

**STRUCTURE, FUNCTION AND AGGREGATION KINETICS
IN SALT-INDUCED PROTEIN PRECIPITATION**

Thesis by
Todd Michael Przybycien

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1989

(Submitted March 10, 1989)

© 1989

Todd Michael Przybycien

All Rights Reserved

To my parents
Edward and Diana,
with love.

I'm not aware of too many things

I know what I know if you know what I mean

Choke me in the shallow water

before I get too deep

What I am is what I am

are you what you are or what?

Don't let me get too deep

- What I am, *Edie Brickell and the*

New Bohemians

Keep your tips up!

- *Mammoth Mountain*

ACKNOWLEDGEMENTS

I firmly believe that the individual is largely the cumulative product of interactions with others. To this end, I offer my sincere appreciation to a few of the many special people who provided inspiration for the present work.

A number of people were instrumental in preparing me for the rigors of Caltech. My parents instilled in me a love of the pursuit of knowledge and the drive to persevere. I draw great strength from them and to them this thesis is dedicated. Grant Carey was among a particularly devoted group of teachers at Churchville-Chili High School who reinforced my parents' guidance. My eyes and mind were opened as an undergraduate at Washington University; I thank the Engineering School for providing the wherewithal to pursue my education. My advisor, Dr. Robert Powell, fueled my interest in chemical engineering and in graduate study.

At Caltech, my development as an independent investigator and the present work derive from the support of my thesis advisor, Dr. Jay Bailey. To be afforded the opportunity to collaborate with a man of vision is a great honor. Helpful insight to sundry questions and general enthusiasm concerning my research were provided by the complement of my thesis committee, Drs. Frances Arnold, Sunney Chan, and Jack Richards.

My peers at Caltech have been a source of amusement and amazement. It has been a pleasure working with the post-docs and pre-docs of BIT. Office-mates Roger (Rog-man) Hart and Mike (Mike-man) Kosinski have been excellent sounding boards and have endured an endless fusillade of banter. Hiking (surviving), canoeing (drowning) and skiing with them have renewed my appreciation (fear) of nature. I also owe my skills at dartsmanship and cravings for jalapeños to the Rog-man. Rog and Mike, non illēgitimus carborundum. Anne (Oh my

God!) McQueen, the reigning champion of the seven-course petit-déjeuner, has also been a very supportive and valued friend.

Roommates, John (Johnny my Baby) Brewer and Scott (Scotch) King, and I have spent innumerable good times playing 9-5-2, bowling, irritating American Airline stewardesses, darting, and goofing-off in general. John has never let the fact that he's Canadian hold him down; and I still maintain that Scott can mix a libation capable of blowing your expletive off. I can only hope that I have been an inspiration in the proper use of colorful adjectives. With roommates like these, who needs the IRS? Guest appearances were put in by Holden (Opie) Thorpe and David (The Huang Man for the Job) Huang.

Special mention is due Frank (Frankster) Doyle and the rest of the St. Francis Street crew and assorted (sordid?) hangers-on who provided a little bit of Venice Beach in San Gabriel. I have enjoyed being part of the gang. Go ∇^2 !

Thanks beyond words are reserved for Valerie (Poopie, Bear-feed, various) Patrick. Valerie has been a model of diligence as a fellow graduate student in chemical engineering and a constant source of love and support. She has taught me that manamana is not to be put off until mañana. I look forward to starting the next phase of my life with her.

ABSTRACT

Salt-induced precipitation is a biological separation technique that exposes proteins to unnatural environments. Macromolecular-scale issues of activity, structure, and aggregation have been addressed as a function of governing parameters.

The effects of salt type and concentration on protein solubility and recoverable activity were studied using α -chymotrypsin (α CT) as a model protein and five salts spanning the lyotropic series. Unaccounted for salt-protein interactions and changes in protein physical properties were the likely source of discrepancies between the experimental and theoretical solubility behavior. Active protein recovery was a function of salt type, but not concentration. A salting-out performance parameter was identified; an optimum salt may exist for a particular protein.

α CT precipitates from the solubility-activity study were examined for perturbations in secondary structure via Raman spectroscopy and in active site tertiary structure via electron paramagnetic resonance spectroscopy. NaBr, KBr, and KSCN-induced precipitates had increased β -sheet and decreased α -helix contents; these changes were correlated with active protein yields. Spectra of spin-labelled precipitates indicated that the active site remains intact. Molecular modelling was used to estimate changes in the dipole moment and hydrophobic surface area for the altered precipitates. A general mechanism for the precipitation of globular proteins was proposed.

The generality of secondary structure changes was explored for twelve different proteins via Raman spectroscopy. KSCN-induced precipitates exhibited increased β -sheet and decreased α -helix contents; structural changes for Na₂SO₄-induced precipitates were less significant. The β -sheet increase may occur at the

expense of α -helix segments. β -sheet increases were correlated with the fraction of charged residues and the surface area of the native protein. α -helix decreases were correlated with the dipole moment and helical content of the native protein.

The effects of temperature, protein concentration, salt type, and salt concentration on α CT aggregation kinetics were studied. Stopped-flow turbidimetry indicated that temperature and salt concentration effects are exerted through changes in protein solubility. Protein concentration effects are well-described by Smoluchowski's collision equation. The aggregation of partially inhibited α CT demonstrated poisoning behavior. Solute particle radius distributions determined by dynamic laser light scattering indicated that aggregation depends on the supersaturation. A detailed population balance model, accounting for specific and nonspecific quaternary interactions, was developed.

TABLE OF CONTENTS

| | |
|--|-----------|
| Acknowledgements | v |
| Abstract | vii |
| List of Tables | xv |
| List of Figures | xvii |
| | |
| CHAPTER 1 Introduction | 1 |
| 1.1 Background | 1 |
| 1.2 Bioprocess Separation Operations | 1 |
| 1.3 Formation and Recovery of Protein Precipitates | 5 |
| 1.4 Thesis Scope | 6 |
| 1.5 References | 10 |
| | |
| CHAPTER 2 Solubility-Activity Relationships in the Inorganic Salt-Induced Precipitation of α-Chymotrypsin | 13 |
| 2.1 Abstract | 14 |
| 2.2 Introduction | 15 |
| 2.3 Theoretical Background | 16 |
| 2.3.1 <i>Thermodynamics of precipitate formation</i> | 15 |
| 2.3.2 <i>Empirical salting-out equation</i> | 20 |
| 2.4 Materials and Methods | 20 |
| 2.4.1 <i>Materials and experimental conditions</i> | 20 |
| 2.4.2 <i>Preparation of samples for solubility analysis</i> | 21 |
| 2.4.3 <i>Solubility assay</i> | 21 |
| 2.4.4 <i>Preparation of precipitate for activity analysis</i> | 22 |

| | | |
|---|--|----|
| 2.4.5 | <i>Activity assay</i> | 23 |
| 2.4.6 | <i>Active fraction assay</i> | 23 |
| 2.5 | Results and Discussion | 24 |
| 2.5.1 | <i>Theoretical predictions</i> | 24 |
| 2.5.2 | <i>Solubility determinations</i> | 27 |
| 2.5.3 | <i>Comparison of empirically and theoretically derived salting-out constants</i> | 29 |
| 2.5.4 | <i>Precipitate activity determinations</i> | 33 |
| 2.5.5 | <i>Optimization strategy</i> | 35 |
| 2.6 | Conclusions | 37 |
| 2.7 | Acknowledgement | 38 |
| 2.8 | Nomenclature | 39 |
| 2.9 | References | 41 |
| 2.10 | Tables | 46 |
| 2.11 | Figures | 48 |
| CHAPTER 3 Structure-Function Relationships in the Inorganic Salt-Induced Precipitation of α-Chymotrypsin | | |
| 3.1 | Abstract | 59 |
| 3.2 | Introduction | 60 |
| 3.3 | Materials | 61 |
| 3.4 | Methods | 62 |
| 3.4.1 | <i>Preparation of precipitate samples</i> | 63 |
| 3.4.2 | <i>Functional characterization of precipitate</i> | 64 |
| 3.4.3 | <i>Raman spectroscopy</i> | 64 |

| | | |
|------------------|--|------------|
| 5.1 | Abstract | 138 |
| 5.2 | Introduction..... | 139 |
| 5.3 | Materials and Methods..... | 142 |
| | 5.3.1 <i>Materials</i> | 142 |
| | 5.3.2 <i>Inhibition of αCT with TPCK</i> | 142 |
| | 5.3.3 <i>Stopped-flow turbidimetry</i> | 143 |
| | 5.3.4 <i>Solubility determinations</i> | 145 |
| | 5.3.5 <i>Dynamic laser light scattering measurements</i> | 146 |
| 5.4 | Experimental Results..... | 147 |
| | 5.4.1 <i>Apparent aggregation rate</i> | 147 |
| | 5.4.2 <i>Particle size distributions</i> | 150 |
| 5.5 | Kinetics Modelling | 153 |
| | 5.5.1 <i>Model development</i> | 153 |
| | 5.5.2 <i>Kinetic rate expressions</i> | 156 |
| | 5.5.3 <i>Method of solution</i> | 158 |
| | 5.5.4 <i>Modelling results</i> | 160 |
| 5.6 | Discussion..... | 163 |
| 5.7 | Acknowledgements | 167 |
| 5.8 | Nomenclature | 168 |
| 5.9 | References..... | 170 |
| 5.10 | Tables..... | 176 |
| 5.11 | Figures..... | 177 |
| CHAPTER 6 | Conclusions | 185 |
| 6.1 | Summary of Results | 186 |
| 6.2 | New Directions..... | 191 |

| | | |
|--|--|-----|
| 6.2.1 | <i>Mixing effects on protein solubility</i> | 191 |
| 6.2.2 | <i>Secondary structure estimation via Raman spectroscopy</i> | 195 |
| 6.2.3 | <i>Aggregation kinetics</i> | 198 |
| 6.2.4 | <i>Protein denaturation in biological separation processes</i> | 199 |
| 6.3 | References..... | 201 |
| Appendix A Thermostated Sample Stage for Raman Spectrometer | | |
| | | 204 |
| Appendix B Raman Spectral Analysis Package (RSAP) | | |
| | Documentation | 206 |
| B.1 | Introduction..... | 207 |
| B.2 | Package Architecture | 208 |
| B.3 | Running RSAP | 209 |
| B.4 | Program Description | 210 |
| | <i>B.4.1 RSAP operating programs</i> | 210 |
| | <i>B.4.2 RSAP analysis programs</i> | 210 |
| | <i>B.4.3 Housekeeping programs</i> | 214 |
| B.5 | References..... | 215 |
| B.6 | Table | 216 |
| B.7 | Figure..... | 217 |
| B.8 | RSAP Source Code Listings | 218 |
| Appendix C Stopped-Flow Device..... | | |
| | | 241 |
| C.1 | Schematic Diagram Description..... | 242 |

| | |
|--|------------|
| C.2 Operating Instructions | 242 |
| C.3 Figure: Stopped-Flow Schematic..... | 244 |
| Appendix D Shimadzu UV-260 Data Acquisition Program | 245 |
| D.1 Source Code Listing | 246 |
| Appendix E Aggregation Kinetics Model..... | 254 |
| E.1 Source Code Listing | 255 |
| Appendix F Immunoaffinity Chromatography Project | 262 |
| F.1 Introduction..... | 263 |
| F.2 Coupling Chemistry | 264 |
| F.3 NMR Spectroscopy | 266 |
| F.4 Proposed Synthetic Route | 267 |
| F.5 Experimental Apparatus..... | 269 |
| F.6 Proposed Experiments..... | 269 |
| F.7 References..... | 272 |
| F.8 Figures..... | 275 |

LIST OF TABLES

Chapter 2

- 1 Salt cuts with corresponding α CT solubility ranges 46
- 2 Estimated parameters of salting-out equation 46
- 3 Dehydrated hydrophobic surface areas computed for α CT from
experimental salting-out constants 47

Chapter 3

- 1 Salt cuts with corresponding α CT solubility ranges 92
- 2 salt molal surface tension increments 92
- 3 Control analyses for the Raman Spectral Analysis Package 93
- 4 Raman analysis of α CT precipitate samples 94
- 5 Calculated physical properties for native and altered α CT
structures 95

Chapter 4

- 1 Proteins studied 131
- 2 Precipitate secondary structure estimates 132
- 3 Correlation between secondary structure estimates and spectral
noise 134
- 4 Precipitate structure changes and physical properties of native
proteins 135
- 5 Correlation between estimated secondary structure changes and
elements of primary, secondary, and tertiary structure 136

Chapter 5

- 1 Molar heats of solution for α CT in various salt solutions 176
- 2 Rate constant estimates for population balance model 176

Appendix B

1 RSAP file extension codes 216

LIST OF FIGURES

Chapter 2

- 1 Protein relative surface hydrophobicity as a function of the frequency of charged groups 51
- 2 Comparison of experimental data with the theoretical prediction of the salting-out constant as a function of the salt molal surface tension increment.....52
- 3 α CT solubility curves for KSCN, KBr, NaBr, NaCl, and Na_2SO_4 solutions.....53
- 4 Apparent salting-out constants for NaBr solutions as a function of the dilution factor 54
- 5 Calculated salt interaction parameter as a function of salt cut ... 55
- 6 α CT precipitate activity, active fraction, and specific activity as a function of salt cut for KSCN, KBr, NaBr, NaCl, and Na_2SO_4 solutions.....56
- 7 Precipitate active fraction as a function of salt type 57
- 8 Salting-out performance parameter as a function of salt type 58

Chapter 3

- 1 Raman amide I band spectrum of native α CT.....99
- 2 Chemical structure of α CT active site spin label 100
- 3 Solid phase Raman amide I band control spectra: lysozyme, trypsin, and elastase.....101
- 4 Subtracted Raman amide I band spectra of first cut α CT precipitate samples 102
- 5 Secondary structure content of α CT precipitate samples as a

| | |
|--|-----|
| function of salt type | 103 |
| 6 Correlation of whole α CT precipitate active fraction with change in total β -sheet content | 104 |
| 7 Conventional EPR spectra of free spin label and spin labelled α CT in solution | 105 |
| 8 Solid phase conventional EPR spectra of spin-labelled α CT and KSCN and Na_2SO_4 -induced precipitates | 106 |
| 9 Solid phase ST-EPR spectra of spin-labelled α CT and KSCN and Na_2SO_4 -induced precipitates | 107 |
| 10 C^α backbone of native α CT | 108 |
| 11 C^α backbone of modified α CT | 109 |
| 12 Proposed precipitation mechanism | 110 |

Chapter 5

| | |
|---|-----|
| 1 α CT aggregation turbidity trajectories as a function of KSCN concentration | 179 |
| 2 α CT aggregation turbidity trajectories as a function of temperature | 180 |
| 3 α CT and TPCK- α CT aggregation turbidity trjectories as a function of protein concentration | 181 |
| 4 Apparent Smoluchowski rate constant versus the initial supersaturation ratio | 182 |
| 5 Reduced solute weight-average radius distribution versus α CT concentration | 183 |
| 6 Detailed population balance reaction scheme | 184 |

Appendix A

| | |
|---|-----|
| 1 Thermostatted sample stage for Raman spectrometer | 205 |
|---|-----|

Appendix B

1 RSAP directory structure 217

Appendix C

1 Stopped-flow schematic diagram 244

Appendix F

1 Synthetic scheme for methylphosphoric-*bis*-imidazolide 275

2 90 MHz ³¹P NMR spectrum of methylphosphonic-*bis*-
imidazolide 276

3 Synthetic scheme for trial matrix activation 277

4 90 MHz ³¹P NMR spectrum of derivatized Sepharose 278

5 300 MHz ³¹P NMR spectrum of derivatized Sepharose 279

6 Synthetic scheme for immunosorbent preparation 280

7 *in situ* NMR chromatography column design 281

Chapter 1

Introduction

1.1 Background

Salt-induced precipitation is the oldest and perhaps most widely used protein purification procedure in the separations repertoire of the bioprocess industry. The reduction of protein solubility in aqueous solutions accompanying the addition of neutral salts was first described in the late nineteenth century by Hofmeister [1,2]. Yet despite the relative maturity of this process, the salt and protein properties governing solubility remain obscure; the nature and extent of salt-protein interactions in solution are not well understood. Moreover, the dearth of predictive ability for precipitation performance is an impediment to the design and operation of separation facilities. To address this issue, fundamental aspects of protein behavior in high salt environments and physical properties of protein precipitates have been explored in this work as a function of relevant process parameters. The rôle of precipitation in bioprocesses varies and is reflected in the means used to generate and recover protein precipitates; thus, a discussion of the general organization of bioprocesses and the mechanics of precipitation operations germane to the present work follows.

1.2 Bioprocess Separation Operations

The importance of bioprocesses is well established. In 1985, the total world market for the products of biotechnology was approximately two billion dollars; estimates of the future market are fifteen and fifty-six dollars for 1990 and 2000, respectively, assuming a 6% annual inflation rate [3]. Products, typically proteins, are targeted for health care, plant agriculture, energy, and environmental applications [4] and are produced on a wide range of scales [5].

The final cost of a (bio)chemical product is often determined by the efficiency of the separation processes used in its manufacture. This is underscored

by the inverse relationship found to exist between the initial product concentration in the process stream and its market price; the graphical representation of this relationship has been termed the “Sherwood plot.” For products valued at 10 \$ kg⁻¹ or less, separations costs become the dominant factor [6]. In addition, up to 70% of the total capital equipment costs of a bioprocess may be related to separations. Since many conventional separation processes are inappropriate for biological materials, scaled-up bench analytical processes, with suboptimal efficiencies, are often used [5].

A bioprocess can be divided into two major phases: upstream processing, consisting of fermentation and homogenization (for intracellular products); and downstream processing, consisting of separation operations. Upstream processing, for both intracellular and extracellular products, has a significant impact on the downstream processing requirements. Important fermentation variables include the time required to grow the cells, pigment production by the cells, and the ionic strength and composition of the culture media, particularly if complex media are employed [6]. For recombinant proteins, the level of expression is important [7].

Crude extracts resulting from upstream processing of intracellular products are typically dilute aqueous solutions; the desired protein is usually a small fraction of the the total protein present and may be contaminated with nucleic acids, lipids, and carbohydrates [8]. Extracellular products may have far fewer contaminants, but may be significantly more dilute than intracellular products. At this stage, and in all subsequent operations, prudent handling is required. Proteins are often sensitive to extremes of pH, ionic strength, shear, temperature, and reagent concentrations. If processing conditions deviate greatly from those of the protein’s native environment, substantial yield losses in the form of

irreversible denaturation may occur.

Although the organization of downstream processing operations depends heavily on the initial composition of the process stream and the nature of the desired protein [9], it is possible to identify four generic operations common to most separation trains [10]: particulate removal, primary isolation, purification, and final product isolation. In the first stage, cellular debris and unbroken cells are removed and product purity may range from 0.1 to 2.0% . During primary isolation, non-protein contaminants are removed and purity may reach 1.0 to 10% . The performance of primary operations is critical as they determine the nature and extent of subsequent processing steps. Purification involves the elimination of non-product proteins with resulting purities of 50 to 80% ; this step may be particularly difficult if contaminant proteins with properties similar to the desired product are present. The final product isolation stage is composed of concentration and polishing operations. At this point, product purity is dictated by the end-use, ranging in increasing stringency from industrial enzymes, to food proteins, to injectables.

Precipitation is largely employed as a primary isolation procedure [10,11]. It may also be used at the polishing stage as a concentration operation; a net concentration can be achieved by resuspending precipitate with a solvent volume smaller than that of the initial solution. The potential for simultaneous purification and concentration, ease of operation, and relative stability of many precipitate suspensions, allowing the possibility of a holding step, motivate the use of precipitations. Other commonly used primary isolation operations include two-phase liquid-liquid extractions, sorption, and ultrafiltration [10].

1.3 Formation and Recovery of Protein Precipitates

Protein precipitation may be induced by the addition of a number of different reagents acting by different mechanisms [10,13]: high concentrations of neutral salts compete with protein for water of solvation [14]; acid or base addition to adjust the solution pH to the protein isoelectric point reduces electrostatic repulsions between proteins [15]; organic solvents reduce electrostatic repulsions by lowering the solution dielectric constant [16]; nonionic polymers, such as polyethylene glycol, compete with protein for water of solvation via an excluded volume effect [17]; polyelectrolytes interact nonspecifically with charged residues in proteins, serving as flocculating agents [18]; polyvalent metal cations may associate with negatively charged amino acid side chains and/or coordinate to exposed imidazole and thiol groups [19]; and bifunctional ligands may specifically cross-link proteins with multiple cofactor/inhibitor binding sites, much like antigen-antibody interactions, in affinity precipitations [20].

Criteria for the selection of a particular precipitant include the protein structural stability, ease of reagent handling, reagent costs and efficiency, fractionation ability, and the product end-use. The general relationship between reagent concentration and protein solubility may be expressed in power law form [13]

$$\ln S = \ln S_o + KR, \quad (1)$$

where S and S_o are the protein solubilities at reagent concentration R and in the absence of reagent, respectively, and K is an effectiveness parameter often depending on environmental conditions. S_o may also have a dependence on the geometry of the process equipment. Note that precipitation depends on the absolute concentration of protein; if the protein concentration is less than S , then

no precipitation will occur at reagent concentration R . Combinations of different techniques are often used; salting-out operations are typically performed at the isoelectric pH. The empirical nature of equation (1) precludes the *a priori* determination precipitation conditions for different reagents.

Salt-induced precipitation offers advantages relative to other means of generating precipitates as salts are relatively inexpensive, are safe to handle, have some fractionation ability, and tend to minimize non-specific interactions between different proteins [13]. Important process parameters affecting salting-out include pH, temperature, protein concentration, and salt type [21]. Also, the rate of salt addition and the mixing conditions affect both protein solubility and precipitate morphology [8]. Inefficient mixing increases the potential for denaturation through locally high salt concentrations [13]. Rapid reagent addition results in smaller precipitate particle sizes [22].

Precipitates may be recovered via filtration, sedimentation or floatation, and centrifugation [13]. Centrifugation is the method of choice as filtration operations are plagued by fouling problems and sedimentation or floatation operations may require long times to attain equilibrium. The ease and efficiency of solid-liquid separation depends on the particle size, the relative densities of the precipitate and the solvent, and the solvent viscosity [8]. Higher salt concentrations, while increasing the fraction of protein precipitated at a given protein concentration, typically increase the solution viscosity and density, making recovery operations more difficult.

1.4 Thesis Scope

The intent of this work is to examine the microscale implications of the process parameters controlling salt-induced precipitation. A large body of literature

exists on mesoscale phenomena, concerning primarily hydrodynamic effects (see [18] and references in Chapter 5), and is complimentary to the approach taken in the present work. Microscale connotes processes occurring on a molecular scale and mesoscale refers to equipment scale phenomena [3]. An understanding of both regimes is crucial to the rational design and operation of precipitation processes. To this end, this work consists of a concerted study of protein structure, function, and aggregation kinetics in salt-induced precipitation.

Chapter 2 describes a study of the relationship between protein solubility and the yield of active protein on redissolution of precipitate; this relationship was examined as a function of salt type for a series of salts that span the range of salting-out potentials. α -chymotrypsin (α CT) was used as a model globular protein. Experimental data was used to assess the predictive ability of the current equilibrium theory for protein salting-out. The putative trade-off between the amount of salt used to induce precipitation and the specific activity of redissolved precipitate was explored. Finally, a parameter describing the active protein yield per unit of salt was defined and the relative performance of the salts used was assessed on that basis.

The effects of precipitation on protein structure are explored in Chapter 3. Again, α CT was used as the model protein. The solubility measurements from Chapter 2 guided the selection of precipitation conditions. The effects of the type and amount of salt on elements of secondary and tertiary structure were analyzed via spectroscopic techniques. Secondary structure estimates were obtained by an examination of amide I band Raman spectra. Tertiary structure effects on spin-labelled α CT precipitates were probed by conventional and saturation transfer electron paramagnetic resonance. Molecular modelling was used in an attempt to incorporate spectral observations into a structure

for α CT precipitates. Computed changes in protein physical properties for the model precipitate were used to explain the discrepancies between experimental and predicted solubility behavior found in Chapter 2. A generalized mechanism describing the conformational and phase changes associated with precipitation was also suggested.

Secondary structure perturbations in a variety of different protein precipitates are discussed in Chapter 4. A number of different proteins representing a range of native structures were precipitated by a chaotropic salt and a structure preserving salt. Secondary structure content was probed by the Raman spectroscopy technique described in Chapter 3. The generality of the results found for α CT were explored and extended. Structural perturbations were compared with elements of primary, secondary, and tertiary structure to determine the protein properties responsible for salting-out behavior.

Chapter 5 comprises an experimental and theoretical description of the aggregation kinetics in protein precipitation. Stopped-flow-turbidimetry was used to follow the rate of aggregation of α CT as a function of the salt type and concentration, protein concentration, and temperature. The kinetic effects of these parameters were compared on the basis of the initial protein supersaturation ratios. Particle size distributions for α CT solutions and supernatants were estimated via dynamic laser light scattering. A kinetic model was formulated in terms of a detailed population balance; conformational changes and both non-specific and specific interactions were considered. This approach may serve as a paradigm for the investigation of other protein aggregative phenomena.

Chapter 6 is a recapitulation of the salient results of Chapters 2 through 5. The successes and limitations of the present work provide the basis for recommendations for further study included in Chapter 6. The appendixes provide

supporting information and source code listings for the more involved data acquisition, analysis, and modelling aspects of the work.

1.5 References

- 1 Hofmeister, F., "Zur Lehre von der Wirkung der Salze," *Arch. Exp. Path. Pharm.* **24**, 247-250 (1887-88)
- 2 Hofmeister, F., "Ueber die Darstellung von krystallisirtem Eieralbumin und die Krystallisirbarkeit colloider Stoffe," *Z. Physiol. Chem.* **14**, 165-171 (1890)
- 3 National Research Council, *Frontiers in Chemical Engineering, Research Needs and Opportunities*, National Academy Press, Washington D.C. (1988)
- 4 National Science Foundation, *Biotechnology Research and Development Activities in Industry: 1984 and 1985*, NSF 87-311, Washington, D.C. (1987)
- 5 National Research Council, *Separation and Purification, Critical Needs and Opportunities*, National Academy Press, Washington, D.C. (1987)
- 6 Darbyshire, J., "Large Scale Enzyme Extraction and Recovery," *Topics in Enzyme and Fermentation Biotechnology* **5**, 147-186 (1981)
- 7 Gray, P.P., "Product Isolation and Downstream Processing of Recombinant Organisms," in *The University of Queensland Biotechnology Unit, Biotechnology Training Courses*, Chapter 10 (1984)
- 8 Street, G., "Large Scale Industrial Enzyme Production," *CRC Critical Reviews in Biotechnology* **1** 59-85 (1984)
- 9 Ladisch, M.R. and Voloch, M., "A Perspective in Bio-Separation Processes," report prepared for the Workshop in Research Needs in Renewable Materials Engineering, sponsored by the National Science Foundation (1983)
- 10 Bailey, J.E. and Ollis, D.F., *Biochemical Engineering Fundamentals*, 2nd ed., McGraw-Hill, New York, 727-753 (1986)
- 11 Dunnill, P., "Trends in Downstream Processing of Proteins and Enzymes," *Process Biochemistry* **18**, 9-13 (1983)

- 12 Scopes, R.K., *Protein Purification, Principles and Practice*, Springer-Verlag, New York, pp. 48-51 (1982)
- 13 Bell, D.J., Hoare, M., and Dunnill, P., "The Formation of Protein Precipitates and Their Centrifugal Recovery," *Adv. Biochemical Eng.* **26**, 1-72 (1983)
- 14 von Hippel, P.H. and Schliech, T., "The Effects of Neutral Salts on the Structure and Conformational Stability of Macromolecules in Solution," in *Structure and Stability of Biological Macromolecules* (Timasheff, S.N. and Fasman, G.D., Eds.) Marcel Dekker, New York, 505-537 (1969)
- 15 Steindl, F., Jungbauer, A., Wensch, E., Himmler, G., and Katinger, H., "Isoelectric Precipitation and Gel Chromatography for Purification of Monoclonal IgM," *Enzyme Microb. Technol.* **9**, 361-364 (1987)
- 16 Schubert, P.F. and Finn, R.K., "Alcohol Precipitation of Proteins: The Relationship of Denaturation and Precipitation for Catalase," *Biotechnol. Bioeng.* **23**, 2569-2590 (1981)
- 17 Kim, C.W., Someren, J.T., Kirshen, M., and Rha, C., "Steric Exclusion of Salts by Polyethylene Glycol," *Phys. Chem. Liq.* **18**, 11-20 (1988)
- 18 Fisher, R.R., "Protein Precipitation with Acids and Polyelectrolytes: The Effects of Reactor Conditions and Models of the Particle Size Distributions," Ph.D. Thesis, Iowa State University (1987)
- 19 Zaworski, P.G. and Gill, G.S., "Precipitation and Recovery of Proteins from Culture Supernatants Using Zinc," *Anal. Biochem.* **173**, 440-444 (1988)
- 20 Lowe, C.R., "New Developments in Downstream Processing," *J. Biotechnol.* **1**, 3-12 (1984)
- 21 Volesky, B. and Luong, J.H.T., "Microbial Enzymes: Production, Purification and Isolation," *CRC Critical Reviews in Biotechnology* **2**, 119-146

(1985)

- 22 Charm, S.E. and Matteo, C.C. "Scale-Up of Protein Isolation," in *Methods in Enzymology*, XXII (Jakoby, W.B. Ed.) Academic Press, New York, 476-556 (1971)

Chapter 2

Solubility-Activity Relationships in the Inorganic Salt-Induced Precipitation of Chymotrypsin

The material contained in this chapter has been accepted
for publication in *Enzyme and Microbial Technology*

2.1 Abstract

Inorganic salt-induced precipitation is a commonly employed protein separation/concentration technique that is potentially denaturing. The effects of precipitating environments on recoverable activity were investigated using α -chymotrypsin as a model protein. The following salts, in order of increasing denaturation potential and decreasing molal surface tension increment, were studied: Na_2SO_4 , NaCl , NaBr , KBr and KSCN . These salts span the lyotropic series. Solubility measurements gave salting-out constants that disagree with current salting-out theory. Salt-protein interactions and concomitant protein structural perturbations are the likely source of the discrepancies. Activity and active fraction measurements indicated that the fraction of precipitate that is active upon dissolution is a function of the salt molal surface tension increment, but not the salt concentration. The specific activity of soluble precipitate remained essentially constant regardless of salt concentration or type. A salting-out performance parameter was defined and calculations indicated that an optimum salt may exist for a particular protein. The recovered active fraction was not affected by varying the amount of salt added and, correspondingly, the amount of protein precipitated.

2.2 Introduction

Inorganic salt-induced precipitation is a commonly employed protein separation/concentration technique that is potentially denaturing. The salting out process has widespread application as a primary isolation procedure in product recovery trains;¹ representative operations include the recovery of recombinant proteins such as insulin, human growth hormone and interferon,² the separation of plasma proteins³ and the purification of polyclonal and monoclonal antibodies.⁴ As a primary isolation procedure, salt-induced precipitation occupies a key position in the determination of the nature and cost of further downstream processing.

Since the economic viability of bioprocesses depends on the recovery of biologically active products, it is important to investigate the relationships between process yield and protein activity. Past studies have not addressed the issue of activity loss in precipitating environments.

In this work, the relationships between protein solubility in salt solutions and recoverable activity from protein precipitates were investigated. The salt type and concentration were varied as process parameters. Solubility data was used to assess the adequacy of current salting-out theory applied to a conformationally labile protein. Optimization criteria were explored with activity and solubility information .

The model system consisted of a single protein in various buffered salt solutions. The serine protease α -chymotrypsin (α CT) was selected as the model protein due to the facility with which the activity and active fraction may be assayed. Also, the native, three-dimensional structure is well known, and the active site may be specifically spin-labelled. These additional properties were exploited in structural studies performed concurrently on precipitate samples and

reported elsewhere.⁵ The salts used in this work include Na₂SO₄, NaCl, NaBr, KBr and KSCN. These salts span the lyotropic series; Na₂SO₄ is considered a structure-stabilizing salt and KSCN is considered a chaotropic agent. These particular salts were chosen for study because they span a wide range of molal surface tension increments,⁶ they have no apparent specific interactions with α CT, and the α CT solubilities resulting from solutions spanning the solubility limits of these salts lie in a workable range (< 10 mg/mL).

2.3 Theoretical Background

2.3.1 Thermodynamics of Precipitate Formation

Melander and Horvath⁶ developed an expression for the solubility of proteins in ionic media based on an adaptation of the solvophobic theory of Sinanoğlu and coworkers.^{7,8} The solvophobic theory relates the energetics of a nonpolar solute to the surface tension of the solvent. The presence of an electrolyte raises the surface tension of the solvent. The ability of a salt to increase the surface tension is quantified by the molal surface tension increment, σ . The σ value is used to correlate the particular salt used with its effect on protein solubility in the theory presented by Melander and Horvath.

In order to identify the major assumptions and arguments underlying the Melander-Horvath theory, a brief summary of results is given next. This provides a useful context for the analysis and interpretation of experimental data obtained in this study. The molal Gibbs free energy of protein, \bar{G}_2 , in a dilute ideal solution may be expressed as⁹

$$\bar{G}_2 = \bar{G}_2^\circ + RT \ln c_2 \quad (1)$$

where c_2 is the protein concentration in grams per kilogram solvent. \bar{G}_2° is the Gibbs free energy of the protein at unit concentration, the standard state. The

notation of Timasheff and coworkers¹⁰ is used throughout; component 1 is the principle solvent (water), 2 is the protein and 3 is the salt. If \bar{G}_2 is considered relative to a hypothetical protein gas phase at state \bar{G}_2^\ominus , equation (1) may be recast as

$$\Delta\bar{G}_2 = \Delta\bar{G}_2^\circ + RT \ln c_2 \quad (2)$$

where $\Delta\bar{G}_2 = \bar{G}_2 - \bar{G}_2^\ominus$ and $\Delta\bar{G}_2^\circ = \bar{G}_2^\circ - \bar{G}_2^\ominus$. $\Delta\bar{G}_2^\circ$ may be interpreted as the molal Gibbs free energy required to transfer protein from the gas phase into solution at the standard state. For a saturated protein solution in equilibrium with protein precipitate, equation (2) may be rearranged in terms of the protein solubility, S , in grams per kilogram solvent

$$\ln S = \frac{1}{RT}(\Delta\bar{G}_p - \Delta\bar{G}_2^\circ). \quad (3)$$

$\Delta\bar{G}_p$ is the molal Gibbs free energy of the precipitate phase relative to the hypothetical gas phase. $\Delta\bar{G}_2^\circ$ may be expanded as⁶

$$\Delta\bar{G}_2^\circ = \Delta\bar{G}_{\text{cav}}^\circ + \Delta\bar{G}_{\text{es}}^\circ + \Delta\bar{G}_{\text{vdw}}^\circ + \ln(RT/PV). \quad (4)$$

The $\Delta\bar{G}_{\text{cav}}^\circ$, $\Delta\bar{G}_{\text{es}}^\circ$ and $\Delta\bar{G}_{\text{vdw}}^\circ$ terms represent the contributions of cavity formation to accomodate the protein in the solvent, of electrostatic interactions between the protein and the electrolyte, and of the van der Waals interactions respectively to the phase transfer energy $\Delta\bar{G}^\circ$. The quantity $\ln(RT/PV)$ is the change in solvent free volume.¹¹ If only the salt concentration varies and in the absence of specific salt-protein interactions and conformational changes, then only $\Delta\bar{G}_{\text{cav}}^\circ$ and $\Delta\bar{G}_{\text{es}}^\circ$ need be evaluated; the remaining terms are essentially constant.⁶

The molal free energy of cavity formation is given by^{6,11}

$$\Delta\bar{G}_{\text{cav}}^\circ = [\mathcal{N}_A \Phi + 4.8 \mathcal{N}_A^{1/3} (\kappa^e - 1) V^{2/3}] \gamma \quad (5)$$

where Φ is the surface area of a protein molecule that is dehydrated upon precipitation (thought to be the hydrophobic surface area), V and γ are the molar volume and surface tension of the solvent, respectively, and κ^e adjusts the macroscopic surface tension to molecular dimensions and high curvature. The molal surface tension increment, σ , enters through a linear expansion of the surface tension in terms of the salt molality, m_3 ,

$$\gamma = \gamma^\circ + \sigma m_3 \quad (6)$$

where γ° is the surface tension of pure water in contact with air. A more rigorous approach might replace γ° with the somewhat smaller interfacial tension between water and a suitable model hydrocarbon.

The electrostatic interaction term is treated by a combination of Debye-Hückel theory applicable at low ionic strengths and Kirkwood theory valid at high ionic strengths.¹² In the Debye-Hückel theory, the protein is considered to be a uniformly charged, spherical macro-ion. The Kirkwood model used treats the protein as a neutral dipole, a prolate ellipsoid with equal and opposite charges at the foci. The electrostatic free energy is given by

$$\Delta\bar{G}_{es}^\circ = \frac{Z^2\epsilon^2\mathcal{N}_A}{2D} \left(\frac{1}{b} - \frac{\kappa}{1+\kappa a} \right) - \frac{2\pi\mathcal{F}^3g(\lambda_o)\mathcal{N}_A}{D^2RT} \mu\xi m_3 \quad (7)$$

where

$$\xi = \frac{1}{2}(\nu_+Z_+^2 + \nu_-Z_-^2). \quad (8)$$

The first term of (7) represents the Debye-Hückel theory contribution; Z is the net charge of the protein, ϵ is the charge on a proton, D is the dielectric constant of the solvent, b is the radius of the protein, a is the distance of closest approach between the centers of the protein and the salt ions, and κ is the reciprocal electrical double layer thickness. κ depends on the square root of the

ionic strength. However, at high ionic strengths, $\kappa a \gg 1$ and the Debye-Hückel term approaches a constant value β . The second term of (7) is the contribution from Kirkwood theory; \mathcal{F} is the Faraday constant, $g(\lambda_o)$ is a function of the eccentricity of the protein¹³ and μ is the protein dipole moment. The large dipole moments characteristic of proteins¹⁴ indicate that higher order moments of the protein charge distribution may also make a significant contribution to the energetics of the electrostatic interaction; these effects are not accounted for in the present development. The conversion factor from molality to ionic strength is given by ξ . Equation (5) expresses ξ in terms of the stoichiometry of the salt component ions ν_i and the charge on the ions Z_i .

If all terms of (5) and (7) invariant with respect to salt concentration are lumped into a constant designated $\ln S_o$, and, recalling that at high salt concentrations, the Debye-Hückel interaction term is given by a constant β , (3) through (8) may be combined to obtain

$$\ln S = \ln S_o + \beta + \frac{1}{RT} [\mathcal{N}_A \Phi + 4.8 \mathcal{N}_A^{1/3} (\kappa^e - 1) V^{2/3}] \sigma m_3 - \frac{\pi \mathcal{F}^3 g(\lambda_o) \mathcal{N}_A}{(DRT)^2} \mu \xi m_3. \quad (9)$$

This expression gives the protein solubility as a function of the type of salt, reflected in the parameter σ , and the salt concentration, m . Further consolidation is possible if the protein-dependent cavity terms and Kirkwood electrostatic terms are replaced by Ω and Λ , respectively, giving⁶

$$\ln S = \ln S_o + \beta - \Omega \sigma m_3 + \Lambda m_3. \quad (10)$$

Note that Ω is linearly dependent on the protein hydrophobic surface area, Φ , and that Λ is proportional to the protein dipole moment, μ .

2.3.2 Empirical Salting-out Equation

Cohn¹³ found that the solubility of proteins in aqueous salt solutions was well represented by

$$\ln S = \beta' - K_s m_3 \quad (11)$$

where β' is a pH- and temperature-dependent constant and K_s , the salting-out constant, depends only on the particular protein and salt used. Recent work has shown β' to be sensitive to the contacting conditions employed.¹⁵

In view of (10), the constants in the empirical expression are given by

$$K_s = \Omega\sigma - \Lambda \quad (12)$$

and

$$\beta' = \ln S_o + \beta. \quad (13)$$

Hence, the present theory suggests a form for systematizing the data and a means of estimating the ability of a salt to precipitate a given protein.

2.4 Materials and Methods

2.4.1 Materials and Experimental Conditions

The α -chymotrypsin (EC 3.4.21.1, three times crystallized from four times crystallized bovine pancreatic chymotrypsinogen, salt free), benzoyl-L-tyrosine ethyl ester (BTEE), 4-methylumbelliferyl-p-trimethyl ammonium cinnamate chloride (MUTMAC) and 4-methylumbelliferone were purchased from Sigma. All salts and buffer solution components were analytical reagent grade. Water deionized by passage through two Research IonXchanger Model 2 columns supplied by Illinois Water Treatment Co. was used throughout.

All procedures were carried out at $25 \pm 2^\circ\text{C}$ in 50 mM glycine buffer adjusted to pH 3.00 with HCl. α CT stock solutions were 10 mg/mL unless otherwise

noted. At pH 3, α CT is most stable¹⁶ and the autolysis rate is negligible.¹⁷ Aliquots of a 10 mg/mL α CT solution were assayed for activity and active fraction; over a period of 120 hours no significant changes were noted indicating that autolysis is not a factor. Also, α CT solubility in all electrolyte solutions used was lower at pH 3.00 than at 7.00, 7.80 or 9.10. At pH 3 the α CT solubilities were similar to those in unbuffered solution; the pH of a 10 mg/mL α CT solution is about 2.96.

2.4.2 Preparation of Samples for Solubility Analysis

Each salt was added to 5 mL portions of buffered α CT stock solution in 15 mL centrifuge tubes. A range of salt concentrations was used in increments of 1, 5 or 10% of salt saturation. The salt was added while vortexing the solution to reduce local concentration gradients. Tubes were then vortexed continuously for 15 minutes; the lowest speed was used to prevent foaming. The samples were aged by standing 90 minutes. Reproducible results required standardized contacting conditions. Precipitate-containing tubes were spun in a Beckman model TJ-6 table-top centrifuge at 2700 rpm for 70 minutes. Generally 10^5 G-min was sufficient to obtain a clear liquid phase;¹⁸ no measurable differences resulted from longer spins. After spinning, 1-2 mL of the clear liquid phase (location of the precipitate at the bottom or top of the tube depended on the relative densities of the precipitate and solution phases) was withdrawn. The solution was passed through a glass wool plug to remove any precipitate that may have been disturbed by sampling and prepared for solubility analysis.

2.4.3 Solubility Assay

Protein concentrations were determined by a Coomassie Blue G-250 dye binding assay obtained from Bio-Rad Laboratories. The use of this assay as op-

posed to measuring the UV absorbance at 280 nm (A_{280}) was based on the latter's sensitivity to structural perturbations.^{19,20} The dye binds to NH_3^+ groups of the protein,²¹ resulting in an increase in absorbance at 595 nm proportional to the amount of dye bound. With the exception of the N-terminal amino group of Ile₁₆, all of the amino (Lys), guanidinium (Arg) and N-terminal amino groups of αCT are located on the surface of the protein.²² The accessibility of these groups to the dye is not expected to change significantly in the event of protein unfolding. A series of αCT solutions of known concentration (10-0.05 mg/mL) was assayed to construct a calibration curve to extend the useful range of the assay. A_{595} measurements were made on Shimadzu UV-260 and UV-160 spectrophotometers.

Before assays were performed, samples were serially diluted with buffer up to 100-fold and allowed to equilibrate on standing for 15 minutes. It was found that the effects of residual salt in precipitate and/or protein aggregation on the assay could be diluted out.²³ High salt concentrations disrupt the binding of the dye. Multimerization reduces the total protein surface area and leads to an underestimation of actual protein concentration; the presence of small amounts of salt is known to drive the dimerization of αCT at acid pH's.²⁴ $(\text{NH}_4)_2\text{SO}_4$, perhaps the most commonly used salt for precipitations,¹⁸ was not included in this work as it interfered with the Bio-Rad assay to a somewhat greater extent than the other salts studied. $(\text{NH}_4)_2\text{SO}_4$ has a molal surface tension increment that lies within the range spanned by the other salts studied and its omission should not affect the generality of the conclusions drawn.

2.4.4 Preparation of Precipitate for Activity Analysis

Using the αCT solubility data, a set of fractional precipitates was produced

for activity assays. Precipitate samples from a series of salt cuts of the same starting solution were analyzed to determine the effects of increasing salt concentrations on recoverable activity. The protocol for sample preparation was similar to that above, except that the clear solution from one cut was then used as the starting solution for the next cut. A series of up to five cuts were taken of the stock protein solution depending on the salt solubility and salting-out constant. The cuts are defined in terms of the cumulative percent of the protein precipitated from the stock solution. Equivalent salt cuts were taken for each of the salts used. Table 1 describes the protein solubility ranges for these cuts.

After the removal of the clear liquid phase, the precipitate was resuspended in the glycine buffer. The precipitate pellet or cake was dispersed with a stirring rod and then vortexed for 15 minutes. The samples were allowed to stand 15 minutes and were then centrifuged at 2700 rpm for 15 minutes to sediment any undissolved precipitate. Redissolved α CT is called "soluble α CT precipitate" in the following.

2.4.5 Activity Assay

The activity of soluble α CT precipitate was determined via the method of Hummel.²⁵ The technique consists of following the BTEE hydrolytic rate by monitoring the change in A_{256} with time. One unit of chymotryptic activity is equivalent to one micromole of BTEE substrate hydrolyzed per minute at pH 7.8 and 25°C. The measurement gives the activity per gram protein.

2.4.6 Active Fraction Assay

Active fraction measurements of soluble precipitates were made using the suicide substrate MUTMAC with the technique of Gabel.²⁶ The formation of hydrolysis product, 4-methylumbelliferone, was monitored in a Shimadzu RF-

540 spectrofluorimeter coupled to a DR-3 controller. The sample cuvettes were illuminated at 356 nm and the emission was recorded at 451 nm.²⁷ Thorough mixing of the cuvette contents was requisite for a reproducible signal. The fluorescence intensity was calibrated with 4-methylumbelliferone solutions of known concentration; the 4-methylumbelliferone was recrystallized from hot ethanol-water solution before use as a standard. Measurements were performed at pH 7.9 at 25°C. This assay gives the fraction of the total soluble enzyme that is active. The specific activity is obtained by dividing the activity by the active fraction and is expressed in units of activity per gram active protein.

2.5 Results and Discussion

2.5.1 Theoretical Predictions

Equations (9) and (12) may be used to predict the variation of the salting-out constant K_s with the salt used via the molal surface tension increment. Protein properties enter through the hydrophobic surface area Φ in the hydrophobic term Ω and the dipole moment μ in the electrostatic term Λ .

For water at 25°C, the hydrophobic term becomes⁶

$$\Omega = 0.00243\Phi + 0.0292 \quad (14)$$

where Φ is in \AA^2 per molecule. Calculations performed by Melander and Horvath⁶ indicated that an inverse relationship exists between the frequency of charged groups in the amino acid sequence of a protein, CHF , and the relative surface hydrophobicity, RSH . RSH is the nonpolar surface area of the protein that is dehydrated upon precipitation normalized by its molecular weight. Values of RSH for myoglobin, ovalbumin, hemoglobin and albumin were calculated from salting-out constants.⁶ The CHF of α CT was used to determine

RSH and in turn Φ ; the ratio of charged amino acids to the total number of residues gives $CHF = 0.12$ for αCT .¹⁶ A plot of RSH versus CHF using the data from Melander and Horvath⁶ with the estimated value of CHF for αCT is given in *Figure 1*. An estimate of 0.060 \AA^2 per molecule/dal is indicated for RSH . Using a molecular weight of 24.5 kdal, $\Phi \approx 1460 \text{ \AA}^2$ per molecule. This gives $\Omega = 3.6 \text{ cm/dyne}$.

Miller et al.²⁸ computed the solvent accessible surface area of αCT from the crystal structure of Tsukada and Blow.²⁹ For a probe radius of 1.4 \AA , the van der Waals radius of water, the solvent accessible surface area was estimated to be $10,440 \text{ \AA}^2$ per molecule. This calculation includes half a layer of solvent in addition to the protein molecular volume. The surface area was subdivided into contributions from nonpolar, polar and charged surface residues. For αCT , the nonpolar portion was about 59%, giving a nonpolar surface area of about 6160 \AA^2 . This value is over four times greater than the surface area estimated above.

The electrostatic term is

$$\Lambda = 0.00171\mu\xi \quad (15)$$

for water at 25°C , assuming that αCT is a prolate ellipsoid with characteristic dimensions of $51 \times 40 \times 40 \text{ \AA}$.³⁰ The dipole moment μ is in Debye and ξ is given by (8); Λ has dimensions of m^{-1} .

The dipole moment of αCT was calculated from atomic coordinates. A number of high resolution crystal structures have been reported for αCT .^{22,29,31} The coordinates from the study of Blevins and Tulinsky³¹ were chosen for calculations as the crystals were formed at pH 3.5. In addition, at 1.67 \AA , the study represents the highest resolution structure of αCT to date. The actual

coordinates were obtained from the Brookhaven Protein Data Bank.³² α CT crystallizes in the P_{21} space group with two dimers per unit cell. The dimer is the asymmetric unit. Although the individual molecules of the dimer were found to be nearly superimposable, the dipole moment calculations were performed on each molecule separately and then averaged.

The database contains coordinates relative to a cartesian system of axes. The center of mass was determined relative to the origin of this system using

$$\vec{R} = \sum_i m_i \vec{r}_i / \sum_i m_i \quad (16)$$

where m_i is the mass of the i^{th} atom and \vec{R} and \vec{r}_i represent the vectors from the origin to the center of mass and i^{th} atom respectively. Since hydrogen atoms are too light to scatter at detectable levels their masses have been added to the nearest heavy atom.³³ There was no electron density for the last four residues of the A chain except for the N atom of Val₉ (9N); the masses of these residues was summed and placed at the 9N coordinates.

The dipole moment was calculated relative to the center of mass¹⁴

$$\vec{\mu} = \sum_i q_i (\vec{r}_i - \vec{R}) \quad (17)$$

where q_i is the formal charge assigned to the i^{th} atom. Since α CT carries a net positive charge at pH 3, the computed dipole moment depends on the choice of origin; the center of mass was chosen as this approximates the point about which the molecule reorients itself in solution. The values of q_i were assigned using typical values for the pK_a of charged groups in proteins³⁰ and the Henderson-Hasselbalch equation at pH 3. Charged groups interacting over a distance of 5 Å or less²⁴ were considered to carry full charge at pH 3. The localization of the charges relative to the heavy atom(s) involved was assigned as in Barlow and

Thornton.¹⁴ The charge on the carboxyl group at the C-terminus of the A chain (Leu₁₃) was arbitrarily placed at the 9N coordinates.

Using the transformation matrix provided in the database, the vector $\vec{\mu}$ was converted from the cartesian coordinate system to crystallographic coordinates. The magnitude of the dipole moment μ was found using the unit cell parameters of the monoclinic axis system. The average calculated value of the dipole moment for the molecules of the dimer was 223 ± 1 Debye. The center of mass of α CT was estimated to be within 0.8 \AA of the C' atom of Gly₁₉₇.

Having obtained the dipole moment, the electrostatic interaction becomes

$$\Lambda = 0.381\xi. \quad (18)$$

The value of Λ depends on the stoichiometry of the electrolyte through the conversion factor ξ ; $\xi = 1$ for 1-1 salts and $\xi = 3$ for 2-1 salts.

Combining the value for Ω and (18) gives the relationship between the salt used and the resulting K_s value for α CT

$$K_s = 3.6\sigma - 0.381\xi. \quad (19)$$

This is a family of lines in which salts with higher molal surface tension increments are predicted to be more efficient in their salting-out action. Also, salts with higher ionic valences and/or stoichiometries should be less effective on a molal basis than 1-1 electrolytes with similar surface tension increments. *Figure 2* gives a plot of (19) for 1-1 (solid line) and 2-1 (dashed line) salts.

2.5.2 Solubility Determinations

Plots of the natural logarithm of protein concentration as a function of salt molality ideally consist of two linear segments. The segment at low salt

concentrations should be horizontal at the logarithm of the initial protein concentration, C . No protein will precipitate until enough salt is added to reduce its solubility to its initial concentration. The salt concentration at which protein solubility equals the initial concentration is referred to below as m^* . The second segment, starting at m^* , will have a slope equal to K_s , as indicated in equation (11).

Figures 3 a-e give the natural logarithm of the α CT concentration of the clear solution phases as a function of the salt molality for KSCN, KBr, NaBr, NaCl and Na₂SO₄. A series of dilutions of each sample was analyzed as described above. The salting-out portion of each curve was fit with a line by least squares. $K_s^{app}(d)$ and $\beta'(d)$ represent the calculated slopes and intercepts respectively of these lines for each dilution d . For each salt, the slopes of the apparent salting-out constants decrease and approach a constant value as interfering effects are diluted out. Errant points occur where precipitate density is in the neighborhood of the salt solution density, making separation by centrifugation difficult, and where the protein concentration is reduced to detectable limits. Based on observation and salt solution densities, precipitate densities ranged from 1.19 to 1.33 g/cm³. This range brackets typical protein precipitate densities.¹⁸

To arrive at the true salting-out constant for each salt, an empirical relation was fit to the $K_s^{app}(d)$ values and extrapolated to infinite dilution. The following decaying exponential function of dilution

$$K_s^{app}(d) = K_s + ae^{-bd} \quad (20)$$

was found to fit the serial dilution data well. Here d is the dilution factor (d represents the ratio of the total diluted volume to the original sample volume) and a and b are fitted constants. The data were fitted via a weighted nonlinear

least squares procedure;³⁴ weights were assigned using the reciprocal variances of the $K_s^{app}(d)$ values fitted to the solubility data. A representative fit of equation (20) to the data for NaBr is given in *Figure 4*.

Fitted values for the constants a and b for all of the salts considered are given in *Table 2*. The 100-fold dilution salting-out constant $K_s^{app}(100)$ was typically within a few percent of the extrapolated infinite dilution value K_s . This indicates that equation (20) does not introduce any artifacts into the values determined for K_s . The standard deviation of the K_s values was estimated from the variance of the exponential fit.

The critical salt concentration at each dilution $m^{*app}(d)$ was calculated from the corresponding K_s^{app} , β'^{app} and C values

$$m^{*app}(d) = \frac{\beta'^{app} - \ln(C/d)}{K_s^{app}}. \quad (21)$$

As $m^{*app}(d)$ should be independent of dilution, an average value was determined. Using m^* and K_s , the intercept β' was computed. *Table 2* gives a compilation of K_s , β' and m^* values with corresponding standard deviations of the mean for each of the salts.

2.5.3 Comparison of Empirically and Theoretically Derived Salting-Out Constants

The K_s values determined above are plotted as a function of the surface tension increment along with the theoretical relationship in *Figure 2*. With the exception of KSCN, the theory overestimates the salting-out constants. Also, there is an apparent discrepancy in the functional dependence of K_s on σ . This is not surprising since the theory is constrained to conformationally inert proteins. A concurrent structural study indicated that the conformation of α CT in precipitate is perturbed to different extents by the salts used.⁵

The hydrophobic and electrostatic interaction terms Ω and Λ are based on physical properties of the protein that are inherently conformation dependent. Ω is linearly dependent on Φ ; the salt-protein interaction may regulate the actual extent of the protein surface area that is dehydrated on precipitation. The predicted dehydrated surface area is only a fraction of the total hydrophobic surface area. A study of the thermal denaturation of lysozyme in solutions of different tetraalkylammonium bromide salts found that the amount of hydrophobic surface area exposed varied with the salt used.³⁵ Similarly, the electrostatic interaction contribution Λ is proportional to the protein dipole moment μ . The dipole moment of a protein is largely determined by its characteristic dimension as opposed to the distribution of charged groups on its surface.¹⁴ The linear dimensions of a protein are conformation-dependent; size changes accompanying denaturation have been used as a means of fractionating proteins by gel permeation chromatography.³⁶

The linearity of the solubility data at high salt concentrations in *Figure 2* implies that the salt concentration dependence of the theory is correct. This was corroborated by the structural studies. Precipitate conformational perturbations depended only on the particular salt used, not the salt concentration.⁵ Protein physical properties are not changing significantly with salt concentration.

Assuming that hydrophobic interactions dominate the salting-out process, it is possible to explain deviations from ideal behavior by changes in Φ resulting from conformational perturbations. The dehydrated hydrophobic surface area depends on the square of the linear dimension while the dipole moment is proportional to linear dimension; structural changes may alter Φ to a greater extent than μ . Further support for this assumption is lent by the success of the

present theory in explaining salt effects on protein retention factors in hydrophobic interaction chromatography (HIC).^{6,37,38} Also, the coefficients in equations (14) and (15) weight the hydrophobic surface area contribution to the salting-out constant over an order of magnitude more heavily than that of the dipole moment.

Table 3 gives a list of the dehydrated hydrophobic surface areas required of α CT to account for deviations from the theory. These Φ values were computed using the experimental K_s values with equations (12), (14) and (15). The value required of KSCN precipitates is on the order of the total hydrophobic surface area.²⁸ Since the theory overestimates the salting-out constants for the remaining salts, the corresponding Φ values calculated are fractions of the predicted dehydrated hydrophobic surface area. It is unlikely, in view of the structural studies,⁵ that the salts used vary the dehydrated surface area over an order of magnitude as shown in *Table 3*. Protein conformational perturbations may only account for a portion of the deviations from the predicted salting-out abilities.

Salt-protein interaction effects may also be invoked to explain the discrepancies. Extensive studies of the interactions of salts with proteins have been conducted by Timasheff and coworkers.^{10,39,40} Interactions have been expressed in terms of a "preferential interaction parameter", ξ_3 , which is a thermodynamic quantity. ξ_3 is defined by¹⁰

$$\xi_3 = \left(\frac{\partial m_3}{\partial m_2} \right)_{T, G_1, G_3} \quad (22)$$

where m_i represents the molality of component i . The observed interaction can be broken down into contributions from salt binding to the protein and salt exclusion from the protein-solvent interface³⁹

$$\xi_3^{\text{obs}} = \xi_3^{\text{excl}} + \xi_3^{\text{bind}} \quad (23)$$

where the two quantities on the right have opposite signs.

The change in chemical potential of the protein with the addition of salt depends on the preferential interaction parameter via

$$\left(\frac{\partial \bar{G}_2}{\partial m_3}\right)_{T,P,m_2} = -\left(\frac{\partial m_3}{\partial m_2}\right)_{T,\bar{G}_1,\bar{G}_3} \left(\frac{\partial \bar{G}_3}{\partial m_3}\right)_{T,P,m_2} \quad (24)$$

The last term on the right represents the co-solvent interaction term and can be written in terms of the salt activity, a_3 ,

$$\left(\frac{\partial \bar{G}_3}{\partial m_3}\right)_{T,P,m_2} = RT \left(\frac{\partial \ln a_3}{\partial m_3}\right)_{T,P,m_2} \quad (25)$$

The salt activity is in turn a function of the mean ionic activity coefficient γ_{\pm}

41

$$a_3 = (\nu_+^{\nu_+} \nu_-^{\nu_-}) (m_3 \gamma_{\pm})^{(\nu_+ + \nu_-)}. \quad (26)$$

From equations (3) through (8), the derivative of the protein chemical potential with respect to the salt concentration at equilibrium is

$$\left(\frac{\partial \bar{G}_2}{\partial m_3}\right)_{T,P,m_2} = RT \left[(\Lambda - \Omega \sigma) + \left(\frac{\partial \ln S}{\partial m_3}\right)_{T,P,m_2} \right] \quad (27)$$

since

$$\left(\frac{\partial \bar{G}_2}{\partial m_3}\right)_{T,P,m_2} = \left(\frac{\partial \Delta G_p}{\partial m_3}\right)_{T,P,m_2} \quad (28)$$

The first term in the brackets in (27) is the calculated salting out coefficient, K_s^{cal} , and the last term is the experimental salting-out constant, K_s^{exp} . Combining equations (24) through (28) with the definitions above gives an expression for the preferential interaction parameter

$$K_s^{\text{exp}} - K_s^{\text{cal}} = \xi_3 (\nu_+ + \nu_-) \left[\frac{1}{m_3} + \left(\frac{\partial \ln \gamma_{\pm}}{\partial m_3}\right)_{T,P,m_2} \right]. \quad (29)$$

Values of ξ_3 were calculated from the experimental and theoretical salting-out constants and salt activity data.⁴¹ For valid comparisons between the salts, ξ_3

was computed at salt concentrations resulting in equivalent protein solubilities; ξ_3 was calculated at m^* and at salt concentrations corresponding to the cuts listed in *Table 1*. *Figure 5* gives a plot of ξ_3 versus salt cut for each salt. Negative values indicate that α CT is preferentially hydrated vis-à-vis equation (23) and that experimental K_s values are lower than predicted. This is the case for all the salts used in this study with the exception of KSCN. KSCN has positive ξ_3 values representing salt binding to the protein. The magnitude of the interaction generally increases with increasing salt concentration. Qualitatively similar behavior has been seen in the interaction of Na_2SO_4 , NaCl and KSCN at different concentrations and pH's with bovine serum albumin.⁴⁰ However, the ξ_3 values are sensitive to salt type, concentration, pH and the particular protein used.

These results suggest that structural changes or bulk solution properties alone cannot account for the solubility of a conformationally sensitive protein in electrolyte solutions. The deviations from the theoretically predicted behavior are probably the result of salt-specific salt-protein interactions that mediate conformational change. The deviations can be accounted for by changes in physical properties such as Φ and μ , but these changes cannot in turn be attributed to any intrinsic bulk property of the salt. Similarly, in the absence of predictive ability for the salt interaction parameter, ξ_3 as formulated in equation (29) functions as a correction factor.

2.5.4 Precipitate Activity Determinations

The empirically determined values of K_s and m^* were used to calculate the salt concentrations required to precipitate given fractions of the starting protein solution for activity analysis. The fraction of protein precipitated at a particular

salt concentration $F(m_3)$ is given by

$$F(m_3) = 1 - e^{-K_s(m_3 - m^*)} \quad \text{for } m_3 \geq m^*. \quad (30)$$

The number of cuts taken for a particular salt was determined by the above relation and the salt saturation concentration.

The α CT precipitates produced in this study were composed of soluble and insoluble aggregate fractions. Insoluble precipitates were protein aggregates that could not be redissolved in the glycine buffer. These precipitates were considered inactive. However, the insoluble portions were readily solubilized by chaotropic agents such as 6 M urea and 8 M guanidine hydrochloride. The active fraction assays performed on the soluble precipitates indicated that they were composed of both active and inactive protein.

The results of the activity and active fraction assays performed on precipitates from each salt cut with Na_2SO_4 , NaCl, NaBr, KBr and KSCN are given in *Figures 6 a-e*. Note that these values are on a soluble precipitate basis; for example, the active fraction is expressed as the fraction of the soluble portion of the precipitate that was active. The activity and active fraction were generally lower than the native protein values. The salient feature of these plots is the specific activity. The specific activity is a measure of the activity on an active enzyme basis and is found by dividing the activity by the active fraction. For each salt, the specific activity remains essentially constant at the value of the native enzyme. The scatter is likely due to increasing concentrations of residual salt in the precipitates at the higher salt cuts; the enzymatic activity of α CT is affected by the presence of salts and may even exceed that of the native enzyme.⁴² This suggests that any precipitate that is able to recover catalytic activity regains full original activity, regardless of the type or amount of salt

used.

A control study was performed to see if the redissolution-renaturation process was evolving with time. Assays of soluble precipitate concentration, activity and active fraction showed no significant changes over a 24 hour period following redissolution.

If the insoluble precipitate fraction is lumped with the inactive soluble portion, the active fraction can be computed on a total precipitate basis. For a given salt, the active fraction f_{act} is virtually independent of the salt cut. *Figure 7* gives the average active fraction for each salt as a function of the corresponding salt molal surface tension increment. The active fraction data fall in two groups roughly corresponding with the lyotropic series; the lower surface tension increment salts gave lower active fractions.

2.5.5 Optimization Strategy

In this work, two process variables were studied, salt type and concentration. The activity data indicate that recovered specific activity is independent of both variables. There appears to be no optimum cut for a given salt; this would be determined solely on the basis of process purity and concentration requirements and subsequent desalting operations.

A trade-off may exist between the amount of protein precipitated and the amount of trapped salt. The void volume of precipitate aggregates is filled with salt-containing mother liquor. Models are available for the relationship between the number of molecules in a floc and the corresponding void volume.^{43,44} However, the absence of information concerning how the number of molecules in an aggregate are related to the salt concentration prohibits the determination of the optimum, if it exists.

The fraction of precipitate that can be recovered in active form is a function of the salt type only. Therefore, it may be possible to optimize the type of salt used to give the best performance in terms of salting-out efficiency, K_s , and yield of active protein, f_{act} .

The product of the salting-out constant and the average active fraction is used as the performance indicator, giving the expected recovery of active protein for a given increment of salt. *Figure 8* gives a plot of $K_s f_{act}$ as a function of the salt surface tension increment at 25°C. The data is roughly parabolic with a minimum in the vicinity of NaBr. Both Na₂SO₄ and KSCN are good performers, yet they lie at opposite ends of the lyotropic series. Na₂SO₄ is relatively innocuous, long considered a protein structure-stabilizing salt.^{45,46} KSCN is an aggressive chaotropic agent.^{40,45} This surprising result might be attributed to competing kinetic events and/or equilibria. Salts induce both structural changes and aggregation. The unexpected performance of KSCN may result from an aggregation rate that is faster than the rate of structural change. Presumably, once protein molecules are ensconced in an aggregate, they are immune to further salt-imposed structural change.

This hypothesis was tested by performing a limited investigation at a reduced temperature. The solubility of αCT in the same salt solutions at 15°C was measured at a few points where protein was precipitated at significant levels to arrive at K_s^{15} values. Precipitate was collected, redissolved at 15°C and the soluble portion active fraction f_{act}^{15} was assayed and expressed on a total soluble and insoluble precipitate basis. The $K_s^{15} f_{act}^{15}$ products are plotted on *Figure 7* for comparison. The large shift in the behavior of this performance indicator with temperature suggests that kinetic and/or equilibrium phenomena indeed play an important role in the precipitation process. In addition, it is in-

interesting to note that the performance at 15°C is generally poorer than that at 25°C. This has important implications as most industrial and analytical protein precipitations are performed at reduced temperatures.

