformation with N_h consecutive helical residues in a specific helical sequence.

Step 4: Alter the helix length from the previous N_h to the next N_h. N_h can take any value from $N_h=4, 5, 6, \ldots$ until N. Repeat the same calculations as in step 3. Compare the free energy of folding, the set of the successive helical residues with the

lowest $\Delta G^{c\to\alpha}$ is considered as the locally stable helix with the specific possible helix-starting residue.

Step 5: Choose another helix starting residue, repeat steps 1 through 4 and find the possible helical portions of the chain with this specific starting residue for different N_h's.

Step 6: After calculating the $\Delta G^{c\to\alpha}$'s for all the residues, compare all the $\Delta G^{c\to\alpha}$'s for all the cases. The helix with the lowest free energy of folding is deemed as the most stable helix in the chain.

A case study for the most stable conformation of S-peptide in RNase A is shown in Table 8-4. The Gibbs energies of folding for all possible stable local helix formations of the S-peptide are listed. The table indicates that the most stable helix in water at 0°C consists of residues 1 to 11 with the lowest Gibbs free energy of folding $\Delta G^{c\to\alpha}/RT=0.1688$.

The two-dimensional and three-dimensional contour maps for Gibbs free energy of folding of the twenty residue long S-peptide as a function of helix starting residue N_s and helix length N_h are shown in Figures 8-3 and 8-4. It can be seen from Table 8-4 and Figure 8-3 that when the helices start from residues 1, 2, 3, 4 and 5 with helix chain lengths 11, 10, 9, 8 and 7, correspondingly, the conformations with these helical sequences have relatively lower Gibbs energies, which is illustrated as a deeply tinted band parallel to the off-diagonal of the contour map with $N_s=1$ and $N_h=11$ as the lowest free energy helical sequence of S-peptide. In Figure 8-4, an exhaustive search for the lowest Gibbs

free energy of folding is represented over the entire N_s and N_h space, and the figure shows clearly that the lowest Gibbs energy helical sequence of the S-peptide starts at residue 1 of the peptide and the helix is 11 residues long.

Table 8-4. Calculated Local Minimum Gibbs Energy of Folding for S-peptide with Different Helical Lengths

N_h	Residues in Hypothetical Helix	$\Delta G^{c\to\alpha}/RT$
4	5-8	0.8338
5	4-8	0.7553
6	3-8	0.7662
7	5-11	0.2852
8	4-11	0.2068
9	3-11	0.2177
10	2-11	0.2473
11	1-11*	0.1688*
12	1-12	0.2585
13	1-13	0.7424
14	1-14	0.7008
15	2-16	0.9196
16	1-16	0.8412
17	1-17	0.9754
18	1-18	1.1952
19	1-19	1.3477
20	1-20	1.3586

*: The predicted most stable conformation of S-peptide at 0° C

Figure 8-3. Contour Map of Gibbs Energy of Folding of S-Peptide into α-helical Conformation as a Function of Helix Length and Helix Starting Residue

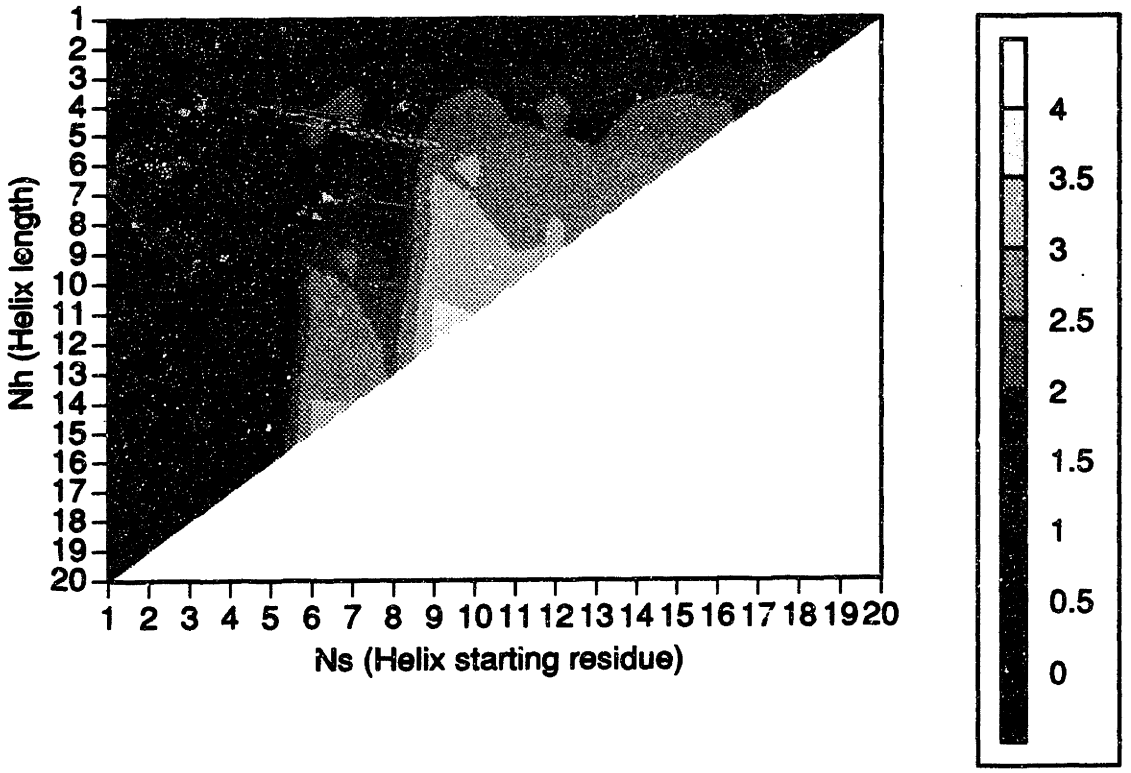
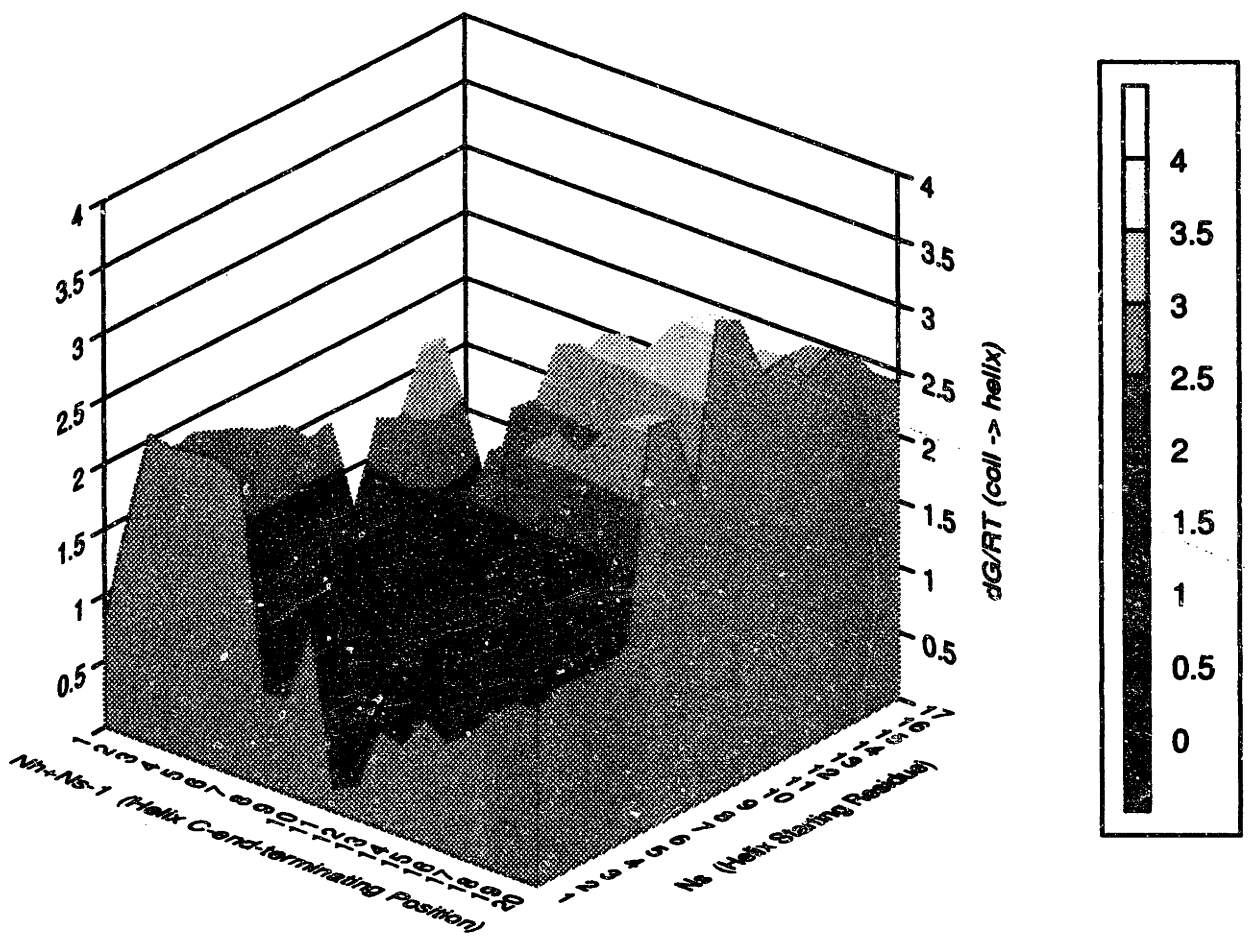


