TWO-STATE CONFORMATIONAL BEHAVIOR IN

PROTEIN ACTIVE CENTERS

by

JEREMY R. LOHMAN

A DISSERTATION

Presented to the Department of Chemistry
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2007

"Two-state Conformational Behavior in Protein Active Centers," a dissertation prepared by Jeremy R. Lohman in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Chemistry. This dissertation has been approved and accepted by:

_____
Dr. J. Andy Berglund, Chair of the Examining Committee

_____ 11/27/07 _____
Date

Committee in Charge: Dr. J. Andy Berglund, Chair
                            Dr. S. James Remington, Advisor
                            Dr. Brian W. Matthews
                            Dr. Tom H. Stevens
                            Dr. Darren W. Johnson

Accepted by:

_____
Dean of the Graduate School

An Abstract of the Dissertation of

Jeremy R. Lohman   for the degree of   Doctor of Philosophy

in the Department of Chemistry   to be taken   December 2007

Title: TWO-STATE CONFORMATIONAL BEHAVIOR IN PROTEIN ACTIVE

CENTERS

Approved: _____

Dr. S. James Remington, Advisor

Cellular processes are carried out by proteins, which often utilize conformational

changes for function. In theory, conformational changes can be harnessed to promote,

prevent or monitor cellular processes. Such changes in protein active centers require

perturbations through interactions with other proteins, small molecules or through energy

input into the system, for example light. The work presented incorporates rational design

and crystallographic elucidation of two-state conformational changes in two proteins,

green fluorescent protein (GFP) and malate synthase (MS).

GFP indicators were previously developed to quantitate the thiol/disulfide redox

status within cells. Cysteine residues were introduced in close proximity on the surface

of GFP and allow the formation of a disulfide bond. The indicators provide a fluorescent

readout of the ambient thiol/disulfide equilibrium, however thermodynamic studies

showed the resulting thiol/disulfide to be unusually stable (-287 mV) in comparison to

the cellular redox buffer glutathione (-240 mV). In order to produce a family of redox indicators suitable for use in less reducing environments, amino acids were inserted near the introduced cysteine pair in order to destabilize the disulfide. The resulting family of redox indicators, termed roGFP-iX, exhibit midpoint potentials in the more desirable range of -229 to -246 mV. Crystallographic analysis indicates that roGFP-iX indicators undergo much larger two-state conformational changes than the original indicators. Surprisingly, a cis-peptide was discovered between the cysteine and the inserted residue which in combination with the conformational changes helps to explain the reduced stability of the disulfide.

Malate synthase is an important virulence factor for certain microbes and carries out the Claisen condensation between glyoxylate and acetyl-CoA to produce malate. Crystal structures of *Mycobacterium tuberculosis* and *Escherichia coli* malate synthase isoform G had previously been determined with substrates or products bound. To determine the conformational changes necessary for substrate binding and product release, crystal structures of *Escherichia coli* malate synthase isoform A were determined in both the apo and acetyl-CoA/inhibitor bound forms. The crystallographic models revealed two-state conformational changes in the part of the active-site loop necessary for substrate binding, which has important implications for drug design.

This dissertation includes my unpublished co-authored materials.

CURRICULUM VITAE

NAME OF AUTHOR: Jeremy Ray Lohman

PLACE OF BIRTH: Kennewick, Washington

DATE OF BIRTH: October 6, 1981

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

> University of Oregon
> Washington State University
> Columbia Basin Community College

DEGREES AWARDED:

> Doctor of Philosophy in Chemistry, 2007
> > University of Oregon
>
> Bachelor of Science in Biochemistry with Biophysics option, 2003
> > Washington State University
>
> Associate of Arts and Science, 2000
> > Columbia Basin Community College

AREAS OF SPECIAL INTEREST:

> Structure and Function of Biomacromolecules
> Protein Engineering
> Novel Protein Chemistry

PROFESSIONAL EXPERIENCE:

> Research Assistant, Department of Chemistry and Institute of Molecular Biology
> > University of Oregon, 2004-2007
>
> Teaching Assistant, Department of Chemistry
> > University of Oregon, 2003-2004

Undergraduate Research Assistant, School of Molecular Biosciences
Washington State University, 2001-2003

Student Research Apprentice
Pacific Northwest National Laboratory, 2000

GRANTS, AWARDS AND HONORS:

American Heart Association Predoctoral Fellowship
"Development of a Green Fluorescent Based Redox Sensor for Use in Oxidizing Cellular Environments"
University of Oregon, 2006-2007

The Protein Society Finn Wold Travel Award
Washington State University, 2003

PUBLICATIONS:

Lohman, J. R., Remington, S. J. (2007) Development of a family of redox-sensitive green fluorescent protein indicators for use in relatively oxidizing subcellular environments. [Manuscript in Preparation]

Lohman, J. R., Olson, A., Remington, S. J. (2007) Atomic resolution structures of *Escherichia coli* and *Bacillis anthracis* malate synthase A: Comparison with isoform G and implications in structure based drug design. [Manuscript in Preparation]

Gloss, L. M., Topping, T. B., Binder, A. K., Lohman, J. R. (2007) Kinetic folding of *Haloferax volcanii* and *Escherichia coli* dihydrofolate reductases: Haloadaptation by unfolded state destabilization at high ionic strength. [Submitted to J. Mol. Biol.]

Wright, D. B., Banks, D. D., Lohman, J. R., Jaqueline, J. L., Gloss, L. M. (2002) Comparison of the effect of salt on the activity and stability of *Escherichia coli* and *Haloferax volcanii* dihydrofolate reductases, *J. Mol. Biol. 323*, 327-344.

Campbell, J. A., Hess, W. P., Lohman, J. R., Goheen, S. C. (2000) Analysis of Hanford-related organics using matrix-assisted laser desorption ionization time-of-flight mass spectrometry, *J. Radioanal. Nucl. Chem. 250*, 247-253.

ACKNOWLEDGMENTS

My doctoral studies were only possible through the help and support of so many. Foremost I would like to thank my advisor, Jim Remington, for his support and especially his advice, which helped me become a better scientist. The Remington lab members have been an integral part of my learning experience. I learned crystallization and other useful techniques from Karen Kallio our wonderful lab technician. My lab mates Nathan Henderson, Xiaokun Shu, Dave Anstrom and Mark Cannon all have been excellent sources of information and scientific discussion. I have had the pleasure of learning with and teaching a number of undergraduate students in the lab including, Leslie Colip, Andrew Olson and Josef Heenan. The Matthews lab members, of which there are too many to name, have been a tremendous source of knowledge for many aspects of my research. My thesis committee members, Andy Berglund, Brian Matthews, Tom Stevens, Darren Johnson, Bruce Branchaud and Bea Darimont all contributed useful ideas and helped kept me on track. The American Heart Association deserves mention for funding my research for two years. I would also like to thank teachers who inspired me to learn about science and to undertake the challenge of graduate school, especially Lisa Gloss.

Many people who did not contribute directly to my work are also in need of thanks. My family has been particularly supportive in my scientific aspirations, and very consoling in times of strife. At times, research can become monotonous, in my friends I constantly found the necessary escape to keep life enjoyable.

Dedicated to my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

OVERVIEW

I was initially interested in the proteins malate synthase (MS) and green fluorescent protein (GFP) for two very different reasons. The enzyme MS is important for the growth and survival of microbes, several of which are pathogenic; therefore the study of this enzyme has implications in drug design and human health. MS also carries out a difficult carbon-carbon bond forming reaction during catalysis, which is an interesting problem in enzyme chemistry. GFP is a wonderfully unique protein that allows molecular and cellular biologists to probe the inner workings of cells with a convenient fluorescent readout. Practically, my interest in GFP was in its utility as a biosensor, specifically in expanding its range as a sensor for cellular thiol/disulfide equilibrium. As X-ray crystallographic studies on these two projects unfolded, the appearance of important two-state conformational changes near the area of chemical activity (active center) in each protein became readily apparent. In this chapter, protein conformational changes are discussed with respect to enzymes and information transducing proteins (biosensors). Also a brief history of MS and GFP is presented.

PROTEIN CONFORMATIONAL CHANGES

When proteins interact with each other or with small molecules, or are perturbed by energy input into the system (e.g. light), the three dimensional structure is sometimes dramatically affected through conformational changes. These changes can involve large parts of the protein such as domains or be localized to particular active centers. Conformational changes are implied by the function of many proteins, even without direct evidence. For example, G protein-coupled receptors (GPCR) bind or "sense" molecules on the outside of the cell and must transmit that signal to the interior of the cell (1). Upon ligand binding, an induced conformational change is transmitted from the extracellular domain through the transmembrane domain and to the cytoplasmic domain producing a cellular response. Without this conformational change, the presence of the ligand would be only felt on the outside of the cell.

Proteins use conformational changes to achieve many different cellular functions, from catalysis to information transduction. An example of signal transduction from outside to inside the cell has already been presented with the GPCRs. Another set of signal transduction proteins are transcription factors such as OxyR, which undergoes a conformational change upon oxidation by H_2O_2 and reactive oxygen species (ROS) to accomplish gene regulation (2, 3). The cellular "energy currency" molecule, adenosine triphosphate (ATP), is produced when ATP synthase converts an electrochemical gradient into mechanical protein conformational changes (4). Many other proteins use ATP as the energy to drive the conformational changes necessary for function, such as for building actin filaments (5) or in the myosin molecular motors that utilize the

filaments to produce movement (*6*). An underlying theme is the interaction of proteins with a small molecule. The interactions with enzymes and signal-transducing proteins will be discussed in detail below.

Many techniques are available to study conformational changes in proteins, each with their strengths and limitations. The highest resolution structural information can be acquired using X-ray crystallography, which describes the endpoints of the conformational changes in atomic detail. X-ray crystallography reveals only the states before and after conformational changes, making the models are inherently two-state. NMR has the added benefit of being able to assign time constants to the changes, but in many cases, it lacks the structural accuracy of crystallography. When proteins are not amenable to crystallization or analysis by NMR, fluorescence resonance energy transfer (FRET) and cryo-electron microscopy can be useful but are limited to large conformational changes (such as domain movement).

ENZYME FUNCTION AND CONFORMATIONAL CHANGES

The observation that proteins can be crystallized introduced the idea that proteins are discrete molecules with well-defined structures. Previously, Emil Fischer in 1894 proposed the lock-and-key model for catalysis, where an enzyme and substrate interact with no change in conformation (*7*). In this model, the substrate and protein have complementary shapes that fit together precisely, which explains enzyme specificity. However, the model lacks in explaining how substrate binding would lead to transition state stabilization, which is necessary for catalysis. In reality proteins are inherently

flexible an idea better captured by Daniel Koshland's 1954 model of induced-fit (*8*).

Briefly, a protein and substrate initially bind weakly and after conformational changes

binding affinity is maximal. These conformational changes can strain a substrate to

promote catalysis, or place the side chains of residues in the proper orientation for

catalysis that may not have been favorable for initial substrate binding. Some enzymes

also use conformational changes to modulate function, as observed with allosteric

binding. In metabolism, many enzymes are inhibited by metabolites downstream of the

products they produce to prevent intermediates from building up. A specific example is

phosphofructokinase-1, whose products feed into the citric acid cycle, yet is inhibited by

citrate binding to a site away from the active site (*9*). In such cases, drugs can be

designed to inhibit enzyme function either through direct binding to the active site

(transition-state analogs) or at allosteric sites, which may not always be naturally

occurring. Indeed, the ultimate purpose of my work on MS has been to improve the

chances that effective inhibitors can be identified.

MALATE SYNTHASE

The ability of some microbes to grow solely on acetate interested some scientists

and lead them to the discovery of malate synthase, which was first purified in 1956 from

Escherichia coli grown in media containing acetate (*10*). The study of MS is important

as it and the enzyme isocitrate lyase form the basis of the glyoxylate cycle, a modified

citric acid cycle that allows microbes to grow on acetate and other two-carbon molecules

(*11*). MS catalyzes a magnesium-dependent carbon-carbon bond-forming reaction

between acetyl-CoA and glyoxylate yielding malate. Citrate synthase (CS) catalyzes a similar carbon-carbon bond forming reaction between acetyl-CoA and oxaloacetate, but is independent of magnesium (*12*). The complete lack of sequence similarity between CS and MS provoked Howard et al. (*13*) to solve the structure of *Escherichia coli* malate synthase isoform G from (ecMSG). These studies were performed to elucidate whether the reaction mechanisms were similar and determine how proteins with so little sequence similarity can carry out a similar reaction.

Previous structural studies showed CS to be almost completely α-helical (*12*). Structural studies by Howard et al. showed that CS and MS structures were indeed very different (*13*). The active site of MS is sandwiched between a mostly α-helical domain and a TIM barrel, which is made of alternating α-helices and β-sheets. This MS structure had glyoxylate bound and showed that CS and MS differ in the mode of activation of the electrophilic substrate, with two histidines in CS replaced by an arginine and magnesium. Further structural and mutagenic studies by Anstrom et al. on ecMSG confirmed that both MS and CS utilize an aspartate for the deprotonation of acetyl-CoA (*14*). While the exact mechanisms of MS and CS differ, the strategies involved remain similar.

Interestingly, CS structural studies showed that there was a large conformational change between the apo and substrate bound structures, whereby the binding of oxaloacetate induces the protein to form a binding site for acetyl-CoA (*12*). Previous small angle X-ray scattering (*15*) and circular dichroism (*16, 17*) experiments on MS indicated that the binding of glyoxylate, like the binding of oxaloacetate in CS, induces conformational changes. To explain the putative conformational change it was proposed

that the important domains of ecMSG would separate to allow the active site to sequester

reactants from solution (*13*). NMR studies of ecMSG later refuted this idea (*18*). Given

that crystals of ecMSG are difficult to grow, I searched for a MS more amenable to

crystallographic studies, in the hope of finding the conformational changes upon substrate

binding thus aiding drug design. The results of the search are reported in chapter III.

SIGNAL TRANSDUCTION AND CONFORMATIONAL CHANGES

The transfer of information in a cell is of utmost importance to maintaining

cellular homeostasis. For a protein to transmit a signal, such as the presence of a

molecule, the protein must first bind the molecule followed by a conformational change,

as without the change there would be very little a protein could do to achieve cellular

responses. Substrate binding often changes the structure in order for the protein to have

another function, such as binding to DNA or other proteins and in the process

information is transferred. Transcription factors such as nuclear hormone receptors are

cytoplasmic and upon hormone binding are translocated into the nucleus where they

change patterns of gene regulation (*19*). Specifically, the hormone receptors are partially

unfolded and require chaperones for solubility. When the hormone binds, the nuclear

receptor undergoes a conformational change from the unfolded to a folded state.

Likewise, calmodulin undergoes a conformational change when it binds to calcium (*20*),

at which point it can bind to many different proteins and regulate their function. In

combination with fluorescent proteins, conformational changes in many proteins have

been exploited to develop sensors.

FLUORESCENT PROTEIN BASED BIOSENSORS

The discovery and cloning of the green fluorescent protein from *Aequorea victoria* has revolutionized cellular biology. GFP is a protein that contains a chromophore that is produced auto-catalytically from a -S-Y-G- sequence in the center of the folded protein (*21*). Since chromophore production occurs without the need of external cofactors, GFP can be fused with other proteins to monitor the cellular localization of the fusion protein in real time. The power of GFP as a cellular marker led to the development and discovery of a whole rainbow of fluorescent proteins (FP).

With multiple FPs available, the development of completely genetically encoded FRET biosensors became possible. Such biosensors employ a donor, usually a blue or cyan FP that is excited at UV-blue wavelengths, and emits near the peak of an acceptor, usually a green or yellow FP. When the FPs are in close proximity, excitation of the donor leads to fluorescence from the acceptor. By contrast, when they are far apart, emission comes from the donor, and the ratio between the emission intensities gives rise to the FRET signal. Conformational changes in various proteins fused between the FPs alter the distance and/or orientation between the donor and acceptor, thus changing the FRET signal. One of the greatest powers of FRET systems is that the signal is inherently ratiometric. Because the emission peaks interconvert, changes in cell thickness and other factors cancel out. One of the first FRET biosensors demonstrated was through a calmodulin fusion, which was developed almost simultaneously by Romoser et al. (*22*) and Miyawaki et al. (*23*) in 1997. The number of these sensors has grown quickly since their first implementation. One group reports in a symposium abstract the production of

"redoxifluor", using of the redox sensitive domain of a yeast transcription factor, Yap1p,

to produce a FRET sensor, but as of now no further reports have been made (24).

Another group produced a redox-sensitive FRET sensor using the redox regulated

domain of HSP-33 and appropriately coined it HSP-FRET (25). Studies of HSP-FRET in

cardiomyocytes showed that after oxidation, reduction of the probes was a very slow

process, which limits the utility of the system. Fortunately, FRET based systems are not

the only way to make a fluorescent biosensor.

The excitation spectrum of GFP displays two peaks, which correspond to the

neutral and anionic forms of the chromophore (21). This makes GFP an inherently two-

state protein. Our job is to manipulate the inherent chromophore equilibrium of GFP to

produce a biosensor. In many cases biosensors have been produced using GFP which

preserve both excitation peaks, and in some cases perturbation results in the

interconversion of the peaks which makes these sensors ratiometric. Some groups have

utilized circularly permuted variants of GFP to produce sensors, for example inserting

calmodulin into the barrel of GFP next to the phenolic end of the chromophore produced

a sensor whose excitation spectra was dependent upon calcium binding (26). Similarly,

OxyR and its conformational change were used to make a H_2O_2 sensor (27). All of the

previously described sensors utilize the conformational changes in other proteins linked

to FPs to change the signal. Modification of GFP through mutagenesis can produce

simpler sensors.

Some of the first attempts to change the excitation and emission spectra of GFP

led to variants whose chromophore protonation state was pH dependent. Some of these

mutants, discovered by Hanson et al., exhibited ratiometric dual color emission properties and were named deGFP for short (*28*). Crystal structure analysis of deGFP revealed that a strand near the chromophore undergoes a large conformational change between high and low pH. Redox sensitive GFP (roGFP) developed by Hanson et al. has a very small conformational change before and after its reaction with small redox molecules, which is partly responsible for its thiol/disulfide equilibrium properties (*29*). The second chapter reports the development of a roGFP with different thiol/disulfide equilibrium properties due to much larger engineered conformational changes.

