Stability and Structure of *Escherichia coli* Citrate Synthase

Ayeda Ayed

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Department of Chemistry
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

STABILITY AND STRUCTURE OF *Escherichia coli* CITRATE SYNTHASE

BY

AYEDA AYED

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Acknowledgments

I am grateful to my supervisor Dr. H. W. Duckworth for his guiding advice, patience, and encouragement throughout my graduate work.

I thank the members of my advisory committee, Dr. J. C. Jamieson, Dr. Joe O'Neil, and Dr. B. Worobec.

I have had the pleasure of collaborating with Andrew Krutchinsky and Dr. K. G. Standing of the Department of Physics, and for that I extend my appreciation.

My sincere thanks are due to Dr. Lynda Donald for her encouragement, help, and friendship. I would also like to acknowledge past and present students of Dr. Duckworth's group, especially Gillian Sadler for her friendship, Lorraine Lamaga and Christos Konstantatos for their technical assistance.

Dr. L. Krucszynski was most helpful in interfacing a fluorimeter and a spectropolarimeter with computers for data acquisition, and for that I thank him.

Finally, I acknowledge The University of Manitoba for financial assistance.

Abstract

Escherichia coli citrate synthase (CS) is a hexamer of identical subunits which is allosterically inhibited by NADH. Although its three-dimensional structure has not yet been determined, a satisfactory model of the subunit fold can be constructed using the homologous vertebrate CS, a dimer whose structure is known, as a guide. Equilibrium unfolding of *E. coli* CS by chaotropic agents or heat, whether probed by circular dichroism or by fluorescence, occurs in two stages, with an intermediate (I state) that is stable between 2.5 and 5.5 M urea, approximately. The two transitions are associated with 67% and 33% loss of structure, respectively. Only the second transition is reversible. Urea-gradient gel electrophoresis and electrospray ionization-time-of-flight mass spectrometry (ESI-TOF MS) were used to show that the I state is a collection of aggregates. *Acinetobacter anitratum* CS, also allosteric and hexameric, unfolds via one transition, unlike *E. coli* CS. Unfolding experiments with CS chimeras, in which one structural domain came from *E. coli* CS and the other from *A. anitratum* CS, suggested that the existence of an I state is associated with an *E. coli* large domain. Cavity-creating mutations were introduced into the cores of the two domains of *E. coli* CS. The mutant proteins also showed biphasic unfolding. Some mutants exhibited a stabilized first transition, a behaviour that can be explained by destabilization of the I state. The secondary structure features in which these mutations are located form the cores of the two domains. A model of the I state constructed from these results shows a mostly hydrophobic surface, consistent with the tendency of the intermediate to aggregate.

ESI-TOF MS was used to study the quaternary structure of CS and also CS-ligand interactions. CS exhibits an equilibrium between dimers and hexamers that is pH and protein-concentration dependent. The association constant (K_A) for the dimer-hexamer equilibrium was determined, and it was shown that NADH binds strongly only to hexamer and converts dimers to hexamers. The dissociation constants (K_D) for NADH binding to dimer and hexamer were calculated from the mass spectrometric data. This study establishes a relationship between quaternary structure and NADH binding in *E. coli* CS.

Table of Contents

List of Abbreviations

AcCoA	Acetyl coenzyme A
ACI::eco	Chimera with *A. anitratum* large domain and *E. coli* small domain
ANS	8-Anilino-naphthalene sulfonic acid
ATP	Adenosinetriphosphate
bp	Base pair
CD	Circular dichroism
CoA	Coenzyme A (same as CoASH)
CoASH	Coenzyme A (same as CoA)
CS	Citrate synthase
DEAE	Diethyl amino ethyl
DTNB	5,5'-Dithiobis-(2-nitrobenzoic acid)
ECO::aci	Chimera with *E. coli* large domain and *A. anitratum* small domain
EDTA	ethylenediaminetetraacetic acid
ESI	Electrospray ionization
$FADH_2$	Flavin adenine dinucleotide, reduced form
K_A	Association constant
K_D	Dissociation constant
m	Slope of unfolding transition
m/z	mass-to-charge ratio
MALDI	Matrix assisted laser desorption
MG	Molten globule
MRE	Mean residue ellipticity
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut-off
NADH	Nicotinamide adenine dinucleotide, reduced form
NMR	Nuclear magnetic resonance
OAA	Oxaloacetate
RNA	Ribonucleic acid
RNase	Ribonuclease
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TCA	Tricarboxylic acid
TFBA	1,1,1-Trifluoro-1-bromoacetone
TM	Transition midpoint
TOF	Time-of-flight
TRIS	Tris-(hydroxymethyl)aminomethane
u	units
UV	Ultraviolet
WT	Wild type

List of Figures

List of Tables

CHAPTER ONE Citrate Synthase

Introduction

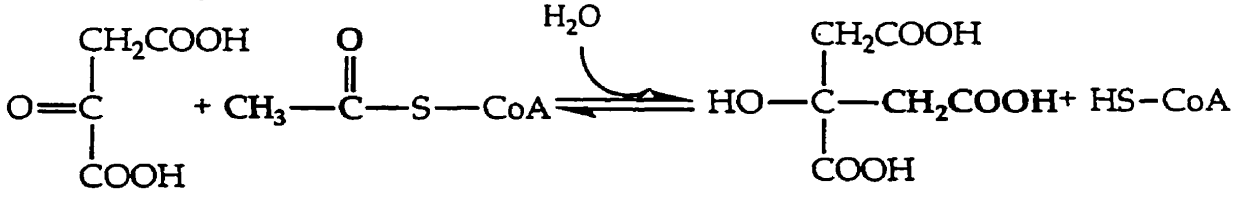
General introduction

The subject of this thesis is citrate synthase (CS) from *Escherichia coli*, an allosteric, multisubunit enzyme. Although no three-dimensional structure is available for this enzyme, it is homologous to the dimeric, non-allosteric class of CS for which several crystal structures are available. In those structures, the subunits in the dimer fold into two structural domains, and this information was used previously (Molgat *et al.*, 1991) to construct chimeric proteins involving the *E. coli* and *A. anitratum* enzymes. Studies on the unfolding of the chimeras, as well as the parent proteins, suggested that a fuller study of the unfolding of *E. coli* CS would be of interest. The first project described in this work (Chapter 2) involved the study of the equilibrium unfolding of this protein together with mutants of the *E. coli* CS designed to understand its biphasic unfolding behaviour. The main conclusion of this work was that unfolding of *E. coli* CS involves a self-associating intermediate which retains structured elements from the cores of the two domains of the molecule. The second project (Chapter 3) was the study of the quaternary structure of *E. coli* CS and its relationship to allosteric inhibition by NADH. Binding of NADH and substrates to CS was observed using electrospray ionization-time-of-flight mass spectrometry (ESI-TOF MS), and constants for NADH binding to dimer and hexamer were determined from the mass spectra. This study established a clear relationship between the quaternary structure of the enzyme and allosteric binding and inhibition. This chapter introduces CS, placing emphasis on the regulation of

the *E. coli* enzyme and on structural features as they can be inferred from X-ray crystallography studies of homologous citrate synthases.

Introduction to citrate synthase

Citrate synthase catalyzes the condensation of an acetyl group from acetyl Coenzyme A (AcCoA) with oxaloacetate (OAA) to form citrate:



Oxaloacetate Acetyl Coenzyme A Citrate Coenzyme A

This step is the point of entry of two carbons into the tricarboxylic acid cycle (TCA, Figure 1.1). The acetyl group arises from glycolysis, and β–oxidation of fatty acids. The TCA cycle is the second stage of aerobic respiration, serving catabolic and biosynthetic functions as it completely oxidizes the acetyl group to CO_2. First, it stores chemical energy in the forms of reduced nicotinamide adenine dinculeotide (NADH), reduced falvin dinonucleotide ($FADH_2$), and adenosine 5'-triphosphate (ATP). NADH and $FADH_2$ are processed by the electron transport chain, the third step of aerobic respiration, where substrate level phosphorylation produces ATP. Second, the intermediates of the TCA cycle serve as building blocks for biosynthesis of amino acids and other materials.

CS is also the first enzyme of the glyoxylate shunt used in some organisms (plants and bacteria) as an anaplerotic pathway, shown in Figure 1.2. This pathway ensures the production of TCA intermediates which are important for biosynthesis and gluconeogenesis (the biosynthesis of carbohydrates).

In *E. coli*, a variant, non-cyclic form of the TCA cycle is found that operates in some facultative anaerobes (Figure 1.3). In this pathway, the enzyme fumarate reductase, expressed mainly under anaerobic conditions,

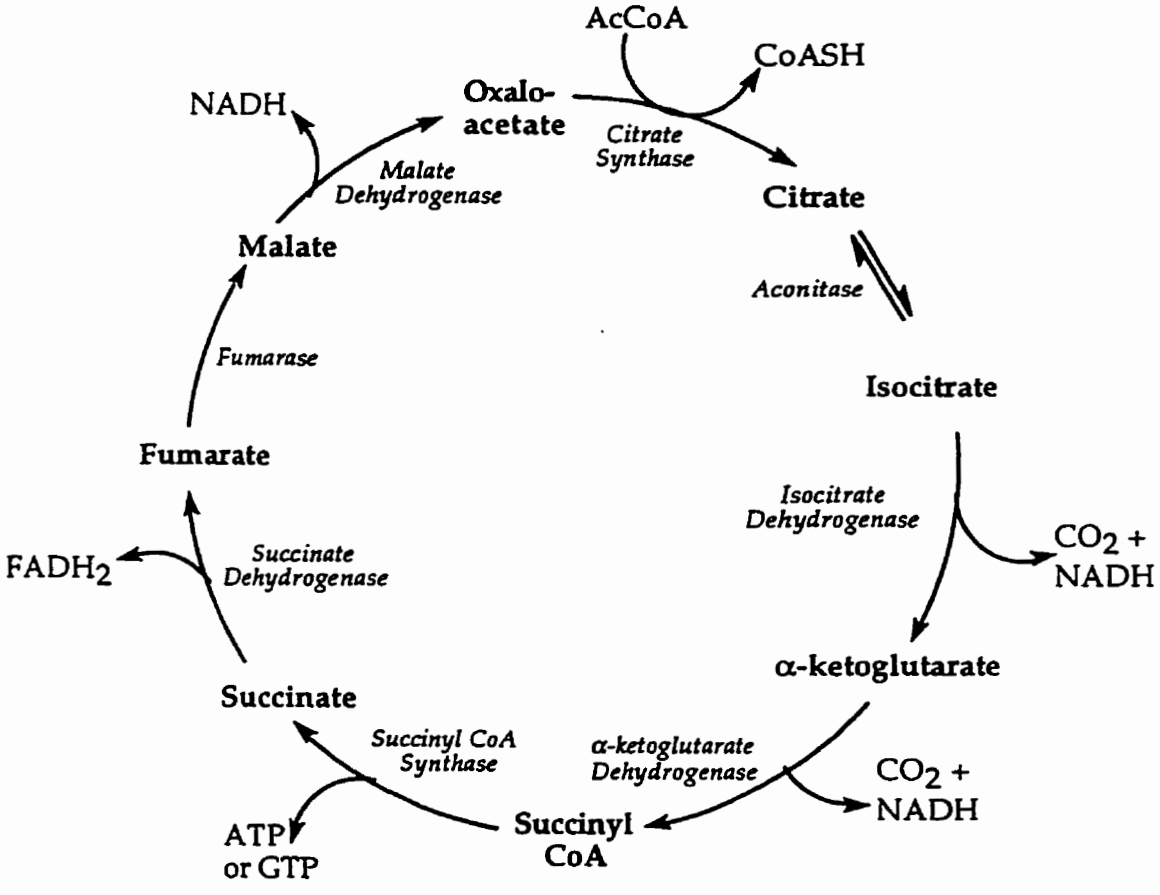


Figure 1.1 The tricarboxylic acid cycle (TCA). The enzyme names are shown in italics. One turn of the cycle oxidizes the acetyl group to two CO_2 molecules, producing 3 NADH, 1 $FADH_2$ and 1 ATP.

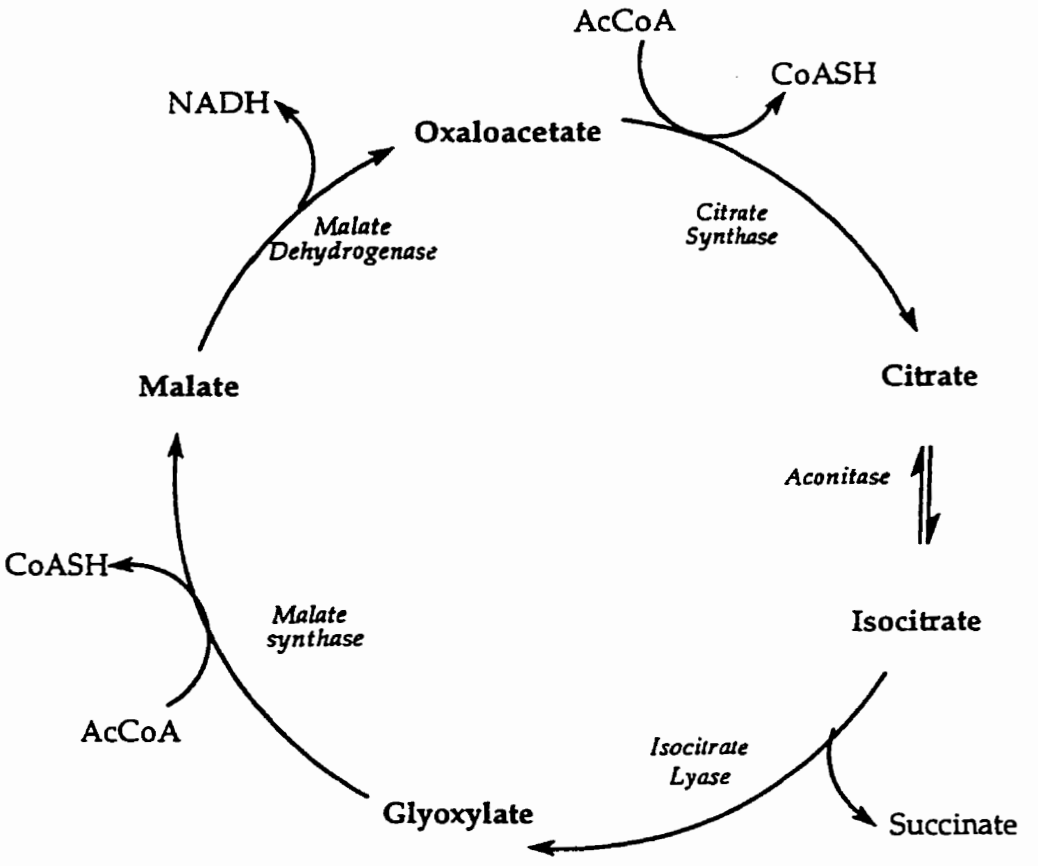


Figure 1.2 The glyoxylate cycle. Enzyme names are shown in italics. This pathway acts mainly anaplerotically, producing little energy.

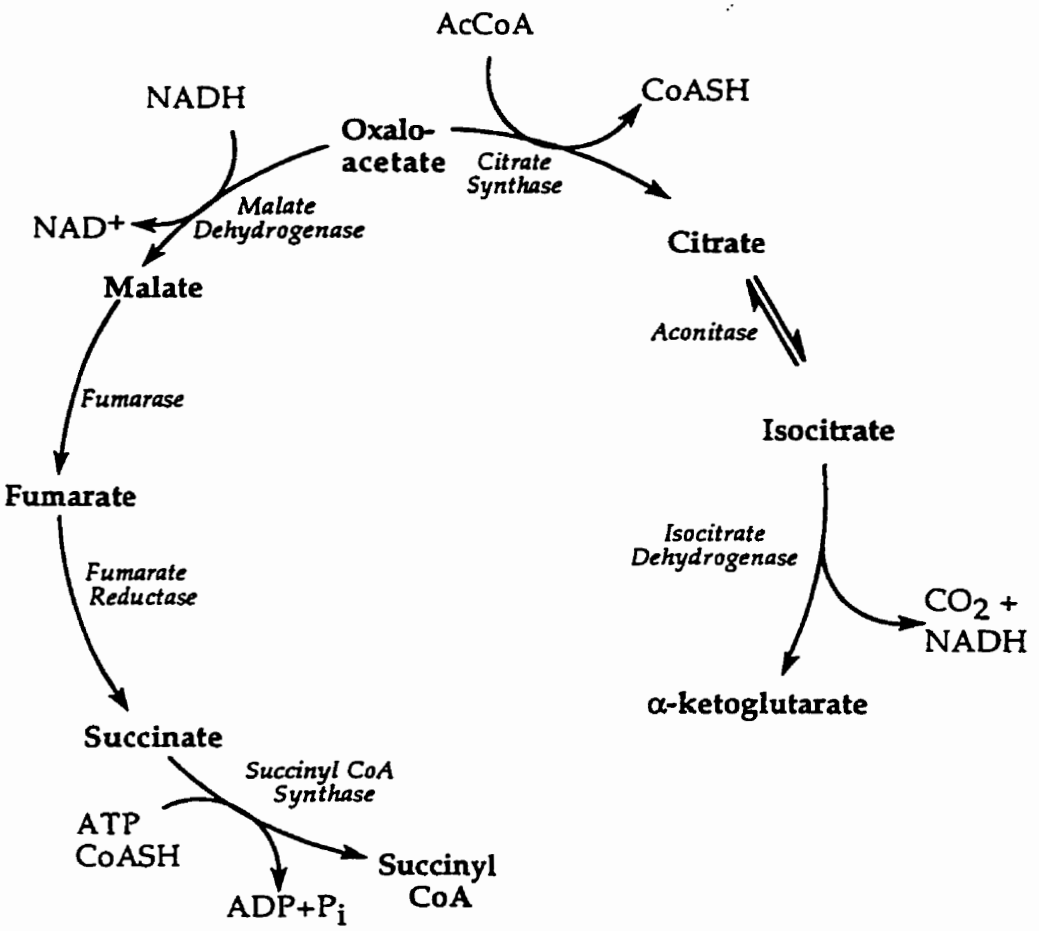


Figure 1.3 The non-cyclic TCA cycle. Enzyme names are italicized.

reverses part of the TCA cycle. This pathway functions anaplerotically, allowing the formation of TCA intermediates required for biosynthesis.

In view of the importance of citrate synthase in the above pathways, it is not a surprise that it is subject to regulation by a variety of molecules. ATP, an indicator of a surplus in cell energy, inhibits citrate synthase in eukaryotic organisms by binding to the CoA binding site, thereby acting as an isosteric inhibitor. Similarly, α-ketoglutarate, an analogue of OAA, inhibits CS isosterically in some Gram-negative bacteria. NADH, a product of the TCA cycle, inhibits citrate synthase of Gram negative bacteria by an allosteric mechanism.

Regulation of Citrate Synthase

Classes of citrate synthase

The citrate synthases have been classified according to their mode of regulation and also their subunit structure. These two features are intimately related. Classically, citrate synthases are divided into two main groups (Weitzman & Danson, 1976, Weitzman, 1966, Weitzman & Jones, 1968): the non-allosteric, dimeric enzymes, and the allosteric CSs which form higher order aggregates (usually hexamers) of the basic dimeric structure. Generally, CS from eukaryotes and archaea are insensitive to NADH inhibition and are dimeric, belonging to the former class, while those from Gram negative bacteria exhibit allosteric inhibition by NADH, and are hexameric, belonging to the latter class. *E. coli* is a Gram negative bacterium and its CS, encoded by *gltA* (see Appendix II) (Guest, 1981, Ner *et al.*, 1983), belongs to the allosteric, hexamer-forming class of enzyme (Weitzman & Danson, 1976), for which NADH inhibition is strong (Weitzman & Jones, 1968). On the other hand, pig

CS, for which several three-dimensional structures are available (see later) is non-allosteric and dimeric.

Recently, several organisms were found to have second and third CS genes, the sequences of which more closely resemble the non-allosteric dimeric class of CS. For example, the recent completion of the *E. coli* CS genome resulted in the discovery of a gene which encodes a protein with a sequence 29.5 % identical to the *gltA*-encoded CS. All the residues which were identified by crystallography and sequence alignments as essential for substrate binding and catalysis were conserved. Prior to the completion of the *E. coli* genome sequence, evidence had been found for a second CS enzyme which was dimeric and immunologically distinct from the *gltA* product (Patton *et al.*, 1993). Later, it was shown that this enzyme, the product of CSII as revealed by the *E. coli* genome sequence, is in fact not a citrate synthase, but a methylcitrate synthase, which utilizes propionyl CoA as a substrate (Textor *et al.*, 1997).

Other organisms such as *Pseudomonas aeroginosa, Bacillus subtilis,* and yeast have been shown to have second and third CS genes; they are classified currently as citrate synthases in the database entries. It is possible that they are all citrate synthases with different regulatory properties or expression conditions. However, the discovery that the *E. coli* "second CS" is really a methylcitrate synthase may indicate that a larger class of synthase enzymes may be present which utilize different CoA substrates or are subject to different regulatory modes.

Citrate synthase sequences

The amino acid sequences of citrate synthases from many organisms are available. The Swiss-protein database (up to release October 1997) has over

60 different CS sequences from eukaryotes, prokaryotes and archaea (see Appendix I). Most are 380 to 430 amino acids long, making the average subunit molecular weight about 40 to 50 KDa. Those sequences in the Swiss-protein database as at November 1997 are shown in Appendix I as an alignment designed to show homologies. The large number of partial CS sequences are from the *Bartonella* and *Rickettsia* genera, as *gltA* has been used extensively in the identification of different strains of these organisms in clinical situations (Birtles & Raoult, 1996).

The alignment in Appendix I was generated by Maxhom, a program for sequence analysis (Rost & Sander, 1993). The order of the sequences is such that those of highest sequence identity to *E. coli* CS are near the top of the page, while those of lowest identity are nearest the bottom of the page. Those sequences of closest identity to the *E. coli* enzyme tend to be those from prokaryotic species, while those of lowest identity to CS are from eukaryotes or the archaea. This pattern is consistent with the class assignment based on the mode of regulation of CS, as discussed above. Also of interest are those sequences which are encoded by "second" or "third" CS genes present in some species, which were included in the alignment and show low identity with the *gltA*-encoded CS from the same organisms.

The alignment shows several regions where obviously low identity occurs, usually associated with the introduction of gaps to optimize the alignment. Obvious regions include the N-terminus, residues 200-230, 280-300, and 330-360. However, all of the residues which have been identified as catalytically important or essential for substrate binding by crystallography and other methods are 100% conserved in all sequences, including those of CSII and CSIII which may be synthases which utilize different CoA substrates, as

was shown for *E. coli* CSII (Textor *et al.*, 1997). Those residues involved in binding and catalysis will be discussed in more detail later for the *E. coli* and pig citrate synthases, as these represent the best studied examples, both structurally and functionally, from each of the two classes of the enzyme.

In conjunction with the known three-dimensional structures of homologous CSs (described below), the sequence alignment was used in this study to identify conserved, hydrophobic, and buried residues which may be important for the stability of *E. coli* CS. Those chosen on the basis of the sequence alignment and examination of the homologous three-dimensional structure were replaced with alanine by site-directed mutagenesis to destabilize *E. coli* CS deliberately in specific regions.

Three-dimensional Structures

To study *E. coli* CS folding and to perform site-directed mutagenesis studies, information about its three-dimensional structure would be helpful. However, a structure is not yet available for *E. coli* CS, or any of the allosteric CS enzymes; only dimeric CS three-dimensional structures are available. The existing homologous structures, therefore, are used as models of the *E. coli* CS structure. This section describes some features of the known structures of CS.

Crystallographic citrate synthase structures

Three-dimensional structures of citrate synthases from four different species have been elucidated by X-ray crystallography: *Sus scrofa* (pig) heart (Remington *et al.*, 1982), *Gallus gallus* (chicken) (Liao *et al.*, 1991), *Thermoplasma acidophilum* (a thermophilic archaeon)(Russell *et al.*, 1994), and *Pyrococcus furiosus* (a hyperthermophilic archaeon) (Russel *et al.*, 1997). All structures determined to date are homodimers related by a two-fold axis, with an overall fold that is classified as an all alpha motif (Figure 1.4). Each

monomer in the dimeric unit folds into two distinct domains, with the active site situated in a deep cleft between them. As described in later sections, the small domain moves relative to the large domain when substrates are present; CS is often regarded as a classic induced-fit enzyme due to the large conformational change that it undergoes upon substrate binding.

The marked similarity of the X-ray crystal structures of CS from eukaryotes to those recently determined for *T. acidophilum* and *P. furiosus* (from archaea), despite sequence identities of only 20% between the two classes of organisms, suggests that the use of the pig CS structure as the basis for a model for *E. coli* CS, is valid (Duckworth *et al.*, 1987); the *E. coli* and pig CS sequences are about 30% identical. Therefore, for the description of the CS structure in the following section, the pig CS structure is used; the three-dimensional structure was used to identify buried residues in the hydrophobic core for site-directed mutagenesis of the homologous residues in *E. coli* CS.

Secondary structural elements

All CS crystal structures determined to date are homodimers. For the pig enzyme, each monomer is composed of 20 α-helices, with 315 residues out of 437 being involved in secondary structure. A 12-residue antiparallel βsheet (residues 56-69) is the only non α-helical secondary structure present. The remaining regions form extended loops which appear to be shortened in the thermophilic organisms (Russell *et al.*, 1994). Figure 1.5 shows the secondary structure elucidated for three CSs based on the crystal structures. The *E. coli* sequence, aligned against the pig CS and the predicted secondary structure (using the PHD algorithm for secondary structure prediction (Rost &

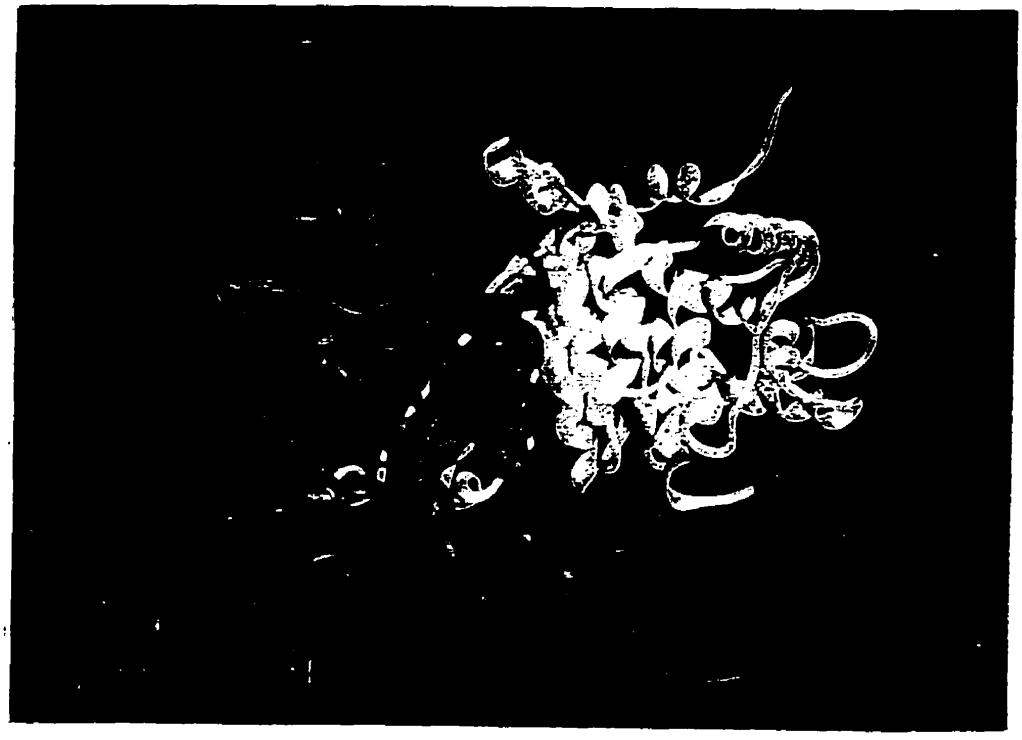


Figure 1.4 Ribbon diagram of the pig CS dimer, with the two-fold axis perpendicular to the page surface. Large domains are coloured yellow and purple, small domains green and red.

Sander, 1993)), is shown in Figure 1.6. The predicted secondary structure is in good agreement with those deduced previously by simple alignment with the pig structure (not shown). In pig CS, 72% of the residues are involved in secondary structure. For the *E. coli* CS model, the corresponding value is 63% based on the alignment (270 residues out of 426).

The large domain and dimer contact

The large domains (residues 1-274 and 381-437 in pig CS) mediate dimer contact. The structure of the large domains in the dimeric CS can be described as three layers made up of antiparallel pairs of α-helices. The innermost monomer-monomer contact layer is composed of four antiparallel pairs of helices, forming an 8-helical sandwich (Figure 1.7). The helices mediating this interaction, F G M L, interact with the two-fold related helices F′ G′ M′ L′ by hydrophobic residue packing, a few salt bridges and hydrogen bonds (the prime used here and throughout this thesis refers to the other subunit). The dimeric structure of CS is the minimal functional unit, and the thermophilic counterparts have stabilized this interface by the introduction of additional electrostatic interactions and isoleucine clusters (Russell, 1997).

The next layer is a pair of bent antiparallel helices (S and I), which wrap around either side of the helical sandwich (Figure 1.7). Helix S is bent smoothly, but Helix I has a kink due to a proline residue. The interaction between each pair of helices S and I and the 8-helical sandwich is mediated by hydrophobic interactions. The I helix forms the main hydrophobic core of the large domain, as it is composed of mainly hydrophobic residues.

The outermost layer is composed of the remaining large domain helices A C D E J K, which are mostly amphipathic helices with the polar side chains forming the surface of the protein, and nonpolar ones interacting with

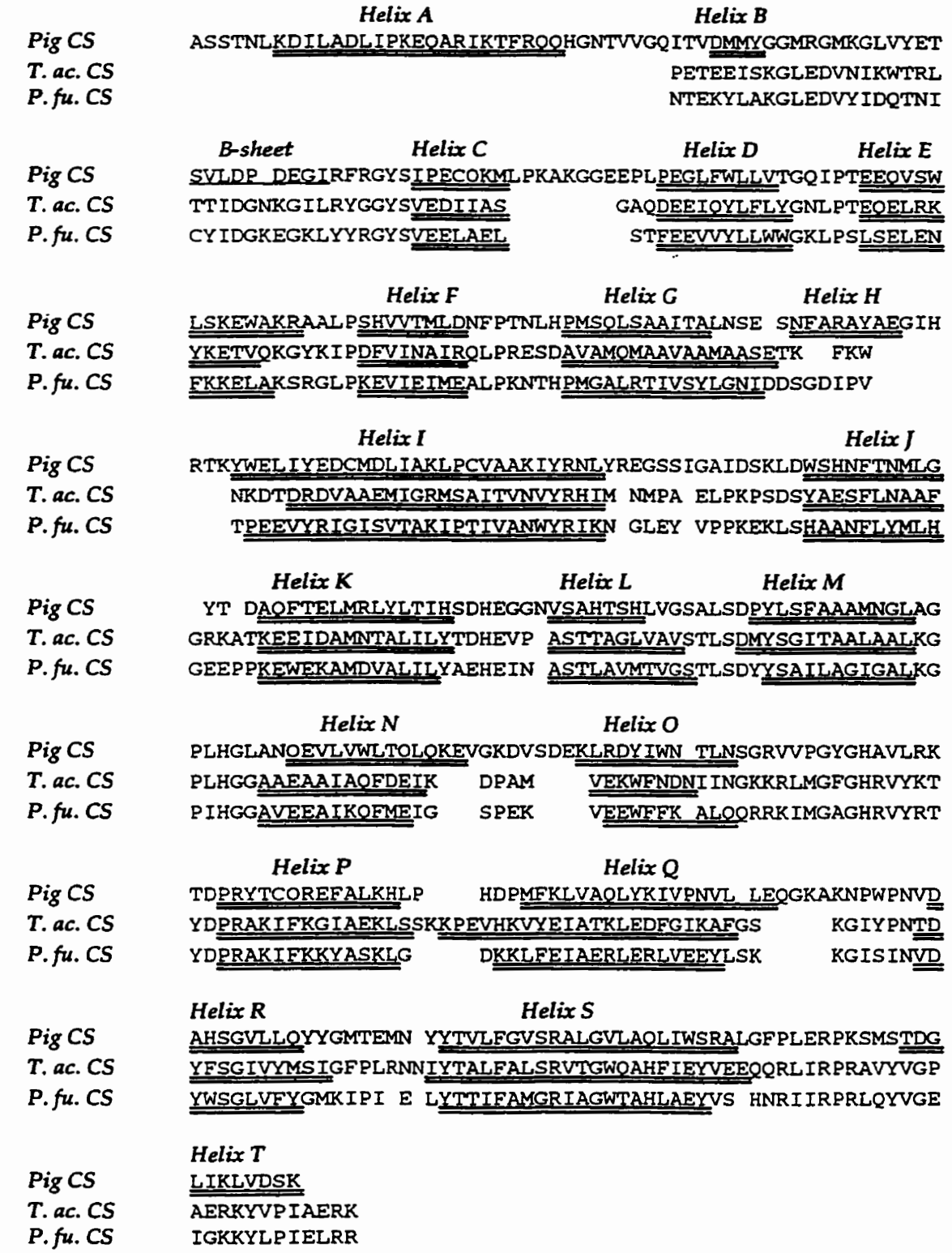


Figure 1.5 Structure-based sequence alignment for pig, *Thermoplasma acidophilum*, and *Pyrococcus furiosus* citrate synthases and their associated secondary structure (adapted from Russel *et al.*, 1997).

```
                          Helix A                                  Helix B
Pig CS          ..ASSTNLKDILADLIPKEQARIKTFRQQHGNT.VVGQITV.DMMY 42
E. coli CS      ADTKAKLTLNGDTAVELDVLKGTLGQDVIDIRTLGSKGV...FTFD 43

                                                  Helix C
Pig CS      GGMRGMKGLVYETSVLDPDEGIRF.RGYSIPECQKMLPKAKGGE..... 85
E. coli CS  PGFTSTASCESKITFIDGDEGILLHRGFPIDQLATDSN.........--- 81

                Helix D          Helix E                Helix F
Pig CS      EPLPEGLFWLLVTGQIPTEEQVSWLSKEWAKRAALPSHVVTMLDNFPTNL 135
E. coli CS  ...YLEVCYILLNGEKPTQEQYDEFKTTVTRHTMIHEQITRLFHAFRRDS 128

              Helix G              Helix H                        Helix I
Pig CS      HPMSQLSAAITALNSESNFARAYAEG.IHRTKYWELIYEDCMDLIAKLPC 184
E. coli CS  HPMAVMCGITGALAAFYHDSLDVNNPR.HREIAAFR.......LLSKMPT 170

                                  Helix J
Pig CS      VAAKIYRNLYREGSSIGAIDSKLDWSHNFTNMLGY.TDAQ.........FT 225
E. coli CS  MAAMCYKYSIGQPFVYPRNDLSYAG..NFLNMM.FSTPCEP.YEVNPILE 216

                Helix K              Helix L              Helix M
Pig CS      ELMRLYLTIHSDHEGGNVSAHTSHLVGSALSDPYLSFAAAMMGLAGPLHG 275
E. coli CS  RAMDRILILHADHEQ.NASTSTVRTAGSSGANPFACIAAGIASLWGPAHG 265

                  Helix N                  Helix O
Pig CS      LANQEVLVWLTQLQKEVGKDVSDEKLRDYIWNTLNSGRVVPGYGHAVL.R 324
E. coli CS  GANEAALKMLEEISSVKH.....IPEFVRRAKDKNDSFRLMGFGHRVY.K 309

                  Helix P                          Helix Q
Pig CS      KTDPRYTCQRETA...LKHLPHDPMFK L VAQLYKIVPNVLLEQGKAKN 369
E. coli CS  NYDPRATVMRETCHEVLKELG.TKDD LLEVAMELENIALNDPY.FIEKK 356

                  Helix R                          Helix S
Pig CS      PWPNVDAHSGVLLQYYGMT.EMNYYTVLFGVSRALGVLAQLIWSRALGFP 418
E. coli CS  LYPNVDFYSGIILKAMGIP..SSMFTVIFAMARTVGWIAHWSEMHSDGM. 403

                  Helix T
Pig CS      .LERPKSMSTDGLIKLVDSK          437
E. coli CS  KIARPRQLYTGYEKRDFKS..DIKR    426
```

Figure 1.6 Sequence alignment of pig and *E. coli* citrate synthase sequences. The secondary structure shown for pig was determined crystallographically and that for *E. coli* was generated using a secondary structure prediction algorithm, (Rost & Sander, 1993).

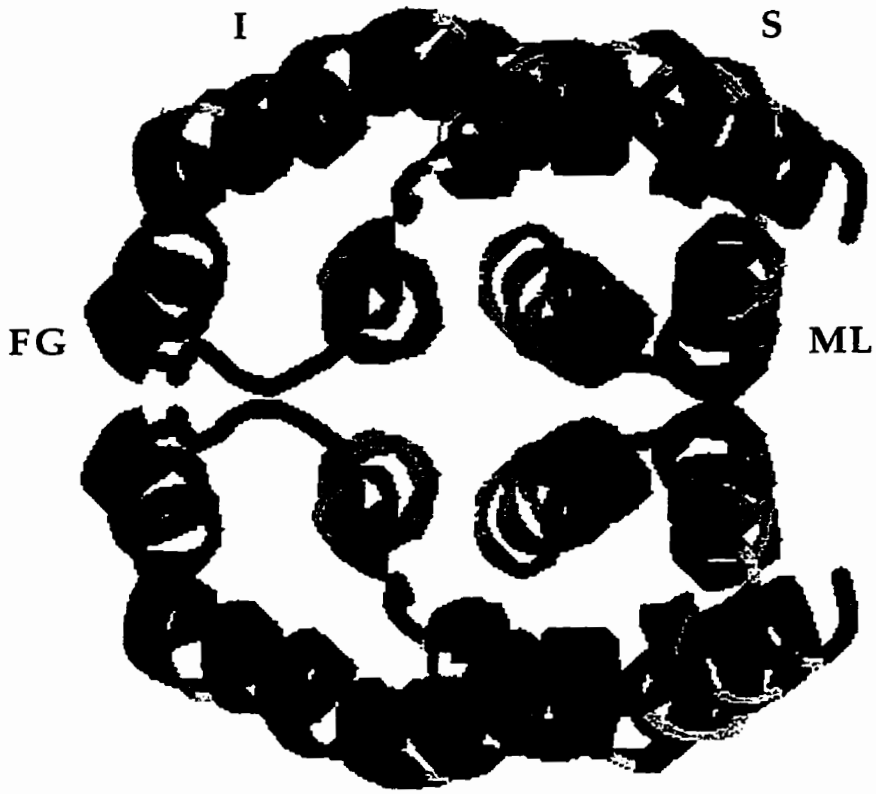


Figure 1.7 The CS dimer interface is made up of four antiparallel pairs of helices (F G M L and the two fold related counterparts). Helices S and I wrap around the dimer interface.

the helices of the dimer interface and the SI helical layer. The short B and T helices interact with the large domain of the second subunit, with their axes roughly parallel to the two-fold axis relating the two monomers. These two helices are absent in the archaebacterial CS structures, as is the rather long A helix of the pig CS. The predicted secondary structure for *E. coli* CS using the PHD algorithm (Rost & Sander, 1993) bears resemblance to the archaebacterial ones, with the A, B, and T helices being absent.

The small domain

The small domain of pig CS is composed of helices N O P Q and R (residues 275 to 380) and is illustrated in Figure 1.8. Helix R forms the core of the domain, essentially being surrounded by 2 antiparallel helix pairs: N-O, and P-Q. An algorithm designed to analyze proteins of known three-dimensional structures for hydrophobic folding units identified this part of the protein as an "autonomous folding unit"; these units are thought to contain the elements that initiate protein folding (Tsai & Nussinov, 1997). The definition of "domain" here is based on the structural observation of two, fairly independent structures, and as described later, the observation of movement of the domains relative to each other. In some proteins, structural domains identified in this manner or by other criteria (such as proteolytic cleavage, for example) were found to unfold independently (Jaenicke, 1996), although many multidomain proteins unfold in a single cooperative step.

The *P. furiosus* and *T. acidophilum* CS small domains exhibit shorter loops than the pig counterpart. *P. furiosus* especially has much shorter loops, which are thought to reduce the flexibility of the protein at the very high temperatures preferred by this organism (Russell *et al.*, 1997).

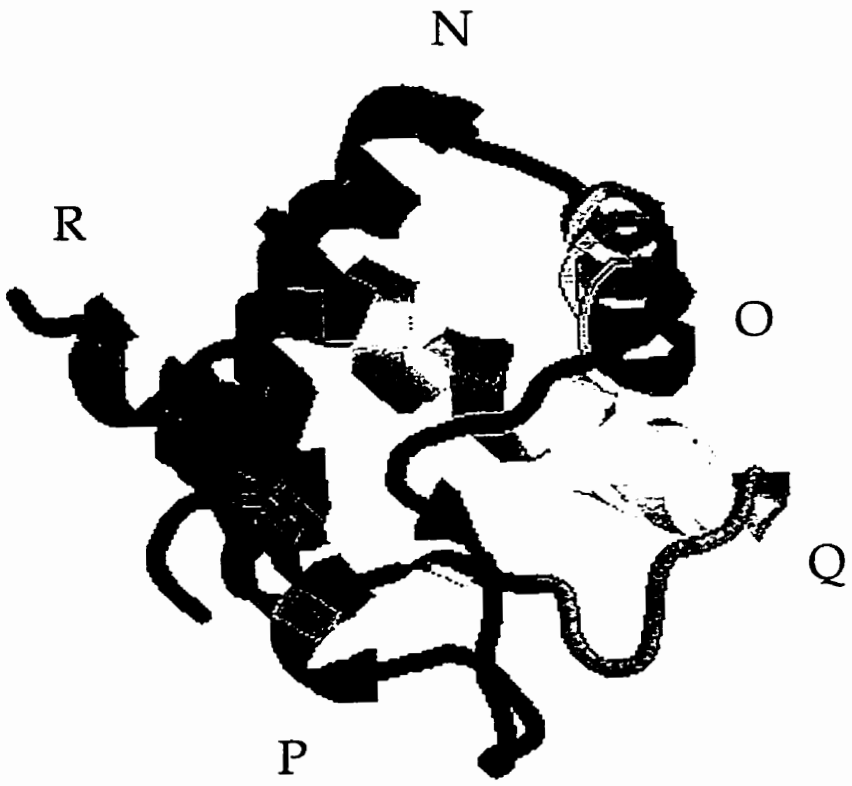


Figure 1.8 The small domain of pig CS (residues 275 to 380)

Substantial inter-domain hydrophobic contacts do exist between the two structural domains of CS, specifically between helices K-P, K-R, S-R, and S-N within each monomer. The implication of this observation is that the folding of the two structural domains may not be independent, as is sometimes observed for some multidomain proteins. On the other hand, the inter-domain interactions may form after independent folding of the two domains as an assembly step, not unlike the assembly of multisubunit proteins (Jaenicke, 1991).

It is because of this extensive interdomain interface that Lesk and Chothia reasoned that a rigid body rotation used by Remington's group (Remington *et al.*, 1982) to describe the movement of the small domain relative to the large domain is inadequate, and studied the differences between the open and closed forms more carefully (Lesk & Chothia, 1984). The open conformation occurs in the absence of bound ligand, or in the presence of citrate, and the closed conformation occurs in the presence of OAA, or citrate and CoA (see Figure 1.9). Using the two conformations they identified a core structure that remains unchanged between the two conformations. Then the shifts and rotations of the secondary structural elements in the small domain relative to the core structure, and relative to each other, were calculated. It was discovered that there are significant rotations and shifts of helices of the small domain relative to each other and relative to the core of the large domain, with the cumulative effect of a large rotation of the O-P loop towards the large domain of the other subunit; this loop contains the backbone atoms that interact with the adenine moiety of the AcCoA molecule. This conformational change was classified as a shear

Figure 1.9 Space filling representations of the pig CS in the closed and open conformations.

mechanism, distinct from a hinge mechanism for domain movement (Lesk & Chothia, 1984).

The presence of two structural domains in CS makes it an interesting subject for folding and stability studies. As described later, the study of unfolding of chimera proteins produced by domain swapping between the *E. coli* and *A. anitratum* CS enzymes initiated the project described in Chapter 2 of this thesis.

Activity of Citrate Synthase

Activity and function are intimately related to structure and stability. Some of the mutants constructed to probe stability of CS resulted in altered function, while others which were constructed to probe function exhibited altered stabilities; the relationship between function and stability is often inverse, as they represent apparently opposing features of a protein (Shoichet *et al.*, 1995). In this section, a description of the active site of CS sets the stage for understanding the relationship between stability and functionality.

Oxaloacetate and Acetyl CoA binding

The active site of CS is located in the cleft between the two structural domains (Figure 1.10) and involves residues from both subunits in the dimer. Therefore the minimal functional unit of CS is a dimer with two active sites. The first CS structure was obtained with bound citrate and CoA (Remington *et al.*, 1982), and this was used to deduce the residues involved in binding of the substrates OAA and AcCoA. Later, additional structures obtained complexed with malate, OAA, and CoA analogues carboxymethyl-CoA and S-acetonyl-CoA confirmed the deduced ligands for substrate binding and catalysis (Karpusas *et al.*, 1991, Liao *et al.*, 1991). The information gained from

Figure 1.10 Ribbon representation of CS monomer showing the active site residues in the cleft between the large and small domains. The active site residues here are coloured blue and represented by stick models, and the large and small domains are coloured gray and yellow, respectively.

those studies has been combined for the following description of the binding of the substrates, OAA and AcCoA.

OAA binds such that all of the hydrogen bond acceptors are liganded, as is shown in the schematic representation of the active site of CS in Figure 1.11 which uses the residue numbers from the *E. coli* CS sequence. OAA is bound by three arginines (R387, R314, and R407', with the last originating from the second subunit), and two histidines (H305 and H229). H305 is the residue which polarizes the carbonyl group of OAA, making it more susceptible to condensation. Site-directed mutagenesis of this residue in *E. coli* (Pereira *et al.*, 1994) and pig CS (Evans *et al.*, 1996) confirmed its importance in OAA binding, catalysis, and substrate specificity.

AcCoA (Figure 1.12) binds only after OAA induces the conformational change to the closed form, since the AcCoA binding site does not exist in the open conformation. As with the case of OAA, every potential hydrogen bond donor or acceptor is liganded. The 5' diphosphate salt-bridges with two arginines from one subunit (pig CS R324 and R46), and another from the other subunit (pig CS R164'). The closed conformation brings the O-P loop closer to the active site cleft as a result of OAA binding and concomitant shifts and rotations of helices of the small domain. This loop, which shifts by 6.1 Å and rotates by 28° relative to the core of CS (Lesk & Chothia, 1984), contains the adenine recognition loop (pig CS residues 312-320), the backbone atoms of which hydrogen bond with the adenine moiety, wrapping around the ring edgewise. In addition to hydrogen bonding, various hydrophobic interactions are involved in the binding of AcCoA, most notably V314 of the pig CS (probably equivalent to L300 in *E. coli*). At the active site, the acetyl group interacts with H264 and D362, the residues

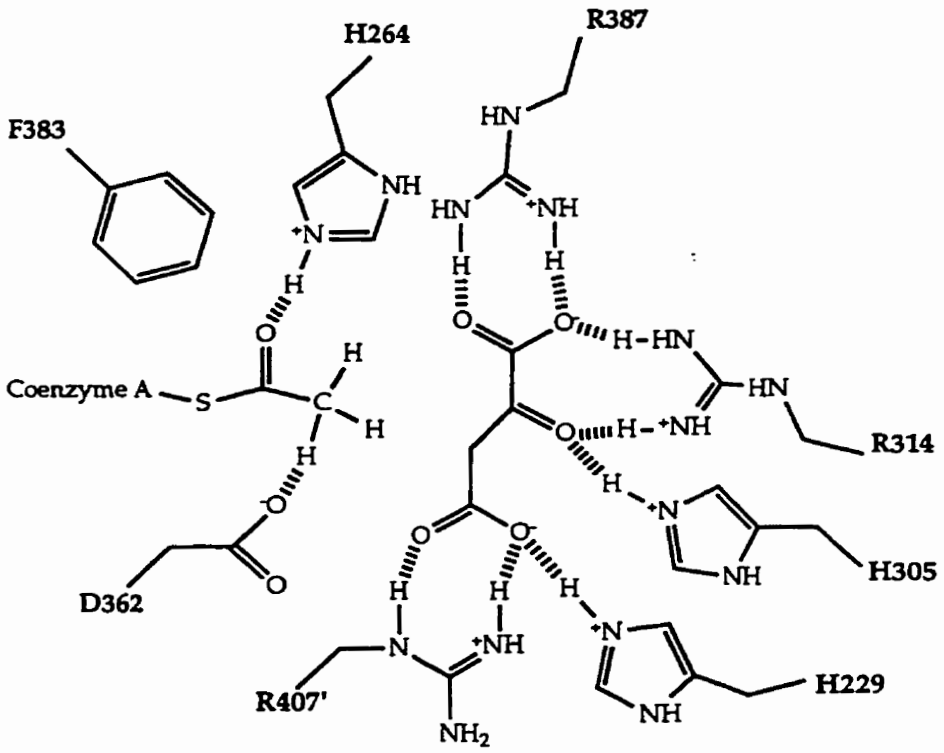


Figure 1.11 Oxaloacetate and AcCoA binding at the active site of citrate synthase. The residue numbers here correspond to the *E. coli* enzyme sequence. Histidine 305 is the residue which polarizes the OAA carbonyl prior to the condensation reaction. The acetyl of AcCoA bound at the active site. Both *E. coli* CS residues D362 and H264 are catalytically involved in the CS mechanism. F383 is present in all CS and methylcitrate synthase sequences. Figure adapted from Pereira *et al.* (1994).

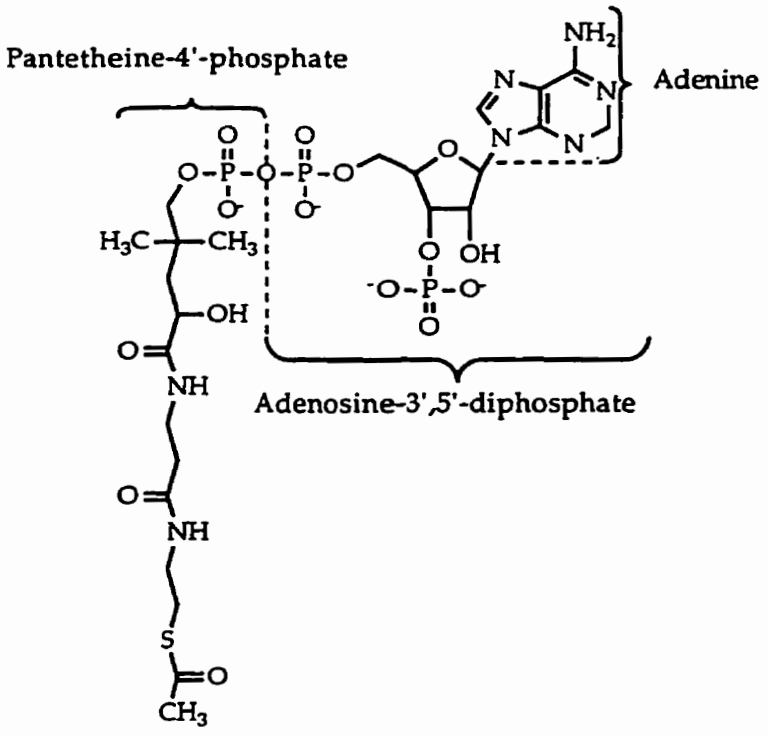


Figure 1.12 Structure of Acetyl Coenzyme A.

involved in the catalytic mechanism of CS. F383 nearby makes an edgewise interaction with the acetyl group and a mutation study showed that it plays a role in the affinity of CS for AcCoA, and in the allosteric equilibrium of *E. coli* CS (Pereira *et al.*, 1994). The F383A mutant was included in my work on the stability of CS (Chapter 2).

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 residues which are directly involved in catalysis are H305, H264, and D362 (*E. coli* CS numbering), identified by site-directed mutagenesis and three-dimensional structures. The current working mechanism (Evans *et al.*, 1996) is shown in Figure 1.13. First, OAA binds in the cleft between the large and small domain and induces a conformational change to the closed conformer. H305 polarizes the carbonyl bond of OAA, increasing the positive charge on the carbon. This creates a species more susceptible to condensation with AcCoA. Second, AcCoA binding to the CS-OAA complex takes place, and H264 and D362 act in concert in an acid-base reaction to enolize the AcCoA. This is thought to be the overall rate-limiting step in the mechanism. Third, the enol nucleophile attacks the polarized OAA at the carbonyl carbon in a condensation reaction to form citryl CoA. Finally, hydrolysis of the citryl CoA yields citrate and CoA, whereupon a conformational change to the product-release, open conformation takes place.

Many studies, mostly site-directed mutagenesis in *E. coli* and pig CS of residues identified by sequence alignments and by examination of the three-dimensional structures, are consistent with the above current view of the

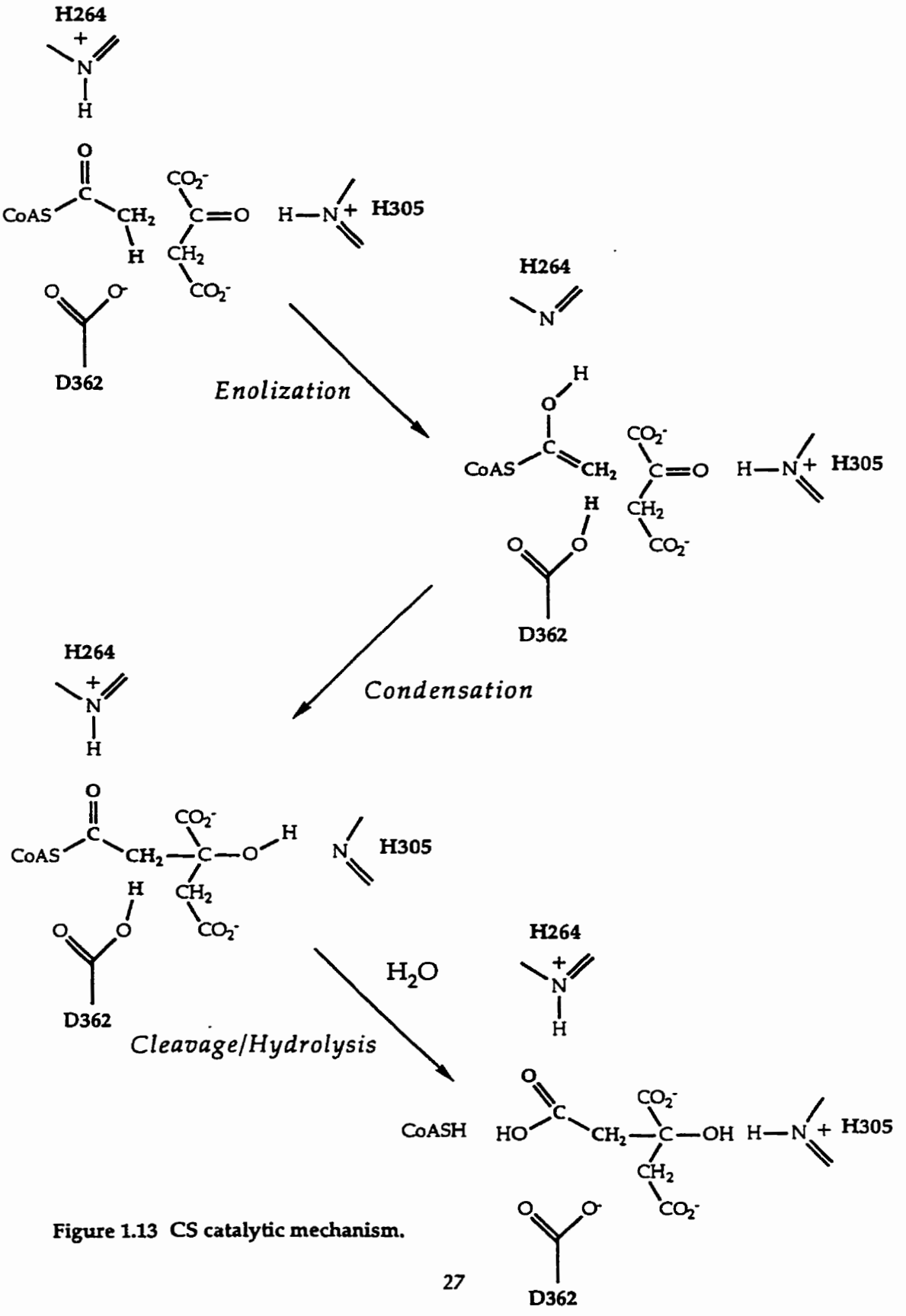


Figure 1.13 CS catalytic mechanism.

mechanism. Additional information about the roles of residues involved in binding and recognition of substrates is also available and demonstrates the remarkable specificity of this enzyme for its substrates.

Escherichia coli Citrate Synthase

E. coli CS belongs to the allosteric subgroup of these enzymes which has two distinguishing features from the non-allosteric class. First, its dimers associate to form higher order aggregates, with hexamer being predominant. Second, it is subject to strong allosteric inhibition by NADH (Figure 1.14), a product of the TCA cycle. The presence of these two features in this class and their absence in the non-allosteric class has led to the speculation that the oligomeric state of CS is important in the allosteric regulation. However, a clear and definitive relationship between quaternary structure and NADH inhibition has not been established for this or any of the other Gram negative CSs.

Subunit structure

Tong and Duckworth (1975) studied the subunit structure of *E. coli* CS in detail using analytical ultracentrifugation under various conditions of pH and KCl concentrations. They discovered that CS is capable of various association-dissociation equilibria under most conditions of pH, except at pH 9 where only dimer was found. For example, at pH 7, species ranging in mass from monomer to decamer were observed. The addition of KCl resulted in a more defined population of dimers and hexamers which could be shifted completely to hexamer with higher salt concentrations (0.1 M). Similar evidence was obtained for the predominance of dimers and hexamers by experiments using cross-linking with dimethyl suberimidate followed by SDS PAGE, although higher even-numbered and

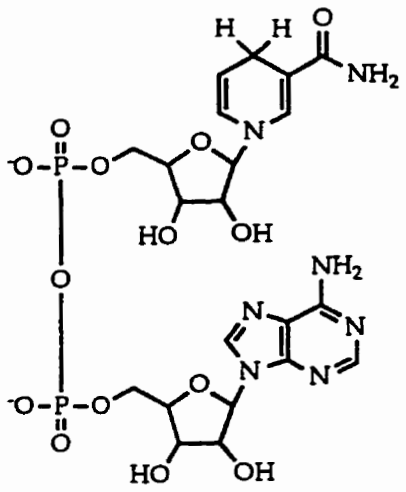


Figure 1.14 The structure of the reduced form of nicotinamide adenine dinucleotide (NADH).

likely nonspecific aggregates were also observed. A study of CS subunit structure was not undertaken in the presence of NADH using the ultracentrifuge.

Molgat *et al.* (1992) applied non-denaturing gel electrophoresis to the study of CS subunit structure. At pH 8.1 three bands were observed, two that were hexamers of differing charge (probably conformational isomers) and one corresponding to dimer. Three electrophoretically distinguishable bands had also been observed previously at pH 8.9 by Danson and Weitzman (1973), and they recognized that these correspond to three forms of CS in equilibrium since re-electrophoresis of any one of the bands resulted in the same pattern. The molecular weight measurements by Molgat *et al.* (1992) strengthen the case for an equilibrium for *E. coli* CS where dimers associate to form hexamers. Again, no corresponding study was undertaken in the presence of NADH. The presence of an equilibrium between two oligomeric forms under near-physiological conditions is usually an indication that it may be physiologically significant for regulation, and that the components may represent active and inactive forms (Traut, 1994).

NADH inhibition and binding

Weitzman (1966) first showed that *E. coli* CS is inhibited by NADH, an end product of the TCA cycle, and that this inhibition is reversible. The extent of inhibition is pH dependent, with a weaker inhibitory effect at higher pH; at pH 9.2, the enzyme is no longer subject to inhibition by NADH. The finding that only dimer exists at pH 9 by the ultracentrifugation experiments (Tong & Duckworth, 1975) indicated that hexamers may be needed for NADH to exert its inhibition. The sensitivity to inhibition was also abolished by the addition of KCl which acts as an allosteric activator.

A more thorough study of the NADH binding properties of CS was undertaken by Duckworth and Tong (1976) where fluorescence enhancement and gel filtration techniques were applied. Confirming the previous observation that NADH is a less efficient inhibitor at higher pH, they found that both affinity and the number of binding sites per subunit drop as higher pH values are reached. At pH 6, a maximum of 0.65 sites per subunit was found with a K_D approaching 0.28 μM, whereas at pH values closer to 9, 0.25 binding sites per subunit were found. Gel filtration methods confirmed the maximal occupancy of ~0.6 sites per subunit at pH 7. The pH dependence of the number of available sites was suggested to be a reflection of the quaternary structural changes that take place in the pH 7 to 9 range, perhaps with different oligomeric forms having different affinities for the inhibitor. It would be difficult, if not impossible, to use conventional techniques (such as fluorescence or ultracentrifugation) to determine binding affinities for NADH to the different oligomeric forms under similar or identical conditions, since those techniques give average properties. I have been able to use mass spectrometry, however, to determine such values for NADH binding to the dimeric and hexameric CS oligomers, as described in Chapter 3.

Enzymology

Citrate synthase from *E. coli* exhibits a sigmoid AcCoA saturation curve (Faloona & Srere, 1969), but the activator KCl yields Michaelis menten kinetic behaviour instead, with a 40-fold activation of the enzyme when the non-saturating AcCoA concentration of 100 μM is used. KCl also abolishes NADH inhibition; this suggests that the *E. coli* CS allosteric R state binds AcCoA and KCl well and is active, and that the allosteric T state binds NADH and is inactive (the nomenclature of Monod, *et al.* (1965) is used here for the

allosteric states). OAA binds equally well to T and R states and its binding to CS is hyperbolic under all conditions tested (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 more 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.