Figure 8-4. Gibbs Energy Surface of Folding of S-peptide into α-helix Conformation



An important conclusion can be drawn from Table 8-4 that according to the model prediction, a polypeptide might have multiple helical conformations. In other words, the folded form of the peptide chain is a mixture of peptides with different helical sequences such as 1-11, 2-11, 3-11, 4-11, and 5-11. It suggests that the experimentally observed properties may be an averaged value of the conformational mixture of the peptides.

8.7 Helix Population and Helicity

The conformational transition of a polypeptide chain can be considered as a quasi-chemical equilibrium. The Gibbs energy of folding the coiled chain into a helix conformation with a specific helical sequence is, in fact, the equilibrium constant between the two states, the specific helical sequence state of the chain and the coil state as given below:

$$\Delta G^{c\to\alpha}/RT = -\text{Ln } K = -\text{Ln } \{[\alpha\text{-helix}]/[\text{coil}]\} \tag{8-20}$$

Where, [α-helix] and [coil] are defined as the percentage of polypeptide molecules in helical and coil conformations, respectively. Therefore, the "two-state model" helicity is defined as:

$$[\text{ helicity }] = \frac{e^{-\frac{\Delta G^{c\to\alpha}}{RT}}}{1 + e^{-\frac{\Delta G^{c\to\alpha}}{RT}}} \tag{8-21}$$

The lower the Gibbs energy of folding, the higher the helix content with that specific helical sequenced conformation. For example, the model-predicted most probable helix

conformation of S-peptide at 0°C is the [1-11] helical sequence in the twenty residue chain with Gibbs energy of folding $\Delta G^{c\to\alpha}/RT = 0.1688$, which means that the helix content of this specific conformation is around 46%, comparable to the 30% helicity value indicated by Kim and Baldwin (1984) for S-peptide and measured by Bierzynski, et al., (1982) for C-peptide at near 0°C and pH 5.

Considering the dispersity of the helix population, we can also define the averaged helicity of a polypeptide in solution. According to statistical thermodynamics, the probability of a given polypeptide solution system in state j, in analogy to a quantum state of the canonical ensemble (Prausnitz, 1986), with the Gibbs energy $\Delta G^{c\to\alpha}/RT$ with respect to its reference state is:

$$p_j = \frac{e^{-\frac{\Delta G_j^{c\to\alpha}}{RT}}}{1 + \sum_{k}^{\text{all }\alpha} e^{-\frac{\Delta G_k^{c\to\alpha}}{RT}}} \qquad (8\text{-}22)$$

Here, by state j, we mean a state that the polypeptide chain is folded into a specific helical sequence $N_h(j)$ in its solvent. The summation here is over all possible helical sequence states of the polypeptides. Then, the average α-helicity for all the possible helical conformations (with various helical lengths and sequences) is defined by equation (8-23).

Here, the summation upper limit, "all α" stands for the summation over all possible α-helical conformations of the polypeptides in solvent water under given conditions. It should also be pointed out that the average helicity from this population dispersity

$$[\text{ helicity }] = \frac{\sum_{k}^{\text{all }\alpha} p_k}{1 + \sum_{k}^{\text{all }\alpha} p_k} \qquad (8\text{-}23)$$

formulation for S-peptide at 0°C is 49%. This value is very close to the 46% helicity of the "two state model" calculation as shown above, and it also differs from the reported experimental value of 30% helicity due to different definitions of the helicity. According to Bierzynski, et al. (1982), their experimental "helicity" was defined as the ratio of the $[\theta]_{222}$ for the peptide to the $[\theta]_{222}$ for complete helix formation. They treated the $[\theta]_{222}$ (from CD Spectra) of 100% helicity as an averaged value of $[\theta]_{222}$ when unfolded S-peptide recombines with folded S-protein if the change is attributed entirely to the 10 peptide groups of the helix formed by residues 3-12 and the $[\theta]_{222}$ provided by the α-helix basis spectrum compiled from 15 proteins with an average helix length of around 10 residues. They assume that the helix is completely melted out at high temperatures and use $[\theta]_{222}$ at 45 °C as a base line for 0% helicity.

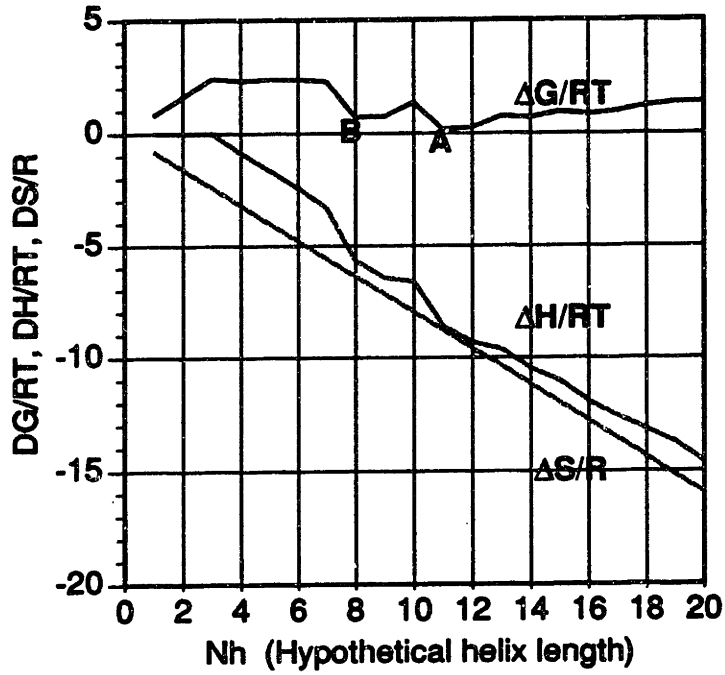
Table 8-5 and Figure 8-5 show the Gibbs energy of folding of the S-peptide, and the individual components, $\Delta H^{c\to\alpha}/RT$ and $\Delta S^{c\to\alpha}/R$ (all are dimensionless), versus the specific helical sequence N_h with helix starting from residue number 1. Each dot on a line represents a possible single helix starting from residue number one (n=1) to the residue corresponding to the dot under consideration (or residue N_h for this case). For instance, point A on the line $\Delta G/RT$ corresponds to the helical sequence N_h=11 (from

residues 1 to 11) of the S-peptide, and the dot B corresponds to the hypothetical helix $N_h = 8$ spanning from residues 4 to 11 of the peptide.

