UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Structural Insights on E. coli Phosphoenolpyruvate Carboxykinase structures, conformations, and ligand specificity from Integrated X-ray Methods

Permalink

https://escholarship.org/uc/item/5cs1f070

Author

Tang, Henry Yue Hin

Publication Date

2017

Peer reviewed|Thesis/dissertation

Structural Insights on *E. coli* Phosphoenolpyruvate Carboxykinase
structures, conformations, and ligand specificity from Integrated X-ray Methods

By

Henry Yue Hin Tang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John A. Tainer, Co-chair
Professor Jamie H. D. Cate, Co-chair
Professor David E. Wemmer
Professor Andreas Martin

Summer 2017

Structural Insights on *E. coli* Phosphoenolpyruvate Carboxykinase
structures, conformations, and ligand specificity from Integrated X-ray Methods

Copyright 2017

by

Henry Yue Hin Tang

Abstract

Structural Insights on *E. coli* Phosphoenolpyruvate Carboxykinase
structures, conformations, and ligand specificity from Integrated X-ray Methods

by

Henry Yue Hin Tang

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor John A. Tainer, Co-chair
Professor Jamie H. D. Cate, Co-chair

Phosphoenolpyruvate carboxykinase (PCK) is a key metabolic enzyme responsible for catalyzing the first committed step of gluconeogenesis, the conversion of oxaloacetate to phosphoenolpyruvate and carbon dioxide, in all living organisms. PCK has been implicated in diabetes mellitus along with tumor growth under glucose-depleted conditions. Despite its critical role in metabolism and health, certain key aspects of this enzyme remain uncharacterized. The work in this dissertation focuses on understanding the relationship of structures, conformations, and ligand specificity for the *E. coli* isozyme of PCK through structural X-ray methods.

Because PCK directly utilizes CO_2, the CO_2 binding pocket can be engineered to accommodate different substrates to enzymatically form carbon-carbon bonds. Chapter 2 details results for six crystal structures with mutations in or near the CO_2 binding site that lead to binding of the nonnative ligands thiosulfate and methanesulfonate as defined by macromolecular crystallography. Even though these ligands both contain a sulfonate moiety, mutants are found to bind only one or the other, proving to dissect aspects of ligand recognition. The orientation of these ligands also are altered by the hydrogen bonding network present in the pocket of interest. Through computational simulations, these phenomena are rationalized through the displacement of loosely bound water molecules in the binding pocket.

Conformational changes are crucial to the catalytic cycle of many enzymes, and PCK is no exception. Chapter 3 discusses my efforts to better understand how PCK undergoes conformational changes in different circumstances. First, insertion mutants in new intermediate crystallographic conformations were characterized and a metric to measure the closedness of the enzyme was developed. Crystallographic models provide accurate atomic information but are not comprehensive for conformational states, so I go beyond crystallographic models to examine the solution state conformations of PCK. This more accurately reflects biological conditions when compared to models obtained from crystal structures because the solution state is free of influences such as crystal packing and the experimental conditions are more realistic relative to those used to grow crystals. To examine solution state conformations, small angle X-ray scattering (SAXS) data were collected on EcPCK and 10 of its mutants under various reaction conditions resulting in 33 experimental profiles. Interestingly, the apo state in solution did not correspond to its crystal structure; instead, it adopted a more closed conformation. Furthermore,

the ATP-bound state of EcPCK has only been found in the cap-open state, where the cap is a flexible loop that closes over the active site cleft, while the SAXS data supports a closed-cap conformation. Given these different results in solution and the value of comparing them comprehensively to crystal structures, I present a novel way of visualizing SAXS data in high-throughput, using PCK as the model system.

Together, the results and concepts from this work provide insight into a key region of the EcPCK active site cleft and to the behavior of the system as a whole in solution. These data and ideas further our understanding of this enzyme, and to pave the way toward characterizing enzymes in their solution states as they undergo changes during catalysis.

Table of Contents

List of figures

List of tables

List of abbreviations and symbols

AMBER	Assisted model building with energy refinement
ASU	Asymmetric unit
ATP	Adenosine-5'-triphosphate
C-C bond	Carbon-carbon bond
CC	Correlation coefficient
CPCC	Closed protein active site cleft, closed cap
CPOC	Closed protein active site cleft, open cap
E. coli	*Escherichia coli*
EcPCK	*E. coli* phosphoenolpyruvate carboxykinase
FRET	Förster resonance energy transfer
GAFF	Generalized AMBER force field
GTP	Guanosine-5'-triphosphate
kDa	Kilodalton
LB	Lysogeny broth
MD	Molecular dynamics
MM-GB/SA	Molecular mechanics – Generalized Born/Surface Area
MnSOD	Manganese superoxide dismutase
NADH	Nicotinamide adenine dinucleotide (reduced)
NDP	Nucleoside diphosphate
Ni-NTA	Nickel-nitrilotriacetic acid
NTP	Nucleoside triphosphate
OAA	Oxaloacetate
OPCC	Open protein active site cleft, closed cap
OPOC	Open protein active site cleft, open cap
OWAB	Occupancy weighted average B-factor
PCK	Phosphoenolpyruvate carboxykinase
PCK1	Phosphoenolpyruvate carboxykinase (cytosolic)
PCK2	Phosphoenolpyruvate carboxykinase (mitochondrial)
PDB	Protein Data Bank
PEG	Polyethylene glycol
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PEPCK	Phosphoenolpyruvate carboxykinase
PEPCK-C	Phosphoenolpyruvate carboxykinase (cytosolic)
PEPCK-M	Phosphoenolpyruvate carboxykinase (mitochondrial)
P(r)	Electron pair distance distribution function
RESP	Restrained electrostatic potential
R_g	Radius of gyration
RMS	Root mean square
RSCC	Real space correlation coefficient
RSR	Real space R value
RSZO	Real space observed density z-score

SAXS	Small angle X-ray scattering
shRNA	Short hairpin RNA
SIBYLS	Structurally Integrated BiologY for the Life Sciences (beamline)
siRNA	Small interfering RNA
T. cruzi	*Trypanosoma cruzi*
THJ	Thiosulfate
V_r	Volatility of ratio
WAT	Water
WT	Wild type

Acknowledgements

I would like to first thank my advisor, Professor John Tainer, for his invaluable support, mentorship, and guidance over the years. It has been an honor to work with John and to delve into the field of structural biology with one of the top experts in this area. I will forever be grateful for the opportunity to work in the Tainer lab and the SIBYLS beamline.

Secondly, I thank my dissertation committee, Professor Cate, Professor Wemmer, and Professor Martin for their contributions in shaping this work and helping me develop my skills as a scientist.

Next, I am indebted to many of my colleagues for their support. I would like to thank Greg Hura and Steve Yannone for their scientific mentorship, which was vital to the success of this work. The Tainer Group would not be the same without Jane Tanamachi, who has been a great resource in all aspects of graduate school life. I thank the SIBYLS beamline scientists Dr. Greg Hura, Dr. Michal Hammel, and Dr. Scott Classen for their key roles in synchrotron data collection, which provided the technical basis for all the work presented here. I thank Dr. David Shin for his role in teaching me macromolecular crystallography. I thank Dr. Tadeusz Ogorzalek and Dr. Soumya Govinda Remesh for their stimulating scientific conversations and support in lab. I thank Kevin Dyer and Kathryn Burnett for their assistance with SAXS data collection. Finally, I thank Camille Schwartz and Monica Kuzdovich for the roles they played in maintaining the wet lab.

I have been very fortunate to be able to garner support from the friends I have made over the years in Berkeley. Thank you, Anna Susa, Ben Cotts, Connor Bischak, Jenny Lin, Luke Latimer, Ben Kriegel, Andrew Wong, and Jonathan Thirman, for the dinners at Imm Thai and other restaurants, hangouts, and other miscellaneous shenanigans.

Finally, none of this would have been possible without the support of my parents and brother. Thank you for supporting my endeavors, no matter where they may lead me.

1 Introduction

1.1 EcPCK Background

Phosphoenolpyruvate carboxykinase (abbreviated PCK or PEPCK) is a key enzyme in the gluconeogenesis pathway. PCK is the enzyme that catalyzes the first committed step of gluconeogenesis: the conversion of oxaloacetate (OAA) into phosphoenolpyruvate (PEP). The enzyme was first isolated from chicken liver (1) and characterized (2, 3) by Utter and Kurahashi in 1954, when it was first termed oxaloacetic carboxylase. Since then, countless studies have been performed on this system, with 80 crystal structures from various organisms complexed with different ligands deposited in the PDB as of time of writing. These structural and biochemical results provide the foundation for the work presented here.

1.1.1 The significance of phosphoenolpyruvate carboxykinase in gluconeogenesis

Gluconeogenesis is the pathway by which glucose is generated in all living organisms, from animals to fungi to microorganisms. An organism can then take the glucose resulting from gluconeogenesis and use it for a myriad of purposes, such as storing the glucose as glycogen for future energy needs, or constructing cell walls out of this monomeric building block. As glucose has maintained its role as a primary energy source throughout the course of evolution, the gluconeogenesis pathway is an essential pathway for all of life.

Figure 1.1 shows PCK in the context of the gluconeogenesis and glycolysis metabolic pathways. Although PEP can be converted into pyruvate during glycolysis in only one step with pyruvate kinase, the reverse reaction cannot happen intracellularly as it is energetically infeasible. Also, to maintain metabolic regulation, nature uses a second set of enzymes to catalyze a reverse reaction. Thus, a two-step process involving different enzymes is required: first, the conversion of pyruvate into OAA by pyruvate carboxylase, followed by the conversion of OAA to PEP by PCK.

1.1.2 The reaction mechanism of PCK

The actual mechanism behind the PCK reaction takes place in two steps: the decarboxylation of OAA to form the enolate of pyruvate intermediate, followed by phosphotransfer of the γ-phosphate from NTP (either ATP or GTP, depending on the organism and PCK isozyme) onto the α-carbon of the enolate of pyruvate. Figure 1.2 illustrates the electron-pushing mechanism just described. Since one high-energy phosphate compound PEP is generated in the reaction while another (NTP) is hydrolyzed, the free energy change of the reaction is close to 0 and the reaction is reversible under cellular conditions; regulation is thus controlled by the location of the various substrates. The mechanism for the GTP-form of PCK has the same mechanism, substituting the ATP for a GTP.

Because the reverse reaction can readily occur in vitro (1), it means that the enzyme is capable of forming carbon-carbon bonds. Reversing the mechanism, PEP first transfers its phosphate group onto NDP, generating the enolate of pyruvate intermediate. Then the enolate intermediate attacks the central carbon of CO_2, forming OAA and a new C-C bond in the process. Since C-C bond formation in the realm of organic chemistry is still a challenging area, and this enzyme can form C-C bonds, it is of academic and possible practical interest for enzyme engineering purposes.

In order for PCK to be catalytically active, it requires divalent cation cofactors. Mg^{2+} is absolutely critical for the proper functioning of this enzyme, while optimal catalytic rates are only achieved when Mn^{2+} is also present. Work involving the possible role of Ca^{2+} in catalysis and as an allosteric inhibitor of PCK has been performed (4, 5) but the inhibitory effect was never discovered as no allosteric Ca^{2+} binding site had been discovered in subsequent work by the same group (6); that work will be discussed in Section 1.2.1.

There is obvious conformational change during catalysis as inferred from crystal structures of the apo state versus that of the $ATP/Mg^{2+}/Mn^{2+}$ holoenzyme, described in more detail in Section 1.1.3, so it is possible that conformational change has a role to play in catalysis, for example, bringing catalytically important sidechains into position, or changing the shape of a substrate binding pocket to eject the product upon the completion of catalysis. The structural mechanistic underpinnings of the chemical reaction mechanism at this point in time, especially in solution, are still enigmatic and will be further explored in Chapter 3 of this work.

1.1.3 Structural overview of EcPCK

The following section describes some key aspects of the *E. coli* PCK (EcPCK) structure; Figure 1.3 provides a visualization of the EcPCK structure in its open and closed states to aid in seeing the description presented. EcPCK is a 60 kDa protein consisting of 540 amino acids. Present in solution as a monomer, it has a bilobate architecture, consisting of an N-terminal and C-terminal domain. Threading the sequence through the two lobes, the main chain will cross between the two domains three times. The active site is located in between the two lobes, and will be referred to as the active cleft. The cleft consists of distinct regions for binding the various substrates: ATP, along with OAA, or PEP and CO_2. The two divalent cations are involved in ATP binding: Mg^{2+} is located between the β- and γ-phosphate of ATP and coordinates to Thr255 of the protein, while Mn^{2+} coordinates to the γ-phosphate of ATP, Lys213, His232, and Asp269. In the GTP-binding isozymes of PCK, the Mn^{2+} coordination sphere has been implicated in OAA/PEP substrate binding as well, but that has not been observed in EcPCK. There is also a pi-pi stacking interaction between the adenine of ATP and Arg449. CO_2 binding is through hydrogen bonding mediated through two residues: Arg65 and Tyr 207 (7).

Upon substrate binding, there are two major changes in the structure that occur: the two lobes of the enzyme come together in a clamshell-like manner, bending at the hinge, and a loop termed the cap or Ω-loop above the active site (residues 391-400 for the loop itself, residues 386-406 when including the bases of the loop) closes over the cleft. The conformational changes have been studied through crystallographic models complexed with various small molecules and are described in Sections 1.2.1 and 1.2.2, but no solution state structural experiments have been

performed to our knowledge. The cap has been determined to be crucial in catalysis through mutation experiments in the GTP isozyme (8, 9). Additional details regarding past experiments of EcPCK and other members of the PCK family are discussed in Section 1.2.

1.1.4 Significance in health

Due to its role in gluconeogenesis, there was a lot of interest initially into the role of PCK in diabetes mellitus (10). Inhibition of this enzyme should lead to decreased blood glucose levels since one source of glucose has been cut off. However, at the time of writing, no treatments of diabetes have been found through inhibition of PCK. Delving into more recent literature, the idea that PCK is the sole enzyme for regulating the gluconeogenesis pathway became less likely. For one, this goes against metabolic control theory (11). A study by Burgess et al. (12) showed that in mouse livers with 90% decreased PCK content versus normal mouse livers, gluconeogenic flux only decreased by 40%, suggesting that PCK in liver only has a weak effect on gluconeogenesis. PCK expression was closely tied to energy output by the TCA cycle, however. Results such as these show that biology is complex: even a well-studied process such as gluconeogenesis may not be as simple as it appears on first glance.

Although PCK has been studied in the past extensively because of its central significance in metabolism, there has been renewed interest recently due to its newly identified role in cancer. The majority of work done in the past has focused on the eukaryotic cytosolic variant of PCK (PEPCK-C or PCK1), but recent work has highlighted the importance of the mitochondrial isozyme (PEPCK-M or PCK2) (13) and implicated its role in tumor growth (14–17). Cancer cells grow rapidly and require a major energy input to sustain growth. Vascularization of a tumor does not keep up with its growth, leading to constraints in nutrient availability, especially glucose. To adapt to these environments where glucose is a limiting factor, cancer cells can utilize other energy sources such as lactate and amino acids to feed into the gluconeogenesis pathway in which PCK plays a crucial role (18).

The role that PCK2 plays in cancer cells is the subject of several studies recently. PCK2 was found to be overexpressed in tumor tissue by Leithner et al. (15), whereas PCK1 expression levels have been found to be low. Vincent et al. have found that tumor cells with PCK2 silenced through siRNA and shRNA failed to grow when glucose was not present in the medium (16). The synthesis of nucleotides from glutamine in the absence of glucose requires either PCK1 or PCK2; this is carried out by taking the PCK reaction product PEP into the pentose phosphate pathway and serine/glycine metabolic pathways into purine synthesis (19).