2.6 Conclusions

The study of protein solubility and the recoverable activity of precipitates in salt solutions has provided insight into the salting-out process. The current salting-out theory⁶ cannot predict the performance of a given salt for a conformationally labile protein such as α CT. Account must be taken of the interaction of the salt with the protein and its concomitant effects on the physical properties of the protein. Bulk electrolyte solution properties alone appear to have a small role if any in determining the nature and extent of salt-protein interactions. Arakawa and Timasheff^{40,47} have determined values for a preferential interaction parameter that is specific for a given salt and protein, but as yet there is no explanation of what particular salt and protein properties control the nature and magnitude of the interaction.

Recovered activity data implies that the critical factor in precipitations is the nature of the salt used, not the amount used. The fraction that recovered activity was essentially invariant to the salt cut taken for a given salt. Protein denaturation in precipitation is an all-or-nothing phenomenon; enzymatic activity was either fully recovered or lost completely. There may be a structural threshold beyond which the protein cannot properly refold on dissolution. Alternatively, the portion of protein precipitate that is soluble and is able to recover activity may depend on the physical packing of the aggregate.

It may be possible to optimize a precipitation with respect to the type of salt used. In a concentration procedure this may be straightforward; the

construction of a plot similar to *Figure 8* would guide the choice of salt. However, in separations, differing interactions with other proteins present would likely complicate the selection.

A significant remaining question is the extent to which α CT can be considered a model protein. The structural studies performed on α CT precipitates gave results that mimic those seen in studies of other protein aggregates;⁵ this lends support to the use of α CT as a model protein. It is unlikely that α CT represents a pathological case. However, the extent of salt-protein interaction and its implications for structural perturbation and recoverable activity are surely protein dependent.

The salting-out process is still not completely understood. Melander and Horvath⁶ have added considerable insight into the particular protein properties involved, namely the dehydrated hydrophobic surface area and the dipole moment. Questions remain concerning the mechanism by which a salt influences the physical properties of a protein. Seen in a larger context, this work indicates that varying degrees of salt-protein interaction may favor the choice of one salt over another in a precipitation. As the nature of these interactions is revealed, it may be possible to optimize a precipitation step and move from recipe-based process designs to a rational design basis.

2.7 Acknowledgement

This research was supported by the National Science Foundation.

2.8 Nomenclature

| | |
|---------------------------|--|
| a | distance of closest ion approach to protein, Å |
| b | protein radius, Å |
| c_2 | protein concentration, g/kg solvent |
| C | initial protein concentration, mg/mL |
| d | dilution factor |
| D | solvent dielectric constant |
| f_{act} | active fraction of precipitate at 25°C |
| f_{act}^{15} | active fraction of precipitate at 15°C |
| $F(m_3)$ | fraction of protein precipitated at salt concentration m_3 |
| \mathcal{F} | Faraday constant |
| \bar{G}_i | chemical potential of component i |
| \bar{G}_2° | chemical potential of protein in solution at standard state |
| \bar{G}_2^\ominus | chemical potential of hypothetical protein gas phase |
| $\Delta\bar{G}_i$ | chemical potential of i^{th} protein phase relative to gas phase |
| $\Delta\bar{G}_i^\ominus$ | i^{th} interaction contribution to the molal Gibbs free energy of transferring protein from gas phase into solution |
| K_s | salting-out constant, molality ⁻¹ |
| $K_s^{app}(d)$ | apparent salting-out constant at dilution d , molality ⁻¹ |
| m_i | mass of i^{th} atom, molality of component i |
| N_A | Avogadro's number |
| P | pressure |
| q_i | charge on i^{th} atom |
| \vec{r}_i | vector from origin to i^{th} atom |
| R | gas constant |
| \vec{R} | vector from origin to center of mass |

| | |
|-------|---------------------------------------|
| S | protein solubility, g/kg solvent |
| T | temperature |
| V | solvent molar volume |
| Z | net charge on protein |
| Z_i | charge on i^{th} salt component ion |

Greek symbols

| | |
|----------------|--|
| β | Debye-Hückel interaction at high ionic strengths |
| β' | intercept of empirical salting-out equation |
| γ | solvent surface tension, dyne·cm ⁻¹ |
| γ° | pure water surface tension, dyne·cm ⁻¹ |
| γ_{\pm} | salt mean ionic activity coefficient |
| ϵ | protonic charge |
| κ | reciprocal electrical double layer thickness, Å ⁻¹ |
| κ^e | surface tension correction factor |
| λ_o | reciprocal ellipsoidal eccentricity |
| Λ | electrostatic contribution to salting-out constant, molality ⁻¹ |
| μ | magnitude of protein dipole moment, Debye |
| $\vec{\mu}$ | protein dipole moment vector, Debye |
| ν_i | stoichiometric coefficient of salt component ion of type i |
| ξ | conversion factor from molality to ionic strength |
| ξ_3 | salt interaction parameter |
| σ | salt molal surface tension increment, dyne·g·cm ⁻¹ ·mol ⁻¹ |
| Φ | protein hydrophobic surface area, Å ² per molecule |
| Ω | hydrophobic contribution to salting-out constant, cm·dyne ⁻¹ |

2.9 References

- 1 Bailey, J.E. and Ollis, D.F., *Biochemical Engineering Fundamentals*, 2nd ed., McGraw-Hill, New York, 727 (1986)
- 2 McGregor, W.C., "Large-scale Isolation and Purification of Proteins from Recombinant *E. coli*," *Ann. N.Y. Acad. Sci.* **413**, 231-237 (1983)
- 3 Steinbuch, M., "Protein Fractionation by Ammonium Sulphate, Rivanol and Caprylic Acid Precipitation," in *Methods of Plasma Protein Fractionation* (Curling, J.M., Ed.) Academic Press, London, 33-56 (1980)
- 4 Steward, M.W., *Antibodies: Their Structure and Function*, Chapman and Hall, New York, 13-17 (1984)
- 5 Przybycien, T.M. and Bailey J.E., "Structure-Function Relationships in the Inorganic Salt-Induced Precipitation of α -Chymotrypsin," *Biochim. Biophys. Acta* (1989) in press
- 6 Melander, W. and Horvath, C., "Salt Effects on Hydrophobic Interactions in Precipitation and Chromatography of Proteins: An Interpretation of the Lyotropic Series," *Arch. Biochem. Biophys.* **183**, 200-215 (1977)
- 7 Sinanoğlu, O. and Abdunur, S., "Effect of Water and Other Solvents on the Structure of Biopolymers," *Fed. Proc.* **24**, 12-23 (1965)
- 8 Halicioğlu, T. and Sinanoğlu, O., "Solvent Effects on Cis-Trans Azobenzene Isomerization: A Detailed Application of a Theory of Solvent Effects on Molecular Association," *Ann. N.Y. Acad. Sci.* **158**, 308-317 (1969)
- 9 Denbigh, K., *Principles of Chemical Equilibrium*, 4th ed., Cambridge University Press, New York, 278 (1981)
- 10 Lee, J.C., Gekko, K. and Timasheff, S.N., "Measurements of Preferential Solvent Interactions by Densimetric Techniques," *Methods Enzymol.* **61**, 26-49 (1979)

- 11 Horvath, C., Melander, W. and Molnar, I., "Solvophobic Interactions in Liquid Chromatography with Nonpolar Stationary Phases," *J. Chromatogr.* **125**, 129-156 (1976)
- 12 Edsall, J.T. and Wyman, J., *Biophysical Chemistry*, Vol. 1, Academic Press, New York, 289, 303-305 (1958)
- 13 Cohn, E.J. and Edsall, J.T., *Proteins, Amino Acids and Peptides*, Reinhold, New York, 282, 603 (1943)
- 14 Barlow, D.J. and Thornton, J.M., "The Distribution of Charged Groups in Proteins," *Biopolymers* **25**, 1717-1733 (1986)
- 15 Foster, P.R., Dunnill, P. and Lilly, M.D., "The Kinetics of Protein Salting-Out: Precipitation of Yeast Enzymes by Ammonium Sulfate," *Biotechnol. Bioeng.* **28**, 545-580 (1976)
- 16 Desnuelle, P., "Chymotrypsin", in *The Enzymes*, Vol. 4, (Boyer, P.D., Lardy, H. and Myrback, K., Eds.), Academic Press, New York, 93 (1960)
- 17 Kumar, S. and Hein, G.E., "Concerning the Mechanism of Autolysis of α -Chymotrypsin," *Biochemistry* **9**, 291-297 (1970)
- 18 Scopes, R.K., *Protein Purification Principles and Practice*, Springer-Verlag, New York, 48, 51 (1982)
- 19 Donovan, J.W., "Ultraviolet Difference Spectroscopy - New Techniques and Applications," *Methods Enzymol.* **27**, 497-525 (1973)
- 20 Herskovits, T.T., "Difference Spectroscopy," *Methods Enzymol.* **11**, 748-775 (1967)
- 21 Fazekas de St. Groth, S., Webster, R.G. and Datyner, A., "Two New Staining Procedures for Quantitative Estimation of Proteins on Electrophoretic Strips," *Biochim. Biophys. Acta* **71**, 377-391 (1963)
- 22 Birktoft, J.J. and Blow, D.M., "Structure of Crystalline α -Chymotrypsin V.

- The Atomic Structure of Tosyl- α -chymotrypsin at 2Å Resolution," *J. Mol. Biol.* **68**, 187-240 (1972)
- 23 Stoesz, J.D. and Lumry, R.W., "Refolding Transition of α -Chymotrypsin: pH and Salt Dependence," *Biochemistry* **17**, 3693-3699 (1978)
- 24 Aune, K.C. and Timasheff, S.N., "Dimerization of α -Chymotrypsin. I. pH Dependence in the Acid Region," *Biochemistry* **10**, 1609-1617 (1971)
- 25 Hummel, B.C.W., "A Modified Spectrophotometric Determination of Chymotrypsin, Trypsin and Thrombin," *Can. J. Biochem.* **37**, 1393-1399 (1959)
- 26 Gabel, D., "Active Site Titration of Immobilized Chymotrypsin with a Fluorogenic Reagent," *FEBS Lett.* **49**, 280-281 (1974)
- 27 Clark, D.S. and Bailey, J.E., "Structure-Function Relationships in Immobilized Chymotrypsin Catalysis," *Biotechnol. Bioeng.* **25**, 1027-1047 (1983)
- 28 Miller, S., Janin, J., Lesk, A.M. and Chothia, C., "Interior and Surface of Monomeric Proteins," *J. Mol. Biol.* **196**, 641-656 (1987)
- 29 Tsukada, H. and Blow, D.M., "Structure of α -Chymotrypsin Refined at 1.68Å Resolution," *J. Mol. Biol.* **184**, 703-711 (1985)
- 30 Stryer, L., *Biochemistry*, W.H. Freeman, New York, 60, 160 (1981)
- 31 Blevins, R.A. and Tulinsky, A., "The Refinement and the Structure of the Dimer of α -Chymotrypsin at 1.67Å Resolution," *J. Biol. Chem.* **260**, 4264-4275 (1985)
- 32 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., "The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures," *J. Mol. Biol.* **112**, 535-542 (1977)
- 33 Sundaram, K. and Viswanadham, V.N., "Moment of Inertia Ellipsoids of

- Various Proteins of Known Crystalline Structure," *Physiol. Chem. Phys.* **12**, 187-191 (1980)
- 34 Kuester, J.L. and Mize, J.H., *Optimization Techniques with FORTRAN*, McGraw-Hill, New York, 240-250 (1973)
- 35 Badley, R.A., Carruthers, L. and Phillips, M.C., "Hydrophobic Free Energy and the Denaturation of Proteins," *Biochim. Biophys. Acta* **495**, 110-118 (1977)
- 36 Corbett, R.J.T. and Roche, R.S., "Use of High-Speed Size Exclusion Chromatography for the Study of Protein Folding and Stability," *Biochemistry* **23**, 1888-1894 (1984)
- 37 Melander, W.R., Corradini, D. and Horvath, C., "Salt-Mediated Retention of Proteins in Hydrophobic Interaction Chromatography," *J. Chromatogr.* **317**, 67-85 (1984)
- 38 Fausnaugh, J.L. and Regnier, F.E., "Solute and Mobile Phase Contributions to Retention in Hydrophobic Interaction Chromatography of Proteins," *J. Chromatogr.* **359**, 131-146 (1986)
- 39 Arakawa, T. and Timasheff, S.N., "Mechanism of Protein Salting In and Salting Out by Divalent Cation Salts: Balance between Hydration and Salt Binding," *Biochemistry* **23**, 5912-5923 (1984)
- 40 Arakawa, T. and Timasheff, S.N., "Preferential Interactions of Proteins with Salts in Concentrated Solutions," *Biochemistry* **21**, 6545-6552 (1982)
- 41 Robinson, R.A. and Stokes, R.H., *Electrolyte Solutions*, Butterworths, London, Appendix 8.10 (1959)
- 42 Kahana, L. and Shalitin, Y., "Salt Effects on the Properties of α -Chymotrypsin I. Effects on the Enzyme Activity of Chymotrypsin," *Israel J. Chem.* **12**, 573-589 (1974)

- 43 Lagvanker, A.L. and Gemmell, R.S., "A Size-Density Relationship for Floccs," *J. Am. Water Works Assoc.* **9**, 1040-1046 (1968)
- 44 Bell, D.J., Heywood-Waddington, D., Hoare, M. and Dunnill, P., "The Density of Protein Precipitates and Its Effect on Centrifugal Sedimentation," *Biotechnol. Bioeng.* **24**, 127-141 (1982)
- 45 Carrea, G., Bovara, R. and Pasta, P., "The Effect of Hofmeister Anions and Protein Concentration on the Activity and Stability of Some Immobilized NAD-Dependent Dehydrogenases," *Biotechnol. Bioeng.* **24**, 1-7 (1982)
- 46 Ikeda, K., Kunugi, S. and Ise, N., "Specific and Non-Specific Effects of Inert Anions and Cations on the Dimerization of α -Chymotrypsin and the Catalytic Activities of Monomeric and Dimeric α -Chymotrypsins," *J. Biochem.* **91**, 347-355 (1982)
- 47 Arakawa, T. and Timasheff, S.N., "Abnormal Solubility Behavior of β -Lactoglobulin: Salting-In by Glycine and NaCl," *Biochemistry* **26**, 5147-5153 (1987)

2.10 Tables

Table 1 Salt cut α CTsolubility ranges

| salt cut | α CT solubility | | percent α CT precipitated |
|----------------|------------------------|-------|-------------------------------------|
| | change (mg/mL) | | |
| 1 | 10 | → 8 | 20 |
| 2 | 8 | → 6 | 40 |
| 3 | 6 | → 4 | 60 |
| 4 | 4 | → 2 | 80 |
| 5 ^a | 2 | → 0.5 | 95 |

a The last cut taken depends on salt saturation, m^* and K_s values.

Table 2 Salting-out equation fitted values^a

| salt | n^b | a | b | K_s | m^* | β' | $\sigma \times 10^3$ |
|---------------------------------|-------|------|-------|--------------|-----------|-----------|--|
| | | | | (m^{-1}) | (m) | | ($\frac{\text{dyn}\cdot\text{g}}{\text{cm}\cdot\text{mol}}$) |
| KSCN | 6 | 12.2 | 0.021 | 6.83±0.92 | 0.05±0.03 | 2.64±0.23 | 0.45 |
| KBr | 6 | 1.35 | 0.155 | 0.67±0.01 | 2.35±0.24 | 3.89±0.16 | 1.31 |
| NaBr | 6 | 1.09 | 0.221 | 1.05±0.06 | 2.20±0.12 | 4.61±0.18 | 1.32 |
| NaCl | 4 | 0.79 | 0.345 | 0.82±.00 | 3.95±0.07 | 5.52±0.06 | 1.64 |
| Na ₂ SO ₄ | 4 | 2.78 | 0.053 | 3.43±0.08 | 1.48±0.03 | 7.38±0.16 | 2.73 |

a Salting-out equation parameters are reported with corresponding standard deviation of the mean

b The number of data points used to fit parameters a and b from equation (20) is given by n .

Table 3 Dehydrated hydrophobic surface areas computed from experimental salting-out constants^a

| salt | Φ |
|---------------------------------|--------------------------------|
| | (\AA^2 per molecule) |
| KSCN | 6580 |
| KBr | 319 |
| NaBr | 434 |
| NaCl | 288 |
| Na ₂ SO ₄ | 677 |

a These Φ values correspond to the protein surface area that would have to be dehydrated for the theoretical K_s value, from equations (14), (18) and (19), to match the experimental value. The value of Φ computed from the correlation in Figure 1 is 1460 \AA^2 per molecule.

2.11 Figures

Figure 1 Protein relative surface hydrophobicity RSH as a function of the frequency of charged groups CHF . Data for myoglobin, ovalbumin, hemoglobin and albumin were taken from Melander and Horvath.⁶ The dashed horizontal line represents the RSH estimated for αCT using $CHF = 0.12$.

Figure 2 Comparison of experimental data with the theoretical prediction of the relationship between the salting-out constant and the salt molal surface tension increment. Symbols: + KSCN, \square KBr, \circ NaBr, \triangle NaCl and \times Na_2SO_4 . The solid line is the theoretical prediction for 1-1 salts, the dashed is for 2-1 salts. Error bars on the experimental values represent the standard deviation of the mean.

Figures 3 a-e αCT solubility curves for KSCN, KBr, NaBr, NaCl and Na_2SO_4 solutions given in diagrams a through e respectively. The solubility curves are plotted as the natural logarithm of the αCT concentration in grams per kilogram solvent versus the salt molality. The 100-fold dilution data were obtained from the Bio-Rad micro assay procedure; the others used the standard assay. Dilution factors are given by: $\triangle \times 1$, $\square \times 2$, $\times \times 5$, $+ \times 10$, $\circ \times 20$, $\blacksquare \times 100$. The lines represent least square fits to high salt concentration data.

Figure 4 Weighted least squares fitting of NaBr serial dilution salting-out constants to determine the actual salting-out constant. Apparent salting-out constants are plotted versus the correspond-

ing dilution factor. The solid line represents the fit of equation (20) to the data and the dashed line is the fitted salting-out constant. Error bars represent the standard deviation of the mean.

Figure 5 Calculated salt interaction parameters versus salt cut. Symbols: + KSCN, \square KBr, \circ NaBr, \triangle NaCl and \times Na₂SO₄.

Figures 6 a-e α CT activity, active fraction and specific activity as a function of salt cut for KSCN, KBr, NaBr, NaCl and Na₂SO₄ solutions respectively. Note that all data is on a soluble precipitate basis and is expressed in terms of the fraction of the native value. For native α CT, the activity was 47.3 ± 0.9 units, the active fraction was 0.351 ± 0.007 and the specific activity was 135 ± 4 units per gram active protein; limits represent the standard deviation of the mean. Symbols: \triangle activity, \square active fraction and \circ specific activity.

Figure 7 Fraction of precipitate that recovered activity on redissolution as a function of the salt type. Salt type is expressed in terms of the corresponding molal surface tension increment. Symbols: + KSCN, \square KBr, \circ NaBr, \triangle NaCl and \times Na₂SO₄. Error bars represent the standard deviation of the mean.

Figure 8 Salting-out performance parameter as a function of the type of salt. The type of salt is quantified in terms of the corresponding molal surface tension increment. The performance parameter is the product of the salting-out constant and the active fraction and is a measure of the efficiency of a salt in

precipitating protein that is able to recover activity. Symbols: ■ 25°C and ○ 15°C. The salts, in order of increasing σ values, are KSCN, KBr, NaBr, NaCl and Na₂SO₄. Error bars denote the standard deviation of the mean. The solid line is a least squares parabolic fit for the 25°C data; the dotted is at 15°C.

Figure 1 Protein relative surface hydrophobicity versus frequency of charged groups

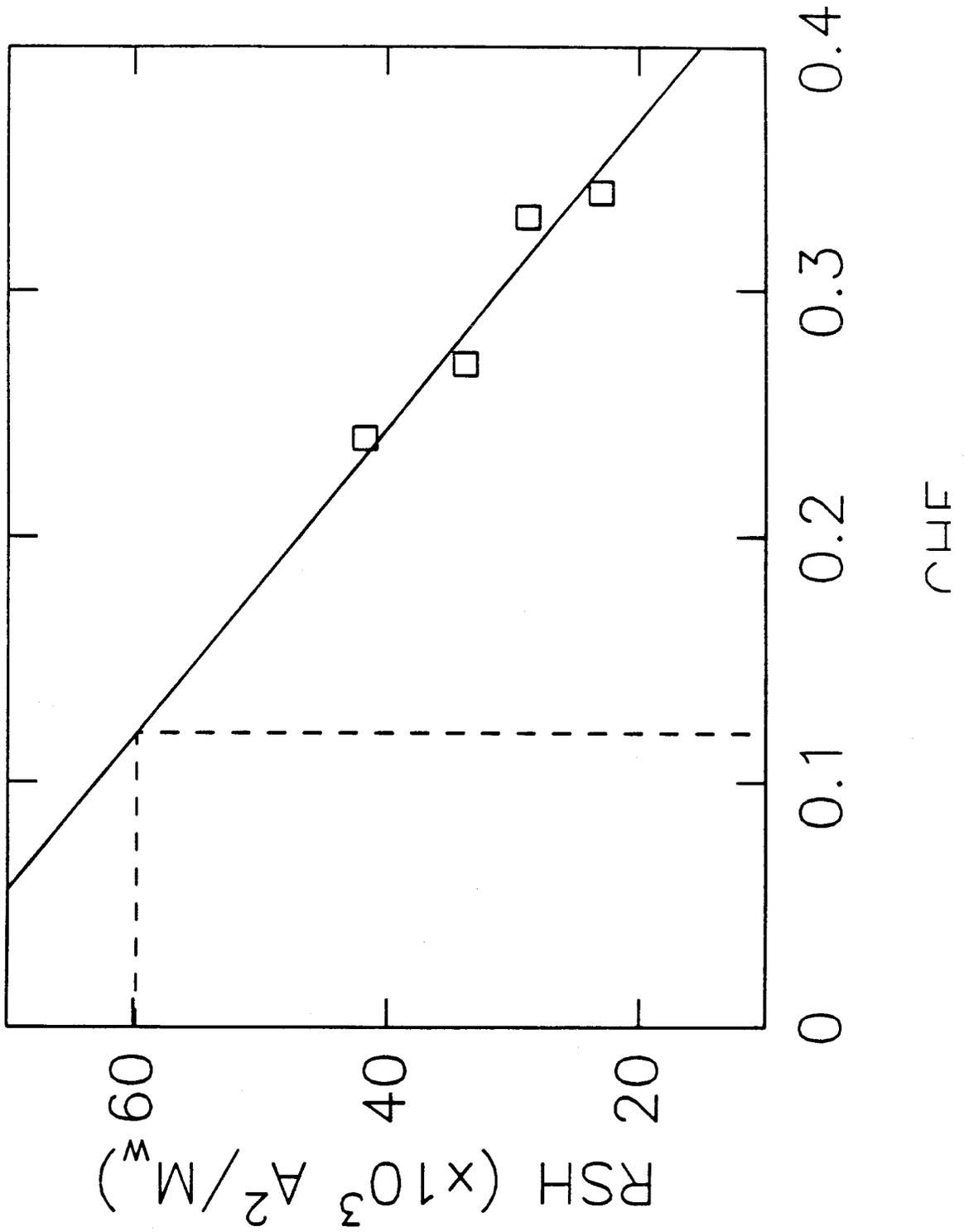


Figure 2 Salting-out constant as a function of salt type

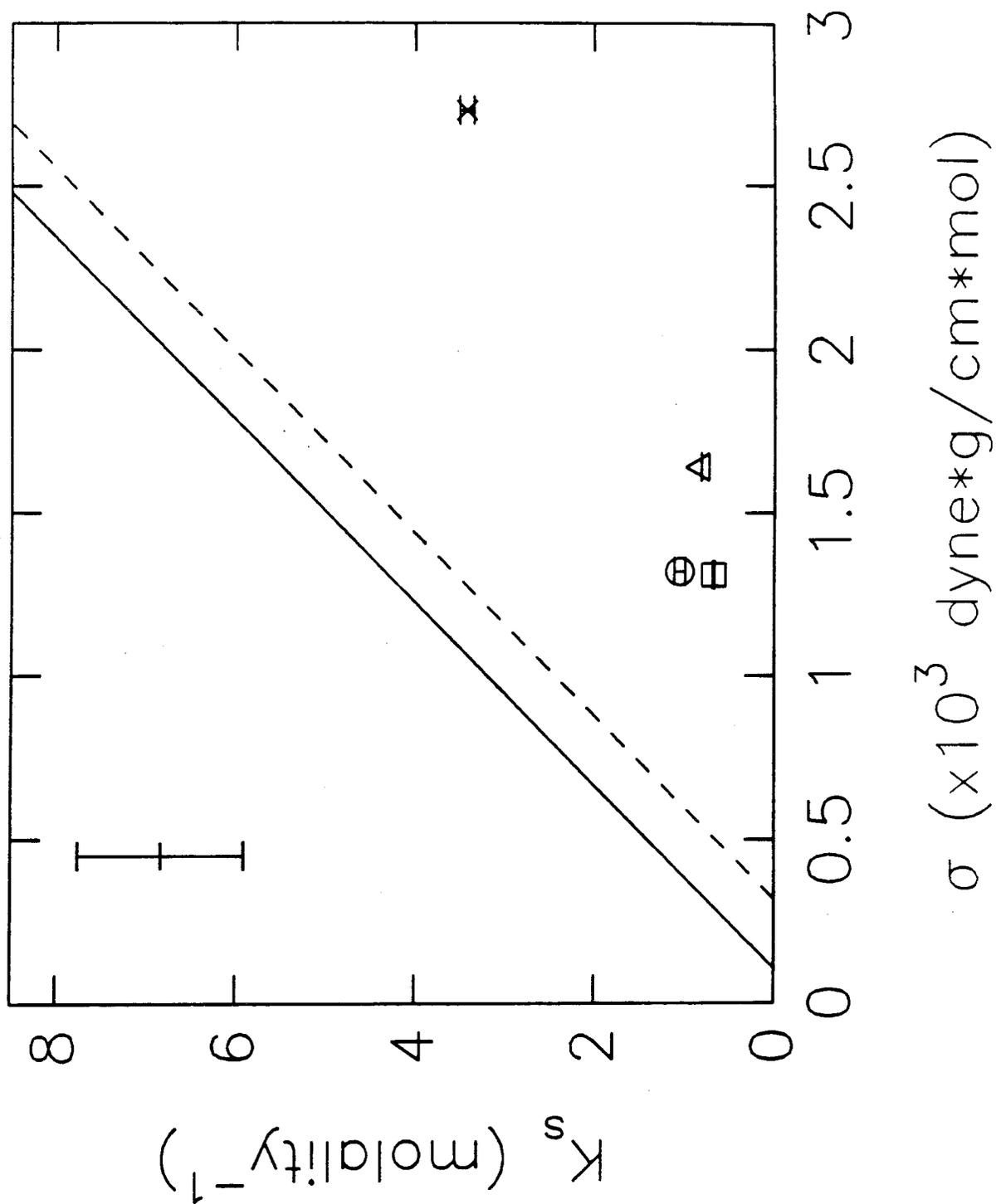


Figure 3 α CT solubility curves

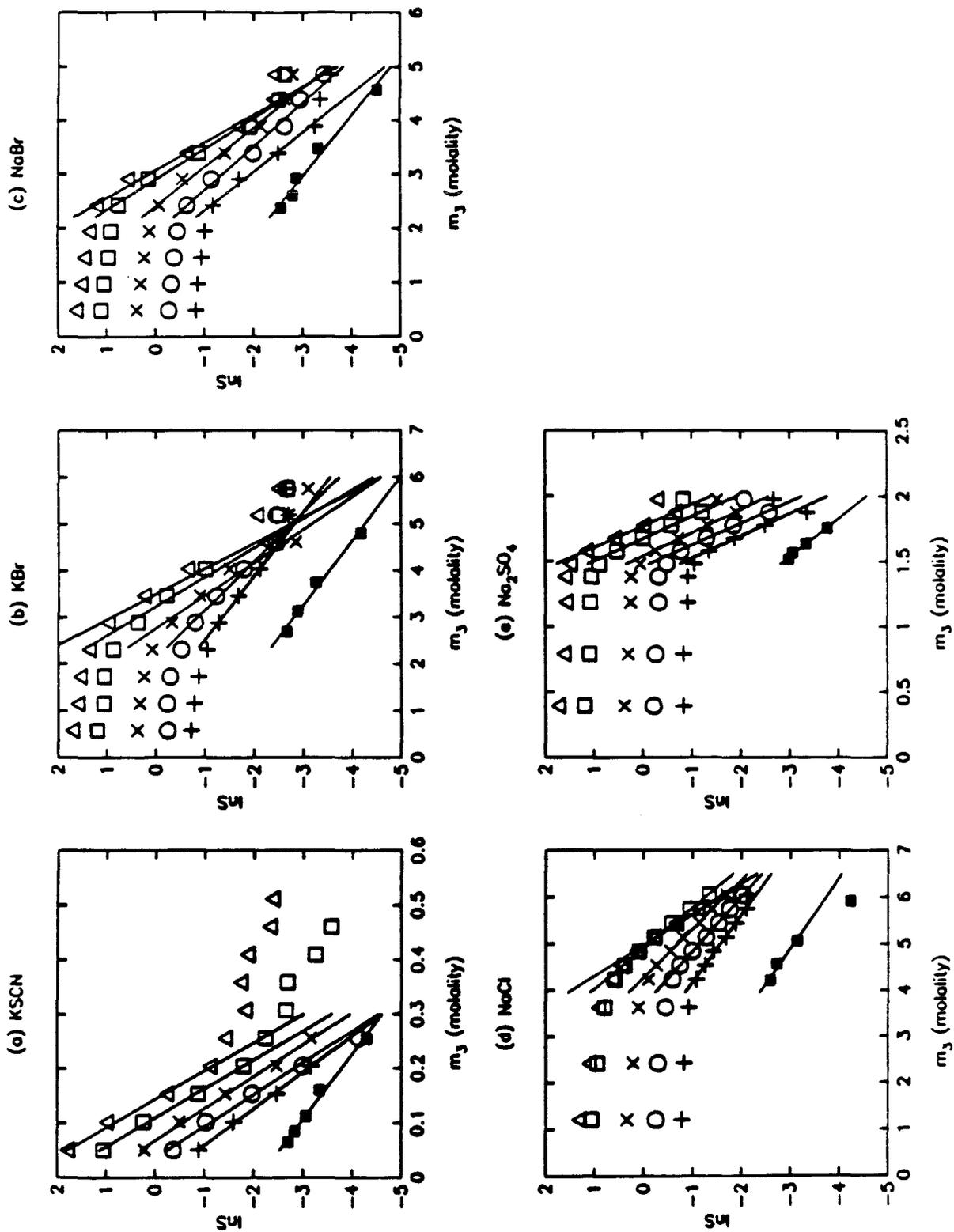


Figure 4 Representative serial dilution data

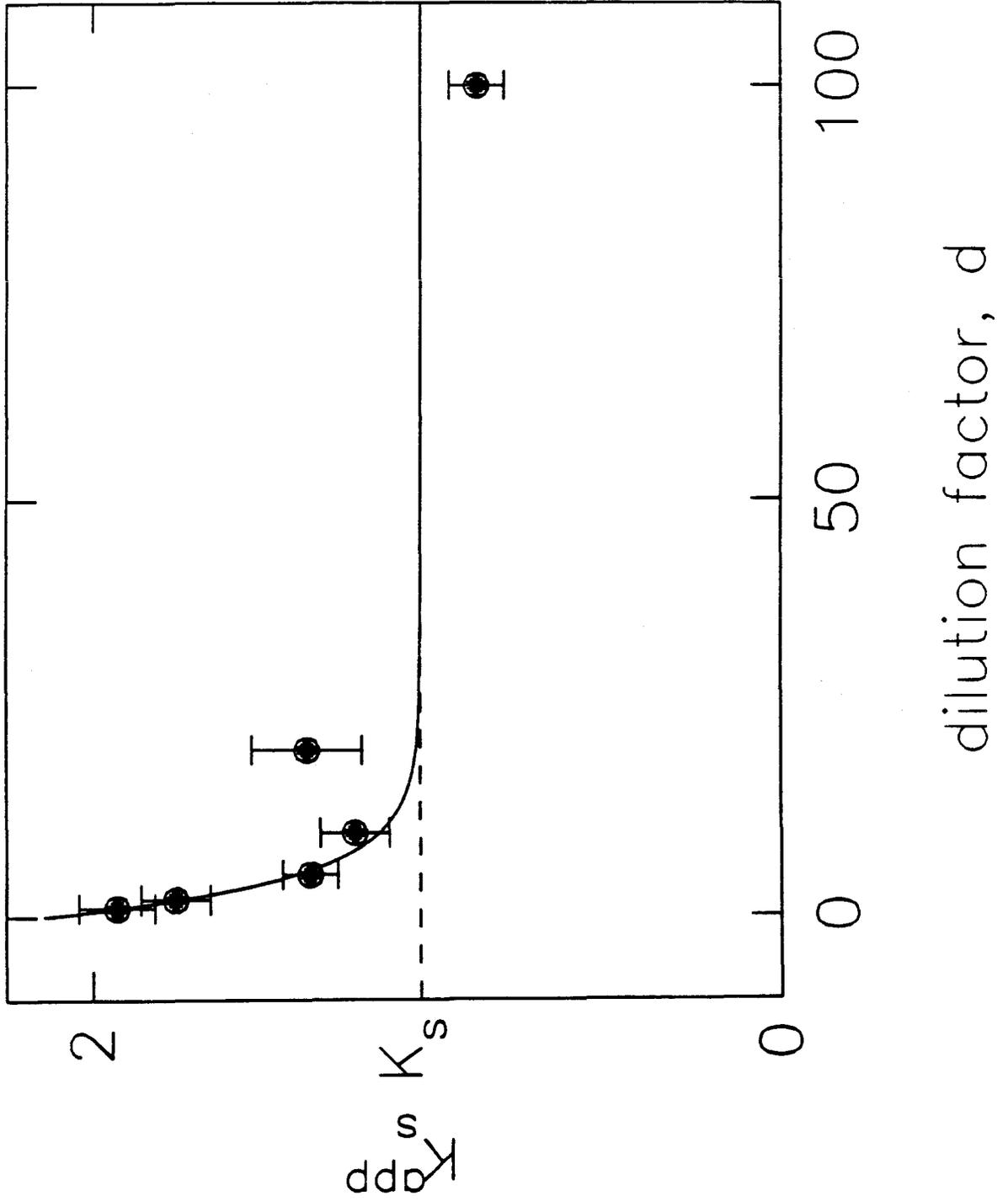


Figure 5 Salt interaction parameters versus salt cut

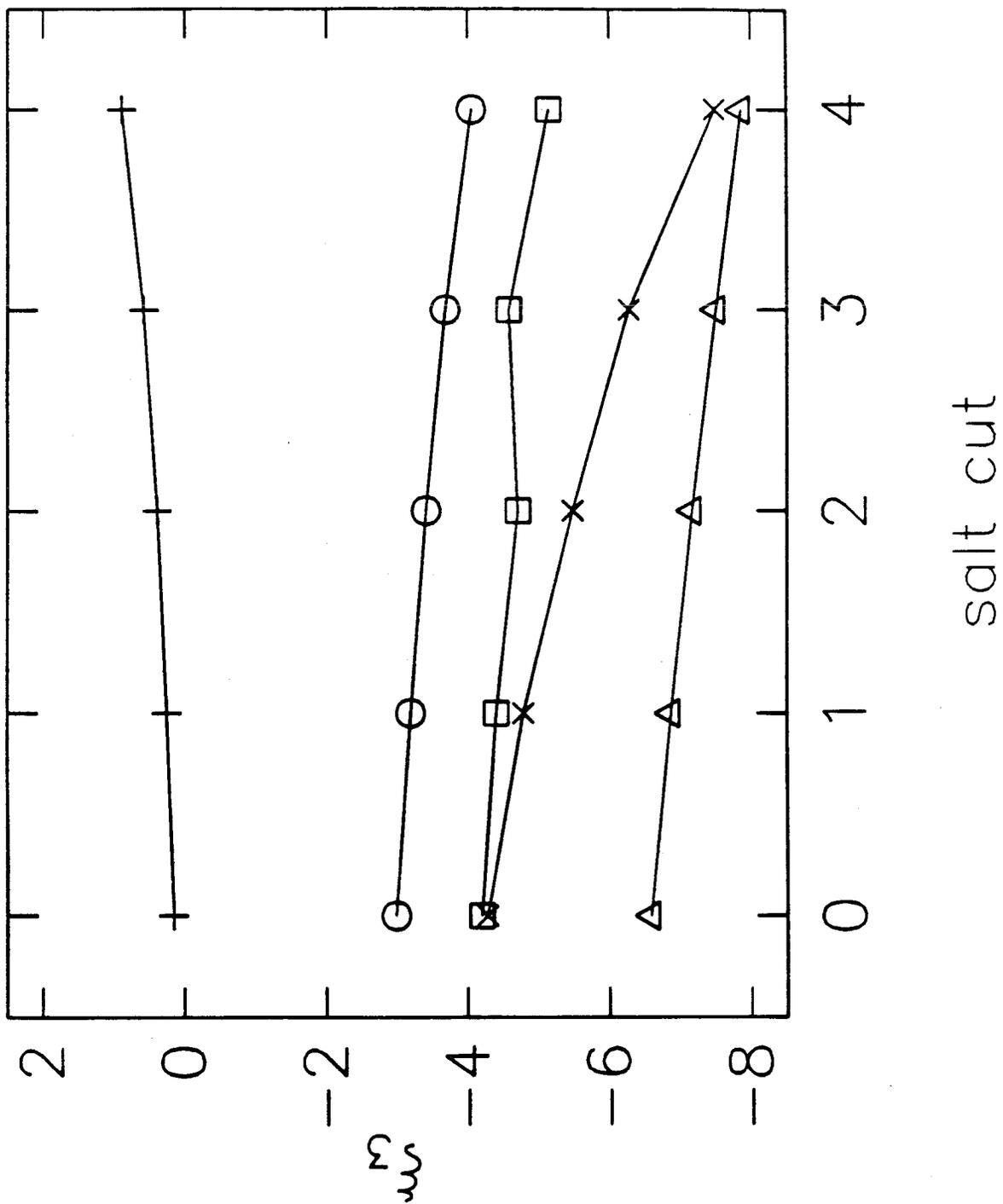


Figure 6 α CT precipitate activity, active fraction and specific activity versus salt cut

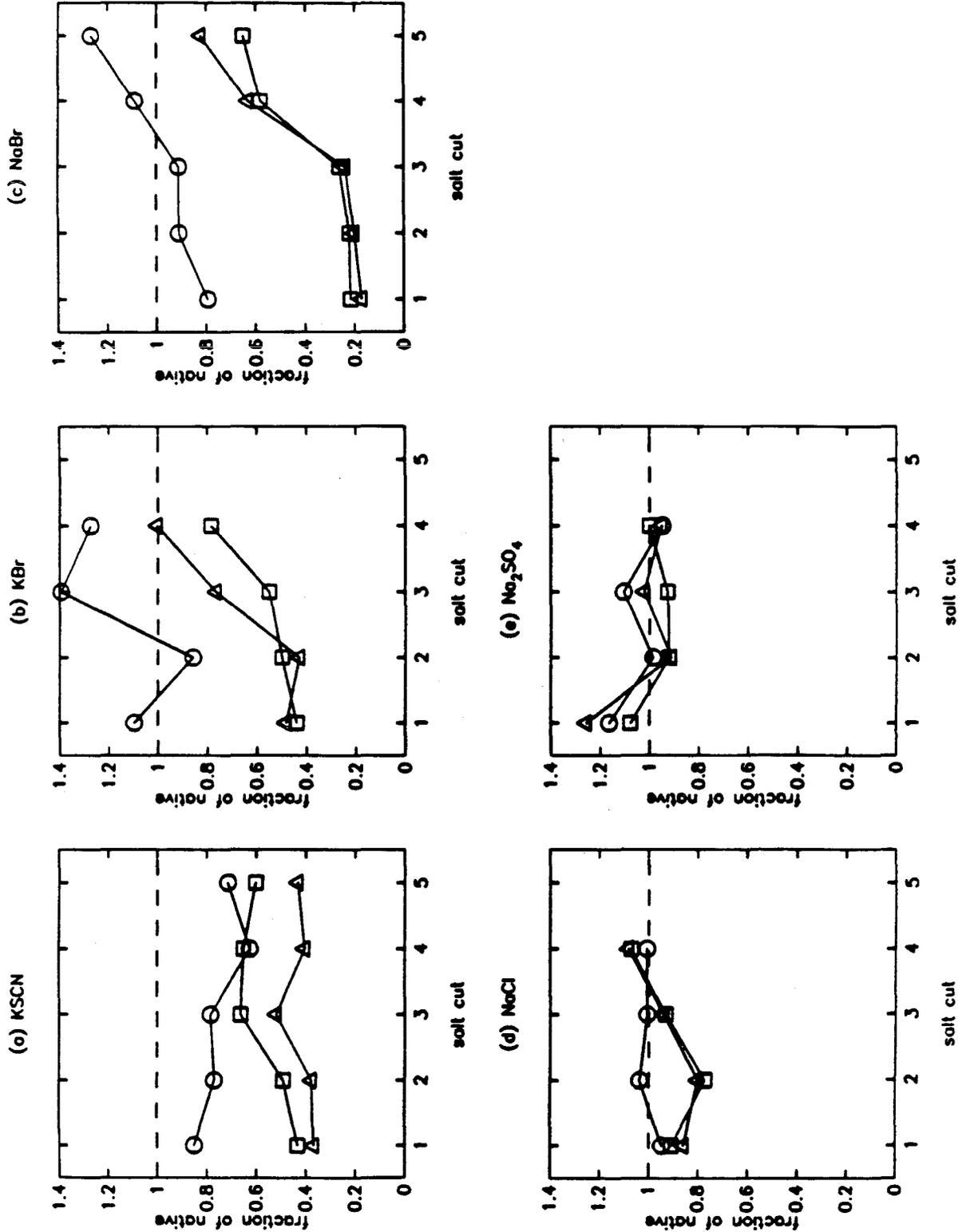


Figure 8 Salting-out performance parameter as a function of salt type

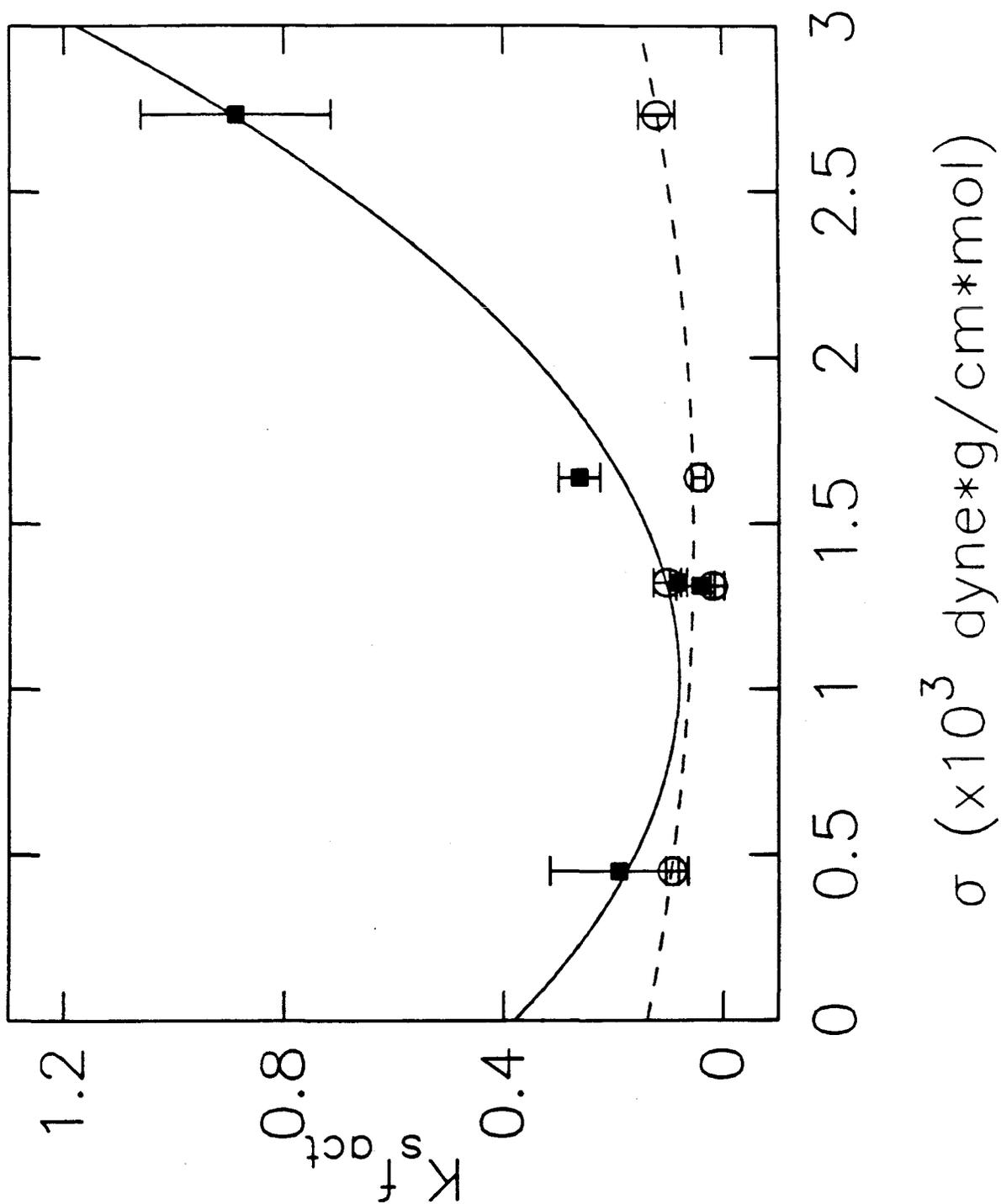
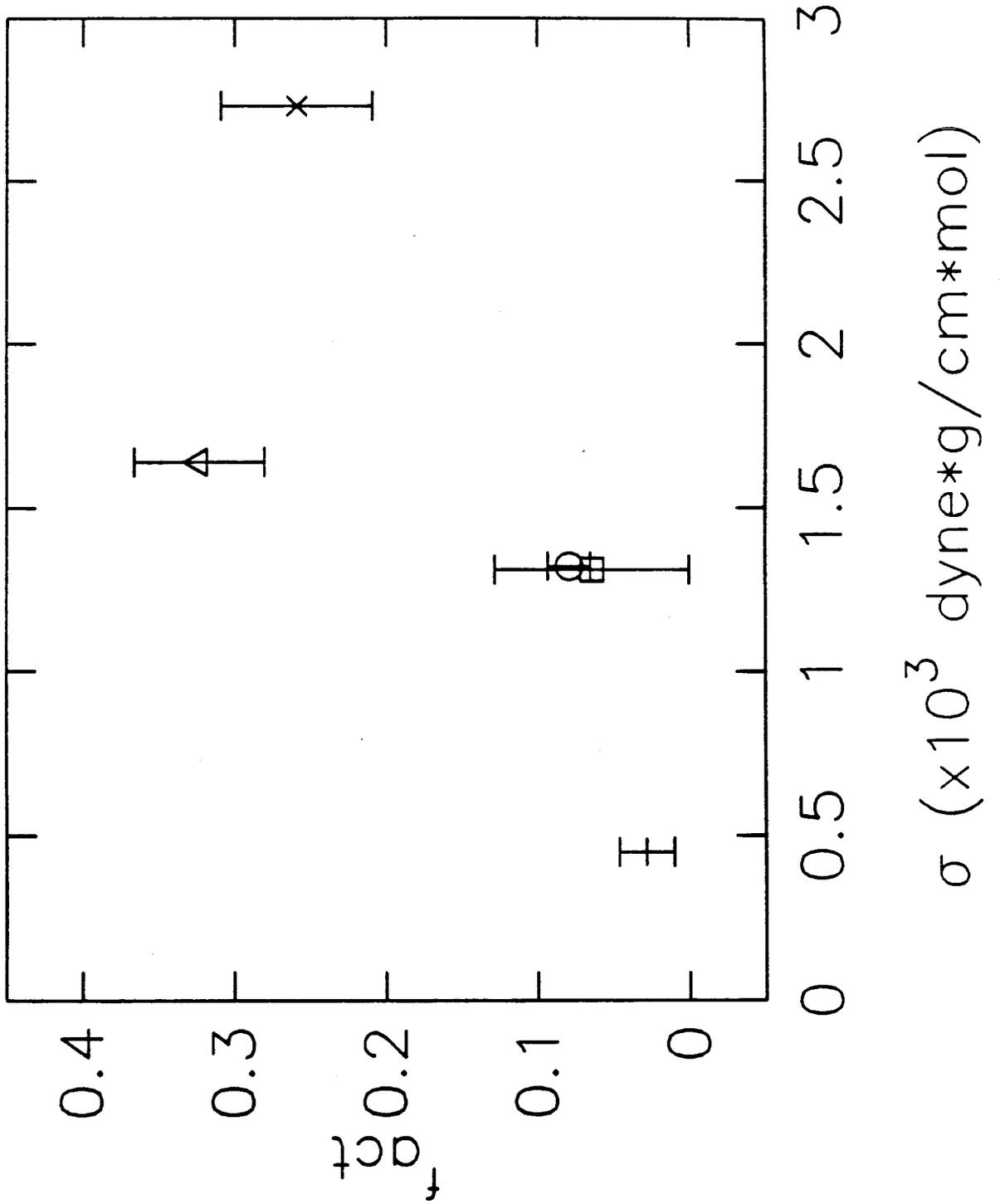


Figure 7 Active protein recovery as a function of salt type



Chapter 3

Structure-Function Relationships in the Inorganic Salt-Induced Precipitation of Chymotrypsin

The material contained in this chapter has been accepted
for publication in *Biochimica et Biophysica Acta*

3.1 Abstract

α -Chymotrypsin (α CT) was used as a model protein to study the effects of salt-induced precipitation on protein conformation. Process parameters investigated included the type and amount of salt used to induce precipitation. The salts studied included Na_2SO_4 , NaCl , NaBr , KBr and KSCN . Precipitate secondary structure content was examined via laser Raman spectroscopy. Conventional and saturation transfer electron paramagnetic resonance spectroscopy were employed to probe the tertiary structure of the active site in spin-labelled α CT precipitates. As the molal surface tension increment of the inducing salt increased, the β -sheet content increased and the α -helix content decreased. There was no significant variation in secondary structure with the amount of salt used. The fraction of precipitate that recovered activity on redissolution was correlated with the change in secondary structure content. Spin-labelled precipitate spectra indicated that the active site remains unaltered during precipitation. Molecular modelling was employed to investigate how physical properties of α CT were affected by these types of conformational change. Estimated physical property changes could not account entirely for observed deviations from current equilibrium theory for salt-induced precipitation. The spectroscopic observations were also combined with activity/solubility results to propose a mechanism for the salt-induced precipitation of globular proteins.

3.2 Introduction

Protein precipitation induced by the addition of inorganic salts is a biological separation/concentration process that exposes proteins to unnatural environments. This may cause alterations in native, active molecular structure [1] with concomitant losses in biological activity [2]. Processing costs in terms of diminished activity of the desired protein, can be significant.

In this work, the nature and extent of protein conformational changes accompanying the salting-out process have been examined. The effects of precipitation on elements of secondary and tertiary protein structure were studied as functions of the type and amount of salt used. Correlations were developed between the observed structural perturbations and the findings of a solubility-activity investigation reported elsewhere [2]. A general mechanism for the salt-induced precipitation of globular proteins has also been proposed.

The model system consisted of a single protein, α -chymotrypsin (α CT). The selection of α CT was motivated by the availability of high resolution crystal structure data [3-5], the ability to selectively spin-label the enzyme active site [6], and the facility with which the concentration, activity and active fraction may be assayed [2]. The salts employed include Na_2SO_4 , NaCl , NaBr , KBr and KSCN . These salts span the lyotropic series, representing a range of structure-stabilizing and chaotropic properties [7]. $(\text{NH}_4)_2\text{SO}_4$, a commonly used precipitant, was not included in this study due to interference with the protein concentration assay [2]; since the lyotropic series is based on the nature of the anion, the effects of $(\text{NH}_4)_2\text{SO}_4$ should be well-represented by the Na_2SO_4 data.

Laser Raman spectroscopy was used to estimate the secondary structure content [8,9] of α CT precipitate samples. Amide I band spectra were analyzed in terms of spectral data from proteins with known secondary structure [10,11].

The relative number of residues belonging to various elements of secondary structure was assigned via a constrained superposition of protein reference spectra; the bulk average fraction of ordered and disordered α -helix, antiparallel and parallel β -sheet, reverse turn and random coil was fit for each sample.

Conventional and saturation transfer electron paramagnetic resonance spectroscopies (EPR, ST-EPR) were used in conjunction with spin-labelled α CT (SL- α CT) precipitates to probe tertiary structure perturbations at the active site [12,13]. The techniques are sensitive (on different time scales) to the rotational mobility of the spin label. The spin label mobility is quantified by the rotational correlation time τ_c and is indicative of the “openness” or extent of unfolding of the labelled region. The conventional technique is sensitive to fast motions $10^{-11} \leq \tau_c \leq 10^{-7}$ s [12]; ST-EPR responds to slower motions, $10^{-7} \leq \tau_c \leq 10^{-3}$ s [14,15].

Secondary and tertiary structure information were combined in efforts to model the conformation of α CT in precipitate phases. The physical properties of a plausible structure were calculated and compared to those for the native enzyme.

3.3 Materials

α -chymotrypsin (EC 3.4.21.1, three times crystallized from four times crystallized α -chymotrypsinogen, salt free), lysozyme (EC 3.2.1.17, three times crystallized, dialyzed and lyophilized from chicken egg white), trypsin (EC 3.4.21.4, dialyzed, lyophilized from bovine pancreas, salt free), elastase (EC 3.4.21.36, purified by affinity chromatography, lyophilized), 4-methylumbelliferone, and 4-methylumbelliferyl-p-trimethylammonium cinnamate chloride (MUTMAC) were purchased from Sigma. The organic free radicals 1-oxyl-2,2,6,6-tetramethyl-

4-piperidinol (TEMPO-4-OH) and 1-oxyl-2,2,6,6-tetramethyl-4-piperidinyl ethoxyphosphorofluoridate (TEMPO-4-EPF) were obtained from Aldrich. Centricon-10 centrifugal microconcentrators with a molecular weight cutoff of 10,000 were purchased from Amicon. Sephadex G-25 was a product of Pharmacia Fine Chemicals. Protein concentrations were determined using a Coomassie Blue G-250 dye binding assay from Bio-Rad. Na_2SO_4 , NaCl, NaBr, KBr and KSCN, and all buffer solution components were analytical reagent grade. Water deionized by passage through two Research IonXchanger Model 2 columns supplied by Illinois Water Treatment Co. was used throughout.

3.4 Methods

3.4.1 Preparation of Precipitate Samples

A series of fractional precipitates was produced for spectroscopic analysis using α CT-salt solution solubility data from [2]. Precipitate from a set of salt cuts of the same starting solution were analyzed to investigate the effects of increasing salt concentrations on protein structure. The details of the sample preparation are described elsewhere [2]. Four or five cuts were taken depending on the salt used; the decrease in α CT solubility corresponding to each salt cut is given in Table 1. All precipitates were produced from 10 mg/mL α CT stock solutions in 50 mM glycine buffer, adjusted to pH 3.00 with aq. HCl, at $25 \pm 2^\circ\text{C}$.

Soluble and insoluble precipitate fractions were produced as in [2] with additional manipulation noted here. Precipitate pellets were lyophilized following centrifugation. α CT was spin-labelled before precipitation for EPR and ST-EPR analysis. Samples were packed into Kimax melting point tubes for spectroscopic work.

3.4.2 Functional Characterization of Precipitate

Precipitate samples prepared and examined in this work are referred to by a classification scheme developed in [2]. Whole α CT precipitate samples were divided into soluble and insoluble fractions by resuspending the precipitate in the glycine buffer. Precipitate which did not dissolve was also considered inactive. Further categorization of soluble precipitate was possible on the basis of activity and active fraction assays. Activity was assessed by monitoring the hydrolytic rate of a chromophoric substrate [16]. The active fraction assay employed the fluorogenic suicide substrate MUTMAC [17] and allowed the classification of soluble precipitate into active and inactive portions. The "total active precipitate fraction" refers to the soluble, active precipitate expressed as a function of the whole precipitate.

3.4.3 Raman Spectroscopy

All spectra were taken with a SPEX 14018 double monochromator with a 2400 line/mm grating and a Hamamatsu multi-alkali, cooled photomultiplier tube. The excitation source was a Spectra-Physics 170 argon ion laser. A custom interference filter placed ahead of the sample stage was used to remove the plasma lines. The photomultiplier was coupled to a SPEX DPC-2 digital photon counter and electrometer. Data retrieval and storage was facilitated by a SPEX SCAMP computer. A thermostatted sample stage was designed and machined for the spectrometer; samples were maintained at 10 °C to dissipate heat generated by the impinging laser beam [10].

Typical operating conditions were as follows: the number of counts full scale was 10^5 , the integration time was 0.2 s, the frequency range scanned was 1500-1800 cm^{-1} (amide I band) and the frequency increment was 1 cm^{-1} . The

entrance and exit slits were set at $540 \mu\text{m}$ and the stray light slits were set at $1080 \mu\text{m}$ for a resolution of 6 cm^{-1} at $20,492 \text{ cm}^{-1}$. The 488 nm line of the Ar laser was used at a power of 0.5 W. Between 25 and 200 scans were collected and co-added for each sample to obtain sufficient signal-to-noise (S/N) levels. All spectra were collected in the backscattering mode.

Precipitate samples were exposed to the laser beam for a maximum of 5 hours. The first 20 to 60 minutes of exposure were used to quench sample fluorescence [18]. Samples were visually inspected for beam damage. In addition, late scans were compared with early scans to ensure that spectral features were not evolving in the laser beam. S/N ratios averaged about 150 for the control spectra and 40 for precipitate spectra after smoothing.

A representative Raman spectrum for a solid sample of native αCT is given in Figure 1. The spectrum consists of five peaks in the $1500\text{-}1800 \text{ cm}^{-1}$ region. Four fairly sharp peaks between 1500 and 1630 cm^{-1} arise from the aromatic vibrational modes of tryptophan, tyrosine and phenylalanine side-chains. The largest feature is the amide I band centered at 1670 cm^{-1} . These five peaks lie on a broad, intense fluorescence background signal. Spectra were examined for glycine artifacts as precipitate samples were lyophilized from glycine buffer mother liquor. Glycine has a sharp peak at 1670 cm^{-1} and another at 1570 cm^{-1} . No peak at 1570 cm^{-1} was evident in sample spectra; the glycine contribution to the amide I band was considered negligible.

A Raman Spectral Analysis Package (RSAP) was written in IBM Professional FORTRAN to carry out data analysis on an IBM PC/XT. RSAP consists of a smoothing routine, a subtraction and normalization program, and a structure fitting routine. The algorithms described by Williams [11] motivated the creation of RSAP. This package may be obtained from the authors. In order to

validate the use of Williams’ spectral database [11] in this study and to highlight differences between the analytical methods used here and in Williams [11], a brief summary of RSAP is given next.

Each spectrum was 5-point smoothed via the algorithm of Savitsky and Golay [19]. The sloping baseline, fluorescence and aromatic peaks were fit simultaneously and subtracted from the smoothed spectra. The aromatic side chain bands between 1500 and 1630 cm^{-1} were fit with Gaussian-Lorentzian product functions. The fluorescence band was approximated by a Gaussian with a bandwidth of 140.1 cm^{-1} and a center frequency of 1616 cm^{-1} [11]. In RSAP, the function fit to the experimental spectra, $S_{\text{cal}}(\nu)$, was

$$S_{\text{cal}}(\nu) = \alpha\nu + \beta + \sum_{i=1}^4 A_i \mathcal{G}_i^{(1-R)} \mathcal{L}_i^R + A_f \mathcal{G}_f \quad (1)$$

where

$$\mathcal{G}_i = \frac{2\sqrt{\pi \ln 2}}{B_i} \exp \left[\frac{-4 \ln 2 (\nu - \nu_{oi})^2}{B_i^2} \right] \quad (2)$$

and

$$\mathcal{L}_i = \frac{2B_i}{\pi B_i^2 + 4\pi(\nu - \nu_{oi})^2}. \quad (3)$$

The first two terms of (1) represent the sloping baseline with slope α and intercept β . The summation term represents the aromatic peak contributions with the Gaussian and Lorentzian lineshape functions of unit area \mathcal{G}_i and \mathcal{L}_i defined in equations (2) and (3), respectively. In (2) and (3), B_i is the bandwidth at half height, ν_{oi} is the center frequency and A_i is the amplitude of the i^{th} peak. The Lorentzian-Gaussian ratio R was fixed at 0.7143, giving the fitted peaks approximately 70% Lorentzian character [20]. The last term of equation (1) represents the Gaussian fluorescence peak.

The values of α , β , ν_{oi} , B_i and A_i were fit simultaneously to experimental spectra, $S_{\text{exp}}(\nu)$, at 1 cm^{-1} intervals using a least squares objective function, Φ

$$\Phi = \sum_{\nu} (S_{\text{exp}}(\nu) - S_{\text{cal}}(\nu))^2. \quad (4)$$

Φ was minimized using a modified nonlinear Marquardt algorithm [21]. The summation over ν in equation (4) is from 1500 to 1630 and 1720 to 1800 cm^{-1} , excluding the amide I band. A representative fit of the individual elements of equation (1) along with $S_{\text{cal}}(\nu)$ to an experimental spectrum is given in Figure 1.

Subtracted spectra were normalized by the sum of the observed intensities between 1615 cm^{-1} and 1710 cm^{-1} at 5 cm^{-1} intervals [11]. Noise levels were estimated by the standard deviation of the spectral intensities between 1730 and 1760 cm^{-1} also at 5 cm^{-1} intervals [11]. S/N ratios of the smoothed, subtracted spectra were calculated as 2.5 times the ratio of the amide I peak amplitude to three times the noise standard deviation; these values were useful in assessing the sensitivity of the structure fits to spectral noise.

Normalized spectra from proteins with known secondary structure contents were used to estimate sample structure contents from the subtracted, normalized amide I band. An experimental spectrum \mathbf{b} , a vector consisting of intensity data from 1630 to 1700 cm^{-1} at 5 cm^{-1} intervals, can be deconvoluted in terms of contributions from reference spectra

$$\mathbf{Ax} \simeq \mathbf{b} \quad (5)$$

where \mathbf{A} is the matrix of reference spectra (15 frequencies \times 13 proteins) and \mathbf{x} is a superposition vector [11]. The sample secondary structure fractions \mathbf{f} may then be calculated by

$$\mathbf{Fx} = \mathbf{f} \quad (6)$$

where \mathbf{F} is the matrix of reference protein structure content (6 structure types \times 13 proteins) [11]. Structure content was assigned in RSAP by combining equations (5) and (6) as

$$\mathbf{A}\mathbf{F}^+\mathbf{f} \simeq \mathbf{b} \quad \text{subject to} \quad \|\mathbf{f}\|_1 = 1.0 \quad \text{and} \quad \mathbf{f} \geq \mathbf{0}. \quad (7)$$

\mathbf{F}^+ indicates the pseudoinverse of \mathbf{F} . The constraints bound the space of physically meaningful solutions; the structure fractions must be non-negative and must sum to unity. The constrained system (7) was solved using the LDP least distance programming procedure developed by Lawson and Hanson [22]. LDP allows both constraints to be incorporated into the problem solution. Matrices \mathbf{A} and \mathbf{F} were taken from reference data in Williams [11].

Because of the sensitivity of system (7) to shifts in spectral frequencies, care was taken to ensure proper monochromator alignment. During each spectrometer start-up, the monochromator was calibrated by running a spectrum of lysozyme and adjusting the frequency of the most intense aromatic band to 1553 cm^{-1} as per the lysozyme calibration spectrum given by Williams [11].

One caveat accompanying the use of system (7) to assign \mathbf{f} is that the matrix $\mathbf{A}\mathbf{F}^+$ represents an averaging of individual protein spectral contributions to a particular structure type. $\mathbf{A}\mathbf{F}^+$ is essentially a matrix of spectra representing pure classes of secondary structure. However, this approach was preferred to the singular value decomposition method proposed by Williams [11]. The singular value decomposition technique requires user intervention in the selection of the best solution from a set of candidate solutions; this method was avoided so that any trends in structure fit results would not be subject to bias.

3.4.4 EPR and ST-EPR Spectroscopy

αCT was spin-labelled with TEMPO-4-EPF, pictured in Figure 2. The

procedure of Morrisett and Broomfield [23] was used, substituting acetonitrile for benzene [24]. The extent of labelling was determined via active fraction assays using the MUTMAC suicide substrate [2]. Typically, more than 97% of the active enzyme was labelled.

The separation of free spin label from the labelled enzyme was performed in two steps. The first consisted of a 4-fold concentration/purification of SL- α CT using Centricon-10 centrifugal microconcentrators. The labelling reaction mixture was transferred to the concentrators and spun at 5 °C at 6500 rpm for 1 hour ($300,000 \times \text{g}\cdot\text{min}$); a JA-20 rotor in a Beckman TJ-25 centrifuge was used. The concentrators were then inverted and spun at 3000 rpm for approximately 3 min ($3000 \times \text{g}\cdot\text{min}$) to collect the retentate. Retentate was pooled and passed through a column of Sephadex G-25 ($3 \text{ cm}^2 \times 45 \text{ cm}$) in the second step. Column pre-equilibration and elution was performed at 5 °C using 50 mM glycine buffer, pH 3.00. 5 mL fractions were collected and assayed for protein content using the Bio-Rad dye binding assay.