Table 8-5. Calculated Thermodynamic Parameters for the S-peptide (0° C)

Chain length N_h	$\Delta G^{c \to \alpha}(i)/RT$	$\Delta H^{c \to \alpha}(i)/RT$	$\Delta S^{c \to \alpha}(i)/R$
1	0.7981		-0.7981
2	1.5962		-1.5962
3	2.3942		-2.3942
4	2.3158	-0.8766	-3.1923
5	2.3454	-1.6450	-3.9904
6	2.3563	-2.4322	-4.7885
7	2.2778	-3.3087	-5.5866
8	0.7174	-5.6673	-6.3847
9	0.7470	-6.4357	-7.1827
10	1.3617	-6.6191	-7.9808
11	0.1688	-8.6101	-8.7789
12	0.2585	-9.3185	-9.5770
13	0.7424	-9.6327	-10.3751
14	0.7008	-10.4724	-11.1732
15	0.9506	-11.0207	-11.9712
16	0.8412	-11.9282	-12.7693
17	0.9754	-12.5920	-13.5674
18	1.1952	-13.1703	-14.3655
19	1.3477	-13.7265	-15.1636
20	1.3586	-14.6030	-15.9617

Figure 8-5. Thermodynamic Parameters of Folding of S-peptide As a Function of Helix Chain Length



8.8 Prediction of Polypeptide Conformation

Simple helix-forming synthetic and natural peptides provide excellent systems for testing the proposed model. For example, according to Kim and Baldwin (1984), the C-peptide helix may function as an autonomous folding unit, and helix formation in the isolated S-peptide (residues 1-20) is localized in a manner resembling the intact protein. Therefore, it should be interesting to see how the predicted locations of helical segments agree with experimental data of these and other polypeptides. Section 8.6 has briefly shown the prediction results for S-peptide in comparison with Kim and Baldwin's work, more detailed discussions will be given in the following Sections.

Table 8-6 lists the amino acid sequences for over thirty polypeptides studied in this paper, and Table 8-7 gives the Gibbs energy of folding of these natural and synthetic polypeptides and the comparison of model predicted conformations with experimental observations documented in the literature. Generally speaking, the model prediction of the favored chain conformation for the 33 peptides investigated at the experimental temperatures (0, 1, and 3°C, respectively) and pH (7.0) are consistent with the conformations reported in the literature.

It should be noted that the calculations are carried out at neutral pH. Though most of the ionizable groups on the side chain of the amino acid residues and the amino and carboxyl groups at the chain ends will be charged at this pH, excellent match between the model predictions and the experimental observations was obtained without rigorously taking into account the charge effect.

Table 8-6. Amino Acid Sequences of Some Natural and Synthetic Polypeptides

	Peptides	Sequences	Note	References
1	C-Peptide	K-E-T-A-A-A-K-F-E-R-Q-H-M	1-13 of RNase A	a
2	S-Peptide	K-E-T-A-A-A-K-F-E-R-Q-H-M-D-S-S-T-S-A-A	1-20 of RNase A	a
3	Pα	N-N-F-K-S-A-E-D-C-M-R-T-A-G-G-A	43-58 of BPTI	b
4	Pα5	S-A-E-D-C-M-R-T-A-G-G-A	47-58 of BPTI	c
5	3K(I)	A-A-A-A-K-A-A-A-A-K-A-A-A-A-K-A	Alanine based	d
6	3K(II)	A-K-A-A-A-A-K-A-A-A-A-K-A-A-A-A	Alanine based	d
7	4K	A-K-A-A-K-A-A-A-A-K-A-A-A-A-K-A	Alanine based	d
8	6K(I)	A-K-A-A-K-A-K-A-A-K-A-K-A-A-K-A	Alanine based	d
9	6K(Ii)	A-K-A-A-A-K-K-A-A-A-K-K-A-A-A-K-A	Alanine based	d
10	3E	A-E-A-A-A-A-E-A-A-A-A-E-A-A-A-A	Alanine based	d
11	(I+3)E,K	A-E-A-A-K-A-E-A-A-K-A-E-A-A-K-A	Alanine based	d,f
12	(I+4)E,K	A-E-A-A-A-K-E-A-A-A-K-E-A-A-A-K-A	Alanine based	d,f
13	C-RN16	K-E-T-A-A-A-K-F-L-R-A-H-A	C-peptide analog	e
14	C-RN21	A-E-T-A-A-A-K-F-L-R-A-H-A	C-peptide analog	e
15	C-RN23	A-A-T-A-A-A-K-F-L-R-A-H-A	C-peptide analog	e
16	C-RN28	A-E-T-A-A-A-K-F-L-A-A-H-A	C-peptide analog	e
17	C-RN54	A-A-T-A-A-A-K-F-L-A-A-H-A	C-peptide analog	e
18	C-RN25	A-R-T-A-A-A-K-F-L-E-A-H-A	C-peptide analog	e
19	C-RN26	A-D-T-A-A-A-K-F-L-R-A-H-A	C-peptide analog	e
20	C-RN80	A-E-T-A-A-A-K-Y-L-R-A-H-A	C-peptide analog	e
21	C-RN121	A-A-T-A-A-A-K-Y-L-R-A-H-A	C-peptide analog	e
22	C-RN119	A-E-T-A-A-A-K-Y-L-A-A-H-A	C-peptide analog	e
23	C-RN120	A-A-T-A-A-A-K-Y-L-A-A-H-A	C-peptide analog	e
24	C-RN84	A-E-T-A-E-A-K-Y-L-R-A-H-A	C-peptide analog	e
25	(I+4)K,E	A-K-A-A-A-E-K-A-A-A-E-K-A-A-A-E-A	de novo design	f
26	(I+3)K,E	A-K-A-A-E-A-K-A-A-E-A-K-A-A-E-A	de novo design	f
27	3ALA	Y-K-A-A-A-A-K-A-A-A-A-K-A-A-A-A-K	Alanine-based	g
28	3LEU	Y-K-A-A-L-A-K-A-A-L-A-K-A-A-L-A-K	Alanine-based	g
29	3PHE	Y-K-A-A-F-A-K-A-A-F-A-K-A-A-F-A-K	Alanine based	g
30	3ILE	Y-K-A-A-I-A-K-A-A-I-A-K-A-A-I-A-K	Alanine-based	g
31	1VAL	Y-K-A-A-A-A-K-A-A-V-A-K-A-A-A-A-K	Alanine-based	g
32	2VAL	Y-K-A-A-V-A-K-A-A-V-A-K-A-A-A-A-K	Alanine-based	g
33	3VAL	Y-K-A-A-V-A-K-A-A-V-A-K-A-A-V-A-K	Alanine-based	g

a: Kim & Baldwin, 1984
b: Cas & Kim, 1988
c: Goodman & Kim, 1989
d: Marqusee, et al., 1989
e: Fairman, et al., 1990
f: Marqusee, et al., 1987
g: Padmanabhan, et al., 1990

Table 8-7. Comparison of the Model Predicted and Observed Conformations of Polypeptides