The significance of PEPCK-M has only been recently discovered, and it poses a promising target for cancer treatments. Additional studies on PCK2 are necessary to comprehend the full extent to which this enzyme impacts eukaryotic life. The results presented in this dissertation that pertain to understanding the ligand binding interactions in the active site, along with characterizing the bound water molecules in the active site may ultimately translate into approaches for drug design and inhibitor development.
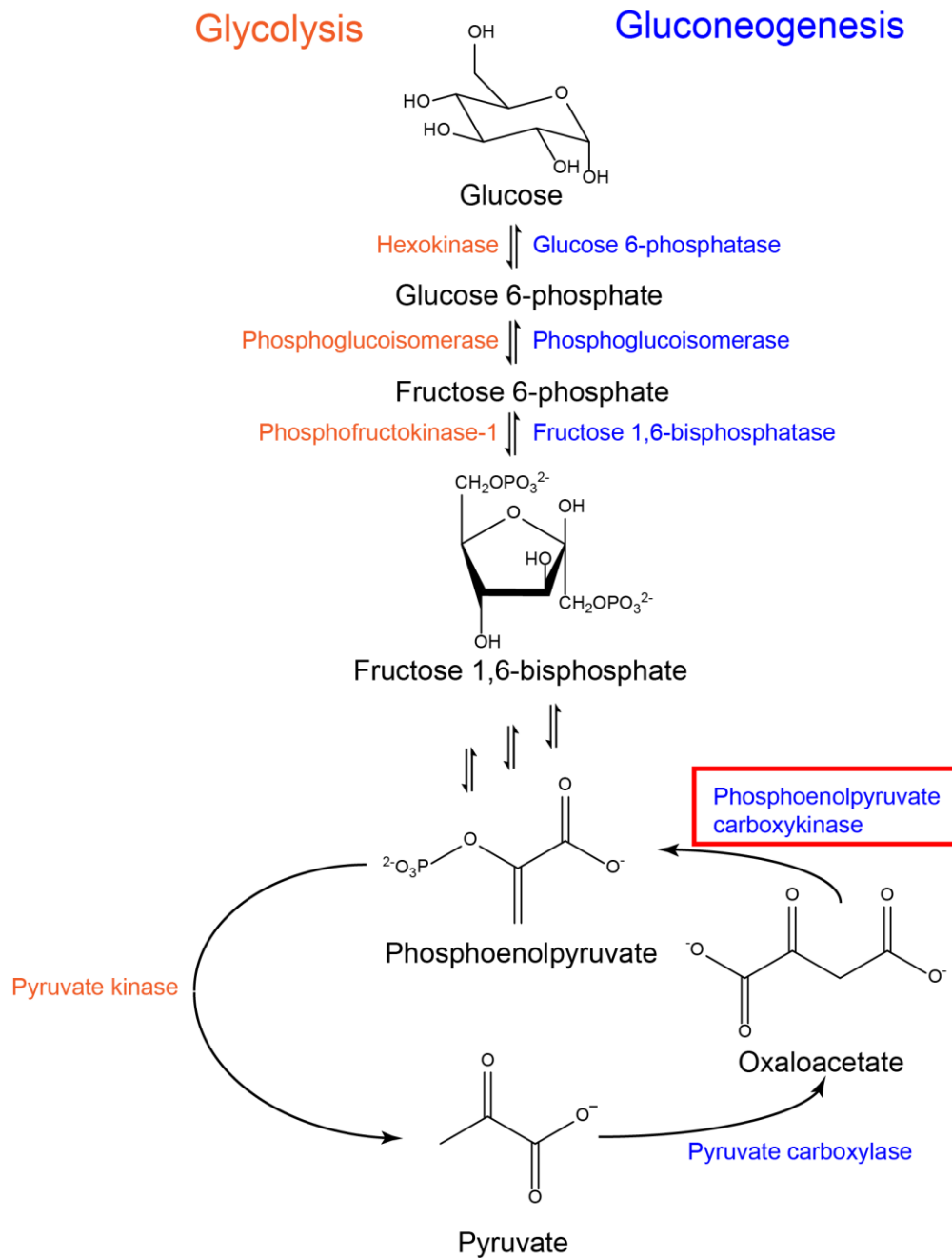
Figure 1.1. An abbreviated summary of the glycolysis and gluconeogenesis pathways in metabolism highlighting the role of phosphoenolpyruvate carboxykinase (PCK). Unidirectional arrows illustrate the direction of the reaction in vivo. Processes and enzymes involved in glycolysis are on the left side in orange, while gluconeogenesis-related enzymes are on the right in blue. Reactions between fructose 1,6-bisphosphate and phosphoenolpyruvate are omitted for brevity and are represented by the three equilibrium arrows. Adapted from Lehninger (20).
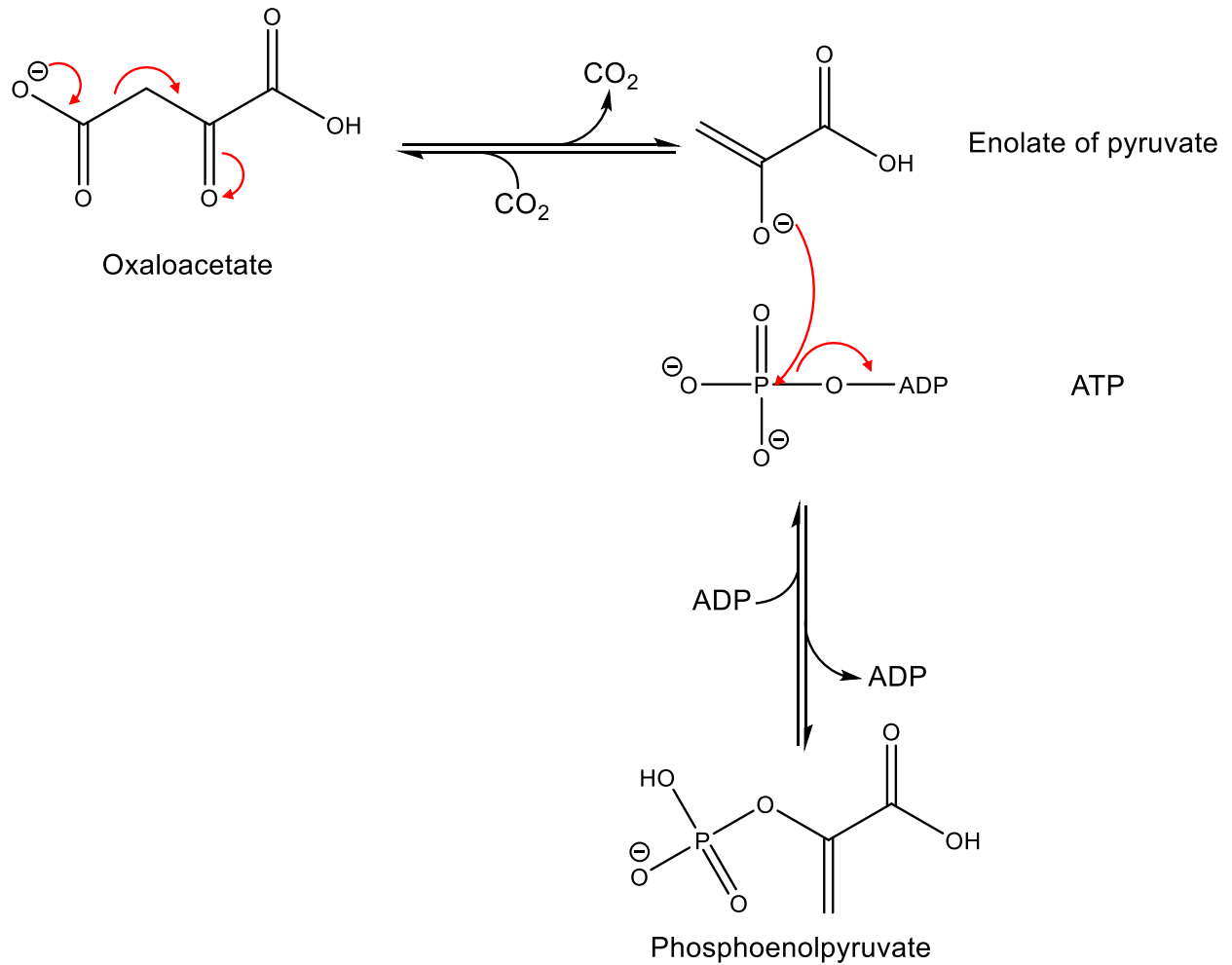
Figure 1.2. The reaction mechanism of EcPCK, the enzyme that catalyzes the first committed step of gluconeogenesis: the conversion of oxaloacetate to phosphoenolpyruvate. The reaction proceeds first by a decarboxylation of oxaloacetate, forming the enolate of pyruvate intermediate. The enolate then attacks the γ-phosphate of ATP, generating phosphoenolpyruvate.

N-terminal Domain **Cap** **C-terminal Domain**

Open Conformation

Closed Conformation
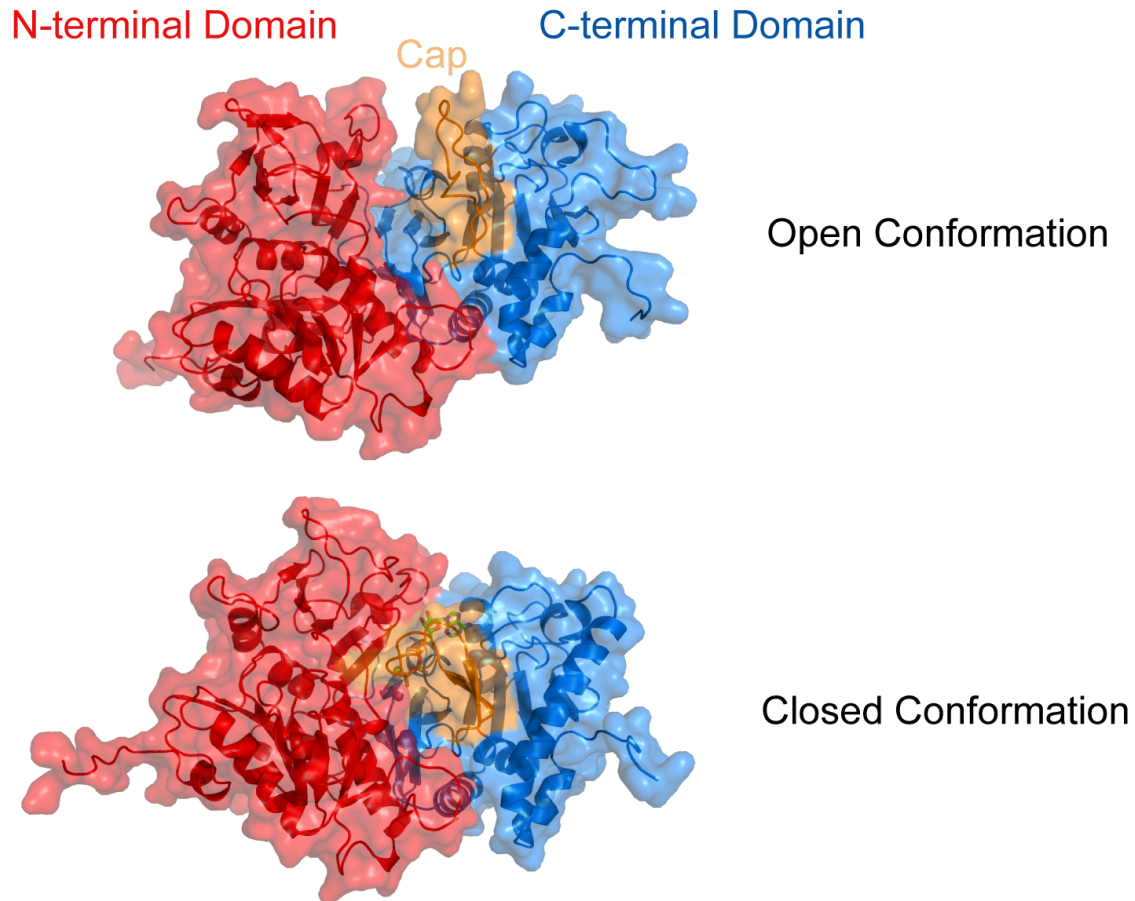
Figure 1.3. An overview of EcPCK crystal structures in its open apo state (top) and closed state complexed with ATP/Mg^{2+}/Mn^{2+} (bottom). The N-terminal domain (red) of the enzyme is positioned to the left side, and the C-terminal domain (blue) on the right. Structures were aligned to the C-terminal domains. The cap which closes over the active site cleft is depicted in orange. ATP in the closed conformation is shown with green sticks, Mg^{2+} as the green sphere, and Mn^{2+} as the purple sphere underneath the cap. Upon ligand binding, the domains undergo a rotation and close in a clamshell-like fashion.

1.2　Previous studies of PCK

Because PCK plays an important role in gluconeogenesis and thus all of life, it is not surprising that there are many prior studies on this system, including structural studies involving protein crystallography. This section touches upon previous work done on EcPCK as well as its eukaryotic variants, and provides motivation for using EcPCK as a model system for understanding other members of the PCK superfamily.

1.2.1　*E. coli* PCK structural work

A large amount of work regarding the structure of the PCK family has been accomplished by Goldie and Delbaere and colleagues. EcPCK served as the isozyme from which subsequent work was derived. The first EcPCK crystals were grown in 1991 (5), but the structure was not solved until 1996 (21). The story of this enzyme highlights the difficulties encountered when trying to overcome the phase problem in X-ray crystallography. Phasing was finally accomplished by multiple isomorphous replacement, through a series of heavy atom soaking experiments interwoven with computational methods such as solvent flattening. It is interesting to think about whether structures such as EcPCK could have been solved faster with more modern methods, but that would be a topic of discussion more relevant for crystallographic history and technique development. Nevertheless, it was exciting for me to see how much, and how little, the field has changed in the past three decades through this one particular case.

The initial PCK structure described by Matte et al. (21) provided a great deal of insight into the overall structure and the purported mode of nucleotide binding. The structure was crystallized with only acetate bound the active site, and not nucleotide. Because it was the first solved structure of the PCK family, it was seen that the nucleotide binding fold discovered in EcPCK was found to be different than existing P-loops. There was a large amount of reference in this work to previous attempts to elucidate the role of Ca^{2+} in the mechanism.

Subsequent work by Tari et al. (22) involved crystallization of the ATP/Mg^{2+}/oxalate complex, where oxalate served as a mimic to the reaction intermediate, the enolate of pyruvate. With nucleotide bound, they were then able to observe the interactions between the protein and the ATP molecule, as well as the global conformational changes in the protein as a result of ATP binding. Domain closing has been described as a motion involving a 20° rotation between the N- and C-terminal domains. In this crystallized complex, it was also observed that the ATP was in a syn conformation, with the adenine over the ribose instead of in a more energetically anti configuration where the adenine and ribose are staggered. The authors also stated that this was the only structure where ATP was found to be in the syn conformation. Mechanistic insights were made from this structure, with evidence suggesting a direct phosphoryl transfer from the ATP to the oxaloacetate and not progressing through a phosphoryl-histidine intermediate.

Other highlights from the PCK work done by Sudom et al. (23) involve gleaning more mechanistic insight into the phosphoryl transfer mechanism through crystallization of the ADP/AlF_3 as it mimics the transition state complex. The alignment of the various players in the reaction's transition state suggests that transfer of the phosphate between ATP and the enolate of

pyruvate happens directly, without an intermediate where phosphate is transferred onto a sidechain, resulting in a two-step mechanism.

One interesting point to note is that in these crystal structures, there has never been evidence of PEP binding. OAA binding is suggested in one of the structures, but upon closer inspection of the PDB deposit containing the model and maps with the supposed OAA binding, PDB ID 2PXZ, a close up around the OAA modeled in showed poor electron density around the ligand: for the carboxylate functionalities on either end of the molecule, there was only density supporting one of the oxygens. Since the electrons at the carboxylate functionality undergo delocalization, this brings into question whether the OAA molecule modeled in this position is real, or an over interpretation of the density maps. Since there is little to no evidence of binding of these substrates, mechanistic insights of the *E. coli* isozyme of this enzyme have to be drawn from other methods.