An apparent contradiction in the above description of the *E. coli* CS allostery and subunit structure is the fact that KCl is an allosteric activator but also induces hexamerization of CS. The effect of NADH is maximal at a pH where more hexamer is present. It is possible that two functionally and perhaps structurally distinct hexameric forms of CS exist. However, this point remains unresolved.

The allosteric binding site and dimer association to form hexamers

To date, neither the NADH binding site nor the hexamerization elements have been identified for the *E. coli* or any of the other allosteric CSs; the lack of a three-dimensional structure of CS of the allosteric class precludes the identification of these two important features. However, there is some evidence that the NADH binding site is near cysteine 206 of the *E. coli* enzyme, since the modification of this residue by alkylating agents such as 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB, also known as Ellman's reagent) (Talgoy *et al.*, 1979), or 1,1,1-trifluoro-1-bromoacetone (TFBA, used for ^{19}F-NMR work) (Donald *et al.*, 1991) abolishes NADH sensitivity and binding, but not CS activity. This residue is in a putative loop between Helices J and K in the large domain, which exhibits low level of identity between the allosteric and non-allosteric classes of CS (see alignment in Figure 1.6 and

Appendix I). This location is about 30Å away from the active site in the pig CS structure (Donald *et al.*, 1991).

A. anitratum and chimera citrate synthases

Acinetobacter anitratum is a strictly aerobic, Gram negative bacterium. *A. anitratum* CS has been purified and characterized (Morse & Duckworth, 1980), and the *gltA* gene cloned (Donald & Duckworth, 1987). The allosteric properties of this CS differ from the *E. coli* CS in several ways. First, large quantities of NADH are required to inhibit this CS and this inhibition is easily reversed by the addition of ATP. Second, the substrate saturation kinetics differ in that hyperbolic saturation by AcCoA is observed in the absence of KCl, instead of the sigmoid saturation observed in *E. coli*. Third, the extents to which hexamer is formed are different: *E. coli* CS shows a mixture of dimers and hexamers, while *A. anitratum* CS is hexameric, as shown by non-denaturing gel electrophoresis (Molgat *et al.*, 1992). The above features of the *A. anitratum* CS have been interpreted as being a consequence of its allosteric equilibrium, which is thought to be shifted towards the R or active state compared to that of the *E. coli* enzyme.

An additional difference between the *A. anitratum* and *E. coli* CSs is their stabilities against urea denaturation and proteolysis. Morse and Duckworth (1980) showed that *E. coli* CS is inactivated by urea at low concentrations after short incubation times, whereas *A. anitratum* CS is resistant to inactivation by urea at concentrations up to 3 M and at long incubation times. It is not obvious from the sequences of these two enzymes why *A. anitratum* CS is so much more stable than *E. coli* CS, nor why there are marked differences in enzyme kinetics and allosteric behavior.

Molgat *et al.* (1992) studied the *A. anitratum* and *E. coli* CS in order to understand the relationship between their allosteric properties and their domain and subunit structures. Using the homologies of these two enzymes to the known pig CS structure, chimera proteins were produced which contained the large domain from one CS and small domain from the other. The two resultant functional chimeras were subject to allosteric inhibition by NADH. The main conclusion of this work was that the observed differences in stabilities, allosteric properties and subunit interactions that lead to hexamerization are conferred by the large domains of these enzymes.

Thesis objectives

The objective of this thesis was to study the unfolding of *E. coli* and *A. anitratum* CS, and their chimeras. Since a three state urea unfolding profile was obtained for the *E. coli* enzyme, site-directed mutagenesis was used in order to clarify the origin of this behavior. Other approaches were used to characterize the intermediate state in the *E. coli* CS unfolding profile.

An unanswered question, as already mentioned, was how the quaternary structure of CS and the NADH inhibition are related. The objective of the second investigation (Chapter 3) was to use mass spectrometry to study the subunit structure of CS and the binding of NADH and substrates to the enzyme.

CHAPTER TWO Unfolding of Citrate Synthase

Introduction

The protein folding problem

Protein folding is described as a problem because of our inability to deduce a three-dimensional structure of a protein given its genetically prescribed primary sequence. Christian Anfinsen's Nobel-prize winning work with ribonuclease (RNase) showed that the sequence of a protein contains all of the information necessary for a folded, functional conformation (Anfinsen, 1973). Furthermore, Cyrus Levinthal reasoned that the folding process cannot be achieved by random searching of all possible conformations (Levinthal, 1968). The process of folding is often dubbed as the "second half of the genetic code" (Kolata, 1986) in the literature, since it completes the flow of information from the DNA level to function, as shown schematically in Figure 2.1.

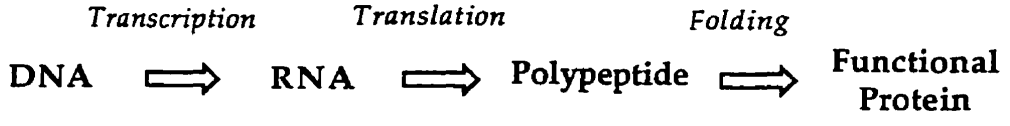


Figure 2.1 The central dogma: protein fold is genetically prescribed by the DNA sequence.

The solution to the problem of protein folding remains elusive despite concentrated efforts on many proteins for decades and despite the availability of almost 6000 three-dimensional protein structures in the protein data bank (PDB) growing at the rate of one every five hours! The number of gene sequences that code for proteins of unknown function is growing exponentially, including entire genome sequences for more than a dozen

organisms, making the solution to the protein folding problem all the more urgent. The three-dimensional structures that we do know have provided information about the interactions that maintain a folded conformation: hydrogen bonds that stabilize secondary and tertiary structure, hydrophobic interactions between sidechains in the interior of proteins, salt bridges between polar and charged residues and backbone atoms, and van der Waals contacts between all types of atoms in the protein. We have also learned that protein structure is hierarchical: primary (sequence), secondary (local arrangement of backbone), tertiary (non-local interactions) and quaternary levels (domain and assembly). The observation that some proteins which are different in sequence may adopt very similar folds must necessarily imply that a limited number of protein folds are possible; it has been suggested that perhaps 1000 different folds exist (Chothia, 1992), and that a code could be deciphered from the knowledge of these types of folded structures.

The protein folding problem can be divided into several related issues: kinetics (the events and pathways during the process of folding), stability (the forces that lead to a stable folded conformation), and prediction (the use of statistical knowledge to predict the fold of a protein). The main objective of this project was to study the stability of *E. coli* citrate synthase, as it relates to structure and function, and to identify the reason why its unfolding transition is biphasic.

The stability of proteins

The folded states of proteins are marginally more stable than the unfolded states, with conformational stabilities (ΔG) of about -5 to -10 kcal.mol^{-1}; these values are on the order of a few hydrogen bonds (Creighton, 1993). This stability is the net effect of a large number of weak stabilizing and

destabilizing factors. On the one hand, folded proteins are destabilized by a large contribution from conformational entropy of the unfolded state. On the other, folded conformations bury hydrophobic side chains which, if exposed, would destabilize the unfolded state because they increase the ordering of water (the hydrophobic interaction). While hydrogen bonding contributes significantly to the secondary structural arrangements in folded conformations, the total number of such bonds might be expected to be similar in the folded and unfolded states, the only difference being the hydrogen bonding partners. It has been long held that the hydrophobic interaction is the major force which drives folding and therefore is the major determinant of stability of folded conformations (Dill, 1990). Other ideas are being considered, however, especially the notion that both the hydrophobic interaction and hydrogen bonding make large but comparable contributions to the stability (Pace, 1995, Pace *et al.*, 1996).

The latter hypothesis, that hydrogen bonds contribute as much as hydrophobic interactions to the stability of a protein, is based on the observation that the burial of the peptide group within the interior of a protein is very unfavorable but that a compensating effect is achieved if the group can form an intramolecular bond, an interaction which is thought to be more stable than an intermolecular backbone hydrogen bond (Pace *et al.*, 1996). This point has been difficult to test directly, however, since it has not been possible to measure the contribution to stability of buried backbone atoms involved in hydrogen bonds. Hydrogen bonds involving particular sidechains, however, may be removed by site-directed mutagenesis, and some estimates to stability have been made of the contributions of sidechain hydrogen bonds. Interest in evaluating the contribution of intramolecular

hydrogen bonding to stability is mounting in the literature, and recently, a group was successful in devising a methodology to test this. In this approach, backbone amide linkages in an α-helix of T4 lysozyme were replaced with ester linkages. Removal of two hydrogen-bonding interactions in this manner destabilized the protein by 1.7 kcal.mol^{-1} (Koh *et al.*, 1997). On the other hand, NMR-determined amide coupling constants, which are a probe for protein secondary structure, indicated that an intraprotein NH-OC hydrogen bond in ubiquitin is weaker than the hydrogen bonds between these groups and bulk water (Juranic *et al.*, 1995).

The hydrophobic effect has received, by far, the most attention as the potential major force in determining stabilities of folded proteins. Since the first study using site-directed mutagenesis to probe the contribution of hydrophobic residues to stability of tryptophan synthase (Yutani *et al.*, 1984), an explosion of data on other proteins indicated that the effects are varied, with changes in stabilities of 0 to 2.5 kcal.mol^{-1} for each -CH_2- group (Pace, 1992). These observed changes have not always been in agreement with transfer free energy values obtained from systems which model the transfer of hydrophobic side chains from aqueous environments to the non-polar interior of a protein.

The hydrophobic interaction

One of the more striking features of folded proteins, discovered with the solution of the first protein structure, that of myoglobin, is the burying of hydrophobic residues in the interior, and the remarkable rarity of cavities. In fact, about 80 to 85% of non-polar side chains are buried in the interior of globular proteins, out of contact with water (Lesser & Rose, 1990). It is because of this observation that the hydrophobic interaction has been hailed as the

major force for protein folding and stability. The importance of hydrophobic interactions in proteins was predicted by Bernal in 1939 (1939), long before Kauzman (1959) and Tanford (1962) popularized the idea among protein chemists by describing of the hydrophobic interaction in terms of water structure.

The interaction of hydrophobic groups with water is unfavorable; water is a poor solvent for non-polar molecules, and if present in aqueous solutions, hydrophobic groups interact preferentially with one another. In other words, the hydrophobic interaction is a consequence of the strong interaction of the water molecules with each other.

In proteins, the non-polar amino acids are A, C, F, I, L, M, P, V, and W. The magnitude of the contribution of the hydrophobic interaction to the stability of folded protein conformations has been estimated using models of transfer between aqueous solution and non-polar solvents. In such an approach, a non-polar solvent is the model for the protein's hydrophobic interior, although there are great differences in the properties between the two. In particular, it would be difficult to assign energy terms for the difference in entropy (sidechains are fixed in a protein's interior) and the difference in dispersion forces (they are greater in the protein's interior) (Pace, 1992).

Nevertheless, several hydrophobicity scales have been calculated using different models for amino acids and the interior of proteins. The scales vary somewhat, but the most widely accepted one currently is that for the transfer of N-acetyl amides from water to n-octanol (Fauchére & Pliska, 1983) corrected for differences in volume between the solvents and the solutes (Sharp *et al.*, 1991). The estimate of the strength of the hydrophobic interaction according

to this scale is 28 cal.mol^{-1}.Å^{-2}, in agreement with the generally accepted value of 25 cal.mol^{-1}.Å^{-2} derived from solubility and partition experiments (Creighton, 1993).

Hydrophobic residue replacements

Another, more direct method of measuring the magnitude of the hydrophobic interaction in proteins is by site-directed mutagenesis. Yutani *et al.* (1984) were the first to apply site-directed mutagenesis to the study of the role of hydrophobic residues in the stability of tryptophan synthase. Substituting E49, a buried residue, with other more hydrophobic amino acids, they found that the conformational stability of the mutants increased linearly with the hydrophobicity of the substituting residue, consistent with the importance of the hydrophobic interaction in protein stability. However, replacements of buried hydrophobic residues with alanine in several proteins showed variable results, with destabilizations of 1.6 to 5.8 kcal.mol^{-1} for different L to A substitutions in *Staphylococcus* nuclease (Shortle *et al.*, 1990), and 2.7 to 5 kcal.mol^{-1} for T4 lysozyme (Eriksson *et al.*, 1992). These values implied that the strength of the hydrophobic interaction in proteins may be higher than that deduced from model systems. Furthermore, the magnitudes of the destabilizations were highly variable from one site to another.

The explanation for the above two points was revealed upon examination of T4 lysozyme L to A mutants by X ray crystallography; it was noted that a cavity always remained after the substitution, but that the size was variable, indicative of different responses of the protein to the replacement at the different sites. A linear relationship was found between the size of the cavity created and the loss of conformational stability, with the following empirical relationship:

$$\Delta\Delta G = a + b.\Delta V \qquad [1]$$

where $\Delta\Delta G$ is the observed change in stability of the mutant protein from the WT protein, a (the y-intercept) is a constant free energy term which is dependent only on the identities of the amino acids involved (see below), b is the free energy contribution per $Å^3$ of volume, and ΔV is the change in volume observed from the X ray crystal structures. For the series of L to A mutations of T4 lysozyme, a was found to be -1.9 $kcal.mol^{-1}$, and b was -0.024 $kcal.mol^{-1}.Å^{-3}$. The value -1.9 $kcal.mol^{-1}$ agrees remarkably well with the transfer free energy difference between L and A calculated from the n-octanol transfer free energy values ($\Delta\Delta G_{tr}$=-1.9) (Fauchére & Pliska, 1983). The variable term $b.\Delta V$ depends on the response of the protein to the mutation, with the value for b being consistent with previous estimates for the strength of the hydrophobic interaction when expressed in terms of surface area, 20 $cal.mol^{-1}.Å^{-2}$ (Eriksson *et al.*, 1992, Matthews, 1996).

Similar responses to hydrophobic residue replacements were found in other systems, most notably with Barnase, another small monomeric protein (Buckle *et al.*, 1996). Again, a linear relationship was found between stability change and the volume of the created cavity. Surprisingly, however, a buried water molecule was found in the smallest of those cavities (I76A), forming a single hydrogen bond with the carbonyl of the backbone. Inspection of other cavities in this and other proteins, including the T4 mutants, revealed that some are lined with a proportion of polar atoms, capable of forming at least one hydrogen bond with a water molecule, which may not be necessarily observable crystallographically. It was suggested that the presence of this cavity-bound water molecule in the Barnase I76A mutant must be stabilizing, since this mutant was the least destabilized of the barnase mutants studied;

the cost of removing this water molecule from bulk solvent may be offset by the hydrogen bond and favourable van der Waals interactions (Buckle *et al.*, 1996).

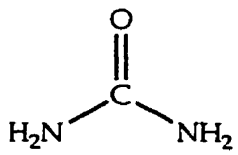
Thus, in general, the $\Delta\Delta G$ values obtained for hydrophobic replacements depend on the context in which they are made: whether the residues involved are buried or exposed, and the rigidity or flexibility of the protein structure in its vicinity. It is interesting to note that mutations of exposed hydrophobic residues on the surface of the protein do not affect the stability significantly, leading Matthews to speculate that only a small number of residues in a given protein are sufficient for stability and folding (Matthews, 1996, Shoichet *et al.*, 1995) .

Assessment of protein conformational stability

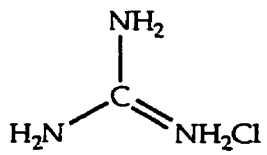
The conformational stability (that is, the free energy difference between the folded and unfolded states) of a protein has been classically determined by equilibrium denaturation curves, in which the native structure is gradually perturbed and changes in conformation are monitored. Common perturbants include chaotropic agents (urea, guanidine hydrochloride), high or low temperature, pH changes, and even pressure. Global or local properties of a protein may be monitored as a function of perturbants at equilibrium; optical techniques such as fluorescence or circular dichroism are most widely used, but NMR and mass spectrometry have been also applied. Other classical, but not often utilized methods include measurements of hydrodynamic properties by gel filtration, ultracentrifugation, gel electrophoresis, or susceptibility to proteolysis. Several different techniques are often used in the same study to follow unfolding of a protein in order to obtain different, but

complementary information about the details of unfolding (and folding, if the denaturation is fully reversible).

The best known and most commonly used denaturants are urea and guanidine hydrochloride:



Urea



Guanidine hydrochloride

These denaturants are thought to hydrogen bond with water at high concentrations, and in the case of urea, leading to the disruption of the hydrogen bonding pattern of water and reducing the strength of the hydrophobic interaction. However it is believed that urea and guanidine hydrochloride act by increasing the solubility of proteins, interacting preferentially with the surface. Urea and guanidine hydrochloride increase the solubilities of both polar and nonpolar molecules, interacting with their surfaces more favorably than water does. It is attractive to think of urea and guanidine hydrochloride as ligands which bind more favourably to the unfolded state, thus increasing its stability.

If chaotrope induced unfolding of a protein is two-state, with a fully folded Native (*N*) and fully Unfolded (*U*) states, then the conformational stability may be calculated using the linear extrapolation method (Pace *et al.*, 1988, Pace, 1986). For the following reaction,

$$N \underset{K_{fold}}{\overset{K_{unfold}}{\rightleftharpoons}} U$$

the free energy of N (ΔG_{fold}) can be defined as

$$\Delta G_{fold} = -RT \ln K_{fold} = -RT \ln \frac{[N]}{[U]} \qquad [2]$$

where R is the gas constant and T is the temperature in degrees kelvin. Using the denaturation data, the ratio $[N]/[U]$ may be calculated, and in the transition region, the apparent free energy difference relative to the unfolded state, ΔG_{app} varies linearly with the concentration of denaturant, and extrapolation to conditions in the absence of denaturant gives an estimate of ΔG_{fold}, or the stability of the N state over the U state

$$\Delta G_{app} = \Delta G_{fold} + m[denaturant] \qquad [3]$$

where m, the slope of the transition, is a measure of the dependence of the stability on the denaturant concentration. This value of m is thought to be proportional to the difference between the numbers of denaturant binding sites in the native and unfolded states, and thus appears to be related to the the difference in exposed surface area between the two states.

The use of chaotropic agents and the linear extrapolation model to estimate conformational stability have been the subject of debate, particularly whether the extrapolation to 0 M concentration is valid for such estimates. Evidence presented recently using three different chaotropic agents to denature α-chymotrypsin indicated that ΔG_{fold} values obtained by linear extrapolation are independent of denaturant (Santoro & Bolen, 1988). Other evidence obtained using a thermodynamic cycle involving urea and pH

titration to denature RNase also strongly suggests that the ΔG value obtained is the property of the protein alone (Yao & Bolen, 1995).

Equilibrium intermediates of protein folding

Kinetic intermediates that might be encountered during the folding process are difficult to study due to their transient nature and their conformational heterogeneity. Although elegant techniques have been devised for the detection and characterization of kinetic intermediates (methods such as pulse labelling for NMR experiments, for instance), there is considerable interest in intermediates which are stable and significantly populated at equilibrium. Such equilibrium intermediates, found for some proteins under mildly denaturing conditions, are thought to resemble those which may be transiently stable during the course of folding, and thus may provide clues about folding mechanisms (Ptitsyn, 1995).

There is increasing evidence that equilibrium folding and unfolding transitions of most proteins are not adequately described by two-state models, where only the native or unfolded states are stable. These findings have changed the classical view that most proteins unfold in a two-state, cooperative manner, and that the presence of stable intermediates is the exception rather than the norm. Many proteins studied exhibit equilibrium intermediates called molten globules, structures which retain significant secondary structure but exhibit transient or fluctuating tertiary interactions (Kuwajima, 1989). That a significant number of proteins unfold to a molten globule state may indicate that hierarchical assembly of structure is a general mechanism for protein folding.

If a protein has two or more structural domains, as is likely if it is large, they may undergo separate unfolding if they are structurally independent;

the unfolding reaction may be characterized as a multistep transition (called "transitions with separable stages" by Tanford (Tanford, 1968)). Independent unfolding of the domains may be shown by separating them (by cleavage for example) and demonstrating their autonomous unfolding transitions. For example, urea-induced equilibrium unfolding of γ-II-crystallin from calf eye lens, a two-domain protein, has been shown to involve a three state transition, with an intermediate in which the N-terminal domain is still folded and the C-terminal domain unfolded (Jaenicke, 1991, Jaenicke, 1996). Many other proteins which exhibit multistep equilibrium denaturation transitions have been reported in the literature (Conejero-Lara *et al.*, 1996, Gokhale *et al.*, 1996, Guha & Bhattacharyya, 1995, Hayashi-Iwasaki *et al.*, 1996, Sánchez del Pino & Fersht, 1997, Sherman *et al.*, 1995), although the general preference for more detailed work appears to be small, single domain proteins (100 - 150 residues) with known three-dimensional structures, and which display a two-state profile.

There is accumulating evidence that the kinetic molten globule states of proteins are similar to their equilibrium molten globule states (Ptitsyn, 1995). The molten globule state also appears to be a general intermediate in protein folding, to judge from the large number of proteins for which kinetic and/or equilibrium molten globules have been identified (Ptitsyn, 1992). The best studied example of a molten globule undertaken is with α-lactalbumin (Kuwajima, 1996), a small protein which contains two subdomains: an α-helical domain and a β-sheet domain. Its molten globule is observed as an equilibrium unfolding intermediate at low concentrations of urea or guanidine hydrochloride, at low pH, or upon the removal of bound calcium at neutral pH (Kuwajima, 1989, Kuwajima, 1996). NMR studies revealed that

the molten globule of α-lactalbumin has a heterogeneous structure where the α-helical domain is highly structured and contains loose hydrophobic interactions involving three aromatic residues and an unfolded β-sheet domain. The kinetic molten globule observed during refolding of this protein is identical with the equilibrium molten globule state, indicating that molten globules which are stable at equilibrium may provide important clues about kinetic intermediates and the mechanism of protein folding (Kuwajima, 1996).

The general features of molten globules are as follows (Christensen & Pain, 1994, Kuwajima, 1996). First, they retain native-like secondary structure in all or parts of the protein molecule. Second, they are relatively compact, only 10 to 20% more expanded than the native molecule. Third, the tertiary packing is transient or non-specific. These features imply that molten globule intermediates exhibit exposed hydrophobic surfaces that may result in aggregation. In fact, the binding and enhanced fluorescence of hydrophobic dyes such as 8-anilinonaphthalene sulfonic acid (ANS) has become a diagnostic technique for the presence of molten globule intermediates (Semisotnov, 1991). Another implication of exposed hydrophobic surfaces is the possibility that chaperonins target molten globules as their substrates, binding them through mainly hydrophobic interactions (Christensen & Pain, 1994, Kuwajima, 1996). The three-dimensional structure of the chaperonin, groEL, contains two heptameric cylinders with hydrophobic interiors; partially unfolded proteins are thought to interact with the interiors by hydrophobic interactions (Braig *et al.*, 1994).

A general diagnostic of a stable molten globule at equilibrium is the observation of non-coincident unfolding transitions obtained by different

techniques, a criterion which was identified earlier as an indicator of the presence of a stable intermediate (Tanford, 1968). For example, non-cooperative breakdown of secondary and tertiary structure in an unfolding transition (as probed by circular dichroism and intrinsic fluorescence, respectively), is evidence that such an intermediate may be present. Given the features of a molten globule, then, one would expect tertiary interactions to be lost at lower denaturant concentrations than secondary interactions.

Folding studies on citrate synthases

Pig CS folding and stability has been studied in some detail. The two-state guanidine hydrochloride-induced transition has a midpoint of 2.5 M when monitored by CD or fluorescence (West *et al.*, 1990). The authors claim that the tertiary and secondary interactions appear to break down cooperatively. However, the fluorescence-monitored transition appeared to occur at slightly lower denaturant concentrations than that monitored by CD. The denaturation is not reversible upon dilution (West *et al.*, 1990), but slow dialysis (Kelly & Price, 1992), or the addition of chaperonins (Buchner *et al.*, 1991, Zhi *et al.*, 1992) appear to improve the recovery of the enzyme. CS in fact is a model protein used for studies of chaperonins and heat shock proteins, as it undergoes thermal aggregation (Ehrnsperger *et al.*, 1997). The stabilities of some pig CS active site mutants have been assessed by thermal denaturation, and an inverse relationship was found between specific activity and stability; mutants exhibiting the largest increases in thermal stability were those which had the lowest activity (Zhi *et al.*, 1991), consistent with the observation of an inverse relationship between stability and function (Shoichet *et al.*, 1995).

Objectives

CS from *E. coli* is a large multisubunit, multidomain protein, an enzyme with features which makes it an interesting subject for protein folding and stability studies. The enzyme from *A. anitratum,* while almost 70% identical in sequence to the *E. coli* protein, is far more resistant to urea denaturation (Morse & Duckworth, 1980). CS chimeras constructed by swapping the two structural domains identified on the basis of the homologous three-dimensional structure of the pig enzyme were shown to exhibit regulatory properties dictated by the source of the large domain (Molgat *et al.*, 1992). The study of the stability of the chimeras and the parent proteins formed the original experiments of this project, and the discovery of a biphasic denaturant-induced unfolding transition for *E. coli* CS shaped the rest of my research towards understanding the reason for the biphasic behaviour and characterizing the partially unfolded intermediate state of *E. coli* CS.

Experimental

Site-directed mutagenesis

Site-directed mutagenesis by the modified Zoller and Smith method (Kunkel *et al.*, 1987, Zoller & Smith, 1983) was used. The M13mp18 clone of the 2168-bp EcoRI-SalI fragment of pES*gltA* containing the citrate synthase coding region was used, and was described elsewhere (Anderson & Duckworth, 1988). Synthetic oligonucleotides of 17 or 25 bases complementary to the coding strand (see Appendix II for DNA sequence of *gltA* gene) were used to introduce the mutations, and are listed in Table 2.1. Mutants were identified and verified by DNA sequencing of the entire coding region for CS using the single-stranded M13mp18 clone. The *Eco*RI-*Sal*I fragment from a restriction digest of the replicative form of the M13mp18 clone was then subcloned into pBR322 to yield the expression plasmid pES*gltA*, which was sequenced across the mutation to verify it. All DNA preparation, manipulation and sequencing methods are standard laboratory procedures (Maniatis *et al.*, 1982).

Choice of Mutants

Several criteria were used in choosing the mutants of CS to be constructed. The object of the study was to create internal cavities so as to destabilize the folded protein in selected regions. Large, hydrophobic residues

TABLE 2.1 Mutagenic oligonucleotides used to construct mutant citrate synthases

Mutant	Oligonucleotide (5' to 3')
F121A	AAGCATGGCCAGACGG
M131A	TGACTGCCGCTGGATGC
L141A	ACGCCGCCGCCGCGCCG
L164A	TCGACAGCGCGCGGAAC
L259A	GTCCCCACGCTGAAGCA
L275A	CTGATTTCTTCCGCCATTTTCAGCG
L369A	CCCATCGCTTTCGCGATGATACCAG
W391A	CAGTGGGCGATCGCGCCAACGGTAC
W395A	TGCATTTCGCTCGCGTGGGCGATCC
M274A	CTTCAGCGCTTTCAGC
L326A	GCTCTTTCGCCACTTCA
V338A	CCATAGCCGCTTCCAGC
W260A	CAGGTCCCGCCAGTGAA

that were conserved in the CS alignment (see Appendix I) were identified. Then the equivalent residues were examined in the known three-dimensional structure of pig CS to assess the extent of burial and proximity to other hydrophobic residues. In this way, residues which were highly conserved, and were likely to contribute to structurally important hydrophobic cores, were identified. A few such residues were chosen in each of the two structural domains of CS. Table 2.2 lists the mutants and the homolgous pig CS residues, and the extent of conservation of each among CS sequences.

Protein Expression and Purification

Wild type and mutant pES*gltA* plasmids were expressed in MOB154, an *E. coli* strain with a stable mutation in the *gltA* gene (Wood *et al.*, 1983). Three litres of LB medium containing 25 μg/mL ampicillin were sufficient for the preparation of approximately 100-150 mg of CS, although larger amounts were realized for stable mutants and smaller amounts for less stable ones (data not shown). The purification procedure was adapted from that reported by Duckworth and Bell (1982). Briefly, cells were harvested after growth for 18-24 hours by centrifugation and disrupted by an Aminco French Press in standard CS buffer (20 mM Tris, 1 mM EDTA, pH 7.8). 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 2 litres of standard buffer to remove unbound material. A linear KCl gradient (50 to 300 mM, using 1 L) was used to elute the bound protein. CS activity was usually found in a prominent peak that eluted at about 100 mM KCl (see Figure 2.2 A for a typical elution profile), and was judged to be ~90% pure by SDS PAGE.

TABLE 2.2 Location of *E. coli* CS residues chosen for site-directed mutagenesis and their equivalents in the pig CS sequence

E. coli CS Residue	Location helix, domain	Pig CS Residue
F121	F, large	L128
L141	G, large	L148
L164	I, large	L178
M131	G, large	M138
L259	M, large	L269
W260	M, small	A270
M274	N, small	W284
L275	N, small	L285
L326	P, small	L338
L336	Q, small	L349
V338	Q, small	V350
L369	R, small	L382
W391	S, large	V405
W395	S, large	L409

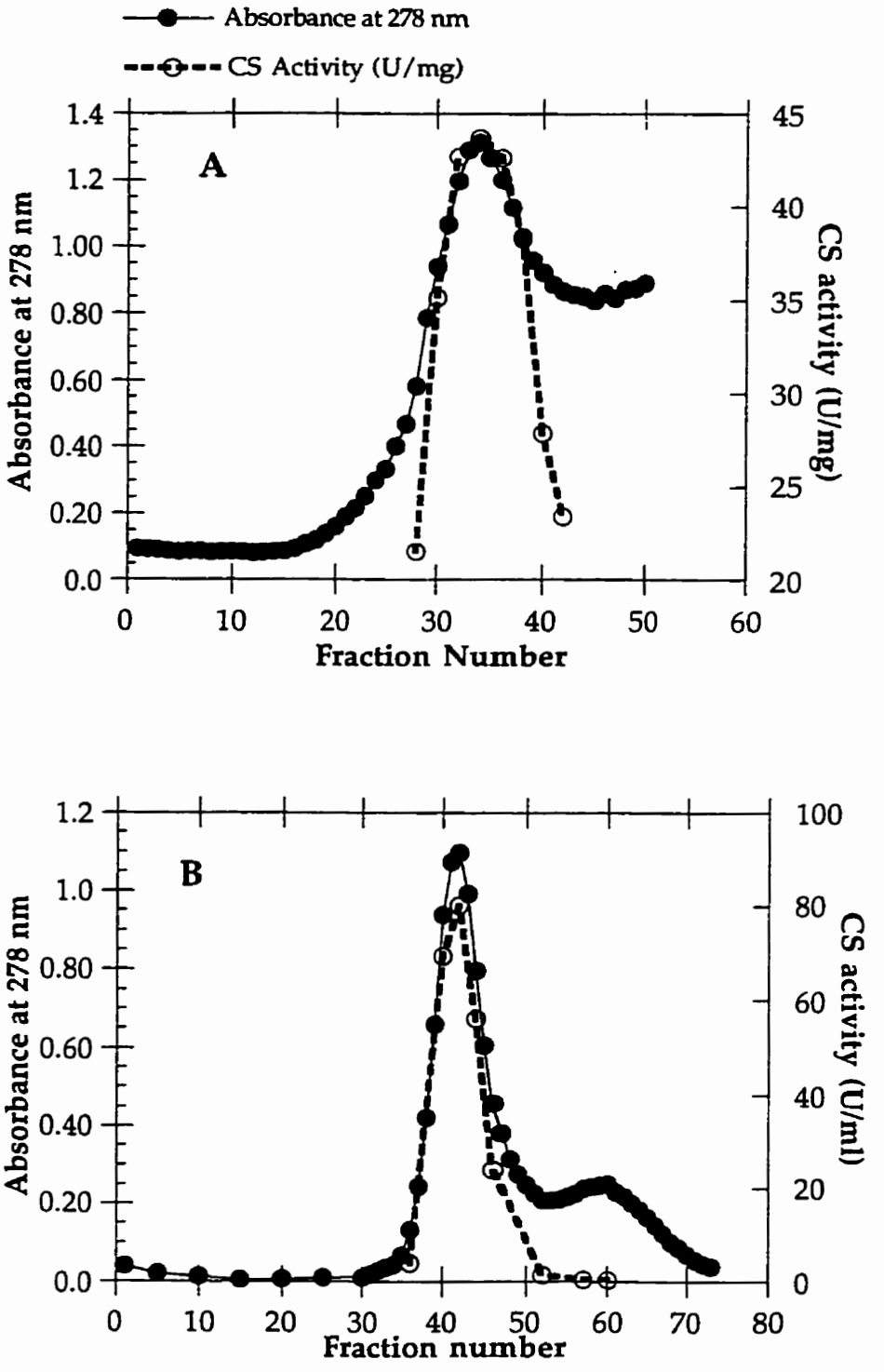


Figure 2.2 Typical elution profiles of CS from (A) DEAE-cellulose anion exchange column with a gradient from 50 to 300 mM KCl and (B) from Sephadex G200. Fraction size is about 12 mL (200 drops).

Fractions were assayed for citrate synthase using the method of Srere *et al.* (1963) and those with the highest specific activity were pooled and concentrated by ultrafiltration. A final gel filtration step (Figure 2.2 B) using a Sepharose 6B (separation range 10^4 - 4 X 10^6 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 100 mg, which is judged to be 99+% pure by SDS PAGE and/or electrospray ionization time-of-flight mass spectrometry (ESITOF MS). CS was usually stored highly concentrated (50 to 200 $mg.mL^{-1}$) in standard buffer at 4°C and was stable for months. Mutant proteins were routinely checked for their purity and molecular mass by ESI-TOF MS under denaturing conditions (see Chapter 3). Molar extinction coefficients were determined for all mutants (Table 2.3) and wild type CS by the method of Edelhoch (Edelhoch, 1967), and were typically around 48 000 $M^{-1}cm^{-1}$, except for the tryptophan mutants which have coefficients of about 3 700 $M^{-1}cm^{-1}$. The standard deviation for WT CS extinction coefficient measurements was about 2075 $M^{-1}cm^{-1}$, or about 5%.

Alkylation of CS at C206 using TFBA

Wild type CS was alkylated as described (Donald *et al.*, 1991) using 1,1,1-trifluorobromoacetone (TFBA). Briefly, CS (5 $mg.mL^{-1}$, 100 μM) in 20 mM sodium phosphate buffer, 0.1 mM KCl, 1 mM EDTA, pH 7.8 was incubated at 37°C for 5 minutes. To this TFBA in acetonitrile was added to a final concentration of 10 mM TFBA, and the mixture was incubated at 37°C for one hour, which is sufficient time for complete alkylation of cysteine 206 (Talgoy *et al.*, 1979) The alkylation reaction is shown below:

TABLE 2.3 Extinction coefficients for mutant, chemically modified and wild type citrate synthases

PROTEIN	Extinction Coefficient at 278 nm ($M^{-1}cm^{-1}$)	Tryptophan/ Tyrosine Ratio*
F121A	n/a	n/a
F383A	45547	0.222
L141A	n/a	n/a
L164A	n/a	n/a
L259A	45 091	0.219
L275A	44 492	0.220
L326A	43 100	0.228
L336A	43 597	0.222
L369A	45 757	0.216
M131A	47 493	0.236
M274A	42 710	0.224
R387L	42 711	0.210
TFBA-CS	41 576	0.223
V338A	44 852	0.221
W260A	35 867	0.148
W391A	38 389	0.150
WTCS	47 699	0.215

**CS has a tryptophan/tyrosine ratio of 0.187. For tryptophan mutants the ratio is 0.125.*

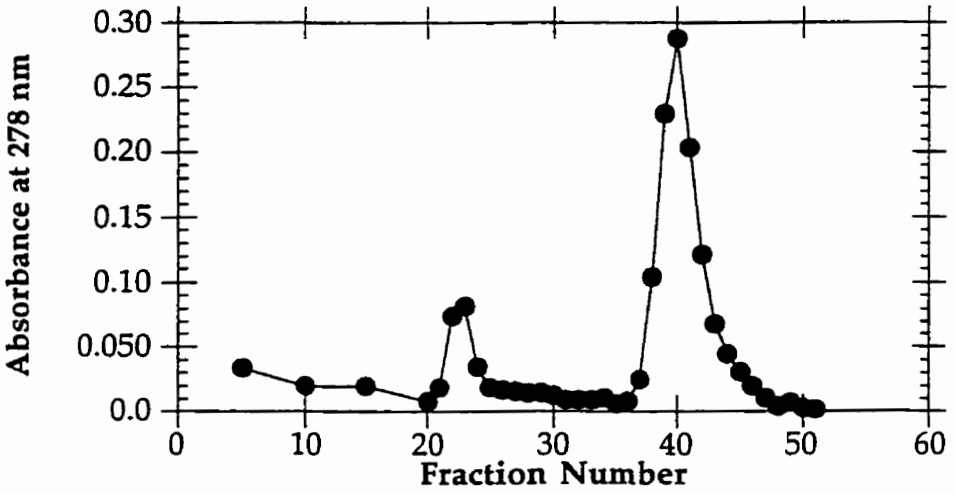
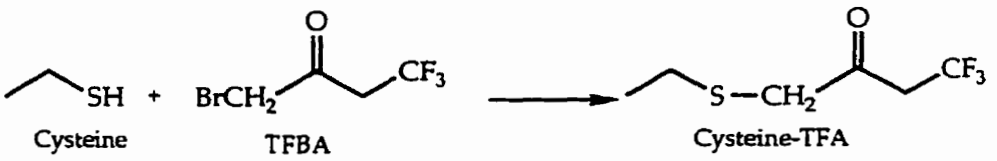


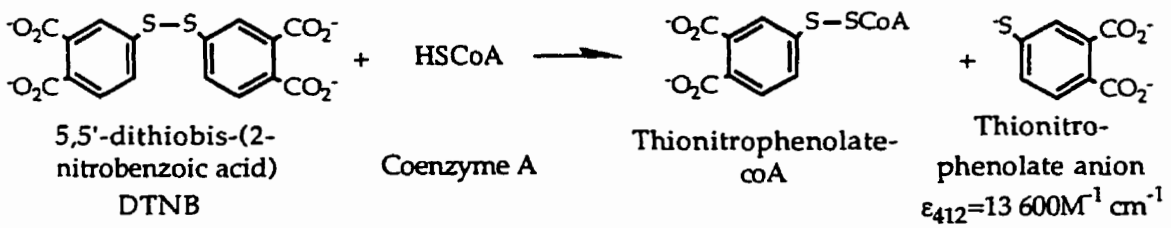
Figure 2.3 Purification of TFBA-CS using size exclusion chromatography on a Sepharose 6B column using 20 mM Tris, 1 mM EDTA, 50 mM KCl, pH 7.8. Fractions are about 12 mL each. The earlier-eluting peak was also observed on a Sephadex G200 column.



The reaction was quenched with dithiothreitol and the mixture was loaded onto a Sepharose 6B column for purification. The purification step described here is an improvement on previous work, and a typical elution profile is shown in Figure 2.3. A peak eluting much earlier than the main CS peak at around fraction 23 was likely aggregated CS. The second peak, corresponding to modified CS, elutes at a volume characteristic of that for unmodified, hexameric CS. The resulting protein was analyzed on a non-denaturing polyacrylamide gel as described later and studied by urea denaturation.

CS activity assays and kinetic measurements

CS catalyzes the condensation of OAA and AcCoA to produce citrate and CoA. The free SH of the CoA product is the basis for the CS assay as described by Srere *et al.* (1963), where DTNB was used to produce 1 mole of the thionitrophenolate (TNB^-) anion per mole of CoA produced:



This species absorbs strongly at 412 nm an extinction coefficient of 13 600 $M^{-1}cm^{-1}$, 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 µmol CoA per minute at room temperature. The standard CS assay contains 100 µM AcCoA, 100 µM OAA, and 50 µM

DTNB in 100 mM KCl, 20 mM Tris, 1 mM EDTA, pH 7.8, and the assay volume was typically 1 mL.

Limited kinetic measurements were carried out on some of the mutants that exhibited reduced enzyme activity in the presence of KCl. Typically two substrate saturation curves were obtained (one for OAA and another for AcCoA) in the presence of KCl in order to determine whether the reduced activity was due to defective substrate binding (a K_m effect) or whether the catalytic rate was affected (a K_{cat} effect).

Preparation of Urea Denaturation Curves

Stock urea solutions (9 to 10 M) in 20 mM Tris, 1 mM EDTA, pH 7.8, were prepared fresh and used within three days. Calculation of urea concentration (C_{urea}) using refractive index measurements of urea solutions by the equation

$$C_{urea} = 117.66(\Delta N) + 29.753(\Delta N)^2 + 185.56(\Delta N)^3 \qquad [4]$$

where ΔN is the difference in refractive index between the urea solution and buffer gave values within 1.5% of the concentration calculated from the weight of urea (Creighton, 1988). The exact volume of stock urea solution was calculated from the density of the solution, calculated by:

$$d = 1 + 0.2658W + 0.0330W^2 \qquad [5]$$

where d is the density of the urea solution and W is the weight fraction denaturant in the solution (Creighton, 1988).

Twenty to 30 different urea concentrations were typically used to define each denaturation curve by mixing the appropriate volumes of urea, buffer and CS solutions for a total solution volume of 1 mL. Unless otherwise

stated, CS concentrations of 0.1 mg.mL^{-1} (2.09 μM) were used, and were determined spectrophotometrically for the stock solution. Solutions were prepared and incubated overnight at room temperature before circular dichroism or fluorescence measurements.

Circular Dichroism and Fluorescence Spectroscopies

A JASCO A500 spectropolarimeter equipped with a 486 PC (running a BASIC program written by Dr. L. Kruczynski) was used for all circular dichroism (CD) measurements in a 0.1 cm quartz cell (Hellma). When needed, temperature was regulated by a Haake circulating water bath. To minimize the contribution of instrument drift during lengthy experiments, baseline measurements were taken at least every hour. The mean residue ellipticity at 222 nm (MRE_{222}) in units of deg.cm^2.dmol^{-1} was calculated for each point in the denaturation curve by the equation:

$$MRE_\lambda = {}^{\theta_\lambda M_{avg}}\!/_{10LC} \qquad [6]$$

where $\theta\lambda$ is the measured ellipticity in millidegrees at wavelength λ, M_{avg} is the average mass per amino acid of CS (112.4 u), 10 is a scaling factor, L is the pathlength of the sample cell in cm, and C is the concentration of the protein in mg.mL^{-1}. The error on the MRE_{222} obtained for WT CS was evaluated by pooling data from various separate experiments and determination of the standard deviation.

Fluorescence measurements were made on a Fluoro IV Fluorimeter interfaced with a 386 PC running a BASIC program written by Dr. L. Kruczynski. Samples were excited at 280 or 295 nm and emission spectra collected. A 1 cm quartz cuvette was used for all measurements and precautions were taken to minimize dust and particulate matter, by filtering

all buffers and stock solutions through a 0.2 μM cellulose acetate filter (Nalgene).

Reversibility of CS denaturation

Samples preincubated at different urea concentrations were dialyzed against 0 and 4 M urea to test the reversibility of the two transitions in CS unfolding. CD and fluorescence measurements of the samples were carried out before and after dialysis. Since native extinction coefficients would not be useful under denaturing conditions, the Bradford protein assay was used for protein determination (Bradford, 1976). The samples were also checked for recovery of citrate synthase activity.

ANS fluorescence enhancement

ANS, 8-anilinonaphthalene sulfonic acid (Figure 2.4) has been used extensively as marker for exposed hydrophobic surfaces, a method originally introduced by Semisotnov (Semisotnov, 1991). Ten μM ANS was included in each of the urea solutions in the CS denaturation curve to determine whether CS exhibited exposed hydrophobic surfaces. ANS emission was monitored at 470 nm upon excitation at 380 nm. ANS fluorescence was not enhanced by urea when the experiment was repeated in the absence of protein.

Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel electrophoresis (PAGE) experiments were carried out on all mutant proteins as described (Molgat *et al.*, 1992). Briefly, 15.7 X 18 cm slab gels with a thickness of 0.1 cm were used, with acrylamide concentrations of 6-8%. Tris-glycine buffer 33 mM, pH 7.5, was used for the running buffer and a stacking gel of 5% acrylamide was used to focus the protein bands. Coomassie Blue in 10% acetic acid, 25% methanol was used to visualize the protein bands. CS exhibited 3 bands; two of these are

hexamers and the other dimer (Molgat *et al.*, 1992). This experiment was used to determine whether a given mutation caused a change in the association/dissociation of the dimers to form hexamers.

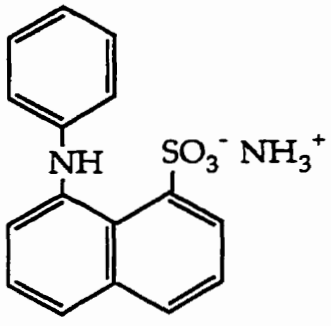


Figure 2.4 Structure of the ammonium salt of 8-anilino-naphthalene sulfonic acid.

Urea gradient gel electrophoresis

Urea gradient acrylamide gels were prepared by a method similar to that described (Goldenberg, 1988) where the urea concentration gradient is perpendicular to the direction of electrophoresis. Gels (15.7 x 18 cm and 0.1 cm thick) were cast from high to low urea concentration, pouring the acrylamide with a gradient former from the top of the gel. Acrylamide concentrations were typically 5 to 10%, and a reverse acrylamide gradient (1-2% difference) was employed to minimize the reduced mobility of protein due to viscosity differences between high and low concentrations of urea. The success of the gradient formation was evaluated by including bromophenol blue in the concentrated urea solution and inspecting the colour gradation by eye. Typically a sample of 0.1 to 0.3 mg CS in 20 mM Tris, 1 mM EDTA pH 7.8 was loaded and was run for 5 hours at 200 volts (25 mA) in a 33 mM Tris-glycine buffer, pH 7.5. The gels were stained with Coomassie Blue as described for nondenaturing gel electrophoresis.

Results

Urea denaturation monitored by Circular Dichroism

CD in the far UV (about 220 nm) is a measure of chirality of the peptide bond in proteins, and as such provides information on the secondary structure content. This property was used to follow the equilibrium unfolding of CS by urea. The CD spectrum of CS (Figure 2.5) is typical of an all α-helix protein (Adler, 1973), with a double minimum at 208 and 220 nm. CS from *E. coli* is about 30% identical in sequence to the pig CS for which several X-ray three-dimensional structures are available. From those structures, the pig CS is classified as an all α-helix protein, with a fold that is unique to CS.

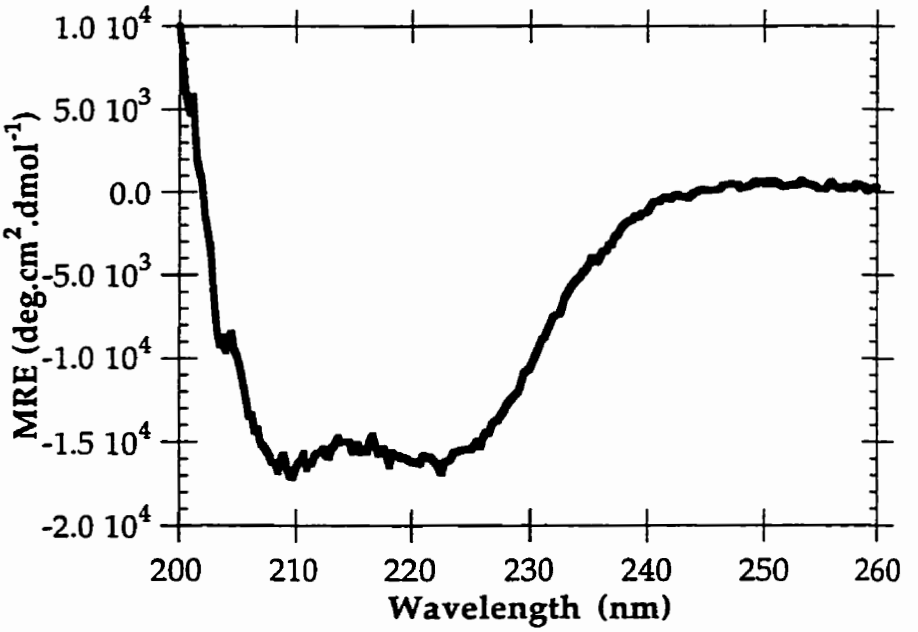


Figure 2.5 Circular dichroism spectrum of CS (2.09 μM) in 20 mM Tris, 1 mM EDTA, pH 7.8. Path length used was 0.1 cm.

The mean residue ellipticity for WT CS at 222 nm for 0.1 $mg.mL^{-1}$ sample in a 0.1 cm pathlength cuvette ranged between -14 162 and -20 424 $deg.cm^2.dmol^{-1}$ with an average of -16 885 ± 1 600 $deg.cm^2.dmol^{-1}$. Therefore the error on the CD measurements was about 10%. This error may have been due to a combination of inaccuracy in protein determination and imprecise instrument calibration. The instrument was routinely calibrated with standard CD signal calibrants, androsterone or camphorsulfonic acid (Adler, 1973).

As shown in Figure 2.6 A, as urea was added the breakdown of secondary structure of CS occured in two transitions. In the first transition (Transition I), about 60% of the total ellipticity was lost, with a midpoint of 1.75 M urea. An intermediate state, somewhat resistant to further denaturation, persisted in the range of 3.5 to 5 M urea and contained the remaining ellipticity, which was lost in the second transition (Transition II) with a midpoint of 6.5 M urea. Representative CD spectra of CS at 0, 4, and 9 M urea are shown in Figure 2.6 B. Guanidine hydrochloride, a somewhat more effective chaotropic agent, also denatured CS in a biphasic manner (Figure 2.7), with transition midpoints of about 1 and 3 M.

Denaturation of TFBA-modified CS

Denaturation of TFBA-modified CS was also biphasic (Figure 2.8), with apparent midpoints of transitions of 1.5 M and 6.5 M, almost identical to the unmodified citrate synthase; Transition I midpoint was at a slightly lower urea concentration than for unmodified CS. The experiment was carried out to determine if and how hexamerization affects the denaturation profile of CS. TFBA-modified CS does not form hexamers under most conditions, as found by ESI-TOF MS as discussed later in Chapter 3 and by non-denaturing

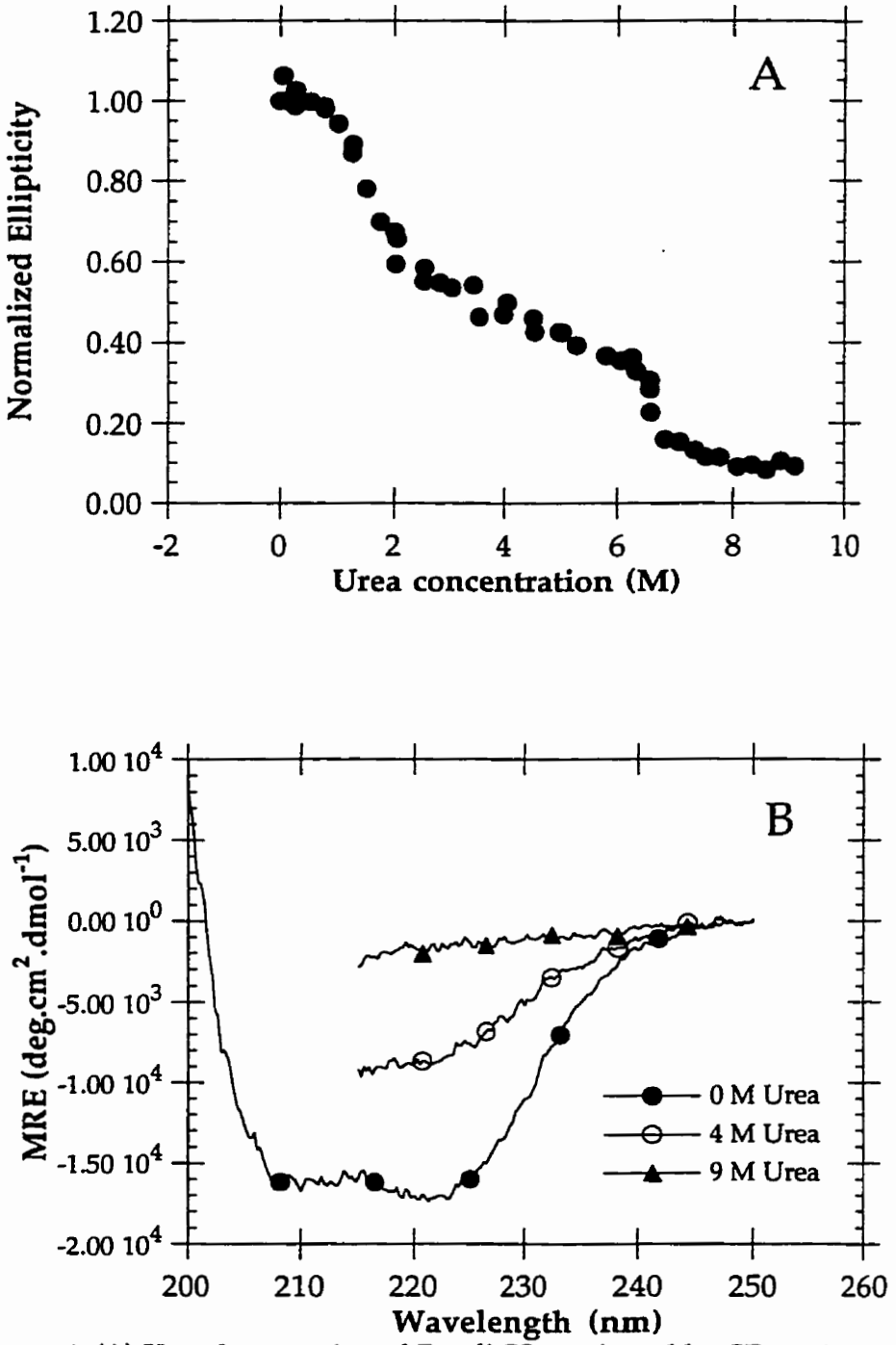


Figure 2.6 (A) Urea denaturation of *E. coli* CS monitored by CD at 222 nm. The ellipticity was normalized against the value for CS in the absence of denaturant. The data were pooled from several experiments. (B) CD spectra of CS at 0, 4, and 9 M urea. The spectra in the presence of urea were not measured below 215 nm as urea absorbs strongly in this region. MRE stands for mean residue ellipticity.

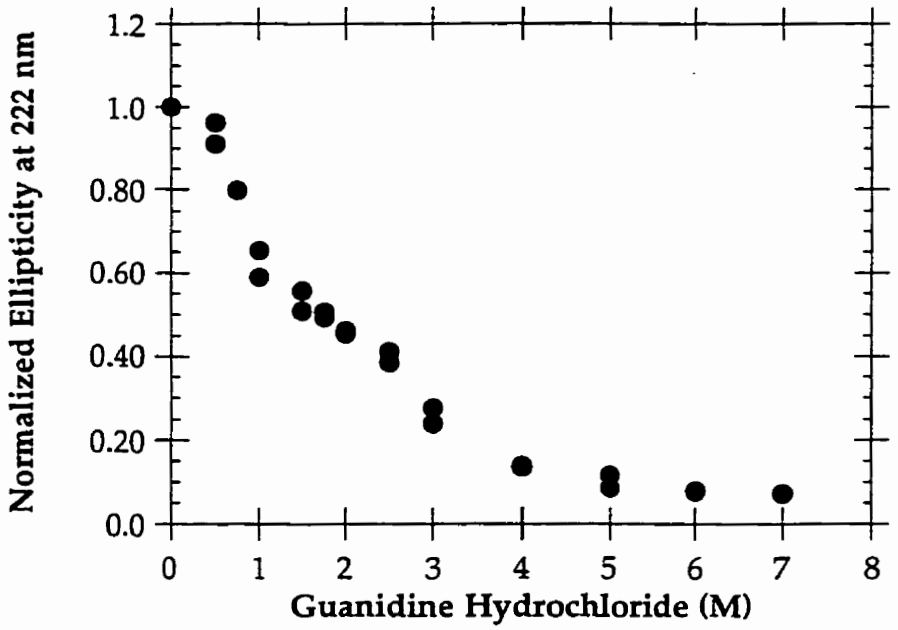


Figure 2.7 Guanidine hydrochloride denaturation of CS as monitored by CD. Protein concentration was 0.1 mg.mL^{-1}. The points represent data from two experiments.

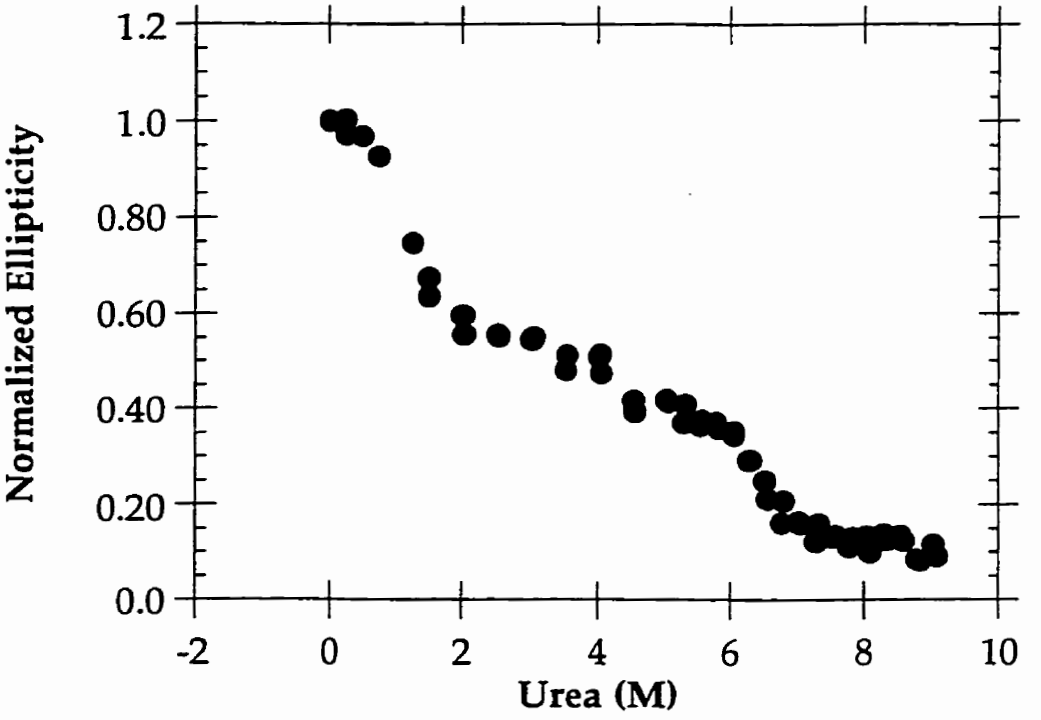


Figure 2.8 CD-monitored urea denaturation of TFBA-CS. The results shown are from two separate experiments.

gel electrophoresis (data not shown). The slight decrease in the midpoint of Transition I may be a reflection of decreased stability due to the absence of hexamers.

Unfolding of E. coli CS monitored by fluorescence spectroscopy

E. coli CS contains three tryptophans at positions 260, 391, and 395, and 16 tyrosines. The tryptophans usually dominate the fluorescence emission spectra of proteins since energy transfer usually occurs from the tyrosines to tryptophans (Creighton, 1988, Lakowicz, 1983), and as shown in Figure 2.9, this is the case with CS. At 280 nm, both tyrosine and tryptophan residues are excited, whereas at 295 nm or higher, only tryptophans are selectively excited (albeit not at the maximum absorbance wavelength). Spectra of CS excited at 280 and 295 nm were similar in shape, with emission maxima at 320 and 325 nm, respectively, and a shoulder peak at about 360 nm. The emission wavelength of tryptophans is strongly dependent on the polarity of its environment; in a polar environment, such as aqueous solvent when a protein is unfolded, tryptophan emission is around 360 nm. In an apolar environment, such as the interior of a protein, the emission maximum is blue-shifted. While the emission maxima obtained for CS were indicative of buried tryptophans, the presence of a shoulder peak at 360 nm suggests that a fraction of the population of tryptophans is solvent exposed, and correspondingly red-shifted. As shown later, the mutant W260A does not exhibit this shoulder peak in its emission spectrum.

Both fluorescence intensity and emission maxima can give information about the environment around tryptophans, a convenient measure of tertiary interactions in proteins (Creighton, 1988). This feature was used to monitor the unfolding of *E. coli* CS by urea. The relative change

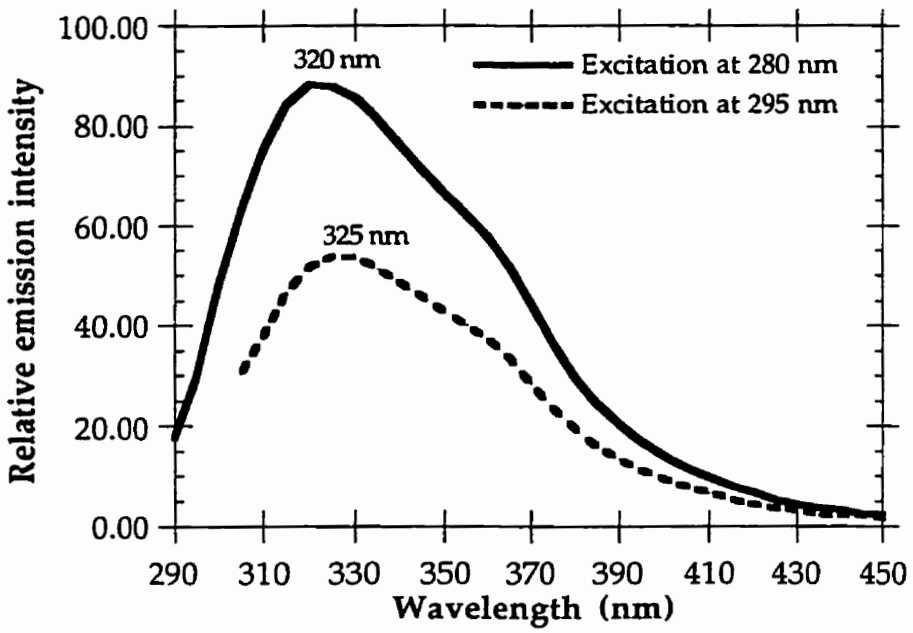


Figure 2.9 Emission spectra of CS at two different excitation wavelengths. At 280 nm, both tryptophans and tyrosines are excited, but at 295 nm, only the tryptophans contribute to the emission spectrum. CS concentration was 2.09 μM in 20 mM Tris, 1 mM EDTA, pH 7.8. A 1 cm quartz cuvette was used.