SL- α CT precipitate samples were prepared at 25 ± 2 °C and lyophilized as above. Only the first cut (see Table 1) was taken with each salt. Capillary tubes containing SL- α CT precipitate were placed in Wilmad 704 PQ sample tubes for conventional EPR work or used directly for ST-EPR analysis.

All conventional spectra were taken at the X band using a Varian E-Line Century Series EPR spectrometer interfaced to a COMPAQ PC microcomputer. A field set of 3395 G with a scan range of 100 G was used. The modulation amplitude was 4 G at a frequency of 100 kHz. Microwave power at 9.512 GHz did not exceed 1.0 mW. Spectra were recorded using an 8 minute scan time with a 0.128 s time constant. All EPR work was performed at room temperature.

Spin label motion affects the lineshapes of EPR spectra. The rotational

mobility of the spin label is a function of the available free volume; if the active site unfolds, the spin label's motion will be less hindered, resulting in faster motion and smaller rotational correlation times. In the fast motion regime, τ_c may be analytically related to spectral features. From the development of Stone et al. [25], τ_c in seconds is given by

$$\tau_c = 6.52 \times 10^{-10} \Delta H_0 \left[\sqrt{\frac{I_0}{I_{+1}}} + \sqrt{\frac{I_0}{I_{-1}}} - 2 \right] \quad (8)$$

where ΔH_0 is the width in Gauss of the central hyperfine component and I_{m_I} is the peak to peak amplitude of the hyperfine resonance corresponding to the nitrogen nuclear spin quantum state m_I . The constant in equation (8) was computed from typical nitroxyl radical hyperfine splitting and g tensor values [26]. Equation (8) is strictly valid for $\tau_c < 5$ ns [27].

Saturation-transfer electron paramagnetic resonance spectra were taken with an IBM ER200D spectrometer with a TE₁₀₂ cavity. Spectra were recorded at the X band with a modulation amplitude of 5 G. The microwave power was 63 mW. The second-harmonic absorption out-of-phase signal was observed using a modulation frequency of 50 kHz and phase-sensitive detection at 100 kHz, 90° out-of-phase. All ST-EPR spectra were taken at room temperature.

3.5 Results

The salts used in this study represent a range of salting-out abilities [2]. Melander and Horvath [28] have used the salt molal surface tension increment, σ , to correlate a particular salt with its salting-out efficiency. The σ -values follow the order KSCN < KBr ~ NaBr < NaCl < Na₂SO₄ and are listed in Table 2; σ is used here as the variable representing salt type. Salts with greater denaturing potential have smaller σ -values.

Many α CT precipitate samples were composed of soluble and insoluble fractions. Precipitate solubility behavior was assessed by attempting redissolution in glycine buffer. Salts with smaller surface tension increments gave precipitates with more insoluble material [2]. Both original whole precipitate and separated, insoluble precipitate samples were analyzed via Raman spectroscopy.

3.5.1 Secondary Structure Estimation

A series of control spectra were analyzed using RSAP to determine the suitability of the spectrometer scanning conditions and the validity of the structure fits. RSAP estimates the relative amounts of ordered and disordered α -helix, parallel and antiparallel β -sheet, reverse turn and random coil from Raman spectra.

The control for the algorithm employed to solve system (7) consisted of equating the column of **A** corresponding to the lysozyme reference spectrum [11] to the experimental spectrum **b**. The resulting structure assignments were within one structure percent of the lysozyme reference values contained in **F**; small deviations may be attributed to round-off error and the incorporation of the constraints. This demonstrates that RSAP calculations are not introducing artifacts into the structure estimates.

Figure 3 gives original and subtracted amide I band spectra for solid samples of native lysozyme, elastase and trypsin. The spectrum for native α CT is given in Figure 1. The RSAP structure assignments for these proteins and the structure estimates from x-ray crystal structure data are presented in Table 3.

X-ray structure assignments for elastase and trypsin were made from the analyses by Levitt and Greer [29] and the corresponding crystallographic work [3,30,31]. The assignments for lysozyme were taken from the reference protein

matrix **A** in Williams [11]. Residues forming the classic 3.6₁₃ helix were assigned as ordered helix. Disordered helix encompassed 3.0₁₀ and α_{II} helical segments along with two residues from each terminus of 3.6₁₃ segments. In addition, helical residues identified by either Levitt and Greer [29] or in the original x-ray work, but not both, were assigned as disordered helix. β -sheet classifications were made according to chain direction and Levitt and Greer's H-bond and combined H-bond, C $^{\alpha}$ -C $^{\alpha}$ methods [29]. The x-ray structures were used to estimate reverse turn content. The remaining structure fraction was classified as random coil.

The fit to the experimental lysozyme spectrum was passable (see Table 3). The only significant discrepancies were in the underestimation of ordered α -helix content and the overestimation of antiparallel β -sheet content. These discrepancies are reflected in the total helix (H_T) and total sheet (S_T) estimates. The absence of parallel sheet structure is accurately reproduced. The variation of the experimental fit from the reference determination is likely due to lower spectral S/N ratios than in Williams' work [11]; sample exposure to the laser beam was minimized to prevent damage, hence, fewer scans were taken.

Spectra of trypsin, elastase and α CT were also analyzed (see Table 3). Note that none of these proteins is in the reference set. These proteins are structurally homologous and were examined to determine how well RSAP performed with proteins with this particular distribution of secondary structure. Although there is a consistent overestimation of β -sheet content and a smaller underestimation of helix content, the fits are reasonable.

The results from the analysis of lysozyme, α CT, elastase and trypsin give some degree of uncertainty in the absolute assignment of secondary structure content. However, Raman spectroscopy has been shown to be sensitive to confor-

mational perturbations induced in proteins by salts [32] and the RSAP analysis is suitable for assessing relative changes in bulk average secondary structure content.

Two additional control RSAP estimates were run. A lyophilized sample of native α CT was analyzed to determine if the sample work-up perturbed the structure. The structure fit compares favorably with that of native α CT (see Table 3). Lyophilization does not appear to affect the secondary structure content of α CT. Finally, the spectrum of α CT was shifted up and down by one wavenumber and analyzed. The RSAP results are given in Table 3. The average change in structure assignment for a 1 cm^{-1} shift was one structure percent. The antiparallel β -sheet content proved to be the most sensitive structure class.

Raman spectra of first cut whole α CT precipitate induced by Na_2SO_4 , NaCl , NaBr , KBr and KSCN are shown in Figure 4. The more denaturing salts produced subtle changes in the amide I band. Proceeding down the lyotropic series from Na_2SO_4 to KSCN , the amide I band shifts about 5 cm^{-1} to higher frequency and the bandwidth decreases by approximately 13 cm^{-1} . As conditions become increasingly chaotropic, the distribution among classes of secondary structure narrows, favoring more ordered structures. These effects were more pronounced in spectra of insoluble KSCN -induced precipitate samples.

The secondary structure content of a number of α CT precipitate samples for each salt was assigned via RSAP. The results of these structure fits are expressed in Table 4. Precipitate samples from the salt cuts listed in Table 1 were analyzed for KSCN , KBr and NaCl . There was no salt concentration dependence. Precipitate structural perturbations were independent of the amount of salt used. Based on this observation, only first cut samples of NaBr and Na_2SO_4 precipitates were analyzed to economize spectrometer time.

The precipitate structure fits exhibited a significant dependence on the type of salt for salts with molal surface tension increments below a threshold between NaCl and NaBr. Helix content was diminished and may be completely lost in some samples as more denaturing salts were used. Total β -sheet content increased with decreasing σ -values. Again, these effects were more pronounced in insoluble KSCN-induced precipitate samples. Figure 5 gives a plot of the total helix and total sheet contents as a function of the molal surface tension increment of the salt. Where more than one salt cut sample for a given salt was analyzed, an average value is plotted with error bars representing the standard deviation of the mean; for single samples, error bars were estimated. The total helix and sheet contents were used as they were shown to have the best agreement between the Raman and x-ray diffraction estimates [11] and the standard deviations of the entire class contents were smaller. Only the changes from the native fitted values for the NaBr, KBr and KSCN precipitates are significant. Also there is no significant difference between the values estimated for the soluble precipitates induced by these three salts. The KSCN whole precipitates had an average absolute increase in β -sheet content of about 10 structure percent; insoluble KSCN precipitates averaged a 15 structure percent increase.

The Euclidean norm of the residual vector, RNORM, is given by [11]

$$\text{RNORM} = \|\mathbf{A}\mathbf{F}^+\hat{\mathbf{f}} - \mathbf{b}\|_2 \quad (9)$$

where $\hat{\mathbf{f}}$ is the solution vector of system (7). RNORM gives an estimate of how well the structure fit describes the experimental spectrum and is reported for each fit in Tables 2 and 3. For the precipitate spectra, RNORM averaged 0.040 giving a root mean square fitting error of about 0.016 structure percent for each structure type. The RNORM values were negatively correlated with the spectral

S/N ratio.

A few of the structure estimates are beyond the range of contents found in the reference protein set for individual elements (see Table 2); this occurs because the superposition vector \mathbf{x} is implicitly unconstrained in system (7). Results of this type represent an extrapolation of reference data rather than an interpolation. The extrapolation arises in a few fits where the fraction of antiparallel β -sheet contents exceeds 0.65.

Secondary structure perturbations are correlated with changes in the active fraction on a whole precipitate basis. Active fraction data were taken from Figure 7 of [2]. Figure 6 gives a plot of the active fraction as a function of the β -sheet content; quantities have been normalized by the corresponding native protein values. The active fraction of α CT incubated at pH 3.00 and assayed at its activity maximum at pH 7.9 is approximately 35% [2,33]. Precipitates with the greatest β -sheet levels had the lowest total active fractions.

3.5.2 Active Site Tertiary Structure

EPR spectra of TEMPO-4-OH and SL- α CT in solution are shown in Figure 7. In order to eliminate apparent spectral shifts due to variations in the microwave frequency from experiment to experiment, spectra are presented in terms of g -values

$$g = 714.46\nu/B \quad (10)$$

where ν is the microwave frequency in GHz and B is the magnetic field strength in Gauss [34]. Rotational correlation times were computed via equation (8). τ_c 's were approximately 23 ps and 2.2 ns for solutions of free spin label and SL- α CT respectively. The τ_c for SL- α CT is similar to that reported by Clark and Bailey [12]. The spectrum of TEMPO-4-OH is representative of the class of nitroxyl

radicals to which TEMPO-4-EPF belongs and indicates that SL- α CT samples are not contaminated with residual free spin label from the labelling reaction.

Since the Raman analysis gave evidence that structural perturbations were not a function of the amount of salt used to precipitate α CT, only first cut SL- α CT precipitates were examined with EPR. The irreversible inhibition of α CT with the spin label did not appreciably affect the solubility behavior of the protein; SL- α CT solubilities were well-represented by correlations obtained for α CT [2].

Solid SL- α CT precipitate samples produced rigid limit ($\tau_c > 0.1 \mu\text{s}$) EPR spectra. Figure 8 gives representative EPR spectra of lyophilized SL- α CT and Na_2SO_4 and KSCN-induced SL- α CT precipitates. The spectra are virtually identical. Spin label motion is restricted to similar extents even though secondary structure content is altered. Strongly hindered spectra imply that the active site does not unfold during precipitation. EPR spectra are sensitive to unfolding; spectra of SL- α CT immobilized on CNBr-activated Sepharose 4B beads sharpened upon addition of a denaturant (n-propanol) [24] as did spectra of SL- α CT treated with urea [6]. If the active site had unfolded on precipitation, spin label rotation would be less hindered, and the decreased correlation times would have resulted in sharper resonances.

An examination of the separation of $m_I = -1$ and $m_I = +1$ hyperfine resonances eliminated the possibility that the spin label is immobilized in different environments at the active site. The position of the outer hyperfine resonances is indicative of the polarity of the spin label environment. Assuming the anisotropic hyperfine splitting tensor is axially symmetric, the separation of the extrema is quantified by $2A_{\parallel}$ (see Figure 8). The average value of $2A_{\parallel}$, computed from first cut SL- α CT precipitate spectra for all five salts studied and the

lyophilized SL- α CT spectrum, was 68.5 ± 0.4 G (mean \pm S.D.). Precipitation does not appear to bring about any changes in the polarity of the active site environment.

To probe for active site perturbations at slower time scales, the same first cut SL- α CT precipitate samples were analyzed via ST-EPR spectroscopy. ST-EPR spectra of Na₂SO₄ and KSCN induced precipitate samples are presented in Figure 9 along with a spectrum of lyophilized SL- α CT. The precipitate spectra for all the salts used were similar to that of the native, labelled enzyme. If the spin label was sampling altered active site environments in the precipitate samples, the resulting variation in spin-lattice relaxation times would affect the lineshape of the ST-EPR spectra [35,36].

The ST-EPR spectra are indicative of strongly hindered spin label motion. Analysis of spectral features with reference graphs from maleimide spin-labelled hemoglobin are consistent with τ_c values in the millisecond regime [36]. If there are any perturbations of active site structure on precipitation, they are not resolvable in the slow motion limit of ST-EPR spectroscopy.

3.5.3 Molecular Modelling

The secondary and tertiary structure content information was combined in an effort to propose a plausible structure for α CT precipitate phases. The KSCN whole precipitates were selected for modelling work as they had the most perturbed conformations. The BIOGRAF molecular modelling system was used [37]. An analysis of the physical properties of the altered conformation may aid the understanding of the protein-associated parameters that are important in the precipitation process.

Since α CT has 241 amino acids, and in view of the average structure frac-

tions assigned to the KSCN whole precipitates (see Table 4, line 26) and the native protein (see Table 3, line 9), 22 residues assume the β -sheet conformation while 17 residues of helix are lost on the average relative to the native structure. Rounding errors gave fractions of secondary structure that sum to 1.02 for native α CT; given the necessarily speculative nature of this modelling approach, only 5 residues of undefined structure were presumed lost in order to close the structure balance. The modelling procedure consisted of building these altered structure elements into the native protein using BIOGRAF. The resulting structure is proposed as a paradigm for the bulk average protein conformation found in whole precipitate. The crystal structure of Blevins and Tulinsky [5] was used as the starting point; the coordinates were obtained from the Brookhaven Protein Data Bank [38].

Strands of β -sheet were inserted into the protein structure at the expense of regions of undefined structure. The locations of additional β -sheet strands were selected using the following criteria: the target segment must have an extended structure initially, existing β -sheet and reversed turn elements remain undisturbed, the active site must retain its native conformation and the target segment must have β -sheet-forming potential based on the heuristics of Chou and Fasman [39]. The selected segments include Cys₁-Pro₈, Val₁₁₈-Ser₁₂₅ and Asp₁₂₈-Gly₁₃₃, all of which are random coil in the native structure; the chymotrypsinogen A numbering system has been retained. A total of 22 residues were chosen. The amide bond dihedral angles Φ and Ψ of these segments were set to -139° and $+135^\circ$ respectively [40].

The C-terminal helix, Ala₂₃₃-Asn₂₄₅, was converted to random coil. The dihedral angles Φ and Ψ of the helix were arbitrarily bounded and initial angles were assigned to each residue using random numbers between 0 and 1 to inter-

polate between the bounds [41]. The bounds were set using the Ramachandran diagram of α CT in the neighborhood of the allowed values for right-handed α -helix [3]; the limits were $-120^\circ \leq \Phi \leq -30^\circ$ and $-70^\circ \leq \Psi \leq -25^\circ$.

In the altered structure, there was no conversion of α -helix directly to β -sheet. Random coil segments were consumed to form β -sheets and produced by the disruption of the C-terminal helix. The helix and sheet forming potentials of the primary structure [39] of the C-terminal helix imply amorphous character [42]. However, since β -sheets are greatly extended structures when compared to helices with the same number of residues, transforming the C-terminal helix to a β -sheet results in a radically altered structure. This option was less favorable energetically in the minimizations described below. This does not preclude the unexplored possibility that a portion of the C-terminal helix acquires β -sheet character. The internal helix extending from Asn₁₆₇ to Ile₁₇₆ was not considered because of the topological requirements of this transition.

The modified structure was constructed in two steps. The first consisted of excising the target segment and altering the dihedral angles as determined above. The altered segment was then docked and new amide bonds were made. For the β -sheet segments, the extended structure necessitated the formation of non-planar, lengthened amide bonds. The altered structure was subjected to a constrained energy optimization to arrive at more reasonable amide bonds at the ends of the altered segments and to remove any unfavorable side chain interactions introduced in the second step. The optimization was initiated with 100 steps of energy minimization. The resulting structure was allowed to transcend local energy minima by running straight dynamics at 300 K for 10.0 ps with time steps of 0.002 ps. The optimization was concluded with another 100 steps of energy minimization to regularize any bonds that may have been lengthened in

the dynamics run. All the optimization procedures employed the Dreiding force field [37] and held the pre-existing strands of β -sheet and residues at the active site fixed in space. The torsional angles Φ and Ψ of the altered segments were constrained so that the optimization procedure would not corrupt the desired conformation. Dihedral angle torsions corresponding to the new β -sheet segments were assigned a force constant of 100 kcal/mol and those relating to the randomized C-terminal segment were assigned a force constant of 10 kcal/mol; the default value of the force constants for the remaining dihedral angle torsions was 20 kcal/mol. Thus the new β -segments were rigidly constrained while the new random coil residues were allowed a somewhat greater freedom to relax from the initial assignments.

The α -carbon backbone of the native protein structure is given in Figure 10. Residues belonging to anti-parallel β -sheet, α -helix and active site structures have been highlighted. The β -sheet segments form two β -barrel structures with approximately orthogonal axes of revolution. A similar rendering of the altered, optimized α CT molecule is given in Figure 11. It is noteworthy that the three β -strands constructed based on the heuristics summarized above are located in a cluster on the surface opposite the active site.

Protein physical properties thought relevant to the salting-out process include the hydrophobic surface area and the dipole moment [28]. These properties were computed for the modified molecule and compared to those found for the native enzyme. The hydrophobic surface area was computed with BIOGRAF using a probe radius of 1.4 Å (the van der Waals radius of water). The dipole moment was calculated as in [2]. The results are compiled in Table 5. The hydrophobic surface area of the modified structure was 26% larger than that of the native protein; the dipole moment decreased by 22%. Using the Melander and

Horvath [28] approach, the modified structure gives a salting-out constant that is about 33% greater than that predicted for the native protein. However, this new value is still less than the experimentally determined salting-out constant.

3.6 Discussion

The structure of salt-induced α CT precipitates is dictated by the particular salt employed. The salt cut, or the salt concentration used to produce a given amount of protein precipitate, does not appear to impact the conformation of α CT precipitates. Precipitate secondary structure assignments were invariant to the salt cut. This corroborates results obtained in solubility-activity studies; the fraction of α CT precipitate that recovered activity upon redissolution was essentially independent of the amount of salt used for a given salt [2].

α CT has been shown to be a salt-dependent equilibrium mixture of active and inactive conformers in solution [43-45]. Cuppet et al. [44] found increased α -helix content in α CT as KCl concentrations increased from 0.0 to 0.5 M. However, Raman spectra of α CT precipitates exhibited no coherent trends as a function of salt cut for a given salt. Precipitates induced by KSCN, KBr and NaBr, the more denaturing salts, had reduced helix contents. The protein conformational perturbations associated with precipitation may dominate those observed in dilute salt solution.

The secondary structure fits indicate that α CT precipitation involves the formation of additional β -sheet structures. These results are similar to those obtained from Raman, infrared and circular dichroism spectroscopic studies of other protein aggregates. Lysozyme gels formed in methanol and ethanol were found to have increased β -sheet contents and decreased α -helix contents [46]. This increase in β -sheet structure in lysozyme was attributed to intermolecu-

lar rather than intramolecular strand interactions. Thermally induced globular protein gels of α CT, bovine serum albumin, insulin, glucagon and β -lactoglobulin were also found to have increased sheet and decreased helix contents [47]. The agreement of these observations with those reported here lends credibility to the use of α CT as a representative globular protein. The formation of β -sheet strands may be a fundamental phenomenon in self-associating protein systems.

It is possible that the augmented β -sheet content in α CT precipitates is due to intermolecular strand associations [46,47]. Some of the fitted structures had increased parallel rather than antiparallel β -sheet contents; this may depend on the orientation of the individual protein molecules in the quaternary structure of the precipitate. Alternatively, the similarity of the structural reference spectra for pure antiparallel and pure parallel β -sheet in the matrix \mathbf{AF}^+ (see equation 7) [10] may render the distinction between the two structures somewhat artificial [48]. For the reference set used in this work (13 proteins from Table 5 of [11], excluding avidin, fd phage and the helical form of poly-L-lysine), the condition number of \mathbf{AF}^+ was approximately 22.2 and the smallest singular value was about 0.0303. Thus, \mathbf{AF}^+ is relatively well-conditioned and computations should be numerically stable. However, there exists some degree of degeneracy between the columns of \mathbf{AF}^+ representing antiparallel and parallel β -sheet; the dot product of the corresponding normalized column vectors is about 0.911. The distinction between the two structure types may be further trivialized by the lower S/N ratios associated with some of the precipitate spectra.

The total β -sheet content appears to be invariant to salt cut for a given salt type. Insoluble KSCN-induced precipitate samples had the highest β -sheet contents. The extent of these structures determines the solubility of the precipitate. Chaotropic agents such as 6 M guanidine-HCl or 8 M urea are needed to

disrupt these precipitates [2]. The intermolecular H-bonds formed by interacting β -sheet strands stabilize protein-protein interactions in precipitate aggregates.

Since recovered activities remained essentially constant at the native protein values [2], protein in the precipitate phase may be inactivated sterically by physical association with the aggregate. The correlation of secondary structure content with the inactive precipitate fraction suggests that salts may have two distinct modes of interaction. Salts may be either chaotropic or structure-stabilizing.

α CT precipitation does not result in a gross unfolding event at the active site. EPR and ST-EPR spectra of protein precipitates induced by Na_2SO_4 , NaCl, NaBr, KBr and KSCN were nearly identical; the active site spin label was hindered rotationally in environments of similar polarity in all precipitate samples. At acidic pHs, α CT dimerizes in the presence of salts [49-51]. Dimerization may be the first step in the aggregation process [52]. The active site region forms the interface between the molecules of dimeric α CT in crystals formed from the native enzyme at low pH [4,5]. Since dimerization only occurs for the active form [43], the active site may be protected from conformational change. This may account for the generally complete recovery of original activity observed in the soluble precipitate fractions [2].

Physical properties calculated for a hypothetical precipitate phase monomer may be used to characterize the nature of the discrepancy between the theoretical and experimental performance of salts in a salting-out process. Using the values of the hydrophobic surface area and the dipole moment calculated for the native and altered α CT structures, salting-out constants for KSCN were computed and compared with the experimental value [2] (see Table 5). Although the changes in the physical properties gave a salting-out constant closer to the experimental

value, the magnitudes were not large enough to account for the discrepancy. The arbitrary structural changes introduced into the native structure may not reflect the actual extent of the perturbations and/or an additional salt-protein interaction has not been accounted for in the theory [2].

A general mechanism for the salt-induced precipitation and precipitate redissolution behavior of globular proteins is proposed in Figure 12; the mechanism is based on the spectral data and solubility-activity data [2] collected for α CT. The relative magnitudes of the associated rate and equilibrium constants for different conditions are also given. \mathcal{N} refers to the native solution conformation of the protein. \mathcal{RD} and \mathcal{ID} represent a lumping of distinguishable reversibly and irreversibly denatured solution conformations, respectively. A subscript p indicates the corresponding precipitate phase. Horizontal reactions represent conformational changes and are a function of the salt type or σ only. If α CT is representative of globular proteins, β -sheet content increases with a concomitant decrease in α -helix content proceeding from left to right. Vertical reactions involve phase changes accompanying aggregation and redissolution and depend on the protein supersaturation. The supersaturation is determined by the salt type, salt cut and protein concentration. This mechanism is similar to that proposed for the alcohol-induced precipitation of catalase [53].

As indicated in Figure 12, the rate and equilibrium constants are inherently dependent on the type and strength of the electrolyte solution; this dependence eliminates the apparent net flux to the insoluble, irreversibly denatured precipitate phase $\mathcal{ID}_p^{\text{insol}}$. Under precipitating conditions, the redissolution rates are negligible resulting in large phase equilibrium constants (K_{pi}). Further, owing to the size and presumed heterogeneous nature of the precipitate particles, only a fraction of the precipitate remains in communication with the bulk solution.

The phase equilibria are valid only at the particle-solution interface. Irreversible denaturation is significant only when chaotropic salts are employed. During redissolution, the phase equilibria shift to favor solubilization and the driving force for denaturation is removed.

Solubility and activity measurements on α CT precipitates indicated that a portion of the soluble precipitate was not able to recover activity and that for the more denaturing salts, insoluble precipitates were formed [2]. In KSCN-induced precipitates, for example, soluble and insoluble, irreversibly denatured protein coexisted with soluble protein that recovered full activity. Spectral data indicated that conformational change accompanies precipitation. How far protein proceeds horizontally in the scheme depicted in Figure 12 depends on the interaction of the particular salt with the protein. The aggregation rate may be conformation-dependent; the conformation may affect the number of possible intermolecular interaction sites. If the β -sheet interactions are involved in stabilizing aggregates, the extent of sheet formation will affect the aggregation rate. The coexistence of many different states of protein in precipitate implies that the rate of conformational change may be comparable to that of aggregation.

3.7 Acknowledgements

The authors are indebted to Dr. Robert Williams of the Naval Research Laboratory for providing us with the source code for his Raman spectral analysis programs and to Dr. Larry Dalton of the University of Southern California for taking the ST-EPR spectra and assisting with their interpretation. This research was supported by the National Science Foundation.

3.8 References

- 1 Von Hippel R.H. and Scheich T. "The Effects of Neutral Salts on the Structure and Conformational Stability of Macromolecules in Solution," in Structure and Stability of Biological Macromolecules (Timasheff, S.N. and Fasman, G.D., Eds.) Marcel Dekker, New York 417-574 (1969)
- 2 Przybycien, T.M. and Bailey, J.E. "Solubility-Activity Relationships in the Inorganic Salt-Induced Precipitation of α -Chymotrypsin," *Enz. and Micro. Tech.*, (1989) in press
- 3 Birktoft, J.J. and Blow, D.M., "Structure of Crystalline α -Chymotrypsin V. The Atomic Structure of Tosyl- α -chymotrypsin at 2Å Resolution," *J. Mol. Biol.* **68**, 187-240 (1972)
- 4 Tsukada, H. and Blow, D.M., Structure of α -Chymotrypsin Refined at 1.68Å Resolution," *J. Mol. Biol.* **184** 703-711 (1985)
- 5 Blevins, R.A. and Tulinsky, A., "The Refinement and the Structure of the Dimer of α -Chymotrypsin at 1.67Å Resolution," *J. Biol. Chem.* **260**, 4264-4275 (1985)
- 6 Berliner, L.J., "Urea Denaturation of Active-Site Spin-Labeled α -Chymotrypsin," *Biochemistry* **11**, 2921-2924 (1972)
- 7 Carrea, G., Bovara, R. and Pasta, P., "The Effect of Hofmeister Anions and Protein Concentration on the Activity and Stability of Some Immobilized NAD-Dependent Dehydrogenases," *Biotechnol. Bioeng.* **42**, 1-7 (1982)
- 8 Spiro, T.G. and Gaber, B.P., "Laser Raman Scattering as a Probe of Protein Structure," *Ann. Rev. Biochem.* **46**, 553-572 (1977)
- 9 Peticolas, W.L., "Classical and Resonance Raman Spectroscopy of Biological Macromolecules," *NATO Adv. Study Inst. Ser. A (Struct. Mol. Biol.)***45**, 237-267 (1982)

- 10 Williams, R.W. and Dunker, A.K., "Determination of the Secondary Structure of Proteins from the Amide I Band of the Laser Raman Spectrum," *J. Mol. Biol.* **152**, 783-813 (1981)
- 11 Williams, R.W., "Estimation of Protein Secondary Structure from the Laser Raman Amide I Band," *J. Mol. Biol.* **166**, 581-603 (1983)
- 12 Clark, D.S. and Bailey, J.E., "Structure-Function Relationships in Immobilized Chymotrypsin Catalysis," *Biotechnol. Bioeng.* **25**, 1027-1047 (1983)
- 13 Hyde, J.S. and Dalton, L., "Very Slowly Tumbling Spin Labels: Adiabatic Rapid Passage," *Chem. Phys. Lett.* **16**, 568-572 (1972)
- 14 Dalton, L.R., Robinson, B.H., Dalton, L.A. and Coffey, P., "Saturation Transfer Spectroscopy," *Adv. Magn. Reson.* **8**, 149-249 (1976)
- 15 Hyde, J.S. and Thomas, D.D., "Saturation Transfer Spectroscopy," *Ann. Rev. Phys. Chem.* **31**, 293-317 (1980)
- 16 Hummel, B.C.W., "A Modified Spectrophotometric Determination of Chymotrypsin, Trypsin, and Thrombin," *Can. J. Biochem.* **37**, 1393-1399 (1959)
- 17 Gabel, D., "Active Site Titration of Immobilized Chymotrypsin with a Fluorogenic Reagent," *FEBS Lett.* **49**, 280-281 (1974)
- 18 Loader, J., *Basic Laser Raman Spectroscopy*, Heyden and Son, Ltd., London, 7 (1970)
- 19 Savitzky, A. and Golay, M., "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.* **36**, 1627-1639 (1967)
- 20 Williams, R.W., program FITOV.FOR
- 21 Kuester, J.L. and Mize, J.H., *Optimization Techniques with FORTRAN*, McGraw-Hill, New York, 240-250 (1973)
- 22 Lawson, C.L. and Hanson, R.J., *Solving Least Squares Problems*, Prentice-

- Hall, Englewood Cliffs, New Jersey 158-173 (1974)
- 23 Morrisett, J.D. and Broomfield, C.A., "A Comparative Study of Spin-Labeled Serine Enzymes: Acetylcholinesterase, Trypsin, α -Chymotrypsin, Elastase, and Subtilisin," *J. Biol. Chem.* **247**, 7224-7231 (1972)
 - 24 Clark, D.S. and Bailey, J.E., "Deactivation of α -Chymotrypsin and α -Chymotrypsin-CNBr-Sepharose 4B Conjugates in Aliphatic Alcohols," *Biochim. Biophys. Acta* **788**, 181-188 (1984)
 - 25 Stone, T.J., Buckman, T., Nordio, P.L. and McConnell, H.M., "Spin-Labelled Biomolecules," *Proc. Nat. Acad. Sci. US* **54**, 1010-1017 (1965)
 - 26 Nordio, P.L., "General Magnetic Resonance Theory," in *Spin Labeling* (Berliner, L.J., Ed.) Academic Press, New York, 5-52 (1976)
 - 27 Kivelson, D., "Theory of ESR Linewidths of Free Radicals," *J. Chem. Phys.* **33**, 1094-1106 (1960)
 - 28 Melander, W. and Horvath, C., "Salt Effects on Hydrophobic Interactions in Precipitation and Chromatography of Proteins: An Interpretation of the Lyotropic Series," *Arch. Biochem. Biophys.* **183**, 200-215 (1977)
 - 29 Levitt, M. and Greer, J., "Automatic Identification of Secondary Structure in Globular Proteins," *J. Mol. Biol.* **114**, 181-293 (1977)
 - 30 Sawyer, L., Shotton, D.M., Campbell, J.W., Wendell, P.L., Muirhead, H. and Watson, H.C., "The Atomic Structure of Crystalline Porcine Pancreatic Elastase at 2.5Å Resolution: Comparisons with the Structure of α -Chymotrypsin," *J. Mol. Biol.* **118**, 137-208 (1978)
 - 31 Kreiger, M., Kay, L.M. and Stroud, R.M., "Structure and Specific Binding of Trypsin: Comparison of Inhibited Derivatives and a Model for Substrate Binding," *J. Mol. Biol.* **83**, 209-230 (1974)
 - 32 Chen, M.C., Lord, R.C. and Mendelsohn, R., "Laser-Excited Raman Spec-

- troscopy of Biomolecules. V. Conformational Changes Associated with the Chemical Denaturation of Lysozyme," *J. Amer. Chem. Soc.* **96**, 3038-3042 (1974)
- 33 Fersht, A.R., "Conformational Equilibria in α - and δ -Chymotrypsin, The Energetics and Importance of the Salt Bridge," *J. Mol. Biol.* **64**, 497-509 (1972)
- 34 Swartz, H.M., Bolton, J.R. and Borg, D.C. in *Biological Applications of Electron Spin Resonance*, Wiley-Interscience, New York, 23 (1972)
- 35 Hemminga, M.A., "Qualitative Explanation of Saturation Transfer ESR Effects Using the Concept of Hole Burning," *J. Mag. Res.* **33**, 215-218 (1979)
- 36 Robinson, B., Thomann, H., Beth, A., Fajer, P. and Dalton, L., "The Phenomenon of Magnetic Resonance: Theoretical Considerations," in *EPR and Advanced EPR Studies of Biological Systems* (Dalton, L.R., Ed.) CRC Press, Boca Raton, Florida, 11-110 (1985)
- 37 BIOGRAF/III version 1.40: BIOGRAF was developed at the California Institute of Technology by S.L. Mayo, B.D. Olafson and W.A. Goddard III and is supported by BioDesign, Inc., 199 So. Los Robles Avenue, Suite 270, Pasadena, CA 91101.
- 38 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., "The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures," *J. Mol. Biol.* **112**, 535-542 (1977)
- 39 Chou, P.Y. and Fasman, G.D., "Prediction of Protein Conformation," *Biochemistry* **13**, 222-245 (1974)
- 40 IUPAC-IUB Commission on Biochemical Nomenclature, "Abbreviations

- and Symbols for the Description of the Conformation of Polypeptide Chains. Tentative Rules (1969)," *Biochemistry* **9**, 3471-3479 (1970)
- 41 Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions*, Dover, New York, 991-995 (1972)
- 42 Dunker, A.K., Fodor, S.P.A. and Williams, R.W., "Lipid-Dependent Structural Changes of an Amphomorphic Membrane Protein," *Biophys. J.* **37**, 201-203 (1982)
- 43 Stoesz, J.D. and Lumry, R.W., "Refolding Transition of α -Chymotrypsin: pH and Salt Dependence," *Biochemistry* **17**, 3693-3699 (1978)
- 44 Royer, G., Wildnauer, R., Cuppett, C.C. and Canady, W.J., "Kinetic and Thermodynamic Evidence for the Existence of at Least Two Salt-dependent Forms of α -Chymotrypsin at Slightly Acid pH," *J. Biol. Chem.* **246**, 1129-1134 (1971)
- 45 Cuppett, C.C., Resnick, H. and Canady, W.J., "Physical Evidence for the Existence of at Least Two Salt-dependent Forms of α -Chymotrypsin at Slightly Acid pH," *J. Biol. Chem.* **246**, 1135-1141 (1971)
- 46 Nemoto, N., Kiyohara, T., Tsunashima, Y. and Kurata, M., "Raman, IR and CD Spectroscopy of Aggregates and Gels of Lysozyme in Alcohols," *Bull. Inst. Chem. Res. Kyoto Univ.* **61**, 203-213 (1983)
- 47 Clark, A.H., Saunderson, D.H.P. and Suggett, A., "Infrared and Laser-Raman Spectroscopic Studies of Thermally-Induced Globular Protein Gels," *Int. J. Peptide Res.* **17**, 353-364 (1981)
- 48 Berjot, M., Marx, J. and Alix, A.J.P., "Determination of the Secondary Structure of proteins from the Raman Amide I Band: The Reference Intensity Profiles Method," *J. Raman Spectrosc.* **18**, 289-300 (1987)
- 49 Aune, K.C. and Timasheff, S.N., "Dimerization of α -Chymotrypsin. I. pH

- Dependence in the Acid Region," *Biochemistry* **10**, 1609-1617 (1971)
- 50 Aune, K.C., Goldsmith, L.C. and Timasheff, S.N., "Dimerization of α -Chymotrypsin. II. Ionic Strength and Temperature Dependence," *Biochemistry* **10**, 1617-1622 (1971)
- 51 Ikeda, K., Kunugi, S. and Ise, N., "Specific and Non-Specific Effects of Inert Anions and Cations on the Dimerization of α -Chymotrypsin and the Catalytic Activities of Monomeric and Dimeric α -Chymotrypsins," *J. Biochem.* **91**, 347-355 (1982)
- 52 Tellam, R. and Winzor, D.J., "Self-Association of α -Chymotrypsin at Low Ionic Strength in the Vicinity of its pH Optimum," *Biochem. J.* **161**, 687-694 (1977)
- 53 Schubert, P.F. and Finn, R.K., "Alcohol Precipitation of Proteins: The Relationship of Denaturation and Precipitation for Catalase," *Biotechnol. Bioeng.* **23**, 2569-2590 (1981)

3.9 Tables

Table 1 Salt Cuts and α CT Solubility Ranges^a

| salt cut | α CT solubility change (mg/mL) | percent α CT precipitated |
|----------------|---|-------------------------------------|
| 1 | 10 \rightarrow 8 | 20 |
| 2 | 8 \rightarrow 6 | 40 |
| 3 | 6 \rightarrow 4 | 60 |
| 4 | 4 \rightarrow 2 | 80 |
| 5 ^b | 2 \rightarrow 0.5 | 95 |

a The salt cuts taken were the same as in [2].

b The last cut taken depends on salt saturation and salting-out constant values.

Table 2. Salt Molal Surface Tension Increments^a

| Salt | σ ($\times 10^3$ dyn \cdot g/cm \cdot mol) |
|---------------------------------|---|
| KSCN | 0.45 |
| KBr | 1.31 |
| NaBr | 1.32 |
| NaCl | 1.64 |
| Na ₂ SO ₄ | 2.73 |

a Obtained from Melander and Horvath [28].

Table 3. RSAP Control Analyses

| Sample ^a | Analysis ^b | Types of secondary structure ^c | | | | | | | | | RNORM ^d |
|---------------------|-----------------------|---|----------------|----------------|----------------|----|----|----------------|----------------|----|--------------------|
| | | H _o | H _d | S _a | S _p | T | R | H _T | S _T | U | |
| min | RSAP | 0 | 2 | 0 | 0 | 8 | 7 | 2 | 0 | 16 | |
| max | RSAP | 62 | 29 | 65 | 24 | 22 | 17 | 85 | 65 | 34 | |
| LZM | RSAP | 8 | 23 | 38 | 0 | 21 | 10 | 31 | 38 | 31 | 0.038 |
| | x-ray ^e | 24 | 22 | 19 | 0 | 22 | 13 | 46 | 19 | 35 | |
| ELS | RSAP | 0 | 10 | 57 | 0 | 28 | 5 | 10 | 57 | 33 | 0.030 |
| | x-ray ^f | 6 | 9 | 46 | 0 | 26 | 13 | 15 | 46 | 39 | |
| TPN | RSAP | 0 | 11 | 58 | 0 | 20 | 11 | 11 | 58 | 31 | 0.037 |
| | x-ray ^g | 4 | 9 | 55 | 0 | 24 | 8 | 13 | 55 | 32 | |
| α CT | RSAP ^h | 0 | 13 | 58 | 0 | 15 | 16 | 13 | 58 | 31 | 0.040 |
| | x-ray ⁱ | 2 | 12 | 50 | 0 | 17 | 19 | 14 | 50 | 36 | |
| lyo α CT | RSAP ^h | 0 | 15 | 56 | 0 | 17 | 14 | 15 | 56 | 31 | 0.034 |
| +1cm ⁻¹ | RSAP ^h | 0 | 13 | 60 | 0 | 12 | 17 | 13 | 60 | 29 | 0.045 |
| -1cm ⁻¹ | RSAP | 0 | 14 | 55 | 0 | 16 | 15 | 14 | 55 | 31 | 0.037 |

a Sample abbreviations: min/max, bounds of structure space spanned by reference protein set; LZM, lysozyme; ELS, elastase; TPN, trypsin; lyo α CT, lyophilized α CT; +1 cm⁻¹ and -1 cm⁻¹, shifted α CT spectra.

b X-ray analysis refers to structure assignments based on the crystal structure and the method of Levitt and Greer [29].

c Structure types: H_o, ordered helix; H_d, disordered helix; S_a, antiparallel β -sheet; S_p, parallel β -sheet; T, reverse turn; R, random coil; H_T, total helix; S_T, total sheet; U, undefined structure (U = T + R); all expressed as a percent of total structure.

d RNORM is the Euclidean norm of the residual vector as defined in equation (9).

e Structure assignment from Williams [11].

f Crystal structure from Sawyer et al. [30].

g Crystal structure from Kreiger et al. [31].

h Rounding error gives sum of fractions exceeding 100% .

i Crystal structure from Birktoft and Blow [3].

Table 4. RSAP Analysis of α CT Precipitate Samples^a

| Sample | Cut | Types of secondary structure ^c | | | | | | | | | RNORM |
|---------------------------------|----------------|---|----------------|----------------|----------------|----|----|----------------|----------------|----|-------|
| | | H _o | H _d | S _a | S _p | T | R | H _T | S _T | U | |
| Na ₂ SO ₄ | 1 | 0 | 10 | 58 | 0 | 22 | 10 | 10 | 58 | 32 | 0.033 |
| NaCl | 1 | 0 | 9 | 61 | 0 | 20 | 10 | 9 | 61 | 30 | 0.039 |
| | 2 | 0 | 10 | 59 | 0 | 19 | 12 | 10 | 59 | 31 | 0.034 |
| | 3 ^b | 0 | 16 | 56 | 0 | 17 | 14 | 16 | 56 | 31 | 0.033 |
| | 4 ^b | 0 | 16 | 58 | 0 | 18 | 12 | 16 | 58 | 30 | 0.039 |
| | avg | 0 | 13 | 59 | 0 | 19 | 12 | 13 | 59 | 31 | |
| | | | ±4 | ±2 | | ±1 | ±2 | ±4 | ±2 | ±1 | |
| NaBr | 1 | 0 | 8 | 65 | 0 | 10 | 17 | 8 | 65 | 27 | 0.052 |
| NaBr (ins) | 4 | 0 | 5 | 56 | 10 | 19 | 10 | 5 | 66 | 29 | 0.036 |
| | 5 | 0 | 10 | 56 | 7 | 18 | 9 | 10 | 63 | 27 | 0.036 |
| | avg | 0 | 8 | 56 | 9 | 19 | 10 | 8 | 65 | 28 | |
| | | | ±4 | ±0 | ±2 | ±1 | ±1 | ±4 | ±2 | ±1 | |
| KBr | 1 | 0 | 4 | 68 | 0 | 17 | 11 | 4 | 68 | 28 | 0.056 |
| | 2 | 0 | 0 | 48 | 22 | 17 | 13 | 0 | 69 | 31 | 0.084 |
| | 3 | 0 | 9 | 64 | 0 | 12 | 15 | 9 | 64 | 27 | 0.048 |
| | 4 | 3 | 4 | 66 | 0 | 13 | 14 | 7 | 66 | 27 | 0.044 |
| | avg | 1 | 4 | 62 | 6 | 15 | 13 | 5 | 67 | 28 | |
| | | ±2 | ±4 | ±9 | ±11 | ±3 | ±2 | ±4 | ±2 | ±2 | |
| KBr (ins) | 1 | 0 | 6 | 60 | 6 | 18 | 10 | 6 | 66 | 28 | 0.034 |
| | 2 | 0 | 6 | 61 | 4 | 18 | 11 | 6 | 65 | 29 | 0.042 |
| | 3 | 0 | 6 | 70 | 0 | 15 | 9 | 6 | 70 | 24 | 0.040 |
| | 4 | 0 | 0 | 70 | 4 | 14 | 12 | 0 | 74 | 26 | 0.042 |
| | avg | 0 | 5 | 65 | 4 | 16 | 11 | 5 | 69 | 27 | |
| | | ±3 | ±6 | ±3 | ±2 | ±1 | ±3 | ±4 | ±2 | | |
| KSCN | 1 | 0 | 10 | 60 | 5 | 15 | 10 | 10 | 65 | 25 | 0.033 |
| | 2 | 0 | 4 | 60 | 9 | 22 | 5 | 4 | 69 | 27 | 0.028 |
| | 3 | 0 | 6 | 60 | 6 | 20 | 8 | 6 | 66 | 28 | 0.031 |
| | 4 | 0 | 5 | 61 | 7 | 19 | 8 | 5 | 68 | 27 | 0.033 |
| | 5 | 0 | 5 | 55 | 11 | 21 | 8 | 5 | 66 | 29 | 0.037 |
| | avg | 0 | 6 | 59 | 8 | 19 | 8 | 6 | 67 | 27 | |
| | | ±2 | ±2 | ±2 | ±3 | ±2 | ±2 | ±2 | ±2 | ±2 | |
| KSCN (ins) | 1 | 0 | 0 | 74 | 1 | 18 | 7 | 0 | 75 | 25 | 0.034 |
| | 2 | 0 | 3 | 54 | 22 | 17 | 4 | 3 | 76 | 21 | 0.035 |
| | 3 | 0 | 3 | 71 | 0 | 16 | 10 | 3 | 71 | 26 | 0.039 |
| | 4 ^b | 0 | 5 | 72 | 0 | 14 | 11 | 5 | 72 | 22 | 0.036 |
| | 5 | 0 | 8 | 60 | 10 | 14 | 8 | 8 | 70 | 22 | 0.033 |
| | avg | 0 | 4 | 66 | 7 | 16 | 8 | 4 | 73 | 24 | |
| | | ±3 | ±9 | ±7 | ±2 | ±3 | ±3 | ±3 | ±3 | ±2 | |

a Abbreviations are as defined in Table 3.

b Rounding error gives sum of structure fractions exceeding 100% .

Table 5. Calculated Physical Properties for Native and Altered α CT Structures

| Structure | hydrophobic surface area ^a (\AA^2) | Dipole moment ^b (Debye) | Salting-out constant ^c (m^{-1}) |
|-----------------------|--|--|---|
| native | 2470 | 223 | 2.33 |
| modified ^d | 3100 | 174 | 3.11 |

a Calculated via BIOGRAF [37].

b Computed as in [2].

c The experimental value is 6.83 m^{-1} [2].

d The model for KSCN whole precipitate was used as the modified structure.

3.10 Figures

Figure 1 Raman amide I band spectrum of native α CT. The solid line is the original spectrum. The dotted lines represent the individual fitted peaks for subtraction and the sum of the fitted peaks described in equation (1).

Figure 2 Chemical structure of 1-oxyl-2,2,6,6-tetramethyl-4-piperidinyl ethoxyphosphorofluoridate (TEMPO-4-EPF) shown attached to the hydroxyl moiety of Ser₁₉₅ at the α CT active site.

Figure 3 Solid phase Raman amide I band control spectra; (a) lysozyme, (b) trypsin and (c) elastase. Both the raw (solid line) and subtracted (dotted line) spectra are shown.

Figure 4 Subtracted Raman amide I band spectra of first cut α CT precipitate samples.

Figure 5 The total α -helix (H_T) and total β -sheet (S_T) structure content fractions of solid α CT precipitate samples as a function of the molal surface tension increment of the inducing salt (see Table 4). Open and filled symbols represent whole precipitate samples and the corresponding insoluble portions respectively. Dashed lines give the native protein structure fits (see Table 3). The salts in order of increasing σ -values are: KSCN, KBr, NaBr, NaCl and Na₂SO₄. Error bars represent \pm S.D. of the mean; error bars for NaBr and Na₂SO₄ whole precipitates were estimated to be ± 0.03 structure % .

Figure 6 Correlation of whole precipitate active fraction with the change in total β -sheet content for the salts used. Quantities are normalized by native α CT values. The active fraction of native α CT after incubation at pH 3.00 was 0.35 ± 0.02 [2]. Symbols: \star , KSCN; \blacksquare , KBr; \bullet , NaBr; \blacklozenge , NaCl; \blacktriangle , Na_2SO_4 . Error bars denote \pm S.D. of the mean.

Figure 7 Conventional EPR spectra of free TEMPO-4-OH (top) and SL- α CT (bottom) in solution. Each spectrum represents a scan of 100 G.

Figure 8 Conventional EPR spectra of SL- α CT, and KSCN and Na_2SO_4 induced SL- α CT precipitate. All samples were solid phase. $2A_{\parallel}$ denotes the separation of the outer hyperfine resonances. Each spectrum represents a scan of 100 G.

Figure 9 ST-EPR spectra of SL- α CT, and KSCN and Na_2SO_4 induced SL- α CT precipitate. Spectra were manually digitized and appear smoother than the original data. Each spectrum represents a scan of 100 G.

Figure 10 C^α backbone of native α CT. Dashed lines represent segments of ordered secondary structure, i.e. α -helix and β -sheet. Reverse turns and unordered segments are shown as dotted lines. The N and C termini are labelled as is every tenth residue. Note that two orthogonal β -barrels each composed of five strands of antiparallel β -sheet form the dominant ordered structural motif. In this view,

the axis of the β -barrel on the left half of the diagram runs from top to bottom and that of the β -barrel on the right is perpendicular to the plane of the page. Also visible are the two helical segments; the C-terminal helix is located at the bottom of the figure and an internal helical segment from Asn₁₆₇ to Ile₁₇₆ [5] is shown on the lower left side. The active site is located at the back side of the diagram.

Figure 11 C $^{\alpha}$ backbone of modified α CT. The solid lines represent segments of the backbone that have been altered to mimic the secondary structure content of the KSCN-induced whole precipitate. The labels and dashed and dotted lines correspond to those in Figure 10. Three strands of β -sheet have been added and the C-terminal helix has been disrupted. Note that this orientation is rotated a few degrees towards the top of the page with respect to Figure 10.

Figure 12 Proposed precipitation mechanism. \mathcal{N} , \mathcal{RD} and \mathcal{ID} represent native, reversibly denatured and irreversibly denatured solution conformations respectively. A subscript p denotes the corresponding precipitate phase. The relative magnitudes of the associated rate and equilibrium constants as a function of the solution conditions are also given.

Figure 1 Raman amide I band spectrum of native α CT

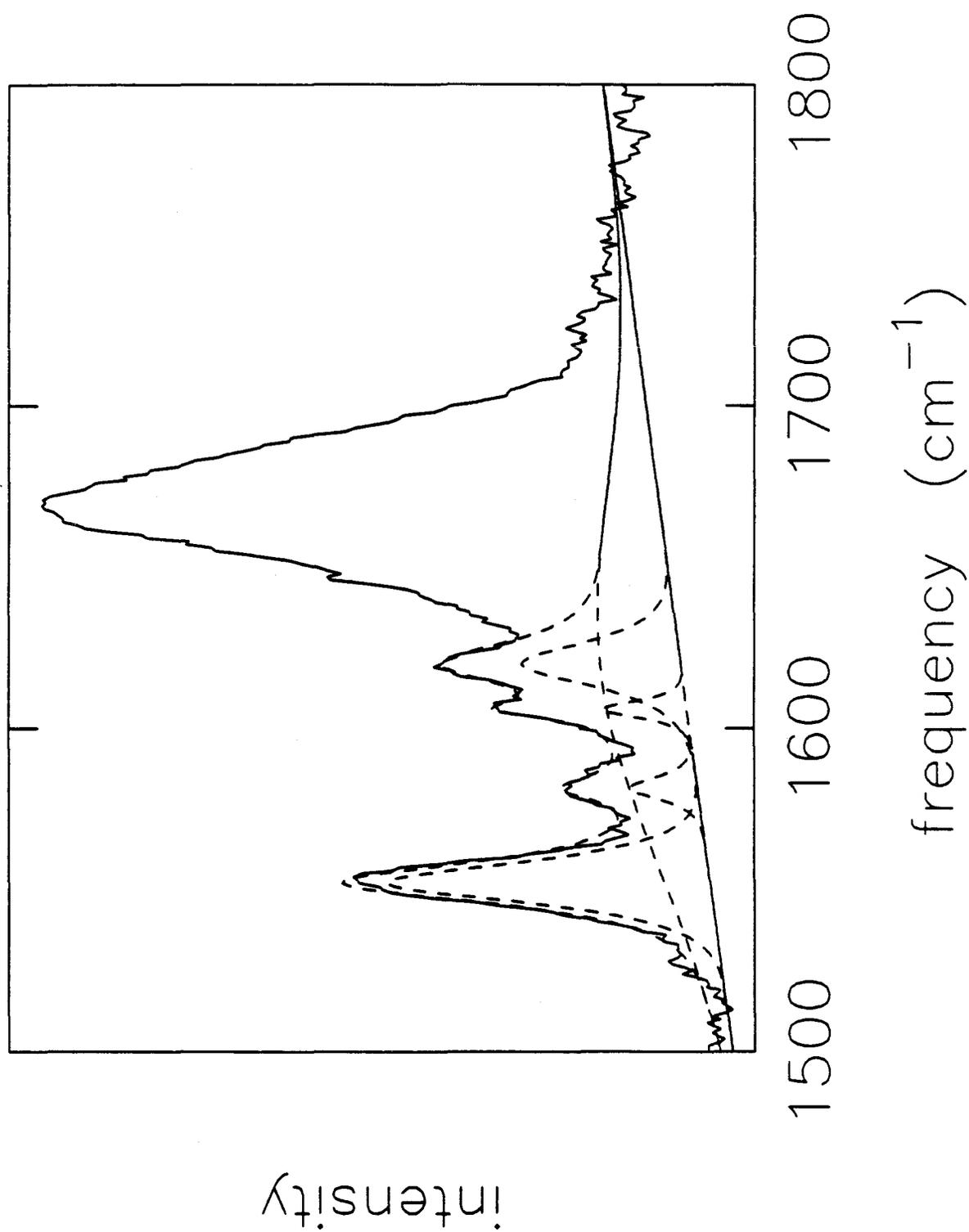


Figure 2 Chemical structure of active site spin label

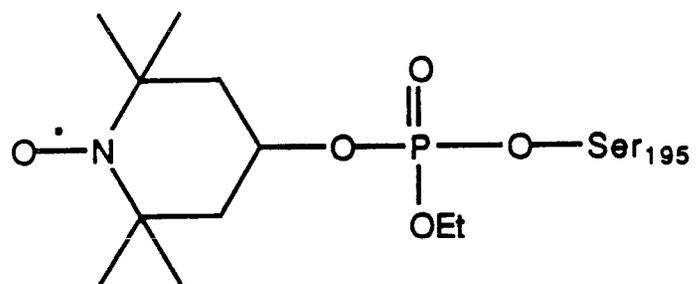


Figure 3 Raman amide I band spectra of control proteins

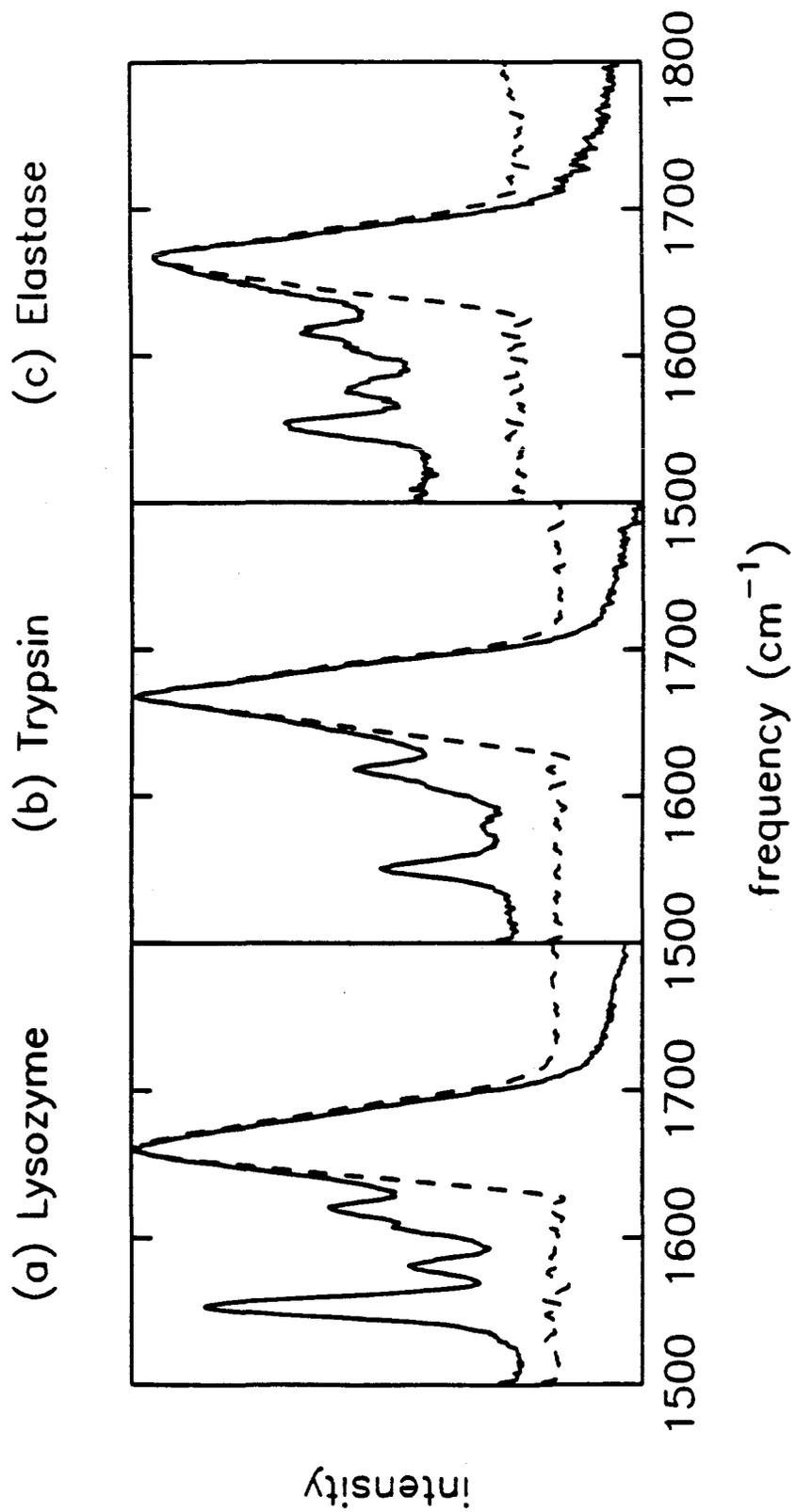


Figure 4 Subtracted Raman amide I band spectra of precipitate samples

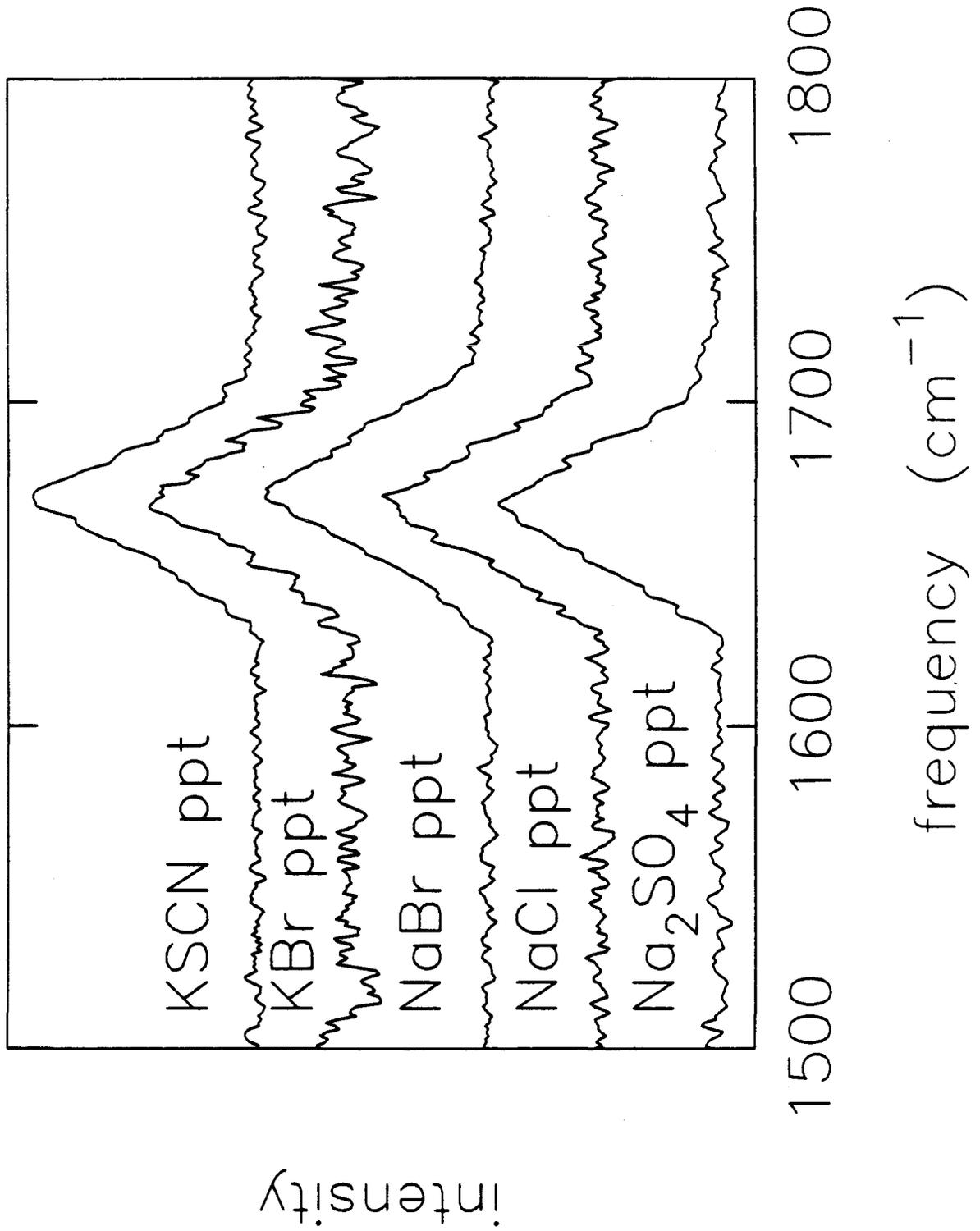


Figure 5 α -helix and β -sheet contents of precipitate samples as a function of salt type

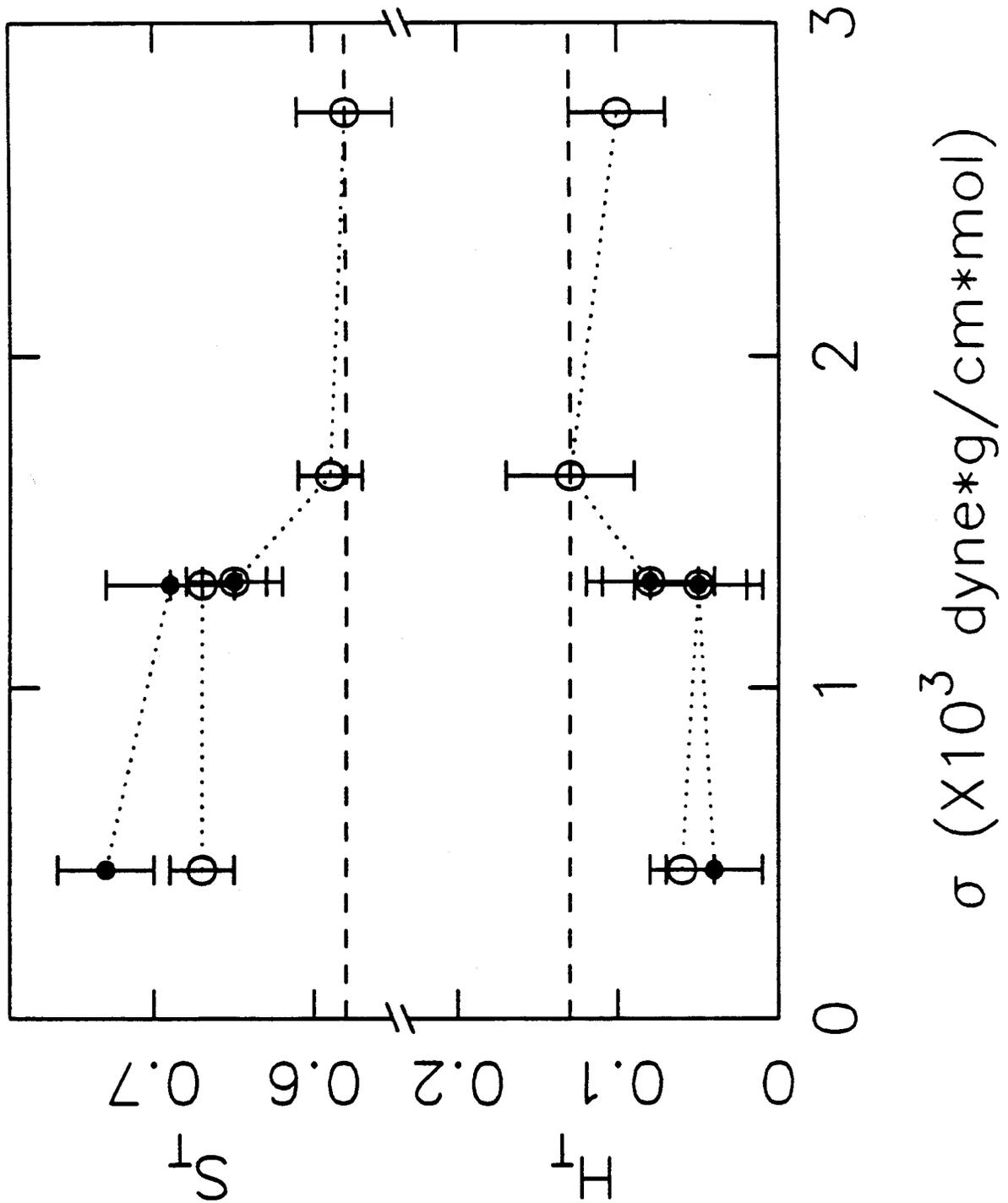


Figure 6 Correlation of whole precipitate active fraction with change in β -sheet