No	Peptides	Min $\Delta G^{c \to \alpha}$ /RT	Helical Residues	
			Model	Literature*
		Natural Peptides		
1	C-Peptide	0.1688 (0° C)	1-11	3-12
2	S-Peptide	0.1688 (0° C)	1-11	3-12
3	Pα	0.5356 (0° C)	45-52	43-58
4	Pα5	1.9964 (0° C)	48-52	47-58
		Synthetic Peptides		
5	3K(I)	0.6481 (1° C)	1-16	16 residue helix
6	3K(II)	0.6481 (1° C)	1-16	16 residue helix
7	4K	0.4656 (1° C)	1-16	16 residue helix
8	6K(I)	0.2925 (1° C)	1-16	16 residue helix
9	6K(II)	0.9566 (1° C)	1-17	17 residue helix
10	3E	1.1866 (1° C)	1-16	16 residue helix
11	(I+3)E,K	0.8201 (1° C)	1-16	16 residue helix
12	(I+4)E,K	1.4951 (1° C)	1-17	17 residue helix
13	C-RN16	-1.2660 (3° C)	1-12	13 residue helix
14	C-RN21	-1.3613 (3° C)	1-12	13 residue helix
15	C-RN23	-1.5635 (3° C)	1-12	13 residue helix
16	C-RN28	-2.2111 (3° C)	1-13	13 residue helix
17	C-RN54	-2.4133 (3° C)	1-13	13 residue helix
18	C-RN25	-2.0180 (3° C)	4-12	13 residue helix
19	C-RN26	-1.2545 (3° C)	4-12	13 residue helix
20	C-RN80	1.5576 (3° C)	1-9	13 residue helix
21	C-RN121	1.3555 (3° C)	1-9	13 residue helix
22	C-RN119	1.0328 (3° C)	1-13	13 residue helix
23	C-RN120	1.8307 (3° C)	1-13	13 residue helix
24	C-RN84	0.2066 (3° C)	1-13	13 residue helix
25	(I+4)K,E	1.4951 (1° C)	1-17	17 residue helix
26	(I+3)K,E	0.8201 (1° C)	1-16	16 residue helix
27	3ALA	0.7000 (0° C)	2-17	17 residue helix
28	3LEU	-2.6856 (0° C)	2-17	17 residue helix
29	3PHE	-6.5404 (0° C)	2-17	17 residue helix
30	3ILE	-2.6856 (0° C)	2-17	17 residue helix
31	1VAL	-0.2261 (0° C)	2-17	17 residue helix
32	2VAL	-1.1523 (0° C)	2-17	17 residue helix
33	3VAL	-1.6904 (0° C)	2-17	17 residue helix

*: References listed in Table 8-6.

8.9 Natural Polypeptides

Peptide Pα in BPTI is one of the two subdomains important for the first crucial intermediates in the folding of BPTI (bovine pancreatic trypsin inhibitor). It contains 16 residues corresponding to residues [43-58]:

Asn-Asn-Phe-Lys-Ser-Ala-Glu-Asp-Cys-Met-Arg-Thr-Ala-Gly-Gly-Ala
43 58

which include the C-terminal and a short beta strand in BPTI (Oas and Kim, 1988). Goodman and Kim (1989) reported that at low temperature (0°C), peptide Pα5, corresponding to residues 47-58 involved in the C-terminal α-helix region of BPTI, can form an α-helix in isolation as well.

It has also been observed that short protein fragments C-peptide and S-peptide (residues 1-13 and 1-20, respectively) in ribonuclease A form observable helix content (Kim and Baldwin, 1984). C-peptide and S-peptide contain the first 13 and 20 residues of bovine pancreatic ribonuclease A (RNase A), respectively:

Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-Gln-His-Met-Asp-Ser-Ser-Thr-Ser-Ala-Ala
1 13 20

Their results show that the helix formed by S-peptide is localized to certain residues. Five residues (14-18) with low helical propensity occur in a stretch next to the 3-13 helix. At 3°C and pH 5, the C-peptide was found to have ~25% helix content (Bierzynski, et

al., 1982).

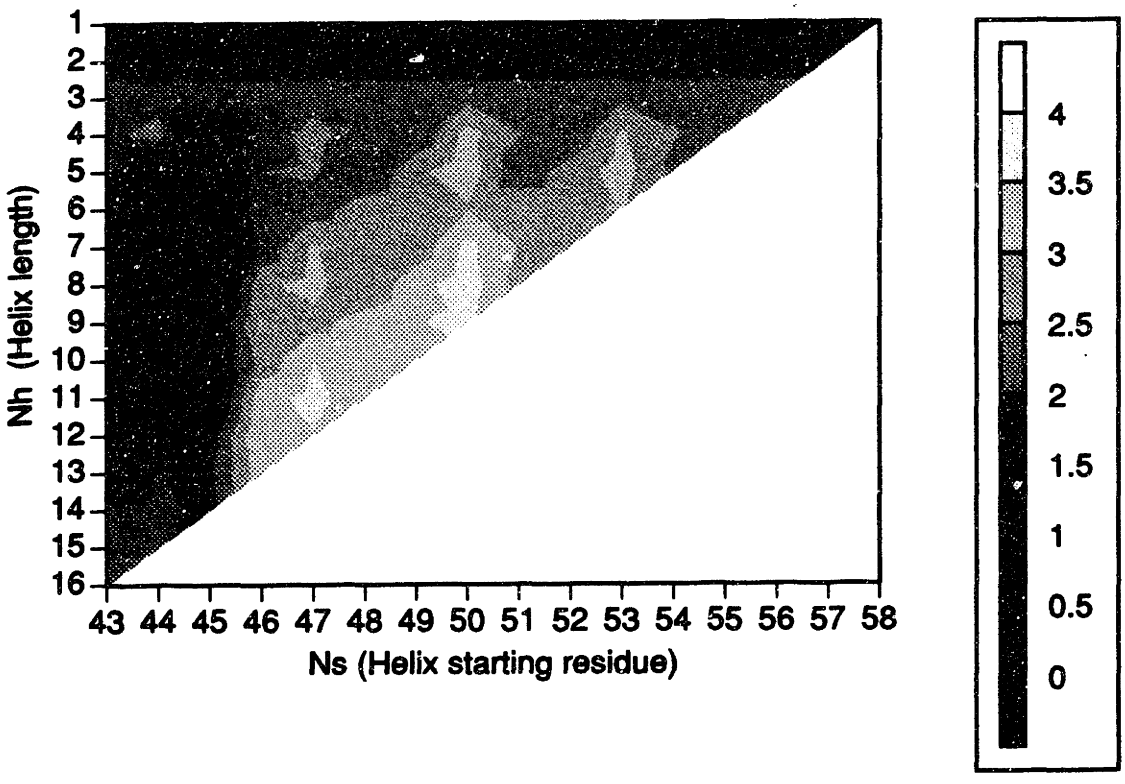
The partial helix formation in C- and S-peptide at 0°C is a typical example to study residue-residue interactions. As seen in Table 8-7, the model predictions indicate that the helix in C- and S-peptides are limited to a certain region of the chain, and that helix termination signal exists in the isolated peptides, which are in good agreement with the experimental work of Kim and Baldwin (1984). According to the values predicted in column $\Delta G_i^{c\to\alpha^*}/RT$(Intrinsic) of Table 8-3, there exist a series of residues with higher intrinsic helix-forming potentials in the chain (residues 1 to 9 except residue 3, Thr), which will favor helical structures in aqueous solution due mainly to hydrophobic interactions.