Chapter II and III include previously unpublished co-authored materials.

CHAPTER II

DEVELOPMENT OF A FAMILY OF REDOX-SENSITIVE GREEN FLUORESCENT PROTEIN INDICATORS FOR USE IN RELATIVELY OXIDIZING SUBCELLULAR ENVIRONMENTS

Experimental data presented in this chapter are my own with some data collection done by Josef Heenan. I prepared the original manuscript. Jim Remington provided direction, support and extensively edited the manuscript.

SUMMARY

Green fluorescent protein (GFP) indicators have previously been developed that rapidly and quantitatively respond to changes in the thiol/disulfide equilibrium within subcellular compartments. In these indicators, surface exposed cysteines residues were introduced so as to form a redox-active disulfide that in turn controls the emission properties of the internal chromophore. The biosensors have been shown to be effective reporters of the thiol/disulfide status within reducing compartments such as the mitochondria and cytosol for several cell types. However, due to the unusual thermodynamic stability of the introduced disulfide bond, the indicators are not useful for quantitative analysis within more oxidizing compartments such as the endoplasmic

reticulum. Here we report the development of a new family of GFP-based redox indicators in which the thermodynamic stability of the disulfide has been substantially lowered by insertion of a single amino acid into the main chain, adjacent to cysteine 147. The insertions result in indicators with midpoint potentials of -229 to -246 mV, and are thus better suited for study of relatively oxidizing subcellular compartments. High resolution crystallographic models of two variants highlight how the structure of the disulfide is perturbed by the presence of the insertion and give new insight into the factors which influence the thermodynamic stability of a reactive disulfide.

INTRODUCTION

Previously developed redox-sensitive green fluorescent proteins are effective indicators of cellular thiol/disulfide equilibrium. The indicators were produced by substitution of cysteines residues on the surface of GFP where they form a disulfide bond without substantially distorting the structure of the protein. On a background of wild-type or S65T GFP, Hanson et al. substituted cysteines residues at positions Ser147 and Gln204 (to create roGFP1 and roGFP2, respectively) or Asn146 and Ser202 (to create roGFP3 and roGFP4, respectively) (*29*). Using the same approach but on the background of the yellow fluorescent GFP variant YFP (λ^{em}_{max} = 525 nm), Ostergaard et al. produced rxYFP^N_C, where N and C indicate the locations of the introduced cysteines (*30*).

In vitro and in vivo, these two families of indicators show two-state changes in fluorescence excitation in response to the thiol/disulfide ratio. As determined by simple titrations against thiol/disulfide redox buffers such as DDT or lipoic acid, the midpoint

potentials of roGFP1, roGFP2 and rxYFP$^{149}_{202}$ (which contain the same reactive

disulfide) have been reported to be -287, -272 (29) and -261 mV (30), respectively.

Compared to the midpoint potential of the GSSG-GSH couple (-240 mV (31)), these

indicators are quite reducing and are thus best suited for use in reducing subcellular

environments. roGFP indicators are ratiometric by excitation, that is, they exhibit two

excitation peaks (~390 and 475 nm), corresponding to the neutral and an anionic

chromophore forms, respectively. Excitation of either peak gives rise to green

fluorescence at about 510 nm. Upon oxidation of the redox-sensitive disulfide, the

population of the neutral chromophore is favored at the expense of the anionic

chromophore, demonstrating two-state behavior. In contrast, rxYFPs have a single

fluorescence excitation peak (at ~512 nm) which varies in amplitude with oxidation and

thus these indicators are not ratiometric. However, most other properties are very similar.

roGFPs and rxYPPs have been expressed in a variety of cell types. roGFP

indicators were used to determine the thiol/disulfide equilibrium within the mitochondria

($E_{pH=8} \sim$ -360 mV) (29) and cytoplasm ($E_{pH=7} \sim$ -320 mV) (32) of single HeLa cells, as

well in Arabidopsis (mitochondria $E_{pH=7.8}$ -362±10 mV and cytoplasm $E_{pH=7}$ -318±13 mV

(33)). The cytosolic thiol/disulfide equilibrium of *Saccharomyces cerevisiae* and

Escherichia coli has been determined using rxYFP$^{149}_{202}$ and shown to be -289 mV and -

259 mV respectively (30, 34). However, the disulfide bonds of all of these probes are too

stable to allow the measurement of the redox potential in more oxidizing environments

such as the endoplasmic reticulum (estimated $E^{\circ\prime}$ -180±6 mV (35)), where all of the

probes are completely oxidized (34, 36, 37). Austin et al. (37) used this feature of

roGFP1 to determine that, contrary to the prevailing misconception, the endocytotic

pathway in human prostate cancer PC3 cells is oxidizing, however a quantitative result

was not reported.

Thus, there is a clear need to extend the range of midpoint potentials over which

redox-sensitive GFPs can be used as quantitative reporters. One attractive approach to

extend this range is to reengineer existing GFP-based indicators so as to lower the

thermodynamic stability of the reactive disulfide. In nature, midpoint potentials of

protein disulfides are found to cover an extremely large range that spans the relevant

physiological states. Considering only those proteins that are catalytically active in

establishing thiol/disulfide equilibrium processes, the midpoint potentials of the

thiol/disulfide oxidoreductases range from -270 mV for *E. coli* thioredoxin (*38*) to -122

mV DsbA (*39*). This range corresponds to a difference of approximately three orders of

magnitude in the equilibrium constants for disulfide formation.

From studies of the thiol/disulfide oxidoreductases (*40*), peptides and small

dithiol-containing molecules (*41, 42*) it is clear that several factors affect the

thermodynamic stability of disulfides, including thiol pK_a, geometric strain and entropic

effects. Approaches to modify disulfide stability could in principle make effective use of

any one or combination of these factors. One approach to reduce disulfide stability is to

decrease the pK_a of one or both of the participating thiols, which stabilizes the reduced

state by making the thiol a better leaving group (*43*). In practice this has been achieved

in proteins by substitution of one or more positively charged residues close to the reactive

disulfide, which in principle would stabilize a negative charge on the thiolate. Such an

approach has been studied in detail for roGFP1 (*44*) and rxYFP$^{147}_{202}$ (*45*). In roGFP1,

substitution of three arginine or lysine residues near the disulfide increased the midpoint

potential by 16mV. In roGFP1, the substitutions had the added effect of increasing the

rates of oxidation and reduction, for a 4.9 fold increase in oxidation rate constant by H_2O_2

and 6.0 fold increase in the DTT reduction rate constant. Similarly, placement of three

arginine residues near the disulfide of rxYFP$^{147}_{202}$ led to a 4.3 fold increase in the rate of

oxidation by β-mercaptoethanol and 13 fold increase in the rate of oxidation by GSSG

(the changes in indicator midpoint potentials were not reported). The latter results

suggest that at least part of the observed rate increase can be attributed to electrostatic

effects on substrate diffusion.

Two other approaches to reduce disulfide stability are to increase geometric strain

in the oxidized state and to change entropic factors associated with the reaction. These

may be linked. For example, in small molecules that undergo ring closure upon disulfide

formation, decreasing the ring size below six atoms increases the midpoint potential

because the disulfide cannot achieve the lowest energy geometry, whereas increasing the

ring size above six increases the entropic barrier to disulfide formation. The latter effect

may also be thought of as lowering the effective concentration of the reacting species (*46,

47*). There is a similar relationship for peptides containing cysteine residues. Cysteines

that are immediately adjacent to each other or separated by one intervening residue do not

readily form disulfides, which is thought to be due to geometric strain, but beyond a

certain size, increasing the length of the peptide decreases disulfide stability due to

entropic effects (*43, 48-50*).

In proteins, the addition or deletion of residues flanking the disulfide may additionally increase or decrease disulfide stability, depending upon how the protein rearranges to accommodate the insertion or deletion. In one example, the deletion of the proline from the active site peptide –CPGC– to produce –CGC– in thioredoxin destabilized the disulfide bond (towards a more oxidizing potential) to ≥-200 mV versus -270 mV for wild-type as determined by the non-equilibrium with $NADH/NAD^+$ via thioredoxin reductase (*51*). The inherent instability of the CGC peptide (-167 mV) is suggested to be the reason behind the large change in midpoint potential of the –CGC– thioredoxin (*51*).

Here we describe efforts to use single amino acid insertions to destabilize the redox-sensitive disulfide bridge of a redox-sensitive GFP, in order to produce a roGFP suitable for relatively oxidizing subcellular environments. An additional requirement was that the roGFP exhibit ratiometric response to changes in the ambient thiol/disulfide equilibrium. Equilibrium and kinetic measurements indicate that for one class of insertion mutants (roGFP1-iX), the disulfide is substantially destabilized, showing higher rates of reduction and lower rates of oxidation compared to the parent molecule roGFP1. Crystal structure analysis of two variants in the oxidized and reduced states verified disulfide formation in the oxidized state and revealed large conformational changes associated with oxidation/reduction. The factors contributing to disulfide stability are discussed in the light of the atomic models.

METHODS AND MATERIALS

Experimental approach and mutational analysis

Initially, we chose to insert a single glycine residue on either side of the cysteine residues (147 and 204) participating in the redox-sensitive disulfide of roGFP1 and roGFP2. Glycine has the fewest restrictions on backbone conformation and thus in principle the insertion could be accommodated by the protein in a variety of ways. Glycine-containing mutants, four for each roGFP1 and roGFP2, were produced and purified; those shown to have spectra sensitive to addition of reductants were characterized further by titration against DTT and lipoic acid. Of those that were characterized, none exhibited a response that was ratiometric by excitation; having only a general increase in fluorescence upon reduction, similar to rxYFP. However one variant, roGFP2 with glycine insertion C-terminal to Cys147 (roGFP2-147CG), was found by titration to have a thermodynamically destabilized disulfide. This variant was then subjected to random and directed mutagenesis with two objectives in mind. The first objective was to produce a variant with ratiometric response to changes in the ambient thiol/disulfide equilibrium, while the second objective was to investigate the consequences of side chain size and charge at the insertion position on the thermodynamic stability of the disulfide linkage.

As it is well known that wild-type GFP does not fold efficiently at 37° C (*21*), before random mutagenesis was initiated, the so-called "folding mutations" [F64L, F99S, M153T, V163A and I167T, see references (*21, 52, 53*)] were incorporated into roGFP2-147CG and roGFP1-147CG to increase the likelihood of detection of random mutants

which retain fluorescence. Residues near the putative disulfide and known to influence

the protonation state of the chromophore (Gly147a, His148, Thr203 and Ser205) were

chosen for randomization in those backgrounds. Bacterial colonies expressing visibly

fluorescent protein were picked, sequenced, expressed in small quantities overnight and

finally subjected to SDS-PAGE analysis to assess aggregation under oxidizing

conditions. Those variants which formed dimers or higher aggregation states were

rejected. Variants which survived this selection were then further characterized by

fluorescence spectroscopy.

Gene construction, mutagenesis and protein expression/purification

Fluorescent proteins in plasmid pQE-30 are expressed constitutively in *E. coli*

DH5α cells, for that reason roGFP1 (GFP with C48S/S147C/Q204C) in pRSET$_B$ was

PCR amplified using primers containing BamHI and HindIII cut sites and cloned into

pQE-30 containing an N-terminal His$_6$ tag. To produce the initial insertions and all other

non-random mutations the QuikChange (Stratagene) protocol was utilized.

Random mutants were produced using the Stratagene Multi-Site Directed

Mutagenesis kit and transformed into DH5α cells. Two primers were utilized one

randomizing positions Gly147a and His148 simultaneously and the other primer

randomizing positions Thr203 and Ser205. Colonies expressing random mutants were

screened by placing the plates on a UV-transilluminator and selecting colonies which

were visibly fluorescent. These colonies were used to inoculate 2 mL of growth media,

after growth overnight the bacteria were harvested and plasmids purified. DNA sequencing of the GFP coding region was used to identify all mutants.

roGFP1 and the initial insertion mutants were prepared from pRSET$_B$ plasmids as described previously (29, 44) and require expression at 18-22°C. Random insertion mutants in plasmid pQE-30 were expressed in E. coli strain DH5α by allowing cultures to reach an OD$_{600}$ of between 0.4 and 0.9 growing at 37°C, at which point the temperature was reduced to 25°C overnight, no IPTG was needed for efficient protein production. Cells were pelleted by centrifugation and resuspended in 50mM HEPES (pH 7.9), 300mM NaCl and 10% glycerol. The cells were frozen, thawed and then sonicated for 5 min. Cell lysate was centrifuged and the supernatant was applied to a column of Ni^{2+}-agarose resin (Qiagen) for purification. Samples were dialyzed into 20mM Tris (pH 8.0), 1mM EDTA and further purified using anion exchange. Pure fractions were concentrated by filtration and buffer exchanged into 20mM HEPES (pH 7.9).

Midpoint potential determinations

To minimize O$_2$-mediated oxidation all solutions and samples were prepared/incubated in an anaerobic glove-box. Protein samples were prepared at ~1 µM in 100 mM HEPES (pH 7.0), 300 mM NaCl, 1 mM EDTA with 2.5 mM total lipoic acid (mixture of oxidized and reduced forms) and allowed to equilibrate for 3-4 hours. After equilibration, samples were removed and fluorescent spectra were determined at 25°C. Apparent redox midpoint potentials for the mutants were determined as described by Hanson et al (29). However, because oxidized lipoic acid absorbs near the 395nm

excitation maximum of GFP, the intensity ratio of 505 nm fluorescence for excitation at

the isobestic point (~425 nm) and 465 nm was used in the analysis. The midpoint

potential ($E^{\circ'}$) of lipoic acid was assumed to be -0.290 V (*42*). All titrations were

performed at least twice.

Oxidation/reduction kinetics

Reaction rates for reduction of redox-sensitive GFPs by DTT and oxidation by

oxidized cysteine (cysteine or CSSC) were determined by addition of a large excess of

reagent (1 mM) and monitoring fluorescence over time. Solutions were degassed prior to

use and experiments were conducted at 25°C, with ~1 µM protein in 100 mM HEPES

(pH 7.0), 300 mM NaCl and 1 mM EDTA. Prior to oxidation by CSSC, concentrated

protein samples (~50 µM) were incubated with ~1 mM DTT for one hour to reduce the

protein. Removal of the remaining small amount DTT using a gel filtration spin column

was found to have no significant effect on the rate of oxidation. The pseudo first-order

rate constant for the reactions were determined by fitting the reaction progress curves to a

single-exponential function using KaleidaGraph (Synergy Software).

Spectroscopy

Fluorescence excitation/emission scans and kinetic traces were determined using

a LS55 fluorescence spectrophotometer (PerkinElmer Life-sciences) or a FluoroMax-3

fluorescence spectrophotometer (Jobin Yvon Inc.). Fluorescence quantum yields were

determined by preparing protein samples so that absorbance at 400 nm equaled that of the

standard 9-aminoacridine ($\lambda_{\text{max, abs}}$ = 400 nm). Emission spectra were corrected for photomultiplier wavelength sensitivity and integrated as described previously (*44, 54*). Wild-type GFP and roGFP1 were utilized as controls.

Crystal structure determination

roGFP1-iR mutant (oxidized) was concentrated to ~25 mg/ml in 20mM HEPES (pH 7.9). Crystals grew in 1-2 days by hanging drop vapor diffusion against 0.1 M Tris (pH 8.5), 22-24% PEG 1550 and 0-0.06 M MgCl$_2$. Drops consisted of 1 μl protein and 1μl well solution. Crystals of oxidized roGFP1-iR were transferred to a sitting drop (initially 10 μl of mother liquor) and artificial mother liquor containing 0.1 M Tris (pH 8.0), 30% PEG 1550, 0.05 M MgCl$_2$ and 20mM TCEP was added gradually over the course of two days (final drop volume was 50 μl) to obtain crystals with reduced cysteines.

roGFP1-iE (oxidized) was concentrated to ~25 mg/ml in 20mM HEPES (pH 7.9). Crystals grew in ~7 days by hanging drop vapor diffusion against 0.1 M Tris (pH 8.0), 24% PEG 1550 and 0.04 M sodium acetate. Again, TCEP (20mM) was used to reduce crystals except the final condition was 0.1 M Tris (pH 8.0), 30% PEG 1550, 0.05 M sodium acetate.

X-ray diffraction data were collected from single frozen crystals for both mutants in both the oxidized and the reduced state at ALS beamline 5.0.1. Data sets were indexed and reduced using the HKL2000 suite (*55*). Molecular replacement solutions were found with Molrep using the GFP S65T coordinate file (PDB code 1EMA) as the search model.

ARP/wARP was used to rebuild the model (*56*). Refmac 5 (*57*) was used for further refinement after model building in Coot (*58*). ARP/wARP occasionally introduces water molecules into inappropriate density; therefore, waters were only kept or introduced if they were associated with reasonable electron density peaks and had appropriate proximity to hydrogen bonding partners.

RESULTS

Expression of roGFP1 and roGFP2 mutants that involved insertion of single glycine residues adjacent to and on either side of Cys204 led to soluble protein of the appropriate molecular weight, yet the protein was non-fluorescent; as a result this material was not characterized further. However, single glycine insertions adjacent to Cys147 were more successful and resulted in expression of fluorescent protein. Partial purification and characterization of a protein with an insertion N-terminal to Cys147 (roGFP2-146NG) suggested that this molecule was responsive to changes in the ambient thiol disulfide equilibrium in vitro. However, the estimated midpoint potential $\sim$ -285 mV was close to that of the original roGFP2 and hence this variant was not further characterized. More promising behavior resulted from insertion of glycine C-terminal to Cys147 (roGFP2-147CG). Preliminary tests suggested that a disulfide with significantly decreased stability $\sim$ -255 mV formed under oxidizing conditions. roGFP2-147CG has a single excitation peak at 490 nm and emission peak at 508 nm, the amplitude of which increases upon titration with reduced lipoic acid.

Random mutagenesis was then used to generate a library of variants with different excitation behaviors. Individual random mutants were expressed, purified and checked for aggregation behavior. Monomeric mutants were characterized further by adding DTT to the presumably O_2 oxidized protein to determine the effect on the fluorescence excitation spectra. The majority of the mutants tested lacked the desired ratiometric behavior, the most common response being a general decrease in fluorescence intensity upon reduction. However, in a few cases ratiometric response was identified. As a result of these procedures, a promising family of variants was obtained (roGFP1-147CX/H148S, also containing folding mutations, see Table 2-1), several members of which were further characterized. For the following discussion, we have adopted the shorthand notation roGFP1-iX, where X is the amino acid at insertion position 147a, to identify members of this family.

