ANALYSIS OF MOBILE RESIDUES OF *E. COLI* CITRATE SYNTHASE USING NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

by

Kajal Choudhary

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Ph. D.

Department of Chemistry

University of Manitoba

Winnipeg

ABSTRACT:

E. coli Citrate Synthase (CS) is a large hexameric protein with a molecular weight of 280kDa and belonging to the type II class of citrate synthases. The crystal structure of *E. coli* CS in the T (inactive) state is known. The structure shows that the backbone atoms of 42 residues (1-5, 266-297 and 330-335) have temperature factors about 9 times greater than the average as compared to the rest of the molecule. Residues 266-297, also referred to as the "Mobile loop", are particularly of interest since they form a part of the active site and any rearrangement in the mobile loop can provide useful information about the R(active) state of the protein. In this study, Nuclear Magnetic Resonance (NMR) Spectroscopy has been used to study the flexible regions of *E. coli* CS in solution. The flexible residues have been assigned based on the amino acid type by ^{15}N-specific amino acid labeling, while the residue type has been assigned by site-directed mutagenesis. Changes in the dynamics of the flexible residues, in response to the substrate binding, have been studied using both NMR and Fluorescence Spectroscopy. Also a method to use Mass Spectrometry for accessing the isotopic incorporation in the samples prepared for NMR spectroscopy has been described. The initial hypothesis in this study was that only the mobile loop residues which show significant high B-factors will contribute to the NMR spectrum. However in the NMR spectrum, in addition to the mobile loop, some uncharacterized flexible regions were also observed. We also found that some of the residues show signs of slow conformational exchange resulting in multiple signals in the NMR spectrum. In addition we see that the environment of some flexible residues is changed in the presence of substrates, a few residues were immobilized, but most remained mobile.

ACKNOWLEDGEMENTS:

I am thankful to Dr. Harry Duckworth and Dr. Joe O′Neil for their encouragement and advice throughout the graduate program. I would also like to thank my committee members Dr. Frank Schweizer and Dr. Barbara Triggs-Raine for reading the thesis and providing useful suggestions. Thanks to Dr. Lynda Donald for doing all the Mass Spectrometry work, for her advice and friendship. Thanks to Vic Spencer for developing the simulations for the Mass Spectrometry results, Dr. Kirk Marat for his help with the Spectrometer and NMR software. Finally, I am grateful to my parents and brother for their love and support, my husband Nitin for his patience, encouragement and help with the thesis.

TABLE OF CONTENTS:

1.0 INTRODUCTION 1

1.1 INTRODUCTION TO CITRATE SYNTHASE 2

1.2 Classes of Citrate Synthase 5

1.3 Three dimensional structures of Citrate Synthases 5

1.4 Three-dimensional structure of type II CS 6
Secondary structural elements and subunit-subunit contacts 9
Dimer-Dimer Interactions 13
N-terminal Domain 13
Active Site 13

1.5 Activity of Citrate Synthase 16

1.6 Catalytic Mechanism 20

1.7 *E. coli* Citrate Synthase 22
Enzymology 22
NADH binding and inhibition 23

1.8 Mobile Regions in *E. coli* CS as determined from X-ray crystallography 27

1.9 Thesis Objectives 30

2.0 EXPERIMENTAL PROCEDURES 31

2.1 Site-directed Mutagenesis 31

2.2 Sub-cloning into pET24b^{+} vector 33

2.3 Expression strains for *E. coli* CS 35

2.4 Growth Medium 35
Uniform ^{15}N labeling 36
Uniform ^{15}N, ^{2}H labeling 37
Specific amino acid labeled *E. coli* CS 38

2.5 Purification of *E. coli* CS 38

2.6 CS activity assays and kinetic measurements 41

2.7 Spin labeling 42

3.0 NMR STUDIES OF FLEXIBLE REGIONS OF *E. COLI* CITRATE SYNTHASE 43

3.1 Introdution to NMR Spectroscopy of Biomolecules 43

3.2 Chemical Shift assignment 44

3.3 Problems with Large systems 44

3.4 Different approaches to studying large systems 46
Specific amino acid labeling 46
Spectroscopic Developments 49
Advantages of Deuteration 53

3.5 NMR Spectroscopy of large molecules 54

3.6 Protein Dynamics from NMR 58

3.7 Experimental Procedures 60
NMR Samples 60
NMR Spectroscopy 61
Determination of thermodynamic parameters 63

3.8 Results 64
3.81 Uniform ^{15}N-labeled *E. coli* CS 64

3.82 Uniformly ^{15}N, ^{2}H labeled *E. coli* CS 73

3.83 Specific ^{15}N, ^{13}C labeled *E. coli* CS 88
^{15}N-Phe *E. coli* CS 89
^{15}N-Met *E. coli* CS 92
^{15}N Glycine *E. coli* CS 109
^{15}N$_2$-Tryptophan *E. coli* CS 112
^{15}N-Leucine *E. coli* CS 117
^{15}N-lysine *E. coli* CS 127
^{15}N-His *E. coli* CS 134
Spin labeling of *E. coli* CS 138

3.9 Discussion 142

3.10 Future Work 150

4.0 ESTIMATION OF ISOTOPIC INCORPORATION IN NMR SAMPLES BY MASS SPECTROMETRY 151

4.1 Introduction 151

4.2 Experimental Procedures 152
Protein for mass spectrometry 152
Simulations 153

4.3 Results 153

4.31 Uniform ^{15}N-labeled *E. coli* CS 156

4.32 Specific ^{15}N-labeled *E. coli* CS **158**

4.4 Discussion **164**

5.0 FLUORESCENCE STUDIES OF *E. COLI* CS **165**

5.1 Introduction **165**

5.2 Experimental Procedures **168**
Mutant F287W 168
Fluorescence Spectroscopy 169

5.3 Results **170**

5.4 Discussion **176**

6.0 REFERENCES **180**

APPENDEX I

LIST OF TABLES:

Table 2.1	**Pairs of complimentary oligonucleotides used for different mutants**	**31**
Table 2.2	**PCR conditions for site-directed mutagenesis in pCC gltA**	**32**
Table 2.3	**PCR conditions used for cloning of gltA gene into pET24b+ vector**	**34**
Table 3.1	**Line widths of resolved cross-peaks of U-^{15}N labeled *E. coli* CS, in the presence and absence of 0.1M KCl**	**68**
Table 3.2	**Line widths of cross-peaks from decoupled and coupled HSQC spectrum of ^{15}N-Met *E. coli* CS**	**99**
Table3. 3	**Intensities of cross-peaks in conformational exchange as measured directly from the HSQC spectrum of ^{15}N-Leu *E. coli* CS at different temperatures**	**125**
Table 3.4	**Percentage incorporation of the spin label 3-iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline) in cysteines residues of *E. coli* CS**	**139**
Table 3.5	**Distance of backbone nitrogen of methionines in *E. coli* CS from sulphur atom of Cys-206**	**141**
Table 3.6:	**Thermodynamic parameters for three pairs of peaks involved in conformational exchange**	**147**
Table 4.1	**Tryptic peptides from *E. coli* CS used in the simulations for calculating isotopic incorporation**	**155**
Table 4.2	**Isotopic incorporation of ^{15}N, ^{13}C in samples prepared for NMR Spectroscopy**	**163**
Table 5.1	**Catalytic properties of CS (WT) and mutant F287W**	**170**
Table 5.2	**Fluorescence emission of the wild type, F287W and W260A CS with and without substrate complex**	**172**
Table 5.3	**Dissociation constants of WT and mutant F287W CS as measured from the change in the fluorescence upon addition of substrates**	**176**

LIST OF FIGURES:

Figure 1.1	**The tricarboxylic acid cycle**	**4**
Figure 1.2	**Hexameric structure of type II *E. coli* CS shown using ribbon representation**	**7**
Figure 1.3	**Hexameric structure of *E. coli* CS showing the residues forming the cationic pore**	**8**
Figure 1.4	**Sequence alignments of residues of type I CS from pig, *P. furiosus* and type II CS from *E. coli***	**11**
Figure 1.5	**Ribbon representation of *E. coli* CS dimer showing the subunits interaction**	**12**
Figure 1.6	**Comparison of polypeptide region around coenzyme A binding site in type I CS from *P. furiosus* type II CS from *E. coli* CS**	**15**
Figure 1.7	**Ribbon representation of *E. coli* CS dimer showing the active site residues in the cleft between the large and small domain**	**16**
Figure 1.8	**Binding site of substrates OAA and Ac-CoA of E. coli CS based on pig and chicken heart CS**	**19**
Figure 1.9	**Catalytic mechanism of CS showing condensation of OAA and AcCoA to form Citrate**	**21**
Figure 1.10	**Ribbon representation of hexameric *E. coli* F383A mutant with six NADH bound to each subunit**	**26**
Figure 1.11	**Graph of average main chain thermal (B-factors) versus residue number**	**28**
Figure 1.12	**Comparison of the mobile loop region between WT *E. coli* CS and CS R109L mutant**	**29**
Figure 2.1	**Elution profile of *E. coli* CS off DE52 Column**	**39**
Figure 2.2	**Elution profile of *E. coli* CS off G200 column**	**40**
Figure 3.0	**Pulse sequence of the basic HSQC sequence**	**50**
Figure 3.1	**One-dimensional and two-dimensional [^{1}H, ^{15}N] HSQC spectra of U-^{15}N *E. coli* CS**	**65**
Figure 3.2	**Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of U-^{15}N *E. coli* CS**	**66**
Figure 3.3	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of U-^{15}N *E. coli* CS in the presence of 0.1M KCl**	**67**
Figure 3.4	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of U-^{15}N**	

E. coli CS in the presence of 0.1M KCl and 1.1 molar excess of OAA and CM-CoA 69

Figure 3.5 Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of U-^{15}N *E. coli* CS in the presence of 1.1 molar excess of 0.1M NADH 71

Figure 3.6 Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of U-^{15}N truncated *E. coli* CS 72

Figure 3.7 Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 0.06mM U-^{15}N, ^{2}H *E. coli* CS at 500 MHz 74

Figure 3.8 Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.9mM U-^{15}N, ^{2}H *E. coli* CS at 600 MHz 76

Figure 3.9 Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H *E. coli* CS at 600 MHz 77

Figure 3.10 Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.9mM U-^{15}N, ^{2}H *E. coli* CS at 800 MHz 78

Figure 3.11 Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H *E. coli* CS at 800 MHz 79

Figure 3.12 Comparison of equivalent regions in two-dimensional [^{1}H, ^{15}N] TROSY spectra of U-^{15}N, ^{1}H and U-^{15}N, ^{2}H *E. coli* CS 80

Figure 3.13 New cross-peaks that were observed in the two-dimensional [^{1}H, ^{15}N] TROSY spectrum of U-^{15}N, ^{2}H, but not in U-^{15}N, ^{1}H *E. coli* CS 81

Figure 3.14 Two-dimensional [^{1}H, ^{15}N] CRINEPT-TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS 82

Figure 3.15 Overlay of two-dimensional [^{1}H, ^{15}N] TROSY spectra of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS in the presence and absence of substrates 84

Figure 3.16 Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured after one month of amide back-exchange 85

Figure 3.17 Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured after two months of amide back-exchange 86

Figure 3.18 Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured after one year of amide back-exchange 87

Figure 3.19 Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Phe

	labeled *E. coli* CS	**90**
Figure 3.20	**B-factors of backbone nitrogen for phenylalanine residues in *E. coli* CS**	**91**
Figure 3.21	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Met labeled *E. coli* CS**	**93**
Figure 3.22	**B-factors of backbone nitrogen for methionine residues in *E. coli* CS**	**94**
Figure 3.23	**Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Met labeled *E. coli* CS in the presence and absence of substrates**	**95**
Figure 3.24	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{13}C-Lys-^{15}N-Met labeled *E. coli* CS, with no decoupling between carbonyl carbon of lysine and alpha nitrogen of methionine**	**97**
Figure 3.25	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{13}C-Lys-^{15}N Met labeled *E. coli* CS, with coupling between carbonyl Cabon of lysine and alpha nitrogen of methionine**	**98**
Figure 3.26	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Met labeled M274L *E. coli* CS**	**101**
Figure 3.27	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Met M168L *E. coli* CS**	**102**
Figure 3.28	**Regions of two-dimensional HSQC spectrum of ^{15}N-Met *E. coli* CS at different temperatures and the variation of chemical shifts with temperature**	**106**
Figure 3.29	**Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of ^{15}N-Met labeled *E. coli* CS**	**108**
Figure 3.30	**B-Factors of backbone nitrogen for glycine residues in *E. coli* CS**	**109**
Figure 3.31	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Gly labeled *E. coli* CS**	**110**
Figure 3.32	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Gly labeled *E. coli* CS in the presence and absence of substrates**	**111**
Figure 3.33	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Trp labeled *E. coli* CS**	**113**
Figure 3.34	**Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Trp labeled *E. coli* CS**	**114**
Figure 3.35	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Trp labeled *E. coli* CS after overnight dialysis**	**115**

Figure 3.36	**Overlay of 1D slice of HSQC spectra of $^{15}N_2$-Trp labeled *E. coli* CS wild-type spectrum and W260A mutant**	**116**
Figure 3.37	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Leu labeled *E. coli* CS**	**117**
Figure 3.38	**B-factors of backbone nitrogen for leucine residues in *E. coli* CS**	**118**
Figure 3.39	**Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Leu labeled *E. coli* CS in the presence and absence of substrates**	**119**
Figure 3.40	**Overlay of two-dimensional [^{15}N, ^{1}H] HSQC spectra of ^{15}N-leu labeled *E. coli* CS at different temperatures**	**120**
Figure 3.41	**Variation in the proton chemical shift of cross-peaks of ^{15}N-Leu labeled *E. coli* CS with temperatures**	**121**
Figure 3.42	**Two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Leu labeled *E. coli* CS at 10 ºC and 5 ºC**	**123**
Figure 3.43	**Regions of ^{15}N-Leu labeled *E. coli* CS showing conformational exchange between two peaks at different temperatures**	**124**
Figure 3.44	**Plot of ln(intensity) vs 1/T to calculate ΔH and ΔS of the cross-peaks in slow conformational exchange of ^{15}N-Leu labeled *E. coli* CS**	**126**
Figure 3.45	**Comparison of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Leu labeled *E. coli* wild type CS and L275A CS mutant**	**128**
Figure 3.46	**Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Lys labeled *E. coli* CS**	**129**
Figure 3.47	**Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Lys labeled *E. coli* CS in the presence and absence of substrates**	**130**
Figure 3.48	**B-factors of backbone nitrogen for lysine residues in *E. coli* CS**	**131**
Figure 3.49	**Comparison of two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Lys labeled *E. coli* wild type CS and truncated CS mutant**	**133**
Figure 3.50	**B-factors of backbone nitrogen for histidine residues in *E. coli* CS**	**135**
Figure 3.51	**Two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-His labeled *E. coli* CS**	**136**
Figure 3.52	**Two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-His labeled *E. coli* CS in the presence of 1M urea**	**137**
Figure 3.53	**Overlay of two-dimensional HSQC spectra of ^{15}N-Met labeled *E. coli* CS in the presence and absence of spin labeling**	**140**
Figure 4.1	**Simulated ion distribution for a peptide at natural abundance and for 50 to 100% incorporation of ^{15}N in a single nitrogen atom**	**154**

Figure 4.2	**Labeling efficiency of U-^{15}N labeled *E. coli* CS**	**157**
Figure 4.3	**Labeling efficiency of ^{15}N-Phe labeled *E. coli* CS**	**159**
Figure 4.4	**Labeling efficiency of ^{15}N-Gly labeled *E. coli* CS**	**161**
Figure 4.5	**Labeling efficiency of ^{15}N-His labeled *E. coli* CS**	**162**
Figure 5.1	**Wire-frame representation of *E. coli* CS showing the three tryptophans**	**167**
Figure 5.2	**Fluorescence spectra of *E. coli* WT CS, mutant F287W and mutant W260A**	**171**
Figure 5.3	**Fluorescence spectra of *E. coli* WT CS, mutant F287W and mutant W260A in the absence and presence of substrates**	**173**
Figure 5.4	**Fluorescence monitored titration of mutant F287W with CM-CoA and OAA**	**174**
Figure 5.5	**Fluorescence monitored titration of WT CS with CM-CoA and OAA**	**175**

LIST OF ABBREVIATIONS:

Ac-CoA	**Acetyl Coenzyme A**
ATP	**Adenosine triphosphate**
Ala	**Alanine**
Arg or R	**Arginine**
Asp	**Aspartic Acid**
$CaCl_2$	**Calcium Chloride**
CS	**Citrate Synthase**
CM-CoA	**Carboxymethyl Coenzyme A**
CoA	**Coenzyme A**
Cys	**Cysteine**
CRIPT	**cross relaxation-induced polarization transfer**
CRINEPT	**cross-correlated relaxation-enhanced polarization transfer**
DEAE	**Diethyl amino ethyl-cellulose**
DTNB	**5,5′-dithiobis-(2-nitrobenzoic acid)**
DSS	**2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt**
E. coli	***Escherichia coli***
EDTA	**Ethylenediaminetetraacetic acid**
ESI-TOFMS	**Electrospray ionization time-of-flight mass spectrometry**
Glu or E	**Glutamic Acid**
Gln	**Glutamine**
Gly	**Glycine**
His	**Histidine**
HSQC	**Heteronuclear single quantum correlation**
Hz	**Hertz**
Ile	**Isoleucine**
IPTG	**Isopropyl β-D-1-thiogalactopyranoside**
KH_2PO_4	**Potassium dihydrogen phosphate**
KCl	**Potassium Chloride**
LB	**Luria-Bertani**
Leu or L	**Leucine**
Lys	**Lysine**
$MgSO_4$	**Magnesium Sulphate**
Met or M	**Methionine**
MnSO4	**Manganese Sulphate**
MALDI	**Matrix-Assisted Laser Desorption/Ionization**
NADH	**Nicotinamide adenine dinucleotide**
NMR	**Nuclear Magnetic Resonance**
NTP	**Nucleotide triphosphate**
Na_2HPO_4	**Disodium hydrogen phosphate**
NH_4Cl	**Ammonium Chloride**
NaCl	**Sodium Chloride**
NaOH	**Sodium hydroxide**
NaN_3	**Sodium Azide**
OAA	**Oxaloacetic Acid**
PCR	**Polymerase Chain Reaction**
Phe or F	**Phenylalanine**
Pro	**Proline**
Ser	**Serine**
SH	**Sulphydryl**
TCA	**Tricarboxylic Acid**
Tyr	**Tyrosine**

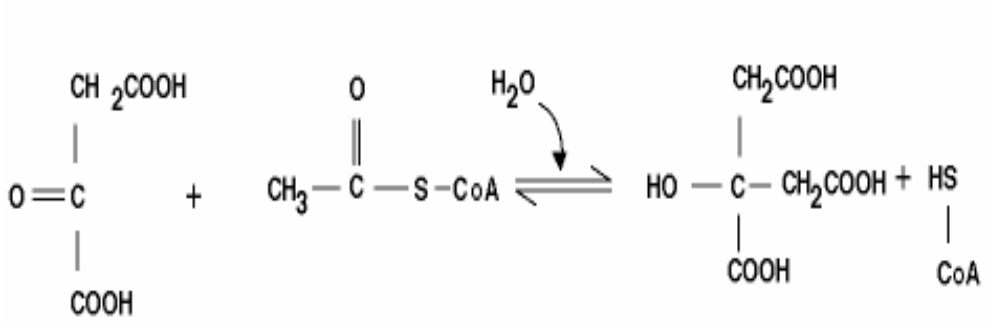
Thr	**Threonine**
TNB	**2-nitro-5-thiobenzoic acid**
Trp or W	**Tryptophan**
Tris	**tris (hydroxymethyl) aminomethane**
TROSY	**Transverse Relaxation Optimized Spectroscopy**
T_1	**spin-lattice relaxation time**
T_2	**spin-spin relaxation time**
P. furiosus	***Pyrococcus furiosus***

1.0 INTRODUCTION

The enzyme studied in the thesis is citrate synthase (CS) from *Escherichia coli*, an allosteric, multisubunit enzyme. Although three-dimensional structures of *E. coli* CS are available, no structures are available with substrates bound with the enzyme. In the structures of *E. coli* CS, a region comprising residues 266-297 the "Mobile loop", is close to the active site and shows the interesting property of very high flexibility. The dynamics of the mobile loop was proposed to be a key part of the allosteric mechanism of *E. coli* CS by the X-ray crystal structure studies [Nguyen, et al., 2001; Stokell, et al., 2003]. To further investigate these highly flexible residues in solution, Nuclear Magnetic Resonance (NMR) spectroscopy studies were initiated. The results and discussion are described in Chapter 3 of the thesis. During the course of above investigation, we found that interpretation of the NMR results, depended largely upon correct knowledge of heavy isotope being incorporated in the NMR samples. A collaborative work, which uses Mass Spectrometry as a technique, was used to analyze the isotopic incorporation in NMR samples. The importance of the method used and its application to the labeled samples of *E. coli* CS are described in Chapter 4. Along with NMR, fluorescence spectroscopy was used to obtain information about the dynamics of residues in the mobile loop. The results of fluorescence spectroscopy of the two mutants, F287W and W260A along with WT CS to study the dynamics of tryptophans are described in Chapter 5.

1.1 Introduction to Citrate Synthase

Citrate Synthase, found in all eukaryotes and in most of the microorganisms, is responsible for catalyzing the entry of two carbon units into the tricarboxylic acid cycle (TCA) (Figure 1.1). The above function is done by catalyzing the condensation of an acetyl group from acetyl Coenzyme A (AcCoA) with oxaloacetate (OAA) to form citrate:



Oxaloacetate Acetyl Coenzyme A Citrate Coenzyme A

In most organisms the TCA cycle is an amphibolic pathway serving the dual functions of energy production and production of metabolites for biosynthesis [Danson, et al., 1973]. Energy is produced in the form of NADH that may be coupled through oxidative phosphorylation to the production of ATP or may be used as a source of reducing hydrogen in biosynthesis [Weitzman & Jones, 1968].

CS is also involved in the glyoxylate cycle that is utilized in some organisms (plants and bacteria) as a method of replenishing intermediates of the TCA cycle and gluconeogenesis. In many bacteria α-ketoglutarate

dehydrogenase is not produced during anaerobic growth and thus an incomplete TCA cycle is found. This variation of the TCA cycle is not responsible for the production of energy but leads to biosynthesis of important biomolecular precursors such as α-ketoglutarate. Production of α-ketoglutarate is important because it is enzymatically converted to glutamate: this reaction is one of the primary pathways for the formation of α-amino groups directly from ammonia [Kornberg HL, 1966].

CS is the first enzyme in the above pathways and thus is a possible target for regulation at the enzyme level by feedback inhibition. Feedback inhibitors are usually products of the reaction such as, in this case, ATP, NADH or α-ketoglutarate. NADH inhibition is restricted to Gram-negative bacteria [Weitzman & Jones, 1968; Weitzman PDJ, 1966] and CS from these organisms forms a separate class of citrate synthase.

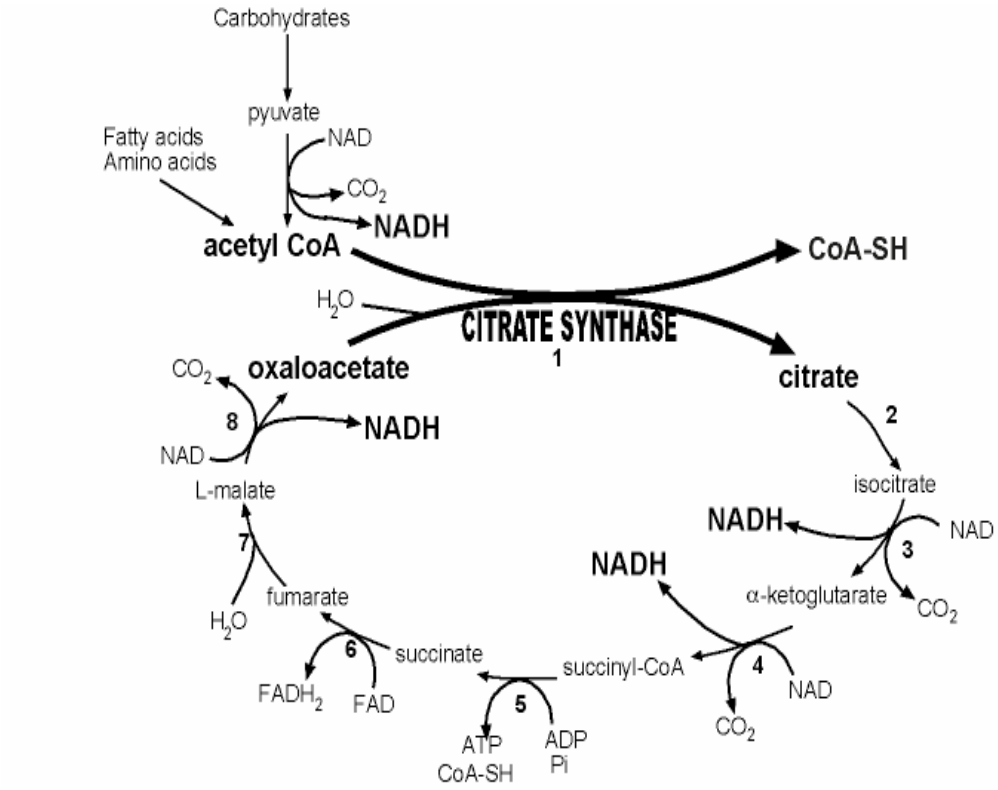


Figure 1.1: The Tricarboxylic Acid Cycle

Enzymes involved in the cycle are: 1) Citrate Synthase 2) Aconitase 3) Isocitrate Dehydrogenase 4) α-ketoglutarate Dehydrogenase 5) Succinyl CoA Synthase 6) Succinate Dehydrogenase 7) Fumarase 8) Malate Dehydrogenase [Nguyen, et al., 2001]

1.2 Classes of Citrate Synthase

The citrate synthases have been classified according to their subunit structure and mode of regulation. Classically, citrate synthases are divided into two main groups [Weitzman & Danson, 1976; Weitzman, 1966, Weitzman & Jones, 1968]: Type I CS, found in eukaryotes, archaea and gram-positive bacteria, are non-allosteric, homodimers while Type II from gram-negative bacteria are allosteric and form predominantly hexamers, by trimerization of the basic dimeric structure. One of the unique properties by which these two classes of citrate synthases differ is the strong allosteric regulation of hexameric CS by NADH, which is not seen in dimeric CS.

In several Gram-negative organisms, a second CS gene has been found, the sequence of which resembles more closely the non-allosteric dimeric class of CS. In *E. coli* and other organisms this type-I like dimeric CS was found to be methylcitrate synthase , which uses propionylcoenzyme A as a substrate [Gerike, et al., 1998]

1.3 Three dimensional structures of Citrate Synthases

Three-dimensional structures from *Sus scrofa* (pig) [Remington, et al., 1982], *Gallus gallus* (chicken) [Liao, et al., 1991], *Thermoplasma acidophilum* (a thermophilic archaeon) [Russell, et al., 1994], *Pyrococcus furiosus* [Russell, et al., 1997] and *Sulfolobus solfataricus* (thermophilic Archaeon) [Bell, et al., 2002] have been determined by X-ray crystallography. All structures of type I CS

are homodimers related by a two-fold axis, with an overall fold that is classified as an alpha motif. Each monomer in the dimeric unit folds into two distinct domains, with the active site situated in a deep cleft between them.

1.4 Three-dimensional structure of type II CS

Three-dimensional structures of type II CS from *E. coli* WT CS [Nguyen, et al., 2001] and mutant proteins [Maurus, et al., 2003; Stokell, et al., 2003] have also been determined by X-ray crystallography. The *E. coli* CS structure consists of a hexameric complex of three dimer units arranged about a central 3–fold axis (Figure 1.2). There is a strong 2-fold relationship between the monomer units in each dimer, with a few regions adopting different conformations in the two monomers, especially around residues 141-164. This is different from the dimeric type I enzymes where a defined 2-fold relationship is found between the monomer units and the conformations of the two monomers in a dimer unit are identical.

Cationic pore

The interactions holding the *E. coli* CS hexamer complex together are localized about a cationic porelike structure that runs along the central 3-fold axis relating the dimer units. All six subunits in *E. coli* CS contribute to the interactions in the pore region. The majority of such contacts involve residues 111-119, 122-127, 178-190 and 204-207. The inner surface of the pore is highly positively charged as it is lined with arginine side chains. The side chains of Arg 119, Arg 125, and Arg 126 (18 in total from the six monomer units)

contribute to the cationic pore and form an extensive network of hydrogen bonds and charged interactions (Figure 1.3). Most of the amino acids forming the cationic pore are strongly conserved among Type II CSs but the equivalent region is very different in Type I enzymes.

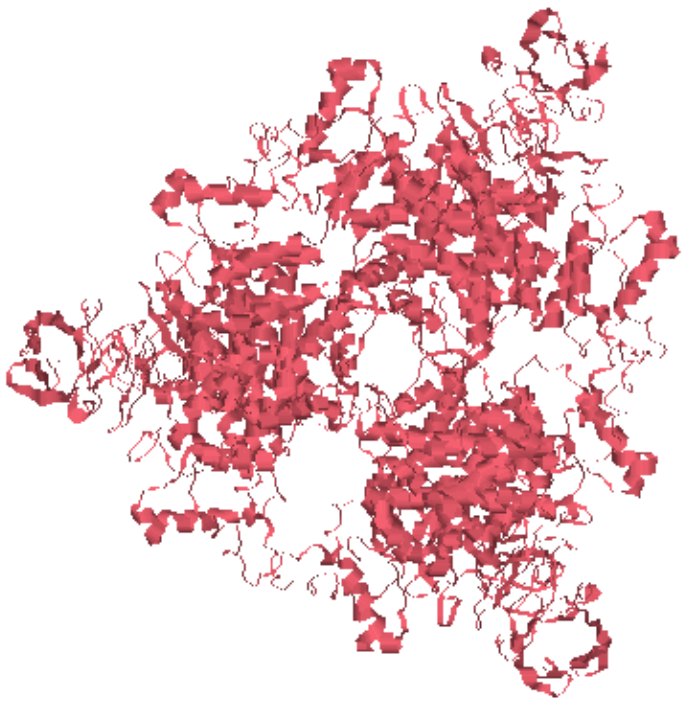


Figure 1.2: Hexameric structure of type II *E. coli* CS shown using ribbon representation. The three equivalent dimer units which are related by central 3-fold axis are coloured red [Nguyen, et al, 2001].

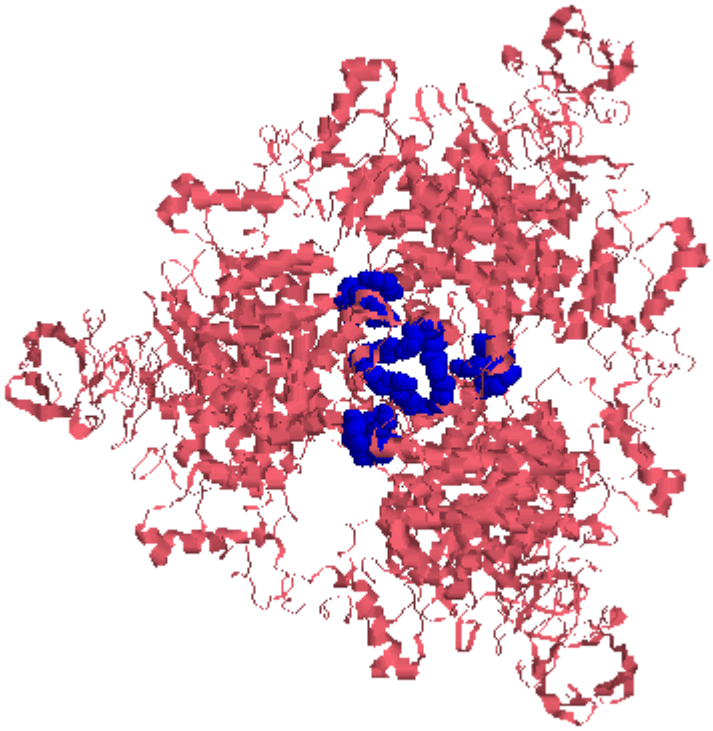


Figure 1.3: Hexameric structure of type II *E. coli* CS with the three dimer units shown in red in ribbons. Residues forming the cationic pore R119, R125, R126 from each subunit are coloured blue in space-fill representation [Nguyen, et al, 2001].

Secondary structural elements and subunit-subunit contacts

E. coli CS is mostly helical consisting of 14 α-helical segments extending throughout the protein sequence. The β-sheet segment is restricted to the first 52 residues of the N-terminal region. Comparison of the sequences in Figure 1.4 shows that all the helices found in *E. coli* CS have counterparts in type I CS. The main differences between these structures occur at the interhelical loops.

The subunit consists of three domains, namely residues 1-52 (N-terminal β-sheet domain); 53-262 and 375-426 (large domain); and 263-374 (small domain). The larger domain contains 12 helices (C-M and S). The subunit-subunit contact is mediated by four antiparallel pairs of helices FF′, GG′, LL′, MM′ (the prime indicates the other subunit) forming an eight helical sandwich. The dimer interface is wrapped over either side by the bent antiparallel helices S and I (Figure 1.5). Helix I is kinked due to the presence of a proline residue whereas helix S is smoothly bent. The interactions between each pair of helices S and I and the 8-helical sandwich are mostly mediated by the hydrophobic interactions.

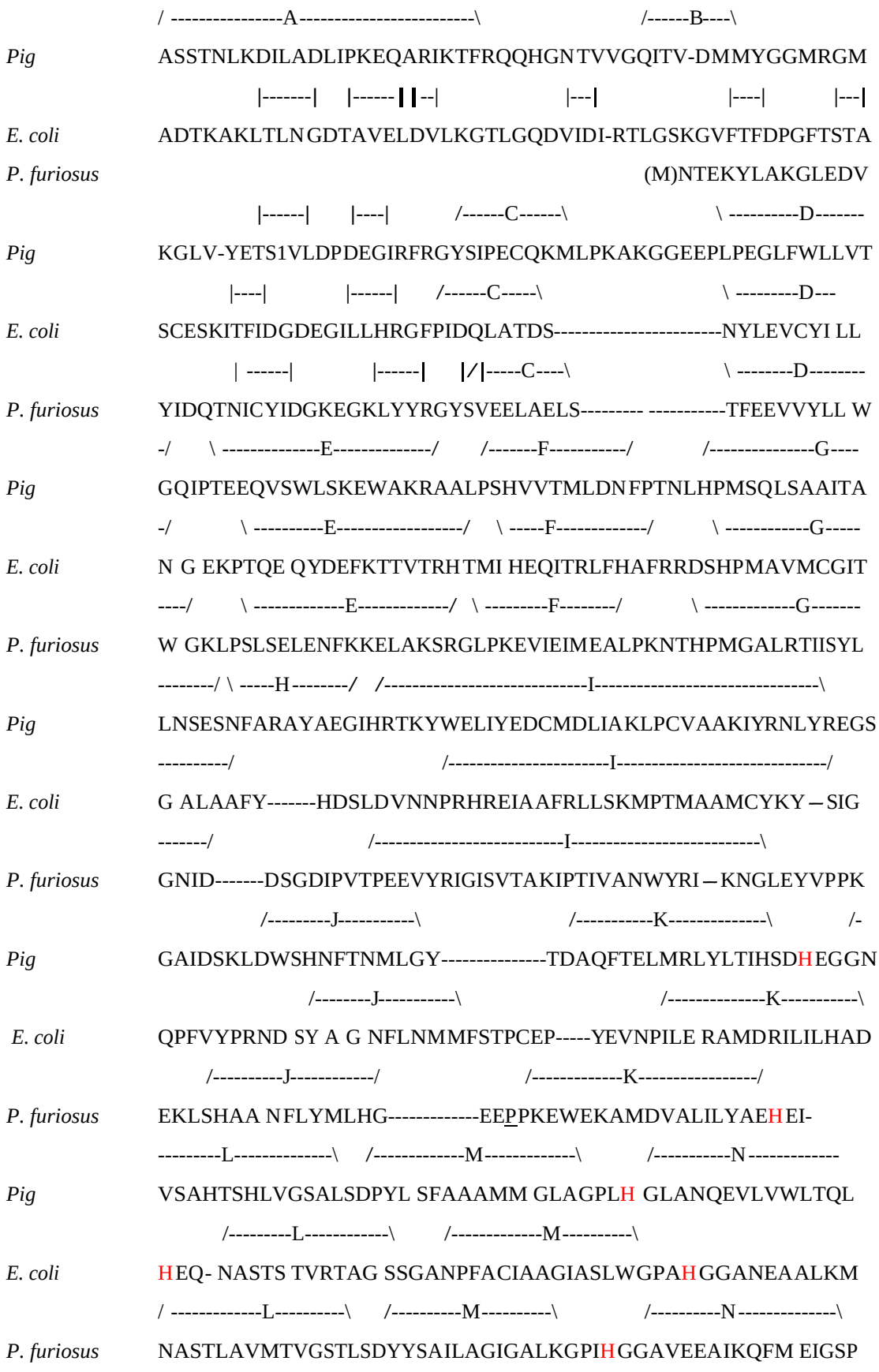

/ ----------A----------\ /------B----\
Pig ASSTNLKDILADLIPKEQARIKTFRQQHGNTVVGQITV-DMMYGGMRGM
|-------| |------||--| |---| |----| |---|
E. coli ADTKAKLTLNGDTAVELDVLKGTLGQDVIDI-RTLGSKGVFTFDPGFTSTA
P. furiosus (M)NTEKYLAKGLEDV
|------| |----| /------C------\ \ ----------D-------
Pig KGLV-YETS1VLDPDEGIRFRGYSIPECQKMLPKAKGGEEPLPEGLFWLLVT
|----| |------| /------C-----\ \ ---------D---
E. coli SCESKITFIDGDEGILLHRGFPIDQLATDS----------NYLEVCYI LL
| ------| |------| |/|-----C----\ \ --------D--------
P. furiosus YIDQTNICYIDGKEGKLYYRGYSVEELAELS--------- ----------TFEEVVYLL W
-/ \ ----------E----------/ /-------F----------/ /----------G----
Pig GQIPTEEQVSWLSKEWAKRAALPSHVVTMLDNFPTNLHPMSQLSAAITA
-/ \ ----------E----------/ \ -----F----------/ \ ----------G-----
E. coli N G EKPTQE QYDEFKTTVTRHTMI HEQITRLFHAFRRDSHPMAVMCGIT
----/ \ ----------E----------/ \ ---------F--------/ \ ----------G-------
P. furiosus W GKLPSLSELENFKKELAKSRGLPKEVIEIMEALPKNTHPMGALRTIISYL
--------/ \ -----H--------/ /----------I----------\
Pig LNSESNFARAYAEGIHRTKYWELIYEDCMDLIAKLPCVAAKIYRNLYREGS
----------/ /----------I----------/
E. coli G ALAAFY-------HDSLDVNNPRHREIAAFRLLSKMPTMAAMCYKY—SIG
-------/ /----------I----------\
P. furiosus GNID-------DSGDIPVTPEEVYRIGISVTAKIPTIVANWYRI—KNGLEYVPPK
/---------J----------\ /----------K----------\ /-
Pig GAIDSKLDWSHNFTNMLGY----------TDAQFTELMRLYLTIHSDHEGGN
/--------J----------\ /----------K----------\
E. coli QPFVYPRND SY A G NFLNMMFSTPCEP-----YEVNPILE RAMDRILILHAD
/----------J----------/ /----------K----------/
P. furiosus EKLSHAA NFLYMLHG----------EEPPKEWEKAMDVALILYAEHEI-
---------L----------\ /----------M----------\ /----------N----------
Pig VSAHTSHLVGSALSDPYL SFAAAMM GLAGPLH GLANQEVLVWLTQL
/---------L----------\ /----------M----------\
E. coli HEQ- NASTS TVRTAG SSGANPFACIAAGIASLWGPAHGGANEAALKM
/ ----------L----------\ /----------M----------\ /----------N----------\
P. furiosus NASTLAVMTVGSTLSDYYSAILAGIGALKGPIHGGAVEEAIKQFM EIGSP

Figure 1.4: Sequence alignment of type I *pig* [Bloxham, et al., 1980; Bloxham, et al., 1981] and *pyrococcus furiosus* [Russell, et al., 1997] citrate synthases and type II *E. coli* [Bhayana & Duckworth, 1984] citrate synthase. Regions of α-helix are indicated above the sequences as a series of dashes, with the first and last residues labeled with oblique strokes. The letters of the helix follow the convention of Remington et al., 1982β strands are indicated similarly with the first and last residues labeled with vertical bars. Active site residues in the three sequences are colored red.

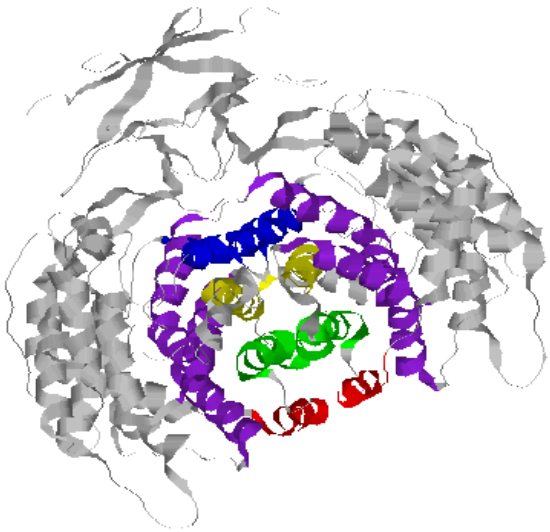


Figure 1.5: Ribbon representation of *E. coli* CS dimer showing the interaction between the subunits interaction. The CS dimer interface is made up of four antiparallel pairs of helices F (red), G (green), M (yellow), L (blue) and the related counterparts in the other subunit. Helices S and I (purple) wrap around the dimer interface [Nguyen, et al., 2001].

Dimer-Dimer Interactions

The dimer-dimer interaction is mediated by residues localized in a small region of the large domain. The residues involved are 112 (from the corner EF), residues 114-116, 119, 122 (Helix F), residues 125-127 (corner GF), residues 181-183, 187, 189 (corner IJ) and residues 205-208 (corner JK). Most dimer-dimer interactions are van der Waals contacts, with one hydrogen bond between the carbonyl oxygen of Cys 206 in one dimer and the main chain amide nitrogen of Asn 189 of the other.

N-terminal Domain

One of the unique features of *E. coli* CS is the 52 N-terminal amino acids not present in non-mammalian type I CS (*P. furiosus*, Figure 1.4) Eukaryotic type I CS does have an N-terminal extension but this is folded completely differently, largely as a long helix that lies on the outer surface of the subunit. The N-terminal in *E. coli* CS is formed by the contribution of both subunits of the dimers to form a single compact and independently folded domain rich in β-sheet structure. The function of the N-terminal domain is not known, but its location in the hexameric *E. coli* CS structure is remote from both the dimer-dimer contact regions and the catalytic sites.

Active Site

A three-dimensional structure of *E. coli* CS with substrates bound is not yet available. On the other hand, the amino acids associated with substrate binding by Type I CS have been defined by extensive structural

studies employing a number of complexes with citrate, OAA and various coenzyme A analogues [Wiegand & Remington, 1986; Karpusas, et al., 1990; Remington, 1992]. Equivalent residues were recognized in structural studies of Type I CS from extremophiles [Russell, et al., 1994; Russell, et al., 1997; Russell, et al., 1998]. Sequence alignments show that all the key residues of the active site are also fully conserved among Type II citrate synthases (residues shown in red in Figure 1.4). The active site in *E. coli* CS is therefore located by comparison with the residues in Type I CS. Many studies in *E. coli* CS, mostly by site-directed mutagenesis, have verified the roles of these active sites residues in the CS catalytic mechanism [Anderson, et al., 1988; Pereira, et al., 1994; Handford, et al., 1988, Man, et al., 1994].

On comparing the active site regions between *P. furiosus* type I and *E. coli* type II CS differences were observed relating to the binding of substrate acetyl-CoA. The most important difference is in the position of one of the active site residues, H264, a key residue involved in the binding of acetyl-CoA. In *E. coli* CS the imidazole group of H264 is ~ 11Å away from its required location for catalysis. The second difference is in residues 299-303, required for binding the adenine moiety of coenzyme A, which in *E. coli* CS are removed from the position needed to form the binding site. These two differences mean that the residues are not oriented correctly to form the binding site of acetyl-CoA in *E. coli* CS. A further difference is the presence of high mobility of main chain atoms of residues 266-297 in *E. coli* CS. The B-

factors [Carugo & Argos, 1998], which provide an indication of flexibility, of these residues are as much as 9 times the average of other residues in *E. coli* CS.

The regions around the coenzyme A binding site in *P. furiosus* Type I and *E. coli* type II CS are shown in Figure 1.6. Comparison shows that the coenzyme A binding site is well folded in Type I CS with coenzyme A bound. In *E. coli* Type II CS, with no coenzyme A bound the polypeptide adjacent to (residues 266-297), and within, the acetyl-CoA binding site is in a more extended or open conformation.

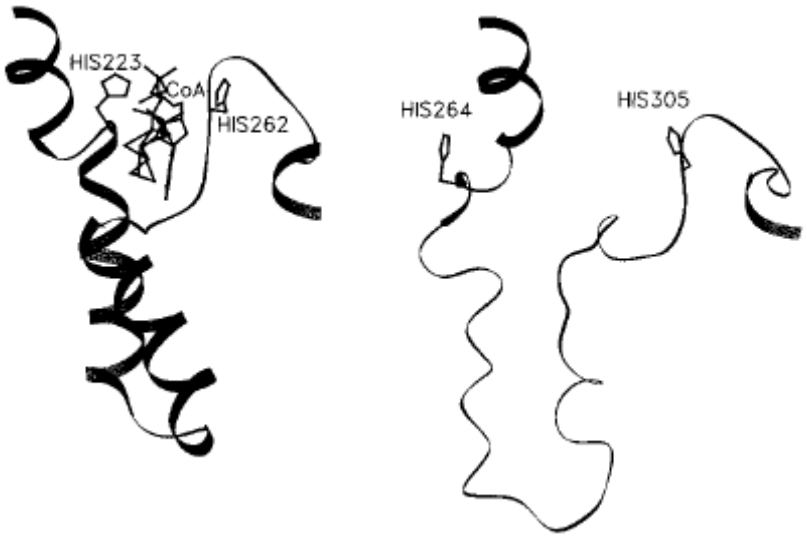


Figure 1.6: Polypeptide region around the coenzyme A binding site in type I CS from *P. furiosus* with bound CoA and type II CS from *E. coli* CS (no substrates bound). Substantial refolding is required in Type II *E. coli* CS for the region consisting of residues 266-297 to form the coenzyme A binding site. Figure adapted from Nguyen, et al., 2001.

1.5 Activity of Citrate Synthase

As described earlier, each subunit in *E. coli* CS consists of three domains: the N-terminal, a large domain and a small domain. The active site of CS is located in the cleft between the large and small domains (Figure 1.7) and involves residues from both subunits in the dimer. Therefore the minimal functional unit of CS is a dimer with two active sites. The individual subunit active sites are formed predominantly by residues within a particular subunit except that Arg407, which assists in OAA binding, is contributed from the other subunit of the dimer.

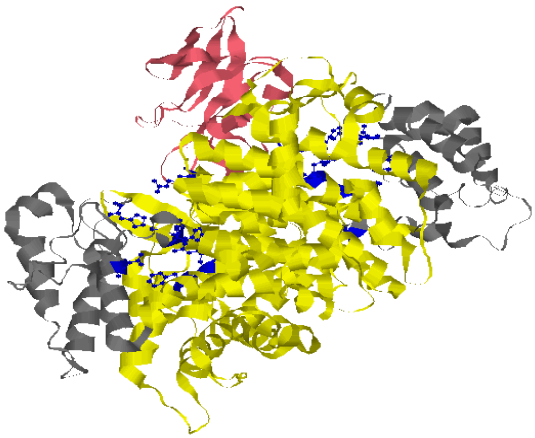


Figure 1.7: Ribbon representation of *E. coli* CS dimer showing the active site residues in the cleft between the large and small domains. The active site residues in both subunits are coloured blue and represented by ball-stick model. The N-terminal, large and small domains are coloured red, yellow and gray respectively.

Crystal studies of type I CS together with product, substrate or substrate analogue complexes show that the enzyme is capable of existing in two most populated conformations [Remington, et al., 1982; Liao, et al., 1991; Karpusas, et al., 1991]. One is a "closed" form of the enzyme which occurs in the presence of OAA or citrate and CoA, while the second "open" form occurs in the absence of bound ligands. The conformational difference between the two forms is described by an 18.5º rotation of the small domain relative to the large domain. After the conformational transition, the small domain makes contact with the large domain of the other subunit [Remington, 1982]. Lesk & Chothia, 1984 described domain closure as the result of small relative motions of packed helices within each domain as well as at the interface. In that study it was suggested that significant rotations and shifts of the small domains relative to each other and relative to the large domains, together with the large movement of the OP loop towards the large domain of the other subunit are responsible for domain closure in CS. This OP loop contains the backbone atoms that interact with the adenine moiety of the acetyl-CoA [Lesk & Chothia, 1984].