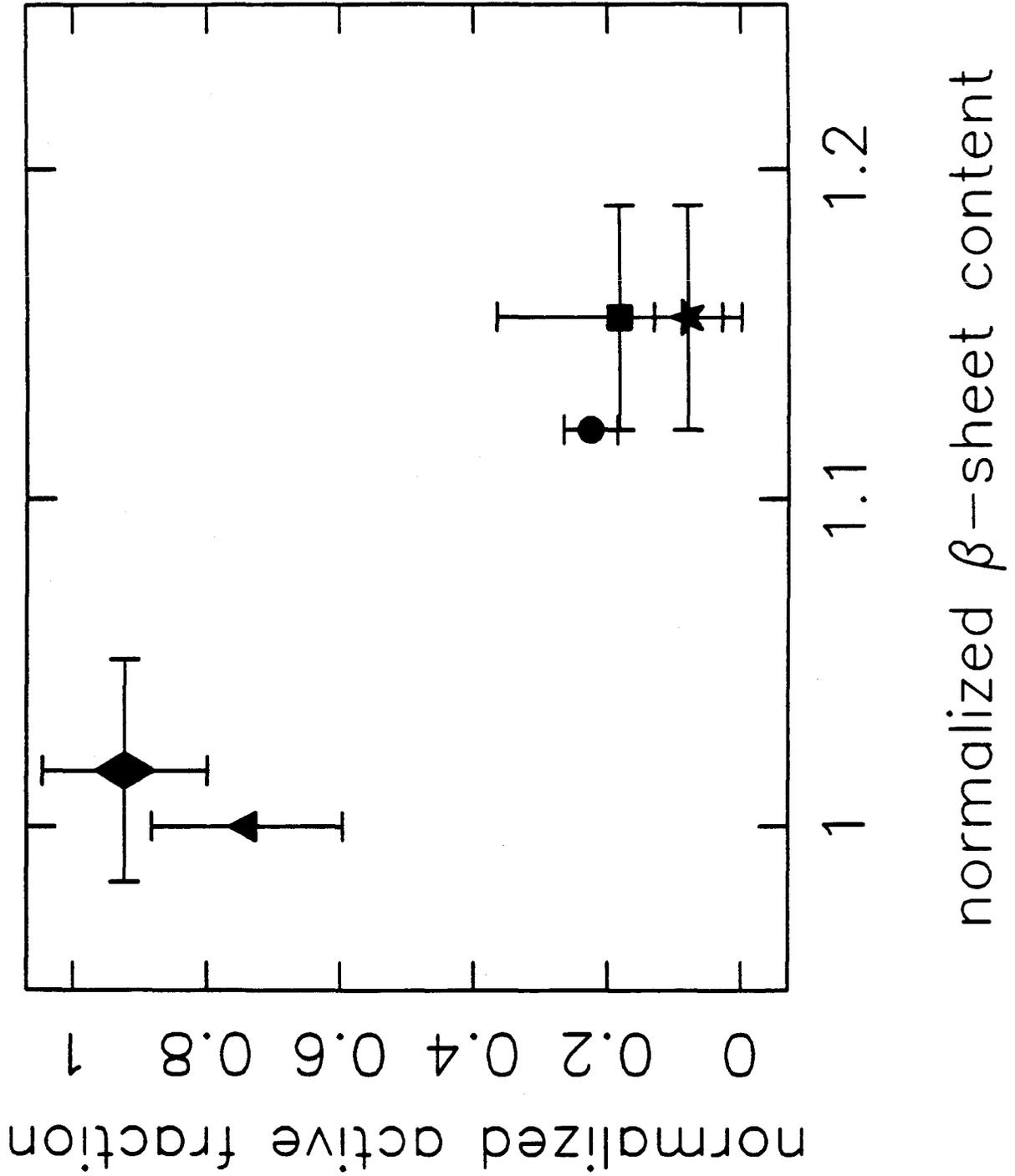


Figure 7 EPR spectra of free spin label and spin-labelled α CT

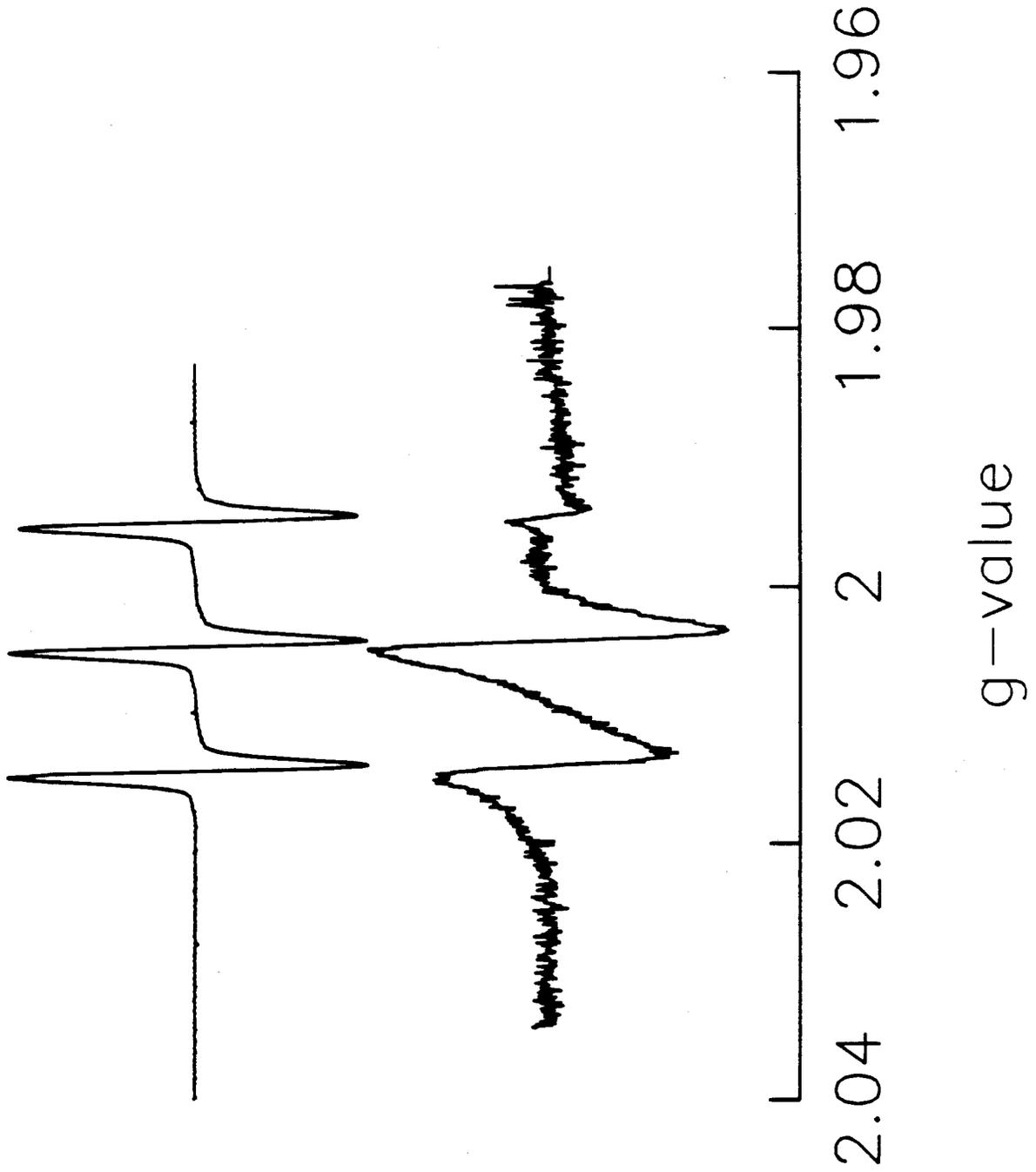


Figure 8 EPR spectra of spin-labelled α CT precipitates

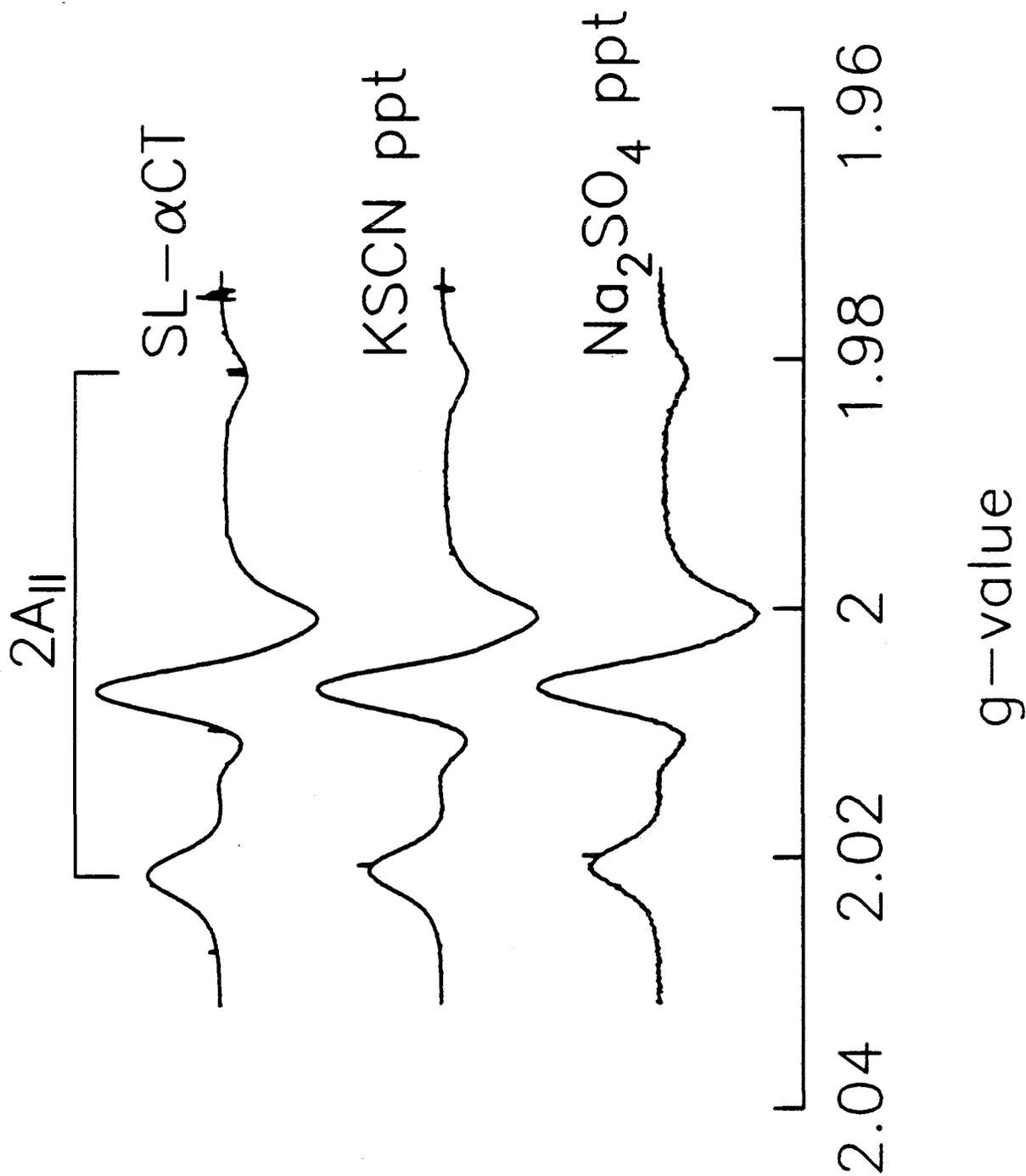


Figure 9 ST-EPR spectra of spin-labelled α CT precipitates

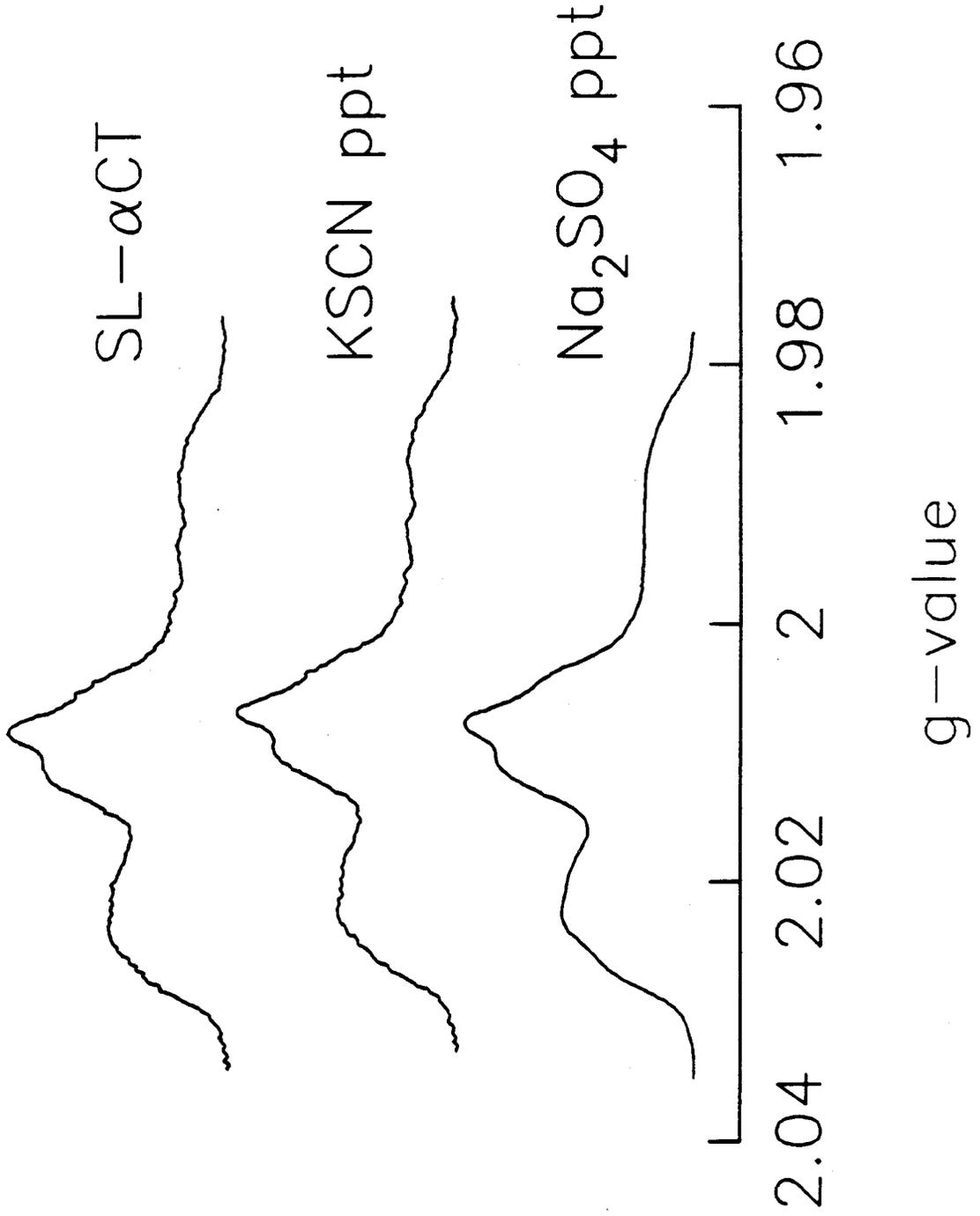


Figure 10 C^α backbone of native αCT

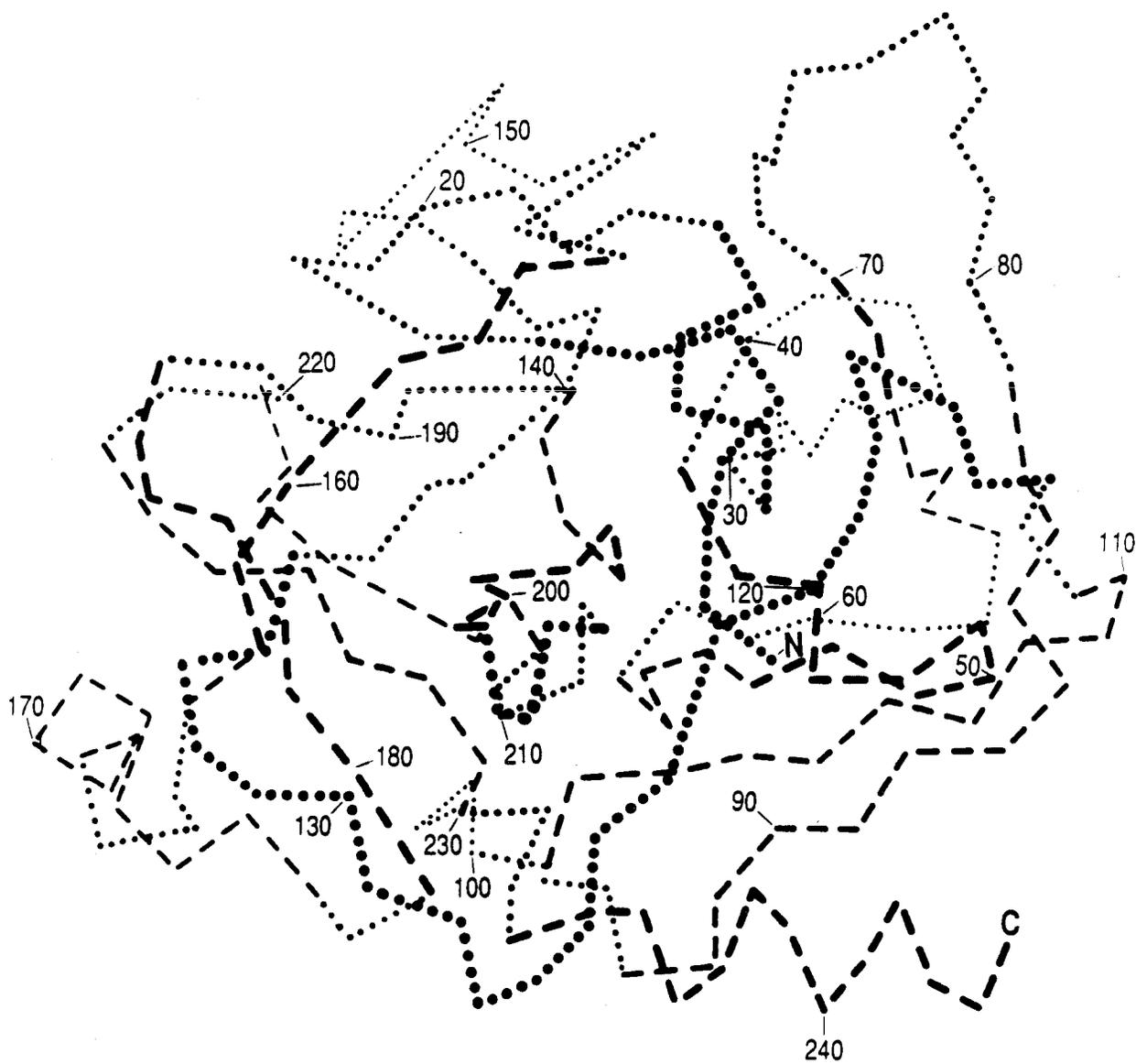


Figure 11 C α backbone of model KSCN-induced α CT precipitate

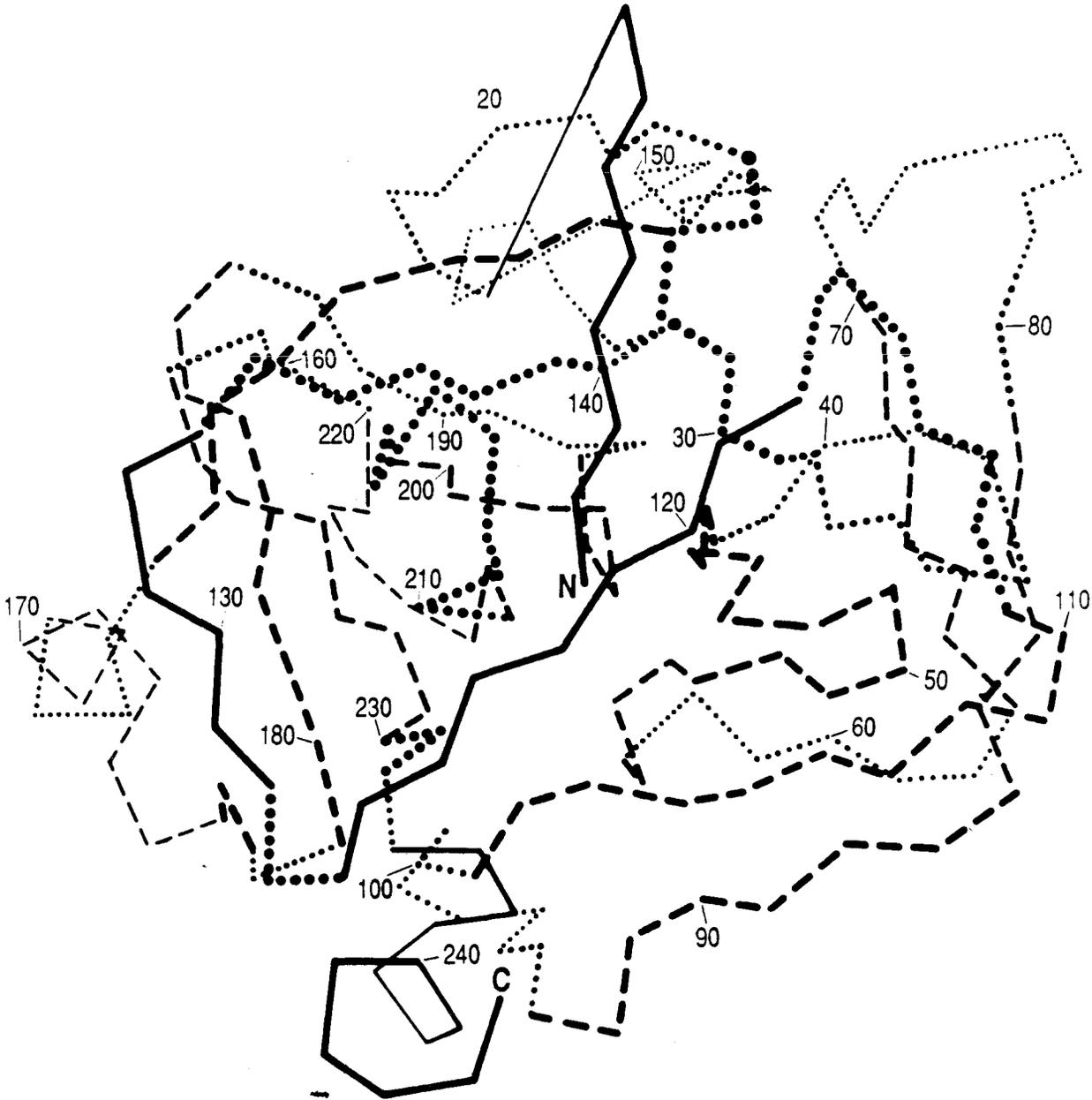
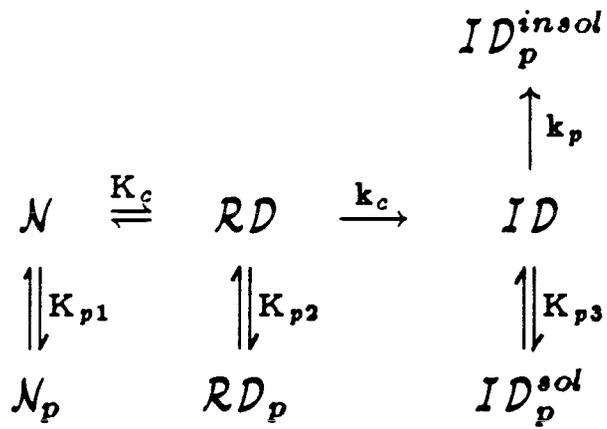


Figure 12 Proposed precipitation mechanism



| conditions | salt used | K_c | k_c | K_{pi} | k_p |
|---------------|----------------|---------|----------|----------|----------|
| precipitation | large σ | > 1 | ~ 0 | > 1 | ~ 0 |
| | small σ | $\gg 1$ | > 0 | $\gg 1$ | > 0 |
| redissolution | none | < 1 | 0 | $\ll 1$ | 0 |

Chapter 4

Secondary Structure Perturbations In Salt-Induced Protein Precipitates

The material contained in this chapter has been prepared
for submission to the *Journal of Biological Chemistry*

4.1 Abstract

The secondary structure implications of precipitation induced by a chaotropic salt, KSCN, and a structure stabilizing salt, Na₂SO₄, were studied for twelve different proteins. α -helix and β -sheet content of precipitate and native structures were estimated from the analysis of amide I band Raman spectra. A statistical analysis of the estimated perturbations in the secondary structure contents indicated that the most significant event is the formation of β -sheet structures with a concomitant loss of α -helix on precipitation with KSCN. The effects of elements of primary, secondary, and tertiary structure existing in the native protein on the conformational changes for each protein were also analyzed. In the native protein, primary structure was quantified by the fractions of hydrophobic and charged amino acids, secondary structure by x-ray estimates of α -helix and β -sheet contents, and tertiary structure by the dipole moment and solvent-accessible surface area. For the KSCN precipitates, native protein attributes affecting β -sheet content in the corresponding precipitates included the fraction of charged amino acids in the primary sequence and the surface area. Changes in α -helix content in the precipitates were influenced by the initial helical content and the dipole moment of the native proteins. The enhanced β -sheet contents of precipitates found in this work parallel protein structural changes occurring in other aggregative phenomena.

4.2 Introduction

Salt-induced protein precipitation is a widely used primary isolation technique in the bioprocess industry. However, high salt environments are potentially denaturing and may diminish active protein yields. Perturbations in protein structure occurring during precipitation may also be representative of those accompanying aggregative phenomena in general; yet, it is not certain whether these perturbations are the cause or the effect of self-association events. In order to design precipitation operations for active protein recovery rationally, an understanding of the structural consequences, at the molecular level, of variations in process parameters is crucial.

In this work, the structural implications of precipitation have been investigated for a variety of proteins to determine if correlations exist with elements of native protein structure. A set of twelve proteins including α -chymotrypsin, trypsin, alcohol dehydrogenase, subtilisin, superoxide dismutase, immunoglobulin G, hexokinase, adenylate kinase, lysozyme, concanavalin A, ribonuclease A, and cytochrome C were studied. This set of proteins contains members of the serine protease, dehydrogenase, kinase, and immunoglobulin homologous structure groups. Conformational changes induced by precipitation were analyzed for correlations with elements of primary, secondary, and tertiary structure. The relative importance of the different levels of organization on protein behavior during precipitations was assessed on this basis. The salts used include Na_2SO_4 , a structure preserving salt, and KSCN, a chaotropic salt.

Previous work with α -chymotrypsin indicated that as the inducing salt becomes more denaturing, the β -sheet contents of the associated precipitates increased relative to the native contents [1]. Other investigators have reported β -sheet augmentation in a number of different aggregating protein and peptide

systems [2-12]. The generality of this phenomena is explored in this study.

Secondary structure content was estimated from amide I band Raman spectra. The fractions of ordered and disordered α -helix, parallel and antiparallel β -sheet, β -reverse turn, and random coil structures were assigned via an analysis routine [1] based on the algorithms and reference data of Williams [13]. Estimates of the secondary structure contents of precipitates were considered relative to the estimated values for the corresponding lyophilized protein samples.

4.3 Materials and Methods

4.3.1 Materials

The proteins used in this study were obtained from Sigma. Table 1 lists the proteins, grouped according to common structural motifs, with the corresponding Enzyme Commission number (where appropriate), acronym, source and nature of preparation. All the proteins, except ribonuclease A and cytochrome C, were obtained as lyophilized preparations. Ribonuclease A and cytochrome C were lyophilized from 10 mg mL⁻¹ solutions before use. Na₂SO₄, KSCN, and all buffer solution components were analytical reagent grade. Distilled, deionized water processed by a Corning Mega-Pure system was used throughout.

4.3.2 Preparation of Precipitate Samples

Precipitates were generated at 25°C from 10 mg mL⁻¹ protein stock solutions in 50 mM glycine buffer adjusted to pH 3.00 with aq. HCl. Sufficient salt was added to 1 mL portions of stock solution to bring concentrations to 2.0 M and 1.8 M for KSCN and Na₂SO₄, respectively. Details of the precipitate work-up protocol are described elsewhere [14]. Supernatant was withdrawn and discarded following centrifugation and the precipitate pellets were

lyophilized. Ribonuclease A was not precipitated by KSCN under these conditions; the KSCN-ribonuclease A solution was lyophilized. The lyophilized pellets were broken up and packed into Kimax melting point tubes for spectroscopic analysis.

4.3.3 Raman Spectroscopy

Spectroscopic work was performed in a similar fashion to Przybycien and Bailey [1]. All spectra were obtained with a SPEX 1401 double monochromator with an 1800 line/mm grating and a Products For Research cooled photomultiplier tube. A Coherent Innova 90 Ar⁺ laser served as the excitation source. Data acquisition and control was performed with a SPEX DM3000R AT-compatible microcomputer running the DM3000 software package.

Precipitate spectra were collected in the backscattering mode; samples were maintained at 10°C via a thermostatted sample stage. The amide I band was scanned from 1500 to 1800 cm⁻¹ at intervals of 1 cm⁻¹. The integration time was 2 s/cm⁻¹. Entrance and exit slits were set at 450 μm and the stray light slit at 900 μm for a resolution of 6 cm⁻¹ at 20492 cm⁻¹. The 488 nm line of the Ar⁺ laser was used with powers ranging from 25 to 100 mW at the sample. Reasonable signal-to-noise, S/N, ratios were obtained within 3 to 20 scans. Obtaining signals for some of the samples was extremely difficult; since spectra are collected via surface scattering, the precipitate morphology is an important factor in determining signal intensity. Samples spent a maximum of four hours in the beam and were visually inspected for damage after use.

The spectral data were analyzed with the Raman Spectral Analysis Package (RSAP) described elsewhere [1]; RSAP implements the smoothing, subtraction, normalization, and structure estimation algorithms described by Williams [13].

RSAP provides estimates of the ordered and disordered α -helix, parallel and antiparallel β -sheet, β -reverse turn, and random coil contents via a constrained superposition of reference data from proteins with known structures. There was no evidence of glycine peaks in the protein spectra.

4.4 Results

The RSAP secondary structure estimates of KSCN and Na_2SO_4 -induced precipitates and the corresponding native proteins are given in Table 2. It was not possible to obtain a Raman signal for Na_2SO_4 precipitates of every sample. In view of recent work concerning the spectral similarities of parallel and antiparallel β -sheet[15,16], the assignment of the six individual structure types may be somewhat artificial. The ordered and disordered α -helix, parallel and antiparallel β -sheet, and the β -reverse turn and random coil estimates have been combined into total helical, H_T , total sheet, S_T , and undefined, U , fractions, respectively. The x-ray structure assignments were taken from Williams [13] and Levitt and Greer [17] and are listed in Table 2. The total structure class fractions for the native proteins agree more closely with the x-ray assignments than the individual structure estimates and are used for comparative purposes below. The RNORM values in Table 2 values represent the Euclidean norm of the residual fitted structure vectors and give an indication of the ability of the reference protein set to describe the experimental structure [13,14]. Some spectra were of marginal quality, as reflected in the low S/N ratios; these spectra typically had extremely high fluorescence background levels and in most cases corresponded with precipitate samples.

Although the native protein S_T values tend to overestimate the β -sheet contents determined from the x-ray structures, almost all of the precipitate samples

had larger S_T values. α -helix contents in the native proteins were generally underestimated and corresponding precipitate H_T estimates were typically smaller. It may be inferred that β -sheet formation consumes segments of α -helix. The relationship between precipitation-induced conformational perturbations and the initial structure content of the protein is not clear; changes in proteins with similar secondary structures such as α -chymotrypsin and trypsin, subtilisin and alcohol dehydrogenase, and superoxide dismutase and immunoglobulin G were generally in the same direction, but of different magnitudes (see Table 4).

A series of statistical analyses were performed on the RSAP estimates to determine the significance of the results in the face of variations in spectral quality and fitting ability. The estimated conformational changes on precipitation were also compared with elements of primary, secondary, and tertiary structure existing in the native state in an effort to determine what protein properties control salting-out behavior. These analyses rely on parametric significance tests for small samples based on Student's t -distribution [18].

To assess whether or not the structure fits were correlated with spectral quality or the structure space spanned by the reference protein set, the hypothesis that the H_T and S_T values for all the samples were uncorrelated with the S/N and RNORM values, $\chi_o(\rho_{XY} = 0)$, where ρ_{XY} is the correlation coefficient of the populations X and Y , was tested. The results of the test are given in Table 3. The subdiagonal entries in the matrix are the computed sample correlation coefficients, r_{XY} , between the row and column variables, X and Y , with the number of samples, n , noted in parentheses below. r_{XY} is computed via [18]

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where x_i and y_i are the individual variable values and \bar{x} and \bar{y} are the sample

means of the variable values. For $\rho_{XY} = 0$, the statistic t_o , given by [18]

$$t_o = \frac{r_{XY}}{\sqrt{1 - r_{XY}^2}} \sqrt{n - 2}, \quad (2)$$

has the Student's t -distribution with $n - 2$ degrees of freedom. The superdiagonal entries in Table 3 are the corresponding probabilities, $P(|t| > t_o)$, of committing an error of the first kind by rejecting the assumption \mathcal{H}_o that the variables are uncorrelated; these values are conservative estimates from tabulated values of $P(|t| \geq t)$ for the Student's t -distribution [18].

\mathcal{H}_o cannot be rejected for the total helix fits H_T and the spectral S/N ratio or the total sheet fits S_T and the S/N ratio. The spectral quality does not appear to affect the outcome of the RSAP analysis; although the acceptance of \mathcal{H}_o does not necessarily imply that \mathcal{H}_o is correct, the small values of the corresponding sample correlation coefficients lend considerable support to this assumption. At the 1% significance level, \mathcal{H}_o can be rejected for the variation of both structure fits with RNORM and with each other, implying that these quantities are correlated. The structure estimates are somewhat limited by the span of the reference set. H_T and S_T are strongly inversely correlated, samples with large S_T estimates tend to have small H_T estimates.

The next logical step is to determine if the structure fits are adequate. The significance of the deviations of the fitted structure values for the lyophilized samples from the x-ray values was estimated by testing the hypotheses $\mathcal{H}_o(\bar{D} = 0)$ where \bar{D} is the mean deviation. At the 5% significance level, \mathcal{H}_o can be rejected for both the α -helix and the β -sheet fits. Thus, there is a significant deviation between the RSAP estimates and the x-ray values. This is not surprising recalling the correlation found between the structure estimates and RNORM. To validate the RSAP results, the consistency of this fitting error was examined;

the hypothesis $\chi_o(d_i = \bar{D})$ where d_i is the deviation of an individual lyophilized sample structure estimate from the x-ray value was tested. χ_o cannot be rejected at the 60% significance level for H_T and at greater than the 80% level for S_T , implying that the deviation is consistent. Assuming χ_o , the structural changes estimated for precipitate samples are meaningful when considered relative to the native sample estimates; the relatively constant fitting deviation is eliminated via subtraction.

We have argued that the spectral quality does not introduce artifacts in the RSAP estimates and that considering precipitate structure relative to the lyophilized structure eliminates the systematic errors in the RSAP analysis. The estimated structural perturbations imposed by precipitation with Na_2SO_4 and KSCN were analyzed to determine if these changes were statistically meaningful; since ribonuclease A was not precipitated by KSCN, it was eliminated from consideration. The hypothesis $\chi_o(\bar{D}' = 0)$ was tested where \bar{D}' is the average deviation between H_T or S_T for precipitates and the corresponding lyophilized samples. At the 1% significance level, χ_o could be rejected for both the β -sheet increases and the α -helix decreases found with the KSCN precipitates. The structural changes found for the Na_2SO_4 precipitates are of questionable significance; χ_o cannot be rejected at the 80% significance level. This is commensurate with the characterization of KSCN as a chaotropic salt and Na_2SO_4 as a structure preserving salt.

Given that the protein conformational changes induced by KSCN are significant and those by Na_2SO_4 are uncertain, an analysis of the changes as a function of the physical properties of the particular proteins was performed. Primary structure effects have been quantified by the fraction of hydrophobic residues, f_{hyd} , and the fraction of charged residues, f_{chg} ; f_{hyd} has been used

with some success in the correlation of the relative hydrophobicity of proteins [19] and f_{chg} has been found to be inversely related to the hydrophobic surface area of proteins [20]. The secondary structure contents determined by x-ray crystallography, S_{T_x} and H_{T_x} , have been used to estimate native structure contributions to high salt behavior. Melander and Horvath [20] have identified the protein dipole moment, μ , and the hydrophobic surface area, Φ , as the protein physical properties relevant to salting-out; these properties are determined by the tertiary structure of the protein. Estimates of μ [21] and of the solvent accessible surface area, A_s , [22,23] for the protein set have been used to quantify the role of tertiary structure; Miller and coworkers [22] have found that Φ is a nearly constant fraction of A_s for a set of 37 different proteins. The values for ΔS_T and ΔH_T , where

$$\Delta H_T = H_T - H_{T_x} \quad (3)$$

and

$$\Delta S_T = S_T - S_{T_x} \quad (4)$$

for precipitates from both salts are presented with a compilation of f_{hyd} , f_{chg} , S_{T_x} , H_{T_x} , μ , and A_s in Table 4.

To determine if relationships exist between the structural changes and the structural properties listed above, the hypothesis $\mathcal{H}_o(\rho_{XY} = 0)$ was tested for each of the possible parameter pairs. The sample correlation coefficient, r_{XY} , defined in equation (1) was computed along with the probability $P(|t| \geq t_o)$ and is presented in Table 5. For the KSCN precipitates (excluding ribonuclease A), ΔH_T is found to be negatively correlated with ΔS_T , H_{T_x} , and μ ; \mathcal{H}_o can be rejected at the 10% confidence level. At the same level, ΔS_T is correlated with f_{chg} and A_s . The strong negative correlation between the change in β -sheet

content and the change in α -helix content supports the earlier conjecture that β -sheet formation occurs at the expense of existing α -helix structures during precipitation by a chaotrope. The β -sheet structures formed are thought to be of an intermolecular nature [4,5]. The negative correlation between change in α -helix content and initial α -helix content suggests that helices are not equally susceptible to disruption. The relationship between ΔH_T and μ may be attributed to the fact that helices may have significant dipole moments [24]. The strong dependence of the β -sheet increase on the charged group frequency is somewhat surprising. A possible explanation for this may be the inverse relationship found between the charged group frequency and the hydrophobic surface area [20]; if aggregation in precipitation events requires the interaction of hydrophobic patches, then those proteins with smaller hydrophobic surface areas, and consequently, greater charged group frequencies, will be obliged to undergo more extensive changes to fulfill the surface area needs. In Table 5, there is a slight negative correlation between f_{chg} and A_s .

At the 10% significance level, χ_o may be rejected for the change in α -helix content with ΔS_T and H_{T_x} for the Na_2SO_4 precipitates; recall that the significance of the changes for these samples is not clear and that the number of samples is smaller. The correlation between ΔH_T and ΔS_T for Na_2SO_4 - induced precipitates is weaker than that for KSCN. The relationship between ΔH_T and H_{T_x} is the opposite of that found for the KSCN precipitates.

With the exception of the correlation between the change in α -helix and the initial helical content, the correlations of the structure changes with the structural properties considered here differ only in magnitude for KSCN and Na_2SO_4 . Other than the inverse relationship between the changes in helix and sheet contents, the most significant correlation found was that between the in-

crease in β -sheet contents for KSCN-induced precipitates and the charged group frequency. It is interesting to note that \mathcal{M}_o may be rejected at the 1% significance level for μ and A_s , and that μ and f_{chg} do not appear to be related. This is consistent with the findings of Barlow and Thornton [22]; the distribution of charged groups on the protein's surface makes a larger contribution to the dipole moment than the actual number of charges.

4.5 Discussion

The conformational perturbations imposed by precipitation are significant for KSCN. Extrapolating from the correlation analysis, hydrophobic forces dominate changes in β -sheet while electrostatic forces are operative in changes in α -helix. This assertion is based on the positive correlation found between the fraction of charged amino acids and the change in β -sheet content and the inverse relationship between charged group frequency and hydrophobic surface area [20]. This analysis complements the results of investigations of proteins and peptides in other aggregative processes.

Studies of the thermal denaturation of ovalbumin found an increase in both the exposed hydrophobic residues and the β -sheet content [25]. The aggregation of peptides, corresponding to sequences of hydrophobic amino acids found in naturally occurring proteins, in polar solvents is accompanied by β -sheet formation [8]; further, if alanines in the peptides were replaced with α -aminoisobutyric acid, which disrupts β -sheets and promotes α -helix formation, the solubility increased. Shimada and Matsushita [26] used f_{hyd} to distinguish between gelation-type and coagulation-type proteins.

The only relationship found between structural change and initial secondary structure content was that between ΔH_T and H_{T_x} . In studies of thermally-

induced aggregation, Hegg [27] determined that gel-forming potential was unrelated to the protein native secondary structure content. The generally low level of correlation between the structural changes and the elements of structure in the native proteins may be the result of inappropriate choices for the important protein physical properties considered and/or the concerted action of a number of properties.

Conformational perturbations observed in this work, in particular the formation of β -sheet structures, are consistent with those found in the aggregation of other proteins. Fuhrhop and coworkers [9] found that mixtures of helical poly-L-lysine and poly-D-lysine spontaneously form precipitates with the β -sheet structure. A number of other workers have found β -sheet increases accompanying thermally-induced aggregation [4-6] and solvent-induced aggregation [2,3,7]. Koenig's group has proposed mechanisms for thermal denaturation and intermolecular β -sheet formation [4,5].

This work has also demonstrated that α -helix structures may be sacrificed to form β -sheet structures. Many of the amino acids identified as good α -helix formers in the secondary structure prediction algorithm of Chou and Fasman [28] also appear high in the ranking of β -sheet formers. Environments of high ionic strength may disrupt charged-group interactions stabilizing helical conformations to an extent that sheet conformations become energetically favorable. Since helices may be amphipathic [29], correlations with the fraction of hydrophobic or charged amino acids in the protein may be diminished.

The body of work concerning the structural changes associated with protein aggregation prompts the question, 'Why β -sheet?'. In strongly denaturing environments such as high concentrations of chaotropic salts, high temperature, or low dielectric constant solvents, why is the formation of an ordered struc-

ture favored and why is β -sheet the structure of choice? The answer may lie in the study of subunit interactions in oligomeric proteins and in the hydration properties of secondary structures.

Many quaternary protein structures involve regions of β -sheet at the subunit contact sites [30]. Concanavalin A [31] and liver alcohol dehydrogenase [32] subunits interact through hydrogen bonding of adjacent antiparallel β -sheet strands. Similar interactions are seen at specific aggregation sites of monomeric proteins such as β -lactoglobulin [12]. β -sheets have relatively large surface areas and present opportunities for ordered hydrogen bonding.

Solvation effects may also play an important role. The structure of α -chymotrypsin is related to the extent of hydration [33]. Barlow and Poole [34] found that the strength of water binding to the backbone carbonyl moieties was correlated with the type of secondary structure present; the binding strength for β -sheet structures was lower than that for α -helix due to differences in the geometry of the water-carbonyl group interactions. Monte Carlo simulations of the structure of water surrounding periodic polypeptides in the β -sheet conformation with and without ions indicated that ions weakened water-water and water-amino acid interactions; water molecules became polarized towards the ions and away from each other and the amino acids [35]. The effects of salts on the structure of water and differences in the hydration of different secondary structures may account for the preferential formation of β -sheet structure in protein precipitates.

4.6 Acknowledgements

The authors are indebted to Drs. Helen Thompson and John Freeman of the Monsanto Corporate Research Laboratories for providing access to their Raman

spectroscopy facility and for useful discussions. This work was supported by the National Science Foundation.

4.7 References

- 1 Przybycien T.M., and Bailey, J.E., "Structure-Function Relationships in the Inorganic Salt-Induced Precipitation of α -Chymotrypsin," *Biochim. Biophys. Acta*, (1989) in press
- 2 Timasheff, S.N., and Stevens, L., "Infrared Spectra and Protein Conformations in Aqueous Solutions II. Survey of Globular Proteins," *J. Biol. Chem.* **242**, 5467-5473 (1967)
- 3 Herskovits, T.T., Gadegbeku, B., and Jaillet, H., "On the Structural Stability and Solvent Denaturation of proteins I. Denaturation by the alcohols and Glycols," *J. Biol. Chem.* **245**, 2588-2598 (1970)
- 4 Lin, V.J.C., and Koenig, J.L., "Raman Studies of Bovine Serum Albumin," *Biopolymers* **15**, 203-218 (1976)
- 5 Painter, P.C., and Koenig, J.L., "Raman Spectroscopic Study of the Proteins of Egg White," *Biopolymers* **15**, 2155-2166 (1976)
- 6 Clark, A.H., Saunderson, D.H.P., and Suggestt, A., "Infrared and Laser-Raman Spectroscopic Studies of Thermally-Induced Globular Protein Gels," *Int J. Peptide Protein Res.* **17**, 353-364 (1981)
- 7 Nemoto, N., Kiyohara, T., Tsunashima, Y., and Kurata, M., "Raman, IR and CD Spectroscopy on Aggregates and Gels of Lysozyme in Alcohols," *Bull. Inst. Chem. Res. (Kyoto)* **61**, 203-213 (1983)
- 8 Narita, M., Chen, J-Y., Sato, H., and Kim, Y., "Critical Peptide Size for Insolubility Caused by a β -Sheet Aggregation and Solubility Improvement in Hydrophobic Peptides by Replacement of Alanine Residues with α -Aminoisobutyric Acid Residues," *Bull. Chem. Soc. Jpn.* **58**, 2494-2501 (1985)
- 9 Fuhrhop, J-H., Krull, M., and Büldt, G., "Precipitates with β -Pleated

- Sheet Structure by Mixing Aqueous Solutions of Helical Poly(D-lysine) and Poly(L-lysine)," *Angew. Chem. Int. Ed. Eng.* **26** 699-700 (1987)
- 10 Maeda, H., Fukada, K., and Ikeda, S., "Kinetics of Chain Folding of Model Polypeptides into the β -Structure," in *Ordering and Organisation in Ionic Solutions* (Ise, N., and Soga, I., Eds) World Scientific, New Jersey, 193-202 (1987)
 - 11 Wong, P.T.T., and Heremans, K., "Pressure Effects on Protein Secondary Structure and Hydrogen Deuterium Exchange in Chymotrypsinogen: A Fourier Transform Infrared Spectroscopic Study," *Biochim. Biophys. Acta* **957**, 1-9 (1988)
 - 12 Casal, H.L., Köhler, U., and Mantsch, H.H., "Structural and Conformational Changes of β -Lactoglobulin B: An Infrared Spectroscopic Study of the Effect of pH and Temperature," *Biochim. Biophys. Acta* **957**, 11-20 (1988)
 - 13 Williams, R.W., "Estimation of Protein Secondary Structure from the Laser Raman Amide I Spectrum," *J. Mol. Biol.* **166**, 581-603 (1983)
 - 14 Przybycien, T.M., and Bailey, J.E., "Solubility-Activity Relationships in the Inorganic Salt-Induced Precipitation of α -Chymotrypsin," *Enz. Microb. Technol.*, (1989) in press
 - 15 Susi, H., and Byler, D.M., "Fourier transform Infrared Study of Proteins with Parallel β -Chains," *Arch. Biochem. Biophys.* **258**, 465-469 (1987)
 - 16 Berjot, M., Marx, J., and Alix, A.J.P., "Determination of the Secondary Structure of Proteins from the Raman Amide I Band: The Reference Intensity Profiles Method," *J. Raman Spectrosc.* **18**, 289-300 (1987)
 - 17 Levitt, M., and Greer, J., "Automatic Identification of Secondary Structure in Globular Proteins," *J. Mol. Biol.* **114**, 81-293 (1977)

- 18 Fisz, M., *Probability Theory and Mathematical Statistics*, 3rd ed., Wiley, New York, 427-433 (1963)
- 19 Bigelow, C.C., "On the Average Hydrophobicity of Proteins and the Relation between it and Protein Structure," *J. Theoret. Biol.* **16**, 187-211 (1967)
- 20 Melander, W., and Horvath, C., "Salt Effects on Hydrophobic Interactions in Precipitation and Chromatography of Proteins: An Interpretation of the Lyotropic Series," *Arch. Biochem. Biophys.* **183**, 200-215 (1977)
- 21 Barlow, D.J., and Thornton, J.M., "The Distribution of Charged Groups in Proteins," *Biopolymers* **25**, 1717-1733 (1986)
- 22 Miller, S., Janin, J., Lesk, A.M., and Chothia, C., "Interior and Surface of Monomeric Proteins," *J. Mol. Biol.* **196** 641-656 (1987)
- 23 Miller, S., Lesk, A.M., Janin, J., and Chothia, C., "The Accessible Surface Area and Stability of Oligomeric Proteins," *Nature (London)* **328**, 834-836 (1987)
- 24 Van Belle, D., Couplet, I., Prevost, M., and Wodak, S.J., "Calculations of electrostatic Properties in Proteins, Analysis of Contributions from Induced Protein Dipoles," *J. Mol. Biol.* **198**, 721-735 (1987)
- 25 Kato, A., and Takagi, T., "Formation of Intermolecular β -Sheet Structure during Heat Denaturation of Ovalalbumin," *J. Agric. Food Chem.* **36**, 1156-1159 (1988)
- 26 Shimada, K., and Matsushita, S., "Relationship between Thermocoagulation of proteins and Amino Acid Compositions," *J. Agric. Food Chem.* **28** 413-417 (1980)
- 27 Hegg, P-O., "Conditions for the Formation of Heat-Induced Gels of Some Globular Food Proteins," *J. Food Sci.* **47** 1241-1244 (1982)

- 28 Chou, P.Y., and Fasman, G.D., "Prediction of Protein Conformation," *Biochemistry* **13**, 222-245 (1974)
- 29 Cornette, J.L., Cease, K.B., Margalit, H., Spogue, J.L., Berzofsky, J.A., and DeLisi, C., "Hydrophobicity Scales and Computational techniques for Detecting Amphipathic Structures in Proteins," *J. Mol. Biol.* **195**, 659-685 (1987)
- 30 Liljas, A., and Rossmann, M.G., "X-Ray Studies of Protein Interactions," *Ann. Rev. Biochem.* **43**, 475-507 (1974)
- 31 Reeke, Jr., G.N., Becker, J.W., and Edelman, G.M., "The Covalent and Three-Dimensional Structure of Concanavalin A IV. Atomic Coordinates, Hydrogen Bonding, and Quaternary Structure," *J. Biol. Chem.* **250**, 1525-1547 (1975)
- 32 Brändén, C-I., Eklund, H., Nordström, B., Boiwe, T., Söderlund, G., Zeppezauer, E., Ohlsson, I., and Åkeson, Å., "Structure of Liver Alcohol Dehydrogenase," *Proc. Nat. Acad. Sci. USA* **70**, 2439-2442 (1973)
- 33 Lüscher, M., and Rüegg, M., "Thermodynamic Studies of the Interaction of α -Chymotrypsin with Water I. Determination of the Isosteric Enthalpies and Entropies of Water Binding to the Native Enzyme," *Biochim. Biophys. Acta* **533**, 428-439 (1978)
- 34 Barlow, D.J., and Poole, P.L., "The Hydration of Protein Secondary Structures," *FEBS Lett.* **213**, 423-427 (1987)
- 35 Liegener, C-M., Otto, P., Chen, R., and Ladik, J., "On the Electronic Band Structure of Periodic β -Pleated Sheet Polypeptides in the Presence of Water and Ions," *Theor. Chim. Acta* **73**, 449-458 (1988)
- 36 Schmidt, J.J., and Colowick, S.P., "Chemistry and Subunit Structure of Yeast Hexokinase Isoenzymes," *Arch. Biochem. Biophys.* **158**, 458-470

(1973)

- 37 Kishi, F., Maruyama, M., Tanizawa, Y., and Nakazawa, A., "Isolation and Characterization of cDNA for Chicken Muscle Adenylate Kinase," *J. Biol. Chem.* **261**, 2942-2945 (1986)

Table 2 Precipitate secondary structure estimates

Structure fractions expressed as percent of total structure content. Symbols used: H_o , ordered α -helix; H_d , disordered α -helix; S_a , antiparallel β -sheet; S_p , parallel β -sheet; T , β -reverse turn; R , random coil; H_T , total α -helix content ($H_T = H_o + H_d$); S_T , total β -sheet content ($S_T = S_a + S_p$); U , undefined structure ($U = R + T$); RNORM, Euclidean norm of residual structure fit vector; and S/N, signal-to-noise ratio of amide I band.

| protein | sample | H_o | H_d | S_a | S_p | T | R | H_T | S_T | U | S/N | RNORM |
|---------|--|-------|-------|-------|-------|-----|-----|-------|-------|-----|-----|-------|
| CHA | x-ray | 2 | 12 | 50 | 0 | 17 | 19 | 14 | 50 | 36 | | |
| | lyo ^a | 0 | 13 | 58 | 0 | 15 | 16 | 13 | 58 | 31 | 150 | 4.0 |
| | Na ₂ SO ₄ ↓ ^a | 0 | 10 | 58 | 0 | 22 | 11 | 10 | 58 | 32 | 40 | 3.3 |
| | KSCN↓ ^a | 0 | 6 | 59 | 8 | 19 | 8 | 6 | 67 | 27 | 40 | 3.2 |
| TPN | x-ray | 4 | 9 | 55 | 0 | 24 | 8 | 13 | 55 | 32 | | |
| | lyo ^a | 0 | 11 | 58 | 0 | 20 | 11 | 11 | 58 | 31 | 175 | 3.7 |
| | KSCN↓ | 0 | 13 | 60 | 1 | 10 | 16 | 13 | 61 | 26 | 32 | 4.9 |
| SUB | x-ray | 20 | 10 | 7 | 30 | 22 | 11 | 30 | 37 | 33 | | |
| | lyo ^a | 0 | 13 | 58 | 0 | 15 | 16 | 13 | 58 | 31 | 150 | 4.0 |
| | Na ₂ SO ₄ ↓ ^a | 0 | 10 | 58 | 0 | 22 | 11 | 10 | 58 | 32 | 40 | 3.3 |
| | KSCN↓ ^a | 0 | 6 | 59 | 8 | 19 | 8 | 6 | 67 | 27 | 40 | 3.2 |
| ADH | x-ray | 18 | 12 | 30 | 8 | 19 | 12 | 29 | 37 | 34 | | |
| | lyo ^b | 5 | 12 | 53 | 0 | 18 | 13 | 18 | 53 | 30 | 55 | 3.5 |
| | Na ₂ SO ₄ ↓ | 0 | 14 | 43 | 16 | 14 | 14 | 14 | 48 | 28 | 37 | 3.9 |
| | KSCN↓ ^b | 0 | 5 | 76 | 0 | 6 | 15 | 5 | 76 | 20 | 59 | 5.6 |
| SOD | x-ray | 0 | 0 | 52 | 0 | 35 | 13 | 0 | 52 | 48 | | |
| | lyo ^b | 0 | 13 | 61 | 0 | 5 | 18 | 13 | 61 | 23 | 44 | 6.3 |
| | Na ₂ SO ₄ ↓ | 0 | 0 | 34 | 41 | 15 | 10 | 0 | 76 | 24 | 6 | 9.4 |
| | KSCN↓ | 0 | 7 | 77 | 0 | 0 | 16 | 7 | 77 | 16 | 20 | 7.3 |
| IGG | x-ray | 1 | 2 | 62 | 3 | 19 | 13 | 3 | 65 | 32 | | |
| | lyo | 0 | 6 | 71 | 0 | 13 | 10 | 6 | 71 | 23 | 59 | 3.2 |
| | Na ₂ SO ₄ ↓ | 0 | 8 | 53 | 12 | 16 | 14 | 8 | 65 | 27 | 31 | 2.9 |
| | KSCN↓ | 0 | 10 | 66 | 1 | 10 | 13 | 10 | 67 | 23 | 53 | 4.6 |

Table 2 (cont'd)

| protein | sample | H_o | H_d | S_a | S_p | T | R | H_T | S_T | U | S/N | RNORM |
|---------|-----------------------------------|-------|-------|-------|-------|-----|-----|-------|-------|-----|-----|-------|
| HEK | x-ray | 29 | 6 | 11 | 6 | 25 | 23 | 35 | 17 | 48 | | |
| | lyo ^b | 0 | 16 | 35 | 18 | 22 | 10 | 16 | 53 | 32 | 6 | 7.9 |
| | KSCN↓ | 0 | 17 | 58 | 9 | 3 | 13 | 17 | 67 | 16 | 14 | 5.3 |
| ADK | x-ray | 40 | 22 | 0 | 19 | 15 | 5 | 62 | 19 | 19 | | |
| | lyo | 4 | 15 | 49 | 0 | 23 | 9 | 19 | 49 | 32 | 23 | 3.8 |
| | KSCN↓ ^b | 0 | 0 | 71 | 5 | 14 | 10 | 0 | 75 | 25 | 38 | 4.7 |
| LYZ | x-ray | 24 | 22 | 19 | 0 | 22 | 12 | 46 | 19 | 35 | | |
| | lyo ^a | 8 | 23 | 38 | 0 | 21 | 10 | 31 | 38 | 31 | 150 | 3.8 |
| | Na ₂ SO ₄ ↓ | 18 | 27 | 33 | 0 | 15 | 6 | 46 | 32 | 22 | 10 | 4.2 |
| | KSCN↓ | 17 | 23 | 39 | 0 | 0 | 21 | 40 | 39 | 21 | 19 | 4.9 |
| CNA | x-ray | 0 | 2 | 65 | 0 | 22 | 10 | 2 | 65 | 33 | | |
| | lyo | 0 | 3 | 76 | 2 | 14 | 6 | 3 | 78 | 19 | 39 | 2.8 |
| | KSCN↓ | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 100 | 0 | 3 | 14.5 |
| RNS | x-ray | 11 | 12 | 46 | 0 | 21 | 10 | 23 | 46 | 31 | | |
| | lyo | 0 | 13 | 58 | 0 | 16 | 13 | 13 | 58 | 29 | 150 | 3.8 |
| | KSCN↓ ^c | 3 | 22 | 40 | 0 | 0 | 35 | 24 | 40 | 35 | 9 | 8.5 |
| CYT | x-ray | 32 | 16 | 0 | 10 | 18 | 24 | 48 | 10 | 42 | | |
| | lyo | 78 | 0 | 0 | 0 | 6 | 16 | 78 | 0 | 22 | 9 | 46.1 |

^a RSAP results taken from Przybycien and Bailey [14]

^b Rounding error gives sum of structure contents exceeding 100%

^c Sample prepared by lyophilization from KSCN solution

Table 3 Correlation between secondary structure estimates and spectral noise

The statistical significance of the hypothesis that structure estimates obtained from amide I band spectra via the RSAP analysis are uncorrelated with the quality of the individual spectra, $\mathcal{H}_o(\rho_{XY} = 0)$, is given below in matrix form. The subdiagonal entries are the sample correlation coefficients r_{XY} as defined in equation (1) with the sample size in parentheses below. Superdiagonal elements are the significance of \mathcal{H}_o from tables of the Student's t -distribution, $P(|t| > t_o)$. The symbols used are: H_T , the total α -helix RSAP estimate; S_T , the total β -sheet RSAP estimate; S/N, the spectral signal-to-noise ratio; and RNORM, the Euclidean norm of the residual fitted structure vector.

| | H_T | S_T | S/N | RNORM |
|-------|---------------|---------------|---------------|-----------|
| H_T | | $\ll 0.001$ | < 0.8 | < 0.001 |
| S_T | -0.92 (28) | | < 0.8 | < 0.01 |
| S/N | -0.08 (28) | -0.05 (28) | | < 0.2 |
| RNORM | -0.70 (28) | -0.50 (28) | -0.28 (28) | |

Table 4 Precipitate structure changes and physical properties of native proteins

The secondary structure changes in the total α -helix and total β -sheet contents of the protein precipitates studied are listed with quantities that represent primary, secondary, and tertiary structure properties in the native proteins. The symbols used are: ΔH_T^x and ΔS_T^x , the relative change in α -helix and β -sheet content for precipitates induced by KSCN ($x = K$) or Na_2SO_4 ($x = N$), respectively; f_{hyd} , the hydrophobic amino acid fraction; f_{chg} , the charged amino acid fraction; H_{Tx} and S_{Tx} , the total α -helix and β -sheet contents assigned via the x-ray structure; μ , dipole moment; A_s , solvent accessible surface area.

| protein ^a | ΔH_T^K | ΔS_T^K | ΔH_T^N | ΔS_T^N | f_{hyd}^b | f_{chg}^b | H_{Tx} (De) | S_{Tx} (Å ²) | μ^c | A_s^d |
|----------------------|----------------|----------------|----------------|----------------|--------------------|--------------------|------------------|-------------------------------|---------|---------|
| CHA | -7 | 9 | -3 | 0 | 39 | 12 | 14 | 50 | 223 | 10,440 |
| TPN | 2 | 3 | | | 34 | 13 | 13 | 55 | 356 | 9,930 |
| SUB | -7 | 14 | | | 44 | 12 | 30 | 37 | 508 | 10,390 |
| ADH | -13 | 23 | -4 | 5 | 44 | 23 | 29 | 37 | 823 | 13,610 |
| SOD | -6 | 16 | -13 | 15 | 34 | 28 | 0 | 52 | 245 | 6,760 |
| IGG | 4 | -4 | 2 | -6 | 36 | 14 | 3 | 65 | | 37,800 |
| HEK | 1 | 14 | | | 36 | 14 | 3 | 65 | | 18,200 |
| ADK | -19 | 26 | | | 34 | 33 | 62 | 19 | | 8,500 |
| LYZ | -2 | 4 | 15 | -6 | 34 | 22 | 49 | 19 | 111 | 6,620 |
| CNA | -3 | 22 | | | 38 | 27 | 2 | 65 | 187 | 8,300 |
| RNS | 11 | -18 | | | 30 | 23 | 23 | 46 | 481 | 6,790 |
| CYT | | | | | 31 | 35 | 48 | 10 | 310 | 5,570 |

a Acronyms defined in Table 1.

b Computed from sequences given in Levitt and Greer [17], with the exception of HEK [36] and ADK [37]

c Given in Barlow and Thornton [21]

d Given in Miller et al. [22,23]; TPN and IGG computed from $A_s = 5.3M_w^{0.76}$ [22]; SOD and HEK computed from $A_s = 6.3M_w^{0.73}$ [23]

Table 5 Correlation between estimated secondary structure changes and elements of primary, secondary, and tertiary structure

Statistical significance of the hypothesis that the salt-induced changes in protein secondary structure on precipitation are uncorrelated with the primary, secondary, and tertiary structure of the corresponding native conformation, $\mathcal{H}_o(\rho_{XY} = 0)$. The format of the matrix is the same as in Table 3 and the symbols used are the same as in Table 4. RNS was not included in the calculations concerning KSCN precipitates. RNS and CYT were included in the computation of the correlation terms relating the structural properties of native proteins.

| | ΔH_T^K | ΔS_T^K | ΔH_T^N | ΔS_T^N | f_{hyd} | f_{chg} | H_{Tz} | S_{Tz} | μ | A_s |
|------------------|----------------|----------------|----------------|----------------|------------------|------------------|---------------|--------------|--------------|--------|
| ΔH_T^K | | < 0.01 | < 0.4 | < 0.4 | < 0.8 | < 0.4 | < 0.1 | < 0.4 | < 0.1 | < 0.2 |
| ΔS_T^K | -0.79 (10) | | < 0.4 | < 0.2 | < 0.4 | < 0.05 | < 0.4 | < 0.4 | < 0.4 | < 0.1 |
| ΔH_T^N | 0.55 (5) | -0.57 (5) | | < 0.1 | < 0.8 | < 0.8 | < 0.1 | < 0.4 | < 0.8 | > 0.8 |
| ΔS_T^N | -0.58 (5) | 0.76 (5) | -0.87 (5) | | > 0.8 | < 0.2 | < 0.6 | < 0.8 | < 0.8 | < 0.6 |
| f_{hyd} | -0.14 (10) | 0.29 (10) | -0.27 (5) | 0.12 (5) | | < 0.6 | < 0.8 | > 0.8 | < 0.2 | < 0.6 |
| f_{chg} | -0.29 (10) | 0.62 (10) | -0.29 (5) | 0.69 (5) | -0.24 (12) | | < 0.2 | < 0.05 | > 0.8 | < 0.4 |
| H_{Tz} | -0.53 (10) | 0.31 (10) | 0.75 (5) | -0.44 (5) | -0.12 (12) | 0.45 (12) | | < 0.001 | < 0.8 | < 0.4 |
| S_{Tz} | 0.40 (10) | -0.30 (10) | -0.59 (5) | 0.17 (5) | 0.07 (12) | -0.59 (12) | -0.92 (12) | | > 0.8 | < 0.4 |
| μ | -0.68 (7) | 0.48 (7) | -0.33 (4) | 0.26 (4) | 0.48 (9) | -0.09 (9) | 0.12 (9) | -0.05 (9) | | < 0.02 |
| A_s | 0.47 (10) | -0.56 (10) | 0.09 (5) | -0.47 (5) | 0.24 (12) | -0.26 (12) | -0.33 (12) | 0.34 (12) | 0.67 (12) | |

Chapter 5

Aggregation Kinetics in Salt-Induced Protein Precipitation

The material contained in this chapter has been prepared
for submission to the *AIChE Journal*

5.1 Abstract

The kinetics of protein aggregation in salt-induced precipitation processes were studied as a function of salt type, salt concentration, temperature and protein concentration. α -chymotrypsin (α CT) was used as a model protein. Stopped-flow turbidimetry was used to monitor the progress of precipitations. Analysis of the linear portions of the turbidity trajectories indicate that temperature and salt concentration effects are related to protein solubility; the protein concentration dependence is well-described by the Smoluchowski collision equation. The aggregation kinetics of partially inhibited α CT exhibit poisoning behavior, underscoring the importance of dimerization and monomer addition. Solute particle radius distributions determined via dynamic laser light scattering for low salt and supernatant α CT solutions demonstrate indicated that significant aggregation does not occur in the absence of supersaturation. A detailed population balance model was proposed that accounted for specific and non-specific interactions and monomer addition. The model is expected to find general application to protein aggregation phenomena, in particular, for proteins that have specific quaternary interactions.