Overall, there has been substantial study into the mechanistic underpinnings of EcPCK with a strong focus on understanding the phosphoryl transfer process for PCK. However, little is known about the other parts of the mechanism, including the role of the CO_2 in the reaction and the ingress and egress of substrate and product. With largely crystallographic data to work with, conformational changes have been difficult to elucidate and evaluate.

1.2.2 Studies on other PCK isozymes

In addition to the work done on the *E. coli* isozyme of PCK, this enzyme from several other organisms has been characterized as well. Out of the invertebrate organisms, the structures of PCK from *Thermus thermophilus* (24), *Actinobacillus succinogenes* (25), *Anaerobispirillum succiniciproducens* (26), *Corynebacterium glutamicum* (27) as well as from pathogenic organisms such as *Trypanosoma cruzi* (28), *Mycobacterium tuberculosis* (29) have been solved. On the vertebrate side, structures of the rat (*Rattus norvegicus*) and human (*Homo sapiens*) cytosolic variants along with the chicken (*Gallus gallus*) mitochondrial variant have been solved.

The PCK structures come from a diverse set of organisms yet the overall fold is conserved. This is unsurprising for a crucial metabolic enzyme that most likely evolved from a common ancestor. This origin would logically result in conservation with a divergence from a core structure, which is what we see in PCK. Comparison of the structures show that the catalytic core remains largely the same, with key catalytic residues present in the same positions across the various isozymes. Figure 1.4 illustrates the high degree of structural similarity between the prokaryotic variants (Panel a) and between EcPCK and the eukaryotic variants (Panel b). The overall positions of the helices remain extremely consistent, with most of the differences in tertiary structure located in the peripheral loops of the enzyme. Section 1.2.3 explores the homology of the PCK superfamily further.

Various isozymes of PCK from vertebrates, including rat, chicken, and human, have been studied extensively by Holyoak et al. (8, 9, 30–34) These structures were crystallized in complex with various combinations of nucleotide and 3- and 4-carbon substrates OAA and PEP or its analogs. It can be seen from the above work that while the structures of EcPCK did not appear to crystallize in complex with OAA or PEP, both these substrates were present in some of

8

these structures. The presence of these substrates greatly aids in understanding the mechanism of PCK, as it provides concrete evidence for the position and orientation.

Beyond solving the structures, the effects of the Ω-loop which closes over the active cleft was studied by Johnson and Holyoak (8, 9). This 10-residue flexible loop is purported to close over the active site to facilitate catalysis by excluding bulk solvent, positioning residues for catalysis, and providing an energetically favorable conformation upon ligand binding. First, an alanine residue in the center of the loop, A467, was mutated to a glycine (8). This increased the flexibility of the loop, causing a larger loss in entropy upon lid closing. Also, an interaction between Y235 and the enolate intermediate was disrupted, resulting in disruption of catalysis. Subsequently, additional effects of the lid were probed by mutating the entire lid sequence to 1, 2, or 3 glycine residues (9). In all of these mutants, the enzyme was unable to fully reach the closed state, leading to a loss of catalysis, further cementing the importance of the lid domain in catalysis.

Inhibition of mammalian cytosolic PCK has been studied with various ligands that compete with OAA and PEP in the active site cleft (34). The ligands studied were largely bifunctional acids, including dicarboxylates, phosphonyl and sulfonyl carboxylates, diphosphoryls and disulfonates, and aromatic molecules. This work revealed the protein-ligand interactions involved in inhibitor binding and also native substrate binding as well, paving the way for subsequent inhibitor design. Subsequent work with inhibitors probed allosteric inhibition sites, leading to the discovery of 3-mercaptopicolinic acid for inhibition (31). This work had a large biological significance, tying into earlier knowledge that picolinic acid and various tryptophan metabolites act as inhibitors to gluconeogenesis. The data on PCK inhibitors provides us with a better understanding of the mechanistic underpinnings of the enzyme along with possible avenues for treating diseases where knockdown of PCK activity could lead to therapeutic effectiveness.

1.2.3 Structural homology of EcPCK to other members of the PCK superfamily

The existence of at least two forms of PCK was alluded to in previous sections. These forms can be broadly separated into two classes of the PCK superfamily: those which utilize GTP, and the others which utilize ATP. The PCK isozymes that utilize ATP are most commonly found in prokaryotes, while the GTP isozymes are predominantly in both the cytosol and mitochondria of eukaryotes (35) . Analysis of the protein-nucleotide interactions in ATP-utilizing EcPCK and the GTP-utilizing rat cytosolic PCK show that there is hydrogen bonding between the protein and the C6-carbonyl on guanine that would not be possible with adenine (32), precluding substitution of the two nucleotides.

The sequence homology between these two forms is not high; however, the structures of these two forms are similar, as can be seen in Figure 1.4. Upon overlaying structures from the PCK superfamily, there is no doubt about their similarity. The PCK superfamily highlights the importance of structure-function relationships: if one were looking purely at the sequence data, it would be hard to make the link between the ATP and GTP isozymes of this superfamily. This observation suggests that although high-throughput genomic studies in bioinformatics are important for the understanding of certain biological phenomena, structural information is still

irreplaceable for defining similarity. (In Chapter 3, I will introduce new metrics for defining structural similarity in solution by SAXS).

Structural insights from one member of the superfamily can be carried over to the next because of two features: the aforementioned structural homology, and the fact that key residues between the ATP and GTP forms are conserved even as the remaining sequence changes (30). The work reported here focuses on the *E. coli* variant of PCK (EcPCK) due to its robustness and amenability for overexpression in an *E. coli* host system, and the fact that the results in the behavior of PCK in general can be extrapolated from this system.
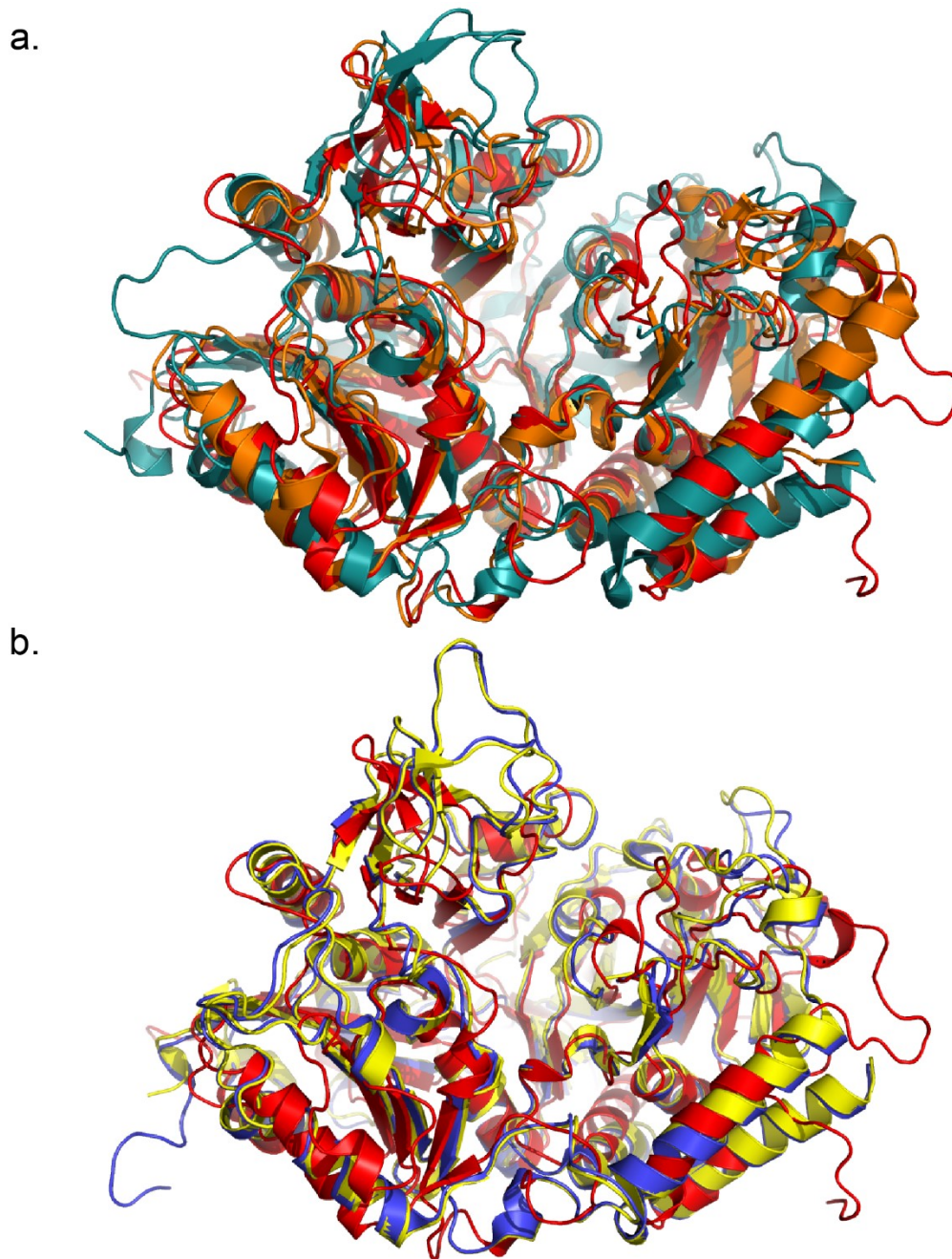
Figure 1.4. Structural homology between members of the PCK superfamily. Overlaying the PCK structures from different organisms shows that while the core is structurally conserved, the peripheral loops can vary widely across the isozymes. a) The PCK structures from three prokaryotic organisms, *E. coli* (red, Y207K model discussed in Ch. 3), *Mycobacterium tuberculosis* (teal, PDB ID 4WIE), and *Thermus thermophilus* (orange, 1J3B). b) Structural comparison of EcPCK (red) with *Homo sapiens* (yellow, 1KHG) and *Rattus norvegicus* (blue, 2QEW) cytosolic PCK show high similarity even though these organisms are drastically different.

1.3   Motivation for studies performed in this dissertation

Although there has been a great deal of work done on EcPCK and its isozymes, several aspects of this enzyme remain largely unstudied. This dissertation is directed toward characterization of the CO_2 binding pocket in EcPCK and understanding ligand-binding events in PCK relevant to enzymes in general, along with characterizing enzymatic conformational changes both crystallographically and in solution through integrated X-ray methods.

CO_2 binding has been explored in EcPCK through crystal structures of the complex pressurized under CO_2, but exploration of this pocket has not been performed. Here, several sites in this region of the active cleft have been identified and mutated to achieve nonnative ligand binding. These results are detailed by X-ray crystallography in Chapter 2.

In addition to engineering that pocket, the conformational changes that PCK undergoes during its catalytic cycle have been partially elucidated through crystallographic models. Yet, these results may be biased by the nature of crystallographic experiments. For example, in the *T. cruzi* PCK structure, the enzyme crystallized in the closed form even in the absence of substrate, indicating that crystal packing could have a strong influence on PCK conformation. Information gleaned from solution state data would provide valuable insight into the PCK mechanism under conditions that are more biologically relevant. Solution small angle X-ray scattering (SAXS) experiments have been performed on EcPCK in both its apo and holoenzyme forms to determine its conformation in an aqueous environment. The results arising from these experiments are described in Chapter 3. Finally, Chapter 4 provides a summary of the results along with future prospects that can build upon these results.

To accomplish the work in this dissertation, I designed 30 point mutants and identified 4 regions for multiple mutation and insertion mutants resulting in the generation of over 1000 mutants, crystallized and collected datasets of 160 crystals across 30 mutants of EcPCK, and performed over 300 SAXS experiments. The focus of this dissertation will be on 8 crystal structures and 12 SAXS profiles. Each data chapter begins with an introduction that takes a more in-depth look at these topics; the descriptions provided here below serve as a preview for the reader.

1.3.1   Crystallography of engineered PCK complexed with nonnative ligands

EcPCK has been crystallized extensively over the years to better understand how this enzyme works, providing a rich source of foundational structural information. While the existing crystallographic efforts have focused on mechanistic insight into the phosphotransfer mechanism of EcPCK, the work reported here leverages off the existing structures to rationally engineer nonnative ligand binding into the existing CO_2 pocket. The motivation for my work stems from the demand for catalysts capable of carbon-carbon bond formation. In the reverse PCK reaction, the enolate of pyruvate attacks CO_2 to form a new carbon-carbon bond and makes OAA. The structural knowledge to understand and manipulate the ligand binding and catalysis events provides the foundations for engineering binding of a ligand that is capable of a methyl transfer reaction, leading to enzymatic formation of a novel carbon-carbon bond.

Nonnative ligand binding is characterized in Chapter 2 of this work mainly by X-ray crystallography, with simple kinetic assays to measure the activity of the native reverse PCK reaction of various mutants to compare with the wild type. WT EcPCK and over 30 mutants were crystallized in the presence of various small molecules. The crystallographic results show two nonnative ligands, thiosulfate and methanesulfonate, bound in the CO_2 pocket of PCK and several of the mutants. Several single and double point mutants were found to bind to these ligands, but each with a different position and orientation of the ligand. One of these mutants did not alter the pocket directly: this structure uncovers the effects of neighborhood mutations in ligand binding. Interestingly, even though these ligands are similar in size and chemical properties, all the mutants bound either thiosulfate or methanesulfonate, but never both. The high-resolution crystal structures were paired with computational results to unveil the role of bound, but loosely bound water molecules in ligand recognition.

1.3.2   Exploration of conformational space spanned by EcPCK

Crystallography provides atomic resolution information into protein structure, but at the cost of providing conformational information under a small subset of chemical conditions. Solution-based structural techniques sample macromolecular conformations under biologically relevant conditions without the influence of crystallographic packing, but at the cost of missing the atomic details. Chapter 3 delves into characterization of EcPCK conformational states beginning from the crystallographic models then comparing those results to those obtained by solution-based biological small angle X-ray scattering (SAXS).

From the crystal structures of apo EcPCK and the ATP/Mg^{2+}/Mn^{2+} complex, I was able to model the open and closed domain states. Combined with crystal structures of the cap in the open and closed positions, these results provided solid experimental atomic resolution data to generate chimeric models of the four extreme states. Calculating the theoretical SAXS profiles of these states and comparing them with a metric known as the volatility of ratio (36), the conformational space spanned by the various states can be drawn out on a force plot to show the extent of the conformational state spanned by the EcPCK crystal structures. Experimental SAXS data is then added to visualize a large amount of data at once and to draw conclusions. Finally, fitting the experimental SAXS data with crystallographic models yields subtle but important differences providing insight on the solution state behavior of EcPCK. The SAXS experiments show that the capabilities of this technique easily exceed what researchers are using it for today.

2 Investigations of the EcPCK CO_2 binding site through mutations and characterization of nonnative ligand binding

2.1 Introduction

2.1.1 The EcPCK active site

The enzyme active site is where the two most visible steps of an enzyme-catalyzed reaction occur: substrate binding and catalysis. In the PCK superfamily, the active site is located in the cleft between the two domains of the protein (21): hence the more descriptive name for it is the active site cleft. The active site cleft is divided into several discrete regions for binding its substrates. In EcPCK, starting from the most straightforward region, ATP binding is accomplished through coordination of two metal ions, Mg^{2+} and Mn^{2+}, as well as through a pi-pi stacking interaction between adenine of ATP and Arg449 of the protein (22). The ATP γ-phosphate points towards the rest of the active cleft as it gets transferred during the reaction. The PCK phosphotransfer mechanism was heavily investigated in previous studies (23, 32, 37–39) and will not be described here. There is also a binding site for OAA, which serves to bind PEP and CO_2 in the reverse direction and is described further below.