Spectroscopic properties of roGFP1-iX

The characterized members of the roGFP1-iX family have similar excitation and emission spectra, with excitation peaking at ~395nm (band A) and ~465nm (band B) and fluorescence emission peaking at ~505 nm. The relative amplitudes of the excitation peaks depend strongly on the presumed oxidation state of the 147-202 disulfide. In the oxidized state, excitation band A is favored over the band B; however, upon reduction, band B becomes dominant (Figure 2-1). The behavior is two-state with an isobestic point at ~ 425 nm. Fluorescence quantum yields for excitation at 400nm are somewhat lower than wild-type GFP or roGFP1 and range from 0.24 (roGFP1-iD) to 0.33 (roGFP1-iH),

Table 2-1. roGFP-iX redox potential, kinetic and spectroscopic properties

Xins[a]	$E^{o\prime}$ (mV)	reduction rate[b]	oxidation rate[c]	Φ^d	δ^e	$\lambda_{em\ 400nm\ ex}$	$\lambda_{iso\ 505nm\ em}$
L	-229±5	0.77	0.22	0.30	7.2	505	425
E	-236±7	0.33	0.14	0.31	4.5	503	425
Q	-239±6	0.33	0.21	0.31	4.1	505	425
H	-238±4	0.47	0.22	0.33	3.1	505	425
R	-237±5	0.27	0.22	0.24	2.9	502	425
S	-240±3	0.48	0.24	0.31	2.4	505	425
D	-246±1	0.23	0.18	0.24	2.8	502	425
roGFP1	-287	0.18	0.33	0.64	6.5	508	425

a - mutations roGFP1 (C48S/S147C/Q204C) + X insertion/H148S/F64L/F99S/M153T/V163A/I167T

b - reduction by 1mM DTT units of min^{-1} standard deviation less than ±0.05 (at least two trials)

c - oxidation by 1mM cystine units of min^{-1} standard deviation less than ±0.05 (at least two trials)

d - fluorescence quantum yield from excitation at 400 nm

e - dynamic range is the maximum observed δ-fold change in excitation peak ratio

Table 2-1 summarizes the spectral and biochemical properties of the variants. Dynamic

range (defined as the maximum observed ratio of excitation peak ratios) measurements

indicate that the insertion mutants have fluorescence changes on the order of roGFP1

(6.5) with dynamic ranges spanning from 2.4 (roGFP1-iS) to 7.2 (roGFP1-iL).
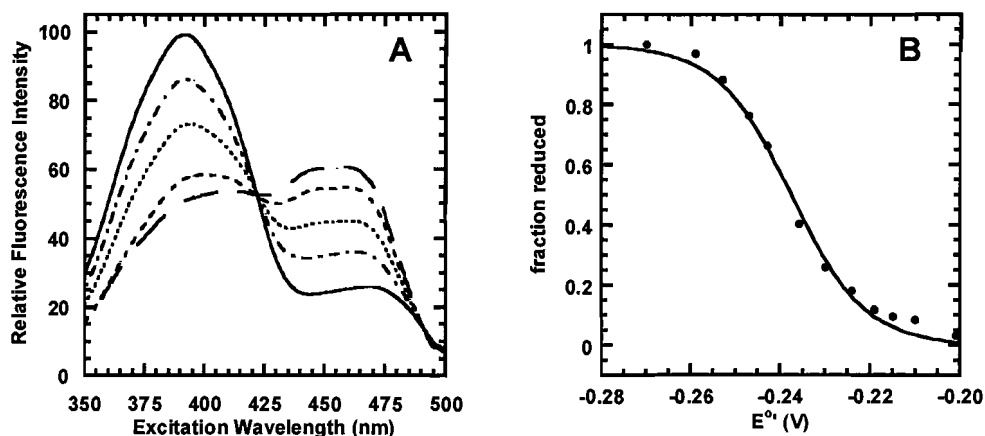
Figure 2-1. Redox titration of roGFP1-iE. (A) Fluorescence excitation spectra while
monitoring emission intensity at 505 nm at the following redox potential values: -181
(solid line), -224 (dashed and dotted line), -236 (dotted line), -247 (short dash line) and -
390 mV (long dashed line). (B) Fluorescence emission ratios were normalized to the
highest and lowest intensities for each scan to produce a fraction reduced. The solid line
represents the best fit to a theoretical two-state titration with a midpoint of -238mV.

While wild-type GFP is insensitive to solution pH over the range of about 4-10

(21, 59, 60); most mutations close in space to and including His148 confer pH sensitivity

and roGFP1-iX mutants are no exception. The excitation spectra of oxidized roGFP1-iE

is quenched at low pH, and at pH 6.0 the fluorescence intensity is approximately half that

at pH 7.0 (Figure 2). At high pH, beginning around pH 8.0, the spectra shifts from

favoring the 395 nm peak to favoring the 465 peak. Similarly, in the reduced state at pH

5.5, the fluorescence intensity is half that as at pH 7.0 and above pH 9.5 the 465 nm peak

becomes favored at the expense of the 395 nm peak. The 465/395 ratio is essentially

constant for both oxidized and reduced spectra between pH 6.0 and 8.0 indicating that

this would be a useful range where pH artifacts are negligible (Figure 2-2). The pH

dependence of roGFP1-iL was found to vary similarly to that of roGFP1-iE.

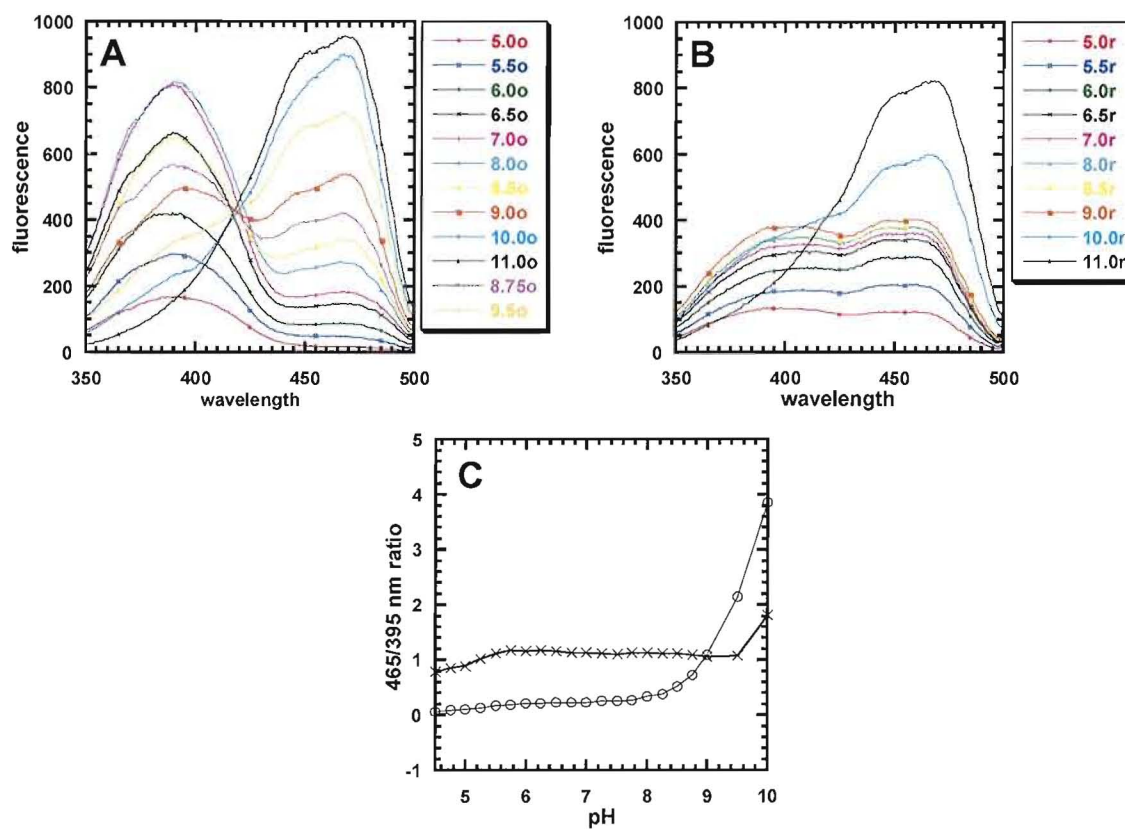**Figure 2-2.** pH titrations of roGFP1-iE. (A) Oxidized spectra at various pHs. (B) Reduced spectra at various pHs. (C) Fluorescence excitation ratio versus pH. Circles are oxidized data and Xs are reduced data. Lines are drawn to guide the eye.

Redox equilibrium and reaction kinetics

Redox midpoint potentials were determined from the equilibrium constant of the reaction between roGFP1-iX and lipoic acid ($E^{\circ\prime}$ assumed to be -290 mV, (*42*)), see Figure 1 for representative data. Redox midpoint potentials ranging from -229 mV (roGFP1-iL) to -246mV (roGFP1-iD) are summarized in Table 2-1.

Pseudo-first order rate constants for the reduction of the variants were determined by monitoring fluorescence over time after the addition of a large excess of DTT (1mM) at pH 7.0, Figure 2-3. Rate constants for the insertion mutants ranged from k_{DTT} 0.23 min^{-1} (roGFP1-iD) to k_{DTT} 0.77 min^{-1} (roGFP1-iL). All rate measurements had an error less than ±0.05 min^{-1} determined by at least two experiments, see Table 2-1. The oxidation rates were determined under similar conditions using 1mM cysteine (CSSC). Rate constants for the oxidation reactions varied from k_{CSSC} 0.14 min^{-1} (roGFP1-iE) to k_{CSSC} 0.24 min^{-1}(roGFP1-iS). Reversibility of the titration was verified by comparisons of spectra obtained before and after oxidation/reduction kinetic trials, normalized to the isobestic point to account for small changes in indicator concentration. By comparison, roGFP1 was found to be reduced more slowly (k_{DTT} 0.18 min^{-1}, close to the previously reported value of 0.13 min^{-1} (*44*)) than any of the roGFP1-iX. roGFP1 also oxidized more rapidly (k_{CSSC} 0.33 min^{-1}) than any roGFP1-iX.
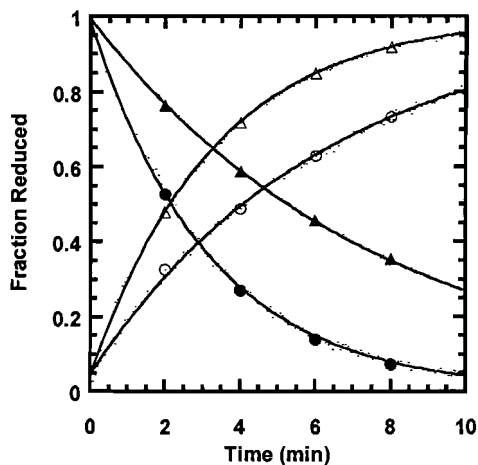
Figure 2-3. Oxidation and reduction kinetics of roGFP1-iE (triangles) and roGFP1 (circles). The fraction reduced over time is shown after addition of DTT (open symbols) or oxidized cysteine (solid symbols). Pseudo first-order rate constants were determined by fitting the rate curves (solid lines) to a single exponential.

Crystal structure analysis of roGFP1-iR and roGFP1-iE

Both variants crystallized in a $P2_12_12_1$ lattice, which is approximately

isomorphous to GFP-S65T (PDB ID code 1EMA), and contains one molecule per

asymmetric unit. Data collection and refinement statistics can be found in Table 2-2 for

all atomic models described below. As expected, each model is very similar to that of

wild-type or S65T GFP and consist of an 11 stranded beta barrel with an internal helix

from which the chromophore forms. However, as detailed below, there are significant

differences localized to the site of the amino acid insertion (position 147a) that are

strongly influenced by the state of oxidation of the 147-204 disulfide.

Table 2-2. roGFP1-iX data collection and refinement statistics

Crystal	iE oxidized	iE reduced	iR oxidized	iR reduced
Data collection				
Total observations	180,776	125,308	190,447	136,444
Unique reflections	52,564	35,231	54,553	41,093
Cell dimensions (a, b, c) (Å)	51.2, 62.4, 68.8	51.2, 62.3, 69.6	51.1, 62.7, 70.1	51.0, 62.4, 68.8
Resolution (Å)	50.0-1.31	50.0-1.50	50.0-1.31	50.0-1.40
Highest resolution shell (Å)	1.36-1.31	1.55-1.50	1.36-1.31	1.45-1.40
Completeness[a] (%)	98.6 (92.4)	97.2 (95.1)	99.5 (99.8)	93.3 (93.6)
Average I/s [a]	20.4 (1.9)	28.1 (2.1)	29.8 (9.8)	20.0 (1.6)
R_{merge} [a,b]	0.051 (0.429)	0.040 (0.472)	0.039 (0.098)	0.044 (0.449)
Refinement				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
No. of molecules[c]	1	1	1	1
No. of protein atoms[c]	1853	1790	1843	1821
No. of solvent atoms[c]	268	242	303	228
Resolution range (Å)	50.0-1.31	50.0-1.50	50.0-1.31	50.0-1.40
Crystallographic R-factor[d]	0.164	0.188	0.152	0.185
R-free	0.186	0.210	0.170	0.208
Average B-factors ($Å^2$)				
Protein atoms	14.0	17.9	9.8	14.6
Solvent	25.9	29.1	20.2	24
Root mean square deviations from ideality				
Bond lengths (Å)	0.007	0.010	0.007	0.010
Bond angles (degrees)	1.451	1.450	1.363	1.437

[a] Values in parentheses indicate statistics for the highest resolution shell.

[b] $R_{merge} = \Sigma |I - <I>| / \Sigma <I>$, where I is the observed intensity, and $<I>$ is the average of intensities obtained from multiple observations of symmetry-related reflections.

[c] Per asymmetric unit.

[d] R-factor $= \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$, where F_o and F_c are observed and calculated structure amplitudes, respectively.

Oxidized roGFP1-iR. The oxidized form of roGFP1-iR diffracted to better than 1.31 Å resolution at the ALS beamline 5.0.1, with the data resolution being limited by detector geometry. The final model consists of residues 3-231 with none in the disallowed regions of the Ramachandran plot. The final model has an R-factor of 0.152 and an R-free of 0.170 at 1.31 A.

Residues Asn146, Cys147, Arg147a, Ser148 and Cys204 are all clearly defined by electron density maps. For representative electron density, see Figure 2-4. Residues Cys147 and Cys204 form a disulfide that spans two adjacent segments of beta sheet and is in the **pg⁻t** (positive χ_{ss} torsion angle with a "gauche⁻" χ^{1n} value of ~ −60° and "trans" χ^{1c} value of ~±180° (61)) conformation (see Table 3). For comparison, the torsion angles describing the disulfides found in other redox sensitive GFPs are included in Table 2-3. As described in the study of disulfide conformations by Srinivasan et al. (61), there are relatively few disulfides in this conformation (2 of the 72 studied), the closest are immunoglobulin Fab (PDB ID code 1FB4, H chain residues Cys101 and Cys104) and human lysoszyme (PDB ID code 1LZ1, residues Cys77 and Cys95). Remarkably, the peptide bond between Cys147 and Arg147a is in the rare cis conformation (62). The side chain of Arg147a spans a gap between the strands containing the disulfide and forms hydrogen bonds with Ser202 and the backbone carbonyl of Phe223. The hydroxyl of Ser148 forms a hydrogen bond to the chromophore hydroxyl group. The chromophore is slightly bent from planarity with a twist and tilt of 1.3 and 9.6, respectively (63).

Table 2-3. Comparison of disulfide torsion angles

	χ^{1n}	χ^{1c}	χ^{2n}	χ^{2c}	χ^{3}	Cα-Cα Å	Cβ-Cβ Å	DSE*
roGFP2 average	-67	-64	-100	-79	116	4.0	4.2	19.5
roGFP1-R7 average	-58	-58	-104	-78	105	4.0	4.1	15.5
rxYFP$^{147}_{202}$ average	-58	-64	-99	-83	103	3.9	4.1	14.7
roGFP1-iR	-80	180	-150	45	86	4.9	3.8	11.7
roGFP1-iE	-59	-178	-157	46	68	4.6	3.5	8.0

For χ angles, n refers to N-terminal cysteine and c for C-terminal cysteine
* Disulfide strain energy kJ/mol

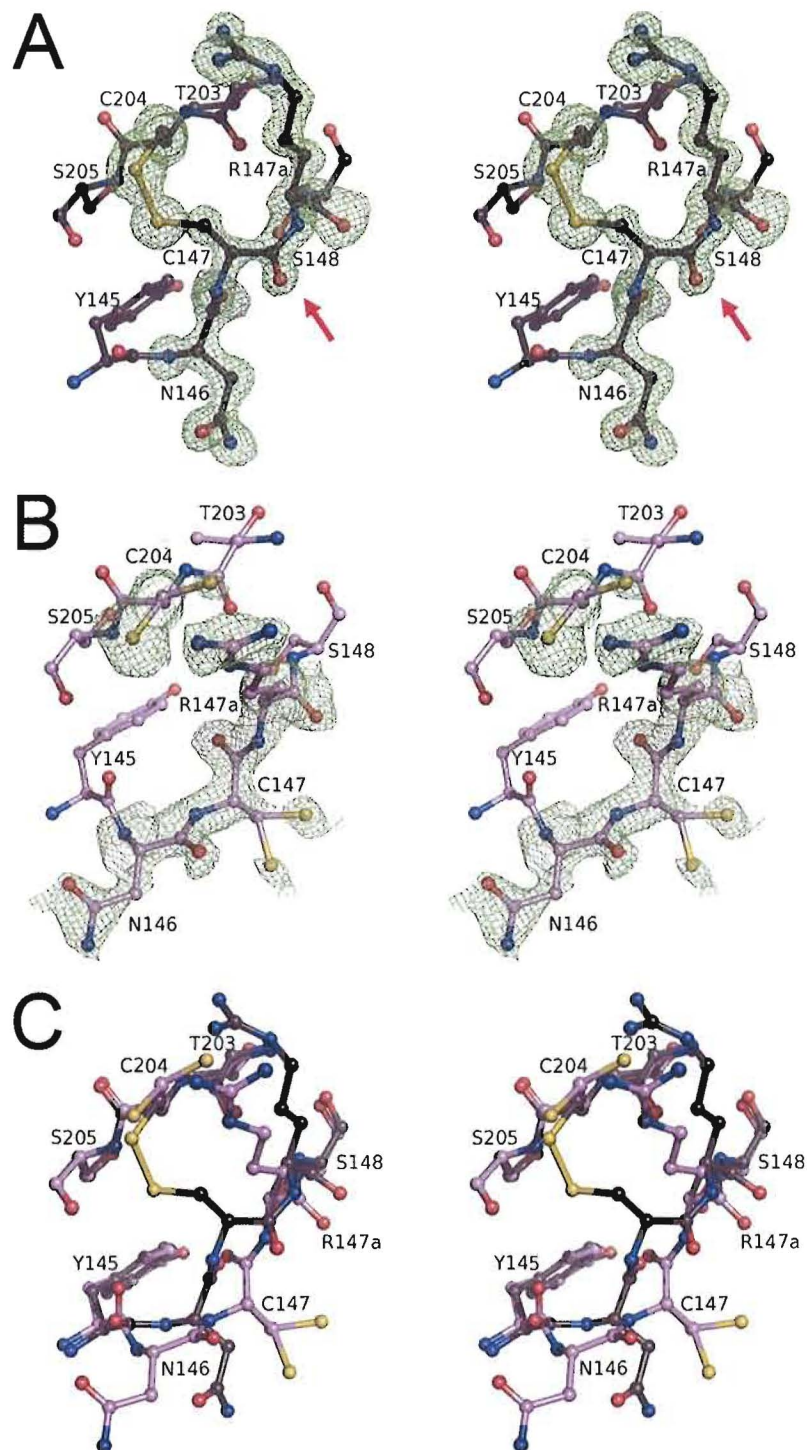**Figure 2-4**. roGFP1-iR in oxidized and reduced states. (A) Stereo image of oxidized state with error weighted F_o-F_c omit map density at 3σ in green. Red arrow denotes rare non-proline cis-peptide bond. (B) Stereo image of reduced state with error weighted F_o-F_c omit map density at 2.5σ in green (C) Overlay of roGFP1-iR in the oxidized (black) and reduced (lavender) states.