A schematic representation of the active site of *E. coli* CS, based on crystal structures of pig and chicken heart, CS is shown in Figure 1.8. OAA is bound by hydrogen bonding to three arginines (R387, R314 and R407′; where R407′ is contributed from the second subunit) and two histidines (H305 and

H229). H305 is responsible for polarizing the carbonyl group of OAA, making it more susceptible to condensation.

The conformational change induced by binding of OAA leads to the closed form of the enzyme and thus creates the binding site for acetyl-CoA. Residues from the small and large domains of both subunits participate in binding acetyl-coenzyme A. The adenine moiety of the coenzyme A is recognized by a loop (pig CS residues 314-320), forming three hydrogen bonds and wrapping around the ring edgewise. The three charged phosphate moieties of the coenzyme form salt bridges with arginine residues. The 5′ diphosphates interact with one subunit (pig CS Arg 324 and Arg 46), and the 3′-phosphate interacts with the other subunit (pig CSArg164′). In the *E. coli* CS active site, the acetyl group interacts with H264 and D362, the residues involved in the catalytic mechanism of CS. F383, which is a highly conserved residue in all citrate synthases, makes an edge-on interaction with the acetyl group of acetyl-coA. Mutation of this residue [Pereira, et al., 1994] showed that it plays a role in the affinity of CS for AcCoA and in the allosteric equilibrium of *E. coli* CS.

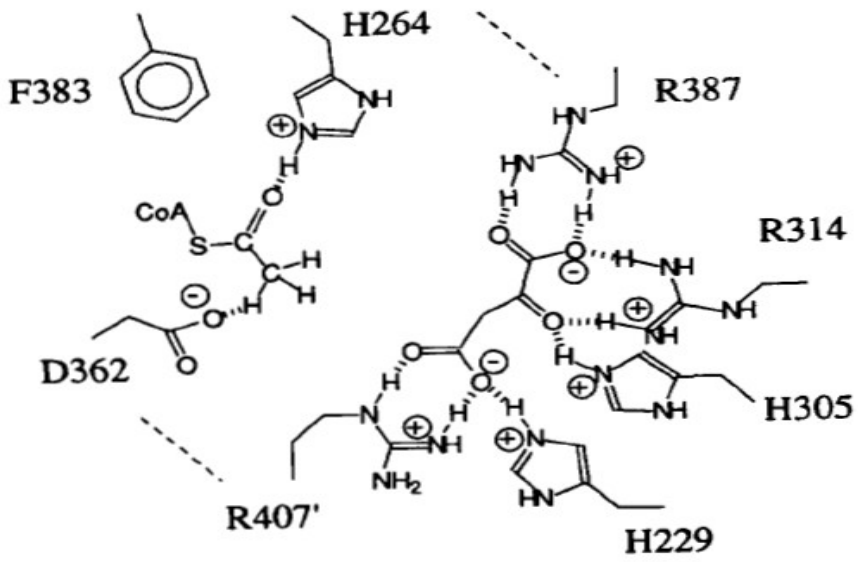


Figure 1.8: Binding of OAA and AcCoA at the active site of *E. coli* citrate synthase based on pig and chicken heart citrate synthases. Figure adapted from Pereira, et al., 1994.

1.6 Catalytic Mechanism

The reaction catalyzed by CS can be divided into three chemical events: enolization of the acetyl group of AcCoA, condensation of OAA and AcCoA to form citryl CoA and hydrolysis of citryl CoA into CoA and citrate.

The current working mechanism of the CS catalyzed reaction [Evans, et al., 1996] is shown in the Figure 1.9. OAA binds in the cleft between the large and the small domains and induces a conformational change to the closed conformer. Binding of CS-OAA with AcCoA is facilitated by the presence of a positive charge created by polarization of the carbonyl bond of OAA by H305 and also by H264 and D362 which act in concert in an acid-base reaction to enolize the AcCoA. The nucleophile then condenses stereospecifically with the *si*-face of the activated OAA to give citryl–CoA. The newly formed citryl CoA induces additional structural changes in the enzyme. The active site becomes completely closed, and an aspartate residue is recruited to enable a trapped water molecule to hydrolyze the thioester bond. CoA leaves the enzyme, followed by citrate, and the enzyme returns to the open conformation.

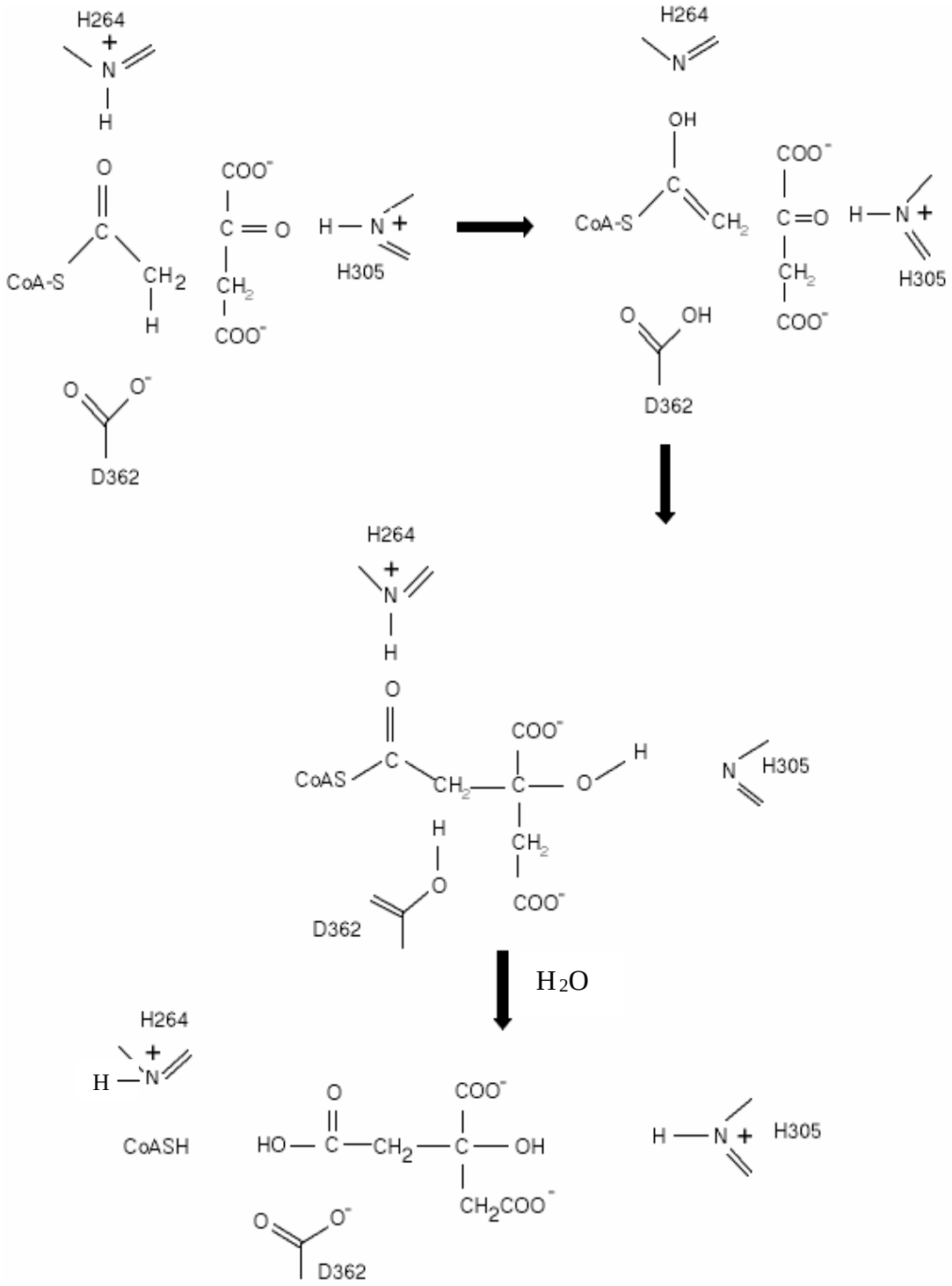


Figure 1.9: Catalytic Mechanism of CS showing condensation of OAA and AcCoA to form Citrate [Evans, et al., 1996].

1.7 *E. coli* Citrate Synthase

Enzymology

E. coli CS is subjected to allosteric inhibition by NADH [Bhayana & Duckworth, 1984; Weitzman & Danson, 1976]. The structural basis for this inhibition has two features, an allosteric binding site for NADH, distinct from the active site [Danson & Weitzman, 1973; Talgoy, et al., 1979]; and a conformational equilibrium between two conformational states, the T state and the R-state (the terminology of Monod, et al., 1965 is used here for the allosteric states). The T state in *E. coli* CS binds NADH and is inactive, the R-state is active and does not bind NADH [Duckworth & Tong, 1976].

CS from *E. coli* exhibits a sigmoid AcCoA saturation curve, but KCl an allosteric activator of the enzyme [Weitzman, 1966] improves the binding of Ac-CoA and converts the saturating curve from sigmoid to hyperbolic [Faloona & Srere, 1969; Morse & Duckworth, 1980]. KCl also abolishes NADH inhibition [Duckworth & Tong, 1976]; this suggests that the *E. coli* CS allosteric R state binds Ac-CoA and KCl well and is active. OAA binds equally well to T and R states and its binding is hyperbolic [Anderson, et al., 1991]. The above features imply that *E. coli* CS, in the absence of any ligands, is shifted towards an inactive state (allosteric T state) since NADH binding is favoured and substrate binding is difficult. In the presence of KCl, the allosteric activator, the equilibrium is shifted towards the allosteric R state, since NADH binding is difficult while substrate binding is not.

Subunit structure

Equilibrium ultracentrifugation studies [Tong & Duckworth, 1975] showed that aggregates greater than a dimer are favoured at low pH. In the presence of KCl, an activator of the enzyme, oligomerization states are much simplified. In the presence of 0.05 M KCl, molecular species corresponding to dimers and hexamers are present whereas only hexamer is present in the presence of 0.1M KCl. Later, ESI-TOFMS studies done on *E. coli* CS [Ayed, et al., 1998] further confirmed the previous finding that at low ionic strength, only dimer and hexamer are present in significant amounts. Theoretical masses of CS dimer (95,770 Da) and hexamer (287,310 Da) were calculated using ESI-TOFMS. The dimer-hexamer equilibrium was also shown to be affected by the concentration of the enzyme, hexamer being favoured at high concentrations. Under the conditions used by Ayed et al., 1998, at a protein concentration of 20μM in the presence of 5mM NH_4Cl, equal amounts of dimer and hexamer were present whereas at 100 μM enzyme predominantly hexamer was found. NMR experiments with *E. coli* CS, to be described in this thesis, were carried out at millimolar concentrations and therefore predominantly hexamer is expected under these conditions.

NADH binding and inhibition

The specific inhibition of Gram–negative bacterial citrate synthases by NADH is an example of an allosteric phenomenon which also distinguishes them from Type I CS. NADH inhibition of CS has been studied in a number

of organisms, but most extensively in *E. coli* and strictly aerobic genera *Acinetobacter* and *Pseudomonas*. The first evidence of NADH inhibition came from studies on the *E. coli* enzyme (Weitzman et al., 1966). The inhibition was shown to be desensitized in the presence of 0.2 M KCl or at mildly alkaline pH (~9).

NADH binding to *E. coli* CS was first studied by Duckworth & Tong, 1976 using fluorescence enhancement and gel filtration techniques. Inhibition of *E. coli* CS by NADH was found to be strongest at pH values below 7.5 and fell off rapidly as the pH was raised as discovered previously [Weitzman, 1966]. At pH 6, a maximum of 0.65 binding sites per subunit was found with a K_D approaching 0.28μM, whereas at pH values closer to 9, 0.25 binding sites per subunit were found. The dependence of the number of NADH binding sites on pH was attributed to the different oligomeric forms having different binding affinities for NADH. Later, NADH binding sites to the dimeric and hexameric CS oligomers were calculated using ESI-TOFMS [Ayed, et al., 1998]. The study confirmed the previous suggestions that NADH shifts the oligomeric ratio towards hexamer. Binding to the hexamer involves two classes of sites, one with tighter binding (six sites per hexamer) and the second with loose binding (at least 12 sites). The tight and loose binding corresponds to specific and non-specific NADH binding respectively. Weak, non-specific binding with two sites per dimer was also reported [Ayed, et al., 1998].

Three-dimensional crystal structures of NADH-bound *E. coli* CS F383A further show six distinct NADH binding sites [Maurus, et al., 2003], quite remote from the active sites, in the hexameric *E. coli* CS complex [Figure 1.10]. The binding sites are located at the interfaces between dimer subunits, so that most of each site is formed by one subunit but some key residues are drawn from the adjacent dimer. This explains the previous finding that NADH binding induces a shift to hexamer in the dimer-hexamer equilibrium. NADH is bound to the protein via contributions of nine amino acid side chains by hydrogen bonds and salt bridges. The pyrophosphate moiety of NADH binds to R163, K167, R109 and H110. The adenine end of NADH interacts at the dimer-dimer interface with residues 182-191 of one dimer and residues 204-211 of the other. This arrangement explains the NADH regulation present only in hexameric CS rather than dimeric CS.

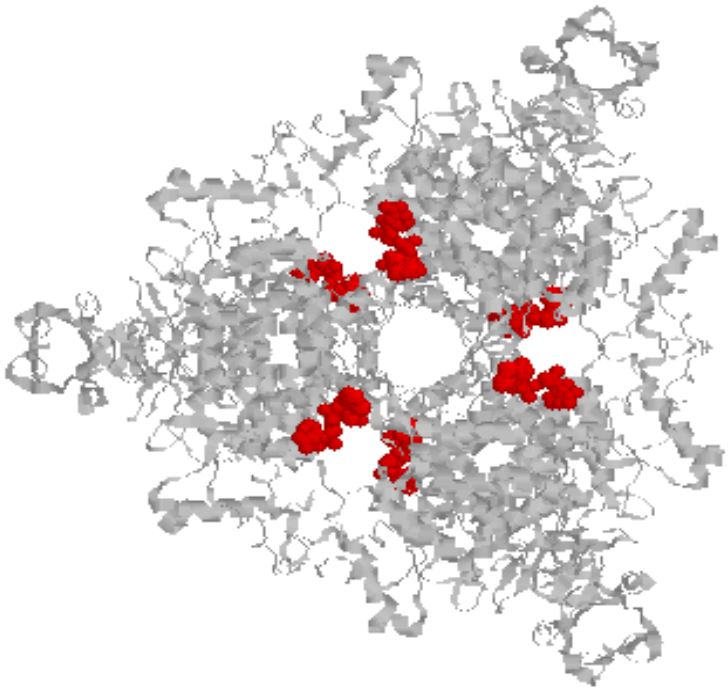


Figure 1.10: Ribbon representation of hexameric *E. coli* CS F383A mutant with all subunits coloured as gray. Six NADH molecules bound to the enzyme, are shown in red in space-fill representation. The NADH binding sites lie at the dimer-dimer interface showing that allosteric regulation is possible only in the hexameric form of CS [Maurus, et al., 2003].

1.8 Mobile Regions in *E. coli* CS as determined from X-ray crystallography

Three-dimensional structures are available for *E. coli* CS wild type and mutant proteins, with [Maurus, et al., 2003; Stokell, et al., 2003] and without [Nguyen, et al., 2001] allosteric inhibitor NADH. The essential features of the structure have been explained in the previous section.

The three-dimensional crystal structure of *E. coli* CS shows some regions of high mobility as indicated by their high B-factors. B-factors indicate the precision of atom positions, so higher B-factors indicate higher mobility. Residues 1-5 in the N-terminal region, residues 266-297 and residues 330-335 show B-factors which are about 9 times the average value of the rest of the residues (Figure 1.11) [Nguyen, et al., 2001].

Residues 266-297 which will be referred to as the mobile loop form a part of the active site region, loosely folded on the outer surface of the protein. As discussed previously, His 264 (active site residue) in *E. coli* CS, at one end of the mobile loop is in the wrong position for catalysis. From a comparison with the equivalent region in type I CS, the X-ray crystal structure of *E. coli* CS suggested that partial refolding of the mobile loop is required to form the binding site for the substrate acetyl-CoA in *E. coli* CS [Nguyen, et al., 2001]. Residues 330-335 will be referred to as "helical linker",

these act as a link to connect the active site region including the mobile loop residues to the allosteric NADH binding site [Stokell, et al., 2003].

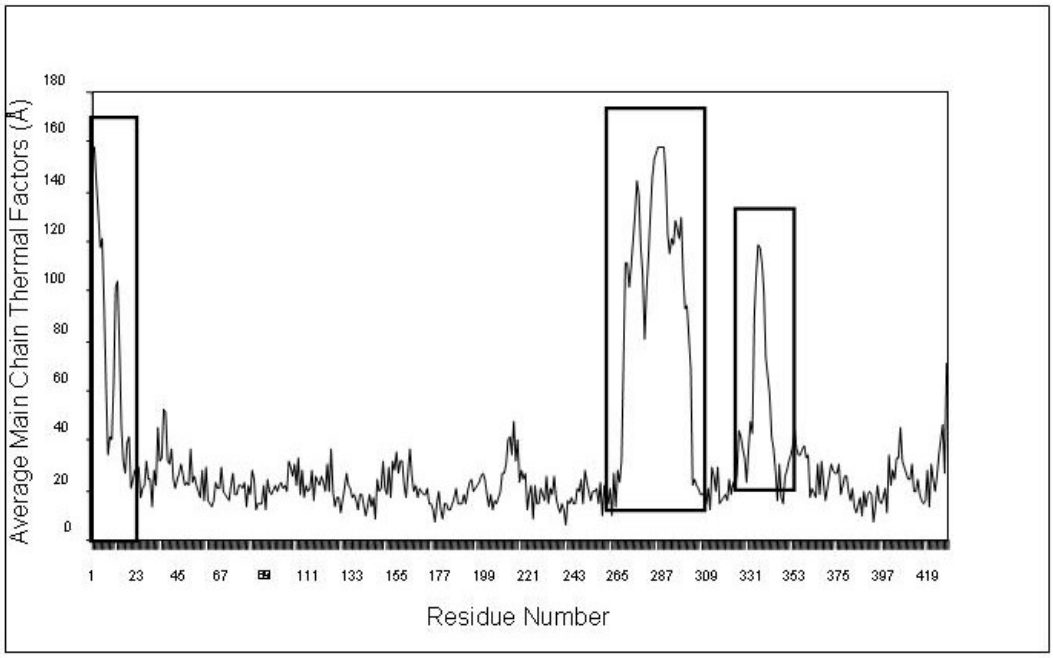


Figure 1.11: Graph of Average Main Chain Thermal (B-factors) versus Residue Number of *E. coli* CS. The peaks marked with boxes with high B-factors belong to residues in the N-terminal region, mobile loop residues "266-297" and helical linker residues "330-335". Figure adapted form Nguyen, et al., 2001.

In one mutant protein, R109L, the mobility of the residues 262-298 and 316-342 is greatly reduced [Stokell, et al., 2003]. Residues 262-298, which constitute mostly the residues in the mobile loop, are partially refolded in the mutant protein and the conformation appears to be more towards that required for binding acetyl-CoA [Figure 1.12]. This again is based on the comparison of equivalent residues in Type I CS structures. Reduction in the mobility of the second region, residues 316-342, helps provide a more rigid structure for the packing of residues of 266-297. The overall arrangement shows that residues 316-342, structurally link the refolded residues (266-297) to the NADH binding site [Stokell, et al., 2003].

(a) **(b)**

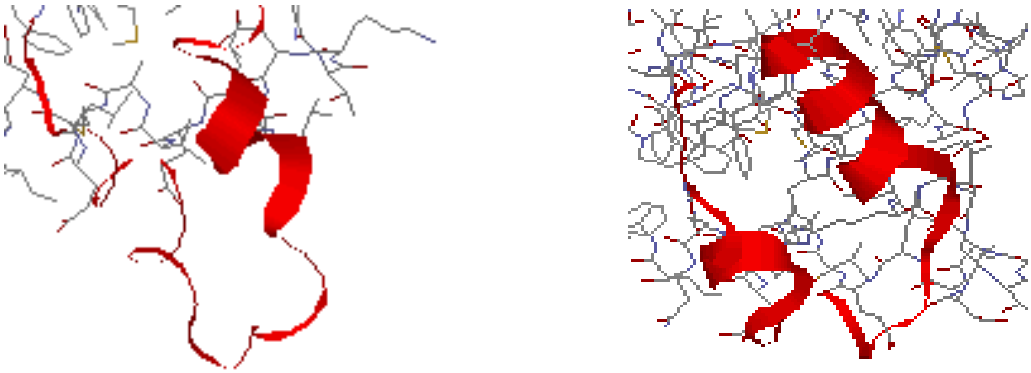


Figure 1.12: Comparison of the mobile loop region in (a) *E. coli* CS WT [Nguyen, et al., 2001] (b) *E. coli* CS R109L [Stokell, et al., 2003]. Residues "266-297" of the mobile loop are coloured in red and are shown in ribbon representation. The mobile loop residues in the *E. coli* R109L CS show a partially refolded structure compared to that in *E. coli* CS WT.

1.9 Thesis Objectives

The objective of the thesis is to study the mobile regions of *E. coli* CS in solution using the technique of Nuclear Magnetic Resonance spectroscopy (NMR). Attempts will be made to identify the mobile residues using different approaches to see if these residues are involved in the function of the protein as suggested by crystal structures. Previous work on the enzyme using NMR spectroscopy includes ^{19}F NMR, to study the environment around reactive cysteine, Cys-206 of *E. coli* CS [Donald, et al., 1991].

During the course of the above investigation, different isotopic labeled *E. coli* CS samples were prepared. The methodology normally used to measure the extent of isotopic incorporation is not suitable for large proteins such as *E. coli* CS. The objective of the second investigation (Chapter 4) was to use mass spectrometry to quantify the amount of isotopic incorporation in different samples prepared for NMR spectroscopy.

A third objective is to study conformational changes in the mobile loop in *E. coli* CS using fluorescence spectroscopy.

2.0 EXPERIMENTAL PROCEDURES

2.1 Site-directed Mutagenesis

CS mutants were prepared using the Quick change Site-directed Mutagenesis Kit obtained from Stratagene. The template used was the pCCgltA plasmid (pUC18 plasmid containing the *E. coli* CS gene, gltA, insert via Sal1 and EcoR1 restriction sites) created by Charlton Cooper. Pairs of complementary oligonucleotides were obtained from Genosys (Sigma) (Table 2.1). Three design parameters were kept in mind while designing the primers: 1) the mutations should be located in the centre of the oligonucleotides to provide stable binding on both sides; 2) guanine or cytosine bases should be found at the ends; 3) melting temperature of the primers was ~72-78ºC based on the equation 2.1 [Rychilik, et al., 1989].

$$T_m = 2^oC \text{ X (number of A′s and T′s in the primer) + } \quad 2.1$$
$$4\ ^oC \text{ X (number of G′s and C′s in the primer)}$$

Table 2.1: Pairs of complimentary oligonucleotides used for different mutants

Mutation	Primers (5′-3′)
M168L UP	CTGCTGTCGAAACTGCCGACCATGGCC
M168L DN	GGCCATGGTCGGCAGTTTCGACAGCAG
M274L UP	GCGCTGAAACTGCTGGAAGAAATCAG
M274L DN	CTGATTTCTTCCAGCAGTTTCAGCGC
R109E UP	CTACGGTGACCGAACATACCATGATC
R109E DN	GATCATGGTATGTTCGGTCACCGTAG
R109K UP	CTACGGTGACCAAACATACCATGATC
R109K DN	GATCATGGTATGTTTGGTCACCGTAG
F287W UP	CATTCCGGAATGGGTTCGTCGTGCG
F287W DN	CGCACGACGAACCCATTCCGGAATG

Primer stock was made by first adding 100µl of nanopure water to the oligonucleotide stock tubes and then taking aliquots to get a concentration of 100ng/µl. The reaction was carried out using 1µl dNTPs (supplied with the kit), 130ng of the two primers (stocks of 100 ng/ µl), 5-50ng template (stock of 7-10ng/µl), 1 µl Pfu DNA polymerase, 5µl Pfu DNA polymerase buffer and nanopure water to a total volume of 50µl. The PCR conditions used for the mutagenesis are given in Table 2.2.

Table 2.2: PCR conditions for Site-directed Mutagenesis in pCC gltA

Cycle	Temperature	Time
1	95 ºC	1 min
2	95 ºC	1 min
3	55 ºC	1 min
4	68 ºC	11 min
	16 times cycle 2	
5	68 ºC	11 min
6	4 ºC	End

The PCR reaction produced a small quantity of homogenous mutant plasmid DNA containing staggered nicks. The DNA was then digested for 1hour at 37ºC with the addition of 1µl of Dpn1 restriction endonuclease (which digests only methylated template DNA). The undigested mutant DNA was then transformed into Epicurian Coli XL1-Blue cells (1ul DNA/ 50µl cells), which take up the DNA and repair the staggered nicks. Cells were initially incubated on ice for 30 minutes following the DNA addition. Cells were then heat shocked for 2 minutes in a 42ºC water bath, followed by two minutes of incubation on ice. NZY^+ broth equilibrated to

42ºC was added (0.5ml), followed by a 1 hour incubation at 37ºC [Quick-Change Site-Directed Mutagenesis kit]. Alternatively, transformation of the PCR products followed by Dpn1 digestion works equally well with $CaCl_2$ Competent MOB154 cells. The entire culture was then spread on several LBA plates and incubated overnight at 37ºC. DNA preparations were done as standard laboratory procedures [Maniatis, et al., 1982]. The presence of only the desired change was verified either by DNA sequencing or by analysis of tryptic fragments of mutant protein by Mass Spectrometry.

2.2 Sub-cloning into pET24b$^+$ vector

E. coli CS gene, *gltA*, is regulated by its native promoter [Duckworth, et al., 1982], but it produces very little CS protein when the host cells are grown in minimal salts glucose medium (M. Bidinosti, unpublished observations) as required for the isotope labeling of the protein. Consequently we used the polymerase chain reaction to introduce new restriction sites into the original plasmid, pES*glt*A [Anderson, et al., 1988] and sub cloned the gene into vector pET 24b+ (Novagen), which places the *gltA* gene under control of the T7 expression system.

PCR amplification was carried out using 32pmoles of each primer, pCC / pES gltA plasmid as the template stock with the desired mutation, 1ul dNTPs (20mM of each nucleotide), 20ul of Pfx DNA polymerase buffer, 2ul of 50mM $MgSO_4$, 0.3ul of Pfx DNA polymerase and nanopure water to a total of 100ul. The PCR conditions used are given in Table 2.3.

Table 2.3: PCR conditions used for cloning of gltA gene into pET24b+ vector

Cycle	Temperature	Time
1	94ºC	1 min
2	94ºC	1 min
3	55ºC	1 min
4	68ºC	2min 30 sec
5	25 times cycle 2	
6	68ºC	11 min
7	4ºC	End

The product of the PCR was checked on an agarose gel (0.8%), to verify the size of the insert DNA (~1200bp). Following the PCR reaction, 10µl of insert DNA was cleaned with the nucleotide removal kit from Qiagen. The insert DNA was then digested using the restriction enzymes 1) Hind III (1µl), incubated at 37ºC for ~1 hour followed by Qiagen clean up and resuspension of DNA into 30µl nanopure water 2) Nde1, incubation at 37 ºC for ~1.5 hours followed by Qiagen clean up and resuspension of DNA into 30µl nanopure water. Three 1µl additions of Nde1 were added at half hour intervals to ensure complete digestion. The pET-24b+ plasmid (vector) was digested in the same way as described above. 1ul of Calf Intestinal Alkaline Phosphatase (CIAP) was added during Nde1 digestion in order to dephosphorylate the 5′-termini and prevent self ligation. The digested vector was taken into 30µl double distilled water and was stored at -30ºC for further use. Different concentrations of insert and vector were tried for the ligation reaction. The

reaction was incubated overnight at 14ºC followed by transformation into $CaCl_2$ competent JM103 cells. Small scale DNA preparation was carried out on single colonies, followed by digestion of DNA with restriction enzymes Hind III and Nde I and verification of insert size by gel electrophoresis.

2.3 Expression strains for *E. coli* CS

For preparing unlabeled protein, pCC plasmids were expressed in the gltA$^-$ *E. coli* host strain MOB154 and grown in LB medium. For growth in minimal media, plasmids pCC / pES gltA were first cloned into pET24b+ vector as described above followed by transformation into *E. coli* host strain BL21 (DE3) plysS (Novagen). ^{15}N-lysine labeled *E. coli* CS was produced in a lysine auxotrophic KL334 (DE3) strain. This strain was generated from *E. coli* strain KL334 [Birge, et al., 1974] and using λDE3 lysogenization kit (Novagen). U-^{15}N and ^{15}N-Lys labeled truncated (first 40 amino acids missing) *E. coli* CS was produced using expression strain in BL21 (DE3) plysS prepared by David Stokell.

2.4 Growth Medium

For preparing U-^{15}N labeled *E. coli* CS, minimal M9 medium was used as a base. Na_2HPO_4 (6g), KH_2PO_4 (3g), NH_4Cl (1g) and NaCl (0.5g) were dissolved per liter [Maniatis, et al., 1982] and autoclaved. Stock solutions of 20% Glucose, 1M $MgSO_4$, 0.1M $CaCl_2$ and 1% thiamine were made. These

solutions were autoclaved separately except thiamine which was filter sterilized.

For preparing selectively ^{15}N-amino acid labeled *E. coli* CS, modified M9 medium was used. Na_2HPO_4 (6g), KH_2PO_4 (3g) and NaCl (0.5 g) were dissolved per liter and autoclaved. Stock solutions of 40% glucose, 1M $MgSO_4$ and 5M NaOH were made separately and autoclaved. Stock solutions of unlabeled amino acids, sodium acetate, succinic acid, nucleotides and other supplements ($CaCl_2$, $ZnSO_4$, $MnSO_4$, L-Tryptophan, Thiamine, Niacinamide , Biotin) were made and filter sterilized. The amounts per liter were as follows: Ala(0.50g), Arg(0.40g), Asp(0.40g), Cys(0.05g), Glu(0.65g), Gln(0.40g), Gly(0.55g), His(0.10g), Ile(0.23g), Leu(0.23g), Lys(0.42g), Met(0.25g), Phe(0.13g), Pro(0.10g), Ser(2.0g), Thr(0.23g), Tyr(0.16g), Val(0.23g), adenine(0.5g), guanosine(0.65g), thymine(0.20g), uracil(0.5g), cytosine(0.2g), $CaCl_2$(2mg), $ZnSO_4$(2mg), $MnSO_4$(50mg), L-Tryptophan (50mg), Thiamine(50mg), Niacinamide (50mg), Biotin(1mg) [LeMaster, et al, 1985].

Just before the inoculation, 1ml of Chloramphenicol (34mg/ ml in EtOH) and 1ml of Kanamycin (25mg/ml) were added to each liter of media.

2.5 Protein Preparation

Uniform ^{15}N labeling

For uniform ^{15}N labeling, a single colony was transferred into 2mL of Luria-Bertani Broth (LB), and grown overnight. This culture was directly inoculated into 100mL complete minimal glucose medium with unlabeled

NH_4Cl. The cells were collected by centrifugation and transferred into 3 x 1L of minimal glucose with unlabeled NH_4Cl. When the culture reached an A_{600} value of ~0.5, the cells were collected by centrifugation and re-suspended into 2 x 1L minimal glucose medium with $^{15}NH_4Cl$ [Lian & Middleton, 2001]. After 1 hour, expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to give 1 mM final concentration. Cells were harvested after 8-10 hours further growth [Choudhary, et al., 2006]. The same procedure was followed for preparing U-^{15}N truncated *E. coli* CS.

Uniform ^{15}N, ^{2}H labeling

For uniform ^{15}N, ^{2}H labeling, a single colony was transferred into 2mL of LB as above. The next day the culture was inoculated into 100mL of LB and grown for about two hours. The cells were collected by centrifugation and re-suspended into 100mL complete minimal glucose medium in water with unlabeled NH_4Cl. When the culture reached an A_{600} value of ~0.5, the cells were collected by centrifugation and re-suspended into 250mL complete minimal media in D_2O with unlabeled NH_4Cl. Again when the culture reached an A_{600} value of ~0.5, the cells were directly inoculated into 2X 1L minimal glucose medium in D_2O without NH_4Cl. After 1 hour, $^{15}NH_4Cl$ was added to the medium and expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to give 1mM final concentration [Gardner & Kay, 1998]. Cells were harvested after overnight growth.

Specific amino acid labeled *E. coli* CS

For single amino acid labeling, a single colony was transferred into 2mL LB as above and grown overnight. The next day, this culture was inoculated into 100mL of minimal glucose medium containing the desired ^{15}N-labeled amino acid amid a pool of 19 unlabeled amino acids and other components [LeMaster, et al.,1985], but not containing NH_4Cl [Ramesh, et al., 1994]. The cells were collected by centrifugation and resuspended in 2 x 1 L minimal glucose medium lacking NH_4Cl and the labeled amino acid, but containing the 19 unlabeled amino acids and other components as required. When A_{600} ~ 0.3, the ^{15}N-labeled amino acid was then added at 50% of the normal requirement, a rate-limiting concentration [McIntosh, et al., 1990] and then expression was induced at A_{600} ~ 0.5 by addition of IPTG to 1 mM as above. The culture was harvested after 8-10 hours further growth at 37°C. The pool of 19 unlabeled amino acids ensures growth and should repress transaminase activities in the cell [McIntosh, et al., 1990].

2.5 Purification of *E. coli* CS

The purification procedure was adapted from that reported by [Duckworth, et al., 1982]. Briefly, cells were harvested after growth for 8-10 hours by centrifugation and disrupted by an Amicon French Press in standard CS buffer (20mM Tris, 1mM EDTA, pH7.8). After centrifugation to remove cellular debris, the crude extract was passed through a diethyl amino ethyl-cellulose (DEAE) column (8 X 5 cm) which had been pre-equilibrated

with the standard buffer. The column was washed with at least 1 liter of standard buffer to remove unbound material. A linear KCl gradient (50 to 300mM, 1L) was used to elute the bound protein. CS activity was usually found in a prominent peak that eluted at about 100mM KCl (Figure 2.1).

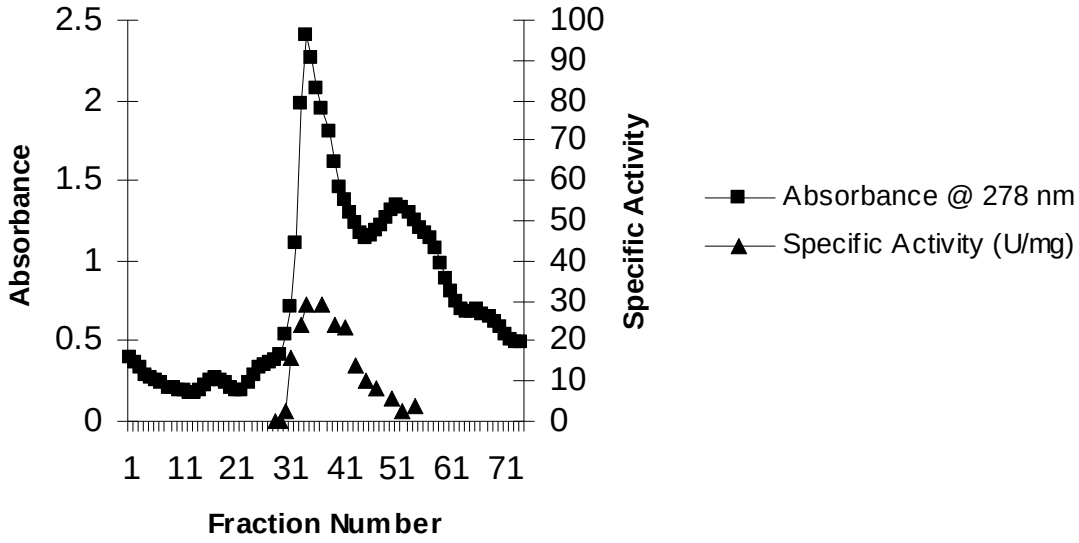


Figure 2.1: Elution Profile of *E. coli* CS off DE52 Column. Fractions are pooled based on activity; for example fractions 32-44 were pooled in this preparation.

Fractions were assayed for Citrate Synthase using the method of [Srere, et. al. 1963] and those with the highest activity were pooled and concentrated by ultrafiltration. A final gel filtration step (Figure 2.2) using a Sepharose 6B (Separation range 10^4 – $4X10^6$ Da) or Sephadex G200 (separation range $5 X 10^3$ -$6 X 10^5$ Da) column (110 X 4 cm) gave a CS preparation of 50 to 100mg. Purity of the samples was checked by matrix-assisted laser desorption ionization mass spectrometry.

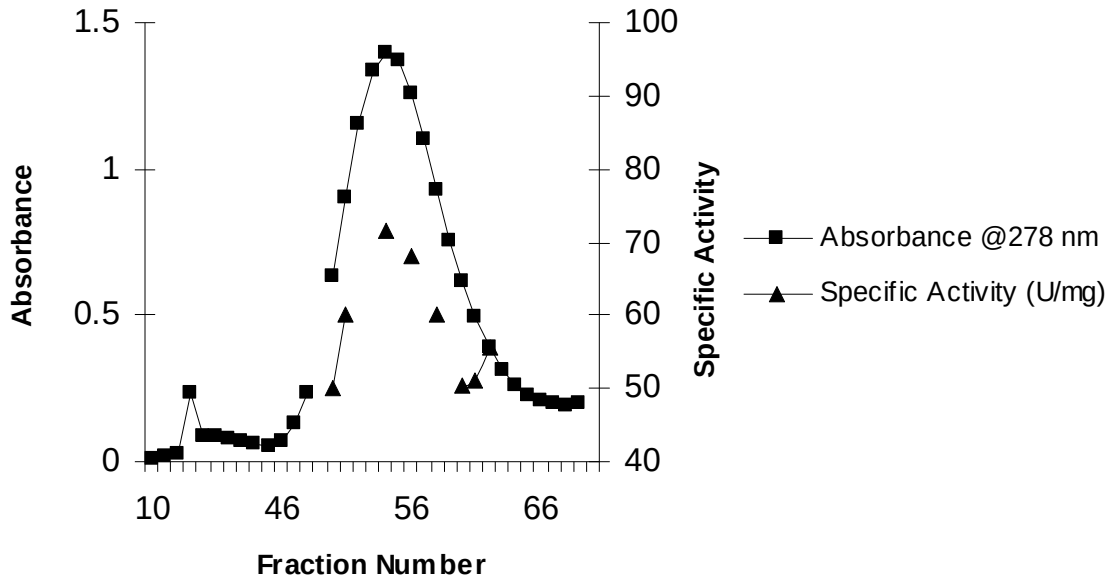
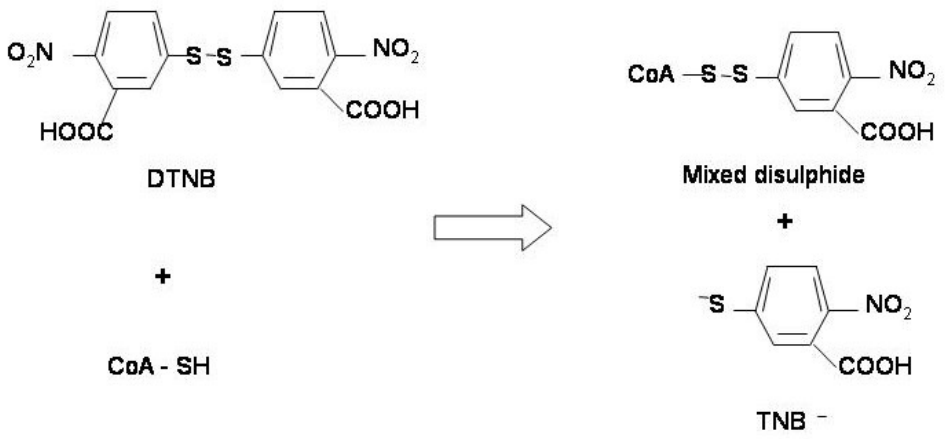


Figure 2.2: Elution profile of *E. coli* CS from a G200 column. Again fractions are pooled based on activity. For example, fractions 51-58 were pooled in this preparation.

2.6 CS activity assays and kinetic measurements

CS catalyzes the condensation of OAA and AcCoA to produce citrate and CoA. Measurement of the appearance of the free SH of the CoA product is the basis for the CS assay [Srere, et. al., 1963]. This reacts with DTNB, producing 1 mole of the thionitrophenolate (TNB^-) anion per mole of CoA produced:



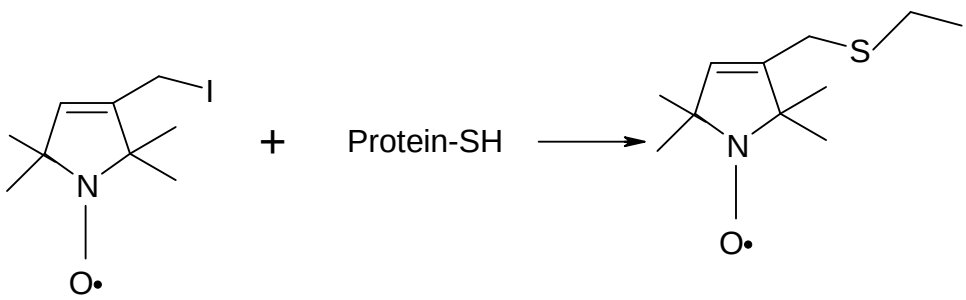
TNB is the coloured species in the reaction which absorbs strongly at 412 nm with an extinction coefficient of 13600 M^{-1} cm^{-1}, and its rate of production was measured spectrophotometrically to obtain CS reaction rates using a Milton Roy Spectrophotometer. One unit of CS activity is defined as the amount of the enzyme required to produce 1 µmole CoA per minute at room temperature. The standard CS assay [Duckworth & Bell, 1982] contains

100µM AcCoA, 100µM OAA and 50µM DTNB in 100mM KCl, 20mM Tris, 1mM EDTA, pH 7.8, and the assay volume was typically 1ml.

Limited kinetic measurements were carried out on some of the mutants in the presence of KCl. Typically two substrate saturation curves were obtained (one for OAA and another for AcCoA) in the presence of KCl.

2.7 Spin labeling

The reagent 3-Iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline) for spin labeling of reactive cysteine was purchased from Sigma. The reagent was dissolved in methanol and stored at -30ºC. Modification of *E. coli* CS was carried out by adding 20-50 fold excess of the reagent. The reaction was allowed to proceed for 12 hours at room temperature. Excess reagent was removed by dilution and concentrated using an Amicon filtration unit. The incorporation of the label was determined using mass spectrometry, by the analysis of tryptic fragments containing cysteine residues.



3-Iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline)

modified side- chain

3.0 NMR STUDIES OF FLEXIBLE REGIONS OF *E. COLI* CITRATE SYNTHASE

3.1 Introdution to NMR Spectroscopy of Biomolecules

NMR spectroscopy is now well established as a powerful technique for the determination of the three-dimensional structures of proteins in solution. Studies by R. R. Ernst (described in Ernst, et al., 1987) and Wüthrich, 1986 demonstrated that it is possible to obtain high resolution structures of small proteins, with molecular masses less than approximately 10kDa (100 amino acids). The work of Ernst provided the framework for the extension of the NMR technique from one to two and three frequency dimensions. Subsequent developments in NMR spectroscopy have had a further significant impact on solution studies of proteins (Bax, 1994; Kay, 1995). The improvements in the technology have been several fold and include (i) the increase in the dimensionality of experiments from two to three and four, providing improved resolution for the complex spectra; (ii) the uniform incorporation of ^{15}N, ^{13}C and ^{2}H labels into the biological system of interest coupled with the development of sophisticated NMR pulse schemes to transfer magnetization between scalar (through bond) and dipolar (through space) coupled spins; and (iii) significantly improved radiofrequency (RF) electronics and increased magnetic field strengths as well as the development of commercially available hardware such as pulsed field gradients.

3.2 Chemical Shift assignment

The first step in any NMR study of molecular structure or dynamics is to assign the frequencies at which energy absorption occurs for each of the specific NMR active nuclei in the sample. The frequency of absorption is directly related to a parameter termed the chemical shift. The Chemical shift of a particular nucleus can be thought of as its "nuclear signature" and once such information is available the spectroscopist has a handle on probing structure and dynamics at that site in the molecule. In the standard protocol for protein structure determination by NMR spectroscopy, sequence-specific resonance assignment plays a pivotal role (Wüthrich, 1986). Based on the known amino acid sequence, the sequence specific assignment of a protein is established by observing cross-peaks between sequentially neighbouring amino acids in multidimensional NMR spectra. Several different assignment strategies are available, and the standard approach to obtain complete assignments (Wüthrich et al., 1982) for proteins up to about 30kDa in size involves uniform $^{13}C/^{15}N$-labeling and using triple resonance experiments (Montelione, et al., 1989; Montelione, et al., 1990; Kay, et al., 1990; Ikura, et al., 1990)

3.3 Problems with Large systems

As the molecular weight increases, the number of cross peaks in NMR spectra also increases leading to overlapping of resonances. The second,

more serious issue with large macromolecules is the rapid decay of the NMR signal that occurs during the several transfer steps in a complex NMR experiment [Fernández, et al., 2003]. A typical multidimensional NMR experiment can be schematized as follows:

A-------B------C-------D--------E----------Z

where the transfer of magnetization proceeds from A to Z via B,C,D etc. The amount of time required to transfer magnetization along each link in the chain, for example from B to C, is a function of the strength of the coupling between the participating links. During this transfer, the signal decays via relaxation processes which for macromolecules, increase linearly with molecular weight. The decay time of the signal varies depending on the type of nucleus i.e. whether A = ^{1}H, ^{13}C, ^{15}N. If the decay rate is the same order of magnitude as the transfer rate, a significant decrease of the signal is observed in the spectrum [Kay, 1997].

The relaxation of nuclei is caused by fluctuating magnetic fields that are the result of the overall molecular tumbling in solution as well as internal dynamics [Kay, 1997]. T_1 relaxation determines the length of the experiment whereas T_2 relaxation determines the width of the lines in the spectrum. The width of a line in the spectrum is inversely proportional to the T_2, which depends on the size of the molecule; the larger the molecular mass the shorter T_2 becomes and the broader are the lines in the spectrum [Venters, et al.,

2002]. Faster transverse relaxation of the nuclei during the delays and during acquisition limits the use of triple-resonance experiments with large proteins.

3.4 Different approaches to studying large systems

Specific amino acid labeling

An important step towards solving the three-dimensional structure of a protein by NMR is the sequential assignment of spectra. Uniform labeling of the protein offers tremendous benefits to the assignment process, but in the case of large proteins selective labeling of amino acids can simplify a complex spectrum considerably. In order to $^{15}N/^{13}C$ label amino acids specifically, a minimal growth medium is supplemented with a mixture of unlabeled and labeled amino acids [LeMaster, et al., 1985]. In order to ^{13}C label amino acids at specific sites, a minimal growth medium is supplemented with suitably labeled metabolites acting as the precursors and sole ^{13}C source for the appropriate amino acids [Lian & Middleton, 2001].

Cross-labeling and isotopic dilutions as a result of transamination are the main concerns where selective labeling is desired. Transaminases catalyse the reversible transfer of the amino groups between α-amino acids and α-ketoacids. Essentially all of the amino acids derive their α-amino groups directly or indirectly from transaminase-catalyzed reactions involving glutamate and glutamine as the nitrogen donor making selective labeling of these two amino acids very difficult. Amino acid biosynthesis is regulated by feed-back inhibition and hence supplementing the growth with high

concentration of unlabeled amino acids can repress and inhibit the biosynthetic pathway thereby decreasing these concerns [McIntosh, et al., 1990]. Another approach is to use auxotrophic host strains that have been modified to contain the appropriate genetic lesions to control the biosynthesis of a particular amino acid. However, the yield of a soluble protein can vary widely between different host strains and a host strain may have to be converted to the appropriate genotypes for selective labeling. When auxotrophic strains are not used, incorporation of labeled amino acids like Leu, Val, or Met which terminate a metabolic pathway, is a more efficient strategy for avoiding cross-labeling and/ or isotopic dilution than is the case for residues such as Ser, Asp and Glu [McIntosh, et al., 1990; Lian & Middleton, 2001]. The protocol used for labeling a protein selectively with natural amino acids can also be used for incorporation of non-natural amino acids such as fluorinated amino acids [Son, et al., 2006; Vaughan, et al., 1999].

One strategy for specific assignments in large proteins of molecular mass greater than 100kDa is based on combining selective amino acid labeling with site-directed mutagenesis. The comparison of cross-peaks in selective labeled WT protein and the absence of a particular peak in a selective labeled mutant spectrum enable a specific residue to be identified. This approach has been used in the assignment of the isoleucine residues in ClpP tetradecameric complex of molecular weight 300kDa [Sprangers, et al., 2005]. In another example, a tyrosine residue of a membrane protein KcsA of

molecular weight 60kDa (embedded in 60kDa detergent micelles) was assigned by comparing the ^{15}N-Tyr labeled KcsA TROSY spectrum with ^{15}N-Tyr labeled Y78F KcsA TROSY spectrum [Tzakos, et al., 2006].