The results in Table 8-7 show that the helix in both C- and S- peptides proceeds from residues 1 to 11, which agrees with Kim and Baldwin's report that the S-peptide helix terminates before residue Thr17 and it is not possible to tell whether Met 13 is contained in the helix formed by isolated S-peptide (1-20), though in native RNase A, Met 13 is the last residue in the α-helix. From our calculation, both the solvation energy and the residue-residue interactions of residues 5 through 8 (Ala5-Ala6-Lys7-Phe8) favor α-helical conformation (see the third column in Table 8-3). Lys1 also likes to be in helical state since solvation and side-chain interactions both favor helical structure. While overally, by considering the end effect and all the binary interactions, the residues 1-11 are involved in the formation of the α-helical sequence in S-peptide. Residues 12-20 are excluded in the helix sequence of S-peptide either due to the unfavorable solvation

energy (or, residues are inherently helix breaking) and unfavorable side-chain interactions. In particular, Ala 19 and 20 do not participate in the helix formation, which is consistent with the conclusion of Kim and Baldwin (1984). Kim and Baldwin did not found measurable helix shifts of Ala19 and Ala20, which suggests that the α-helix formed by isolated S-peptide (1-20) is localized. Although Met13 and Asp14 favors helical structure, these two consecutive residues can not even make one turn of a helix, so it is impossible for residues 13 and 14 to form another helical sequence in S-peptide. In fact, residues 3-13 are helical in RNase A (Wlodawer and Sjolin, 1983), the discrepancy between our results and the conformation of C- and S-peptides in RNase A might be due to the binding of some residues of the helix to the other parts of the protein chain in RNase A.

Pα helix of BPTI represents another example of local helix formation. Table 8-6 shows that, according to the model, Pα can form an α-helix spanning from residues 45-52 at 0°C, which is very close to the conformation of the model peptide Pα in BPTI (Oas and Kim, 1988). According to our model, the net contributions of Asn43 and Asn44 to the free energy of folding is unfavorable at 0°C, so Asn43 and Asn44 are considered as helix-breakers here, and they are not in the helical region of the Pα fragment. The contour map of the Gibbs energy of folding of Pα peptide into α-helical conformation as a function of helical length and helical starting residue is given in Figure 8-6.

Figure 8-6. Contour Map of Gibbs Energy of Folding of Pα-peptide in BPTI into α-helical Conformation as a Function of Helical Length and Helix Starting Residue



Goodman and Kim (1989) reported that Pα5, corresponding to the α-helical region of BPTI (residues 47-58) forms partial helical structure at 0 °C and neutral pH. While our calculation results indicate that only residues 48-52 in Pα5 are included in the helix at 0°C. As shown in Table 8-6 that two cysteine residue (Cys51 and Cys55) in this region of native BPTI have been replaced with Ala during the calculation in accordance with the experimental work of Goodman and Kim.

In order to understand the specificity of the protein Heparinase, Sasisekharan, et al. (1991) employed an approach to be able to isolate a peptide which encompasses the binding site of the enzyme and binds substrate in a manner identical to the native protein due to the extreme difficulties to study the whole enzyme-substrate complex by solving the x-ray structure of the complex. The understanding of the stable structure of the peptide and the interactions which govern the formation and the stabilization of this structure is essential to the study of the recognition of specific enzyme-substrate binding sites. Sasisekharan et al. have synthesized HBP-1 peptide, the desired fragment of Heparinase which binds to heparin, in aqueous solution at a temperature of 37°C. The sequence of the peptide is:

H_2N-I F K K N I A H D K V E K K D K D G-COOH

Based on the sequence information, we carried out a case study on the conformation of this peptide. The model prediction results show that the 18-residue long

peptide HBP-1 forms α-helix in aqueous solution at the experimental temperature 37°C. The lowest free energy conformation corresponds to an α-helical sequence spanning from residue 1 to residue 16 of the peptide with the free energy of folding: $\Delta G^{c\to\alpha}/RT$ = -0.2549. Asp17 and Gly18 are not favored to be part of the helix at this particular temperature. The helical conformation of HBP-1 and the relative positions of the 16 residues in the helical sequence of the 18 residue peptide are illustrated in Figure 8-7. The search for the lowest Gibbs energy conformation among all the possible conformations of the polypeptide chain in the aqueous system is shown in the free energy surface contour map of HBP-1 in Figure 8-8.

8.16 Synthetic Polypeptides

Numerous recent efforts have been focused on short, synthetic peptides for helix formation studies. These peptides can serve as good model compounds for mutation studies as well. The model and the binary interaction parameters presented above not only allow us to estimate the changes in local interaction energy when a residue is folded into an α-helical conformation for polypeptides, but also can be applied to the prediction of the change in conformational free energies associated with site-directed mutation in α-helix. Around thirty synthetic peptides are chosen to test the validity of the model in this study (see Table 8-6).

Figure 8-7. Predicted HBP-1 Peptide Conformation in Water (at 310 K)

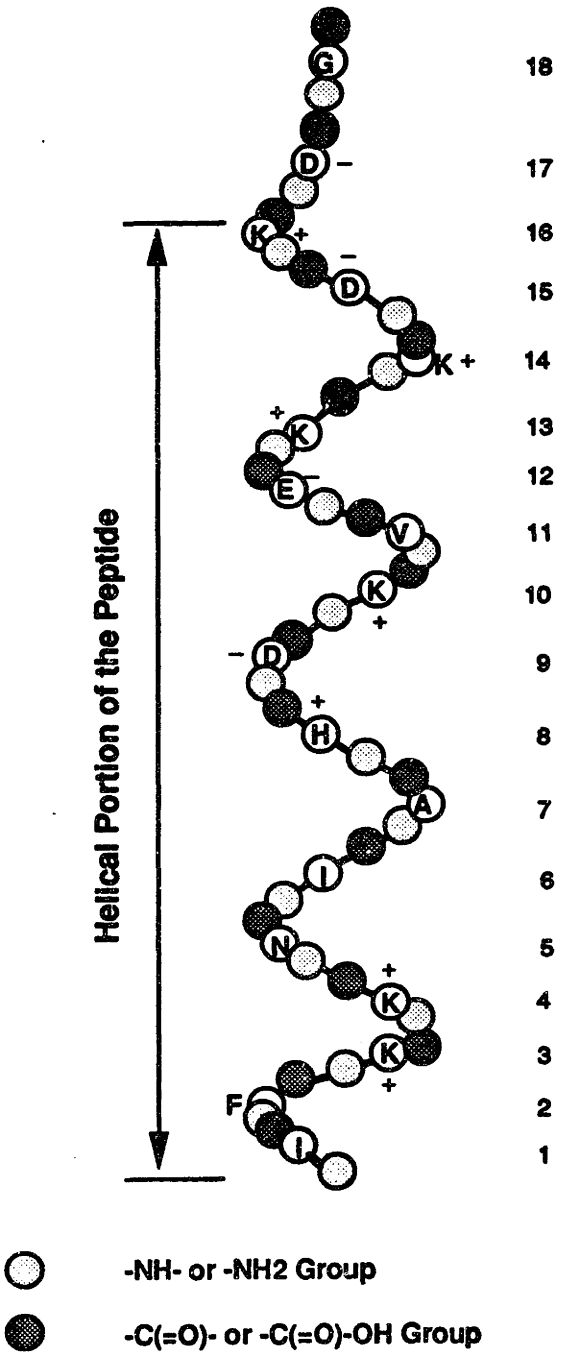
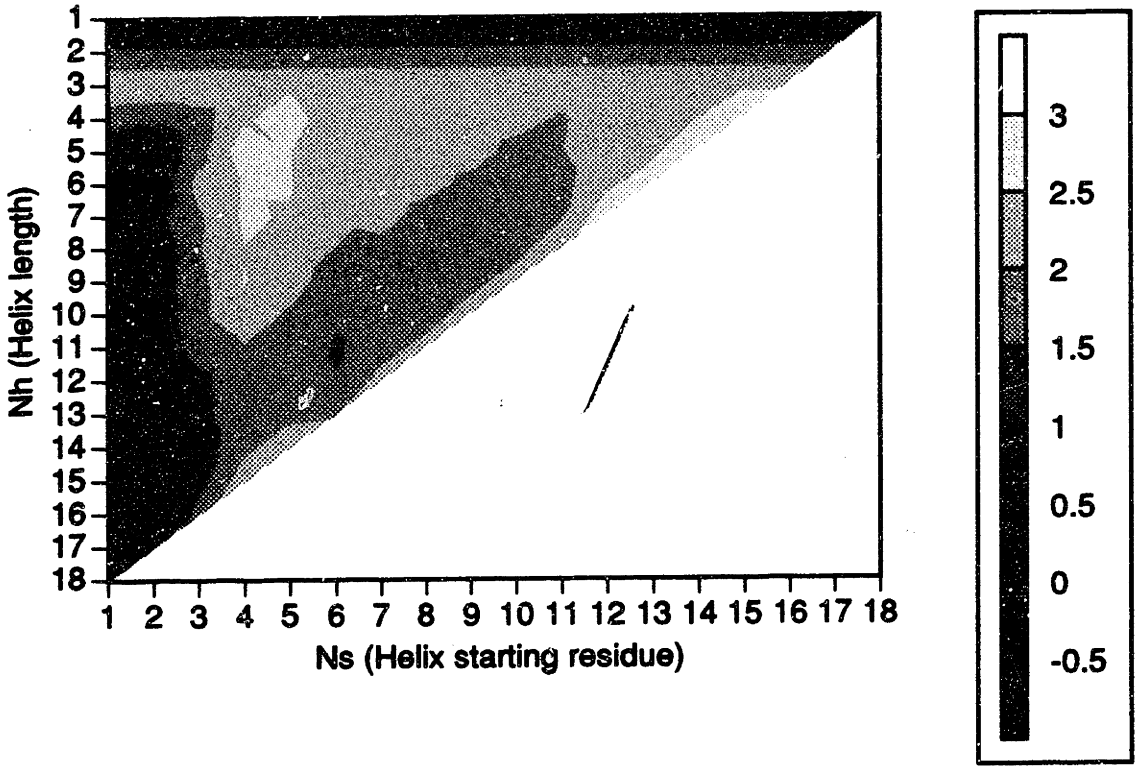