5.2 Introduction

Inorganic salt-induced precipitation is often used as a primary isolation procedure in bioprocess recovery operations (Bailey and Ollis, 1986). The formation of protein precipitates involves the facilitation or enhancement of protein-protein interactions, via physical or chemical means, resulting in aggregation; the kinetics of protein self-association are intimately linked with environmental conditions. An understanding of the nature and extent of the effects of fundamental process parameters on the intrinsic precipitation kinetics may have implications for the development of design and operation guidelines. In addition, precipitation may be a useful paradigm for protein aggregation occurring in other systems such as the gelation of food proteins (Hegg, 1982), the self-association of pharmaceuticals during compounding operations (Massey et al., 1988), and the formation of inclusion bodies (Kane and Hartley, 1988; Gardiner et al., 1987).

In this work, we investigated environmental factors affecting perikinetic aggregation induced by salt addition. Perikinetic aggregation describes protein-protein interactions driven by the removal of hydration and/or electrostatic barriers; the macroscopic or orthokinetic regime is characterized by the domination of hydrodynamic forces (Bell et al., 1983). α -chymotrypsin (α CT), a serine protease, was used as a model globular protein. Process parameters studied include protein concentration, temperature, and the type and amount of inducing salt. The salts used span the lyotropic series representing a range of salting-out and chaotropic potentials. Na_2SO_4 , NaCl, NaBr, KBr and KSCN were among the salts used.

The kinetics of the precipitation process were followed by stopped-flow turbidimetry. This technique gives the weight-average molecular weight of the

solution as a function of time. Dynamic laser light scattering (DLS) was used to estimate the initial particle size distributions and determine the aggregative state of protein in supernatant samples. A general kinetic scheme incorporating aspects of protein chemistry peculiar to α CT, including the structural effects of precipitation (Przybycien and Bailey, 1989), was developed to model the stopped-flow data. The model consists of simultaneous salt-mediated conformational change and dimerization followed by the indefinite association of altered dimers augmented by monomer addition. A detailed population balance, based on Smoluchowski aggregation theory, was formulated in terms of the moments of the particle size distribution (Ziff, 1984). Relevant model rate constants were estimated as a function of the process conditions.

This work is similar in spirit to studies of the association of caseins in the batch coagulation of milk. Dalgleish and co-workers have investigated the kinetics of CaCl_2 and chymosin-induced aggregation of caseins extensively using light scattering techniques (Dalgleish et al., 1981; Dalgleish, 1980; Parker and Dalgleish, 1977). Carlson et al. (1987) have further developed kinetic expressions for the coagulation of milk. The aggregation kinetics have been adequately described by the hydrolysis of micellar casein via Michaelis-Menten kinetics and subsequent Smoluchowski aggregation of the proteolyzed micelles.

The microscopic view taken here and in the milk coagulation investigations are complementary to the studies of orthokinetic aggregation in continuous-flow stirred-tank reactors (CSTR) and plug-flow reactors (PFR) by other investigators. Dunnill and Hoare have characterized particle size distributions and developed kinetic expressions for protein precipitations in process equipment; alcohol dehydrogenase and fumarase were precipitated from a clarified yeast extract in a CSTR using ammonium sulfate (Foster et al., 1976) and soya protein precip-

itated by a variety of reagents in a PFR (Chan et al., 1986). Shear effects on the kinetics of polyacrylic acid-induced precipitation of lysozyme from egg white have been investigated by Fisher and Glatz (1988 a,b). The flow fields in these studies are characterized by high average shear rates ($>100 \text{ s}^{-1}$) and Reynolds numbers (>500). Consequently, the precipitation kinetics are dominated by shear-induced collisions resulting in both particle aggregation and breakup.

Turbidity data was analyzed as a function of the initial supersaturation ratio via Smoluchowski's equation to gain insight into the parameters controlling α CT aggregation kinetics. The protein concentration dependence was well described by the collision equation. Temperature and salt concentration had essentially identical effects on the association constant; both of these parameters affect the supersaturation level via the absolute protein solubility. Experiments with dimerization-incompetent protein indicated that dimerization is a fundamental step in the aggregation of α CT and also that the rate of monomer addition to aggregating dimers is significant. α CT did not aggregate until a condition of supersaturation was created. Light scattering measurements of low salt concentration protein solutions and supernatant samples consisted essentially of monomers and dimers.

The kinetic model incorporates the structural perturbations associated with aggregation, treats particles created by the specific interaction (dimerization) as the primary aggregable species, and includes monomer addition to the aggregate. Rate constants were fit to the turbidity data; the fits were scattered due to the numerical realities of the solution techniques used, but the estimates fell within reasonable ranges and the gross features of the raw data were well-represented. A simulation of the inhibited α CT aggregation kinetics using rate constants fitted from control runs had the same qualitative behavior as the experimental

data, lending support to the veracity of the proposed model. The model may find application to aggregating protein systems that have specific quaternary interactions in addition to nonspecific interactions such as insulin (Jeffrey et al., 1976) and β -lactoglobulin (Casal et al. 1988).

5.3 Materials and Methods

5.3.1 Materials

The α -chymotrypsin (EC 3.4.21.1, three times crystallized from four times crystallized bovine pancreatic chymotrypsinogen, salt free) and 4-methylumbelliferyl-*p*-trimethyl ammonium cinnamate chloride (MUTMAC) were purchased from Sigma. L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK) was obtained from Boehringer Mannheim. All salts and buffer solution components were analytical reagent grade. Distilled, deionized water processed by a Corning Mega-Pure system was used throughout. All α CT and salt solutions were made up in 50 mM glycine buffer adjusted to pH 3.00 with HCl (Przybycien and Bailey, 1989a) unless otherwise noted.

5.3.2 Inhibition of α CT with TPCK

Inhibited α CT was prepared by alkylating histidine 57 with TPCK. TPCK inhibition renders α CT incapable of dimerization (Horbett and Teller, 1973; Neet and Brydon, 1970). The procedure of Schoellmann and Shaw (1963) was used with the following modifications. The starting enzyme solution was prepared in 100 mM phosphate buffer adjusted to pH 7.2 with NaOH. A 15-fold molar excess of TPCK was dissolved in acetonitrile before addition to the α CT solution; TPCK is sparingly soluble in aqueous media and its addition to the α CT solution resulted in noticeable turbidity. Active fraction assays using the MUTMAC suicide substrate (Przybycien and Bailey, 1989a) were employed to

assess the extent of inhibition. After 7.5 hours, the active fraction of the enzyme was reduced to $19.4 \pm 0.5\%$ from its initial value of $78.3 \pm 1.1\%$; the confidence limits represent the standard deviation of the mean of triplicate assays. Unreacted TPCK was removed from the enzyme solution via centrifugal ultrafiltration followed by gel permeation chromatography as in Przybycien and Bailey (1989b). Water was used as the eluent. Fractions were assayed for protein content with the Bio-Rad dye binding assay. α CT-containing fractions were pooled and lyophilized. This mixture of inhibited and unreacted α CT is referred to below as TPCK- α CT.

5.3.3 Stopped-flow turbidimetry

The kinetics of the salt-induced aggregation process were monitored by stopped-flow turbidimetry. In classical or Rayleigh light scattering, the absorbance, A_λ , of a solution of macromolecules is given by

$$A_\lambda = \frac{16\pi}{6.909} \frac{K(\lambda)C_o\bar{M}_w}{M_1} \quad (1)$$

where $K(\lambda)$ is a wavelength-dependent optical constant; M_1 is the monomeric solute molecular weight; C_o is the total concentration of the macromolecule in mg mL^{-1} and \bar{M}_w is the weight-average molecular weight of the solute (Cantor and Schimmel, 1980). Equation (1) is valid for dilute solutions where multiple scattering events are insignificant, and species that are small with respect to the wavelength of the incident light. This relationship holds for larger solutes as well, provided that the distribution is not significantly polydisperse (Carlson et al., 1987). The constant $K(\lambda)$ was determined by measuring the absorbances of a series of α CT solutions ranging from 2.5 to 100 mg mL^{-1} in glycine buffer; the absorbance at 488 nm, A_{488} , was used to facilitate comparisons with dynamic light scattering experiments. α CT in solution is an equilibrium mixture

of monomers and dimers, and as such, \bar{M}_w in equation (1) is a function of the total protein concentration and the dimerization equilibrium constant, K_D , giving

$$A_\lambda = \frac{16\pi}{6.909} \frac{K(\lambda)C_o}{M_1} \left(\frac{\sqrt{1 + 8K_D C_o / M_1}}{1 + \sqrt{1 + 8K_D C_o / M_1}} \right). \quad (2)$$

Gilleland and Bender (1976) reported a value of $4.97 \times 10^3 \text{ M}^{-1}$ for K_D at pH 3.89. Using this value, a nonlinear least squares fit of A_{488} data as a function of C_o to equation (2) gave $K(\lambda) = 2.56 \pm 0.09 \times 10^{-5} \text{ mL g}^{-1}$; the confidence interval represents an estimate of the standard deviation of the mean.

The stopped-flow system consisted of an Aminco-Morrow device customized for use with a Shimadzu UV-260 spectrophotometer. The major components of the fluid delivery system were a pair of pneumatically actuated drive syringes, a mixing tee, a Wilmad WG-41MH micro flow cuvette and a receiving syringe with an adjustable mechanical stop. The flow-cell had a volume of 0.21 mL, a 10 mm path length, and a metal outer jacket. Both the drive syringes were mounted in a thermostatted aluminum block using a recirculating water bath for temperature control. The flow-cell was placed in a Shimadzu CPS-260 thermoelectric holder/cell positioner. Samples were placed in a water bath at the temperature to be studied prior to use in the flow device. Absorbance was monitored at 488 nm with a resolution of 1 nm. The receiving syringe stop was set to allow approximately three flow-cell volumes to flow. Between runs, the cell was flushed to $A_{488} = 0.0$ with 0.1 M HCl and water. Data acquisition and control was facilitated by an IBM PC/XT interfaced with the spectrometer. The fastest data sampling rate available with our apparatus was 13.33 Hz.

The mixing efficiency of the stopped-flow device was tested by monitoring a pH indicator reaction. The protonation/deprotonation reactions of indicator solutions are essentially diffusion-limited. The corresponding color changes pro-

vide a facile means for assessing the duration of mixing transients with UV-vis detection. A standard bromocresol green indicator solution (Weast, 1980) was used; on acidification, the solution changes from blue to yellow. The rate of disappearance of the intense absorption band at 613 nm was followed after mixing with 0.1 M HCl in the device. Complete mixing ($A_{613} = 0.0$) occurred within 1 s after actuating the drive syringes.

For α CT aggregation experiments, the relationship between A_{488} and \bar{M}_w expressed in equation (1) is expected to hold up to $A_{488} = 1.26$. Formisano et al. (1978) conducted experiments concerning the self-association of glucagon and were successful using data up to $A_{310} = 0.8$; the cutoff value for this work was scaled-up by the wavelength. In the modelling work discussed below, only data between the end of the mixing transient and the time at which $A_{488} = 1.26$ were considered.

5.3.4 Solubility Determinations

To estimate kinetic parameters associated with precipitation, the protein solubility for a given salt type, salt concentration and temperature must be known. Measurements were performed in triplicate using the procedure of Przybycien and Bailey (1989a) with the following modifications. α CT and salt solutions in the glycine buffer were mixed using the stopped-flow device with a 15 mL centrifuge tube as the receiving vessel; the solubility of proteins in salt solutions is linked to the contacting conditions employed (Foster et al., 1976). Determinations were made at 25°C and in a cold room at 5°C. Precipitates were aged by standing 1 h and 5 h at 25 and 5°C, respectively. Supernatant aliquots were withdrawn and diluted 100-fold to minimize salt interference and protein aggregation effects. At this point, the 5°C samples were allowed to warm to

room temperature. Protein content was assayed with the Bio-Rad dye binding assay.

5.3.5 Dynamic Laser Light Scattering Measurements

The dynamic laser light scattering (DLS) apparatus included a Lexel 95 Ar⁺ laser, a custom-made goniometer, an EMI photon counting photomultiplier tube and a Brookhaven BI30 autocorrelator. α CT solutions were pipetted into precleaned glass sample tubes having a volume of about 0.8 mL. Sample cells were placed into a thermostatted paraffin oil index matching bath. The homodyne scattered intensity autocorrelation function was measured with a 90° scattering angle and a laser output of approximately 200 mW at the 488 nm line. The beam was attenuated with a 0.6 O.D. neutral density filter. All measurements were performed at 23°C.

Typical operating conditions consisted of a 1 μ s sample time and an accumulation time of 15 minutes. Baselines ranged from 10^7 to 10^9 . The autocorrelation functions were inverted to obtain particle size distributions using the constrained regularization program CONTIN 2DP (Provencher, 1982 a,b) on a VAX 11/780. The form factor for solid spheres was assumed. The refractive indexes of the various glycine buffer-salt solutions were determined with a Carl Zeiss Abbe refractometer at 23.5°C. The corresponding kinematic viscosities of these solutions was measured with model 25 A643 and 50 M724 Cannon-Fenske Routine Viscometers at 23°C.

α CT solutions were made up in 0.2 μ m filtered glycine buffer. Supernatant samples from precipitations were prepared by mixing the appropriate salt and α CT solutions in filtered buffer in the stopped-flow device. The precipitate phase was pelleted by spinning the 15 mL receiving tubes at 2700 rpm for 70 minutes

(about 10^5 G min) in a Beckman TJ-6 table-top centrifuge. The supernatant was allowed to remain in equilibrium with the precipitate and sampled prior to light scattering experiments.

5.4 Experimental Results

5.4.1 Apparent Aggregation Rate

The stopped-flow system was used to study the effects of protein concentration, salt concentration, salt type and temperature on the kinetics of precipitate formation. Each of these variables affects the supersaturation of the solution. The turbidity curves were typically sigmoidal with a substantially linear mid-section; the linear portion of the trajectories often extended well beyond the presumed spectrometer linear response limit at $A_{488} = 1.26$. At initial supersaturation ratios near unity, the mode of the curves shifted, resembling a first-order response. Representative turbidity curves for α CT-KSCN solutions are presented in Figures 1 through 3; the altered parameters are the salt concentration, the temperature, and the protein concentration in Figures 1, 2, and 3, respectively. Turbidity curves are used to present the aggregation kinetics since the relationship between A_{488} and \bar{M}_w expressed in equation (1) also depends on C_o . In the linear response range of the turbidity curves, the molecular weight increases by a factor of about 10^3 . All protein and salt concentrations quoted are after two-fold dilution by mixing in the stopped-flow device.

To facilitate the presentation and analysis of the raw data, an apparent Smoluchowski aggregation rate constant, k_{app} , was fit to the linear region of the turbidity curves. If N represents the solute particle number concentration, then (Bell, et al., 1983)

$$\dot{N} = -k_{app}N^2 \quad (3)$$

where the ($\dot{}$) notation is used to represent time differentiation. The integrated form of equation (3) may be expressed in terms of the particle number-average molecular weight, \bar{M}_n (Bell, et al., 1983)

$$\bar{M}_n(t) = \bar{M}_n(0)(1 + k_{app}C_0t/M_1). \quad (4)$$

The dimensions of k_{app} are $M^{-1} s^{-1}$. Thus, a linear increase in the solute weight average molecular weight with time is predicted for a purely collisional aggregation mechanism. The linear portion of the experimental data was fit via a least squares technique and the resulting slope estimates were converted to k_{app} values using equations (1) and (4). The initial supersaturation ratio, s_o , is the initial α CT concentration divided by the α CT solubility and is used to quantify the driving force for precipitation. The experimental data for each salt is summarized in a plot of k_{app} as a function of s_o in Figure 4; this plot provides a common basis for comparing the effects of the parameters studied on the aggregation kinetics of α CT.

The empirical relationship between the supersaturation ratio and the salt type and concentration is given by a rearrangement of the Cohn equation (Cohn and Edsall, 1943)

$$s = C_o \exp\{K_s m - \beta'\} \quad (5)$$

where K_s is the salting-out constant, m is the salt molality and β' is a constant. The dilution introduced by the stopped-flow device limits the salt concentration range available for study; the range is bounded at the upper end by half the salt saturation level and at the lower end by the salt concentration giving $s_o = 1$. At an α CT concentration of 10 mg mL^{-1} and 25°C , as the salt concentration and hence the initial supersaturation ratio increased, the apparent aggregation constant increased for KSCN, KBr and NaBr. These results are shown in Figure

4. The limitations on the concentration ranges for Na₂SO₄ and NaCl precluded corresponding experiments at reasonable αCT concentrations for these salts.

Temperature effects were probed in the range of 5 to 30°C for a given salt type at fixed salt and αCT concentrations. Assuming that the molal heat of solution of αCT, $\Delta\bar{H}_o$, is constant over this temperature range, the supersaturation ratio is given by (Denbigh, 1981)

$$\frac{s_{o1}}{s_{o2}} = \exp \left\{ \frac{\Delta\bar{H}_o}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \right\} \quad (6)$$

where R is the gas constant in cal mol⁻¹ K⁻¹ and s_{o1} and s_{o2} are the initial supersaturation ratios at temperatures T_1 and T_2 respectively. Values for $\Delta\bar{H}_o$ at a fixed salt concentration for each salt were estimated from solubility measurements at 5 and 30°C and are presented in Table 1. Using equation (6) it is possible to estimate initial supersaturation ratios for solutions with identical salt concentrations at intermediate temperatures. Increasing temperatures resulted in decreased αCT solubilities for each salt, giving increased initial supersaturation ratios. As the supersaturation increased, the apparent aggregation rate constant generally increased as shown in Figure 4.

αCT concentrations from 1.0 to 60 mg mL⁻¹ were also used to vary the initial supersaturation ratio. At fixed salt concentrations and 25°C, or constant αCT solubility, the estimated aggregation rate constants were essentially invariant to αCT concentration. Comparisons of the fitted Smoluchowski rate constants from salt to salt are not valid as the αCT solubility levels were different. These results are indicated in Figure 4 by the horizontal lines.

The effects of temperature and salt concentration, manifested by the initial supersaturation ratio, appear to be indistinguishable. Both variables alter the absolute αCT solubility and the apparent kinetic constant. Thus, the activation

energy of the association must be small. The sensitivity of the apparent aggregation rate constant to changes in the α CT solubility, evidenced by the slopes of the temperature and salt concentration data in Figure 4, varies from salt to salt. The rate of precipitation also varies with the α CT concentration, however, the apparent kinetics do not appear to be affected.

It is well known that α CT dimerizes in acidic environments (Aune and Timasheff, 1971; Gilleland and Bender, 1976). To determine the role of dimerization in α CT precipitation kinetics, stopped-flow experiments were performed using TPCK- α CT, holding conditions fixed at 0.5 m KSCN and 25°C. As a control, an equivalent concentration of α CT at the same conditions was also run. Since the inhibition of α CT with TPCK was not complete, a further control run using a concentration of α CT equivalent to the uninhibited enzyme concentration was also performed at the same conditions. The turbidity curves of these three experiments performed in duplicate are given in Figure 3.

The TPCK- α CT sample aggregated much more slowly than either of the control samples. The inhibited monomer has a poisoning effect on the protein that is still competent for dimerization; dimerization and monomer addition are important steps in the precipitation of α CT. The variation of the turbidity curves between duplicate runs is the likely result of fluctuations in the mixing efficiency (Walton, 1967).

5.4.2 Particle Size Distributions

Dynamic laser light scattering measurements gave insight into the initial and final aggregation state of soluble α CT in the precipitation experiments. Effects on the particle size distribution as manifested in the estimated weight-average particle radius, \bar{r}_w , were explored as a function of the total protein

concentration, C_o . Results were examined with respect to the predicted behavior.

For equilibrium mixtures of monomers and dimers, the solute weight-average molecular weight, \bar{M}_w , is given by

$$\frac{\bar{M}_w}{M_1} = \frac{A + 4A_2}{A + 2A_2} \quad (7)$$

where A and A_2 are the molar concentrations of monomer and dimer, respectively. For globular protein aggregates, \bar{r}_w is related to \bar{M}_w via (Parker and Dalglish, 1977)

$$\bar{M}_w = 1.88\bar{r}_w^3. \quad (8)$$

Using the equilibrium relationship between A and A_2 and the mass balance, \bar{r}_w may be expressed as a function of C_o

$$\frac{\bar{r}_w}{r_1} = \left[1 + \frac{(\sqrt{1 + 8K_D C_o/M_1} - 1)^2}{8K_D C_o/M_1} \right]^{1/3} \quad (9)$$

where r_1 is the radius of the monomer. Equation (9) is represented by the solid curve in Figure 5 using $M_1 = 25000$, $K_D = 4.97 \times 10^3 \text{M}^{-1}$, and $r_1 = 19.6 \text{ \AA}$. αCT is an oblate ellipsoid with dimensions $35 \times 58 \times 45 \text{ \AA}$ (Matthews et al., 1967); r_1 was computed as the radius of the sphere with an equivalent volume.

Experimentally determined \bar{r}_w values normalized by the monomer radius are also presented in Figure 5. For most of the samples, the particle size distributions were unimodal. Where multimodal distributions were encountered, \bar{r}_w values from the peak representing the smallest particles were reported; these peaks incorporated at least two orders of magnitude more protein than the next largest peak in the distribution and the moments were not significantly different from the moments of the entire distribution. The span of the error bars represents

the standard deviation of the weight-average radius distribution, σ_w , computed using (Herdan, 1960)

$$\sigma_w = (\bar{r}_z \bar{r}_w - \bar{r}_w^2)^{1/2} \quad (10)$$

where \bar{r}_z is the z-average particle radius. σ_w gives a quantitative indication of the polydispersity of the sample.

α CT solutions made up in the glycine buffer gave radii consistent with monomeric and dimeric particles. Neat aqueous α CT solutions had \bar{r}_w values lower than were expected. The buffered α CT solutions were reasonably described by equation (9); given the accuracy of the data, the use of the literature K_D value appears justified.

Particle size distributions were computed for α CT-salt solutions in which the salt concentration was not sufficient to induce precipitation. The thick vertical bar at 10 mg mL^{-1} in Figure 4 represents the range of α CT radii found for ten different such solutions. Salt concentrations were between 0.005 m and 0.1 m for each of the salts studied. No significant aggregation is induced until the minimum amount of salt resulting in $s_o > 1$ is added.

Supernatant α CT solutions from salt-induced precipitations were examined to assess the final aggregation state of soluble protein after precipitation was complete. Radii for supernatant solutions from 2.35 m KBr, 2.5 m NaCl, and 1.0 m Na_2SO_4 induced precipitations at 25°C were commensurate with particle populations consisting largely of monomers and dimers; these results are plotted at the corresponding α CT solubility level. The Na_2SO_4 sample had a distinctly bimodal population distribution and the normalized radius and standard deviation are given for the entire distribution. Samples examined from KSCN and NaBr-induced precipitations did not leave enough soluble protein for meaningful results.

5.5 Kinetics Modelling

5.5.1 Model Development

A reaction scheme was developed for the salt-induced aggregation of α CT and is presented in Figure 6. This mechanism is an interpretation of stopped-flow turbidity trajectories, particle size distributions, and secondary and tertiary structure effects (Przybycien and Bailey, 1989b). We have endeavored to incorporate all of the observed phenomena while eschewing complexity in the formulation of a plausible kinetic model.

As intimated above, native α CT is an equilibrium mixture of monomers and dimers at the pH and concentrations relevant to this work. This is indicated by the equilibrium between the native forms A and A_2 in Figure 6 and has been verified by the DLLS data shown in Figure 5. There is some degree of confusion concerning both the temperature and ionic strength sensitivity of the dimerization. At moderate ionic strengths, Aune et al. (1971) report a temperature-dependent enthalpy of dimerization at pH 4.12 while Tellam and Winzor (1971) consider the equilibrium to be temperature-independent at pH 7.9; the equilibrium constant, K_D , found by Tellam and Winzor at pH 7.9 agrees with that determined by Gilleland and Bender (1976) at pH 3.89. The type and concentration of electrolyte also affects K_D (Ikeda et al., 1982). Agapow and Winzor (1988) indicate that these effects diminish with increasing ionic strength. We have assumed that, at the high salt concentrations and narrow temperature range considered in this work, the forward and reverse dimerization rate constants k_{+d} and k_{-d} reported by Gilleland and Bender (1976) at pH 3.89 are adequate and remain constant.

Secondary structure estimates indicated that conformational changes may

occur during precipitation; however, the tertiary structure of the active site remains intact (Przybycien and Bailey, 1989b). This conformational change from the native to an altered state is indicated in Figure 6 by the parallel transformations of A to B and A_2 to B_2 . The conformational change is assumed to be first order; first order kinetics have proved an adequate description of structure in homopolypeptides resulting from temperature and pH jumps (Maeda et al., 1987; Auer and Miller-Auer, 1986). The same rate constant, k_1 , has been assigned to both the monomer and dimer reactions. Dimerization is known to occur about the active site (Matthews et al., 1967) and since the active site does not appear to play an active role in the conformational change, this may be a valid approximation. Thus, we have assumed that conformational change and dimerization occur at mutually exclusive sites on the α CT molecule.

This conformational change is assumed to occur only for the supersaturation α CT concentration. As salt is added to a α CT solution, competition for water of solvation results in the dehydration of a portion of the protein (the supersaturation). This process may be responsible for inducing the observed conformational change. In the model, α CT in excess of the soluble concentration level undergoes structural change and is removed from the native protein pool. Once the combined concentration of A and A_2 is reduced to the soluble level, no further reaction occurs; A and A_2 remain free to equilibrate in solution. This is supported by the DLLS measurements which have shown that at protein concentrations less than the solubility level dictated by the salt concentration, only monomers and dimers exist; further, supernatant samples from precipitations indicate that the soluble protein is also primarily composed of monomeric and dimeric species. Currently, however, there is no direct evidence that soluble α CT remains in the native state.

In this scheme, the conformational change creates protein that is competent for aggregation. The altered dimer, $\{B_2\}_1$, is the primary aggregable species. Tellam and Winzor (1977) found that the polymerization of α CT at low ionic strengths in the neighborhood of pH 8 is best described by the isodesmic self-association of dimers. In Figure 6, the aggregation rate constant (or kernel) k_3 is considered to be constant. This is formally equivalent to assuming a linear polymerization model where each interacting species has only two reactive sites (Ziff, 1984); these sites likely correspond to the location(s) of the conformational change.

The stopped-flow experiments with TPCK- α CT shown in Figure 3 imply that not only is dimerization a crucial step in α CT aggregation, but that rendering a portion of the α CT incapable of dimerization results in a poisoning effect. One interpretation of this phenomena is that the rate of monomer addition to dimer aggregates is significant, consuming aggregable sites. In the model in Figure 6, altered monomer, B , is allowed to interact with altered dimer aggregates with the same intrinsic kinetic constant, k_3 , as the dimer association. In order for the mixed multimeric species $\{B\{B_2\}_i\}$ to rejoin the dimeric aggregate pool $\{B_2\}_i$, another molecule B must interact with the mixed aggregate via a dimerization reaction. The rate constant k_2 is assigned to this dimerization and is distinct from the rate constant k_{+d} linking the native species. This is consistent with the assumption of independent sites for dimerization and self-association. It was further assumed that the essence of the monomer addition rate could be adequately described by a single addition event.

The aggregative reactions involving k_2 and k_3 are considered to be irreversible. Investigations of aggregation phenomena by Brown and Glatz (1986) and Brakalov (1987) have indicated that the dominant fragmentation mech-

anisms are hydrodynamically driven. In the stopped-flow experiments, high shear flow fields exist primarily in the mixing tee and are transient; the flow cuvette contents have no significant velocity gradients during the course of a measurement.

5.5.2 Kinetic Rate Expressions

Rate expressions were derived with mass action kinetics for each of the species considered in the kinetics scheme of Figure 6 and formulated in terms of a detailed population balance:

$$\dot{A} = -k_1(1 - s^{-1}(t))A - k_{+d}A^2 + k_{-d}A_2, \quad (7)$$

$$\dot{A}_2 = -k_1(1 - s^{-1}(t))A_2 + \frac{1}{2}k_{+d}A^2 - \frac{1}{2}k_{-d}A_2, \quad (8)$$

$$\begin{aligned} \dot{B} = & k_1(1 - s^{-1}(t))A - k_2B^2 - B \sum_{i=1}^{\infty} k'_3\{B_2\}_i \\ & - k_2B \sum_{i=1}^{\infty} \{B\{B_2\}_i\}, \end{aligned} \quad (9)$$

$$\begin{aligned} \{\dot{B}_2\}_i = & \left(k_1(1 - s^{-1}(t))A_2 + \frac{1}{2}k_2B^2\right)\delta[i - 1] + \frac{1}{2} \sum_{j=1}^{i-1} k'_3\{B_2\}_j\{B_2\}_{i-j} \\ & - \{B_2\}_i \sum_{j=1}^{\infty} k'_3\{B_2\}_j - k'_3B\{B_2\}_i \\ & + k_2B\{B\{B_2\}_{i-1}\}H[i - 2], \end{aligned} \quad (10)$$

$$\{B\{\dot{B}\}_i\} = k'_3B\{B_2\}_i - k_2B\{B\{B_2\}_i\} \quad (11)$$

where $s(t)$ is the supersaturation ratio as a function of time, $\delta[i]$ is the Kronecker delta function and $H[i]$ is the Heaviside unit step function. Equation (10) resembles Smoluchowski's population balance equation for indefinite association with time varying source and sink terms (Hendriks and Ziff, 1985).

The supersaturation ratio dependence of the conformational change terms appearing in equations (7) through (10) arises from the following considerations.

If the protein in the native state is lumped and the conformational change is first order in the supersaturation concentration, then

$$\dot{A} + 2\dot{A}_2 = -k_1(A + 2A_2 - (A_s + 2A_{2s})) \quad (12)$$

where A_s and A_{2s} are the soluble, equilibrium concentrations of native monomer and dimer, respectively. However, analogous terms in the individual equations for \dot{A} and \dot{A}_2 are not valid as A_s and A_{2s} are functions of time and as the protein solubility approaches the initial protein concentration, it is possible that $A(0) < A_s(0)$ in the stopped-flow experiments. This anomaly is removed if the conformational change terms are expressed as functions of the supersaturation ratio. Equation (12) may be rewritten in terms of $s(t)$ as

$$\dot{A} + 2\dot{A}_2 = -k_1(1 - s^{-1}(t))(A + 2A_2). \quad (13)$$

The conformational change terms in equations (7) through (10) are assumed to have the same dependence on $s(t)$. $s(t)$ is obtained by dividing equation (13) by $A_s + 2A_{2s}$ and integrating, thus

$$s(t) = 1 + (s_o - 1)e^{-k_1 t}. \quad (14)$$

The aggregation rate constant k'_3 was expanded in terms of the number of possible ways for particles of α CT to interact with each other. Assuming each altered α CT molecule has one site for aggregative interaction requires

$$k'_3 = \begin{cases} 4k_3, & \text{for dimer-dimer association and} \\ 2k_3, & \text{for monomer-dimer association.} \end{cases} \quad (15)$$

k_3 is the intrinsic association constant between aggregation sites as a result of this explicit probabilistic definition (Ziff, 1980).

5.5.3 Method of Solution

Equations (10) and (11) give rise to an infinite set of coupled differential equations. This set of equations can be closed, however, if they are transformed in terms of moments. (Ziff, 1984). The n^{th} moments of the distributions of $\{B_2\}_i$ and $\{B\{B_2\}_i\}$ are defined as

$$\phi_n = \sum_{i=1}^{\infty} i^n \{B_2\}_i \quad (16)$$

and

$$\psi_n = \sum_{i=1}^{\infty} i^n \{B\{B_2\}_i\}, \quad (17)$$

respectively. Thus, equations (10) and (11) may be recast as

$$\begin{aligned} \dot{\phi}_n = & k_1(1 - s^{-1})A_2 + \frac{1}{2}k_2B^2 + 2k_3 \sum_{\nu=0}^n \binom{n}{\nu} \phi_{\nu} \phi_{n-\nu} - 4k_3\phi_n\phi_0 \\ & + k_2B \sum_{\nu=0}^n \binom{n}{\nu} \psi_{\nu} - k_3B\phi_n \end{aligned} \quad (18)$$

and

$$\dot{\psi}_n = 2k_3B\phi_n - k_2B\psi_n, \quad (19)$$

respectively, where the binomial coefficient $\binom{n}{\nu}$ is given by

$$\binom{n}{\nu} = \frac{n!}{\nu!(n-\nu)!}. \quad (20)$$

Stopped-flow turbidity data gives the weight average molecular weight of the particles (see equation (1)) as a function of time. In terms of the species included in the model, the reduced weight-average molecular weight is

$$\frac{\bar{M}_w}{M_1} = \frac{A + 4A_2 + B + 4\phi_2 + 4\psi_2 + 4\psi_1 + \psi_0}{A + 2A_2 + B + 2\phi_1 + 2\psi_1 + \psi_0}. \quad (21)$$

By conservation of mass, the denominator in equation (21) may be replaced by $(A_o + 2A_{2o})$, the total initial concentration of native protein. Since the

aggregation kernel k'_3 is constant, only the first three moments ϕ_n and ψ_n are required to compute \bar{M}_w .

To solve for $\bar{M}_w(t)$, the equations for \dot{A} , \dot{A}_2 , \dot{B} , $\dot{\phi}_n$, and $\dot{\psi}_n$ where $n = 0, 1, 2$ must be solved simultaneously. By integrating equation (13), A may be substituted by a function of A_2 and the differential equation for \dot{A} can be eliminated. \dot{A}_2 becomes

$$\dot{A}_2 = \frac{1}{2}k_{+d}\left(\frac{S}{M_1}s\right)^2 - \left(k_1(1-s^{-1}) + \frac{1}{2}k_{-d} + 2k_{+d}\frac{S}{M_1}s\right)A_2 + 2k_{+d}A_2^2 \quad (22)$$

where S is the protein solubility in mg mL^{-1} . The mass balance may be solved for B to eliminate the equation for \dot{B} ; since $B(0) = \phi_n(0) = \psi_n(0) = 0$,

$$B = \frac{S}{M_1}(s_o - s) - 2\phi_1 - 2\psi_1 - \phi_0. \quad (23)$$

Thus, the simultaneous solution of equations (18) and (19) for $n = 0, 1, 2$ and equation (22) given the substitution for B in equation (23) is sufficient to determine $\bar{M}_w(t)$.

The goal of the modelling work was to use the turbidity data to arrive at reasonable estimates for rate constants k_1 , k_2 , and k_3 . The nonlinear least squares Marquardt algorithm MRQMIN (Press et al., 1986) were used to fit these rate constants via equation (21) to the turbidity trajectories (see equation (1)). The convergence criterion was based on a weighted sum of square deviations (χ^2) objective function with weights inversely proportional to the magnitude of the data; this allowed more uniform fits as the molecular weight data span several orders of magnitude (Beckman and Farmer, 1987). Solution of Equation (21) was achieved at each iteration of the Marquardt algorithm by the numerical integration of the corresponding rate equations. These equations are quite stiff (Gear, 1971); the fitted rate constants control the stiffness, complicating the

numerical solution. A predictor-corrector algorithm, DDASSL (Petzold, 1983), was used to perform the integrations. At each step in the integration, the species concentrations were checked for negative values. The concentration of B was also monitored as it is computed implicitly in the integration via equation (25). Negative excursions on the order of the error tolerance in the integration scheme were allowed. Turbidity data between the mixing transient and the $A_{488} = 1.26$ cut-off were used for fitting purposes. The combined optimization-integration routine was implemented in VAX-11 FORTRAN on a VAX 11/780 mainframe computer; the optimization routine was in single precision and the integration routine was in double precision.

5.5.4 Modelling Results

Fits of the population balance model to the turbidity trajectories as a function of salt concentration, temperature and protein concentration are given in Figures 1 through 3 for KSCN-induced precipitations. The three parameter model was able to describe the experimental data closely in all cases except at low temperatures where the shape of the trajectory no longer appears sigmoidal. The fitted rate constants were somewhat scattered. Difficulties in the numerical integration due to the stiffness of the equation set were compounded in the optimization by the extremely "shallow" nature of the χ^2 hypersurface ($\frac{\partial \chi^2}{\partial k_i} \ll 1$). The fitted parameters exhibited a dependence on the initial guesses and specifying a smaller convergence criterion resulted in diverging estimates. Table 2 gives a summary of the estimated rate constants k_1 , k_2 , and k_3 averaged for each salt type; in this compilation, obvious outliers were discarded. For all the salts, the nonspecific interaction constant k_3 was larger than the specific interaction constant k_2 . In general, there are no significant differences between

the estimates for the salts studied.

The scattered nature of the fitted rate constants precludes an analysis of variations with salt concentration, temperature and protein concentration. No significant trends were seen for a given salt. This implies that on a gross scale the model does a reasonable job of accounting for the dependence of the aggregation on these parameters. The salt concentration effects are expressed through the solubility. The activation energies of the reactions governed by k_1 , k_2 , and k_3 must be low; the temperature dependence of the aggregative constants may be largely due to solubility effects as seen with the apparent Smoluchowski constants, k_{app} . The absence of trends with the protein concentration indicates that the molecularity of the reactions has been correctly rendered.

A simulation of the TPCK- α CT data was undertaken to examine the veracity of the proposed monomer addition mechanism. TPCK- α CT was shown to have a poisoning effect on the aggregation kinetics. In the polymerization termination reactions given in Figure 6, inhibited native monomer, A^* , undergoes the same conformational change as A and A_2 . The same kinetic constant, k_1 is employed as we have assumed that the dimerization (active) site and the structural change site are distinct. Altered, inhibited monomer, B^* , acts by removing altered dimers from the pool of aggregable protein; the nonspecific association rate constant k_3 has been used as interaction of B^* with $\{B_2\}_i$ presumably occurs via the same aggregation site. A single addition event is used to be consistent with the monomer addition mechanism proposed for the uninhibited enzyme. The $\{B^*\{B_2\}_i\}$ species represents the terminal aggregate.

Species balances for B^* and $\{B^*\{B_2\}_i\}$ are given by

$$\dot{B}^* = k_1 \frac{A_o^*}{s_o} (s - 1) - 2k_3 B^* \phi_0 \quad (24)$$

and

$$\dot{\psi}_n^* = 2k_3 B^* \phi_n \quad (25)$$

where

$$\psi_n^* = \sum_{i=1}^{\infty} i^n \{B^* \{B_2\}_i\}. \quad (26)$$

The supersaturation ratio is now a function of $A(0)$, $A_2(0)$, and $A^*(0)$. The equation for $\dot{\phi}_n$ must also be modified to reflect the depletion of aggregable dimers

$$\begin{aligned} \dot{\phi}_n = & k_1(1 - s^{-1})A_2 + \frac{1}{2}k_2 B^2 2k_3 \sum_{\nu=0}^n \binom{n}{\nu} \phi_\nu \phi_{n-\nu} - 4k_3 \phi_n \phi_0 \\ & + k_2 B \sum_{\nu=0}^n \binom{n}{\nu} \psi_\nu - k_3(B + B^*)\phi_n. \end{aligned} \quad (27)$$

Inclusion of the inhibited and terminal aggregate species in the weight-average molecular weight equation (21) gives

$$\frac{\bar{M}_w}{M_1} = \frac{A + 4A_2 + A^* + B + B^* + 4\phi_2 + 4\psi_2 + 4\psi_1 + \psi_0 + 4\psi_2^* + 4\psi_1^* + \psi_0^*}{A + 2A_2 + A^* + B + B^* + 2\phi_1 + 2\psi_1 + \psi_0 + 2\psi_1^* + \psi_0^*} \quad (28)$$

where by conservation of mass, the denominator may be replaced by $(A(0) + 2A_2(0) + A^*(0))$.

The set of 11 differential equations was solved numerically using the average values of k_1 , k_2 , and k_3 fit to the control data sets with the equivalent total concentration of uninhibited protein. The simulated turbidity trajectory is presented with the experimental TPCK- α CT data in Figure 3. The simulated turbidity curve increases very slightly. If the constant turbidity background of the TPCK- α CT experiments (which was present before the addition of salt) is subtracted, the simulation is able to adequately describe the poisoning effect. The simulated curve increases somewhat more slowly, however; this may be an inherent limitation of the assumption of a single monomer addition mechanism.

5.6 Discussion

The kinetics of salt-induced α CT aggregation and precipitation have been examined as a function of salt type, salt concentration, temperature, and protein concentration. The linear portions of the turbidity trajectories were analyzed in terms of Smoluchowski kinetics via equation (4). All of the parameters studied affect the initial protein supersaturation ratio. The apparent Smoluchowski aggregation rate constant was invariant with respect to the protein concentration, supporting the bimolecular collision mechanism. Increasing temperatures and salt concentrations had similar effects on k_{app} ; the activation energy of the association is small. Pure Smoluchowski kinetics are incapable of describing the initiation or nucleation events preceding the bulk aggregation process, however (Parker and Dalgleish, 1977). This initial stage corresponds to the kinetics of the salt-water-protein interaction responsible for the creation of the condition of supersaturation. Experiments with inhibited α CT indicated that specific interactions via dimerization are also important in aggregation.

α CT aggregation does not commence until the initial supersaturation ratio exceeds unity as evidenced by the DLLS measurements. α CT solutions at low salt concentrations at pH 3 and supernatant solutions at equilibrium were substantially free of aggregates larger than dimers. However, it is not clear if soluble protein is structurally perturbed at high salt concentrations and/or if the dimerization equilibrium constant is significantly changed. The DLLS determinations indicate merely that soluble protein is sufficiently hydrated to remain a mixture of monomers and dimers.

A detailed population balance model incorporating structural change and monomer-dimer interactions was developed in efforts to model both the initial transient and the bulk aggregation behavior of α CT. The initial step in the

proposed mechanism is the formation of aggregative sites via conformational change. Secondary structure analysis of α CT precipitates indicated that precipitates induced by chaotropic salts had increased β -sheet contents. The locations of these structural changes may serve as sites for intermolecular association through hydrogen bonding; a more definitive determination of the temperature dependence of the association constants may provide energetic evidence to support this hypothesis. This phenomenon appears to be widespread in aggregating protein systems (Przybycien and Bailey, manuscript in preparation; Nemoto et al., 1983; Clark et al., 1981). The conformation of α CT has been shown to be a function of the extent of hydration (Lüscher and Rüegg, 1978). Yet, the causative relationship between electrolyte, water structure, and protein conformation has not been established; salt addition may disrupt the water structure with the resulting competition for water of solvation driving the conformational change or the interactions between charged groups on the protein surface may be disrupted by the charge screening effect of the salt ions, allowing conformational relaxation (von Hippel and Schleich, 1969).

The primary aggregable species in the model is the structurally altered dimer. α CT has been shown to aggregate by the reversible association of dimers at low ionic strength near the pH optimum (Tellam and Winzor, 1977); the TPCK- α CT turbidity curves corroborate the importance of dimerization in salt-induced aggregation at pH 3. The importance of pre-existing specific interaction sites in addition to the presumed non-specific sites formed during aggregation has been demonstrated in the polymerization of insulin (Jeffrey et al., 1976) and β -lactoglobulin (Casal et al., 1988).

Monomer addition was also explicitly included in the kinetic scheme. Aggregation at uninhibited α CT concentrations equivalent to that present in the

TPCK- α CT solution was much more rapid than the inhibited solution at the same salt concentration and temperature. Thus, the inhibited monomer must prematurely terminate the polymerization of the uninhibited species; monomer addition to aggregate species is significant. The single addition mechanism employed in the model may be inadequate to describe the full extent of monomer addition as evidenced by the enhanced level of poisoning in the simulation in comparison with the experimental data of Figure 3.

The values estimated for the rate constants k_1 , k_2 , and k_3 in Table 2 are in good qualitative agreement with other values reported in the literature. Rate constants for pH and thermally-induced β -sheet formation in homopolypeptides range from 10^{-2} to 10^2 s^{-1} (Maeda et al., 1987). Smoluchowski rate constants between 4×10^4 and 2.5×10^7 $M^{-1}s^{-1}$ have been determined for aggregating casein micelles. k_2 and k_3 estimates lie within this range of magnitudes; however, the nonspecific interaction represented by k_3 was typically stronger than the specific interaction represented by k_2 . At the current level of resolution, the significance of the variations in the fitted rate constants from salt to salt is not clear.

The constant kernel used in the model is a first approximation to the size dependence of aggregative interactions. Shimada and Matsushita (1980) have identified α CT as a gelation type as opposed to a coagulation type protein. The constant kernel is incapable of predicting the formation of an infinite mass or gel of protein. Leyrvaz (1984) has identified the particle size dependence of the aggregation kernel sufficient for a solution to exist for all time and to exhibit gelation behavior as

$$k'_3(i, j) \sim (ij)^\omega \quad \text{where} \quad \frac{1}{2} < \omega \leq 1 \quad (29)$$

for the association of an i -mer with a j -mer. If we generalize our model to allow for nonspecific protein interaction at f sites, k'_3 may be given by (Ziff et al., 1982)

$$k'_3 = k_3((f-2)i+2)((f-2)j+2) \quad (30)$$

which satisfies the minimum requirements for gelation described in expression (29). Other aggregation kernels accounting for various physical aspects of associating particle systems have been described (Beckman and Farmer, 1987; Drake, 1972). Note that for the more complicated kernels, the $i - j$ dependence may result in moment equations which are no longer closed; Hulbert and Katz (1964) have suggested expanding higher order moments ($n > 2$) in terms of lower order moments in a truncated Laguerre series as an approximate solution to this difficulty. However, the convergence properties of the parameter estimation routine must be improved before adding rigor to the model.

This model should find application to aggregating protein systems in which nonspecific associations are supplemented by specific interactions. This has been shown for α CT, insulin and β -lactoglobulin and is likely the case for oligomeric proteins. In the absence of specific interactions, analytical solution of the model equations for the constant aggregation kernel is possible; the reduced weight-average molecular weight is given by

$$\frac{\bar{M}_w}{M_1} = 1 + \frac{k_3 A_o}{k_1} ((1 - e^{-2k_1 t}) + 4(1 - e^{-k_1 t}) + 2k_1 t). \quad (31)$$

As a greater understanding of the relationships between process parameters and the intrinsic kinetic events in protein association is gained, the ultimate goal of the *a priori* specification of aggregative (or nonaggregative) environmental conditions may be realized.

5.7 Acknowledgements

The authors are indebted to Dr. Folim Halaka of the Monsanto Corporate Research Laboratories for providing access to the DLLS apparatus and for performing the data inversions. This research was supported by the National Science Foundation.

5.8 Nomenclature

| | |
|--------------------|--|
| A | native monomer species (concentration), (M) |
| A_2 | native dimer species (concentration), (M) |
| A_λ | optical absorbance at wavelength λ , AU |
| A^* | inhibited native monomer species (concentration), (M) |
| B | altered monomer species (concentration), (M) |
| B^* | inhibited, altered monomer species (concentration), (M) |
| $\{B_2\}_i$ | altered dimer i -meric aggregate species (concentration), (M) |
| $\{B\{B_2\}_i\}$ | altered mixed i -meric aggregate species (concentration), (M) |
| $\{B^*\{B_2\}_i\}$ | poisoned aggregate species (concentration), (M) |
| C_o | total initial protein concentration, mg mL ⁻¹ |
| $H[i]$ | Heaviside unit step function |
| $\Delta\bar{H}_o$ | α CT molal heat of solution, cal mol ⁻¹ |
| k_{app} | apparent Smoluchowski rate constant, M ⁻¹ s ⁻¹ |
| k_1 | conformational change rate constant, s ⁻¹ |
| k_2 | aggregative dimerization rate constant, M ⁻¹ s ⁻¹ |
| k_3 | intrinsic dimer aggregation rate constant, M ⁻¹ s ⁻¹ |
| k'_3 | dimer aggregation rate constant, M ⁻¹ s ⁻¹ |
| k_{+d} | native monomer association rate constant, M ⁻¹ s ⁻¹ |
| k_{-d} | native dimer dissociation rate constant, s ⁻¹ |
| $K(\lambda)$ | optical constant at wavelength λ , mL g ⁻¹ |
| K_D | native α CT dimerization equilibrium constant, M ⁻¹ |
| K_s | salting-out constant, molality ⁻¹ |
| m | salt concentration, molality |
| \bar{M}_n | solute number-average molecular weight, Daltons |
| \bar{M}_w | solute weight-average molecular weight, Daltons |

| | |
|-------------|--|
| M_1 | solute monomeric molecular weight, Daltons |
| N | solute particle number concentration, M |
| \bar{r}_w | solute weight-average particle radius, Å |
| \bar{r}_z | solute z-average particle radius, Å |
| r_1 | solute monomeric radius, Å |
| s | supersaturation ratio, dimensionless |
| s_o | initial supersaturation ratio, dimensionless |
| S | protein solubility, mg mL ⁻¹ |
| T | temperature, Kelvin |

Greek symbols

| | |
|-------------|--|
| β' | salting-out equation constant, dimensionless |
| $\delta[i]$ | Kronecker delta function |
| σ_w | standard deviation of solute weight-average radius distribution, Å |
| ϕ_n | n^{th} moment of dimer aggregate distribution, M |
| ψ_n | n^{th} moment of mixed aggregate distribution, M |
| ψ_n^* | n^{th} moment of poisoned aggregate distribution, M |

5.9 References

- Agapow, P-M. and Winzor, D.J., "Allowance for Effects of Electrostatic Repulsion on Protein Dimerization," *Biochem. Biophys. Acta* **953**, 197 (1988).
- Aune, K.C. and Timasheff, S.N., "Dimerization of α -Chymotrypsin. I. pH Dependence in the Acid Region," *Biochemistry* **10**, 1609 (1976).
- Aune, K.C., Goldsmith, L.C. and Timasheff, S.N., "Dimerization of α -Chymotrypsin. II. Ionic Strength and Temperature Dependence," *Biochemistry* **10**, 1617 (1971).
- Auer, H.E. and Miller-Auer, H., "Dynamics of the Disordered - Beta Transition in Poly-L-Tyrosine Determined by Stopped-Flow Spectrometry," *Biopolymers* **25**, 1607, (1986).
- Bailey, J.E. and Ollis, D.F., *Biochemical Engineering Fundamentals*, 2nd ed., McGraw-Hill, New York, 727 (1986).
- Beckman, J.R. and Farmer, R.W., "Bimodal CSD Barite Due to Agglomeration in an MSMPR Crystallizer," *AIChE Symp. Ser. 253* **83** 85 (1987).
- Bell, D.J., Hoare, M. and Dunnill, P., "The Formation of Protein Precipitates and Their Centrifugal Recovery," *Adv. in Biochemical Eng.* **26**, 1 (1983).
- Brakalov, L.B., "A Connection Between the Orthokinetic Coagulation Capture Efficiency of Aggregates and Their Maximum Size," *Chem. Eng. Sci.* **42**, 2373 (1987).
- Brown, D.L. and Glatz, C.E., "Aggregate Breakage in Protein Precipitation," *Chem. Eng. Sci.* **42**, 1831 (1987).
- Cantor, C.R. and Schimmel, P.R., *Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function*, W.H. Freeman,

New York, 838 (1980).

Carlson, A., Hill, C.G. and Olson, N.F., "Kinetics of Milk Coagulation: II. Kinetics of the Secondary Phase: Micelle Flocculation," *Biotechnol. Bioeng.* **29**, 590 (1987a).

Carlson, A., Hill, C.G. and Olson, N.F., "Kinetics of Milk Coagulation: III. Mathematical Modelling of the Kinetics of Curd Formation Following Enzymatic Hydrolysis of κ -Casein - Parameter Estimation," *Biotechnol. Bioeng.* **29**, 601 (1987b)

Chan, M.Y.Y., Hoare, M. and Dunnill, P., "The Kinetics of Protein Precipitation by Different Reagents," *Biotechnol. Bioeng.* **28** 387 (1987).

Clark, A.H., Saunderson, D.H.P. and Suggett, A., "Infrared and Laser-Raman Spectroscopic Studies of Thermally-Induced Globular Protein Gels," *Int. J. Peptide Protein Res.* **17**, 353 (1981).

Cohn, E.J. and Edsall, J.T., *Proteins, Amino Acids and Peptides*, Reinhold, New York, 282 (1943).

Dalgleish, D.G., Brinkhuis, J. and Payens, T.A.J., "The Coagulation of Differently Sized Casein Micelles by Rennet," *Eur. J. Biochem.* **119**, 257 (1981).

Dalgleish, D.G., "A Mechanism for the Chymosin-Induced Flocculation of Casein Micelles," *Biophys. Chem.* **11**, 147 (1980).

Denbigh, K., *Principles of Chemical Equilibrium*, 4th ed., Cambridge University Press, New York, 266 (1981).

Drake, R.L., "A General Mathematical Survey of the Coagulation Equation," in *Topics in Current Aerosol Research*, vol. 3, part 2 (Hidy and Brock, Eds.) Pergamon Press, New York, 202 (1972).

Fisher, R.R. and Glatz, C.E., "Polyelectrolyte Precipitation of Proteins: I.

- The Effects of Reactor Conditions," *Biotechnol. Bioeng.* **32**, 777 (1988a).
- Fisher, R.R. and Glatz, C.E., "Polyelectrolyte Precipitation of Proteins: II. Models of the Particle Size Distributions," *Biotechnol. Bioeng.* **32**, 786 (1988b).
- Formisano, S., Johnson, M.L. and Edelhofer, H., "Effects of Hofmeister Salts on the Self-Association of Glucagon," *Biochemistry* **17**, 1468 (1978).
- Foster, P.R., Dunnill, P. and Lilly, M.D., "The Kinetics of Protein Salting-Out: Precipitation of Yeast Enzymes by Ammonium Sulfate," *Biotechnol. Bioeng.* **18**, 545 (1976).
- Gear, C.W., *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 209 (1971).
- Gardiner, S.A.M., Olbrich, R., Marston, F.A.O., Fish, N.M., and Hoare, M., "Process Design for Recombinant Protein Recovery from Inclusion Bodies," in *Proceedings of the 4th European Conference on Biotechnology* (M. Neijssel, R.R. van der Meer, and K.Ch.A.M. Luyben, Eds.) Elsevier, New York, p. 541 (1987).
- Gilleland, M.J. and Bender, M.L., "Kinetics of α -Chymotrypsin Dimerization," *J. Biol. Chem.* **251**, 498 (1976).
- Hegg, P-O., "Conditions for the Formation of Heat-Induced Gels of Some Globular Food Proteins," *J. Food. Sci.* **47**, 1241 (1982).
- Hendriks, E.M. and Ziff. R.M., "Coagulation in a Continuously Stirred Tank Reactor," *J. Coll. Int. Sci.* **105**, 247 (1985).
- Herdan, G., *Small Particle Statistics*, 2nd ed., Academic Press, New York, 264 (1960).
- Horbett, T.A. and Teller, D.C., "Low pH Dimerization Behavior of Active-Site Derivatives of Bovine α -Chymotrypsin," *Biochemistry* **12**, 1349 (1973).

- Hulbert, H.M. and Katz, S., "Some problems in Particle Technology, A Statistical Mechanical Formulation," *Chem. Eng. Sci.* **19**, 555 (1964).
- Ikeda, K., Kunugi, S., and Ise, N., "Specific and Non-Specific Effects of Inert Anions and Cations on the Dimerization of α -Chymotrypsin and the Catalytic Activities of Monomeric and Dimeric α -Chymotrypsins," *J. Biochem.* **91**, 347 (1982).
- Jeffrey, P.D., Milthorpe, B.K., and Nichol, L.W., "Polymerization Pattern of Insulin at pH 7.0," *Biochemistry* **15**, 4660 (1976).
- Kane, J.F. and Hartley, D.L., "Formation of Recombinant Protein Inclusion Bodies in *Escherichia coli*," *Trends. Biotechnol.* **6**, 95 (1988).
- Leyrvaz, F., "Critical Exponents in the Smoluchowski Equations of Coagulation," in *Kinetics of Aggregation and Gelation* (F. Family and D.P. Landau, Eds.) Elsevier, New York, 201 (1984).
- Lüscher, M. and Rüegg, M., "Thermodynamic Studies of the Interaction of α -Chymotrypsin with Water. I. Determination of the Isothermic Enthalpies and Entropies of Water Binding to Native Enzyme," *Biochim. Biophys. Acta* **533**, 428 (1978).
- Maeda, H., Fukada, K., and Ikeda, S., "Kinetics of Chain Folding of Model Polypeptides into the β -Structure," in *Ordering and Organisation in Ionic Solutions* (N. Ise and I. Suga Eds.) 193 (1987).
- Massey, E.H., Tensmeyer, L.G., and Sheliga, T.A., "Aggregation of Human Insulin (HI) in Isophane Suspension Formulations (NPH)", Symposium on Formulation Development and Drug Delivery Systems for Pharmaceutical Proteins, 196th ACS National Meeting, Los Angeles, September 26, 1988.
- Matthews, B.W., Sigler, P.B., Henderson, R., and Blow, D.M., "Three-dimensional Structure of Tosyl- α -chymotrypsin," *Nature* **214**, 652 (1967).

Neet, K.E. and Brydan, S.E., "Sedimentation of Chemically Modified Chymotrypsin," *Arch. Biochem. Biophys.* **136**, 223 (1970).

Nemoto, N., Kiyoharo, T., Tsunashima, Y., and Kurata, M., "Raman, IR and CD Spectroscopy on Aggregates and Gels of Lysozyme in Alcohols," *Bull. Inst. Chem. Res. (Kyoto)* **61**, 203 (1983).

Parker, T.G. and Dalgleish, D.G., "The Use of Light-Scattering and Turbidity Measurements to Study the Kinetics of Extensively Aggregating Proteins: α_s -Casein," *Biopolymers* **16**, 2533 (1977).

Petzold, L.R. "The DDASSL Differential/Algebraic System Solver," Sandia National Laboratory, (1983).

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., *Numerical Recipes, The Art of Scientific Computing*, Cambridge University Press, New York, pp. 521-528 (1986).

Provencher, S.W., "A Constrained Regularization Method for Inverting Data Represented by Linear Algebraic or Integral Equations," *Comput. Phys. Commun.* **27**, 213 (1982a).

Provencher, S.W., "CONTIN: A General Purpose Constrained Regularization Program for Inverting Noisy Linear Algebraic and Integral Equations," *Comput. Phys. Commun.* **27**, 229 (1982b).

Przybycien, T.M. and Bailey, J.E., "Solubility-Activity Relationships in the Inorganic Salt-Induced Precipitation of α -Chymotrypsin," *Enz. Microb. Tech.* (1989a) in press.

Przybycien, T.M. and Bailey, J.E., "Structure-Function Relationships in the Inorganic Salt-Induced Precipitation of α -Chymotrypsin," *Biochim. Biophys. Acta* (1989b) in press.

Shimada, K. and Matsushita, S., "Relationship between Thermocoagulation

of Proteins and Amino Acid Compositions," *J. Agric. Food Chem.* **28**, 413 (1980).

Schoellmann, G. and Shaw, E., "Direct Evidence for the Presence of Histidine in the Active Center of Chymotrypsin," *Biochemistry* **2**, 252 (1963).

Tellam, R. and Winzor, D.J., "Self-Association of α -Chymotrypsin at Low Ionic Strength in the Vicinity of its pH Optimum," *Biochem. J.* **161**, 687 (1977).

von Hippel, P.H. and Schleich, T., "The Effects of Neutral Salts on the Structure and Conformational Stability of Macromolecules in Solution," in *Structure and Stability of Biological Macromolecules*, (S.N. Timasheff and G.D. Fasman, Eds.) 421 (1969).

Walton, A.G., *The Formation and Properties of Precipitates*, Interscience, New York, 66 (1967).

Weast, R.C., . *CRC Handbook of Chemistry and Physics*, CRC Press, D-150 (1980).

Ziff, R.M., "Aggregation Kinetics via Smoluchowski's Equation," *Kinetics of Aggregation and Gelation* (F. Family and D.P. Landau, Eds.) Elsevier, 191 (1984).

Ziff, R.M., Hendriks, E.M., and Ernst, M.H., "Critical Properties for Gelation: A Kinetic Approach," *Phys. Rev. Lett.* **49**, 593 (1982).

Ziff, R.M., "Kinetics of Polymerization," *J. Stat. Phys.* **23**, 241 (1980).

5.10 Tables

Table 1. Molar heats of solution for α CT in various salt solutions

| salt | salt concentration (m) | $\Delta\bar{H}_o^a$ (cal mol ⁻¹) | α CT concentration (mg mL ⁻¹) |
|---------------------------------|------------------------|--|--|
| KSCN | 0.5 | -21500 ± 600 | 10 |
| KBr | 2.35 | -6800 ± 490 | 10 |
| NaBr | 2.6 | -6200 ± 310 | 10 |
| NaCl | 2.5 | -1990 ± 240 | 25 |
| Na ₂ SO ₄ | 1.0 | -2120 ± 330 | 25 |

a limits represent ± S.D. of the mean

Table 2 Rate constant estimates for population balance model

| salt | k_1^a (s ⁻¹) | N^b | k_2 (M ⁻¹ s ⁻¹) | N | k_3 (M ⁻¹ s ⁻¹) | N |
|---------------------------------|-----------------------------|-------|--|-----|--|-----|
| KSCN | 2.3 ± 5.5 × 10 ⁰ | 21 | 0.7 ± 1.4 × 10 ⁷ | 20 | 0.6 ± 1.3 × 10 ⁸ | 20 |
| KBr | 1.0 ± 1.3 × 10 ¹ | 11 | 1.4 ± 1.8 × 10 ⁴ | 15 | 7.6 ± 7.5 × 10 ⁴ | 15 |
| NaBr | 2.7 ± 2.3 × 10 ¹ | 10 | 3.9 ± 2.4 × 10 ⁴ | 13 | 1.1 ± 1.6 × 10 ⁶ | 12 |
| NaCl | 1.1 ± 1.5 × 10 ¹ | 7 | 3.3 ± 2.6 × 10 ⁴ | 7 | 1.7 ± 3.2 × 10 ⁶ | 7 |
| Na ₂ SO ₄ | 4.7 ± 4.6 × 10 ¹ | 5 | 3.0 ± 2.9 × 10 ⁴ | 5 | 0.8 ± 1.4 × 10 ⁷ | 5 |

a error limits represent ± S.D. of the mean

b number of data sets used in computing the mean, the total number fit was:

KSCN, 22; KBr, 15; NaBr, 13; NaCl, 8; Na₂SO₄, 5

5.11 Figures

Figure 1 α CT aggregation turbidity trajectories as a function of KSCN concentration. The protein concentration and temperature were fixed at 10 mg mL^{-1} and 25°C , respectively. The KSCN concentrations are noted on the figure. The experimental data is given by the solid lines; dashed lines represent the corresponding fits of the population balance model. The dotted horizontal line delimits the linear response regime.

Figure 2 α CT aggregation turbidity trajectories as a function of temperature. The protein concentration and KSCN concentration were held constant at 10 mg mL^{-1} and 0.5 m , respectively. The temperatures are noted on the figure. The solid lines represent the experimental data; dashed lines give the corresponding fits of the population balance model. The dotted horizontal line delimits the linear response regime.

Figure 3 α CT and TPCK- α CT aggregation turbidity trajectories as a function of protein concentration. The KSCN concentration and temperature were fixed at 0.5 m and 25°C , respectively. Experimental data for α CT is given by the solid lines and that for TPCK- α CT by the dotted lines. The protein concentrations are noted on the figure; experiments were performed in duplicate. Dashed lines give the corresponding fits of the population balance model. The dot-dash curve gives the result of the poisoning simulation using average rate constants fitted to the 3.57 mg mL^{-1} control runs. The dotted horizontal line delimits the linear response regime.

Figure 4 The apparent Smoluchowski constant, k_{app} , versus the initial su-

persaturation ratio, s_o , for each of the salts studied. The symbols representing the parameters used to manipulate s_o , temperature, salt concentration, and α CT concentration, are indentified in the figure legend. The fitted lines are given as visual aids.

Figure 5 Reduced solute weight-average particle radius, \bar{r}_w/\bar{r}_1 versus α CT concentration, C_o . The theoretical relationship expressed by equation (9) is given by the solid line; dashed lines show the monomer and dimer values of the reduced radius. Symbols give DLLS results for salt-free and supernatant α CT solutions. The heavy vertical bar at $C_o = 10 \text{ mg mL}^{-1}$ gives the range of reduced radius values found for low salt concentrations ($< 0.1 \text{ m}$) for each salt; the bar represents results of ten different samples. Error bar shows the standard deviation of the weight-average reduced radius distributions. $\bar{r}_1 \approx 19.6 \text{ \AA}$.

Figure 6 Detailed population balance model reaction scheme. Aggregative interactions are given by the upper scheme; the lower scheme depicts the poisoning mechanism.