Another approach used in resonance assignment of large proteins uses a ^{13}C-^{15}N double-labeling strategy. The strategy combines ^{13}C labeling of an amino acid at its carbonyl carbon position with ^{15}N labeling of the next amino acid in the sequence at the α-position [Klein-Seetharaman, et al., 2002; Tzakos, et al., 2006]. Depending upon the number of such pairs, the cross-peaks in the ^{15}N-dimension in HSQC/ TROSY spectra will display splitting due to $^{1}J_{C'N}$ coupling [Weigelt, et al., 2002], and only one peak will appear in an HNCO [Grzesiek, et al., 1992] type correlation spectrum.

Spin labeling

Spin labels are stable free-radical molecules containing one or more unpaired electrons and a reactive group able to covalently modify protein side chains. The common labels contain a single electron in the paramagnetic nitroxide moiety. The attachment of a nitroxide spin label at a defined site in the protein induces nuclear relaxation within a radius of up to 20Å from the electron spin, which is manifested in the NMR spectrum as a selective broadening or removal of resonance lines [Cutting, et al., 2004].

The sites for spin labeling in a protein are dependent on the availability and number of reactive side groups capable of being modified by the labeling reagent. Typically, spin labels contain isothiocyanate, maleimide

or methanethiosulphonate functionalities that are reactive toward cysteine or lysine side groups [Kosen, 1989]. Ideally, native cysteines are removed by targeted mutagenesis and a unique cysteine is engineered into a protein at a specific residue to provide a single reactive site for the spin label in a defined region [Battiste, et al., 2000]. Where a single Cys variant is not available, it may be possible to spin label the wild-type protein selectively by relying on differences in the accessibility and reactivity of the Cys side groups, although quantification of labeling is not straightforward. Spin labels are rather bulky molecules and it is important to consider their effect on local or global conformation of the protein.

Spin labeling has been used for NMR spectral simplification and assignment by removing overlapping cross-peaks from protons close to the labeled site. The advantages of spin labeling in NMR spectroscopy are best illustrated by some recent applications [Roosild, et al., 2005, Cutting, et al., 2004].

Spectroscopic Developments

Heteronuclear Single Quantum Correlation (HSQC)

This experiment correlates heteronuclei such as ^{15}N or ^{13}C with ^{1}H in proteins. The pulse sequence for a basic HSQC experiment [Bodenhausen & Ruben, 1980] is shown in Figure 3.0. The experiment employs two insensitive nuclei enhanced by polarization transfer (INEPT) [Morris & Freeman, 1979]

magnetization transfers. The first one creates antiphase heteronuclear coherence and the second is used to convert this coherence back to observable magnetization.

Briefly, the first INEPT step creates a proton antiphase magnetization (2IxSz) during τ. The chemical shifts during this period are refocused by the 180º pulses on both spins. Coherence is subsequently transferred to the directly attached heteronucleus (^{15}N or ^{13}C) by the simultaneous 90º pulses on both spins (2IzSy). The S-spin (^{15}N or ^{13}C) coherence is frequency-labeled during the t_1 period. The 180º pulse on the I-spin (^{1}H) in the middle of t_1 refocuses the evolution of J_{IS} coupling. A 90º pulse on both spins transfer the magnetization back to proton as antiphase I-spin magnetization (2IySz). The final spin-echo period converts this antiphase term into in-phase proton magnetization (2IySz→ Ix).

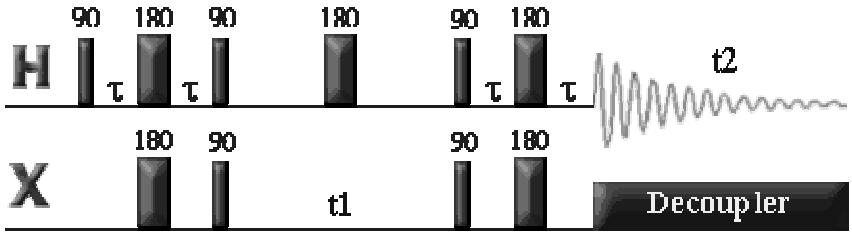


Figure 3.0: The basic elements for the 2D ^{15}N-^{1}H HSQC experiment. The delay τ is set to 1/ ($2J_{IS}$). X is either ^{15}N or ^{13}C. t_1 and t_2 are the evolution and detection periods respectively.

Transverse Relaxation Optimized Spectroscopy (TROSY)

Transverse relaxation has a large impact on the quality of NMR spectra. With increasing molecular weight, transverse relaxation becomes faster. Consequently, the resonance lines become broader and the sensitivity decreases. The TROSY [Pervushin, et al., 1997] technique which was introduced a few years ago has found widespread applications [Fernandez, et al., 2003], since better spectra are readily obtained when working with molecular sizes above 15-20kDa. The TROSY technique can reduce the effective relaxation of the measured signal during the pulse sequence and data acquisition resulting in increased spectral resolution and improved effective sensitivity.

Briefly, if we consider the ^{15}N-^{1}H amide groups of proteins, both nuclei have a spin-1/ 2. These two nuclei can classically be described by a magnetic dipole having two orientations: up and down. In the ^{1}H-NMR spectrum a doublet is observed originating from the protons attached to ^{15}N nuclei with spin up and spin down respectively. These doublets are routinely collapsed into a single, centrally located line by decoupling with the expectation of obtaining a simplified spectrum and improved sensitivity. However, in the spectrum of a large protein the individual multiplet components have different line widths [Pervushin, et al., 1997]. Decoupling averages the different relaxation rates that deteriorate the averaged signal for large molecules studied at high magnetic fields. In TROSY, no decoupling is

applied, and only the narrowest, most slowly relaxing line in each multiplet is selected. TROSY disregards half of the potential signal of an amide moiety. This loss is compensated in large molecules by the slower relaxation of the selected resonance line.

The two processes causing significant relaxation in the amide moiety are the dipole-dipole (DD) and chemical shift anisotropy (CSA) interactions. The DD and CSA interactions are present simultaneously and their magnetic fields add or subtract depending on the direction of the attached nucleus. These interactions can add, leading to larger, or subtract, to give smaller z-components. Where the interference between the two relaxation rates subtracts it can reduce the rate of relaxation of a transition. Consequently, the relaxation rates of the two transitions are potentially different. The interference between the two mechanisms is possible because of a similar dependence on the rotational motion of the molecules [Goldman, 1984; Pervushin, et al., 1997]. The field created by the nuclear dipole depends on $(3\cos^2\theta-1)$, where θ is the angle between the main magnetic field and the direction of the H-N bond in an amide moiety, all other factors being constant. For the magnetic field caused by CSA, the same angular dependence is obtained when an axially symmetric CSA tensor with its main axis along the H-N bond is assumed [Pervushin, et al., 1997]. With increasing strength of the main magnetic field, CSA relaxation increases whereas DD relaxation stays constant. The field dependent interference between the two

processes leads to an optimum field for detection of NMR signals for large proteins. The optimal TROSY effect can thus be obtained by choosing the appropriate field strength. For the amide moiety the TROSY effect for both the ^{1}H and the ^{15}N nuclei has an optimum at about the same magnetic field strength corresponding to a proton frequency of approximately 1000 MHz.

In nearly all multidimensional NMR experiments for studies of biological macromolecules in solution, magnetization is transferred from one type of nucleus to another via scalar spin-spin couplings using INEPT (insensitive nuclei enhanced by polarization transfer) pulse sequences [Morris & Freeman, 1979]. With increasing molecular weight, transverse relaxation during INEPT transfer periods becomes a limiting factor that severely compromises sensitivity. On the other hand performance of cross-correlated relaxation induced polarization transfer (CRIPT) improves for larger molecules [Bruschweiler & Ernst, 1992]. CRINEPT combines CRIPT and INEPT [Riek, et al., 1999]. CRIPT and CRINEPT make use of the same cross-correlated relaxation, which is the basis for TROSY. Thus for amide groups the optimal magnetic field strength for these polarization transfer mechanisms corresponds to 1GHz proton resonance frequency.

Advantages of Deuteration

In solution-state NMR spectroscopy, protons are replaced by deuterons for two reasons—to simplify the spectrum and to reduce or eliminate relaxation pathways. Because of the lower gyromagnetic ratio of ^{2}H,

compared to ^{1}H ($\gamma[^{2}H]/\gamma[^{1}H]= 0.15$), replacement of protons with deuterons reduces proton line widths by removing contributions from dipolar ^{1}H-^{1}H relaxation and ^{1}H-^{1}H scalar couplings [Goto & Kay, 2000]. The line-width narrowing effect is one of the prime reasons why NMR can now be used to study large proteins with sizes of 100kDa and above [Fiaux, et al, 2004].

At the high field strengths used for macromolecular studies, long macromolecular correlation times result in very efficient transverse relaxation of the active spins. The effectiveness of deuteration in lengthening T2 relaxation times is most significant for the $13C^{\alpha}$ [Yamazaki, et al., 1994] and $^{1}H^{N}$ nuclei [Markus, et al., 1994], and the least for ^{15}N (due to the presence of directly attached exchangeable protons) and ^{13}CO (due to the predominant effect of the carbonyl chemical shift anisotropy). The level of deuteration required is dependent on spectral quality, the size of the system studied and the type of NMR experiments performed [Gardner & Kay, 1998]. Different protocols to achieve deuteration have been reported [Gardner & Kay, 1998; Leiting, et al., 1998; Lian & Middleton, 2001; Fiaux, et al., 2004].

3.5 NMR Spectroscopy of large molecules

The developments in solution NMR spectroscopy have led to the study of large molecules which were previously considered inaccessible to the technique. Complete sequential assignment of 723-residue monomeric *E. coli* malate synthase G (MSG) of MWT 81.4kDa and 121 residues in 7, 8-dihydroneopterin aldose (DHNA) from *Staphylococcus aureus* of MWT 110kDa

have been obtained using TROSY based triple and four dimensional NMR experiments [Tugarinov, et al., 2002, Salzmann, et al., 2000]. TROSY and CRIPT/CRINEPT-TROSY based ^{15}N-^{1}H correlation experiments have been used to study protein-protein interactions of large molecules such as the GroES-GroEL complex of MWT ~900kDa. The chaperonin complex consists of a bigger unit, GroEL, of molecular mass 800kDa, which interacts with the smaller unit, GroES, of molecular mass 72kDa, in the presence of ATP. In order to study the complex, backbone assignments for GroES were obtained using TROSY based triple resonance experiments. Residues 17-32 in GroES belong to a mobile loop, that forms contact with GroEL and in doing so undergoes a large conformational change, as predicted by the crystallographic results. The mobile loop residues of GroES showed random coil chemical shifts indicating disordered conformation. But after binding to the GroEL, large chemical shift dispersion of the mobile loop residues was observed indicative of the ordered conformation of the mobile loop. In addition, the mobile loop residues showed variable dynamic properties in the complex. In the complex some of the mobile loop residues were immobilized whereas others retained a high degree of mobility. The study also showed that it was possible to observe only the flexible regions in the complex of MWT 900kDa when using TROSY whereas both flexible parts and structured parts could be observed by using CRIPT/ CRINEPT based experiments [Fiaux, et al., 2002].

After the studies of GroES-GroEL complex attempts were made towards the NMR analysis of GroEL of molecular mass 800kDa [Riek, et al., 2002]. TROSY experiments of U- ^{15}N, ^{2}H labeled GroEL showed about 20 resonances which were tentatively assigned to the flexible residues of the C-terminal. CRINEPT/ CRIPT-TROSY experiments gave an improvement over the TROSY spectrum because of relaxation optimization throughout the pulse sequence which increases the chances of observing residues from the structured parts of large proteins. But even with the use of CRIPT-TROSY in combination of high deuteration (at least 85%) of the protein, only 20% of the expected ^{15}N-^{1}H correlation peaks were observed. Probable reasons for missing residues were attributed to incomplete H-D back exchange and very broad line widths of the residues [Riek, et al., 2002]. Further studies were carried out by preparing specific amino acid labeled GroEL with high deuteration background. The labeling scheme was carried out so that almost all the backbone amides were protonated. Under such conditions also 36% of the expected ^{15}N-^{1}H correlation peaks of a particular amino acid were observed. The results showed that in the case of large proteins, the reason for missing resonances can not only be the spectral overlap or incomplete $^{2}H^{N}$/ $^{1}H^{N}$ exchange but other factors like conformational averaging which can broaden line widths beyond detection [Fiaux, et al., 2004].

In the case of membrane proteins, the effective size of the protein to be studied by NMR is increased because the protein has to be studied as micelles

in detergent or in a detergent-phospholipid mixture. NMR spectroscopy has been done on ^{15}N-Lys and $^{15}N_2$-Trp labeled Rhodopsin of molecular mass in the range of 100-120kDa [Klein-Seetharaman, et al., 2004; Klein-Seetharaman, et al, 2002]. Rhodopsin contains 11 lysine residues but the HSQC spectra of ^{15}N-Lys labeled Rhodopsin shows one strong peak at 4ºC corresponding to the highly mobile lysine near the C terminus. Further, a total of more than the expected 11 signals with weak intensity were observed at temperatures higher than 20ºC. These weak signals were attributed to the microsecond to millisecond conformational exchange processes [Klein-Seetharaman, et al, 2002]. In the HSQC spectrum of ^{15}N-Trp labeled Rhodopsin, 10 signals corresponding to backbone nitrogen and 5 signals corresponding to the indole nitrogen were observed. More than the expected signals of the backbone nitrogen (total 5 tryptophans in the protein) were attributed to the microsecond to millisecond conformational exchange processes [Klein-Seetharaman, et al, 2004].

In addition to the backbone protein spins, side chain methyl groups also offer a useful alternative spectroscopic probe in large systems. Methyl groups give rise to intense signals as a result of rapid rotation of the three protons about the methyl symmetry axis and also due to the favourable relaxation properties. These properties of methyl moieties have been used in the complete assignments of Ile, Leu, Val in *E. coli* malate synthase G of MWT 81.4kDa [Tugarinov, et al., 2003, Kay, 2005]. A similar approach was used in

the analysis of the methyl groups of heteroheptamer Arp2/ 3 actin nucleation complex of MWT 280kDa. The ^{1}H-^{13}C-HSQC spectrum of one of the subunits of Arp2/ 3 complex showed resonances from about half of the Leu, Val, and Ile methyl groups. The quality of the spectrum obtained for the complex was quite low which is attributed partly to micromolar concentrations of the subunits in the complex and partly to involvement of the subunits in chemical exchange dynamics [Kreishman-Deitrick, et al., 2003].

3.6 Protein Dynamics from NMR

It is well recognized that internal protein dynamics play an important role in the function of proteins [Boehr, et al., 2006; Kay, 2005; Schnell, et al., 2004]. Conformational dynamics on the micro-to millisecond time scale (μs-ms) are biologically very important, because they are close to the time scales on which protein folding, allosteric transitions and product release take place and therefore these are associated with functional processes [Vendruscolo & Dobson, 2006; Kalodimos, et al., 2004]. Important developments have been made towards method development to extract information about the dynamic processes, which can provide insight into the relation between motion and enzymatic function [Kay, 1998, Mettermaier & Kay, 2006; Ishima & Torchia, 2000].

A key property that is exploited in the functioning of the protein is molecular flexibility. A study done by Freedberg, et al., 1998, used this property to understand the details of a protease-inhibitor interaction. HIV-1

protease is a homodimeric protein essential for the function of the AIDS virus. X-ray studies show that flaps of the inhibited enzyme (residue 46-56 and 146-156) are closed over the inhibitor and are linked to the inhibitor by a network of hydrogen bonds involving bridging water molecules. Studies indicated that the flexibility of the protease flaps permits access to the active site and is therefore important for the function. Freedberg, et al., 1998, used NMR to show that all the residues that experience motion on the chemical exchange time scale are concentrated in either the tips of the flaps or are near the terminal loop. The study concluded that the tip of the flaps form a dynamic structure that stabilizes the complex with H-bonds and permits the flexibility required facilitating product release following catalysis.

Fructose-1, 6-bisphospahatase (FBPase) which catalyzes the conversion of fructose 1, 6-biphosphate to fructose 6-phosphate and phosphate provides another example of an enzyme which employs flexible loops in its biochemical function. A combined X-ray and fluorescence study of this protein showed that the structural loop responsible for allosteric regulation of catalysis is disordered in the T-state and becomes ordered in the R-state [Nelson, et al., 2000]. Similar examples which show the importance of flexibility on biologically relevant time scales in performing the biological functions are known [Kalodimos, et al., 2004; Ogata, et al., 1996; McElheny, et al., 2005].

Conformational dynamics also plays an important role in the regulation of allosteric proteins since allosteric processes and dynamics are closely linked with each other. Allosteric processes are closely associated with ligand induced conformational changes that propogate between the allosterically coupled binding sites. The allosteric molecules thus change its coordinates as a function of time, which constitutes dynamics [Kern & Zuiderweg, 2003]. There is also an increasing realization that proteins should be treated as a dynamic ensemble of conformational states. Recent observations show that conformational events that occur during protein function are already present before ligands bind. In contrast to being folded in a single native conformation, proteins have evolved to sample multiple defined conformations that are critical for function [Eisenmesser, et al., 2005; Kern & Zuiderberg, 2003].

3.7 Experimental Procedures

NMR Samples

NMR samples were prepared in phosphate buffer (20mM phosphate, 1mM EDTA, 0.04%NaN_3, pH 7.0) by overnight dialysis. Labeled Citrate Synthase was then concentrated to 0.6 – 2.0mM using Centricon concentrating units (Millipore, MWT cutoff 50 x 10^3) to a final volume of 500 – 700µl. The final NMR sample was then prepared by adding D_2O to 10% for spin locking and 1µl of 2, 2-dimethyl-2-silapentane-5-sulfonate (DSS) (stock of 36.0mM) for reference in the ^1H dimension. Stock solutions of substrates (200mM

Oxaloacetic acid in phosphate buffer acid, pH 7.0 was adjusted using stock solution of 10M NaOH, 20mM CM-CoA in water) were made and added in 1:1.1 (protein :substrate) molar ratio in the presence of 0.1M KCl. Stock solutions of OAA were made just before addition to the NMR sample.

NMR Spectroscopy

NMR Spectroscopy was carried out on AMX-500 Bruker and Varian INOVA-600 Spectrometers, operating at 500 and 600 MHz, respectively. A few experiments were conducted on the INOVA – 800 Varian Spectrometer at NANUC, University of Alberta. For most experiments, the Varian INOVA-600 MHz spectrometer was used which was equipped with a triple-resonance probe head with an actively shielded z-gradient coil and three radio frequency (rf) channels for generating ^{1}H, ^{13}C and ^{15}N rf pulses plus deuterium lock. One dimensional proton and two-dimensional [^{15}N, ^{1}H]-Heteronuclear Nuclear Quantum Coherence (HSQC) [Kay et al., 1992] and / or Transverse Relaxation Optimized Spectroscopy (TROSY) [Pervushin, et al., 1997] experiments were carried out on *E. coli* citrate synthase samples labeled uniformly with ^{15}N, or specifically with one ^{15}N-labeled amino acid, or specifically with one ^{15}N- and one ^{13}C-labelled amino acid. Cross-Relaxation Induced Polarization Transfer (CRINEPT) Spectroscopy [Riek, et al., 1999] was carried out on the U-$^{15}N/^{2}H$ labeled *E. coli* CS. The experiments were carried out at 25ºC except where temperature dependence studies were carried out.

For the temperature dependence study a series of ^{1}H-^{15}N HSQC spectra were taken on ^{15}N-Met and ^{15}N-Leu labeled *E. coli* CS in the temperature range of 5-40 ºC usually in 5 ºC increments. After each temperature change, the sample was allowed to equilibrate for 30 min. The probe head was tuned and shimmed at each temperature. Each spectrum was referenced to (DSS) by collecting a ^{1}H one-dimensional spectrum using the same ^{1}H spectral width as for the 2D-[^{1}H, ^{15}N] HSQC spectrum.

NMR data were processed and displayed using Spin-Works (K. Marat). Both F_2 (^{1}H) and F_1 (^{15}N) dimensions were multiplied by a 90º-shifted squared sine-bell function, and zero-filled before Fourier transformation. The dimension of the resulting processed data sets was 4096X512 for the ^{1}H-^{15}N HSQC/ TROSY experiments. Line widths were measured on the (^{1}H or ^{15}N) slices from the spectra on which no apodization function was applied. The [^{1}H,^{15}N] HSQC/ TROSY spectra were recorded usually with a data size of 1024 X 64 complex points in the ^{1}H and ^{15}N dimensions respectively; the number of scans accumulated per (t_1,t_2) was 512 and a recycle delay of 0.7s was used to give a total measuring time of 15 hours/ spectrum on the 600MHz Spectrometer. The number of scans accumulated varied from 256 to 800 for some experiments depending on the S/ N ratio in the first increment of the two dimensional experiment. For the specific $^{15}N/^{13}C$ labeled *E. coli* CS the data size collected was 1024 X 128 complex points to get better resolution in the nitrogen dimension. Chemical shifts are reported from DSS for the

proton dimension, using 4.76 ppm from DSS as the chemical shift of water at 25°C. ^{15}N referencing was done indirectly relative to DSS as recommended [Wishart, et al., 1995].

Determination of thermodynamic parameters

To extract thermodynamic parameters for the conformational exchange based on the temperature dependence data, the natural logarithm of the ratio the of normalized intensities (ln[intensity]) versus the reciprocal of the absolute temperature (K) (1/T) was plotted and fitted to the Arrhenius equation (Equation 3.1);

$$\ln \textit{intensity} = \ln Z - E_a / RT \qquad \text{Equation 3.1}$$

where E_a is the energy of activation, R is the gas constant, and Z is the preexponential factor. ΔH and ΔS were calculated as described [Fan, et al., 2000].

According to Eyring's transition-state theory, the temperature dependence of a rate is given by equation 3.2

$$k_{ex} = \frac{k_B T}{h} \exp\left(\frac{-\Delta G}{RT}\right) = \frac{k_B T}{h} \exp\left(\frac{-\Delta H}{RT}\right) \exp\left(\frac{\Delta S}{R}\right) \qquad \text{Equation 3. 2}$$

where k_B is the Boltzmann's constant, h is Planck's constant, R is the gas constant, and ΔG, ΔH and ΔS are the free energy, enthalpy and entropy of the exchange.

From equations 3.1 and 3.2: $\Delta S = R\left(\ln Z - \ln \frac{k_B}{h}\right)$ and $\Delta H = E_a - RT$

Thus ΔH and ΔS can be determined from the Arrhenius plots.

3.8 Results

3.81 Uniform ^{15}N-labeled *E. coli* CS

To establish NMR conditions, we initially acquired spectra of the 280kDa U-^{15}N labeled *E. coli* CS in the 500MHz spectrometer. The HSQC spectrum of *E. coli* CS revealed 29 well resolved signals plus a region of many overlapping cross-peaks with chemical shifts typical of helical residues, presumably corresponding to the flexible parts of the protein (Figure 3.1). The overlapping peaks made it impossible to gain information about the residues giving rise to these signals. The use of TROSY (Figure 3.2) [Pervushin, et al., 1997] did not improve the spectrum significantly.

Substrates were added to see if the regions giving rise to the cross peaks in the HSQC spectrum change their environment. Addition of substrates was done in a sequential manner with the addition of 0.1M KCl first, followed by a 1.1 molar excess of each of oxaloacetate (OAA) and the substrate analogue carboxymethyl CoA (CM-CoA). Two-dimensional HSQC spectra were recorded after addition of KCl and in the presence of both substrates (Figure 3.3 & 3.4). There was no change in the chemical shifts of the resolved peaks and any change which might have occurred in the overlapped region between 7.9 ppm and 8.6 ppm was not observable to us. Although there was no change in the chemical shifts, broadening in the line widths of some resolved peaks was observed after addition of 0.1M KCl (Table 3.1).

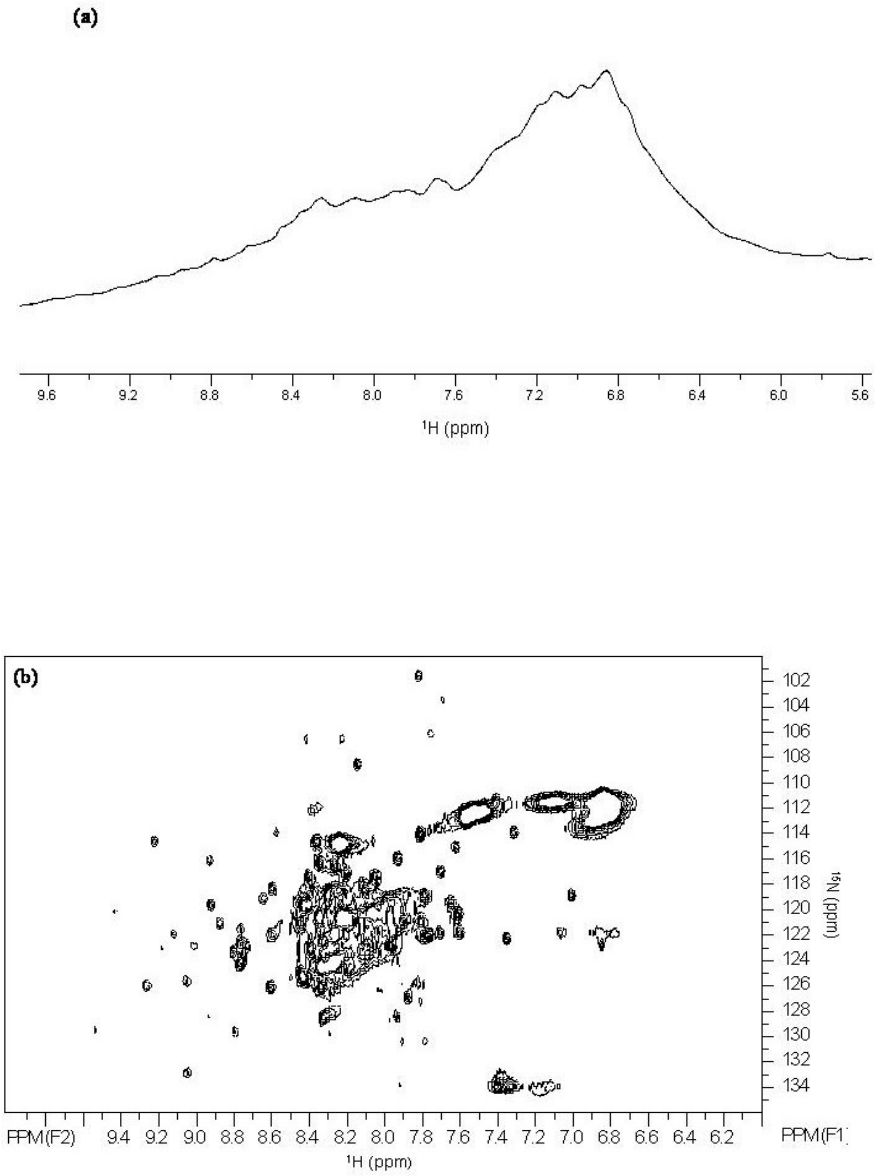


Figure 3.1: (a) One-Dimensional proton spectrum, showing the amide region of 1.2mM U-^{15}N *E. coli* CS. (b) Two-Dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.2mM U-^{15}N *E. coli* CS measured at 25ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

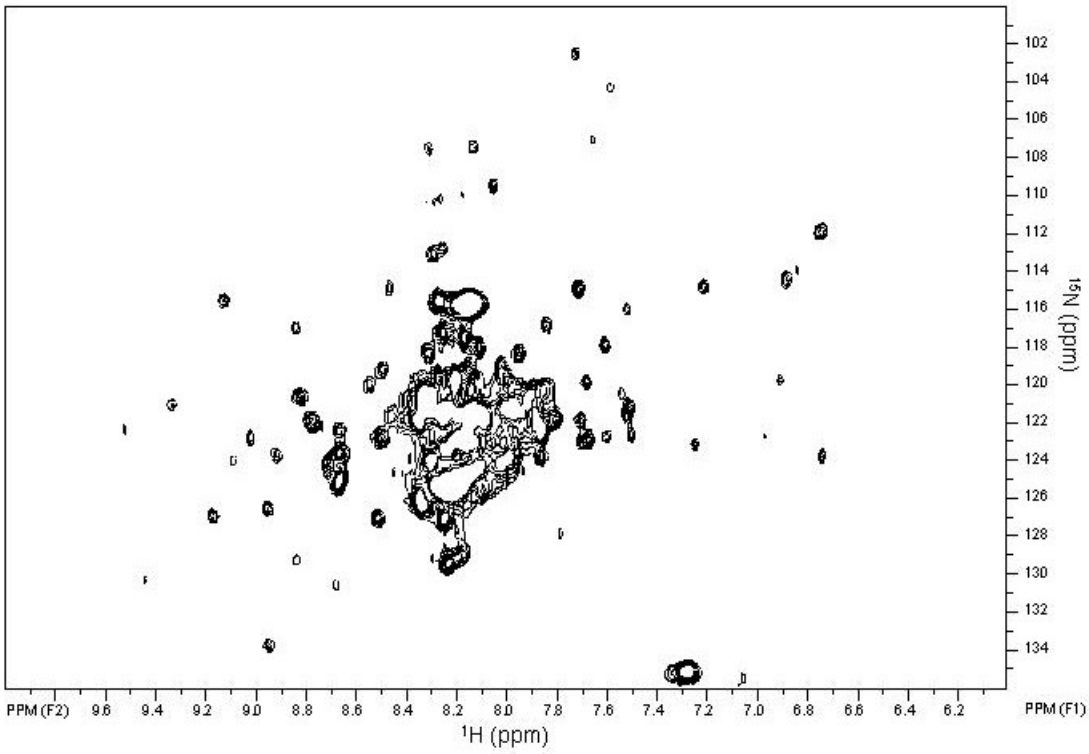


Figure 3.2: Two-Dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.2mM U-^{15}N *E. coli* CS measured at 25ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

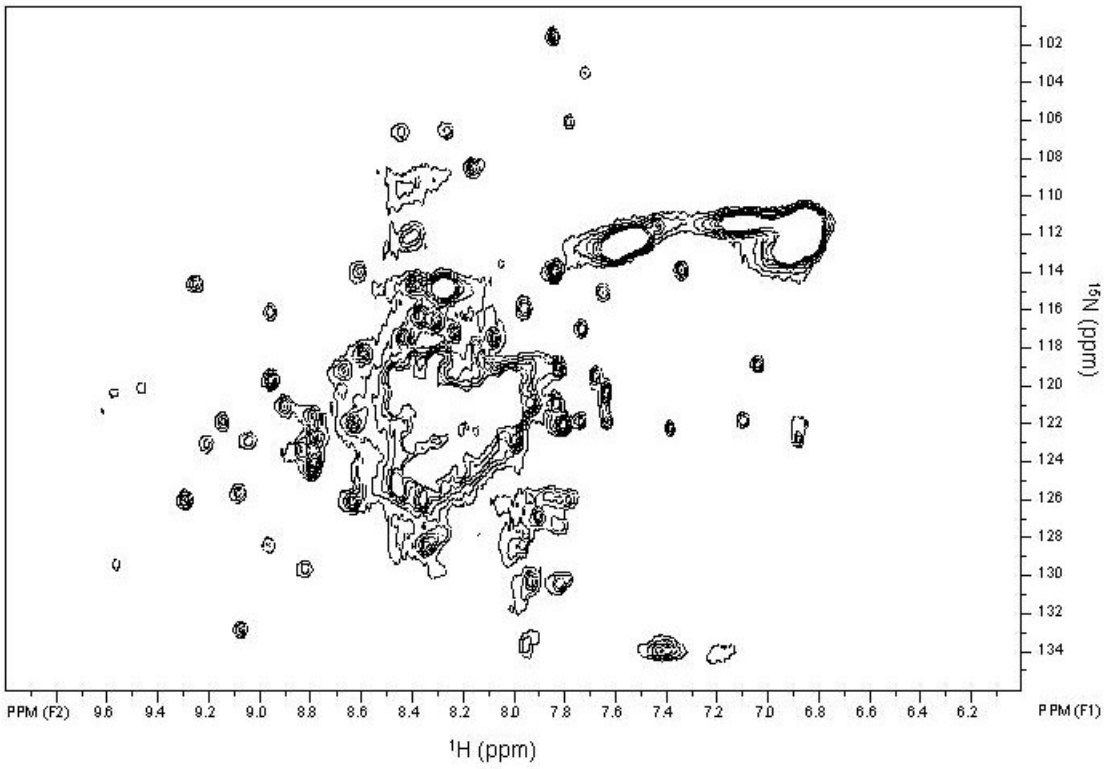


Figure 3.3: Two-Dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.2mM U-^{15}N *E. coli* CS in the presence of 0.1M KCl, measured at 25ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

Table 3.1: Line widths (^{1}H) of resolved cross-peaks of U-^{15}N labeled *E. coli* CS, with and without the addition of 0.1M KCl

^{1}H (ppm)	^{15}N (ppm)	Line width (^{1}H) (Hz) No KCl	Line Width (^{1}H) (Hz) With KCl
9.05	132.91	23.0	25.6
8.80	129.15	19.8	38.2
8.94	128.23	20.0	36.2
9.05	125.26	22.6	30.1
9.27	126.04	23.8	27.1
9.53	129.46	23.4	24.2
9.19	123.05	27.3	31.8
9.12	122.09	13.8	22.9
9.04	123.32	23.3	24.8
9.01	122.98	24.3	36.9
8.93	115.90	21.4	24.8
8.92	119.80	19.1	28.6
8.87	121.20	21.1	30.5
7.82	101.70	17.3	21.7
7.69	103.59	21.8	23.3
7.78	106.23	22.0	26.6
8.15	108.71	18.3	29.8
8.23	106.72	19.0	31.4
8.41	106.65	20.1	40.3
8.57	114.14	28.2	36.3
7.92	133.89	19.5	27.7
7.91	130.50	18.2	32.3
7.79	130.46	25.4	36.9
7.01	119.04	18.9	24.6
7.06	121.95	17.6	26.9
7.35	122.21	19.2	18.3
7.31	114.15	21.1	24.6
7.62	115.21	20.6	26.9
7.81	114.08	25.3	23.2

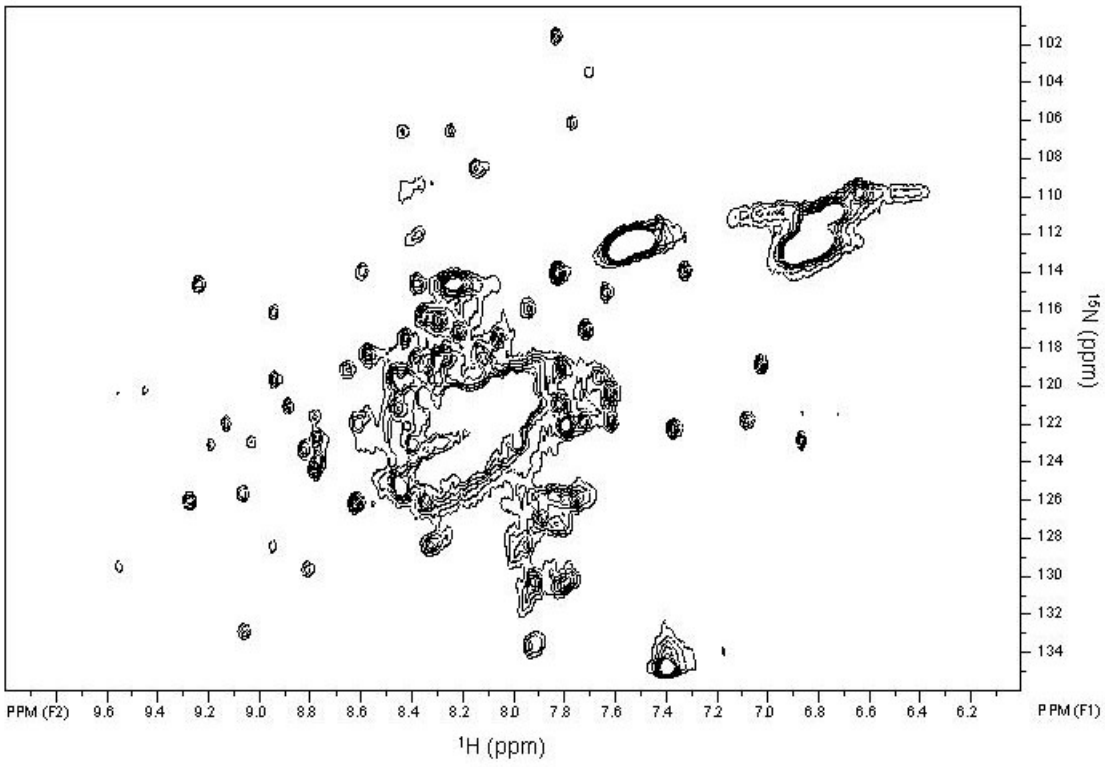


Figure 3.4: Two-Dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.2mM U-^{15}N *E. coli* CS in the presence of 0.1M KCl and 1.1 molar excess of OAA and CM-CoA, measured at 25ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

In a separate experiment, NADH, the allosteric inhibitor, was added to the U-^{15}N labeled *E. coli* CS. No changes were observed in the chemical shifts and line widths of the cross-peaks in the two-dimensional HSQC spectrum of *E. coli* CS as a result of this addition (Figure 3.5).

In the published crystal structures, the first six amino acids in the N-terminal region display B factors comparable to residues of the mobile loop, implying considerable flexibility. To test for the contribution of the flexible N-terminal region to the cross-peaks in the uniform labeled spectrum, U-^{15}N truncated (first 40 amino acids missing) *E. coli* CS were prepared. The HSQC spectrum (Figure 3.6) showed almost the same spectrum as U-^{15}N labeled WT CS, confirming that the overall folding of the truncated polypeptide is similar to the wild type, but with a few differences. In the regions marked with boxes in Figure 3.6, five peaks were observed in the wild type spectrum, which are not present in the truncated protein, indicating that those peaks are coming from the N-terminal region of the protein. An additional broad peak (marked with an arrow) was observed at (8.5ppm, 108ppm) in the two-dimensional HSQC spectrum of truncated *E. coli* CS, but not in the wild type (Figure 3.6, compare Figure 3.5). This additional peak may arise because of a change in the exchange kinetics of the residue, as a result of the N-terminal truncation.

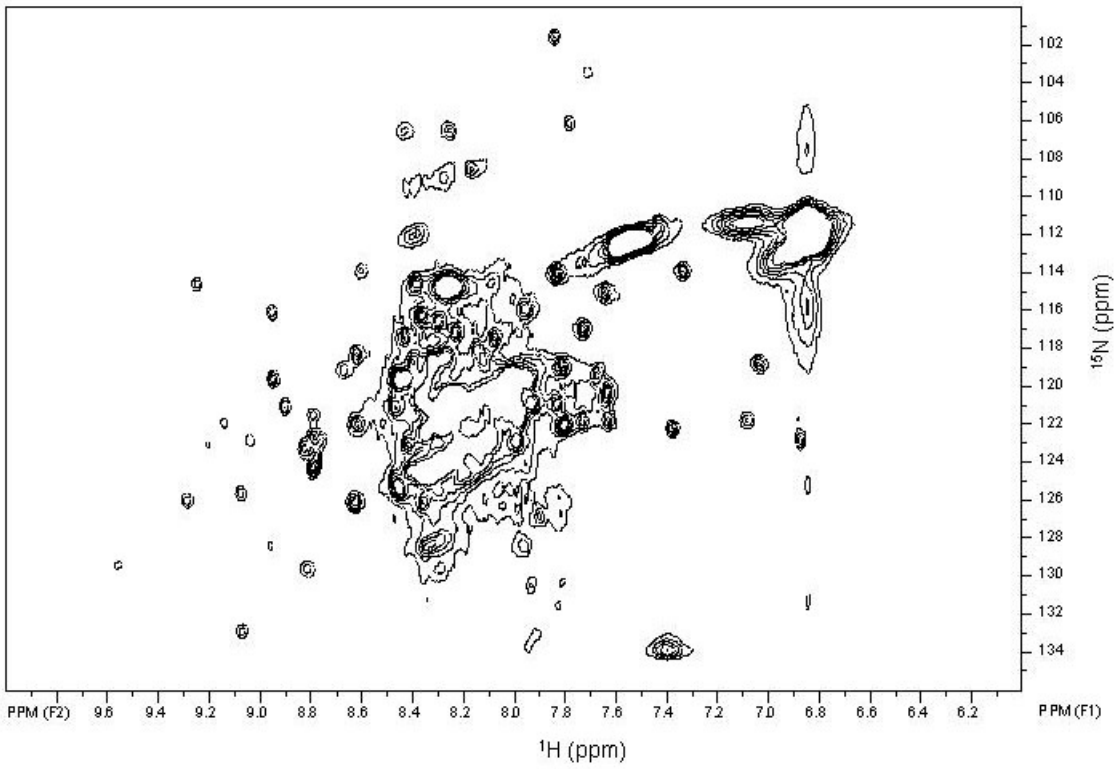


Figure 3.5: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.2mM U-^{15}N labeled *E. coli* CS in the presence of 1.1 molar excess of 0.1M NADH measured at 25 ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

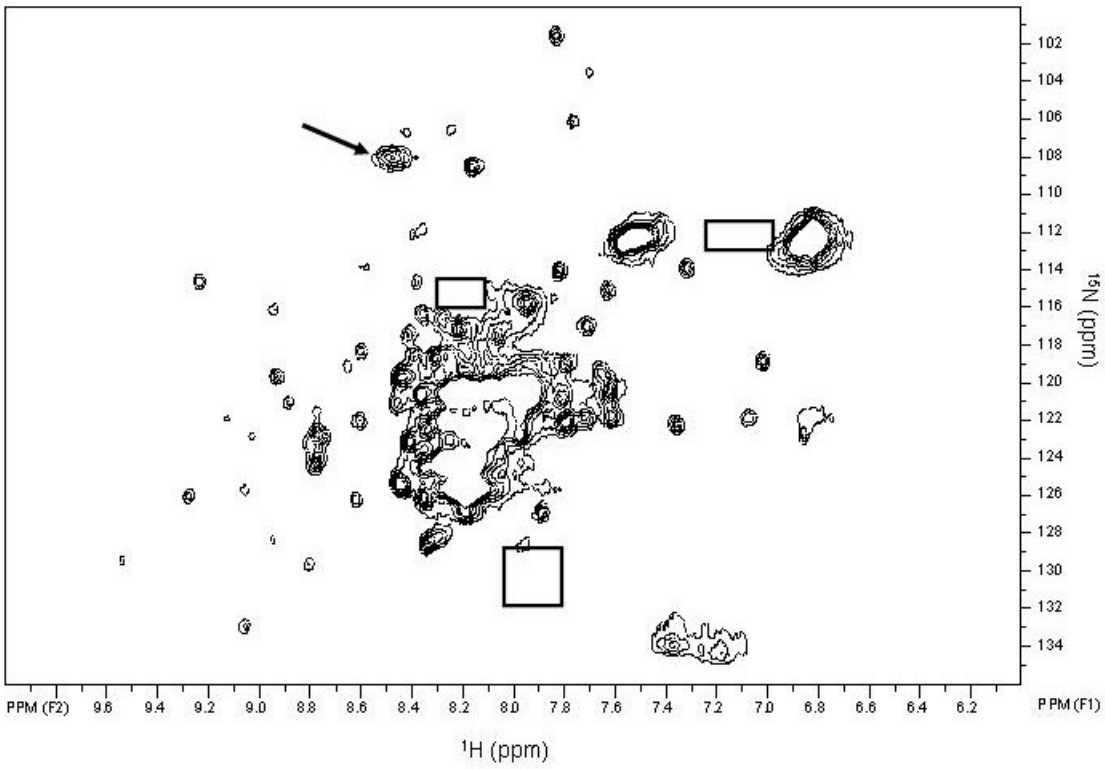


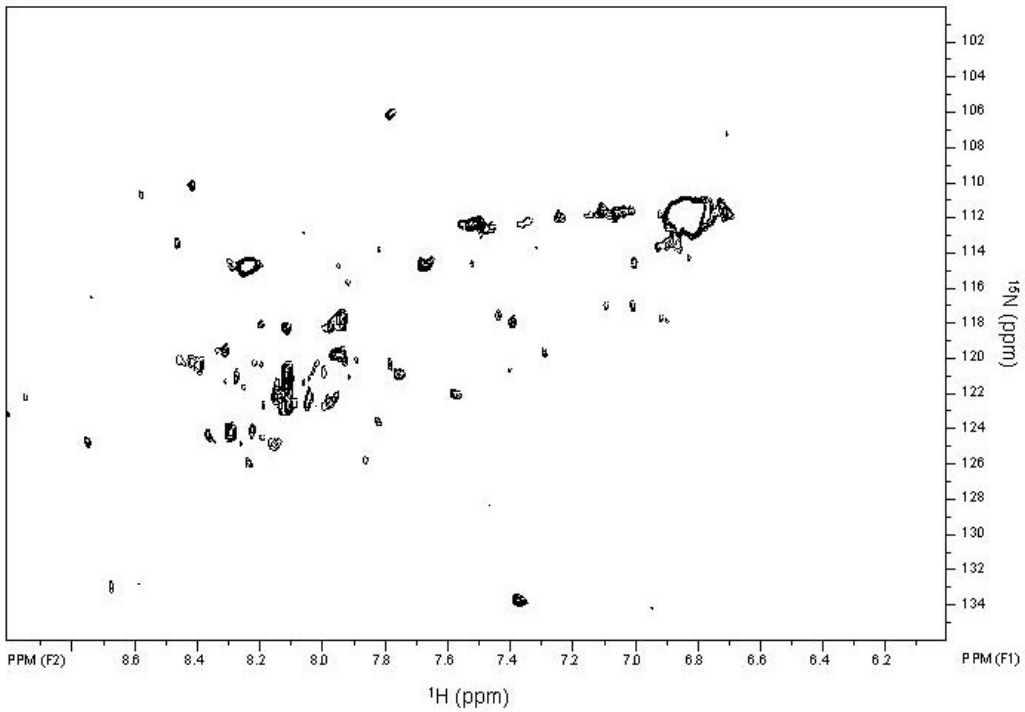
Figure 3.6: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 0.95mM U-^{15}N labeled *E. coli* truncated (first 40 amino acids missing) CS measured at 25 ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours. The arrow indicates a new peak, not present in the wild type spectrum (Figure 3.1), and the three boxes indicate regions where peaks have disappeared.

3.82 Uniformly ^{15}N, ^{2}H labeled *E. coli* CS

The main advantage of deuteration in proteins is to reduce transverse relaxation, to obtain sharp lines in the NMR spectrum. This is particularly useful for large proteins where line widths are broadened to a great extent, and often are not even observable, due to transverse relaxation.

The first attempt to produce *E. coli* U-^{15}N, ^{2}H CS was made by inoculating cells initially into a large volume (2L) of unlabelled medium followed by resuspension and induction in a small amount (250ml) of U-^{15}N, D_2O medium [Marley, et al., 2001]. This strategy, which was intended to maximize the use of the labeling medium, did not produce enough protein for NMR experiments. The two-dimensional HSQC spectrum (Figure 3.7), shows cross-peaks with a very low signal to noise ratio after 24 hours of acquisition.

In the second attempt to produce U-^{15}N, ^{2}H (70%) labeled CS the cells were grown in 2L of medium containing D_2O (70%) followed by addition of $^{15}NH_4Cl$ and induction. Protein yield in this case was about three times more than the previous preparation. Experiments on this sample were mainly done at 600MHz, with a few experiments done on the 800MHz spectrometer.



3.7: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 0.06 mM U-^{15}N, ^{2}H labeled *E. coli* CS measured at 25ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

Two dimensional HSQC and TROSY spectra of the sample from the U-^{15}N, ^{2}H (70%) medium, at both 600 and 800 MHz, are shown in Figures 3.8 - 3.11. Cross-peaks in the proton chemical shift region between 7.8 ppm and 8.3ppm were observed are and better resolved than in the sample with no deuterium enrichment. On the other hand, most of the well isolated peaks seen in the undeuterated sample (Figure 3.1 and Figure 3.2) were not seen after deuteration, and those that were still detected (four) were greatly reduced in intensity. These changes in relative intensity must reflect another aspect of deuterated samples: deuterons on peptide nitrogen must back-exchange to hydrogens before the corresponding cross peaks can be seen. Back-exchange should be relatively fast for nitrogens in random coil regions, so the cross-peaks between 7.8 ppm and 8.3 ppm are intense. The isolated peaks with larger dispersion, whose chemical shifts suggest special environments, are weak probably because back-exchange occurs much more slowly in those environments or are not observable due to exchange broadening of the peak. The effect of deuteration in improving the quality of NMR spectra in large proteins can be seen in the cross-peaks, shown in Figure 3.12a, which are quite broad in the TROSY spectrum of protonated protein (Figure 3.2) but are well defined peaks in the TROSY spectrum of deuterated protein (Figure 3.12b). A few cross-peaks shown in Figure 3.13 are observable only in the spectrum of the U-^{15}N/^{2}H sample. These cross-peaks must belong to the rigid parts of the protein, where fast transverse relaxation rates lead to

broadening, but this effect now has been reduced by deuteration. Both at 600MHz and 800MHz, there was a sensitivity gain when using TROSY as compared to the HSQC experiment. Such gains were achieved only when the protein had been deuterated.