Reduced roGFP1-iR. Reduced roGFP1-iR crystals diffracted well and data statistics were acceptable to 1.40 Å. The final model consists of residues 2-231 with none in the disallowed regions of the Ramachandran plot. The final model has an R-factor of 0.185 and an R-free of 0.208. The backbone of residues Asn146, Cys147, Arg147a and Ser148 have clear density, and while the side-chains are found in slightly weaker density, a satisfactory model could be constructed, see Figure 2-4. A large conformational change is observed for residues 146-147a relative to the oxidized state and the peptide bond between residues 147 and 147a is in the normal trans configuration (see later section for a more detailed comparison). Both the Cys147 and Cys204 side chains are disordered and are modeled in two conformations. Ser148 is hydrogen bonded to the chromophore, whereas Arg147a forms a salt bridge with Asp117 of a symmetry related molecule. The chromophore twist and tilt components are 0.7 and 11.2 respectively (*63*).

roGFP1-iE. Diffraction data for oxidized roGFP1-iE were collected to 1.31 Å resolution with good statistics. The final model consists of residues 3-147 and 148-231 with none in the disallowed regions of the Ramachandran plot and an R-factor and R-free of 0.164 and 0.186, respectively. Electron density is clear for all groups adjacent to the disulfide except for the inserted amino acid Glu147a. Although weak density was present for the main chain at position 147a, it was not sufficiently clear to permit modeling of a single conformation and thus this entire residue was omitted from the model, see Figure 2-5. In contrast, Asn146, Cys147 and Ser148 could be satisfactorily modeled. The disulfide of roGFP1-iE shares the same **pg⁻t** conformation as roGFP1-iR.
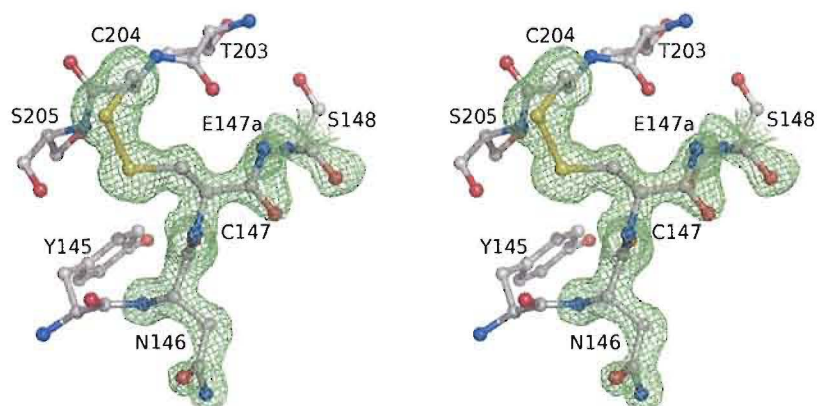
Figure 2-5. roGFP1-iE in the oxidized state. Stereo image with error weighted F_o-F_c omit map density at 3σ in green.

The 1.50 Å model of reduced roGFP1-iE consist of residues 3-231, with non in the disallowed regions of the Ramachandran plot, and a R-factor and R-free of 0.188 and 0.210, respectively. A $2F_o$-F_c electron density map indicates weak density for the backbone atoms of Asn146 and side chain. There is weak F_o-F_c density for two conformations of the Glu147a Cγ but not enough density to model the carboxyl groups. Additional density near ($\sim$2.0 Å) the Sγ of Cys204 in one conformation is present and is unlikely to indicate an oxidation product due to geometric considerations, however the density is likely to be a water molecule corresponding to the alternate conformation of Cys204. Cys147 is also observed in two conformations with weak F_o-F_c density.

Structural comparison of oxidized and reduced roGFP1-iX

Comparisons of the oxidized and reduced states of the variants reveal surprising result. In each case oxidation/reduction induces a dramatic conformational change whereby the segment comprising residues 146, 147, 147a adopts entirely different conformations in the two states. Furthermore, it is clear from the roGFP1-iR variant that the peptide bond between 147-147a undergoes cis-trans isomerization during the open-closed transition, with the cis configuration formed in the oxidized state. Interestingly, the most extreme motion of any atom in the loop 146-147a is exhibited by Sγ of Cys147, which moves approximately 8 Å between the two states (see Figure 2-4).

The Asn146/Cys147/Xaa147a segment is anchored at one end by Tyr145, whose side chain is buried in the core of the protein, and at the other end by Ser148, which is pinned by backbone hydrogen bonds to the adjacent β-sheet and a hydrogen bond to the chromophore. As judged from the refined thermal parameters (B-factors), the open (reduced disulfide) loop conformation is characterized by a higher degree of flexibility than observed for this same segment in the closed (disulfide) state. For oxidized roGFP1-iR, the main chain B-factor average of Asn146/Cys147/Arg147a atoms is 15.2 Å^2, which is larger than the average for the whole structure (residues 7-227) of 8.9 Å^2. Similarly, in the reduced state the B-factors of the loop are 22.6 Å^2 while the average is 12.9 Å^2. roGFP1-iE shows similar behavior with respect to the B-factors with the oxidized and reduced loops having average B-factors of 20.4 Å^2 and 34.1 Å^2 and overall averages of 12.9 Å^2 and 17.0 Å^2 respectively. In any case, the open-closed conformational change is

well described by a two state model, in which only the atoms of the segment 146-147a

are seen to move.

In contrast to the segment 146-147a, the peptide containing Cys204 is essentially

unresponsive to changes in the oxidation state of Cys204. This peptide is locked in place

by virtue of the regular backbone hydrogen bonds to the neighboring β-strand (residues

221 – 225) and by side chains buried within the protein. However, the Cα-Cβ torsion

angle of Cys204 is variable, with the Cys204 Sγ occupying at least two conformations

(Figure 2-4).

X-ray diffraction data were collected with short exposures to minimize x-ray

induced damage. Nevertheless, the weaker density for Cys147 in the reduced states of

roGFP1-iR and roGFP1-iE may be a result of x-ray damage, it may also result from the

flexibility of the loop. Cys204 in the roGFP1-iR structure appears to have no oxidation

artifacts associated and little negative density in the Fo-Fc map as modeled indicating

minor radiation damage.

A feature of oxidized roGFP1-iR is that the hydrophobic portion of the Arg147a

side chain is extended and lies adjacent to the disulfide bridge, partially shielding this

group from solvent. Moreover, the Arg147a conformation is secured by a hydrogen bond

between Nε of the guanidinium group and Oγ of Ser202. We think that this interaction is

important and helps to rationalize the anomalous ordering of midpoint potentials within

the roGFP1-iX series (see Discussion). In the reduced state, this interaction is lost, and

the Arg147a side chain makes polar interactions with a neighboring molecule in the

crystal. For the Glu147a variant, such a stabilizing interaction is either not possible or

energetically unfavorable, as the glutamate side chain is not ordered in either the oxidized or reduced form.

Comparison of roGFP1-iR and roGFP2

Crystal structures are available for roGFP2, which differs from roGFP1 by only a single methyl group (resulting from the Ser65 →Thr substitution). roGFP2 thus serves as a good model for roGFP1, which is the parent for the roGFP1-iX variants described herein. Both oxidized and reduced forms of roGFP2 were reported by Hanson et al. (29). Comparing the two roGFP2 structures, oxidation/reduction results in only very small changes in the positions of the cysteine side chains, thus they are strongly localized by the protein fold. Furthermore, according to a statistical database (61), compiled from available protein structures and calculation of the disulfide energy strain (DSE) (64) (Table 2-3), the geometry of the disulfide linkage in oxidized roGFP2 is highly strained, which is assumed to be a consequence of rigidity in the β-barrel fold of the protein. In strong contrast, the largest motion observed upon oxidation/reduction of roGFP1-iX variants is rearrangement of Cys204, which moves by up to 8 Å suggesting that a great deal of flexibility is imparted by the insertion.

Furthermore, the protein backbone accommodates the insertion between residues 147 and 148 in different ways dependent upon the oxidation state (Figure 2-4). Two positions, Tyr145 and Ser148, pin the ends of the segment containing the insertion. Tyr145 is fixed as part of the hydrophobic core of the protein, whereas Ser148 is oriented by hydrogen bonds in the regular β-sheet around that position. However, the intervening

peptide conformation varies greatly depending upon the oxidation state of the thiol/disulfide. In the oxidized state, formation of the disulfide forces Arg147a to bulge out of the barrel (Figure 2-6). In the reduced state, Arg147a appears to be in a position similar to that of the Cys147 in roGFP1 and instead, Asn146 is bulged out. Hence, the actual dislocation resulting from the insertion translocates in response to the side chain oxidation state, much as observed for insertions within the backbone of helical regions in T4 lysozyme (65). For both wild-type GFP and roGFP2 in the reduced state the Asn146/Ser147/His148 strand has B-factors that are comparable with surrounding residues, which is in contrast to roGFP1-iR where the B-factors are much larger than average for this loop. This suggests that in addition to two-state conformational flexibility (i.e. static disorder), the insertion also confers to the backbone greater microscopic heterogeneity (dynamic disorder).
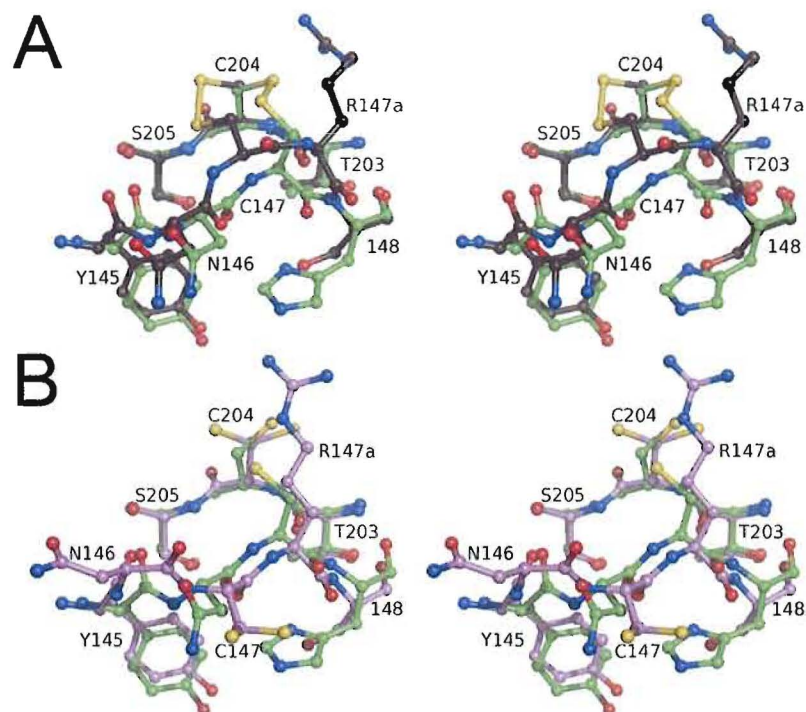
Figure 2-6. roGFP1-iR and roGFP2 comparison (A) Overlay of oxidized roGFP2 (green) and roGFP1-iR (black). (B) Overlay of reduced roGFP2 (green) and roGFP1-iR (lavender).

DISCUSSION

Two-state indicators that depend upon chemical equilibrium are useful for quantitative analysis over a rather narrow range, centered upon the conditions where the equilibrium constant for interconversion of the two indicator species is one. Inevitable measurement errors introduce further restrictions on this range. roGFPs 1–6 exhibit midpoint potentials that range from -272 mV (roGFP2) to -299 mV (roGFP3). In practice, roGFPs are most useful for measuring the thiol/disulfide equilibrium within ~35 mV of the indicator midpoint, corresponding to ~ 10-90% oxidized. roGFP1 (midpoint ~ -290 mV), for example, is most useful for quantitative measurements within a range of ~

-255 to -325 mV at pH 7.0. However a much wider range is found within cells from around -360 mV for the mitochondria (*29, 33*) to -180mV for the ER (*35*).

Partly in effort to extend this range limitation, several groups have introduced surface positive charges such as arginine and lysine residues into the vicinity of the reactive disulfide. In theory, these modifications favor formation of the cysteine thiolate, which in turn destabilizes the disulfide. This has two important consequences, one of which is to significantly accelerate the rate constants for oxidation and reduction (by increasing the concentration of the reactive thiolate), while the second is to shift the midpoint potential toward less negative values. Dooley et al. (*32*), Cannon and Remington (*44*) and Hansen et al. (*45*) successfully applied this approach.

Of the roGFP1-Rn variants described by Cannon and Remington (*44*), roGFP1-R12 and -14 (each with three basic substitutions) exhibit the least negative midpoint potentials of –265 mV and –263 mV, respectively. Introduction of three basic groups thus leads to rather modest shifts of 22-24 mV, compared to roGFP1 at about -287 mV.

Despite the apparent success of this approach and the general agreement with theory, the unusual stability of the disulfide in roGFP1/2 is puzzling because the high resolution crystallographic studies of roGFP2 revealed it to be geometrically quite strained (*29*). Given that our goal is to achieve more dramatic shifts in the midpoint potential, a more invasive approach to destabilize the disulfide linkage is required. To this end, we inserted single amino acid residues within the polypeptide backbone, adjacent to the reactive disulfides. A further requirement is that the indicator response be ratiometric; in this case, the response should entail a shift in equilibrium between two

forms excited at ~ 400 or ~ 475 nm. Initial experiments using glycine insertions produced soluble but non-fluorescent protein (for insertion around Cys204) or indicators that were not ratiometric (insertion around Cys147). However, for one promising mutant, ratiometric behavior was restored utilizing random mutagenesis at certain selected sites. This effort produced a family of redox-sensitive fluorescent proteins termed roGFP1-iX, where the distinguishing feature is the identity "X" of the amino acid inserted at position 147a, adjacent to Cys147.

As shown in Table 2-1, these variants have midpoint potentials in the range of ~ -229 to ~ -246 mV. Compared to the parent roGFP1, this corresponds to shifts of ~ 50-60 mV toward less negative midpoint potentials, indicating that the apparent equilibrium constant for disulfide formation in roGFP-iX is reduced by approximately an order of magnitude. All the variants have significantly increased rates of reduction by DDT over that of the parent roGFP1, however the rate of oxidation by cystine (CSSC) is similar for all variants. Closer inspection of Table 2-1 reveals several interesting and counterintuitive results. Consider, for example, the variants roGFP-iL, roGFP-iR and roGFP-iE which have leucine, arginine and glutamate at position 147a, respectively. From the discussion above, it is expected that the positive charge on the arginine side chain would destabilize the disulfide compared to leucine or glutamate at this position, but the reverse is true. To rationalize these results, we compared the crystal structures of roGFP-iR and roGFP-iE in both the oxidized and reduced states. In the following section, we attempt to explain the apparent reduction in disulfide stability in terms of geometric, electrostatic and entropic contributions.

Geometric strain

Comparison of the 147/204 disulfide geometry (Table 3) with the statistical database compiled by Srinivasan et al. (*61*) suggests that the disulfide linkage is only slightly atypical, for example, the disulfide χ^2 values of ~ 180° are infrequently observed. Thus, strain in the disulfide linkage itself is not likely to be a significant factor accounting for the reduced thermodynamic stability compared to roGFP1 or 2. Indeed, on a geometrical basis alone, the disulfide in roGFP1-iR should be thermodynamically more stable that that seen in roGFP2. However, geometric strain in this system is apparent, which is illustrated by the presence of the rare and unfavorable cis-peptide bond between Cys147 and Arg147a (see Figure 2-4). Upon reduction of the disulfide, the cis peptide bond isomerizes to the more relaxed trans form. We conclude that trans to cis peptide isomerization is required in order to form an energetically favorable disulfide linkage, and that this component of the overall energetic status substantially reduces the apparent equilibrium constant for disulfide formation. The role of non-proline cis-trans peptide isomerizations in the active center of a protein has been demonstrated before in concanavalin A, suggesting these isomerizations are not completely novel (*66*).