Figure 1 α CT aggregation turbidity trajectories as a function of KSCN concentration

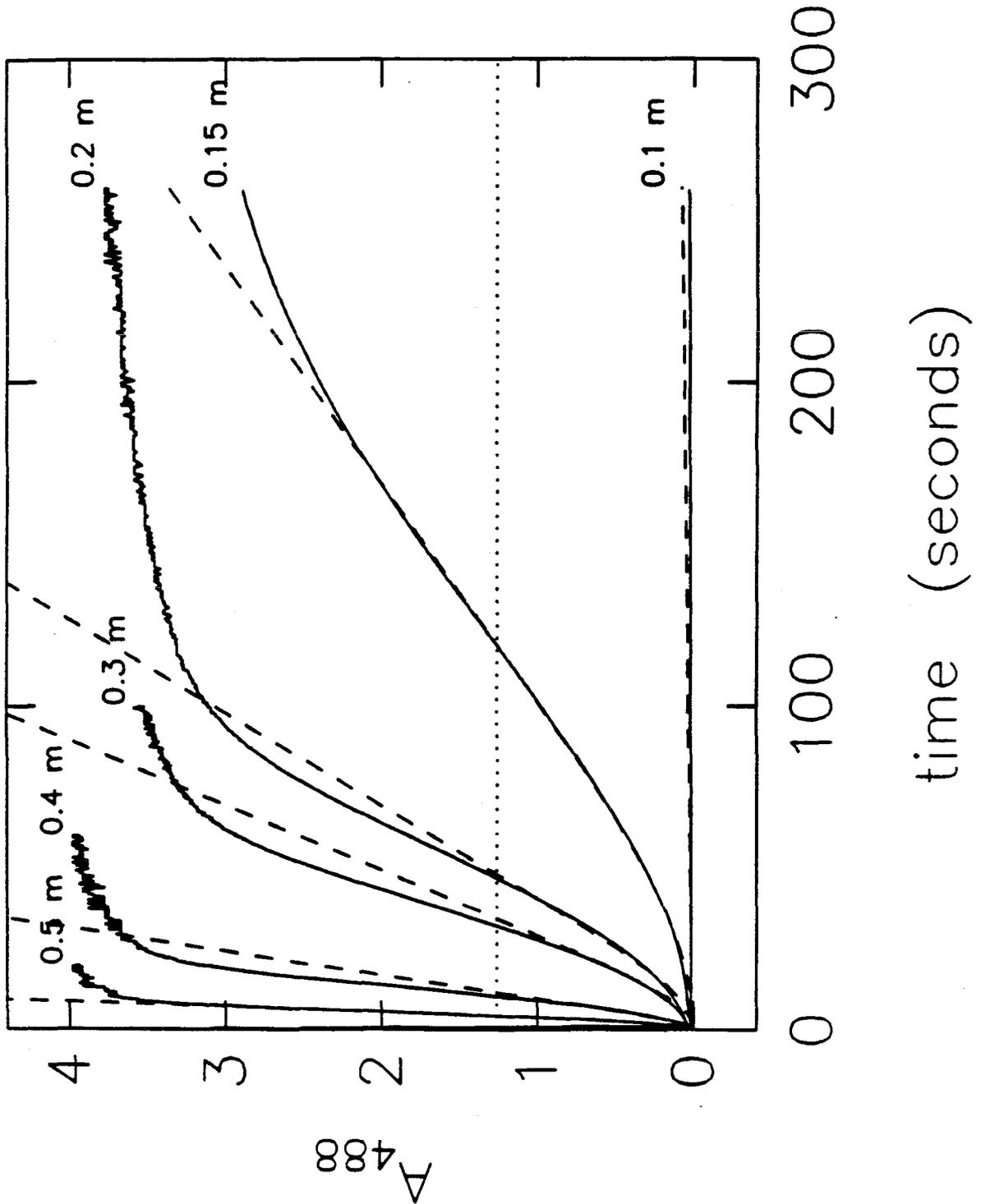


Figure 2 α CT aggregation turbidity trajectories as a function of temperature

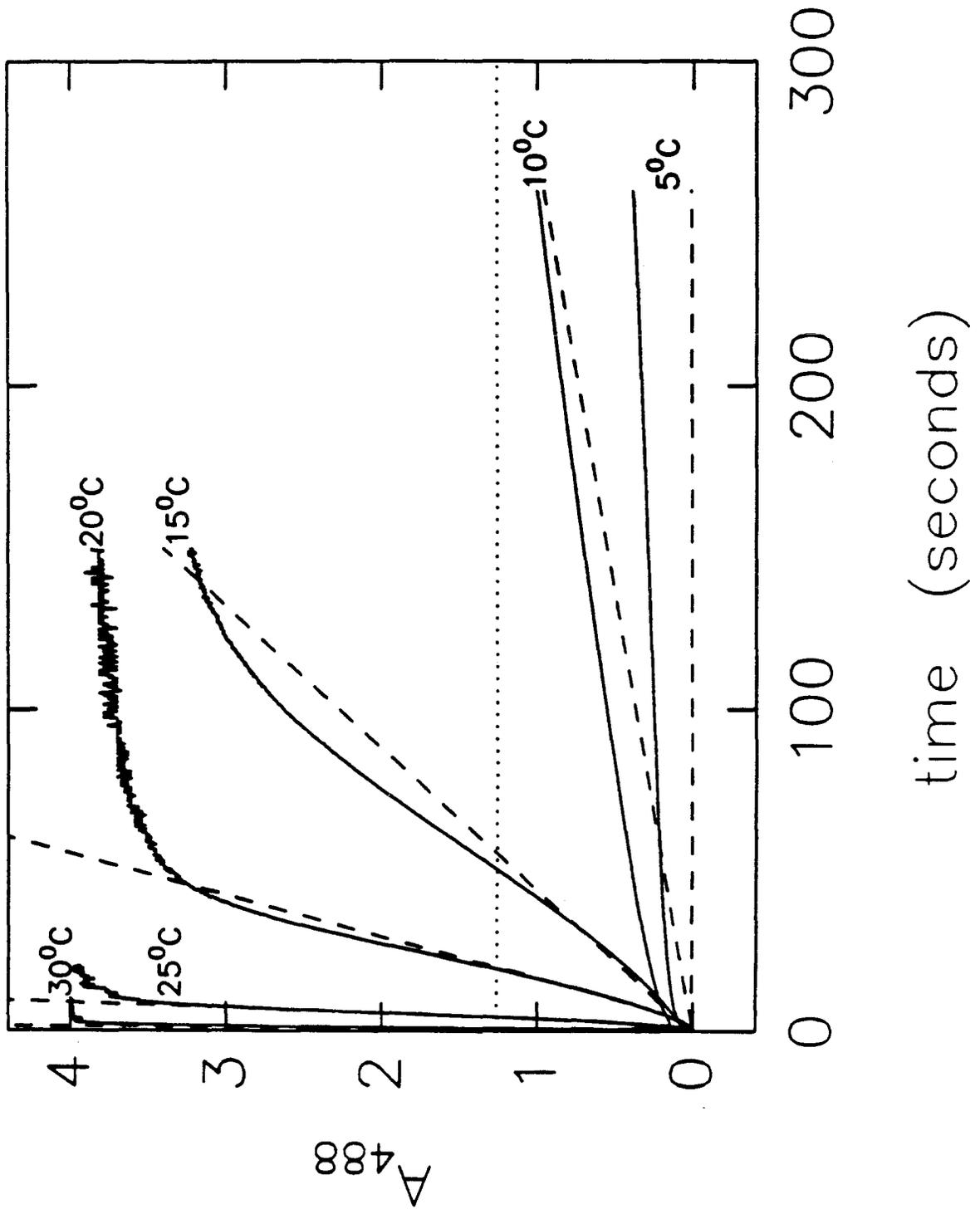


Figure 3 α CT and TPCK- α CT aggregation turbidity trajectories

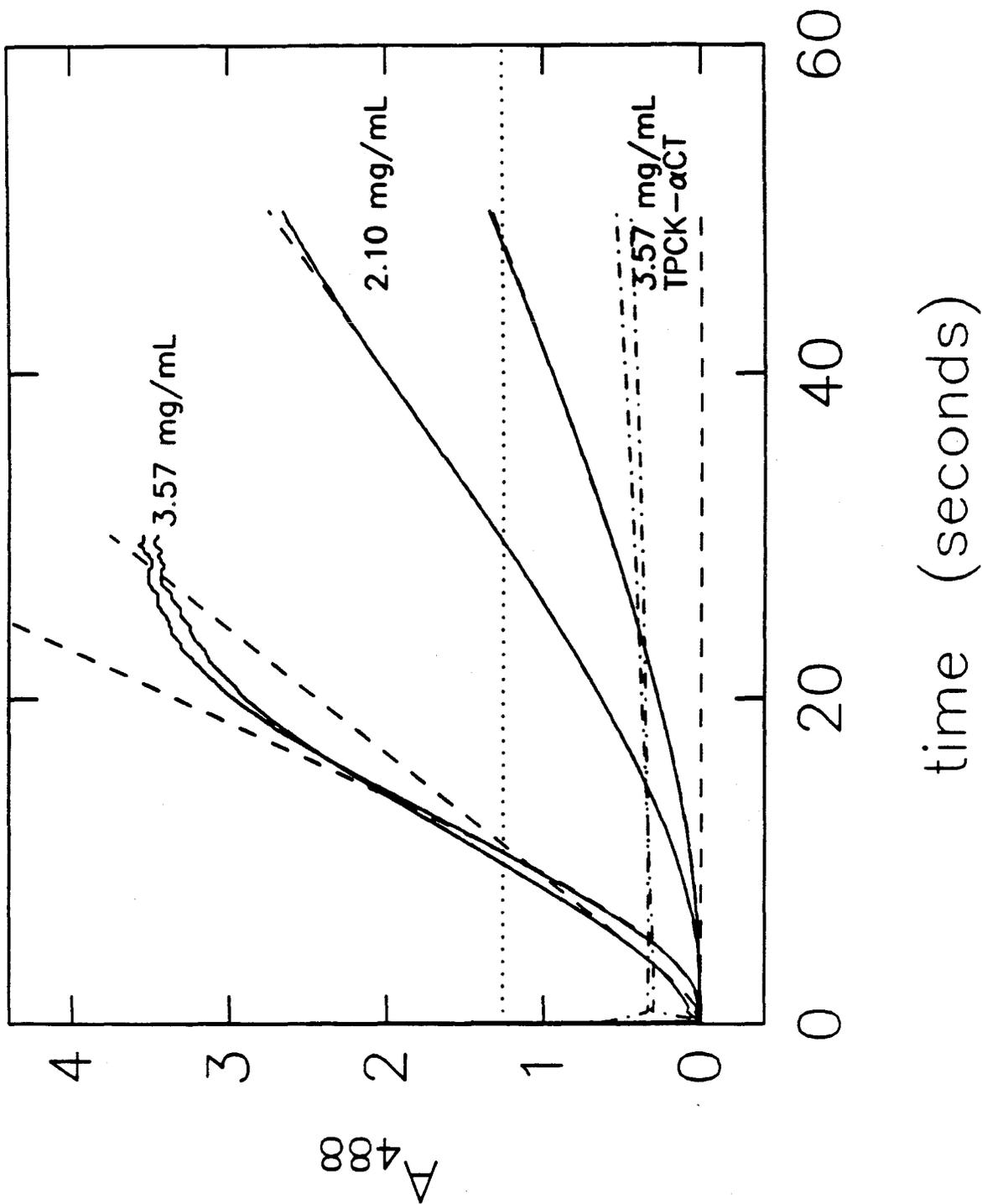


Figure 4 Apparent Smoluchowski rate constant versus initial supersaturation ratio

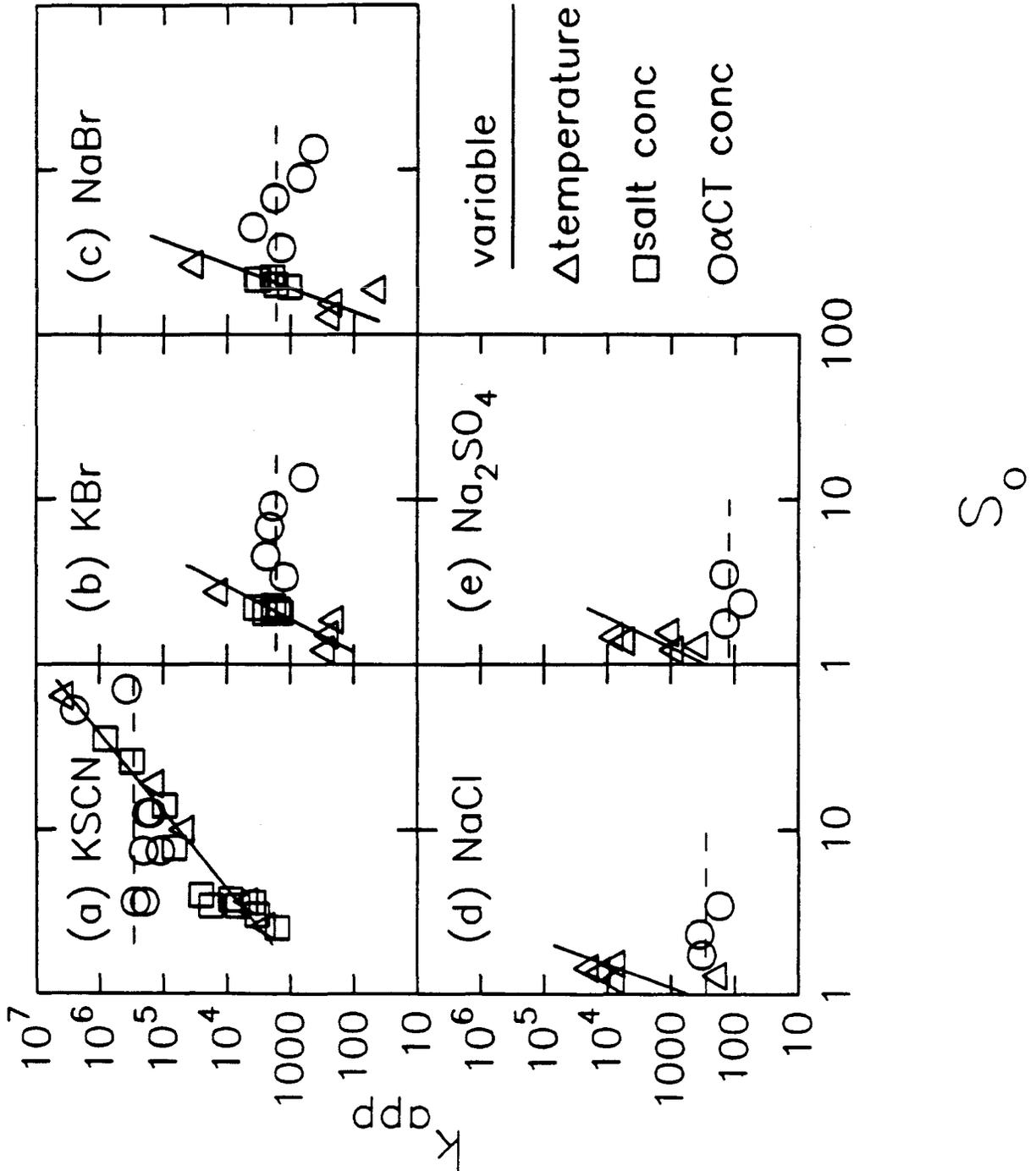


Figure 5 Reduced solute weight-average particle radius versus α CT concentration

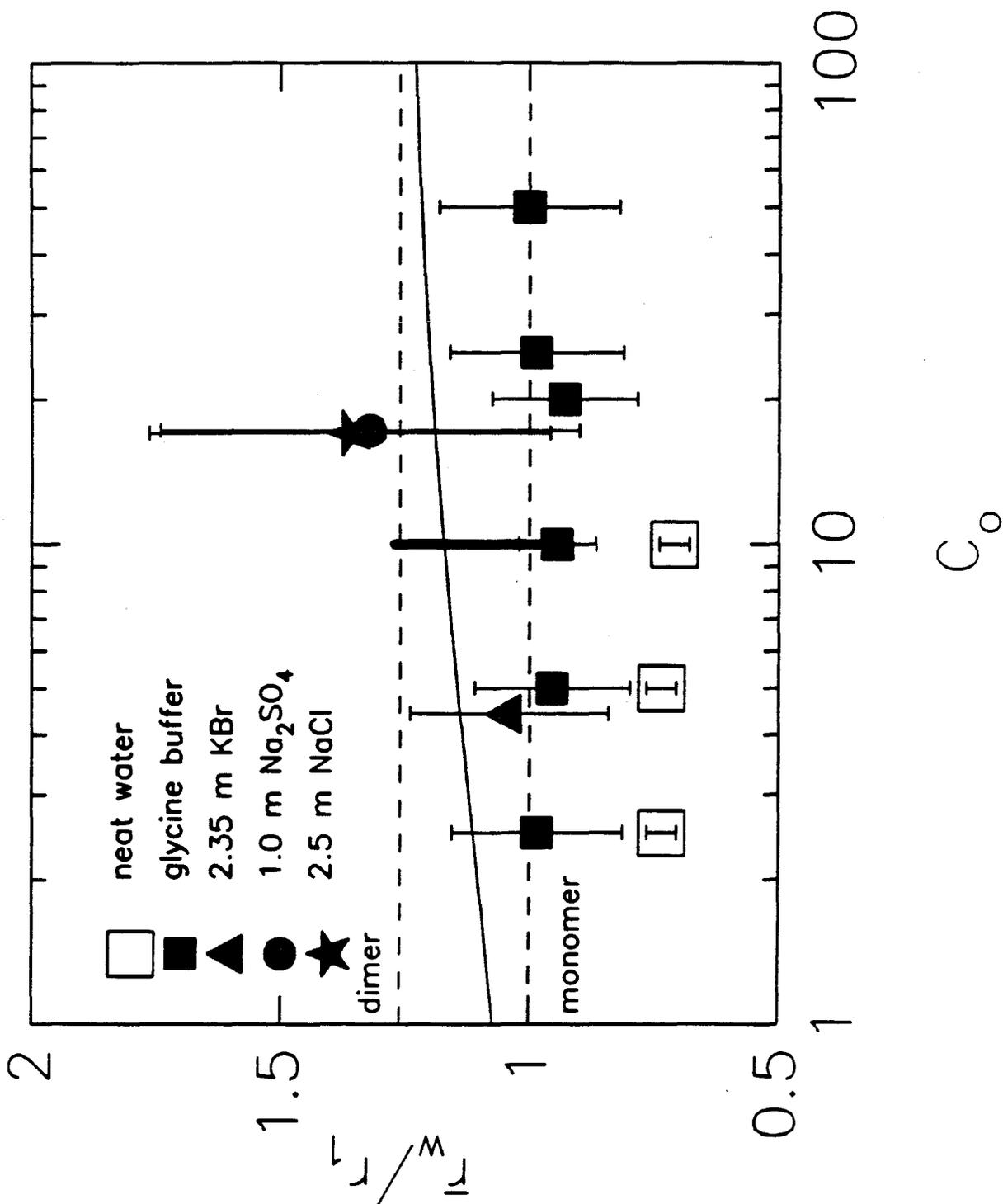
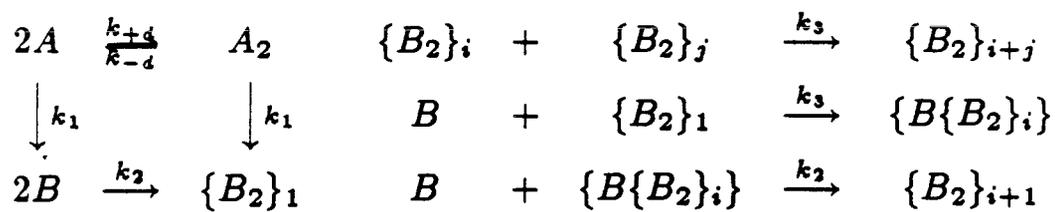


Figure 6 Detailed population balance model reaction scheme

Aggregation



Poisoning

Chapter 6

Conclusions

6.1 Summary of Results

The motivation for this work stems from the paucity of information pertaining to protein behavior, on the molecular level, in separations environments. Proteins differ from conventional chemical products in that they must have, in addition to the correct covalent structure, the right conformation. It is this added structural requirement that complicates biological separations processes. Design and operating guidelines are largely based on heuristics as the conformational implications of processing can not be predicted with certainty.

This work has addressed issues of protein activity, structure, and aggregation in salt-induced precipitation as a function of important process parameters. The insights gained pertaining to the salting-out process may also be cogent to other bioprocesses where high ionic strengths and/or aggregative environments are encountered; in aqueous two-phase extractions, for example, precipitates may form in salt phases or at the salt-polymer phase interface. The techniques used may find application in investigations of protein behavior in other processes as well.

In Chapter 2, the relationships between protein solubility and recoverable activity were explored as a function of salt type, salt concentration, and temperature using α CT as the model protein. The inability of the current equilibrium theory to correctly predict the solubility behavior of α CT may be attributed to either changes in protein physical properties, such as the dipole moment and hydrophobic surface area, due to structural perturbations, or specific salt-protein interactions not accounted for by the theory or both. There was no clear correlation between the discrepancies and salt properties such as activity or hydration number.

Activity losses were due to changes in the protein active fraction; specific

activity determinations indicated that protein that was able to recover activity on redissolution regained full native activity. The protein active fraction was sensitive to the salt type, but not the salt concentration. The salt concentrations required to cause precipitation in 1 to 10 mg mL⁻¹ α CT solutions are likely above some threshold value(s) where the concentration effects become saturated. The salts could be clearly divided into chaotropic (KSCN, KBr, NaBr) and structure-preserving (NaCl, Na₂SO₄) classes on the basis of their affect on the protein active fraction.

The product of the salting-out constant and the redissolved precipitate active fraction was indicative of the performance of a given salt. Na₂SO₄ was the best performer, while the bromide salts were the worst. The performance of KSCN rivaled that of NaCl; this was unexpected as these salts have different modes of interaction with proteins. Also, the performance was poorer at lower temperatures for each of the salts studied. These findings challenge the dogmatic adherence to traditional modes of operation employing non-denaturing salts and low temperatures; chaotropic salts and moderate temperatures may prove efficacious in some processes.

Chapter 3 is a companion study to Chapter 2; the effects of precipitation on elements of secondary and tertiary structure in α CT were probed. Raman amide I band spectra indicated that the α CT precipitates induced by chaotropic salts had increased β -sheet contents and decreased α -helix contents. This effect was greater for the insoluble portion of these precipitates. β -sheet formation has been found in other aggregating protein and peptide systems. These β -sheet structures may form the interaction sites for protein aggregates. The conformational changes depended only on the type of salt, not the concentration, mimicking the active fraction dependence found previously. As expected,

the correlation between the active fraction and the extent of structural change was strong. The inactivated protein may correspond to that which is extensively crosslinked by β -sheet interactions vis-à-vis the extent of the structural perturbations estimated for insoluble, inactive precipitates.

Conventional and saturation transfer electron paramagnetic resonance spectra of spin-labelled α CT precipitates indicated that the active site is not affected by precipitation. The active site forms the interface for dimerization in α CT; dimerization may protect the active site from denaturation. This complements the specific activity results and lends support to the conclusion that inactivation may be due to quaternary interactions in the precipitate phase rather than secondary structure changes.

A plausible molecular model was constructed for the structurally altered precipitate samples. The model was based on the secondary and tertiary information derived from the spectroscopic analyses. An estimation of the dipole moment and hydrophobic surface area of the altered structure was able to explain part of the discrepancy between the theoretical and experimental solubility behavior.

A general mechanism for the salt-induced precipitation of globular proteins, based on the solubility, activity, and structural data for α CT, was proposed. In the mechanism, the type of salt controls the extent of the structural perturbation while both the type and amount of salt control the equilibrium between precipitate and soluble protein. Since soluble, active precipitate was found to coexist with insoluble, inactive precipitate, the rates of the phase and conformational change reactions must be comparable. The extent to which α CT can be considered a true model protein in these investigations depends on the generality of the secondary structural perturbations found; evidence is given in Chapter 4

and in the work of other investigators (see references in Chapters 3 and 4) that β -sheet formation during aggregation is a widespread phenomenon.

Chapter 4 details the secondary structural implications of precipitation with KSCN and Na₂SO₄ for a variety of different proteins. The Raman amide I band analysis described in Chapter 3 was used. For the KSCN-induced precipitates, the β -sheet contents increased and the α -helix contents decreased; β -sheet formation may occur at the expense of α -helix. The structural changes estimated for the Na₂SO₄ precipitates were less significant.

The secondary structure changes were examined with respect to elements of primary, secondary, and tertiary structure found in the native proteins. The increase in β -sheet contents was positively correlated with the fraction of charged residues and negatively correlated with the solvent accessible surface area. The hydrophobic surface area may be the ultimate factor; the hydrophobic surface area is inversely proportional to the fraction of charged groups and is a nearly constant fraction of the solvent accessible surface area. The formation of β -sheet structures in protein aggregates mimics the subunit interactions of oligomeric proteins such as concanavalin A and alcohol dehydrogenase and may be the result of a secondary structure dependent hydration free energy.

The aggregation kinetics of α CT were described in Chapter 5. The process parameters studied included the type and amount of salt, the temperature, and the protein concentration. The progress of the aggregation reaction was monitored via stopped-flow turbidimetry. Apparent Smoluchowski rate constants were derived from the linear portion of the sigmoidal turbidity trajectories. The protein concentration had no significant effect on the apparent rate constants found for each of the salts; the protein concentration dependence of aggregation was adequately described by a collision type equation. The temperature and

salt concentration had nearly identical effects on the apparent rate constant. The temperature and salt concentration determine the protein solubility; the similarity of their effects on the apparent aggregation rate constant implies that the activation energy for the self-association process in precipitation is small.

Precipitations performed with partially inhibited α CT demonstrated the importance of the dimerization reaction and monomer addition to the aggregation process. The partially inhibited α CT aggregated much more slowly than a control precipitation reaction consisting of an amount of α CT equivalent to the concentration of the uninhibited enzyme. The inhibited monomer must poison the polymerization of the protein that remains competent for aggregation; the dimer is the primary aggregable species.

The particle size distributions of various α CT solutions were estimated by dynamic laser light scattering. α CT solutions from 1 to 50 mg mL⁻¹ were equilibrium mixtures of monomers and dimers. Both low salt- α CT solutions and supernatant solutions consisted essentially of monomers and dimers. Aggregation does not occur unless a condition of supersaturation exists; soluble protein is free of aggregates.

A detailed population balance model was developed to describe the aggregation behavior of α CT. This model incorporated the structural change effects discussed in Chapters 3 and 4, the specific interaction (dimerization) site, and monomer addition. The model was fit to the stopped flow data and rate constants for the conformational change, nonspecific aggregation, and aggregative dimerization were estimated. The fitted rate constants fell within reported ranges for analogous reactions of other proteins and peptides. Rate constants fit to the inhibition control experiments were able to model the qualitative behavior of the inhibited α CT aggregations. The model is unique in that it takes

structural perturbations, and specific and nonspecific quaternary interactions occurring in aggregation into account. The model may have general utility in the description of protein aggregative processes, especially where specific interaction sites must be considered.

6.2 New Directions

During the course of the present work, a number of areas meriting further investigation have become apparent. These areas include extensions of completed work such as the effects of mixing on protein solubility, improved secondary structure estimation via Raman spectroscopy, and a more rigorous treatment of aggregation kinetics. In addition, this work may be considered a case study under the banner "Protein Denaturation in Bioseparations" and used as a model approach to other bioprocesses where protein activity losses are a concern. A brief accounting of proposed research strategies is given below.

6.2.1 *Mixing Effects on Protein Solubility*

Although solubility is governed by thermodynamics, the solubility of proteins in solutions of high ionic strength depends on the contacting conditions between the salt and protein streams [1-3]. The solubility of a protein, S , as a function of the salt concentration, m , is given by the empirical relation

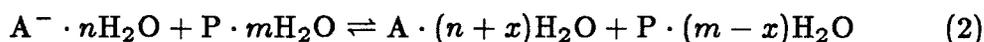
$$\ln S = \beta' - K_s m \quad (1)$$

where K_s is the salting-out constant which depends on the physical properties of both the salt and the protein and β' is a constant incorporating mixing effects as well as temperature and pH effects. For example, for KSCN and α -chymotrypsin at 25 °C, in the precipitations carried out via the protocol of Chapter 2, $K_s = 6.83 \pm 0.92 \text{ m}^{-1}$ and $\beta' = 2.64 \pm 0.23$; corresponding solubility

determinations performed with the stopped-flow device in Chapter 5 yielded $K_s = 6.76 \pm 0.41 \text{ m}^{-1}$ and $\beta' = 1.88 \pm 0.10$. Thus, the effect can be significant and appears to follow the form of equation (1).

The sensitivity of β' to mixing conditions has an obvious impact on the ability to scale-up precipitation operations correctly. In order to develop predictive ability for β' , the controlling process parameters must be identified. A semi-empirical approach linking turbulent mixing variables with the kinetics of the salt-protein interaction in precipitation is proposed.

The mechanism of the salt-induced protein solubility reduction is thought to involve the partial dehydration of the protein surface by the electrolyte anions. A possible representation of this action may be given by the equilibrium relationship [4]



where the salt anions, A^- , compete with the protein, P , for water of hydration. Some salts may compete more effectively than others; the salt type dependence of protein solubility may appear in both β' and K_s in equation (1). It is very likely that the kinetics of this equilibrium are rapid and, hence, may be a function of the mixing conditions [5,6]; the value of x in equation (2) may depend on the local salt concentration.

For turbulent mixing, the concentration of the i^{th} reactant, C_i , at any point in the reactor may be described by [7]

$$C_i = \bar{C}_i + c_i \quad (3)$$

where \bar{C}_i is the bulk average concentration of component i in the reactor and c_i is the corresponding fluctuation term. The mixing conditions may be characterized

by the intensity or degree of segregation, I_s , defined as [7]

$$I_s = \overline{c_i^2} / \overline{c_{i0}^2} \quad (4)$$

where $\overline{c_{i0}^2}$ and $\overline{c_i^2}$ are the average square fluctuation concentrations of component i initially and at time t , respectively. At $t = 0$, $I_s = 1$ and the mixture is completely segregated; as time increases and the mixture approaches uniformity, I_s drops to zero. I_s gives a measure of the degree of completion of diffusive processes.

Toor [5] showed that for the reaction



the fractional conversion of the limiting reagent, F , was related to the intensity of segregation via

$$F = 1 + (\beta + 1) \left[1 + g(I_s^{1/2} \overline{c_i^2}^{1/2} / \overline{c_{i0}}) \right]. \quad (6)$$

In equation (6), β is related to the stoichiometry of the reaction by

$$\beta = aC_{A_o} / bC_{B_o} \quad (7)$$

where the subscript o denotes the initial concentration. The c_t terms in equation (6) refer to the total concentration fluctuation

$$c_t = ac_A + bc_B \quad (8)$$

and the function $g(x)$ is given by

$$g(x) = \frac{x}{\sqrt{2}} \text{ierfc} \frac{1}{\sqrt{2}x} \quad (9)$$

where $\text{ierfc}(y)$ is the first integral of the complementary error function.

In a well-mixed reactor, the intensity of segregation may be related to the mean residence time, τ' , and the correlation time for fluid motions, τ , in a straight forward fashion via [7]

$$I_s = \frac{1}{1 + \tau'/\tau}. \quad (10)$$

τ , in turn, is related to the Schmidt number, N_{Sc} , the length scale of segregation, L_s , and the velocity energy dissipation per unit mass, ϵ , by [7]

$$\tau = \frac{1}{2} \left[3 \left(\frac{5}{\pi} \right)^{2/3} \left(\frac{L_s^2}{\epsilon} \right)^{1/3} + \left(\frac{\nu}{\epsilon} \right)^{1/2} \ln N_{Sc} \right] \quad (11)$$

where ν is the kinematic viscosity. This expression is valid at high Schmidt numbers; for aqueous electrolyte and protein solutions, $N_{Sc} > 10^3$ in general.

The variables in equation (11) are experimentally accessible. L_s may be estimated by integrating the Eulerian concentration correlation function [7]

$$L_s = (1/\bar{c}_i^2) \int \overline{c_i(\mathbf{x})c_i(\mathbf{x} + \mathbf{r})} d\mathbf{r} \quad (12)$$

where \mathbf{x} represents the coordinates of a fixed location in the reactor and \mathbf{r} denotes the separation distance between the fixed point and an arbitrary second point. ϵ is related to the power input, P , by [7]

$$\epsilon = \eta P/M \quad (13)$$

where η is the efficiency of turbulence production and M is the mass of fluid.

If we consider the approach to the true thermodynamic equilibrium in reaction (2) and if no precipitate is initially present, equations (6), (10) and (11) may be applied to estimate the fractional conversion of protein to precipitate. The fractional conversion, in turn, may be used to relate the observed solubility with the true solubility. If this method is successful in explaining the solubility

behavior as a function of the contacting conditions, equation (11) suggests that the appropriate scale-up parameter combination is L_s^2/ϵ . This project is similar in spirit to a study of micromixing effects on the precipitation of BaSO_4 by Pohorecki and Baldyga [8].

6.2.2 Secondary Structure Estimation via Raman Spectroscopy

Analysis techniques for Raman amide I band spectra are evolving, utilizing more sophisticated, physically satisfying approaches. The constrained superposition method of Williams [9,10], a variant of which was used in Chapters 3 and 4, has been criticized by Berjot and coworkers [11]; in some cases, the computed reference spectra (\mathbf{AF}^+ , see Chapter 3) have multiple bands and negative intensity excursions. Further, the distinction between parallel and antiparallel β -sheet may be artificial. FTIR amide I band spectra of the all-parallel β -sheet proteins flavodoxin and triose phosphate isomerase indicate that parallel and antiparallel β -sheet backbone vibrations occur at essentially the same frequencies [12].

The Reference Intensity Profile (RIP) method developed by Berjot et al. [11] circumvents these problems by considering fewer classes of secondary structure and arriving at single bands, or RIPs, to describe the amide I mode vibrations of pure classes of structure. However, this method is flawed by the method used in computing the RIPs as different proteins were used for each structure type. In addition, the spectral data were manipulated to improve the results of the method; "... the RIPs were slightly modified to obtain a better general fit ..." [11]. This amounts to an arbitrary enhancement of spectral features.

In order to address the shortcomings of both of these structure estimation techniques, the following approach is proposed. Ideally, the columns of William's [10] reference spectra matrix \mathbf{AF}^+ would be Gaussian-Lorentzian lineshapes

centered at the backbone vibrational frequency of each pure structure class. In reality, due to the limited number of proteins in the reference set, the lineshapes in each column of \mathbf{AF}^+ are not well-developed. It should be possible, however, to deconvolute \mathbf{AF}^+ in terms of a given lineshape function for each structure class; this would allow the raw reference matrix \mathbf{AF}^+ to be replaced by an ideallized equivalent, \mathbf{L} , composed of these pure lineshape functions. Assuming the columns of \mathbf{AF}^+ and \mathbf{L} are normalized and in the limit where the bandwidth of each lineshape is much less than the frequency interval in the discretized matrix representation

$$\mathbf{L}^T \mathbf{AF}^+ = \mathbf{I}_n \quad (14)$$

where \mathbf{I}_n is the $n \times n$ identity matrix and n is the the number of structure types considered. Equation (14) suggests a criterion for the formation of \mathbf{L} from \mathbf{AF}^+ .

Formally, the problem is given the raw reference matrix \mathbf{AF}^+ , find the equivalent set of Lorentzian-Gaussian product functions, \mathbf{L} ,

$$\mathbf{L} = [\mathbf{l}_i(\nu_{01}, bw_1) : \dots : \mathbf{l}_i(\nu_{0i}, bw_i) : \dots : \mathbf{l}_n(\nu_{0n}, bw_n)] \quad (15)$$

where $\mathbf{l}_i(\nu_{0i}, bw_i)$ is a column vector of intensities of the lineshape function \mathbf{l}_i with center frequency ν_{0i} and bandwidth bw_i at discrete frequencies. For the \mathbf{l}_i ,

$$\mathbf{l}_i = \mathcal{L}^R \mathcal{G}^{(1-R)} \quad \text{such that} \quad \|\mathbf{l}_i\|_2 = 1.0 \quad (16)$$

where \mathcal{L}_i and \mathcal{G}_i are Lorentzian and Gaussian lineshapes, respectively, (as defined in equations (2) and (3) of Chapter 3) and R is the Lorentzian-Gaussian ratio.

To deconvolute \mathbf{AF}^+ , we must find the set of n center frequencies and bandwidths, and the ratio R that best represents \mathbf{AF}^+ ; that is

$$\text{find } \nu_{0i}, bw_i, R \quad \text{such that} \quad \|\mathbf{L}^T \mathbf{AF}^+ - \mathbf{I}_n\|_2 \quad \text{is minimized} \quad (17)$$

Berjot and coworkers [11] met with the greatest success by considering four structure types: disordered and ordered α -helix, total β -sheet, and undefined structure (including β -reverse turn and random coil). By using the literature values for the center frequencies, only $n + 1$ parameters need to be estimated using the criterion in expression (17). An inherent assumption in the formation of both \mathbf{AF}^+ and \mathbf{L} is that the lineshape for each structure type has the same intensity. Williams and Dunker [9] have argued that this assumption is reasonable. The number of bands \mathbf{AF}^+ is deconvoluted into may be determined by a statistical comparison of the residuals from expression (17) as a function of n .

Having obtained \mathbf{L} , the vector of structure fractions, \mathbf{f} , may be estimated from a sample spectrum, \mathbf{b} , using

$$\mathbf{L}\mathbf{f} = \mathbf{b} \quad \text{subject to} \quad \mathbf{f} \geq \mathbf{0} \quad \text{and} \quad \|\mathbf{f}\|_1 = 1.0. \quad (18)$$

This method alleviates the problems of non-physical reference spectra, non-distinct lineshapes, and inconsistent treatment of reference spectra.

Modified experimental protocols may also enhance the reliability of structure estimation by improving spectral signal-to-noise ratios. The Raman amide I band is perched atop a large fluorescence background signal; the background fluorescence is largely due to trace impurities. In Chapter 4, it was noted that commercial preparations of the various proteins studied had significantly lower background levels than the corresponding precipitate samples. Buffers should be purified by passage through a column of activated charcoal before use [13]. Protein solutions may also be treated by dialysis against purified buffer or distilled water, followed by lyophilization. Contaminant proteins will not be removed by this procedure, but the small fluorescent compounds, introduced during commercial preparation, may be substantially reduced.

6.2.3 Aggregation Kinetics

The modelling work presented in Chapter 5 concerning α -chymotrypsin aggregation kinetics incorporated a number of simplifying assumptions to facilitate parameter estimation. The credibility of the model may be enhanced by additional experimental work to verify assumptions and by increasing the level of physical rigor adhered to. Specifically, the assumption that soluble protein is in the native conformation should be confirmed experimentally and the restriction to constant aggregation kernels should be relaxed.

Dynamic laser light scattering data presented in Chapter 5 for supernatant samples of α -chymotrypsin precipitations indicated that soluble protein was largely composed of monomeric and dimeric species. This does not constitute sufficient evidence, however, to verify the assumption that soluble protein remains in the native state. Raman amide I band spectroscopy of supernatant samples can potentially provide this information. Complications arise because of the difficulty of obtaining high quality solution phase protein Raman spectra; protein concentrations in excess of 50 mg mL^{-1} may be required for reasonable acquisition times. Also, solvent subtraction is necessitated as the water stretching mode at 1665 cm^{-1} falls in the middle of the amide I region. Structure perturbation estimates, using the technique of Chapter 3, should be considered relative to the solvent subtracted solution spectra of native the proteins. The results of the analysis may require the restructuring of the model as the conformational changes may not be limited to the supersaturation.

The constant aggregation kernel, k_3 , while allowing the closure of the moment equations, may not be an accurate physical representation of protein aggregation. The investigations of Shimada and Matsushita [14] led to the identification of α -chymotrypsin as a gelation-type protein as opposed to a coagulation-

type protein. In order to account for gelation phenomena in the model, the aggregation kernel should include a particle size dependence of the the form [15]

$$k_3 \sim (ij)^\omega \quad \text{where,} \quad \frac{1}{2} < \omega \leq 1 \quad (19)$$

for the interaction of an i -mer with a j -mer. A variety of aggregation kernels, including a number that incorporate relation (19), have been proposed in the literature [16,17]. An approximate method for eliminating the closure problem posed by the size dependence of these aggregation kernels has been suggested by Hulbert and Katz [18]. The monomer addition mechanism of the model may also require modification to reflect physical assumptions inherent in different kernels regarding the number of nonspecific interaction sites considered.

A Coulter counter may be used to determine the evolution of particle size distributions with time for slowly aggregating systems [18]; less denaturing salts, low protein concentrations, and low temperatures slow the rate of aggregation. The particle size distributions may provide a means of discriminating between competing aggregation kernels. Turbidity measurements can only provide information about the integrated behavior of the particle system.

6.2.4 Protein Denaturation in Biological Separation Operations

The approach taken in the current work with protein precipitation may be a useful paradigm for studies of other separation operations employing potentially denaturing environmental conditions. Spectroscopic probes and activity assays may be used to determine structure-function relationships as a function of processing conditions. The resulting information has important implications for both the design and operation of separation processes as it may facilitate the prediction of active protein yields.

Entirely analogous studies may be performed with other precipitation processes and with various chromatographic techniques. Section 1.3 lists a wide range of reagents for inducing protein precipitation varying in both efficacy and chaotropic potential; comparison of effect on precipitate activity and structure for different precipitants may guide the selection process. The conformational state of protein absorbed on an ion-exchange or hydrophobic interaction resin may also be amenable to analysis via Raman spectroscopy.

The use of proteins as separating agents in bioprocesses may also be studied. Recycle and/or reuse of these materials may be crucial to the economics of the process. A convenient example are the immobilized antibodies used in immunoaffinity chromatography. Immobilization chemistries may be unstable and elution conditions may be denaturing; this results in column capacity losses and antibody leakage. A specific program for the investigation of these phenomena with preliminary feasibility studies is presented in Appendix F; this study can provide information complimentary to that obtained through hapten labelling techniques [20].

6.3 References

- 1 Foster, P.R., Dunnill, P., and Lilly, M.D., "The Kinetics of Protein Salting-Out: Precipitation of Yeast Enzymes by Ammonium Sulfate," *Biotechnol. Bioeng.* **28** 545-580 (1976)
- 2 Bell, D.J., Hoare, M., and Dunnill, P., "The Formation of Protein Precipitates and Their Centrifugal Recovery," *Adv. in Biochemical Eng.* **26**, 1-72 (1983)
- 3 Street, G., "Large Scale Industrial Enzyme Production," *CRC Critical Reviews in Biotechnology* **1**, 59-85 (1984)
- 4 Scopes, R.K., *Protein Purification Principles and Practice* Springer-Verlag, New York, 48 (1982)
- 5 Toor, H.L., "Mass Transfer in Dilute Turbulent and Non-turbulent Systems with Rapid Irreversible Reactions and Equal Diffusivities," *AIChE J.* **8**, 70-78 (1962)
- 6 Toor, H.L., "The Non-premixed Reaction: $A + B \rightarrow \text{Products}$," in *Turbulence in Mixing Operations, Theory and Application*, (Brodkey, R.S. Ed.) Academic Press, New York, 123-163 (1975)
- 7 Brodkey, R.S., "Fluid Motion and Mixing," in *Mixing Theory and Practice*, (Uhl, V.W. and Gray, J.B. Eds.) Academic Press, New York, 7-110 (1966)
- 8 Pohorecki, R. and Bałdyga, J., "The Effects of Micromixing and the Manner of Reactor Feeding on Precipitation in Stirred Tank Reactors," *Chem. Eng. Sci.*, **43**, 1949-1954 (1988)
- 9 Williams, R.W. and Dunker, A.K., "Determination of the Secondary Structure of Proteins from the Amide I Band of the Laser Raman Spectrum," *J. Mol. Biol.* **152**, 783-813 (1981)
- 10 Williams, R.W., "Estimation of Protein Secondary Structure from the Laser

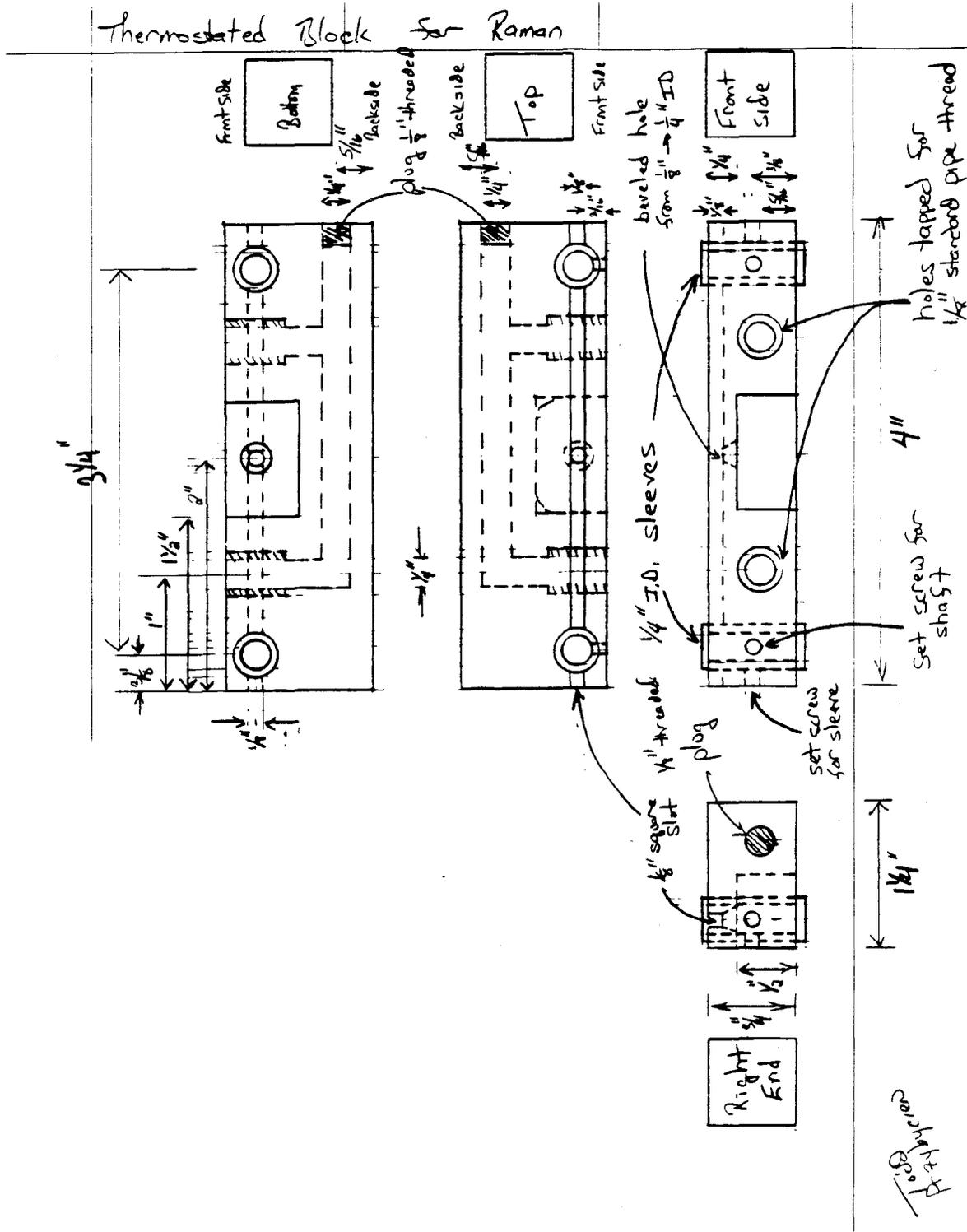
- Raman Amide I Spectrum", *J. Mol. Biol.* **166**, 581-603 (1983)
- 11 Berjot, M., Marx, J., and Alix, A.J.P., "Determination of the Secondary Structure of Proteins from the Raman Amide I Band: The Reference Intensity Profiles Method," *J. Raman Spectrosc.* **18**, 289-300 (1987)
 - 12 Susi, H. and Byler, D.M., "Fourier Transform Infrared Study of Proteins with Parallel β -Chains," *Arch. Biochem. Biophys.* **258**, 465-469 (1987)
 - 13 Freeman, J.J., personal communication
 - 14 Shimada, K. and Matsushita, S., "Relationship Between Thermocoagulation of Proteins and Amino Acid Compositions," *J. Agric. Food Chem.* **28** 413-417 (1980)
 - 15 Leyrvaz, F., "Critical exponents in the Smoluchowski Equations of Coagulation," in *Kinetics of Aggregation and Gelation* (Family, F. and Landau, D.P. Eds.) 201-204 (1984)
 - 16 Drake, R.L., "A General Mathematical Survey of the Coagulation Equation," in *Topics in Current Aerosol Research*, vol. 3, part 2 (Hidy and Brock, Eds.) Pergamon Press, New York, 202-376 (1972)
 - 17 Beckman, J.R. and Farmer, R.W., "Bimodal CSD Barite Due to Agglomeration in an MSMPR Crystallizer," *AIChE Symp. Ser.* **253** **83**, 85-94 (1987)
 - 18 Hulbert, H.M. and Katz, S., "Some Problems in Particle Technology, A Statistical Mechanical Formulation," *Chem. Eng. Sci.* **19**, 555-574 (1964)
 - 19 Treweek, G.P. and Morgan, J.J., "Size Distributions of Flocculated Particles: Application of Electronic Particle Counters," *Env. Sci. Technol.* **11**, 707-714 (1977)
 - 20 Fernandez, E.J., Fernandez, F.B., Jagoda, R.B., and Clark, D.S., "Electron Paramagnetic Resonance Spectroscopy Studies of Immobilized Monoclonal

Antibody Structure and Function," *ACS Symp. Ser.* **31** (Asenjo, J.A. and Hong, J., Eds.) 208-214 (1986)

Appendix A

Thermostatted Sample Stage for the
Raman Spectrometer

A.1 Figure Thermostatted Sample Stage for Raman Spectrometer



Appendix B

Raman Spectral Analysis Package
(RSAP)
Documentation: Version 1.0

B.1 Introduction

RSAP is set of computer programs for the post-analysis of Raman amide I band protein spectra. The amide I band is sensitive to protein secondary structure content; the $C^\alpha-C'$ and $N-C^\alpha$ dihedral angles determine both the conformation of the protein backbone and the energetics of the amide I vibrational mode. The relative amounts of ordered and disordered α -helix, parallel and antiparallel β -sheet, reverse turn and random coil are assigned by RSAP. The sample assignments are made by deconvoluting experimental spectra in terms of contributions from reference spectra of proteins with known secondary structure contents.

Smoothing, subtraction and normalization operations must be performed on experimental spectra before the deconvolution step. Raman spectra of proteins often lie on an intense, sloping, fluorescence background that varies from sample to sample. Aromatic side chains also contribute peaks that lie near the amide I band. Solvent contributions to the amide I band must be removed for solution samples.

The routines that perform the smoothing, subtraction and deconvolution steps were written *de novo*. Subtraction criteria, generalized algorithms and reference protein data were obtained from work by Dunker [1] and Williams [2].

RSAP is written predominately in IBM Professional FORTRAN for use on a IBM PC/XT, PC/AT or compatible computer. The system requirements are a hard disk and a math coprocessor. RSAP is contained on two $5\frac{1}{4}$ inch double-sided, double-density floppy disks; the first contains the executable image and batch files and the second contains the source code. In many of the individual programs of the package, "canned" subroutines that are available publicly have been incorporated both with and without modifications.

B.2 Package Architecture

RSAP is operated by a batch program that calls the executable programs in succession, placing intermediate results in subdirectories along the way. This accounts for the extended structure of the package; the intermediate results are available for comparison with the raw data after each step of the analysis procedure. All data files containing spectra are 2E15.6 formatted to facilitate the use of plotting routines.

Figure 1 is a diagram of the RSAP subdirectory system. The contents of each subdirectory is also given. All of the core package executable files and batch routines reside in the root directory \RAMAN along with batch programs designed to configure diskettes for the storage of RSAP results and to perform the information transfers. The \RAWDATA subdirectory contains the raw experimental spectra. The \SMOOTH subdirectory contains smoothed spectra. For solution samples, solvent subtracted spectra are placed in the \SOLVENT subdirectory. The \PARAMTR subdirectory contains the statistical results of the subtraction. The \AROMATIC subdirectory contains normalized amide I spectra after fluorescence background and aromatic peak subtraction. This subdirectory in turn has two subdirectories, \FITPEAKS and \PARAMTR. The individual fitted peaks, baseline and subtracted spectra are placed in \FITPEAKS. The frequency, bandwidth and amplitude of the fitted peaks corresponding to those in \FITPEAKS resides in \PARAMTR along with statistical calculations of the fitting procedure. \STRUCTUR contains the secondary structure fit results and the \REF subdirectory. The reference data matrices from Williams [2] are stored in \REF. For a given sample, the intermediate data files will all have the same filename but different extensions corresponding to the operation performed; the format is *filename.ext* where *filename* is an eight character code

describing the sample and *ext* is a three character code describing how the data was manipulated. The file extension codes are listed in Table 1. Source code listings are placed in the \SRCECODE subdirectory which is not normally accessed by the analysis programs.

B.3 Running RSAP

The commands associated with RSAP were designed so that little user intervention is required. The following commands are available:

ANALYZE

RSAP analysis is initiated from the \RAMAN subdirectory by this command. The program prompts the user for the name of a sample to analyze and then asks if solvent subtraction is necessary; both solid and solution phase samples may be analyzed. For solid phase samples, it is assumed that a 2E15.6 file containing intensity and frequency data (from 1500 to 1800 cm^{-1} at 1 cm^{-1} intervals) has been placed into the \RAWDATA subdirectory. Solution phase samples require additional data (from 2500-2505, 2650-2655, 2800-2805 and 3100-3105 cm^{-1} at 1 cm^{-1} intervals) with corresponding data for the neat solvent for the solvent subtraction routine. User intervention will only be necessary if the aromatic/fluorescence subtraction routine requires excessive iterations.

INSTALL

This command installs RSAP on the hard disk of the computer. It is invoked by inserting the executable RSAP floppy into drive A and setting the default directory to A:.

DATACOPY

The set of files corresponding to one sample is copied onto a suitably for-

matted diskette in drive A. This command is available in the \RAMAN directory.

DATADISK

Subdirectories corresponding to the RSAP structure are set up on a diskette in drive A. This command is available in the \RAMAN subdirectory and is used in conjunction with the DATACOPY command.

B.4 Program Descriptions

B.4.1 RSAP Operating Programs

Two batch files, ANALYZE.BAT and RUNPKG.BAT, and one FORTRAN program, SAMPLE.EXE control the analysis package. ANALYZE.BAT erases the control routine from any previous run, calls SAMPLE.EXE and then RUNPKG.BAT. SAMPLE.EXE prompts the user for the name of the experimental spectrum to be analyzed and whether or not solvent subtraction will be required. The sample code is written to a short file, SAMPLE.COD, which is used to pass the name of the sample to each of the individual analytical programs. The batch file RUNPKG.BAT is also created by SAMPLE.EXE. RUNPKG.BAT executes the individual analysis programs in the appropriate order and places intermediate results in subdirectories. In this way, the order of execution of the analysis programs can be varied depending on whether or not solvent subtraction is required.

B.4.2 RSAP Analysis Programs

The first analytical routine, SMOOTH.EXE, smooths the raw spectral data. The 5-point smoothing algorithm proposed by Savitzky and Golay [3] was used to be consistent with the approach of Williams [2]. Smoothing is performed first so

that the subsequent aromatic/fluorescent background fitting routine converges in fewer iterations. Smoothed sample spectra were placed in *filename*.SMO files.

If the sample requires solvent subtraction, SOLVENT.EXE is executed. This program subtracts the solvent spectrum from the sample spectrum using the criteria described by Williams and Dunker [1]. A multiple of the solvent spectrum is subtracted so that baseline in the 2500-2800 cm^{-1} region is collinear with that of the 3100-3105 cm^{-1} region. The algorithm consists of solving the following equation

$$S_{\text{exp}}(\nu) - w S_{\text{sol}}(\nu) = \alpha\nu + \beta \quad (1)$$

where $S_{\text{exp}}(\nu)$ and $S_{\text{sol}}(\nu)$ represent the experimental and solvent spectrum as a function of the frequency ν respectively, w is the weight assigned to the solvent spectrum for subtraction, and α and β are the slope and intercept of the resulting baseline between 2500 and 3105 cm^{-1} . The coefficients w , α and β were determined by formulating (1) as a multiple linear regression problem

$$\begin{pmatrix} n & \sum_{\nu} S_{\text{sol}}(\nu) & \sum_{\nu} \nu \\ \sum_{\nu} \nu & \sum_{\nu} S_{\text{sol}}(\nu) \cdot \nu & \sum_{\nu} \nu^2 \\ \sum_{\nu} S_{\text{sol}}(\nu) & \sum_{\nu} (S_{\text{sol}}(\nu))^2 & \sum_{\nu} S_{\text{sol}}(\nu) \cdot \nu \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \\ w \end{pmatrix} = \begin{pmatrix} \sum_{\nu} S_{\text{exp}}(\nu) \\ \sum_{\nu} S_{\text{exp}} \cdot \nu \\ \sum_{\nu} S_{\text{exp}} \cdot S_{\text{sol}} \end{pmatrix}. \quad (2)$$

Equation (2) was solved via an LU decomposition subroutine [4]. Solvent subtracted spectra were assigned to *filename*.SUB files.

AROMATIC.EXE fit a sloping baseline, the fluorescence background and the aromatic peaks simultaneously. The fluorescence was found [2] to be well represented by a Gaussian with a bandwidth of 140.1 cm^{-1} and a frequency of 1616 cm^{-1} . The aromatic peaks were fit with Gaussian-Lorentzian product functions. The function fit to the experimental spectra was

$$S_{\text{cal}}(\nu) = \alpha\nu + \beta + \sum_{i=1}^4 A_i \mathcal{G}_i^{(1-R)} \mathcal{L}_i^R + A_f \mathcal{G}_f \quad (3)$$

where

$$\mathcal{G}_i = \frac{2\sqrt{\pi \ln 2}}{B_i} \exp\left\{\frac{-4 \ln 2(\nu - \nu_{oi})^2}{B_i^2}\right\} \quad (4)$$

and

$$\mathcal{L}_i = \frac{2B_i}{\pi B_i^2 + 4\pi(\nu - \nu_{oi})^2}. \quad (5)$$

The first two terms of (3) represent the sloping baseline with slope α and intercept β . The summation term is the contribution from the aromatic peaks with Gaussian and Lorentzian lineshape functions \mathcal{G}_i and \mathcal{L}_i as defined in equations (4) and (5) respectively. B_i is the bandwidth at half height, ν_{oi} is the frequency and A_i is the amplitude of the i^{th} peak respectively. The Gaussian-Lorentzian ratio R was fixed at 0.7143 [5]. The last term of equation (3) represents the fluorescence contribution. Equation (3) contains 13 free parameters that were fit simultaneously via a bounded nonlinear least squares Marquardt algorithm subroutine [6]. This subroutine was translated into structured FORTRAN from its FORTRAN66 listing [7]. The algorithm is iterative and is the rate determining step in RSAP. In general, the lower the spectral signal-to-noise ratio, the more iterations were required for convergence. AROMATIC.EXE executed at about 1 iteration per minute on a PC/XT; a typical spectrum with S/N = 40 may require ~30 minutes for complete analysis. Provision was made for the user to intervene after every hundred iterations for more problematic experimental spectra. The fitted peaks were subtracted from the experimental spectrum. The subtracted spectrum was normalized by the sum of intensities from 1615 to 1710 cm^{-1} at 5 cm^{-1} intervals. Noise levels were estimated from the standard deviation of the intensities from 1730 to 1760 cm^{-1} at 5 cm^{-1} intervals. The normalized amide I band (1630 to 1700 cm^{-1} at 5 cm^{-1} intervals) was given the file extension NOR. The individual fitted peaks (four aromatic peaks

and the fluorescence peak), the baseline, the fitted function and the subtracted spectrum were placed in the \FITPEAKS subdirectory with the extensions PK1 through PK5, PKB, PKS and PKM respectively. The values of the 13 fitted constants were placed in a parameter file along with statistical information in the \PARAMTR subdirectory.

The final operation is the deconvolution of the normalized amide I band in terms of the protein reference spectra by the program DSTRUCT.EXE. This was done via a constrained least squares fit of a matrix of reference spectra, \mathbf{A} , and corresponding x-ray secondary structure assignments, \mathbf{F} , from [2] using

$$\mathbf{AF}^+\mathbf{f} = \mathbf{b} \quad \text{subject to} \quad \|\mathbf{f}\|_1 = 1.0 \quad \text{and} \quad \mathbf{f} \geq \mathbf{0} \quad (6)$$

where \mathbf{AF}^+ represents a matrix of spectra of pure secondary structures, \mathbf{f} is the vector containing the fractional contribution of each structure class to the total secondary structure and \mathbf{b} is the normalized experimental spectrum. The matrix \mathbf{F} had a condition number of ~ 19 and was stable with respect to psuedoinversion. All one-time reference matrix manipulations were carried out using PC-MATLAB [8]. The constraints from system (6) were incorporated into a matrix inequality

$$\mathbf{Gf} \geq \mathbf{h} \quad (7)$$

where

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{h} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The problem (6) with inequality (7) was recast as a least distance problem via a singular value decomposition of \mathbf{AF}^+

$$\mathbf{AF}^+ = \mathbf{USV}^T = (\mathbf{U}_1 \quad \mathbf{U}_2) \begin{pmatrix} \mathbf{S} \\ \mathbf{0} \end{pmatrix} \mathbf{V}^T \quad (8)$$

and an orthogonal change of variables

$$\mathbf{x} = \mathbf{SV}^T \mathbf{f} - \mathbf{U}_1^T \mathbf{b}. \quad (9)$$

The singular value decomposition of \mathbf{AF}^+ was computed with PC-MATLAB [8] and indicated that the columns of \mathbf{AF}^+ were linearly independent. The least distance problem becomes

$$\text{minimize } \|\mathbf{x}\|_2 \quad \text{subject to} \quad \mathbf{GVS}^{-1} \mathbf{x} \geq \mathbf{h} - \mathbf{GVS}^{-1} \mathbf{U}_1^T \mathbf{b} \quad (10)$$

and was solved via subroutine LDP [9]. The solution vector $\hat{\mathbf{f}}$ is obtained from the solution of (10), $\hat{\mathbf{x}}$, using (9). The fitted structure results were reported to *filename*.STR along with the Euclidean norm of the residual vector $\|\mathbf{AF}^+ \hat{\mathbf{f}} - \mathbf{b}\|_2$. The reference matrices required by DSTRUCT.EXE were located in the \REF subdirectory.

B.4.3 Housekeeping Programs

The housekeeping programs are batch files that install RSAP, set up directories mimicing the RSAP structure on data diskettes and copy data from the hard disk into the appropriate directories on a data diskette. The installation routine resides only on the executable RSAP diskette. DATADISK.BAT and DATACOPY.BAT are in the \RAMAN directory.