Figure 8-8. Contour Map of Gibbs Energy of Folding of HBP-1 Peptide in Heparinase into α-Helix as a Function of Helix-starting Residue



It can be seen from Table 8-7 that, in general, the model predictions for the most stable conformations of the synthetic polypeptides in aqueous solutions agree very well with documented experimental evidence of Marqusee, et al. (1987, 1989) and Padmanabhan, et al. (1990) for alanine based peptides, and of Fairman, et al. (1990) for C-peptide analogs. The model predicts both the partial helix formation and the correct helix lengths of helix sequences for almost all of these synthetic peptides.

For example, Marqusee, et al. (1989) reported that short, 16-residue, alanine-based peptides show stable α-helix formation in water, and the helices are monomolecular. The alanine-based peptides are solubilized by insertion of three or more residues of a single charge type, lysine(+) or glutamic acid(-). The 16-residue, alanine-based peptides were designed such that the results cannot be explained by ion-pair formation or by charged group-helix dipole interactions, therefore, these peptides are appropriate model systems to test our theory and the obtained interaction parameters in this work. Table 8-7 shows that the model predicts helix formation for these peptides in aqueous solution at reported experimental temperatures.

8.11 Mutation Studies of Synthetic Polypeptides

Peptide mutation is an important problem, it lays foundations for further studies in identifying active sites of enzymes and proteins. A good example here is the study on the helix stability upon amino acid substitution for the 13-residue long C-peptide analogs investigated experimentally by Fairman, et al. (1990). As seen from Table 8-7, the model indicates that, at 3°C, peptide C-RN16 helix is less stable than C-RN21 after the

substitution of Lys1 for Ala1 of the reference peptide C-RN16, which matches well with the experimentally reported helix-destabilizing replacement. The model predicted helix stability increase for the replacement of Arg10 (peptide C-RN23) to Ala10 (C-RN54) in an Ala2 background also agrees with Fairman, et al.'s measurement. According to the model, the substitution of Ala2 and Arg10 in C-RN23 by Glu2 and Ala10 (peptide C-RN28) increases the α-helicity of the peptide; the replacement of Glu2 and Ala10 in C-RN119 by Ala2 and Arg10 (C-RN121) in a Tyr8 background lowers the helicity; and the mutation of Ala2 in C-RN120 by Glu2 (C-RN119), also in Tyr8 background, escalates the helix stability. All these predictions are in good agreement with the observations of Fairman et al.

Padmanabhan, et al. (1990) synthesized a 17-residue alanine-based peptide 3Ala (see Table 8-6) as a reference for further mutation studies. The 17-residue peptides were in the same range as those of helical segments in globular proteins. By substituting each of the five different apolar amino acids (Ala, Ile, Leu, Phe, Val) in turn for alanine in a 17-residue alanine-based peptide and determining the extent of α-helix formation, they measured and derived the relative helix-forming tendencies of the five apolar amino acids. Their rank order is:

Leu > Ala > Ile > Phe > Val.

While our model, without considering the steric hinderance due to the limitation of

UNIFAC method, gives the following rank order for the five nonpolar side chains.

Phe > Leu = Ile > Val > Ala.

In summary, the model gives interesting predictions on the stable helix conformation and the length of each helical sequence in the twenty nine synthetic polypeptides investigated. The model is also able to produce promising results for the mutation effects for these polypeptides.

8.12 Temperature Effect

Generally, the magnitude of the interaction energy parameter τ's decreases with temperature, since the higher the temperature, the weaker the interactions between any two species. As temperature is raised to a certain level, all the systems may essentially become an ideal system with τ's equal to zero. The temperature dependence of τ's are given as follows:

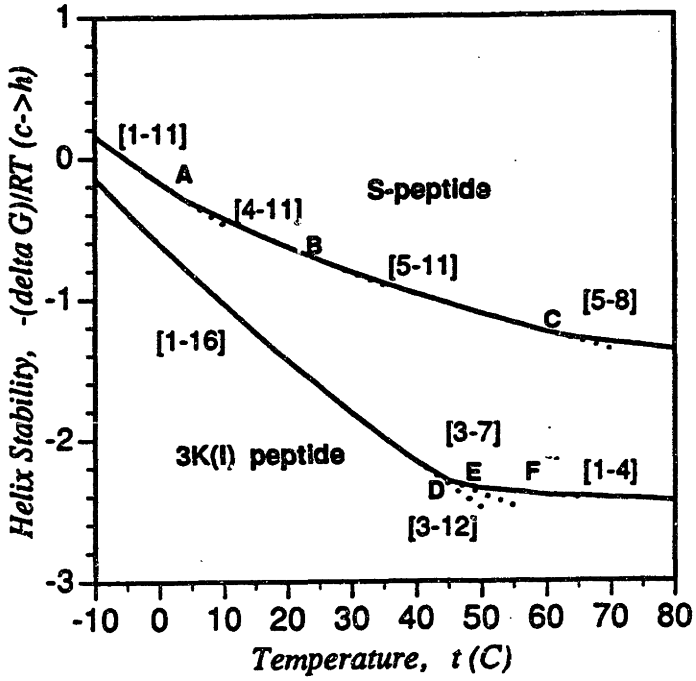
$$\tau_{ij} = a_{ij} + b_{ij}/T \tag{8-22}$$

Here, a_{ij} and b_{ij} are constants regressed within the temperature range under consideration. For simplicity, we assume that binary interaction parameters are inversely proportional to temperature (or, set a_{ij} equal to zero). So, the model suggests that the driving force for helix formation will become smaller with the increase of temperature.