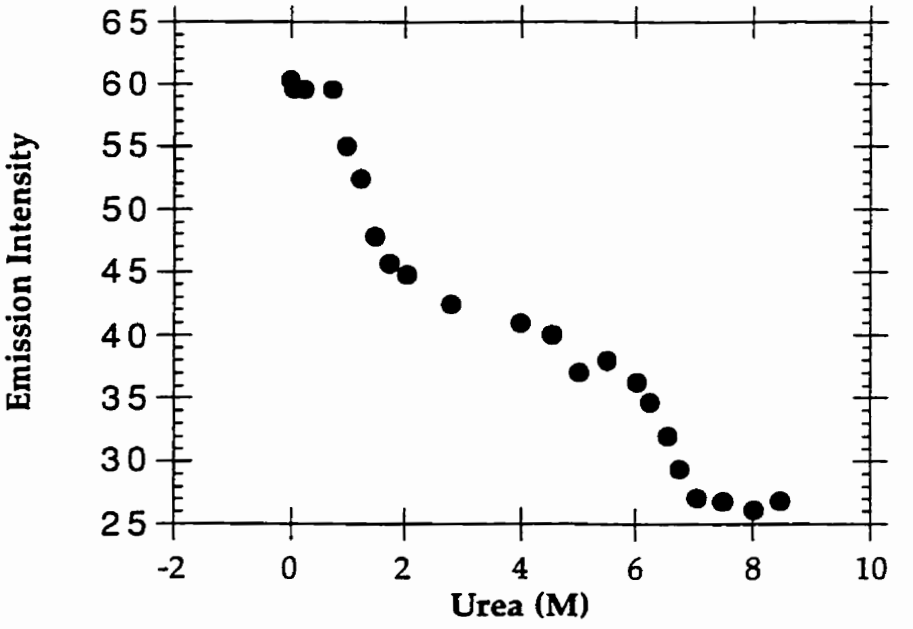


Figure 2.10 Fluorescence monitored urea denaturation of CS (2.09 μM) in 20 mM Tris, 1 mM EDTA, pH 7.8. Excitation wavelength was 280 nm and emission was monitored at 330 nm. Excitation at 295 nm gave an identical denaturation profile (not shown).

in the intensity of emission due to unfolding by urea is shown in Figure 2.10. As with the CD-monitored unfolding, fluorescence also reveals a biphasic breakdown of tertiary interactions. The proportion of structure loss in each transition was comparable to that lost as monitored by CD, suggesting that the tryptophan emission is a measure of global tertiary interactions rather than local interactions around single tryptophan residues.

A plot of the emission maximum wavelength against urea concentration (Figure 2.11) shows that the change in the emission maximum due to unfolding was biphasic as well. In Transition I, the shift was small (10 nm), from 320 to 330 nm. This maximum remained unchanged over the concentration range 2 to 6 M urea, and then a further 30 nm shift occured in Transition II. Representative fluorescence spectra of folded, intermediate, and unfolded CS are shown in Figure 2.12, illustrating the change in the maximum of emission as well as the shape of the spectra due to unfolding. In the unfolded spectrum, a small peak at about 305 nm was observed, which corresponds to the emission of tyrosine residues. Tyrosines are not subject to changes in maximum emission wavelength due to environmental changes as tryptophans are, and thus a peak was revealed in the unfolded spectrum due to the shift of the emission maximum of the tryptophan residues.

Inactivation of CS activity by urea

Urea inactivates CS activity completely by 1 M urea (Figure 2.13). At 50% inactivation, the urea concentration was 0.6 M, well below the transition midpoint obtained by fluorescence or CD and at which the amount of secondary or tertiary structure loss was virtually unchanged. This observation suggests that the mechanism of inactivation may be either the dissociation of the minimal unit for activity, the dimer, into monomers, or

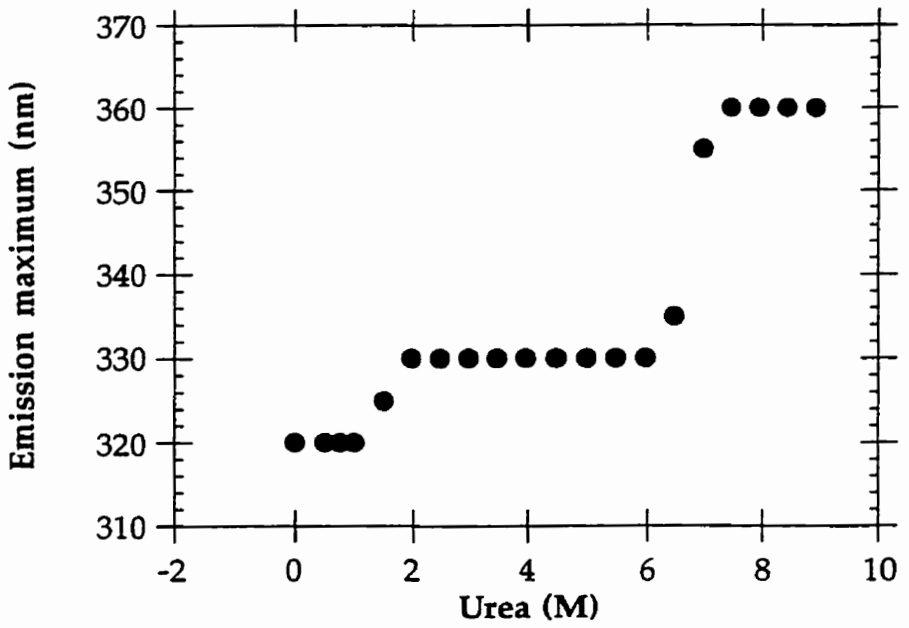


Figure 2.11 The shift in emission maximum of CS during urea denaturation. Excitation wavelength was 280 nm.

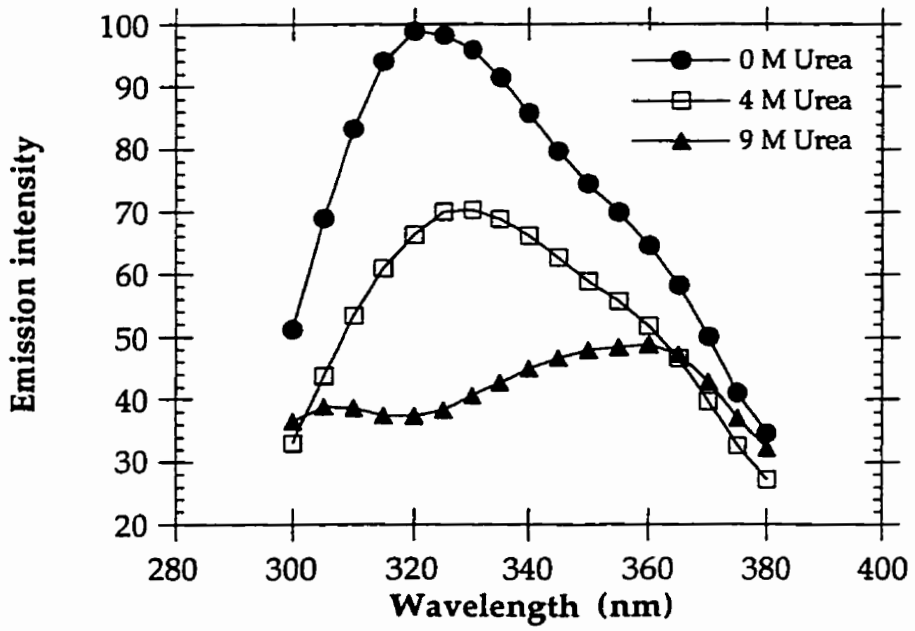


Figure 2.12 Fluorescence spectra of CS at in 0, 4, and 9 M urea. Excitation wavelength was 280 nm.

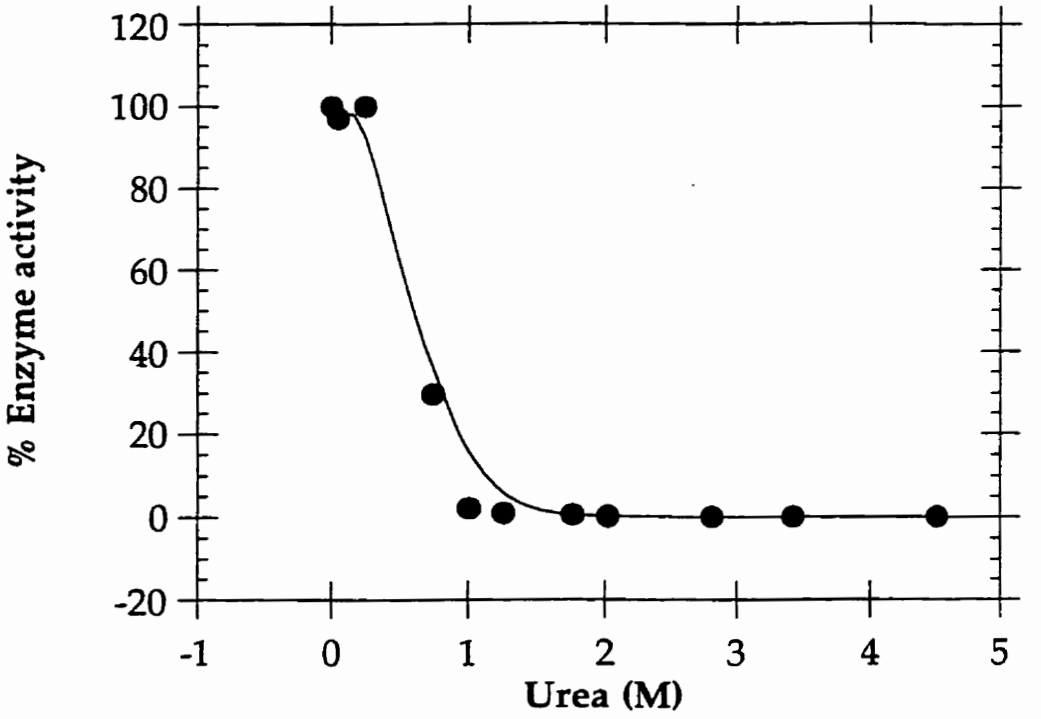


Figure 2.13 Inactivation of CS by urea. The enzyme activity was completely abolished at just under 1 M urea.

the preferrential binding of urea at the active site, which would prevent substrate binding.

Reversibility of Urea Unfolding of CS

Attempts to refold CS from 9 M urea by dialysis resulted in formation of a precipitate, indicating that irreversible aggregation had taken place. To test the reversibility of Transition II, samples that had been preincubated in 0, 4, and 9 M urea were dialyzed against 4 M urea and their ellipticity measured before and after the dialysis (Figure 2.14). Dialysis of the unfolded CS sample (in 9 M) against 4 M urea resulted in the recovery of ellipticity that was characteristic of CS in 4 M urea, suggesting that the Transition II is reversible. The control CS samples that had been preincubated at 4 and 0 M urea also showed the same amount of ellipticity after dialysis against 4 M urea. Samples in 4 M urea dialyzed against buffer retained the same ellipticity and did not exhibit CS activity (not shown), suggesting that the Transition I was not reversible.

Thermal unfolding of CS

The CD-monitored thermal unfolding of CS was biphasic, as shown in a representative experiment in Figure 2.15. A stable intermediate persisted up to 70°C and was only completely unfolded at temperatures near 90°C. No refolding was observed upon cooling after thermal denaturation, and generally, the protein precipitated. It was difficult to obtain consistent thermal denaturation data in separate trials; the inconsistency may lie in the fact that the denaturation curve seemed to depend on the rate of heating, which was not easy to control manually. Nevertheless, the data indicated that thermal induced equilibrium unfolding was biphasic as was observed with

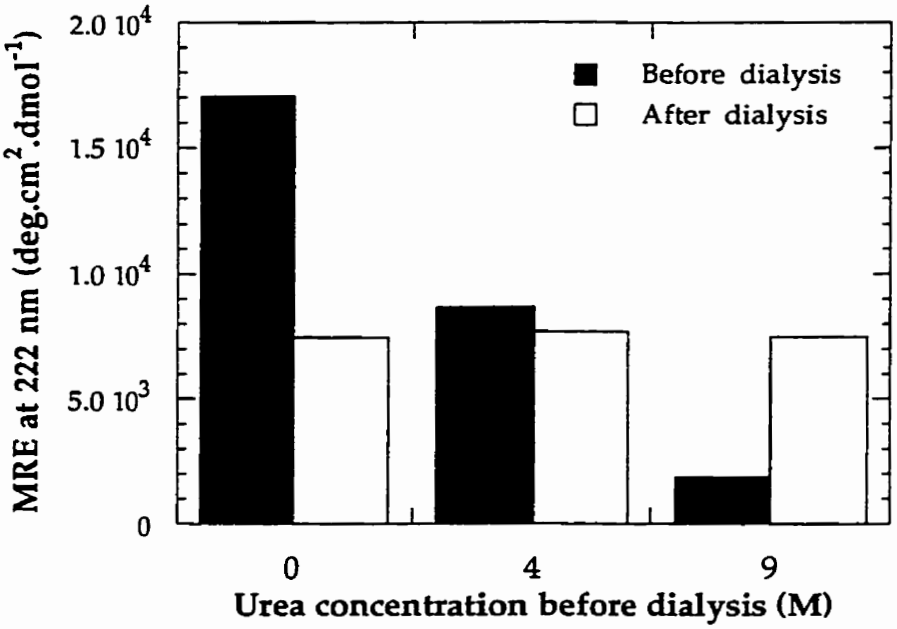


Figure 2.14 Dialysis of native, partially unfolded, and unfolded CS against 4 M urea to test for reversibility of the second transition. The native and 4 M samples were tested in parallel for comparison. The ellipticity at 222 nm was regained for the unfolded sample. The absolute ellipticities are plotted.

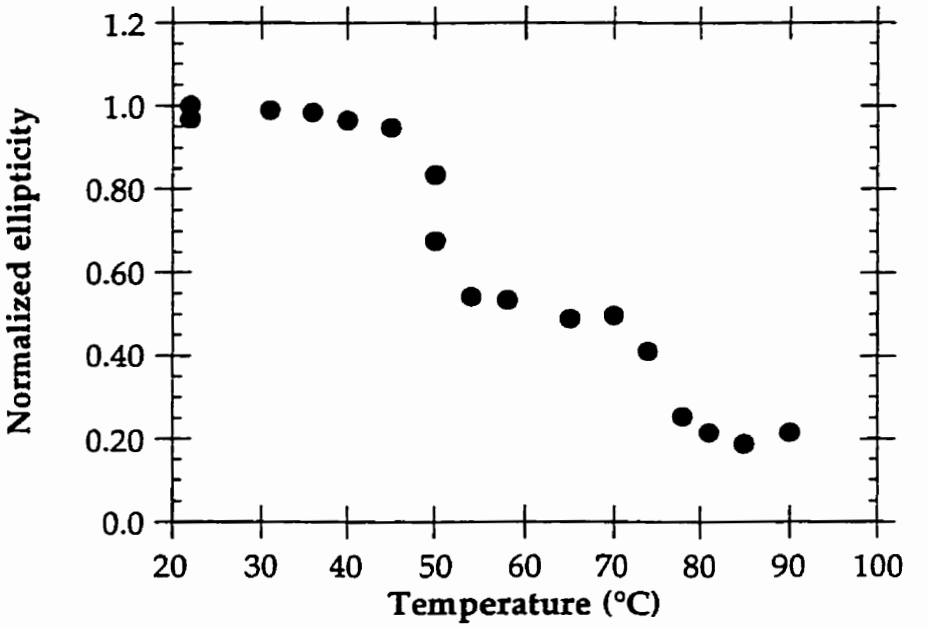


Figure 2.15 Thermal unfolding of CS as monitored by CD. The ellipticity was normalized against the signal obtained at room temperature (22°C).

chaotropic agent-induced unfolding, suggesting a similar unfolding mechanism.

Fluorescence is highly dependent on temperature (Schmid, 1988), and thus reliable fluorescence-monitored thermal denaturation data could not be obtained.

ESI-TOF MS-monitored thermal denaturation was used initially as an attempt to identify when subunit dissociation occurs in the thermal unfolding transition of CS. TFBA-modified CS does not form hexamers, and was used to eliminate the contribution of the hexamer to dimer dissociation. Heating of TFBA-modified CS to temperatures where the stable intermediate is present resulted in a mass spectrum that contains monomeric CS and aggregates from dimer to octamer, as shown in Figure 2.16. The dimer present is not likely to be the native dimer, as the number of charges it acquired in the ESI source is larger than the native dimer, indicating that it is not in the same conformation as the native species (the relationship between the number of charges obtained and the conformation is described in the Introduction to Chapter 3).

ANS fluorescence enhancement

ANS binds native CS at two molecules per subunit (Talgoy *et al.*, 1979). ANS was used in this study to probe for the presence of exposed hydrophobic surfaces due to partial unfolding. Figure 2.17 shows the fluorescence intensity change of ANS at 470 nm as a function of urea concentration in the presence and absence of CS. ANS fluorescence was not affected by urea alone, as illustrated by Figure 2.17. However in the presence of CS, ANS fluorescence decreases initially, reaching a minimum at 0.75 M urea. This is then followed by a large increase in ANS fluorescence between 1 and 2.5 M urea, indicating

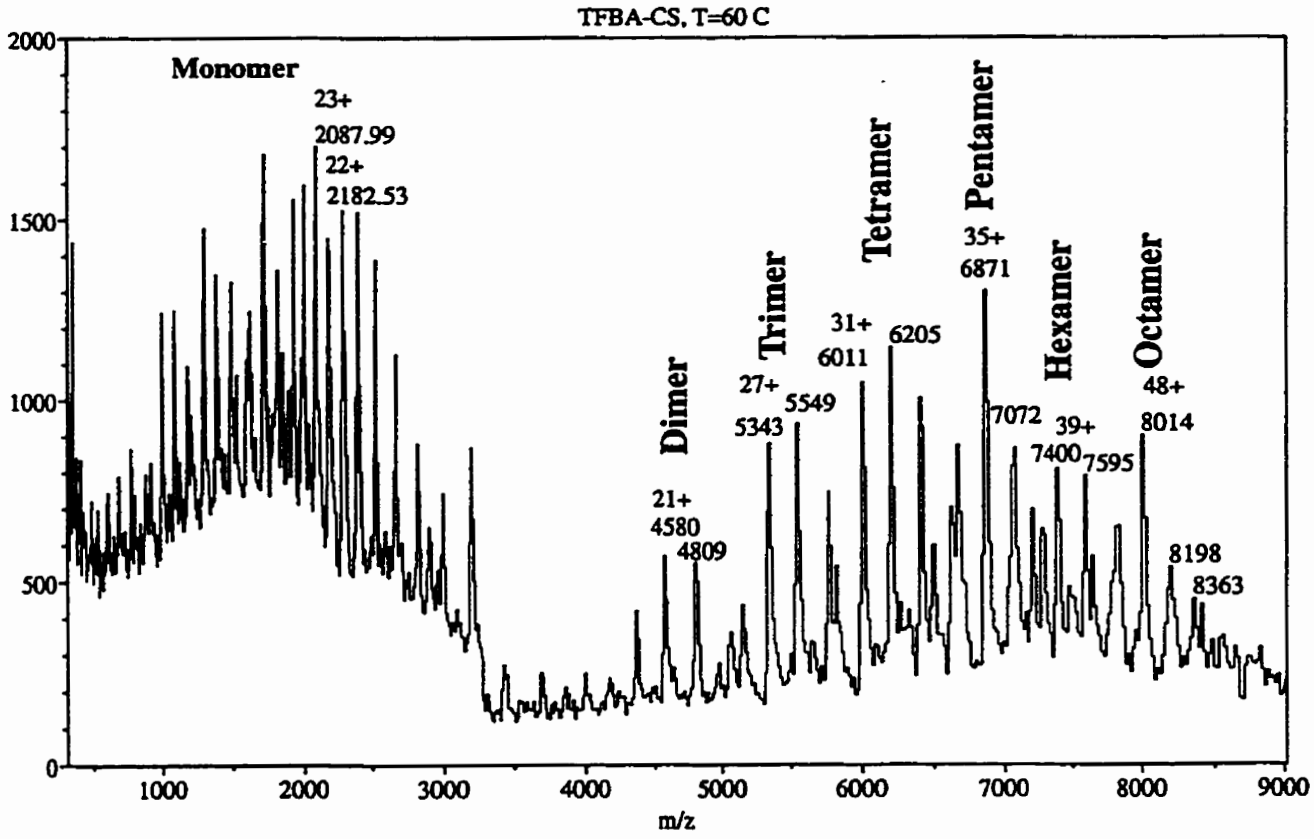


Figure 2.16 ESI-TOF m/z spectrum of TFBA-CS at 60°C. Sample concentration was 10 μM in 5 mM ammonium bicarbonate, pH 7.5. Sample heating was done prior to electrospraying. Monomer and aggregates up to octamers are observed.

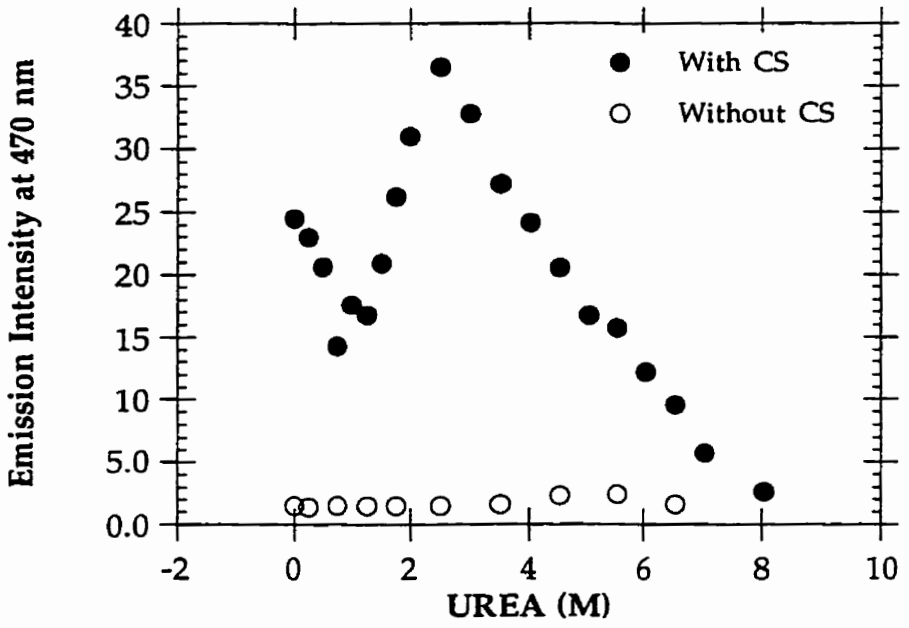


Figure 2.17 ANS fluorescence at 470 nm in the presence and absence of CS (2.09 μM) as a function of urea concentration. The excitation wavelength was 380 nm, and 10 μM ANS was used.

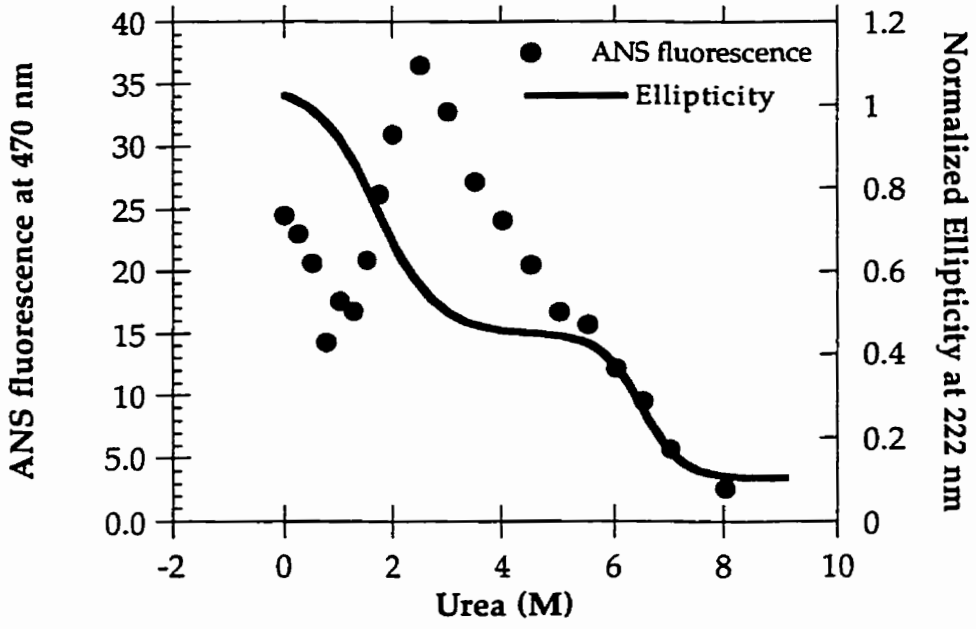


Figure 2.18 A comparison of the ANS fluorescence profile with the CD monitored denaturation profile of CS. The line representing the denaturation curve was drawn to show the trend of the data. The maximum ANS fluorescence coincides with the second half of Transition I of the denaturation curve (2.5 M urea).

an increase in the amount of ANS binding, perhaps due to the presence of exposed hydrophobic surfaces. At concentrations higher than 2.5 M, ANS fluorescence drops gradually until 9 M urea is reached. The concentration at which ANS fluorescence is maximal (2.5 M) corresponds to the second half of the Transition I (Figure 2.18 is an overlay plot of the CD-monitored urea unfolding of CS and ANS binding for comparison). The decrease in ANS fluorescence in a urea concentration range where little change occurs according to the CD and fluorescence measurements indicates that hydrophobic surfaces are becoming inaccessible to ANS, but perhaps by a process other than unfolding.

ANS fluorescence was also used to probe for exposed hydrophobic surface as a function of thermal denaturation of CS (data not shown). Fluorescence was found to be at a maximum at 50°C, and decreased gradually thereafter, a behaviour which is similar to that described above for urea unfolding. This finding with ANS binding is further evidence that the equilibrium unfolding of CS by temperature and urea are similar, and that the intermediate state can be achieved using either method.

Transverse urea gradient gel electrophoresis

In an attempt to identify the point at which the subunits of CS dissociate in the unfolding transition, urea gradient gel electrophoresis was used. In this experiment, a sample of CS was electrophoresed in a direction perpendicular to the urea gradient, such that the CS experienced different urea concentrations across the gel. The rate of migration versus urea concentration should give information about the hydrodynamic properties of the protein, resulting in a faster moving species for a compact native structure and retarded mobility as unfolding occurs at higher urea concentrations. CS

dimer and hexamer species are readily separated on non-denaturing gels as one band corresponding to dimer and two hexameric bands, so it may be possible to detect the dissociation of these species in the presence of urea. A gradient gel at low urea concentrations (0 to 1 M) revealed that subunit dissociation occurs at very low concentrations, approximately <0.25 M urea (data not shown). This is well below the concentration at which maximal ANS binding occurs, suggesting that the enhanced ANS binding is not due to subunit dissociation. It is also consistent with the loss of activity (dependent on the dimeric structure) at about 0.6 M urea mentioned already. Using a wider urea concentration range, it was discovered that a number of bands appear between 3 to 6 M urea (Figure 2.19). The mobility of these bands decreased as higher urea concentrations were reached, and eventually, the bands coalesced to form a single or double band at high urea concentrations. These observations were reproducible. The presence of the multiple bands at intermediate urea concentrations indicated that the intermediate state observed at 4 M urea by fluorescence and CD is a collection of aggregates, in agreement with the observation of aggregates by ESI TOF MS as presented in Figure 2.16.

Urea denaturation of A. anitratum and Chimera CS

In contrast with urea denaturation of *E. coli* CS, *A. anitratum* CS, another gram negative CS (Morse & Duckworth, 1980), exhibited a single transition, with a midpoint of 5.5 M urea, as shown in Figure 2.20 A. The enzyme is resistant to denaturation up to 4 M urea, making it much more stable than the *E. coli* enzyme as reported previously (Molgat *et al.*, 1992).

Chimeras of CS constructed by Gilles Molgat (Molgat *et al.*, 1992) were also denatured by urea. The chimeras were constructed by exchange of the

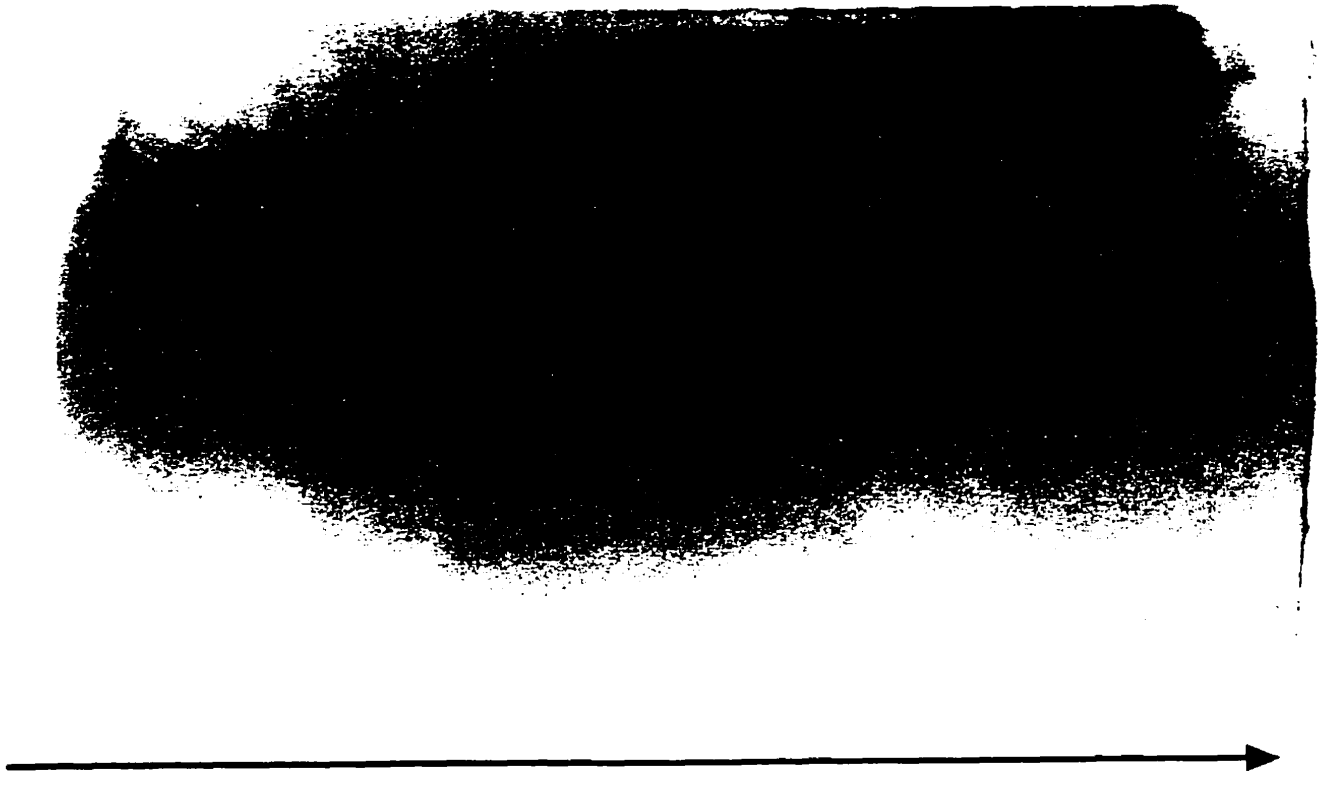


Figure 2.19 Transverse urea gradient gel electrophoresis of *E. coli* CS (0.3 mg in 20 mM Tris 1 mM EDTA, pH 7.8). The arrow indicates the direction of increasing urea concentrations (about .5 to 9 M). A reverse acrylamide gradient from 6 to 4.8 % was used. The gel was run (top to bottom) towards the positive electrode in glycine buffer for 3.5 hours at 150 volts (25 mA) as described in the text.

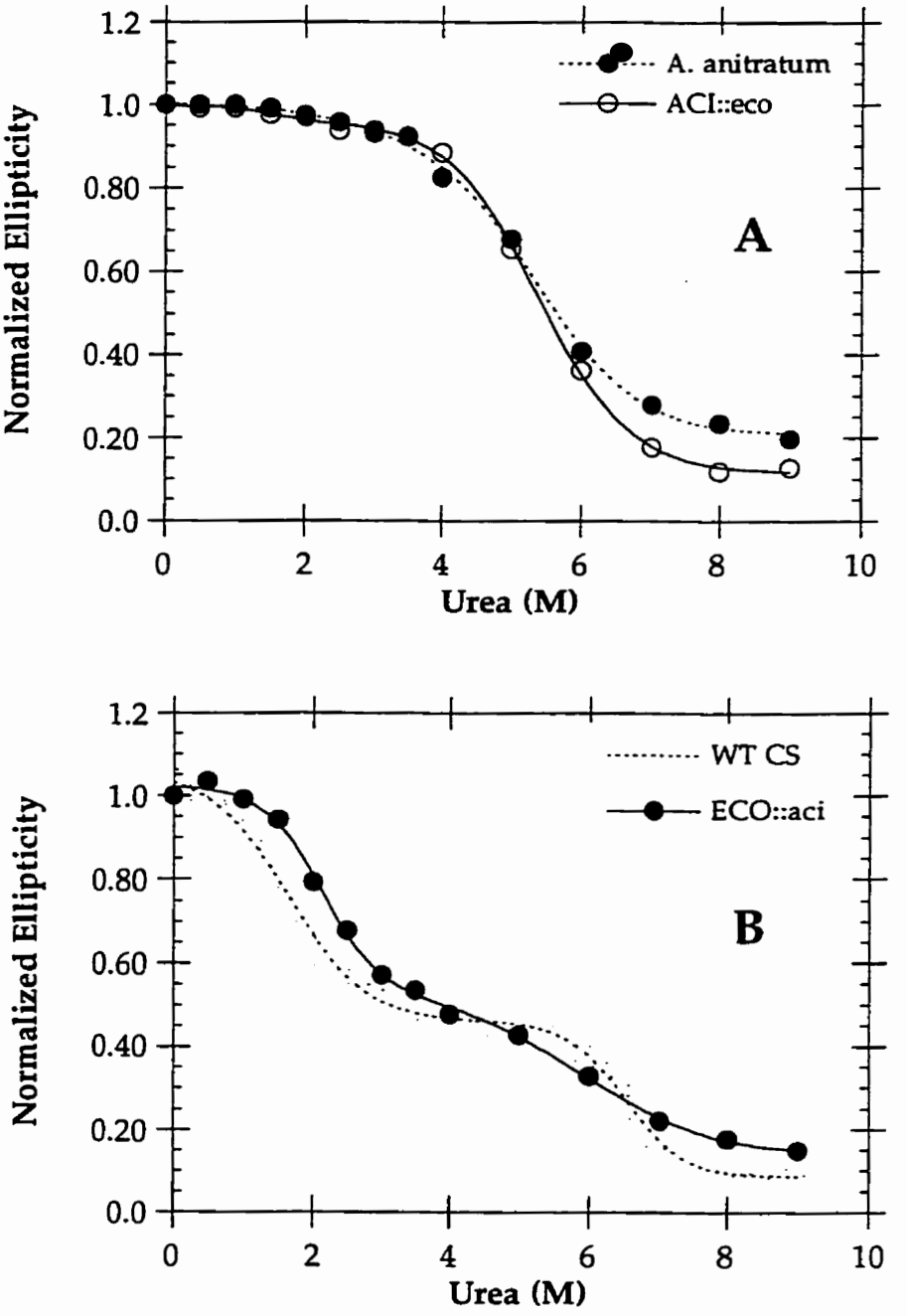


Figure 2.20 Denaturation of CS parent proteins (*E. coli* and *A. anitratum*) and chimera proteins of the structural domains (ACI::eco and ECO::aci). (A) Denaturation of the *A. anitratum* CS and the ACI::eco chimera containing the large domain from *A. anitratum*. Unlike *E. coli* CS, denaturation is a single transition. (B) Denaturation of the *E. coli* CS and the ECO::aci chimera containing the *E. coli* CS large domain. Both proteins exhibited biphasic denaturation. Lines are drawn to show the trend of the data.

structural domains of CS (small: residues 264 to 374, and large: residues 1 to 263 and 375-426) between the two enzymes. The data are also presented in Figure 2.20. The ACI::eco chimera, where the large domain is derived from *A. anitratum* and small one from *E. coli*, had a denaturation curve similar to the *A. anitratum*, that is, a single transition. Similarly, the ECO::aci chimera, where the large domain is derived from the *E. coli* and small domain from *A. anitratum*, exhibited a biphasic denaturation curve similar to that for *E. coli* CS, except that the Transition I appears to have been stabilized relative to the wild type *E. coli* CS. The data suggest that the source of the large domain dictates whether the denaturation profile is biphasic or not.

The experiments in Figure 2.20 actually formed the foundation for a part of the present thesis project, and a hypothesis was formulated based on these results: that the biphasic denaturation observed for *E. coli* CS and ECO::aci CS is a result of independent unfolding of the two structural domains of CS, with the large, less stable domain from *E. coli* unfolding first. Two points are in support of this hypothesis: first, the breakdown of secondary structure (monitored by CD) and tertiary structure (monitored by intrinsic fluorescence) are coincident, implying that the unfolding is cooperative in each transition (i.e. tertiary and secondary structure breakdown were simultaneous). Second, the amount of structural loss in each transition corresponds approximately to the amount of structure associated with each of the two domains of CS based on the three-dimensional model. The approach chosen to test this idea was to destabilize each of the two domains selectively by single site-directed mutants in the hydrophobic cores and assess the stability changes due to the mutations. Characterization of CS mutants constructed for this purpose is reported in the following sections.

Purification of mutant citrate synthases

All mutants chosen, except W395A, were isolated in good yields using the procedure outlined for wild type CS. With some mutants, yields were better than for wild type (data not shown), and can perhaps be attributed to their increased stability. Circular dichroism of all of the mutants showed that they contain essentially the same amount of secondary structure as wild type protein. Mean residue ellipticities for the mutant proteins are compared with the value for wild type in Table 2.4. The standard deviation for a series of values for wild type CS was 1 600 deg.cm^2.dmol^{-1}, which corresponds to a 10% error. Storage of some of the mutants for extended periods of time in the usual conditions (4°C, in CS buffer, and concentrated) resulted in the formation of precipitates; the mutant L164A, a large domain mutant, was especially prone to this.

Citrate Synthase Activities of Mutants

The specific activities of the mutant CSs are reported in Table 2.5. One mutant had a specific activity similar to wild type values (L326A), but the majority exhibited marked changes. Those that had very reduced activities (10 times or more) include M131A, L259A, M274A and L275A. F383A and R387L which were previously constructed and studied (Pereira *et al.*, 1994) have very low activities. Most of the remaining mutants exhibited some reduced activities, including F121A, L141A, W260A, L336A, V338A, and L369A. One mutant L164A exhibited elevated activity.

Enzyme kinetic measurements on selected mutants

Some kinetic measurements were done on mutants with significantly altered enzyme activity, to understand the reason for the change. Possible reasons include a shift in the allosteric equilibrium, which would reduce the

TABLE 2.4 CD Properties of Citrate Synthase Mutants

Citrate Synthase	MRE_{222} deg.cm^2.dmol^{-1}
WT	-16 885
TFBA CS	-15 586
F121A	-17 890
M131A	-16 646
L141A	-18 015
L164A	-17 100
L259A	-15 872
W260A	-14 243
M274A	-18 569
L275A	-18 491
L326A	-17 345
L336A	-17 256
V338A	-15 879
L369A	-16 743
F383A	-17 164
R387L	-16 348
W391A	-16 135

TABLE 2.5 Specific Activities of Wild Type and Mutant Citrate Synthases

Citrate Synthase	Specific Activity U/mg
WT	70
F121A	12
M131A	1
L141A	30
L164A	90
L259A	3
W260A	10
M274A	5
L275A	0.06
L326A	70
L336A	30
V338A	30
L369A	20
F383A	0.03
R387L	0.002
W391A	1

TABLE 2.6 Kinetics of Selected Mutants in the Presence of KCl

Citrate synthase	K_{AcCoA} (μM)	K_{OAA} (μM)	Comments
WT	120 ± 20	26 ± 5	from Pereira *et al.*, (1994)
M131A	360 ± 175	159 ± 56	
L164A	68 ± 11	26 ± 4	
L259A	842 ± 180	11 ± 1	
L275A	1000*	380 ± 93	Sigmoid AcCoA curve
L369A	688 ± 138	19 ± 1	
F383A	2000 ± 100	47 ± 4	from Pereira *et al.*
R387L	ND	ND	from Pereira *et al.*

* Obtained from the sigmoid graph at $C_{50\%}$.

affinity of the mutant enzyme for its substrates, or the removal of residues important for substrate binding (as is the case for R387L, previously studied (Pereira *et al.*, 1994)). Table 2.6 lists citrate synthases and their K_M values for AcCoA and OAA from substrate saturation curves obtained in the presence of a fixed concentration of the second substrate. These values were obtained by fitting the hyperbolic curves to binding hyperbolas, with the exception of mutants which exhibited sigmoid saturation. Values obtained for WT and the mutants F383A and R387L are from previous work (Pereira *et al.*, 1994) and are included for comparison. L164A was the only mutant which showed enhanced CS activity, and this appears to be due to enhanced affinity for AcCoA.

The most dramatic example of a mutant with reduced activity was L275A, which showed sigmoid AcCoA saturation even in the presence of KCl; in the WT enzyme, KCl shifts the allosteric equilibrium toward the R state (active state) and converts the saturation curve for AcCoA from sigmoid to hyperbolic. L275A, then, appears to have an equilibrium that is shifted towards the allosteric T state, as was observed previously with the F383A mutant. This mutant also exhibited greatly reduced affinity for OAA, which may also affect the AcCoA binding; in CS, OAA binds first and increases the affinity for AcCoA.

The L259A and M131A mutants also showed very low activities. The pig residue homologous to L259 interacts with the active site residue F383 in the three-dimensional structure. The L259A mutant had a lowered affinity for AcCoA, consistent with the importance of this residue in maintaining the active site structure for optimal AcCoA binding. The OAA binding affinity for this mutant increased slightly. The M131 homolog in the pig structure, in

turn, interacts with the L259 residue described above, in what appears to be a hydrophobic network. Replacement of this residue with alanine also resulted in the reduction of affinity for AcCoA, also consistent with the idea that these residues are important as scaffolding for the active site.

The last of the mutants shown in Table 2.6, L369A, also has lowered affinity for AcCoA, but it is not obvious from the three-dimensional CS structure why this should be so. The L369 homolog in the pig structure is in the R helix, which is in close proximity to the S helix in which some of the active site residues reside. It is possible that the conformational changes that take place in the small domain during binding and catalysis involve this L369. Alternatively, adjustments of the L369A protein in response to the mutation may ultimately lead to a less efficient active site. The R helix is the core helix of the small domain where substantial rotations and shifts take place upon changing from the open to the closed conformations of CS (Lesk & Chothia, 1984).

Urea Denaturation of CS mutants and Data fitting

With the exception of W391A, mutant citrate synthases showed the same biphasic urea denaturation as the wild type enzyme. Their denaturation data are reported following this section. To compare the resultant denaturation curves to that of wild type CS, the data were fitted assuming a three state model with native (*N*), intermediate (*I*), and denatured (D) states, and two equilibrium constants K_1 and K_2:

$$N \underset{}{\overset{K_1}{\rightleftharpoons}} I \overset{K_2}{\rightleftharpoons} D$$

For any of the points in the denaturation curve, only the *N*, *I*, or *D* states are present at significant concentrations, and the sum of their fractions is 1,

$$f_N + f_I + f_D = 1 \quad [7]$$

where f_N, f_I, and f_D represent the fraction of protein present in the native, intermediate and denatured states, respectively.

Then K_1 and K_2 are defined as

$$K_1 = f_I / f_N \quad [8]$$

$$K_2 = f_D / f_I \quad [9]$$

The total measured ellipticity (or any other property), y_{obs}, at any given urea concentration may be defined as the sum of contributions of each of the states with each having its characteristic ellipticity y_N, y_I, and y_D

$$y_{obs} = y_N f_N + y_I f_I + y_D f_D \quad [10]$$

Using equations 7 and 8 and 9 and substituting into equation 10 yields the measured ellipticity in terms of equilibrium constants and characteristic ellipticities of the native, intermediate and denatured states:

$$y_{obs} = \frac{(y_N + K_1 y_I + K_1 K_2 y_D)}{(1 + K_1 + K_1 K_2)} \quad [11]$$

According to Pace (Pace *et al.*, 1988, Pace, 1986), the free energy change observed at each urea concentration, $\Delta G_{observed}$ is linearly related to the slope

of the transition, m, and the urea concentration, and extrapolation to 0 M urea yields the free energy change for the unfolding , $\Delta G_{unfolding}$

$$\Delta G_{observed} = \Delta G_{unfolding} - m[urea] \quad [12]$$

Since $\Delta G = -RTlnK$ (where R is the gas constant, 1.987 cal.deg^{-1}.mol^{-1}, T is the absolute temperature in degrees kelvin, and K is the equilibrium constant), K_1 and K_2 may be defined as

$$K_1 = e^{\left(\Delta G_1 - m_1[urea] / -RT\right)} \quad [13]$$

$$K_2 = e^{\left(\Delta G_2 - m_2[urea] / -RT\right)} \quad [14]$$

and substitution of equations 13 and 14 into equation 11 yields the final equation used for fitting denaturation data.

$$y_{obs} = \frac{\left(y_N + y_I\left(e^{\left(\frac{\Delta G_1 - m_1[urea]}{-RT}\right)}\right) + y_D\left(e^{\left(\frac{\Delta G_1 - m_1[urea]}{-RT}\right)}\right)\left(e^{\left(\frac{\Delta G_2 - m_2[urea]}{-RT}\right)}\right)\right)}{\left(1 + \left(e^{\left(\frac{\Delta G_1 - m_1[urea]}{-RT}\right)}\right) + \left(e^{\left(\frac{\Delta G_1 - m_1[urea]}{-RT}\right)}\right)\left(e^{\left(\frac{\Delta G_2 - m_2[urea]}{-RT}\right)}\right)\right)} \quad [15]$$

Since seven parameters are involved in fitting a denaturation curve, a pool of data from denaturation curves of mutant and wild type citrate synthases was fitted with the equation in order to obtain best values of three parameters, the normalized ellipticities characteristic of native, intermediate

and denatured states (y_N, y_I, and y_D, respectively). Those values were then held constant in subsequent fittings of individual data sets to obtain the free energy of unfolding and slope parameter for each transition (ΔG_1,ΔG_2 and m_1, m_2, respectively). The fitted pool of denaturation data is shown in Figure 2.21 with the resultant values in Table 2.7. The fitted parameters for the pooled data indicate, on average, that unfolding in Transition I accounts for 67% of the structure, while Transition II accounts for the remaining 33%.

While the irreversibility of the unfolding of CS and the presence of aggregates puts into question the validity of this kind of analysis, the data fitting serves as a method for comparing the mutant denaturation curves to the wild type denaturation, and at the very least, it allows the calculation of the transition midpoint (the concentration of urea at middle of each transition, $[urea]_{1/2}$). At the midpoint of the transition, the ΔG_{obs}=0, so equation 12 becomes

$$0 = \Delta G_{unfolding} - m[urea]_{1/2} \qquad [16]$$

and the concentration of urea at the midpoint, can be calculated,

$$[urea]_{1/2} = {}^{\Delta G_{unfolding}}\!/_{m} \qquad [17]$$

The free energies of unfolding, slopes, and midpoints associated with each transition, for all mutant and wild type citrate synthases, are summarized in Tables 2.8 and 2.9.

Large errors are associated with some of the ΔG_2 values. During data fitting, large variations in the assigned slopes were noted, which usually led to large changes in the ΔG_2 values obtained. The uncertainty may be due to

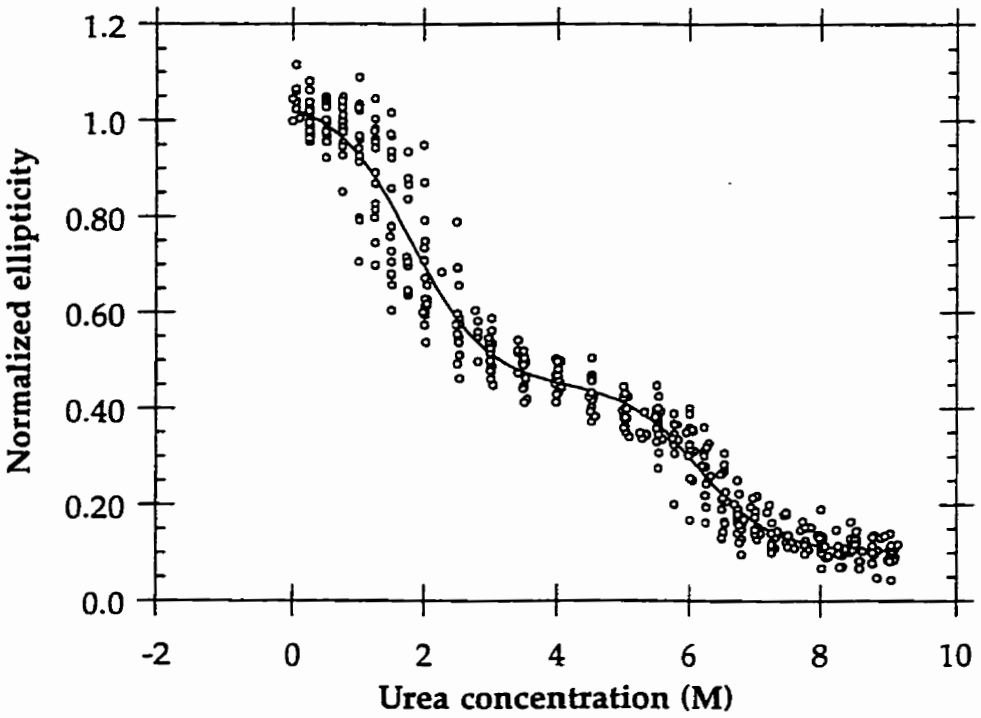


Figure 2.21 Data fitting to obtain values of yN, yI, and yD, the normalized ellipticity characteristic of the Native, Intermediate, and Denatured states. The results of the fit to equation 15 are in Table 2.7.

TABLE 2.7 Values obtained from the fit of the data in Figure 2.21 to equation 15. yN, yI, and yD were constrained to those values for subsequent data fits.

Parameter	Fitted value
yN	1.05 ± 0.02
ΔG_1	1.83±0.19 kcal.mol^{-1}
m_1	-1.01±0.10 kcal.mol^{-1}M^{-1}
yI	0.45 ± 0.02
yD	0.10 ± 0.01
ΔG_2	6.46 ± 0.90 kcal.mol^{-1}
m_2	-1.05±0.14 kcal.mol^{-1}M^{-1}

two reasons. First, Transition II is short, and only a limited number of points in a small concentration range are available. Second, the length of the extrapolation for the determination of ΔG_2 is long. Both points reduce the accuracy with which a slope is assigned, leading to a ΔG_2 value with a large uncertainty. Perhaps an additional factor is the contribution of scatter; the data at ellipticities of lower intensity would have a larger error associated with them, if the error in CD measurements is a constant value rather than a percentage of the ellipticity value.

TFBA-modified CS

Cysteine 206, the residue modified by TFBA, is in the large domain of the model of CS. It resides in a region of the sequence where there exists very little homology between the allosteric citrate synthases and the non allosteric class, and this residue was identified as a marker for the NADH binding site (Donald *et al.*, 1991). As discussed in Chapter 3, mass spectrometry was used to show that this modified CS is dimeric in the absence of KCl (see Figure 3.7 in Chapter 3). Non-denaturing gel electrophoresis also showed dimer only (not shown). Figure 2.22 shows the fitted denaturation curves for WT CS and TFBA-modified CS. The Transition I is destabilized, probably because of loss of hexamer interactions in this protein.

F121A CS

F121 is a large domain residue in Helix F of the CS model (Figure 2.23). The residue is highly conserved in the prokaryotic sequences, and is only replaced with hydrophobic residues in some prokaryotes and most eukaryotes. Data fitting of the denaturation curve of this mutant (Figure 2.24) indicates that Transition I is stabilized (by 0.28 kcal/mol) and the second destabilized. Transition midpoints, however, occur at lower concentrations;

TABLE 2.8 Summary of free energies of unfolding and slopes of the two unfolding transitions of wild type, mutant, and TFBA modified citrate synthases obtained by data fitting using equation 15 with y_N, y_I, and y_D held fixed at the values shown in Table 2.7.

Protein	ΔG_1 *kcal.mol*$^{-1}$	ΔG_2 *kcal.mol*$^{-1}$	m_1 *kcal.mol*$^{-1}$·*M*$^{-1}$	m_2 *kcal.mol*$^{-1}$·*M*$^{-1}$
WT CS	1.80±0.11	10.5±1.65	-1.05±0.06	-1.61±0.25
TFBA	1.33 ± 0.10	7.45 ± 0.76	-0.99 ± 0.07	-1.18 ± 0.12
F121A	2.08±0.18	5.76±0.86	-1.36±0.11	-0.99±0.14
M131A	2.62±0.27	8.47±1.65	-1.24±0.13	-1.40±0.27
L141A	1.62±0.23	8.65±2.10	-1.52±0.20	-1.43±0.34
L164A	1.25±0.11	7.89±1.17	-0.92±0.07	-1.25±0.18
L259A	2.59±0.23	8.15±1.26	-1.31±0.12	-1.34±0.21
W260A	2.74±0.32	5.51±0.66	-1.75±0.19	-0.88±0.11
M274A	3.81±0.48	11.4±2.65	-2.15±0.28	-1.84±0.43
L275A	1.93±0.17	5.25±0.58	-1.49±0.13	-0.85±0.09
L326A	1.52±0.08	6.19±0.53	-0.99±0.05	-0.98±0.08
L336A	3.13±0.39	5.21±0.91	-1.39±0.17	-0.91±0.15
V338A	3.85±0.43	10.2±1.92	-1.63±0.18	-1.65±0.31
L369A	2.00±0.14	9.07±1.72	-0.95±0.07	-1.47±0.28
F383A	3.72±0.63	10.7±3.47	-2.01±0.35	-1.93±0.60
R387L	4.23±0.51	4.05±0.45	-1.68±0.20	-0.64±0.07
W391A	1.20±0.11	3.75±0.62	-0.75±0.08	-0.73±0.11

TABLE 2.9 Summary of transition midpoints for urea denaturation curves of mutant, chemically modified and wild type citrate synthases.

Citrate Synthase	*midpoint of first transition (M)*	*midpoint of second transition (M)*
WTCS	1.71	6.49
TFBA CS	1.34	6.31
F121A	1.53	5.82
M131A	2.11	6.05
L141A	1.07	6.05
L164A	1.36	6.31
L259A	1.98	6.08
W260A	1.57	6.26
M274A	1.77	6.18
L275A	1.30	6.18
L326A	1.54	6.32
L336A	2.25	5.73
V338A	2.36	6.20
L369A	2.11	6.17
F383A	1.93	5.56
R387L	2.52	6.33
W391A	1.60	5.14

Values calculated from parameters in Table 2.8 and equation 17.

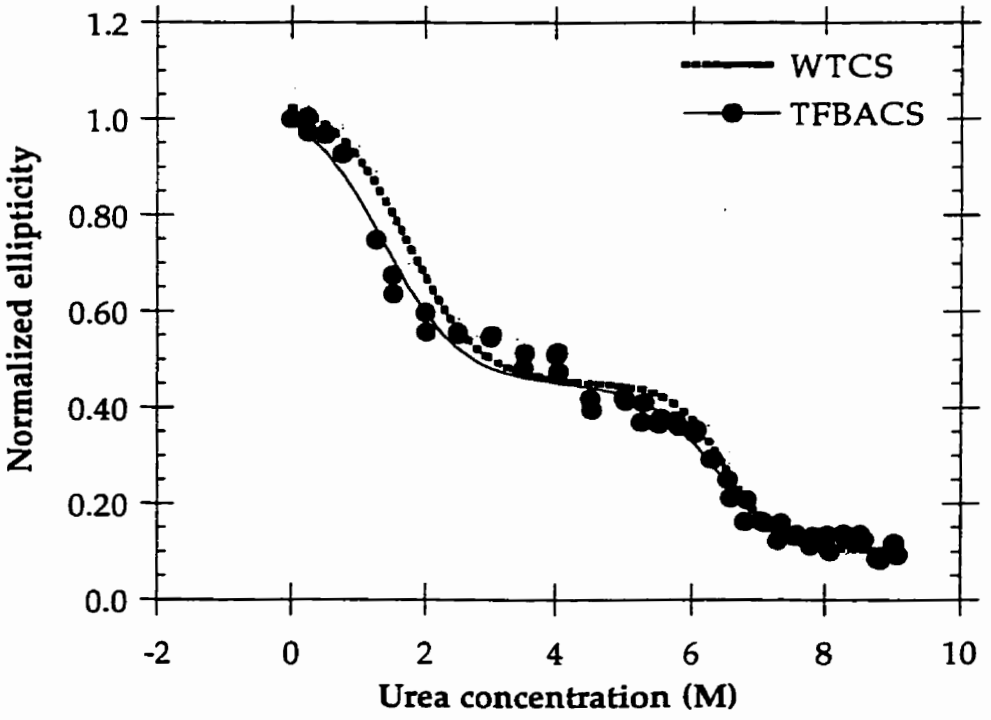


Figure 2.22 Urea denaturation of TFBA CS and WT CS monitored by CD. Lines are fits to the data using equation 10, as described in the text.

```
                    121        131       141                    164
              < Helix F  >     <  Helix G   >  <Helix H>     < Helix I...
CISY_ECOLI TMIHEQITRLFHAFRRDSHPMAVMCGITGALAAFYHDSLDVNNPRHREIAAFRLLSKMPT
CISY_PSEAE TMVHEQLKTFFNGFRRDAHPMAVMCGVIGALSAFYHDSLDITNPKHRQVSAHRLIAKMPT
CISY_ACIAN TMVHDQVSRFFNGFRRDAHPMAIMVGVVGALSAFYHNNLDIEDINHREITAIRLIAKIPT
CISY_ACEAC TLLHEQIRNFFNGFRRDAHPMAILCGTVGALSAFYPDANDIAIPANRDLAAMRLIAKIPT

CISY_PIG   AALPSHVVTMLDNFPTNLHPMSQLSAAITALNSESNFARAYAEGIHRTKDCMDLIAKLPC
CISY_CHICK AALPSHVVTMLDNFPTNLHPMSQLSAAITALNSESNFARAYAEGILRTKSAMDLIAKLPC
CISY_YEAST SEIPEHVIQLLDSLPKDLHPMAQFSIAVTALESESKFAKAYAQGVSKKEDSLDLLGKLPV
CISY_CITMA ATVPDYVYKAIDALPVSAHPMTQFASGVMALQVQSEFQEAYEKGIHKSKDSLNLIARVPV
```

Figure 2.23 Partial sequence alignment of four prokaryotic (top group) and four eukaryotic (bottom group) CS sequences, showing the locations of the large domain mutations. The secondary structure indicated is from A. Chu's model of CS that is based on the pig structure (Duckworth *et al.*, 1987). The residues chosen for mutation are indicated by their number in the *E. coli* sequence and in bold lettering. For the full sequence alignment, see Appendix I.

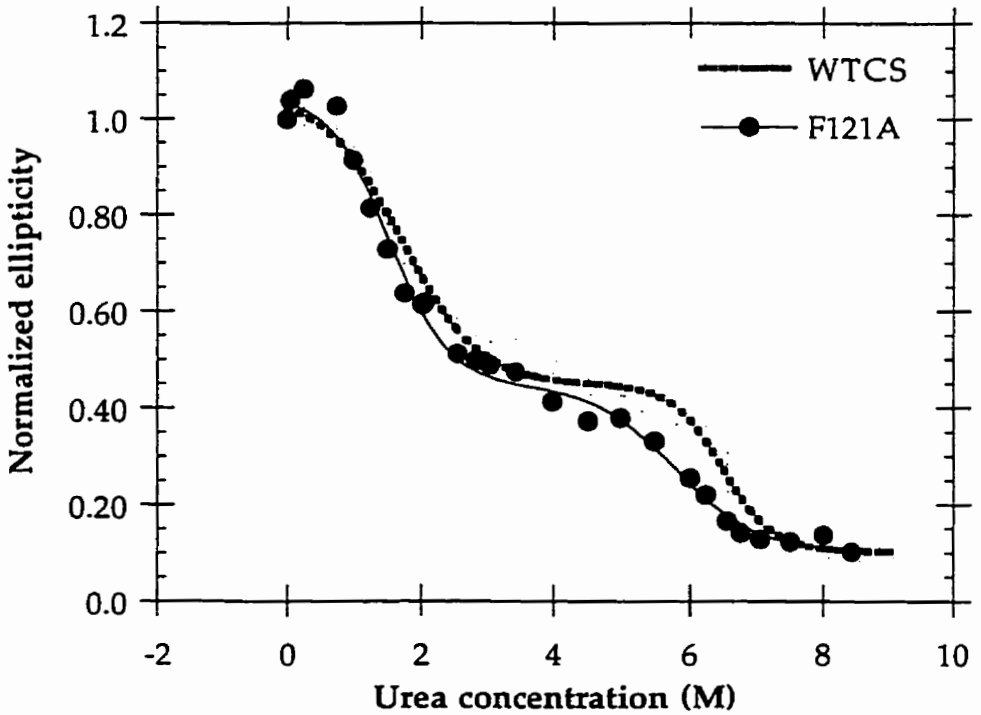


Figure 2.24 Urea denaturation and data fit for F121A CS compared to WT CS.

the apparent increase in $\Delta G1$ can be attributed to the change in the slope of this transition.

M131A CS

M131 is highly conserved large domain residue in the N-terminus of the G helix of the CS model (see alignment in Figure 2.23). The equivalent pig residue, M138, is involved in hydrophobic interaction with L269, a residue which is in contact with the active site residue F397. The greatly reduced activity of the M131A mutant may indicate that the methionine is part of a hydrophobic network which maintains the integrity of the active site. Reduced affinity for both substrates is observed with this mutant, with a greater effect on the OAA binding. Denaturation of this mutant (Figure 2.25) showed that the first transition is stabilized and the second transition destabilized, a trend common among other mutants with reduced activity.