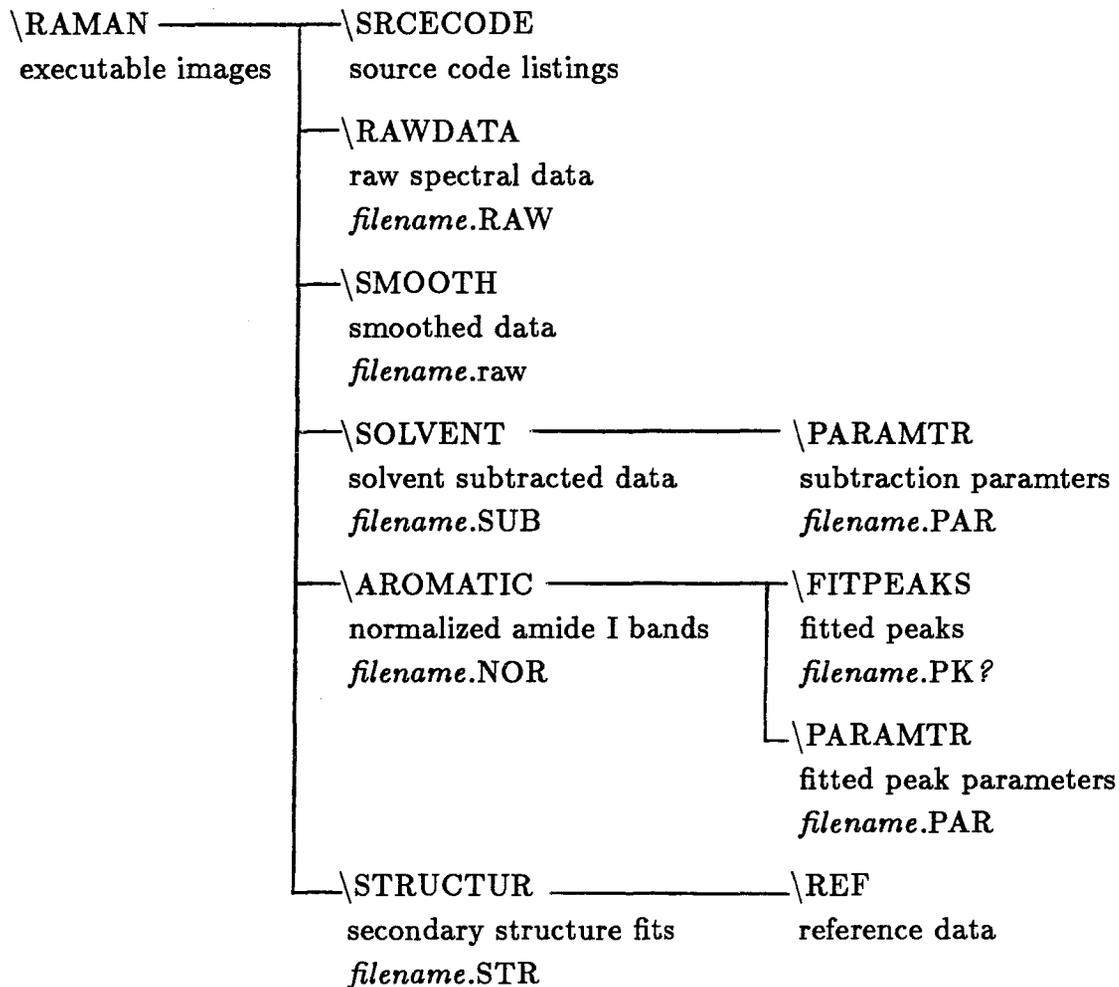
B.5 References

- 1 Williams, R.W. and Dunker, R.K., "Determination of the Secondary Structure of Proteins from the Amide I Band of the Laser Raman Spectrum," *J. Mol. Biol.* **152** 783-813 (1981)
- 2 Williams, R.W., "Estimation of Protein Secondary Structure from the Laser Raman Amide I Spectrum," *J. Mol. Biol.* **166** 581-603 (1983)
- 3 Savitzky A. and Golay, M., "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.* **36** 1627-1639 (1967)
- 4 Strang, G., *Linear Algebra and its Applications* 2nd ed., Academic Press, New York, Appendix C (1980)
- 5 Williams, R.W., program FITOV.FOR
- 6 Kuester, J.L. and Mize, J.H., *Optimization Techniques with FORTRAN* McGraw-Hill, New York, 240-250 (1973)
- 7 a translated version of subroutine BSOLVE was obtained from K.D. Wittrup
- 8 PC-MATLAB, version 2.2, is a product of The Math Works, Inc., Sherborn, MA 01770
- 9 Lawson, C.L. and Hanson, R.J., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, New Jersey, 158-173 (1974)

B.6 Table File extension codes

| extension | file type |
|-------------------------------------|--|
| directory: \RAMAN | |
| EXE | executable FORTRAN routines |
| BAT | executable batch programs |
| COD | sample code file |
| directory: \RAMAN\RAWDATA | |
| RAW | raw spectral data, format: 2E15.6 |
| directory: \RAMAN\SMOOTH | |
| SMO | 5 point smoothed spectra, format: 2E15.6 |
| directory: \RAMAN\SOLVENT | |
| SUB | solvent subtracted spectra, format: 2E15.6 |
| directory: \RAMAN\SOLVENT\PARAMTR | |
| PAR | fitted coefficients for multiple linear regression, R^2 |
| directory: \RAMAN\AROMATIC | |
| NOR | subtracted, normalized amide I band, format: 2E15.6 |
| directory: \RAMAN\AROMATIC\FITPEAKS | |
| PK i | curves for i^{th} fitted peak, format: 2E15.6 |
| PKB | curve for fitted baseline, format: 2E15.6 |
| PKS | curve for sum of fitted peaks and baseline, format: 2E15.6 |
| PKM | subtracted spectra, format: 2E15.6 |
| directory: \RAMAN\AROMATIC\PARAMTR | |
| PAR | fitted peak parameters, sum of square error, S/N ratio |
| directory: \RAMAN\STRUCTUR | |
| STR | assigned secondary structure content, norm of residual |
| directory: \RAMAN\STRUCTUR\REF | |
| REF | reference structure matrices for DSTRUCT.EXE |

B.7 Figure RSAP directory structure



B.8 RSAP Source Code Listings

```
REM ***** ANALYZE.BAT *****
ECHO OFF
REM This batch file runs the program SAMPLE.EXE which stores the
REM the name of the sample for the executable programs to access
REM and then calls the batch file RUNPKG.BAT to run the analysis
REM package
CLS
ERASE RUNPKG.BAT
SAMPLE
RUNPKG
```

B.8 RSAP Source Code Listings (cont'd)

```

C *****
C *
C *                               SAMPLE.FOR                               *
C *
C *                               Todd M. Przybycien  8/6/87                    *
C *
C *   This program creates the file SAMPLE.COD which contains
C *   the name of the sample spectrum and whether the sample requires
C *   solvent subtraction or not and also creates the batch file
C *   RUNPKG.BAT which runs the analysis package and supplies the
C *   sample name to all the files the package creates.
C *
C *****
C
C   Variables
C
C   CHARACTER SAMPLE*8,SOLVNT*8,TYPE,EXT*3,UNK*7,FMT*9
C
C   Constants
C
C   UNK = 'UNKNOWN'
C   FMT = 'FORMATTED'
C
C   Main program
C
WRITE (*,10)
10 FORMAT (15X,'IMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM;',
& /15X,':',
& /15X,':',          RAMAN SPECTRAL ANALYSIS PACKAGE
& /15X,':',
& /15X,':',          by
& /15X,':',
& /15X,':',          Todd M. Przybycien  8/6/87
& /15X,':',
& /15X,':',          California Institute of Technology
& /15X,':',          Department of Chemical Engineering
& /15X,':',          Pasadena, California  91125
& /15X,':',
& /15X,'MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM<',
& //1X,'Enter an 8 character or less sample code: ',
& /1X,'Blanks will be filled with B's')
READ (*,20) SAMPLE
20 FORMAT (A8)
DO 30 I=1,8
IF (ICHAR(SAMPLE(I:I)).EQ.32) SAMPLE(I:I) = 'B'
30 CONTINUE
WRITE (*,40)
40 FORMAT (1X,'Does this sample require solvent subtraction? ',
& '(Y/[N]) ')
READ (*,50) TYPE
50 FORMAT (A1)
IF ((TYPE.EQ.'Y').OR.(TYPE.EQ.'y')) THEN
WRITE (*,60)
60 FORMAT (1X,'Enter the name of solvent spectrum: ')
READ (*,70) SOLVNT
70 FORMAT (A8)
EXT = 'SUB'
ELSE
EXT = 'SMO'
ENDIF
OPEN (1,FILE='SAMPLE.COD', FORM=FMT, STATUS=UNK)
WRITE (1,80) SAMPLE
80 FORMAT (A8)
WRITE (1,90) EXT

```

B.8 RSAP Source Code Listings (cont'd)

```
90 FORMAT (A3)
   IF ((TYPE.EQ.'Y').OR.(TYPE.EQ.'y')) THEN
100   WRITE (1,100) SOLVNT
      FORMAT (A8)
   ENDIF
   CLOSE (1)

C
C      Create the batch file that will run the package
C
OPEN (1,FILE= 'RUNPKG.BAT', FORM=FMT, STATUS=UNK)
   IF ((TYPE.EQ.'Y').OR.(TYPE.EQ.'y')) THEN
      WRITE (1,110) SAMPLE,SOLVNT,SAMPLE,SOLVNT,SAMPLE,SAMPLE,
        SAMPLE,SAMPLE,SAMPLE,SAMPLE,SAMPLE,SAMPLE
110   FORMAT (1X,'ECHO OFF',
      /1X,'CLS',
      /1X,'CD \RAMAN\SMOOTH',
      /1X,'COPY \RAMAN\RAWDATA\ ',A8,'.RAW',
      /1X,'COPY \RAMAN\RAWDATA\ ',A8,'.RAW',
      /1X,'COPY \RAMAN\SAMPLE.COD',
      /1X,'PATH \RAMAN',
      /1X,'SMOOTH',
      /1X,'ERASE *.RAW',
      /1X,'ERASE SAMPLE.COD',
      /1X,'CD \RAMAN\SOLVENT',
      /1X,'COPY \RAMAN\SMOOTH\ ',A8,'.SMO',
      /1X,'COPY \RAMAN\SMOOTH\ ',A8,'.SMO',
      /1X,'COPY \RAMAN\SAMPLE.COD',
      /1X,'PATH \RAMAN',
      /1X,'SOLVENT',
      /1X,'COPY ',A8,'.PAR \RAMAN\SOLVENT\PARAMTR',
      /1X,'ERASE *.PAR',
      /1X,'ERASE *.SMO',
      /1X,'ERASE SAMPLE.COD',
      /1X,'CD \RAMAN\AROMATIC',
      /1X,'COPY \RAMAN\SAMPLE.COD',
      /1X,'COPY \RAMAN\SOLVENT\ ',A8,'.SUB',
      /1X,'PATH \RAMAN',
      /1X,'AROMATIC',
      /1X,'COPY ',A8,'.PAR \RAMAN\AROMATIC\PARAMTR',
      /1X,'COPY ',A8,'.PK? \RAMAN\AROMATIC\FITPEAKS',
      /1X,'ERASE *.SUB',
      /1X,'ERASE *.PAR',
      /1X,'ERASE *.PK?',
      /1X,'ERASE SAMPLE.COD',
      /1X,'CD \RAMAN\STRUCTUR\REF',
      /1X,'COPY \RAMAN\SAMPLE.COD',
      /1X,'COPY \RAMAN\AROMATIC\ ',A8,'.NOR',
      /1X,'PATH \RAMAN',
      /1X,'DSTRUCT',
      /1X,'COPY ',A8,'.STR \RAMAN\STRUCTUR',
      /1X,'COPY ',A8,'.CAL \RAMAN\STRUCTUR',
      /1X,'ERASE ',A8,'.*',
      /1X,'ERASE SAMPLE.COD',
      /1X,'CD \RAMAN',
      /1X,'CLS',
      /1X,'JOB DONE',
      /1X,'PATH C:')
   ELSE
      WRITE (1,120) SAMPLE,SAMPLE,SAMPLE,SAMPLE,SAMPLE,SAMPLE,
        SAMPLE,SAMPLE
120   FORMAT (1X,'ECHO OFF',
      /1X,'CLS',
      /1X,'CD \RAMAN\SMOOTH',
      /1X,'COPY \RAMAN\RAWDATA\ ',A8,'.RAW',
```

B.8 RSAP Source Code Listings (cont'd)

```

&          /1X,'COPY \RAMAN\SAMPLE.COD',
&          /1X,'PATH \RAMAN',
&          /1X,'SMOOTH',
&          /1X,'ERASE *.RAW',
&          /1X,'ERASE SAMPLE.COD',
&          /1X,'CD \RAMAN\AROMATIC',
&          /1X,'COPY \RAMAN\SAMPLE.COD',
&          /1X,'COPY \RAMAN\SMOOTH\',AS,'.SMO',
&          /1X,'PATH \RAMAN',
&          /1X,'AROMATIC',
&          /1X,'COPY ',AS,'.PAR \RAMAN\AROMATIC\PARAMTR',
&          /1X,'COPY ',AS,'.PK? \RAMAN\AROMATIC\FITPEAKS',
&          /1X,'ERASE *.SMO',
&          /1X,'ERASE *.PAR',
&          /1X,'ERASE *.PK?',
&          /1X,'ERASE SAMPLE.COD',
&          /1X,'CD \RAMAN\STRUCTUR\REF',
&          /1X,'COPY \RAMAN\SAMPLE.COD',
&          /1X,'COPY \RAMAN\AROMATIC\',AS,'.NOR',
&          /1X,'PATH \RAMAN',
&          /1X,'DSTRUCT',
&          /1X,'COPY ',AS,'.STR \RAMAN\STRUCTUR',
&          /1X,'COPY ',AS,'.CAL \RAMAN\STRUCTUR',
&          /1X,'ERASE ',AS,'.*',
&          /1X,'ERASE SAMPLE.COD',
&          /1X,'CD \RAMAN',
&          /1X,'CLS',
&          /1X,'JOBDONE',
&          /1X,'PATH C:')
&
&      ENDIF
&      CLOSE (1)
&      WRITE (*,130)
130  FORMAT (1X,'Batch program RUNPKG.BAT created.')
&      STOP 'Normal termination of SAMPLE.EXE'
&      END
```

B.8 RSAP Source Code Listings (cont'd)

```
REM ***** RUNPKG.BAT *****
REM This batch program runs the elements of RSAP by establishing
REM appropriate paths and placing data files in the required
REM directories
ECHO OFF
CLS
CD \RAMAN\SMOOTH
COPY \RAMAN\RAWDATA\NACL60BB.RAW
COPY \RAMAN\SAMPLE.COD
PATH \RAMAN
SMOOTH
ERASE *.RAW
ERASE SAMPLE.COD
CD \RAMAN\AROMATIC
COPY \RAMAN\SAMPLE.COD
COPY \RAMAN\SMOOTH\NACL60BB.SMO
PATH \RAMAN
RAMANFIT
COPY NACL60BB.PAR \RAMAN\AROMATIC\PARAMTR
COPY NACL60BB.PK? \RAMAN\AROMATIC\FITPEAKS
ERASE *.SMO
ERASE *.PAR
ERASE *.PK?
ERASE SAMPLE.COD
CD \RAMAN\STRUCTUR\REF
COPY \RAMAN\SAMPLE.COD
COPY \RAMAN\AROMATIC\NACL60BB.NOR
PATH \RAMAN
DSTRUCT
COPY NACL60BB.STR \RAMAN\STRUCTUR
COPY NACL60BB.CAL \RAMAN\STRUCTUR
ERASE NACL60BB.*
ERASE SAMPLE.COD
CD \RAMAN
JOB00NE
```

B.8 RSAP Source Code Listings (cont'd)

```
C ***** SMOOTH.FOR *****
C *
C *           Todd M. Przybycien
C *
C *           10/14/88
C *
C *           This program performs 5 point smoothing of raw spectral
C * data. The algorithm is based on that proposed by Savitzky and
C * Golay; Anal. Chem., 36, 1627 (1964).
C *
C *****
C
C           Variables
C
C           INTEGER  CHR,NPTS
C           REAL    NU(400),INT(400),SINT(400)
C           CHARACTER FMT*9,ST1*3,ST2*7,ST3*3,SAMPLE*8,SOLVNT*8,INFIL*8,EXT*3
C
C           Constants
C
C           FMT = 'FORMATTED'
C           ST1 = 'OLD'
C           ST2 = 'UNKNOWN'
C           ST3 = 'NEW'
C
C           Main program
C
C           WRITE (*,10)
C           10 FORMAT (1X,'RAMAN SPECTRAL ANALYSIS PACKAGE: ',
C           &          'Data Smoothing Routine')
C
C           Read sample code and open raw data file
C
C           OPEN (1, FILE= 'SAMPLE.COD', FORM=FMT, STATUS=ST1)
C           READ (1,20) SAMPLE
C           20 FORMAT (A8)
C           READ (1,30) EXT
C           30 FORMAT (A3)
C           IF (EXT.EQ.'SUB') THEN
C             READ (1,40) SOLVNT
C           40   FORMAT (A8)
C           ENDIF
C           CLOSE (1)
C           INFIL = SAMPLE
C           50 OPEN (1,FILE=INFIL//'.RAW', FORM=FMT, STATUS=ST1)
C           IF (EXT.EQ.'SUB') THEN
C             NPTS = 325
C           ELSE
C             NPTS = 301
C           ENDIF
C           READ (1,60) (NU(I),INT(I),I=1,NPTS)
C           60 FORMAT (2E15.6)
C           CLOSE (1)
C
C           5 point smooth spectrum with quadratic/cubic convolutes,
C           will not smooth the first and last two points
C
C           Smooth only amide I band of solution spectra since
C           the solvent subtraction program should function
C           independently of the smoothing of the spectrum
C
C           DO 70 I=3,299
C             SINT(I)=-3*(INT(I-2)+INT(I+2))+12*(INT(I-1)+INT(I+1))
C             &          +17*INT(I)
```

B.8 RSAP Source Code Listings (cont'd)

```
          SINT(I)=SINT(I)/35
70 CONTINUE
  SINT(1)=INT(1)
  SINT(2)=INT(2)
  SINT(300)=INT(300)
  SINT(301)=INT(301)
  IF (NPTS.EQ.325) THEN
    DO 80 I=302,325
      SINT(I)=INT(I)
80    CONTINUE
  ENDIF
  WRITE (*,90)
90  FORMAT (/1X,'Smoothing completed')
C
C      Place smoothed spectrum in file [sample code].SMO
C
  OPEN (1, FILE=INFIL//'.SMO', FORM=FMT, STATUS=ST2)
  WRITE (1,100) (NU(I),SINT(I),I=1,NPTS)
100 FORMAT (2E15.6)
  CLOSE (1)
  IF ((INFIL.EQ.SAMPLE).AND.(EXT.EQ.'SUB')) THEN
    INFIL = SOLVNT
    GOTO 50
  ENDIF
  STOP 'Normal termination of SMOOTH.EXE'
  END
```


B.8 RSAP Source Code Listings (cont'd)

```

CLOSE (1)
CLOSE (2)
C
C      Set up matrices for multiple linear regression solvent
C      subtraction and correlation coefficient calculation
C
DO 50 I=1,3
  DO 40 J=1,3
    A(I,J)=0.
40  CONTINUE
    B(I)=0.
50  CONTINUE
    A(1,1)=24.
    R(4)=0.
    DO 60 I=302,NPTS
      A(1,2)=A(1,2)+INT(2,I)
      A(1,3)=A(1,3)+NU(I)
      A(2,2)=A(2,2)+(INT(2,I))**2
      A(2,3)=A(2,3)+INT(2,I)*NU(I)
      A(3,3)=A(3,3)+(NU(I))**2
      B(1)=B(1)+INT(1,I)
      B(2)=B(2)+INT(2,I)*INT(1,I)
      B(3)=B(3)+NU(I)*INT(1,I)
      R(4)=R(4)+(INT(1,I))**2
60  CONTINUE
    A(2,1)=A(1,2)
    A(3,1)=A(1,3)
    A(3,2)=A(2,3)
    R(1)=B(1)
    R(2)=B(2)
    R(3)=B(3)
C
C      Solve matrix problem Ax=b by LU factorization
C      to get regression coefficients, check A to make
C      sure it's invertible
C
CALL DECOMP (NDIM,N,A,IPVT,DET)
IF (DET.EQ.0.) THEN
  WRITE (*,70)
70  FORMAT (/1X,'The matrix A is non-invertible')
  GO TO 999
ENDIF
CALL SOLVE (NDIM,N,A,IPVT,B)
C
C      Carry out subtraction over both water and amide I bands
C
DO 80 I=1,NPTS
  INT(3,I)=INT(1,I)-B(2)*INT(2,I)
80  CONTINUE
C
C      Create subtracted data file for the amide I band:
C      [sample code].SUB
C
OPEN (1, FILE= SAMPLE//'.SUB', FORM= FMT, STATUS= ST2)
WRITE (1,90) (NU(I),INT(3,I),I=1,301)
90  FORMAT (2E15.6)
CLOSE (1)
WRITE (*,100)
100 FORMAT (1X,'Solvent subtraction completed')
C
C      Compute correlation coefficient
C
RSQ=(B(1)*R(1)+B(2)*R(2)+B(3)*R(3)-((R(1)**2)/24.))
&  / (R(4)-((R(1)**2)/24.))

```

B.8 RSAP Source Code Listings (cont'd)

```
C
C      Write solution and correlation coefficient to file
C      [sample code].PAR
C
OPEN (1, FILE= SAMPLE//'.PAR', FORM=FMT, STATUS=ST2)
WRITE (1,110) SAMPLE
110 FORMAT (/1X,'Multiple linear regression solution for ',A8)
WRITE (1,120) B(3),B(1)
120 FORMAT (/6X,'linear baseline coefficients anu + b '
      & /6X,'a = ',E15.6,' b = ',E15.6,/)
      WRITE (1,130) B(2)
130 FORMAT (/6X,'solvent subtraction coefficient,    x = ',E15.6,/)
      WRITE (*,140) RSQ
140 FORMAT (/6X,'regression correlation coefficient, R} = ',F8.6)
CLOSE (1)
999 STOP 'Normal termination of SOLVENT.EXE'
END
```

B.8 RSAP Source Code Listings (cont'd)

```
C *****
C *
C *                               AROMATIC.FOR                               *
C *
C *                               Todd M. Przybycien 9/3/87                       *
C *
C * This program prepares raw spectral data for structural                      *
C * analysis via the following operations:                                       *
C * 1. flat baseline subtraction so that intensities near 1500                *
C *    and 1800 cm-1 are near zero                                             *
C * 2. simultaneous subtraction of aromatic side chain bands                  *
C *    using fit Gaussian-Lorentzian product functions                        *
C * 3. Gaussian fluorescence background subtraction                          *
C * 4. normalization of spectra by division by the sum of                     *
C *    intensities between 1615 and 1710 cm-1 at 5 cm-1                      *
C *    intervals                                                                *
C * 5. noise level computation by measuring standard deviation                 *
C *    of points between 1730 and 1760 cm-1 at 5 cm-1 intervals              *
C *
C * This program assumes intensity data has been collected at                  *
C * 1 cm-1 (real) intervals from 1500 to 1800 cm-1 (real)                    *
C *
C *****
C
C                               Variables
C
C   INTEGER      I, J, KK, NN, ICON, INDEX, MODE, FL
C   REAL         Y1(1500:1800), INTCAL(1500:1800), S(1500:1800),
C   &            N(1500:1800), AMP(5), BW(5), NU0(5), R(5), SUM, MEAN, STDEV,
C   &            MAX, SN, YI, SLOPE,
C   &            B(15), BMAX(15), BMIN(15), BC(15), BV(15), B1(212),
C   &            Y(212), ZC(212), Z(301), X1(301), X(212)
C   CHARACTER    SAMPLE*8, EXT*3, FMT*9, ST1*3, ST2*7
C   COMMON      SLOPE, YI, INTCAL
C
C                               Constants
C
C   FMT         = 'FORMATTED'
C   ST1         = 'OLD'
C   ST2         = 'UNKNOWN'
C
C                               Main program
C
C   WRITE (*,10)
C 10 FORMAT (1X, 'RAMAN SPECTRAL ANALYSIS PACKAGE:', /
C   &        1X, 'Aromatic Side Chain and Fluorescence Subtraction ',
C   &        'and Normalization Routine')
C
C                               Read sample code from file SAMPLE.COD
C
C   OPEN (1, FILE= 'SAMPLE.COD', FORM=FMT, STATUS=ST1)
C   READ (1,20) SAMPLE
C 20 FORMAT (A8)
C   READ (1,30) EXT
C 30 FORMAT (A3)
C   CLOSE (1)
C
C                               Read data from [sample].SMD if a solid sample
C                               or [sample].SUB if a solution sample
C
C   OPEN (1, FILE= SAMPLE//('.'//EXT), FORM=FMT, STATUS=ST1)
C   READ (1,40) (X1(I-1499), Y1(I), I=1500, 1800)
C 40 FORMAT (2E15.6)
C   CLOSE (1)
```

B.8 RSAP Source Code Listings (cont'd)

```
C
C      Obtain an estimate for flat baseline parameters using
C      intensities at either end of the spectrum
C
SLOPE = (Y1(1800)-Y1(1500))/300.
YI    = (Y1(1500) - SLOPE*1500. + Y1(1800) - SLOPE*1800.)/2.
C
C      Approximate peak location, bandwidth and Lorentzian/Gaussian
C      ratios for the four aromatic bands - a priori
C
C      William's program gives a Lorentzian/Gaussian ratio of 2.5
C      this implies  $2.5x + x = 1.0$ ,  $x = .2857$ ,  $R = 3x = .7143$ 
C
NU0(1) = 1550.0
BW(1)  = 20.0
R(1)   = 0.7143
NU0(2) = 1580.0
BW(2)  = 20.0
R(2)   = 0.7143
NU0(3) = 1605.0
BW(3)  = 20.0
R(3)   = 0.7143
NU0(4) = 1620.0
BW(4)  = 20.0
R(4)   = 0.7143
C
C      Fluorescence peak parameters
C
NU0(5) = 1616.0
BW(5)  = 140.1
R(5)   = 0.0
C
C      Compute individual peak amplitude estimates based on the
C      peak parameters above and the flat baseline
C
DO 50 I=1,4
AMP(I) = 0.6*AMPCAL((Y1(INT(NU0(I)))-(SLOPE*NU0(I)+YI)),
&          BW(I),R(I))
50 CONTINUE
C
C      Use a fraction of the intensity at 1616 cm-1 to estimate the
C      fluorescence peak intensity
C
AMP(5) = 0.25*(AMPCAL((Y1(1616)-(SLOPE*1616+YI)),BW(5),R(5)))
C
C      Set up vectors and solution constraints for the
C      Marquardt multivariable nonlinear regression
C      routine
C
C      B contains the initial parameter estimates
C
FL = 15
DO 60 I=1,4
B((3*I)-2) = AMP(I)
B((3*I)-1) = BW(I)
B(3*I)     = NU0(I)
60 CONTINUE
B(13) = AMP(5)
B(14) = SLOPE
B(15) = YI
C
C      BMIN and BMAX contain the parameter constraints
C
DO 70 I=1,4
```

B.8 RSAP Source Code Listings (cont'd)

```

      BMIN((3*I)-2) = 0.025*B((3*I)-2)
      BMIN((3*I)-1) = 5.
      BMIN(3*I) = NU0(I)-7.
      BMAX((3*I)-2) = 10*B((3*I)-2)
      BMAX((3*I)-1) = 50.
      BMAX(3*I) = NU0(I)+7.
70 CONTINUE
      BMIN(13) = 0.25*B(13)
      BMAX(13) = 4*B(13)
C      BMIN(14) = 0.1*B(14)
C      BMAX(14) = 10*B(14)
C      BMIN(16) = Y1(1800) - BMAX(14)*1800.
C      BMAX(15) = Y1(1800) - BMIN(14)*1800.
      BMIN(14) = -1.0E5
      BMAX(14) = 1.0E5
      BMIN(15) = -1.0E7
      BMAX(15) = 1.0E7

C      Initialization of algorithm parameters
C      Set BV(I)=1 for all I to select numerical derivatives
C
      KK = FL
      NN = 212
      FNU = 10.0
      FLA = 0.01
      TAU = 0.001
      EPS = 0.00002
      DO 80 J=1, KK
         BC(J) = B(J)
         BV(J) = 1
80 CONTINUE
      CALL FUNC (KK, BC, X, NN, ZC)
      PH = 0.0
      DO 90 I=1, 131
         X(I) = 1499 + I
         Y(I) = Y1(1499+I)
90 CONTINUE
      DO 100 I=132, 212
         X(I) = 1588 + I
         Y(I) = Y1(1588+I)
100 CONTINUE
      DO 110 J=1, NN
         PH = PH+(ZC(J)-Y(J))*2
         Z(J) = ZC(J)
110 CONTINUE
      ICON = KK
      ITER = 0
      WRITE (*,*) 'Starting non-linear least squares fitting routine'
120 IF (ICON.GT.0) THEN
      &      CALL BSOLVE (ICON, KK, NN, X, B, Z, Y, PH, FNU, FLA, TAU, EPS, BV, BMIN,
      &      BMAX, BC, ZC)
         ITER=ITER+1
         WRITE(*,130) ITER, ICON, PH
130      &      FORMAT (/1X, 'Iter No. ', I4, ' ICON = ', I2, ' d(error)} = ',
      &      E15.8)
         GO TO 120
      ELSEIF (ICON.EQ.-1) THEN
         WRITE(*,*) 'NO FUNCTION IMPROVEMENT POSSIBLE'
      ELSEIF (ICON.EQ.-4) THEN
         WRITE(*,*) 'CORRECTIONS SATISFY CONVERGENCE REQUIREMENTS,'
         WRITE(*,*) 'BUT, LAMDA FACTOR (FLA) STILL LARGE'
      ENDIF
      DO 135 I=1, 15
         IF (B(I).EQ.BMAX(I)) THEN

```

B.8 RSAP Source Code Listings (cont'd)

```

      WRITE (*,137) I,I
137     FORMAT (1X,'B(',I2,') = BMAX(',I2,')')
      PAUSE '<enter> continues run'
      ELSEIF (B(I).EQ.BMIN(I)) THEN
      WRITE (*,139) I,I
139     FORMAT (1X,'B(',I2,') = BMIN(',I2,')')
      PAUSE '<enter> continues run'
      ENDIF
135 CONTINUE
C
C      Convert BSOLVE vector notation back to discrete notation
C
      DO 140 I=1,4
      AMP(I) = B((3*I)-2)
      BW(I) = B((3*I)-1)
      NU0(I) = B(3*I)
140 CONTINUE
      AMP(5) = B(13)
      SLOPE = B(14)
      YI = B(15)
C
C      Redefine BSOLVE variables to use Subroutine FUNC
C      for the subtraction
C
      CALL FUNC (15,B,X1,301,Z)
      MAX = 0.
      DO 150 I=1500,1800
      S(I) = Y1(I) - Z(I-1499)
      IF (S(I).GT.MAX) MAX = S(I)
150 CONTINUE
C
C      Normalize the aromatic side chain subtracted
C      spectrum between 1615 and 1710 cm-1 at 5cm-1 intervals
C
      SUM = 0.
      DO 160 I=1615,1710,5
      SUM = SUM + S(I)
160 CONTINUE
      MAX = MAX/SUM
      DO 170 I=1615,1700,5
      N(I) = S(I)/SUM
170 CONTINUE
C
C      Create data file [sample code].NOR with normalized spectral
C      intensities from 1630 to 1700 cm-1
C
      OPEN (1, FILE= SAMPLE//'.NOR', FORM=FMT, STATUS=ST2)
      WRITE (1,180) (REAL(I),N(I),I=1630,1700,5)
180 FORMAT (2E15.6)
      CLOSE (1)
C
C      Calculate the noise level of spectrum by the
C      standard deviation of the normalized spectral
C      intensities between 1730 and 1760 cm-1
C
      SUM = 0.
      DO 190 I=1730,1760,5
      SUM = SUM + N(I)
190 CONTINUE
      MEAN = SUM/7.
      STDDEV = 0.
      DO 200 I=1730,1760,5
      STDDEV = STDDEV + (N(I)-MEAN)**2
200 CONTINUE
```


B.8 RSAP Source Code Listings (cont'd)

```

FUNCTION AMPCAL (INTENS,BW,R)
REAL    INTENS,BW,R,PI,LN2
PI      = 22./7.
LN2     = ALOG(2.)
AMPCAL = 0.5*INTENS*BW*(PI**(0.5*(R+1)))*(LN2**(0.5*(R-1)))
RETURN
END

C
C***** Subroutine BSOLVE *****
C
SUBROUTINE BSOLVE(ICON,KK,NN,X,B,Z,Y,PH,FNU,FLA,TAU,EPS,BV,
& BMIN,BMAX,BC,ZC)
INTEGER  KK,NN,ICON,J1,J2
REAL    A(15,17),AC(15,17),X(212),DEN,
& B(15),Z(212),Y(212),BV(15),BMIN(15),BMAX(15),
& DERIV(15,212),BC(15),ZC(212)
C
C WRITE(*,*) ICON,KK,NN,PH,FNU,FLA,TAU,EPS
K = KK
N = NN
KP1 = K+1
KP2 = KP1+1

C
C      Calculate derivative matrix
C
DO 200 J1=1,K
  IF(BV(J1).LT.0.) THEN
C
C      User must supply analytical derivative subroutine if desired
C
CALL DERV(K,B,N,Z,BC,J1,JTEST)
ELSE
  DEN = .001*ABS(BC(J1))
  BC(J1) = BC(J1) + DEN
  IF (BC(J1).GT.BMAX(J1)) THEN
    BC(J1) = BC(J1)-2*DEN
    DEN = -DEN
  ENDIF
  CALL FUNC(K,BC,X,N,ZC)
  DO 100 J2=1,N
    DERIV(J1,J2) = (ZC(J2)-Z(J2))/DEN
100  CONTINUE
    BC(J1) = BC(J1) -DEN
  ENDIF
200 CONTINUE

C
C      Calculate A matrix
C
DO 600 J1=1,K
  A(J1,KP1) = 0.0E0
  DO 300 J2=1,N
    A(J1,KP1) = A(J1,KP1)+DERIV(J1,J2)*(Y(J2)-Z(J2))
300  CONTINUE
  DO 400 J2=1,K
    A(J1,J2) = 0.0E0
    DO 400 J3 =1,N
      A(J1,J2) = A(J1,J2)+DERIV(J1,J3)*DERIV(J2,J3)
400  CONTINUE
  IF(A(J1,J1).LT.1.E-20) THEN
    DO 500 J2=1,KP1
      A(J1,J2) = 0.0E0
500  CONTINUE
    A(J1,J1) = 1.0
  ENDIF
600 CONTINUE

```

B.8 RSAP Source Code Listings (cont'd)

```
C
C      Scale equations
C
GN = 0.
DO 700 J1=1,K
    GN = GN+A(J1,KP1)*A(J1,KP1)
700 CONTINUE
DO 800 J1=1,K
    A(J1,KP2) = SQRT(A(J1,J1))
800 CONTINUE
DO 900 J1=1,K
    A(J1,KP1) = A(J1,KP1)/A(J1,KP2)
    DO 900 J2=1,K
        A(J1,J2) = A(J1,J2)/(A(J1,KP2)*A(J2,KP2))
900 CONTINUE
FL = FLA/FNU
C
C      Add lambda to the diagonals of the A matrix
C
1000 DO 1200 J1=1,K
    DO 1100 J2=1,KP1
        AC(J1,J2) = A(J1,J2)
    1100 CONTINUE
    AC(J1,J1) = AC(J1,J1)+FL
1200 CONTINUE
C
C      Solve equations
C
DO 1500 L1=1,K
    L2 = L1+1
    DO 1300 L3=L2,KP1
        AC(L1,L3) = AC(L1,L3)/AC(L1,L1)
    1300 CONTINUE
    DO 1500 L3=1,K
        IF (L1-L3.NE.0) THEN
            DO 1400 L4=L2,KP1
                AC(L3,L4) = AC(L3,L4)-AC(L1,L4)*AC(L3,L1)
            1400 CONTINUE
        ENDIF
    1500 CONTINUE
C
C      Increment parameter values and calculate lambda
C
DN = 0.
DG = 0.
DO 1600 J1=1,K
    AC(J1,KP2) = AC(J1,KP1)/A(J1,KP2)
    BC(J1) = AMAX1(BMIN(J1),AMIN1(BMAX(J1),B(J1)+AC(J1,KP2)))
    DG = DG + AC(J1,KP2)*A(J1,KP1)*A(J1,KP2)
    DN = DN+AC(J1,KP2)*AC(J1,KP2)
    AC(J1,KP2) = BC(J1)-B(J1)
1600 CONTINUE
COSG = DG/SQRT(DN*GN)
JGAM = 0
IF(COSG.LT.0.) THEN
    JGAM = 2
    COSG = -COSG
ENDIF
COSG = AMIN1(COSG,1.0)
GAMM = ARCCOS(COSG)*180./(3.14159265)
IF(JGAM.GT.0) GAMM=180-GAMM
C
C      Calculate new objective function
C
```

B.8 RSAP Source Code Listings (cont'd)

```
      CALL FUNC(K,BC,X,N,ZC)
      PHI = 0.
      DO 1700 J1=1,N
        PHI = PHI+(ZC(J1)-Y(J1))*2
1700 CONTINUE
C
C      Check for object function improvement,change lambda if
C      necessary
C
      IF (PHI.GE.PH) THEN
        IF (FL.GT.1.0E9) THEN
          ICON = -1
          RETURN
        ENDIF
        FL = FNU*FL
        GO TO 1000
      ENDIF
C
C      Calculate convergence criteria
C
      ICON = 0
      DO 1800 J1=1,K
        IF (ABS(AC(J1,KP2))/(TAU+ABS(B(J1))).GT.EPS) ICON=ICON+1
1800 CONTINUE
      IF (ICON.NE.0.AND.FL.GT.1.0.AND.GAMM.GT.90.0) ICON=-1
      IF (ICON.EQ.0.AND.FL.GT.1.0.AND.GAMM.LE.45.0) ICON=-4
      IF (PHI.LT.1.E-10) ICON=0
      FLA = FL
      DO 1900 J2=1,K
        B(J2) = BC(J2)
1900 CONTINUE
      DO 2000 J2=1,N
        Z(J2) = ZC(J2)
2000 CONTINUE
      PH = PHI
      RETURN
      END
C
C***** Function ARCOS *****
C
C      Subprogram to calculate arccos
C
      FUNCTION ARCOS(Z)
      X=Z
      KEY=0
      IF (X.LT.(-1.))X=-1.
      IF (X.GT.1.)X=1.
      IF (X.GE.-1..AND.X.LT.0.)KEY=1
      IF (X.LT.0.)X=ABS(X)
      IF (X.EQ.0.)GO TO 10
      ARCOS=ATAN(SQRT(1.-X*X)/X)
      IF (KEY.EQ.1)ARCOS=3.14159265-ARCOS
      GO TO 999
10  ARCOS=1.5707963
999 RETURN
      END
C
C***** Subroutine FUNC *****
C
      SUBROUTINE FUNC (K,B,X,N,Z)
      INTEGER I,J,K,N
      REAL X(301),B(15),LN2,PI,SLN2PI,R(5),SQDIFF,G,L,
      * EXPMAX,Z(301),SLOPE,YI,AMP(5),NU0(5),BW(5),SQBW(5)
      LN2 = ALOG(2.)
```

B.8 RSAP Source Code Listings (cont'd)

```

PI      = 22./7.
SLN2PI = SQRT(LN2/PI)
SLOPE  = B(14)
YI     = B(15)
B(14)  = 140.1
B(15)  = 1616.
DO 10 I=1,5
      I3      = I*3
      NU0(I)  = B(I3)
      BW(I)   = B(I3-1)
      SQBW(I) = BW(I)**2
      AMP(I)  = B(I3-2)
      R(I)    = 0.7143
10 CONTINUE
R(5)   = 0.0
DO 30 I=1,N
      Z(I) = 0.
      DO 20 J=1,5
        SQDIFF = (X(I)-NU0(J))**2
        EXPMAX = AMAX1(-4*LN2*SQDIFF/SQBW(J),-71.0)
        G      = (2*SLN2PI/BW(J))*EXP(EXPMAX)
        L      = BW(J)/((PI/2.)*SQBW(J)+2*PI*SQDIFF)
        Z(I)   = Z(I)+AMP(J)*(L**R(J))*(G**(1-R(J)))
20 CONTINUE
      Z(I) = Z(I) + SLOPE*X(I) + YI
30 CONTINUE
B(14) = SLOPE
B(15) = YI
RETURN
END
C
C***** Subroutine GLPROD *****
C
C      Calculation of intensities and sum of square error from
C      weighted Gaussian-Lorentzian product functions for a
C      single peak
C
SUBROUTINE GLPROD (NU0,BW,AMP,R)
INTEGER      I
REAL         INTCAL(1500:1800),NU0,BW,AMP,R,PI,LN2,
&            G,L,SQDIFF,EXPMAX,SLN2PI,SQBW
COMMON      SLOPE,YI,INTCAL
PI          = 22.0/7.0
LN2         = ALOG(2.0)
SLN2PI     = SQRT(LN2/PI)
SQBW       = BW*BW
DO 10 I=1500,1800
      SQDIFF = (I - NU0)**2
      EXPMAX = AMAX1(-4*LN2*SQDIFF/SQBW,-71.0)
      G      = (2*SLN2PI/BW)*EXP(EXPMAX)
      L      = BW/((PI/2.)*SQBW+2*PI*SQDIFF)
      INTCAL(I) = AMP*(L**R)*(G**(1-R)) + SLOPE*I + YI
10 CONTINUE
RETURN
END

```

B.8 RSAP Source Code Listings (cont'd)

```

C
C ##### Double Precision Version #####
C
C *****
C *
C *          DSTRUCT.FOR
C *
C *          Todd M. Przybycien  4/11/86
C *
C * This program computes the fraction of ordered and disordered
C * helix, parallel and antiparallel pleated sheet, reverse
C * turn and unordered structure present in samples analyzed via
C * laser Raman spectroscopy. Subtracted spectral data are
C * obtained from the output data file created by program
C * RAMANFIT.FOR. This program uses a constrained least distance
C * program LDP.FOR from C.L. Lawson and R.J. Hanson from "Solving
C * Least Squares Problems", Prentice-Hall, 1974. The fitted
C * structure fractions are constrained to be positive and must sum
C * to unity. This subroutine calls subroutines NNLS, H12, G1, G2
C * and DIFF from the above source also. Protein reference
C * information has been obtained from R.W. Williams, J. Mol. Biol.,
C * 166, 581 (1983). This information is called from reference
C * data files which contain matrices derived from reference
C * compound spectral data and x-ray structure data.
C *
C *****
C
C          Variables
C
C          INTEGER INDEX(13), MODE
C          REAL NU(15),INT
C          DOUBLE PRECISION B(15),X(6),W(66),RNORM,F(6),
C          & FTOT,SHELIX,SSHEET,
C          & AFINV(15,6),G(6,6),GVSU(6,15),VS(6,6),U1T(6,15),
C          & H(6),HTMP,XNORM,U1TB(6),XU1TB(6),AFINV(15)
C          CHARACTER SAMPLE*8,FMT*9,ST1*3,ST2*7,ST3
C
C          Constants
C
C          ST1 = 'OLD'
C          ST2 = 'UNKNOWN'
C          ST3 = 'NEW'
C          FMT = 'FORMATTED'
C          WRITE (*,16)
C 16 FORMAT (//1X,'RAMAN SPECTRAL ANALYSIS PACKAGE:',
C          & /1X,'Secondary Structure Estimation Routine')
C
C          Read sample code from file SAMPLE.COD
C
C          OPEN (1,FILE= 'SAMPLE.COD', FORM=FMT, STATUS=ST1)
C          READ (1,26) SAMPLE
C 26 FORMAT (A8)
C          CLOSE (1)
C
C          Read normalized sample spectrum, convert to double precision
C
C          OPEN (1,FILE= SAMPLE//'.NOR', FORM=FMT, STATUS=ST1)
C          DO 40 I=1,15
C          READ (1,36) NU(I),INT
C          FORMAT (2E15.6)
C          B(I)=DBLE(INT)
C 40 CONTINUE
C          CLOSE (1)
C

```

B.8 RSAP Source Code Listings (cont'd)

```
C          Load constraint and reference matrix files
C
OPEN (1,FILE= 'AFINV.REF',FORM= FMT, STATUS= ST1)
READ (1,50) ((AFINV(I,J),J=1,6),I=1,15)
50 FORMAT (1X,6D23.14)
CLOSE (1)
OPEN (1, FILE= 'H.REF', FORM= FMT, STATUS= ST1)
READ (1,60) (H(I),I=1,8)
60 FORMAT (1X,D23.14)
CLOSE (1)
OPEN (1, FILE= 'GVSU.REF', FORM= 'FORMATTED', STATUS= 'OLD')
READ (1,70) ((G(I,J),J=1,6),I=1,8)
70 FORMAT (1X,6D23.14)
CLOSE (1)
OPEN (1, FILE= 'GVSU.REF', FORM= 'FORMATTED', STATUS= 'OLD')
READ (1,80) ((GVSU(I,J),J=1,15),I=1,8)
80 FORMAT (1X,15D23.14)
CLOSE (1)
OPEN (1, FILE= 'VS.REF', FORM= 'FORMATTED', STATUS= 'OLD')
READ (1,90) ((VS(I,J),J=1,6),I=1,6)
90 FORMAT (1X,6D23.14)
CLOSE (1)
OPEN (1, FILE= 'UIT.REF', FORM= 'FORMATTED', STATUS= 'OLD')
READ (1,100) ((UIT(I,J),J=1,15),I=1,6)
100 FORMAT (1X,15D23.14)
CLOSE (1)

C          Calculate elements of vector H
C
DO 120 I=1,8
  HTMP = 0.
  DO 110 J=1,15
    HTMP = HTMP + GVSU(I,J)*B(J)
110  CONTINUE
  H(I) = H(I)-HTMP
120 CONTINUE

C          Call Least Distance Programming Subroutine
C
CALL LDP (G,8,8,6,H,X,XNORM,W,INDEX,MODE)
IF (MODE.EQ.1) THEN
  WRITE (*,*) 'LDP routine successful'
ELSEIF (MODE.EQ.2) THEN
  WRITE (*,*) 'Bad dimension, N=8'
  GO TO 999
ELSEIF (MODE.EQ.3) THEN
  WRITE (*,*) 'Excessive iterations in subroutine NNLS'
  GO TO 999
ELSE
  WRITE (*,*) 'Constraints are incompatible'
  GO TO 999
ENDIF

C          Calculate solution vector F
C
DO 140 I=1,8
  UITB(I)=0.
  DO 130 J=1,15
    UITB(I) = UITB(I)+UIT(I,J)*B(J)
130  CONTINUE
  XU1TB(I) = X(I)+UITB(I)
140 CONTINUE
FTOT = 0.
DO 160 I=1,8
```

B.8 RSAP Source Code Listings (cont'd)

```

      F(I) = 0.
      DO 150 J=1,6
        F(I) = F(I)+VS(I,J)*XU1TB(J)
150    CONTINUE
      FTOT = FTOT + F(I)
160  CONTINUE
C
C      Calculate residual norm and fit spectrum, A(F+)f
C
      RNORM = 0.
      DO 180 I=1,15
        AFINVF(I)=0.
        DO 170 J=1,6
          AFINVF(I) = AFINVF(I)+AFINV(I,J)*F(J)
170    CONTINUE
      RNORM = RNORM + (B(I)-AFINVF(I))**2
180  CONTINUE
      RNORM = DSQRT(RNORM)
C
C      Calculate structure class fractions
C
      SHELIX = F(1)+F(2)
      SSHEET = F(3)+F(4)
      SRNDM  = F(5)+F(6)
C
C      If successful, create output data file, echo to monitor
C
      OPEN (1, FILE= SAMPLE//'.STR', FORM=FMT, STATUS=ST2)
      WRITE (1,190) SAMPLE
      WRITE (*,190) SAMPLE
190  FORMAT (/1X,'Structural content of sample: ',A8,/)
      WRITE (1,200) F(1),F(2),SHELIX,F(3),F(4),SSHEET,F(5),F(6),
& SRNDM,FTOT
      WRITE (*,200) F(1),F(2),SHELIX,F(3),F(4),SSHEET,F(5),F(6),
& SRNDM,FTOT
200  FORMAT (1X,'IMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM;',
& /1X,': structure type 3 fraction 3 type total :',
& /1X,'LMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM9',
& /1X,': ord -helix ',F8.4,' :',
& /1X,': unord -helix ',F8.4,' :',
& /1X,': total -helix ',F8.4,' :',
& /1X,'GDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD6',
& /1X,': antipar a-sheet ',F8.4,' :',
& /1X,': par a-sheet ',F8.4,' :',
& /1X,': total a-sheet ',F8.4,' :',
& /1X,'GDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD6',
& /1X,': a-reverse turn ',F8.4,' :',
& /1X,': unord structure ',F8.4,' :',
& /1X,': total random coil ',F8.4,' :',
& /1X,'LMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM9',
& /1X,': d structure fractions ',F8.4,' :',
& /1X,'HMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM<')
      WRITE (1,210) RNORM
      WRITE (*,210) RNORM
210  FORMAT (//1X,'Euclidean norm of residual vector = ',E15.6/)
      CLOSE (1)
C
C      A(F+)f represents the calculated spectrum, write to file
C      sample.CAL
C
      OPEN (1,FILE= SAMPLE//'.CAL', FORM= FMT, STATUS= ST2)
      WRITE (1,220) (NJ(I),AFINVF(I),I=1,15)
220  FORMAT (2E15.6)
      CLOSE (1)

999  STOP 'Normal termination of DSTRUCT.EXE'
      END
```

B.8 RSAP Source Code Listings (cont'd)

```
C *****
C *
C *                               JOB DONE.FOR                               *
C *                               Todd M. Przybycien  8/10/87                    *
C *                               This program displays a message indicating that the *
C *                               program is finished executing.                  *
C *
C *****
C      WRITE (*,10)
C      10 FORMAT (6X, '[',
C      &          /6X, ' Raman spectral analysis completed. Computer',
C      &          /6X, ' may be turned off or used by someone else.',
C      &          /6X, ' [',
C      &          /6X, ' ////////////////////////////////////////////////////////////////////',
C      &          /6X, ' ]')
C      STOP
C      END
```

```
CLS
ECHO ON
REM Batch file DATADISK.BAT: Creates data directories on a diskette
PAUSE Insert a formatted diskette into drive A
A:
MD RAWDATA
MD SMOOTH
MD SOLVENT
CD SOLVENT
MD PARAMTR
CD \
MD AROMATIC
CD AROMATIC
MD FITPEAKS
MD PARAMTR
CD \
MD STRUCTUR
C:
```

```
REM ***** SETUP.BAT *****
REM This batch file sets up the appropriate directories
REM and copies the executable image files to the hard disk
REM for the RAMAN SPECTRAL ANALYSIS PACKAGE
C:
MD RAMAN
CD RAMAN
MD RAWDATA
MD SMOOTH
MD SOLVENT
CD SOLVENT
MD PARAMTR
CD \RAMAN
MD AROMATIC
CD AROMATIC
MD FITPEAKS
MD PARAMTR
CD \RAMAN
MD STRUCTUR
CD STRUCTUR
MD REF
CD REF
COPY A:*.REF
CD \RAMAN
COPY A:*.EXE
COPY A:ANALYZE.BAT
REM RAMAN SPECTRAL ANALYSIS PACKAGE: Installed
```

Appendix C

Stopped-Flow Device

C.1 Schematic Diagram Description

The following parts are labelled on the schematic diagram of the stopped-flow device:

- 1 drive ram pneumatic actuator button
- 2 drive ram pneumatic reset button
- 3 flow volume adjustment micrometer
- 4 pneumatic drive ram
- 5 recieving syringe
- 6 drive syringe
- 7 drive syringe
- 8 drain valve, open (arrow up)/closed
- 9 reactant valve, open (arrow up)/closed
- 10 reactant valve, open (arrow up)/closed
- 11 compressed gas fitting (keep pressure \leq 15 psi to avoid breaking Wilmad WG-41MH flow cell)

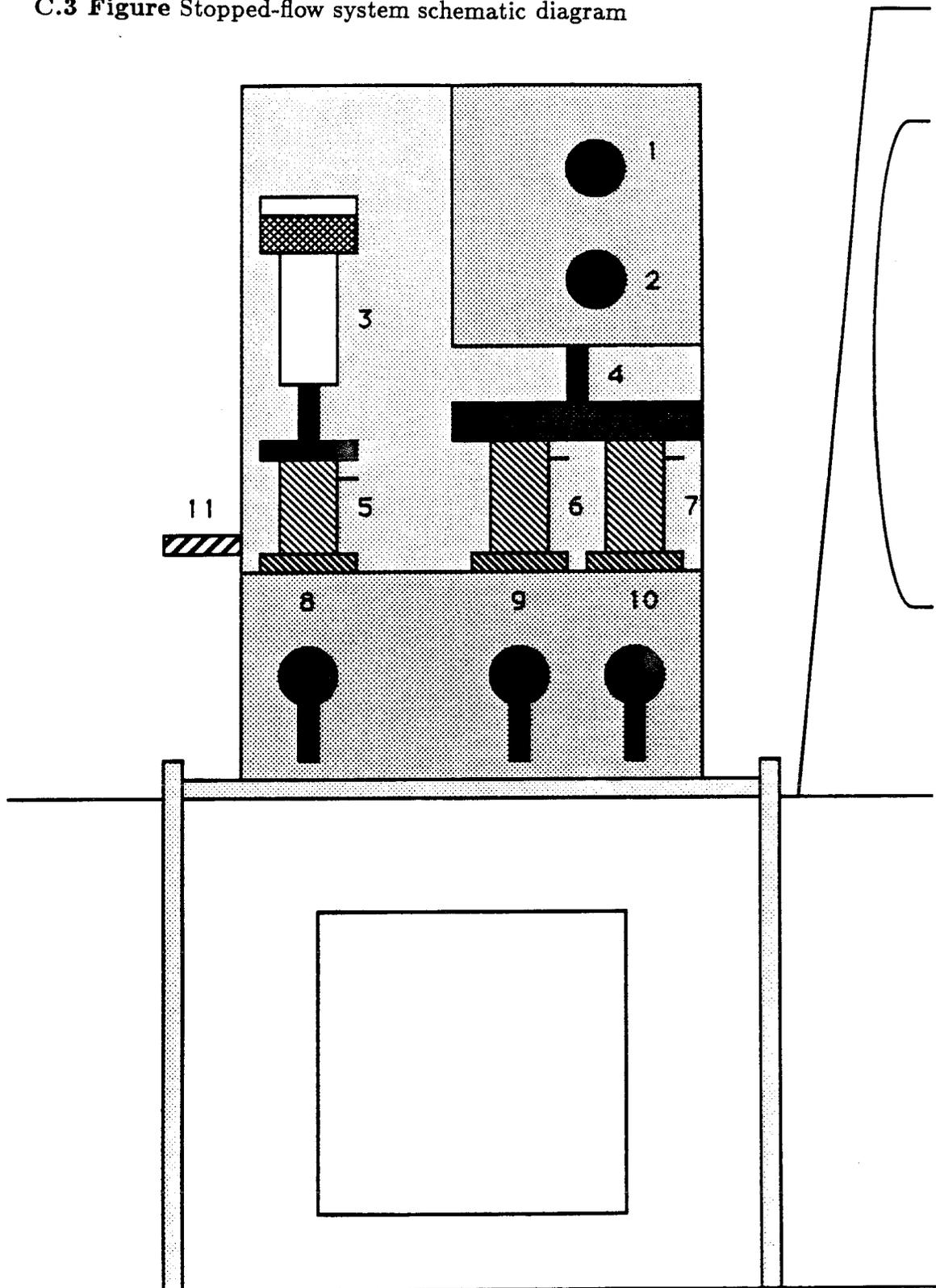
The following parts are not shown on the diagram: Hamilton 5 mL luer-lock syringe barrels may be attached directly to valve bodies (9) and (10) in front of the drive syringe assemblies (6) and (7), tubing may be attached to valve body (8) in front of recieving syringe assembly (5) to drain the system, the mixing tee is mounted on the bottom of the device bed, and inlet and outlet fittings on the back of the syringe block allow the drive syringes to be thermostatted via a recirculating bath.

C.2 Operation Instructions

Basic operation procedures consist of three steps: loading reagents, shooting reagents, and flushing the flow cell. To load the reagents, retract the drive ram

(4) by depressing and holding reset button (2). Plungers for syringes (5), (6), and (7) should be pushed all the way down into the barrels and valves (8). (9), and (10) should be closed. Fill the Hamilton syringe barrels with reactants. Open valve (9) and draw reactant from the open Hamilton barrel into the drive syringe (6) by withdrawing the plunger; close valve (9) when done. Repeat for the other reactant with valve (10) and drive syringe (7). To shoot the reactants into the flow cell, depress and hold actuator button (1) until receiving syringe plunger (5) contacts the adjustable mechanical stop (3). Acquire data with spectrometer. To flush the mixture, open valve (8), depress receiving syringe plunger (5) and close valve (8).

C.3 Figure Stopped-flow system schematic diagram



Appendix D

Shimadzu UV-260 Data Acquisition Program
Source Code Listing

D.1 Source Code for UV-260 Data Acquisition

```
10 '*****
20 '*
30 '          UVBLOCK.BAS
40 '*
50 '          Todd M. Przybycien    8/9/88
60 '*
70 '          This program reads data from the UV-260 8Kb memory
80 '* (0-3999 2-byte data points) and writes it to a file. IEEE-488
90 '* communications protocol based on SCAN450.BAS by Ken Reardon.
100 '* Halo graphics display based on TIMEI.BAS by Jorge Galazzo.
110 '*
120 '*****
130 '
140 '-----
150 ' HALOI.BAS program statements necessary for HALO in interpreted
160 ' BASIC
170 '-----
171 KEY OFF:CLS
172 GOSUB 8000 ' Load Halo offsets
175 DEF SEG = &H1FC0
180 BLOAD "HALOI.BIN",0
185 DEVICES="HALOHERC.DEV"
190 CALL SETDEV(DEVICES)
195 MODEX=0:CALL INITGRAPHICS(MODEX)
200 MODEX=1:CALL SETSCREEN(MODEX)
205 SFONTS="HALO202.FNT":CALL SETFONT(SFONTS)
210 COX=1:FIX=0:CALL SETSTCLR(COX,FIX)
215 HEX=70:AS=1:PA%=0:CALL SETSTEXT(HEX,AS,PA%)
220 XT%=190:YT%=200:CALL MOVTCURABS(XT%,YT%)
225 FT%="UV-260":CALL STEXT(FT%)
230 XT%=120:YT%=240:CALL MOVTCURABS(XT%,YT%)
235 HEX=25:CALL SETSTEXT(HEX,AS,PA%)
240 FT%="Block data transfer to PC-XT":CALL STEXT(FT%):CALL DELTCUR
245 FOR DELAY=1 TO 500:NEXT DELAY
250 CALL CLOSEGRAPHICS
290 KEY OFF:CLS
400 '-----
410 ' Display spectrometer turn-on message
420 '-----
430 LOCATE 5,5:PRINT "This program initiates a scan on the UV-260"
440 LOCATE 6,5:PRINT "and writes the data to a disk file."
445 LOCATE 8,5:PRINT "Wavelength and time based scanning are supported."
450 LOCATE 10,5:PRINT "Data file names are updated to allow a series of"
455 LOCATE 11,5:PRINT "scans to be taken."
470 LOCATE 15,20:PRINT "HIT ANY KEY WHEN READY TO CONTINUE"
480 AS=INKEY$:IF AS="" THEN 480
490 KEY ON
500 '-----
510 ' Initialize scan parameters and display parameter menu
520 '-----
525 FCNT=1:FCNTS="1":FILES="UVDATA"
530 CLS
532 SCANS="t":SCANX=-1:SCANAS="w":MS="3":MAS="1"
534 SCALEX=5:SCALE$=RIGHT$(STR$(SCALEX),1):S=40.0
536 INTVL=0.075:INTVL$=".075"
540 STRT$="0":STRT=0.0:STPS="10":STP= 10.0:MAX$="300":MAX=300.0
547 INTXT$=" sec":INTXTAS=" nm "
545 SCTXT$="time (seconds)":SCTXTAS="wavelength (nm)"
550 OPER$="A":OPERAS="T":OPERX=1
560 LOCATE 2,5:PRINT CHR$(201)+STRING$(29,205)+CHR$(187)
570 LOCATE 3,5:PRINT CHR$(186):LOCATE 3,35:PRINT CHR$(186)
580 LOCATE 4,5
590 PRINT CHR$(186)+" Scan Conditions Set-up Menu "+CHR$(186)
600 LOCATE 5,5:PRINT CHR$(186):LOCATE 5,35:PRINT CHR$(186)
```

D.1 Source Code for UV-260 Data Acquisition (cont'd)

```
810 LOCATE 8,5:PRINT CHR$(200)+STRING$(29,205)+CHR$(180)
820 LOCATE 9,5:PRINT "1. Scan mode (w=wavelength, t=time):"
830 LOCATE 10,5:PRINT "2. Scale (1=2.5,2=5,3=10,4=20,5=40 nm/cm)"
835 LOCATE 10,40:PRINT "interval =" +INTVL$+INTXT$
840 LOCATE 11,5:PRINT "3. Starting "+SCTXT$+":"
850 LOCATE 12,5:PRINT "4. Ending "+SCTXT$+" max= "+MAX$+" : "
860 LOCATE 13,5:PRINT "5. Operation mode (A=abs, T=%T):"
870 LOCATE 14,5:PRINT "6. Output file (6 char [##.DAT]):"
880 LOCATE 9,70:PRINT SCANS
890 LOCATE 10,70:PRINT SCALES
900 LOCATE 11,70:PRINT STRTS
910 LOCATE 12,70:PRINT STPS
920 LOCATE 13,70:PRINT OPER$
930 LOCATE 14,70:PRINT FILES+FCNT$+".DAT"
940 LOCATE 20,5:PRINT "Enter number of field to change: "
950 LOCATE 22,5:PRINT "F1 quits set-up menu"
960 '-----
970 ' Interactive scan parameter selection
980 '-----
810 KEY 1, ""
820 LOCATE 20,38:PRINT CHR$(95):LOCATE 20,38
830 A$=INKEY$:IF LEN(A$)=0 GOTO 830
840 LOCATE 20,38:PRINT CHR$(32)
850 IF LEN(A$)=2 THEN QUIT$=RIGHT$(A$,1) ELSE GOTO 870
860 QUIT=ASC(QUIT$):IF QUIT=59 GOTO 1280 ELSE GOTO 820
870 A=VAL(A$):IF (A<1) OR (A>6) GOTO 820
880 ON A GOTO 890,920,960,1010,1040,1070
890 SCAN$=-SCAN$:SWAP SCANS,SCANAS:SWAP SCTXT$,SCTXTAS:SWAP MS,MAS
900 SWAP INTXT$,INTXTAS
905 LOCATE 9,70:PRINT SCANS
910 GOTO 4000
920 LOCATE 10,70:PRINT CHR$(32):LOCATE 10,70
930 INPUT "",SCALE$:SCALE$=RIGHT$(STR$(SCALE$),1)
940 IF (SCALE$<1) OR (SCALE$>5) GOTO 920
950 GOTO 4000
960 LOCATE 11,70:PRINT STRING$(3,32):LOCATE 11,70
970 INPUT "",STRT$:STRT=VAL(STRT$)
980 IF (SCAN=1) AND ((STRT<191) OR (STRT>900)) GOTO 960
990 IF (SCAN=-1) AND (STRT<>0) GOTO 960
1000 GOTO 4000
1010 LOCATE 12,70:PRINT STRING$(4,32):LOCATE 12,70
1020 INPUT "",STP$:STP=VAL(STP$)
1030 GOTO 4000
1040 OPER$=-OPER$:SWAP OPER$,OPERAS
1050 LOCATE 13,70:PRINT OPER$
1060 GOTO 820
1070 LOCATE 14,70:PRINT STRING$(6,95):LOCATE 14,70
1080 INPUT "",FILES:FCNT$=STR$(FCNT):FCNT$=RIGHT$(FCNT$,LEN(FCNT$)-1)
1090 LOCATE 14,70:PRINT STRING$(6,32)
1100 LOCATE 14,70:PRINT FILES+FCNT$+".DAT"
1110 GOTO 820
1240 '-----
1250 ' Initialize the PC(>488) interface
1260 '-----
1280 DEF SEG = &HC000
1290 INITIALIZE=0:TRANSMIT=3:RECEIVE=6
1300 SYS.CNTRLR%=0:PC.ADDR%=6:ADDR%=9
1310 CALL INITIALIZE (PC.ADDR%,SYS.CNTRLR%)
1320 STS%=0
1330 HDR$="MTA LISTEN 9 DATA '":EOIS="" END":TALK$="MLA TALK 9"
1340 '-----
1342 ' Set the UV-260 memory data capture flag, operation mode and
1344 ' wavelength scale.
1346 '-----
```

D.1 Source Code for UV-260 Data Acquisition (cont'd)

```
1350 IO%=1:CMD%=HDR% + "G44" + EOIS:GOSUB 6000
1360 IO%=2:CMD%=HDR% + "K" + SCALES + EOIS:GOSUB 6000
1370 IO%=3:CMD%=HDR% + "U" + MS + EOIS:GOSUB 6000
1380 '-----
1382 ' Display spectrometer set-up message, start scan
1384 '-----
1390 KEY OFF:CLS
1400 LOCATE 5,5:PRINT "Check the status of the following:"
1410 LOCATE 6,10:PRINT "- UV-260 is ON"
1420 LOCATE 7,10:PRINT "- ext. c. set to 9"
1430 LOCATE 8,10:PRINT "- all scan parameters set manually on the"
1440 LOCATE 9,10:PRINT " UV-260 as the computer only starts the scan"
1450 LOCATE 10,10:PRINT "- diskette inserted into drive A"
1460 LOCATE 11,5:PRINT "Note: Set data format to 2 bytes/point in the"
1470 LOCATE 12,5:PRINT "data proc. mode"
1480 LOCATE 20,40:PRINT "To start scan, hit any key."
1490 A%=INKEY$: IF LEN(A%)=0 GOTO 1490
1500 IO%=4:CMD%=HDR% + "A" + EOIS:GOSUB 6000
1510 CLS:LOCATE 7,5:PRINT "Spectrometer is now scanning."
1511 BEEP
1512 '-----
1514 ' Time scan mode, interrupt scan after end of scan interval
1516 '-----
1520 IF SCAN%=1 GOTO 1590
1530 TIMEOUT=TIMER+STP:WHILE TIMER<TIMEOUT:WEND
1550 IO%=5:CMD%=HDR% + "A" + EOIS:GOSUB 6000
1560 GOTO 1770
1570 '-----
1572 ' Wavelength scan mode, test for end of scan
1574 '-----
1590 SCANNING=1
1600 WHILE SCANNING
1610     TIMEOUT = TIMER + 5
1620     WHILE TIMER<TIMEOUT:WEND
1630     IO%=6:CMD% = HDR% + "B" + EOIS:GOSUB 6000
1670     IF LEFT$(A%,2)=(CHR$(6)+"0") THEN SCANNING=0
1720 WEND
1730 '-----
1740 ' Compute the number of points to be transferred.
1760 '-----
1770 LOCATE 7,5:PRINT "Data transfer to program memory in progress."
1780 BEEP
1810 NPTS%=INT(SCAN%*(STRT-STP)/INTVL)+1:NPTS%=STR$(NPTS%)
1820 NPTS%=RIGHT$(NPTS%, (LEN(NPTS%)-1)):NPTS%=NPTS%-1
1830 '-----
1840 ' Reset the memory flag, request block data transfer, clear
1850 ' data space and transfer data to program
1860 '-----
1890 DIM D(NPTS%+10)
1900 IO%=7
1910 FOR I = 0 TO NPTS%
1920     PTR%=RIGHT$(STR$(I), LEN(STR$(I))-1)
1930     CMD%=HDR% + "<" + PTR% + "=" + EOIS:GOSUB 6000
1940     D%=RIGHT$(A%,9):D(I)=VAL(D%)
1960 NEXT I
1970 IO%=8:CMD% = HDR% + "G46" + EOIS:GOSUB 6000
2000 '-----
2070 ' Compute wavelength/time increments, convert ASCII data to
2080 ' binary and write it to the data file, erase data array so that
2090 ' it may be redimensioned if another run is desired. Error
2095 ' trapping is invoked to avoid loss of data due to absent or
2097 ' full data diskette.
2100 '-----
2110 IF OPER%=1 THEN NORM=8.0/65536.0 ELSE NORM=400.0/65536.0
```