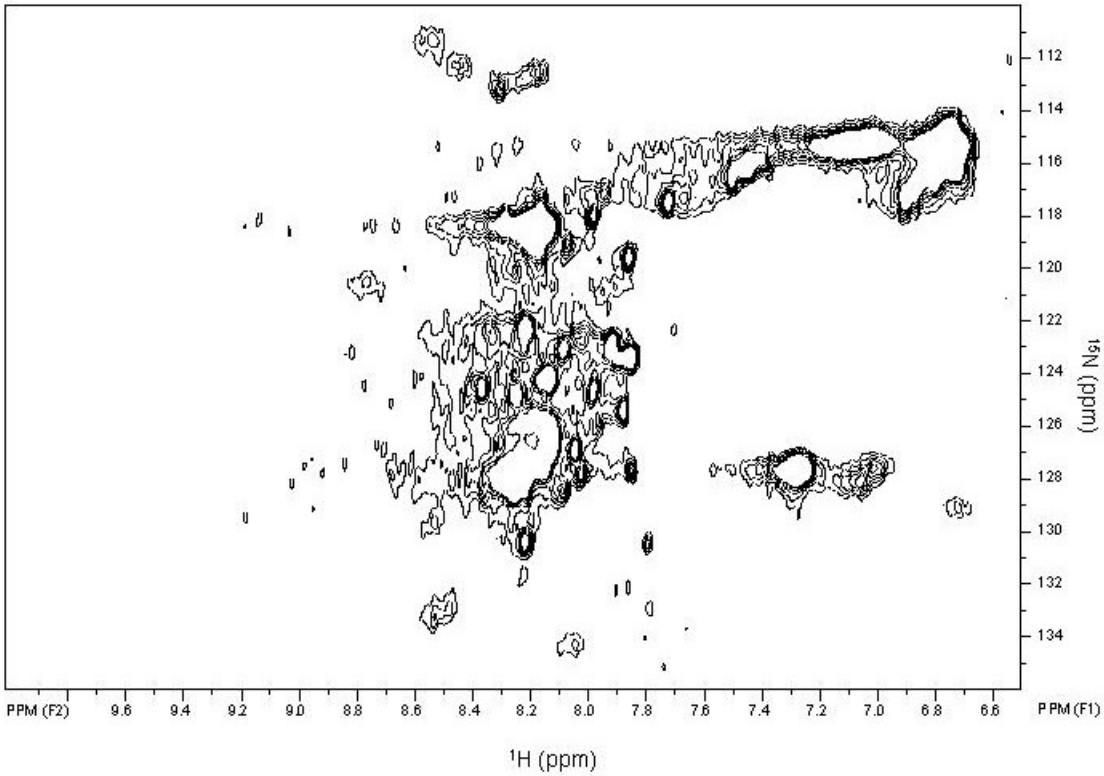


Figure 3.8: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured at 25 ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

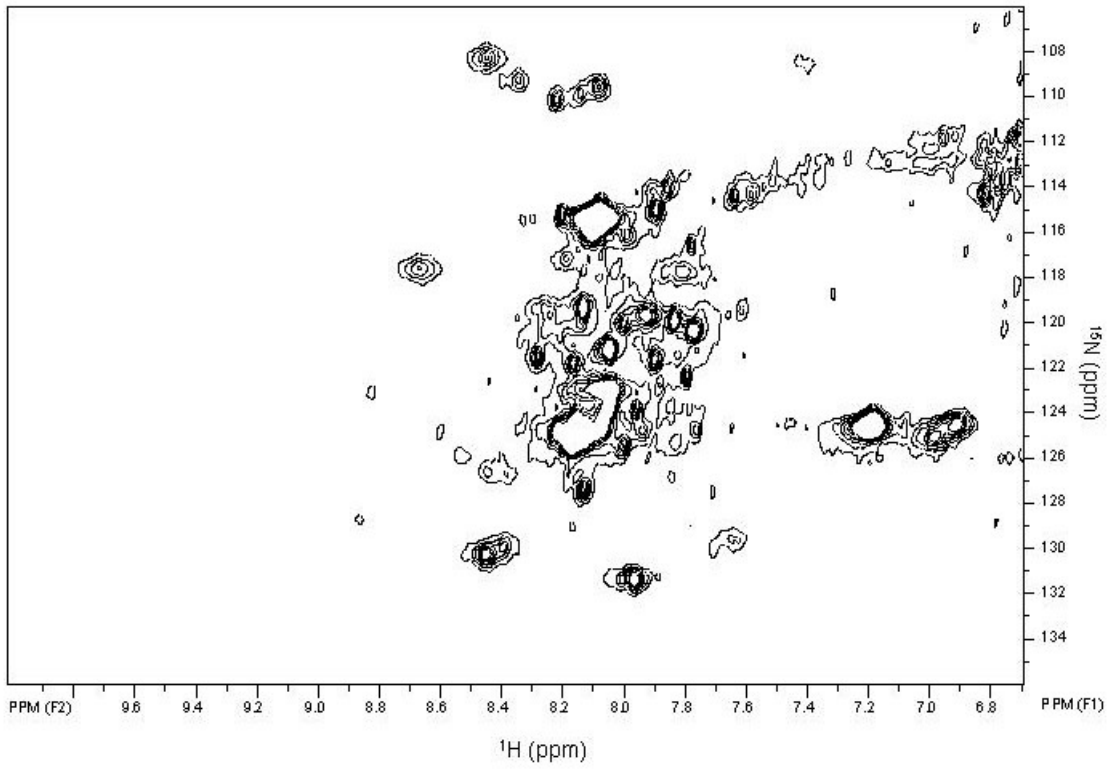


Figure 3.9: Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured at 25 ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

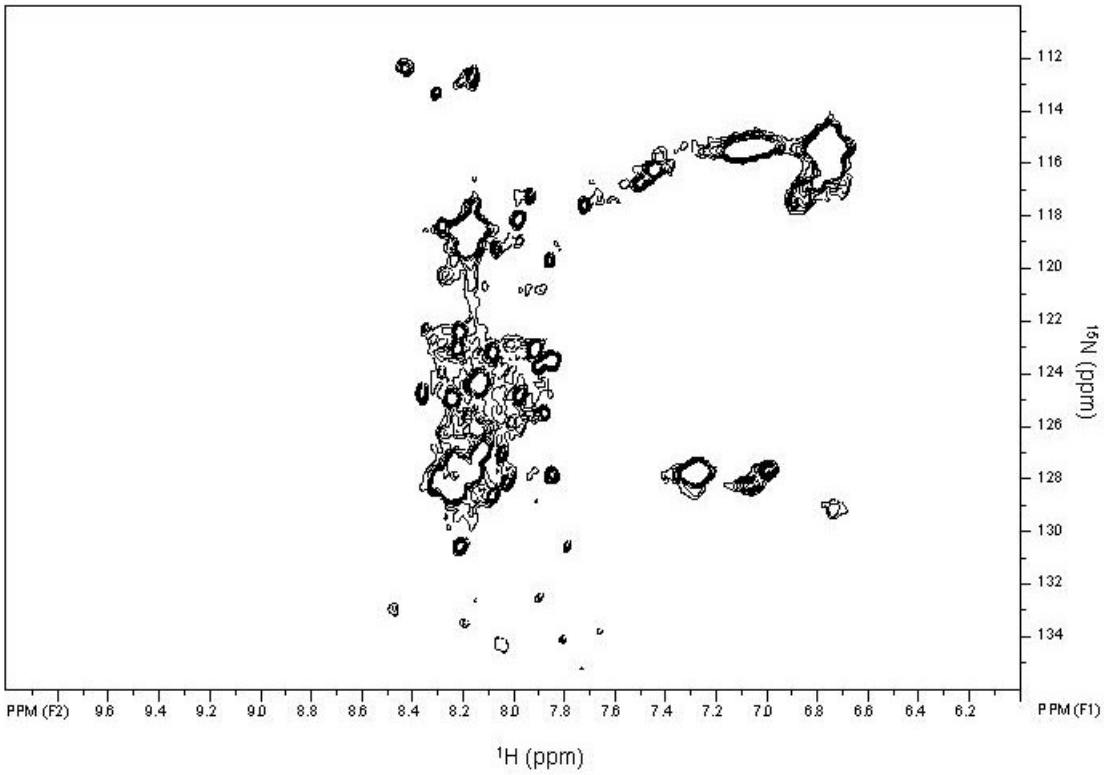


Figure 3.10: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured at 25 ºC at a ^{1}H frequency of 800 MHz. Total acquisition time was 2.5 hours.

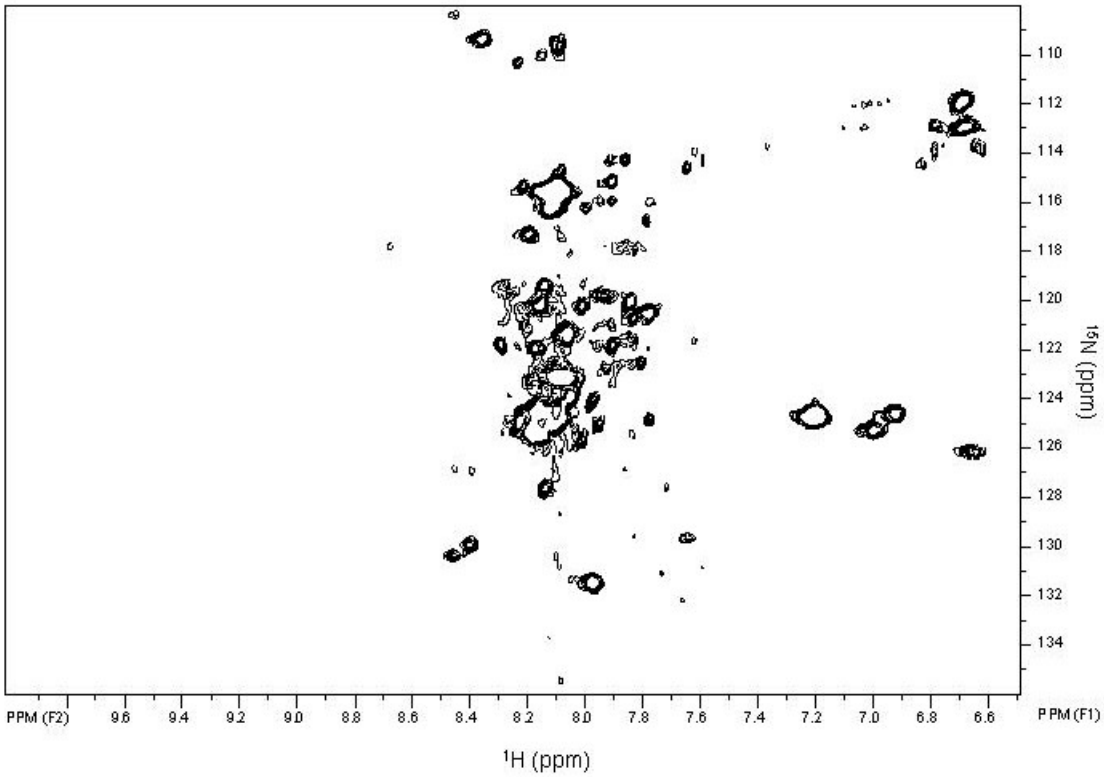


Figure 3.11: Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured at 25 ºC at a ^{1}H frequency of 800 MHz. Total acquisition time was 2.5 hours.

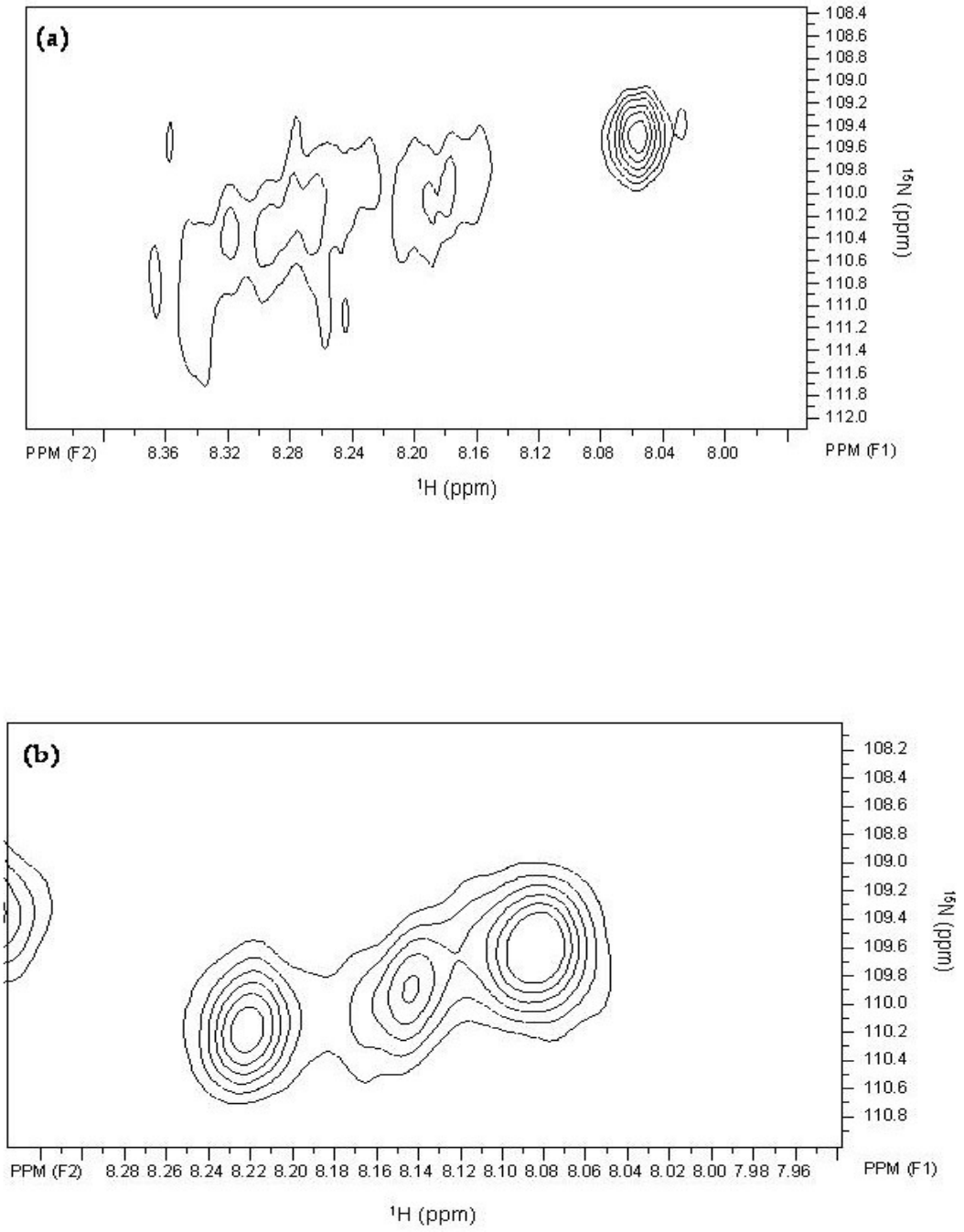


Figure 3.12: Comparison of equivalent regions in two-dimensional [^{1}H, ^{15}N] TROSY spectra of (a)U-^{15}N, ^{1}H (Figure 3.2)and (b)U-^{15}N, ^{2}H *E. coli* CS(Figure 3.9).

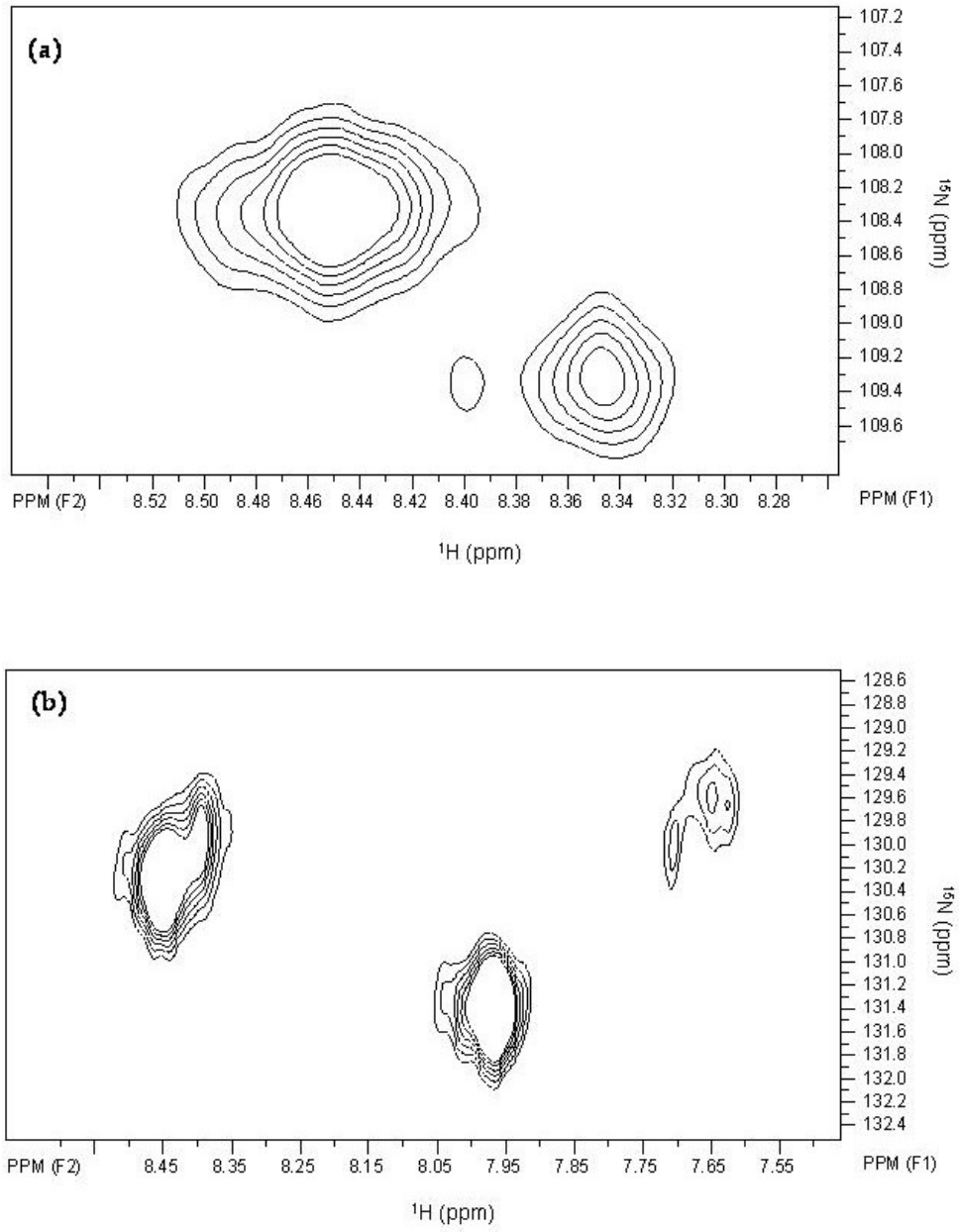


Figure 3.13: New cross-peaks that were observed in the two-dimensional [^{1}H, ^{15}N] TROSY spectrum of U-^{15}N, ^{2}H, but not in U-^{15}N, ^{1}H *E. coli* CS. (a) and (b) taken from different regions from Figure 3.2.

CRINEPT-TROSY overcomes relaxation during magnetization transfer periods, which becomes limiting for very large systems and which is not suppressed by TROSY [Riek, et al., 1999]. This approach allows spectra to be obtained from molecules even above 200kDa [Rudiger, et al., 2002]. A CRINEPT-TROSY spectrum of U-^{15}N, ^{2}H *E. coli* CS (Figure 3.14) at a ^{1}H frequency of 800 MHz showed additional signals which were not observed in the TROSY spectrum. However in CRINEPT-TROSY, for flexible regions of a large molecule a second faster relaxing component is retained, causing doubling of the corresponding signals [Riek, et al., 2002]. In the case of U-^{15}N, ^{2}H *E. coli* CS, it was not possible to take full advantage of CRINEPT-TROSY in looking at the rigid regions due to incomplete back-exchange of deuterium in those parts of the enzyme.

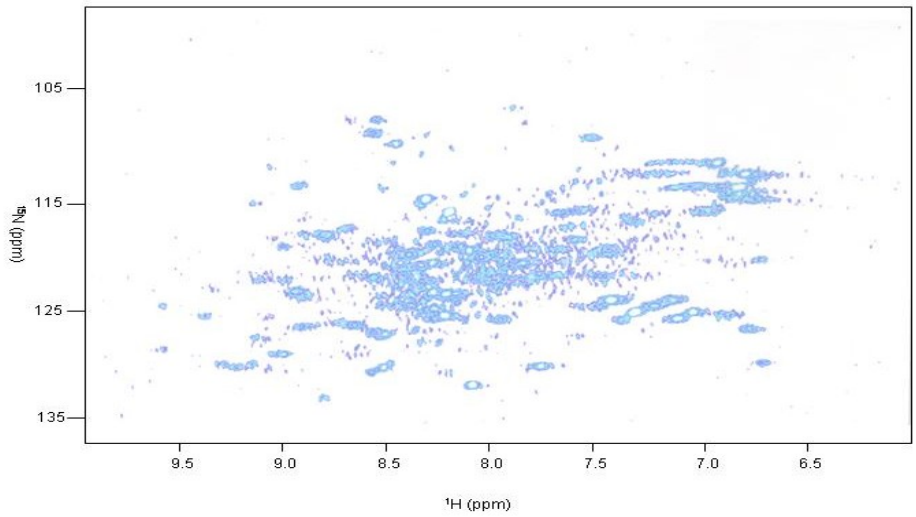


Figure 3.14: Two-dimensional [^{1}H, ^{15}N] CRINEPT-TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured at 25 ºC at a ^{1}H frequency of 800 MHz.

On addition of substrates in the presence of 0.1M KCl (Figure 3.15), ten peaks as marked in Figure 3.15 were not observable and at least four new peaks were observed. These results indicate a change in environment of these flexible residues in the presence of substrates. Once it was realized that a number of cross-peaks displayed incomplete back-exchange during the purification process in water, it was decided to follow the back-exchange process over longer times, by two-dimensional NMR spectroscopy. After one month of further back-exchange under these conditions, five new cross-peaks marked by boxes in Figure 3.16 were observed in the TROSY spectrum. After two months, the intensity of these new cross-peaks increased, but no more peaks appeared (Figure 3.17). No further changes were noted at longer times (Figure 3.18). A few cross-peaks (marked with arrows in Figure 3.18), were broadened possibly due to faster transverse relaxation by the nearby back-exchanged protons.

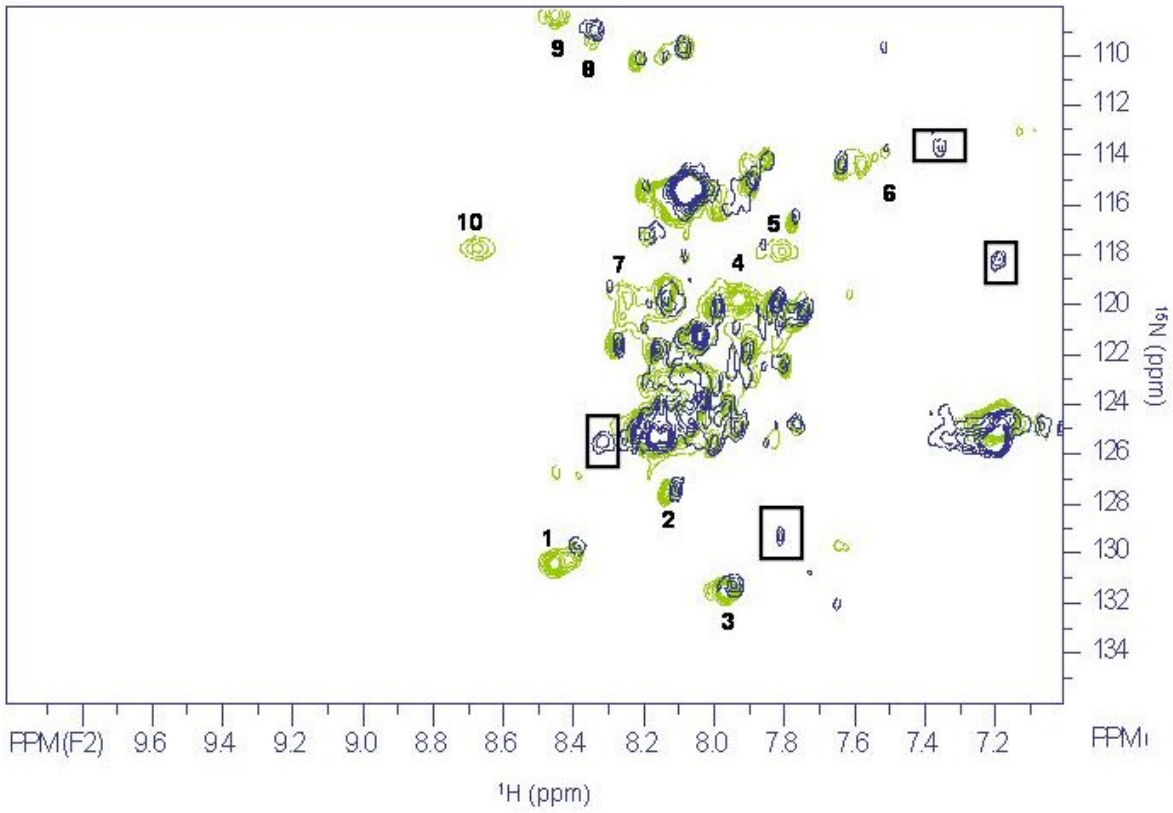


Figure 3.15: Overlay of two-dimensional [^{1}H, ^{15}N] TROSY spectra of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS with (shown in blue) and without substrates (shown in green) measured at 25 ºC at a ^{1}H frequency of 600 MHz. Total acquisition time for each spectrum was 15 hours. Cross-peaks affected by the substrates are marked with numbers. New cross-peaks marked with boxes were observed only after addition of the substrates.

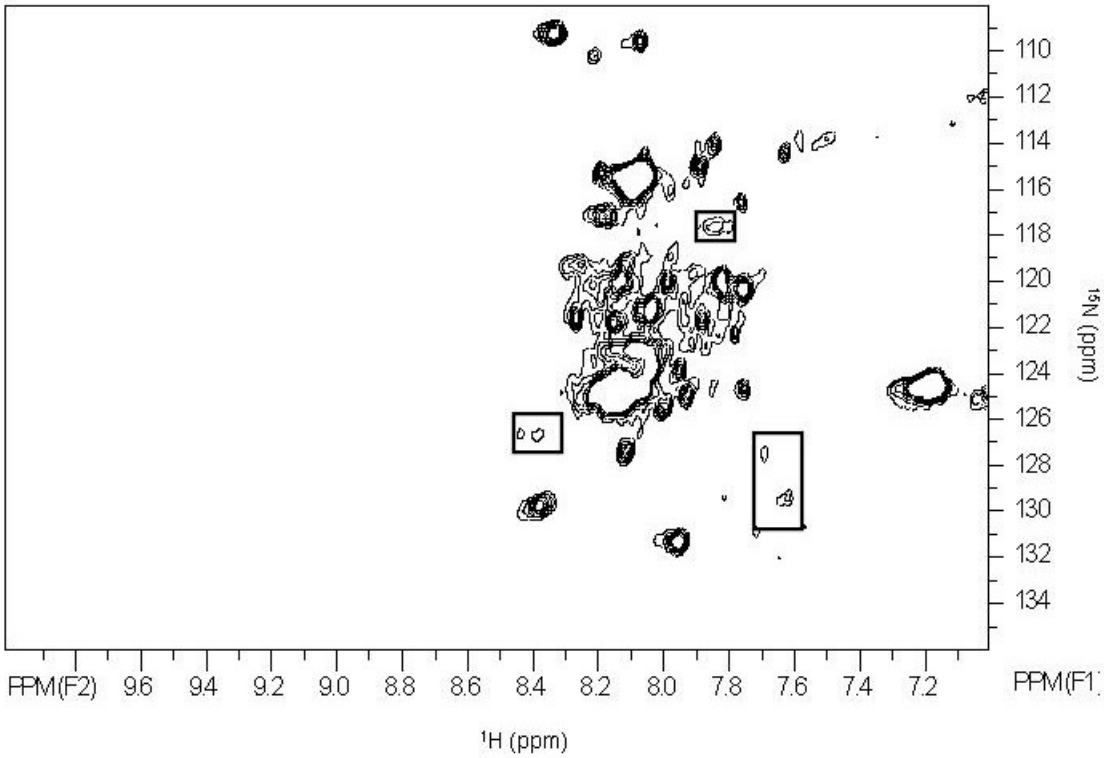


Figure 3.16: Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N labeled *E. coli* CS measured after one month of amide back-exchange, at 25 ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours. New cross-peaks appearing as a result of back-exchange are marked with boxes.

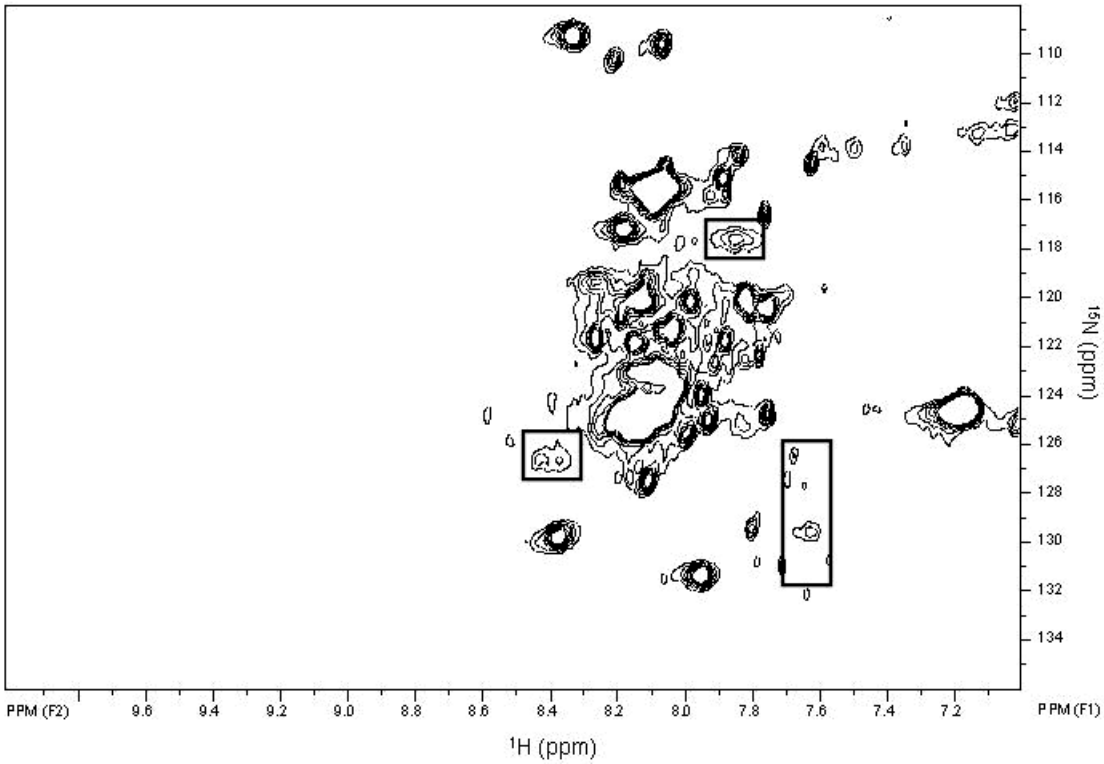


Figure 3.17: Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured after two months of amide back-exchange, at 25 ^{o}C at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours. Back-exchanged cross-peaks with increased intensity are marked with boxes.

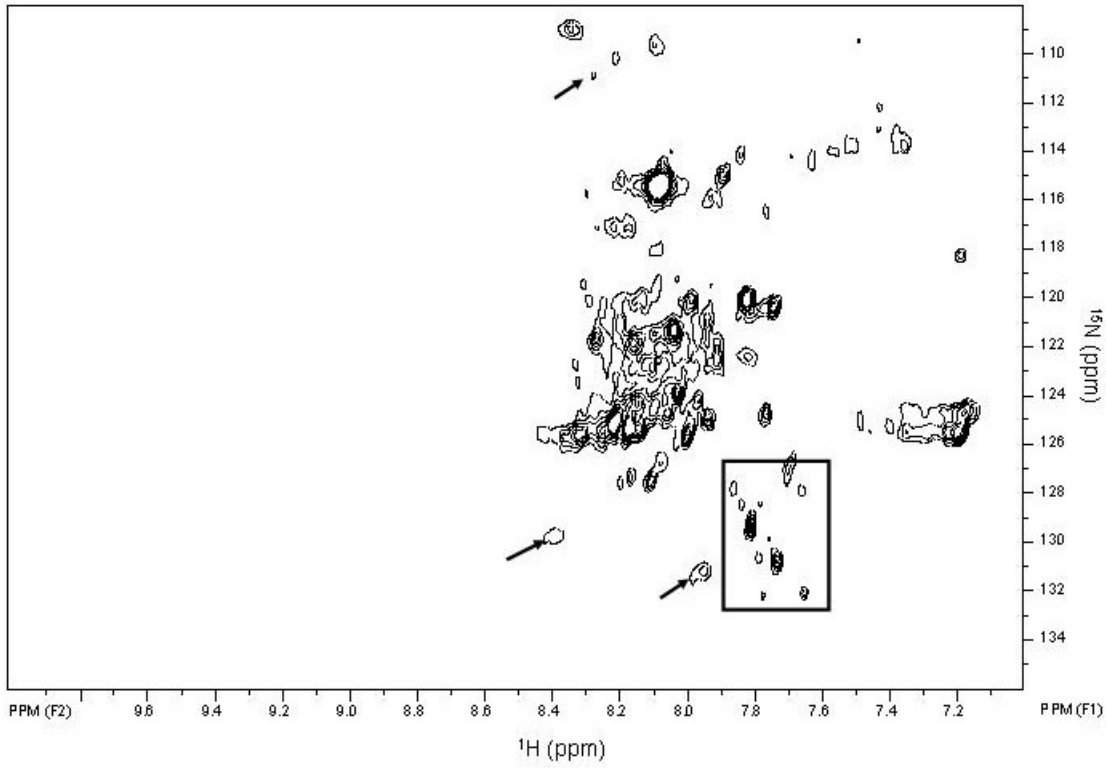


Figure 3.18: Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.9mM U-^{15}N, ^{2}H labeled *E. coli* CS measured after one year of amide back-exchange, at 25 ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours. Back-exchanged cross-peaks with further increased intensity are marked with boxes. Peaks affected by time are marked by arrows.

3.83 Specific ^{15}N, ^{13}C labeled *E. coli* CS

From the studies with U-^{15}N labeled *E. coli* CS it is clear that the crowded region between 7.8 ppm and 8.3 ppm has to be simplified to get further information about the system. To perform this task, samples of *E. coli* CS were prepared which had been labeled with specific ^{15}N-amino acids since the focus was to study the "mobile loop" residues 266-297, the amino acids initially chosen were those for which there were only one or two in the mobile loop sequence.

To minimize conversion of the label to other amino acids, amino acids which are known to be key precursors to others, such as glutamate, aspartic acid, serine and those known to be substrates for transaminases were avoided. In *E. coli* four main transaminases with overlapping substrate specificity have been found. Out of four transaminases, one is non-repressible and is involved in the transamination between Glu↔Asp, Phe and Tyr with highest affinity for Asp. The other three transaminases can be repressed with the addition of unlabeled amino acids in the medium. In addition there is an uncharacterized transaminase which is involved in Glu↔Ala transamination [McIntosh et al., 1990]. Based on the above criteria, specific ^{15}N-amino acid labeling with Ala, Glu, Asp, Asn and Ser in the mobile loop was avoided. Protein samples were prepared using ^{15}N-labels for all of the other amino acids in the mobile loop except Val, Ile and Arg.

^{15}N-Phe *E. coli* CS

The first attempt at preparing specifically ^{15}N-labeled *E. coli* CS was carried out using ^{15}N-phenylalanine. This sample failed to show any signals in the 2D-HSQC spectrum. A second sample was then prepared, and the HSQC spectrum of this sample showed a peak envelope consisting of two clusters of peaks with different intensities. The first peak of higher intensity was observed at a proton chemical shift of 9.5 ppm while a weak, poorly resolved multiplet appeared in the region of 8.0 to 8.2 ppm (Figure 3.19). There are 22 phenylalanine residues present in *E. coli* CS, whereas only one residue, F287 is present in the mobile loop (Figure 3.20). Only one signal was expected in the HSQC spectrum. The isotopic incorporation of ^{15}N label in the two samples as measured by mass spectrometry was ~50 and ~ 60% respectively. The improvement in incorporation of ^{15}N-label in the second attempt was attained by changing the growth and induction conditions. In the first attempt, cells were induced at A_{600} ~ 1.8 and ^{15}N-Phe was added at A_{600} ~ 1.6 whereas, in the second attempt, cells were induced at A_{600} ~ 0.5 and ^{15}N-Phe was added at A_{600} ~ 0.1. In this particular labeling the low level of isotopic incorporation is a possible reason for the weak spectrum.

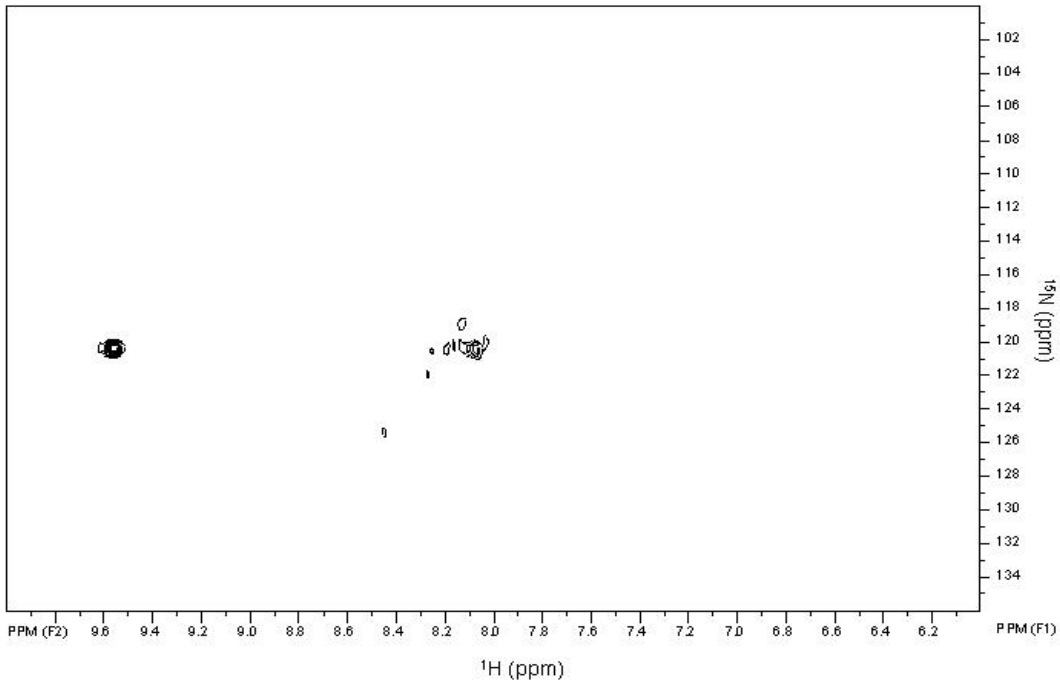


Figure 3.19: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 0.80mM ^{15}N-Phe labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

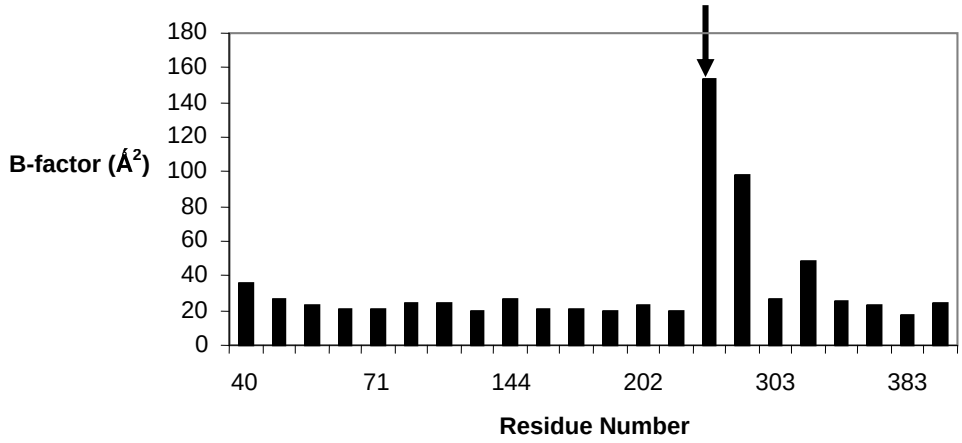


Figure 3.20 B-factors of backbone nitrogen for phenylalanine residues in *E. coli* CS [Nguyen, et al., 2001]. The bar marked with arrow corresponds to F287 of the mobile loop.

^{15}N-Met *E. coli* CS

In the HSQC spectrum of ^{15}N-Met *E. coli* CS, a total of 16 amide signals with variable intensities were observed at proton chemical shifts between 7ppm and 9ppm (Figure 3.21). Since only one of the 18 methionine residues in *E. coli* CS is in the mobile loop (Figure 3.22), only one cross-peak was expected in the 2D-HSQC spectrum. The spectrum actually obtained is puzzling. We do not expect to observe nearly all the methionines, some of which are present in the rigid parts, in nondeuterated protein of MWT 280kDa [Pervushin, et al., 1997]. The intensity of the cross-peaks vary in the spectrum, and the reason might be the difference in the mobilities of the residues (some being mobile than others) or the involvement of residues in a conformational exchange. For example the peak 16 in Figure 3.21 is significantly broader than the rest of the peaks, indicating involvement of that particular methionine in intermediate conformational exchange. The intensity will be higher in cases where the cross-peak observed is arising from the methionine residue which has the same environment in all the six subunits.

Further, mass spectrometry analysis of the sample showed no traces of scrambling into other amino acids that could account for more than the expected peaks. The isotopic incorporation of ^{15}N label as measured by mass spectrometry was estimated to be 95%.

On addition of substrates, OAA and substrate analogue CM-CoA in the presence of 0.1M KCl, one of the 16 Met cross-peaks shifted indicating a change in the environment of the particular methionine residue (Figure 3.23).

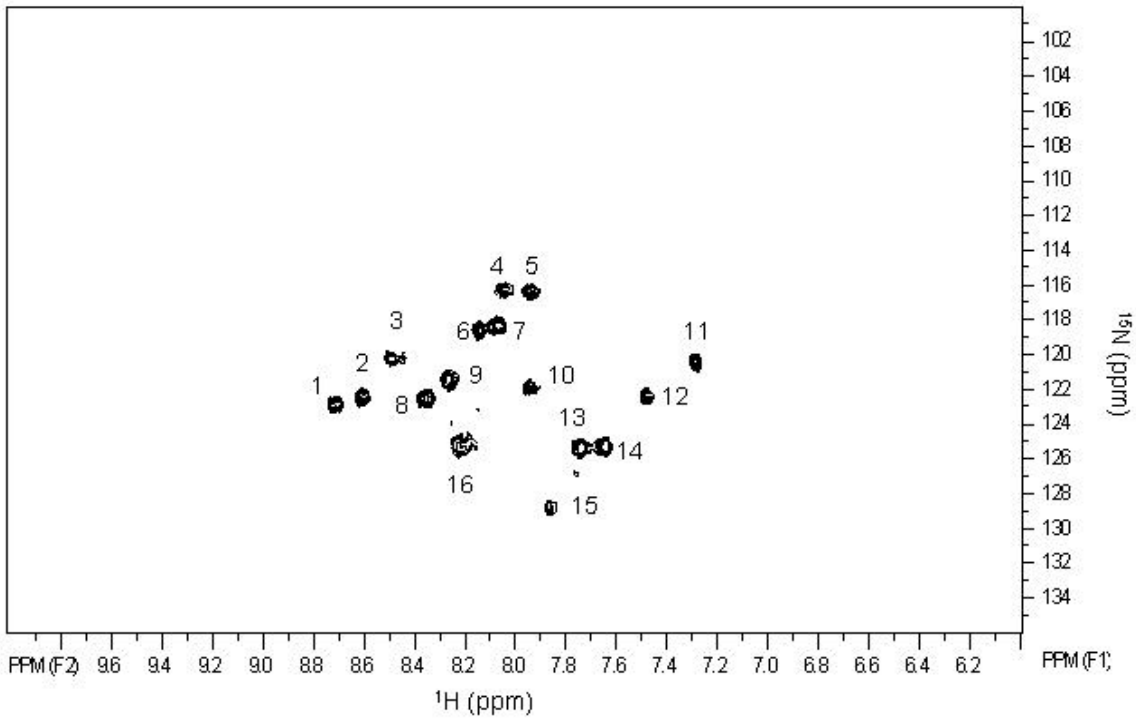


Figure 3.21: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 2.1mM ^{15}N-Met labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 500 MHz. Total acquisition time was 24 hours.

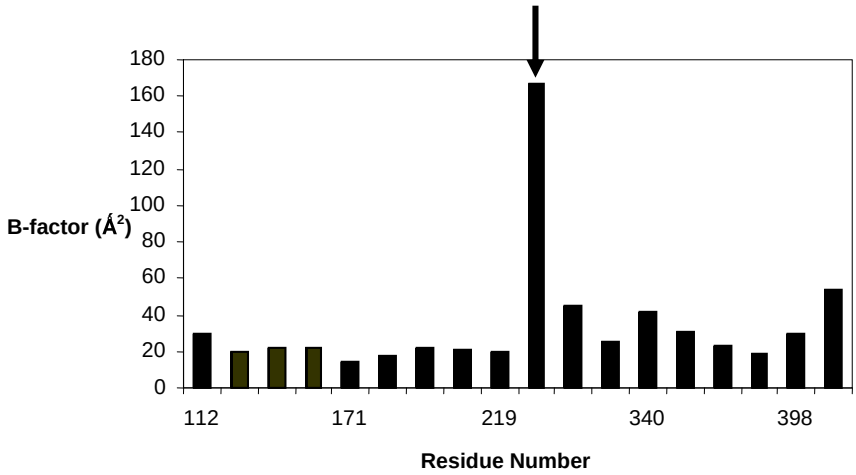


Figure 3.22: B-factors of backbone nitrogen for methionine residues in *E. coli* CS [Nguyen, et al., 2001]. The bar marked with arrow corresponds to M274 of the mobile loop.

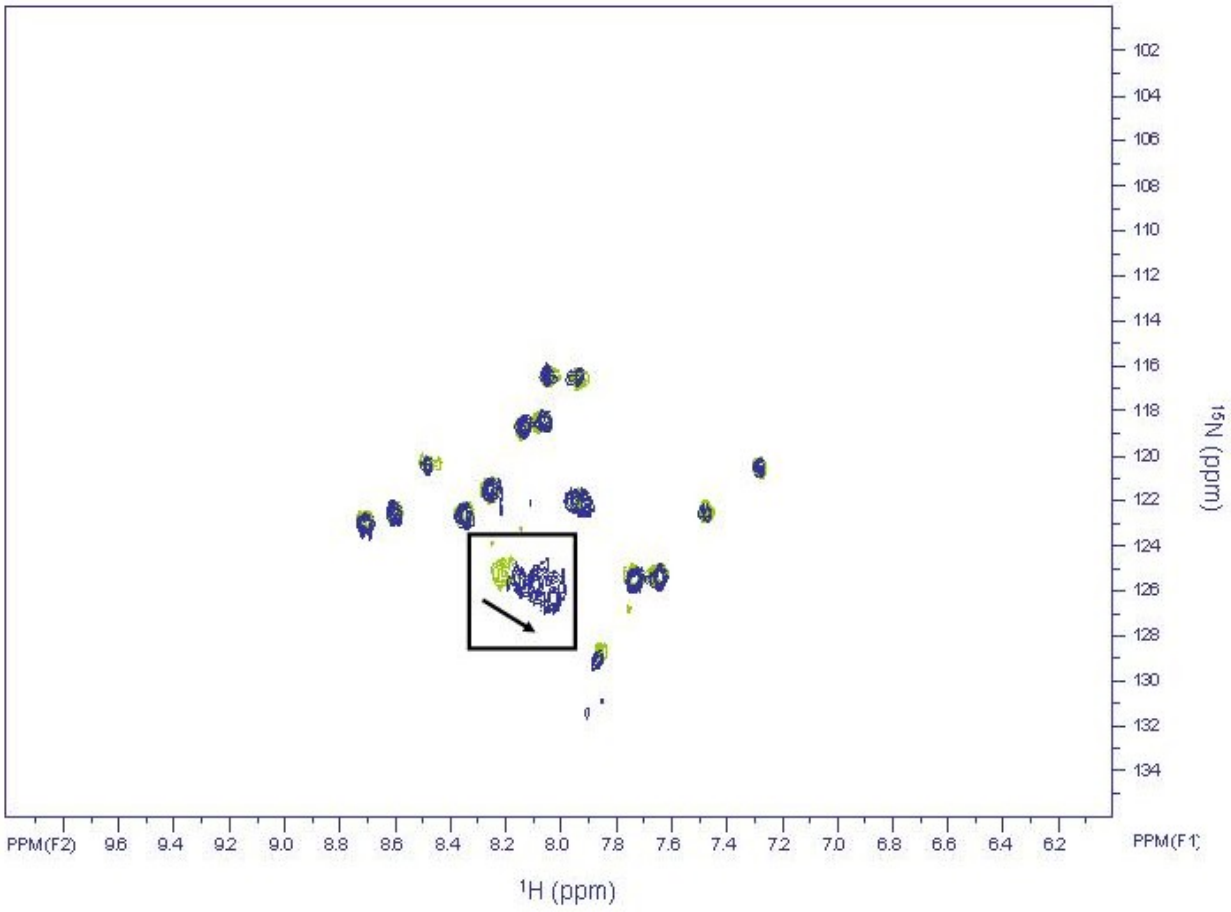


Figure 3.23: Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Met labeled *E. coli* CS, with (shown in blue) and without substrates (shown in green), at 25ºC at a ^{1}H frequency of 500 MHz. The change in the chemical shift of the one cross-peak is marked with an arrow.