Electrostatic effects

In theory, placement of a positive charge adjacent to a disulfide should decrease thermodynamic stability due to stabilization of the cysteine thiolate. This has been experimentally verified by several groups (*41, 44, 67, 68*), even to the level of good quantitative agreement with nonlinear Poisson-Boltzmann theory [Cannon & Remington

(*44*)]. However, comparing the midpoint potentials of roGFP1-iR and roGFP1-iE or

roGFP1-iL (Table 2-1), the variant with the positively charged arginine side chain has the

most stable disulfide. On the basis of the crystal structures, we suggest that this ordering

is due to a specific interaction found only in the oxidized state of roGFP1-iR, where the

guanidinium group forms a hydrogen bond to Ser202 Oγ. In comparison, the Glu147a

side chain is disordered in either state of roGFP1-iE. This specific interaction,

preferentially stabilizing only the oxidized state, is proposed to offset the generalized

electrostatic effects. The large loop movement towards the open state also removes the

147a side chain from the immediate vicinity of Cys204, reducing electrostatic influences

at this reactive center.

Entropic effects

The solution response of roGFP-iX indicators to redox titration is accurately

modeled by a theoretical curve assuming two states, open and closed, in chemical

equilibrium Figure 2-1. However, the crystallographic results indicate that the open form

of loop 146-147a has high flexibility as the B-factors are higher than average. In the

closed state of roGFP-iR, this loop, including the side chain of Arg147a, is well ordered.

This suggests that entropic considerations would favor the open, reduced state due to the

increased number of degrees of freedom. Nevertheless, a two-state structural model that

includes vibrational parameters is an adequate model for the transition, so the effect

cannot be expected to result in a large reduction in the thermodynamic stability of the

disulfide. The situation is less pronounced in the roGFP-iE variant, as both the main and

side chain of Glu147a are to some extent disordered in either the reduced or oxidized states, however it is still true that main chain vibrational parameters are relatively higher in the open conformation than in the closed.

On the other hand, the large conformational change of the 147 loop, with a consequently large physical separation between the reactive species in the reduced state, can be rationalized in terms of a low effective concentration of the reactive species. Alternatively, there are undoubtedly many loop conformations on the reaction pathway that reduces the probability of disulfide formation. This is in contrast to roGFP1, in which the reduced cysteines are held in close proximity by the protein backbone and thus only minor rearrangements are required for disulfide formation.

While it is difficult to estimate the magnitude of the loop entropy contribution to the thermodynamic stability of the disulfide, the effect is unlikely to be negligible. In concert with the geometric strain argument advanced previously, the two contributions are consistent with the observed reduction in disulfide stability. If correct, this argument suggests that larger insertions near the 147 position, possibly even containing the reactive cysteine, into the backbone of roGFPs could result in probes suitable for use in even more oxidizing environments. GFP is known to accommodate the insertion of entire proteins, such as calmodulin (26), in the vicinity of residue 146, so further exploration of our "insertion approach" has a high probability of success.

Suitability of roGFP-iX probes for use *in vivo*

As described previously the range of activity of the redox indicators is around ±35 mV from the midpoint potential, and for the most oxidizing indicator, roGFP1-iL, the range would be from -265 to -195 mV. Assuming the midpoint potential of the ER is -180 mV as reported from isolation of glutathione levels, roGFP1-iL falls short of being able to determine the redox potential in that cellular environment. There are discrepancies in the reported values of the cytosolic redox potential between the methods of glutathione isolation (-232 mV (*35*)) and use of fluorescent redox sensors (-320 mV (*32*)), and as such the ER could be more reducing than previously indicated. What is known is that the ER redox potential is greater than -255 mV based on use of roGFP1 in that environment (*36*). The pH of the ER and Golgi have been reported to be 7.4 and 6.2 (*69*) which is in the range of function of roGFP1-iL or roGFP1-iE suggesting they should work fine in these environments. It is recommended based on the lack of intermolecular disulfide formation, spectral properties and highest redox potentials that roGFP1-iE and especially roGFP1-iL be used for *in vivo* experimentation.

CHAPTER III

ATOMIC RESOLUTION STRUCTURES OF *ESCHERICHIA COLI* AND

BACILLUS ANTHRACIS MALATE SYTHASE A: COMPARISON WITH

ISOFORM G AND IMPLICATIONS FOR STRUCTURE BASED DRUG DESIGN

Experimental data presented in this chapter are my own with some data collection

done by Andrew Olson. I prepared the original manuscript. Jim Remington provided

direction, support and edited the manuscript.

SUMMARY

Enzymes of the glyoxylate shunt are important for the virulence of pathogenic

organisms such as *Mycobacterium tuberculosis* and *Candida albicans*. Two isoforms

have been identified for malate synthase, the second enzyme in the pathway. Isoform A,

found in fungi and plants, comprises ~530 residues, whereas isoform G, found only in

bacteria, is larger by ~200 residues. Crystal structures of malate synthase isoform G from

Escherichia coli and *Mycobacterium tuberculosis* were previously determined at

moderate resolution. Here we describe crystal structures of *Escherichia coli* malate

synthase A (MSA) in the apo form (1.04 Å resolution) and in complex with acetyl-

coenzyme A and a competitive inhibitor, possibly pyruvate or oxalate (1.40 Å

resolution). In addition, a crystal structure for *Bacillus anthracis* MSA at 1.70 Å

resolution is reported. The increase in size between isoforms A and G can be attributed

primarily to an inserted α/β domain of unknown function. Upon binding of inhibitor or

substrate, several active site loops in MSA undergo large conformational changes.

However, in the substrate bound form, the active sites of isoforms A and G from

Escherichia coli are nearly identical. Considering that inhibitors bind with very similar

affinities to both isoforms, MSA is as an outstanding platform for high-resolution

structural studies and drug design efforts.

INTRODUCTION

Malate synthase (MS) and isocitrate lyase make up the glyoxylate shunt (*11*),

which ultimately converts isocitrate to malate. The actions of these enzymes bypass two

steps of the citric acid cycle that evolve CO_2, permitting bacteria, fungi and plants to

utilize C2 carbon sources such as acetate and fatty acids for net biosynthesis. Enzymes of

the glyoxylate shunt have been implicated for the virulence and persistence of many

different pathogens, especially *Mycobacterium tuberculosis* (mTB) (*70-72*) and *Candida*

albicans (*73-75*). As mammals do not have genes encoding either glyoxylate shunt

enzyme, they have become attractive targets for drug design (*76*).

Two distinct isoforms of MS, A (MSA) and G (MSG) (*77*) have been identified.

Members of the isoform G family share >50% identity and are found only in bacteria

(*78*). Likewise, MSA isoforms share high identity but are found in fungi and plants as

well as bacteria. The eukaryotic MSAs form homomultimers, which distinguishes them

from prokaryotic MSAs (79). Both isoforms are represented in pathogenic organisms, for

example, mTB utilizes a G isoform, whereas *Candida albicans* utilizes isoform A.

Escherichia coli is thus far unique in that it has genes for both MSA and MSG,

encoded by the genes *aceB* and *glcB* respectively (80). The proteins are differentially

expressed in *E. coli* with isoform A expression being regulated by fatty acids and acetate

while isoform G expression is regulated by glycolate (81). *E. coli* MSA (ecMSA)

comprises 533 amino acids whereas *E. coli* MSG (ecMSG) comprises 723 amino acids.

The two isoforms are very distantly related. Sequence comparisons suggest that the

larger size of the G isoform can be attributed to the presence of one or more insertions

(13, 78, 80), while the conserved segments of the amino acid sequence show only ~18%

identity. Crystal structures of malate synthase G from *E. coli* (13, 14) and from mTB (78,

82) have been determined, but unfortunately the diffraction resolution is modest, which

limits structure-based drug design efforts (83).

Structural studies of mTB MSG and ecMSG reveal four structural domains (13,

14, 18, 78, 82, 84). A α8/β8 (TIM) barrel is centrally located and contains the active site.

It is buttressed on one side by an N-terminal α-helical clasp and on the other by an α/β-

domain consisting of two insertions into the barrel. The C-terminal segment, which

comprises several α-helices, caps the active site. As of yet, no function has been

attributed to the α/β-domain. Sequence alignments proposed by a few groups disagree

with respect to which domains are missing from the structure of MSA. Alignments from

Smith (78) and Howard (13) suggest the N-terminal clasp domain is missing whereas the

alignment of Molina (*80*) predicts the absence of the α/β-domain, however both alignments agree that the C-terminal cap and the TIM barrel are well conserved.

In the second step of the glyoxylate shunt pathway, the condensation and subsequent hydrolysis of glyoxylate and acetyl-CoA is performed by MS to produce malate and CoA. The modes of substrate binding are known in some detail from crystallographic and NMR structural studies (*13, 14, 18, 78, 82, 84, 85*) and a catalytic mechanism has been proposed. Briefly, bound glyoxylate is activated towards nucleophilic attack by salt bridges to an essential Mg^{2+} ion and by hydrogen bonds to the protein backbone and a conserved arginine. The acetyl-CoA terminal methyl group is deprotonated by the catalytic aspartate residue. The proposed enolate intermediate is stabilized by interaction with the catalytically essential arginine.

The binding site for the substrate glyoxylate is deep within the protein, connected by a channel to the solvent-exposed acetyl-CoA binding site. Prior to the structural studies, the results of small angle x-ray scattering experiments suggested "opening" or separation of domains so that the active site could sequester substrates from the solvent (*15*). Substrate-induced conformational changes were also consistent with circular dichroism data on MS isolated from maise and yeast (*16, 17*). In contrast, data from subsequent NMR studies of apo ecMSG were inconsistent with domain motions, but did suggest disorder in a loop (Thr451-Asp455) that could control access to the active site (*18*).

We report here crystal structures of ecMSA and *Bacillus anthracis* MSA (baMSA). These structures reveal that the conserved portions of MSA and MSG are very

similar and that the difference in size is due to the α/β-domain appended to the G isoform.

Comparisons of *E. coli* MSA apo and ternary complex structures reveal significant loop

movements, which we suggest are necessary for substrate binding and product release.

Active site comparisons of the two substrate-bound isoforms, combined with kinetic

studies, suggest that either isoform could be used as a basis for drug discovery.

METHODS AND MATERIALS

Cloning, mutagenesis, overexpression and purification

The gene for ecMSA was amplified from genomic DNA using primers containing

restriction endonuclease cut sites. The fragments after digestion were cloned into a

modified pBAD expression vector containing a TEV protease site in place of the

enterokinase cleavage site. The ecMSA C438S mutant was prepared using the

QuickChange Site-Directed Mutagenesis Kit (Stratagene) and the presence of the

mutation was verified by DNA sequencing. A synthetic gene coding for the *Bacillus*

anthracis MSA gene was produced by Blue Heron Biotechnology and sub-cloned

similarly to the ecMSA gene. The ecMSG C617S mutant was prepared as described

previously (*14*).

The pBAD vector containing ecMSA or baMSA was transformed into *E. coli*

TOP10 cells. These cells were cultured overnight in 100 mL Luria-Bertani media

supplemented with ampicillin (100 μg/mL). This culture was used to inoculate 4 L of

media in a fermenter at 37°. Once the culture reached an A_{600} of 0.6-0.8, 2 g arabinose

was added to start overexpression. After 4-hours the culture was harvested by

centrifugation, disrupted by sonication with the addition of lysosyme and DNAse, followed by centrifugation to remove insoluble material.

Purification was performed using a Ni-NTA column. The His-tag was removed using 0.15 mg TEV protease per 20 mg purified protein and checked for cleavage using MALDI/TOF MS. After complete cleavage the protein was further purified using size exclusion or anion exchange chromatography. The protein was dialyzed into 50 mM Tris pH 8.0, 20 mM KCl, 5 mM MgCl2, 1 mM TCEP, 0.1 mM EDTA then quantified using the calculated ε_{280} 81.8 mM^{-1} cm^{-1} based on the method of Gill and von Hippel (86), and concentrated to approximately 20 mg/mL prior to crystallization.

Crystallization, derivatization and diffraction data collection

Large, diffraction quality crystals of ecMSA were readily obtained from many conditions by using the hanging drop vapor diffusion technique. The optimized condition for apo ecMSA was 0.1 M Tris pH 7.5, 0.2 M calcium acetate, 22% PEG 8K (1:1, protein: reservoir solution, 2 μL drops). A heavy atom derivative was prepared by soaking a crystal for one hour in HgCl$_2$ saturated mother liquor (87). Crystals were flash frozen and data were collected on a Rigaku rotating anode generator with a Raxis 4 detector. To produce the Mg^{2+} bound ecMSA structure, crystals were placed in artificial mother liquor where Ca^{2+} was replaced by Mg^{2+} and stored for 7 days before data collection data at the Advanced Light Source (Berkley, CA), beam line 5.0.2 using the ADSC-Q315 detector at a wavelength of 0.90 Å and temperature of 100K. To obtain the ecMSA acetyl-CoA/inhibitor ternary complex, protein was mixed with 0.1 M acetyl-CoA

and 0.2 M pyruvate, which was then again subjected to crystallization trials. The ternary

complex crystallized from 0.1 M PIPES pH 6.5, 0.1 M ammonium sulfate, 27% PEG 8K.

Crystals of baMSA grew from 0.2 M sodium citrate, 0.1 M Tris pH 8.0, 30% PEG 1550

from which two crystal forms are found, small needle clusters and large plates. Large

plates of baMSA were used for data collection. It was later found that inclusion of 30-70

mM NiCl$_2$ in the crystallization buffer encouraged formation of the plates over the

needles. Synchrotron radiation at ALS beamline 5.0.2 (wavelength 1.00 Å) was used for

data collection on the ecMSA inhibitor/acetyl-CoA complex and for the baMSA crystal.

Data reduction and structure determination

The home source diffraction data for apo ecMSA were processed and scaled with

the HKL suite (55). Initial phases were obtained and modified with SHELXC/D/E using

the SIRAS method (88, 89). Five heavy atom sites were found. ARP/wARP was then

used to build the model into the experimentally phased electron density map (56). Five

hundred of the five hundred thirty three amino acids were automatically built. The

phases from ARP/wARP were extended to 1.68 Å using REFMAC 5 (57). Additional

amino acids were built using COOT (58). Water molecules were placed using COOT

with the combined criteria of a peak of greater than 2.5 σ in the (F$_o$-F$_c$) difference map, or

1.0 σ in the (2F$_o$-F$_c$) map and reasonable intermolecular interactions. Synchrotron data

were processed using the HKL2000 suite (55). Molecular replacement, using MOLREP

(90), was performed to solve the structures of the acetyl-CoA/inhibitor ecMSA and

baMSA structures using the apo ecMSA structure as a starting point. ARP/wARP was

also used to rebuild the models of the acetyl-CoA/inhibitor bound ecMSA and baMSA

(*56*). The magnesium bound ecMSA, acetyl-CoA/inhibitor bound ecMSA, and baMSA

structures were all refined using REFMAC 5 (*57*).

Enzyme activity assay

The reaction of acetyl-CoA and glyoxylate catalyzed by ecMSA and baMSA was

quantified by measuring the decrease in absorbance at 232 nm due to cleavage of the

thioester bond of acetyl-CoA as described previously (*14, 79*). The reaction conditions

were 0.2 mM acetyl-CoA, 0.5 mM glyoxylate, 50 mM Tris pH 8.0 and 10 mM $MgCl_2$.

The reaction was initiated by the addition of enzyme solution. Kinetic parameters for

substrates were determined using the above conditions with varying substrate

concentrations. Least squares fits for kinetic parameters were obtained using the program

Enzfitter (Biosoft). Inhibition constants for parabanic acid and 4-amido-2,3,4-

pyrrolidenetrion (which absorb in the vicinity of 232 nm where the thioester linkage of

acetyl CoA absorbs), were determined using an assay that monitors the appearance of

free CoA via its reaction with 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) and the

concomitant rise in absorbance at 412 nm (*91*). However, there was a concern that

Cys438, a conserved residue in the active site of MS, might react with DTNB to block

catalysis, so Cys438 was mutated to serine prior to kinetic analyses. The assay

conditions were similar to above with a 1 mM concentration of DTNB.

RESULTS

We successfully cloned ecMSA from genomic DNA. Since genomic DNA was not available for baMSA, the gene was synthesized commercially. Overexpression yielded protein for both ecMSA and baMSA. Kinetic and inhibition constants were obtained for both ecMSA and baMSA allowing comparison with ecMSG. Both ecMSA and baMSA crystallized readily, but ecMSA produced crystals of much higher diffraction quality.

Kinetics and inhibition of malate synthase

The kinetic properties of ecMSA and baMSA are very similar and in addition are similar to those of ecMSG as shown in Table 3-1. ecMSA Cys438, corresponding to Cys617 in ecMSG, was mutated to Ser similarly to as performed by Anstrom et al. (*14*). Inhibition by the glyoxylate competitive inhibitors pyruvate, parabanic acid and 4-amido-2,3,4-pyrrolidenetrione were very similar between the isoforms, indicating that either isoform should be useful for drug discovery. The binding of pyruvate and parabanic acid are relatively weak with low to sub millimolar inhibition constants, whereas 4-amido-2,3,4-pyrrolidenetrione binds better with mid-micromolar inhibition constants and could be improved. Parabanic acid evidently undergoes base and/or enzyme catalyzed hydrolysis under the conditions used in the assay, however, the reaction is slow enough that inhibition constants could be obtained. While the affinity of parabanic acid for the enzyme is modest, in theory it is readily functionalized, therefore parabanic acid could be an excellent lead compound for drug discovery.