In EcPCK, information about binding of substrates other than ATP or CO_2 is not available due to the lack of substrate-bound crystal structures. However, these can be inferred chemically, compared with other isozymes that did crystallize with the substrate of interest, or understood through substrate analogs. Oxalate and pyruvate are two commonly used analogs for OAA and PEP. Chemical inference suggests that because the reverse reaction involves PEP and CO_2, while the forward reaction only takes OAA, PEP binding must take place where the oxalate portion of OAA binds, as well as the γ-phosphate position of ATP, while CO_2 binding should be located at the acetate end of where OAA binds. Section 2.6.2 characterizes OAA binding in further detail.

CO_2 binding is a unique characteristic to PCK that is not available to most enzymes. Rather, reactions typically proceed through the bicarbonate (HCO_3^-) ion because the diffusion-limited nature of carbonic anhydrase means that conversion between CO_2 and HCO_3^- happens instantaneously on biologically-relevant timescales, and the availability of CO_2 in solution is limited by its lower solubility compared to bicarbonate. For example, the conversion of PEP to OAA in vivo is accomplished by the enzyme phosphoenolpyruvate carboxylase (PEPC). While this reaction looks like the reverse reaction of PCK, it uses bicarbonate as its carbon source rather than CO_2. It appears that PCK did not evolve to use CO_2 in its mechanism; rather, since in vivo it catalyzes the decarboxylation of oxaloacetate to generate CO_2, and the reverse direction also has to proceed, CO_2 binding and catalysis is a byproduct of its native function. Better characterization of this CO_2 binding pocket will lead to an improved understanding of this

14

enzyme's mechanism. This knowledge is biologically important and has promise for synthetic biology and practical applications including carbon sequestration.

In addition to understanding how CO_2 binds in this pocket, the possibility of replacing CO_2 for another tetrahedral-like substrate was explored. Because existing structures are available, the CO_2 pocket was modified through exploration of not only the sequence, but also through rational structural mutations in three-dimensions. Enzyme engineering has been long regarded as the Holy Grail in the field of protein science. As biological catalysts, their ability to speed up reactions often rivals or surpasses chemical catalysts with the possibility of much lower than cost and milder conditions, especially compared to transition metal catalysts. The field of chemistry would be revolutionized if one could efficiently design an enzyme to catalyze a desired reaction. While our scientific knowledge has not reached that point yet, growing our understanding of ligand binding in enzyme active sites is merited as the foundation knowledge to modify enzymes for science and technology.

My work investigated the effects of mutations of two structural elements in and around the CO_2 binding pocket. Mutants were characterized for their loss of native activity, and novel ligand binding modes were sought out by protein crystallography, as further described below.

2.1.2 Overview of carbon-carbon bond formation in chemistry

Carbon-carbon (C-C) bond formation remains one of the most important reactions in chemistry. The C-C bond itself is extremely strong, with C-C bond dissociation energies ranging from ~80-140 kcal/mol. Due in part to this strength, C-C bonds are central to a myriad of molecules, ranging from synthetic polymers to pharmaceutically-relevant small molecules to biofuels. There are a variety of purely chemical reactions to generate this bond, including many covered in introductory organic chemistry courses such as the Grignard, Diels-Alder, and Wittig Reactions, just to touch on a few. The 2010 Nobel Prize in Chemistry was awarded to Richard Heck, Ei-ichi Negishi, and Akira Suzuki for their work in carbon-carbon coupling reactions catalyzed by transition metals.

While science has made great strides in devising chemical catalysts and reactions to make carbon-carbon bonds, they pale in comparison to biological catalysts devised by nature and optimized through billions of years of evolution. Compared to chemical catalysts, enzymes catalyze reactions more specifically, and more cost-effectively when transition metals such as palladium are otherwise required. Enzymatic C-C bond formation is a common occurrence in metabolism; enzymes that carry this process out include aldolases, carboxylases, and various synthases. The reader is directed to works by Fessner (40) and Bugg (41) for a more thorough discussion of enzymatic C-C bond formation. EcPCK C-C bond formation is similar to that of the carboxylases, reacting with CO_2 through a carbanion intermediate. Thus, insights gained from this system should be broadly applicable to other carboxylases, expanding the repertoire of enzymes for use in enzyme engineering and the knowledge base to redesign their desired ligand binding in predetermined ways.

2.2 Materials and methods

2.2.1 Identifying regions of interest for mutations

To engineer novel ligand specificity into the CO_2 binding pocket, previous studies were considered in identifying the regions of interest. CO_2 interacts with the WT enzyme through Arg65 and Tyr207. Initially, both these residues were targeted with single point mutants Arg65 is located on a flexible loop while Tyr207 is located just before an α-helix. Alteration of the hydrogen bond network in this area was hypothesized to change ligand specificity, so mutations were made to residues 63-70 of the loop, and to the α-helix following Tyr207 (residues 209-216) as they form the sides of the CO_2 pocket. Figure 2.1 shows the regions where the mutations were targeted. One to four residues across these two regions were mutated to a combination of amino acids in the set (A, C, E, F, I, L, M, Q, S, V).

2.2.2 Expression and purification of EcPCK

The gene coding for EcPCK was PCR-amplified from genomic *E. coli* DNA. A C-terminal 6xHis-tag was added through the reverse primer. The amplified fragment was TOPO-cloned into a pET-21b vector and verified by restriction digests and sequencing. The resulting plasmid was transformed into Rosetta 2 (DE3) pLysS E. coli competent cells (Novagen).

EcPCK point mutants were generated using gene Splicing by Overlap Extension (SOEing) and cloned using Sequence- and Ligation-Independent Cloning (SLIC) (42). G209S K212C and K212I F216V double mutants were isolated from shuffled DNA libraries generated using GRAMMR® technology for high-resolution genetic reassortment (Novici Biotech LLC). Briefly, sequence-verified plasmids containing the specified mutations were shuffled using GRAMMR to re-assort into various combinations so that, on average, each clone contains unique combinations of 2 mutations at only the desired positions with the specified residue substitutions.

Overnight cultures of transformed Rosetta 2 cells were grown at 37°C with shaking at 250 rpm in LB supplemented with 100 μg/mL ampicillin. Overnight cultures were transferred into larger volume cultures and grown until $OD_{600} = 0.500$, then induced with 1 mM IPTG and grown for an additional 16 h at 16°C. Cells were then harvested by centrifugation at 8000 x g for 20 min at 4°C, and frozen at -20°C until use.

Frozen pellets were resuspended in buffer A (20 mM Tris pH 8.0, 200 mM NaCl) and cells were lysed by sonication. Cellular debris was cleared from the lysate by centrifugation at 40,000 x g for 30 min at 4°C.

PCK was batch-bound to Ni-NTA resin (5 PRIME) by rocking at 4°C for 2 h (2 mL 50:50 resin slurry per L culture). Resin was then washed with 10 x CV Buffer A, then 5 x CV of Buffer A + 25 mM imidazole. PCK was eluted with 5 x CV Buffer A + 250 mM imidazole.

Ni-NTA column eluate was loaded onto an S-100 16/60 gel filtration column (GE Healthcare) pre-equilibrated with Buffer A. Fractions containing the main peak were pooled and

concentrated in a Vivaspin 20 concentrator with 10K MWCO (GE Healthcare). 50 μL aliquots were plunged into liquid nitrogen and frozen at -80°C until ready for use.

2.2.3   Enzyme kinetics

Kinetics measurements of the purified EcPCK protein variants were carried out through a coupled enzymatic assay. To probe the activity in the reverse direction, the conversion of PEP to OAA, malate dehydrogenase was used as a readout enzyme. A standard curve of NADH was created at 340 nm. Next, a 100 μL reaction consisting of 100 mM $NaHCO_3$, 10 mM ADP, 2.5 mM $MgCl_2$, 2.5 mM $MnCl_2$, 100 mM Tris pH 7.5, 1 U malate dehydrogenase, and 1 mM NADH was created. 0.5 μg of purified EcPCK enzyme was added, except in the no protein negative control. The reactions were then started by the addition of PEP to a final concentration of 5 mM, save for the no substrate negative control. Spectrophotometric readings at 340 nm were collected on a BMG Labtech PolarSTAR Omega UV/visible plate reader every 20 seconds for 5 minutes. Experiments were performed in triplicate. The resulting absorption versus time plots showed a linear decrease over time until a minimum was reached, usually around 3 minutes. The linear portion was fit with a linear regression, and the slope of the line converted to a specific activity (μmol min^{-1} (mg enzyme)$^{-1}$).

2.2.4   Protein crystallography

PCK crystals (unless specifically mentioned below) were grown by hanging-drop vapor diffusion. Solutions of 10-15 mg/mL protein, 20 mM Tris pH 8.0, 200 mM NaCl, 5 mM $MgCl_2$, 5 mM $MnCl_2$, 5 mM ADP, 5 mM PEP, and one of 100 mM sodium thiosulfate, sodium methanesulfonate, sodium ethanesulfonate, or sodium 2-mercaptoethanesulfonate were combined with an equal amount of precipitant containing 24% PEG 3,350, 100 mM BIS-TRIS pH 5.5, 400 mM NaCl and placed into a 20°C incubator. Rod-like crystals formed in approximately 3 days.

Before data collection on the G209S K212C double mutants, crystals were derivatized with xenon to identify hydrophobic pockets. Crystals were pressurized to 200 psi in a xenon derivatizer (Hampton Research) for 2 minutes and flash frozen.

Crystals of the K212I F216V mutant were initially grown in the same conditions as above, but only fine needles too small for diffraction experiments were observed. Crystals of the mutant were grown by hanging-drop vapor diffusion with a new well solution consisting of 0.16 M ammonium sulfate, 0.08 M sodium acetate pH 4.8, 20% PEG 4000, and 15% v/v glycerol. Resulting crystals were in large plates that diffracted with acceptable statistics.

Crystallographic data were collected at the SIBYLS beamline at the Advanced Light Source in Berkeley, CA (43). Before data collection, crystals were cryoprotected by sitting for one minute in the precipitant solution with 20% glycerol. The crystals diffracted to 1.79-1.12 Å and images were recorded on an ADSC Q-315 detector. 200 degrees of frames were collected for the majority of crystals. Images were processed using XDS (44). The structures were solved by

molecular replacement using Phaser (45) and refined using phenix.refine, both in the PHENIX software suite(46). Fitting was performed in Coot (47). Figures were generated in PyMOL (48) and LigPlot+ (49).

2.2.5 Computational prediction of mutant-methanesulfonate binding and solvent analysis

Thirty mutants were prepared using MOE (50) suite of programs based on G209S-methanesulfonate crystal structures. Methanesulfonate was docked into mutants followed by single point MM-GB/SA rescoring calculations, all of which were performed using MOE. Ten top-ranked methanesultonate binding poses for each mutant were saved as the initial structures for MD simulations, resulting in a total of 300 mutant-methanesultonate systems. The AMBER ff99SB force field (51) was used to model the protein system, while the parameters to model methanesulfonate were obtained from the generic AMBER force field (GAFF) for small molecules (52). The charge for methanesulfonate was derived following the restrained electrostatic potential (RESP) procedure (53, 54) using Gaussian 03 (55). Also, in order to improve the modeling the Mn^{2+} complex, MCPB.py was used to generate the force constant parameters (56). For each mutant-methanesulfonate system, a geometry minimization was conducted prior to slowly heating the system to 300 K and relaxed at 1 atm. After that, a 25 ns NPT production simulation was performed. Subsequently, for each system, MMGBSA.py was implemented to calculate the methanesulfonate binding free energy against the entire MD trajectory. For each mutant, the average binding free energy of 10 systems was calculated and ranked in order to make predictions. In addition, because the small size of the ligand presents a challenge to free energy evaluation using docking and/or MM-GB/SA, for top ranked mutant-methanesulfonate complexes, a conformational search was conducted using the LowModeMD function of MOE, and the interaction energy between ligand and receptor was calculated for each obtained conformation. The interaction energy and the average binding free energy, calculated using MMGBSA.py, were used to cross-screening mutants in order to improve prediction accuracy.

In order to understand the impact of ordered solvent to the ligand preference between WT and G209S, we also implemented the solvent analysis using MOE. Potential water sites within 10 Ångstrom from the ligand site were calculated for each receptor (assuming no ligand binding yet).

2.3   Ligand fitting in macromolecular crystallography

Protein crystallography may seem like a routine task in this current day and age, especially with synchrotron data collection and highly automated data processing software with point-and-click interfaces. However, while modeling protein into electron density can be straightforward and may not require much skill from the crystallographer, building and modeling ligands is an entirely different scenario. Requiring both chemical fundamentals and crystallography knowledge, ligand fitting in macromolecular crystallography is not a trivial task. Mistakes in ligand fitting can lead to massive differences in interpretation of results with possible downstream consequences. Indeed, there have been many structures deposited into the PDB with questionable ligands (57, 58). This section serves to walk the reader through the process of ligand fitting, and to highlight common mistakes made during this process.

Before beginning the ligand fitting process, the resolution of the dataset should justify ligand placement. Because ligands are small molecules consisting of < 100 atoms, medium to high resolution data is required to fit the ligand. Low resolution data lacks many of the features required to correctly position the ligand in the density: with diffraction data extending to 3.0 Å resolution or less, electron density becomes increasingly tube-like. Not only does it become hard to interpret or even observe ligand binding, the protein itself becomes harder to build as sidechains become less distinct from the main chain.

If the resolution is sufficient, the ligand fitting process begins after rounds of refinement and fitting results in a complete protein model. After all density that can be attributed to protein is placed, large unexplained patches of density could possibly be ligand. A list of all possible ligands present in the crystallization condition should be made, including possible reaction or degradation products, along with chemicals present in the well solution to induce crystal growth, such as precipitant or buffer. Hopefully, one of the molecules present in the crystallization experiment fits into the density.

Once a possible molecule has been identified, an initial model of the ligand, and its corresponding restraints, must be built before attempting to model it into the density. If the ligand is common or has been found in an existing crystal structure that has already been deposited into the PDB, most of the work has already been done as that existing ligand model can be used. If such a thing does not exist, one will have to generate the model and restraints using software such as eLBOW (59) in the PHENIX suite or tools built into Coot (60). While the importance of the model is self-evident, the restraints are extremely crucial to building a proper model: during refinement, the model is altered to better fit the electron density. If the ligand restraints are poor, it is possible to deform the ligand to physically unrealistic conformations in order to attain a "better" fit into the electron density; ligand restraints ensure that the refinement software knows what conformations are allowed and what aren't.

Once the ligand is built into the density, the occupancy often needs to be refined as ligands are rarely present in every copy of the unit cell in any given crystal. Accomplishing this in modern refinement software is usually as simple as changing the starting occupancy to a value of less than 1.0, then telling the software to refine the occupancies as well during a given macrocycle of refinement. In this case, it is important that the temperature factor is reasonable. One option is to set it to be similar to surrounding atoms.

In the end, it is important to verify that the ligand that was modeled into experimental electron density makes chemical sense. Validation is accomplished through several metrics (61), and can be split into two categories: global measures and local measures (62). Global measures mainly deal with reciprocal space and thus assess the overall quality of the model and not just the ligand. Metrics that fall under this category include resolution, R-work and R-free, and CC*. While these metrics by themselves do not tell us much about the quality of the ligand fit, the overall quality of the model is important for determining whether conclusions can be safely drawn from the data. Local metrics can further be broken down into two groups: electron density-based metrics, and stereochemistry-based metrics.

Electron density-based metrics such as the real space correlation coefficient (RSCC), and Real Space R-value (RSR) deal with how well the ligand fits in the real space electron density: they compare the observed density with the calculated density, just like the traditional R-values. However, since RSCC and RSR do not account for disorder, this is reflected in the B-factor of these atoms. Two other metrics partly address disorder, the Real Space Observed Density Z-score (RSZO) and Occupancy-weighted Average B-factor (OWAB) that take disorder into account in different ways. In the case of OWAB, occupancy is also taken into account.