In an attempt to identify the cross-peak arising from the single mobile loop methionine, M274, a ^{13}C-^{15}N double labeling approach was tried. The Lys-Met dipeptide appears twice in *E. coli* CS, K273-M274 which belongs to the mobile loop, and K167-M168 located around the NADH binding site. The strategy combines ^{13}C labeling of an amino acid at its carbonyl carbon position with ^{15}N labeling of the next amino acid in the sequence at the α-position. The combination of the two specific labels should result in splitting of cross-peaks due to the $^{1}J_{C'N}$ coupling in the ^{15}N-dimension in a HSQC/TROSY spectrum.

In the HSQC spectrum of ^{13}C [Lys]-^{15}N [Met] labeled *E. coli* CS, fifteen cross-peaks were observed in the proton chemical shift range of ~7.0 to 9.0 ppm. The number of cross-peaks observed is one less than observed in single ^{15}N-labeled *E. coli* CS. The broad peak with chemical shift of 8.3 and 124 ppm in the HSQC spectrum of the single ^{15}N-labeled sample is very weak in the HSQC spectrum of the ^{13}C [Lys]-^{15}N-[Met] labeled sample. A combination of exchange broadening and the fact that the double labeled sample was only half the concentration of the single ^{15}N-Met labeled sample probably accounts for the weak signal.

On comparing the HSQC spectra of decoupled (no C′N coupling) (Figure 3.24) and coupled (C′N coupling allowed) (Figure 3.25) ^{13}CLys-^{15}N-Met labeled *E. coli* CS, it can be seen that peak splitting was not observed in any of the cross-peaks in the ^{15}N dimension of the coupled spectrum. A

comparison of the line widths of the cross-peaks in the decoupled and coupled spectra is given in Table 3.2. ^{15}N line widths of the methionines are in the range of 12-18 Hz comparable to the expected coupling constant (15Hz) and therefore any splitting in the peaks would be very difficult to observe. Line broadening in the ^{15}N dimension was observed for two peaks marked with a box in Figure 3.24. The above experiment did not provide clear cut results for the assignment of the methionine residues of interest.

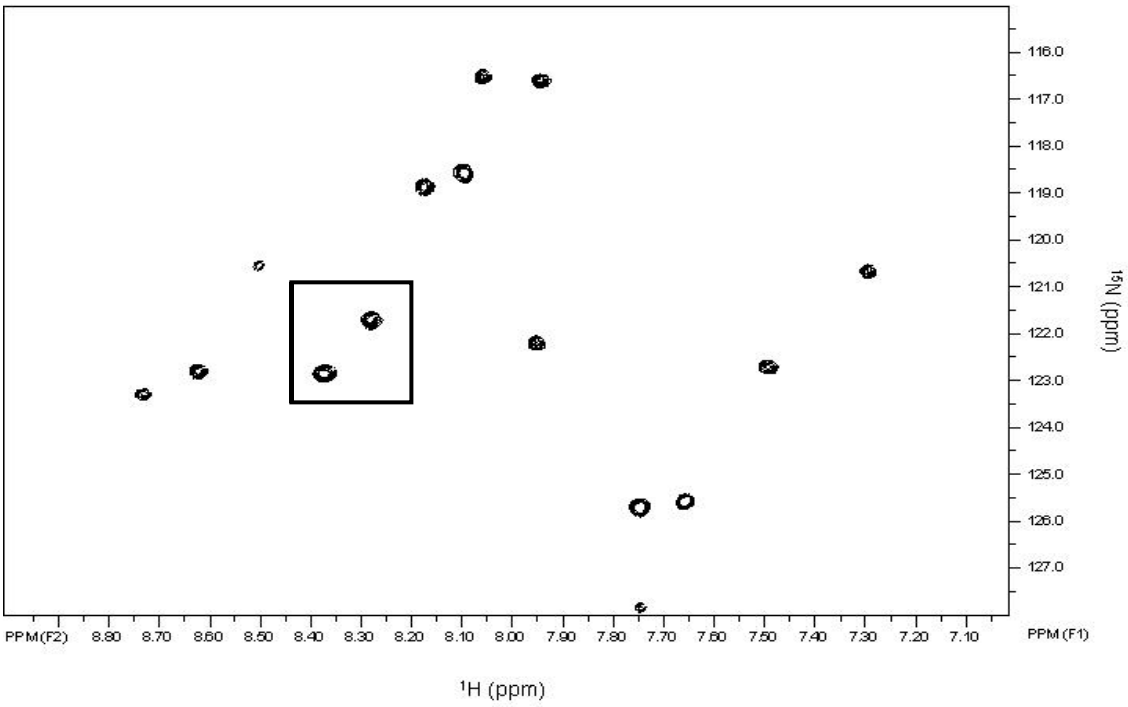


Figure 3.24: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 0.9mM ^{13}CLys-^{15}N-Met labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. No decoupling was allowed between the carbonyl carbon of lysine and the α-nitrogen of methionine. Line broadening was observed for cross-peaks marked with box (See Table 3.2).

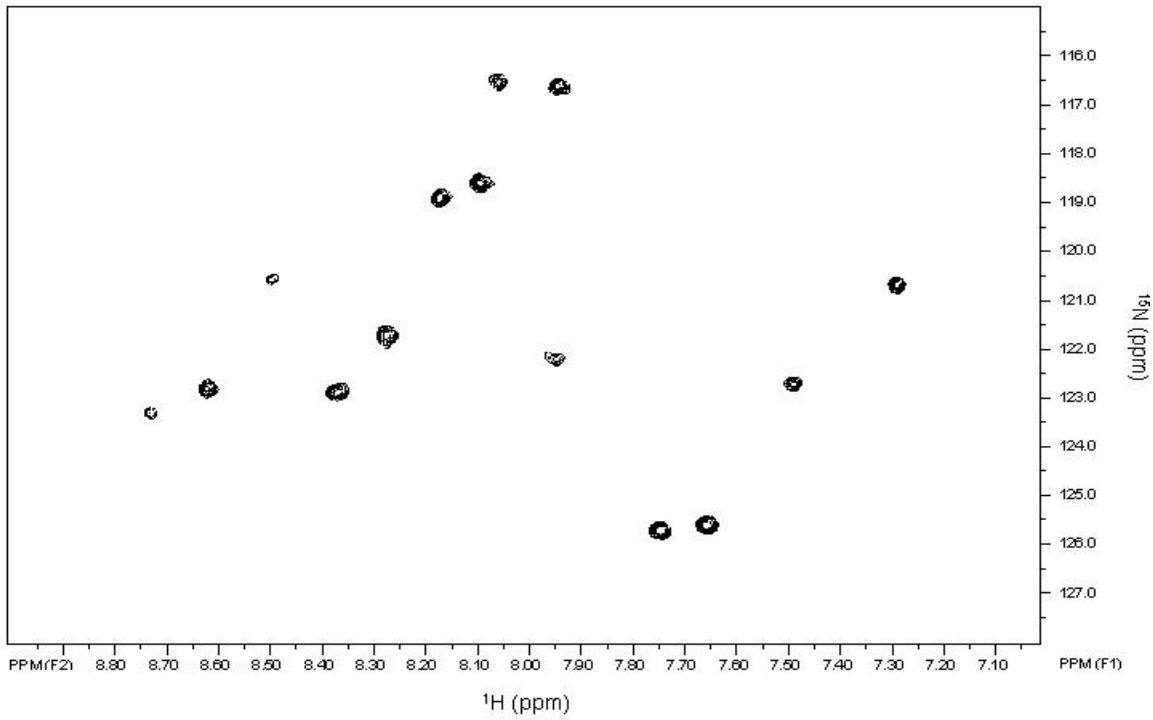


Figure 3.25: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 0.9mM ^{13}CLys-^{15}N Met labeled *E. coli* CS at a 25ºC on ^{1}H frequency of 600 MHz spectrometer. Coupling was allowed between the carbonyl carbon of lysine and the α-nitrogen of methionine.

Table 3.2: Line widths of cross-peaks from decoupled and coupled HSQC spectrum of ^{15}N-Met *E. coli* CS

^{1}H (ppm)	^{15}N (ppm)	^{1}H Line widths (Hz)		^{15}N Line widths (Hz)	
		Decoupled	Coupled	Decoupled	Coupled
7.75	125.79	22.0	23.0	14.0	14.0
7.65	125.63	19.0	18.0	18.0	18.0
7.49	122.74	24.0	18.0	16.0	14.0
7.29	120.72	22.0	17.0	16.0	16.0
7.95	122.24	26.0	21.0	18.0	20.0
7.94	116.67	24.0	18.0	16.0	16.0
8.06	116.59	26.0	14.0	16.0	14.0
8.28	121.75	23.0	27.0	14.0	21.0
8.37	122.90	25.0	25.0	12.0	16.0
8.73	123.32	30.0	22.0	12.0	13.0
8.62	122.86	21.0	21.0	14.0	14.0
8.50	120.59	21.0	17.0	15.0	14.0
8.17	118.86	18.0	11.0	16.0	15.0
8.09	118.65	19.0	23.0	14.0	14.0

Error in measuring line widths: ± 2 Hz

Line widths were measured without application of any line broadening in both dimensions.

Site-specific mutants M274L and M168L were made, to be used for the assignment of cross-peaks in the HSQC spectrum of ^{15}N-Met *E. coli* CS. First, unlabeled protein purification was carried out in LB media for both mutant proteins. Both proteins were checked for the right mutation using mass spectrometry of the tryptic fragments. Limited enzyme kinetic experiments were carried out and the results showed normal kinetics for both mutants. For mutant M274L, values of K_{OAA} and $K_{Ac\text{-}CoA}$ were determined to be 46±4 µM and 158±50 µM as compared to respective values of 26±5 µM and 120±20 µM for wild type. M168 is close to the NADH binding site and hence NADH inhibition was determined for mutant M168L and compared to the wild type. $K_{i,\ NADH}$ for M168L was determined to be 2.0±0.2 µM as compared to 2.8±0.4 µM for the wild type.

^{15}N-Met labeled M274L and M168L *E. coli* CS were then prepared for NMR experiments. The two-dimensional HSQC spectrum of ^{15}N-Met M274L *E. coli* CS (Figure 3.26) showed just one strong cross-peak and a few weak peaks as compared to the 16 cross-peaks seen in the wild type spectrum. The possible reasons may be the concentration of the M274L NMR sample which was half that of the wild type or a drastic change in the conformations of the methionines as a result of the mutation.

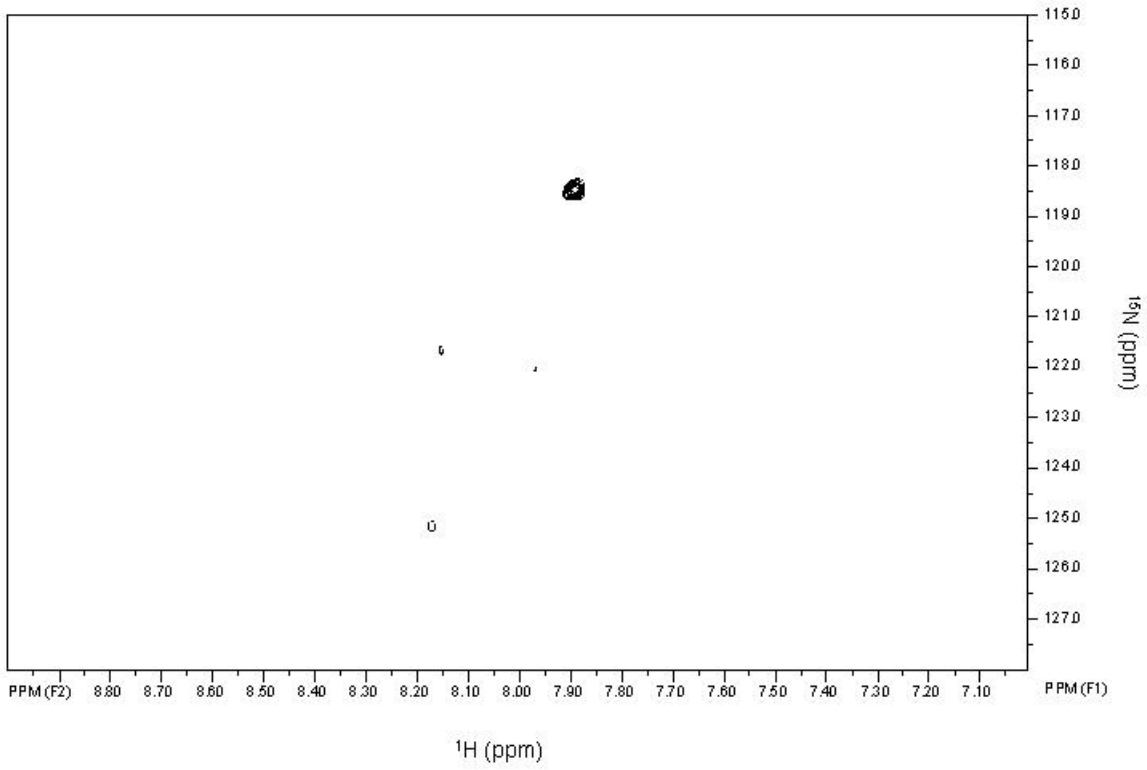


Figure 3.26: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 2.3mM ^{15}N-Met labeled M274L *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

A two-dimensional HSQC spectrum of ^{15}N-Met M168L *E. coli* CS (Figure 3.27) showed almost the same number of peaks as wild type, but with different chemical shifts. This result shows that the mutation has changed the environments of the methionines observable in the NMR spectrum.

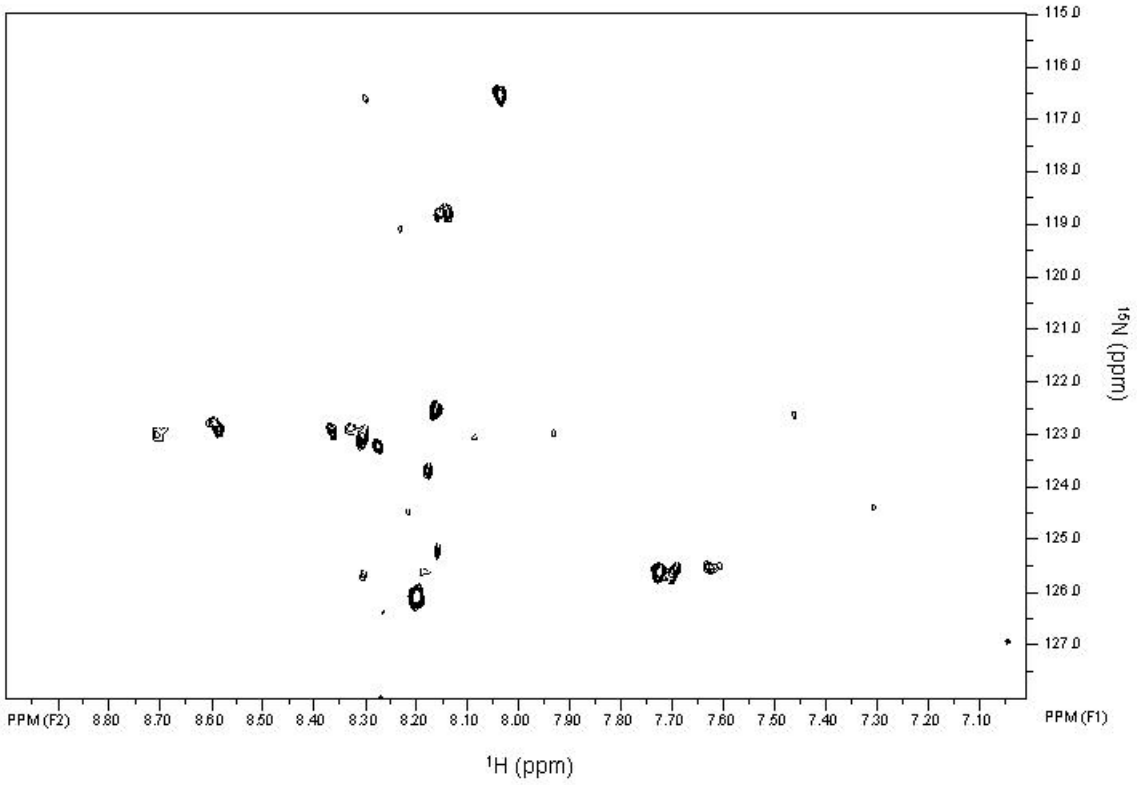
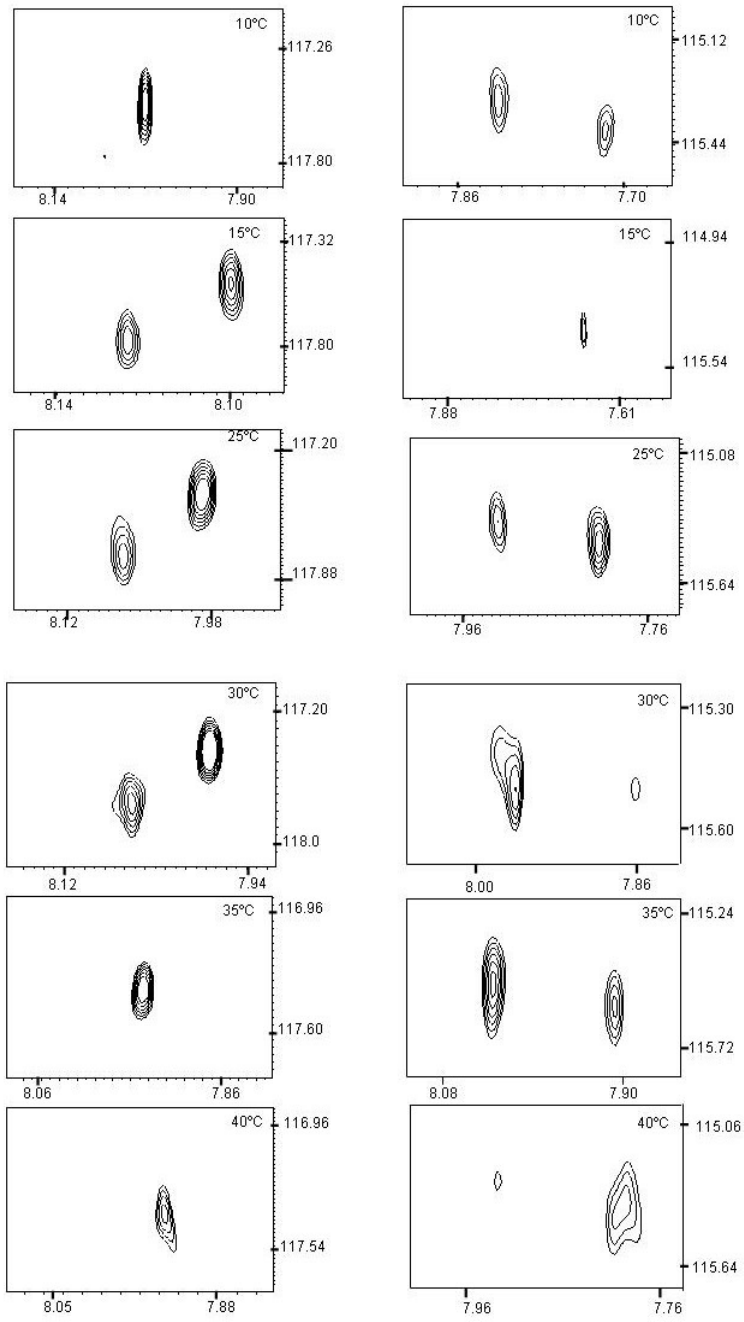


Figure 3.27: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.6mM ^{15}N-Met M168L *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

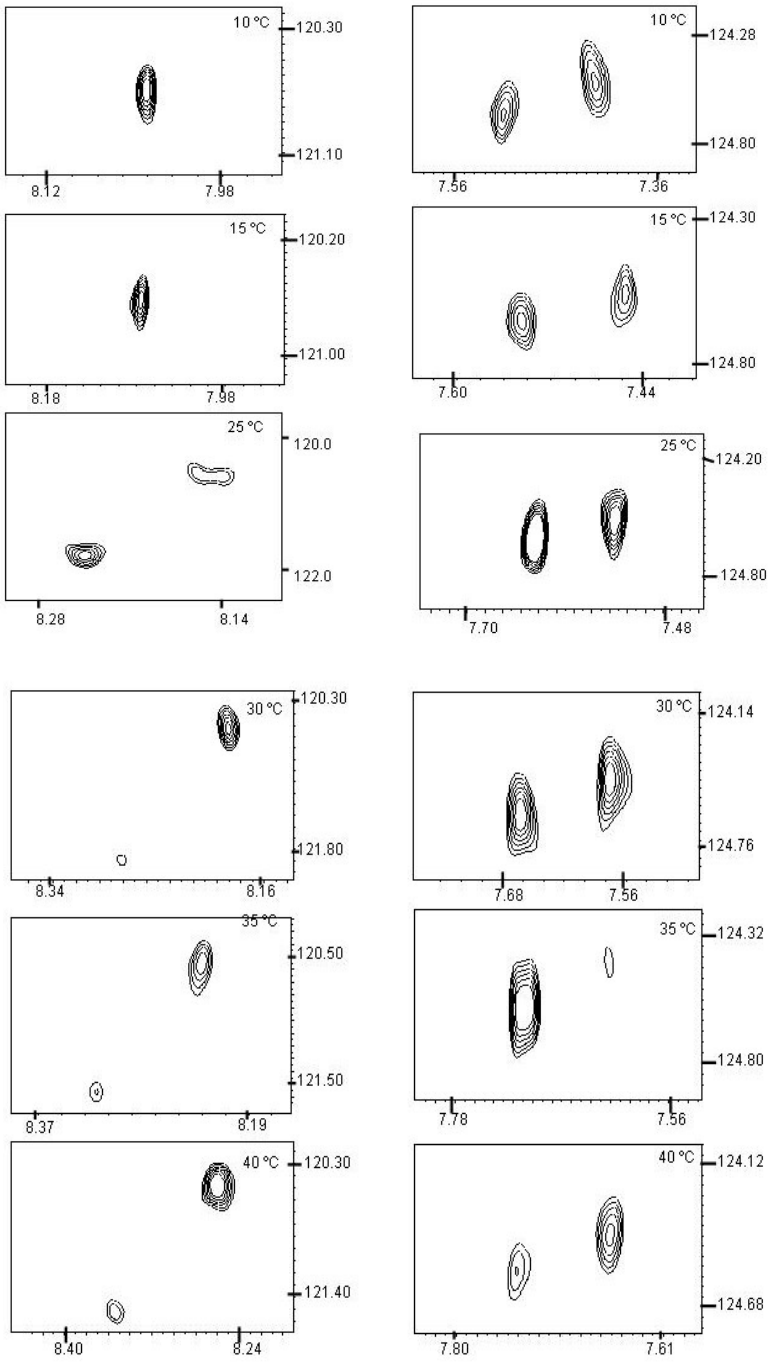
The results from both mutants M274L and M168L are confusing; the completely different NMR spectra suggest that the mobility of the methionine residues has been greatly affected as a result of the mutations.

The presence of 16 cross-peaks out of a total of 18 methionine residues present in CS suggested the presence of multiple conformations in solution. One way to study the behaviour of these conformations is by temperature dependence studies [Hitchens, et al., 2001; Klein-Seetharaman, et al., 2002]. Temperature dependence studies, in the range of 5-40 ºC, were carried out on ^{15}N-Met labeled *E. coli* CS. The maximum temperature was restricted to 40 ºC as the protein starts to denature significantly at higher temperatures [Ayed, et al., 1999]. As the temperature was varied, the relative intensities of a few peaks changed together with the nearest neighbour peak within a chemical shift difference of ~ 0.1ppm in the ^{1}H dimension. This suggests that these pairs of peaks which are in the chemical shift range of ~0.1 ppm of each other, represents multiple conformations of the same residue. The changes observed in the intensities of the cross-peaks were reversible. Control spectra were taken at 25ºC after lowest and highest temperatures. Figure 3.28 shows four pairs of cross-peaks representing two conformations that are in slow exchange for the same methionine residue and one cross peak which might be involved in intermediate exchange to give a broad peak.

Pair 1 Pair 2


10°C
117.26
117.80
8.14
7.90
10°C
115.12
115.44
7.86
7.70
15°C
117.32
117.80
8.14
8.10
15°C
114.94
115.54
7.88
7.61
25°C
117.20
117.88
8.12
7.98
25°C
115.08
115.64
7.96
7.76
30°C
117.20
118.0
8.12
7.94
30°C
115.30
115.60
8.00
7.86
35°C
116.96
117.60
8.06
7.86
35°C
115.24
115.72
8.08
7.90
40°C
116.96
117.54
8.05
7.88
40°C
115.06
115.64
7.96
7.76

Pair 3 Pair 4


10 °C
120.30
121.10
8.12
7.98
10 °C
124.28
124.80
7.56
7.36
15 °C
120.20
121.00
8.18
7.98
15 °C
124.30
124.80
7.60
7.44
25 °C
120.0
122.0
8.28
8.14
25 °C
124.20
124.80
7.70
7.48
30 °C
120.30
121.80
8.34
8.16
30 °C
124.14
124.76
7.68
7.56
35 °C
120.50
121.50
8.37
8.19
35 °C
124.32
124.80
7.78
7.56
40 °C
120.30
121.40
8.40
8.24
40 °C
124.12
124.68
7.80
7.61

Peak 9

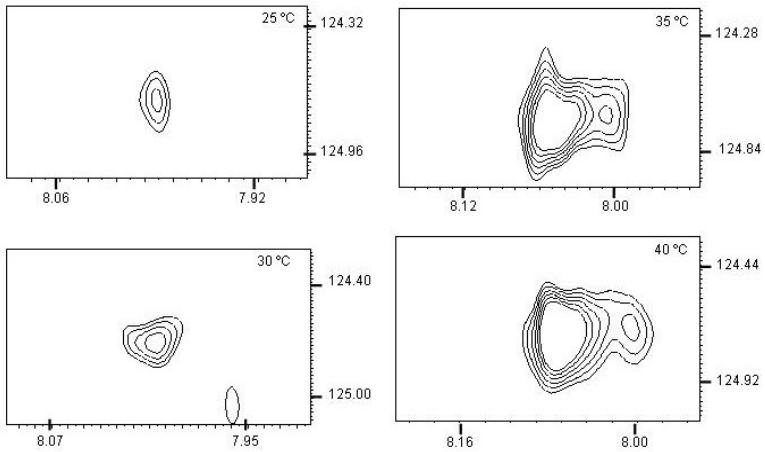
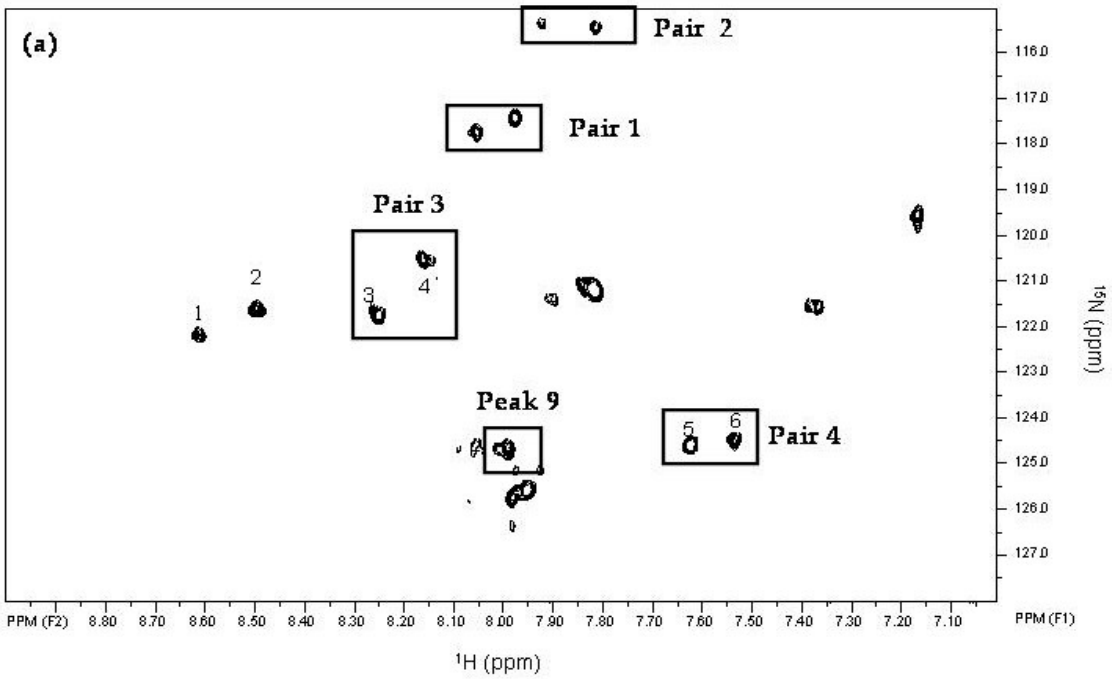


Figure 3.28: Regions of two-dimensional TROSY spectrum of ^{15}N-Met *E. coli* CS at different temperatures. Each inset contains two peaks which are in conformational exchange with each other.

For two pairs of cross-peaks (Peaks 1&2; Peaks 5&6 in Figure 3.29a), values of ΔH = 16±1kJ/ mole; 42±10 kJ/ mole and ΔS = 450±5 J/mole. K; 410±6 J/ mole.K respectively were extracted from the Arrhenius plots (Example of calculations are shown in Table 3.3 and Figure 3.44). The method for calculating the thermodynamic parameters relies on the ratio of intensities of the two peaks. Therefore if the two peaks were observable throughout the temperature range, the ratio of the intensities could be calculated. In cases where the two cross-peaks in a pair merge to give a single peak, sufficient data was not available to do the calculations.

The changes in the ^1H chemical shifts of six cross-peaks with temperature are shown in Figure 3.29b. Where points are missing, the peak was not observed at the lower temperatures. Amide proton temperature coefficients are measured by the slopes of the curves in Figure 3.29b. Amide proton temperature coefficients have been correlated with the presence of hydrogen bonds, with the general rule that gradients more positive than -4.6ppb/ K arise from residues involved in hydrogen bonds [Cierpicki, et al., 2001]. Proton temperature coefficients for all cross-peaks in Figure 3.29b are more positive than -4.6ppb/ K, indicating that the residues being studied are all hydrogen bonded.



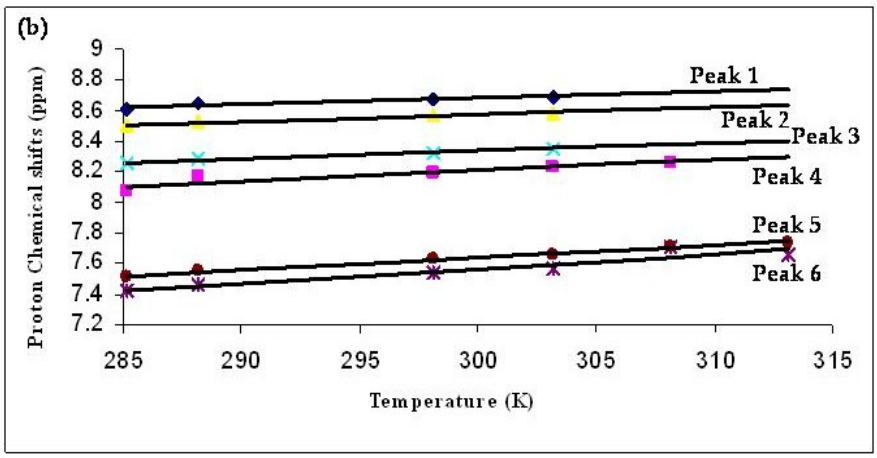


Figure: 3.29: (a) Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of ^{15}N-Met labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours. (b) Chemical shift changes of Methionine cross-peaks with temperature.

^{15}N Glycine *E. coli* CS

Out of a total of 28 glycine residues in *E. coli* CS, two are present at the beginning of the mobile loop (G265 & G266), one is in the helical linker (G330) with high B-factors and six glycine residues are in the N-terminal region (Figure 3.30). In the HSQC spectrum of ^{15}N-Gly *E. coli* CS, ten well resolved cross-peaks were observed with proton chemical shifts between 8.0-8.3 ppm (Figure 3.31). The intensity of the cross-peaks varies in the spectrum. Cross-peaks such as 8 and 9 in Figure 3.31 show very low intensity as compared to the other peaks. These broad peaks reflect their involvement in the conformational exchange over intermediate time scale.

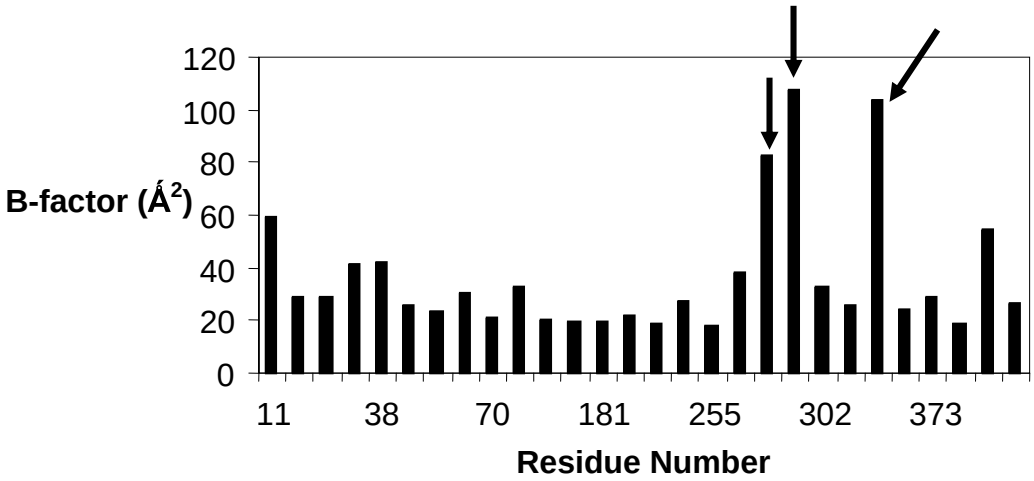


Figure 3.30: B-factors of backbone nitrogens for glycine residues in *E. coli* CS [Nguyen, et al., 2001]. The bars marked with arrows correspond to G265, G267 at the beginning of the mobile loop and G330 in the helical linker region.

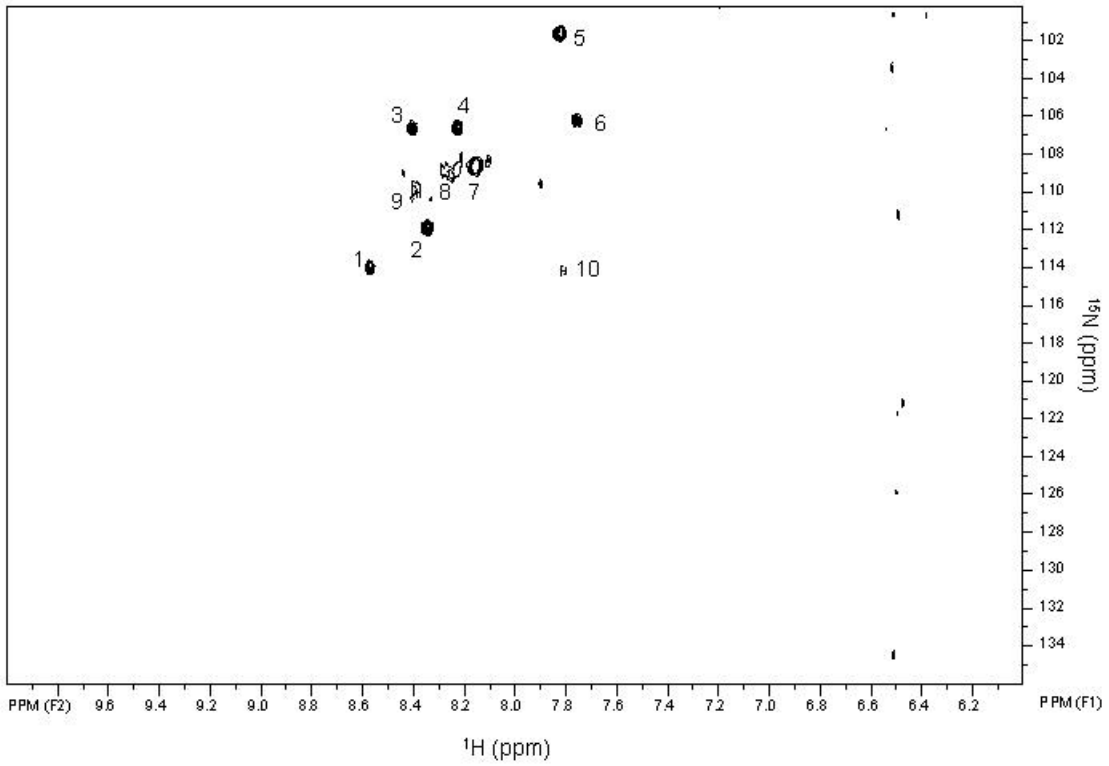


Figure 3.31: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.0mM ^{15}N-Gly labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

On addition of substrates in the presence of 0.1 M KCl, there was change in the chemical shifts of two peaks marked with boxes (Peak 3 & Peak 7) in Figure 3.32. There was a further decrease in the intensity of two broad peaks (Peak 8 & Peak 9) indicating that these peaks which seem to be in conformational exchange are also affected, either in their mobility.

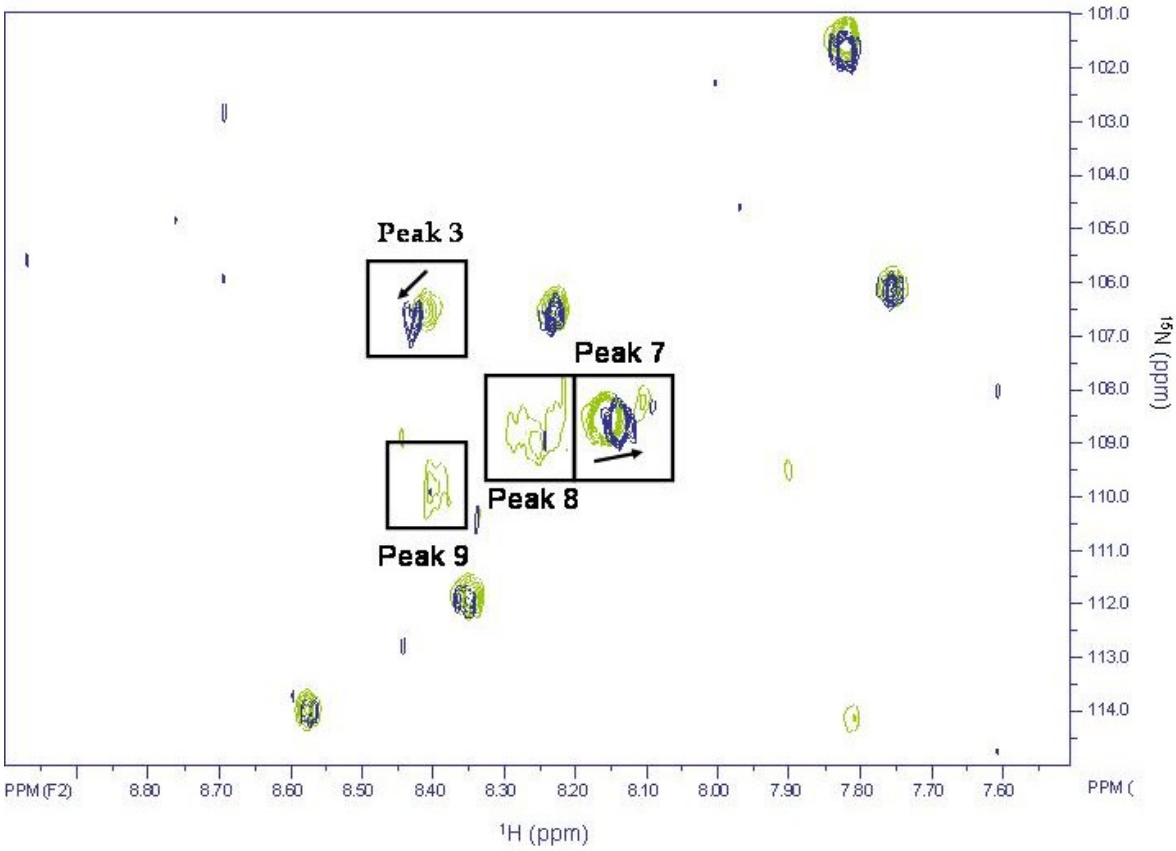


Figure 3.32: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Gly labeled *E.coli* CS with (shown in blue) and without substrates (shown in green), at 25ºC at a ^{1}H frequency of 600 MHz. Cross-Peaks affected by substrates are marked with boxes and follow the same number scheme as Figure 3.31

$^{15}N_2$-Tryptophan *E. coli* CS

For specific labeling with ^{15}N-tryptophan, both the alpha and the indole nitrogens contained the heavy isotope. In the HSQC spectrum of $^{15}N_2$-Tryptophan *E. coli* CS, a broad peak envelope was observed, with proton chemical shifts of ~9.8-10.2 ppm corresponding to the indole nitrogen of tryptophan (Figure 3.33). *E. coli* CS contains three tryptophans, W260, W391 and W395, only W260 is close to the mobile loop. The B-factors of backbone and side chain nitrogens are quite low suggesting rigid environment as compared to the residues in the mobile loop. The broad peak envelope was not resolved enough to tell how many tryptophans are contributing to the signal. No cross-peaks were observed in the α-N region between ~ 7-9 ppm suggesting that the backbone nitrogens of the three tryptophans are in rigid environments.

Addition of substrates to $^{15}N_2$-tryptophan labeled *E. coli* CS led to disappearance of the signal from indole nitrogen (Figure 3.34). The peak envelope was observable again with the same chemical shifts once the sample was dialyzed into phosphate buffer to take out the substrates (Figure 3.35). Thus, binding of substrates causes the tryptophan side chains to become more rigid.

To gain insight into which tryptophans contribute to the HSQC spectrum, I used the mutant protein W260A, which had been previously made and studied [Ayed, et al., 1999]. The W260A mutant clone was

subcloned into pET24b^{+} as usual for the NMR experiments. In the two-dimensional HSQC spectrum of the $^{15}N_2$Trp W260A mutant, a peak and a shoulder in the broad peak envelope have disappeared, as compared to the HSQC spectrum of wild type *E. coli* CS (Figure 3.36).

My NMR results suggest that any conformational changes in the active site region, affect W260 and other tryptophans W391 and/or W395.