Table 3-1. Kinetic and inhibition constants for malate synthases

Enzyme	ecMSA WT	ecMSA C438S	ecMSG WT[b]	ecMSG C167S[b]	baMSA
specific activity[a]	2500±100	1850±50	3400±35	3000±60	2300±20
% of wild type activity	100	74	100	88	100
Km glyoxylate (μM)	55±7	58±7	21±2	50±4	85±2
Km acetyl-CoA (μM)	22±3	56±2	9.0±1.1	46±3	23±3
Ki pyruvate (mM)	-	0.600±0.008	1.0±0.2	1.00±0.09	0.750±0.025
Ki parabanic acid (μM)	-	370±50	-	550±60	-
Ki pyrrolidinetrione (μM)	-	40±7	-	113±12	-

(a) units of mol malate/min./mol protein

(b) values taken from Anstrom et al. 2003, except data regarding parabanic acid and pyrrolidinetrione

ecMSA apo form structure

E. coli MSA crystallized readily from calcium acetate/PEG 8K in space group $P2_1$ with a single monomer per asymmetric unit and unit cell dimensions of $a = 52.0$ Å, $b = 74.1$ Å, $c = 71.9$ Å with $\beta = 97.8°$. These crystals diffracted well and data were collected using the home source to 1.68 Å resolution with excellent statistics (see Table 3-2). The structure was solved using the SIRAS method using a $HgCl_2$ heavy atom derivative data set which diffracted to 1.80 Å. The final model has a crystallographic R factor of 0.172 ($R_{free} = 0.200$). It consists of residues 5-533 with two Ca^{2+} ions and one molecule each of glycerol and acetate. Only amino acid, Glu119, is found in the generously allowed region of the Ramachandran diagram, with none in disallowed regions.

Table 3-2. Malate synthase A data collection and refinement statistics

Crystal	ec native	ec derivative	ec Mg^{++} soaked	ec inhibitor	baMSA
Data collection					
Total observations	214,408	186,790	766,437	785,650	138,489
Unique reflections	59,002	48,087	195,926	229,489	43,852
Cell dimensions a (Å)	52.0	51.9	52.1	75.2	97.8
b (Å)	74.1	73.7	73.9	71.5	52.0
c (Å)	71.9	71.9	71.7	103.3	89.1
$\alpha=\gamma=90°$	$\beta=97.8°$	$\beta=97.7$	$\beta=98.0$	$\beta=90.1$	$\beta=98.9$
Resolution (Å)	50-1.68	50-1.80	30-1.04	30-1.40	50-1.70
Highest resolution shell (Å)	1.74-1.68	1.86-1.80	1.08-1.04	1.45-1.40	1.76-1.70
Completeness[a] (%)	96.9	97.6	89.5	97.1	89.9
	(97.8)	(94.6)	(50.9)	(94.7)	(62.3)
Average I/σ^a	26.1	19.8	12.2	25.1	19.1
	(2.6)	(2.3)	(2.15)	(1.8)	(2.4)
R_{merge}[a,b]	0.055	0.081	0.078	0.057	0.057
	(0.427)	(0.454)	(0.399)	(0.551)	(0.280)
Refinement					
Space group	P2$_1$	P2$_1$	P2$_1$	P2$_1$	C2
No. of molecules[c]	1	1	1	2	1
No. of protein atoms[c]	4236		4341	8504	4010
No. of solvent atoms[c]	516		641	1159	180
Resolution range (Å)	30-1.68		30-1.04	30-1.40	30-1.70
Crystallographic R-factor[d]	0.177		0.153	0.173	0.201
R-free (reflections)	0.208		0.166	0.200	0.245
Average B-factors (Å^2)					
Protein atoms (main chain)	14.3		10.2	19.4	24.6
Solvent	21.5		19.0	29.5	30.1
Root mean square deviations					
Bond lengths (Å)	0.013		0.006	0.009	0.019
Bond angles(°)	1.36		1.07	1.21	1.78

[a] Values in parentheses indicate statistics for the highest resolution shell.

[b] $R_{merge} = \Sigma I <I>/ \Sigma <I>$, where I is observed intensity, $<I>$ is average intensities from multiple observations of symmetry-related reflections.

[c] Per asymmetric unit.

[d] R-factor $= \Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure amplitudes, respectively.

ecMSA ternary complex structure

Solutions containing ecMSA, acetyl-CoA and pyruvate in ammonium sulfate/PEG 4K yielded ternary complex crystals in space group $P2_1$ with two protomers per asymmetric unit. The unit cell dimensions are $a = 75.2$ Å, $b = 71.5$ Å, $c = 103.3$ Å with $\beta = 90.1°$. These co-crystals diffracted well and data with good statistics were collected to 1.40 Å (Table 2). The final model was solved by molecular replacement using the apo-form model, and has a final crystallographic R factor of 0.173 ($R_{free} = 0.200$). The final ternary complex model consists of residues 9-533, one Mg^{2+} ion, either pyruvate or oxalate and one molecule of acetyl-CoA for each protein chain. 1159 water molecules were placed. Alignment of the A and B molecules in the asymmetric unit shows them to be essentially identical with a rmsd of 0.05 Å. A single residue, Glu119 in both monomers, was in a generously allowed region of the Ramachandran diagram, and no residues are in the disallowed regions. However, the unusual conformation of Glu119 is supported by clear electron density.

Overall structure

E. coli malate synthase A is fashioned from an N-terminal α-helical clasp (Domain I, residues 5-60) which interfaces with a α8/β8 TIM barrel (Domain II, residues 92-383) and a C-terminal α-helical plug (Domain III, residues 414-533), as shown in Figure 3-1, panel A. Domains I and II are linked by a loop region (resides 61-91) containing a single α-helix of approximately the diameter of the protein, covering the "bottom" of the TIM barrel. Residues 384-413 form a similar helical region that extends the diameter of the protein to connect domains II and III.

Figure 3-1. Structure of ecMSA with acetyl-CoA and inhibitor bound. Panel A. Stereo-view of overall structure. TIM barrel (green), C-terminal plug (red), N-terminal clasp (blue). CoA fragment and inhibitor shown in lavender ball-and-stick with magnesium as orange sphere. Black ball-and-stick residues from left-to-right are R276, W277, M330, P369 and M102. Panel B. Stereo-view of active site inhibitor with omit Fo-Fc density shown at 3σ for the inhibitor, magnesitum and a few residues. Black dashes represent hydrogen bonds with the red dashes showing the "bad contact" with pyruvate or 2.7 Å hydrogen bond to oxalate which is modeled. Based on comparisons with MSG the blue dashes represent the hypothetical interaction of Asp447 and the terminal methyl group of acetyl-CoA. Panel C. Stereo-view of the CoA binding site with omit Fo-Fc density shown at 2σ for CoA and Cys438. There is some density for the panthothenic acid moiety but it is too broken to accurately model. The waters at the top of the panel interacting with D447 represent the hypothetical location of the terminal acetyl group.

A

B

C
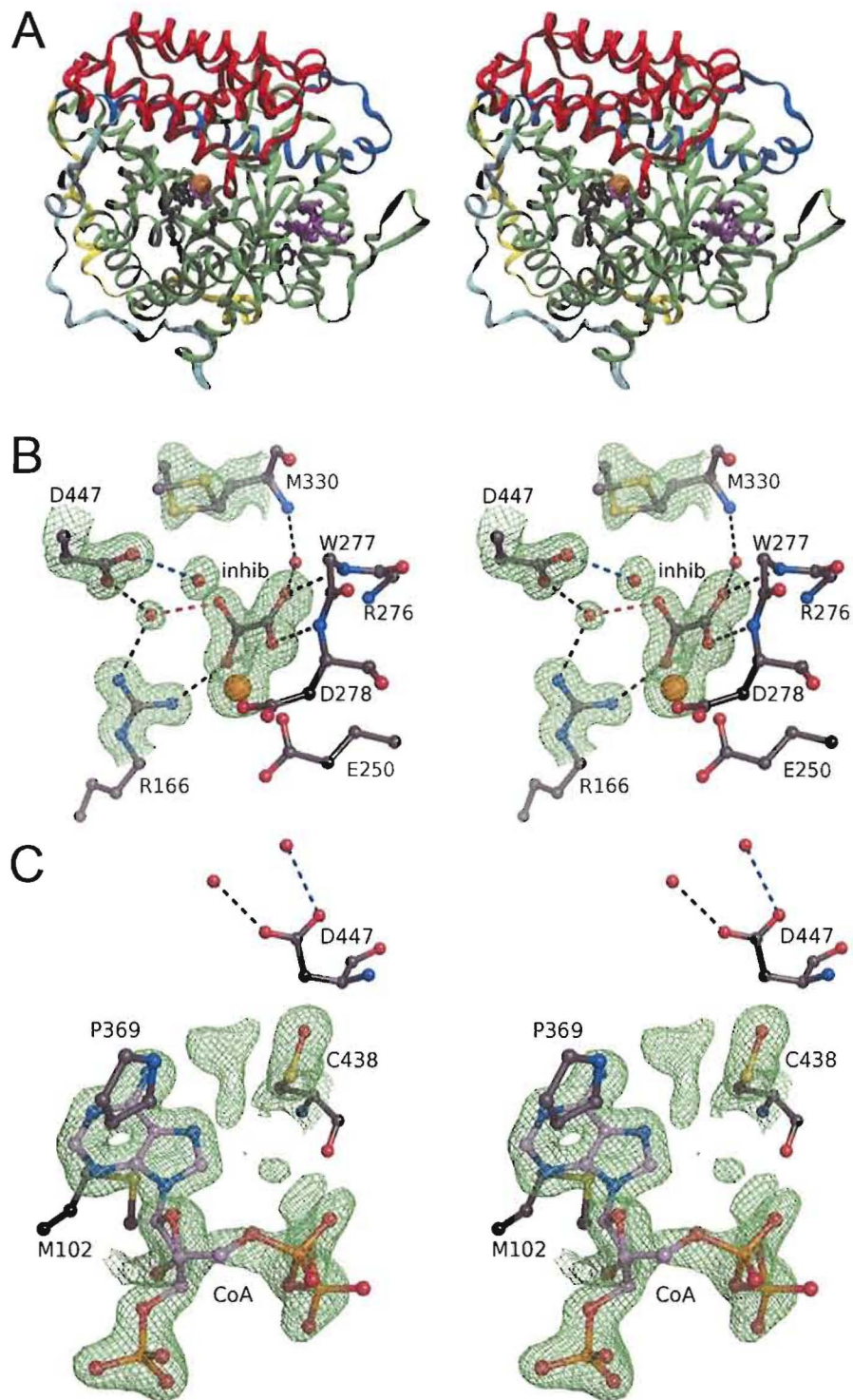
Active site of apo ecMSA

The active site of apo ecMSA has two channels leading to the solvent, in an approximately perpendicular orientation. Residues 439-446 in the C-terminal plug form a strand-β-turn-strand, within which residues 443-445 interface with residues 332-334 in the TIM barrel to separate the channel openings. The active site contains a large peak in the electron density map that was initially modeled as an octahedrally coordinated magnesium atom ligated by four waters and the carboxylates of residues Glu250 and Asp278. However, significant positive peaks in the (F_o-F_c) difference electron density map at the metal site suggested that instead a calcium ion is bound. Indeed, calcium is the major di-cation in the crystallization medium and metal-ligand distances (average 2.33Å) are more consistent with calcium (2.39-2.36 Å) than magnesium (2.07 Å) (*92*).

Crystals identical to those described above were transferred to artificial mother liquor in which Ca^{2+} was replaced by Mg^{2+} and synchrotron diffraction data were collected to 1.04 Å resolution (Table 3-2). The final atomic model consists of residues 5-533, 4 Mg^{2+} ions and 641 water molecules. Many carboxylate and sulfur containing side chains have lowered occupancy, presumably due to radiation damage. An anomalous scattering difference map exhibited no significant peaks at the presumed metal site, indicating that calcium is not present in the magnesium-substituted crystals (data not shown). After refinement of the atomic model, the average metal-ligand coordination distances is 2.10Å, consistent with the replacement of Ca^{++} with Mg^{++} (*92*). The remainder of the Mg^{++}-containing model is essentially identical to the Ca^{++}-containing

model (rmsd 0.09 Å) with the exception of residues 335-343. These form a flexible loop

with B-factors that are much higher than the average for both structures.

Active site of ecMSA ternary complex

The active site of ecMSA is buried between the TIM barrel and the C-terminal

plug. The electron density map was easily interpreted as containing an octahedrally

coordinated magnesium ion ligated by two water molecules and one molecule of pyruvate

or oxalate (see below and Discussion), see Figure 3-1, panel B. Despite the fact that

pyruvate was included in the crystallization mixture, the electron density map clearly

shows that the bound species is not planar as expected. The dihedral angle defined by the

oxygen atoms bound to the metal and intervening carbons refines to a value of -27°. The

bound species makes hydrogen bonds to the backbone amides of Arg276, Trp277,

Asp278 and Mg^{++} as well as to the guanidinium group of Arg166. Three water molecules

contact the bound species. One is also in contact with the amide nitrogen of Met330

while the others contact the guanidinium group of Arg166 and the carboxylate of the

proposed catalytic base Asp477. Of particular interest are the two interactions from

bound water molecules to the group corresponding to the methyl carbon of pyruvate,

which is out of the molecular plane. One involves a very close approach (2.74 Å) and is

thus either a severe steric clash (if a methyl group) or a normal hydrogen bond (if an

oxygen atom). A second presumed water also makes a contact with this group (3.10 Å).

Since both interactions would be favorable for an oxygen atom, we interpret the bound

species to be oxalate, rather than pyruvate (see Discussion). If the bound group were

indeed pyruvate, there is no obvious explanation for the observed deviation from planarity or close contacts of the presumed methyl group with bound solvent molecules.

Acetyl-CoA binding pocket

(Acetyl-) coenzyme A is clearly bound to the enzyme, as there is strong, well resolved electron density for the adenine ring and ribose. However, density is weaker for the phosphate groups, see Figure 3-1, panel C. The adenine ring is sandwiched between residues Pro369 and Met102. Hydrogen bonds are made between the N6 position of the adenine ring and backbone carbonyls of Thr95 and Ala367. Asn105 makes contacts with the ribose hydroxyl and 3'-phosphate. The 3'-phosphate is also bound by Lys101 and Tyr154. The 5'-phosphate is bound by His368. However, density for the panthothenic acid tail of acetyl-CoA is too weak to interpret in either protomer within the asymmetric unit. This may result from partial hydrolysis or multiple binding conformations (see Discussion). Cys438 in the acetyl-CoA binding pocket is clearly oxidized to a cysteine-sulfenic acid.

Bacillus anthracis apo form model

Crystals of *Bacillus anthracis* malate synthase A (baMSA) grew from ammonium sulfate/PEG 4K in space group C2 with unit cell dimensions of a = 97.8 Å, b = 52.0 Å, c = 89.1 Å; β = 98.9°. Data reduction statistics were reasonable to 1.70 Å resolution. The structure was solved using molecular replacement with the magnesium bound ecMSA model after truncation of non conserved residues. The model was constructed using

ARP/wARP. The final model has a crystallographic R factor of 0.194 (R_{free} = 0.240).

There is one protomer per asymmetric unit containing residues 7-14, 19-325, 345-495,

497-529, two magnesium ions and 180 water molecules. One of the modeled magnesium

ions lies in the active site while the other is involved in a crystal contact. The metal ion

at the crystal contact may actually be a weakly bound Ni^{2+} ion, as addition of $NiCl_2$ to the

crystallization medium promotes crystal growth. Structural comparison with apo ecMSA

using DALI results with a rmsd of 1.1 Å. Residues 326-344 could not be modeled due to

a complete lack of density, and in the crystal cannot adopt the conformation of the

corresponding region of apo ecMSA (residues 329-346) as they would clash with a

symmetry mate.

Comparison of apo and ternary ecMSA structures

Overall comparison. Structural alignment of apo and ternary complex ecMSA

results in rmsd of 0.69 Å for all α-carbons. Upon close inspection, the structures differ

mostly in the C-terminal plug region. Alignment of the TIM barrel and N-terminal clasp

portions results in a rmsd of 0.56 Å with the C-terminal plug having an α-carbon rmsd of

0.71 Å. The differences in the positions of the C-terminal plug are very likely the result

of conformational changes due to the presence of an inhibitor in the ternary complex due

to slight rotations/translations between the domains.

Loop movement near the active site upon substrate/inhibitor binding. Three loops

that cluster around the active site exhibit pronounced conformational changes upon

substrate/inhibitor binding (Figure 3-2). Residues 275-278 form a loop at the pyruvate

binding site which switches from an "open" conformation in the apo structure to a "closed" conformation with bound inhibitor, shielding the bound species from solvent. In the apo form, the backbone atoms of this strand cannot make appropriate interactions with the electrophilic substrate, as seen in Figure 3-2. The "open" conformation is stabilized by a pocket, formed by the backbone carbonyls of residues 298, 299 and 301, in which the guanidine moiety of Arg276 is bound. Upon transition from the "open" to "closed" conformation, the guanidine moiety of Arg276 moves ~7 Å to a new position, where it forms salt bridges with Asp352 and Glu356. As a result, the backbone atoms of Arg276 and Trp277 move about 2 Å, forming interactions with the bound inhibitor. In the closed conformation, Trp277 makes a hydrogen bond to the backbone carbonyl of Met330, which moves 1.8 Å to meet it.