The proper visualization of the maps is also important for correct interpretation. Maps are calculated using the observed and model structure factors. Because there are errors in measurement as well as in the phases, map coefficients should be weighted to generate better maps, known as sigma-A weighted or 2mFo-DFc maps, where m is the figure of merit for each reflection, and D is the sigma-A weighting term (63). This will serve to accentuate features that are missing in the model. Modern refinement software should generate these maps by default. Other alternative maps that may prove useful would be the 3mFo-2DFc map which would further accentuate incorrectly modeled features. As model bias can skew the map to make an improperly modeled ligand appear acceptable (the model contributes to the Fc terms), removing the model bias would provide a clearer picture of the ligand density (64). Composite omit maps are generated by partitioning the unit cell into small regions, removing all model atoms from a particular region and calculating the local map (usually 2mFo-DFc), and then combining all the regions into a map of the entire unit cell (65). This is especially valuable for building and validating ligands, and thus, all maps of the bound ligands presented in this chapter are 2mFo-DFc composite omit maps.

Whatever the metrics indicate, it is vital to investigate the ligand visually and see if it makes chemical sense. Things to consider include bond angles and lengths, packing interactions, and hydrogen bonding distances to the protein and bound solvent. Does this new model agree with existing knowledge, such as biochemical data, or analog binding? If something deviates from what is known or even expected, it is worth investigating and testing further.

2.4 Point mutations in the active site often result in loss of native activity

Mutant enzymes containing a point mutation targeting residues largely in the CO_2 binding pocket were generated and characterized for native activity in the reverse direction (OAA formation) as described in the Materials and Methods, Section 2.2.3. The conditions were based on previous work performed by Goldie and Sanwal (4), but modifications were made to increase the buffering capacity of the working solution and to increase the divalent cation concentrations. The results arising from these modified experiments still produced measurements that yielded a linear portion of the reaction curve (data not shown). For clarification, this was not a classical study done to obtain kinetics constants as that was not the goal: the experiment aimed to tell us whether certain mutations disrupted the ability of the enzyme to catalyze the native reaction, thus giving insight on the importance of target residues in the pocket. Because of this, the change in activity should be evident even with the simple assay presented here with reactions performed at one concentration, resulting in a comparative specific activity of the mutant. The goal is more comprehensive testing of the active site neighborhood rather than more fully investigating a single residue or mutation. The simple assay makes this more efficient and cost-effective.

The results from the kinetics assays are presented in Figure 2.1. From the assay results, it is evident that both the position of the mutations and the identity of the mutation play roles in influencing and disrupting native activity. There is a range of effects from this small collection of point mutations, with some mutations showing minimal to no impact on native reverse activity (Y207L, Y286F), six mutants showing ~three-fold reduction in activity (Y207F, T394S, T394V) to six-fold reduction in activity (Y207N, F413Q, F413W), and the remaining mutants showing minimal to negligible activity relative to the no protein control.

The two CO_2 hydrogen bonding positions, R65 and Y207, show markedly different effects when mutated, even though both sidechains are implicated in CO_2 binding. CO_2 has been shown to hydrogen bond to NE of R65 and OH of Y207, so disruption of either of these hydrogen bonds should lead to similar degrees of loss of activity. Looking at the kinetics data, this is not the result. The three R65 mutations (R65H, R65K, R65Q) all resulted in loss of activity, even though these mutations should be able to replicate, or at least mimic, the H-bonding that is required for CO_2 binding and catalysis. Looking at the Y207 results, the results are a lot more varied. Y207L showed a comparable level of activity to WT even though leucine is not capable of hydrogen bonding. Y207F is also not capable of hydrogen bonding and is structurally very similar to tyrosine, save for the hydroxyl group, yet this mutation results in a threefold decrease in activity relative to WT. Y207N shows higher than baseline activity but is almost an order of magnitude reduced from WT. The other two mutations, Y207I and Y207K, show minimal activity relative to WT.

The next three residues in the sequence that were mutated, G209, F216, and S250, all resulted in a drastic loss of activity. Chapter 2.6 will show that mutations at G209 have interesting effects on ligand binding. S250 is positioned more towards the ATP binding area which binds the divalent cations, so loss of activity for these mutants might be due to loss of ADP/ATP affinity rather than through CO_2 interactions.

Y286 is in close proximity to Y207, near the CO_2 binding pocket, but it is not directly involved in H-bonding to CO_2, so the Y286F mutation retains packing in the active site and does

not alter activity. The Y286A and Y286W mutations do not retain this packing, as seen from the loss of activity.

T394 is located on the cap which closes over the active site (Chapter 1.1.3) and is not directly positioned in the active site when the cap is open. However, previous studies have shown that cap mutations can have substantial effects on PCK kinetics (8), and we see that as well in these mutants. We are not altering the flexibility as much as in the previous work, so the mutations must be altering the kinetics in some other fashion. Detailed analysis of the closed cap state of EcPCK in Chapter 3 provides additional information on these residues.

F413 is present in the CO_2 pocket of the active site as well, but not on the flexible section of the cap: the position of the Phe sidechain does not appear to change whether the protein domains are open or closed. Partial activity is retained when F413 is mutated to Q or W, but native activity is lost completely when F413 is mutated to Y. In terms of sterics, the F413Y mutation would be closest in size to the original sidechain, yet that showed the greatest deviation from WT activity. From the subset of mutations made, all the sidechains are neutral at physiological pH yet there is a wide range of impact on activity. Further exploration of mutations at this position is required to understand the role of F413.

Overall, the kinetics data of the various point mutants provide complementary information to the crystal structures. Flexibility combined with hydrogen bounding and electrostatics all play roles in controlling ligand binding and orientation. The change in native reaction rate due to mutation of certain residues highlight catalytically crucial regions and also those that are of interest for engineering efforts. The next section characterizes select mutants that led to novel ligand binding.
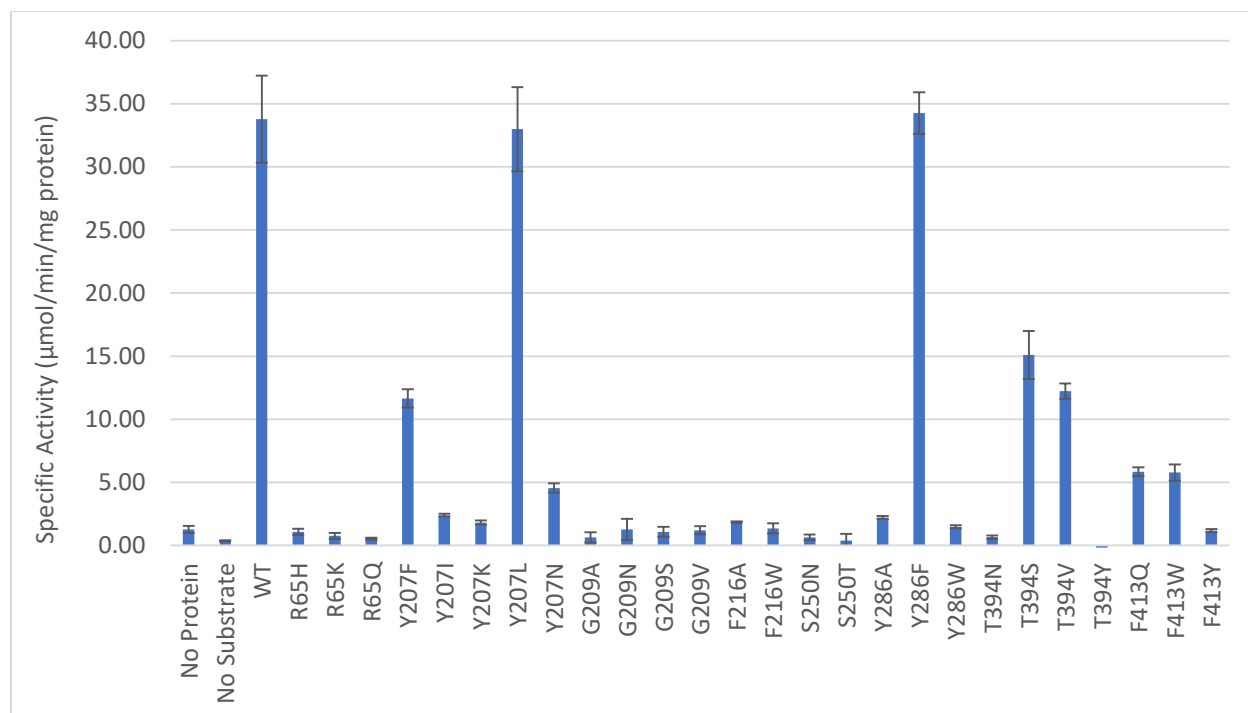
Figure 2.1. Specific activities of various EcPCK mutants in the reverse reaction direction. EcPCK and its mutants were assayed for their propensity for reaction in the reserve direction of the native reaction (OAA formation) through a coupled spectrophotometric assay. Mutants listed in order of residue number. Each measurement was taken in triplicate. Error bars indicate ± 1 standard deviation in the measurements. Wild type specific activity under this assay's reaction conditions was determined to be 34 ± 3 µmol min^{-1} mg^{-1} of purified EcPCK enzyme. The point mutants studied show different degrees of impact on native activity, ranging from no effect on activity (Y207L, Y286F), several-fold loss of activity (Y207F, T394S, T394V), to complete loss of activity. T394Y yielded a slightly negative value of -0.17 and is not visible on this plot. Results suggest that certain positions are more sensitive to others with respect to native activity, and the identity of the mutation also plays a role in loss of activity. Assays completed in collaboration with Sharon Borglin (Lawrence Berkeley National Laboratory).

2.5　Thiosulfate binding in wild type EcPCK

EcPCK has been crystallized in the past, but the protocol was laborious, requiring seeding to obtain crystals suitable for diffraction (5). Thus, the published method was not amenable to the many mutations and ligands that required screening for this work, so my crystallization efforts began by exploring new crystallization conditions to obtain nonnative ligand binding and catalysis. A new condition that did not require seeding was found, containing 24% PEG 3,350, 0.4 M NaCl, and 0.1 M Bis-Tris pH 5.5. Wild type EcPCK was crystallized with the nonnative ligand thiosulfate (THJ), yielding crystals in the $P2_12_12_1$ space group with two molecules per asymmetric unit (ASU) and unit cell parameters a = 94.72 Å, b = 103.61 Å, c = 119.18 Å that diffracted to 1.33 Å with 243952 unique reflections and 92% overall completeness. The model is complete except for the first five residues of the N-terminus and the C-terminal 6xHis affinity tag. Crystallographic data collection and model statistics are presented in Table 2.2 located after Section 2.6.

Electron density indicative of THJ was identified in both the 2mFo-DFc and mFo-DFc maps at the ATP binding site in both copies of the protein as well as the CO_2 binding pocket in chain A. Figure 2.2 shows the 2mFo-DFc electron density from a composite omit map contoured to 2σ surrounding the THJ ligand. The sulfonate (-SO_3) portion of the molecule did not have symmetric density, and density surrounding one of the oxygen atoms was much larger than the others, suggesting that thiosulfate in WT EcPCK binds in the CO_2 binding pocket with possibly more than one orientation. However, upon refinement with the additional conformations, the resulting occupancies in the subsequent orientations were low (< 20%).

The structural information reported here is for the primary orientation of thiosulfate in chain A. Occupancy of the thiosulfate ligand in the CO_2 binding pocket is 68%. Hydrogen bonds between the oxygen atoms of the sulfonate (-SO_3) moiety on THJ and side chains in the active site mediate the protein-ligand interaction (Figure 2.2 and Figure 2.3). The two residues previously reported to play a role in H-bonding, R65 and Y207 (7), are involved in binding of thiosulfate too. Thiosulfate interacts with NE of R65 (2.85 Å) and OH of Y207 (2.67 Å). This can be attributed to the similar O-O distances between CO_2 and the O-S-O bond on thiosulfate: the ideal O=C=O bond length is 2.32 Å, but observations in a previous PCK crystal structure (PDB ID 2OLQ) had a bond length of 2.36 Å, while the O-S-O distance for the atoms interacting with R65 and Y207 in this thiosulfate molecule is 2.50 Å. The position of the native CO_2 from PDB ID 2OLQ is shown in Figure 2.2 (transparent sticks) and further illustrates the similarity of the bond lengths. The thiosulfate oxygen atoms also hydrogen bond with the –NH_3 group of K212 mediated through a water, the carbonyl oxygen of T394 mediated through a bound water as well, the –NH_3 group of K213 (3.07 Å), and NH1 of R333 (2.92 Å) (Figure 2.3). These interactions lead to THJ in the WT CO_2 pocket to be oriented with the sulfur atom pointing to the center of the N-terminal lobe of PCK.

Overall, the enzyme adopts a closed conformation that overlays well with WT EcPCK with ATP bound. This extends and corrects previous studies that stated nucleotide is required to reach the closed conformation, as no ATP was present in this crystallization condition: THJ was present in the ATP binding site and appears to be sufficient for closing the system.

24

Crystallizing WT EcPCK in the presence of both THJ and ATP along with Mn^{2+} and Mg^{2+} required for ATP binding yielded crystals that only contained THJ. THJ was present in 20-fold greater concentration than ATP (100 mM vs. 5 mM), but it was still interesting to see THJ preferred over ATP in its native binding pocket. Aligning the two models yielded an RMSD of only 1.08 Å. Yet, there was one key noticeable difference: the cap over the active site was in the closed position when THJ was bound, but in the open position in the ATP/Mg^{2+}/Mn^{2+} complex. The significance of the cap and conformational change upon thiosulfate binding is detailed in Chapter 3.