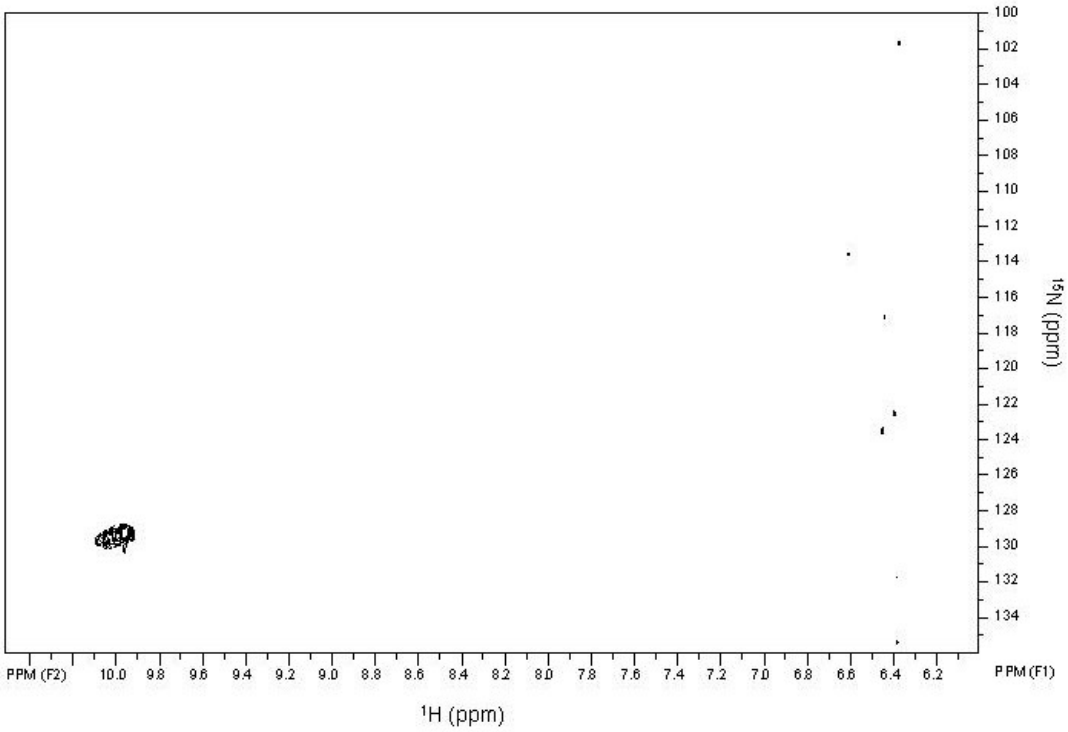


Figure 3.33: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 1.5mM ^{15}N-Trp labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

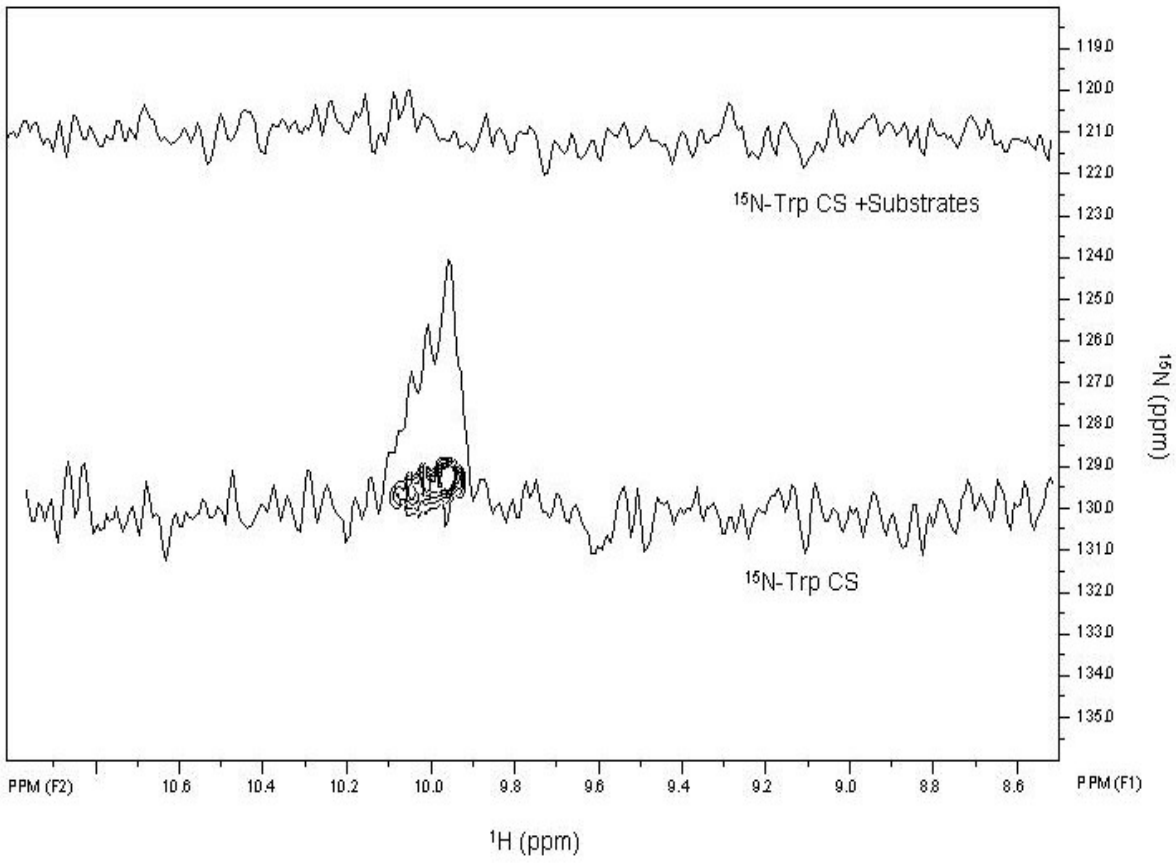


Figure 3.34: Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra of $^{15}N_2$-Trp labeled *E. coli* CS showing the effect of substrates, at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

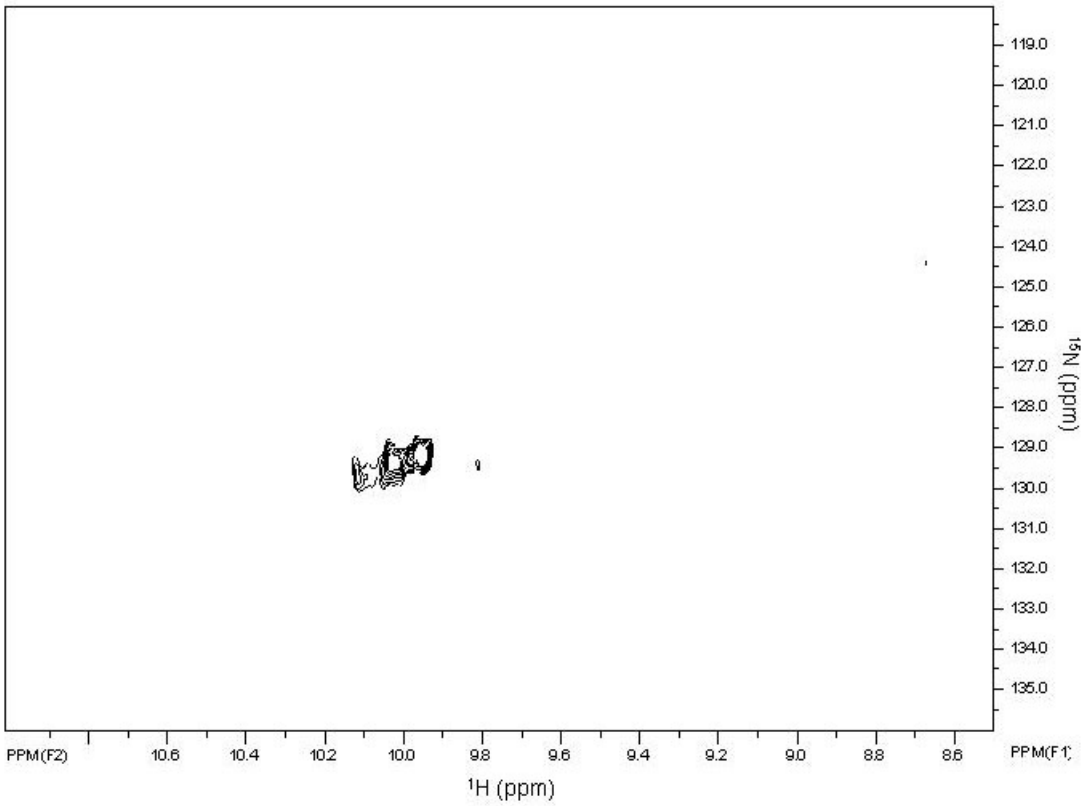


Figure 3.35: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Trp labeled *E. coli* CS after overnight dialysis to take out substrates, at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

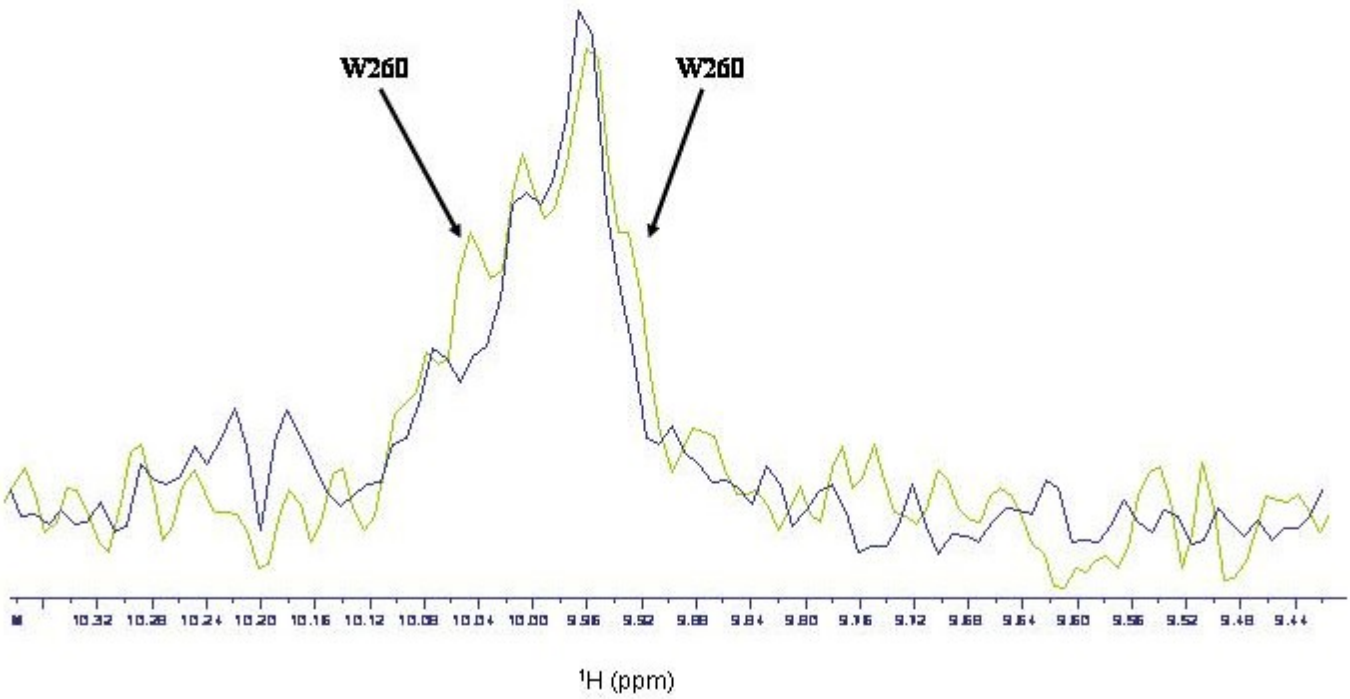


Figure 3.36: Overlay of 1D slices of HSQC spectrum of $^{15}N_2$-Trp labeled *E. coli* CS wild-type spectrum (shown in green) and W260A mutant (shown in blue).

^{15}N-Leucine *E. coli* CS

In the HSQC spectrum of ^{15}N-leucine *E. coli* CS, twelve peaks were observed with proton chemical shifts of 7.4-9.0 ppm (Figure 3.37). Out of a totals 35 leucine residues, two residues L272, L275 are present in the mobile loop and one residue L335 is in the helical linker, are characterized by high B-factors. There are six leucine residues in the N-terminal region, the B-Factor of L7 being the highest among the six residues (Figure 3.38).

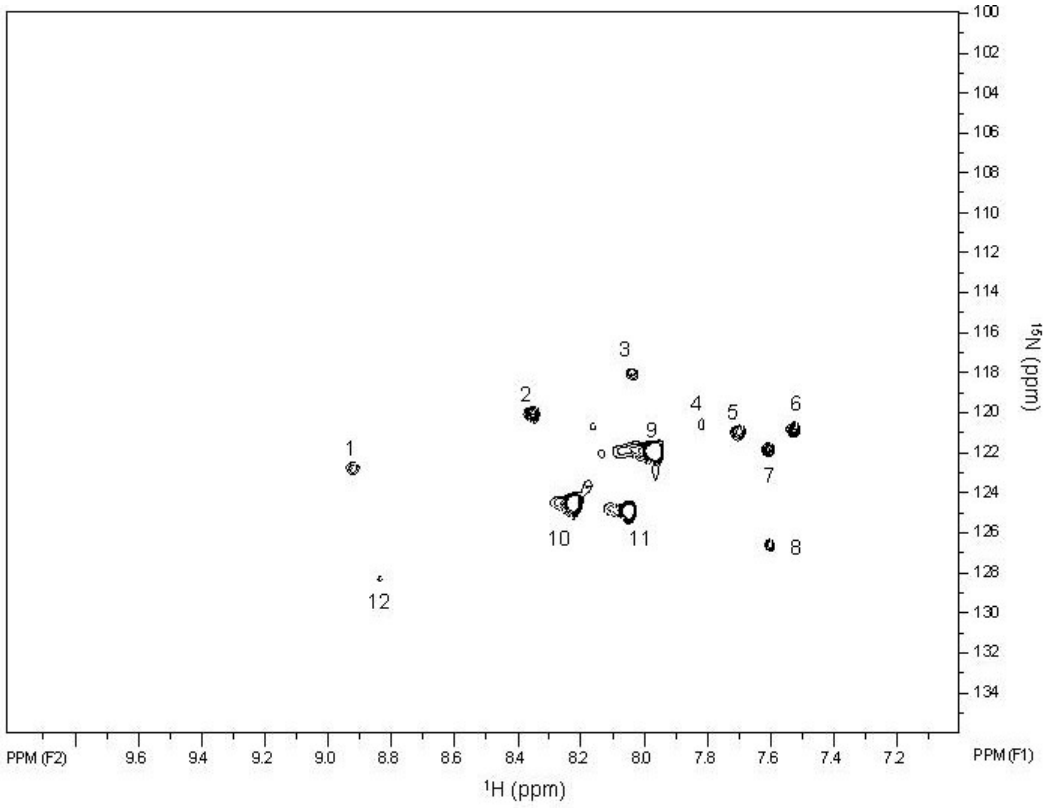


Figure 3.37: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of 0.9mM ^{15}N-Leu labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

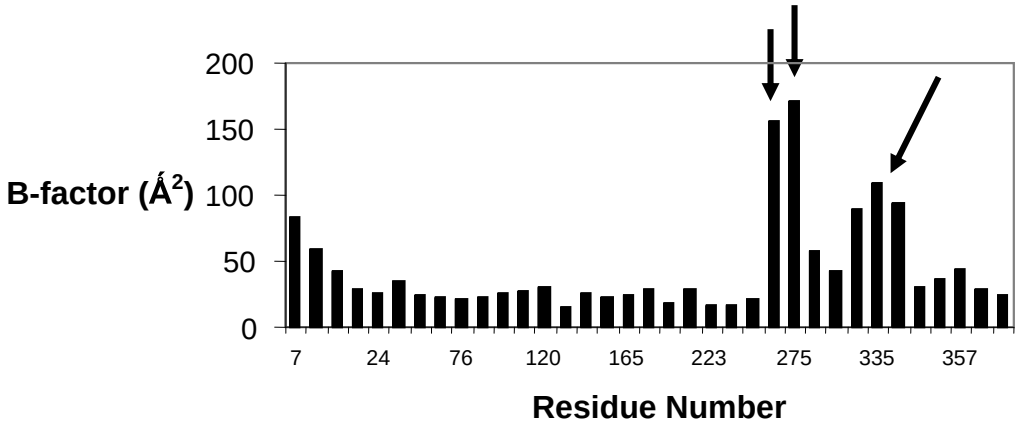


Figure 3.38: B-factors of backbone nitrogen for leucine residues in *E. coli* CS [Nguyen, et al., 2001]. The bars marked with arrow correspond to L272, L275 and L335 from the mobile loop and helical linker region.

On addition of substrates small changes were observed in the cross peaks marked with box in Figure 3.39.

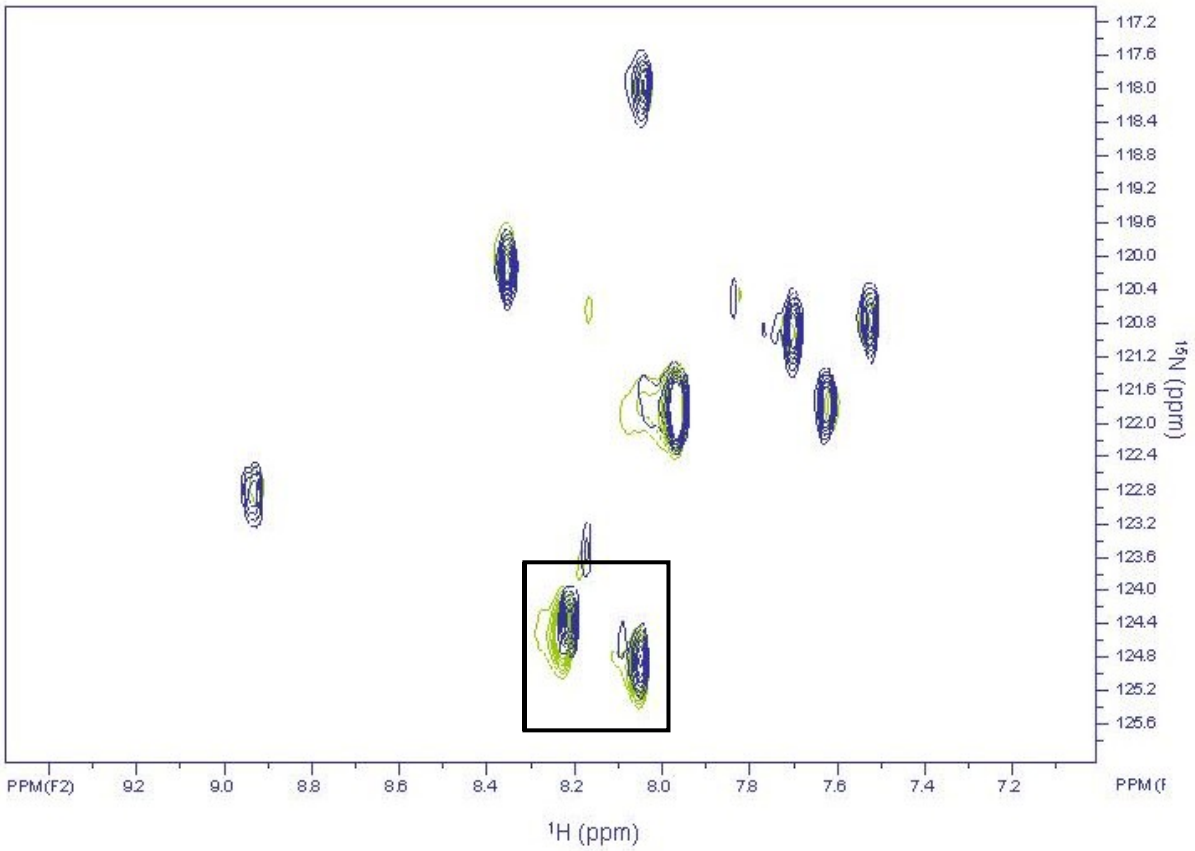


Figure 3.39: Two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Leu labeled *E.coli* CS with (shown in blue) and without (shown in green) substrates, at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time for each spectrum was 15 hours.

The HSQC spectrum of ^{15}N-leucine labeled *E. coli* CS was studied in the temperature range of 5ºC-35ºC. All the cross-peaks were clearly observable in the whole temperature range studied here, and exhibited linear downfield proton shifts at higher temperatures (Figure 3.40, 3.41).

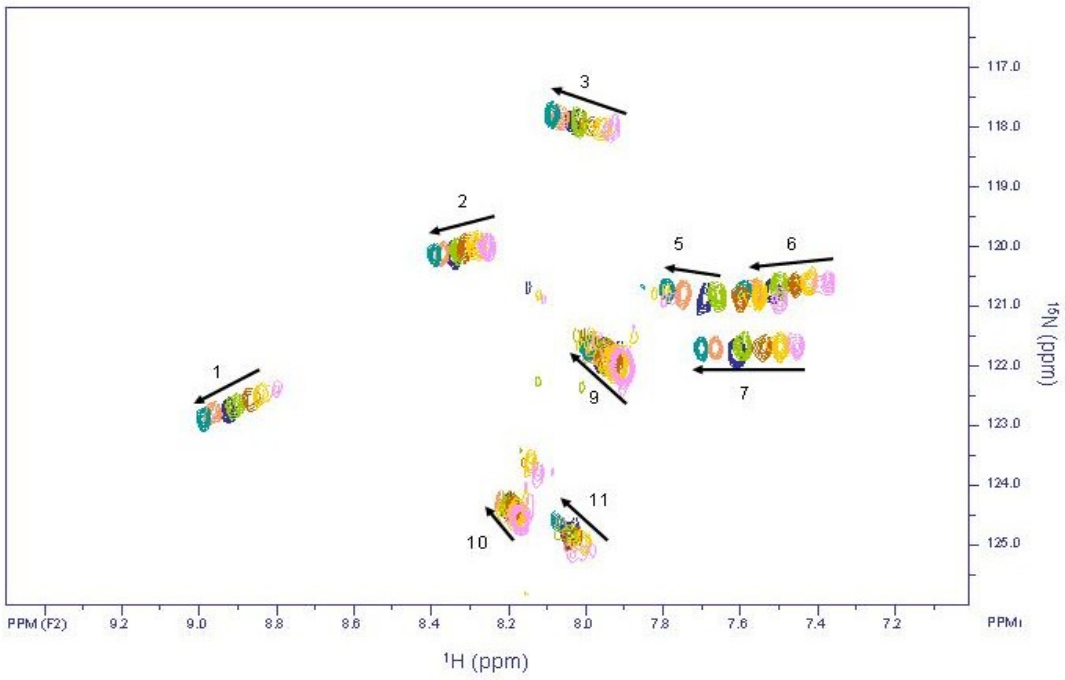


Figure 3.40: Overlay of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-leu labeled *E. coli* CS at different temperatures, measured at a ^{1}H frequency of 600 MHz. The arrow denotes the movement of the cross-peaks with increasing temperatures from 5-35 ºC in 5ºC intervals. Total acquisition time for each spectrum was 15 hours. Peaks are numbered based on Figure 3.37.

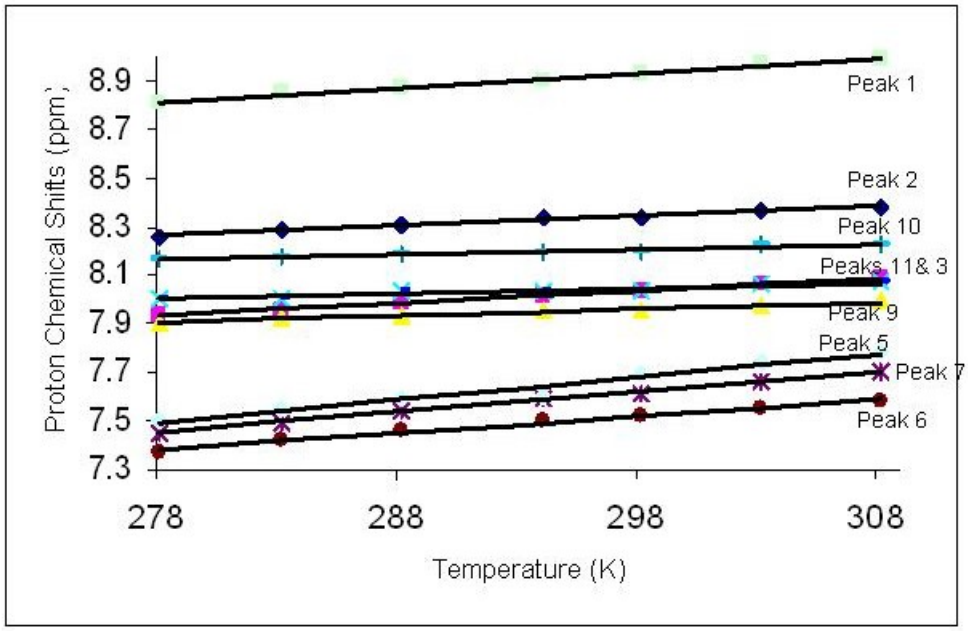


Figure 3.41: Linear change in the proton chemical shift of cross-peaks with increase in temperatures. Peaks are numbered based on Figure 3.37

A few cross-peaks in ^{15}N-Leu *E. coli* CS were observed only at lower temperatures, 5-10ºC, but were not present at higher temperatures suggesting that the conformational exchange enters the slow exchange regime at the lower temperatures (Figure 3.42).

The cross-peaks with ^{1}H,^{15}N chemical shifts of (8.11 ppm, 124.38 ppm) and (8.039ppm, 124.48ppm) varied in intensity as the temperature was changed (Figure 3.43) indicating the presence of conformational exchange between them. At the lowest temperature studied here one of the conformational states is much more populated, whereas at intermediate temperature both conformational states are equally populated, and exchanging slowly on the NMR time scale. At 35ºC the two peaks merged together to give a single peak indicating that the conformational exchange now enters the fast exchange regime.

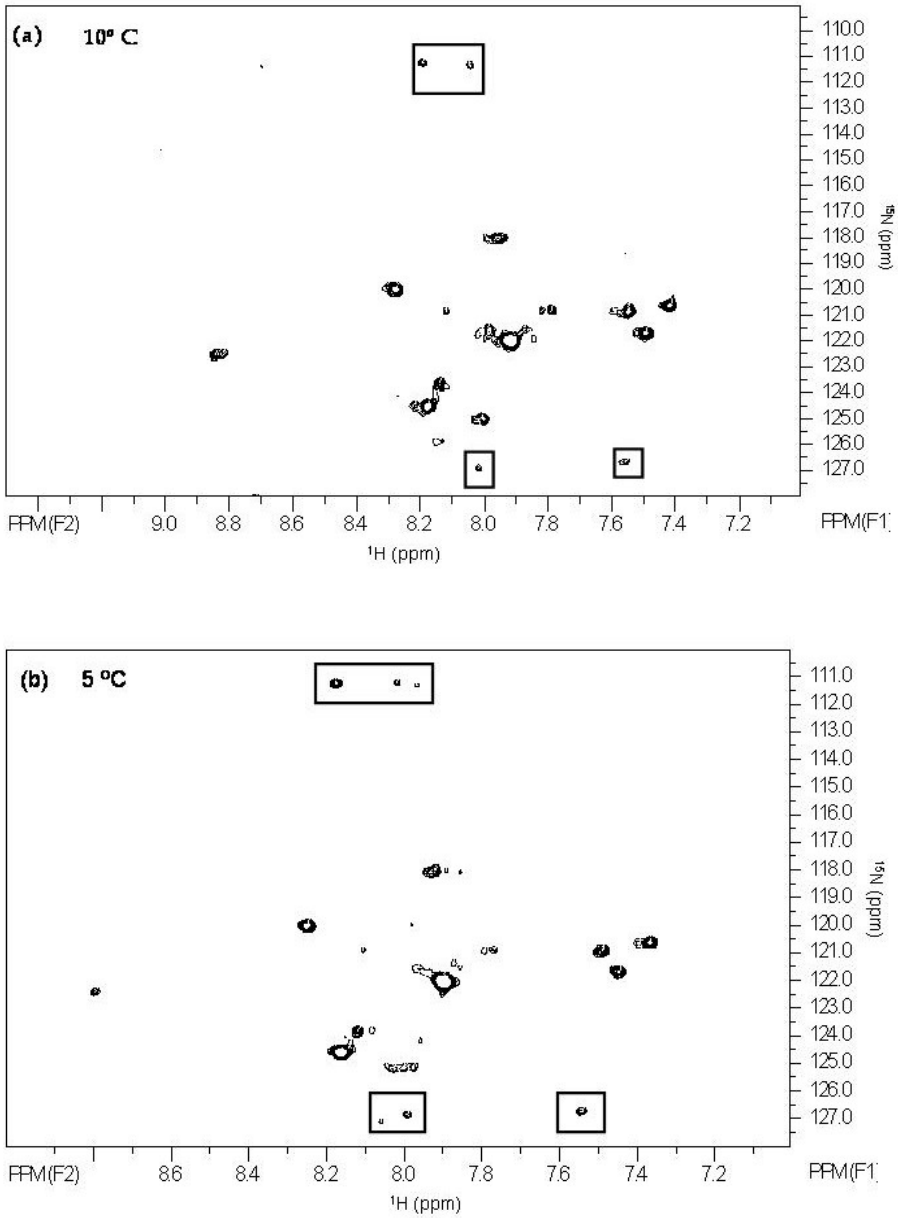


Figure 3.42: Two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Leu labeled *E. coli* CS at 10 ºC and 5 ºC measured at a ^{1}H frequency of 600 MHz. The cross-peaks marked in boxes are only observed at these temperatures and not at higher temperatures. Total acquisition time for each spectrum was 15 hours.

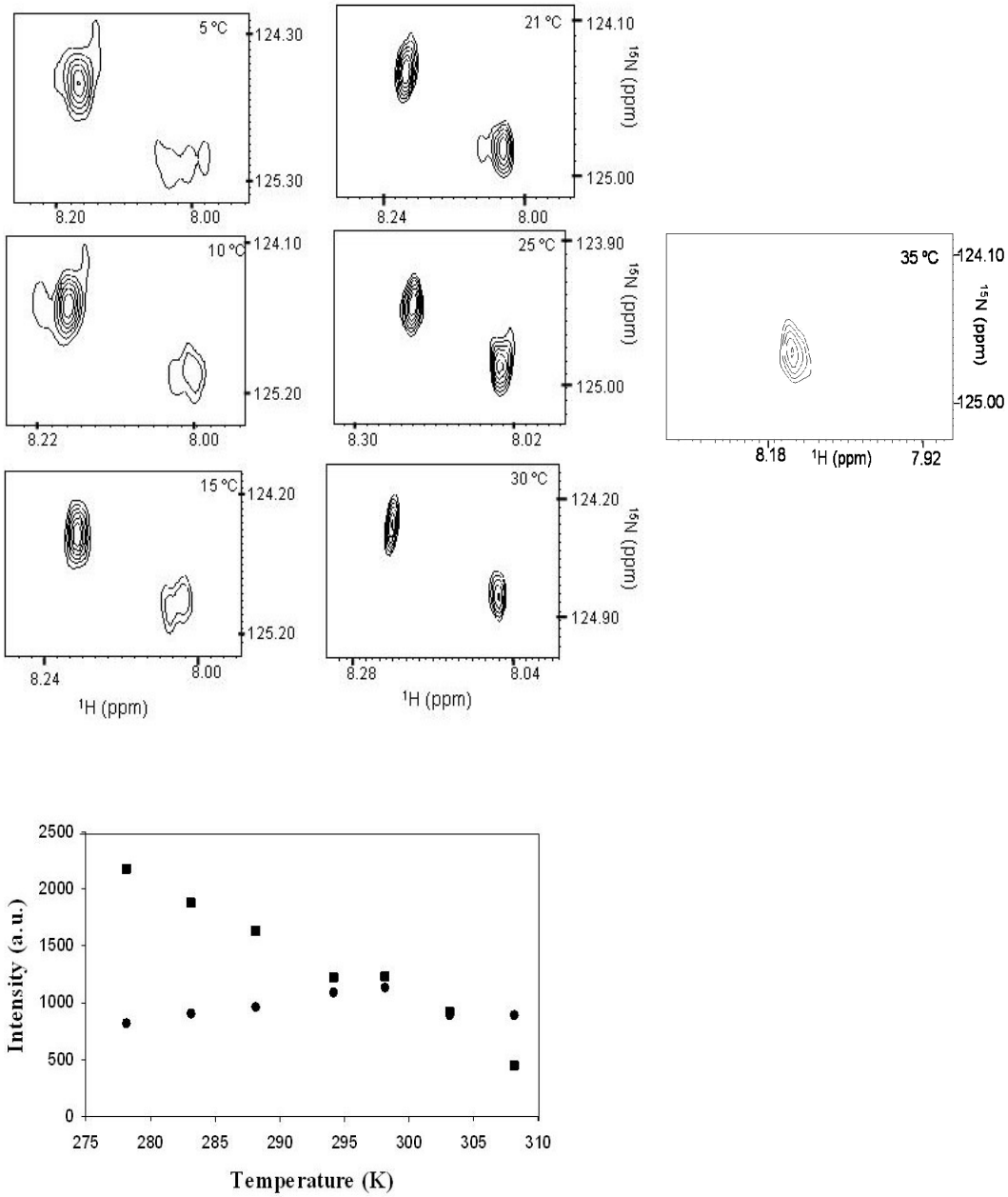


Figure 3.43: Cross-peaks showing conformational exchange between two peaks at different temperatures. The graph shows the variation in intensity of the two peaks as temperature is varied. • denotes peak A ■ denotes peak B

Table 3.3 shows the intensities of the two cross-peaks in conformational exchange in Figure 3.43, as an example of the data used for determining thermodynamic parameters. ΔH and ΔS were calculated as described in Experimental Procedures in section 3.4. Assuming a linear relationship for the Arrhenius plot (Figure 3.44), ΔH and ΔS were determined with values of 28±4 kJ/ mole and 498±14 J/mole. K respectively for the two peaks in conformational exchange at 25ºC.

Table 3.3: Intensities of cross-peaks in conformational exchange as measured directly from the HSQC spectrum of ^{15}N-Leu *E. coli* CS at different temperatures

1/T (K)	Intensity (A)	Intensity (B)	Normalized Intensity(A) N_A	Normalized Intensity (B) N_B	N_B/N_A	$LN(N_B/N_A)$
0.00359	2176	821	1	1	1	0
0.00353	1866	902	0.8575	1.0986	1.2811	0.2477
0.00347	1633	952	0.7504	1.1595	1.5451	0.4351
0.00339	1219	1082	0.5602	1.3179	2.3525	0.8555
0.00335	1232	1124	0.5661	1.3690	2.4180	0.8829
0.00329	919	881	0.4223	1.0730	2.5408	0.9324
0.00324	436	885	0.2003	1.0779	5.3798	1.6826

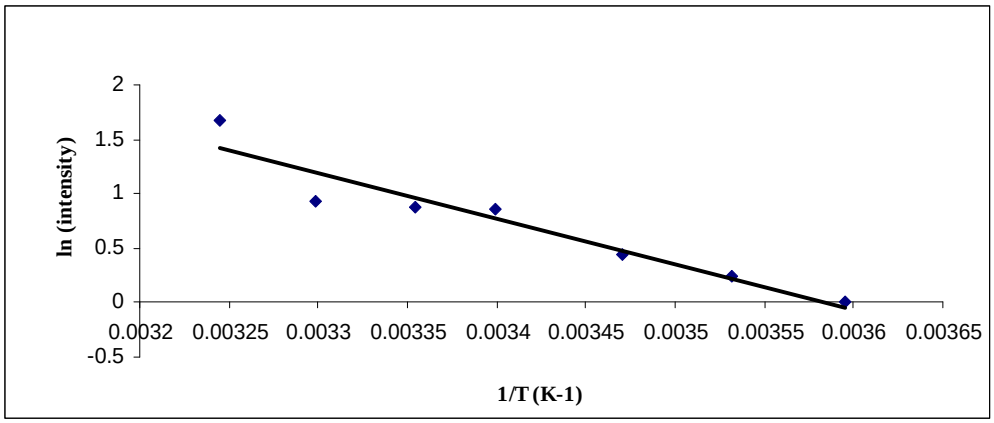


Figure 3.44: Plot of ln (N_B/N_A) vs 1/ T to calculate ΔH and ΔS of the cross-peaks in slow conformational exchange (Table 3.3). Intensities were obtained directly from the HSQC spectra at different temperatures (Table 3.37).

Mutation L275A, which had been made previously [Ayed, et al., 1997], was subcloned into the pET 24b$^+$ vector as usual for NMR experiments. The two-dimensional [^{15}N, ^{1}H] HSQC spectra of wild type and L275A CS, both labeled with ^{15}N-leucine are shown in Figure 3.45. L275 belongs to the mobile loop. The comparison of the two spectra enables the identification of the L275 peak owing to its absence in the mutant spectrum. Two cross-peaks are missing in the mutant spectrum which we know from temperature dependence studies corresponds to a single residue exchanging over multiple conformations. The residual peaks corresponding to L275, in the mutant spectrum are due to the presence of small amount of labeled WT CS from the host strain (as determined by mass spectrometry on a different sample). Another peak affected as a result of mutation is marked with an arrow in Figure 3.45. The probable reason might be the reduction in the mobility or change in the exchange dynamics of the residue as a result of mutation. The remaining peaks in the L275A spectrum were observed at the same chemical shifts for WT CS indicating no change in their environment as a result of mutagenesis.

^{15}N-lysine *E. coli* CS

In the TROSY spectrum of ^{15}N-Lys *E. coli* CS, fourteen well resolved cross-peaks with proton chemical shifts between 7.0- 9.0 ppm were observed (Figure 3.46). There was a significant change in the chemical shifts of three cross-peaks on addition of substrates (Figure 3.47).

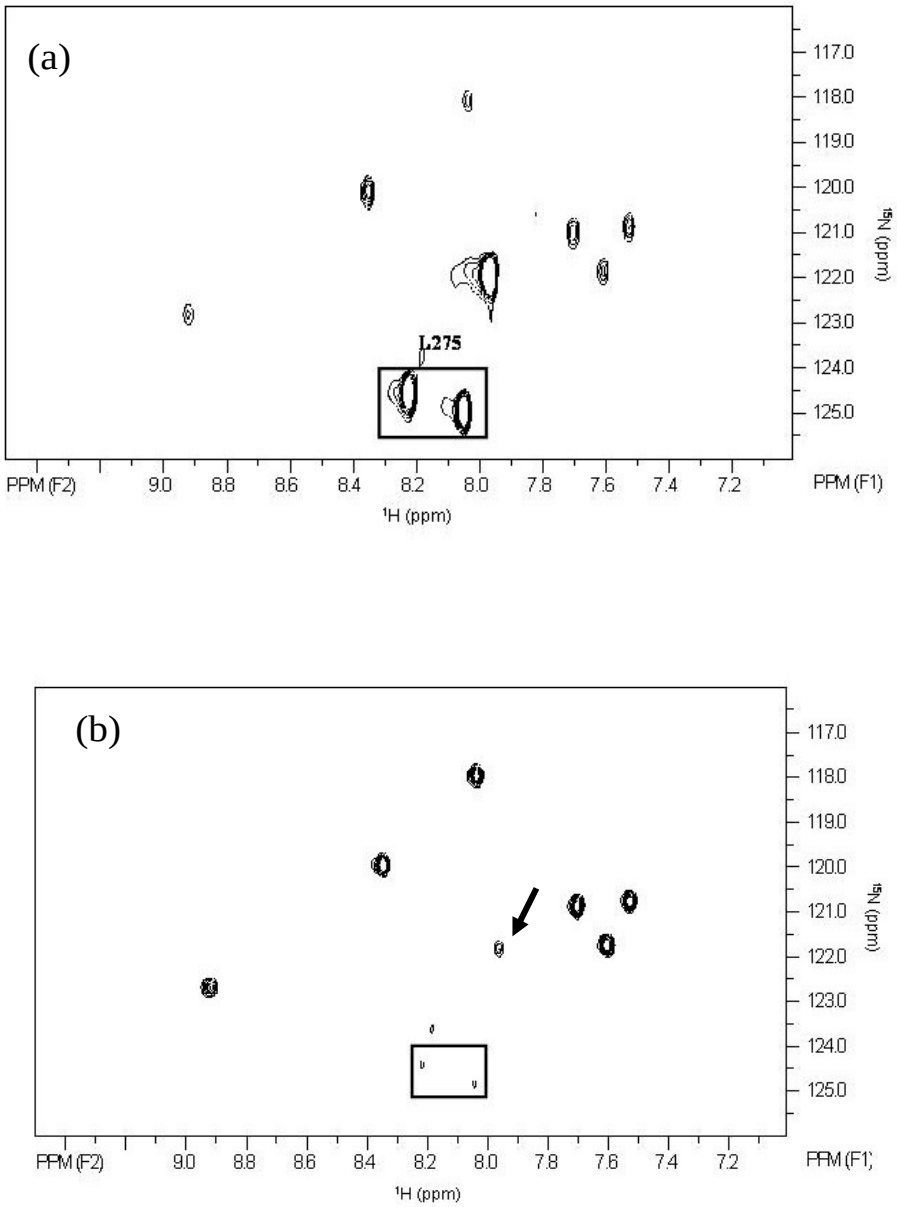


Figure 3.45: Comparison of two-dimensional [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-Leu labeled *E. coli* CS (a) wild type (b) L275A mutant measured at 25 ºC at a ^{1}H frequency of 600 MHz. The two cross-peaks attributed to L275 are marked with box. Another peak affected by substrate is marked with an arrow. Total acquisition time for each spectrum was 15 hours.

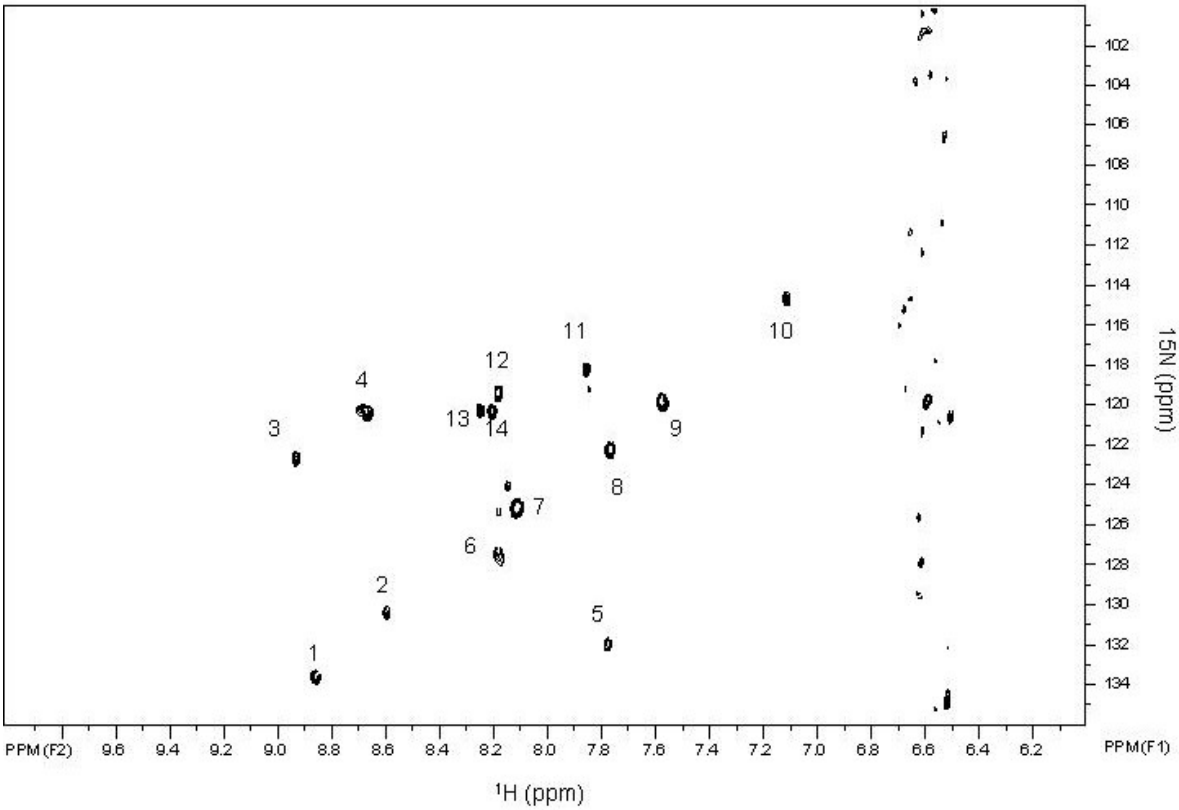


Figure 3.46: Two-dimensional [^{1}H, ^{15}N] TROSY spectrum of 1.0mM ^{15}N-Lys labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

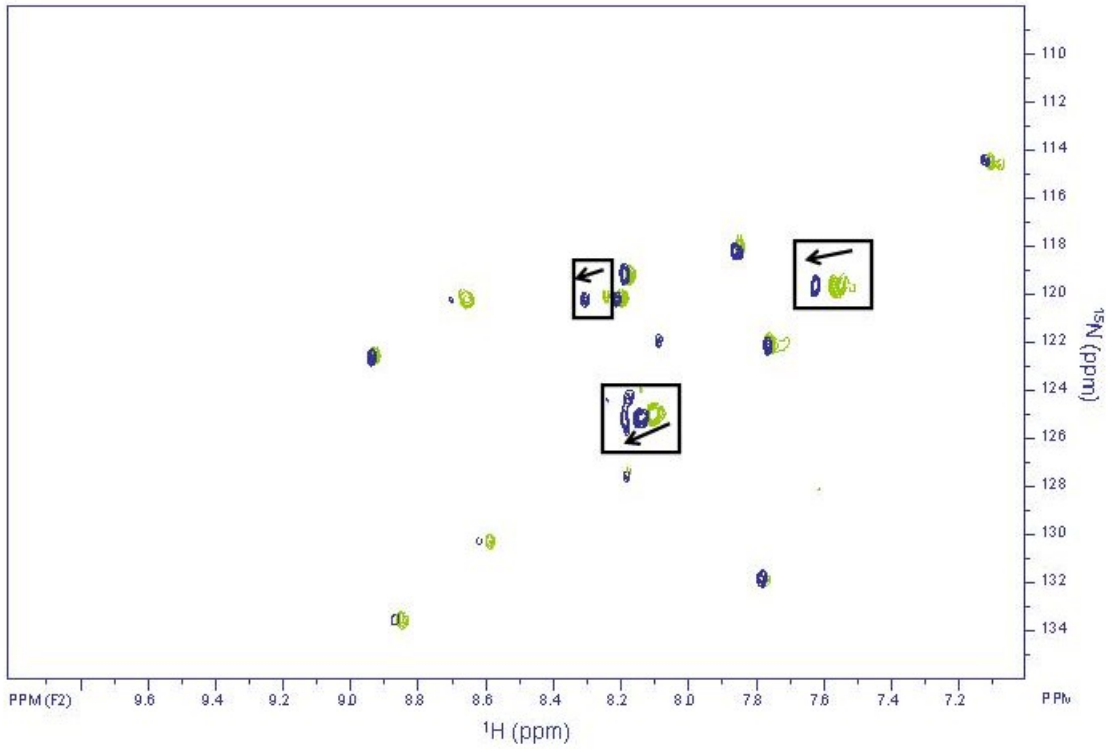


Figure 3.47: Overlay of two-dimensional [^{1}H, ^{15}N] TROSY spectra of ^{15}N-Lys labeled *E. coli* CS with substrates (shown in blue) and without substrates (shown in green), at 25ºC at a ^{1}H frequency of 600 MHz**.** The arrows indicate the downfield shifts of three cross-peaks after addition of substrates.

Out of total 23 lysine residues in *E. coli* CS , four lysine residues in the mobile loop(K273, K282, K292, K294), one in the helical linker (K332) and four in the N-terminal region (K4, K6, K21, K37). The B-factors of backbone nitrogen for lysine residue in *E. coli* CS are shown in Figure 3.48.

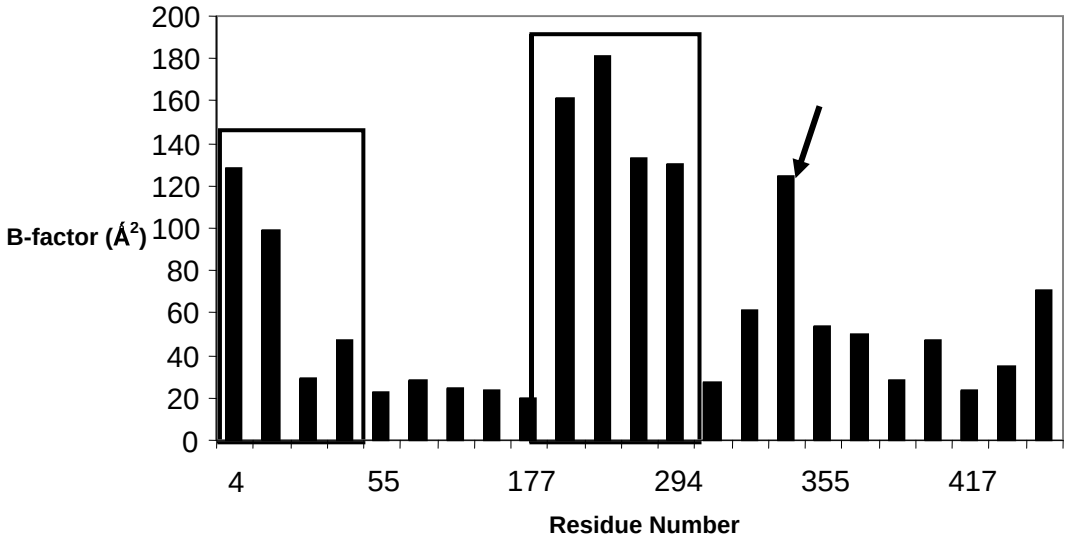


Figure 3.48: B-factors of backbone nitrogen for lysine residues in *E. coli* CS [Nguyen, et al., 2001]. The bars marked with boxes correspond to K4, K6, K21, K37 from the N-terminal region and K273, K282, K292, K294 from the mobile loop. The bar marked with an arrow corresponds to K332 from the helical linker region.

The TROSY spectrum of ^{15}N-lysine labeled truncated *E. coli* CS (first 40 residues removed) shows cross-peaks with identical chemical shifts as WT except for two peaks missing in the truncated spectrum (Figure 3.49). The two missing lysine peaks may probably be attributed to Lys-4 and Lys-6, a region which shows greater flexibility as shown by their high B-factors as compared to the other two lysine residues in the N-terminal region (Figure 3.48). The two peaks contributed from the N-terminal region vary from each in their intensities and ^{1}H chemical shifts (~ 0.9 ppm) indicating that these are not originating from a single residue. Previous results with ^{15}N-Met and ^{15}N-Leu showed that multiple signals are observed in the ^{1}H chemical shifts range of ~0.1ppm (Figure 3.28&3.43). The difference in the intensities can be due to different environments of the two lysines in the N-terminal, K6 being more solvent exposed than K4 might give rise to weak cross-peak due to exchange with water. Three new cross-peaks were also observed in the TROSY spectrum of ^{15}N-Lys truncated *E. coli* CS, probably from a lysine residue which has become more mobile as a result of truncation of the N-terminal.

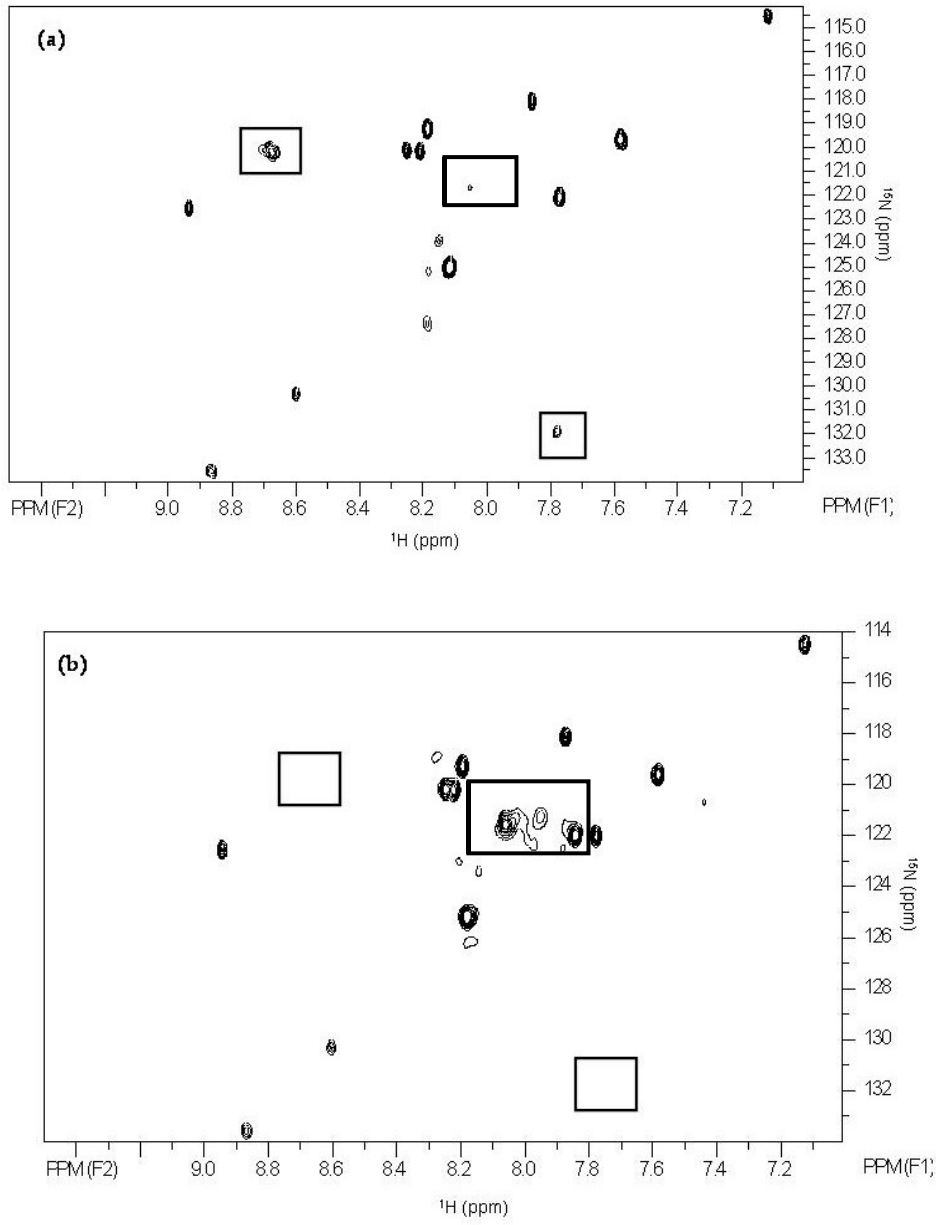


Figure 3.49: Comparison of two-dimensional [^{1}H, ^{15}N] HSQC spectra of ^{15}N-Lys labeled *E. coli* CS (a) wild type (b) Truncated mutant. Two missing peaks and new peaks as compared to the wild-type are marked with boxes in the truncated spectrum. Spectra were taken at 25ºC at a ^{1}H frequency of 600MHz. Total acquisition time for each spectrum was 15 hours.