Since both the interaction parameter τ's and the configurational entropy are temperature dependent, $\Delta G^{c \to \alpha}$ is a function of temperature. Figure 8-9 shows that temperature has profound influence on the conformational stability of polypeptides. The model predictions in Figure 8-9 suggest that the thermal unfolding of polypeptides does not follow the two-state mechanism, rather, the polypeptide chains unfold gradually with temperature. For example, the stable S-peptide helix [1-11] unfolds to a less stable helix [4-11] at a temperature near 5°C (point A on the first curve), and [4-11] unfolds into an shorter, much less stable helix [5-11] (point B) at 30°C, and this short helix unfolding occurs at 65°C (point C of the curve). The [5-8] helical sequence unfolds further at elevated temperatures. This is consistent with the experimental results of Kim and Baldwin (1984) that helix formation is strongly dependent on temperature, and S-peptide α-helix is essentially absent at 47°C, and is populated at 0°C, and this is also in good agreement with the experimentally obtained thermal unfolding curve for C-peptide lactone (Bierzynski, et al., 1982).

Similar thermal unfolding phenomenon is also predicted for the synthetic peptide 3K(I), a 16-residue long alanine-based peptide. The unfolding conformational transitions occur at 45°C, 50°C, and 60°C, corresponding to points D, E and F on the lower curve of the Figure, respectively. It is very important to note that the model prediction results match the experimentally observed thermal unfolding curve (Marqusee, et al., 1989) of 3K(I) peptide excellently.

Figure 8-9. Predicted Thermal Stability of Natural and Synthetic Polypeptides in Water



The same conclusions for the temperature effect on helix stability can be drawn from the comparison between the results in Table 8-7 and Table 8-8. According to the model prediction, the helix formation of peptide (I+4)E,K is temperature driven, with maximum helicity at low temperature, unfolding occurs with increasing temperature. This agrees with the experimental results of Marqusee, et al. (1987) as well.

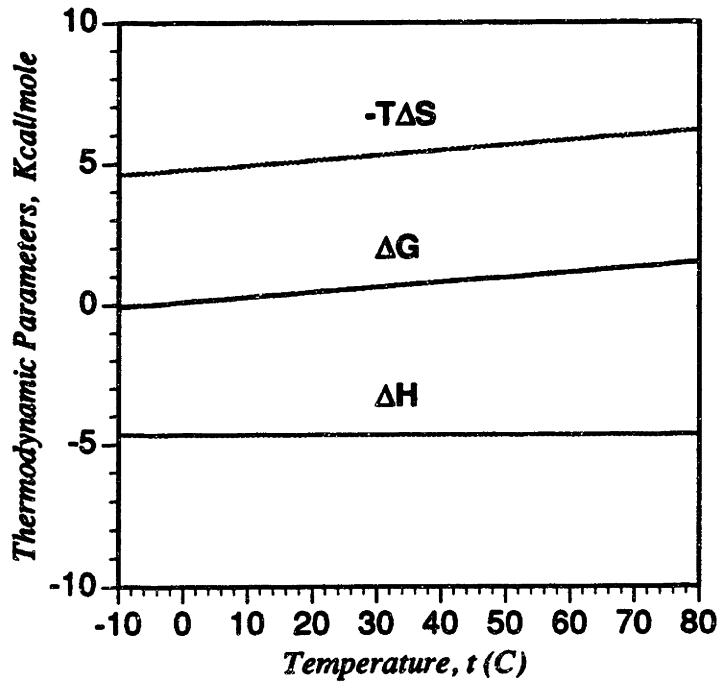
It is also interesting to see from Figure 8-9 that the model predicted thermal unfolding of the alanine-based peptides is very broad, spanning from 0°C to 70°C as Marqusee, et al. (1989) reported experimentally. Our results, however, disagree with the conclusion of Marqusee et al. that helix formation is an enthalpy driving process, the model predictions show that it is an entropy driving process instead (Figure 8-10). Unfolding increases with temperature due to the increase of the entropy term, and the magnitude of the interaction energies decreases with temperature, while the enthalpy of folding stays as constant essentially due to the simplified relationship of τ_{ij} to temperature.

Table 8-8. Model Predicted Temperature Effect on the Stable Conformations of Polypeptides

No	Peptides	$\Delta G^{c\to\alpha}/RT$	Helical Residues	
			Model	Literature*
2	S-Peptide	1.0635 (47°C)	Small helix content 5-11/20	coil
4	Pα5	2.2849 (60°C)	Small helix content 6-9/12	coil
25	(I+4)E,K	2.4179 (70°C)	Small helix content 1-4/17	Small fraction of helix

*: References listed in Table 8-6.

Figure 8-10. Model Predicted Temperature Dependence of Thermodynamic Parameters for the Folding of S-peptide in RNase A



Notation

a,b	=	temperature coefficients of binary interaction parameters
e	=	unit vector
ΔG	=	Gibbs energy
g	=	interaction energy
Δg	=	molal Gibbs energy
ΔH	=	enthalpy
Δh	=	molal enthalpy
I	=	unit matrix
i	=	any residue in a polypeptide chain
l,m	=	helical sequence index
n	=	position index of a residue in a α-helical sequence
N	=	number of residues in a polypeptide chain
N_1	=	number of solvent molecules
N_2	=	number of polymer chain molecules of x segments
N_h	=	number of residues in a α-helical sequence, or helix length
N_s	=	helix starting residue
r	=	helical sequence index
R	=	gas constant, residues
s	=	conformational index
Δs	=	molal entropy
ΔS	=	entropy
T	=	temperature, K
x	=	number of segments per polymeric chain

Greek Letters

τ	=	NRTL binary interaction energy parameter

Superscripts

α	=	α-helix
β	=	β-sheet
c	=	coil

FH	=	Flory-Huggins term
f	=	free water state (or, pure water state, unbonded to peptide chain)
s	=	solvated state of water (bonded to peptide chain)
T	=	transpose or a vector or a matrix

Subscripts

i,j	=	any species
R	=	peptide unit, or "residue"
W	=	water

Literature Cited

Aspen Technology, ASPEN PLUS User Guide, Cambridge, MA (1988).

Bierzynski, A., P.S. Kim, and R.L. Baldwin, Proc. Natl. Acad. Sci. USA, 79, 2470-2474 (1982).

Blackburn, G.M., T.H. Lilley, and E. Walmsley, "Aqueous Solutions Containing Amino Acids and Peptides," J. Chem. Soc., Faraday Trans. 1 (78), 1641-1665 (1982).

Chen, C.-C., Y. Zhu, and L.B. Evans, "Phase Partitioning of Biomolecules: Solubilities of Amino Acids," Biotechnology Progress, 5, 3, 111-118 (1989).

Chen, C.C., Y. Zhu, and L.B. Evans, "A Molecular Thermodynamic Model for the Gibbs Energy of Folding of Homo-polypeptides in Aqueous Systems," AIChE Annual Meeting, Chicago, 1990.

Cohn, E.J. and J.T. Edsall, Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Reinhold Publishing Corporation, New York, 1943.

Davidson, B., and G.D. Fasman, Biochemistry 6:1616 (1967).

Flory, P.J., J. Chem. Phys., 9, 660 (1941); 10, 51 (1947).

Fredenslund, A., R.L. Jones, and J.M. Prausnitz, AIChE Journal, 21, 1086 (1975)

Fredenslund, A., J. Gmehling, and P. Rasmussen, Elsevier Scientific Publishing Company, Amsterdam, 1977.

Goodman, E. M. and P.S. Kim, Biochemistry 28, 4343-4347 (1989).

Haber, E., and C.B. Anfinsen, J. BIOL. CHEM., 237, 1839 (1962).

Huggins, M.L., J. Phys. Chem., 9, 440 (1941); Ann. N.Y. Acad. Sci., 43, 1 (1942).

Kim, P.S., and R. L. Baldwin, Nature, 307(5949), 329-334 (1984).

Korszun, Z.R. and F.R. Salemme, Proc. Natl. Sci. USA, 74(12), 5244-5247 (1977).

Levitt, M., Ann. Rev. Biophys. Bioeng., 11, 251-271 (1982).