Collectively, these experiments provide novel data for better characterizing the CO_2 binding pocket in EcPCK. The ability for this pocket to accommodate a much larger tetrahedral ligand coupled with a library of mutants with slightly altered pockets also provides a starting point for engineering nonnative ligand binding and catalysis into this enzyme. The next section covers the characterization of several different point mutants bound to thiosulfate as well as a second nonnative ligand, methanesulfonate.
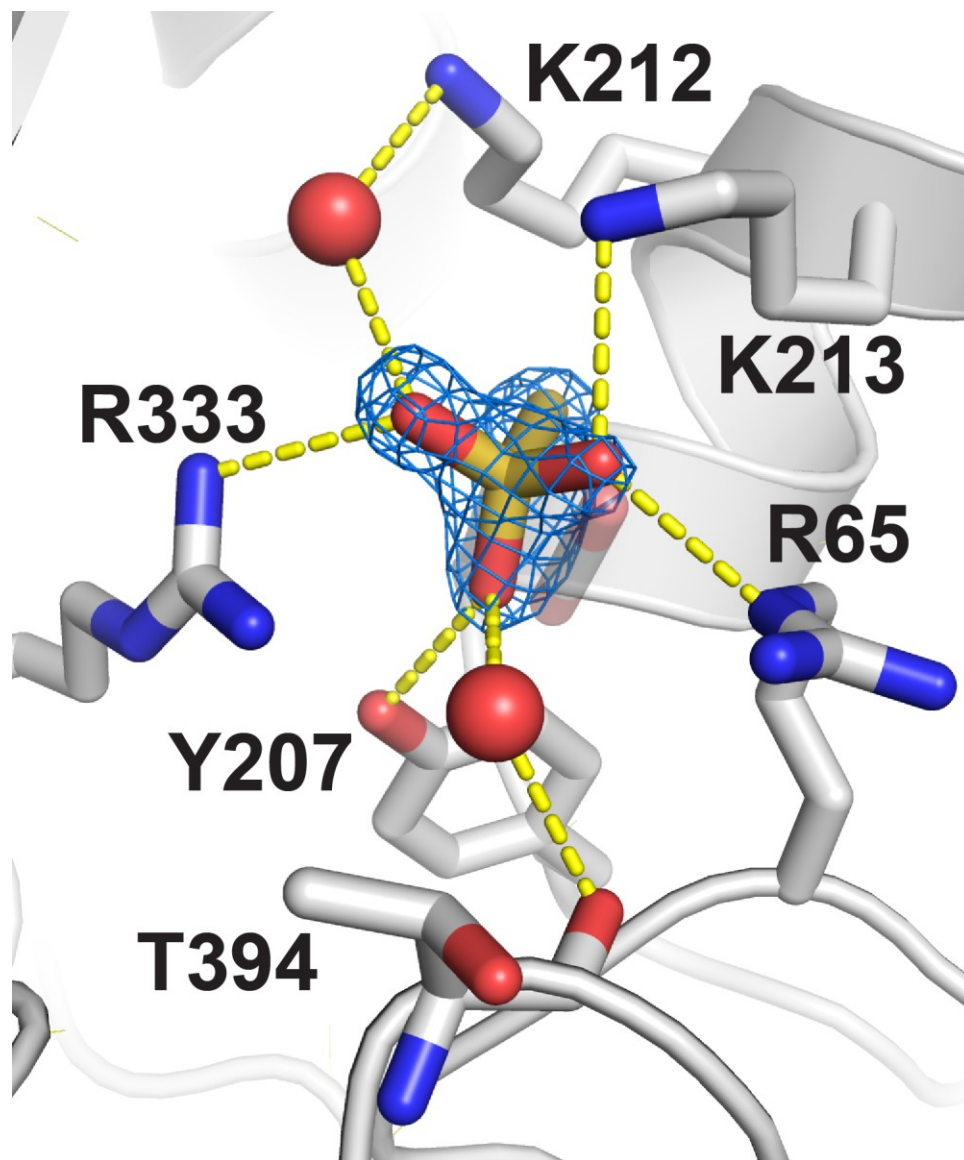
Figure 2.2. Detailed model of the WT EcPCK CO_2 binding pocket with thiosulfate bound. Interactions between the bound nonnative ligand thiosulfate and sidechains are illustrated. The 2mFo-DFc composite omit electron density maps, shown in blue and contoured at 2.0 sigma, surround the nonnative ligand in the pocket. Yellow dashed lines represent hydrogen bonds. Red spheres indicate ordered water molecules. CO_2 from PDB ID 2OLQ shown as transparent sticks to highlight native CO_2 position and orientation.
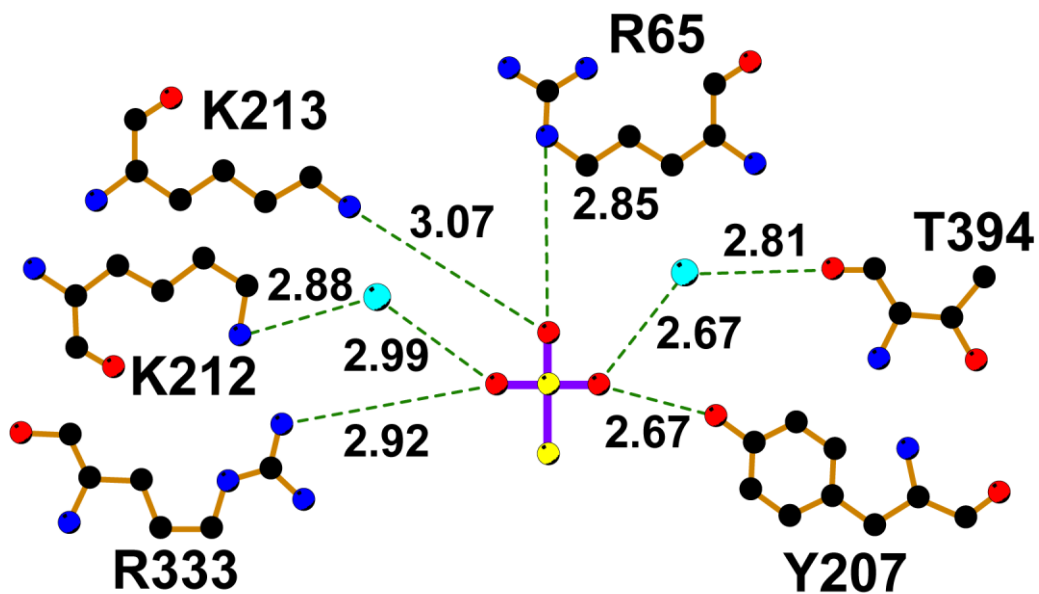
Figure 2.3. 2-D interaction map between thiosulfate and the WT CO_2 binding pocket based upon observed distances in the refined model. Green dashed lines represent hydrogen bonding, while blue spheres represent bound water molecules. Distances are reported in Å.

2.6 Mutations in the CO_2 binding pocket alter ligand specificity

2.6.1 Sidechain-based ligand specificity

From the hydrogen bonding observed with thiosulfate binding in the wild type along with CO_2 binding, we hypothesized that altering the hydrogen bonding network would impart different ligand specificity to the active site of EcPCK. Mutagenesis was performed on previously identified residues responsible for hydrogen bonding in the CO_2 pocket of the native enzyme, R65 and Y207, where Arg65 was mutated to His, Lys, and Gln, and Tyr207 was mutated to Phe, Ile, Lys, Leu, and Asn. The majority of the crystal structures of these mutants did not indicate any presence of thiosulfate in the CO_2 pocket except for the Y207F structure, which is characterized in Section 2.6.2.

Next, instead of directly altering the H-bonding network through changing the existing hydrogen-bonding residues, additional H-bonding capability was conferred through mutation of residues in close proximity to R65 and Y207. A loop spanning T63-K70, which includes hydrogen bonding residue R65, and an α-helix immediately following Y207 from G209-F216 were mutated with 5-10 alternative amino acids in multiple positions in these regions. Several of these mutations were found to bind to nonnative ligands underscoring the role of hydrogen bonds in specificity. These are described in more detail below.

2.6.2 Nonnative ligand binding in the Y207F mutant of EcPCK

The Y207F point mutant was found to bind thiosulfate. As in the WT case, there was thiosulfate found at the ATP binding site. However, the content of the CO_2 pocket varied depending on the identity of the protein chain. Chain A of the Y207F mutant appears to contain OAA, while chain B contains thiosulfate. The crystals of Y207F EcPCK were in the $P2_12_12_1$ space group with unit cell parameters a = 94.63 Å, b = 105.78 Å, c = 120.35 Å and contained 2 copies per ASU. Reflections were observed to 1.46 Å with 209245 unique reflections and 98.7% overall completeness. The first 6 residues of the N-terminus did not have appreciable density and were not modeled in, and density for the 6xHis-tag at the C-terminus was not seen either. Statistics for this dataset are presented in Table 2.2 located at the end of Section 2.6 showing the good quality of the data.

Density resembling oxaloacetate was found near the CO_2 binding pocket of Chain A of the EcPCK Y207F mutant (Figure 2.4). Previously, there has only been one structure of EcPCK with OAA bound, PDB ID 2PXZ. In that model, OAA is located away from the CO_2 binding site and the ATP binding site, towards the base of the Ω-loop (residues 386-406). Yet, this prior structure does not agree with existing knowledge that the $-CO_2^-$ group from oxaloacetate is decarboxylated, since that should place part of OAA in the CO_2 pocket. Density in 2PXZ around the modeled OAA ligand was also not convincing of OAA presence.

OAA binding was also reported in the rat cytosolic PCK (PDB IDs 2QF1 and 2QF2). In those structures, the OAA was coordinated to the Mn^{2+} ion. In 2QF1, OAA was found in two

positions around the manganese ion, while in 2QF2, the density was modeled with both OAA and pyruvate. The orientation modeled was with the CO_2 end of OAA pointed toward the Arg and Tyr involved in H-bonding of CO_2.

In my Y207F mutant, OAA was found to form hydrogen bond interactions through the OAA O2 atom to NE2 of H232 (2.97 Å), and NZ of K213 (3.03 Å), O3 to NH1 of R333 (3.08 Å), and O4 to R65 NE (2.92 Å). The orientation of OAA differs from what was observed in the rat cytosolic model: the $-CO_2^-$ portion that gets removed during the reaction is distal to the two residues involved in H-bonding of CO_2, R65 and Y207. This is most likely explained by the loss of H-bonding capability at the residue 207 position through the Tyr to Phe mutation, forcing OAA to adopt a different conformation to maintain H-bonding through its carbonyl groups through alternate sidechains on the protein. There is no evidence to suggest coordination of Mn^{2+}, even though the mutation is not directly involved in the coordination sphere and Mn^{2+} is present in the crystallization condition.

As previously mentioned, the thiosulfate in the CO_2 binding site of Y207F Chain B is not in the same position as in the WT enzyme, and can be seen in Figure 2.5. While the ligand is near R65, it is no longer interacting with the Phe sidechain now at the residue 207 position. Hydrogen bonding between the ligand and the protein is through R65, H232, S250 and R333 and is depicted in Figure 2.6. NE2 of R65 H-bonds to O01 of THJ4 at a distance of 3.20 Å, the amide N of S250 to O03 2.43 Å away, and OG of S250 to the thio S of THJ 2.67 Å away. O04 of THJ interacts with NE2 of H232 (3.08 Å). Finally, THJ interaction with R333 is through a bound water molecule. Because of the lack of hydrogen bond donors at residue 207, compensatory H-bonds are made with other sidechains, altering the position and orientation of the thiosulfate ligand.
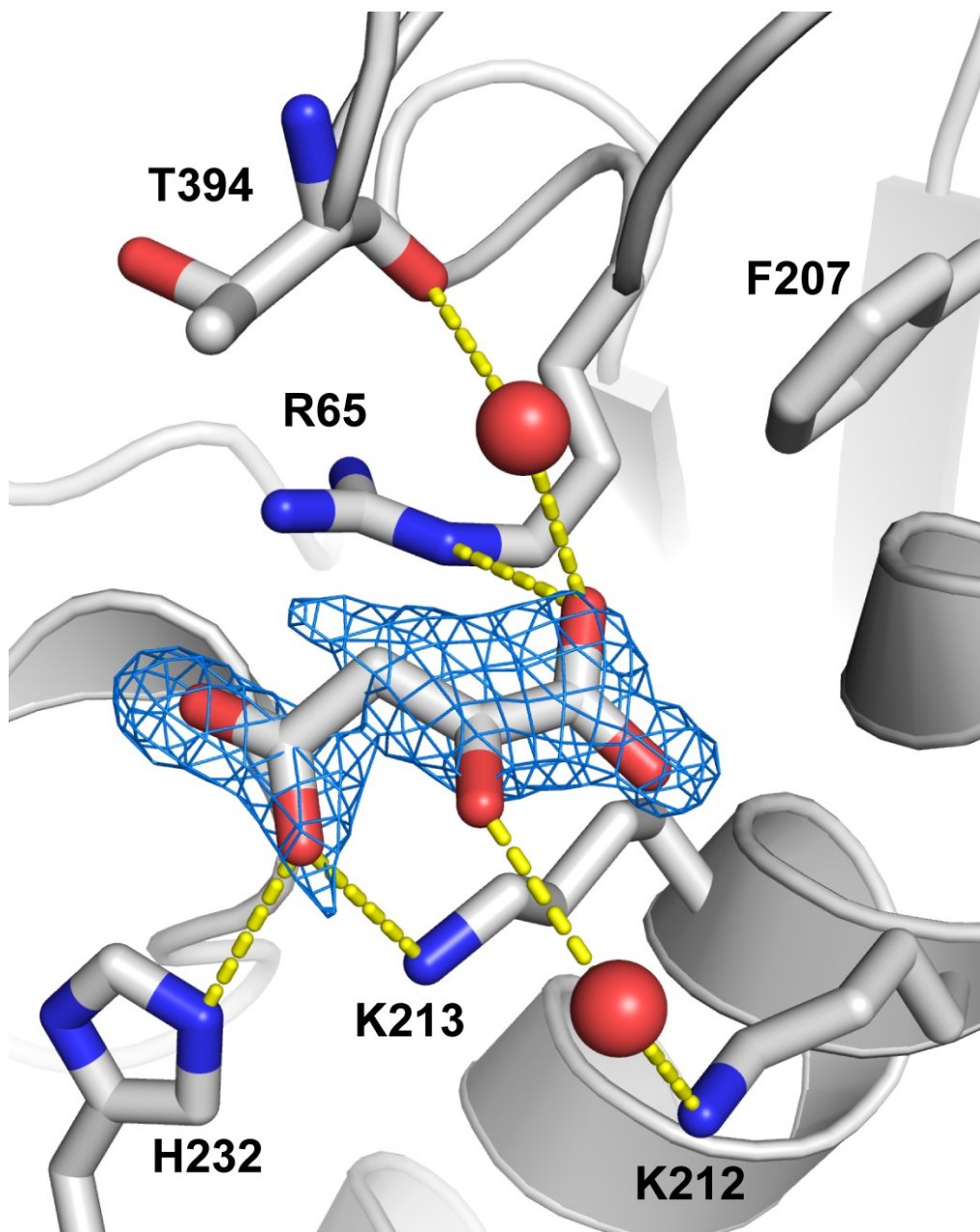
Figure 2.4. Oxaloacetetate binding in CO_2 pocket of EcPCK Y207F mutant. Interactions between the bound ligand oxaloacetate and sidechains are illustrated. The 2mFo-DFc composite omit electron density maps, shown in blue and contoured at 2.0 sigma, surround the ligand in the pocket. Yellow dashed lines represent hydrogen bonds. Red spheres indicate ordered water molecules.
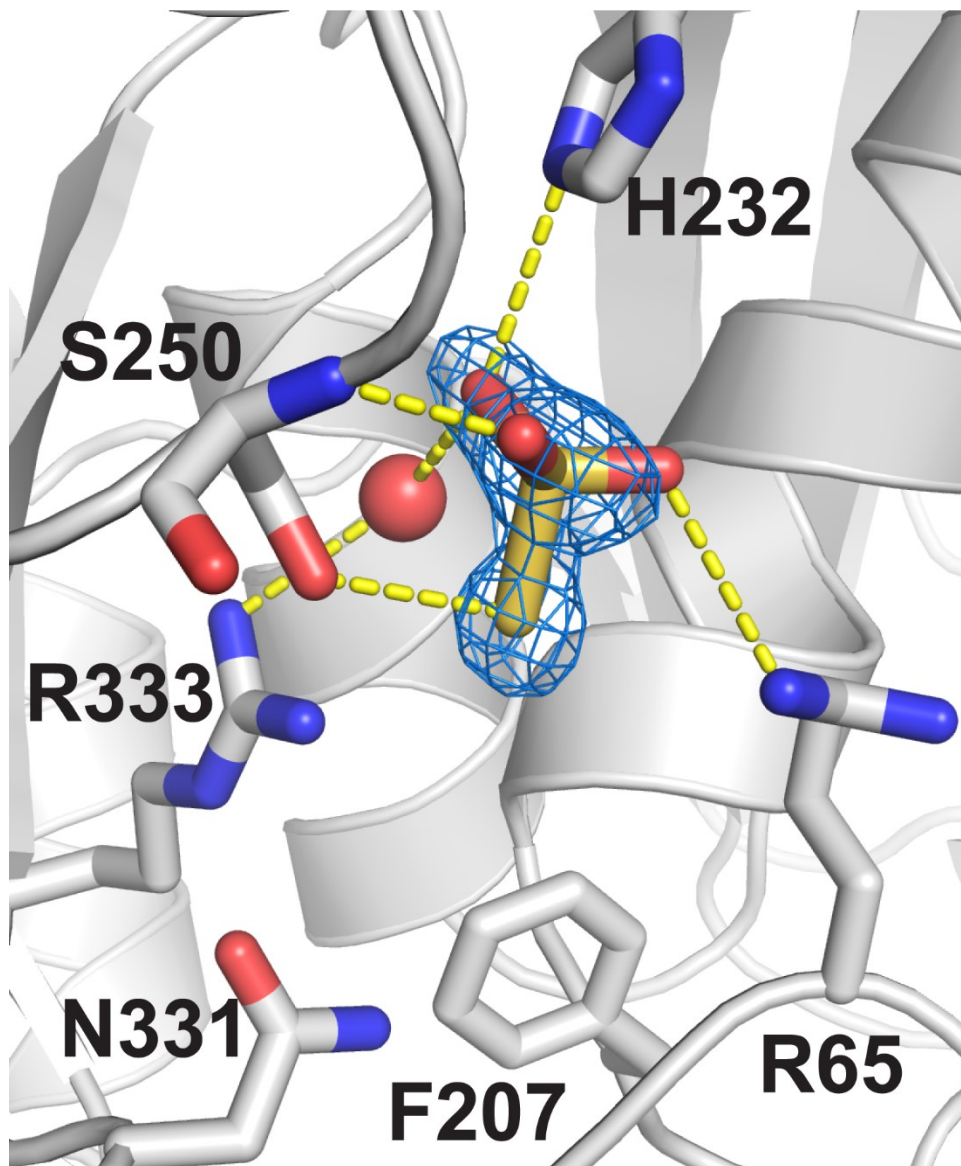
Figure 2.5. Detailed model of the Y207F EcPCK CO_2 binding pocket with thiosulfate bound. Interactions between the bound nonnative ligand thiosulfate and sidechains are illustrated. The 2mFo-DFc composite omit electron maps, shown in blue and contoured at 2.0 sigma, surround the nonnative ligand in the pocket. Yellow dashed lines represent hydrogen bonds. Red spheres indicate ordered water molecules.
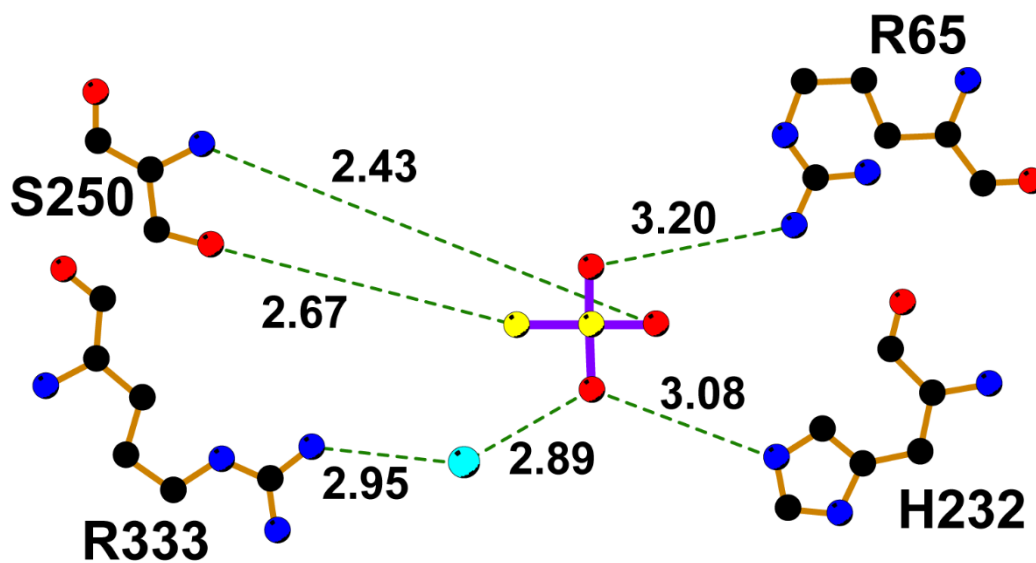
Figure 2.6. 2-D interaction map between thiosulfate and the Y207F CO_2 binding pocket based upon observed distances in the refined model. Green dashed lines represent hydrogen bonding, while blue spheres represent bound water molecules. Distances are reported in Å.