^{15}N-His *E. coli* CS

The X-ray crystal structure of *E. coli* CS shows that one of the active site residues, H264, is not in the correct position to form the binding site. The movement of H264 is likely a key part of substrate binding and hence any information about the mobility is especially interesting.

Out of total 15 histidine residues in *E. coli* CS, only one residue H283 is present in the mobile loop. No histidine residue is present in either the N-terminal region or the helical linker region. The B-factors of the backbone nitrogen for all the histidine residues are quite low except H283 which in the mobile loop (Figure 3.50). In the HSQC spectrum of ^{15}N-His *E. coli* CS, only very weak signals were observed between proton chemical shifts of 8.5 ppm and 7.9 ppm (Figure 3.51), although isotopic incorporation of ^{15}N label as estimated by mass spectrometry was ~95%. The possible reason for the weak signals could be exchange broadening of the histidine residue in the mobile loop and very low mobility of others.

In order to observe strong signals in the HSQC spectrum of ^{15}N-His *E. coli* CS, NMR experiments were carried out in the presence of urea to loosen up the enzyme slightly. Conditions were chosen so that there was only a slight change in the hexamer-dimer equilibrium of enzyme toward smaller units and not complete denaturation. These conditions were chosen based on studies (Ayed, et al., 1999) done on unfolding of *E. coli* CS in the presence of

urea which showed that the presence of 1M urea does not cause significant denaturation of the enzyme.

The ^{15}N-His *E. coli* CS sample was monitored by two-dimensional NMR spectroscopy in the presence of 1M urea. Enzyme activity is decreased only slightly under the conditions used for the experiments. The two-dimensional HSQC spectrum of ^{15}N-His *E. coli* CS in the presence of 1M urea is shown in Figure 3.52. New cross-peaks were observed after only 24 hours of incubation at 25ºC. Three out of five weak peaks disappeared in the presence of urea. The possible reason for this might be the change in the exchange kinetics of the flexible histidine residues in the presence of urea or change in the dimer-hexamer equilibrium of the enzyme.

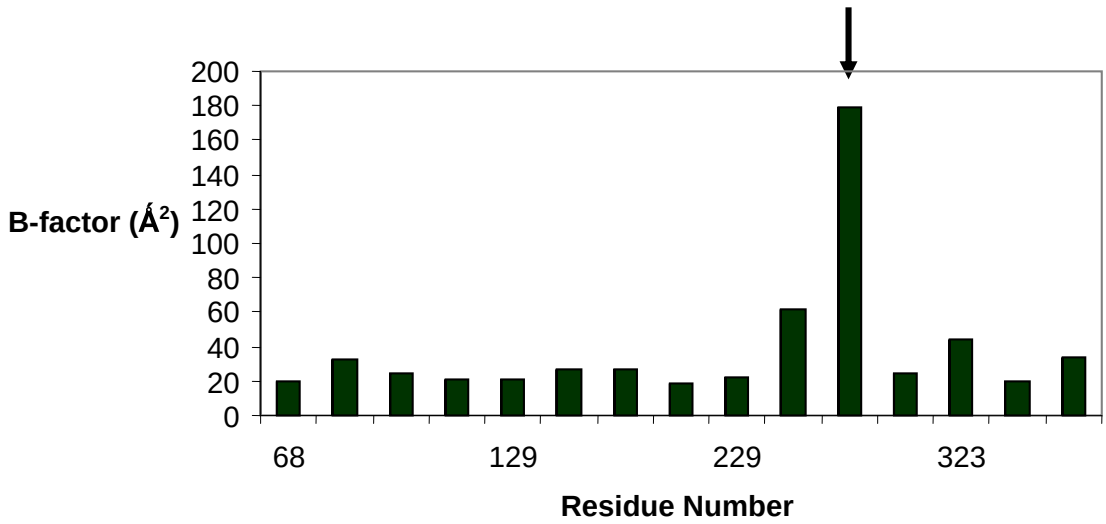


Figure 3.50: B-factors of backbone nitrogen for histidine residues in *E. coli* CS [Nguyen, et al., 2001]. The bar marked with an arrow corresponds to H264 of the mobile loop.

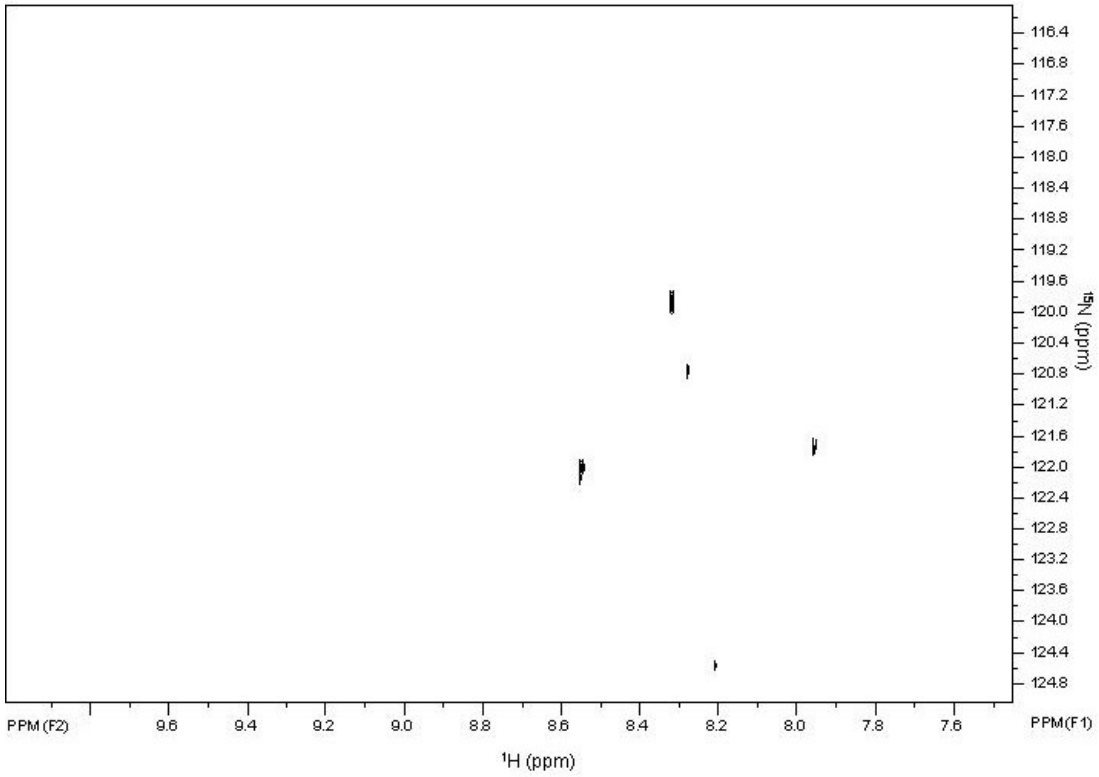


Figure 3.51: Two-dimensional [$^{1}H, ^{15}N$] HSQC spectrum of 0.6mM ^{15}N-His labeled *E. coli* CS at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time was 15 hours.

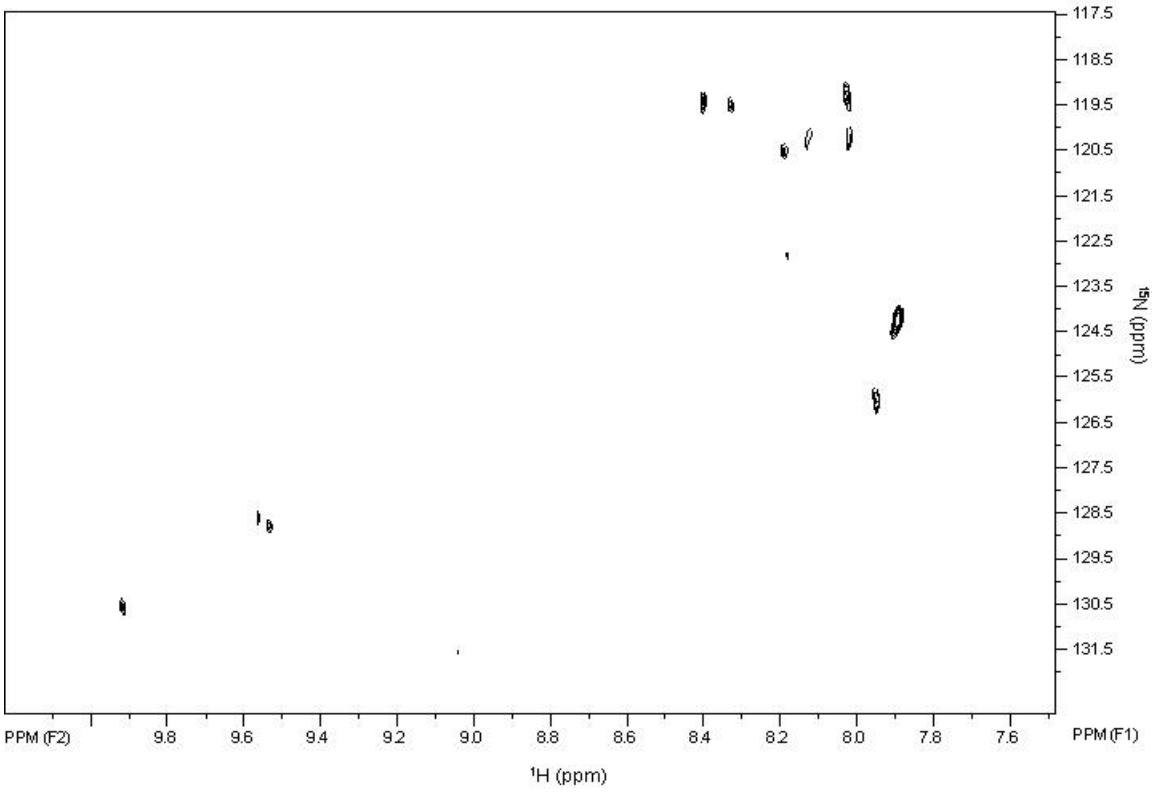


Figure 3.52: Two-dimensional [$^{1}H, ^{15}N$] HSQC spectrum of ^{15}N-His labeled *E. coli* CS in the presence of 1M urea after 24 hours of incubation, measured at 25ºC at a ^{1}H frequency of 600 MHz. Total acquisition time for spectrum was 15 hours.

Spin labeling of *E. coli* CS

Spin labels are paramagnetic moieties which drastically increase T_1 and T_2 relaxation of neighbouring spins. So the effect of spin labeling is broadening or disappearance of cross-peaks in the NMR spectrum, if they arise from nuclei that are close to the spin label. The size of the paramagnetic relaxation enhancement depends on the inverse sixth power of the distance between the nuclear spin and paramagnetic centre, and can be used as a sensitive measure up to distances of 2nm [Cutting, et al., 2004].

Spin labels in proteins are usually attached to reactive cysteines. In *E. coli* CS, one of the six cysteines, Cys-206, has been identified as reactive, which is located at the dimer-dimer contact surface area, 25-30Å away from the active site [Donald, et al., 1991; Nguyen et al, 2001]. Initial experiments were done with unlabeled *E. coli* CS, using a 10 fold molar excess of the spin label 3-iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline) followed by overnight incubation at room temperature. The incorporation of the label was determined using mass spectrometry, by the analysis of tryptic fragments containing cysteine residues. The initial experiment resulted in modification of Cys-206 but modification of other cysteines as well. In the second attempt, a 50 fold molar excess of the spin labeling reagent was used, resulting in an increased modification of Cys-206 as compared to others (Table 3.4).

Table 3.4: Percentage incorporation of the spin label 3-iodomethyl-(1-oxy-2,2,5,5-tetramethylpyrroline) in cysteine residues of *E. coli* CS

Cysteine	Trial 1 10 fold molar excess of spin label	Trial 2 50 fold molar excess of spin label
C86	N/a	24%
C135	17%	24%
C206	55%	67%
C251	19%	26%

N/a indicates insufficient data for analysis

For the NMR experiment, similar labeling conditions, with a 50 fold molar excess of spin label reagent, was used with ^{15}N-Met labeled *E. coli* CS. The 2D-TROSY spectrum of ^{15}N-Met showed that several cross-peaks were displaced from their original position [Figure 3.53]. The result was difficult to analyze because of more than one cysteine being modified.

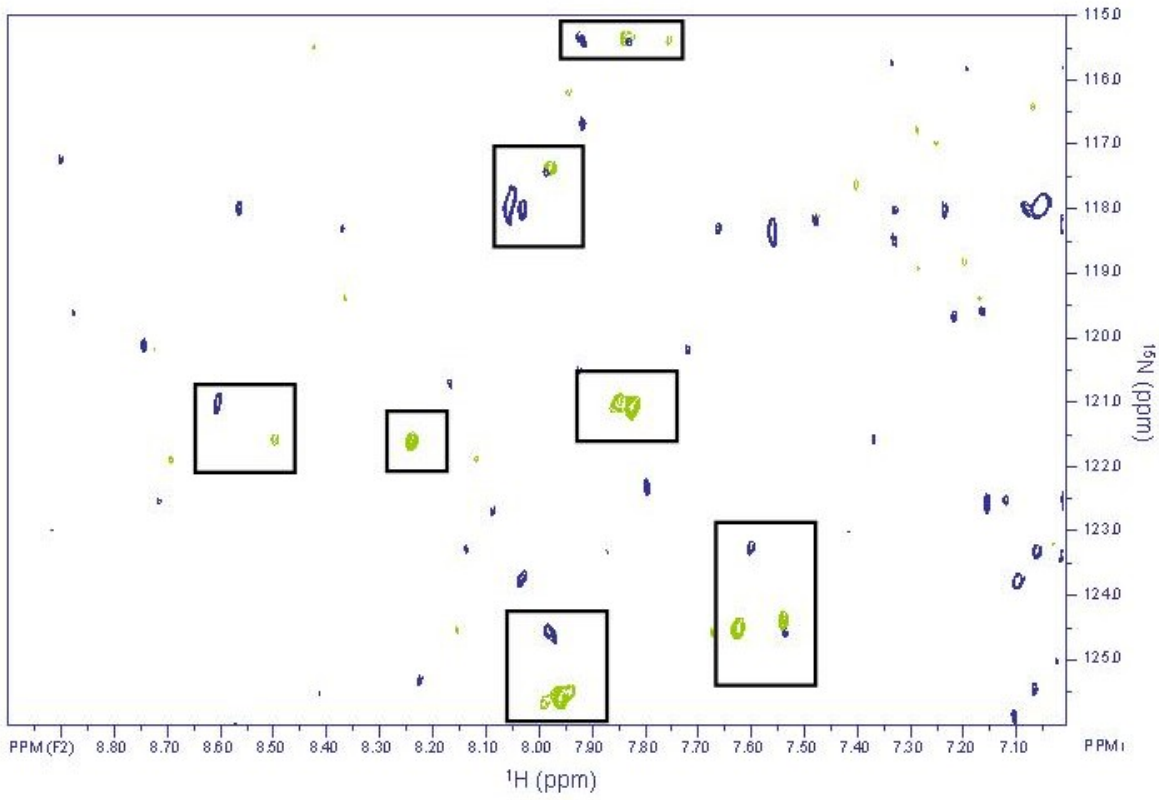


Figure 3.53: Overlay of two-dimensional HSQC spectra of 1.5mM ^{15}N-Met labeled *E. coli* CS with (shown in blue) and without (shown in green) spin labeling.

Table 3.5: Distance of backbone nitrogen of methionines in *E. coli* CS from sulphur atom of Cys-206

Methionine	Distance (nm)	Methionine	Distance (nm)
M112	0.82	M274	3.7
M131	2.4	M301	3.7
M134	3.4	M318	3.1
M168	1.2	M340	4.3
M171	1.1	M372	2.6
M174	1.3	M378	2.3
M200	1.1	M385	2.0
M201	1.4	M398	3.2
M219	4.9	M403	3.4

The modification of Cys-206 by 3-iodomethyl-(1-oxy-2,2,5,5 tetramethyl pyrroline) would have an affect on seven methionine (M112, M168, M171, M174, M200, M201, M385) residues out of total 18 present in *E. coli* CS. Others like M131, M 372, M378 can also be affected, but to a small extent [Cutting et al., 2004].

3.9 Discussion

Large proteins, such as *E. coli* CS, pose a challenge for detailed study by NMR spectroscopy. The *E. coli* CS studied here is a 280kDa hexamer. Large proteins tumble very slowly in solution and hence give rise to broad peaks, to an extent that they are usually unobservable. Regions of high flexibility in large proteins which are moving very fast irrespective of the slow overall motion of the protein, however, give rise to sharp peaks in the NMR spectrum. In the HSQC spectrum of U-^{15}N labeled *E. coli* CS, 23 well resolved relatively narrow cross-peaks and a region of overlapping peaks were observed, indicating these are coming from the flexible parts of the protein. Similar narrow resonances have been observed in U-^{15}N labeled proteins of molecular mass 68kDa and 204kDa [Volkert et al., 1999; Chou et al., 2002]. The narrow resonances in these systems have been assigned to the flexible parts of the protein which are also involved in the function of the protein.

Two-dimensional NMR spectra of *E. coli* CS suggest that there is residual secondary structure in the mobile regions of *E. coli* CS because many chemical shifts deviate from random coil values. This observation suggests that the mobile residues might be in conformational exchange between the unstructured and structured states [Hwang, et al., 2004; Chou, et al., 2002]. This observation is different from the mobile loop in GroES, which displayed random coil chemical shifts as a result of disorderd conformation but showed greater chemical shift dispersion upon GroEL binding [Fiaux, et al., 2002]. In

the case of *E. coli* CS the mobile regions might be moving in a dynamic way with alternative conformations to give a greater chemical shifts dispersion in the NMR spectra.

The origin of cross-peaks observed in the two-dimensional [^{15}N, ^{1}H] HSQC/ TROSY spectra can be speculated from the B-factors of the individual residues. On comparing the number of cross-peaks observed in HSQC/ TROSY spectra of specific ^{15}-N labeled *E. coli* CS, with the B-factors of individual residues, a baseline can be drawn for the residue with lower B-factors contributing to the spectra. In the HSQC spectrum of ^{15}N-Gly labeled E. *coli* CS, eight residues should be observed if residues with B-factor values of 40Å are contributing to the spectrum. In the HSQC spectrum of ^{15}N-Leu labeled *E. coli* CS, eleven residues should be observed if residues with B-factor value of 40 Å are contributing to the spectrum. In the TROSY spectrum of ^{15}N-Lys labeled *E. coli* CS, thirteen residues should be observed if residues with B-factor value of 40 Å are contributing to the spectrum. The number of cross-peaks observed in the [^{15}N, ^{1}H] HSQC/ TROSY spectrum of ^{15}N-Gly, ^{15}N-Leu and ^{15}N-Lys labeled *E. coli* CS were 10, 12 and 14 respectively. For the above three cases, the number predicted matches quite well with number of cross-peaks actually observed. The exception to the above criterion was observed in case of ^{15}N-Met labeled *E. coli* CS, where four peaks should be observed if we consider the above baseline of residues with B-factor values of 40 Å or greater contributing to the spectrum. But sixteen cross-peaks were

observed in the HSQC spectrum of ^{15}N-Met labeled *E. coli* CS. Taking into account four peaks involved in the conformational exchange leading to duplicate peaks; there are still twelve methionine cross-peaks contributing to the HSQC spectrum.

We have combined specific ^{15}N amino acid labeling which give residue type assignments, with site-directed mutagenesis to assign these mobile residues in *E. coli* CS (Appendix I). We have found evidence that at least eight of the mobile residues are contributed by the N-terminal region, one residue, L275, from the mobile loop (residues 266-297) and others from uncharacterized flexible regions of the protein.

Small changes in chemical shifts of some residues, when substrates are added, identify regions of protein that are affected when substrates are bound. The residues affected display cross-peaks even after addition of substrates except in the case of tryptophans. One might expect a reduction in mobility and consequent line broadening for residues in the mobile loop. The most likely explanation might be the existence of a dynamic equilibrium between the T (inactive) and (R) active states of *E. coli* CS. So if there is a population of molecules with high mobility in localized regions and these exchange with molecules in which these regions are less mobile, we would expect to see signals even after addition of substrates [Chou, et al., 2002]. It is tempting to speculate that the residues affected by substrate addition belong to the mobile loop which is closely associated with the active site. Although

experimental evidence from ^{15}N-Trp labeled *E. coli* CS shows that regions other than mobile loop are affected when substrates bind. We show here that the environments of tryptophan side chains from W260 and either W391 and W395 or both are changed. The above result indicates extensive communication within the macromolecule, even between residues relatively distant from the active site.

Another surprising result of this study is the effect of substrate on residue L275 which belongs to the mobile loop, and has been characterized in the NMR spectrum by site-directed mutagenesis. The pair of peaks associated with L275 changes only slightly when substrates are added. In contrast a previous study [Ayed, et al., 1999] which used the same mutant L275A showed that the mutant has a very low activity indicating a much different conformation of the residue when substrates bind.

To account for the presence of more than the expected cross-peaks in the HSQC spectra, temperature dependence studies in the range of 5-40 ºC were done on ^{15}N-Met and ^{15}N-Leu *E. coli* CS. This approach gave evidence that some residues in *E. coli* CS undergo slow exchange over different conformations in solution. After taking into consideration the multiplicity of the peaks in the 2D [^{15}N, ^{1}H] HSQC spectra of ^{15}N-Met and ^{15}N-Leu *E. coli* CS, the numbers of residues contributing to the spectra were modified. In the case of ^{15}N-Met *E. coli* CS, there are 12 methionine residues contributing to total 16 cross-peaks observed in the HSQC spectrum. Four cross-peaks give

duplicates due to their involvement in slow conformational exchange. The remaining eight peaks show no sign of conformational exchange at least in the temperature range of 5-40ºC. In the case of ^{15}N-Leu *E. coli* CS, there are 11 leucine residues contributing to a total of 12 cross-peaks observed in the HSQC spectrum.

Thermodynamic parameters for such residues as listed in Table 3.4 and suggest that the existence of multiple conformational effects, at least for these residues, are local in nature. It was not possible to determine thermodynamic parameters for all pairs of peaks in slow conformational exchange in ^{15}N-Met *E. coli* CS due to insufficient data points. It is difficult to characterize these multiple conformations at this stage but their possible origin might be due to equilibrium of the hexameric protein with a small amount of dimer, heterogeneity within two monomers, inactive or T state – active or R conformational states of the enzyme, and folded-unfolded conformations of the flexible loop giving rise to separate NMR signals.

Table 3.4: Thermodynamic parameters for three pairs of peaks involved in conformational exchange

Peaks	**ΔH kJ/mole**	**ΔS J/mole.K**
^{15}N-Leu	28±4	498±14
^{15}N-Met	16±1	450±5
^{15}N-Met	42±10	410±46

For large proteins, a high level of protein deuteration is essential to acquire high-quality NMR spectra and to fulfill the requisite assignment problem. Deuteration has become a widely used technique for reducing nuclear spin relaxation in large biomolecules, resulting in increased signal intensity and better resolution for multidimensional NMR experiments. However, one problem associated with the expression of protein from deuterated media is the necessity to back-exchange the deuterated amide sites for protons. If the protein of interest can be reversibly denatured, this is a trivial task. However, in the case of the *E. coli* CS, slowly exchanging amide sites in the enzyme core cannot be readily exchanged in this manner while retaining the viable enzyme. In our study an attempt was made to prepare *E. coli* CS in partially deuterated medium. Although deuteration improved the quality of the spectrum in the random coil region, some residues no longer appeared in the HSQC spectrum, perhaps because of incomplete back-exchange of the amides or conformational exchange.

Most experiments involving *E. coli* CS were carried out at a field (600 MHz proton frequency) which is too low to fully exploit relaxation interference effects [Wüthrich, 1998]. Nevertheless, the spectra of selectively ^{15}N-labeled amino acids were of acceptable quality to study only the flexible residues in the enzyme. No significant improvement was observed by using the TROSY scheme when using protonated samples. This observation is expected for a system with exchange contributions to the T_2 relaxation. The

lack of TROSY enhancement could also be due, in principle, to the fact that the samples were not perdeuterated. Perdeuteration can significantly enhance the quality of TROSY spectra of high molecular weight proteins [Pervushin, 2000] and it would have been desirable to use perdeuteration in all the samples of *E. coli* CS in the present work. In one sample, where partial deuteration was accomplished, the combination of TROSY, deuteration and higher fields (800MHz) resulted in improvement in sensitivity and resolution of the spectrum. The fact that nearly the same quality of spectrum was obtained in 32 scans at higher fields as compared to 512 scans at lower fields is encouraging, indicating much more might be achieved if future experiments were carried out at higher fields.

Obtaining a chemical shift assignment in very large systems remains a significant challenge. TROSY-type triple-resonance strategies have been successfully used to obtain backbone assignments in systems up to approximately 110kDa [Salzmann, et al., 2000]. While these strategies are likely to have only limited utility in even larger systems [Riek, et al., 1999] their extension to CRINEPT-based methods may be successful, particularly with the advent of increasingly higher-field magnets. However, even the 2D ^{1}**H**/^{15}N-CRINEPT-TROSY spectra of *E. coli* CS with a mass of 280kDa is not of high enough quality to incorporate into more complex experiments needed for backbone assignment.

In conclusion, the results of the study give a different picture about the mobile regions of the enzyme in solution phase as compared to crystallographic results. The X-ray crystal structure predicted mobility to be associated with the "mobile loop" residues 266-297, residues 330-335, and few residues in the N-terminal region. Cross-peaks observed in the two-dimensional NMR spectra correspond not only to residues in the mobile loop, with very high B-factors but also from various flexible parts of the protein. Some of these flexible regions are contributed from obvious regions such as the N-terminal but others, such as side chains of tryptophans with low B-factors in the X-ray crystal structure, are also visible. Most of the NMR peaks do not change at all when substrates are added; this further suggests that the peaks are arising from different parts of the protein.

These flexible residues also behave differently in response to substrate addition. Some of the flexible residues change their environment when substrates are added but remain mobile as observed in the case of methionine and lysine, while others, like the side chains of tryptophans, are immobilized by substrate. The results from fluorescence studies on *E. coli* CS also showed that there is a change in the environment of the tryptophans in presence of substrates (described in Chapter 5).

3.10 Future Work

In order to define cross-peaks in specific labeled samples of *E. coli* CS more clearly, mutants of particular amino acids in different regions should be studied.

Protocols should be developed to make specific ^{15}N amino acid labeled with highly deuterated background. This will greatly increase the sensitivity of the NMR experiments and will also increase the chances of observing residues from the less flexible or rigid parts of the protein.

4.0 ESTIMATION OF ISOTOPIC INCORPORATION IN NMR SAMPLES BY MASS SPECTROMETRY

4.1 Introduction

Heavy isotope incorporation into proteins is necessary for the use of heteronuclear NMR experiments to study moderate-sized to large proteins (Marius Clore & Gronenborn, 1998). For large proteins where specific labeling with individual amino acids is often used, the protein must also be efficiently labeled. The extent of labeling is an important parameter that must be known in order to interpret the experiment accurately. One other major concern with specific amino acid labeling is the metabolic conversion of one amino acid to the other. This can give false information especially where specific labeling is done in flexible regions in large proteins (Fiaux, et al., 2004).

The most obvious method of determining labeling efficiency is to measure the average mass of the labeled protein by mass spectrometry and to compare it with the value for the unlabeled protein. Either electrospray ionization (Zhao, et al., 2004) or Matrix-Assisted Laser Desorption/ Ionization (MALDI) (Salzmann, et al., 2000; Fiaux, et al., 2004, Markley, et al., 2001) mass spectrometry has been used for this purpose. For uniformly labeled proteins, an average mass measurement of the intact protein is probably adequate. On the other hand, it will be inadequate if the protein in question is labeled with only one amino acid and the mass increase in the intact protein is much

smaller, and often is less than the error in its mass measurement. If incorporation is less than ideal, even a slight heterogeneity broadens the mass peak so that the mass difference cannot be measured with sufficient precision. When only one amino acid is labeled, the mass shift of a peptide is about 1 Da per incidence of that particular amino acid. At any incorporation rate less than 100%, there will be two populations of peptides (labeled and unlabeled) with overlapping ion distributions. In this chapter, a method used for estimating heavy isotopic label in U-^{15}N and specific ^{15}N labeled *E. coli* CS is described. The results in the chapter are described in Choudhary, K. et al., 2006.

4.2 Experimental Procedures

Protein for mass spectrometry

Pure protein prepared for NMR spectroscopy was dialyzed into 50 mM ammonium bicarbonate buffer (Ayed, et al., 1998), diluted to 1mg/mL, and incubated overnight with 1:100 w:w TPCK-trypsin. Later, we found that protein could be diluted directly from the NMR buffer into digest solutions to give a reasonable MALDI spectrum. Aliquots of 0.5 µL each of digest and 2, 5-dihydroxybenzoic acid matrix solution were applied to a metal target and analyzed on a MALDI QqToF instrument in the department of Physics (Ens, et al., 2005). Peptides that covered a wide range in mass were chosen for analysis if they were clearly defined and contained the amino acid of interest. Protein samples for mass spectrometry were prepared by the author, data

collection and analysis of all the results described here were done by Dr. Lynda Donald.

Simulations

To analyze the results of these experiments, a numerical simulation method was devised by Mr. Vic Spicer. In this numerical simulation, the isotope mass distribution of a given peptide was calculated through a number of trials. The peptide mass for each trial was determined by constructing the peptide atom by atom, using a program that produces random numbers between 0 and 100 to select each atom′s mass. For example, for each nitrogen in an unlabeled sample, the mass for ^{14}N is chosen if the random number is less than 99.634; otherwise, ^{15}N is chosen. In the case of ^{15}N enrichment, the opposite approach is taken; at an incorporation rate of 80%, ^{15}N is chosen if the random number is less than 80, and ^{14}N otherwise. The resulting peptide masses from 30,000 trials (1000 for Figure 4.1), are binned at 1 Da resolution.

4.3 Results

The peptides used for analysis are shown in Table 4.1. Figure 4.1 shows simulated ion distributions for a peptide, LMGFGHR ($C_{36}H_{56}N_{12}O_8S$, Peptide 1), first for natural isotopic abundance and then for cases where a single nitrogen atom has been labeled with ^{15}N to various degrees. This simulation applies directly to the labeling of Peptide 1 with leucine (L), methionine (M), or phenylalanine (F), since the peptide contains one of each

residue, and each residue contains one nitrogen atom. The input will be modified if glycine (G) is labeled, since the peptide contains two G residues

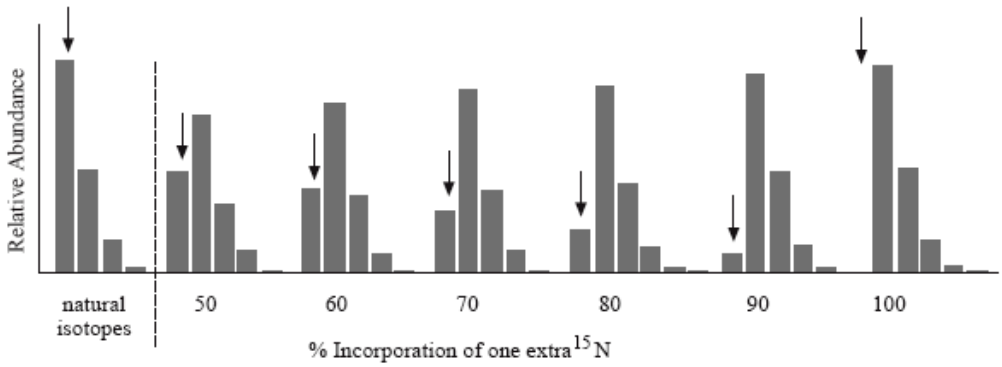


Figure 4.1: Simulated ion distributions (1000 trials) for Peptide 1 (LMGFGHR= $C_{36}H_{56}N_{12}O_8S$), at natural abundance, and for 50 to 100% incorporation of ^{15}N in a single nitrogen atom. The arrows indicate the position of the monoisotopic ions at *m/z* 817 [Choudhary, et al., 2006].

Table 4.1: Tryptic peptides from *E. coli* CS used in the simulations shown in Figures 4.1-4.3

Peptide	**Sequence**	**Site in protein**	**Monoisotopic mass**	**Formula**
1	LMGFGHR	300-306	816.406	$C_{36}H_{56}N_{12}O_8S$
2	HIPEFVRR	283-290	1052.588	$C_{48}H_{76}N_{16}O_{11}$
3	QLYTGYEKR	410-418	1156.588	$C_{52}H_{80}N_{14}O_{16}$
4	HTMIHEQITR	110-119	1264.635	$C_{53}H_{88}N_{18}O_{16}S$
5	YSIGQPFVYPR	178-188	1325.677	$C_{64}H_{91}N_{15}O_{16}$
6	ITFIDGDEGILLHR	56-69	1597.846	$C_{72}H_{115}N_{19}O_{22}$
7	TVGWIAHWSEMHSDGMK	388-404	1970.877	$C_{87}H_{126}N_{24}O_{25}S_2$

4.31 Uniform ^{15}N-labeled *E. coli* CS

For protein prepared with $^{15}NH_4Cl$, comparisons of measured MALDI *m/z* spectra with the simulations for natural isotopic abundance are shown in Figure 4.2 for three peptides, Peptide1, Peptide5 and Peptide 6. The mass shift of the most abundant ion in the three peptides is equal to the number of nitrogens in the peptide.

Labeling efficiency is calculated by a least squares method by comparing the relative abundance of ions from the measured spectrum to those from a set of theoretical spectra. The goodness of fit, y, is calculated from the equation

$$y = \frac{\sum_{i=1}^{n}\left[mass_i\left(theoretical_i - measured_i\right)^2\right]}{N\sum_{i=1}^{n}\left(mass_i\right)}$$

where *i* is any ion, and N is the number of compared points (Snijders, et al., 2005)

For U-^{15}N labeled *E. coli* CS, 96% incorporation of ^{15}N was determined as shown in the Goodness Fit plot in Figure 4.2. Simulations of ion distributions of the three peptides with 96% incorporation of ^{15}N have the same shape as those in the measured spectra (Figure 4.2).

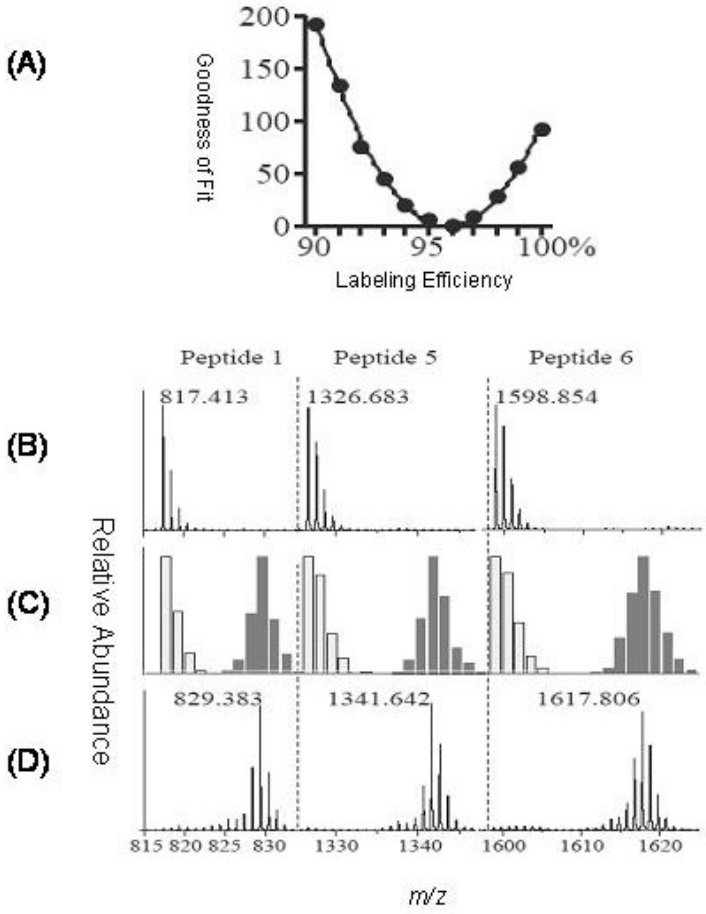


Figure 4.2: Isotopic incorporation in U-^{15}N labeled *E. coli* CS. (A) Labeling efficiency (96%) was calculated by determining the minimum of the Goodness of Fit plot. Comparison of simulated ion distributions with the *m/z* spectra from measured protein digests for three of the peptides: Peptide 1, Peptide 5 and Peptide 6 (Table 4.1). (B) Normalized MALDI spectra from digests of unlabeled CS; (C) Simulations of the ion distributions for natural isotopic abundance • , and for 96% incorporation of ^{15}N into all nitrogens in each peptide ν ; (D) Normalized MALDI spectra from a digest of CS protein labeled with $^{15}NH_4Cl$ [Choudhary, et al., 2006].

4.32 Specific ^{15}N-labeled *E. coli* CS

Incorporation of ^{15}N label in specific labeling with Phe was analyzed using four different peptides, each containing just one phenylalanine. Labeling efficiency of ^{15}N was determined to be 50% (Figure 4.3). The comparison of the measured MALDI *m/z* spectra with the corresponding simulations are shown for peptides in Figure 4.3.

To account for the low incorporation of ^{15}N label, peptides without phenylalanine (Peptides 3, 4, 7) were analyzed. All the peptides had natural abundance isotopic distributions, showing that the ^{15}N label was probably not lost by metabolic transfer to another amino acid during cell growth. Instead, it appears that the protein acquired about half of its phenylalanine from residual unlabeled material carried over from the early stages of cell culture. Changes in cell growth conditions in a subsequent experiment improved the labeling efficiency, but only to 60%.

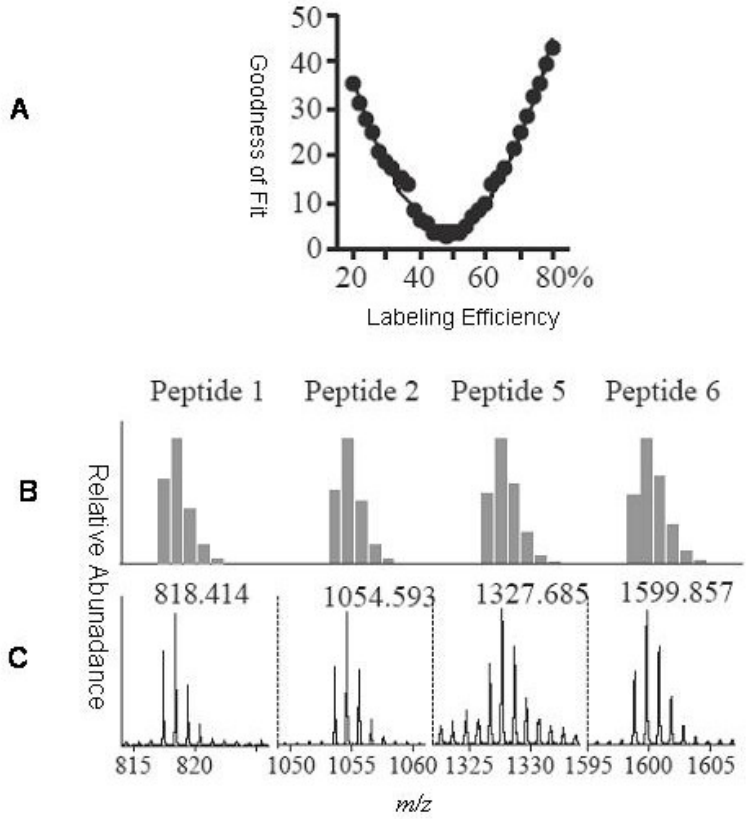


Figure 4.3: Labeling efficiency of ^{15}N in ^{15}N-Phe *E. coli* CS. (A) Labeling efficiency (50%) was calculated by determining the minimum of the Goodness of Fit plot. (B & C) Comparison of the ion distributions from simulations (B) and measured MALDI *m/z* spectra (C) for peptides from protein prepared with ^{15}N-phenylalanine. The simulations were calculated for an incorporation level of 50% [Choudhary, et al., 2006].

Incorporation of ^{15}N label in specific labeling of Gly was analyzed using four different peptides, two of which had single glycines while the other two peptides had two. The algorithm was adjusted for the appropriate number of glycines for the different peptides. The labeling efficiency for ^{15}N was determined to be 72% (Figure 4.4). The comparison of measured MALDI *m/z* spectra with the corresponding simulations matches quite well as shown in Figure 4.4. Metabolic transfer of glycine to serine is a major concern during specific labeling of glycine. In our labeling scheme no significant scrambling was observed from glycine to serine residues. Three peptides which contain serine residues but no glycine residues matched the unlabeled control with no ^{15}N label.

Incorporation of ^{15}N His was analyzed using four different peptides, of which three peptides had single histidines and one peptide had two. In the case of histidine labeling, Peptides 1 and 2 each have a total of three and Peptides 4 and 7 each have a total of six ^{15}N. The incorporation of ^{15}N label was consistent with 98% incorporation (Figure 4.5).

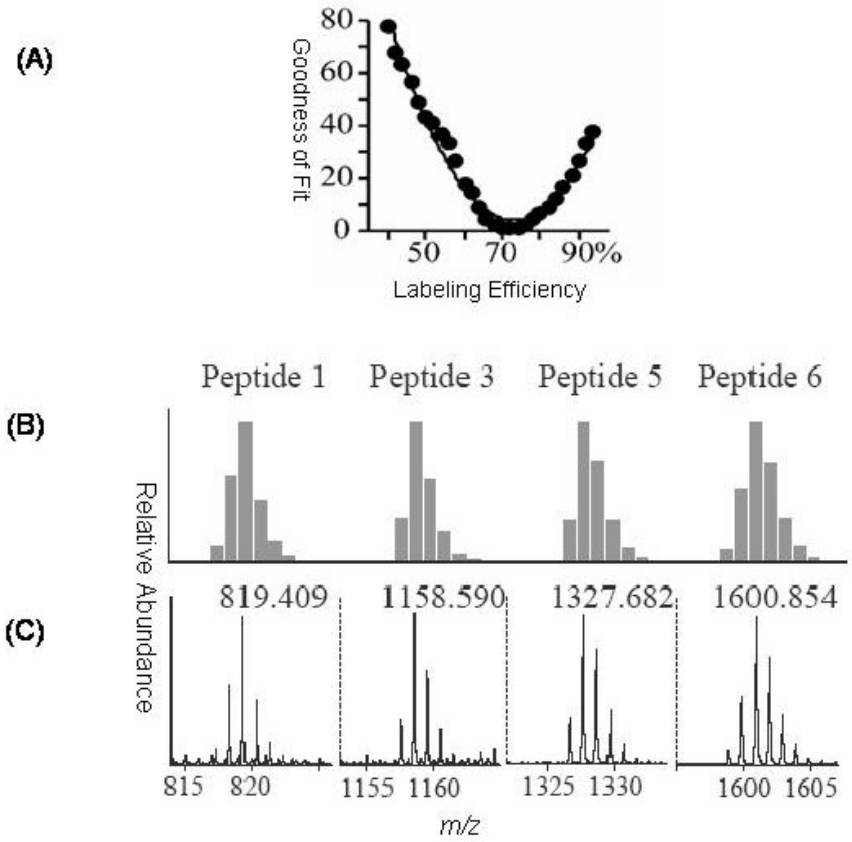


Figure 4.4: Labeling efficiency of ^{15}N in ^{15}N-Gly *E. coli* CS. (A) Labeling efficiency (72%) was calculated by determining the minimum of the plot. (B & C) Comparison of the ion distributions from simulations (B) and measured MALDI *m/z* spectra (C) for peptides for protein prepared with ^{15}N-glycine. The simulations were calculated for an incorporation rate of 72% [Choudhary, et al., 2006].

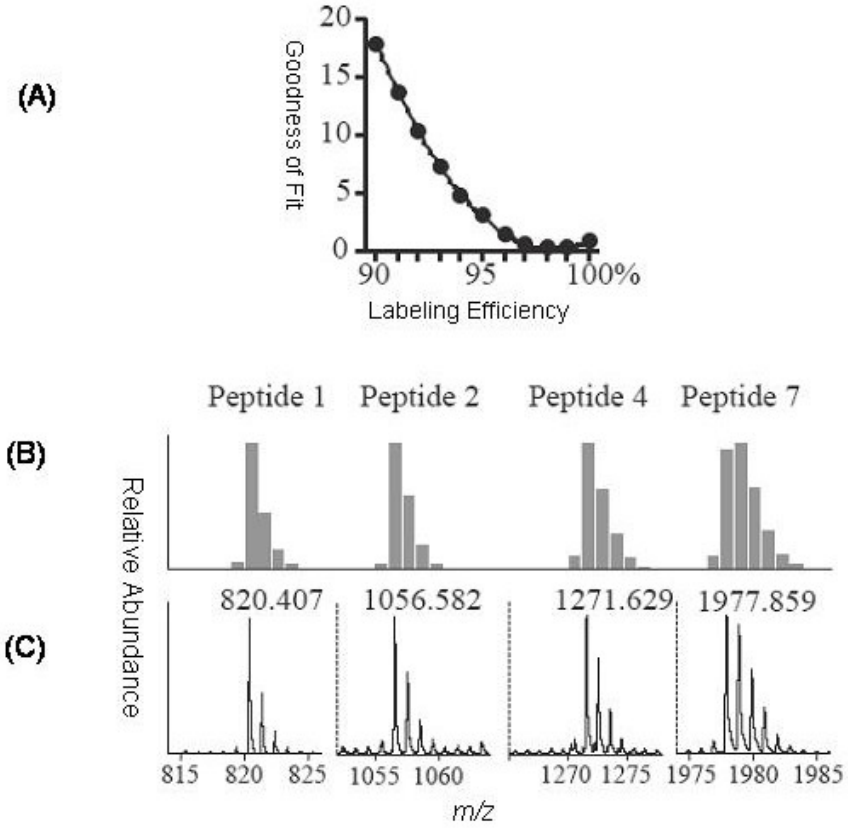


Figure 4.5: Labeling efficiency of ^{15}N in ^{15}N-His *E. coli* CS. (A) Labeling efficiency (98%) was calculated by determining the minimum of the plot. (B & C) Comparison of the ion distributions from simulations (B) and measured MALDI *m/z* spectra (C) for peptides for protein prepared with ^{15}N-histidine. The simulations were calculated for an incorporation rate of 98% [Choudhary, et al., 2006].

Table 4.2: Isotopic incorporation of ^{15}N and ^{13}C in samples prepared for NMR spectroscopy

Labeled Sample	**Isotope Label**	**Isotopic Enrichment (%)**
L-Phe WT CS	C6H5CH2CH(*NH2)COOH	50-60
L-Met WT CS	CH3SCH2CH2CH(*NH2)COOH	95
L-Met M274L CS	CH3SCH2CH2CH(*NH2)COOH	95
L-Met M168L CS	CH3SCH2CH2CH(*NH2)COOH	95
L-Gly WT CS	CH3CH(*NH2)COOH	72
L-Histidine WT CS	C3H4(*N2)CH2(*NH2)COOH	98
L-Lysine WT CS	H2N(CH2)4CH(*NH2)COOH.2HCl	98
L-Lysine WT CS	H2N(CH2)4CH(NH2)*COOH.2HCl	98
L-Lysine Trunc CS	H2N(CH2)4CH(*NH2)COOH.2HCl	98
L-Leucine WT CS	(CH3)2CHCH2CH(*NH2)COOH	75
L-Leucine L275A CS	(CH3)2CHCH2CH(*NH2)COOH	75
L-Tryptophan CS	(C8H5*NH)CH2CH(*NH2)COOH	85
L-Tryptophan W260A CS	(C8H5*NH)CH2CH(*NH2)COOH	85

*N, * C indicates ^{15}N and ^{13}C respectively

4.4 Discussion

In all cases, the amount of labeling was observed to be uniform and specific to the chosen amino acid for any protein preparation. The relative amount of ^{15}N was the same for all peptides having the residue of interest, and was the same as natural abundance for those peptides without that particular amino acid. The peptides span the length of the protein, so this clearly demonstrates that there is no preferential or non-uniform labeling.

In all the single amino acid labeling experiments, the growth medium was supplemented with unlabeled amino acids (McIntosh, et al., 1990), and no NH_4Cl was included (Ramesh, et al., 1994). This should limit dilution of the labeled amino acid by endogenous synthesis or transamination and effectively dilute out any ^{15}N transferred to a different amino acid (Fiaux, et al., 2004; McIntosh, et al., 1990). There is no evidence for transfer of ^{15}N to amino acids other than the intended one. Direct metabolic transfer may be more of a problem with non-bacterial cell lines, (Yabuki, et al., 1998; Klein-Seetharaman, et al., 2002) and amino acids where exchange would be substantial, such as glutamate- glutamine and aspartate-asparagine (Fiaux, et al., 2004; McIntosh, et al., 1990) which were avoided.