C-terminal plug strand-β-turn-strand movment. The loop emanating from Met330 makes contact with the C-terminal plug strand-β-turn-strand that forms part of the acetyl-CoA binding site (Figure 3-2). Ala332 backbone carbonyl makes a hydrogen bond to Met445, which is in the acetyl-CoA binding site. Similarly, the Ile334 backbone nitrogen contacts the Gly443 carbonyl. It appears that substrate binding induces the movement of the loop containing Met445 and Gly443 through the contacts with residues 330-338 thus widening the entrance to the acetyl binding site. The proposed catalytic base (Asp447) which is located at the end of the strand-β-turn-strand, moves about 1Å away from the electrophilic substrate, possibly making room for the acetyl group as well.
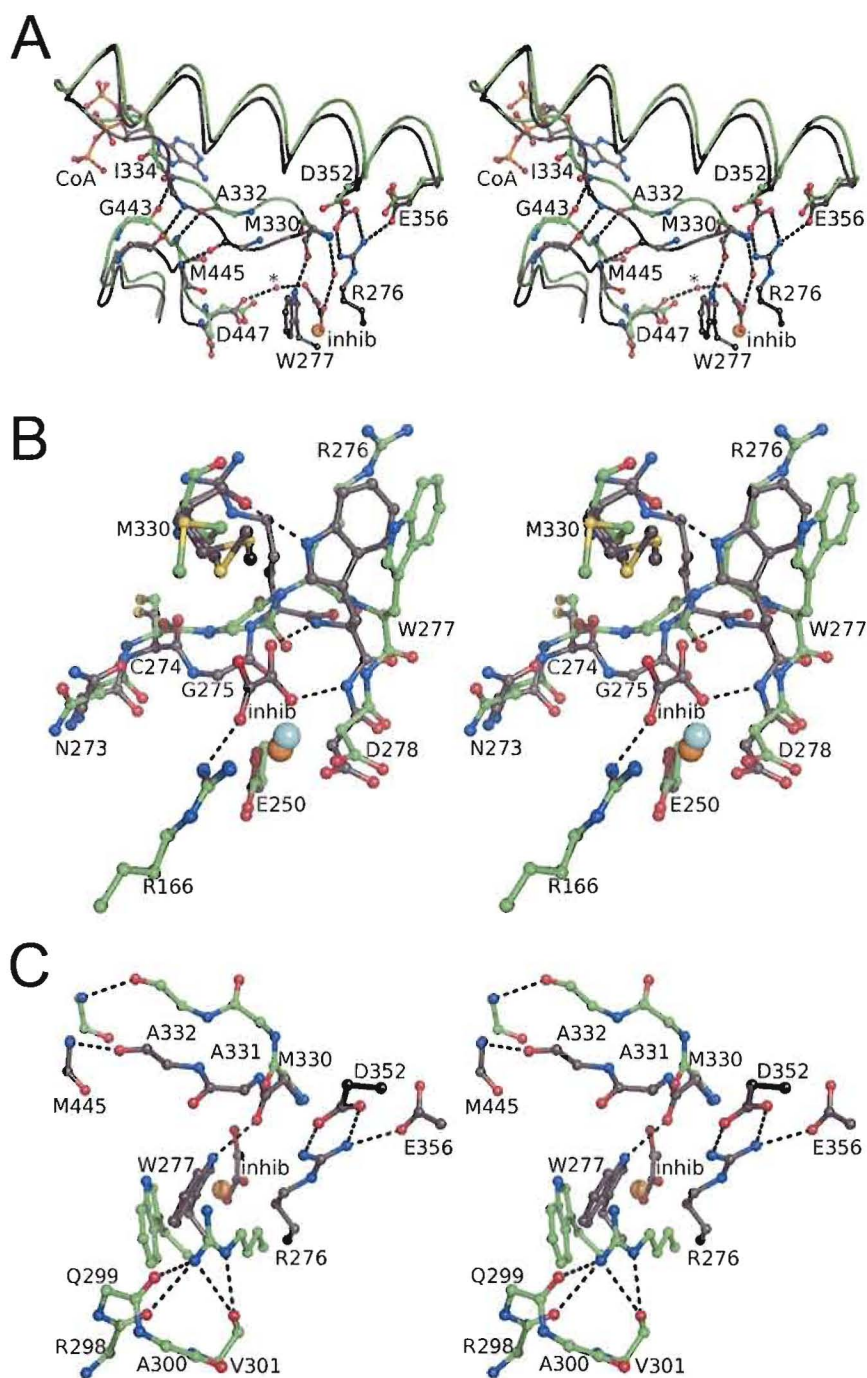
Figure 3-2. Overlay of apo and inhibitor bound ecMSA. Panel A. Overview of the conformational changes described in the text. Inhibitor bound is in black with apo in green. Orange sphere is magnesium. Panel B. Pyruvate binding loop movements. Blue sphere represents calcium in apo structure. Panel C. In the apo structure there is a large opening between the glyoxylate binding loop and the loop containing Met330. Upon inhibitor binding the loop closes up and shields the active site from solvent.

Comparison of ecMSG and ecMSA

Overall structure. The overall fold of MSA is similar to that described previously for MSG. Structural alignment of the MSG (PDB entry 1D8C) and apo ecMSA structures using DALI results in 486 residues aligned with an rmsd of 2.4 Å for α-carbons. The difference in size between MSG and MSA can be attributed largely to the appended α/β-domain, which consists of residues 135-262 and 296-333 in MSG. Residues 135-252 in MSG can be thought of as an insertion between residues 109-110 in MSA and likewise, residues 296-333 form an insertion between residues 158 and 159 in MSA. No function has yet been ascribed to the α/β-domain of MSG yet there are sequence homologies withinin the family that may give clues to its importance (see Discussion). Finally, the N-terminal helical clasp in MSA is shorter by approximately 25 residues, which in MSG forms two short helices.

Catalytic site. The active sites of the ternary structures of ecMSA and ecMSG are very similar. Alignment of 16 active site α-carbons yields an rmsd of 0.32 Å. Notably, residues Cys274, Arg276 and Trp277 in ecMSA (corresponding to Thr451, Phe453 and Leu454 in ecMSG) act to form a gate into the active site. These residues are conserved within each isoform family even though they are different between the isoforms. Interestingly, the active sites of the two isoforms also contain many bound water molecules with conserved positions, as shown in Figure 3-3. The catalytic magnesium is bound by identical residues in both isoforms. Similarly, the catalytic aspartate and arginine are in identical conformations in both isoforms.
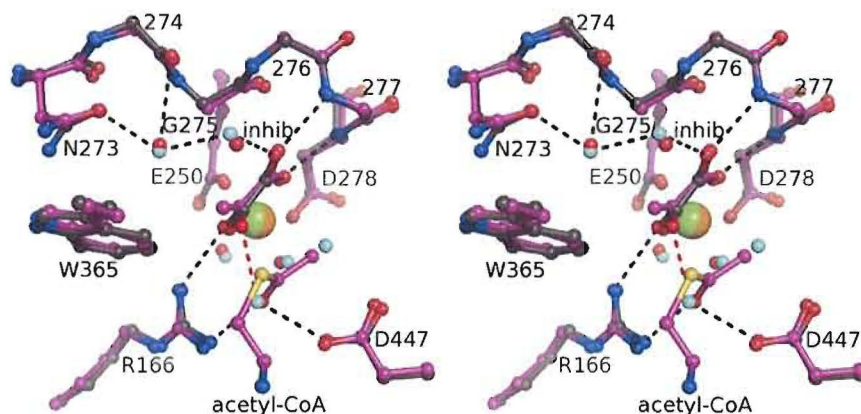
Figure 3-3. Alignment of ecMSA (black) and ecMSG (magenta) with inhibitors bound. Black dashes represent hydrogen bonds and are shown between atoms in ecMSA. The blue spheres represent the waters for ecMSA while the red for ecMSG. The green sphere is magnesium in ecMSG wile the orange sphere is for ecMSA.

Acetyl-CoA binding pocket. It is quite interesting that in both MSG and MSA, coenzyme A derivatives appear to be disordered and/or only fractionally bound (*14, 78, 82*). Acetyl-CoA bound in the ecMSG Anstrom et al. model had B-factors higher than the surrounding residues for the panthothenic acid moiety suggesting flexibility. Similarly, the panthothenic acid moiety of CoA bound in mTB MSG had high B-factors in the models of Smith et al., whereas in the mTB MSG model of Anstrom et al. the CoA occupancy was lowered. In our ecMSA ternary complex the panthothenic acid moiety is not modeled due to weak electron density. Thus, comparison of the binding modes to the two isoforms can only be made for the adenine, ribose and phosphate moities of acetyl-CoA. The defined portion of acetyl-CoA, as modeled in the ecMSA structure, is bound in a very similar manner as seen in ecMSG. An alignment between ecMSA and ecMSG of nine α-carbons in the acetyl-CoA binding site has an rmsd of 0.57 Å. The adenine ring sits on top of a proline in both MS, which caps a TIM barrel α-helix in both isoforms, and is Pro369 and Pro538 for ecMSA and ecMSG respectively. For isoform A, a conserved

methionine at position 102 lies above the ring whereas in ecMSG the residue is Tyr126. The N6 position of the adenine ring is bound by backbone carbonyls in both isoforms. In both isoforms a conserved asparagine, Asn105 in ecMSA, makes a hydrogen bond to the ribose hydroxyl. For MSG the 3'-phosphate is ligated by two arginines whereas in MSA these residues are a lysine and tyrosine. The 5'-phosphate is bound by a histidine in MSA whereas in MSG the diphosphate interacts with the tyrosine capping the adenine ring. One interesting feature is that the conserved ecMSA cys438 in the portion of the acetyl-CoA pocket which interacts with the panthothenic acid moiety is oxidized to a cysteine-sulfenic acid in both ecMSA and ecMSG inhibitor bound structures (*14*).

DISCUSSION

Given that well diffracting crystals of ecMSG and mTB MSG are difficult to obtain, we searched for a MS that would be amenable to structural studies in the hope that this would elucidate the conformational changes upon substrate binding and aid in structure based drug design. *Escherichia coli* malate synthase isoform A was found to crystallize readily and diffract to high resolution, in addition these crystal forms allowed determination of the structure in "open" and "closed" conformations. The "open" conformation reveals two openings to the active site and demonstrates how glyoxylate, inhibitors and malate enter and exit. Inhibitor binding produces the "closed" conformation, for which there is only one opening presumably for acetyl-CoA active site access. The closed conformation for ecMSA is similar to that of MS isoform G and suggests that transition state analogs will inhibit both. The exact structure of the inhibitor

used to obtain the "closed" conformation remains a question, nevertheless this uncertainty does not alter our conclusions concerning conformational changes and the similarities between the active sites of isoforms A and G.

Oxalate binding

The presence of oxalate in the ecMSA ternary structure is unexpected as it was not included in crystallization, but there is anecdotal evidence that as PEG ages it produces oxalate (93-95). Also, Sacchromyces cereviseae MS has a higher affinity for oxalate (K_i = 19 μM) than pryuvate (K_i = 540 μM) (79), which is likely to be similar for ecMSA based on the similarities in pyruvate binding and kinetic constants. Therefore, it is likely that oxalate rather than pyruvate is bound in the reported ecMSA ternary complex active site. The presence in the active site would have consequences such as on the binding of acetyl-CoA. The enolisation reaction of acetyl-CoA by MS has shown to be dependent upon the presence of 2-keto acids such as pyruvate, as without them it takes large quantities of enzyme and high pH to get isotropic exchange with tritiated water. This is taken to mean that without a 2-keto acid inhibitor bound the acetyl group of acetyl-CoA is not bound effectively for proton abstraction. Interestingly it was found that oxalate is deficient in the ability to support the enolization reaction, giving rise to the idea that it prevents the acetyl group from binding.

Weak acetyl-CoA binding

In all malate synthase structures analyzed to date low occupancy and disorder of

bound coenzyme A appears to be a common theme. In the product complex structure of

mTB MSG, the B-factors for CoA are high and CoA is only observed to be bound to only

one of the two proteins in the asymmetric unit (78). A recent mTB MSG products bound

structure also has two protein molecules and only one bound molecule modeled at 70%

occupancy. The suggested explanation is that crystal contacts inhibit binding for mTB

MSG. Full occupancy of acetyl-CoA in the ecMSG structure is modeled, but the B-

factors are much higher than the average protein atoms and may reflect flexibility. The

lack of interpretable electron density for the panthothenic acid moiety of acetyl-CoA in

the ternary ecMSA structure may be due to the presence of oxalate rather than pyruvate

in the active site. For ecMSG with acetyl-CoA and pyruvate bound the pyruvate methyl

group makes hydrophobic interactions with a nearby tryptophan and the terminal sulfur

of acetyl-CoA, and is bound in a planar fashion. A similar conformation of the acetyl-

group and pyruvate would be expected for ecMSA as in ecMSG due to the similar nature

of the active sites. Oxalate binding as described before may prevent the binding of the

acetyl-group of acetyl-CoA, by making hydrogen bonds to the water molecules which are

bound in the place of the acetyl-group. Unable to have the acetyl-group bind the

panthothenic acid tail would be free to adopt many conformations outside the binding

pocket, leading to the weak density seen.

Structural comparisons and loop movement

The loop movements seen for ecMSA are large and affect large portions of the structure setting up "open" and "closed" forms of the enzyme. It appears that these movements are set up to promote the tight binding of glyoxylate before acetyl-CoA. In the "open" conformation the cleft into which the panthothenic acid tail would fit is partially blocked and would obscure binding of the acetyl group. The phospho-adenosine moiety of CoA would still be free to bind in the binding pocket and the tail would be free. The open conformation also allows the free binding of glyoxylate. Closing of the glyoxylate binding loops is presumed to happen contiguously with the rearrangement necessary to open the active site cavity to the acetyl group of acetyl-CoA. It has been shown for *Zea mays* malate synthase (an A isoform) that the affinity for acetyl-CoA is increased almost 50 fold when pyruvate is bound in the active site (*16*). This result suggests that the binding of glyoxylate strongly influences the binding of acetyl-CoA. These structures may demonstrate the conformational changes involved with the increased binding affinity.

The NMR structure of apo ecMSG (*96*) may be in agreement with the crystal structures in the overall trace of the protein backbone but has very different conformations of the side-chains atoms, preventing an accurate comparison to the open conformation of ecMSA. Two important residues in the structure, Phe453 and Leu454, are shown to be disordered in the apo ecMSG NMR structure. It seems likely that Phe453 and Leu454 of ecMSG would undergo a similar conformational change as seen

for Arg276 and Trp277 in ecMSA to open the active site to the binding of glyoxylate and release of malate.

A comparison of the ternary and glyoxylate bound structures of ecMSG revels movements near the acetyl-CoA binding site in particular Met629 which is on the important C-terminal plug strand-β-turn-strand (14). For both isoforms there is a conformational change around this conserved residue (Met445 ecMSA), even though the length of the loop undergoing this movement is different. The movement of the C-terminal plug strand-β-turn-strand opens the active site for acetl-CoA entry. The decreased activity of baker's yeast MS toward malyl-CoA (97), the intermediate in the catalyzed reaction, may be partially due to the lack of accessibility to the active site. Previously the lowered reactivity was suggested to be due to a necessary difficult deprotonation of the 2-hydroxyl of the malyl-CoA prior to binding (13, 98).

Possible α/β-domain function

The function of the α/β-domain of MSG has not been reported, and though some of its residues are close none directly participate in acetyl-CoA binding. The domain may be regulatory or it may serve as an protein-protein interaction scaffold. Dali searches using only the α/β-domain do not return results other than MSG, indicating it is a novel fold. The first thirteen amino acids (ecMSG 134-146) of the domain are conserved between the G isoforms, these residues may be conserved because they are adjacent to the helix which binds the adenine moiety of CoA, or may have another

purpose such as binding regulatory molecules, as there is a cleft that is formed near these

residues.

Inhibition of MS and structural considerations

Substrate analogs have been previously shown to inhibit the catalysis of various

malate synthases. Unfortunately most of these inhibitors are substrates for other enzymes

(glycolate, pyruvate, oxaloacetate or phosphoenolpyruvate) or would harm patients such

as oxalate or fluoroacetate (*16, 78, 79*). One goal of this project is to use structures to

find an effective inhibitor against harmful microbes. Presumed inhibitors designed

against the active site of MSG were analyzed against both MS isoforms and found to be

effective against both isoforms demonstrating that inhibitors designed against one

isoform will work on the other. Further support is the observation that the active sites

with inhibitors/substrates bound are very similar. Since ecMSA crystallizes much more

readily than the MSGs to date or baMSA, it appears that ecMSA is a good choice for

continuing efforts to develop inhibitors based on structural studies. The apo ecMSA

structure also allows the design of inhibitors against the "open" conformation. Initially,

many unsuccessful attempts were made to obtain crystals of ecMSA with substrates or

inhibitors bound by soaking them into crystals of apo ecMSA, the conformational

changes associated with binding of substrate/inhibitors may be inhibited by the crystal.

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

We have successfully introduced large conformational changes into roGFP1 to create the roGFP1-iX family of redox sensors and determined the conformational changes necessary for the catalysis of *Escherichia coli* malate sythase isoform A. Each of these projects has further shown the importance of conformational changes to protein function, especially with regard to two-state conformational changes near active centers.

Previous attempts to raise the redox potential of roGFP had marginal success. The insertion mutation was a rather drastic measure taken to produce a redox biosensor with the thermodynamic properties we needed, and in the end was successful. Insertion of a residue next to the disulfide of roGFP introduced strain into the disulfide, not directly, but through the necessity of the formation of a cis-peptide bond. The formation of the cis-peptide was unexpected, as was the large distance between the cysteines in the reduced state. It is unlikely that if such a structure were seen originally for GFP one would attempt to make a disulfide bond where it currently exists in roGFP1-iX. It appears that the inherent flexibility of the 147 strand containing the disulfide is necessary for the formation of the disulfide bond. This begs the question, what effect will prolines around the disulfide have on the redox potential? The anomalous redox potential of

roGFP1-iR compared to roGFP1-iE was explained by specific interactions in the case of roGFP-iR. Will the redox potential change in the expected direction if the specific interactions are removed? Other questions remain as well. The most important facing the roGFP1-iX indicators is whether they will work in vivo and if so will the be oxidizing enough to measure the thiol/disulfide equilibrium of the endoplasmic reticulum? If not useful in the endoplasmic reticulum, the structures of roGFP-iX can be used to rationally explore additional mutations to develop more oxidizing mutants. Incorporation of Mark Cannon's positive mutations (44) might lead to increases in redox potential and would be beneficial just for the rate enhancement. The information gained from the studies of roGFP1-iX will both allow for the further design of conformational changes in redox biosensors as well as in other biosensors using similar strategies. In the end, this study reveals that conformational changes are a very important part of determining the thiol/disulfide equilibrium in proteins.

Although the goal of the work on malate synthase is to produce drugs to combat microbial infections, elucidation of the structural changes of substrate binding gives insight into the development of such drugs. Parabanic acid was shown to be an inhibitor of malate synthase at micromolar concentrations. In theory, this inhibitor is easily functionalized. Since crystals of apo ecMSA crack when exposed to the substrate glyoxylate, one could functionalize parabanic acid with many different groups and test for tight binding to the closed state using the cracking of crystals as a screening procedure. If this procedure works, the usefulness of conformational changes would be demonstrated, not only from a cellular point of view, but also from one of our own.

BIBLIOGRAPHY

(1)   Wess, J., and Wx. (1997) G-protein-coupled receptors: Molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. *Faseb J. 11*, 346-354.

(2)   Choi, H. J., Kim, S. J., Mukhopadhyay, P., Cho, S., Woo, J. R., Storz, G., Ryu, S. E., and Kj. (2001) Structural basis of the redox switch in the OxyR transcription factor. *Cell 105*, 103-113.

(3)   Zheng, M., Aslund, F., Storz, G., and Zb. (1998) Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science 279*, 1718-1721.