L141A CS

L141A is also a large domain mutant in the C terminus of the G helix (Figure 2.23). The residue is highly conserved, only being replaced with M in some species. In contrast with M131A, both transitions in the denaturation curve of this mutant appear to be destabilized (Figure 2.26), and the effect on enzyme activity was not large.

L164A CS

This mutant is the only one which exhibited elevated enzyme activity; the limited kinetic measurements revealed that AcCoA binding is twofold better than to WT. Denaturation showed that this mutant is destabilized relative to WT (Figure 2.27). This residue, located in the core of the large domain (Helix I, see Figure 2.23), is highly conserved in most species, being replaced only by other hydrophobic residues (M, I). This mutant

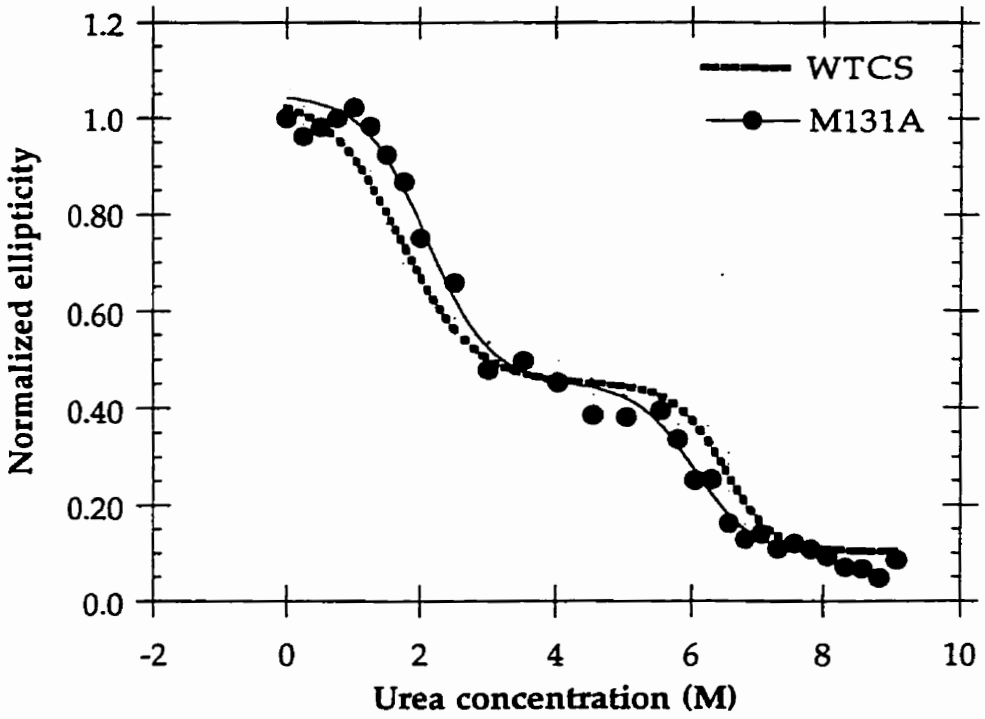


Figure 2.25 Urea denaturation and data fit for M131A CS compared to WT CS.

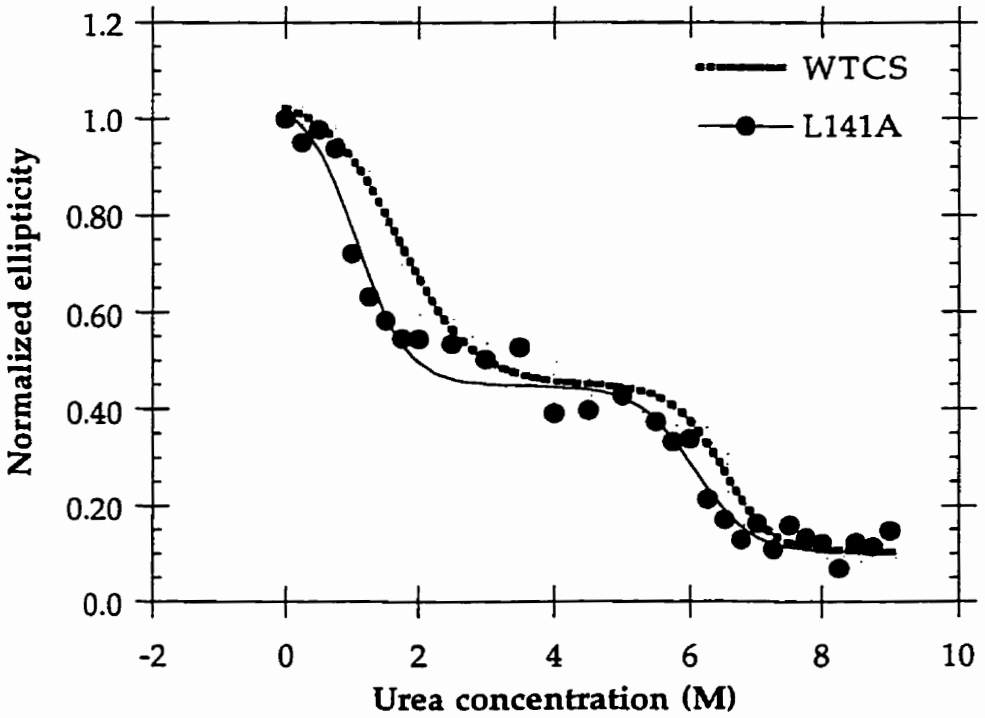


Figure 2.26 Urea denaturation and data fit for L141A CS compared to WT CS.

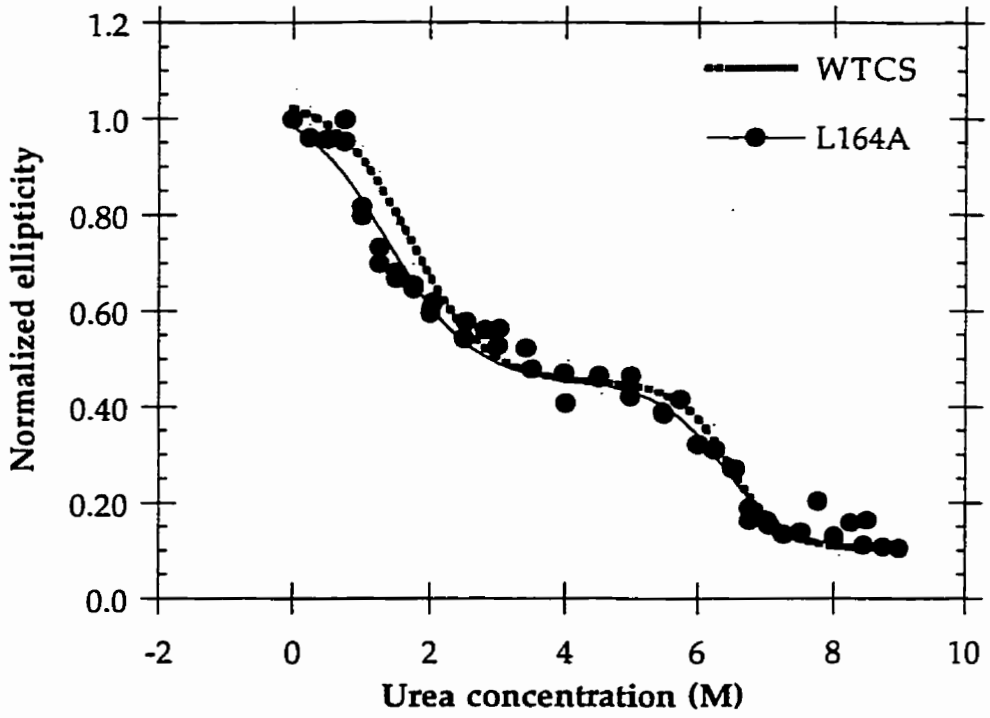


Figure 2.27 Urea denaturation and data fit for L164A CS compared to WT CS.

```
                    259/260        274/275
                  ...Helix M >    <   Helix N   >     < Helix O >
CISY_ECOLI     CIAAGIASLWGPAHGGANEAALKMLEEISSVKHIPEFVRRAKDKNDSFRL
CISY_PSEAE     CIASGIAALWGPAHGGANEAVLRMLDEIGDVSNIDKFVEKAKDKNDPFKL
CISY_ACIAN     CISAGISALWGPAHGGANEAVLKMLDEIGSVENVAEFM..EKVKRKEVKL
CISY_ACEAC     CIAAGIAALWGPAHGGANEAVLKMLARIGKKENIPAFIAQVKDKNSGVKL

CISY_PIG       SFAAAMNGLAGPLHGLANQEVLVWLTQLQ..KEVGKDVSDEKTLNSGRVV
CISY_CHICK     SFAAAMNGLAGPLHGLANQEVLGWLAQLQKAXXXAGADASLRDLNSGRVV
CISY_YEAST     SLAAGLNGLAGPLHGRANQEVLEWLFKLRSKETIEKYLW..DTLNAGRVV
CISY_CITMA     SFLAALNGLAGPLHGLANQEVLLWIKSVVDTEQLKDYVW..KTLNSGKVV
```

Figure 2.28 Partial sequence alignment of four prokaryotic (top group) and four eukaryotic (bottom group) CS sequences, showing the locations of 4 mutants in the M and N helices. The secondary structure indicated is from A. Chu's model of CS that is based on the pig structure. The residues chosen for mutation are indicated by their number in the *E. coli* sequence and in bold lettering. For the full sequence alignment see Appendix II.

is not subject to allosteric inhibition by NADH. This may arise somehow from the proximity to a loop between Helices J and K has been proposed to contain the elements of NADH inhibition/binding and/or allostery since the two classes of CS show low homology in sequence in this region.

L259A CS

L259 is conserved in all known CSs except for CISY_BACSU (a third CS in *Bacillus subtilis*, where methionine is substituted (Figure 2.28 and Appendix I). Its high degree of conservation suggests that it may be important for the function of the enzyme. The equivalent residue in the pig CS structure (L269) makes a hydrophobic contact with the active site residue F397; F397 interacts with the acetyl group of AcCoA. Therefore the role of this residue may be maintenance of the structure of the active site. The observed reduced activity of L259A would suggest that this is the case; the affinity for AcCoA was greatly reduced in this mutant. Denaturation shows that this mutant is modestly stabilized in the first transition and destabilized in the second (Figure 2.29).

W260A CS

W260 is in the M helix close to the interface between the two structural domains, and is well conserved in the prokaryotic sequences (Figure 2.28). The activity of W260A was reduced relative to WT CS, but not to the extent observed with the mutation of the neighboring L259. Again, the first transition appears to be stabilized and the second transition destabilized (Figure 2.30). A striking difference, however, is the large change in the slope of the first transition.

CS fluorescence is dominated by tryptophans, as shown in Figure 2.9 for WT CS. The protein contains three tryptophans, W260, W391, and W395, but

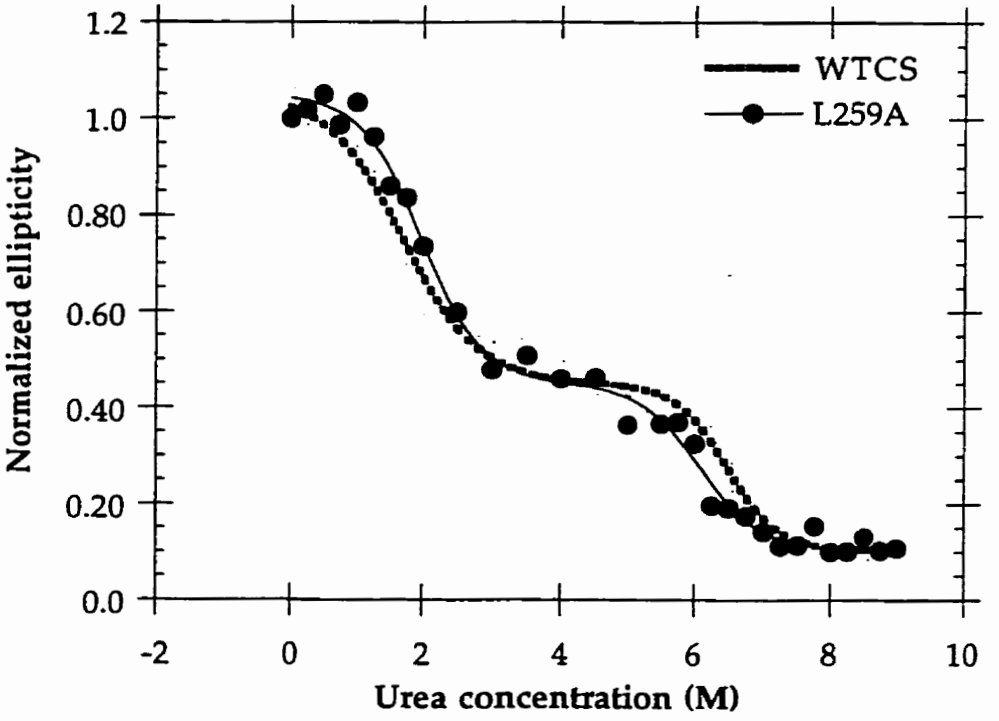


Figure 2.29 Urea denaturation and data fit for L259A CS compared to WT CS.

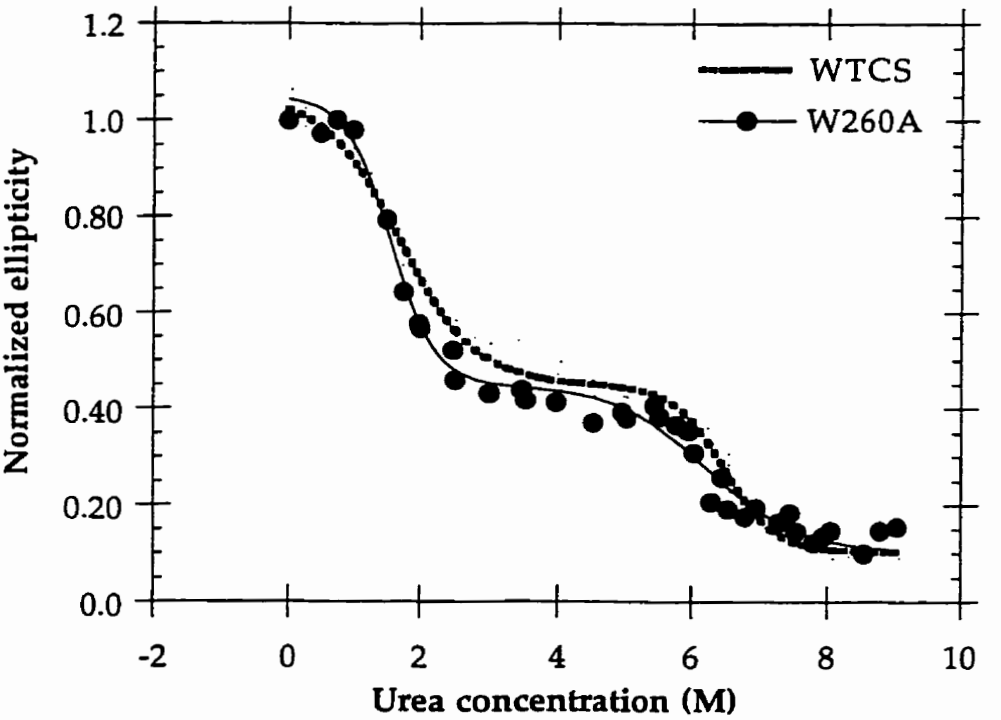


Figure 2.30 Urea denaturation and data fit for W260A CS compared to WT CS.

the contributions of each of those to the fluorescence of the protein was unknown.

The emission spectrum of W260A was observed to have a different shape than the wild type spectrum (Figure 2.31); the shoulder peak in the wild type spectrum is missing for W260A and the emission maximum is shifted by 5 nm to 320 nm. These observations suggest that the residue W260 is the source of the shoulder peak in the wild type spectrum and that it is solvent exposed. If this was the case, then the remaining tryptophan residues, W391 and W395 probably act as a single fluorescence reporter group for CS, since they are located in the same part of the molecule.

Denaturation of W260A followed by fluorescence is also biphasic (Figure 2.32). However the amplitudes of both transitions were lower than the wild type values by about 0.2 (arbitrary fluorescence units). This suggests that fluorescence emission of W260 in the WT CS contributes to both transitions equally.

M274A CS

M274 is a small domain residue in the N helix, and is conserved as a methionine in most prokaryotic CS sequences, but is replaced with tryptophan in the eukaryotic ones (Figure 2.28). Glutamine and aspartic acid replace this residue in archaebacteria and one procaryotic sequence (see Appendix I). While the first transition midpoint seems to be unchanged from the WT value, the slope is steeper and the resultant free energy is twice that for WT (Figure 2.33). The second transition is destabilized. The proximity of this residue to others important for AcCoA binding may explain the reduced activity obtained; the equivalent residue in the pig CS, W284, makes contacts with residues involved in CoA binding.

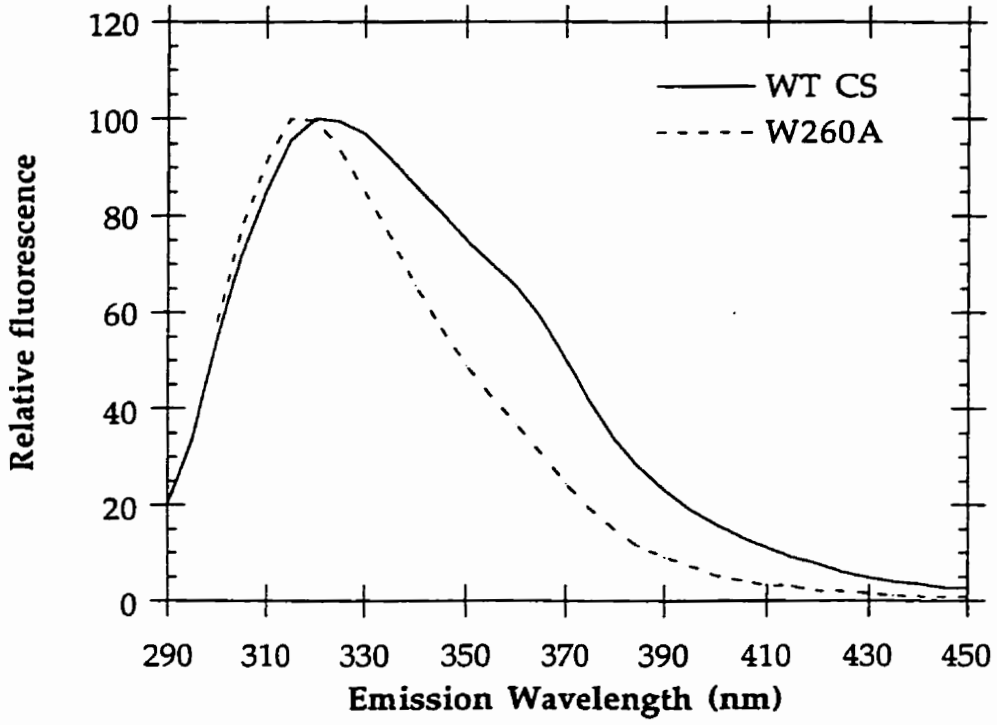


Figure 2.31 Emission spectra of WT and W260A. Excitation was at 280 nm and protein concentration was 2.1 μM.

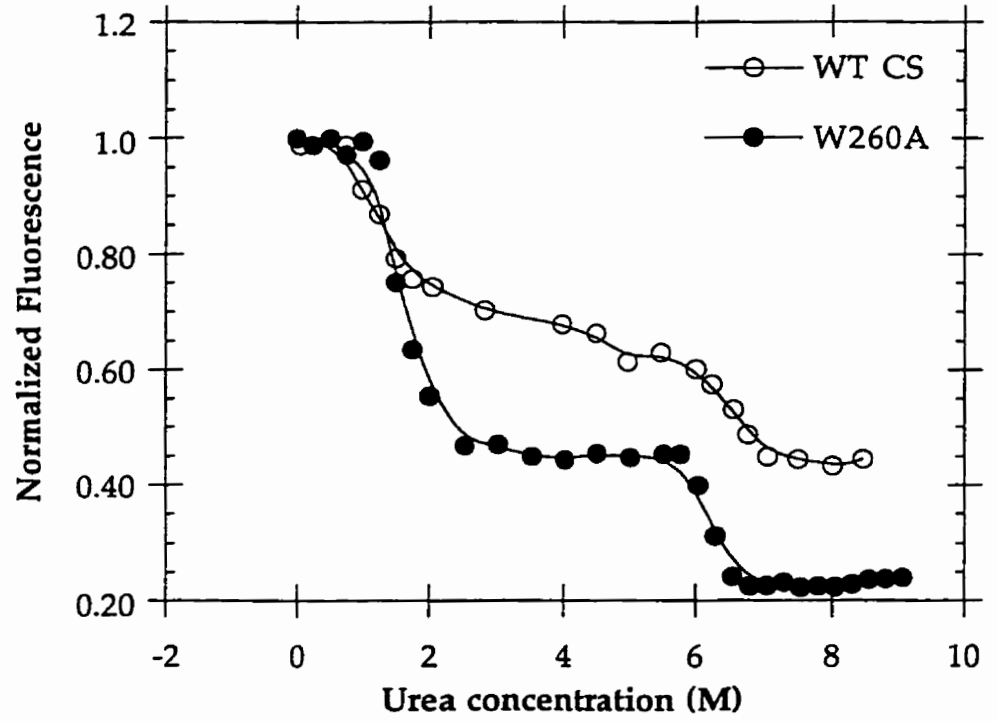


Figure 2.32 Urea denaturation of WT and W260A monitored by fluorescence. The lines are drawn to show the trend of the data.

L275A CS

Unlike M274A, the L275A mutant exhibits a less stable first transition (Figure 2.34). L275 is very well conserved, with only a few sequences showing variation (Figure 2.28); two sequences contain a glutamine in this position, and in both cases they are products of a "second CS gene" in *E. coli* and *Salmonella typhimurium* (Appendix I). The pig equivalent, L285, contacts residues which are important for CoA binding; kinetic measurements show that this mutant is defective in AcCoA binding, exhibiting sigmoid saturation even in the presence of KCl (similar to that observed with the F383A mutant (Pereira *et al.*, 1994)). Oxaloacetate binding affinity was also reduced for L275A.

L326A CS

L326 is in Helix P of the small domain where it is well conserved (Figure 2.35). In a few CSs, the equivalent residue is C, S, or P instead of a hydrophobic residue. The effect of the L326A mutation on the stability was modest; only a small decrease of the stability in the first transition was observed along with the usual decrease in the second transition (Figure 2.36). The enzyme activity of this mutant was normal.

L336A CS

L336 is in the N-terminus of Helix Q in the small domain of CS (Figure 2.35). The effect of replacing this residue with an alanine on the stability is dramatic, with a large increase in the stability of the first transition (+1.33 kcal.mol^{-1}) and a decrease in the second transition (Figure 2.37). The specific activity of this mutant, while somewhat reduced, was unremarkable.

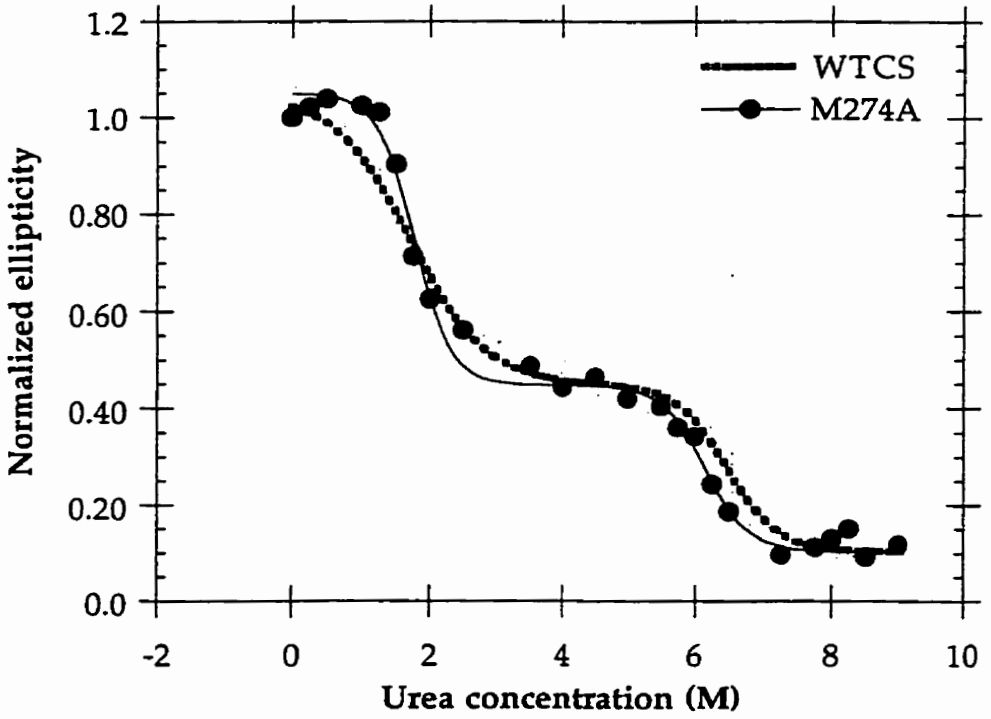


Figure 2.33 Urea denaturation and data fit for M274A CS compared to WT CS.

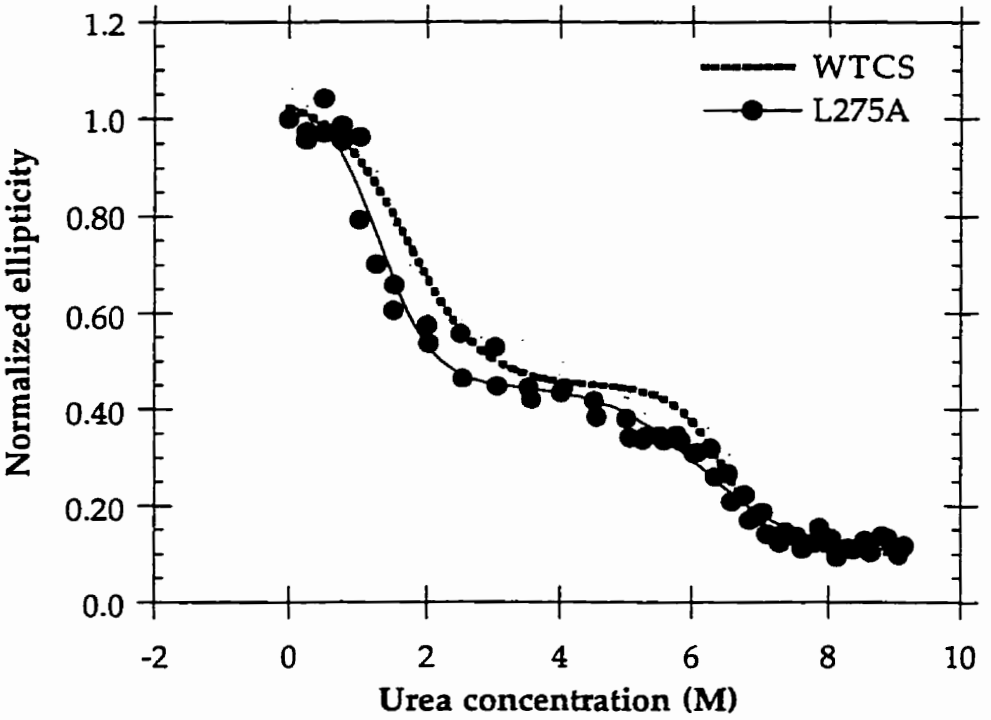


Figure 2.34 Urea denaturation and data fit for L275A CS compared to WT CS.

```
                  326      336/338                          369
            ....Helix P >  <     Helix Q    >       <   Helix R  >
CISY_ECOLI  TCHEVLKELGTKDDLLEVAMELENIALNDPYFIEKKLYPNVDFYSGIILK
CISY_PSEAE  TCDEVLQELGINDPQLELAMKLEEIARHDPYFVERNLYPNVDFYSGIILK
CISY_ACIAN  TCDEVLEALGINDPQLALAMELERIALNDPYFVERKLYPNVDFYSGIILK
CISY_ACEAC  TCHEVLTELGIKDDLLDLAVELEKIALSDDYFVQRKLYPNVDFYSGIILK

CISY_PIG    FALKHLPHDPMFKLVAQLYKIVPNVLLEQG..KAKNPWPNVDAHSGVLLQ
CISY_CHICK  FALKHLPGDPMFKLVAQLYKIVPNVLLEQG..AAANPWPNVDAHSGVLLQ
CISY_YEAST  FALKHFPDYE....LFKLVSTIYEVAVLTKHGKTKNPWPNVDSHSGVLLQ
CISY_CITMA  FALKHLPD....DPLFQLVSKLYEVPILTKLGKVKNPWPNVDAHSGVLLN
```

Figure 2.35 Partial sequence alignment of four prokaryotic (top group) and four eukaryotic (bottom group) CS sequences, showing the locations of 4 mutants in the P, Q, and R helices. The secondary structure indicated is from the model of CS that is based on the pig structure. The residues chosen for mutation are indicated by their number in the *E. coli* sequence and in bold lettering. For the full sequence alignment see Appendix I.

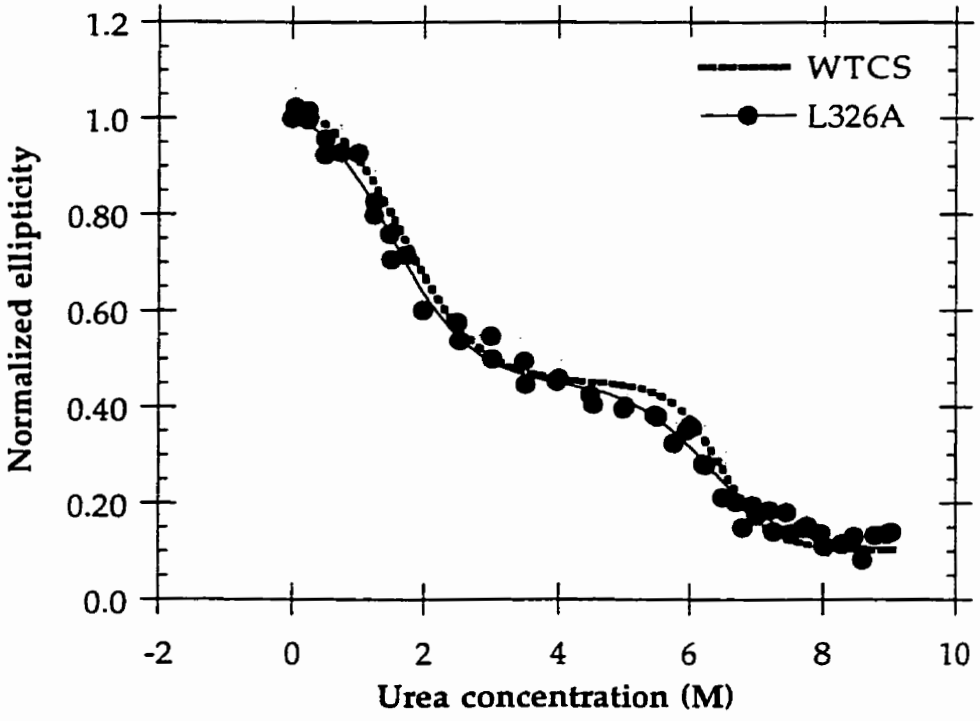


Figure 2.36 Urea denaturation and data fit for L326A CS compared to WT CS.

V338A CS

V338 is in Helix Q of the small domain (Figure 2.35). Mutation of this highly conserved residue resulted in a dramatic increase in the stability of the first transition (+2.05 kcal.mol^{-1}), as evidenced by Figure 2.38. As observed for the L336A mutant, the activity of the V338A mutant was reduced but not to a very large extent.

L369A CS

L369 resides in Helix R of the small domain close to the intersubunit interface, and is well conserved (Figure 2.35). The L369A mutant showed a modest increase and decrease in the stabilities of the first and second transitions, respectively (Figure 2.39). The activity of this mutant was reduced to about 30% of that for WT. Kinetics measurements in the presence of KCl show that the AcCoA affinity was reduced by about 6-fold while that for OAA was unchanged.

F383A CS

F383 is the 100% conserved, active site residue in Helix S which interacts with the acetyl moiety of the AcCoA (Figure 2.40 and Figure 1.11 in Chapter 1). This mutant was constructed and studied by D. Periera, and shows sigmoid AcCoA saturation kinetics in the presence of KCl. The mutant has a dramatically different stability profile compared to WT: the first transition was stabilized by about 2 kcal.mol^{-1} (Figure 2.41). As with the other mutants, the second transition occurs at lower urea concentrations.

R387L CS

R387 is also an active site, Helix S residue (Figure 2.40 and Figure 1.11 in Chapter 1) which is involved in hydrogen bonding with OAA. Its mutation to leucine resulted in a mutant which is essentially inactive (Pereira

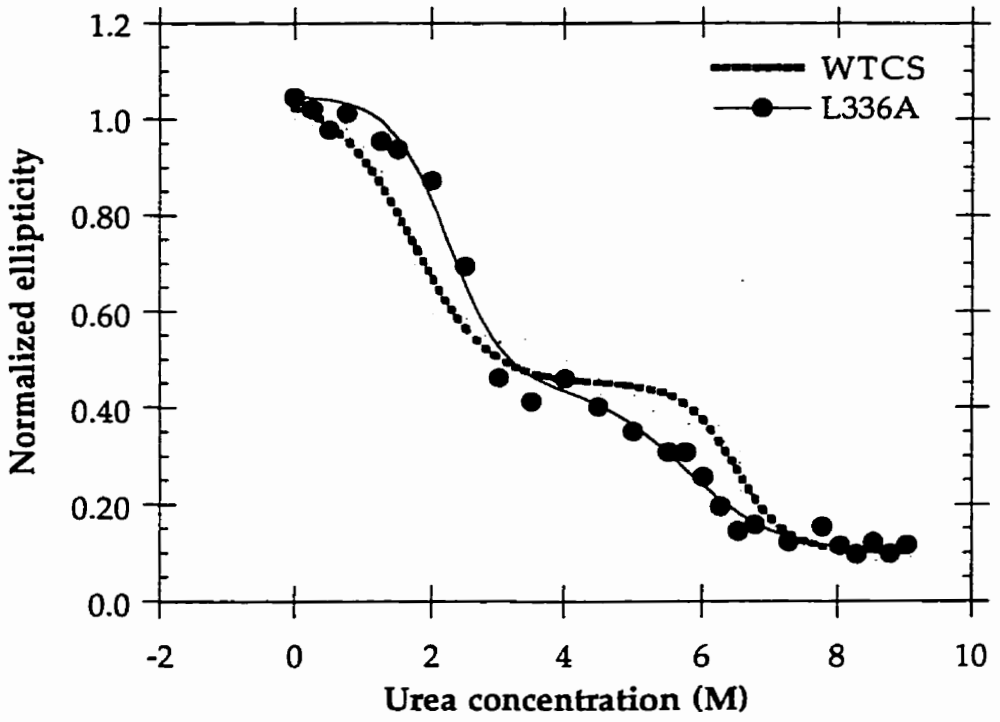


Figure 2.37 Urea denaturation and data fit for L336A CS compared to WT CS.

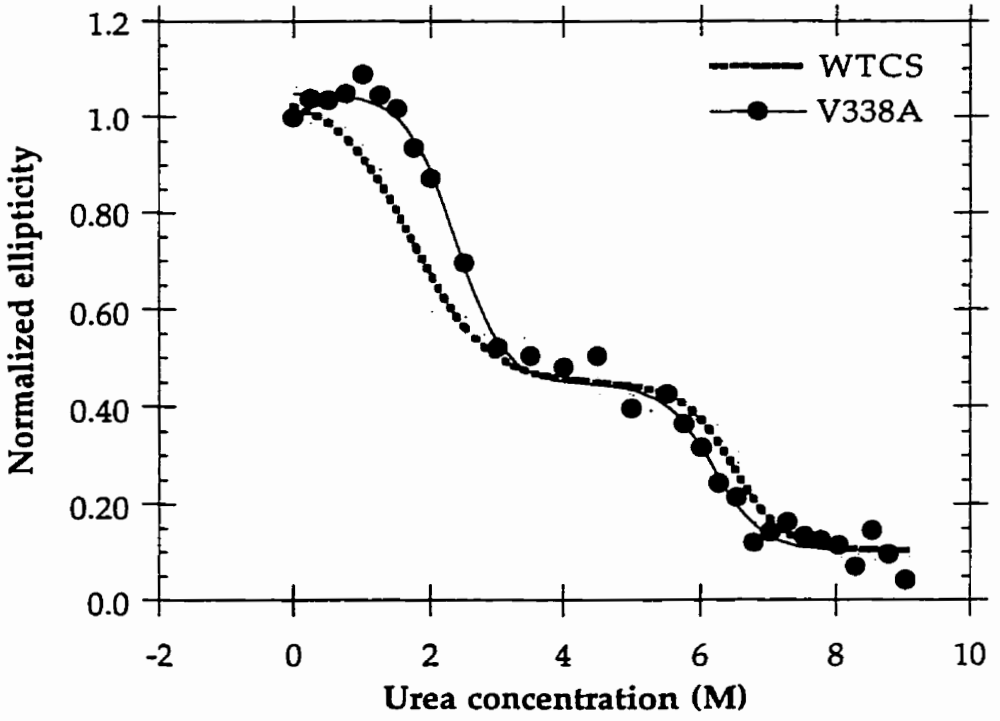


Figure 2.38 Urea denaturation and data fit for V338A CS compared to WT CS.

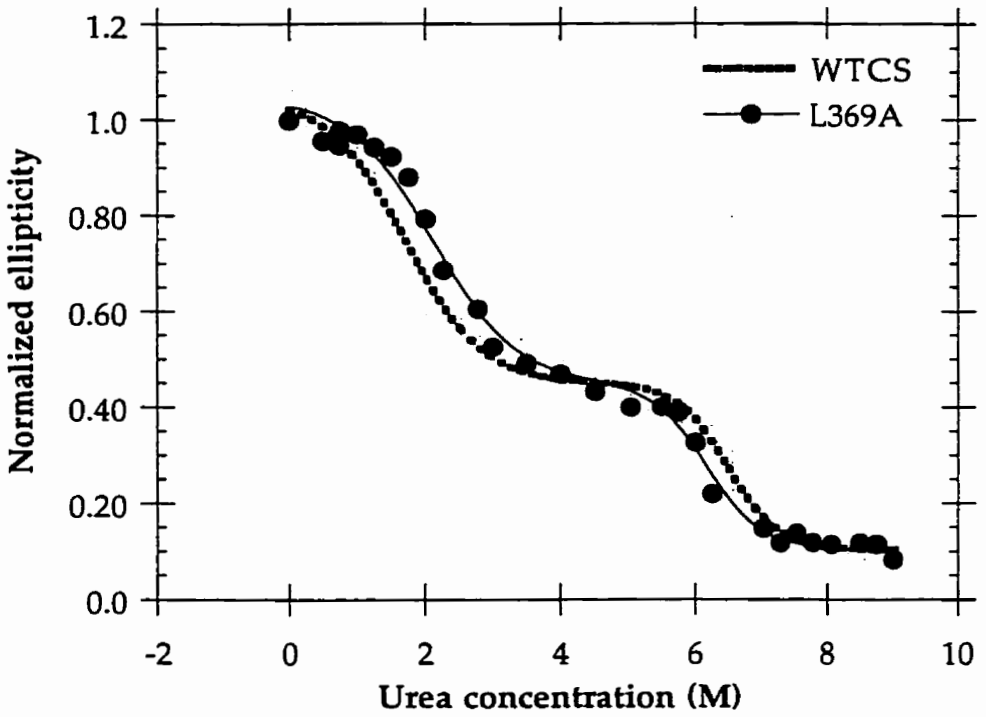


Figure 2.39 Urea denaturation and data fit for L369A CS compared to WT CS.

```
                                383  387  391   395
                             <       Helix S      >
CISY_ECOLI      AMGIPSSMFTVIFAMARTVGWIAHWSEMHS
CISY_PSEAE      AIGIPTSMFTVIFALARTVGWISHWQEMLS
CISY_ACIAN      AIGIPTEMFTVIFALARTVGWISHWLEMHS
CISY_ACEAC      AMGIPTSMFTVLFAVARTTGWVSQWKEMIE

CISY_PIG        YYGMTENYYTVLFGVSRALGVLAQLIWSRA
CISY_CHICK      YYGMTENYYTVLFGVSRALGVLAQLIWSRA
CISY_YEAST      YYGLTESFYTVLFGVARAIGVLPQLIIDRA
CISY_CITMA      HFGLAERYYTVLFGVSRSLGICSQLIWDRA
```

Figure 2.40 Partial sequence alignment of four prokaryotic (top group) and four eukaryotic (bottom group) CS sequences, showing the locations of 4 mutants in the S helix The secondary structure indicated is from the model of CS that is based on the pig structure. The residues chosen for mutation are indicated by their number in the *E. coli* sequence and in bold lettering. For the full sequence alignment see Appendix I.

et al., 1994). The effect on the stability is profound (Figure 2.42): the first transition has been stabilized by about 2.4 kcal.mol^{-1} and the second transition exhibits very low cooperativity. The stabilization might be expected if the burial of a charged group into an active site with other positively charged residues is unfavorable and thus is destabilizing in the WT protein.

W391A CS

W391 is in Helix S close to the active site and the intersubunit interface (Figure 2.40). W391 is highly conserved as a tryptophan among most CS sequences, and where it is replaced, the residue is another aromatic residue (Y) or a leucine (see Appendix I). The large effect of this mutation on the stability and activity was unexpected (Figure 2.43). A helix wheel projection shows that this residue is on the same face as the active site residues F383 and R387 (not shown). Some kinetic measurements indicate that AcCoA binding was greatly reduced and perhaps sigmoidal (data are not shown). The denaturation was not cooperative, and occurred gradually throughout the urea concentration range without an apparent equilibrium intermediate. Numerous efforts to improve the quality of the denaturation data failed.

W395A CS

The W395A mutant, which also resides in the S helix and is conserved (Figure 2.40), was not successfully purified. Although SDS PAGE showed that some W395A CS was present after ion exchange and size exclusion chromatographies, the amount of contaminating protein was very large. It is possible that this mutant does not fold properly *in vivo*, and that it was susceptible to proteolysis perhaps due to its instability. The behaviour of the mutant W391A would indicate that this region of CS is essential for its

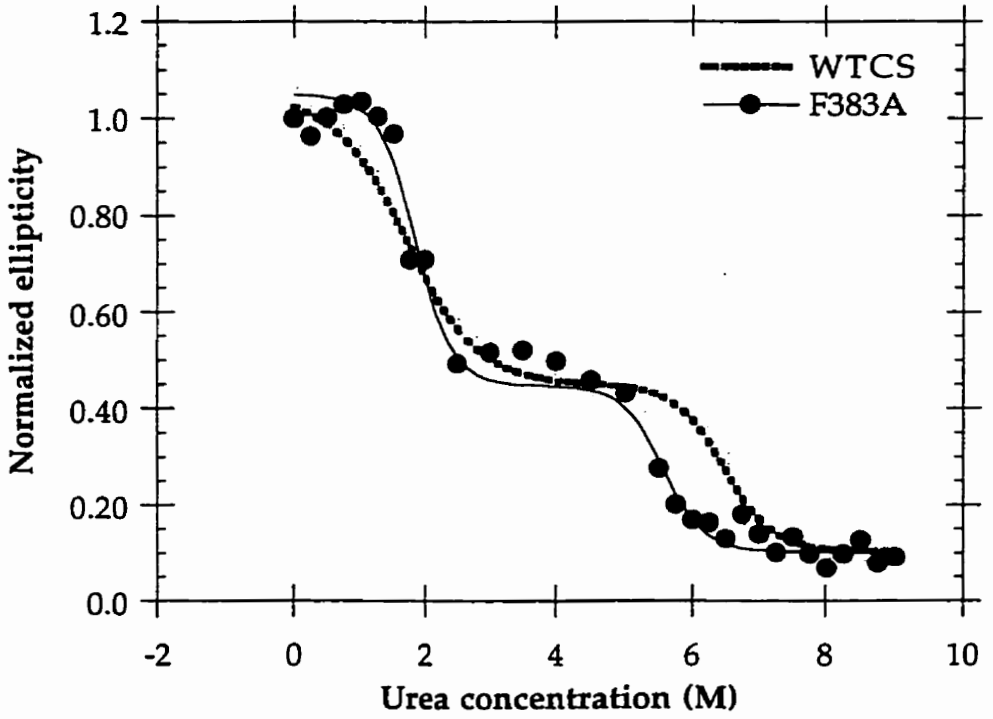


Figure 2.41 Urea denaturation and data fit for F383A CS compared to WT CS.

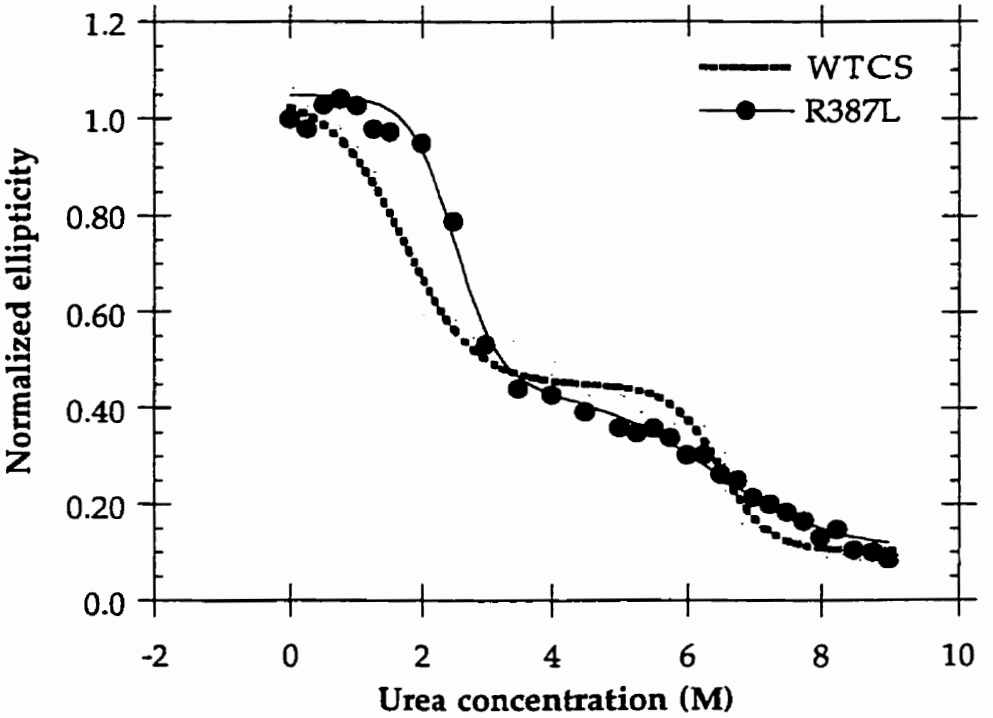


Figure 2.42 Urea denaturation and data fit for R387L CS compared to WT CS.

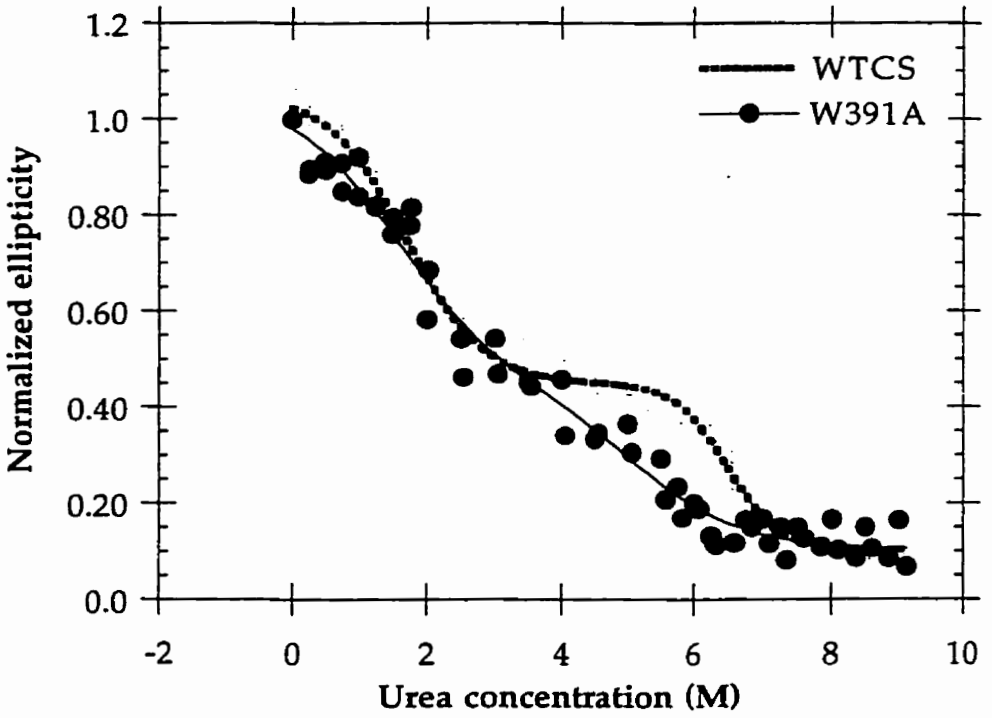


Figure 2.43 Urea denaturation and data fit for W391A CS compared to WT CS.

folding and stability, and so the lack of success in purifying the W395A mutant is not surprising.

Effect of ligands on the unfolding of CS

Citrate, NADH, KCl, and AcCoA were added to the unfolding reaction of WT CS. The denaturation curves were analyzed by the method described for mutant curves by data fitting. Figures 2.44 to 2.48 describe the denaturation curves and Table 2.10 summarizes the results of the data fitting. In all cases, the Transition I had higher $[urea]_{1/2}$ values than wild type protein in the absence of ligand. Transition II had values of $[urea]_{1/2}$ that were reduced, except for the denaturation curves in the presence of KCl and NADH; those exhibited no change in the midpoint of the transition. Both KCl and NADH induce hexamerization, and thus, would be expected to stabilize the protein. KCl may also stabilize salt bridges and NADH binding in its site might contribute to stability by hydrogen bonding to groups that may otherwise be unsatisfied. Ligands in general are found to stabilize the folded conformation of proteins (Creighton, 1993).

Citrate and CoA (active site ligands) affected the denaturation profile in a manner similar to that exhibited by mutants with lowered activity; the Transition I was stabilized apparently at the expense of Transition II.

Patterns and Trends of the Denaturation Data

The bar graphs in Figure 2.49 and 2.50 illustrate the denaturation data graphically. The data are organized such that the mutants of lowest enzyme activity are on the right. Most mutants which exhibited an increase in the stability (ΔG) of the Transition I had lower enzyme activities (notably R387L, F383A, M131A, L259A, M274A, W260A, V338A, and L336A in increasing order

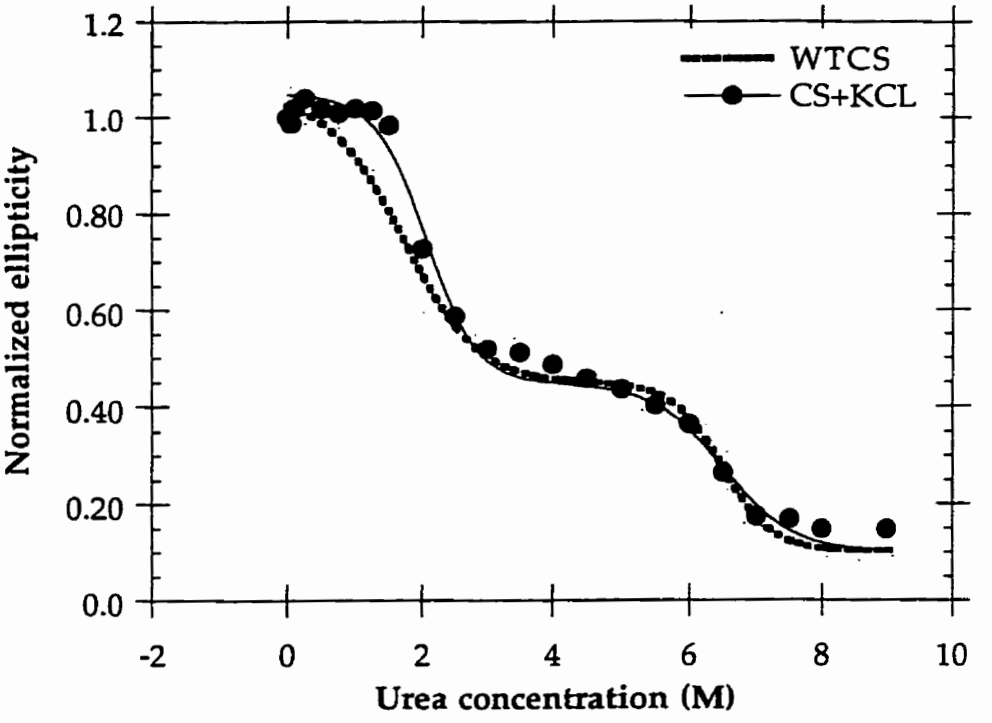


Figure 2.44 Urea denaturation and data fit for CS in the presence and absence of 100 μM KCl.

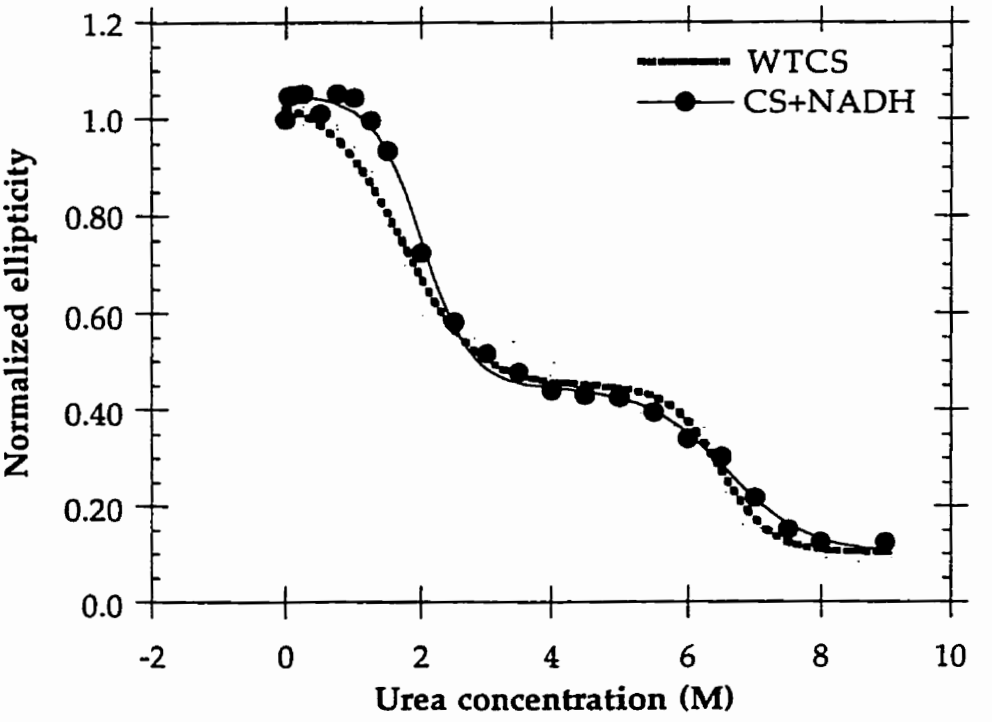


Figure 2.45 Urea denaturation and data fit for CS in the presence and absence of 100 μM NADH.

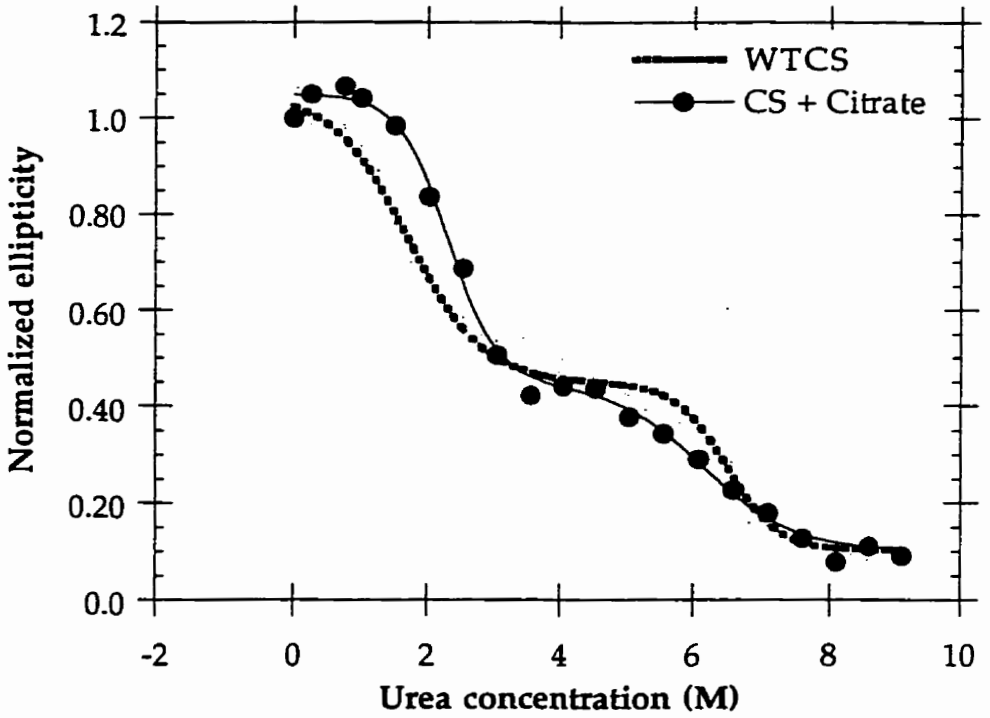


Figure 2.46 Urea denaturation and data fit for CS in the presence and absence of 100 μM citrate.

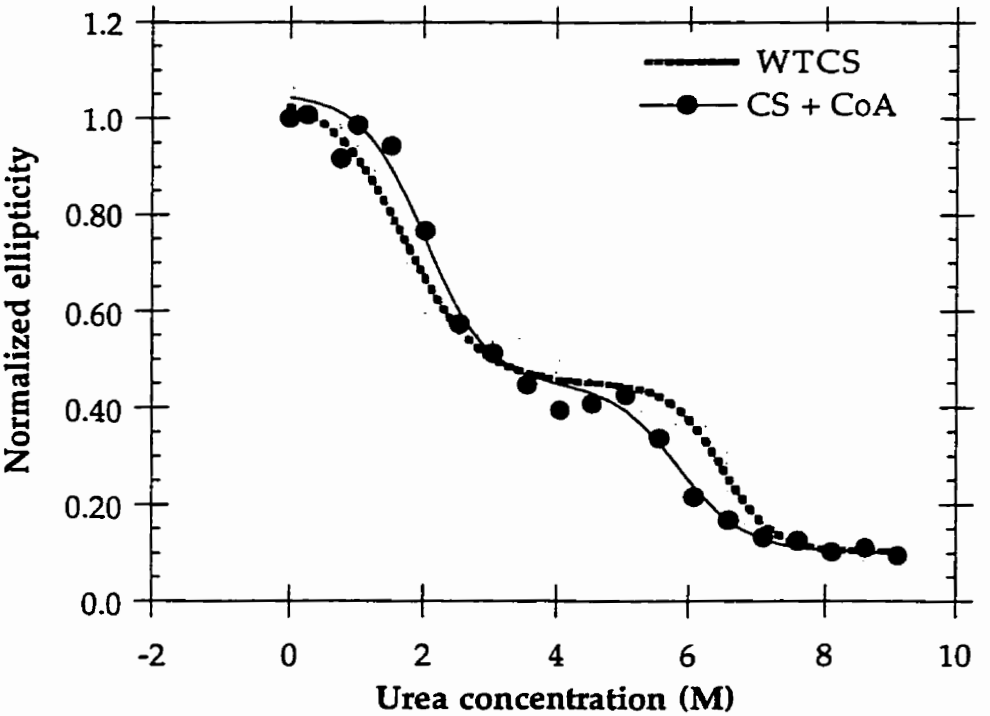


Figure 2.47Urea denaturation and data fit for CS in the presence and absence of 100 μM CoA.

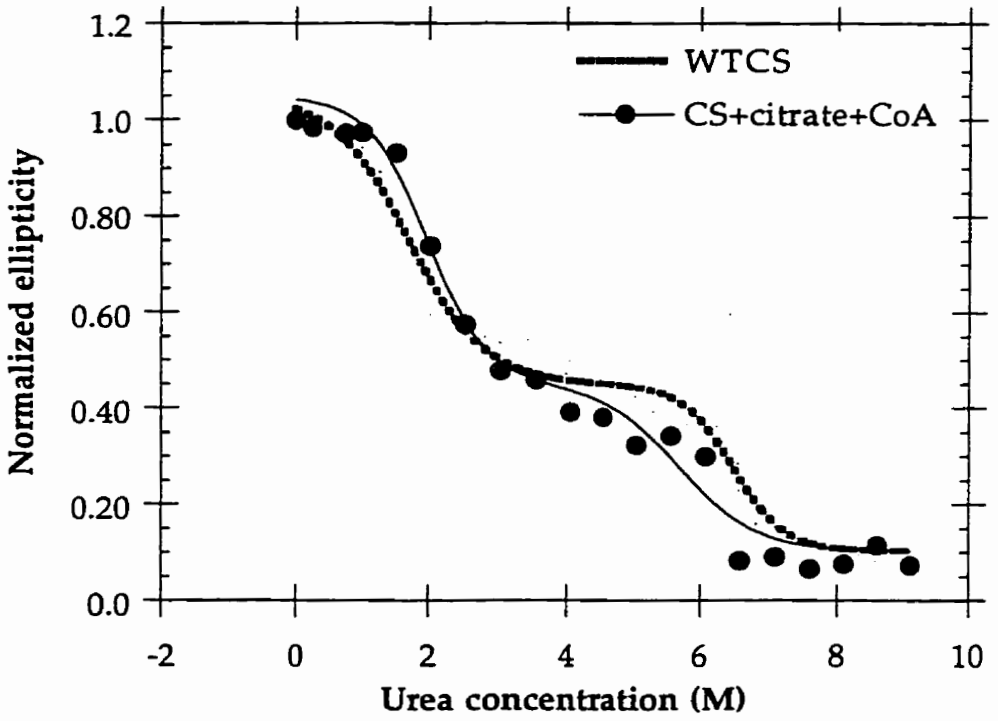


Figure 2.48 Urea denaturation and data fit CS in the presence and absence of 100 μM each of citrate and CoA.