2.6.3 G209N point mutant binds to thiosulfate as well

The EcPCK point mutant G209N was crystallized in the presence of thiosulfate, ADP, PEP, Mg^{2+}, and Mn^{2+}. The resulting crystals with one molecule per ASU were of the C2 space group with a = 125.53 Å, b = 94.66 Å, c = 46.61 Å, β = 96.439°, and diffracted to 1.44 Å with 182990 reflections, resulting in 96.3% overall completeness. The first five N-terminal residues were not observed in the electron density, but 2/6 residues of the C-terminal 6xHis tag were visible. Electron density for a short peptide lying on a special position around residue 25 was also seen. Due to its position away from the active site, it is unlikely to be involved in the catalytic reaction and was not modeled. Statistics can be found in Table 2.2.

Electron density for THJ was identified in the CO_2 binding site (Figure 2.7), while in the nucleotide binding site, ATP, Mg^{2+}, and Mn^{2+} were observed. Occupancy of THJ in this mutant is much higher than in the WT, at 83% occupancy. The THJ molecule in the CO_2 binding site is in a different position and orientation compared to THJ in the wild type owing to the bulkier Asn sidechain now at position 209. Hydrogen bonding of THJ to Y207 is illustrated in Figure 2.8, and is mediated through a water molecule, while R65 and N209 directly H-bond to THJ. The H-bond distance between R65 N and THJ O3 is 2.78 Å, while the G209N ND2 to THJ O1 is 3.02 Å. The larger sidechain at the 209 residue position pushes THJ away from the position it was found in for the WT crystal, and causes THJ O2 to interact with N331 ND2 (3.17 Å). N331 is also in the CO_2 binding pocket but is not involved in WT H-bonding to thiosulfate. Finally, three other sidechain interactions are mediated through bound water molecules, Y207, S250, and R333. The sulfur atom on thiosulfate in this mutant is pointed toward the active site cleft and the C-terminal lobe instead.

ATP was found to be the nucleotide bound in the active site, even though ADP and PEP were added, indicating that this mutant is able to catalyze the first step of the reverse reaction: phosphotransfer from PEP to ADP to generate ATP. The conformation of ATP in this mutant resembles those previously characterized as pound to PCK. Although some of the bond lengths and bond angles deviate from prototypical ATP, this is to be expected since ATP in EcPCK was previously found to be in the less common syn conformation.
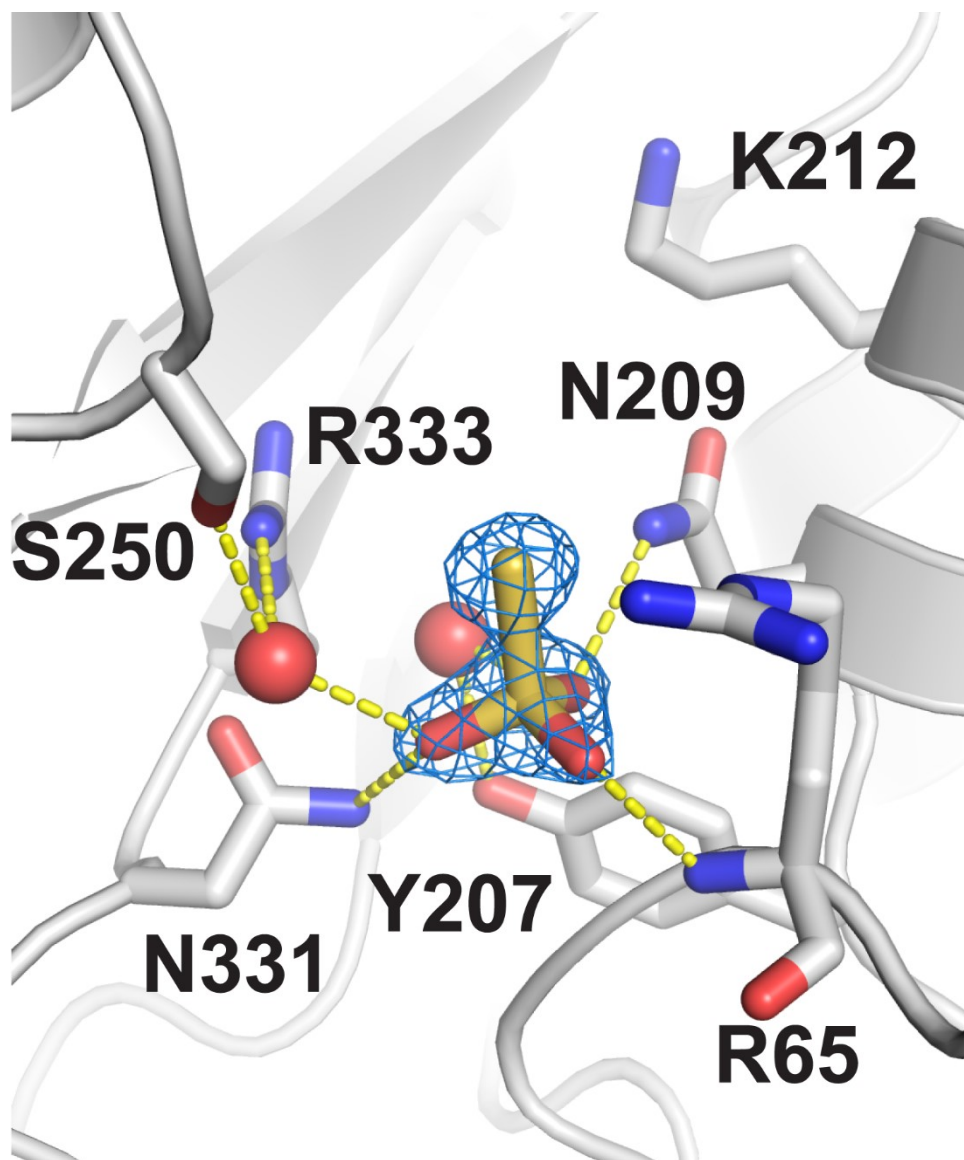
Figure 2.7. Detailed model of the G209N EcPCK CO_2 binding pocket with thiosulfate bound. Interactions between the bound nonnative ligand thiosulfate and sidechains are illustrated. The 2mFo-DFc composite omit electron density maps, shown in blue and contoured at 2.0 sigma, surround the nonnative ligand in the pocket. Yellow dashed lines represent hydrogen bonds. Red spheres indicate ordered water molecules.
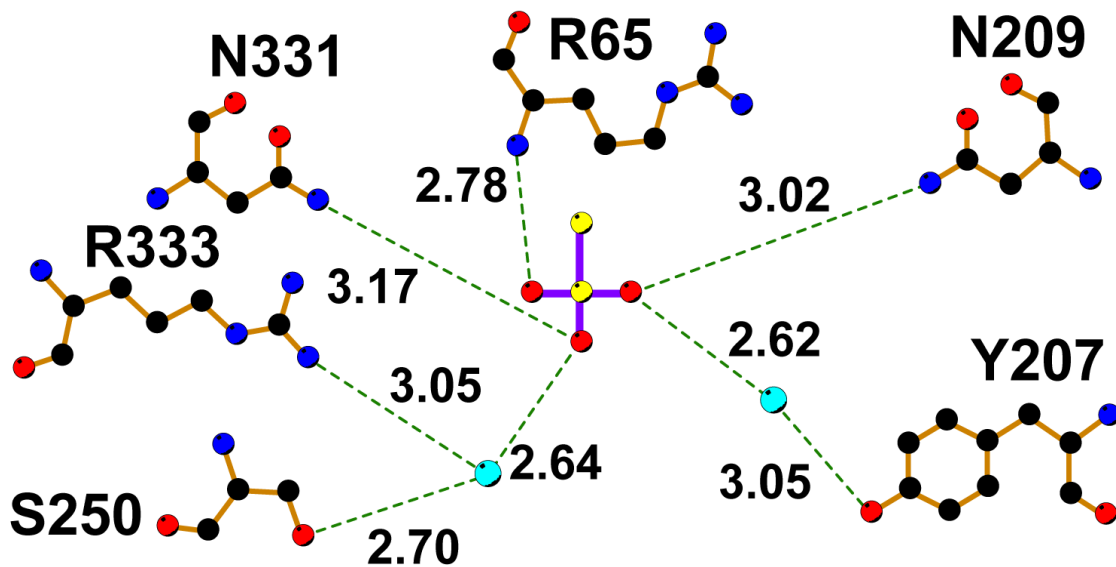
Figure 2.8. 2-D interaction map between thiosulfate and the G209N CO_2 binding pocket based upon observed distances in the refined model. Green dashed lines represent hydrogen bonding, while blue spheres represent bound water molecules. Distances are reported in Å.

2.6.4 G209S mutation changes ligand specificity to methanesulfonate

The G209S mutant crystallized under identical conditions to G209N, giving rise to crystals also with one copy per ASU that are very similar to G209N, with space group C2 and a = 125.84 Å, b = 93.66 Å, c = 46.56 Å, β = 96.234°, diffracting to 1.79 Å with 203413 reflections and an overall completeness of 92.4%. The structure is well-ordered except for the two termini. The first 8 residues of the N-terminus were too disordered in the experimental electron density to be built into the model, while density for only 3/6 of the C-terminal 6xHis tag could be seen. Additional statistics can be found in Table 2.2.

Density for the methanesulfonate ligand can be seen in the CO_2 pocket of this model (Figure 2.9). Post-refinement, the occupancy of the ligand comes to 87%. Whereas density for thiosulfate shows a large sphere for the sulfur atom attached to the sulfonate moiety, the methanesulfonate ligand has a much smaller patch of electron density around the methyl group, agreeing with the smaller number of electrons in a methyl group compared to a sulfur atom, as well as the electronic properties of methanesulfonate (i.e. the electron-withdrawing nature of the oxygen atoms on the sulfonate moiety and the polarization of the C-S bond). This aspect of the density allowed determination of the methanesulfonate ligand orientation in the binding site. H-bonding of the ligand once again involves R65 along with the serine mutation, and is depicted in Figure 2.10. We observe the following interactions: O03 hydrogen bonding to NE of R65 (3.04 Å) and OG of G209S (2.78 Å), O04 bonding to a water molecule in the Mn^{2+} coordination sphere (2.64 Å), and O05 interacting with NH2 (3.17 Å) of R65 and another water in the Mn^{2+} coordination sphere. In this mutant, the methyl group of the methanesulfonate points along the direction of the cleft running between the two lobes.

ATP binding was also observed in this mutant, with the molecule and residues involved in Mg^{2+}/Mn^{2+} coordination overlaying well with the wild type structure. As in the G209N mutant, cocrystallization was performed with ADP and PEP, indicating that the phosphotransfer functionality of the G209S mutant is intact as well (the reverse of the mechanism shown in Figure 1.2).