Marqusee, S. and R. L. Baldwin, Proc. Natl. Acad. Sci. USA, 84, 8898-8902 (1987).

Marqusee, S. and R. L. Baldwin, Proc. Natl. Acad. Sci. USA, 86, 5286-5290 (1989).

Oas, T.G. and P.S. Kim, Nature, 336 (6194), 42-48 (1988).

Padmanabhan, S., S. Marqusee, T. Ridgeway, T.M. Laue and R.L. Baldwin, Nature, 344, 268-270 (1990).

Pauling, L., R.B. Corey, and H.R. Branson. Proc. Natl. Acad. Sci. U.S. 37, 205 (1951).

Renon, H., and J.M. Prausnitz, "Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures," AIChE J., 14, (1), 135 (1968).

Sasisekharan, R., Ph.D. Thesis Proposal, Department of Chemical Engineering, M.I.T., 1991.

Scheraga, H.A., Pure & Appl. Chem., 50, 313-324, 1978.

Sueki, M., S. Lee, S.P. Powers, J.B. Denton, Y. Konishi, and H.A. Scheraga, Macromolecules, 17, 148-155 (1984).

White, F.H., and C.B. Anfinsen Ann. N.Y. Acad. Sci. 81, 515 (1959).

Wlodawer, L., and L. Sj lin, Biochemistry 22, 2720 (1983).

Zhu, Y., L.B. Evans, and C.C. Chen, "Representation of Phase Equilibrium Behavior of Antibiotics," Biotechnology Progress, 266 (1990).

CHAPTER 9

CONCLUSIONS AND FUTURE DIRECTIONS

9.1 Conclusions

Protein phase partitioning and protein folding are the two challenges facing today's biotechnology. The conformation of a protein is very important and closely related to its phase partitioning behavior. Likewise, the physical properties and the phase behavior of the protein depend strongly on its molecular conformation (the folded or unfolded state). It is the same physical interactions working in the biological solution systems that govern both the conformation of a protein and its phase partitioning.

Molecular thermodynamics can facilitate as a useful tool for us to develop an improved understanding of protein phase partitioning and folding phenomena.

Phase Partitioning of Biomolecules (Solubility of amino acids)

Amino acids, peptides, and proteins are ampholytes which exist in solutions partly as neutral dipolar species and partly as cations and anions. The neutral dipolar species are the predominant species in isoelectric solutions of amino acids and proteins. Being dipolar ions or zwitterions, these neutral dipolar species carry dual electric charges even in isoelectric solutions. The ampholyte solution chemistry and the dipolar ionic structure are the key factors characterizing the solubility behavior of the biomolecules.

As a first step toward the representation of the complex mechanisms behind the phase behavior of peptides and proteins, a molecular thermodynamic framework has been established to investigate the solubility behavior of zwitterionic biomolecules, in

particular, the amino acid and small peptides. The framework employs the chemical theory and the Electrolyte NRTL model to account for both the chemical reactions and the physical interactions of the systems. The framework has been used successfully to represent the solubility behavior and activity coefficient of amino acids and small peptides as a function of temperature, pH, solvent, salt and dipolar ion compositions.

Phase Behavior of Antibiotics

β-lactam antibiotics are amino acid derivatives. Antibiotic zwitterions share the same characteristics as amino acids in aqueous solutions. They carry charges like ions, and side chains and the two ring frame like other organic compounds.

The molecular thermodynamic framework developed for amino acids was extended to successfully describe the phase equilibrium behavior of β-lactam antibiotics. Using only binary interaction parameters in computing the activity coefficients of the true species of the antibiotic solutions, the framework successfully correlates and represents the solubilities and the phase partitioning coefficients of β-lactam antibiotics as affected by temperature, ionic strength, solvent compositions, solute compositions, and pH. Furthermore, the framework offers important physical insights in understanding the mechanisms controlling the phase equilibrium behavior of antibiotics and other biomolecules.

Secondary Structure Prediction of Homo-Polypeptides

Under physiological conditions, many polypeptide chains spontaneously fold

into discrete and tightly packed three-dimensional structures. The folded polypeptide chain conformation is believed to represent a minimum Gibbs energy of the system, governed by the weak interactions that operate between the amino acid residues and between the residues and the solvent.

Molecular thermodynamics has been well-established as a practical method to derive semi-empirical expressions for the excess Gibbs energy of mixtures of small molecules. We have successfully applied the molecular thermodynamic approach and developed a model for the Gibbs energy of folding of aqueous homo-polypeptides. The model takes into consideration the residue-residue, residue-solvent and solvent-solvent binary physical interactions along with the local compositions of amino acid residues in aqueous homo-polypeptides upon the chain folding, and the configurational entropy loss accompanying the folding. The UNIFAC group contribution method was used to estimate the binary interaction parameters between water and residues, and ethanol and residues.

The resulting model has been shown to generate critical results that are consistent with the important observations reported in the literature for the folding of the aqueous homo-polypeptides. The model also yields a hydrophobicity scale for the twenty amino acid side chains which compares favorably with established experimental scales.

Modeling of Polypeptide Chain Folding

Polypeptides (both natural and synthetic) represent a large category of biopolymers with both great biomedical interest and theoretical importance as model

compounds for the studies of protein folding. The diversity of the twenty naturally-occurring amino acid side chains makes the phase behavior of polypeptides more complicated than ordinary polymers. In particular, the prediction of polypeptide conformation remains elusive.

The molecular thermodynamic model for aqueous homopolypeptides is generalized for the free energy of folding of polypeptides from coiled conformation into an α-helical conformation in aqueous solution. The model clearly identifies two important enthalpic contributions to the Gibbs energy of folding, which are the intrinsic helix-forming potential of the individual residues in a specific solvent and the cooperative residue-residue interactions in the polypeptide chain of a given amino acid sequence. All the interaction energy parameters between the twenty amino acid residues are also obtained using the UNIFAC group contribution method.

Case study results with the molecular thermodynamic model are in good agreement with the experimental observations in literature that short protein fragments and alanine-based synthetic peptides can fold in aqueous solutions and their structures are stabilized by local interactions (solvation and residue-residue interactions), which indicate that the model and the interaction parameters presented here achieved success in predicting polypeptide chain conformation.

Amino acid substitution simulations are performed to test the cooperativity of the folding and to investigate the sensitivity of the polypeptide conformational stability via site-directed mutations. The model gives satisfactory results for a number of peptides in

this sensitivity study.

The model also successfully predicts thermal unfolding of polypeptides. The results from the model predictions suggest that the thermal unfolding of polypeptide chains is a gradual conformational transition, instead of a two-state scheme.

In summary, we believe molecular thermodynamic modeling provides a sound theoretical approach to help better understand the conformational and phase behaviors of polypeptides and proteins in solutions. The efforts made in this thesis work have improved our understanding on the complexities of protein solutions, and established essential foundations for further studies in modeling protein conformation and phase partitioning. The theories, the methodologies and the interaction parameters developed here can be of importance to protein refolding, denaturation, downstream processing, and related studies.

9.2 Future Directions

The recommendations for the future work after this thesis will be the following:

(1). Incorporate the electrostatic interactions into the model and refine interaction energy parameters.

(2). Apply the model to investigate the effects on polypeptide conformation by solvent such as denaturants (Urea, GuHCl, etc.), surfactants, polymers (PEG, etc.), and other organic solvents, co-solvents, salts, pH, etc.

(3). Extend the model to incorporate disulfide bond formation to predict native

protein conformation, to model protein folding and denaturation.

(4). Extend the model and the methodology developed in this work to study protein-protein interactions and protein-surface interactions (binding of polypeptides and proteins on polymers and other solid surfaces). The model can help better understand the mechanisms behind protein aggregation.

(5). Develop models for phase partitioning of polypeptides and protein molecules.