D.1 Source Code for UV-260 Data Acquisition (cont'd)

```
2130 FLNAMS="A:"+FILES+FCNTS+".DAT"
2140 LOCATE 7,5:PRINT "Data transfer to file ";FLNAMS;" in progress. "
2145 BEEP
2147 ON ERROR GOTO 5000
2150 OPEN FLNAMS FOR OUTPUT AS #1
2160 FOR I = 0 TO NPTS%
2170     Y=NORM*D(I)
2180     X=STRT-(SCAN%*I+INTVL)
2190     PRINT #1,USING"#####.####"    ";Y;X
2200 NEXT I
2210 CLOSE #1
2220 ERASE D
2225 ON ERROR GOTO 0
2230 KEY OFF:CLS
2240 '-----
2250 ' Query for another run, with or without parameter changes
2260 '-----
2265 BEEP:BEEP:BEEP
2270 LOCATE 5,5:PRINT CHR$(201)+STRING$(20,205)+CHR$(187)
2280 LOCATE 6,5:PRINT CHR$(186)+STRING$(20,32)+CHR$(186)
2290 LOCATE 7,5:PRINT CHR$(186)+" Data transfer completed. "+CHR$(186)
2300 LOCATE 8,5:PRINT CHR$(186)+STRING$(20,32)+CHR$(186)
2310 LOCATE 9,5:PRINT CHR$(186)+" Collect more data? ([Y]/N) "+CHR$(186)
2320 LOCATE 10,5:PRINT CHR$(186)+STRING$(20,32)+CHR$(186)
2330 LOCATE 11,5:PRINT CHR$(200)+STRING$(20,205)+CHR$(188)
2340 AS=INKEY$:IF LEN(AS)=0 GOTO 2340
2350 IF AS="N" OR AS="n" THEN CLS:END
2360 FCNT=FCNT+1
2365 FCNTS=STR$(FCNT):FCNTS=RIGHT$(FCNTS,LEN(FCNTS)-1)
2370 LOCATE 9,7:PRINT "Change parameters? (Y/[N])"
2380 AS=INKEY$:IF LEN(AS)=0 GOTO 2380
2390 IF AS="Y" OR AS="y" GOTO 530 ELSE GOTO 1310
3990 '-----
3992 ' Set-up menu parameter updating subroutine
3994 '-----
4000 LOCATE 11,18:PRINT SCTXT$:LOCATE 12,16:PRINT SCTXT$
4002 LOCATE 10,48:PRINT "interval ="*INTVL$+INTXT$+STRING$(3,32)
4004 IF SCAN%=-1 GOTO 4120
4010 ON SCALE% GOTO 4020,4030,4040,4050,4060
4020 S=2.5:INTVL=.025:GOTO 4070
4030 S= 5:INTVL=.05 :GOTO 4070
4040 S= 10:INTVL=.1 :GOTO 4070
4050 S= 20:INTVL=.2 :GOTO 4070
4060 S= 40:INTVL=.2 :GOTO 4070
4070 INTVL$=STR$(INTVL)
4072 LOCATE 10,48:PRINT "interval ="*INTVL$+INTXT$+STRING$(3,32)
4075 IF STRT=0.0 GOTO 900
4080 MAX=STRT-(INTVL*4000):IF MAX<190 THEN MAX=190
4090 MAX$=STR$(MAX):LOCATE 12,37:PRINT MAX$+STRING$(2,32)
4100 IF (STP<MAX) OR (STP>=STRT) GOTO 1010
4110 GOTO 4220
4120 ON SCALE% GOTO 4130,4140,4150,4160,4170
4130 S=2.5:INTVL=1.2 :GOTO 4180
4140 S= 5:INTVL=.6 :GOTO 4180
4150 S= 10:INTVL=.3 :GOTO 4180
4160 S= 20:INTVL=.15 :GOTO 4180
4170 S= 40:INTVL=.075:GOTO 4180
4180 INTVL$=STR$(INTVL)
4182 LOCATE 10,48:PRINT "interval ="*INTVL$+INTXT$+STRING$(3,32)
4190 IF STRT<>0 GOTO 900
4200 MAX=INTVL*4000:MAX$=STR$(MAX):LOCATE 12,37:PRINT MAX$+STRING$(2,32)
4210 IF (STP>MAX) OR (STP<=0) GOTO 1010
4220 LOCATE 10,48:PRINT "interval ="*INTVL$+INTXT$+STRING$(3,32)
4230 LOCATE 10,70:PRINT SCALE$
```

D.1 Source Code for UV-260 Data Acquisition (cont'd)

```
4240 LOCATE 11,18:PRINT SCTXT$:LOCATE 11,70:PRINT STRT$
4250 LOCATE 12,18:PRINT SCTXT$:LOCATE 12,37:PRINT MAX$+STRING$(2,32)
4260 LOCATE 12,70:PRINT STP$
4270 GOTO 820
4970 '
4980 ' Data diskette I/O error processing
4990 '-----
5000 IF ERR=71 GOTO 5020
5010 IF ERR=81 GOTO 5060
5020 LOCATE 9,5:PRINT "Diskette is not ready, hit any key when ready"
5030 A$=INKEY$:IF LEN(A$)=0 GOTO 5030
5040 LOCATE 9,5:PRINT STRING$(45,32)
5050 RESUME 2150
5060 LOCATE 9,5:PRINT "Diskette is full, hit any key when replaced"
5070 A$=INKEY$:IF LEN(A$)=0 GOTO 5070
5080 LOCATE 9,5:PRINT STRING$(45,32)
5090 RESUME 2150
5990 '-----
5992 ' Interface transmit and receive subroutine
5994 '-----
6000 CALL TRANSMIT (CMD$,STS%)
6010 IF STS%<>0 THEN IO$="transmit command status":GOSUB 7000
6020 CALL TRANSMIT (TALK$,STS%)
6030 IF STS%<>0 THEN IO$="transmit listen status":GOSUB 7000
6040 A$=SPACE$(10)
6050 CALL RECEIVE (A$,LENGTH$,STS%)
6060 IF STS%<>0 THEN IO$="receive acknowledgement status":GOSUB 7000
6070 IF LEFT$(A$,1)<>CHR$(8) THEN IO$="acknowledgement":GOSUB 7000
6080 RETURN
6990 '-----
6992 ' Interface error processing subroutine
6994 '-----
7000 CLS
7010 LOCATE 7,5:PRINT "An interface error occurred at location #";IO%
7020 LOCATE 8,5:PRINT "Nature of error: "+IO$
7030 LOCATE 9,5:PRINT "The error status code was: STS% = ";STS%
7040 LOCATE 11,5:PRINT "Consult the CEC PC<>488 manual for details."
7050 END
7990 '-----
7992 ' Halo offset declaration subroutine
7998 '-----
8000 ARC = &H85
8001 BAR = &H28B
8002 BOX = &HA1
8003 CFREEZE = &H2B5
8004 CGRAB = &H2BC
8005 CIR = &H77
8006 CLOSEGRAPHICS = &H2C3
8007 CLR = &H2CA
8008 COMLINE = &H483
8009 CSETALUT = &H2D8
8010 CSETLUT = &H2D1
8011 CSNAP = &H2DF
8012 DEFHATCHSTYLE = &H80
8013 DEFNLNSTYLE = &H532
8014 DELBOX = &H21B
8015 DELCIR = &H222
8016 DELHCUR = &H292
8017 DELLN = &H229
8018 DELTCUR = &H299
8019 DISPLAY = &H2E6
8020 ELLIPSE = &H70
8021 FCIR = &H93
8022 FCLR = &H2ED
```

D.1 Source Code for UV-260 Data Acquisition (cont'd)

| | | | |
|------|--------------|---|-------|
| 8023 | FILL | = | &H9A |
| 8024 | FLOOD | = | &HA8 |
| 8025 | FLOOD2 | = | &HAF |
| 8026 | FTCOLOR | = | &H103 |
| 8027 | FTEXT | = | &H10A |
| 8028 | FTINIT | = | &H111 |
| 8029 | FTLOCATE | = | &H118 |
| 8030 | FTSIZE | = | &H11F |
| 8031 | GPRINT | = | &H1B9 |
| 8032 | GREAD | = | &H1D6 |
| 8033 | GSCAN | = | &H20F |
| 8034 | GWRITE | = | &H1DC |
| 8035 | HALLOC | = | &H48A |
| 8036 | HFREE | = | &H491 |
| 8037 | IMREST | = | &H2FB |
| 8038 | IMSAVE | = | &H2F4 |
| 8039 | INITALPHA | = | &H309 |
| 8040 | INITGRAPHICS | = | &H302 |
| 8041 | INITHCUR | = | &H2A0 |
| 8042 | INITLP | = | &H310 |
| 8043 | INITMARKER | = | &H501 |
| 8044 | INITTCUR | = | &H2A7 |
| 8045 | INQAPAL | = | &H317 |
| 8046 | INQARC | = | &H8C |
| 8047 | INQAREA | = | &H86 |
| 8048 | INQASP | = | &H69 |
| 8049 | INQBKND | = | &H4D |
| 8050 | INQCLR | = | &H62 |
| 8051 | INQCRANGE | = | &H1AB |
| 8052 | INQDEV | = | &H19D |
| 8053 | INQDISPLAY | = | &H326 |
| 8054 | INQDRANGE | = | &H1B2 |
| 8055 | INQERR | = | &H54 |
| 8056 | INQFT | = | &H126 |
| 8057 | INQFTCOLOR | = | &H12D |
| 8058 | INQFUN | = | &H31E |
| 8059 | INQGCUR | = | &H5B |
| 8060 | INQHCUR | = | &H3F |
| 8061 | INQINTERLACE | = | &H32C |
| 8062 | INQLOCATOR | = | &H25A |
| 8063 | INQLPA | = | &H333 |
| 8064 | INQLPG | = | &H33A |
| 8065 | INQMARKER | = | &H508 |
| 8066 | INQMODE | = | &H341 |
| 8067 | INQPAL | = | &H348 |
| 8068 | INQPFLN | = | &H34F |
| 8069 | INQPRN | = | &H1C0 |
| 8070 | INQRGB | = | &H356 |
| 8071 | INQSCAN | = | &H276 |
| 8072 | INQSCREEN | = | &H35D |
| 8073 | INQSTANG | = | &H4F3 |
| 8074 | INQSTAT | = | &H364 |
| 8075 | INQSTEXT | = | &H4DE |
| 8076 | INQSTSIZE | = | &H4EC |
| 8077 | INQTCUR | = | &HF5 |
| 8078 | INQTEXT | = | &HE0 |
| 8079 | INQTSIZE | = | &HFC |
| 8080 | INQVBW | = | &H36B |
| 8081 | INQVERSION | = | &H1A4 |
| 8082 | INQVIEWPORT | = | &H1B8 |
| 8083 | INQWORLD | = | &H1B1 |
| 8084 | LAPPEND | = | &H4AD |
| 8085 | LCLOSE | = | &H4A6 |
| 8086 | LNABS | = | &H15 |

D.1 Source Code for UV-260 Data Acquisition (cont'd)

| | | | |
|------|-------------|---|-------|
| 8087 | LNJOINT | = | &H547 |
| 8088 | LNREL | = | &H1C |
| 8089 | LOPEN | = | &H49F |
| 8090 | LRECORD | = | &H4B4 |
| 8091 | LREST | = | &H55C |
| 8092 | LSAVE | = | &H56A |
| 8093 | LSETUP | = | &H563 |
| 8094 | LSWITCH | = | &H48B |
| 8095 | MAPDTON | = | &H17A |
| 8096 | MAPDTOW | = | &H15E |
| 8097 | MAPNTOD | = | &H173 |
| 8098 | MAPNTOW | = | &H16C |
| 8099 | MAPWTOD | = | &H157 |
| 8100 | MAPWTON | = | &H165 |
| 8101 | MARKERABS | = | &H50F |
| 8102 | MARKERREL | = | &H516 |
| 8103 | MEMCOM | = | &H1E3 |
| 8104 | MEMEXP | = | &H1EA |
| 8105 | MEMMOV | = | &H1F1 |
| 8106 | MONO | = | &H372 |
| 8107 | MOVABS | = | &H23 |
| 8108 | MOVEFROM | = | &H379 |
| 8109 | MOVEFX | = | &H387 |
| 8110 | MOVETO | = | &H380 |
| 8111 | MOVETX | = | &H38E |
| 8112 | MOVHCURABS | = | &H31 |
| 8113 | MOVHCURREL | = | &H38 |
| 8114 | MOVREL | = | &H2A |
| 8115 | MOVTCURABS | = | &HE7 |
| 8116 | MOVTCURREL | = | &HEE |
| 8117 | ORGLOCATOR | = | &H253 |
| 8118 | PAN | = | &H395 |
| 8119 | PEXPAND | = | &H3A3 |
| 8120 | PFNORM | = | &H3B1 |
| 8121 | PIE | = | &H7E |
| 8122 | POLYCABS | = | &H54E |
| 8123 | POLYCREL | = | &H555 |
| 8124 | POLYFABS | = | &H230 |
| 8125 | POLYFREL | = | &H237 |
| 8126 | POLYLNABS | = | &H524 |
| 8127 | POLYLNREL | = | &H52B |
| 8128 | PREAD | = | &H196 |
| 8129 | PROTATE | = | &H39C |
| 8130 | PTABS | = | &H7 |
| 8131 | PTNORM | = | &H3AA |
| 8132 | PTREL | = | &HE |
| 8133 | PWRITE | = | &H18F |
| 8134 | RBOX | = | &H1FF |
| 8135 | RCIR | = | &H214 |
| 8136 | READLOCATOR | = | &H24C |
| 8137 | RECOLOR | = | &H3B8 |
| 8138 | REPPAL | = | &H3BF |
| 8139 | RLNABS | = | &H286 |
| 8140 | RLNREL | = | &H200 |
| 8141 | ROAM | = | &H3C6 |
| 8142 | SAVOYL | = | &H268 |
| 8143 | SCROLL | = | &H3CD |
| 8144 | SETACAM | = | &H3D4 |
| 8145 | SETAMASK | = | &H3DB |
| 8146 | SETAPAL | = | &H3E2 |
| 8147 | SETASP | = | &H2AE |
| 8148 | SETBATTR | = | &H3E9 |
| 8149 | SETBORDER | = | &H3F0 |
| 8150 | SETBW | = | &H3F7 |

D.1 Source Code for UV-260 Data Acquisition (cont'd)

| | | | |
|------|---------------|---|-------|
| 8151 | SETCAMERA | = | &H3FE |
| 8152 | SETCLIP | = | &H13B |
| 8153 | SETCOLOR | = | &H0 |
| 8154 | SETCPAL | = | &H436 |
| 8155 | SETCRANGE | = | &H405 |
| 8156 | SETDEGREE | = | &H51D |
| 8157 | SETDEV | = | &H261 |
| 8158 | SETDRANGE | = | &H40C |
| 8159 | SETFONT | = | &H4C2 |
| 8160 | SETGLOCK | = | &H413 |
| 8161 | SETHATCHSTYLE | = | &HC4 |
| 8162 | SETIEEE | = | &H498 |
| 8163 | SETINTERLACE | = | &H41A |
| 8164 | SETIPAL | = | &H43D |
| 8165 | SETLATTR | = | &H245 |
| 8166 | SETLMODE | = | &H421 |
| 8167 | SETLNSTYLE | = | &H539 |
| 8168 | SETLNWIDTH | = | &H540 |
| 8169 | SETLOCATOR | = | &H23E |
| 8170 | SETMASK | = | &H428 |
| 8171 | SETMATTR | = | &H42F |
| 8172 | SETPATTR | = | &H1CE |
| 8173 | SETPRN | = | &H1C7 |
| 8174 | SETPWD | = | &H4FA |
| 8175 | SETRGB | = | &H444 |
| 8176 | SETSCAN | = | &H27D |
| 8177 | SETSCATTR | = | &H284 |
| 8178 | SETSCREEN | = | &H459 |
| 8179 | SETSEG | = | &H46 |
| 8180 | SETSEG2 | = | &H452 |
| 8181 | SETSTANG | = | &H4E5 |
| 8182 | SETSTCLR | = | &H4D8 |
| 8183 | SETSTEXT | = | &H4C9 |
| 8184 | SETTEXT | = | &HCB |
| 8185 | SETTEXTCLR | = | &HD2 |
| 8186 | SETVBW | = | &H460 |
| 8187 | SETVIEWPORT | = | &H134 |
| 8188 | SETWINDOW | = | &H149 |
| 8189 | SETWORLD | = | &H142 |
| 8190 | SETXOR | = | &H1F8 |
| 8191 | SETXPAL | = | &H44B |
| 8192 | SHIFT | = | &H467 |
| 8193 | STARTGRAPHICS | = | &H46E |
| 8194 | STEXT | = | &H4D7 |
| 8195 | TEXT | = | &HD9 |
| 8196 | VPAN | = | &H475 |
| 8197 | WORLDOFF | = | &H150 |
| 8198 | ZOOM | = | &H47C |
| 9999 | RETURN | | |

Appendix E

Aggregation Kinetics Model
Source Code Listing

E.1 Source code for aggregation kinetics model

```
C ***** KINETICS.FOR *****
C *
C *
C *           Todd M. Przybycien  12/22/88
C *
C *           This program fits rate constants to raw chymotrypsin aggregation
C * turbidity data.  Three rate constants are fit:
C *           k1 - conformational change
C *           k2 - aggregative dimerization
C *           k3 - aggregation
C *
C * The protein available for aggregation is the supersaturation.
C * The rate constants are fit to the raw data using a
C * nonlinear chisquare Marquardt algorithm (MRQMIN and MRQCOF) from
C * "Numerical Recipes." The kinetic rate expressions are integrated for
C * the fitting routine by a modified Gear integration routine (DDASSL) by
C * Linda Petzold of Sandia Labs. For the optimization procedure, the
C * objective function is weighted by the magnitude of the data. Fit
C * statistics are computed with an unbiased objective function, assuming
C * that measurement errors are normally distributed.
C *
C *****
C
C
C           Variables
C
C           REAL          MW,MWS,KD
C           DOUBLE PRECISION  Y,TOUT,T,ATOL,RTOL,RWORK,CHISQ,CHISQ2,CONVRG,
C           &              RPAR,YPRIME
C           DIMENSION      A(11),A488(4888),MW(4888),SIG(4888),
C           &              COVAR(3,3),ATMP(3),ALPHA(3,3)
C           CHARACTER      INFIL*12,OUTFIL*12
C           EXTERNAL      FUNCS
C
C           Constants
C
C           R          = 1.987
C           ACUT      = 2.8
C           ACONV     = 1.3445E8
C
C           A(1)     = conformational change rate constant, k1# (fitted)
C           A(2)     = aggregative dimerization rate constant, k2# (fitted)
C           A(3)     = aggregation rate constant, k3# (fitted)
C           A(4)     = forward dimerization rate constant, k+d
C           A(5)     = reverse dimerization rate constant, k-d
C           A(6)     = solubility in units of molarity, S/M1
C           A(7)     = initial supersaturation ratio, s#
C           A(8)     = initial concentration monomer, A#
C           A(9)     = initial concentration dimer, A2#
C           A(10)    = monomer molecular weight, M1
C           A(11)    = end of mixing transient, index of start time, TSTRT
C
C           A(4)     = 9.4588E3
C           A(5)     = 1.9888E8
C           A(10)    = 2.5888E4
C
C           Read SETUP.FIL to determine the number of data files to be fit
C
C           OPEN (1, FILE= 'SETUP.FIL', FORM= 'FORMATTED', STATUS= 'OLD')
C           READ (1,10) NUM
C           10 FORMAT (1X,I3)
C           DO 180 I = 1,NUM
C             READ (1,20) INFIL,OUTFIL,C#,S,DEGK,DELH,A(11)
C             20 FORMAT (1X,2A12,5E15.6)
C
C           Scale the protein solubility by the temperature of the
```

E.1 Source code for aggregation kinetics model (cont'd)

```

C      experiment, note that the dimerization rate constants
C      should be relatively constant over the temperature range
C      studied because of the small enthalpy of reaction
C
S      = EXP(ALOG(S) - DELH*(1.0/DEGK - 1.0/298.0)/R)
KD     = A(4)/A(5)
A(6)  = S/A(10)
C
C      Calculate initial monomer concentration and initial weight
C      average molecular weight. If Co is the protein concentration
C      in the drive syringe in mg/mL, note that
C
C      and
C      Co      = C1 + C2 = Mw1*(A1o + 2*A2o)
C      Kd      = A2o/(A1o)^2.
C
C      Use initial syringe monomer-dimer equilibrium distribution
C      and just rescale by the final or flow cell supersaturation
C      concentration
C
C      Super/Co = (1/2) - (S/Co).
C
C      Finally, the initial weight average molecular weight becomes
C
C      <Mw>wt = Mw1*(1 + 4*Kd*A1o)/(1 + 2*Kd*A1o).
C
C      To find A0, solve quadratic with numerically stable formula.
C
C      Q      = -0.5*(1.0 + SQRT(1.0 + 8*KD*C0/A(10)))
C      A(8)   = (-C0/A(10))/Q
C      A(9)   = KD*A(8)**2
C
C      Recall dilution by factor of two in stopped-flow experiment.
C
C      A(8)   = 0.5*A(8)
C      A(9)   = 0.5*A(9)
C      A(7)   = (A(8) + 2*A(9))/A(6)
C
C      Read raw data file
C
C      OPEN (2, FILE= INFIL, FORM= 'FORMATTED', STATUS= 'OLD')
C      READ (2,30) (DUMMY,TINT,J=1,2)
30     FORMAT (1X,F9.4,3X,F9.4)
C      REWIND (2)
C      READ (2,30,END=40) (A488(J),TSTOP,J=1,4000)
40     CLOSE (2)
C
C      Find the segment of data between the mixing transient and
C      the absorbance cutoff to be fit, and convert A488 to <Mw>.
C      Recalling that A488=ACONV = (C0/2)*<Mw>, compute the standard
C      deviation of the measured molecular weight data and weight by
C      the square root of the data magnitude
C
C      NPTS = JINT(TSTOP/TINT)
C      IF (A(11).EQ.0.0) A(11) = TINT
C      NSTRT = JINT(A(11)/TINT) + 1
C      NSTOP = NPTS
C      DO 50 J = NSTRT,NPTS
C          MW(J-NSTRT+1) = 2*ACONV*A488(J)/C0
C          T              = A(11) + FLOAT(J-NSTRT)*TINT
C          WRITE (8,53) T,MW(J-NSTRT+1)
C          WRITE (8,53) T,A488(J)
C          FORMAT (1X,2E15.6)
C          IF (A488(J).LT.0.5) THEN
53

```

E.1 Source code for aggregation kinetics model (cont'd)

```

      SIG(J-NSTRT+1) = (2*ACONV*0.002/C0
      *SQR(A488(J)/A488(NSTRT)))
      ELSE
      SIG(J-NSTRT+1) = (2*ACONV*0.004/C0
      *SQR(A488(J)/A488(NSTRT)))
      IF (A488(J).GE.ACUT) THEN
      NSTOP = J
      GOTO 60
      END IF
      END IF
      IF (SIG(J-NSTRT+1).EQ.0.0)
      SIG(J-NSTRT+1) = 2*ACONV*0.002/C0
50  CONTINUE
60  NDATA = NSTOP - NSTRT + 1
C
C      Open output file for fitting results and error messages
C
      OPEN (2, FILE= OUTFIL, FORM= 'FORMATTED', STATUS= 'UNKNOWN')
      WRITE (2,70) INFIL
70  FORMAT (1X,'Kinetic parameter fitting for data file ',A12)
C
C      Set up and execute Marquardt algorithm to fit rate constants
C
      ALAMDA = -1.0
      MFIT   = 3
      MA     = 11
C
C      Initial estimates of fitted parameters
C
      A(1) = 1.0E0
      A(2) = 1.0E4
      A(3) = 1.0E4
      ITER = 1
      CALL MRQMIN (TINT,MW,SIG,NDATA,A,MA,MFIT,COVAR,
      ALPHA,MFIT,CHISQ,FUNCS,ALAMDA,*110)
      WRITE (2,80) ITER,A(1),A(2),A(3),A(4),A(5),CHISQ
80  FORMAT (1X,'ITER= ',I3,'/6X,' k1= ',E13.6,' k2= ',E13.6,
      ' k3= ',E13.6,'/6X,' k+d= ',E13.6,' k-d= ',E13.6,
      ' Chi^2= ',E13.6)
      CHISQ2 = CHISQ
      ALAMD2 = ALAMDA
90  CALL MRQMIN (TINT,MW,SIG,NDATA,A,MA,MFIT,COVAR,
      ALPHA,MFIT,CHISQ,FUNCS,ALAMDA,*110)
      ITER = ITER + 1
      WRITE (2,80) ITER,A(1),A(2),A(3),A(4),A(5),CHISQ
      CONVRG = (CHISQ2 - CHISQ)/CHISQ2
      STEP   = ALAMD2 - ALAMDA
      IF ((CONVRG.LT.0.001).AND.(STEP.GE.0.0)) THEN
      GOTO 110
      ELSE IF (ITER.GT.50) THEN
100  WRITE (2,100) CONVRG
      FORMAT (6X,'50 iterations taken, CONVRG = ',
      E15.6)
      GOTO 110
      ELSE
      CHISQ2 = CHISQ
      ALAMD2 = ALAMDA
      GOTO 90
      END IF
C
C      After fitting parameters, rerun fitting routine with
C      straight statistics
C
110  DO 105 J = NSTRT,NSTOP+1

```

E.1 Source code for aggregation kinetics model (cont'd)

```

      IF (A488(J).LT.0.5) THEN
        SIG(J-NSTRT+1) = 2*ACONV*0.002/C0
      ELSE
        SIG(J-NSTRT+1) = 2*ACONV*0.004/C0
      END IF
105  CONTINUE
      ALAMDA = -1.0
      CALL MRQMIN (TINT,MW,SIG,NDATA,A,MA,MFIT,COVAR,
      &           ALPHA,MFIT,CHISQ,FUNCS,ALAMDA,*125)
      CALL MRQMIN (TINT,MW,SIG,NDATA,A,MA,MFIT,COVAR,
      &           ALPHA,MFIT,CHISQ,FUNCS,ALAMDA,*125)
      ALAMDA = 0.0
      CALL MRQMIN (TINT,MW,SIG,NDATA,A,MA,MFIT,COVAR,
      &           ALPHA,MFIT,CHISQ,FUNCS,ALAMDA,*125)
125  CHISQN = CHISQ/FLOAT(NDATA)
C
C      Compute standard deviations and correlation coefficients
C
      DO 135 J = 1,MFIT
        IF (COVAR(J,J).GT.0.0) THEN
          COVAR(J,J) = SQRT(COVAR(J,J))
        ELSE
          COVAR(J,J) = 0.0
        ENDIF
135  CONTINUE
      DO 150 J = 1,MFIT
        DO 140 K = J+1,MFIT
          IF (COVAR(J,J)*COVAR(K,K).NE.0.0) THEN
            COVAR(J,K) = COVAR(J,K)
            &           / (COVAR(J,J)*COVAR(K,K))
            COVAR(K,J) = COVAR(J,K)
          ELSE
            COVAR(J,K) = 0.0
            COVAR(K,J) = 0.0
          ENDIF
140  CONTINUE
150  CONTINUE
      WRITE (2,160) A(1),A(2),A(3),A(4),A(5),CHISQ,CHISQN,
      &           CONVRG
160  FORMAT (/0X,'k1 (1)           = ',E15.6,' s-1',
      &        /0X,'k2 (2)           = ',E15.6,' s-1 M-1',
      &        /0X,'k3 (3)           = ',E15.6,' s-1 M-1',
      &        /0X,'k+d (4)          = ',E15.6,' s-1 M-1',
      &        /0X,'k-d (5)          = ',E15.6,' s-1',
      &        /0X,'chi square       = ',E15.6,
      &        /0X,'chi square/N     = ',E15.6,
      &        /0X,'cnvrg criterion = ',E15.6,
      &        //0X,'standard deviation and correlation ',
      &        'coefficient matrix')
      WRITE (2,170) ((COVAR(J,K),K=1,MFIT),J=1,MFIT)
170  FORMAT (0X,E10.4,2X,E10.4,2X,E10.4)
      CLOSE (2)
180  CONTINUE
      CLOSE (1)
      STOP
      END
      SUBROUTINE          FUNCS (TINT,A,FMW,FMWDA,MA,NDATA,e)
C
C      Subroutine FUNCS - computes values of the MW and the dMW/dA
C      at each T, called by the Marquardt
C      routine MRQCOF - note conversion to
C      double precision for DDASSL
C
      DOUBLE PRECISION  ATOL,RWORK,RTOL,T,TOUT,Y,YPRIME,RPAR,TOL

```


E.1 Source code for aggregation kinetics model (cont'd)

```
PD(2,3) = 4*RPAR(3)*Y(2) - 2*RPAR(2)*(B + Y(5))
PD(2,5) = 0.5*PD(2,3) + RPAR(2)*B
PD(2,6) = PD(2,3)
C
PD(3,1) = PD(2,1)
PD(3,5) = 2*RPAR(3)*Y(3) - RPAR(2)*(Y(6) + Y(5))
PD(3,3) = 2*PD(3,5) - (2*B*(RPAR(2) + RPAR(3)) + CJ)
PD(3,6) = 2*PD(3,5) - RPAR(2)*B
C
PD(4,1) = PD(2,1)
PD(4,5) = 2*RPAR(3)*Y(4) - RPAR(2)*(Y(7) + 2*Y(6) + Y(5))
PD(4,7) = RPAR(2)*B
PD(4,3) = 2*PD(4,5) + 8*RPAR(3)*Y(3) - 2*PD(4,7)
PD(4,4) = -(2*RPAR(3)*B + CJ)
PD(4,6) = 2*PD(4,5)
C
PD(5,2) = 2*RPAR(3)*B
PD(5,3) = -4*RPAR(3)*Y(2) + 2*RPAR(2)*Y(5)
PD(5,5) = 0.5*PD(5,3) - (PD(4,7) + CJ)
PD(5,6) = PD(5,3)
C
PD(6,5) = RPAR(2)*Y(6) - 2*RPAR(3)*Y(3)
PD(6,3) = 2*PD(6,5) + PD(5,2)
PD(6,6) = 2*PD(6,5) - (PD(4,7) + CJ)
C
PD(7,3) = 2*RPAR(2)*Y(7) - 4*RPAR(3)*Y(4)
PD(7,4) = PD(5,2)
PD(7,5) = 0.5*PD(7,3)
PD(7,6) = PD(7,3)
PD(7,7) = -(PD(4,7) + CJ)
RETURN
END
```

Appendix F

Immunoaffinity Chromatography in situ NMR Project
Feasibility Study

F.1 Introduction

Antibody leakage and capacity losses due to immunosorbent denaturation are problems that plague immunoaffinity chromatography, contributing heavily to the prohibitive expense of this powerful separation technique. Previous work has indicated that in some cases only 0.1% of the antibody retains its antigen binding activity after immobilization on a support matrix [1]. Further activity losses occur on repeated bind-release cycles. There appears to be a "first cycle" effect in which initial capacity falls substantially and smaller capacity losses occur with each subsequent cycle; initial losses of the order of 20% with subsequent losses of 1-2% per cycle have been reported for a variety of immunosorbent systems [2]. As the immunosorbent ages, the optimal operating conditions may evolve [2,3].

To address these issues, a coupling chemistry is proposed that incorporates nuclear magnetic resonance (NMR) -active nuclei. This chemistry affords the opportunity to monitor matrix activation, antibody (ligand) immobilization and leakage *in situ* via NMR. The coupling chemistry is based on the activation of agarose with 1,1'-carbonyldiimidazole [4]; Sepharose CL-4B would be activated with methylphosphonic-*bis*-4-trifluoromethylimidazole. The ^{19}F -labelled moieties serve as a probe of the density of active coupling groups on the matrix. The ^{31}P spectrum of the activated matrix is not sufficient for this purpose as a significant portion of the coupler may form two covalent bonds with the matrix. The ^{31}P chemical shift should reflect the binding of antibody to the activated matrix and may serve as a probe of the ligand density; the chemical shifts of ^{31}P compounds cover a wide range [5] and are environmentally sensitive [6].

A three-phase experimental plan is proposed. The first involves determining the chemical conditions promoting the highest density of active coupling

groups on the matrix. Next, the optimum ligand density, with respect to antigen binding capacity, would be explored. Finally, using the optimized coupling group and ligand densities, antibody leakage may be followed through repetitive bind-release cycles with a variety of eluents. At this stage NMR imaging experiments may be possible, providing an indication of the spatial variation of immunosorbent integrity.

Feasibility studies were undertaken to assess the potential of the proposed coupling chemistry and the NMR sensitivity to labelled, activated Sepharose matrices. These results are presented along with a proposed course of action.

F.2 Coupling Chemistry

Methylphosphoric-*bis*-imidazolide, an analog of the desired linker, was synthesized to demonstrate the effectiveness of the proposed synthetic route. In this synthesis, two equivalents of imidazole were reacted with one equivalent of methylphosphoric difluoride. Guidelines for using methylphosphoric difluoride were based on previous work with α -chymotrypsin spin-labelled inhibitors [7]. Figure 1 gives a description of the reactions performed.

Imidazole (Im) was obtained from Sigma and methylphosphoric difluoride ($\text{OP}(\text{CH}_3)\text{F}_2$) was a product of Alfa. The moisture sensitivity of the $\text{OP}(\text{CH}_3)\text{F}_2$ necessitates anhydrous reaction conditions. A 100 mL three-necked round bottom flask was equipped with a N_2 purge and a CaCl_2 drying tube. A 2.72 g (0.04 mole) portion of Im was pulverized with a mortar and pestle and added to the reaction flask. A 25 mL volume of triethylamine (Et_3N) was then added by passing through a bed of activated 3 Å molecular sieves. The sieve bed was removed and the contents of a 2 g (0.024 mole) ampoule of $\text{OP}(\text{CH}_3)\text{F}_2$ was added to the reaction flask. The stoppered flask was magnetically stirred for 3

h. Residual $\text{OP}(\text{CH}_3)\text{F}_2$ on glassware was killed by rinsing with 5 M sodium acetate buffer, pH 4.5. Due to the sparing solubility of Im in triethylamine, about 30 mL of CH_3CN was added to the flask 90 minutes into the reaction; the remaining Im dissolved immediately.

After 3 h, the reaction was quenched by the addition of 25 mL of 5 M sodium acetate buffer, pH 4.5. A small amount of a white film, presumably $\text{OP}(\text{CH}_3)(\text{OH})_2$, was produced. The resulting solution was reduced to a syrup on a rotary evaporator. The syrup was extracted with 20 mL CH_2Cl_2 ; the addition of a few drops of a saturated aqueous solution of NaCl aided the phase separation. The organic phase was collected and the CH_2Cl_2 stripped off on a rotary evaporator. 0.64 g of a yellow liquid was obtained.

A portion of the CH_2Cl_2 extract was withdrawn and a UV spectrum was taken. The extract had a large absorbance at 285 nm with a shoulder at about 300nm. This is indicative of the Im moiety. Thin layer chromatography (TLC) of the extract on plastic-backed fluorescent silica gel plates gave two spots, $R_f = 0.384$ (moderate) and $R_f = 0.566$ (faint), when developed with ethanol. A small portion of the extract was dissolved in d_6 -acetone and a ^{31}P NMR spectrum was taken. Two sharp resonances were obtained at 15.1 and 41.4 ppm (85% $\text{H}_3\text{PO}_4 = 0$ ppm). This spectrum is given in Figure 2 after 600 scans. Residual triethylamine contaminated the extract rendering the ^1H NMR spectrum inconclusive.

The analytical data are consistent with the presence of the desired product and a side product. From the chemical shifts of analogous compounds the ^{31}P resonance at 15.1 ppm may be tentatively assigned to $\text{OP}(\text{CH}_3)\text{Im}_2$ [6]. The TLC and ^{31}P NMR data indicate that two phosphorous containing compounds were obtained and the UV spectrum demonstrates the presence of the Im group.

This supports the proposed synthetic route.

F.3 NMR Spectroscopy

It was not clear *a priori* that NMR-active nuclei bound to a Sepharose matrix would give reasonable spectra. A ^{31}P -labelled derivative of Sepharose CL-4B was prepared by phosphorylating the hydroxyl residues of the cross-linked agarose with methylphosphoric dichloride; the derivative was prepared in analogy with the method of Bethell and coworkers [4]. The reaction scheme is given in Figure 3. ^{31}P NMR spectra were run in configurations mimicking those proposed for the *in situ* chromatography work.

Sepharose CL-4B, preswollen in 20% ethanol, was purchased from Pharmacia. Methylphosphoric dichloride ($\text{OP}(\text{CH}_3)\text{Cl}_2$) was obtained from Aldrich. About 3 mL of the agarose gel was placed on a Büchner funnel with a sintered glass frit (10-15 μm pores). The gel was washed with approximately 20 mL of each of the following solutions in this order: H_2O ; 7:3 $\text{H}_2\text{O}:\text{CH}_3\text{CN}$; 3:7 $\text{H}_2\text{O}:\text{CH}_3\text{CN}$; 3:7 $\text{Et}_3\text{N}:\text{CH}_3\text{CN}$; 7:3 $\text{Et}_3\text{N}:\text{CH}_3\text{CN}$; Et_3N . All ratios are v/v. The gel was then flushed with another 100 mL of Et_3N and transferred into a glass scintillation vial. Note that at no time was the gel sucked dry.

The vial containing the gel was placed into a glove bag along with the container of $\text{OP}(\text{CH}_3)\text{Cl}_2$ and the bag was flushed with N_2 . Approximately 0.1 g of $\text{OP}(\text{CH}_3)\text{Cl}_2$ was added to the vial; the gel went from translucent to a whitish color at this stage. The sealed vial was removed from the glove bag and placed on an end-over-end rotator for 1 h. The contents of the vial were then emptied into the Büchner funnel and the gel was washed with the same solvents as above in reverse order. At this point the gel had a yellowish tint. The gel was washed a final time with an additional 80 mL of water to ensure that any

remaining chloro moieties were substituted with hydroxyls.

A ^1H decoupled ^{31}P NMR spectrum of the derivatized gel was taken on a 90 MHz spectrometer. The spectrometer was shimmed on a tube containing an equivalent volume of 5% KH_2PO_4 in D_2O . Three resonances at 17.9, 26.1 and 26.5 ppm were resolved. This spectrum is shown in Figure 4. From Figure 3, three phosphorous-containing species are expected: residual, mono-linked and dual-linked methyl phosphate. Based on peak size, resolution and chemical shift data from analogous compounds [6], tentative peak assignments were made. The resonance at 17.9 ppm was assigned to the doubly-bonded species; this species should be present at lower levels and the peak should be broadened due to the hindered mobility. The residual and mono-linked methyl phosphates were assigned to the two downfield peaks. The labelling reaction was deemed successful.

To investigate the quality of the spectral information expected from *in situ* NMR experiments, a spectrum of the derivatized gel was taken in a 10 mm sample tube without spinning in a 300 MHz wide bore instrument. The spectrometer was shimmed on an equivalent volume of Sepharose CL-4B flushed with D_2O . The spectrum is presented in Figure 5. The resolution is somewhat lower; the two peaks at about 26 ppm merged into one resonance and the resonance at 17.9 ppm was broadened considerably. However, gel samples should be chemically stable at prescribed conditions and longer acquisitions may result in spectra of sufficient signal-to-noise ratios to facilitate integrations. It should be possible to obtain significant results with *in situ* experiments.

F.4 Proposed Synthetic Route

The Sepharose-antibody conjugate with the labelled coupling chemistry can

be synthesized in four steps. The first consists of the synthesis of the trifluoromethyl substituted imidazole. Next, the imidazole is reacted with methyl phosphonic dichloride; the methyl group gives a linker that is uncharged. Third, the matrix is activated. Finally, antibody is immobilized on the activated matrix. The overall scheme is given in Figure 6.

The 4-substituted imidazole is produced by a variation of the "formamide synthesis" [8]. 1,1,1-trifluoro-3,3-dichloroacetone is converted to a glyoxal by treatment with aqueous sodium acetate. The glyoxal is reacted *in situ* with formaldehyde and ammonia giving 4-trifluoromethylimidazole ($\text{CF}_3\text{-Im}$). Yields of about 50 % have been obtained [9].

The methyl phosphonic-*bis*-4-trifluoro imidazolide ($\text{OP}(\text{CH}_3)(\text{CF}_3\text{-Im})_2$) may be prepared by analogy with the spin label synthesis given by Morrisett et al. [7]. It is important to maintain rigorously dry conditions for this reaction so the $\text{OP}(\text{CH}_3)\text{Cl}_2$ does not decompose. Use of a glove bag is recommended. Et_3N should be dried in a solvent still before use. The work of Miller et al. [10,11] with $\text{OP}(\text{CH}_3)\text{Im}_2$ as an intermediate in the synthesis of oligodeoxyribonucleoside methylphosphonates may also provide useful synthetic background.

The activation procedure should follow that of Bethell and coworkers [4]. Dioxane is the solvent; again, care must be taken to ensure that the dioxane is dry to prevent decomposition of the $\text{OP}(\text{CH}_3)(\text{CF}_3\text{-Im})_2$ and low activation densities.

The immobilization reaction of Bethell et al. [4] must be modified. The appropriate pH must be selected. High pHs should be avoided as the CF_3 moiety may be oxidized to the carboxylic acid [9,12] and the antibodies may denature; however, since the pK_a of $\text{CF}_3\text{-Im}$ is much lower than that of Im [12], alkaline conditions should not be necessary for the reaction to occur. The

coupler reacts with the primary amine groups of the protein. The amount of antibody immobilized will have to be optimized.

F.5 Experimental Apparatus

The novelty of this approach is that antibody leakage may be monitored in an operating immunoaffinity chromatography column. In order for a column to be placed within the superconducting magnet of the wide bore Brüker 300 MHz NMR spectrometer, a coaxial tube system was designed. This system is shown in Figure 7. The bottom of a 10 mm I.D. NMR tube is removed and a fritted glass disk (14-40 μm nominal pore size) is blown on in its place. The 10 mm I.D. is a standard diameter for chromatography columns; a Bio-Rad Econocolumn flow adaptor may be fitted to the NMR tube so that solutions may be applied directly to the gel bed. Note that the stainless steel inner guide rod should be replaced with an aluminum or plastic rod. The bed height should be at least 4.6 cm so that the receiver coils are covered. This column is placed inside a 20 mm O.D. NMR sample tube and held in place with coaxial tube inserts. Solvent delivery and removal is performed by a two-head peristaltic pump.

F.6 Proposed Experiments

The coupling chemistry described here is not suggested to be an improvement in stability over the currently available linking chemistries. The objective is to arrive at a reasonable immobilization procedure incorporating NMR-active nuclei. Yet, it will still be desirable to maximize the coupling efficiency. By manipulating the conditions of the gel activation procedure, it should be possible to attain linking group densities rivaling those obtained with the 1-1'-carboxydiimidazole technique [4]. This can be determined via NMR.

Once the activation conditions have been determined, the effects of different

ligand densities on column capacity should be explored. It has been shown that capacity is a non-monotonic function of the ligand density [13]; this may be due to a variety of steric interactions [14]. Again, NMR may be used to monitor the ligand density. It may prove possible to distinguish between different modes of ligand attachment (through Lys, Arg, or N-termini NH_3^+ groups) via chemical shifts. This possibility might be explored by taking spectra of immobilized poly-L-lysine and poly-L-arginine; if there is any separation at all in the chemical shifts, deconvolution of the peaks from immobilized antibody samples should be possible.

Having arrived at an efficient coupling procedure, the next stage is the determination of optimal loading, washing and elution conditions and the investigation of the kinetics and mechanism of immunosorbent activity losses associated with operation. The optimum loading conditions may be determined from adsorption breakthrough curves [15]. Spreading in the breakthrough curves gives insights into the flow nonidealities and the mass transfer and sorption rates of the system [16,17]. Many different washing and elution strategies exist in which pH changes, protein denaturants, chaotropic agents, polarity reducing solvents, and temperature effects are employed [3]. Elution requires the disruption of the interaction between the antigen and antibody; as such, the antibody may be subjected to potentially denaturing environments [2,18]. Experiments consisting of treating an eluent equilibrated column with a finite pulse of antigen-containing eluent, may be used to screen potential eluents and to determine optimum operating conditions [19].

Column capacity losses during operation may be ascribed to immunosorbent denaturation and leakage. The *in situ* NMR technique offers the ability to assign the portion of the capacity loss due to leakage in an on-line fashion.

NMR determinations may be verified by enzyme-linked immunoassay (ELISA) techniques [20]. The NMR technique may allow the operating conditions to be updated for each subsequent cycle. If different modes of ligand attachment are distinguishable, perhaps individual rates of bond destruction can be determined. If the NMR signal is of sufficient strength, three dimensional imaging experiments may be the logical extension of this work. Profiles of the ligand density across and down the column may provide insight into the uniformity of sorbent aging and identify flow nonidealities.

This information may lead to the development of robust models able to describe the performance of an immunoaffinity system throughout its lifetime.

F.7 References

- 1 Kasche, V. and Galunsky, B., "Ligand/Ligate Interactions in Heterogeneous Systems: Their Influence on Operational Properties of Affinity Chromatographic and Binding Assay Adsorbents," *Anal. Chem. Symp. Ser.* **9**, 93-110 (1982)
- 2 Eveleigh, J.W. and Levy, D.E., "Immunochemical Characteristics and Preparative Application of Agarose-Based Immunosorbents," *J. Solid-Phase Biochem.* **2**, 45-78 (1977)
- 3 Chase, H.A., "Affinity Separations Utilising Immobilized Monoclonal Antibodies - A New Tool for the Biochemical Engineer," *Chem. Eng. Sci.* **39** 1099-1125 (1984)
- 4 Bethell, G.S., Ayers, J.S., Hancock, W.S., and Hearn, M.T.W., "A Novel Method of Activation of Cross-Linked Agaroses with 1,1'-Carbonyldiimidazole Which Gives a Matrix for Affinity Chromatography Devoid of Charged Groups," *J. Biol. Chem.* **254**, 2572-2574 (1978)
- 5 Harris, R.K., *Nuclear Magnetic Resonance Spectroscopy, A Physicochemical View*, Pitman Publishing, Marshfield, Massachusetts, 188 (1983)
- 6 Mark, V., Dungan, C.H., Crutchfield, M.M., and VanWazer, J.R., "Compilation of P³¹ NMR Data," in *Topics in Phosphorous Chemistry* **5**, 227-457 (1967)
- 7 Morrisett, J.D., Broomfield, C.A., and Hackley, B.E., "A New Spin label Specific for the Active Site of Serine Enzymes," *J. Biol. Chem.* **244**, 5758-5761 (1969)
- 8 Grimmett, M.R., "Advances in Imidazole Chemistry," *Advan Heterocyclic Chem.* **12**, 103-183 (1970)
- 9 Baldwin, J.J., Kasinger, P.A., Novella, F.C., and Sprague, J.M., "4-Trifluoro-

- methylimidazoles and 5-(4-Pyridyl)-1,2,4-triazole, New classes of Xanthine Oxidase Inhibitors," *J. Med. Chem.* **18**, 895-900 (1975)
- 10 Miller, P.S., Agris, C.H., Aurelian, L., Blake, K.R., Murakami, A., Reddy, M.P., Spitz, S.A., and Ts'o, P.O.P., "Control of Ribonucleic Acid Function by Oligonucleoside Methylphosphonates," *Biochimie* **67**, 769-776 (1985)
- 11 Miller, P.S., Reddy, M.P., Murakami, A., Blake, K.R., Lin, S.B., and Agris, C.H., "Solid-Phase Synthesis of Oligodeoxyribonucleoside Methylphosphonates," *Biochemistry* **25**, 5092-5097 (1986).
- 12 Takeuchi, Y., Kirk, K.L., and Cohen, L.A., "Adjacent Lone Pair (ALP) Effects in Heteroaromatic Systems. 2. Isotopic Exchange of Ring Hydrogens in Niro- and Fluoroimidazoles," *J. Org. Chem.* **43**, 3570-3578 (1978)
- 13 Fowell, S.L. and Chase, H.A., "Variation of Immunosorbent Performance with the Amount of Immobilized Antibody," *J. Biotechnol.* **4**, 1-13 (1986)
- 14 Chase, H.A., "Scale-up of Immunoaffinity Separation Processes," *J. Biotechnol.* **1**, 67-80 (1984)
- 15 Arnold, F.H., Blanch, H.W., and Wilke, C.R., "Analysis of Affinity Separations I: Predicting the Performance of Affinity Adsorbers," *Chem. Eng. J.* **30**, B9-B23 (1985)
- 16 Katoh, S., Kambayashi, T., Deguchi, R., and Yoshida, F., "Performance of Affinity Chromatography Columns," *Biotechnol. Bioeng.* **20**, 267-280 (1978)
- 17 Hall, K.R., Eagleton, L.C., Acrivos, A., and Vermeulen, T., "Pore- and Solid-Diffusion Kinetics in Fixed-Bed Adsorption under Constant-Pattern Conditions," *Ind. Eng. Chem. Fund.* **5**, 212-223 (1966)
- 18 Cuatrecasa, P., "Affinity Chromatography of Macromolecules," *Adv. En-*

zymol. **36**, 29-89 (1972)

- 19 Arnold, F.H., Blach, H.W., and Wilke, C.R., "Analysis of Affinity Separations II: The Characterization of Affinity Columns by Pulse Techniques," *Chem. Eng. J.* **30**, B25-B36 (1985)
- 20 Peng, L., Calton, G.J., and Burnett, J.W., "Stability of Antibody Attachment in Immunosorbent Chromatography," *Enz. Microb. Technol.* **8**, 681-685 (1986)

F.8 Figures

Figure 1 Synthetic scheme for methylphosphonic-*bis*-imidazolid

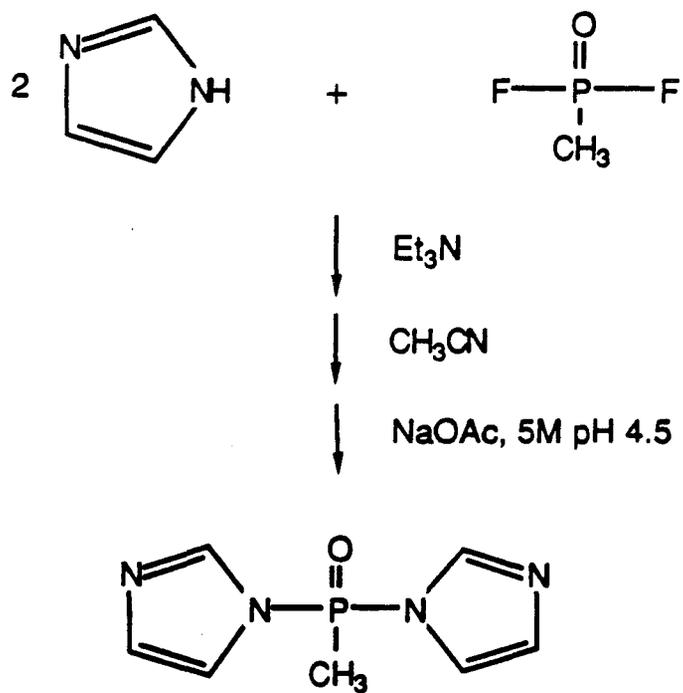


Figure 2 ^{31}P NMR spectrum of methylphosphonic-*bis*-imidazolidine preparation taken in 90 MHz instrument

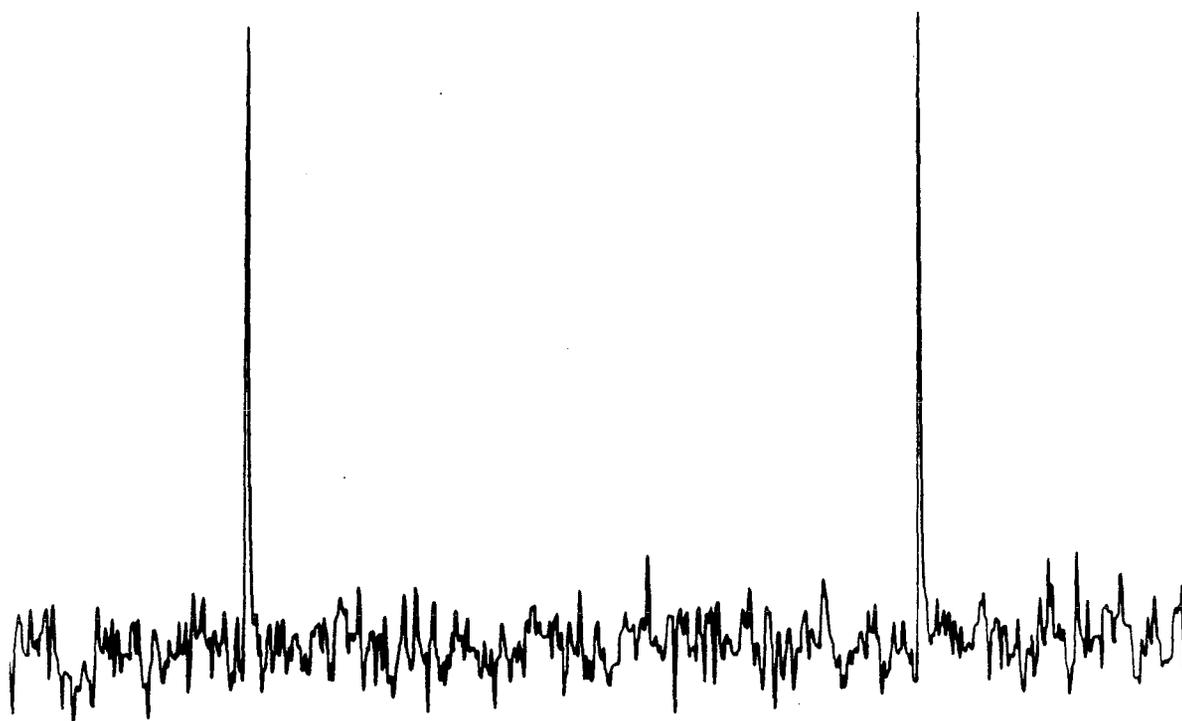


Figure 3 Synthetic scheme for trial matrix activation

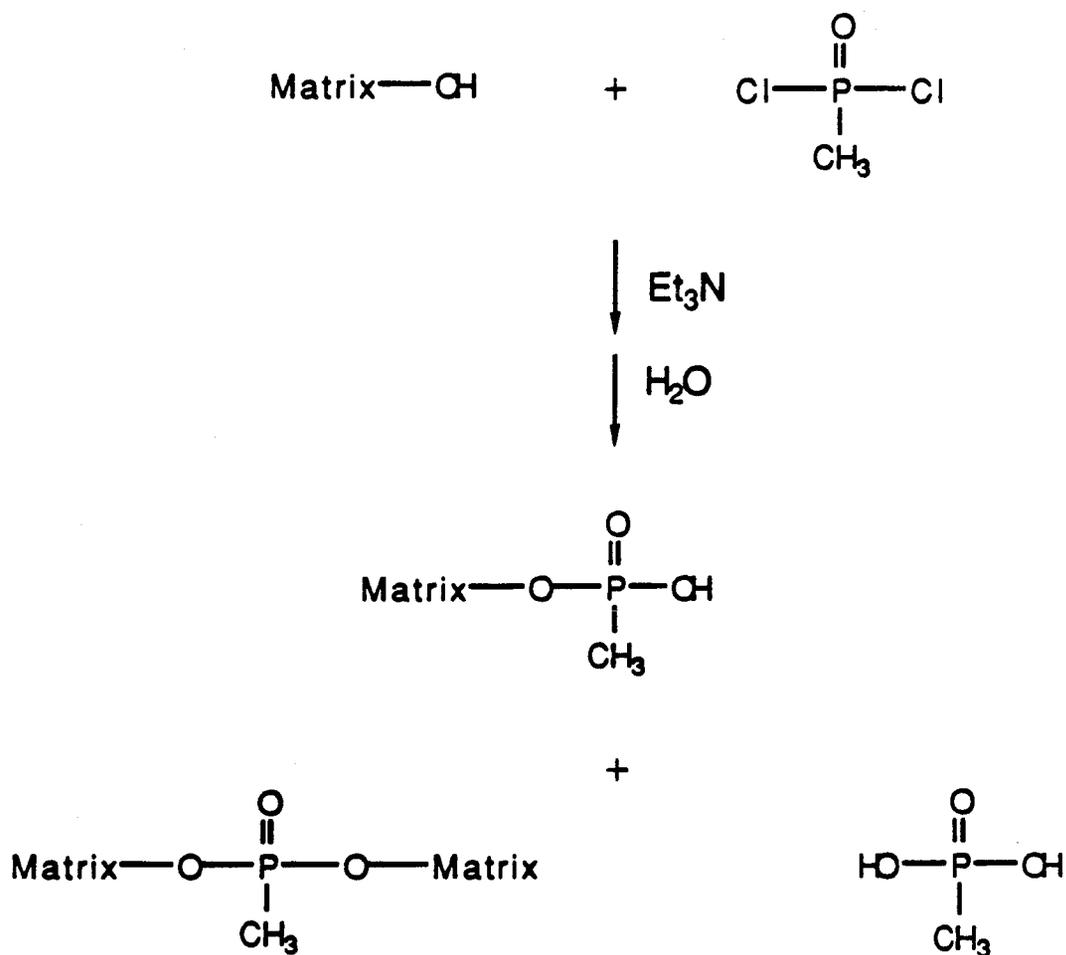


Figure 4 ^{31}P NMR spectrum of derivatized Sepharose taken in 90 MHz instrument

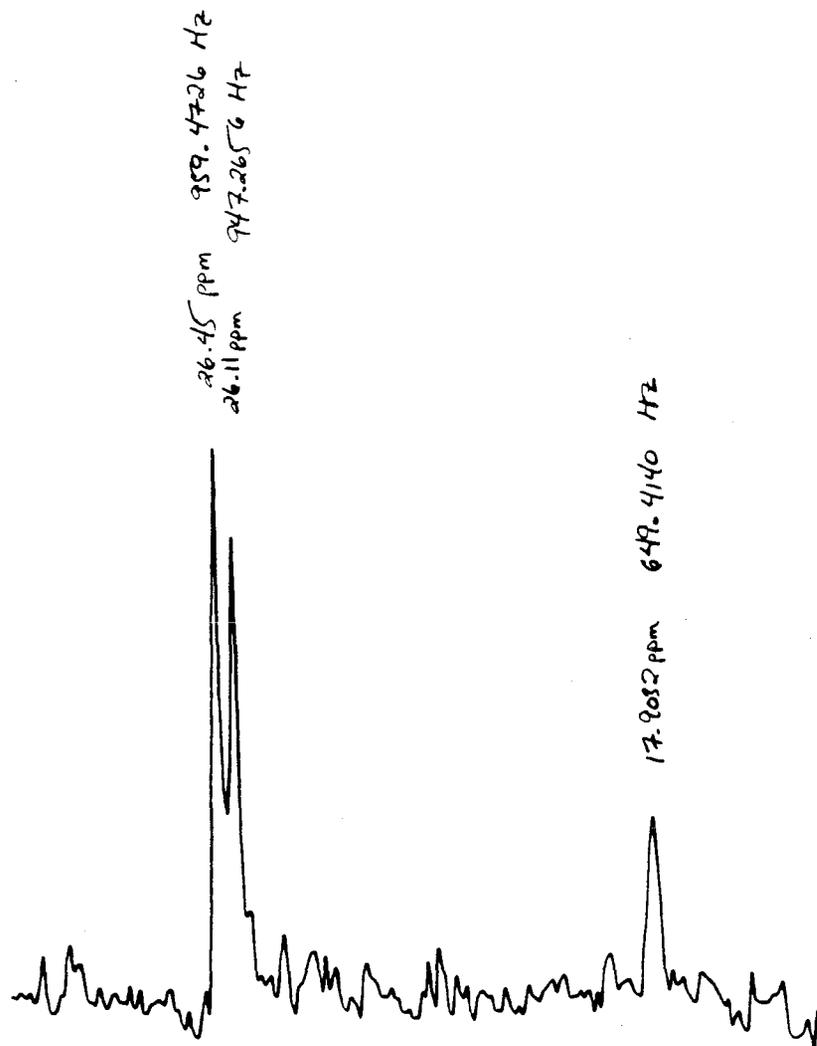


Figure 5 ^{31}P NMR spectrum of derivatized Sepharose taken in wide-bore 300 MHz instrument

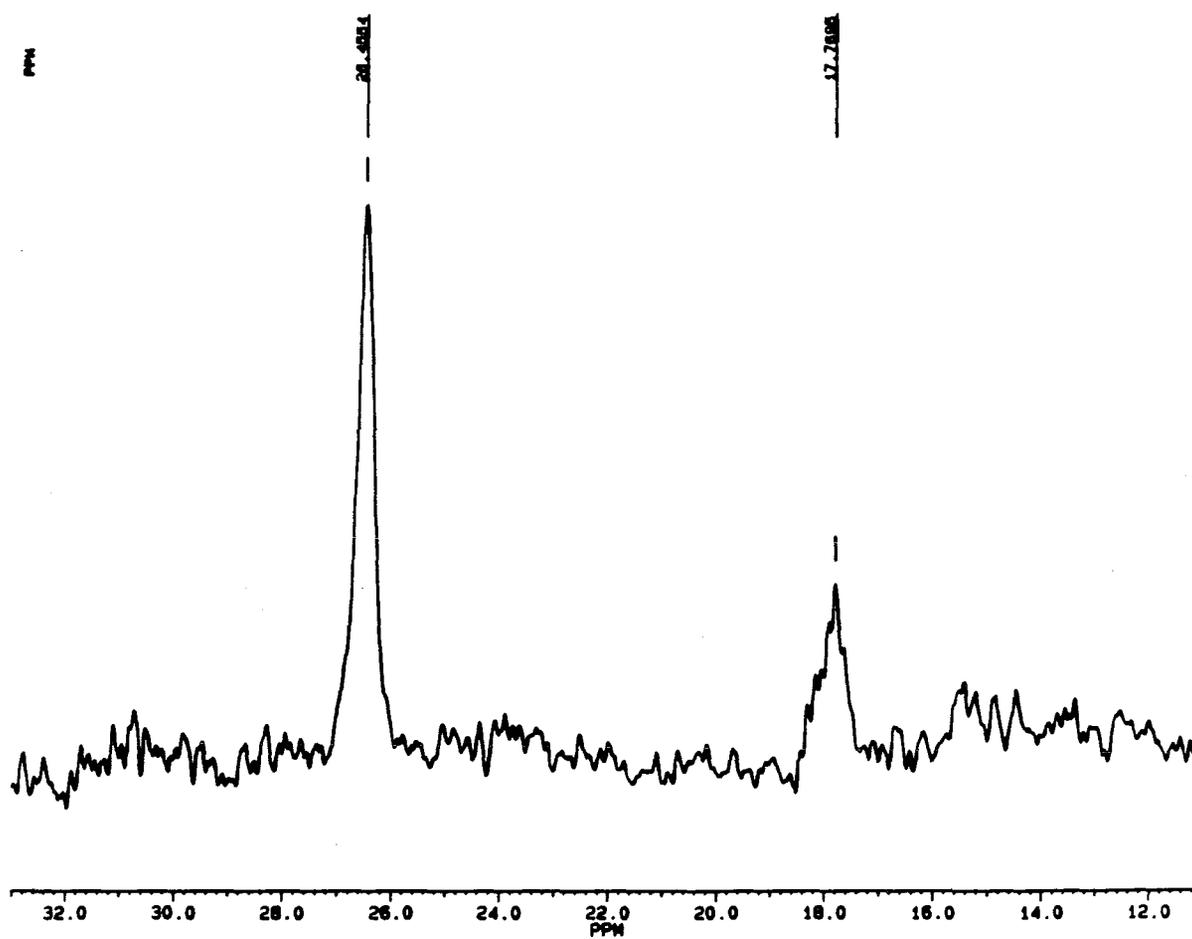
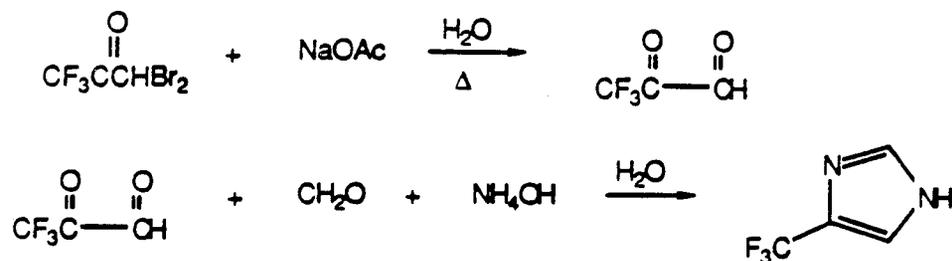
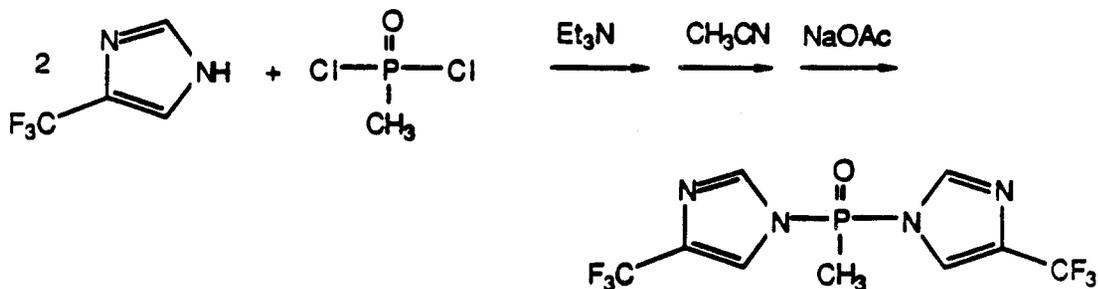


Figure 6 Synthetic scheme for immunosorbent preparation

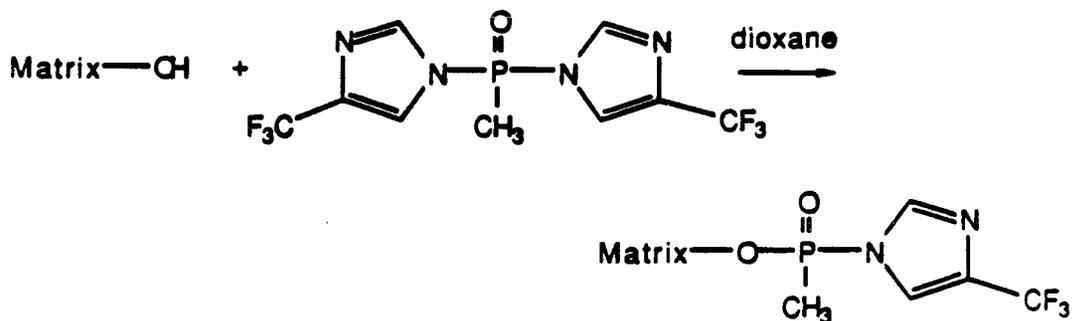
1. Substituted imidazole synthesis



2. Coupler synthesis



3. Matrix activation



4. Ligand immobilization

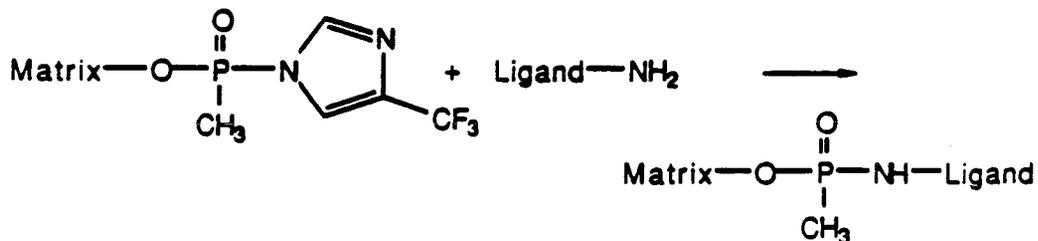


Figure 7 *in situ* NMR chromatography column design