In conclusion, the method described here provides a more precise estimation of isotopic incorpoaration in single amino acid labeled *E. coli* CS as compared to the average mass determination approach.

5.0 FLUORESCENCE STUDIES OF *E. COLI* CS

5.1 Introduction

Fluorescence spectroscopy has been used to study the conformational dynamics of specific regions of proteins [Stewart, et al., 2005, Nelson, et al., 2000]. The aromatic amino acids, tryptophan, tyrosine, and phenylalanine offer intrinsic fluorescent probes of protein conformation and dynamics. Of the three, tryptophan is the most popular probe. The fluorescence of the indole chromophore is highly sensitive to environment which makes it an ideal choice for reporting protein conformational changes. The position of the maximum of the fluorescence spectrum of tryptophan residues in protein varies from 307 to 353 nm depending upon the environment of the indole ring [Ladokhni, 2000]. The blue shift indicates a more hydrophobic environment and the red shift indicates a more polar environment.

Conformational changes in proteins from events such as substrate binding can alter the tryptophan microenviroment resulting in enhancement or quenching of fluorescence and/or in shifts of the spectrum towards blue or red. Trytophan quenching in proteins can occur if the tryptophan is exposed to an aqueous environment. Glutamic acid, aspartic acid, side chain of lysine, glutamine, asparagines, cysteine, cystine (disulphide) also acts as tryptophan quenchers in proteins [Chen & Barkley, 1998].

To gain information about the conformational changes in the mobile loop upon the addition of substrates, a fluorescent probe was introduced replacing phenylalanine, F287 with tryptophan. *E. coli* CS contains three tryptophan residues, W260, W391, W395 as shown in Figure 5.1; none of them belongs to the mobile loop.

The environment of W260 in *E. coli* CS is different from the other two tryptophans (W391 & W395). W260 is in a polar environment whereas the other two tryptophans are in more hydrophobic environment [Ayed, et al., 1999]. W260 is close to the active site residue His264 (Figure 5.1), which has to reorient to form the binding site for the substrate Ac-CoA. The NMR results (Figure 3.29 &3.31) have shown that the environment of W260 along with W391 and W395 is also altered in the presence of substrates. Fluorescence Spectroscopy was carried out on mutant W260A to study the change in the environment of the two tryptophans W391 & W395.

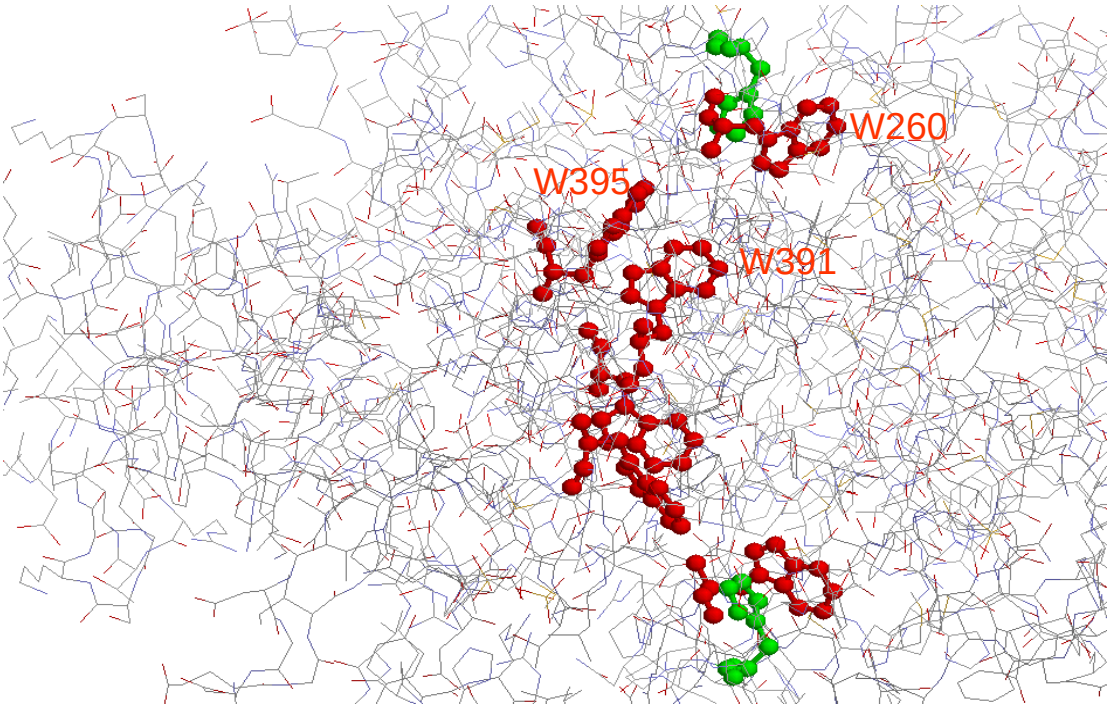


Figure 5.1: Region of hexameric *E. coli* CS in wire-frame representation showing the three tryptophans in two subunits in red and the active site residue H264 in green [Nguyen, et al., 2001]

5.2 Experimental Procedures

Mutant F287W

The mutation F287W was constructed using the Quick change Site-directed Mutagenesis Kit by Stratagene as described in Methods (Chapter 2). Primers used for mutagenesis were:

F287W UP	CATTCCGGAATGGGTTCGTCGTGCG
F287W DN	CGCACGACGAACCCATTCCGGAATG

The mutant plasmid was transformed into MOB154 cells. Expression was carried out by inoculating 2mL of LB medium with a single colony followed by overnight growth in a shaking incubator at 37ºC. The next morning 2mL culture was used to inoculate 100mL LB media. After 9-10 hours of growth, these cells were used to inoculate 2 X 1 L LB media. Just before the inoculation 5mL of ampicillin (25mg/ mL in water) was added to each litre of LB media. Protein purification was carried out as described in Experimental Procedures (Chapter 2). The presence of the mutation was verified by the analysis of tryptic fragments of the protein using mass spectrometry.

Kinetic experiments

Assays were performed on the mutant as described in Experimental Procedures (Chapter 2). Two substrate saturation curves were obtained (one for OAA and another for AcCoA) in the presence of 0.1M KCl.

Fluorescence Spectroscopy

Fluorescence Spectroscopy was carried in a JASCO spectropolarimeter-fluorometer, using a quartz cuvette with a path length of 1cm, excitation bandwith of 8nm, emission bandwidth of 10nm and response time of 0.5sec. The excitation wavelength was 295nm, and the emission spectra were recorded from 310 to 450nm. The concentration of the proteins of the proteins was in the range 0.26-0.38μM in standard CS buffer (20mM tris, 1mM EDTA, pH 7.8). The absorbencies of the samples at 295nm were less than 0.1 to eliminate inner-filter effects. Titrations with the substrates were carried out by successive addition of concentrated stock solution of one substrate in the presence of saturating amounts of the other substrate and 0.1M KCl. Emission spectra were recorded following each addition. Results from substrate titrations were corrected for small dilution effects, baseline corrected by subtracting the solvent spectrum and then analyzed using the Spectra Analysis program (JWSTDA32) on the instrument. Data fitting to theoretical curves was done using KaleidaGraph.

Substrate titration data were analyzed by nonlinear least-square fits, using the following equation:

$$\frac{\Delta F}{F_0} = \frac{(\Delta F_{\max} / F_0)S}{K_d + S} \qquad \text{Equation 5.1}$$

where ΔF is the change in protein fluorescence caused upon addition of substrate S, F_0 is the fluorescence in the absence of substrate and K_d is the dissociation constant [Nelson, et al., 2000].

5.3 Results

Kinetic measurements of mutant F287W CS show tighter binding of substrates OAA and acetyl-CoA as compared to wild type (Table 5.1).

Table 5.1: Catalytic properties of CS (WT) and mutant F287W in the presence of 0.1M KCl

Kinetic parameters	*CS(WT)*	*F287W*
K_{OAA}, µM	26±5	9±2
$K_{Ac\text{-}CoA}$, µM	120±20	33±7

The emission spectra of CS (WT) and the mutants F287W and W260A are shown in Figure 5.2. The difference spectrum in Figure 5.2a shows the slight difference in environment of the inserted tryptophan in mutant F287W. The difference spectrum in Figure 5.2b shows that residue W260 is more solvent exposed as compared to the other two tryptophans.

The emission spectra of CS (WT), mutant F287W and W260A and the changes in the fluorescence emission on addition of 0.5mM OAA in the presence of 100µM CM-CoA and 0.1M KCl are shown in Figure 5.3. The fluorescence spectrum did not change with the addition of 0.1M KCl.

However addition of the substrate OAA, and the substrate analogue CM-CoA, caused a decrease in the fluorescence intensity for all three proteins (Table 5.2, Figure 5.3).

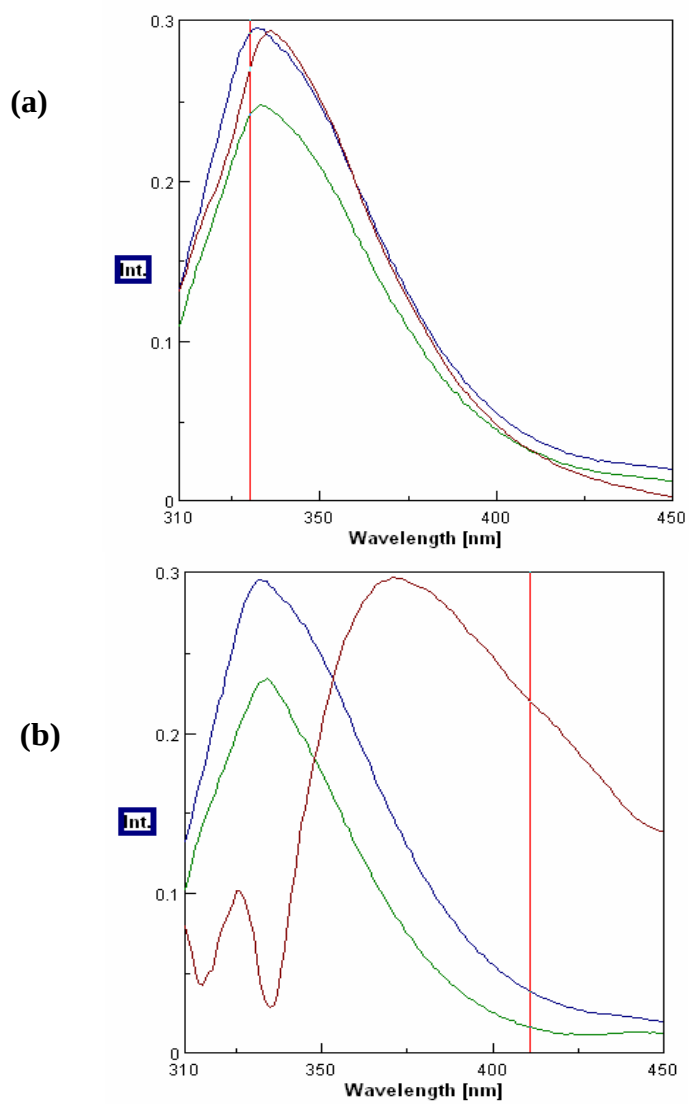


Figure 5.2: (a) Fluorescence spectra of *E. coli* CS (WT) and F287W are shown in blue and green respectively. The difference spectrum is shown in purple. (b) Fluorescence spectra of *E. coli* CS (WT) and W260A are shown in blue and green respectively. The difference spectrum is shown in purple. Spectra were recorded at 25ºC using $\lambda_{excitation}$= 295nm.

Table 5.2: Fluorescence emission of the wild type and mutant proteins with and without substrate complex

	Substrates	F[a]	λ_{max}
CS(WT)	None	1.0	332
	0.1M KCl, 100µM CM-CoA, 0.5mM OAA	0.5	332
F287W	None	1.0	333
	0.1M KCl, 100µM CM-CoA, 0.5mM OAA	0.6	339
W260A	None	1.0	334
	0.1M KCl, 100µM CM-CoA, 0.5mM OAA	0.6	334

[a] Relative fluorescence emission at λ_{max} for each protein

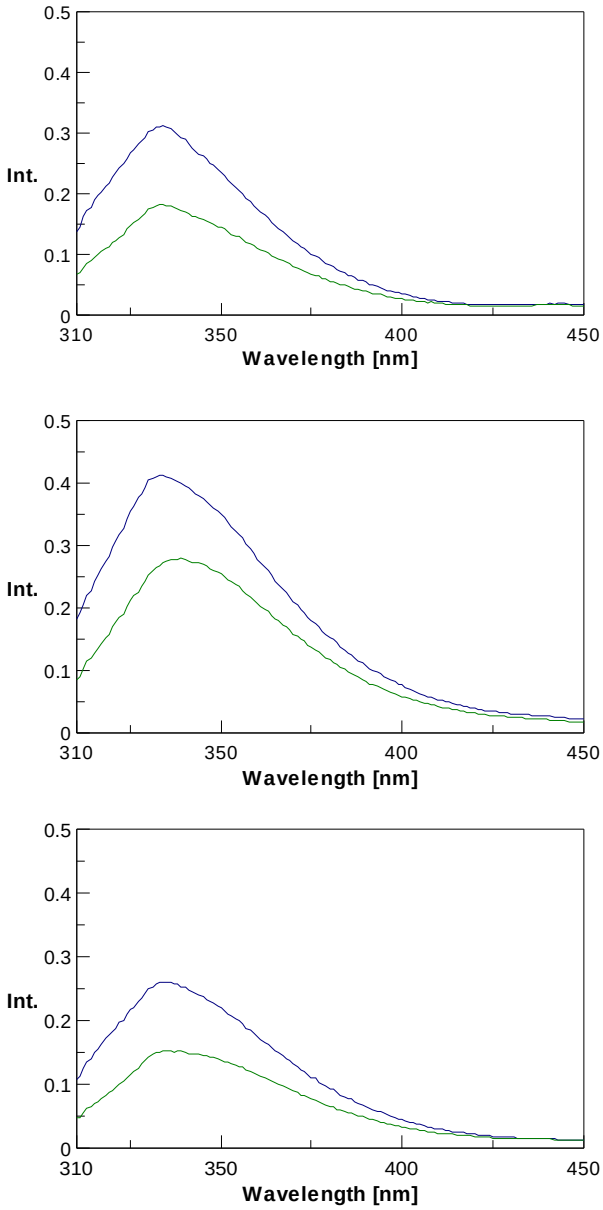


Figure 5.3: Fluorescence spectra of *E. coli* (a) CS (WT) (b) F287W and (c) W260A are shown in blue. The effect of substrates on the fluorescence emission is shown in green in each spectrum. Spectra were recorded at 25ºC using $\lambda_{excitation}$= 295nm. Fluorescence intensity was corrected for dilution.

The changes in the fluorescence in response to the titrations of OAA and CM-CoA allow K_d values to be measured (Table 5.3, Figure 5.4, 5.5)

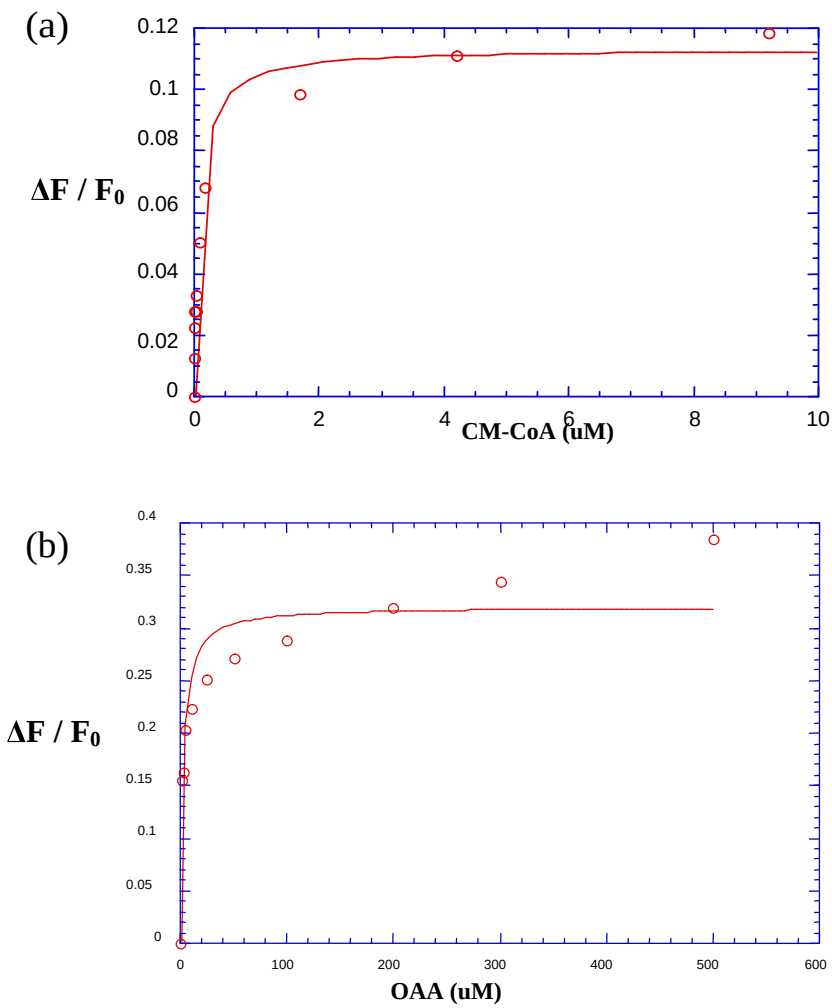


Figure 5.4: Fluorescence monitored titration of mutant F287W with (a) CM-CoA in the presence of 0.1M KCl and 100mM OAA (b) OAA in the presence of 0.1M KCl and 100mM CM-CoA. The changes in the fluorescence intensity were measured after each OAA and CM-CoA addition to a 0.29 μM solution of *E. coli* CS in standard Tris buffer (20mM Tris, 1mM EDTA). On the y-axis, F_0 represents the intensity at the start of the titration, after addition of KCl and saturating amounts of one substrate. The solid line is a least-square fit of the data to Equation 5.1.

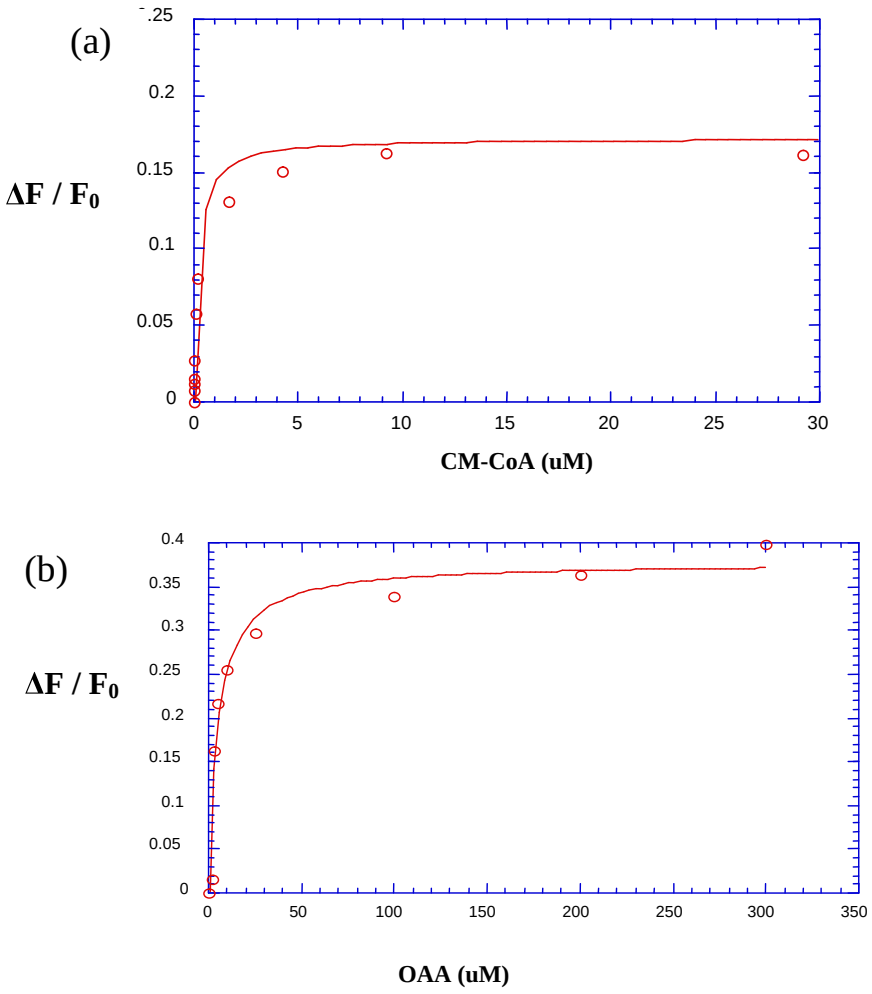


Figure 5.5: Fluorescence monitored titration of wild-type CS with (a) CM-CoA in the presence of 0.1M KCl and 100mM OAA (b) OAA in the presence of 0.1M KCl and 100mM CM-CoA. The changes in the fluorescence intensity were measured after each OAA and CM-CoA addition to a 0.26 μM solution of *E. coli* CS in standard Tris buffer (20mM Tris, 1mM EDTA). On the y-axis, F_o represents the intensity at the start of the titration, after addition of KCl and saturating amounts of one substrate. The solid line is a least-square fit of the data to Equation 5.1.

Table 5.3: Dissociation constants of WT and mutant F287W CS as measured from the change in the fluorescence upon addition of substrates

	CS(WT)	χ^2	F287W	χ^2
$K_{d,}$ OAA (μM)	5±1	0.0030	2.7±0.7	0.0005
K_d, CMCoA(μM)	0.2±0.05	0.0092	0.08±0.01	0.0103

χ^2: Goodness of Fits

5.4 Discussion

The wave-length of maximum fluorescence emission for tryptophan (λ_{max}) in water is 352 nm and decreases in more hydrophobic environments [Lakowicz, 1999]. The λ_{max} for wild-type citrate synthase lies at 332nm indicating that the tryptophans are in a hydrophobic environment. On comparison of the fluorescence spectra of wild-type protein and the W260A mutant, it was observed that W260 is more solvent exposed than the other two tryptophans. The decrease in the fluorescence intensity of wild-type during titrations with the substrates indicates that the environment of some or all of the tryptophans is changing. Furthermore, the fact that the substrates still induce substantial fluorescence quenching in the W260A mutant shows that both W391 and W395 are affected by the substrates.

The extra tryptophan inserted in the mobile loop in the F287W mutant appears to reside in a less hydrophobic environment as the λ_{max} is at 333nm which agrees with the crystal structure that the residue is solvent exposed. The mutant F287W also showed a decrease in fluorescence intensity in the presence of saturating amounts of CM-CoA, OAA and KCl. The decrease in fluorescence intensity is also accompanied by small shift in λ_{max} from 333nm to 339nm. The effect reflects a more polar conformational state for residue 287 in the presence of substrates.

The fluorescence spectra of the W260A protein showed that the tryptophans, W391 and W395 although not very close to the active site are affected by the substrates. The result is consistent with the NMR results of the protein specifically labeled with $^{15}N_2$-Trp. Thus, the effect of substrate binding is not restricted to a few parts of the protein but has global effects.

For the effects locally around the active site, the F287W probe provides a hint of a different conformational state of the mobile loop of *E. coli* CS complexed with substrates. As described previously, it has been proposed that repositioning of active site residue H264 and refolding of the mobile loop have to occur to form the binding site of the substrate Ac-CoA [Nguyen, et al, 2001]. The X-ray crystal structure of wild-type *E. coli* CS does not suggest any changes in solvent exposure of the mobile loop residues. But a comparison of the environment of residues was made between the crystal structure of mutant proteins (R109L and R319L/ R109L) which are more towards active

state (R state) and wild-type CS which is in the inactive (T-state). Residue F287 in the crystal structure of mutants R109L [Stokell et al., 2003] and R319L/R109L is even more solvent exposed as compared to the wild-type protein. The result obtained from fluorescence spectroscopy is consistent with the above comparison that in the presence of substrates, residue 287 moves towards a more polar environment.

To summarize, the work described in the thesis aimed at studying the flexible residues of *E. coli* CS particularly residues of the mobile loop. The dynamics of the mobile loop were proposed to be a key part of the allosteric mechanism of *E. coli* CS by the X-ray crystal structure studies. The mobile loop also showed high flexibility in the crystal structure, a property that has been used to study this region by NMR spectroscopy.

The results from NMR studies showed that in addition to the mobile loop residues there are other regions of high mobility in *E. coli* CS. Some of these other residues are contributed from obvious regions such as the N-terminus while others, like the side chains of tryptophans, were not predicted from the crystal structure. The study showed that the solution state of *E. coli* CS is characterized by multiple conformations of the flexible residues. Four methionine residues and leucine-275 undergoes slow conformational exchange to give multiple signals in the NMR spectrum.

The results from both NMR and fluorescence studies showed changes in the environment of residues when substrates are bound to the enzyme. In

the absence of residue-specific assignments in the NMR spectra, it can just be speculated that residues affected by substrates are coming from the mobile loop. But the fluorescence studies of mutant F287W gave evidence of conformational change of residues in the mobile loop in the presence of substrates. The results from both NMR and fluorescence spectroscopy also showed that the environment of W260 and W391/ W395 (or both), which are neither part of the binding site or the mobile loop are changed in the presence of substrates. The result showed that although conformational changes are expected near the binding site region, the effects of these conformational changes are propagated to distant residues as well.

6.0 REFERENCES

- Anderson, D. H., Donald, L. J., Jacob, M. V. & Duckworth, H. W. A mutant of Escherichia coli citrate synthase that affects the allosteric equilibrium. *Biochem Cell Biol* **69**, 232-8 (1991).
- Anderson, D. H. & Duckworth, H. W. In vitro mutagenesis of Escherichia coli citrate synthase to clarify the locations of ligand binding sites. *J Biol Chem* **263**, 2163-9 (1988).
- Ayed, A. & Duckworth, H. W. A stable intermediate in the equilibrium unfolding of Escherichia coli citrate synthase. *Protein Sci* **8**, 1116-26 (1999).
- Ayed, A., Krutchinsky, A. N., Ens, W., Standing, K. G. & Duckworth, H. W. Quantitative evaluation of protein-protein and ligand-protein equilibria of a large allosteric enzyme by electrospray ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **12**, 339-44 (1998).
- Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-65 (2000).
- Bax, A. Multidimensional nuclear magnetic resonance methods for protein studies. *Curr Opin Struct Biol* **4**, 738-744 (1994).
- Bell, G. S. et al. Stepwise adaptations of citrate synthase to survival at life's extremes. From psychrophile to hyperthermophile. *Eur J Biochem* **269**, 6250-60 (2002).
- Bhayana, V. & Duckworth, H. W. Amino acid sequence of Escherichia coli citrate synthase. *Biochemistry* **23**, 2900-5 (1984).
- Birge, E. A. & Low, K. B. Detection of transcribable recombination products following conjugation in rec+, reCB- and recC-strains of

Escherichia coli K12. *J Mol Biol* **83**, 447-57 (1974).

- Bloxham, D. P., Ericsson, L. H., Titani, K., Walsh, K. A. & Neurath, H. Limited proteolysis of pig heart citrate synthase by subtilisin, chymotrypsin, and trypsin. *Biochemistry* **19**, 3979-85 (1980).
- Bloxham, D. P. et al. Primary structure of porcine heart citrate synthase. *Proc Natl Acad Sci U S A* **78**, 5381-5 (1981).
- Bodenhausen, G. & Ruben, D. Natural abundance nitrogen-15N NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett* **69**, 185-189 (1980).
- Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638-42 (2006).
- Bruschweiler, R. & Ernst, R. R. Molecular dynamics monitored by cross-correlated cross relaxation of spins quantized along orthogonal axes. *J. Chem. Phys* **96**, 1758-1766 (1992).
- Carugo, O. & Argos, P. Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic thermal factors. *Proteins: Structure, Function and Genetics* **31**, 201-213 (1998).
- Chen, Y. & Barkley, M. D. Toward understanding tryptophan fluorescence in proteins. *Biochemistry* **37**, 9976-82 (1998).
- Chou, Y. T., Swain, J. F. & Gierasch, L. M. Functionally significant mobile regions of Escherichia coli SecA ATPase identified by NMR. *J Biol Chem* **277**, 50985-90 (2002).
- Choudhary, K., Spicer, V. L., Donald, L. J., Duckworth, H. W., Ens, W., Lowen, P. C. & Standing, K. G. Method for estimating the isotopic distributions of metabolically labeled proteins by MALDI-TOFMS: Application to NMR samples. *Anal Chem* **78**, 5419-23 (2006).
- Cierpicki, T. & Otlewski, J. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J Biomol NMR* **21**, 249-61 (2001).

- Clore, G. M. & Gronenborn, A. M. Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol* **16**, 22-34 (1998).
- Cutting, B., Strauss, A., Fendrich, G., Manley, P. W. & Jahnke, W. NMR resonance assignment of selectively labeled proteins by the use of paramagnetic ligands. *J Biomol NMR* **30**, 205-10 (2004).
- Danson, M. J. & Weitzman, P. D. Functional groups in the activity and regulation of Escherichia coli citrate synthase. *Biochem J* **135**, 513-24 (1973).
- Donald, L. J., Crane, B. R., Anderson, D. H. & Duckworth, H. W. The role of cysteine 206 in allosteric inhibition of Escherichia coli citrate synthase. Studies by chemical modification, site-directed mutagenesis, and 19F NMR. *J Biol Chem* **266**, 20709-13 (1991).
- Duckworth, H. W. & Bell, A. W. Large-scale production of citrate synthase from a cloned gene. *Can J Biochem* **60**, 1143-7 (1982).
- Duckworth, H. W. & Tong, E. K. The binding of reduced nicotinamide adenine dinucleotide to citrate synthase of Escherichia coli K12. *Biochemistry* **15**, 108-14 (1976).
- Eisenmesser, E. Z. et al. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-21 (2005).
- Ens, W. & Standing, K. G. Hybrid quadrupole/ time-of-flight mass spectrometers for analysis of biomolecules. *Methods Enzymol* **402**, 49-78 (2005).
- Ernst, R. R., Bodenhausen, G. & Wokaun, A. *Principle of magnetic resonance in one and two dimensions,* (Clarendon Press, Oxford, 1987).
- Evans, C. T., Kurz, L. C., Remington, S. J. & Srere, P. A. Active site mutants of pig citrate synthase: effects of mutations on the enzyme catalytic and structural properties. *Biochemistry* **35**, 10661-72 (1996).
- Fan, Y., McPhie, P. & Miles, E. W. Regulation of tryptophan synthase

by temperature, monovalent cations, and an allsoteric ligand. Evidence from arrhenius plots, absorption spectra, and primary kinetic isotope effects. *Biochemistry* **39**, 4692-4703 (2000).

- Faloona, G. R. & Srere, P. A. Escherichia coli citrate synthase. Purification and the effect of potassium on some properties. *Biochemistry* **8**, 4497-503 (1969).
- Fernandez, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* **13**, 570-80 (2003).
- Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207-11 (2002).
- Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. Uniform and residue-specific 15N-labeling of proteins on a highly deuterated background. *J Biomol NMR* **29**, 289-97 (2004).
- Freedberg, D. I., Wang, Y. X., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Kiso, Y. & Torchia, D. A. Flexibility and function in HIV protease: dynamics of the HIV-1 protease bound to the asymmetric inhibitor kynostatin-272 (kni-272). *J Am Chem Soc* **120**, 7916-7923 (1998).
- Gardner, K. H. & Kay, L. E. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **27**, 357-406 (1998).
- Gerike, U., Hough, D. W., Russell, N. J., Dyall-Smith, M. L. & Danson, M. J. Citrate synthase and 2-methylcitrate synthase: structural, functional and evolutionary relationships. *Microbiology* **144 (Pt 4)**, 929-35 (1998).
- Goldman, M. Interference effects in the relaxation of a pair of unlike spin-1/2 nuclei. *J. Magn. Reson* **60**, 437-452 (1984).
- Goto, N. K. & Kay, L. E. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol*

10, 585-92 (2000).

- Grzesiek, S. & Bax, A. Improved 3D triple-resonance NMR techniques applied to a 31kDa protein. *J. Magn. Reson* **96**, 432-440 (1992).
- Handford, P. A., Ner, S. S., Bloxham, D. P. & Wilton, D. C. Site-directed mutagenesis of citrate synthase; the role of the active-site aspartate in the binding of acetyl-CoA but not oxaloacetate. *Biochim Biophys Acta* **953**, 232-40 (1988).
- Hitchens, T. K, Mannerervik, B. & Rule, G.S. Disorder-to-order transition of the active site of human class Pi glutathione transferase, GST P1-1. *Biochemistry* **40**, 11660-9 (2001).
- Hwang, P. M., Bishop, R. E. & Kay, L. E. The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc Natl Acad Sci U S A* **101**, 9618-9623 (2005).
- Ikura, M., Kay, L. E. & Bax, A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659-67 (1990).
- Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nat Struct Biol* **7**, 740-743 (2000).
- Kalodimos, C. G. et al. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386-9 (2004).
- Kalodimos, C. G. et al. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386-9 (2004).
- Karpusas, M., Branchaud, B. & Remington, S. J. Proposed mechanism for the condensation reaction of citrate synthase: 1.9-A structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A. *Biochemistry* **29**, 2213-9 (1990).

- Karpusas, M., Branchaud, B. & Remington, S. J. Proposed mechanism for the condensation reaction of citrate synthase: 1.9-A structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A. *Biochemistry* **29**, 2213-9 (1990).
- Karpusas, M., Holland, D. & Remington, S. J. 1.9-A structures of ternary complexes of citrate synthase with D- and L-malate: mechanistic implications. *Biochemistry* **30**, 6024-31 (1991).
- Kay, L. E. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson* **89**, 496-514 (1990).
- Kay, L. E. Field gradient techniques in NMR spectroscopy. *Curr Opin Struct Biol* **5**, 674-81 (1995).
- Kay, L. E. NMR methods for the study of protein structure and dynamics. *Biochem. Cell Biol.* **75**, 1-15 (1997).
- Kay, L. E. Protein dynamics from NMR. *Nat Struct Biol* **NMR supplement**, 513-517 (1998).
- Kay, L. E. NMR studies of protein structure and dynamics. *J Magn Reson B* **173**, 193-207 (2005).
- Kay, L. E., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* **114**, 10663-10665 (1992).
- Kern, D. & Zuiderweg, E. R. The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* **13**, 748-57 (2003).
- Klein-Seetharaman, J. et al. Solution NMR spectroscopy of [alpha -15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc Natl Acad Sci U S A* **99**, 3452-7 (2002).
- Klein-Seetharaman, J. et al. Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. *Proc Natl Acad*

Sci U S A **101**, 3409-13 (2004).

- Kornberg, H. L. The role and control of the glyoxylate cycle in Escherichia coli. *Biochem J* **99**, 1-11 (1966).
- Kosen, P. A. Spin labeling of proteins. *Methods Enzymol* **177**, 86-121 (1989).
- Kreishman-Deitrick, M. et al. NMR analysis of methyl groups at 100-500 kDa: model systems and Arp2/ 3 complex. *Biochemistry* **42**, 8579-86 (2003).
- Ladokhin, A. S. in *Encyclopedia of Analytical Chemistry* (ed. Meyers, R. A.) 5762-5779 (John Wiley & Sons Ltd, Chichester, 2000).
- Lakowicz, J. R. *Principles of Fluorescence Spectroscopy* (Kluwer Academic Publishers, New York, 1999).
- Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. Characterization of a functionally important mobile domain of GroES. *Nature* **364**, 255-8 (1993).
- Leiting, B., Marsilio, F. & O'Connell, J. F. Predictable deuteration of recombinant proteins expressed in Escherichia coli. *Anal Biochem* **265**, 351-5 (1998).
- LeMaster, D. M. & Richards, F. M. 1H-15N heteronuclear NMR studies of Escherichia coli thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263-8 (1985).
- Lesk, A. M. & Chothia, C. Mechanisms of domain closure in proteins. *J Mol Biol* **174**, 175-91 (1984).
- Lian, L. & Middleton, D. A. Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **39**, 171-190 (2001).
- Liao, D. I., Karpusas, M. & Remington, S. J. Crystal structure of an open conformation of citrate synthase from chicken heart at 2.8-A resolution. *Biochemistry* **30**, 6031-6 (1991).

- Man, W. J., Li, Y., O'Connor, C. D. & Wilton, D. C. The effect of replacing the conserved active-site residues His-264, Asp-312 and Arg-314 on the binding and catalytic properties of Escherichia coli citrate synthase. *Biochem J* **300 (Pt 3)**, 765-70 (1994).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. Molecular Cloning: a laboratory manual. Cold spring Harbor Laboratory, Cold Spring Harbor, NY, 1982
- Markley, J., Lu, M. & Bracken, C. J. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-75 (2001).
- Markus, M. A., Dayie, K. T., Matsudaira, P. & Wagner, G. Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on villin 14T. *J Magn Reson B* **105**, 192-5 (1994).
- Marius Clore, G. & Gronenborn, A. M. Determining the structures of large proteins and protein complexes by NMR *Tibtech* **16**, 22-33 (1998)
- Maurus, R., Nguyen, T. N., Stokell, D. J., Ayed, A., Hultin, P. G., Duckworth, H. W. & Brayer, G. D. Insights into the evolution of allosteric properties. The NADH binding site of hexameric type II citrate synthases. *Biochemistry* **42**, 5555-65 (2003).
- McElheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J. & Wright, P. E. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc Natl Acad Sci U S A* **102**, 5032-7 (2005).
- McIntosh, L. P. & Dahlquist, F. W. Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys* **23**, 1-38 (1990).
- Mittermaier, A. & Kay, L. E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224-228 (2006).
- Monod, D., Wyman, J. & Changeux, J. P. On the nature of allosteric transitions: a plausible model. *J Mol Biol* **12**, 88-118 (1965).

- Montelione, G. T. & Wagner, G. Accurate measurements of homonuclear HN-H(alpha) coupling constants in polypeptides using heteronuclear 2D NMR experiments. *J Am Chem Soc* **111**, 5474-5475 (1989).
- Montelione, G. T. & Wagner, G. Conformation-independent sequential NMR connections in isotope-enriched polypeptides by 1H---13C---15N triple-resonance experiments. *J. Magn. Reson* **87**, 183-188 (1990).
- Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J Am Chem Soc* **101**, 760-762 (1979).
- Morse, D. & Duckworth, H. W. A comparison of the citrate synthases of Escherichia coli and Acinetobacter anitratum. *Can J Biochem* **58**, 696-706 (1980).
- Nelson, S. W., Iancu, C. V., Choe, J. Y., Honzatko, R. B. & Fromm, H. J. Tryptophan fluorescence reveals the conformational state of a dynamic loop in recombinant porcine fructose-1,6-bisphosphatase. *Biochemistry* **39**, 11100-6 (2000).
- Nguyen, N. T., Maurus, R., Stokell, D. J., Ayed, A., Duckworth, H. W., & Brayer, G. D. Comparative analysis of folding and substrate binding sites between regulated hexameric type II citrate synthases and unregulated dimeric type I enzymes. *Biochemistry* **40**, 13177-87 (2001).
- Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y., Ishii, S. & Sarai, A. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat Struct Biol* **3**, 178-87 (1996).
- Pereira, D. S., Donald, L. J., Hosfield, D. J. & Duckworth, H. W. Active site mutants of Escherichia coli citrate synthase. Effects of mutations on catalytic and allosteric properties. *J Biol Chem* **269**, 412-7 (1994).
- Perham, R. N., Duckworth, H. W. & Roberts, G. C. Mobility of

polypeptide chain in the pyruvate dehydrogenase complex revealed by proton NMR. *Nature* **292**, 474-7 (1981).

- Pervushin, K. Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q Rev Biophys* **33**, 161-97 (2000).
- Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-71 (1997).
- Ramesh, V., Frederick, R. Syed, S. E. H., Gibson, C. F., Yang, J-C. & Roberts, G. C. K. The interactions of Escherichia coli trp repressor with tryptophan and with an operator oligonucleotide. NMR studies using selectively 15N-labelled protein. *Eur J Biochem* **225**, 601-8 (1994).
- Remington, S., Wiegand, G. & Huber, R. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 A resolution. *J Mol Biol* **158**, 111-52 (1982).
- Remington, S. J. Structure and mechanism of citrate synthase. *Curr Top Cell Regul* **33**, 209-29 (1992).
- Riek, R., Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. Solution NMR techniques for large molecular and supramolecular structures. *J Am Chem Soc* **124**, 12144-53 (2002).
- Riek, R., Wider, G., Pervushin, K. & Wuthrich, K. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc Natl Acad Sci U S A* **96**, 4918-23 (1999).
- Roosild, T. P. et al. NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* **307**, 1317-21 (2005).
- Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. The crystal structure of citrate synthase from the hyperthermophilic

archaeon pyrococcus furiosus at 1.9 A resolution. *Biochemistry* **36**, 9983-94 (1997).

- Russell, R. J., Hough, D. W., Danson, M. J. & Taylor, G. L. The crystal structure of citrate synthase from the thermophilic archaeon, Thermoplasma acidophilum. *Structure* **2**, 1157-67 (1994).
- Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W. & Taylor, G. L. Structure adaptations of the cold-active citrate synthase from an antartic bacterium. *Structure* **6,** 351-61 (1998).
- Rudiger, S., Freund, S. M., Veprintsev, D. B. & Fersht, A. R. CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90. *Proc Natl Acad Sci U S A* **99**, 11085-90.
- Rychlik, W. & Rhoads, R. E. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res* **17**, 8543-51 (1989).
- Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. NMR assignment and secondary structure determination of an octameric 110kDa protein using TROSY in triple resonance experiments. *J Am Chem Soc* **122**, 7543-7548 (2000).
- Schnell, J. R., Dyson, H. J. & Wright, P. E. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu Rev Biophys Biomol Struct* **33**, 119-40 (2004).
- Snijders, A. P., de Vos, M. G. & Wright, P. C. Novel approach for peptide quantitation and sequencing based on 15N and 13C metabolic labeling. *J Proteome Res* **4**, 578-85 (2005).
- Son, S., Tanrikulu, C. & Tirrell, D. A. Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues. *ChemBioChem* **7**, 1251-1257 (2006).
- Sprangers, R., Gribun, A., Hwang, P. M., Houry, W. A. & Kay, L. E. Quantitative NMR spectroscopy of supramolecular complexes:

dynamic side pores in ClpP are important for product release. *Proc Natl Acad Sci U S A* **102**, 16678-83 (2005).

- Srere, P. A., Brazil, H. & Gonen, L. The citrate condensing enzyme of peigon breast muscle and moth flight muscle. *Acta Chem. Scand* **17**, S129-S134 (1963).
- Stewart, R. C. Analysis of ATP binding to CheA containing tryptophan substitutions near the active site. *Biochemistry* **44**, 4375-85 (2005).
- Stokell, D. J., Donald, L. J., Maurus, R., Nguyen, N. T., Sadler, G., Choudhary, K., Hultin, P. G., Brayer, G. D. & Duckworth, H. W. Probing the roles of key residues in the unique regulatory NADH binding site of type II citrate synthase of Escherichia coli. *J Biol Chem* **278**, 35435-43 (2003).
- Talgoy, M. M., Bell, A. W. & Duckworth, H. W. The reactions of Escherichia coli citrate synthase with the sulfhydryl reagents 5,5'-dithiobis-(2-nitrobenzoic acid) and 4,4'-dithiodipyridine. *Can J Biochem* **57**, 822-33 (1979).
- Tong, E. K. & Duckworth, H. W. The quaternary structure of citrate synthase from Escherichia coli K12. *Biochemistry* **14**, 235-41 (1975).
- Tugarinov, V. & Kay, L. E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125**, 13868-78 (2003).
- Tugarinov, V., Muhandiram, R., Ayed, A. & Kay, L. E. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* **124**, 10025-35 (2002).
- Tzakos, A. G., Grace, C. R., Lukavsky, P. J. & Riek, R. NMR techniques for very large proteins and rnas in solution. *Annu Rev Biophys Biomol Struct* **35**, 319-42 (2006).
- Vaughan, M. D., Cleve, P., Robinson, V., Duewel, H. S. & Honek, J. F.

Difluoromethionine as a novel 19F NMR structural probe for internal amino acid packing in proteins. *J Am Chem Soc* **121**, 8475-8478 (1999).

- Vendruscolo, M. & Dobson, C. M. Structural biology. Dynamic visions of enzymatic reactions. *Science* **313**, 1586-7 (2006).
- Venters, R. A., Thompson, R. & Cavanagh, J. Current approaches for the study of large proteins by NMR. *Journal of Molecular Structure* **602-603**, 275-292 (2002).
- Volkert, T. L., Baleja, J. D. & Kumamoto, C. A. A highly mobile C-terminal tail of the Escherichia coli protein export chaperone SecB. *Biochem Biophys Res Commun* **264**, 949-54 (1999).
- Weigelt, J., van Dongen, M., Uppenberg, J., Schultz, J. & Wikstrom, M. Site-selective screening by NMR spectroscopy with labeled amino acid pairs. *J Am Chem Soc* **124**, 2446-7 (2002).
- Weitzman, P. D. Regulation of citrate synthase activity in escherichia coli. *Biochim Biophys Acta* **128**, 213-5 (1966).
- Weitzman, P. D. & Danson, M. J. Citrate synthase. *Curr Top Cell Regul* **10**, 161-204 (1976).
- Weitzman, P. D. & Jones, D. Regulation of citrate synthase and microbial taxonomy. *Nature* **219**, 270-2 (1968).
- Wiegand, G. & Remington, S. J. Citrate synthase: structure, control, and mechanism. *Annu Rev Biophys Biophys Chem* **15**, 97-117 (1986).
- Wiegand, G. & Remington, S. J. Citrate synthase: structure, control, and mechanism. *Annu Rev Biophys Biophys Chem* **15**, 97-117 (1986).
- Wishart, D. S. et al. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **6**, 135-40 (1995).
- Wuthrich, K. *NMR of proteins and nucleic acids* (Wiley, New York, 1986).
- Wuthrich, K. The second decade--into the third millenium. *Nat Struct Biol* **5 Suppl**, 492-5 (1998).

- Wuthrich, K., Wider, G., Wagner, G. & Braun, W. Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J Mol Biol* **155**, 311-9 (1982).
- Yabuki, T. et al. Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J Biomol NMR* **11**, 295-306 (1998).
- Yamazaki, T., Muhandiram, R. D. & Kay, L. E. NMR experiments for themeasurement of carbon relaxation properties in highly enriched, uniformly 13C, 15N-labeled proteins: application to 13C(alpha) carbons. *J Am Chem Soc* **116**, 8266-8278 (1994).
- Zhao, Q. et al. Production in two-liter beverage bottles of proteins for NMR structure determination labeled with either 15N- or 13C-15N. *J Struct Funct Genomics* **5**, 87-93 (2004).

Appendix I: List of sequence assigned resonances and specific residues assigned in *E. coli* CS

^{1}H (ppm)	^{15}N (ppm)	Sequence Assigned Residues	Specific Residue Assigned
9.57	120.55	Phenylalanine	
8.08	120.55	Phenylalanine	
7.75	125.89	Methionine	
7.66	125.60	Methionine	
7.49	122.26	Methionine	
7.29	120.06	Methionine	
7.9	122.2	Methionine	
7.94	116.52	Methionine	
8.05	116.55	Methionine	
8.27	121.77	Methionine	
8.36	122.87	Methionine	
8.73	123.34	Methionine	
8.62	122.78	Methionine	
8.49	120.52	Methionine	
8.17	118.86	Methionine	
8.09	118.63	Methionine	
8.22	125.57	Methionine	
7.87	129.09	Methionine	
8.92	122.80	Leucine	
8.84	128.29	Leucine	
8.34	120.16	Leucine	
8.03	117.91	Leucine	
7.95	121.86	Leucine	
7.69	120.92	Leucine	
7.61	121.81	Leucine	
7.61	126.74	Leucine	
7.51	120.77	Leucine	
7.82	120.60	Leucine	
8.19	124.38	Leucine	L275
8.03	124.48	Leucine	L275
8.23	106.63	Glycine	
7.82	101.69	Glycine	
7.76	106.35	Glycine	
7.81	114.27	Glycine	
8.16	108.65	Glycine	

8.25	109.06	Glycine	
8.35	112.02	Glycine	
8.41	106.76	Glycine	
8.40	109.98	Glycine	
8.58	114.12	Glycine	
8.86	133.64	Lysine	
8.94	122.58	Lysine	
8.60	130.37	Lysine	
8.68	120.17	Lysine	K 4 or 6
8.25	120.17	Lysine	
8.21	120.27	Lysine	
8.18	119.31	Lysine	
8.12	125.17	Lysine	
8.19	127.48	Lysine	
7.86	118.15	Lysine	
7.58	119.69	Lysine	
7.77	122.18	Lysine	
7.12	114.50	Lysine	
7.78	131.91	Lysine	K 4 or 6
8.31	120.03	Histidine	
8.54	122.08	Histidine	
7.95	121.89	Histidine	
9.97	129.33	Tryptophan	
10.01	129.39	Tryptophan	
10.05	129.60	Tryptophan	W260
10.08	129.70	Tryptophan	