TABLE 2.10 Free energies of unfolding, transition slopes and midpoints for CS urea denaturation in the presence of selected ligands

Description	ΔG_1 $kcalmol^{-1}$	ΔG_2 $kcalmol^{-1}$	m_1 $kcalLmol^{-2}$	m_2 $kcalLmol^{-2}$	TM_1 M	TM_2 M
WTCS	1.80 ± 0.11	10.5 ± 1.65	-1.05 ± 0.06	-1.61 ± 0.25	1.71	6.49
CS + citrate	3.72 ± 0.37	5.63 ± 0.75	-1.61 ± 0.16	-0.92 ± 0.12	2.31	6.12
CS + CoA	2.45 ± 0.33	7.18 ± 1.70	-1.26 ±0.16	-1.23 ± 0.29	2.02	5.84
CS + citrate + CoA	2.64 ± 0.46	6.08 ± 1.73	-1.33 ± 0.23	-1.06 ±0.30	1.98	5.74
CS + KCl	3.17 ± 0.38	7.51 ±1.65	-1.54 ± 0.19	-1.16 ± 0.25	2.06	6.47
CS + NADH	3.24 ± 0.24	6.48 ± 0.81	-1.16 ± 0.12	-0.99 ± 0.12	2.01	6.55

Note TM_1 and TM_2 are the midpoints of Transitions I and II, respectively.

of enzyme activity). There are some exceptions; W391A showed a general destabilization of both transitions and L275A only a slight increase in the stability of the Transition I despite low activity. A clear relationship between the extent of stabilization of Transition I and the reduction of enzyme activity is not apparent from the data; for example, V338A and L336A showed large increases in the stability of Transition I but modest decreases in enzyme activity.

Mutants which exhibited enhanced or unaltered activity (L164A and L326A, respectively) have reduced Transition I stabilities. Matthews (1996) suggested that there is an inverse relationship between stability and function and in general, the observations here are consistent with this view.

The trends observed with changes in the transition midpoints in general parallel those obtained with the change in free energy. In those mutants where this is not the case, the slopes of the transitions have changed significantly from the wild type values, leading to large increases or decreases in the free energy, despite similar midpoint values, or vice versa. In the case of Transition II, large uncertainties are associated with the slopes obtained, mainly due to the amplitude of the transition and the length of the extrapolation that is required to obtain the free energy in the absence of denaturant. However, the data indicate that in all mutants, Transition II occurs at lower urea concentrations than for the WT protein. The effect may be a lower extent of aggregation, or a destabilization of the parts of the protein which unfold in Transition II concurrent with breakdown of aggregates. It is difficult to assess whether either or both reasons for destabilization apply for a given mutant CS.

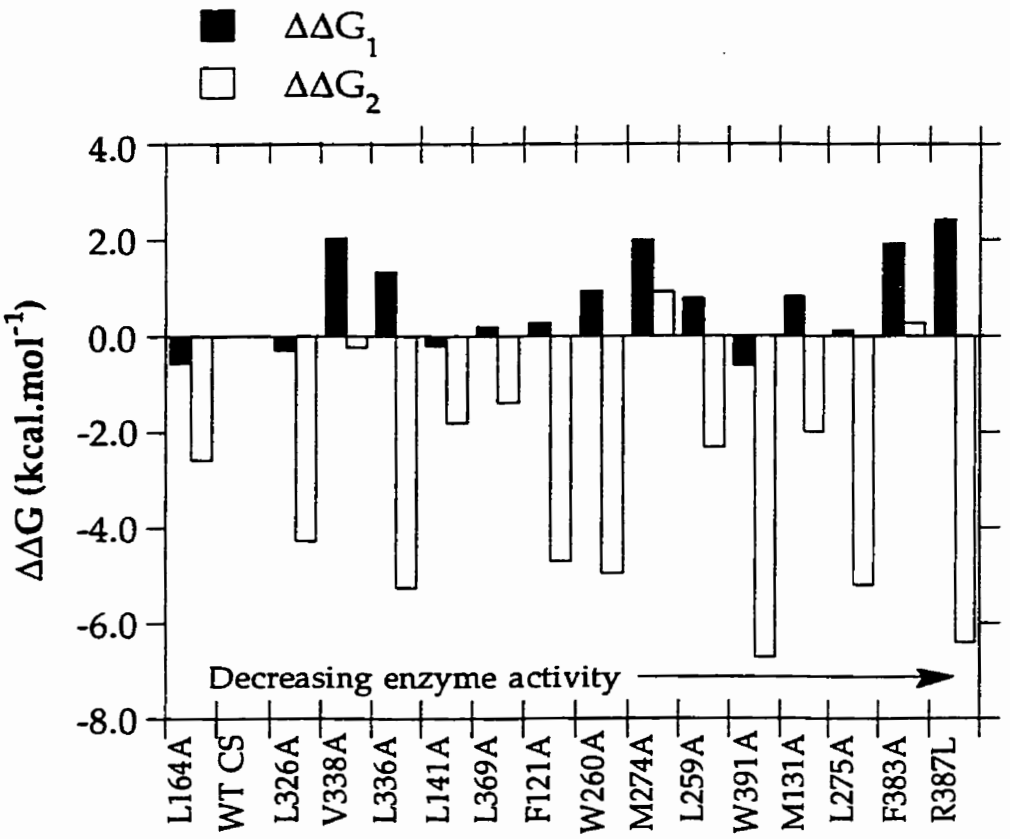


Figure 2.49 The change in the the stabilities (ΔG_1 and ΔG_2) obtained for mutant CSs. These values ($\Delta\Delta G$) were calculated by $\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wildtype}$, so a negative value is a destabilization and a positive one a stabilization. The order of the mutants is from highest to lowest enzyme activity, as indicated by the arrow. Typical errors are 0.4 and 3 kcal.mol^{-1} for $\Delta\Delta G_1$ and $\Delta\Delta G_2$, respectively.

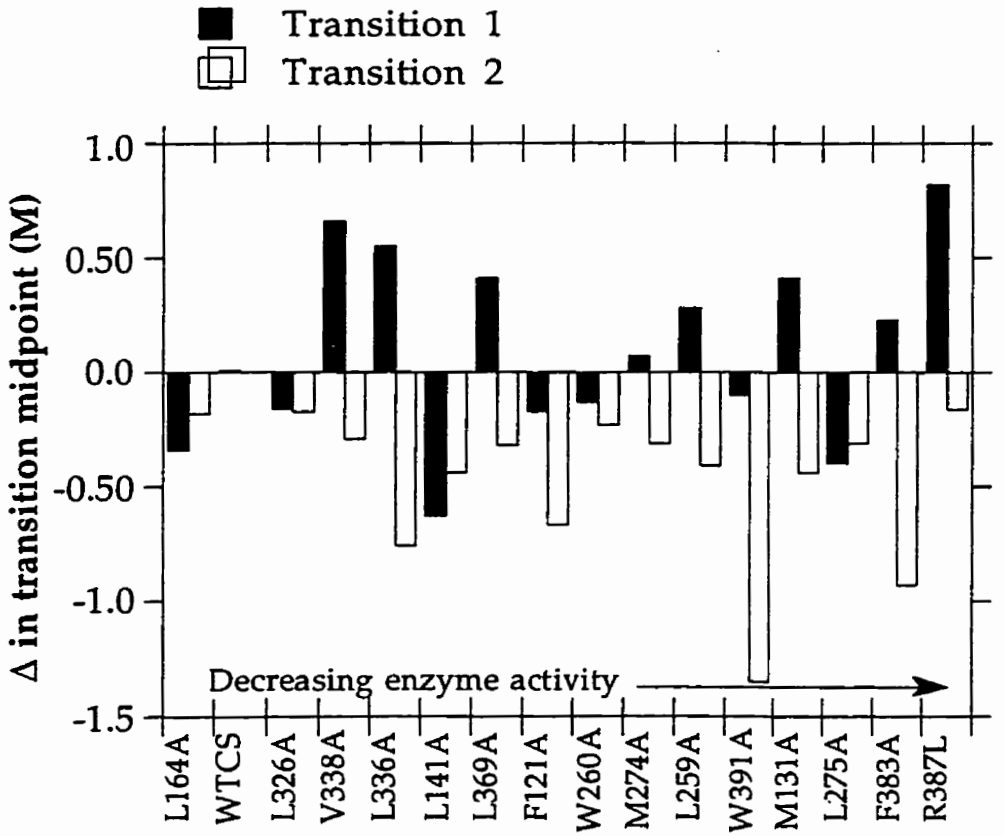


Figure 2.50 The change in transition midpoint, TM, for mutant CSs, calculated by $\Delta TM=TM_{mutant}-TM_{wildtype}$. A positive value is an increase in the TM, and a negative one, a decrease. Errors are typically 0.3 to 0.6 M for TM_1 and 2.5 to 4 M for TM_2.

Ligands increased the concentration of urea at the Transition I midpoint in all cases (Figure 2.51). However, the presence of citrate, CoA, or both decreased the Transition II midpoint, unlike NADH or KCl. The difference may lie in the specific binding regions of the ligand involved: KCl and NADH may bind to parts of the molecule which unfold in Transition I but citrate and CoA may still bind to partially unfolded CS such that aggregation is reduced, with a concomitant reduction in the transition midpoint.

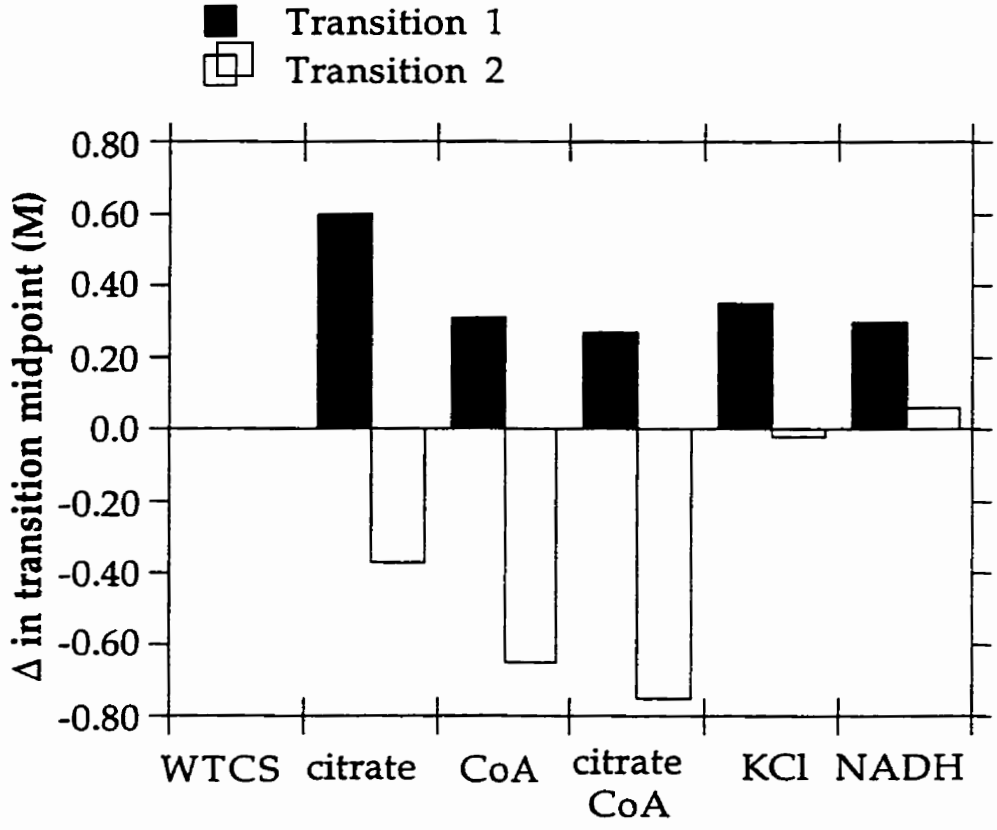


Figure 2.51 The change in transition midpoints of CS denaturation curves in the presence of some ligands. Each ligand was added at 100 μM.

Discussion

Unfolding of wild type E. coli citrate synthase

The unfolding of wild type *E. coli* CS by temperature or chaotropic agents occurs in two stages, with three states: native (N state), an intermediate state which is stable between 4 and 6 M urea (I state) and denatured (D state). Transition I, with a midpoint of 1.7 M urea, is irreversible and accounts for 67% of the structure loss. Transition II is reversible, has a midpoint at 6.5 M urea, and accounts for the remaining structure. Two optical methods, near-UV CD, and intrinsic fluorescence gave coincident unfolding transitions, implying that secondary and tertiary interactions break down cooperatively in each transition (see Figures 2.6 and 2.10). It has been suggested that coincidence of unfolding transitions obtained by different techniques is good evidence, but not proof, for two-state unfolding (Kim & Baldwin, 1982, Tanford, 1968).

Increase in ANS fluorescence, the classic diagnostic for exposed hydrophobic surfaces characteristic of the molten globule, revealed a solvent-exposed hydrophobic species at 2.5 M urea, well below concentrations at which the I state is observed by CD and fluorescence (see Figure 2.18). The ANS fluorescence decreased progressively with urea concentrations higher than 2.5 M. Further investigation was necessary to characterize this ANS-binding species and to exclude subunit dissociation as the source of hydrophobic surfaces detected.

The hexamer-to-dimer dissociation was not found to contribute significantly to the stability of Transition I, as determined from unfolding of TFBA-CS. This chemically modified CS is dimeric in the absence of KCl, and its Transition I is 0.5 kcal.mol^{-1} less stable than the WT CS; it is therefore

unlikely that the ANS binding species arises as a result of hexamer dissociation.

Transverse urea gradient gel electrophoresis (Figure 2.19) revealed that dimer-to-monomer dissociation occurs at urea concentrations below 0.25 M and therefore could not be the source of ANS binding species at 2.5 M urea. Furthermore, a number of bands appeared at progressively higher urea concentrations, reaching a maximum number at around 3 to 6 M urea, suggesting the presence of aggregates. These disaggregate only at much higher urea concentrations.

Thermal denaturation of modified CS also involved an intermediate state which was stable between 50 and 70°C. Electrospray ionization time-of flight mass spectrometry at 70°C also revealed the presence of aggregates up to octamers and including odd numbers of subunits (Figure 2.16). Aggregates corresponding to dimer and hexamer exhibited charge states which differed from those of the native species, consistent with those aggregates being non-native. The increased number of charges for the non-native dimer and hexamer species implied that they are expanded relative to their native counterparts, as the number of charges acquired by a protein in electrospray is related to the number of accessible basic residues and perhaps to surface area as well, as discussed in Chapter 3 (Loo, 1997).

All the observations described above are consistent with the presence of a self-associating intermediate in the equilibrium unfolding of CS. The decrease in ANS binding at urea concentrations leading up to the I state is explained by the burial of hydrophobic surfaces by aggregation. Self association of partly-unfolded intermediates has been observed previously in

other systems and is characteristic of a molten globule (Christensen & Pain, 1994, Kuwajima, 1989).

In summary, *E. coli* CS denaturation is complex: in Transition I, dissociation and unfolding of 67% of the tertiary and secondary interactions takes place, forming the I state, a collection of aggregates which retains the remaining 33% of the structure. Transition II corresponds to the disaggregation and unfolding of the I state to give fully unfolded molecules.

Unfolding of A. anitratum and chimera citrate synthases

The unfolding profile of *A. anitratum* CS differs significantly from that of its *E. coli* counterpart. First, *A. anitratum* CS is far more resistant to denaturation, unfolding only after the urea concentration reaches 4 M. Second, the unfolding of this CS is two-state, and not 3-state as observed for *E. coli* CS. It is not straightforward to attribute these differences to the 30% non-identity between the two sequences.

Unfolding profiles obtained for chimera citrate synthases in which the two structural domains were swapped between the *E. coli* and *A. anitratum* proteins suggested the possibility that the three-state unfolding behavior may be attributed to separate unfolding of the large and small domains. The key observation is that the chimera ECO::aci, in which the large domain was derived from *E. coli* and the small domain from *A. anitratum*, unfolds in a three-state manner, whereas the ACI::eco chimera, in which the large and small domains are derived from *A. anitratum* and *E. coli*, respectively, shows a two-state unfolding transition. These observations strongly suggest that Transition I originates with the large domain, and that the mode of denaturation (whether 2- or 3-state) is dictated by the source of the large domain. Taken alone, these findings point to separate unfolding of the large

and small structural domains in the three-state transitions. This idea was tested by using site-directed mutagenesis to destabilize one structural domain selectively by introducing carefully-placed "cavities" within hydrophobic cores. It is well known that mutations of large hydrophobic sidechains (L, V, M, I, F, W) to smaller ones (A) cause cavity formation in T4 lysozyme and other proteins (Matthews, 1996). To aid in the interpretation of the results actually obtained, which perhaps seem counterintuitive initially (especially with mutants exhibiting a more stable Transition I), it might be useful to understand the consequences of a three-state unfolding transition.

Description of three-state unfolding

In an effort to understand the effects of destabilization of the Native or Intermediate states on a denaturation profile we can use an analysis in which the free energies of the N and I states relative to the U state are plotted against denaturant concentration (Christensen & Pain, 1994, Robson & Pain, 1976). For the purposes of illustration, the parameters obtained from the data fitting of WT CS denaturation are used. From the data fitting, the stability of I relative to U was determined ($\Delta G_{I\text{-}U}$ =-10.45 kcal.mol^{-1}). The stability of N relative to U was not obtained from the data fitting, but it can be calculated (see Appendix III) to be $\Delta G_{N\text{-}U}$ =-12.25 kcal.mol^{-1}. The apparent free energies of N and I relative to U (the reference state at ΔG=0) as a function of urea concentration were then calculated by equations 6 and 7 in Appendix III as shown in Figure 2.52. The midpoints for the N to I and I to U transition correspond to the intersection points of the respective lines in the graph. The relative positions of the midpoints dictate whether or not an intermediate is observed; if the transition midpoints for I to U and N to U are the same, then the intermediate would not be detected

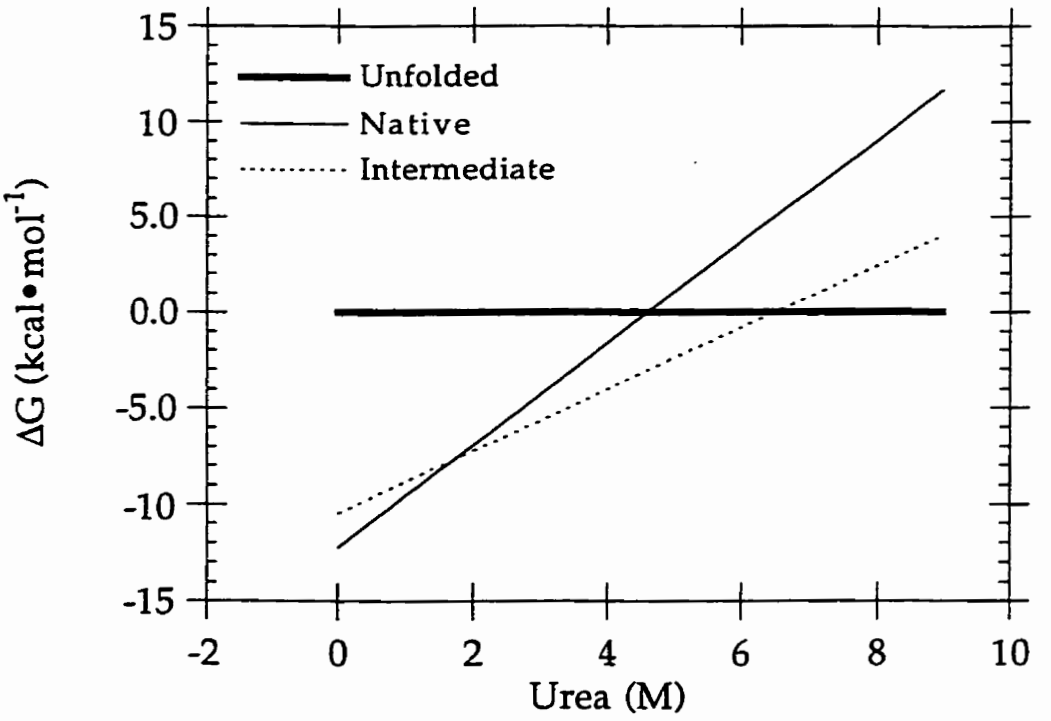


Figure 2.52 The free energies of Native and Intermediate states of WT *E. coli* CS as a function of urea concentration. The method by which the data were generated is described in Appendix III.

since the N and I states unfold simultaneously, and the unfolding would appear to be two-state. This explains why for some proteins intermediates are observed, and for others, they are not (Christensen & Pain, 1994).

Now, consider a hypothetical mutation, located in the part of the molecule that unfolds during Transition II, and causes destabilization of the I state, as shown in Figure 2.53. The graph was calculated using a value $\Delta G_{I\text{-}U}$=9 kcal.mol^{-1}, assuming that urea dependence of this parameter is the same as that for WT CS. This assumption may be valid given that for a mutant protein, the number of binding sites for urea is not altered significantly from that for WT CS; Tanford's denaturant binding model suggests that the slope, ***m*** is related to the difference in the number of binding sites between Native and Unfolded protein. The corresponding graph for the free energy change as a function of urea concentration intersects the Native line at a higher urea concentration (Transition I midpoint increased relative to WT CS), but the point of intersection with the unfolded reference is at a lower concentration (Transition II midpoint decreased relative to WT CS). This analysis predicts that this hypothetical mutant would exhibit an apparently stabilized Transition I and a destabilized Transition II; that is, the midpoints for the N to U and I to U transitions will occur closer together on the urea axis. The overall stability of the N to U transition may be unchanged by this mutation, or the effect may be small enough that it is masked by the large changes in Transition I which arise from a destabilized I state.

If a mutation caused a more stable N state, that would give rise to a similar situation as above, where the Transition I is apparently stabilized while Transition II destabilized. However, since the mutants in this study were chosen to be destabilizing (cavity forming), this scenario is highly

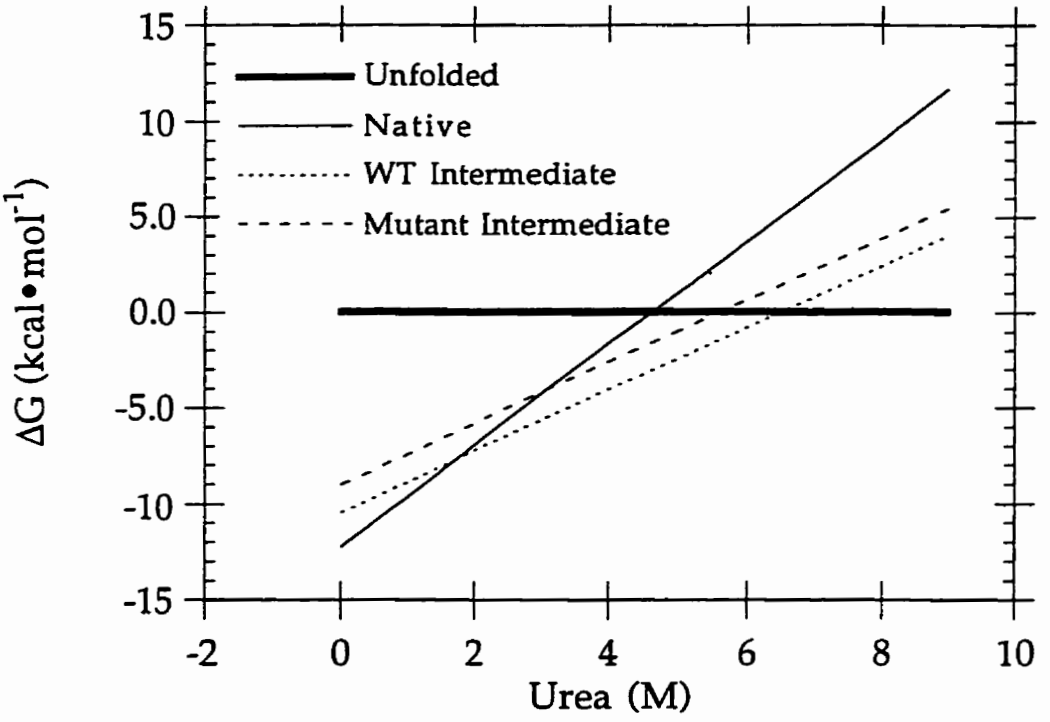


Figure 2.53 The free energies of Native, Intermediate, and a hypothetical destabilized Intermediate of *E. coli* CS as a function of urea concentration.

unlikely. The only possible exceptions may be mutants of residues in the active site, residues which may be optimized not for stability, but for function, as suggested previously (Shoichet *et al.*, 1995).

In another simpler situation, a mutation which caused destabilization of the Transition I should have no effect on the stability of the intermediate, and therefore the midpoint of Transition II should not change (graph not shown). Destabilization of Transition I with no effect on the second, then, is evidence that the mutation is in a region of the protein which unfolds in Transition I. From the above considerations, then, it may be possible to assign mutants to one of two categories: one where Transition I occurs at a higher urea concentration than WT CS indicative of a destabilized I state, and one where the urea concentration is lower, indicative of a destabilized N state.

E. coli CS mutants and the identification of I state structural elements

CS mutants were denatured with urea and their denaturation curves analyzed to obtain the stabilities, slopes, and midpoints for each of the two transitions. However, as mentioned in Results, large uncertainties were associated with parameters obtained for Transition II because the data obtained were not precise enough. The analysis in the preceding section, however, enables us to assign mutants to two categories according to the behaviour of Transition I (Table 2.11). Category A mutants exhibit transition midpoints which are lower than that for wild type CS, and Category B mutants have greater transition midpoints. The secondary structural elements associated with each mutant are listed. This information can be used as an indication of which elements remain folded in the I state.

TABLE 2.11 ***E. coli*** **CS mutants and their associated secondary structural elements**

Category A Reduced Transition I midpoint		**Category B** Increased Transition I midpoint	
Mutant	Location	Mutant	Location
F121A	Helix F	M131A	Helix G
L141A	Helix G	L259A	Helix M
L164A	Helix I	M274A	Helix N
L326A	Helix P	L336A	Helix Q
W260A	Helix M	V338A	Helix Q
L275A	Helix N	L369A	Helix R
		F383A	Helix S
		R387L	Helix S

Note: W391A is not included, but is discussed in a later section.

Most secondary structural elements appear in only one category, with the exception of Helices N, G, and M. It is conceivable, for those mutants which caused a reduced Transition I midpoint, that the overall stabilities of both the N and I states are reduced. As already stated, a mutant will be placed in Category B, if the mutation causes a significant destabilization of the I state and a modest or nil effect on the overall stability of the N state, leading to a remarkable change in the unfolding transition, and, paradoxically, to an apparently more stable Transition I. It may be argued therefore that the mutants in Helices N, G, and M do reside in parts of the molecule which are structured in the I state, but the basic assumptions do not hold true in their cases.

A unifying feature of mutants in Category B and their associated secondary structural elements is the proximity to the inter-subunit interface and to the active site. It is proposed, therefore, that most of the small domain (Helices N, Q, and R) and three helices of the large domain (Helices G, M and S) retain secondary structure in the I state. The proposed structure, shown in Figure 2.54 as a ribbon diagram, contains approximately 115 residues (residues 168-181, 247-281, and 335-400 according to the *E. coli* model). The amount of secondary structure in this proposed intermediate is approximately 34% of the secondary structure in the N state. This is in excellent agreement with the size of the Transition II, 33% of the total mean residue ellipticity observed for CS.

Further evidence for the proposed I state from intrinsic fluorescence studies

Comparison of intrinsic fluorescence spectra of W260A and WT CS revealed that W260A is largely solvent exposed. Therefore the remaining tryptophan residues at positions 391 and 395 dominate the intrinsic

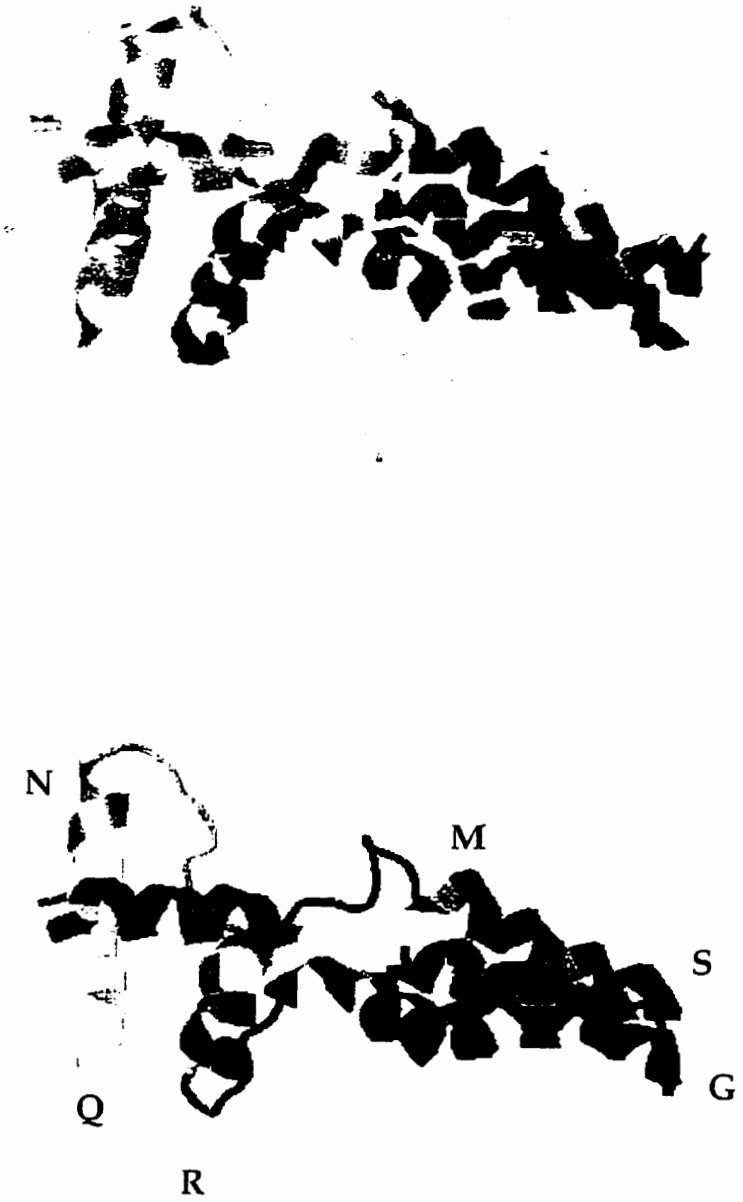


Figure 2.54 Ribbon representations of the proposed I state as it exists in the native pig CS shown as part of the CS monomer (Top) and isolated (Bottom). The colour scheme is a spectrum from blue, representing the N terminus, to red, representing the C terminus.

fluorescence of CS. Since these tryptophans are predicted to be next to each other on the same face of Helix S, the fluorescence spectra are reflective of that specific environment. The emission maximum is a good indicator of polarity of the environment of tryptophans: solvent exposed tryptophans emit at 360 nm, and upon transfer to more apolar environments, the emission shifts to the blue (Lakowicz, 1983). Transition I in WT CS and in W260A CS is associated with only a 10 nm red shift, whereas Transition II exhibits a much larger red shift of 30 nm, indicating that the environment in the vicinity of W391 and W395 does not become appreciably solvent exposed until the end of Transition II. This can be interpreted to mean that at least the C-terminal part of Helix S, is structured in the I state, consistent with the conclusion that Helix S is a part of the proposed intermediate, from the CS mutant studies discussed above.

It is interesting to note that the residues corresponding to the two tryptophan residues in pig CS appear to be on the surface of the proposed *E. coli* intermediate structure. This suggests that, in addition to the hydrophobic interaction between them, aggregation may play a role in maintaining a non-polar environment in their vicinity.

Mutation of residues W395 and W391 to alanine affected the stability of CS profoundly; W391A unfolding was non-cooperative, and the W395A protein was not successfully purified, likely owing to its instability *in vivo*. The fact that these residues are on the same face of a helix (they are 4 residues apart in the sequence) coupled with the observation that these residues are in an apolar environment in the intermediate, suggest that they may be important in the initiation of folding of CS. Tryptophan residues in the fully unfolded yeast phosphoglycerate kinase, and other hydrophobic residues

including aromatic ones in Barnase (Arcus *et al.*, 1994), 434 repressor (Neri *et al.*, 1992), and FK 506 binding protein (Logan *et al.*, 1994) have been shown to have residual structure, even under denaturing conditions. These are thought to act as hydrophobic nucleation sites for folding initiation. It may be of interest to study the two CS tryptophan residues under denaturing conditions to see whether residual interactions exist. That homologous residues in other CSs are aromatic or hydrophobic (see alignment in Appendix I) strengthens the idea that these are critical for folding and/or stability of CS.

Features of the proposed I state

The proposed I state (Figure 2.54), contains elements from both the large and small structural domains of CS. Helices N, Q, and R are the core of the small domain. The large domain helices M and G are those which are involved in the dimer contact region, and thus, along with Helix S, they originate from the central 2 layers of the large domain architecture. It is interesting that the connecting strands between the large and small domains (between Helices R and S and Helices M and N) are part of the proposed structure.

A space-filling model of the I state structure as it exists in the native, folded CS is shown in Figure 2.55. Why such an intermediate would form aggregates is clear from this model - a large proportion of its surface is composed of hydrophobic residues. In solution, however, the structure may be one where tertiary interactions are transient and fluctuating, resembling a molten globule and increasing even more the potential hydrophobic surface available for aggregation.

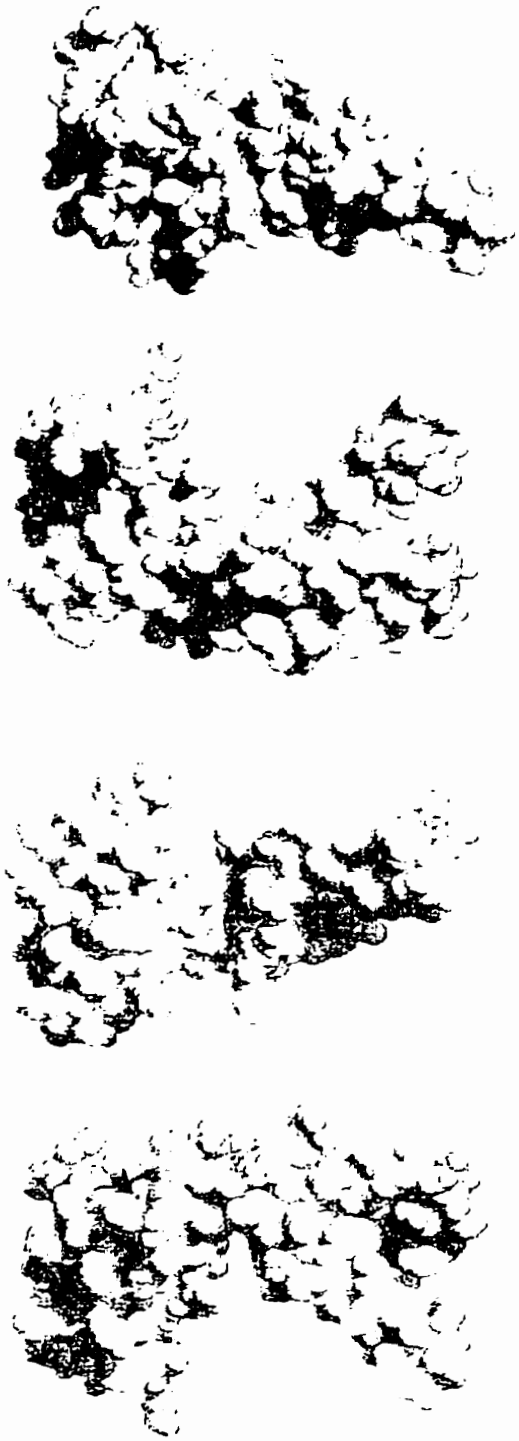


Figure 2.55 Space filling representations of the I state of CS, rotated 90° about the X axis for each figure from the preceding one. The hydrophobic residues are shown in yellow.

Of the residues which were mutated in this proposed I state structure, eight are predicted to be on the surface, perhaps playing a role in the aggregation which stabilizes the I state. The observation of reduced Transition II midpoints in these mutants may be explained as the destabilization of the I state by reduction of aggregation potential. For mutants which are predicted to be buried within the structure (as it exists in the native molecule), it is possible that the source of destabilization of Transition II is formation of a cavity. However, since it is not possible to say conclusively whether or not native tertiary interactions exist in the I state, it is also possible that those buried residues may also be involved in aggregation. All mutants of the I state, therefore, may cause destabilization of the I state through reduced aggregation. A natural conclusion, if the above argument is correct, is that if aggregation does not occur, unfolding of *E. coli*-like CS molecules will approach a two-state transition.

Why is an intermediate not observed in A. anitratum CS unfolding?

Unfolding of *A. anitratum* CS, unlike the *E. coli* enzyme, is a single, two-state transition. A possible explanation for the observed difference is reduced or absent aggregation of the I state in *A. anitratum* CS. Comparison of the sequences of the two enzymes, in the region which are proposed to retain structure in the *E. coli* CS I state, reveals differences at 15 positions (out of a total of 115), with most differences being conservative in nature. Therefore the two enzymes are 90% identical in the I state sequence, and it is unlikely that the observed unfolding differences between the two enzymes can be ascribed to the sequence dissimilarities.

If the observed differences in unfolding are not due to differences in the I state, then they must originate from dissimilarities of the rest of the

molecule (that is, the parts which normally unfold in Transition I for the *E. coli* CS). A simple explanation, then, is that in *A. anitratum* CS, the relative positions of the N to U and N to I transition midpoints are such that the I state is not observed. A slightly more stable N state (relative to *E. coli* CS) would be sufficient to give rise to such a situation. The minimal requirement is illustrated in Figure 2.56: the transition midpoints for N to U and I to U coincide, and only one transition with two states N and U is observed.

Given how little is known about the magnitudes of the different forces that lead to stable protein conformations, and their relationship to sequence information, it is impossible to rationalize the differences observed for *A. anitratum* and *E. coli* CS unfolding in terms of differences in sequence. This is essentially a reiteration of a major issue of the protein folding problem.

Significance of the I state for protein folding

The reversibility of Transition II, when urea concentration is decreased, indicates that the I state may represent the product of an initial collapse in the early folding steps of CS, forming a molten-globule like intermediate that is prone to aggregation. It was previously shown that CS does not fold in *E. coli* strains where the chaperonin GroEl is not expressed, demonstrating the need for assistance in its folding (Horwich *et al.*, 1993). It is attractive to propose that the I state identified here may be the substrate for chaperonin binding. Chaperonins are believed to assist folding by preventing aggregation, preventing pathways leading to misfolded proteins, and assisting in the assembly of multisubunit and multidomain proteins. The observation that

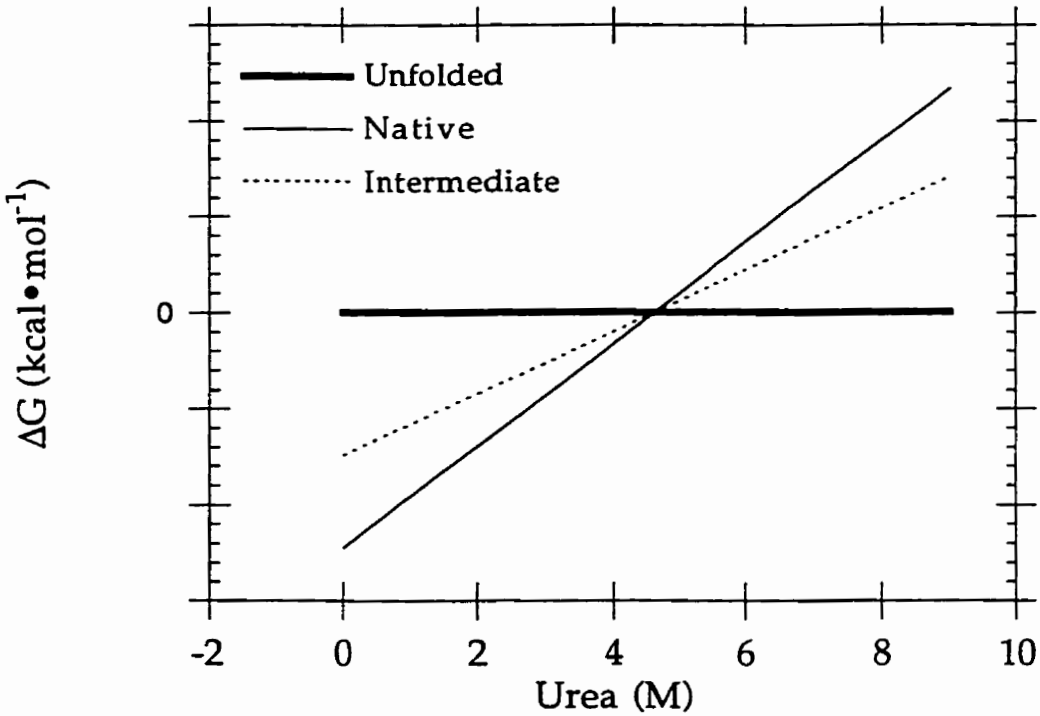


Figure 2.56 Hypothetical free energy profiles where coincidence of the N to U and I to U unfolding transition midpoints leads to cooperative breakdown of both I and N, and therefore I is not observed under equilibrium conditions.

Transition I is not reversible is therefore not surprising given the fact that GroEl is required for CS to fold *in vivo*.

The effect of ligands on CS unfolding

Ligand binding generally stabilizes proteins against denaturation. This effect occurs because the N state binds the ligand more tightly than the U state does, and not necessarily because the N state undergoes a structural change that may impart stability (Creighton, 1993). In the case of CS however, it is well known that binding of substrates, products, or substrate analogues, causes large conformational changes, as discussed in Chapter 1. Similarly, KCl and NADH, the allosteric effectors, cause changes in the oligomeric state of CS (as discussed in Chapter 3). Stabilization of CS arising from the binding of the two classes of ligands (active site ligands, and allosteric ligands) may be a consequence of altered structure in addition to the general effect of ligands on protein stability.

KCl and NADH, ligands which induce hexamerization of CS, caused a stabilization of Transition I but little effect on Transition II (based on transition midpoints). However, citrate and CoA, the products of the CS reaction, caused a stabilization of the Transition I, but a large destabilization of in Transition II. In the latter case, the effect may be that the N state was stabilized sufficiently that the I state is less populated at equilibrium.

Enzymatic activities of CS mutants

Some of the CS mutants constructed exhibited altered CS activity, with most having reduced specific activities under standard conditions. Although correlations have been made previously between stability and activity in proteins such an analysis is not feasible for CS since the changes in stability of the N state induced by some mutations are minimal. Matthews reasoned that

residues which are important for activity in T4 lysozyme may not be optimized for stability, in light of finding an inverse relationship between stability and function (Shoichet *et al.*, 1995).

The mutant L164A CS, which has the replacement in the core of the large domain in Helix I, conforms with the activity/stability relationship. Although the stability of Transition I was reduced by 0.5 kcal.mol^{-1} relative to WT CS, the specific activity of this mutant increased significantly. The instability of this mutant was also apparent from the lower yields obtained after purification, its shortened storage life, and its tendency to form aggregates.

Enzyme kinetics in the presence of KCl for L164A CS revealed that AcCoA binding was enhanced relative to WT. This suggests that the mutation may have shifted the allosteric equilibrium towards the active R state. In support of this idea, the allosteric inhibitor NADH was not found to inhibit the activity of this enzyme. This loss of inhibition cannot be attributed to lack of hexamer formation, a feature essential for NADH binding as shown in Chapter 3; L164A forms hexamers in the presence of KCl based on its elution volume from size exclusion columns and its migration using non-denaturing gel electrophoresis. The L164A mutation, located in Helix I, is in close proximity to a region of low homology between the allosteric and non-allosteric CS sequences. This region near Helix J may be a loop in which residue C206 was previously shown to be a marker for the allosteric NADH binding site (Donald *et al.*, 1991, Talgoy *et al.*, 1979). It is possible that destabilization of the I helix affects NADH binding by a conformational adjustment that shifts the equilibrium to the allosteric R state.

The mutants M274A and L275A both exhibited reduced specific activities, with the latter being severely affected. The pig CS homologues of these two residues interact with a residue involved in hydrophobic interaction with AcCoA, V314 (L300 in *E. coli* CS) which is in the O-P adenine binding loop. Kinetics of L275A CS reveal that the AcCoA saturation curve is sigmoidal even in the presence of KCl. K_{AcCoA} is estimated to be 1 mM, about 10-fold higher than that for WT CS. This finding is similar to that obtained for the F383A mutant, studied by Periera *et al.* (1994). F383A CS had a severely lowered affinity for AcCoA, leading to the suggestion that the allosteric equilibrium was shifted towards the T state. A striking difference between the L275A and the F383A mutants, however, is the K_{OAA} values: F383A has only a slightly reduced affinity for OAA, while L275A has a 10 fold reduced affinity for this ligand. It is possible that L275 is critical for the maintenance of the entire active site architecture, not just for its interaction with residues implicated with CoA binding.

The pig CS homologues of residues L259 and M131 appear to be important in maintaining the conformation of the F383 equivalent; as mentioned above, the latter is involved in AcCoA binding. L259A interacts sideways with the phenyl side chain, and in turn, M131 interacts with L259, forming a network of hydrophobic interactions, spanning three different secondary structural elements of the molecule. Those helices are Helix S (F383), Helix M (L259), and Helix G (M131). The affinity for AcCoA for each of the mutants obtained from AcCoA saturation curves decreases in the order M131A>L259A>F383A, consistent with the importance of the first two in maintaining a proper conformation of F383 for optimal binding of AcCoA. It is interesting to note that the I state structure proposed earlier contains those

three Helices; it is possible that M131, L259, and F383 form a hydrophobic cluster early in the folding of CS.

Future work and concluding remarks

Clearly, three dimensional structures are needed for the *E. coli* WT and mutant CSs to assess the response of the protein to the mutations. Unfortunately, no three dimensional structures are available for the *E. coli* CS nor any of the other gram negative enzymes. However, a collaboration with Nham Nguyen of Dr. G. Brayer's Laboratory at the University of British Columbia has resulted in several crystallography datasets for WT CS in the presence and absence of NADH. Some of the mutants described here and some enzyme-substrate complexes have also been crystallized.

Confirmation of the proposed I state structure by other methods is needed. For example, it may be useful to proteolytically digest CS at 4 M urea and compare the fragments obtained to those obtained for CS in the absence of denaturant. This may, at least, reveal those cleavage sites which are still protected by aggregation and/or folded structure of the I state.

An interesting question is whether the stability of the I state, and thus the ability to observe it, depend on aggregation or not. Denaturation of CS at concentrations lower than those used in this work may reveal whether this is the case; the use of the appropriate combination of a strong denaturant and lower protein concentrations may result in a two-state transition. If a two-state transition can be achieved in this manner, then this would strongly suggest that the stability of the I state is dependent on aggregation.

The mutants which exhibited altered activities warrant further kinetic work. Especially interesting is the L164A mutant which has elevated specific activity, binds AcCoA more easily, but is not subject to NADH inhibition.

These observations suggest that this mutant has an allosteric equilibrium which is shifted to R state, but still has the ability to form hexamers. A more complete study of the enzyme kinetics for this mutant, and others, in the presence and absence of KCl is needed.

CHAPTER THREE Mass Spectrometry of Citrate Synthase

Introduction

General introduction

Mass spectrometry is a technique by which the mass, a simple but important property of all matter, is determined, by separation of gas-phase ions according to their mass-to-charge (m/z) ratios. There are two essential components to the mass spectrometric experiment. First, the analyte must be ionized and evaporated to produce gas-phase ions. Second, the ions must be separated and detected according to their m/z, yielding the information necessary to obtain an accurate measure of molecular mass and the relative abundance of the different ions in the analyte. In conventional mass spectrometry, used extensively for analyses of small compounds, the analyte is ionized and evaporated by methods such as fast atom or ion bombardment, laser or plasma desorption, and chemical ionization (Willard *et al.*, 1988). These ionization methods are often harsh, and fragmentation of the analyte molecules often occurs, a useful feature for learning more about the structures of the molecules. However, macromolecules, such as proteins and nucleic acids, are often difficult to ionize and evaporate using the conventional ionization techniques, and in most instances, their decomposition precedes ionization (Mann & Wilm, 1995). This difficulty was overcome in 1988 by two different ionization techniques: matrix assisted laser desorption (MALDI) (Karash & Hillenkamp, 1988) and electrospray ionization (ESI) (Fenn *et al.*, 1989).

MALDI involves the vaporization and ionization of proteins from a mixture of a light-absorbing matrix and the analyte solution by irradiation with short laser pulses. The nature of this method is such that non-covalent interactions (exemplified by those which maintain the folded conformations

of proteins, subunit interactions and ligand-protein interactions) are destroyed. This method therefore has found utility in molecular mass determinations of pure or mixed preparations, and does not require rigorously purified or desalted samples.

In ESI, the analyte molecules are ionized by directly transferring them from liquid (usually aqueous) to the gas phase. The main advantage of ESI is its gentleness; ions are produced for which non-covalent interactions are preserved. The project described in this chapter involves the ESI method coupled with time-of-flight (TOF) for the mass spectrometric study of CS protein-protein and protein-ligand interactions and equilibria. This was a collaboration project with Andrew Krutchinsky of Dr. Standing's group in the Physics Department.

Electrospray ionization

The ESI technique involves spraying an aqueous analyte solution through a glass or metal capillary which is kept at high electrical potential. The fine mist that emerges is composed of small, highly-charged droplets, which are rapidly evaporated by dry gas, heating, and/or reduced pressure. This process effectively concentrates the charge acquired by the droplets onto the surface of the analyte molecules (such as proteins) to produce gas-phase analyte ions. For MS of proteins, the charging species is usually the proton although negative-mode MS can also be used.

The number of charges acquired by a protein molecule is dependent on the availability of basic sites (such as basic side chains of proteins: lysine, arginine, and the amino terminus), but is thought to be also proportional to the available surface area. The number of attached charges varies among molecules of the same type, and the resulting mass spectrum is composed of a

series of peaks corresponding to the number of attached charges for a single analyte species with m/z values described by

$$\frac{MW + z.M_H}{z} \qquad [1]$$

where MW is the mass of the protein, z is the number of attached charges per molecule, and M_H is the mass of the charging species (1 for proton). A typical spectrum is shown in Figure 3.1. This feature of ESI, called multiple charging, allows the analysis of large ions even with analyzers of limited m/z range. Another advantage of multiple charging is the accuracy with which molecular weights can be calculated from the distribution of the multiply-charged peaks (Siuzdak, 1994) (Hofstadler *et al.*, 1996). A simple algorithm is used to determine the molecular mass from a charge envelope. Assuming that each peak differs from the neighbouring peak by one charge, the charge state for a particular peak (Z_1) may be assigned by

$$Z_1 = \frac{(m/z)_2}{\left(\frac{(m/z)_2 - (m/z)_1}{n}\right)} \qquad [2]$$

where $(m/z)_1$ is the mass to charge ratio of the peak for which the charge state is being determined, $(m/z)_2$ is mass-to-charge ratio of a second peak with a value larger than $(m/z)_1$, and n is the number of peaks separating the two peaks being considered. Once the charge state of one peak has been assigned, the charges on the remaining peaks are directly calculated. Using the m/z ratios and the charge for all the peaks in an envelope, the mass can be calculated for each peak by

$$MW = \left((m/z).z\right) - z.M_H \qquad [3]$$

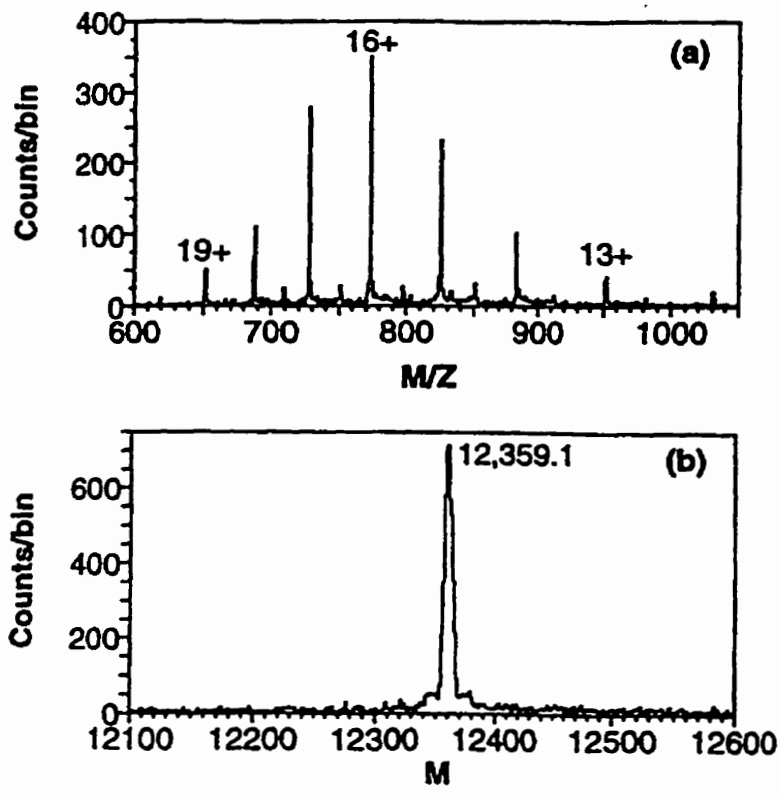


Figure 3.1 ESI MS spectrum of cytochrome c. Taken from (Chernushevich et al., 1997). A. The m/z spectrum. B. Deconvoluted spectrum (mass spectrum).

Since each peak gives an independent mass measurement, a more accurate mass can be obtained over the entire charge envelope. Fortunately, computer algorithms carry out these calculations automatically to yield the molecular mass (see Figure 3.1).

In addition to obtaining accurate measurements of masses of proteins by ESI MS, other important aspects of protein structure may be obtained from the spectra. The most obvious feature is the number of charges acquired: compact proteins, will acquire fewer charges than the extended conformations of unfolded proteins since the number of acquired charges depends on the surface area available for protonation. Thus folded proteins will have, in general, larger m/z ratios than their unfolded counterparts.

Another feature of ESI mass spectra is the width of the charge envelope obtained (i.e. the number of m/z peaks in the charge envelope). This property is related to the number of possible conformations that give rise to different surface areas (that is, conformational heterogeneity), and thus varying extents of accessibility to protonation sites. For a folded population of a protein, one might expect little conformational heterogeneity, whereas unfolded protein populations may exhibit a large number of different interconverting conformations, each with a slightly different surface area. As a consequence of the above, a sample of folded protein would give rise to a few charge states and therefore a narrow charge envelope, in contrast with an unfolded protein sample, which produces many different charge states and thus a wider charge envelope.

The above features were demonstrated first by a study of cytochrome c by ESI MS at a range of pH values over which, as other techniques have shown, it exists in at least three possible conformations (Chowdhury *et al.*,

1990). ESI mass spectra of cytochrome c were used to demonstrate the presence of the three conformations; three charge distributions were observed, with increased numbers of charges, and wider charge envelopes associated with the partially and fully unfolded states (Chowdhury *et al.*, 1990). Heat-induced conformational changes in ubiquitin and cytochrome c were also observed by ESI MS, demonstrating the same relationships between conformational state, the width of the charge envelope and the number of charges attained (Mirza *et al.*, 1993). Most recently, an ESI MS study of the acid-induced unfolding of cytochrome c at different methanol concentrations revealed that the charge distribution is not sensitive to changes in secondary structure, but may be attributed to changes in tertiary structure (Konerman & Douglas, 1997). The same group studied the kinetic unfolding of myoglobin using pH jump experiments (Konermann *et al.*, 1997b), and cytochrome c refolding kinetics (Konermann *et al.*, 1997a) utilizing the information from charge distributions in time-resolved ESI MS experiments.

Non-covalent intermolecular interactions

Since the first reports of the observation of specific non-covalent protein-protein and ligand-protein interactions by ESI MS, a variety of other types of interactions have been studied by this method, including protein-peptide, antibody-antigen, enzyme-substrate, protein-cofactor, and protein-DNA interactions (reviewed by Loo, 1997 and references therein). However, non-specific interactions have also been observed, including adducts with salt ions or solvent species, and the non-specific aggregation between proteins. While more and more evidence is in keeping with the idea that mass spectrometric measurements reflect the interactions in solution, the observations of the non-specific interactions in some cases raises questions

about the reliability and validity of the interactions observed by MS. Therefore an important issue is how non-specific interactions can be differentiated from specific ones. More mass spectrometric studies are needed on systems for which non-covalent interactions have been characterized by conventional techniques such as fluorescence, ultracentrifugation, NMR and X-ray crystallography, so that the results could be compared.

Using ESI MS for the study of protein complexes

A large number of ESI MS studies on protein-protein interactions have been reported, and the findings are usually consistent with quaternary structure information obtained by other methods (Loo, 1997). For example, the tetrameric assemblies of avidin, concavalin A, and human hemoglobin were observed by ESI MS, and are consistent with previous information on their subunit structure (Light-Wahl *et al.*, 1994). Furthermore, two oligomeric species of yeast alcohol dehydrogenase (monomer and tetramer) were observed simultaneously, without the presence of dimer or trimer, also consistent with specific characteristics of the protein as probed by other methods (Loo, 1995).

Some proteins (such as citrate synthase from gram negative bacteria) exhibit equilibria between two oligomeric forms, a feature which is intimately linked to allosteric regulation and function (Traut, 1994). MS should be ideal for the study of such systems; the exact composition of the mixture and the relative abundance of each component could be determined.

The first report of a ligand-protein noncovalent complex studied by ESI MS was an interaction between the cytoplasmic receptor FKBP and two immunosuppressive drugs (Ganem *et al.*, 1991). Using a competitive binding experiment, they demonstrated that the ligand which binds more tightly in

solution complexed more readily with the receptor in ESI mass spectra, producing a peak of higher abundance than with the other ligand (Ganem *et al.*, 1991). More recently in this laboratory, specific binding of the TrpR protein to its operator DNA and to L-tryptophan was demonstrated by ESI-TOF MS (Potier *et al.*, 1997). These and other studies of ligand-protein complexes have suggested that specific binding is observable by ESI MS.

Ultimately, it would be of interest to see whether binding constants could be measured using mass spectrometry, given that gas-phase observations seem to parallel those obtained in solution. Several groups have attempted to use ESI MS measure K_D values for particular complexes. For example, levels of free and ligated bovine serum albumin (with an oligonucleotide ligand) were quantitated directly from their relative abundances in ESI MS spectra, yielding Scatchard plots from which K_D values were obtained; those values agreed with those determined by capillary electrophoresis (Greig *et al.*, 1995). Similarly, Scatchard analysis of ESI MS data for association of vancomycin antibiotics with tripeptide ligands gave dissociation constants which were consistent with solution binding constants (Lim *et al.*, 1995). On the other hand, a study of the association of acyl CoA-binding protein with acyl CoA analogues by ESI MS is an example where the ESI MS data did not reflect solution measurements; in that study, acyl CoA ligands which have very different binding constants in solution appeared to have equal affinity for the binding protein when studied by ESI MS (Robinson *et al.*, 1996). It is thought that different kinds of interactions (hydrophobic, electrostatic) may be stabilizing or destabilizing to the complex in the gas-phase, giving rise to binding energies which differ from those in solution. Hydrophobic interactions, such as those between the acyl CoA binding protein

and its fatty acid ligands, may not play a significant stabilizing role in the gas-phase since the strength of this kind of interaction is dependent on the solvent (Robinson *et al.*, 1996). Loo has suggested that ESI MS may be used to obtain information about the types of interaction that maintain complexes (Loo, 1997).

ESI Time-of-flight mass spectrometry

ESI is a continuous ionization source, and in most ESI experiments, quadropole mass analyzers are used. Quadropole mass spectrometers consist of four rod electrodes situated equidistant from each other, with opposing pairs possessing identical charge. Ions within a narrow m/z ratio are selected by adjusting the alternating current and direct current components of each pair of electrodes, allowing the transmission of the selected ions through to the detector. Those which are outside the m/z range selected strike the quadropole and are not detected (Hofstadler *et al.*, 1996).

In our work, TOF mass analysis was used in a mass spectrometer built at the Physics Department at the University of Manitoba (TOF III) (Chernushevich *et al.*, 1997, Verentchikov *et al.*, 1994). TOF is a pulsed technique for analysis of gas-phase ions, in which all components are accelerated to the same kinetic energy by applying voltage. Each ion with its unique m/z ratio acquires a characteristic velocity which depends on the square root of its m/z ratio, causing the different components to separate during the flight and arrive at the detector at different times. The flight time of a particular ion, therefore, depends on its mass (or m/z), with ions of low m/z arriving at the detector first (Willard *et al.*, 1988). The main advantages of TOF are the ability to detect ions of all masses simultaneously, and the ability to study a mass range which is limited only by the detector

(Chernushevich *et al.*, 1997). The difficulty of coupling a continuous ionization method (ESI) with a pulsed mass analyzer (TOF) was solved by arranging the two components at right angles to each other (Verentchikov *et al.*, 1994).

Technical considerations

The ESI experiment demands that buffer components, salts, and other additives be absent from the sample to be analyzed, since these components will be concentrated onto the analyte molecules during the ESI process and result in uninterpretable spectra. This is especially a problem if the concentrations of the additives are higher than the analyte concentration; in this case, the additives will be more easily observed and will swamp the peaks arising from the analyte. Samples therefore need to be desalted, and the buffering agent must be replaced with low concentrations of an agent that is volatile (such as ammonium bicarbonate, ammonium acetate, or imidazole).

These stringent conditions can present a problem for some systems, where buffer components or concentrations may be essential for stability and maintenance of noncovalent interactions. For example, some proteins may not be stable under low salt conditions; TrpR appears to be partially unfolded in 10 mM ammonium acetate or ammonium bicarbonate buffers which are compatible with electrospray (Potier *et al.*, 1997). Therefore careful choices of buffers must be made such that the native structure essential for function is preserved.