(4)   Abrahams, J. P., Leslie, A. G. W., Lutter, R., Walker, J. E., and Pd. (1994) Structure at 2.8-angstrom resolution of F1-ATPase from bovine heart-mitochondria. *Nature 370*, 621-628.

(5)   Straub, F. B., and Feuer, G. (1950) Adenosine triphosphate, the functional group of actin. *Biochim. Biophys. Acta 4*, 455-70.

(6)   Perry, S. V., and Cotterill, J. (1965) Interaction of actin and myosin. *Nature 206*, 161-3.

(7)   Fischer, E. (1894) Influence of configuration on the action of enzymes. *Berichte der Deutschen Chemischen Gesellschaft 27*, 3479-83.

(8)   Koshland, D. E., Jr. (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A. 44*, 98-105.

(9)   Hofmann, E., and Bt. (1976) Significance of phosphofructokinase to regulation of carbohydrate-metabolism. *Rev. Physiol., Biochem. Pharmacol. 75*, 1-68.

(10)   Wong, D. T. O., and Ajl, S. J. (1956) Conversion of acetate and glyoxalate to malate. *J. Am. Chem. Soc. 78*, 3230-1.

(11)   Kornberg, H. L., and Krebs, H. A. (1957) Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature 179*, 988-91.

(12)   Wiegand, G., and Remington, S. J. (1986) Citrate synthase - structure, control and mechanism. *Annual Review of Biophysics and Biophysical Chemistry 15*, 97-117.

(13)   Howard, B. R., Endrizzi, J. A., and Remington, S. J. (2000) Crystal structure of Escherichia coli malate synthase G complexed with magnesium and glyoxylate at 2.0 angstrom resolution: Mechanistic implications. *Biochemistry 39*, 3156-3168.

(14)   Anstrom, D. M., Kallio, K., and Remington, S. J. (2003) Structure of the Escherichia coli malate synthase G: pyruvate: acetyl-coenzyme A abortive ternary complex at 1.95 angstrom resolution. *Protein Science 12*, 1822-1832.

(15)   Zipper, P., and Durchschlag, H. (1977) Small-angle X-ray studies on malate synthase from bakers-yeast. *Biochem. Biophys. Res. Commun. 75*, 394-400.

(16)   Beeckmans, S., Khan, A. S., Kanarek, L., and Vandriessche, E. (1994) Ligand-binding on to maize (Zea-Mays) malate mynthase - A structural study. *Biochemical Journal 303*, 413-421.

(17)   Schmid, G., Durchschlag, H., Biedermann, G., Eggerer, H., and Jaenicke, R. (1974) Molecular structure of malate synthase and structural changes upon ligand binding to the enzyme. *Biochem. Biophys. Res. Commun. 58*, 419-26.

(18)   Tugarinov, V., Muhandiram, R., Ayed, A., and Kay, L. E. (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: Chemical shift assignments and secondary structure of malate synthase G. *J. Am. Chem. Soc. 124*, 10025-10035.

(19)   Evans, R. M. (1988) The steroid and thyroid-hormone receptor superfamily. *Science 240*, 889-895.

(20)   Ikura, M., Hiraoki, T., Hikichi, K., Mikuni, T., Yazawa, M., Yagi, K., and Qp. (1983) Nuclear magnetic-resonance studies on calmodulin - Calcium-induced conformational change. *Biochemistry 22*, 2573-2579.

(21)   Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem. 67*, 509-544.

(22)   Romoser, V. A., Hinkle, P. M., Persechini, A., and Wz. (1997) Detection in living cells of Ca2+-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence - A new class of fluorescent indicators. *J. Biol. Chem. 272*, 13270-13274.

(23)   Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., Tsien, R. Y., and Xt. (1997) Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature 388*, 882-887.

(24)   Yano, T., Akeyama, N., Yurimoto, H., Sakai, Y., and Lr. (2006) Development of a fluorescent probe for oxidative stress in a living cell using a Yap 1-FRET system and its application to autophagic studies. *Autophagy 2*, 351-351.

(25)   Robin, E., Guzy, R. D., Loor, G., Iwase, H., Waypa, G. B., Marks, J. D., Hoek, T. L. V., and Schumacker, P. T. (2007) Oxidant stress during simulated ischemia primes cardiomyocytes for cell death during reperfusion. *J. Biol. Chem. 282*, 19133-19143.

(26)   Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A. 96*, 11241-11246.

(27)   Belousov, V. V., Fradkov, A. F., Lukyanov, K. A., Staroverov, D. B., Shakhbazov, K. S., Terskikh, A. V., and Lukyanov, S. (2006) Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. *Nature Methods 3*, 281-286.

(28)   Hanson, G. T., McAnaney, T. B., Park, E. S., Rendell, M. E. P., Yarbrough, D. K., Chu, S. Y., Xi, L. X., Boxer, S. G., Montrose, M. H., and Remington, S. J. (2002) Green fluorescent protein variants as ratiometric dual emission pH sensors. 1. Structural characterization and preliminary application. *Biochemistry 41*, 15477-15488.

(29)   Hanson, G. T., Aggeler, R., Oglesbee, D., Cannon, M., Capaldi, R. A., Tsien, R. Y., and Remington, S. J. (2004) Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. *J. Biol. Chem. 279*, 13044-13053.

(30)   Ostergaard, H., Henriksen, A., Hansen, F. G., and Winther, J. R. (2001) Shedding light on disulfide bond formation: engineering a redox switch in green fluorescent protein. *EMBO J. 20*, 5853-5862.

(31)   Schafer, F. Q., and Buettner, G. R. (2001) Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radical Biol. Med. 30*, 1191-1212.

(32)   Dooley, C. T., Dore, T. M., Hanson, G. T., Jackson, W. C., Remington, S. J., and Tsien, R. Y. (2004) Imaging dynamic redox changes in mammalian cells with green fluorescent protein indicators. *J. Biol. Chem. 279*, 22284-22293.

(33)   Jiang, K., Schwarzer, C., Lally, E., Zhang, S. B., Ruzin, S., Machen, T., Remington, S. J., and Feldman, L. (2006) Expression and characterization of a redox-sensing green fluorescent protein (reduction-oxidation-sensitive green fluorescent protein) in Arabidopsis. *Plant Physiology 141*, 397-403.

(34)   Ostergaard, H., Tachibana, C., and Winther, J. R. (2004) Monitoring disulfide bond formation in the eukaryotic cytosol. *J. Cell Biol. 166*, 337-345.

(35)   Hwang, C., Sinskey, A. J., and Lodish, H. F. (1992) Oxidized redox state of glutathione in the endoplasmic-reticulum. *Science 257*, 1496-1502.

(36)   Schwarzer, C., Illek, B., Suh, J. H., Remington, S. J., Fischer, H., and Machen, T. E. (2007) Organelle redox of CF and CFTR-corrected airway epithelia. *Free Radical Biol. Med. 43*, 300-316.

(37)   Austin, C. D., Wen, X. H., Gazzard, L., Nelson, C., Scheller, R. H., and Scales, S. J. (2005) Oxidizing potential of endosomes and lysosomes limits intracellular cleavage of disulfide-based antibody-drug conjugates. *Proc. Natl. Acad. Sci. U.S A. 102*, 17987-17992.

(38)   Krause, G., Lundstrom, J., Barea, J. L., Delacuesta, C. P., and Holmgren, A. (1991) Mimicking the active-site of protein disulfide-isomerase by substitution of proline 34 in *Escherichia coli* thioredoxin. *J. Biol. Chem. 266*, 9494-9500.

(39)   Wunderlich, M., and Glockshuber, R. (1993) Redox properties of protein disulfide isomerase (Dsba) from *Escherichia coli. Protein Science 2*, 717-726.

(40)   Chivers, P. T., Prehoda, K. E., and Raines, R. T. (1997) The CXXC motif: A rheostat in the active site. *Biochemistry 36*, 4061-4066.

(41)   Szajewski, R. P., and Whitesides, G. M. (1980) Rate constants and equilibrium-constants for thiol-disulfide interchange reactions involving oxidized glutathione. *J. Am. Chem. Soc. 102*, 2011-2026.

(42)   Lees, W. J., and Whitesides, G. M. (1993) Equilibrium-constants for thiol disulfide interchange reactions - A coherent, corrected set. *J. Org. Chem. 58*, 642-647.

(43)   Gilbert, H. F. (1990) Molecular and cellular aspects of thiol disulfide exchange. *Advances In Enzymology And Related Areas Of Molecular Biology 63*, 69-172.

(44)   Cannon, M. B., and Remington, S. J. (2006) Re-engineering redox-sensitive green fluorescent protein for improved response rate. *Protein Science 15*, 45-57.

(45)   Hansen, R. E., Ostergaard, H., and Winther, J. R. (2005) Increasing the reactivity of an artificial dithiol-disulfide pair through modification of the electrostatic milieu. *Biochemistry 44*, 5899-5906.

(46) Burns, J. A., and Whitesides, G. M. (1990) Predicting the stability of cyclic disulfides by molecular modeling - Effective concentrations in thiol-disulfide interchange and the design of strongly reducing dithiols. *J. Am. Chem. Soc. 112*, 6296-6303.

(47) Houk, J., and Whitesides, G. M. (1987) Structure reactivity relations for thiol disulfide interchange. *J. Am. Chem. Soc. 109*, 6825-6836.

(48) Zhang, R. M., and Snyder, G. H. (1989) Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of interviening amino acids. *J. Biol. Chem. 264*, 18472-18479.

(49) Pace, C. N., Grimsley, G. R., Thomson, J. A., and Barnett, B. J. (1988) Conformational stability and activity of ribonuclease-T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem. 263*, 11820-11825.

(50) Zhang, R. M., and Snyder, G. H. (1988) Kinetics of disulfide exchange-reactions of monomer and dimer loops of cysteine valine cysteine peptides. *Biochemistry 27*, 3785-3794.

(51) Woycechowsky, K. J., and Raines, R. T. (2003) The CXC motif: A functional mimic of protein disulfide isomerase. *Biochemistry 42*, 5387-5394.

(52) Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology 24*, 1170-1170.

(53) Siemering, K. R., Golbik, R., Sever, R., and Haseloff, J. (1996) Mutations that suppress the thermosensitivity of green fluorescent protein. *Current Biology 6*, 1653-1663.

(54) Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R., and Piston, D. W. (1997) Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical Journal 73*, 2782-2790.

(55) Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307-326.

(56) Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol. 374*, 229-244.

(57) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect D: Biol. Crystallogr. 53*, 240-255.

(58) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect D: Biol. Crystallogr. 60*, 2126-2132.

(59) Bokman, S. H., Ward, W. W., and Mf. (1981) Renaturation of *Aequorea* green-fluorescent protein. *Biochem. Biophys. Res. Commun. 101*, 1372-1380.

(60) Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., Reeves, S. C., and Nv. (1982) Spectral perturbations of the *Aequorea* green-fluorescent protein. *Photochemistry and Photobiology 35*, 803-808.

(61) Srinivasan, N., Sowdhamini, R., Ramakrishnan, C., and Balaram, P. (1990) Conformations of disulfide bridges in proteins. *International Journal of Peptide and Protein Research 36*, 147-155.

(62) Jabs, A., Weiss, M. S., and Hilgenfeld, R. (1999) Non-proline cis peptide bonds in proteins. *J. Mol. Biol. 286*, 291-304.

(63) Quillin, M. L., Anstrom, D. A., Shu, X. K., O'Leary, S., Kallio, K., Chudakov, D. A., and Remington, S. J. (2005) Kindling fluorescent protein from *Anemonia sulcata*: Dark-state structure at 1.38 angstrom resolution. *Biochemistry 44*, 5774-5787.

(64) Schmidt, B., Ho, L., Hogg, P. J., and Fi. (2006) Allosteric disulfide bonds. *Biochemistry 45*, 7429-7433.

(65) Vetter, I. R., Baase, W. A., Heinz, D. W., Xiong, J. P., Snow, S., Matthews, B. W., and Wa. (1996) Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme. *Protein Science 5*, 2399-2415.

(66) Bouckaert, J., Dewallef, Y., Poortmans, F., Wyns, L., Loris, R., and Au. (2000) The structural features of concanavalin A governing non-proline peptide isomerization. *J. Biol. Chem. 275*, 19778-19787.

(67) Huber-Wunderlich, M., and Glockshuber, R. (1998) A single dipeptide sequence modulates the redox properties of a whole enzyme family. *Folding & Design 3*, 161-171.

(68) Mossner, E., Huber-Wunderlich, M., and Glockshuber, R. (1998) Characterization of *Escherichia coli* thioredoxin variants mimicking the active-sites of other thiol/disulfide oxidoreductases. *Protein Science 7*, 1233-1244.

(69) Wu, M. M., Grabe, M., Adams, S., Tsien, R. Y., Moore, H. P. H., and Machen, T. E. (2001) Mechanisms of pH regulation in the regulated secretory pathway. *J. Biol. Chem. 276*, 33027-33035.

(70)   Munoz-Elias, E. J., and McKinney, J. D. (2005) *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nature Medicine 11*, 638-644.

(71)   McKinney, J. D., zu Bentrup, K. H., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., and Russell, D. G. (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature 406*, 735-738.

(72)   Honer Zu Bentrup, K., Miczak, A., Swenson, D. L., and Russell, D. G. (1999) Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *J. Bacteriol. 181*, 7161-7167.

(73)   Lorenz, M. C., Bender, J. A., and Fink, G. R. (2004) Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryotic Cell 3*, 1076-1087.

(74)   Ramirez, M. A., and Lorenz, M. C. (2007) Mutations in alternative carbon utilization pathways in *Candida albicans* attenuate virulence and confer pleiotropic phenotypes. *Eukaryotic Cell 6*, 280-290.

(75)   Lorenz, M. C., and Fink, G. R. (2001) The glyoxylate cycle is required for fungal virulence. *Nature 412*, 83-86.

(76)   Smith, C. V., Sharma, V., and Sacchettini, J. C. (2004) TB drug discovery: addressing issues of persistence and resistance. *Tuberculosis 84*, 45-55.

(77)   Falmagne, P., Vanderwinkel, E., and Wiame, J. M. (1965) Demonstration of two malic synthases in *Escherichia coli*. *Biochim. Biophys. Acta 99*, 246-58.

(78)   Smith, C. V., Huang, C. C., Miczak, A., Russell, D. G., Sacchettini, J. C., and Bentrup, K. H. Z. (2003) Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis*. *J. Biol. Chem. 278*, 1735-1743.

(79)   Durchschlag, H., Biedermann, G., and Eggerer, H. (1981) Large-scale purification and some properties of malate synthase from bakers-yeast. *Eur. J. Biochem. 114*, 255-262.

(80)   Molina, I., Pellicer, M. T., Badia, J., Aguilar, J., and Baldoma, L. (1994) Molecular characterization of *Escherichia coli* malate synthase G - Differentiation with the malate synthase A isoenzyme. *Eur. J. Biochem. 224*, 541-548.

(81)   Vanderwinkel, E., and De Vlieghere, M. (1968) Physiology and genetics of isocitritase and malate synthase in *Escherichia coli*. *Eur. J. Biochem. 5*, 81-90.

(82)   Anstrom, D. M., and Remington, S. J. (2006) The product complex of M-tuberculosis malate synthase revisited. *Protein Science 15*, 2002-2007.

(83)   Anstrom, D. M., Colip, L., Moshofsky, B., Hatcher, E., and Remington, S. J. (2005) Systematic replacement of lysine with glutamine and alanine in *Escherichia coli* malate synthase G: effect on crystallization. *Acta Crystallographica Section F 61*, 1069-1074.

(84)   Tugarinov, V., Choy, W. Y., Orekhov, V. Y., and Kay, L. E. (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. U.S.A. 102*, 622-627.

(85)   Tugarinov, V., and Kay, L. E. (2005) Quantitative C-13 and H-2 NMR relaxation studies of the 723-residue enzyme malate synthase g reveal a dynamic binding Interface. *Biochemistry 44*, 15970-15977.

(86)   Gill, S. C., and Vonhippel, P. H. (1989) Calculation of protein extinction coefficients from amino-acid sequence data. *Analytical Biochemistry 182*, 319-326.

(87)   Sun, P. D., Radaev, S., and Kattah, M. (2002) Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallogr., Sect D: Biol. Crystallogr. 58*, 1092-1098.

(88)   Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr., Sect D: Biol. Crystallogr. 58*, 1772-1779.

(89)   Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. *Zeitschrift Fur Kristallographie 217*, 644-650.

(90)   Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr. 30*, 1022-1025.

(91)   Nakata, M., and Selitrennikoff, C. P. (2002) A method to assay glyoxylate cycle inhibitors for antifungals. *J. Antibiot. 55*, 602-604.

(92)   Harding, M. (2006) Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr., Sect D: Biol. Crystallogr. 62*, 678-682.

(93)   Kryuk, T. V., Mikhal'chuk, V. M., Petrenko, L. V., Nelepova, O. A., and Nikolaevskii, A. N. (2002) Promising inhibitors of poly(ethylene glycol) oxidation in aqueous solutions. *Pharm. Chem. J. 36*, 32.

(94)   Ellis, G. (2006) Degradation of crystalline polyethylene glycol 6000 and its effect on assays for macroprolactin and other analytes. *Clin. Biochem. 39*, 1035-1040.

(95)   Uekawa, N., Endo, M., Kakegawa, K., and Sasaki, Y. (2000) Homogeneous
        precipitation of Cr3+-M2+ (M = Ni, Zn, Co, Cu) oxalate by oxidation of the
        polyethylene glycol-cation complex. *Phys. Chem. Chem. Phys. 2*, 5485-5490.

(96)   Tugarinov, V., and Kay, L. E. (2003) Quantitative NMR studies of high molecular
        weight proteins: Application to domain orientation and ligand binding in the 723
        residue enzyme malate synthase G. *J. Mol. Biol. 327*, 1121-1133.

(97)   Eggerer, H., and Klette, A. (1967) Über das katalyseprinzip der malat-synthase.
        Eur. J. Biochem. 1, 447-475.

(98)   Lill, U., Kollmannkoch, A., Bibinger, A., and Eggerer, H. (1991) Inhibitors of
        metabolic reactions - Scope and limitation of acyl-CoA-analog CoA-thioethers.
        Eur. J. Biochem. 198, 767-773.