Objectives

The main purpose of this work was to apply ESI-TOF MS to study the quaternary structure of CS, and also CS binding to its allosteric inhibitor, NADH, and its substrates (AcCoA and OAA). As discussed in Chapter 1, the

allosteric regulation of CS has been suggested to involve its subunit structure, but no information has been obtained on this issue. The ESI MS technique was also applied to the routine analysis of CS, mutants of CS, and chemically modified CS.

Experimental

Preparation of citrate synthase for mass spectrometry

The *E. coli* CS plasmid containing the *gltA* gene was used for overexpression of the protein (Anderson & Duckworth, 1988) in *E. coli* host strain MOB154 (Wood *et al.*, 1983). Citrate synthase was purified from cell extracts using a two-step purification protocol involving ion exchange and gel filtration as described in Chapter 2. TFBA-modified CS was prepared and purified also as described in Chapter 2.

The final CS preparation is in 20 mM Tris, 1 mM EDTA, 50 mM KCl, pH 7.8, a buffer not suitable for ESI-TOF MS since Tris and salt concentrations exceed that of the protein, and would interfere with the electrospray process. Suitable buffers for electrospray such as ammonium bicarbonate and ammonium acetate are volatile, and thus do not interfere with the electrospray experiment. For the preparation of native CS for functional studies, an aliquot of CS (100 to 200 μL of 1 mg/mL) was washed repeatedly with 20 mM ammonium bicarbonate (Fisher) or 20 mM ammonium acetate (Aldrich) in a Centricon 30 unit of 30 K molecular weight cut-off (MWCO) (Amicon) or dialyzed exhaustively using the waterbug procedure (Orr, 1995) using regenerated cellulose membranes with 12 - 14 K MWCO (Spectrum). For denatured CS samples for mass determinations only, aliquots of CS were dialyzed against deionized water (three times against 4 L). The recovered samples usually contained precipitated protein which was dissolved by the

addition of acetic acid to a final concentration of 5%. Unless otherwise noted, protein concentrations for both functional studies and mass determinations were typically about 10 μM (0.48 mg/mL). The concentration of CS for the functional work was determined by spectrophotometry using the known molar extinction coefficient of 47 700 M^1 cm^{-1} at 278 nm. All chemicals were of reagent grade or better.

Preparation of ligands for mass spectrometry

The disodium salt of NADH (Sigma), and the lithium salt of AcCoA (Pharmacia Biotech) were dialyzed against 20 mM ammonium acetate or ammonium bicarbonate using a 500 MWCO cellulose ester membrane (Spectrum), also utilizing the waterbug procedure. Concentrations of NADH and AcCoA stock solutions were determined spectrophotometrically using their known extinction coefficients at 340 (6 220 $M^{-1}cm^{-1}$) and 260 nm (14 700 $M^{-1}cm^{-1}$), respectively. Oxaloacetate (Sigma) was available as the free acid, and was simply dissolved in the desired buffer without further purification.

Sample preparation for functional studies

A 100 μl sample of the desired combination of CS and ligand was prepared either one day prior to the MS measurements or, in some cases, just a few minutes before spraying. Unless specified, all samples were diluted such that the final buffer concentration was 5 mM and the CS concentration ~ 10 μM. Specific experiments that were carried out include: study of CS dimer-hexamer equilibrium by ESI-TOF MS at different CS concentrations, titration of CS (10 μM) with NADH to study NADH-CS complexes and the effect of NADH on dimer-hexamer equilibrium, and the observation of substrate binding and enzyme activity of CS.

ESI-TOF MS

The TOF mass spectrometer (TOF III) at the University of Manitoba Physics Department with electrospray ion source and orthogonal extraction was used for these experiments, and has been described elsewhere (Verentchikov *et al.*, 1994). Samples were continuously sprayed at 0.4 µl/min. The sample aerosol experiences a countercurrent of hot N_2 and three stages of pumping, with decreasing pressure at each stage for desolvation (Chernushevich *et al.*, 1997). At the second stage, a voltage difference is applied to aid in the declustering of aggregates (usually non-specific), and is usually adjusted such that a balance is found between proper desolvation and fragmentation of the ions. CS was stable at the maximum declusterng voltage (350 V), a value which resulted in good desolvation but no fragmentation of the protein or complexes. The instrument was calibrated using multiply charged ions of Substance P (Sigma) dissolved in a 1:1 mixture of water/methanol and 1% acetic acid.

Results

Denatured Wild Type Citrate Synthase

The results in this section illustrate the utility of ESI-TOF MS for routine molecular weight determinations. Figure 3.2 shows the mass-to-charge and the corresponding deconvoluted mass spectra of WT CS in 5% acetic acid. Under these conditions (pH ~ 2.5) the protein is expected to be denatured, and the results obtained by ESI-TOF MS support this. The m/z spectrum shows a charge envelope that is wide, containing peaks that correspond with ions that are highly charged (+30 to +60). A wide charge envelope of highly-charged ions are diagnostic of denatured protein, since unfolded proteins have a large surface area and exist in many interconverting conformations; a larger number of charges are acquired by the protein than if

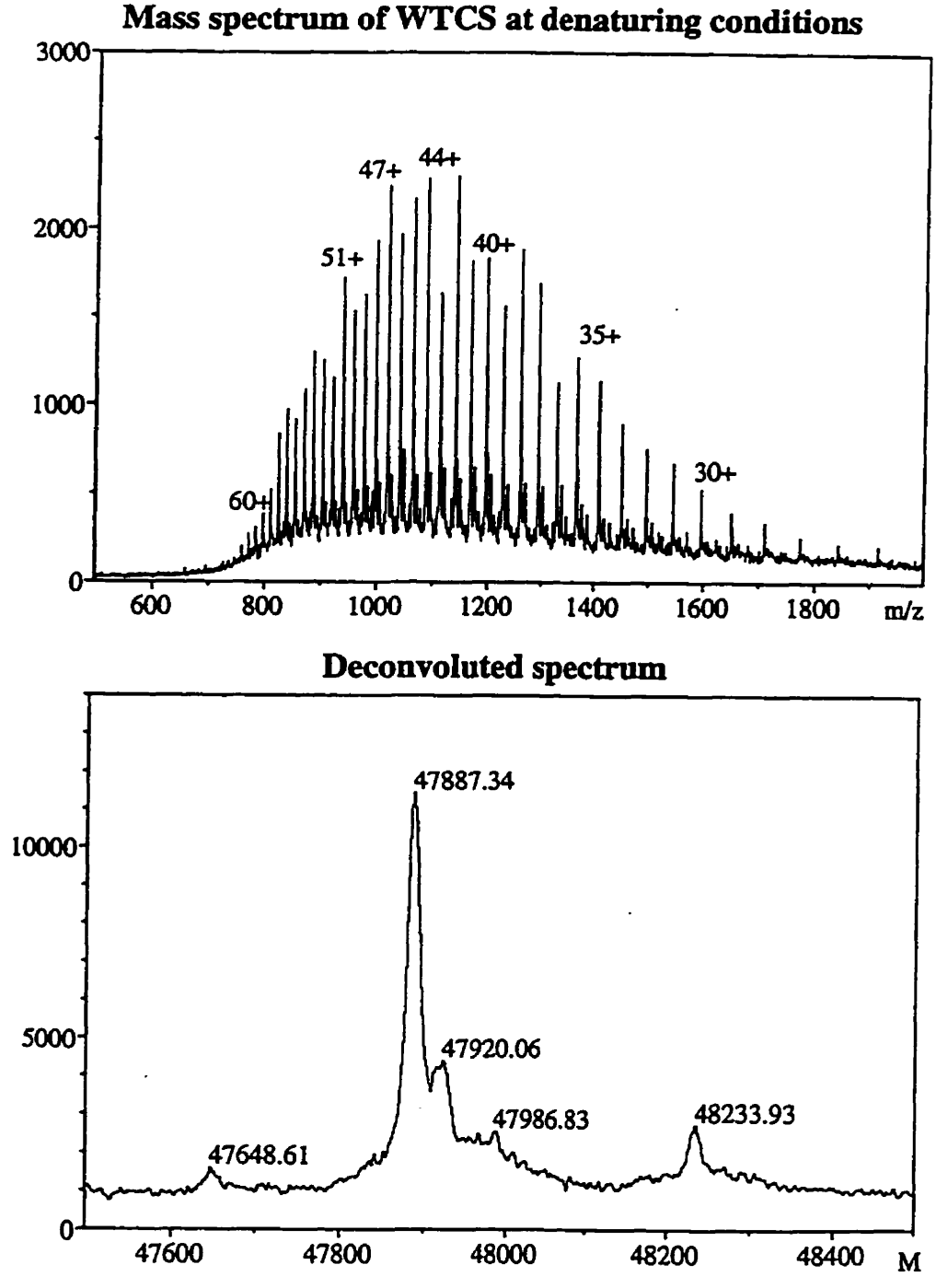


FIGURE 3.2 ESI-TOF m/z and corresponding mass spectra of denatured CS in 5% acetic acid. The multiple charging obtained in the m/z spectrum here is typical of denatured proteins, and seems to depend on the available surface area and the number of basic residues in the molecule. Deconvolution yields a prominent peak for CS with a molecular mass that agrees well with the calculated value. The adduct peak at M=48 233.9 is due to a frequently-encountered impurity introduced by the mass spectrometer.

it was compact, and the conformational heterogeneity leads to a large set of ions of different charges.

The deconvoluted spectrum shows a main peak with a mass of 47 887, as well as two adduct peaks: one with an additional 39 u (likely a potassium adduct) and another with 346 u (an impurity introduced by the quadropole of the instrument). The mass measured here is in good agreement with the calculated mass of CS based on the gene sequence (47 885 u) (Ner *et al.*, 1983) using chemical atomic weights, with the correction of F288 to valine (Anderson & Duckworth, 1988), omission of the C-terminal methionine, and replacement of N10 by aspartate as found by N-terminal sequencing of the protein (Duckworth & Bell, 1982).

Molecular Weight Determinations of CS mutants

Some mutant CS preparations were checked by ESI-TOF MS to confirm the loss of mass due to the mutation. Two examples were chosen for illustration here, and are shown in Figure 3.3: CS L326A and CS V338A. As seen with WT CS, the m/z spectra of these mutants were diagnostic of unfolded protein, exhibiting a wide charge envelope, as well as ions with a large number of charges. The molecular weights obtained for the two mutants and others are listed in Table 3.1, with the expected molecular masses and molecular mass change expected due to the mutation. In most cases, the change in molecular weight observed was within 2 u from that expected.

Confirmation of the alkylation of C206 with TFBA

Cysteine 206 can be easily alkylated, as was previously reported using TFBA for ^{19}F-NMR (Donald *et al.*, 1991). While other methods have been used to confirm the alkylation of this residue such as DTNB reactivity, no information was available as to whether other sites have been modified, nor

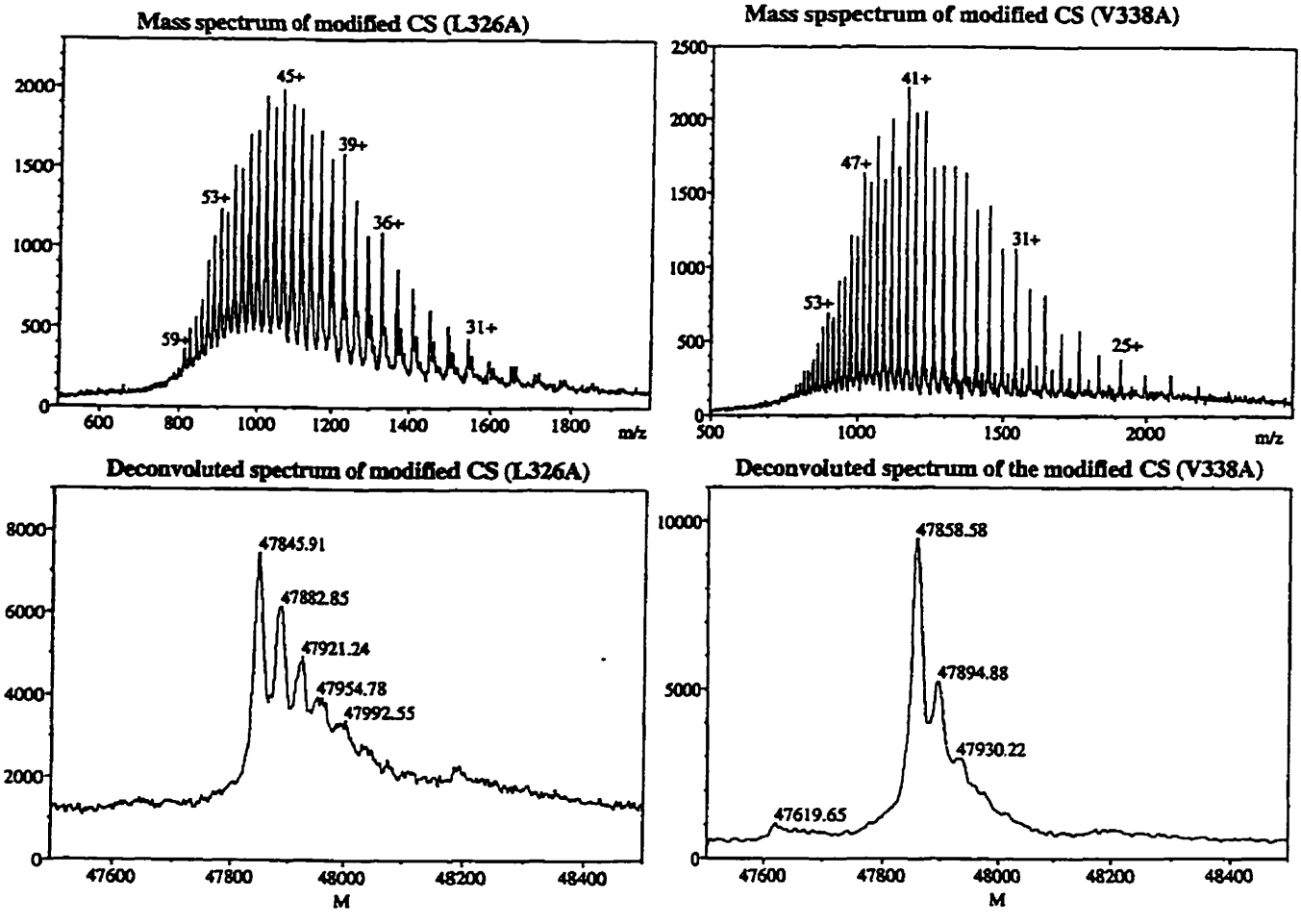


Figure 3.3 ESI-TOF m/z and mass spectra of CS L326A (right) and CS V338A (left), in 5% acetic acid. Protein concentration was approximately 10 μM. Adduct peaks are due to potassium (39 u) which may be present in the sample or introduced during electrospray. The changes in molecular weights from the wild type value observed here are consistent with the loss of mass due to the mutations.

TABLE 3.1 ESI/TOF Determination of Molecular Masses of Citrate Synthases

Sample	Theoretical Mass	Measured Mass	Expected MW Change•	Observed MW Change••
Wild Type	47 885	47 887	-	-
CS D109A*	47 841	47 845	44	42
CS L164A	47 843	47 846	42	40
CS L259A	47 843	47 843	42	44
CS L326A	47 843	47846	42	41
CS V338A	47 857	47 858	28	29
CS-TFBA§	47 995	48 003	110	116

•Based on the theoretical molecular weights of the WT and mutants.
••Based on the measured molecular weights of the WT and mutants.
*** CS D109A was prepared by Jennifer Hacking.**
§CS modified using TFBA at cysteine 206.

whether alkylation was complete. The ESI-TOF mass spectrum in Figure 3.4 shows that the molecular weight of the modified CS has increased by 116 u (Table 3.1), 6 u larger than the expected change of 110 u, and thus consistent with a singly modified CS. However the tailing of the peak may indicate that a small population of CS is being modified at an additional site, or that desolvation was not optimal, thus leading to a broader mass peak. Spectra of the modified CS under non-denaturing conditions, as shown in the next section also support the finding that mainly one site is modified and that the alkylation is essentially complete.

Subunit Structure

ESI-TOF mass spectra of CS under non-denaturing conditions were collected at two different pH values: at pH 6 (in 5 mM ammonium acetate) and at pH 7.5 (in 5 mM ammonium bicarbonate), shown in two typical experiments in Figure 3.5. Both spectra exhibit two m/z envelopes centered around 6 000 and 9 000. Deconvolution into mass spectra for ammonium bicrabonate show that the two charge envelopes correspond to dimeric CS (MW 95 770 ± 10) and hexameric CS (MW 287 322 ± 30), in good agreement with the expected masses of 95 770 and 287 310, for dimer and hexamer respectively. The mass spectrum at pH 7.5 (5 mM ammonium bicarbonate) also contains a small amount of tetrameric CS. Its absence from the spectrum at pH 6 may indicate that it is an artifact perhaps due to the electrospray process and arises from the nonspecific association of dimers. Nevertheless, it seems that the predominant species are dimers and hexamers under both buffer conditions. Previous ultracentrifugation data under similar conditions also show the presence of dimers and hexamers as the main species under some conditions (Tong & Duckworth, 1975).

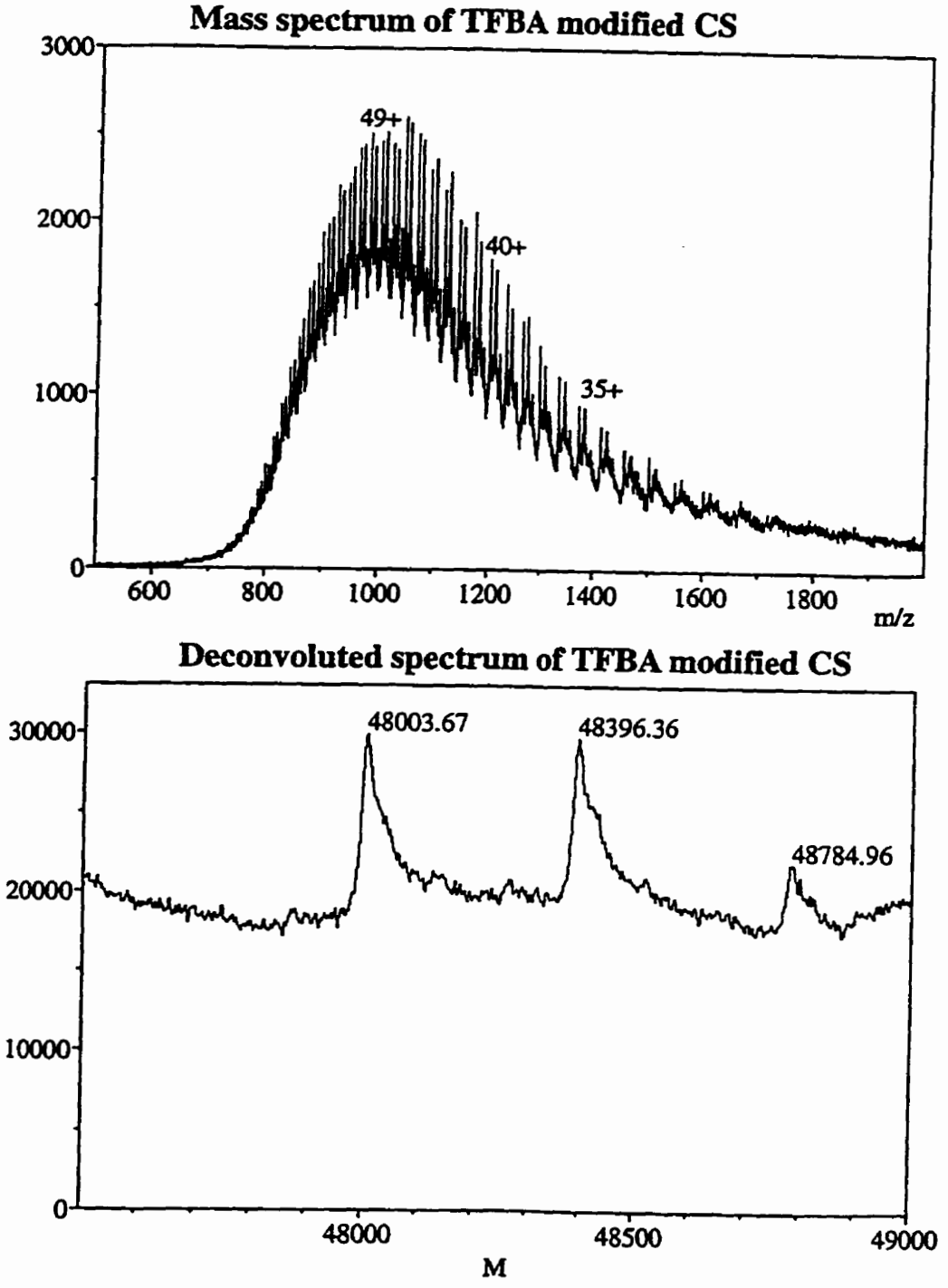


Figure 3.4 ESI-TOF m/z and mass spectra of TFBA-modified CS in 5% acetic acid. The mass increase is consistent with the addition of one TFBA group (110 u) to each CS subunit. Adducts are due to an impurity introduced in the TOF instrument.

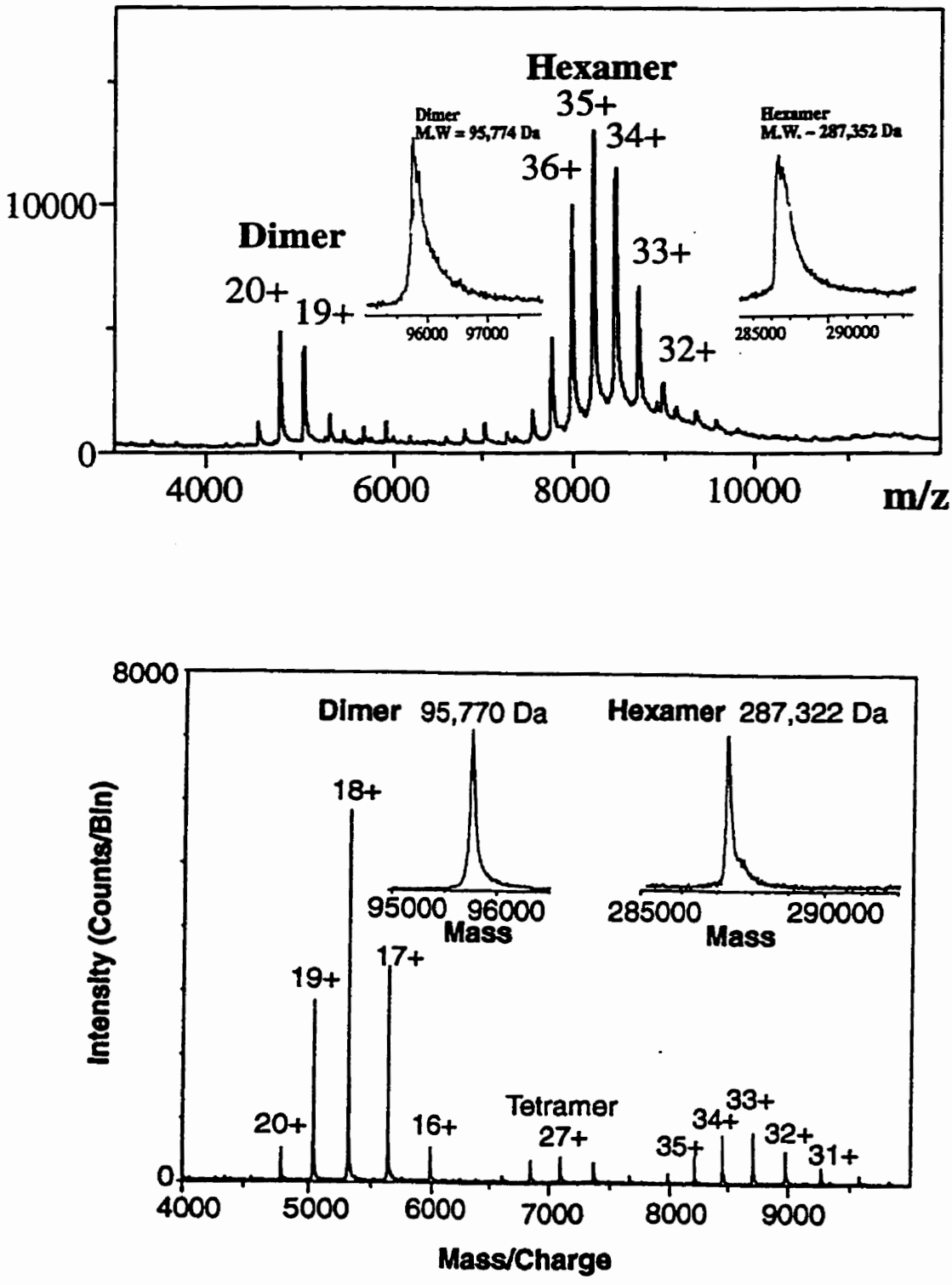


Figure 3.5 ESI-TOF MS spectra of native CS in 5 mM ammonium acetate, pH 6 (top) and 5 mM ammonium bicarbonate, pH 7.5 (bottom). CS subunit concentration was 9 and 10 μM, respectively. Deconvolutions are shown as insets.

A further observation may be made regarding the relative amounts of dimers and hexamers under these conditions; at pH 7.5, the amount of dimer is consistently larger than at pH 6, as quantified by integration of the relevant peaks in the m/z or mass spectra. The equilibrium between hexamers and dimers has been shown to be pH dependent by ultracentrifugation with a similar trend; lower pH values resulted in increased amount of hexamers, consistent with the ESI-TOF MS results here.

CS spectra were also acquired at different protein concentrations, ranging from 4 μM to 100 μM (subunit concentration). In addition to demonstrating the ability of the intstrument to measure mass spectra in a wide range of protein concentrations, the spectra in Figure 3.6 show a change in the relative amounts of dimer and hexamer that is consistent with a simple equilibrium between the two species. A more quantitative assessment of this data is presented in a later section, allowing the determination of the association constant for dimer to hexamer under the conditions used here.

TFBA-modified CS was also studied by MS, and the spectra in Figure 3.7 show that no hexamer is formed under these conditions, a previously unknown effect of the modification. The deconvoluted spectrum shows that the majority of the dimer is doubly modified; small amounts of unmodified and singly modified dimer can been seen in the rather broad dimer peak. An independent method, non-denaturing gel electrophoresis, discussed in Chapter 2 also showed TFBA-modified CS to be mainly dimeric in 20 mM Tris, 1 mM EDTA, pH 7.8. However in the presence of 50 mM KCl, the modified CS elutes on size exclusion columns (Sepharose 6B, Sephadex G200) as a hexamer, indicating that the KCl still shifts the equilibrium to hexamer, as is the case with unmodified CS.

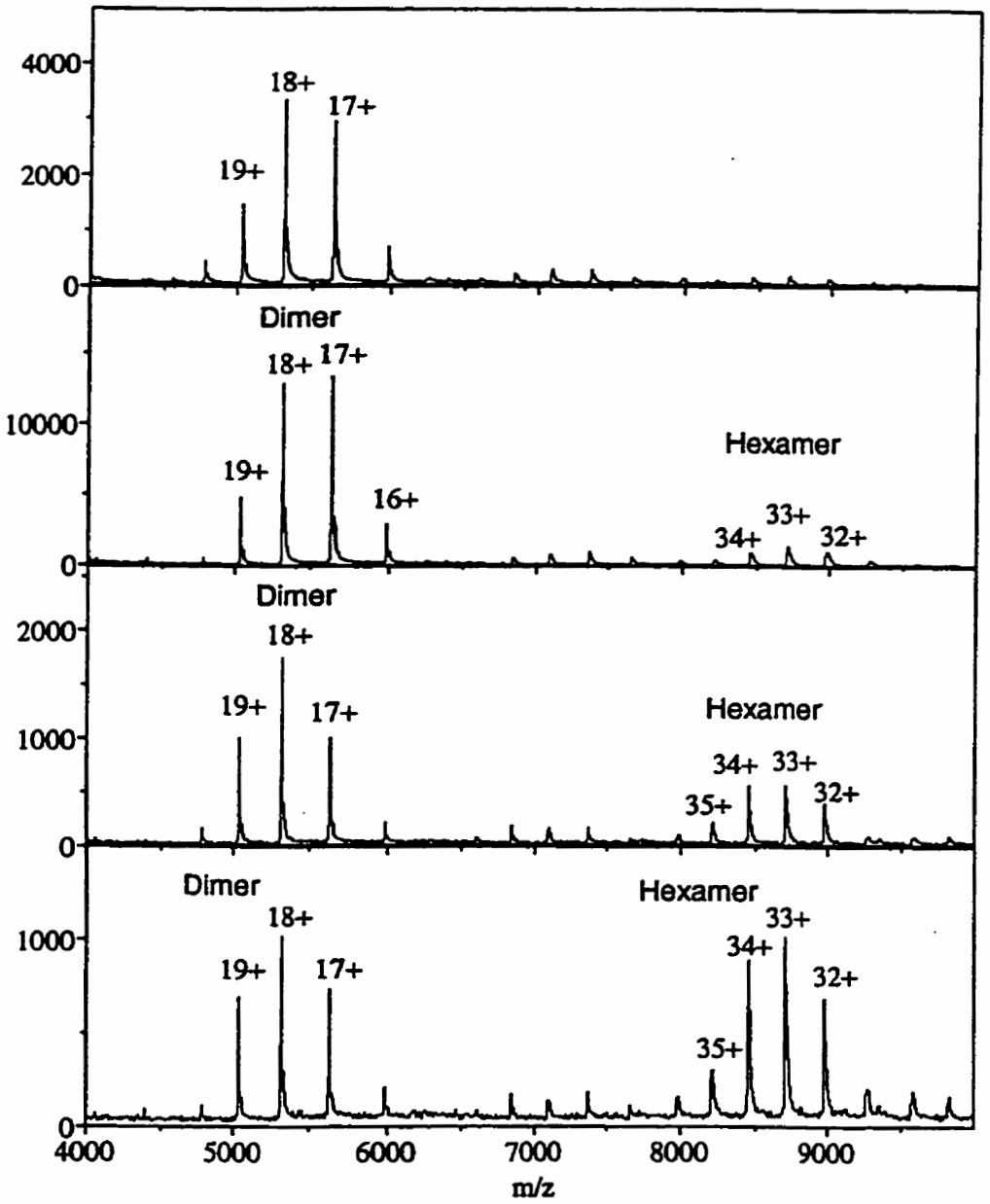


Figure 3.6 Spectra of native CS at different concentrations, in 5 mM ammonium bicarbonate, pH 7.5. The concentrations of CS subunit from top to bottom were 4 μM, 10 μM, 30 μM and 100 μM. The change in the relative amounts of dimers and hexamers is consistent with a simple equilibrium between them, and is quantitatively described in the next section.

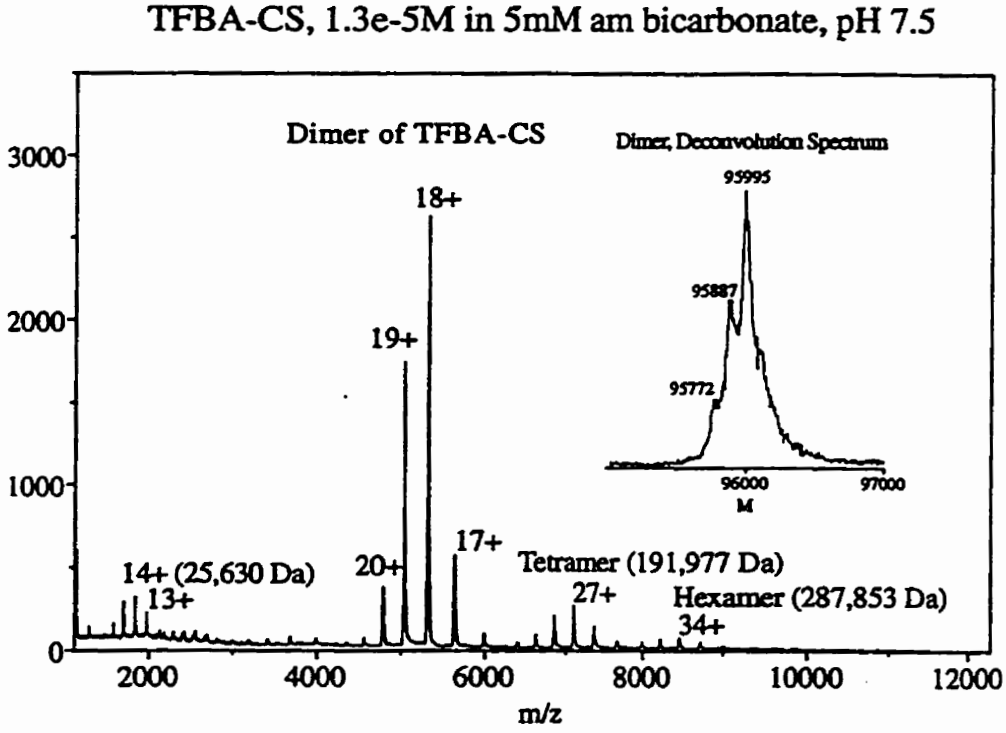


Figure 3.7 ESI-TOF MS spectrum of TFBA-CS in 5 mM ammonium bicarbonate, pH 7.5.

Effect of NADH on CS

Electrospray ionization spectra of CS in the presence of increasing amounts of NADH, the allosteric inhibitor of CS, show several interesting features (Figure 3.8). First, NADH caused a change in the relative amounts of dimers and hexamers, shifting the equilibrium towards hexamer with increasing NADH concentration. Under the conditions used to acquire the spectra in Figure 3.8 (5 mM ammonium bicarbonate pH 7.5), about 100 μM NADH was needed to completely shift all of the CS to the hexameric form. At pH 6 (5 mM ammonium acetate) only 10 μM NADH was required to shift the CS completely to hexamer as shown in Figure 3.9. Second, only NADH-CS hexamer complexes appear at low NADH concentrations, with the various ligated hexameric species fully resolved in the deconvoluted mass spectra. Dimeric CS also binds hexamer, but only at much higher NADH concentrations, as seen in (C) of Figure 3.8 at 18 μM NADH. Third, the distribution of the relative populations of the ligated species differ between dimer and hexamer; for example at 4.5 μM NADH, the singly and doubly ligated hexamer peaks are present in larger amounts than the unligated hexamer, whereas for dimer at 18 μM NADH, the singly and doubly ligated dimer populations are considerably smaller than the unligated dimer. The last two features discussed above indicate, albeit qualitatively, that NADH binding to hexamer is tighter and more specific than that to dimer. A quantitative treatment of the MS data presented above is discussed in a later section, but necessitates the assessment of the sensitivity difference of the ESI-TOF instrument to molar concentrations of dimer and hexamer.

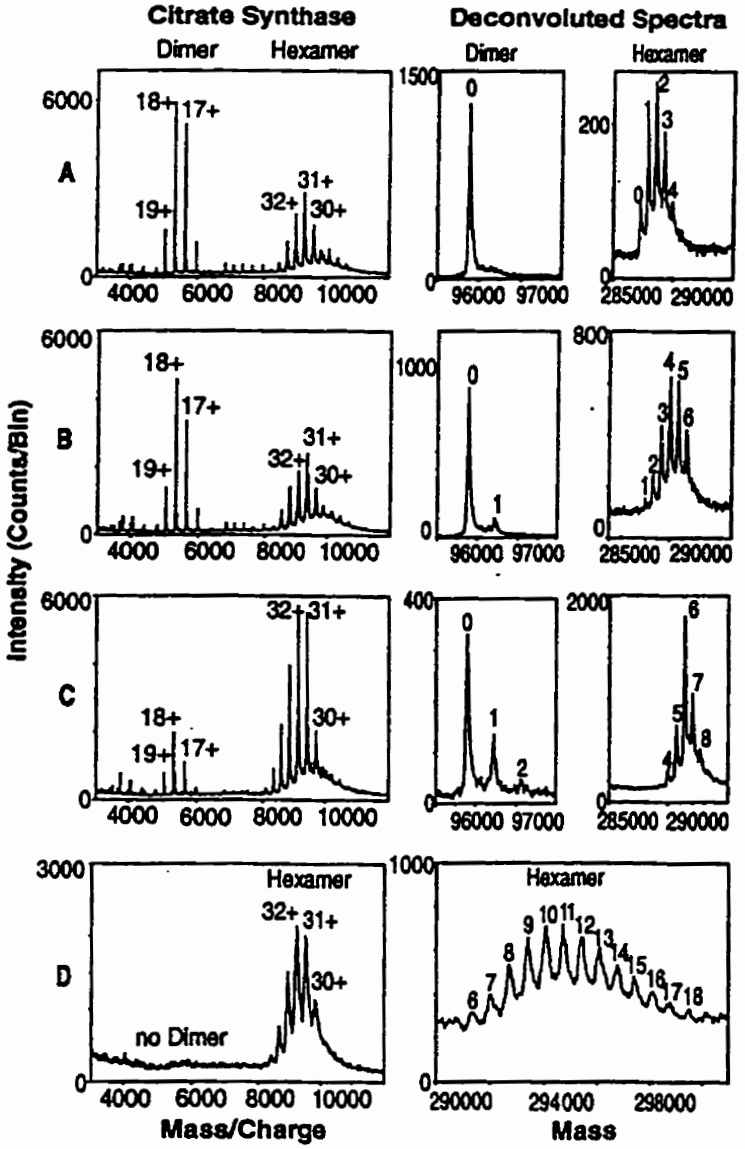


Figure 3.8 Selected ESI-TOF spectra of CS in the presence of increasing amounts of NADH, the CS allosteric inhibitor at 4.5 μM (A), 9.0 μM (B), 18 μM (C), and 108 μM (D) in 5 mM ammonium bicarbonate pH 7.5. The digits labelling the deconvoluted spectra correspond to the number of NADH molecules bound. A quantitative assessment of the binding to both CS oligomeric forms is provided in a later section.

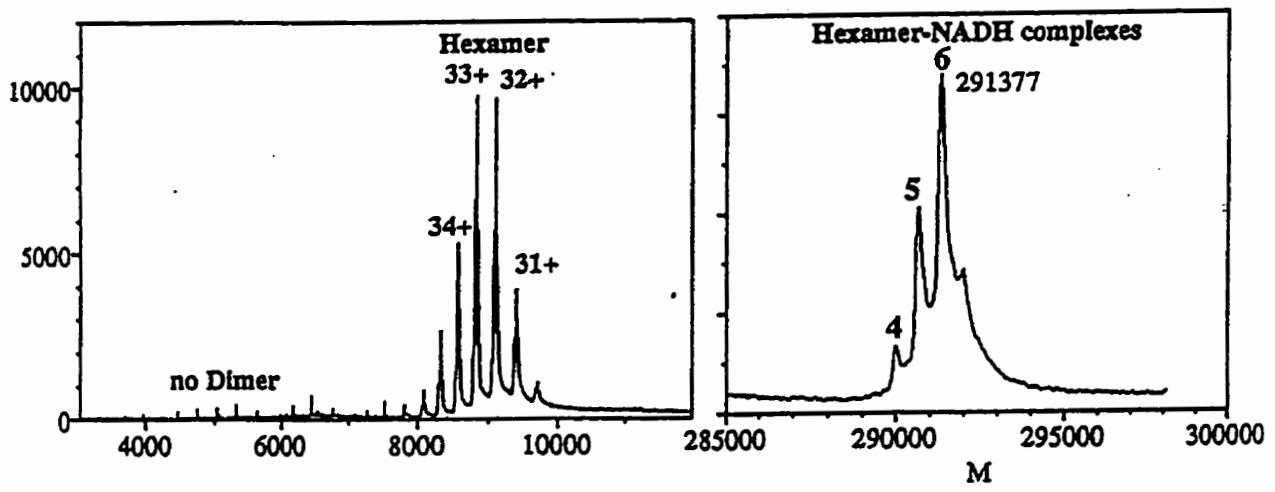


Figure 3.9 ESI-TOF mass spectrum of CS in 5 mM ammonium acetate, pH 6, with 10 μM NADH. The entire CS population is shifted to hexamer by the NADH.

Sensitivity of the ESI-TOF Instrument to Amounts of Dimer and Hexamer

The assessment of the sensitivity of the instrument to relative molar quantities of dimer and hexamer was necessary if meaningful quantitative values, such as the association constant for the dimer-hexamer equilibrium, or the dissociation constant for NADH binding to hexamer or dimer, are to be calculated using the MS data presented thus far.

The instrument was calibrated using a sample of CS in which the amounts of hexamer and dimer were known. The sample used was prepared by using equimolar amounts of the TFBA-modified CS (pure dimer), and the unmodified CS in the presence of excess NADH (pure hexamer). TFBA CS is not subject to hexamer formation in the presence of NADH, and therefore the amount of dimer and hexamer in the sample described is known, and can be used to assess the sensitivity of the instrument.

The results of this experiment are shown in Figure 3.10. Integration of the mass-to-charge or mass peaks corresponding to each species and calculation of the D/H peak area ratio yielded 1.3 ± 0.1 for equimolar dimer and hexamer concentrations. This indicates that the instrument is 1.3 times more likely to detect dimer than hexamer given equal amounts of the two species. This factor was used in subsequent calculations of dimer/hexamer molar ratios based on the ratios of dimer/hexamer peak areas obtained for the NADH binding experiment (Figure 3.8) as well as the mass spectra of CS at different concentrations (Figure 3.6) for the determination of K_A.

Derivation of equations for data fitting to obtain K_A and K_D

The peaks of the mass spectra (usually m/z spectra, although mass spectra may be used also) were integrated to obtain their area, and thus the relative amount of each species. This was performed for the entire charge

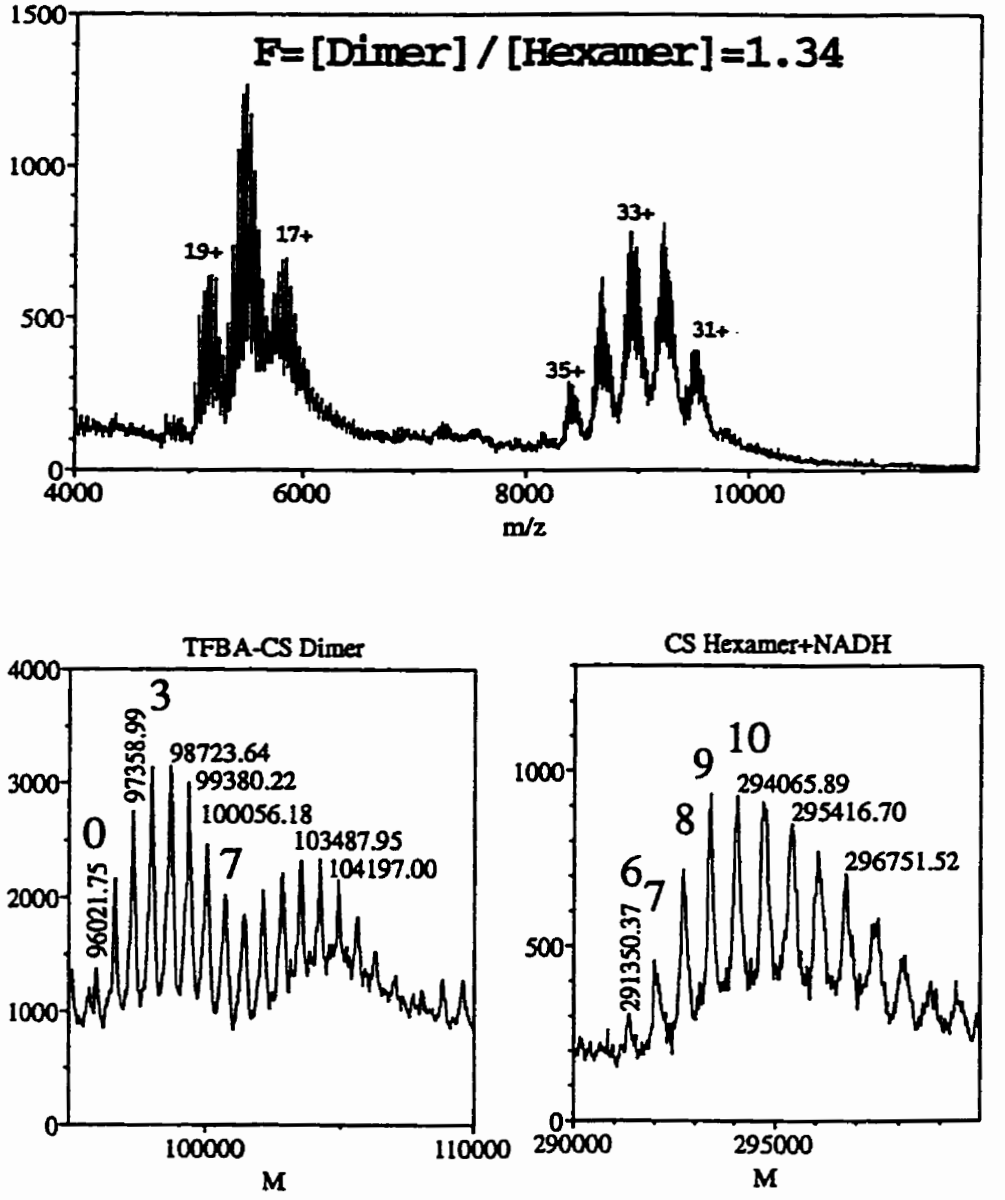


Figure 3.10 ESI-TOF mass spectrum of equimolar concentrations of dimeric CS (TFBA modified CS) and hexameric CS (unmodified CS in the presence of 100 μM NADH) in 5 mM ammonium bicarbonate. The dimer-to-hexamer ratio was 1.3, a slight mass discrimination against hexamer. The digits labelling the deconvoluted spectra correspond to the number of bound NADH. At such a high NADH concentration, TFBA modified CS binds, nonspecifically, a large number of NADH.

envelope for each species, and the ratio of the peak areas corresponding to dimers and hexamers (***D/H***) was calculated and corrected with the determined sensitivity ratio (1.3 for ***D/H***) discussed above. The corrected D/H ratio was taken to be a measure of the molar ratio, R, of the dimer/hexamer in solution

$$R = \frac{[D]}{[H]} \qquad [4]$$

For the dependence of ***[D]/[H]*** on the total CS concentration, an equation was derived, relating the protein concentration, the experimentally determined ***[D]/[H]***, and K_A, the association constant for formation of hexamers from dimers. If the system is described simply as an equilibrium between dimers and hexamers

$$3D \underset{}{\overset{K_A}{\rightleftharpoons}} H$$

then the equilibrium association constant, K_A, is defined as

$$K_A = \frac{[H]}{[D]^3} \qquad [5]$$

The concentrations of dimers and hexamers are related to the total CS subunit concentration by

$$[CS]_{subunit} = 2[D] + 6[H] \qquad [6]$$

Using equations 6 and 4 and rearranging, the total protein concentration can be expressed in terms of hexamer

$$[H] = \frac{[CS]}{(2R+6)} \qquad [7]$$

K_A may be redefined as follows

$$K_A = \frac{[H]^3}{[D]^3[H]^2} \qquad [8]$$

and from equation 4,

$$R^3 = \frac{[D]^3}{[H]^3} \qquad [9]$$

Substituting 9 into 8 yields,

$$K_A = \frac{1}{R^3[H]^2} \qquad [10]$$

Substituting for ***[H]*** using equation 7 into equation 10 and rearranging to solve in terms of CS concentration

$$[CS] = \frac{2R+6}{\sqrt{K_A}\sqrt{R^3}} \qquad [11]$$

Since it would be rather inconvenient to express equation 11 in terms of R, the axes were reversed to obtain a plot of ***[CS]*** versus R and the data fitted to equation 11 to obtain K_A. The data and the fit however, are presented conventionally in Figure 3.11 (after the fit and reversal of the axes).

For the NADH binding experiments, the free NADH concentration, $[NADH]_{free}$, was needed for a plot of binding occupancy (y-axis) versus $[NADH]_{free}$ (x-axis) to obtain K_D for binding to hexamer and dimer. The average number of bound NADH molecules per subunit for dimer and hexamer at each NADH concentration were determined from the deconvoluted mass spectra. The average molecular mass was calculated for each species with its bound NADH over the collection of peaks representing all the complexes with bound NADH molecules (MW_{avg}). This was also performed on the spectrum in the absence of NADH, yielding the molecular weights of the dimer and hexamer in the absence of bound NADH (MW_0). Then using these values obtained for the dimer and hexamer for each spectrum and the molecular weight of NADH (665 u), the average number of bound NADH molecules (B) per subunit, N (2 for dimer, 6 for hexamer) can be calculated by

$$B = \frac{\left(MW_{avg} - MW_0\right)/665u}{N} \qquad [12]$$

Using equations [4] and [6], the molar concentrations of dimers and hexamers may be obtained by

$$[H] = \frac{[CS]}{2R+6} \qquad [13]$$

$$[D] = [CS]\left(\frac{R+2}{R+3}\right) \qquad [14]$$

Then the total bound NADH concentration is subtracted from the known total concentration of NADH to yield the free NADH concentration,

$$[NADH]_{free} = [NADH]_{total} - (B_{dimer}[D] + B_{hexamer}[H]) \qquad [15]$$

Each spectrum in the NADH binding series was treated in this manner and plots of average number of bound NADH for dimer and hexamer (i.e. binding occupancy) versus free NADH concentration were generated. The dissociation constants, K_D, for NADH binding to dimer and hexamer were obtained by fitting (Kaleidagraph package) the data generated above to

$$B = \frac{B_{max} \cdot K_D}{K_D + [NADH]_{Free}} \qquad [16]$$

Where B is the average number of bound NADH molecules per subunit (the y-axis), B_{max} is the maximum occupancy (it would be one if one site per monomer exists), K_D is the dissociation constant, and $[NADH]_{free}$ is the free NADH concentration. For fitting the data for dimer, B_{max} was fixed at 1. For Hexamer, two terms were used to obtain K_D values for specific binding and non-specific binding.

Determination of K_A for Dimer-Hexamer equilibrium

Treatment of the data in Figure 3.6 (dependence of CS concentration on the dimer/hexamer ratio) as described in the preceding section yields the plot

shown in Figure 3.11, where the dimer/hexamer ratio decreases with increasing CS concentration. The extent of association, and therefore amount of hexamer, increases with increasing protein concentration, consistent with an equilibrium between dimer and hexamer. Fitting the data to equation 11, yields a K_A of (6.9 ± 0.7) X 10^{10} M^{-2}. No data under the same conditions are available from previous ultracentrifugation work but at 50 mM KCl, 20 mM Tris, pH 7.8, the K_A obtained was 4.1 X 10^{11} M^{-2}, a slightly larger value than that calculated from the MS data here at somewhat lower pH values and in the absence of salt. KCl is an allosteric activator of CS, and is known to induce hexamerization, and as such, an estimate of K_A in electrospray conditions based on the ultracentrifuge data is not feasible.

Determination of K_D for NADH binding to Dimer and Hexamer

Increasing concentrations of NADH affected the equilibrium between dimer and hexamer, as shown in Figure 3.12. Hexamer formation is favored in the presence of NADH, and the dimer/hexamer ratio seems to level out to about 0.02 at ~30 μM NADH. However, an additional effect of NADH is the apparent higher specificity of binding to the hexameric species, as seen in Figure 3.8 (mass spectra) and quantitatively described in Figure 3.13 (binding curves). In these binding curves, the differences in binding specificity are clarified. The hexamer binds NADH at low concentrations, filling one site per subunit by about 10 μM NADH non-cooperatively. At higher concentrations, bound NADH continues to increase, but gradually. The dimer, on the other hand, exhibits a gradual linear increase in the number of bound molecules throughout the entire concentration range, never reaching one site per subunit. The finding of gradual increases in the number of bound NADH to

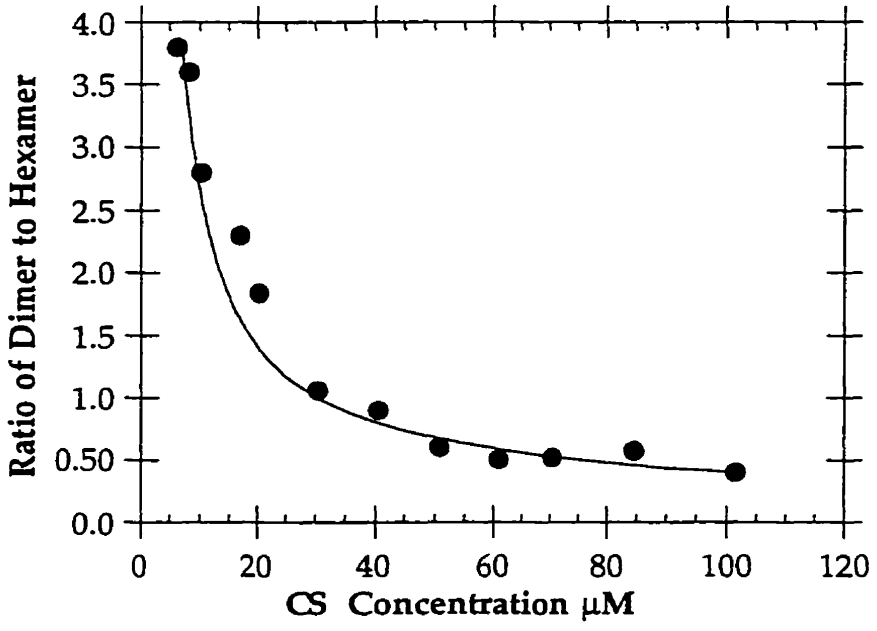


Figure 3.11 Dependence of dimer/hexamer molar concentration ratio, [D]/[H] on CS subunit concentration. The molar ratios were obtained from the ratio of integrated peak areas which were corrected for the sensitivity difference of the instrument for dimer and hexamer. The curve is a fit to the data assuming that the [D]/[H] ratio is described by the equilibrium 3D<->H, with KA=[H]/[D]^3 = (6.9 ± 0.7) X 10^{10} M^{-2}.

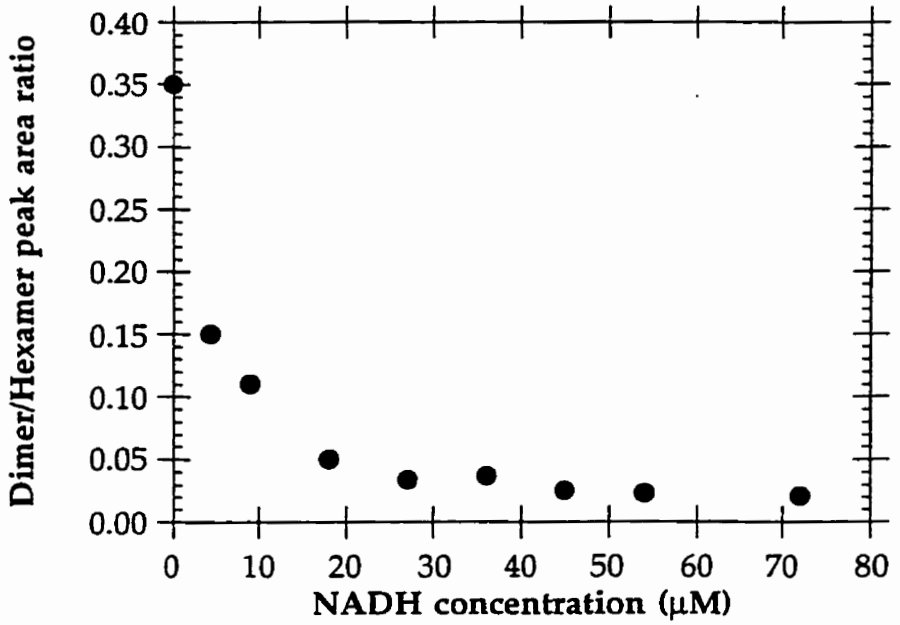


Figure 3.12 The effect of NADH concentration on the molar dimer-hexamer ratio ([D]/[H]).

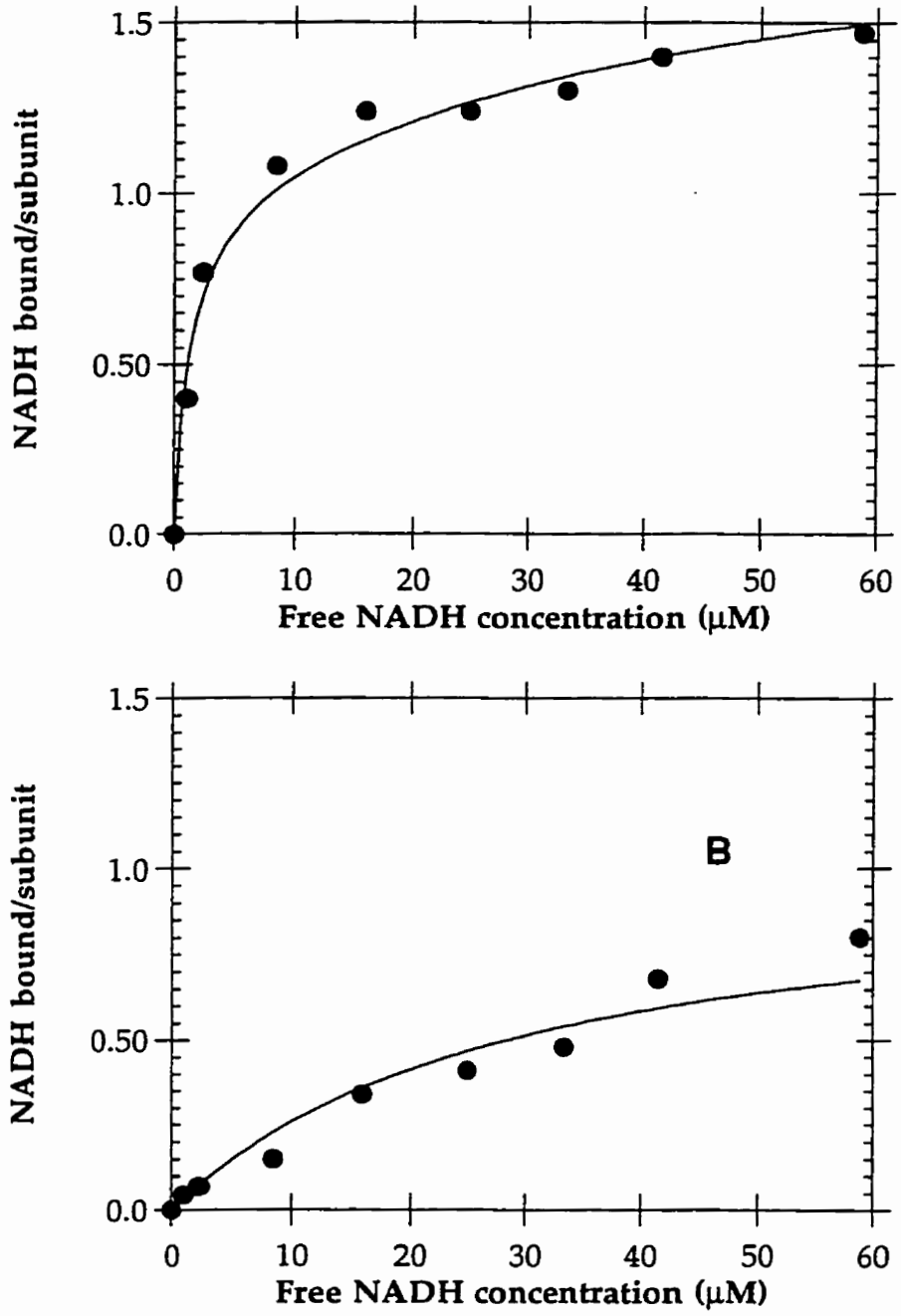


Figure 3.13 NADH binding to hexameric (A) and dimeric (B) CS in 5 mM ammonium bicarbonate, pH 7.5. The lines are fits to the data using equation 16 to obtain K_D values for binding. For NADH binding to hexamer, two terms were used to obtain two K_D values for tight and nonspecific binding.

both oligomeric CS forms may suggest that non-specific binding occurs to both dimers and hexamers.

Fitting of the data to equation 16, as described, affords the determination of K_D for binding to both species. To account for both specific and non-specific binding to hexamer, two terms were used to obtain K_D values for the two classes of binding. Hexamer binds NADH specifically with a K_D of 1.1 ± 0.2 μM, and non-specifically with a K_D 55±3 μM. On the other hand, the dimer binds NADH with a K_D of 28 ± 3μM. The K_D value obtained for CS-NADH binding from fluorescence techniques under similar conditions (20 mM Tris, pH 7.8) was 1.1 μM, in excellent agreement with that obtained here for the hexamer (Duckworth & Tong, 1976).

Study of CS-Substrate Complexes

In attempts to study the effect of NADH on CS-substrate complex formation involving mixtures of CS, NADH, OAA, and AcCoA, it was discovered that the number of possible ligand-CS complexes was far too great to unambiguously determine the exact compositions of the complexes obtained. This was in part due to poor resolution in such spectra. In one example at pH 6 (in ammonium acetate), shown in Figure 3.14, a complex of hexamer apparently with 6 NADH and 4 AcCoA molecules was observed, but poor resolution precluded unambiguous identification of the exact components of this complex. The results here warranted a simplified approach.

To this end, the study of CS with its substrates OAA and AcCoA was undertaken at pH 6 and pH 7.5. At pH 6, a spectrum was obtained (Figure 3.15) where both dimer and hexamer formed complexes with a species of about 800 u, consistent with binding of AcCoA. The ligation pattern is similar

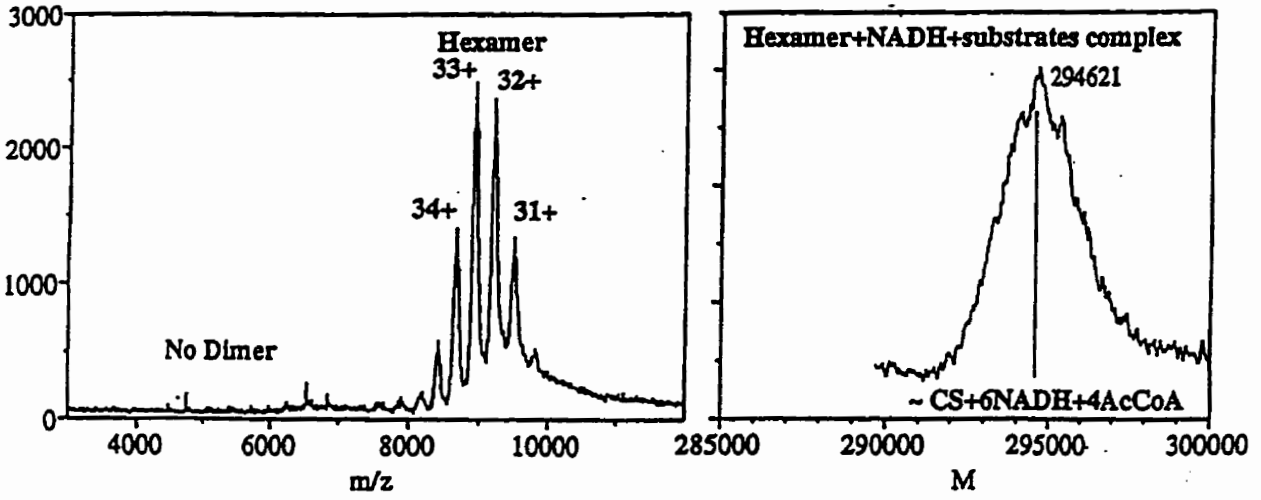


Figure 3.14 ESI-TOF mass spectrum of CS in the presence of equimolar concentrations of NADH, OAA, and AcCoA in ammonium acetate, pH 6. The weight of the complex obtained (294 621) corresponds roughly to a hexamer with 6 and 4 molecules of NADH and AcCoA, respectively. No dimer was observed, consistent with the effect of NADH previously discussed.

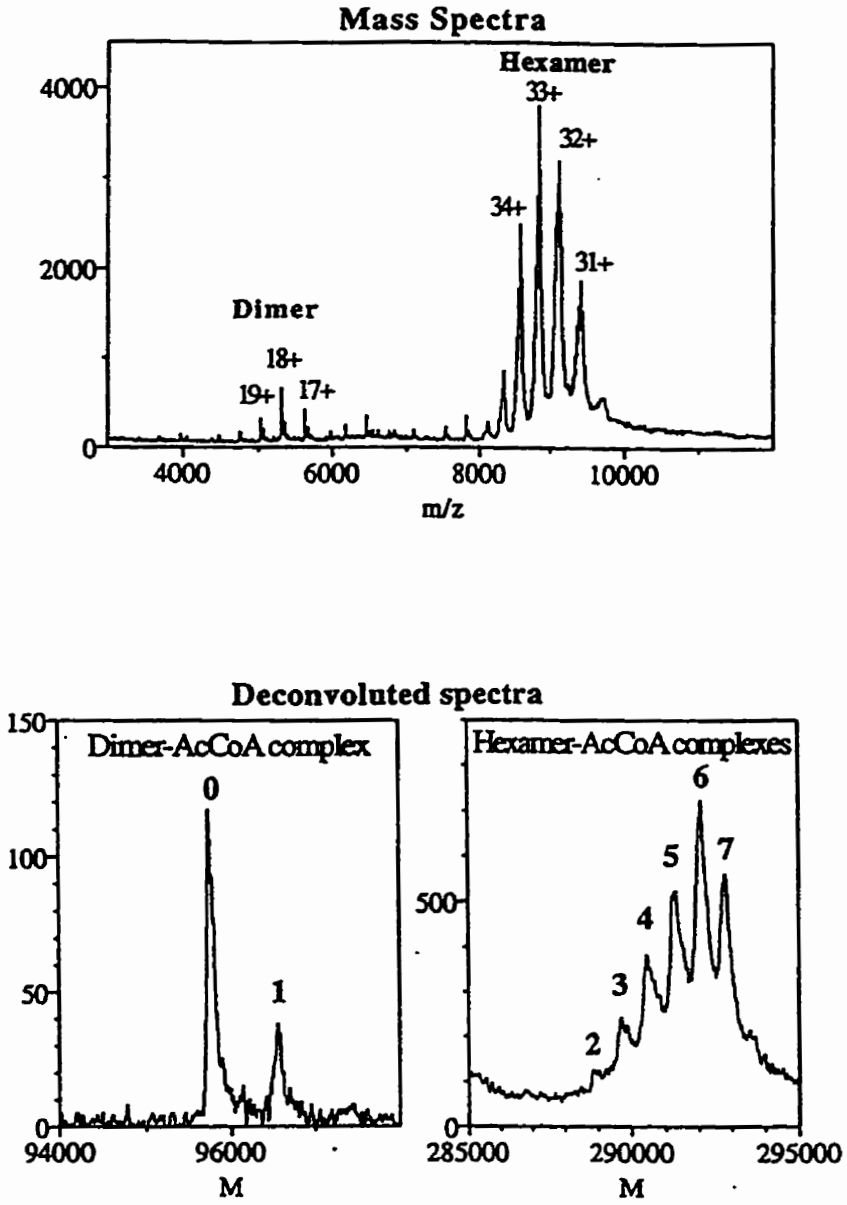


Figure 3.15 Spectrum of CS in the presence of equimolar concentrations of the substrates OAA and AcCoA in 5 mM ammonium acetate, pH 6.

to that obtained with NADH; the hexamer peak is mostly ligated, whereas the dimer is not. This suggests that binding to hexamer is more specific. Upon titration of CS with increasing amounts of substrates, the equilibrium between dimer and hexamer shifted towards the hexameric form, as observed with the addition of NADH (data not shown).

The shift in molecular mass of the hexamer and dimer is consistent with a ligand of 800 u. AcCoA and CoA have masses of 810 and 768, respectively, but owing to broadened peaks the masses obtained from the deconvoluted spectra could not be used to clearly identify which species were bound. The decreased sharpness of the peaks may be due to the presence of complexes with both species, as well as with OAA and citrate.

At pH 7.5, the spectrum acquired soon after mixing the components was too complex (not shown), since CS was more active at this pH and produced a complex mixture of substrates and products, unlike the situation at pH 6. To alleviate this complexity, the mixture was allowed to incubate for about two hours to convert all the substrates to products (citrate and CoA). The resultant spectrum, shown in Figure 3.16, revealed dimer and hexamer complexed with the products; for dimer, a set of peaks corresponding to CoA complexes was present, with two additional sets corresponding to one and two citrate molecules bound; hexamer appeared to exhibit a complex pattern of peaks corresponding to the initial binding of citrate then followed by CoA and an additional citrate. No unligated hexamer was found, unlike the dimer.

Examination of the low m/z region of the two substrate binding experiments at pH 6 and 7.5 revealed the extent of substrate formation; at pH 6, the ratio of AcCoA to CoA was large, even after incubation for two hours

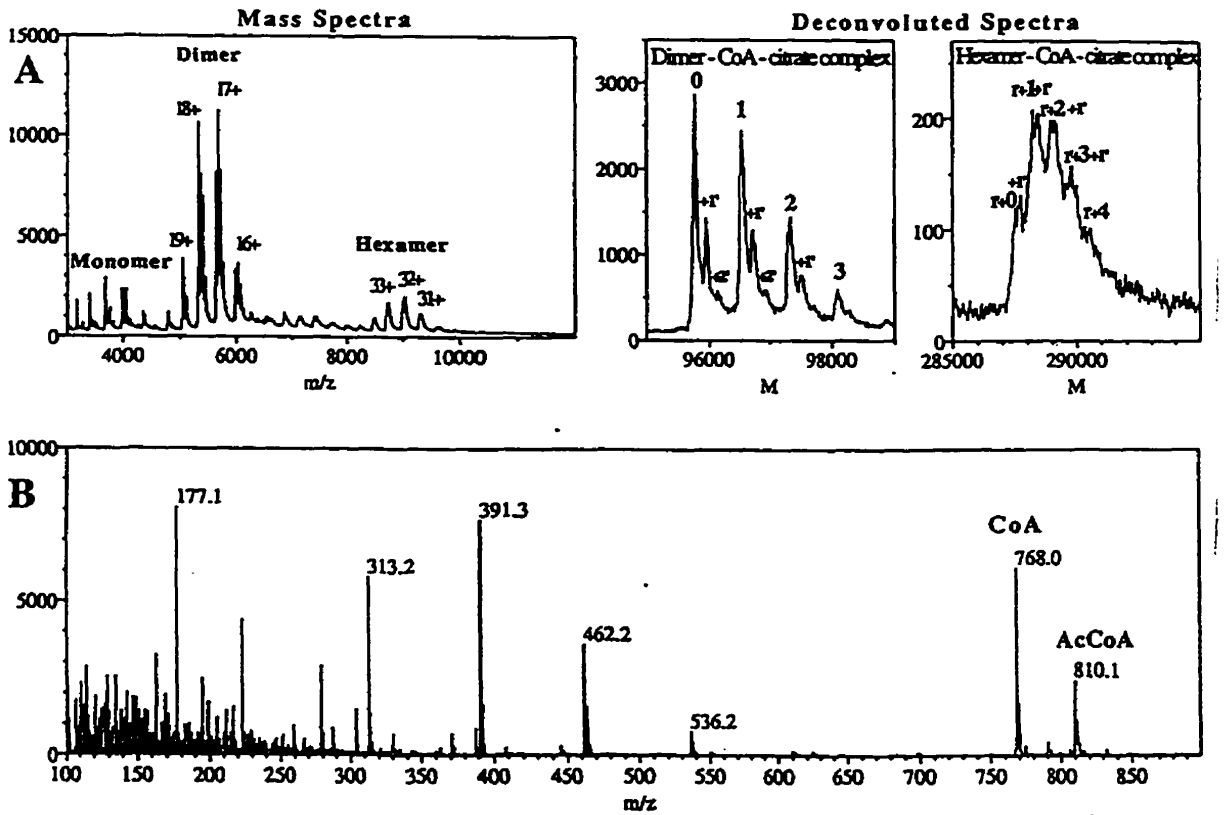


Figure 3.16 (A) CS spectra in the presence of equimolar concentrations of the substrates OAA and AcCoA, in 5 mM ammonium bicarbonate, pH 7.5. The sample was mixed and allowed to incubate for two hours prior to ESI-TOF measurements to allow the conversion of all substrates to products (citrate and CoA). The labels in the deconvoluted spectra refer to bound CoA molecules (numbers without prime) and citrate (numbers with prime). (B) Low m/z region showing the relative amounts of CoA and AcCoA.

(data not shown), whereas at pH 7.5, it was small, as seen in Figure 3.16 B. The difference may be simply be a pH effect; the pH optimum of CS is 8.0 (Faloona & Srere, 1969, Weitzman, 1966).

Further work was not undertaken to determine whether NADH prevents binding of the substrates, since it was expected that many possible combinations of bound ligands would lead to complex, poorly resolved spectra.

Discussion

Molecular mass determinations and charging

Mass spectrometry is unparalleled in its ability to measure masses accurately. Traditional methods for mass measurements of proteins and other macromolecules, such as ultracentrifugation, SDS-PAGE, or gel filtration measure molecular weights to an accuracy of 10% at best. Mass spectrometry on the other hand can give accurate mass determinations to within 0.01%, requiring very little material (Loo, 1997). This kind of accuracy is essential where small changes in molecular weight need to be detected, as in studies of ligand binding, chemical modification, or site-directed-mutagenesis.

The molecular mass determination of denatured CS samples (wild type, chemically modified, or mutant) were all within 0.02% of the expected masses based on the protein sequence. In this case, the accuracy of the mass determinations has allowed the confirmation of the loss of mass due to mutagenesis and the addition of mass due to chemical modification with TFBA. The fact that the measured mass of wild type CS monomer is within 2 u of the mass expected from the DNA sequence suggests that no truncation occurs due to proteolysis. The mass obtained shows that the N-terminal

methionine is not present, a result consistent with N-terminal sequencing of the protein (Ner *et al.*, 1983).

The masses obtained under non-denaturing conditions also agree well with those expected. For example, the CS dimer and hexamer peaks in ammonium bicarbonate correspond to masses of 95 770 ± 10 and 287 322 ± 30. While the dimer molecular weight is exactly as expected from the theoretical value, the hexamer mass is a mere 0.004% larger.

It has been noted by many that the charge envelope in the m/z spectrum of a protein is reflective of the conformation of the protein in solution in two ways: First, the width of the charge envelope (or the number of charge states) appears to be correlated to a global solution structure. Thus for a native compact protein, a narrow charge envelope is observed (typically 3 to 5 charge states), whereas for a denatured protein, the charge envelope is much wider (10 or more charge states). The width of the charge envelope may be correlated to the number of conformational states present: for a native protein, the set is small, but a denatured sample may comprise a large number of many interconverting conformations. Second, the maximum number of charges attained seems to be correlated with the surface area - and thus compactness - of a protein, such that a native, compact conformation acquires fewer charges than the unfolded counterpart. The maximum number of charges obtained in an unfolded protein is also usually related to the number of basic residues available (Loo, 1997).

The above observations made with many other proteins are corroborated by the spectra obtained for CS. In 5% acetic acid (pH ~ 2.5), the charge envelope is wide (about 40 charge states observed), and the number of charges is high (the maximum number of charges obtained is +63, with the

+45 ion being most abundant). Citrate synthase has 63 basic groups, including the N-terminal NH_3^+, so that the maximum number of charges obtained is consistent with a fully denatured protein in which all the basic sites are accessible. In contrast, CS in ammonium bicarbonate or ammonium acetate (pH 7.5 and 6, respectively) exhibits narrow charge envelopes (4 and 6 charge states for dimer and hexamer respectively) with smaller numbers of charges (+20 and +35 for dimer and hexamer, respectively), indicating that a large number of basic residues are inaccessible for protonation presumably because the structures are folded and compact. Assuming that protonation occurs only on the basic sites of the protein, then the dimeric CS makes-inaccessible about 85% of its basic residues and hexamer 91%.

Perhaps the more interesting aspect of ESI spectra is the preservation of non-covalent interactions, providing the opportunity to study protein-protein and protein-ligand interactions. The CS subunit structure is preserved in the electrospray source, yielding mainly dimer and hexamer as already discussed. Furthermore, a simple hexamer-dimer equilibrium is demonstrated by using different CS concentrations, indicating that equilibria present in solution may be reflected in electrospray ionization spectra. These data were used to calculate a plausible association constant for the dimer-hexamer equilibrium.

The effect of pH on the equilibrium between hexamer and dimer parallels that observed previously by ultracentrifugation techniques, further strengthening the case for a correlation between solution and gas-phase observations. This placed us in a position to undertake some work on the functional aspects of CS, especially on the relationship between NADH regulation and subunit interactions.

NADH binding and subunit association

It was proposed for years that the subunit structure of *E. coli* and other gram negative citrate synthases is related to the strong inhibition by NADH. The data presented here clearly identify the effects of NADH on the subunit structure, making it possible to formulate a mechanism for inhibition and to calculate constants that describe the equilibria in the system (presented in Table 3.2).

NADH binding to hexameric CS can be classified into two types: Specific, with a K_D of 1.1 μM, and nonspecific, with a K_D of 55 μM. The nonspecific class of sites (6 additional sites per hexamer) was fitted to account for the gradual increase in the average number of bound NADH at higher NADH concentrations. Dimer binds NADH weakly throughout the NADH concentration range with a K_D of 28 μM, about 25-times more weakly than hexamer does. The weak binding observed for dimer and hexamer at high NADH concentrations may not be a reflection of a true solution property, but may arise as an electrospray artifact; the NADH can be concentrated onto the surface of the CS oligomers nonspecifically during droplet evaporation. Despite the uncertainty of specificity at high ligand concentrations, it appears that specific interactions can be observed and that hexameric CS binds NADH specifically and selectively, with a K_D that is identical to that determined previously by fluorescence methods, and that NADH causes a shift in the dimer-hexamer equilibrium towards hexamer.

Cysteine 206 has been considered the marker for the allosteric NADH binding site based on the observation that its modification by alkylating agents results in an active enzyme that is desensitized to NADH inhibition (Donald *et al.*, 1991). Now it appears that an alternative, and perhaps more likely,

TABLE 3.2 Equilibrium constants for CS dimer-hexamer-NADH system as determined from ESI-TOF MS data in 5 mM ammonium bicarbonate

Parameter	Value
K_A =[hexamer]/[dimer]^3	(6.9 ± 0.7) X 10^{10} M^{-2}
K_D for NADH binding to dimer	28.3 ± 3.4 μM
K_D for NADH binding to hexamer "tight" sites	1.1 ± 0.2 μM
K_D for NADH binding to hexamer, "loose" sites	55 ± 3 μM*

*The K_D value for the second or "loose" set of NADH binding sites was calculated by assuming that there are 6 of these sites per hexamer.

explanation is that the TFBA modified CS does not form the NADH binding hexameric form, resulting in a population of exclusively dimeric CS which appears not to have a high affinity for NADH. While it is still possible that this residue may be in or near the NADH binding site, the observations that only dimer is present in this modified enzyme and that hexamerization does not occur upon the addition of NADH, strongly suggest that a key conformational change accompanied by hexamerization is blocked by the modification. Cysteine 206 is in a region of the sequence where no homology is evident between NADH-sensitive and insensitive citrate synthases, and contains three, proline residues at positions 205, 208, and 213, which are conserved among the gram-negative organisms and may play an important role in such a conformational change. Proline-rich-regions have been implicated with binding processes, especially protein-protein binding, owing to an increased rigidity of the peptide containing the prolines and consequent stabilization of the interaction due to a smaller entropy loss upon binding (Williamson, 1994). While the occurrence of the prolines mentioned above can hardly be classified as a proline rich region (since only three are present), their conservation in NADH sensitive citrate synthases and absence in other CSs, as well as their close proximity to C206, may indicate that they play an important role in NADH regulation and/or subunit interactions. Site-directed mutagenesis of one of these residues, P205, resulted in a CS which is not inhibited by NADH nor is it hexameric (unpublished work of Lin Ye), a behavior similar to the TFBA-modified CS. Our attempts to study this mutant by mass spectrometry failed since it was unstable in the electrospray-compatible buffers used.

Energetics of binding and subunit association

It is still impossible to rationalize the observed affinities of a protein for a ligand given our current poor understanding of the energetics of protein structure, particularly the strengths of basic interactions (Creighton, 1993). However, we do know that the affinities observed depend on the free energies of the complexes and the components. For CS, the main equilibria taking place are depicted in Figure 3.17. Three of the constants in the thermodynamic cycle, K_A, K_D^{dimer}, and $K_D^{hexamer}$ were determined experimentally from the MS results, while the fourth, $K_A^{'}$, was deduced from the linkage relationship:

$$K_A^{'} = K_A \times \left(\frac{K_D^{dimer}}{K_D^{hexamer}} \right) \qquad [17]$$

The value for $K_A^{'}$, 1.75 X 10^{12} M^{-2}, is larger than the association constant in the absence of ligand, implying a shift towards the hexameric state (since $K_A = [H]/[D]^3$).

In linked equilibria, the free energy of coupling, $\Delta G_{coupling}$, may be calculated. In this case, $\Delta G_{coupling}$ is defined as

$$\Delta G_c = -RT \ln \left(\frac{K_A^{'}}{K_A} \right) = -RT \ln \left(\frac{K_D^{hexamer}}{K_D^{dimer}} \right) \qquad [18]$$

and the resulting value is a modest -1.79 kcal.mol^{-1}. This implies that binding of NADH favours tighter subunit interactions, and reciprocally, that tighter subunit interactions favour NADH binding. But does this correspond to the free energy change for the transition from the allosteric R to T states? In other words, can we interpret the scheme in Figure 3.17 in terms of allosteric T and R states?

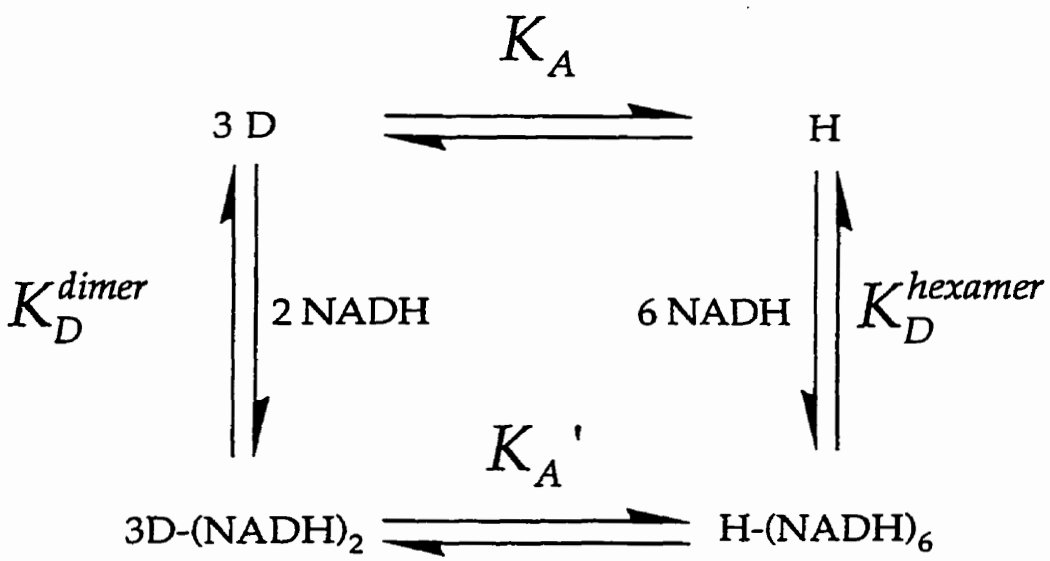


Figure 3.17. Linkage relationship between subunit association and NADH binding. K_A, K_D (dimer), and K_D (hexamer) were determined experimentally from the mass spectrometric measurements. The association constant K_A' was calculated based on the experimentally determined values. The constants are tabulated in Table 3.2.

Without taking into account other observations made about the allosteric behaviour of CS, the scheme would imply that the hexamer is the T state, and dimer is the R state, since hexamerization appears to be necessary for a tighter NADH binding. Also the observation here that NADH binding is hyperbolic, as was determined previously (Duckworth & Tong, 1976) would imply that this enzyme, in the absence of any ligands or salts, is shifted towards the allosteric T state. When previous observations are taken into account, however, the picture is complicated; CS is both activated and hexamerized in the presence of KCl, acting as an allosteric activator and rendering the sigmoid AcCoA saturation curve hyperbolic. These observations contradict the idea that the hexamer is the T state. Furthermore, mass spectrometry results here showed increased hexamerization in the presence of the substrates AcCoA and OAA, also implying that the active state is hexameric. Taken together, the results here and previous observations would suggest two distinct hexameric forms of CS: active (allosteric R state) and inactive (allosteric T state).

But what is the role of the dimeric CS? It is well known from the crystallographic data available on pig and chicken citrate synthases that the dimeric form is the minimum functional unit for catalytic activity, and this is the case here as well, since the dimeric CS is functional (at pH 9 , and TFBA-modified CS), albeit not as efficient as the KCl-induced hexameric form. However, the dimeric form of *E. coli* CS does not seem to have the properties of an allosteric T state owing to its weak binding of NADH. These two observations suggest that the dimer is neither a very good T state, nor an efficient R state enzyme. So does this suggest that the presence of the dimer is artefactual? Perhaps so, but an interesting possibility is that the dimer is an

intermediate necessary for the conformational switch from the hexameric T state to the hexameric R state, as depicted schematically in a theoretical model in Figure 3.18. TFBA-modified CS forms hexamer in the presence of KCl or at sufficiently high protein concentrations, but does not bind NADH nor form hexamers in the presence of the inhibitor, supporting the idea that two distinct hexamers may exist, with the dimer being an intermediate for the conformational switch. Clearly, further evidence is needed to support this hypothetical model; it would be interesting to see whether substrates induce hexamerization of the dimeric TFBA-CS by MS.

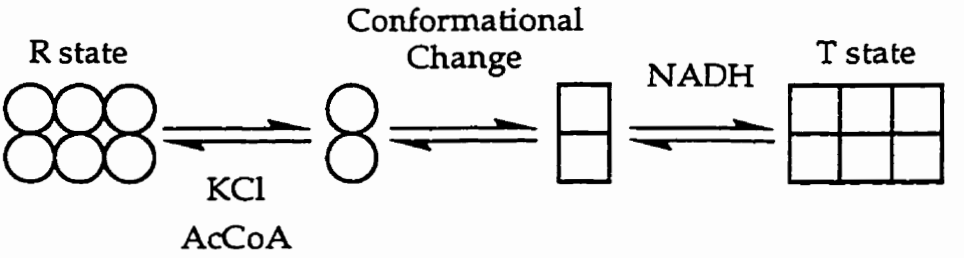


Figure 3.18 Theoretical model for allosteric states of CS. In this model, two conformationally distinct hexamers are proposed: the active R state, induced by KCl or AcCoA, and the inactive T state, present in the absence of ligands in *E. coli* CS and induced by NADH. The role of the dimer is proposed to be a way for the conformational transitions to take place, where dissociation is a necessary step.

Conclusions and future work

CS was an ideal system to test the feasibility of mass spectrometry for functional studies; the enzyme was stable and functional at low ionic strength required for electrospray ionization, and a large body of information on the solution properties was available to make comparisons with the mass spectrometric models deduced. New information on the mechanism of NADH inhibition was obtained. Specifically, it appears that NADH exerts its effect by binding to the hexameric form of CS and depleting the dimer

population. Binding constants were obtained that agree well with solution measurements and some idea of the free energy linking ligand binding and subunit association was obtained. Further questions need to be answered regarding the apparent conflict between an allosteric hexameric T state and the allosteric hexameric R state; it may be possible to resolve the issue by a more thorough study (such as presented for CS-NADH binding) of the substrates-CS equilibria using WT and TFBA-modified CS.

The methods used here may be extended to other systems with multiple equilibria between subunit assemblies and their ligands to gain a better understanding of allosteric systems. Since CS appears to be the first such system studied by mass spectrometry, it is not possible to draw conclusions as to the applicability of the method to other allosteric proteins.

APPENDIX I Sequence Alignment of Citrate Synthases

```
                1                                                  50
CISY_ECOLI      ADTKAKLTLN GDTAVELDVL KGTLGQDVID IRTLGSKGVF TFDPGFTSTA
CISY_BARVB      .......... .......... .......... .......... ..........
CISY_RICAZ      .......... .......... .......... .......... ..........
CISY_BARDO      .......... .......... .......... .......... ..........
CISY_BARVI      .......... .......... .......... .......... ..........
CISY_BAREL      .......... .......... .......... .......... ..........
CISY_BARTA      .......... .......... .......... .......... ..........
CISY_BARQU      .......... .......... .......... .......... ..........
CISY_BARBA      .......... .......... .......... .......... ..........
CISY_BARHE      SKNKAHITVN DK.KIELSVR KGTLGPDVIE IASLYKTDTF TYDPGFTSTA
CISY_PSEAE      ADKKAQLIIE GSAPVELPVL SGTMGPDVVD VRGLTATGHF TFDPGFMSTA
CISY_ACIAN      TGKKAVLHLD GK.EIELPIY SGTLGPDVID VKDVLASGHF TFDPGFMATA
CISY_RHITR      TEQSAKLTW. GEKTVDLPVK TEPSAQSLIL VPFIRTPPLF TYDPGFTSTA
CISZ_RHITR      DNNNACVLVD GH.SAELKLR SSTIGPNVLG IGSLYETKMF TYDPGFTSTA
CISY_ACEAC      ..STATISVD GK.SAEMPVL SGTLGPDVID IRKLPALGVF TFDPGYGETA
CISY_RICAU      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICPA      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICSL      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICSI      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICBE      ....AELKIR DK.IFKLPIL KASIGQDVID ISKVYSADCF TYDPGFMSTA
CISY_ASTRI      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICAK      ....AELKIR GK.IFQLPIL KASIGEDVID ISRVSAVNCF TYDPGFMSTA
CISY_RICAF      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICRH      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICMA      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICCA      ....VELKIR GK.IFKLPIL KASIGEDVID ISRVSSADCF TYDPGFMSTA
CISY_RICCN      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_RICTY      ....AELKIR GK.MFKLPIL KASIGKDVID ISRVSAADCF TYDPGFMSTA
CISY_RICJA      ....AELKIR GK.IFKLPIL KASIGEDVID ISRVSAADCF TYDPGFMSTA
CISY_COXBU      SNRKAKLSFE NQ.SVEFPIY SPTLGKDVID VKTLGNHGAY ALDVGFYSTA
CISY_RICHE      ....AELKIR GK.IFKLPIL KASIGEDVIN ISRVSAADCF TYDPGFMSTA
CISY_RICPR      ....AELKIR GK.LFKLPIL KASIGKDVID ISRVSAADYF TYDPGFMSTA
CISY_MYCTU      TDDTATLRYP GG.EIDLQIV HATEGADGIA LGPLLAKGHT TFDVGFANTA
CISY_CORGL      DNNKAVLHYP GG.EFEMDII EASEGNNGVV LGKMLSTGLI TFDPGYVSTG
CYSZ_CUCMA      ..VKGTLTIV DERQVQVSEE GTIKATDLKK ITTGPNKGLK LYDPGYLNTA
CISY_HELPY      ......VNNE NNERYEFETI ESTRGPKAVD FSKLFETGFF SYDPGYSSTA
CISY_THIFE      .......... .......... .......... .......... NFAPGLEGVA
CISZ_BACSU      .......... .......... .......... .........M TATRGLEGVV
CISZ_STRMU      .......... .......... .......... .......... ....GLKDMI
CISY_STRHY      .......... .......... .......... .......... VVSKGLENVI
CISY_BACCO      .......... .......... .......... .......... QFIPGLEGVI
CISW_BACSU      .......... .......... .......... .......... HYSPGLDGVI
CISY_THEAC      .......... .......... .......... .......... EISKGLEDVN
CISZ_ECOLI      .......... .......... .......... .......... ....ALSGVP
CISY_BACSU      .......... .......... .......... .......... ....GLKGIT
CISZ_SALTY      .......... .......... .......... .......... ....ALSGVP
CISY_MYCSM      .......... .......... .......... MTTATESEAP RIHKGLAGVV
CISY_SCHPO      KDRLAELIPE KQ..AEIKKF RAEHGQDVIG EVTIN..... QMYGGARGVR
CISZ_MYCTU      .......... .......... .......... .......... NFVPGLDGVV
CISY_CANTR      .......EIL PAKAEEVKQL KKDYGKTVIG EVLLE..QAY GGMRGIKGLV
CISY_NEUCR      .......ELL PENIEKIKAL RKEHGSKVVD KVTLD..... QVYGGARGIK
CISY_CHICK      .......... .......DVL AALIPKEQAR IKTLGQITVD MSYGGMRGMK
CISY_CAEEL      ....APLSTS AEGSTNLKVL SKKIPAHNAK VKSFRTVNID MIYGGMRSMK
CISY_YEAST      .......EII PAKAEEIKKF KKEHGKTVIG EVLLE..... QAYGGMRGIK
CISY_EMENI      .......... .....KVKKL RKEHGNKVIG ELTLDQKCLV WEGSVLDSEE
CISY_CITMA      MQSSADLDLH SQLKVKSDLG KAQLGNITID VVIG...... ....GMRGMT
CISY_SYNY3      ..KKVIAEII PQKQAELKEV KEKYGDKVVG QYTVK..... QVIGGMRGMK
CISY_PIG        KDILADLIPK EQ..ARIKTF RQQHGNTVVG QITVD..... MMYGGMRGMK
CISZ_YEAST      .......... ..HAQDVRQF VKEHGKTKIS DVLLE..... QVYGGMRGIP
CISY_ASPNG      .......... .....KVKKL RKEHGSKVIG EVTLDQKCLV WEGSVLDSEE
CISX_YEAST      ..KEALENVI PKKRDAVKKL KACYGSTFVG PITISSQSMF WQGTSLDPEH
CISY_ARATH      .DLKSQLQLI PEHKDRLKKL KSEHGKVQLG NITVD..... MVIGGMRGMT
```

```
                51                                                  100
CISY_ECOLI      SCESKITFID GDEGILLHRG FPIDQLATDS NYLEVCYILL NGEKPTQEQY
CISY_BARVB      .......... .......... .......... .......... ..........
CISY_RICAZ      .......... .......... .......... .......... ..........
CISY_BARDO      ......TYID GDKGILLYRG YPIEQLAEKG DFLESCYLLL YGELPTQQEK
CISY_BARVI      .........D GNKGILLYRG YPIEQLAEKG DFLESCYLLL YGELPTQQEK
CISY_BAREL      .......YID GDKGILLYRG YPIDQLAEKG DFLESCYLLL YGELPTQQEK
CISY_BARTA      ......TYID GNKGILLYRG YPIDQLAERG DFLESCYLLL YGELPTKQEK
CISY_BARQU      ........ID GDKGILLYCG YPIDQLAEKG DFLESCYLLL YGELPTKQEK
CISY_BARBA      .......YID GDEGILLYHG YSIDQLAENG DFLETCYLLL YGELPNKQQK
CISY_BARHE      SCESKITYID GNEGILLYRG YPIDQLAEKG DFLESCYLLL YGELPTKQEK
CISY_PSEAE      SCESKITYID GDKGVLLHRG YPIEQLAEKS DYLETCYLLL NGELPTAAQK
CISY_ACIAN      SCESKITFID GDKGILLHRG YPIDQLATQA DYLETCYLLL NGELPTAEQK
CISY_RHITR      SCESSITFID GDEGVLLHRG YPIEQLAEHG DFLEVCYLLL YGELPTAAQK
CISZ_RHITR      SCESSITFID GDEGVLLHRG YPIEQLAEHG DFLEVCYLLL YGELPTAAQK
CISY_ACEAC      ACNSKITFID GDKGVLLHRG YPIAQLDENA SYEEVIYLLL NGELPNKVQY
CISY_RICAU      SCQSTITYID GDKGILRHRG YNIKDLAEKS DFLEVAYLLI YGELPTIEQY
CISY_RICPA      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICSL      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICSI      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICBE      SCRSTITYID GDQGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPNNKQY
CISY_ASTRI      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICAK      SCKSTITYID GDKGILRHRG YNIKDLAEKS DFLEVAYLLI YGELPSIEQY
CISY_RICAF      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICRH      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICMA      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICCA      SCRSTITYID GDKGILRYRG YDIKDLADKS DFLEVAYLLI YGELPSSEQY
CISY_RICCN      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_RICTY      SCQSTITYID GDKGILWHRG YDIKDLAEKS DFLEVAYLMI YGELPSSEQY
CISY_RICJA      SCQSTVTYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSGEQY
CISY_COXBU      ACESKITFID GEKGILLYRG YPIDQLADKS DYMEVCYLLM YGELPNKGEK
CISY_RICHE      SCQSTITYID GDKGILRHRG YDIKDLAEKS DFLEVAYLLI YGELPSSEQY
CISY_RICPR      SCQSTITYID GDKGILWYRG YDIKDLAEKS DFLEVAYLMI YGELPSSDQY
CISY_MYCTU      AAKSSITYID GDAGILRYRG YPIDQLAEKS TFIEVCYLLI YGELPDTDQL
CISY_CORGL      STESKITYID GDAGILRYRG YDIADLAENA TFNEVSYLLI NGELPTPDEL
CYSZ_CUCMA      PVRSSISYID GDLGILRYRG YPIEELAESS TYVEVAYLLM YGNLPSQSQL
CISY_HELPY      GCQSKISYVN GKKGELYYRG HRIEDLVAKY KYVDVCKLLL TGELPKNQDE
CISY_THIFE      ATQSSISNID GAAGLLSYRG FAIADLAAHS SFEEVALLLL DGVLPGAADL
CISZ_BACSU      ATTSSVSSII DDT..LTYVG YDIDDLTENR SFEEIIYLLW HLRLPNKKEL
CISZ_STRMU      ACDTRISAIK DNK..LSYAG YDIADLMDNT RFEEVIYLLW NLHLPTAIEL
CISY_STRHY      IKVTNLTFID GEKGILRYRG YNIEDLVNYG SYEETIYLML YGKLRTKKEL
CISY_BACCO      ASETKISFLD TVNSEIVIKG YDLLALSKTK GYLDIVHLLL EGTIPNEAEK
CISW_BACSU      AAETHISYLD TQSSQILIRG YDLIELSETK SYLELVHLLL EGRLPEESEM
CISY_THEAC      IKWTRLTTID GNKGILRYGG YSVDIIASGA QDEEIQYLFL YGNLPTEQEL
CISZ_ECOLI      AGNTALCTVG KSGNDLHYRG YDILDLAKHC EFEEVAHLLI HGKLPTRDEL
CISY_BACSU      CVETSISHID GEKGRLIYRG HHAKDIALNH SFEEAAYLIL FGKLPSTEEL
CISZ_SALTY      AGNTALCTVG KSGNDLHYRG YDILDLAEHC EFEEVAHLLI HGKLPTRDEL
CISY_MYCSM      VDTTAISKVV PETNSLTYRG YPVQDLAAQC SFEQVAYLLW HGELPT.DQL
CISY_SCHPO      SLIWEGSVLD PNEGIR.FRG YTIPECQKQP LPESLFWLLV TGEIPTLSQV
CISZ_MYCTU      AFTTEIAEPD KDGGALRYRG VDIEDLVSQR TFGDVWALLV DGNFGSGLPP
CISY_CANTR      WEGSVLDPIE G....IRFRG RTIPDIQKEP LPEALFWLLL TGEVPTEAQT
CISY_NEUCR      CLVWEGSVLD AEEGIRFQEL LPKAPGGKEP LPEGLFWLLL TGEVPSEQQV
CISY_CHICK      GLVYETSVLD PDEGIR.FRG FSIPECQKEP LPEGLFWLLV TGQIPTGAQV
CISY_CAEEL      GMVTETSVLD PEEGIR.FRG YSIPECQKEP LPEAIWWLLC TGDVPSEAQT
CISY_YEAST      GLVWEGSVLD PEEGIRFQRE LPKAEGSTEP LPEALFWLLL TGEIPTDAQV
CISY_EMENI      GIRFRGLTIP ECQKLLPKAP GGEEPLPE.. ...GLFWLLL TGEVPSEQQV
CISY_CITMA      GLLWETSLLD PDEGIR.FRG LSIPECQKEP LPEGLLWLLL TGKVPSKEQV
CISY_SYNY3      GLMSDLSRCD PYQGII.FRG YTIPQLKEEP LPEGIFWLLM TGQLPTHAQV
CISY_PIG        GLVYETSVLD PDEGIR.FRG YSIPECQKEP LPEGLFWLLV TGQIPTEEQV
CISZ_YEAST      GSVWEGSVLD PEDGIR.FRG RTIADIQKDP LPEALFWLLL TGEVPTQAQV
CISY_ASPNG      GIRFRGRTIP ECQELLPKAP GGQEPLPE.. ...GLFWLLL TGEIPTEQQV
CISX_YEAST      GIKFQGLTIE ECQNRLPNTG IDGDNFLPE. ...SMLWLLM TGGVPTFQQA
CISY_ARATH      GLLWETSLLD PEEVFALKAL LPTAQSGGLN HYRRSFVASL NWKGTLAKSK
```

```
                101                                                       150
CISY_ECOLI      DEFKTTVTRH TMIHEQITRL FHAFRRDSHP MAVMCGITGA LAAFYHDSLD
CISY_BARVB      .......... .......... .......... .......... ..........
CISY_RICAZ      .......... .......... .......... .......... ..........
CISY_BARDO      NDFDRRIMQH TMVHEQFSRF FQGFRRDSHP MAVMVACLGA MSAFYHDSID
CISY_BARVI      SDFDRCIMQH TMVHEQFARF FHGFRRDSHP MAVMVACLGA MSAFYHDSID
CISY_BAREL      NDFDRCIMQH TMVHEQFSRF FHGFRRDSHP MAVMVACLGA MSAFYHDSID
CISY_BARTA      HDFDRCIMQH TMVHEQFARF FHGFRRDSHP MAVMVACLGA MSAFYHDSID
CISY_BARQU      IDFDRCIMQH MMVHERFTRF FHGFRRDSHP MAVMVACLGA MSAFYHDSID
CISY_BARBA      IDFDRCIMRH TMVHEQFARF FHGFRRDSHP MAVMVACLGA MSAFYHDSIN
CISY_BARHE      IDFDRCIMQH TMVHEQFARF FHGFRRDSHP MAVMIACLGA MSAFYHDSID
CISY_PSEAE      EQFVGTIKNH TMVHEQLKTF FNGFRRDAHP MAVMCGVIGA LSAFYHDSLD
CISY_ACIAN      VEFDAKVRAH TMVHDQVSRF FNGFRRDAHP MAIMVGVVGA LSAFYHNNLD
CISY_RHITR      KDFDYRVVHH TMVHEQMSRF FTGFRRDAHP MAVMCGCVGA LSAFYHDSTD
CISZ_RHITR      KDFDYRVVHH TMVHEQMSRF FTGFRRDAHP MAVMCGCVGA LSAFYHDSTD
CISY_ACEAC      DTFTNTLTNH TLLHEQIRNF FNGFRRDAHP MAILCGTVGA LSAFYPDAND
CISY_RICAU      NNFTTQVAHH SLVNERLHYL FQAFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICPA      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICSL      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICSI      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICBE      NDFTKKVAHH ALVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_ASTRI      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICAK      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAIGS LSAFYPDLLN
CISY_RICAF      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICRH      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICMA      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICCA      NNFTKKVAVH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICCN      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICTY      HNFTTKIAHH ALVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICJA      NNFTKQVAHH SLVNDRLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_COXBU      EKFVRTIKEH TSVYEQVTKF FNGFHYDAHP MAMVLSTIGA LSAFYHDALD
CISY_RICHE      NNFTKQVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_RICPR      CNFTKKVAHH SLVNERLHYL FQTFCSSSHP MAIMLAAVGS LSAFYPDLLN
CISY_MYCTU      AQFTGRIQRH TMLHEDLKRF FDGFPRNAHP MPVLSSVVNA LSAYYQDALD
CISY_CORGL      HKFNDEIRHH TLLDEDFKSQ FNVFPRDAHP MATLASSVNI LSTYYQDQLN
CYSZ_CUCMA      ADWEFAISQH SAVPQGLVDI IQAMPHDAHP MGVLVSAMSA LSVFHPDANP
CISY_HELPY      SEFELELRHR SFVHESLLNM FSAFPSNAHP MAKLSSGVSI LSTLYSTHQN
CISY_THIFE      ERFDHGLRAH RQVKYNVREI MKFMPVTGHP MDMLHCAVAS LGMFYPQQAE
CISZ_BACSU      EELKQQLAKE AAVPQEIIEH FKSYSENVHP MAAL.RTAIS LLGLLDSEAD
CISZ_STRMU      KQFEEKLRKN YAISDAIEQI LIQSRQHLHP MSVL.RSTVS LLGVYNLKAE
CISY_STRHY      NDLKAKLNEE YEVPQEVLDT IYLMPKEADA IGLLEVGTAA LASIDKNFKW
CISY_BACCO      QHLEETLKQE YDVPDEIIQV LSLLPKTAHP MDALRTGVSV LASFDTELLN
CISW_BACSU      ETLERKINSA SSLPADHLRL LELLPEDTHP MDGLRTGLSA LAGYDRQ.ID
CISY_THEAC      RKYKETVQKG YKIPDFVINA IRQLPRESDA VAMQMAAVAA MAASETK.FK
CISZ_ECOLI      AAYKTKLKAL RGLPANVRTV LEALPAASHP MDVMRTGVSA LGCTLPEKEG
CISY_BACSU      QVFKDKLAAE RNLPEHIERL IQSLPNNMDD MSVVRTVVSA LGE......N
CISZ_SALTY      NAYKSKLKAL RGLPANVRTV LEALPAASHP MDVMRTGVSA LGCTLPEKEG
CISY_MYCSM      ALFSQRERAS RRIDRSMQAL LAKLPDNCHP MDVVRTAISY LGAEDLEE.D
CISY_SCHPO      QALSADWAAR SQLPKFVEEL IDRCPPTLHP MAQFSLAVTA LEAYERGMNK
CISZ_MYCTU      AEPFPLPIHS GDVRVDV... .......QAG LAMLAPIWGY APLLDID..D
CISY_CANTR      RALSEEFAAR SALPKHVEEL IDRSPSHLHP MAQFSIAVTA LESESQFAKA
CISY_NEUCR      RDLSAEWAAR SDVPKFIEEL IDRCPSDLHP MAQLSLAVTA LEHTSSFARA
CISY_CHICK      SWLSKEWAKR AALPSHVVTM LDNFPTNLHP MSQLSAAITA LNSESNFARA
CISY_CAEEL      AAITKEWNAR ADLPTHVVRM LDNFPDNLHP MAQF...IAA IAALNNESRG
CISY_YEAST      KALSADLAAR SEIPEHVIQL LDSLPKDLHP MAQFSIAVTA LESESKFAKA
CISY_EMENI      RDLSAEWAAR SDLPKFIEEL IDRVPSTLHP MAQFSLAVTA LEHESAFAKA
CISY_CITMA      DGLSKELRDR ATVPDYVYKA IDALPVSAHP MTQFASGVMA LQVQSEFQEA
CISY_SYNY3      DALKHEWQNR GTVNQDCVNF ILNLPKDLHS MTML...SMA LLYLQKDSKF
CISY_PIG        SWLSKEWAKR AALPSHVVTM LDNFPTNLHP MSQLSAAITA LNSESNFARA
CISZ_YEAST      ENLSADLMSR SELPSHVVQL LDNLPKDLHP MAQFSIAVTA LESESKFAKA
CISY_ASPNG      RDLSAEWAAR SDLPKFIEEL IDRCPSTLHP MSQFSLAVTA LEHESAFAKA
CISX_YEAST      ASFRKELARG RKLPHYTEKV LSSLPKDMHP MTQL...AIG LASMNKGSLF
CISY_ARATH      LKHCRKTWNR AAVSDYVYNA IDALPSTAHP MTQFASGVMA LQVQSEFQKA
```

```
                151                                                 200
CISY_ECOLI      VNNPRHREIA AFRLLSKMPT MAAMCYKYSI GQPFVYPRND LSYAGNFLNM
CISY_BARVB      .......... .......... .......... .......... ..........
CISY_RICAZ      .......... .......... .......... .......... ..........
CISY_BARDO      ITDPQQRMIA SIRLISKVPT LAAMAYKYSI GQAFVYPHNS LSYAANFLRM
CISY_BARVI      ITDPQQRMIA SVRLISKVPT LAAMAYKYSI GQAFVYPRND LSYAANFLRM
CISY_BAREL      ITDPHQRMIA SVRLISKVPT LAAMAYKYSI GQAFVYPRND LSYAANFLRM
CISY_BARTA      ITDPQQRMIA SVRLISKVPT LAAMAYKYSI GQAFVYPRND LSYAANFLRM
CISY_BARQU      IKDAQQRMIA AIRLISKVPT LAAMAYKYSI GQAFVYPRND LSYAANFLHM
CISY_BARBA      ITDPQQRMIA SIRLISKVPT LAAMAYKYSI GQPFVYPRND LNYATNFLHM
CISY_BARHE      ITDPQQRMIA SIRLISKVPT LAAMAYKYSI GQAFVYPRND LSYAANFLRM
CISY_PSEAE      ITNPKHRQVS AHRLIAKMPT IAAMVYKYSK GEPMMYPRND LNYAENFLHM
CISY_ACIAN      IEDINHREIT AIRLIAKIPT LAAWSYKYTV GQPFIYPRND LNYAENFLHM
CISY_RHITR      ITDPHQRMVA SLRMIAKMPT LAAMAYKYHI GQPFVYPKND LDYASNFLRM
CISZ_RHITR      ITDPHQRMVA SLRMIAKMPT LAAMAYKYHI GQPFVYPKND LDYASNFLRM
CISY_ACEAC      IAIPANRDLA AMRLIAKIPT IAAWAYKYTQ GEAFIYPRND LNYAENFLSM
CISY_RICAU      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICPA      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICSL      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICSI      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICBE      FFKEADYELT AIRMIAKIPT IAAMSYKYSI GQPFVYPDNS LDFTENFLHM
CISY_ASTRI      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNP LDFTENFLHM
CISY_RICAK      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFVYPDNS LDFTENFLHM
CISY_RICAF      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNP LDFTENFLHM
CISY_RICRH      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICMA      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICCA      F.KEADYELT AIRMIAKIPT MAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICCN      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICTY      F.NETDYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICJA      F.KEADYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_COXBU      ITKPADRELS AIRLIAKMPT LAAMSYKYSI GQPFMHPRRA MNYAENFLHM
CISY_RICHE      C.KEADYKLT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_RICPR      F.NETDYELT AIRMIAKIPT IAAMSYKYSI GQPFIYPDNS LDFTENFLHM
CISY_MYCTU      PMDNGQVELS TIRLLAKLPT IAAYAYKKSV GQPFLYPDNS LTLVENFLRL
CISY_CORGL      PLDEAQLDKA TVRLMAKVPM LAAYAHRARK GAPYMYPDNS LNARENFLRM
CYSZ_CUCMA      AKSKQVRDKQ IARIIGKAPT IAAAAYLRLA GRPPVLPSSN LSYSENFLYM
CISY_HELPY      MHTEEDYQTM ARRIVAKIPT LAAICYRNEV GAPIIYPDIA RSYVENILFM
CISY_THIFE      RGNTLHLDAM AMRIIARMPT IVAMWEQMRF GNDPISPRPD LSHAANFLYM
CISZ_BACSU      TMNPEANYRK AIRLQAKVPG LVAAFSRIRK GLEPVEPRED YGIAENFLYT
CISZ_STRMU      ERSVEATYDQ SIQLMAKIPT IIATFARLRQ GLSPIAPRKD LGFAANFLYM
CISY_STRHY      KENDKEK... AISIIAKMAT LVANVYRRKE GNKPRIPEPS DSFAKSFLLA
CISY_BACCO      REHSTNLK.R AYQLLGKIPN IVANSYHILH SEEPVQPLQD LSYSANFLYM
CISW_BACSU      DRSPSANKER AYQLLGKMPA LTAASYRIIN KKEPILPLQT LSYSANFLYM
CISY_THEAC      WNKDTDRDVA A.EMIGRMSA ITVNVYRHIM NMPAELPKPS DSYAESFLNA
CISZ_ECOLI      HTVSGARDI. ADKLLASLSS ILLYWYHYSN GERIQPETDD DSIGGHFLHL
CISY_BACSU      TYTFHPKTEE AIRLIAITPS IIAYRKRWTR GEQAIAPSSQ YGHVENYYYM
CISZ_SALTY      HTVSGARDI. ADKLLASLNS ILLYWYHYSN GERIQPETDD DSIGGHFLHL
CISY_MYCSM      VDTAEANYAK SLRMFAVLPT IVATDIRRRQ GLTPIPPHSQ LGYAQNFLNM
CISY_SCHPO      HDYWKYEYED CMDLIAKTVP IAGRIYRNLY RDGVVAPIQD KDHSYNFANV
CISZ_MYCTU      ATARQQLARA SVMALSYVAQ SARGIYQPAV PQRIIDECS. .TVTARFMTR
CISY_CANTR      YAKGVHKSED SIELLAKLPT IAAKIYRNVF HDGKLPAQID LDYGANLASL
CISY_NEUCR      YAKGINKKED SMDLIAKLPT IAARIYQNVK GGKVAAVQKD KDYSFNFANQ
CISY_CHICK      YAEGILRTKS AMDLIAKLPC VAAKIYRYRA GSSIGAIDSK LDWSHNFTNM
CISY_CAEEL      VAKASYWEYA SMDLLAKLPT VAAIIYRYRD GSAVSVIDPK KDWSANFSSM
CISY_YEAST      YAQGVSKKED SLDLLGKLPV IASKIYRNVF KDGKITSTDP NADYGKNLAQ
CISY_EMENI      TAKGINKKED SMDLIAKLPT IAAKIYRNVF KDGKVAPQKD KDYSYNLANQ
CISY_CITMA      YEKGIHKSKD SLNLIARVPV VAAYVYQRIY KDGKIIPKDD LDYGGNFSHM
CISY_SYNY3      AKLYDEGKID SMDLIAKIPR VAAIIYRHKR DSKLIDSDSK LDWAGNYAHM
CISY_PIG        YAEGIHRTKD CMDLIAKLPC VAAKIYRYRE GSSIGAIDSK LDWSHNFTNM
CISZ_YEAST      YAQGISKQED SLDLLGKLPV IAAKIYRNVF KDGKMGEVDN ADYAKNLVNL
CISY_ASPNG      YAKGINKKED SMDLIAKLPT IAAKIYRNVF KDGKVAPQKD KDYSYNLANQ
CISX_YEAST      ATNYQKGLID SLNLIASLPL LTGRIYSNNE GHPLGQYSEE VDWCTNICSL
CISY_ARATH      YENGIHKSKD CLNLIARVPV VAAYVYRRMY KNGDSIPSDS LDYGANFSHM
```

```
                201                                                  250
CISY_ECOLI      MFSTPCEPYE VNPILERAMD RILILHADHE QNASTSTVRT AGSSGANPFA
CISY_BARVB      .......... .......... .......... .......... ..........
CISY_RICAZ      .......... .......... .......... .......... ..........
CISY_BARDO      CFAVPCEEYQ VNPVLTRAMD RIFILHADHE QNASTSTVRL AGSSGANPFA
CISY_BARVI      CFSVPCEEYK TNPVLTRAMD RIFILHADHE QNASTSTVRL AGSSGANPFA
CISY_BAREL      CFAVPCEEYK INPVLARAMD QIFILHADHE QNASTSTVRL AGSSGANPFA
CISY_BARTA      CFSVPCEGYN INSVLTRAMD RIFTLHADHE QNASTSTVRL AGSSGANPFA
CISY_BARQU      CFSVPCEEYK INPVLSRAMD RICTLHADHE QNASTSTVRL VGSSGANPFA
CISY_BARBA      CFSVPCEEHK ISPVIARAMD RIFTLHADHE QNASTSTVRL AGSSGANPYA
CISY_BARHE      CFSVPCEEYK INPVLTRAMD RIFTLHADHE QNASTSTVRL AGSSGANPFA
CISY_PSEAE      MFNTPCETKP ISPVLAKAMD RIFILHADHE QNASTSTVRL AGSSGANPFA
CISY_ACIAN      MFATPADRYK VNPVLARAMD RIFTLHADHE QNASTSTVRL AGSTGANPYA
CISY_RHITR      CFAVPCEEYV VNPVLARAMD RIFILHADHE QNASTSTVRL AGSSGANPFA
CISZ_RHITR      CFAVPCEEYV VNPVLARAMD RIFILHADHE QNASTSTVRL AGSSGANPFA
CISY_ACEAC      MFARMSEPYK VNPVLARAMN RILILHADHE QNASTSTVRL AGSTGANPFA
CISY_RICAU      MFATPCTKYK VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANAFA
CISY_RICPA      MFATPCTKYT VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICSL      MFATPCTKYT VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICSI      MFATPCTKYT VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICBE      MFATPCEKYK VNPVIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_ASTRI      MFATPCTKYT VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICAK      MFATPCTKYK VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANAFA
CISY_RICAF      MFATPCTKYT VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICRH      MFATPCTKYK VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICMA      MFATPCTKYK VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICCA      MFATPCMKYE VNPVIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICCN      MFATPCTKYT VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICTY      MFATPCTKYK VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICJA      MFATPYTKYT VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_COXBU      LFGTPYEETE PDPVLARAMD RIFILHADHE QNASTTTVRV AGSTGANPFA
CISY_RICHE      MFATPCTKYK VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_RICPR      MFATPCTKYK VNPIIKNALN KIFILHADHE QNASTSTVRI AGSSGANPFA
CISY_MYCTU      TFGFPAEPYQ ADPEVVRALD MLFILHADHE QNCSTSTVRL VGSSRANLFT
CISY_CORGL      MFGYPTEPYE IDPIMVKALD KLLILHADHE QNCSTSTVRM IGSAQANMFV
CYSZ_CUCMA      LDSLGNRSYK PNPRLARVLD ILFILHAEHE MNCSTSAARH LASSGVDVFT
CISY_HELPY      LRGYPYSRVE ITPLEVEAFD KILTLHADHS QNASSTTVRN VASTGVHPYA
CISY_THIFE      LSGR.....E PDPAHTKILD SCLILHAEHT INASTFSVLV TGSTLTNPYH
CISZ_BACSU      LNGE.....E PSPIEVEAFN KALILHADHE LNASTFTARV CVATLSDIYS
CISZ_STRMU      LNGRLPSELE I.....LAMN RALVLHAEHE LNASTFAARV CASTLSDIYS
CISY_STRHY      SFAREPTTDE IN.....AMD KALILYTDHE VPASTTAALV AASTLSDMYS
CISY_BACCO      ITGKKPTELE .....EKIFD RSLVLYSEHE LPNSTFTARV IASTLSDLYG
CISW_BACSU      MTGKLPSSLE .....EQIFD RSLVLYSEHE MPNSTFAARV IASTHSDLYG
CISY_THEAC      AFGRKATKEE ID.....AMN TALILYTDHE VPASTTAGLV AVSTLSDMYS
CISZ_ECOLI      LHGE.....K PSQSWEKAMH ISLVLYAEHE FNASTFTSRV IAGTGSDMYS
CISY_BACSU      LTGE.....Q PSEAKKKALE TYMILATEHG MNASTFSARV TLSTESDLVS
CISZ_SALTY      LHGE.....K PTQSWEKAMH ISLVLYAEHE FNASTFTSRV IAGTGSDVYS
CISY_MYCSM      CFGEVPE... ..PVVVRAFE QSMVLYAEHS FNASTFAARV VTSTQSDIYS
CISY_SCHPO      L.......GFA NNEEFVELMR LYLTIHADHE GNVSAHTGHL VGSALSSPFL
CISZ_MYCTU      WQGEP..... .DPRHIEAID AYWVSAAEHG MNASTFTARV IASTGADVAA
CISY_CANTR      L.......GFG ENKEFLELMR LYLTIHSDHE GNVSAHTTHL VGSALSSPFL
CISY_NEUCR      L.......GFG DNKDFVELLR LYLTIHTDHE GNVSAHTTHL VGSALSSPFL
CISY_CHICK      LGYTDAQ... ....FTELMR LYLTIHSDHE GNVSAHTSHL VGSALSDPYL
CISY_CAEEL      LGYD...... .DPLFAELMR LYLVIHSDHE GNVSAHTSHL VGSALSDPYL
CISY_YEAST      LLGYE..... .NKDFIDLMR LYLTIHSDHE GNVSAHTTHL VGSALSSPYL
CISY_EMENI      L.......GFA DNKDFVELMR LYLTIHSDHE GNVSAHTTHL VGSALSSPML
CISY_CITMA      LGFD...... .DPKMLELMR LYVTIHSDHE GNVSAHTGHL VASALSDPYL
CISY_SYNY3      MGFE...... .QHVVKECIR GYLSIHCDHE GNVSAHTTHL VGSALSDPYL
CISY_PIG        LGYTDAQ... ....FTELMR LYLTIHSDHE GNVSAHTSHL VGSALSDPYL
CISZ_YEAST      IGSKDEDFVD L.......MR LYLTIHSDHE GNVSAHTSHL VGSALSSPYL
CISY_ASPNG      L.......GYG DNNDFVELMR LYLTIHSDHE GNVSAHTTHL VGSALSSPML
CISX_YEAST      LGMTNGTTSQ QSLDFINLMR LYTGIHVDHE GNVSAHTTHL VGSALSDPYL
CISY_ARATH      LGFD...... .DERLKELMR LTSPSTVMHE GNVSAHTGHL VGSALSDPYL
```

```
                 251                                                  300
CISY_ECOLI       CIAAGIASLW GPAHGGANEA ALKMLEEISS VKHIPEFVRR AKDKNDSFRL
CISY_BARVB       .......... ......ANEA CLKMLLEIGS VKRIPEFIAR AKDKNDPFRL
CISY_RICAZ       .......... .......... ..NMLKEIGS SENIPKYIAK AKDKNDPFRL
CISY_BARDO       CIAAGVACLW GPAHGGANEA CLKMLKEIGS VKKIPEFIAR AKDKNDPFRL
CISY_BARVI       CIAAGVACLW GPAHGGANEA CLKMLQEIGS VKRIPEFIAR AKDKNDPFRL
CISY_BAREL       CIAAGVACLW GPAHGGANEA CLKMLQEIGS IKRIPEFIAR AKDKNDPFRL
CISY_BARTA       CIAAGVACLW GPAHGGANEA CLKMLQEIGS VERIPEFIAR AKDKNDPFHL
CISY_BARQU       CIAAGVACLW GPAHGGANEA CLKMLQKIGS VERIPEFIAR AKDKNDPFRL
CISY_BARBA       CIAAGVACLW GPAHGGANEA CLKMLQEIGS VKKIPEFIAR AKDKNDPFRL
CISY_BARHE       CIAAGVACLW GPAHGGANEA CLKMLQEIGS VERIPEFIAR AKDKNDSFRL
CISY_PSEAE       CIASGIAALW GPAHGGANEA VLRMLDEIGD VSNIDKFVEK AKDKNDPFKL
CISY_ACIAN       CISAGISALW GPAHGGANEA VLKMLDEIGS VENVAEFM.. EKVKRKEVKL
CISY_RHITR       CIAAGIACLW GPAHGGANEA ALNMLTEIGT VDRIPEYIAR AKDKNDPFRL
CISZ_RHITR       CIAAGIACLW GPAHGGANER ALNMLTEIGT VDRIPEYIAR AKDKNDPFRL
CISY_ACEAC       CIAAGIAALW GPAHGGANEA VLKMLARIGK KENIPAFIAQ VKDKNSGVKL
CISY_RICAU       CISTGIASLW GPAHGGANEA VINMLKEIGS SENIPKFIAK AKDKNDPFRL
CISY_RICPA       CISTGIASLW GPAHGGANEA VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICSL       CISTGIASLW GPAHGGANEA VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICSI       CISTGIASLW GPAHGGANEA VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICBE       CVSTGIASLW GPAHGGANEA VINMLKEIGS VENIPKYIAK AKDKNDNFRL
CISY_ASTRI       CISTGIASLW GPAHGGANEA VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICAK       CISTGIASLW GPAHGGANEA VINMLKEIGS SENIPKYIAK AKDKSDPFRL
CISY_RICAF       CISTGIASLW GPAHGGANEA VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICRH       CISTGIASLW GPAHGGANEV VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICMA       CISTGIASLW GPAHGGANEV VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICCA       CISTGIASLW GPAHGGANEA VINMLKEIGS SENIPKYIAK AKDKDDPFRL
CISY_RICCN       CISTGIASLW GPAHGGANEA VINMIKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_RICTY       CISTGIASLW GPAHGGANEA VINMLKEIGS SENIPKYVAK AKDKNDPFRL
CISY_RICJA       CISTGIASLW GPAHGGANEA VINMLKEIGS SEYIPKYIAK AKDKNDPFRL
CISY_COXBU       CISAGISALW GPAHGGANEA CLNMLRKIGD EKNIGQYIKK AKDKNDPFRL
CISY_RICHE       CISTGIASLW GPAHGGANEA VINMLKEIGS SENIPKYIAK AKDKNDPFRL
CISY_RICPR       CISTGIASLW GPAHGGANEA VINMLKEIGS SENIPKYVAK AKDKNDPFRL
CISY_MYCTU       SISGGINALW GPLHGGANQA VLEMLEGIRS GDDVSEFVRK VKNREAGVKL
CISY_CORGL       SIAGGINALS GPLHGGANQA VLEMLEDIKS NHDATEFMNK VKNKEDGVRL
CYSZ_CUCMA       ALSGAVGALY GPLHGGANEA VLKMLSEIGT VNNIPEFIEG VKNRKR..KM
CISY_HELPY       AISAGISALW GHLHGGANEK VLLQLEEIGD VKNVDKYIAR VKDKNDNFKL
CISY_THIFE       VIGGAIGTLA GPLHGGANQK VVEMLEEISS VQQVGAYLDR KMANKE..KI
CISZ_BACSU       GITAAIGALK GPLHGGANEG VMKMLTEIGE VENAEPYIRA KLEKKE..KI
CISZ_STRMU       CVTTAIGTLK GPLHGGANER VFDMLREIRE YGDVDSYLQE KLNSKE..KI
CISY_STRHY       SLTAALAALK GPLHGGAAEE AFKQFIEIGD PNRVQNWFN. DKVVNQKNRL
CISY_BACCO       ALTGAVASLK GHLHGGANEA VMEMLQDAQT VEGFKHLLHD KLSKKE..KI
CISW_BACSU       ALTGAVASLK GNLHGGANEA VMYLLLEAKT TSDFEQLLQT KLKRKE..KI
CISY_THEAC       GITAALAALK GPLHGGAAEA AIAQFDEIKD PAMVEKWFN. DNIINGKKRL
CISZ_ECOLI       AIIGAIGALR GPKHGGANEV SLEIQQRYET PDEAEADIRK RVENKE..VV
CISY_BACSU       AVTAALGTMK GPLHGGAPSA VTKMLEDIGE KEHAEAYLKE KLEKGE..RL
CISZ_SALTY       AIIGAIGALR GPKHGGANEV SLEIQQRYET PDEAEADIRK RVENKE..VV
CISY_MYCSM       AVTAAIGALK GSLHGGANEA VMHDMLEIGS AEKAPEWLHG KLSRKE..KV
CISY_SCHPO       SMAASLNGLA GPLHGLANQE VLNFL..ITM KKEIGDDLSE EKLLNSGRVV
CISZ_MYCTU       ALSGAIGAMS GPLHGGAPAR VLPMLDEVER AGDARSVVKG ILDRGE..KL
CISY_CANTR       SLAAGLNGLA GPLHGRANQE VLEWLFKLRS KEAIEKYLW. .DTLNAGRVG
CISY_NEUCR       SVAAGLNGLA GPLHGLANQE VLNWLTEMKK VEAITKYLW. .DTLNAGRVV
CISY_CHICK       SFAAAMNGLA GPLHGLANQE VLGWLAQLQK AXXXAGADAS LRDLNSGRVV
CISY_CAEEL       SFSAAMAGLA GPLHGLANQE VLVFLNKIVG EEQLKEWVW. .KHLKSGQVV
CISY_YEAST       SLAAGLNGLA GPLHGRANQE VLEWLFKLRS KETIEKYLW. .DTLNAGRVV
CISY_EMENI       SLAAGLNGLA GPLHGLANQE VLNWLTEMKK VVGNDLSDQS IKDLNAGRVV
CISY_CITMA       SFLAALNGLA GPLHGLANQE VLLWIKSVVD TEQLKDYVW. .KTLNSGKVV
CISY_SYNY3       SYSAGVNGLA GPLHGLANQE VLKWLLQFIE EKGTKVSDKD IEDYVDHVVV
CISY_PIG         SFAAAMNGLA GPLHGLANQE VLVWLTQLQ. .KEVGKDVSD EKTLNSGRVV
CISZ_YEAST       SLASGLNGLA GPLHGRANQE VLEWLFALKS KDTIEKYLW. .DTLNSGRVI
CISY_ASPNG       SLAAGLNGLA GPLHGLANQE VLNWLTKMKA AEAIKNYLWS TLNAGQ..VV
CISX_YEAST       SYSSGIMGLA GPLHGLAAQE VVRFLIEMNS NISSIAREQE IKDLNSNRVI
CISY_ARATH       SFAAALNGLA GPLHGLANQE VLLWIKSVVS KEQLKEYVW. .KTLNSGKVI
```

```
                301                                                  350
CISY_ECOLI      MGFGHRVYKN YDPRATVMRE TCHEVLKELG TKDDLLEVAM ELENIALNDP
CISY_BARVB      MGFGHRVYKN YDPRAKIMQK TCHEVLKELN INDPLLDIAI ELEKIALNDE
CISY_RICAZ      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QNNPLLQIAI ELEAIALKDE
CISY_BARDO      MGFGHRVYKN YDPRAKIMQK TCHEVLKELN RQDDLLDIAI ELEHIALNDE
CISY_BARVI      MGFGHRVYKN YDPRAKIMQK TCHEVLKELN IQDDLLDIAI ELEKIALNDE
CISY_BAREL      MGFGHRVYKN YDPRAKIMQK TCHEVLKELN IQDDLLDIAI ELEKIALSDE
CISY_BARTA      MGFGHRVYKN YDPRAKIMQK TCHEVLKELN IQDDLLDIAM ELEKIALNDE
CISY_BARQU      MGFGHRVYKN YDPRAKIMQQ TCHEVLKELN IQDDLLDIAI ALENTALNDE
CISY_BARBA      MGFGHRVYKN YDPRAKIMQK TCHEVLQELN IQDDLLDIAM ELEHIALNDE
CISY_BARHE      MGFGHRVYKN YDPRAKIMQQ TCHEVLKELN INDPLLDIAI TLENIALNDE
CISY_PSEAE      MGFGHRVYKN FDPRAKVMKQ TCDEVLQELG INDPQLELAM KLEEIARHDP
CISY_ACIAN      MGFGHRVYKN FDPRAKVMKQ TCDEVLEALG INDPQLALAM ELERIALNDP
CISY_RHITR      MGFGHRVYKN YDPRAKIMQK TAHEVLGELG IKDDLLDIAI ELERIALTDD
CISZ_RHITR      MGFGHRVYKN YDPRAKIMQK TAHEVLGELG IKDDLLDIAI ELERIALTDD
CISY_ACEAC      MGFGHRVYKN FDPRAKIMQQ TCHEVLTELG IKDDLLDLAV ELEKIALSDD
CISY_RICAU      MGFGHRVYKN YDPRAAVLKE TCKAVLKELG QNNPLLQIAI ELEAIALKDE
CISY_RICPA      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_RICSL      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_RICSI      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_RICBE      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_ASTRI      MGFGHRVYKN YDPRAAVLKE TCKEVLKALG QLDNLLQIAI ELEAIALKDE
CISY_RICAK      MGFGHRVYKN YDPRASVLKE TCKEVLKELG QNNPLLQIAI ELEAIALKDE
CISY_RICAP      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_RICRH      MGFGHRIYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_RICMA      MGFGHRIYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_RICCA      MGFGHRVYKN YDPRAAVLKE TCNEVLKELG QNNPLLQIAI ELEAIALKDE
CISY_RICCN      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QLDNFLQIAI ELEAIALKDE
CISY_RICTY      MGFGHRVYKS YDPRAAVLKE TCKEVLNELG QNNPLLQIAI ELEALALKDE
CISY_RICJA      MGFGHRVYKN YDPRAAVLKE TCKEVLKELG QLDNLLQIAI ELEAIALKDE
CISY_COXBU      MGFGHRVYKN YDPRAKVMQK TCYEVLDAVG RHNELFKLAI KLEKIALEDD
CISY_RICHE      IGFGHRVYKN YDPRAAVLKE TCKEVLKELG QNNPLLQIAI ELEAIALKDE
CISY_RICPR      MGFGHRVYKS YDPRAAVLKE TCKEVLNELG QLDNLLQIAI ELEALALKDE
CISY_MYCTU      MGFGHRVYKN YDPRARIVKE QADKILAKLG GDDSLLGIAK ELEEAALTDD
CISY_CORGL      MGFGHRVYKN YDPRAAIVKE TAHEILEHLG .GDDLLDLAI KLEEIALADD
CYSZ_CUCMA      SGFGHRVYKN YDPRAKVIRK LAEEVFSIVG .RDPLIEVAV ALEKAALSDE
CISY_HELPY      MGFGHRVYKS YDPRAKILKG LKDELHKGVK MDERLSEIAA KVEEIALKDE
CISY_THIFE      WGFGHRIYKT RDPRAVILKG MMEDMASHGN LRHSLFEIAI EVERQATE..
CISZ_BACSU      MGFGHRVYKH GDPRANDLKE MSKRLTNLTG .ESKWYEMSI RIEDIVTS..
CISZ_STRMU      MGFGHRVYQT QDPREKYLRE MARELTK..G TEHDIWYQLS KKVEMCMKQ.
CISY_STRHY      MGFGHRVYKT YDPRAKIFKK LALTLIERNA DARRYFEIAQ KLEELGIKQ.
CISY_BACCO      MGFGHRVYMK MDPRAAMMKE ALKELSAVNG ..DDLLLQMC EAGEQIMRE.
CISW_BACSU      MGFGHRVYMK MDPRALMMKE ALQQLCDKAG ..DHRLYEMC EAGERLMEK.
CISY_THEAC      MGFGHRVYKT YDPRAKIFKG IAEKLSSKKP EVHKVYEIAT KLEDFGIKA.
CISZ_ECOLI      IGFGHPVYTI ADPRHQVIKR VAKQLSQEGG ........SL KMYNIALETV
CISY_BACSU      MGFGHRVYKT KDPRAEALRQ KAEEVAGNDR DLD..LALHV EAEAIRLLEI
CISZ_SALTY      IGFGHPVYTI ADPRHQVIKR VAKQLSEEGG SLK.MYHIAD RLETVMWE..
CISY_MYCSM      MGFGHRVYKN GDSRVPTMKV ALEQV.AQVR DGQRWLDIYN TLES.....A
CISY_SCHPO      PGYGHAVLRK TDPRYTAQRE FALEHLP... .KDPMFQLVS RLYEIGVLTE
CISZ_MYCTU      MGFGHRVYRA EDPRARVLRA A....AERLG APRYEVAVAV EQAALSELRE
CISY_CANTR      PGYGHAVLRK TDPRYTAQRE FALKHMPDYE ....LFKLVS NIYEVAPGDQ
CISY_NEUCR      PGYAHAVLRK TDPRYSAQRK FAQEHLPE.. ..DPMFQLVS QVYKIAPKTE
CISY_CHICK      PGYGHAVLRK TDPRYTCQRE FALKHLPGDP MFKLVAQLYK IVPNVLLEQG
CISY_CAEEL      PGYGHAVLRK TDPRYECQRE FALKHLPNDD LFKLVSTLYK ITPGILLEQG
CISY_YEAST      PGYGHAVLRK TDPRYTAQRE FALKHFPDYE ....LFKLVS TIYEVAVLTK
CISY_EMENI      PGYGHAVLRK TDPRYTSQRE FALRKLPD.. ..DPMFKLVS QVYKIAVLTE
CISY_CITMA      PGFGHGVLRK TDPRYTCQRE FALKHLPD.. ..DPLFQLVS KLYEVPILTK
CISY_SYNY3      PGYGHAVLRD TDPRFHHQVD FSKFHLKDDQ MIKLLHQCAD VIPKKLLT..
CISY_PIG        PGYGHAVLRK TDPRYTCQRE FALKHLPHDP MFKLVAQLYK IVPNVLLEQG
CISZ_YEAST      PGYGHAVLRK TDPRYMAQRK FAMDHFPDYE LFK.LVSSIY EVAPGVLTE.
CISY_ASPNG      PGYGHAVLRK TDPRYVSQRE FALRKLPD.. ..DPMFKLVS QVYKIAVLTE
CISX_YEAST      PGYGHAVLRK PDPRFTAMLE FAQKRPIEFE NKNVLLMQKL AEIAPKVLLE
CISY_ARATH      PGYGHGVLRN TDPRYVCQRE .....FALKH HPDDPLFQCC KLMKLASCLT
```

```
                351                                                   400
CISY_ECOLI      YFIEKKLYPN VDFYSGIILK AMGIPSSMFT VIFAMARTVG WIAHWSEMHS
CISY_BARVB      YFVEKKLYPN VDFYSGITLK ALGFPTE... .......... ..........
CISY_RICAZ      YFIERKLYPN VDFYSGIIYK AMGIPS.... .......... ..........
CISY_BARDO      YFIEKKLYPN VDFYSGITLK ALGFP..... .......... ..........
CISY_BARVI      YFVEKKLYPN VDFYSGITLK ALGFPTE... .......... ..........
CISY_BAREL      YFIEKKLYPN VDFYSGITLK ALGFPTE... .......... ..........
CISY_BARTA      YFVEKKLYPN VDFYSGITLK ALGFPTE... .......... ..........
CISY_BARQU      YFIEKKLYPN VDFYSGITLK ALGFPTE... .......... ..........
CISY_BARBA      YFINKKLYPN VDFYSGITLK ALGFPTE... .......... ..........
CISY_BARHE      YFIEKKLYPN VDFYSGITLK ALGFPTEMFT VLFALARSVG WVAQWKEMIE
CISY_PSEAE      YFVERNLYPN VDFYSGIILK AIGIPTSMFT VIFALARTVG WISHWQEMLS
CISY_ACIAN      YFVERKLYPN VDFYSGIILK AIGIPTEMFT VIFALARTVG WISHWLEMHS
CISY_RHITR      YFIEKKLYPN VDFYSGITLK ALGFPTTMFT VLFALARTVG WIAQWNEMIE
CISZ_RHITR      YFIEKKLYPN VDFYSGITLK ALGFPTTMFT VLFALARTVG WIAQWNEMIE
CISY_ACEAC      YFVQRKLYPN VDFYSGIILK AMGIPTSMFT VLFAVARTTG WVSQWKEMIE
CISY_RICAU      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICPA      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICSL      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICSI      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICBE      YFIERKLYPN VDFYSGIIYK AMGIPPQMFT VLFATARTVG WMAQWKEMHE
CISY_ASTRI      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICAK      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICAF      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFVIARTVG WMAQWKEMHE
CISY_RICRH      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICMA      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICCA      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICCN      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICTY      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_RICJA      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_COXBU      YFIEKKLYPN VDFYSGLTLN AIGIPSNMFT VIFALSRTVG WISHWMEMMS
CISY_RICHE      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTIG WMAQWKEMHG
CISY_RICPR      YFIERKLYPN VDFYSGIIYK AMGIPSQMFT VLFAIARTVG WMAQWKEMHE
CISY_MYCTU      YFIERKLYPN VDFYTGLIYR ALGFPTRMFT VLFALGRLPG WIAHWREMHD
CISY_CORGL      YFISRKLYPN VDFYTGLIYR AMGFPTDFFT VLFAIGRLPG WIAHYREQGA
CYSZ_CUCMA      YFVKRKLYPN VDFYSGLIYR AMGFPPEFFT VLFAIPRMAG YLAHWRESLD
CISY_HELPY      YFIERNLYPN VDFYSGTILR ALKIPVRFFT PVFVIGRTVG WCAQLLEHKS
CISY_THIFE      RLGAQGIHAN VDFYSGVLYH EMGIKADLFT PIFAMARSAG WLAHWREQLA
CISZ_BACSU      ...EKKLPPN VDFYSASVYH SLGIDHDLFT PIFAVRRMSG WLAHILE.QY
CISZ_STRMU      ...KKNLIPN VDFYSATVYH VLGIDSSIFT LIFAMSRVSG WIAHIQEQQK
CISY_STRHY      .FSSKGIYPN TDFYSGIVFY ALGFPVYMFT ALFALSRTLG WLAHIIEYVE
CISY_BACCO      ...EKGLFPN LDYYAAPVYW KLGIPIPLYT PIFFSSRTVG LCAHVMEQHE
CISW_BACSU      ...EKGLYPN LDYYAAPVYW MLGIPIPLYT PIFFSARTSG LCAHVIEQHA
CISY_THEAC      .FGSKGIYPN TDYFSGIVYM SIGFPLNIYT ALFALSRVTG WQAHFIEYVE
CISZ_ECOLI      MWESKKMFPN LDWFSAVSYN MMGVPTEMFT PLFVIARVTG WAAHIIEQRQ
CISY_BACSU      YKPGRKLYTN VEFYAAAVMR AIDFDDELFT PTFSASRMVG WCAHVLEQAE
CISZ_SALTY      ...TKKMFPN LDWFSAVSYN MMGVPTEMFT PLFVIARVTG WAAHIIEQRQ
CISY_MYCSM      MFAATRIKPN LDFPTGPAYY LMDFPIESFT PLFVMSRITG WTAHIMEQAA
CISY_SCHPO      HGKTKNPYPN VDSHSGVLLQ YYGLKESFYT VLFGVSRTLG VASQLIWDRA
CISZ_MYCTU      RRPDRAIETN VEFWAAVVLD FARVPANMMP AMFTCGRTAG WCAHILEQKR
CISY_CANTR      HGMTKNPWPN VGSHSGVLLQ YYGLTESFYT VLFGVSRAFG VLPQLILDRG
CISY_NEUCR      HGKTKNPYPN VDAHSGVLLQ HYGLTENYYT VLFGVSRAIG VLPQLIIDRA
CISY_CHICK      ..AAANPWPN VDAHSGVLLQ YYGMTENYYT VLFGVSRALG VLAQLIWSRA
CISY_CAEEL      ..KAKNPWPN VDSHSGVLLQ YFGMTESFYT VLFGVSRALG CLSQLIWARG
CISY_YEAST      HGKTKNPWPN VDSHSGVLLQ YYGLTESFYT VLFGVARAIG VLPQLIIDRA
CISY_EMENI      HGKTKNPYPN VDAHSGVLLQ YYGLTENYYT VLFGVSRALG VLPQLIIDRA
CISY_CITMA      LGKVKNPWPN VDAHSGVLLN HFGLAERYYT VLFGVSRSLG ICSQLIWDRA
CISY_SYNY3      YKKIANPYPN VDCHSGVLLY SLGLTEQYYT VVFAVSRALG CMANLIWSRA
CISY_PIG        ..KAKNPWPN VDAHSGVLLQ YYGMTENYYT VLFGVSRALG VLAQLIWSRA
CISZ_YEAST      HGKTKNPWPN VDAHSGVLLQ YYGLKESFYT VLFGVSRAFG ILAQLITDRA
CISY_ASPNG      HGKTKNPYPN VDAHSGVLLQ YYGLTENYYT VLFGVSRALG VLPQLIIDRA
CISX_YEAST      HGKSKNPFPN VDSASGILFY HYGIRELFFT VIFGCSRAMG PLTQLVWDRI
CISY_ARATH      ELESEEPWPN VDAHSGVLLN HYGLTERYYT VLFGVSRSLG ICSQLIWDRE
```

```
                401                         426
CISY_ECOLI      DGMKIARPRQ LYTGYEKRDF KSDIKR
CISY_BARVB      .......... .......... ......
CISY_RICAZ      .......... .......... ......
CISY_BARDO      .......... .......... ......
CISY_BARVI      .......... .......... ......
CISY_BAREL      .......... .......... ......
CISY_BARTA      .......... .......... ......
CISY_BARQU      .......... .......... ......
CISY_BARBA      .......... .......... ......
CISY_BARHE      DPQKISRPRQ LYTGYAAREY IPIDKR
CISY_PSEAE      GPYKIGRPRQ LYTGHTQRDF TALKDR
CISY_ACIAN      GPYKIGRPRQ LYTGEVQRDI KR....
CISY_RHITR      DPQRIGRPRH VYTGAPLREY VPVSKR
CISZ_RHITR      DPQRIGRPRQ LYTGAPLREY VPLSKR
CISY_ACEAC      EGQRISRPRQ LYIGAPQRDY VPLAKR
CISY_RICAU      DPQKISRPR. .......... ......
CISY_RICPA      DPQKISRPR. .......... ......
CISY_RICSL      DPQKISRPR. .......... ......
CISY_RICSI      DPQKISRPR. .......... ......
CISY_RICBE      DPQKISRPR. .......... ......
CISY_ASTRI      DPQKISRPR. .......... ......
CISY_RICAK      DGQKISRPR. .......... ......
CISY_RICAF      DPQKISRPR. .......... ......
CISY_RICRH      DPQKISRPR. .......... ......
CISY_RICMA      DPQKISRPR. .......... ......
CISY_RICCA      DPQKISRPR. .......... ......
CISY_RICCN      DPQKISRPR. .......... ......
CISY_RICTY      DPQKISRPR. .......... ......
CISY_RICJA      DPQKISRPR. .......... ......
CISY_COXBU      SPHRLARPRQ LYTGETEREV ISLDKR
CISY_RICHE      DPQKISRPR. .......... ......
CISY_RICPR      DPQKISRPRQ LYTGYVHREY KCIVER
CISY_MYCTU      EGSKIGRPRQ IYTGYTERDY VTIDAR
CISY_CORGL      AGNKINRPRQ VYTGNESRKL VPREER
CYSZ_CUCMA      DPTKIIRPQQ VYTGEWLRHY IPPNER
CISY_HELPY      PQARITRPRQ VYVG...... ......
CISY_THIFE      D.NRIFRPTQ VYTGEQDRRY VPVAQR
CISZ_BACSU      DNNRLIRPRA DYTGPDKQKF VPIEER
CISZ_STRMU      N.NKLIRPRS HYTGMRKLRY IPIERR
CISY_STRHY      EQHRLIRPRA LYVGPEYQEY VSIDKR
CISY_BACCO      N.NRIFRPRV LYTG...... ......
CISW_BACSU      N.NRLFRPRV SYMG...... ......
CISY_THEAC      EQQRLIRPRA VYVGPAERKY VPIAER
CISZ_ECOLI      D.NKIIRPSA NYVGPEDRPF VALDKR
CISY_BACSU      N.NMIFRPSA QYTGAIPEEV LS....
CISZ_SALTY      D.NKIIRPSA NYTGPEDRPF VSIDDR
CISY_MYCSM      S.NALIRPLS EYSGQPQRSL ......
CISY_SCHPO      LGLPIERPKS FSTEALKKMV ET....
CISZ_MYCTU      LG.KLVRPSA IYVGPGPRSP ES....
CISY_CANTR      LGMPIERPKS FSTEKYIELV KSIGK.
CISY_NEUCR      VGAPIERPKS YST....... ......
CISY_CHICK      LGFPLERPKS MST....... ......
CISY_CAEEL      MGLPLERPKS HST....... ......
CISY_YEAST      VGAPIERPKS FSTEKYKELV K.....
CISY_EMENI      FGAPIERPKS FSTEAYAKLV GAKL..
CISY_CITMA      LGLPLERPKS VTLDWIEKNC K.....
CISY_SYNY3      FGLPIERPGS ADLKWFHDKY RE....
CISY_PIG        LGFPLERPKS MSTDGLIKLV DS....
CISZ_YEAST      IGASIERPKS YSTEKYKELV KN....
CISY_ASPNG      LGAPIERPKS YSTELSPSLL VLSCK.
CISX_YEAST      LGLPIERPKS L......... ......
CISY_ARATH      LLLALERPKS V......... ......
```

(SWISS-PROT Release 34 and updates up to 10-Oct-1997)
Description of the citrate synthases in the sequence alignment obtained from SWISS-PROT

CISW_BACSU	CS III BACILLUS SUBTILIS
CISX_YEAST	CS 3 SACCHAROMYCES CEREVISIAE (BAKER'S YEAST)
CISY_ACEAC	CS ACETOBACTER ACETI
CISY_ACIAN	CS ACINETOBACTER ANITRATUM
CISY_ARATH	CS MITOCHONDRIAL PRECURSOR ARABIDOPSIS THALIANA
CISY_ASPNG	CS MITOCHONDRIAL PRECURSOR ASPERGILLUS NIGER
CISY_ASTRI	CS (FRAGMENT) ASTRAKHAN RICKETTSIA
CISY_BACCO	CS BACILLUS COAGULANS
CISY_BACSU	CS I BACILLUS SUBTILIS
CISY_BARBA	CS (FRAGMENT) BARTONELLA BACILLIFORMIS
CISY_BARDO	CS (FRAGMENT) BARTONELLA DOSHIAE
CISY_BAREL	CS (FRAGMENT) BARTONELLA ELIZABETHAE
CISY_BARHE	CS BARTONELLA HENSELAE
CISY_BARQU	CS (FRAGMENT) BARTONELLA QUINTANA
CISY_BARTA	CS (FRAGMENT) BARTONELLA TAYLORII
CISY_BARVB	CS (FRAGMENT) BARTONELLAVINSONII BERKHOFII
CISY_BARVI	CS (FRAGMENT) {GENE: GLTA} - BARTONELLA VINSONII
CISY_CAEEL	CS MITOCHONDRIAL PRECURSOR CAENORHABDITISELEGANS
CISY_CANTR	CS MITOCHONDRIAL PRECURSOR CANDIDA TROPICALIS
CISY_CHICK	CS MITOCHONDRIAL GALLUS GALLUS (CHICKEN)
CISY_CITMA	CS MITOCHONDRIAL PRECURSOR CITRUS MAXIMA (PUMMELO)
CISY_CORGL	CS CORYNEBACTERIUM GLUTAMICUM
CISY_COXBU	CS COXIELLA BURNETII
CISY_ECOLI	CS ESCHERICHIA COLI
CISY_EMENI	CS MITOCHONDRIAL PRECURSOR EMERICELLA NIDULANS
CISY_HELPY	CS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI)
CISY_MYCSMCS	MYCOBACTERIUM SMEGMATIS
CISY_MYCTU	CS 1 MYCOBACTERIUM TUBERCULOSIS
CISY_NEUCR	CS, MITOCHONDRIAL PRECURSOR NEUROSPORA CRASSA
CISY_PIG	CS MITOCHONDRIAL PRECURSOR SUS SCROFA (PIG)
CISY_PSEAE	CS PSEUDOMONAS AERUGINOSA
CISY_RHITR	CS CHROMOSOMAL RHIZOBIUM TROPICI

CISY_RICAF	CS (FRAGMENT) RICKETTSIA AFRICAE
CISY_RICAK	CS (FRAGMENT)RICKETTSIA AKARI
CISY_RICAU	CS (FRAGMENT) RICKETTSIA AUSTRALIS
CISY_RICAZ	CS (FRAGMENT) RICKETTSIA AZADI
CISY_RICBE	CS (FRAGMENT) RICKETTSIA BELLII
CISY_RICCA	CS (FRAGMENT) RICKETTSIA CANADA
CISY_RICCN	CS (FRAGMENT) RICKETTSIA CONORII
CISY_RICHE	CS (FRAGMENT) RICKETTSIA HELVETICA
CISY_RICJA	CS (FRAGMENT) RICKETTSIA JAPONICA
CISY_RICMA	CS (FRAGMENT) RICKETTSIA MASSILIAE
CISY_RICPA	CS (FRAGMENT) RICKETTSIA PARKERI
CISY_RICPR	CS RICKETTSIA PROWAZEKII
CISY_RICRH	CS (FRAGMENT) RICKETTSIA RHIPICEPHALI
CISY_RICSI	CS (FRAGMENT) RICKETTSIA SIBIRICA
CISY_RICSL	CS (FRAGMENT) RICKETTSIA SLOVACA
CISY_RICTY	CS (FRAGMENT) RICKETTSIA TYPHI
CISY_SCHPO	CSMITOCHONDRIAL PRECURSOR SCHIZOSACCHAROMYCES POMBE
CISY_STRHY	CS (FRAGMENT) STREPTOMYCES HYGROSCOPICUS
CISY_SULSO	CS SULFOLOBUS SOLFATARICUS
CISY_SYNY3	CS SYNECHOCYSTIS SP. (STRAIN PCC 6803)
CISY_THEAC	CS THERMOPLASMA ACIDOPHILUM
CISY_THIFE	CS THIOBACILLUS FERROOXIDANS
CISY_YEAST	CSMITOCHONDRIAL PRECURSOR SACCHAROMYCES CEREVISIAE
CISZ_BACSU	CS II BACILLUS SUBTILIS
CISZ_ECOLI	POSSIBLE CS 2 ESCHERICHIA COLI
CISZ_MYCTU	PUTATIVE CS 2 MYCOBACTERIUM TUBERCULOSIS
CISZ_RHITR	CS PLASMID RHIZOBIUM TROPICI
CISZ_SALTY	POSSIBLE CS 2 SALMONELLA TYPHIMURIUM
CISZ_STRMU	CS STREPTOCOCCUS MUTANS
CISZ_YEAST	CS PEROXISOMAL SACCHAROMYCES CEREVISIAE (BAKER'S YEAST)
CYSZ_CUCMA	CS GLYOXYSOMAL PRECURSOR CUCURBITA MAXIMA (PUMPKIN)

APPENDIX II *gltA* Sequence

```
          1100
               *          *          *          *          *          *
TTA ACG CTT GAT ATC GCT TTT AAA GTC GCG TTT TTC ATA TCC TGT ATA CAG CTG ACG CGG
AAT TGC GAA CTA TAG CGA AAA TTT CAG CGC AAA AAG TAT AGG ACA TAT GTC GAC TGC GCC
<*   R   K   I   D   S   K   F   D   R   K   E   Y   G   T   Y   L   Q   R   P

                                                                    1200
               *          *          *          *          *          *
ACG GGC AAT CTT CAT ACC GTC ACT GTG CAT TTC GCT CCA GTG GGC GAT CCA GCC AAC GGT
TGC CCG TTA GAA GTA TGG CAG TGA CAC GTA AAG CGA GGT CAC CCG CTA GGT CGG TTG CCA
<R   A   I   K   M   G   D   S   H   M   E   S   W   H   A   I   W   G   V   T

               *          *          *          *          *          *
ACG TGC CAT TGC GAA AAT GAC GGT GAA CAT GGA AGA CGG AAT ACC CAT CGC TTT CAG GAT
TGC ACG GTA ACG CTT TTA CTG CCA CTT GTA CCT TCT GCC TTA TGG GTA GCG AAA GTC CTA
<R   A   M   A   F   I   V   T   F   M   S   S   P   I   G   M   A   K   L   I

                                    1300
               *          *          *          *          *          *
GAT ACC AGA GTA GAA ATC GAC GTT CGG GTA CAG TTT CTT CTC GAT AAA GTA CGG GTC GTT
CTA TGG TCT CAT CTT TAG CTG CAA GCC CAT GTC AAA GAA GAG CTA TTT CAT GCC CAG CAA
<I   G   S   Y   F   D   V   N   P   Y   L   K   K   E   I   F   Y   P   D   N

               *          *          *          *          *          *
CAG CGC GAT GTT TTC CAG CTC CAT AGC CAC TTC CAG CAG GTC ATC CTT CGT GCC CAG CTC
GTC GCG CTA CAA AAG GTC GAG GTA TCG GTG AAG GTC GTC CAG TAG GAA GCA CGG GTC GAG
<L   A   I   N   E   L   E   M   A   V   E   L   L   D   D   K   T   G   L   E

          1400
               *          *          *          *          *          *
TTT CAG CAC TTC ATG GCA GGT TTC ACG CAT TAC GGT GGC GCG CGG GTC GTA ATT TTT GTA
AAA GTC GTG AAG TAC CGT CCA AAG TGC GTA ATG CCA CCG CGC GCC CAG CAT TAA AAA CAT
<K   L   V   E   H   C   T   E   R   M   V   T   A   R   P   D   Y   N   K   Y

                                                                    1500
               *          *          *          *          *          *
CAC GCG GTG ACC GAA GCC CAT CAG GCG GAA AGA ATC ATT TTT GTC TTT CGC ACG ACG AAC
GTG CGC CAC TGG CTT CGG GTA GTC CGC CTT TCT TAG TAA AAA CAG AAA GCG TGC TGC TTG
<V   R   H   G   F   G   M   L   R   F   S   D   N   K   D   K   A   R   R   V

               *          *          *          *          *          *
AAA TTC CGG AAT GTG TTT AAC GGA GCT GAT TTC TTC CAG CAT TTT CAG CGC CGC TTC GTT
TTT AAG GCC TTA CAC AAA TTG CCT CGA CTA AAG AAG GTC GTA AAA GTC GCG GCG AAG CAA
<F   E   P   I   H   K   V   S   S   I   E   E   L   M   K   L   A   A   E   N

                                    1600
               *          *          *          *          *          *
AGC ACC GCC GTG CGC AGG TCC CCA CAG TGA AGC AAT ACC TGC TGC GAT ACA GGC AAA CGG
TCG TGG CGG CAC GCG TCC AGG GGT GTC ACT TCG TTA TGG ACG ACG CTA TGT CCG TTT GCC
<A   G   G   H   A   P   G   W   L   S   A   I   G   A   A   I   C   A   F   P

               *          *          *          *          *          *
GTT CGC ACC CGA AGA GCC AGC GGT ACG CAC GGT GGA GGT AGA GGC GTT CTG TTC ATG GTC
CAA GCG TGG GCT TCT CGG TCG CCA TGC GTG CCA CCT CCA TCT CCG CAA GAC AAG TAC CAG
<N   A   G   S   S   G   A   T   R   V   T   S   T   S   A   N   Q   E   H   D

          1700
               *          *          *          *          *          *
AGC GTG CAG GAT CAG AAT ACG GTC CAT AGC ACG TTC CAG AAT CGG ATT AAC TTC ATA CGG
TCG CAC GTC CTA GTC TTA TGC CAG GTA TCG TGC AAG GTC TTA GCC TAA TTG AAG TAT GCC
<A   H   L   I   L   I   R   D   M   A   R   E   L   I   P   N   V   E   Y   P
```

```
                                                                  1800
                *         *         *         *         *         *
TTC GCA CGG CGT GGA GAA CAT CAT ATT CAG GAA GTT ACC GGC GTA GGA GAG ATC GTT GCG
AAG CGT GCC GCA CCT CTT GTA GTA TAA GTC CTT CAA TGG CCG CAT CCT CTC TAG CAA CGC
<E   C   P   T   S   F   M   M   N   L   F   N   G   A   Y   S   L   D   N   R

                *         *         *         *         *         *
CGG GTA AAC AAA TGG CTG ACC AAT GGA ATA CTT GTA ACA CAT CGC GGC CAT GGT CGG CAT
GCC CAT TTG TTT ACC GAC TGG TTA CCT TAT GAA CAT TGT GTA GCG CCG GTA CCA GCC GTA
<P   Y   V   F   P   Q   G   I   S   Y   K   Y   C   M   A   A   M   T   P   M

                                  1900
                *         *         *         *         *         *
TTT CGA CAG CAG GCG GAA CGC GGC AAT TTC ACG GTG ACG AGG ATT GTT AAC ATC CAG CGA
AAA GCT GTC GTC CGC CTT GCG CCG TTA AAG TGC CAC TGC TCC TAA CAA TTG TAG GTC GCT
<K   S   L   L   R   F   A   A   I   E   R   H   R   P   N   N   V   D   L   S

                *         *         *         *         *         *
GTC GTG ATA GAA CGC CGC CAG CGC GCC GGT AAT ACC ACA CAT GAC TGC CAT TGG ATG CGA
CAG CAC TAT CTT GCG GCG GTC GCG CGG CCA TTA TGG TGT GTA CTG ACG GTA ACC TAC GCT
<D   H   Y   F   A   A   L   A   G   T   I   G   C   M   V   A   M   P   H   S

          2000
                *         *         *         *         *         *
GTC GCG ACG GAA AGC ATG GAA CAG ACG GGT AAT CTG CTC GTG GAT CAT GGT ATG ACG GGT
CAG CGC TGC CTT TCG TAC CTT GTC TGC CCA TTA GAC GAG CAC CTA GTA CCA TAC TGC CCA
<D   R   R   F   A   H   F   L   R   T   I   Q   E   H   I   M   T   H   R   T

                                                                  2100
                *         *         *         *         *         *
CAC CGT AGT TTT AAA TTC GTC ATA CTG TTC CTG AGT CGG TTT TTC ACC ATT CAG CAG GAT
GTG GCA TCA AAA TTT AAG CAG TAT GAC AAG GAC TCA GCC AAA AAG TGG TAA GTC GTC CTA
<V   T   T   K   F   E   D   Y   Q   E   Q   T   P   K   E   G   N   L   L   I

                *         *         *         *         *         *
GTA ACA AAC TTC CAG GTA GTT AGA ATC GGT CGC CAG CTG ATC GAT CGG GAA ACC GCG GTG
CAT TGT TTG AAG GTC CAT CAA TCT TAG CCA GCG GTC GAC TAG CTA GCC CTT TGG CGC CAC
<Y   C   V   E   L   Y   N   S   D   T   A   L   Q   D   I   P   F   G   R   H

                                  2200
                *         *         *         *         *         *
CAG CAA AAT ACC TTC ATC ACC ATC AAT AAA AGT AAT TTT AGA TTC GCA GGA TGC GGT TGA
GTC GTT TTA TGG AAG TAG TGG TAG TTA TTT TCA TTA AAA TCT AAG CGT CCT ACG CCA ACT
<L   L   I   G   E   D   G   D   I   F   T   I   K   S   E   C   S   A   T   S

                *         *         *         *         *         *
AGT GAA GCC TGG GTC AAA GGT GAA CAC ACC TTT TGA ACC GAG AGT ACG GAT ATC AAT AAC
TCA CTT CGG ACC CAG TTT CCA CTT GTG TGG AAA ACT TGG CTC TCA TGC CTA TAG TTA TTG
<T   F   G   P   D   F   T   F   V   G   K   S   G   L   T   R   I   D   I   V

          2300
                *         *         *         *         *         *
ATC TTG ACC CAG CGT GCC TTT CAG CAC ATC CAG TTC AAC AGC TGT ATC CCC GTT GAG GGT
TAG AAC TGG GTC GCA CGG AAA GTC GTG TAG GTC AAG TTG TCG ACA TAG GGG CAA CTC CCA
<D   Q   G   L   T   G   K   L   V   D   L   E   V   A   T   D   G   N   L   T

                                                                  2400
                *         *         *         *         *         *
GAG TTT TGC TTT TGT ATC AGC CAT TT AAG GTC TCC TTA GCG CCT TAT TGC GTA AGA CTG
CTC AAA ACG AAA ACA TAG TCG GTA AA TTC CAG AGG AAT CGC GGA ATA ACG CAT TCT GAC
<L   K   A   K   T   D   A   M
```

APPENDIX III Free Energy of *N* and *I* as a Function of Urea

Using the denaturation data fitting results obtained for WT, we can model the free energy profile of the denaturation in order to understand how different scenarios (such as destabilization of intermediate) affects the denaturation curve. For this purpose, we need the stabilities of the Native and Intermediates expressed relative to one reference state, the Unfolded state. So the stability of the unfolded state is always 0 (Robson & Pain, 1976). We already have determined the stability of the intermediate state relative to the unfolded state from the data fitting. For wild type, ΔG_{I-U} =-10.45 kcal.mol^{-1}. However, we do not have, from the data fitting, a value for ΔG_{N-U}. But this can be calculated, if we convince ourselves as follows:

The equilibrium constants are

$$K_{N-I} = {}^{I}/_{N} \qquad [1]$$

$$K_{I-U} = {}^{U}/_{I} \qquad [2]$$

$$K_{N-U} = {}^{U}/_{N} \qquad [3]$$

From equations 2 and 3, we obtain

$$K_{N-U} = {}^{K_{I-U} \cdot I}/_{N} \qquad [4]$$

and using equation 1, we obtain

$$K_{N-U} = K_{I-U} \cdot K_{N-I} \qquad [5]$$

Therefore from ΔG=-RTln$K$,

$$\Delta G_{N-U} = -RT \ln(K_{I-U} \bullet K_{N-I})$$

or

$$\Delta G_{N-U} = \Delta G_{I-U} + \Delta G_{N-I} = -10.45 + -1.8\ kcal.mol^{-1} = -12.25\ kcal.mol^{-1}$$

The apparent free energy for the Native to Unfolded transition is related to the sum of the slopes of the individual transitions, since,

$$\Delta G_{apparent} = \Delta G_0 - m[urea] \qquad [6]$$

and so for the N to U transition,

$$\Delta G_{apparent} = \Delta G_{N\text{-}I} - m_1[urea] + \Delta G_{I\text{-}N} - m_2[urea]$$

$$\Delta G_{apparent} = (\Delta G_{N\text{-}I} + \Delta G_{I\text{-}N}) - [urea](m_1 + m_2)$$

$$\Delta G_{apparent} = \Delta G_{N\text{-}U} - [urea](m_1 + m_2) \qquad [7]$$

Then the apparent ΔG as a function of urea concentration may be generated for the Native and Intermediate states using equations 7 and 6, respectively, using the stabilities and slopes obtained from the denaturation data fits. The data was generated for wild type CS, and is presented in Figure I.

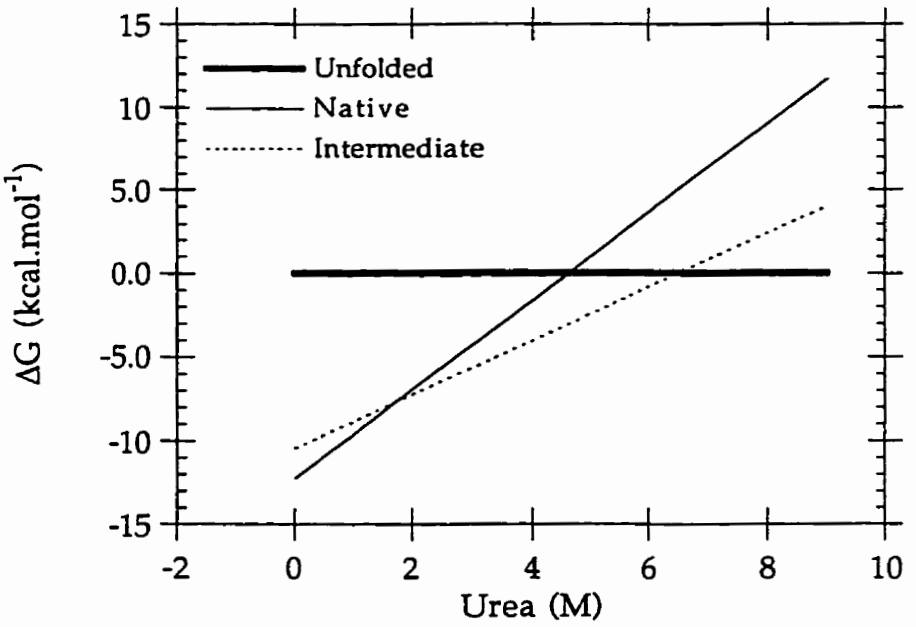


Figure I The free energies of Native and Intermediate states of WT *E. coli* CS as a function of urea concentration.

In Figure I, the points at which the intermediate line intersects the Native and unfolded lines (i.e. when they have equal ΔG values) corresponds

to the transition midpoints (1.8 and 6.49 M urea). The point at which the native line intersects the unfolded reference corresponds to the end and beginning of transitions I and II, respectively. The relative positions of the transition midpoints dictate whether or not an intermediate is observed; if the transition midpoints for I to U and N to U are the same, then the intermediate would not be detected.

REFERENCES

Adler, A., Greenfield, N. J., Fasman, G.D. (1973) *Methods in Enzymlogy 27,* 675-735.

Anderson, D., Donald, L., Jacob, M., & Duckworth, H. (1991) *Biochemistry and Cell Biology 69,* 323-238.

Anderson, D. H., & Duckworth, H. W. (1988) *The Journal of Biological Chemistry 263,* 2163-2169.

Anfinsen, C. B. (1973) *Science 181,* 223-230.

Arcus, V., Vuilleumier, S., Freund, S., Bycroft, M., & Fersht, A. (1994) *Proceedings of the National Academy of Sciences USA 91,* 9412-9416.

Bernal, J. (1939) *Nature 143,* 663-667.

Birtles, R., & Raoult, D. (1996) *International Journal of Systematic Bacteriology 46,* 891-897.

Bradford, M. (1976) *Analytical Biochemistry 72,* 248-254.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D., Joachimiak, A., Horwich, A., & Sigler, P. (1994) *Nature 371,* 578-586.

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F., & Kiefhaber, T. (1991) *Biochemistry 30,* 1586-1591.

Buckle, A., Cramer, P., & Fersht, A. (1996) *Biochemistry 35,* 4298-4305.

Chernushevich, I. V., Ens, W., & Standing, K. G. (1997) in *Electrospray Ionization Mass Spectrometry* (Cole, R. B., Ed.) pp 203-234, John Wiley and Sons, Inc., New York.

Chothia, C. (1992) *Nature 357,* 543-544.

Chowdhury, S., Katta, V., & Chait, B. (1990) *Journal of the American Chemical Society 112,* 9012-9013.

Christensen, H., & Pain, R. (1994) in *Mechanisms of Protein Folding* (Pain, R. H., Ed.) pp 55-89, IRL Press, Oxford.

Conejero-Lara, F., Parrado, J., Azuaga, A., Smith, R., Ponting, C., & Dobson, C. (1996) *Protein Science 5*, 2583-2591.

Creighton, T. (1993) *Proteins. Structure and Molecular Properties*, Second edition ed., W. H. Freeman and Company, New York.

Creighton, T. E. (1988) in *Practical Approach Series*, IRL Press, Oxford.

Dill, K. (1990) *Biochemistry 29*, 7133-7155.

Donald, L., & Duckworth, H. (1987) *Biochemistry and Cell Biology 65*, 930-938.

Donald, L. J., Crane, B. R., Anderson, D. H., & Duckworth, H. W. (1991) *The Journal of Biological Chemistry 266*, 20709-20713.

Duckworth, H., Anderson, D., Bell, A., Donald, L., Chu, A., & Brayer, G. (1987) in *Biochemical Society Symposia* (Kay, J., & Weitzman, P., Eds.) pp 83-92.

Duckworth, H. W., & Bell, A. W. (1982) *Canadian Journal of Biochemistry 60*, 1143-1147.

Duckworth, H. W., & Tong, E. K. (1976) *Biochemistry 15*, 108-114.

Edelhoch, H. (1967) *Biochemistry 6*, 1948-1954.

Ehrnsperger, M., Gräber, S., Gaestel, M., & Buchner, J. (1997) *The EMBO Journal 16*, 221-229.

Eriksson, A., Baase, W., Zhang, X.-J., Heinz, D., Blaber, M., Baldwin, E., & Matthews, B. (1992) *Science 255*, 178-183.

Evans, C., Kurz, L., Remington, S., & Srere, P. (1996) *Biochemistry 35*, 10661-10672.

Faloona, G. R., & Srere, P. A. (1969) *Biochemistry 8*, 4497-4503.

Fauchére, J.-L., & Pliska, V. (1983) *European Journal of Medicinal Chemsitry 18*, 369-375.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989) *Science 246*, 64-71.

Ganem, B., Li, Y.-T., & Henion, J. (1991) *Journal of the American Chemical Society 113*, 6294-6296.

Gokhale, R., Agarwalla, S., Santi, D., & Balram, P. (1996) *Biochemistry 35*, 7150-7158.

Goldenberg, D. P. (1988) in *Protein structure. A practical approach* (Creighton, T. E., Ed.), IRL press at Oxford University Press, New York.

Greig, M., Gaus, H., Cummins, L., Sasmor, H., & Griffey, R. (1995) *Journal of the American Chemical Society 117*, 10765-10766.

Guest, J. R. (1981) *Journal of General Microbiolgy 124*, 17-23.

Guha, S., & Bhattacharyya, B. (1995) *Biochemistry 34*, 6925-6930.

Hayashi-Iwasaki, Y., Numata, K., Yamagishi, A., Yutani, K., Sakurai, M., Tanaka, N., & Oshima, T. (1996) *Protein Science 5*, 511-516.

Hofstadler, S., Bakhtiar, R., & Smith, R. (1996) *Journal of Chemical Education 73*, A83-A88.

Horwich, A. L., Low, K. B., Fenton, W. A., Hirshfield, I. N., & Furtak, K. (1993) *Cell 74*, 909-917.

Jaenicke, R. (1991) *Biochemistry 30*, 3147-3161.

Jaenicke, R. (1996) *FASEB Journal 10*, 84-92.

Juranic, N., Ilich, P., & Macura, S. (1995) *Journal of the American Chemical Society 117*, 405-410.

Karash, M., & Hillenkamp, F. (1988) *Analytical Chemistry 60*, 2299-2301.

Karpusas, M., Holland, D., & Remington, S. (1991) *Biochemistry 30*, 6024-6031.

Kauzmann, W. (1959) *Advances in Protein Chemistry 14*, 1-63.

Kelly, S., & Price, N. (1992) *International Journal of Biochemistry 24*, 627-630.

Kim, & Baldwin. (1982) *Annual Review of Biochemistry 51,* 459-489.

Koh, J., Cornish, V., & Schultz, P. (1997) *Biochemistry 36,* 11314-11322.

Kolata, G. (1986) *Science 233,* 1037-1039.

Konerman, L., & Douglas, D. (1997) *Biochemistry 36,* 12296-12302.

Konermann, L., Collings, B., & Douglas, D. (1997a) *Biochemistry* , 5554-5559.

Konermann, L., Rosell, F., Mauk, A., & Douglas, D. (1997b) *Biochemistry 36,* 6448-6454.

Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods in Enzymology 154,* 367-382.

Kuwajima, K. (1989) *Proteins 6,* 87-103.

Kuwajima, K. (1996) *FASEB Journal 10,* 102-109.

Lakowicz, J. R. (1983) *Principles of Fluorescence Spectroscopy,* Plenum Press, New York.

Lesk, A., & Chothia, C. (1984) *Journal of Molecular Biology 174,* 175-191.

Lesser, G., & Rose, G. (1990) *Proteins: Structure Function and Genetics 8,* 6-13.

Levinthal, C. (1968) *J. Chim. Phys. 65,* 44-45.

Liao, D.-I., Karpusas, M., & Remington, S. (1991) *Biochemistry 30,* 603`-6036.

Light-Wahl, J., Schwartz, B., & Smith, R. (1994) *Journal of the American Chemical Society 116,* 5271-5278.

Lim, H.-K., Hsieh, Y., Ganem, B., & Henion, J. (1995) *Journal of Mass Spectrometry 30,* 708-714.

Logan, T., Thériault, Y., & Fesik, S. (1994) *Journal of Molecular Biology 236,* 637-648.

Loo, J. (1995) *Journal of Mass Spectrometry 30,* 180-183.

Loo, J. A. (1997) *Mass Spectrometry Reviews 16,* 1-23.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular cloning. A laboratory manual,* Cold Spring Harbor, New York.

Mann, M., & Wilm, M. (1995) *Trends in Biochemical Sciences 20,* 219-224.

Matthews, B. (1996) *FASEB Journal 10,* 35-41.

Mirza, U., Cohen, S., & Chait, B. (1993) *Analytical Chemistry 65,* 1-6.

Molgat, G. F., Donald, L. J., & Duckworth, H. W. (1992) *Archives of Biochemistry and Biophysics 298,* 238-246.

Monod, J., Wyman, J., & Changeux, J.-P. (1965) *Journal of Molecular Biology 12,* 88-118.

Morse, D., & Duckworth, H. (1980) *Canadian Journal of Biochemistry 58,* 696-706.

Ner, S. S., Bhayana, V., Bel, A. W., Giles, I. G., Duckworth, H. W., & Bloxham, D. P. (1983) *Biochemistry 22,* 5243-5249.

Neri, D., Billeter, M., Wider, G., & Wüthrich, K. (1992) *Science 257,* 1559-1563.

Orr, A., Ivanova, VS, Bonne, WM. (1995) *Biotechniques 19,* 204-206.

Pace, C. (1992) *Journal of Molecular Biology 226,* 29-35.

Pace, C. (1995) *Methods in Enzymology 259,* 538-554.

Pace, C., Shirley, B., & J.A., T. (1988) in *Protein structure* (Creighton, T., Ed.) pp 311-330, IRL press, Oxford.

Pace, C., Shirley, B., McNutt, M., & Gajiwala, K. (1996) *FASEB Journal 10,* 75-83.

Pace, C. N. (1986) *Methods in Enzymology 131,* 266-280.

Patton, A., Hough, D., Towner, P., & Danson, M. (1993) *European Journal of Biochemistry 214,* 75-81.

Pereira, D. S., Donald, L. J., Hosfield, D. J., & Duckworth, H. W. (1994) *The Journal of Biological Chemistry 269,* 412-417.

Potier, N., Donald, L., Chernushevich, I., Ayed, A., Ens, W., Arrowsmith, C., Standing, K., & Duckworth, H. (1997) *Protein Science.* In Press.

Ptitsyn, O. (1992) in *Protein Folding* (Creighton, T., Ed.) pp 243-300, W. H. Freeman and Company, New York.

Ptitsyn, O. (1995) *Trends in Biochemical Sciences 20,* 376-379.

Remington, S., Wiegand, G., & Huber, R. (1982) *Journal of Molecular Biology 158,* 111-152.

Robinson, C., Chung, E., Kragelund, B., Knudsen, J., Aplin, R., Poulsen, F., & Dobson, C. (1996) *Journal of the American Chemical Society 118,* 8646-8653.

Robson, B., & Pain, R. H. (1976) *Biochemical Journal 155,* 331-344.

Rost, B., & Sander, C. (1993) *Journal of Molecular Biology 232,* 584-599.

Russel, R., Ferguson, J., Hough, D., Danson, M., & Taylor, G. (1997) *Biochemistry 36,* 9983-9994.

Russell, R., Hough, D., Danson, M., & Taylory, G. (1994) *Structure 2,* 1157-1167.

Sánchez del Pino, M., & Fersht, A. (1997) *Biochemistry 36,* 5560-5565.

Santoro, M., & Bolen, D. (1988) *Biochemistry 27,* 8063-8068.

Schmid, F. (1988) in *Protein structure* (Creighton, T., Ed.) pp 251-286, IRL Press, Oxford.

Semisotnov, G., Rodionova, NA, Razgulyaev, OI, Uversky, V. N., Gripas, A. F., & Gilmanshin, R. I. (1991) *Biopolymers 31,* 119-128.

Sharp, K., Nicholls, A., Friedman, R., & Honig, B. (1991) *Biochemistry 30,* 9686-9697.

Sherman, M., Beechem, J., & Mas, M. (1995) *Biochemistry 34,* 13934-13942.

Shoichet, B., Baase, W., Kuroki, R., & Matthews, B. (1995) *Proceedings of the National Academy of Science USA 92,* 452-456.

Shortle, D., Stites, W., & Meeker, A. (1990) *Biochemistry 29,* 8033-.

Siuzdak, G. (1994) *Proceedings of the National Academy of Sciences USA 91*, 11290-11297.

Srere, P., Brazil, H., & Gonen, L. (1963) *Acta Chemica Scandinavia 17*, S129-S134.

Talgoy, M. M., Bell, A. W., & Duckworth, H. W. (1979) *Canadian Journal of Biochemistry 57*, 822-833.

Tanford, C. (1962) *Journal of the American Chemical Society 84*, 4240-4247.

Tanford, C. (1968) *Advances in Protein Chemistry 23*, 121-282.

Textor, S., Wendisch, V., De Graff, A., Müller, U., Linder, M., Linder, D., & Buckel, W. (1997) *Archives of Mircobiology 168*, 428-436.

Tong, E. K., & Duckworth, H. W. (1975) *Biochemistry 14*, 235-241.

Traut, T. (1994) *Critical Reviews in Biochemistry and Molecular Biology 29*, 125-163.

Tsai, C.-J., & Nussinov, R. (1997) *Protein Science 6*, 24-42.

Verentchikov, A. N., Ens, W., & Standing, K. G. (1994) *Analytical Chemistry 66*, 126-133.

Weitzman, P., & Danson, M. (1976) *Current Topics in Cellular Regulation 10*, 161-204.

Weitzman, P. D. J. (1966) *Biochemical Journal 101*, 44C-45C.

Weitzman, P. D. J., & Jones, D. (1968) *Nature 219*, 270-272.

West, S., Kelly, S., & Price, N. (1990) *Biochemica et Biophysica Acta 1037*, 332-336.

Willard, H., L. L. , M., Dean, J., & Settle, F. (1988) *Instrumental methods of analysis*, Seventh ed., Wadsworth Publishing Company, Belmont.

Williamson, M. P. (1994) *Biochemical Journal 297*, 249-260.

Wood, D. O., Atkinson, W. H., Sikorski, R. S., & Winkler, H. H. (1983) *Journal of Bacteriology 155,* 412-416.

Yao, M., & Bolen, D. (1995) *Biochemistry 34,* 3771-3781.

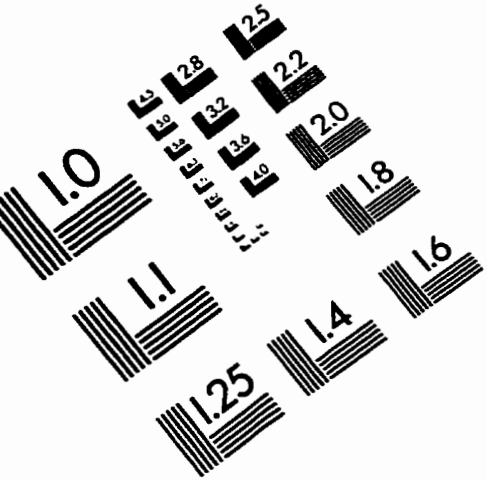
Yutani, K., Ogasahara, K., Aoki, K., Kakuno, T., & Sugino, Y. (1984) *The Journal of Biological Chemistry 259,* 10476-14081.

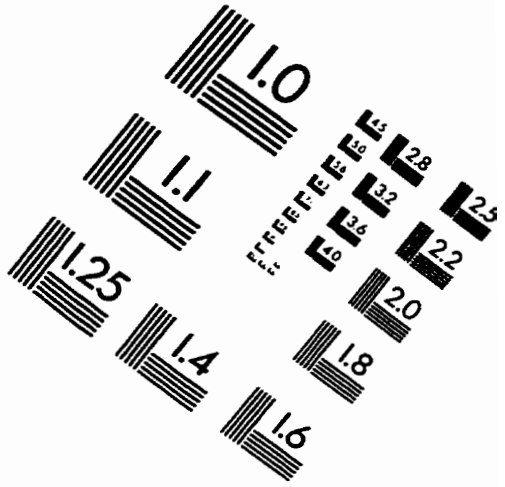
Zhi, W., Landry, S., Gierasch, L., & PA, S. (1992) *Protein Science 1,* 522-529.

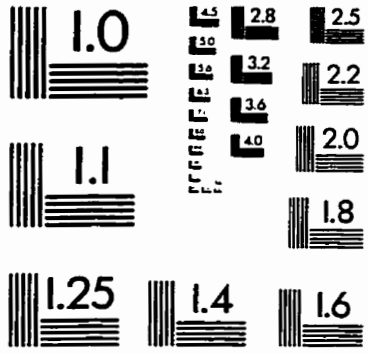
Zhi, W., Srere, P., & Evans, C. (1991) *Biochemistry 30,* 9281-9286.

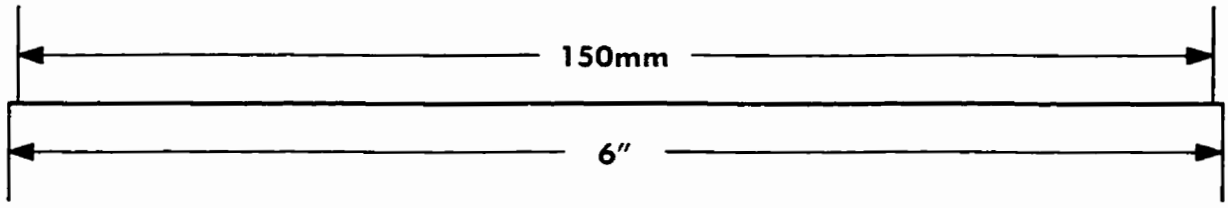
Zoller, M., & Smith, M. (1983) *Methods In Enzymology 100,* 468-500.

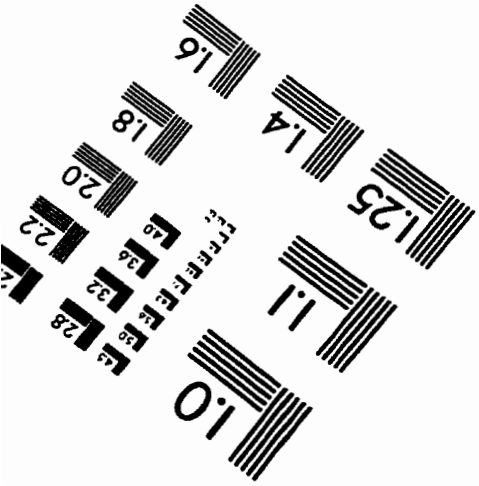
IMAGE EVALUATION
TEST TARGET (QA-3)

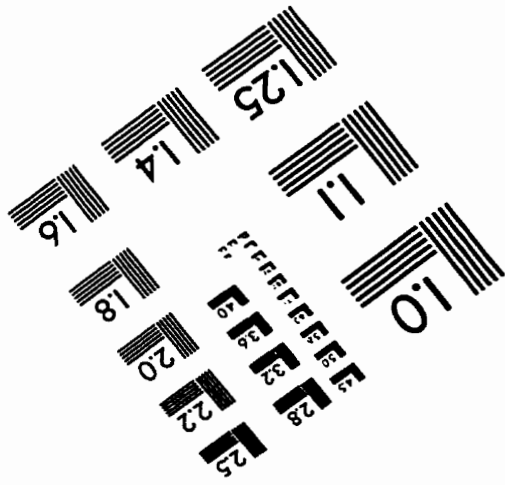












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989