G209S mutant crystallized in the presence of thiosulfate yielded crystals with an empty CO_2 pocket (data not shown), indicative of the ligand selectivity exhibited by this enzyme. Efforts to crystallize WT and G209N structures with methanesulfonate bound also yielded crystals with empty CO_2 binding pockets.
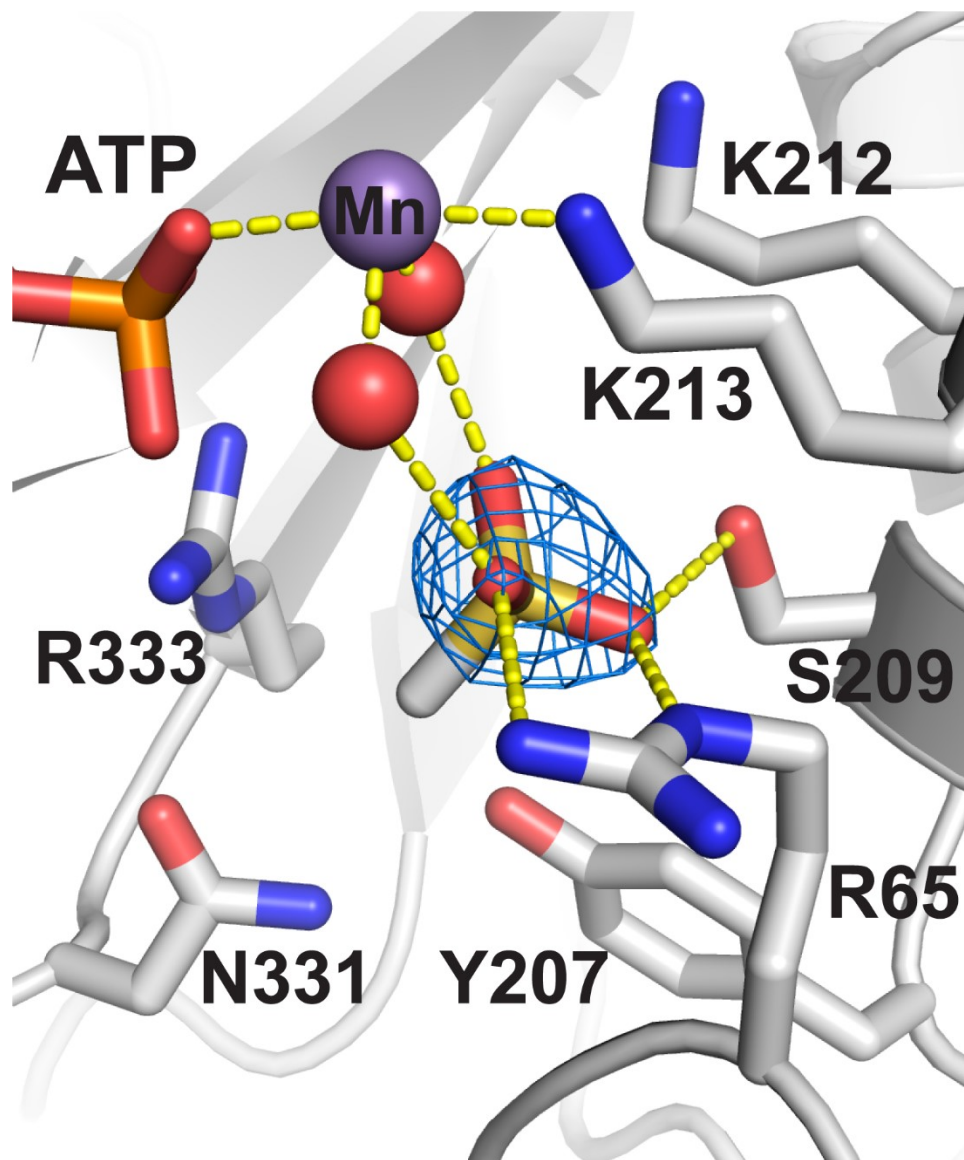
Figure 2.9. Detailed model of the G209S EcPCK CO_2 binding pocket with methanesulfonate bound. Interactions between the bound nonnative ligand methanesulfonate and sidechains are illustrated. The 2mFo-DFc composite omit electron density maps, shown in blue and contoured at 2.0 sigma, surround the nonnative ligand in the pocket. Yellow dashed lines represent hydrogen bonds. Red spheres indicate ordered water molecules. Purple sphere represents the manganese ion. Two ligands in the Mn coordination sphere are omitted for clarity and are shown in Figure 2.13.
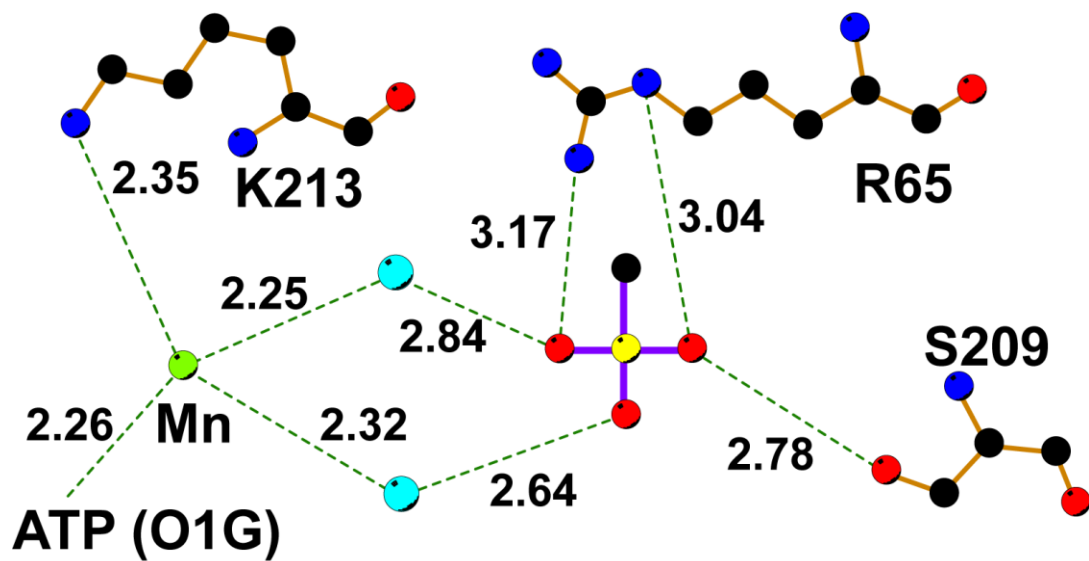
Figure 2.10. 2-D interaction map between thiosulfate and the G209S CO_2 binding pocket based upon observed distances in the refined model. Green dashed lines represent hydrogen bonding, while blue spheres represent bound water molecules. ATP (O1G) represents the O1G atom on ATP which is coordinated to Mn^{2+} (green sphere) and whose structure is not shown. Distances are reported in Å.

2.6.5 G209S K212C double mutant reverts specificity back to thiosulfate and shows a change in manganese (II) coordination

Crystals of the G209S K212C double mutant were obtained under the same conditions as the previous mutants, leading to a crystal with 1 copy/ASU in the C2 space group. This double mutant was crystallized in the presence of thiosulfate and methanesulfonate in two separate crystal trials, but only the thiosulfate condition yielded electron density in the binding pocket of interest (Figure 2.11). Density for the model's N-terminus began at Thr8, and 3/6 of the residues in the C-terminal 6xHis-tag were seen.

ATP binding in this mutant had one change: the Mn^{2+} coordination sphere was no longer octahedral as previously observed: K213 is now pointing away from Mn^{2+} and instead hydrogen bonds to the nonnative ligand thiosulfate. The resulting manganese coordination sphere now has five members in a square pyramidal geometry.

The addition of the K212C mutation to the G209S active site reverted the ligand selectivity back to thiosulfate, even though the single G209S mutant had methanesulfonate bound. Occupancy of thiosulfate after refinement was 0.72. Binding of thiosulfate in this mutant was achieved through hydrogen bonding between NE of R65 to O01 (2.95 Å) of THJ, OG of S209 (2.57 Å) and O04 of THJ, and between NZ of K213 and O03 of THJ (2.92 Å) (Figure 2.12). S05 interacts with 2 bound water molecules, one of which is in the Mn^{2+} coordination sphere. O03 of THJ is also interacting with another water in the Mn^{2+} coordination sphere. As such, the thio S of thiosulfate points toward ATP in the active site.

The G209S K212C mutant influenced the ATP binding region. In the G209S crystal structure (Figure 2.13a) as well as the wild type structure, Mn^{2+} has an octahedral coordination sphere consisting of an oxygen atom from the γ-phosphate of ATP, Lys213, His232, Asp269 and 2 water molecules, while in the mutant (Figure 2.13b), Lys213 is no longer coordinating. Instead, Lys213 is now hydrogen bonding to a bound water molecule positioned where the amino group on the Lys would be. Instead of the previously seen octahedral coordination sphere, a coordination sphere with a square pyramidal configuration was observed. Although it seems like the other oxygen on the sidechain of Asp269 could complete the octahedral sphere, the 2.81 Å distance between that oxygen and the Mn^{2+} does not suggest that coordination between those two atoms is occurring.

Comparing the Mn coordination sphere in EcPCK to two other manganese-containing enzymes, manganese superoxide dismutase (MnSOD) and Mre11, highlights differences in manganese ion coordination. MnSOD cycles through Mn^{2+} and Mn^{3+} states (66), and the structure reveals a drastically different coordination sphere for the Mn^{3+} ion (67). The Mn^{3+} coordination sphere in PDB ID 1N0J is organized in a trigonal bipyramidal geometry, with 3 His, 1 Asp, and 1 water completing the coordination sphere. While both MnSOD and the G209S K212C mutant of EcPCK contain pentacoordinate manganese ions, the geometry difference between the two suggests the EcPCK manganese retains its 2+ oxidation state since the double mutant largely retains its octahedral characteristic rather than the trigonal bipyramidal organization. Looking at manganese binding in Mre11 (68), where it is also in the 2+ oxidation state, the two Mn^{2+} ions coordinate to His, Asp, and Asn residues along with water molecules. Neither Mn^{2+} coordination sphere in Mre11 seems to be complete: while the ligands are

organized in an octahedral geometry, there are only 4 or 5 ligands instead of the expected 6. The geometry of the Mn coordination sphere in EcPCK has more ideal geometry than the coordination spheres of Mn in MnSOD and Mre11. From the coordination spheres of MnSOD and Mre11, Lys coordination to Mn does not seem to be a common occurrence, so the EcPCK Mn coordination sphere should provide novel data of interest to researchers dealing with manganese-bound enzymes.

Crystals of this particular mutant were derivatized with Xe gas to probe for possible hydrophobic pockets in the enzyme. A xenon atom was found, but not in the active site cleft. Figure 2.14a highlights the binding site of the xenon atom relative to the overall structure. It is located in a hydrophobic pocket in the C-terminal domain from the opposite side of the active site. The pocket is formed by two α-helices consisting of mainly hydrophobic residues. Figure 2.14b shows the local environment of the Xe atom and explicitly shows the hydrophobic sidechains located within 5 Å of this heavy atom.
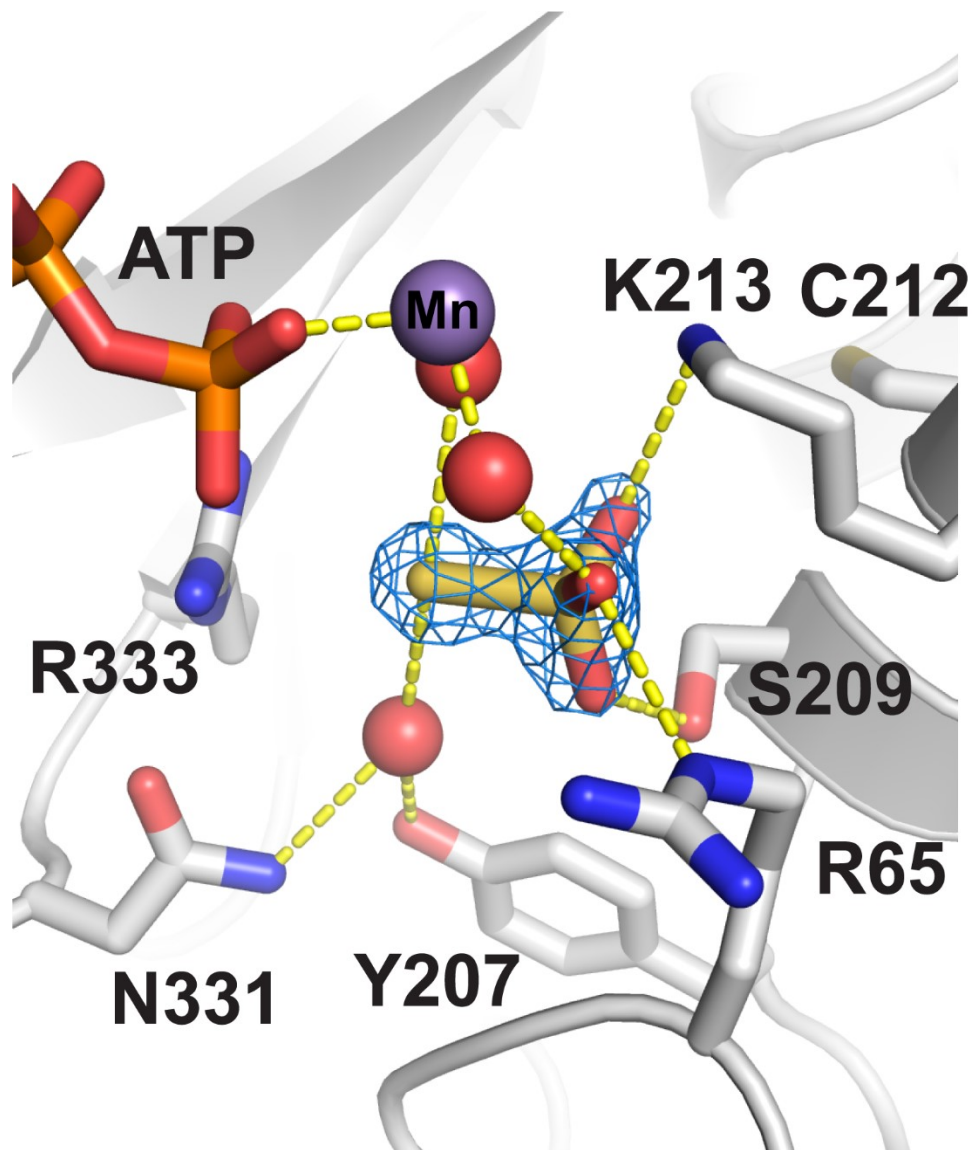
Figure 2.11. Detailed model of the G209S K212C EcPCK CO_2 binding pocket with thiosulfate bound. Interactions between the bound nonnative ligand thiosulfate and sidechains are illustrated. The 2mFo-DFc composite omit electron density maps, shown in blue and contoured at 2.0 sigma, surround the nonnative ligand in the pocket. Yellow dashed lines represent hydrogen bonds. Red spheres indicate ordered water molecules. Purple sphere represents the manganese ion.
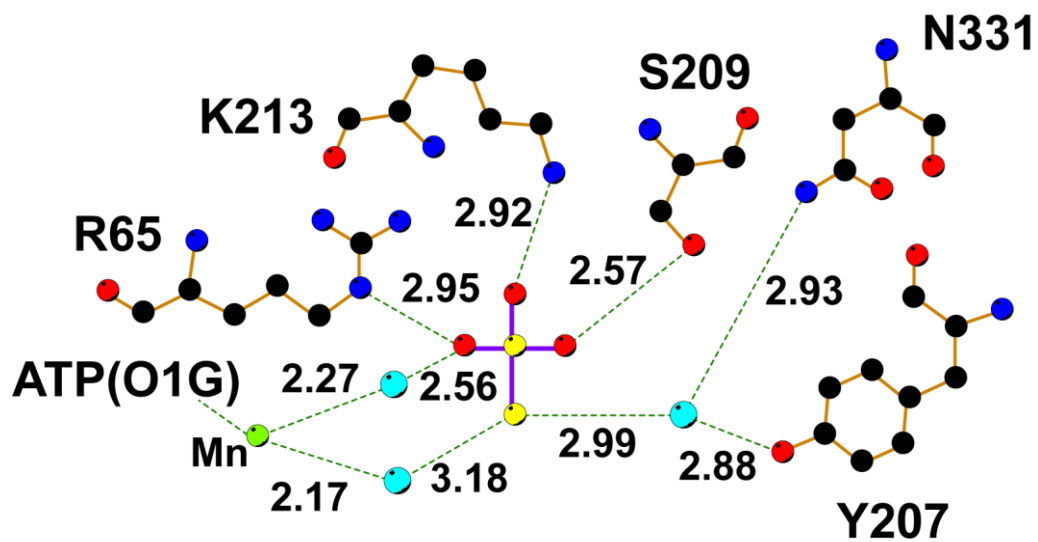
Figure 2.12. 2-D interaction map between methanesulfonate and the G209S K212C CO_2 binding pocket based upon observed distances in the refined model. Green dashed lines represent hydrogen bonding, while blue spheres represent bound water molecules. ATP (O1G) represents the O1G atom on ATP which is coordinated to Mn^{2+} (green sphere) and whose structure is not shown. Distances are reported in Å.
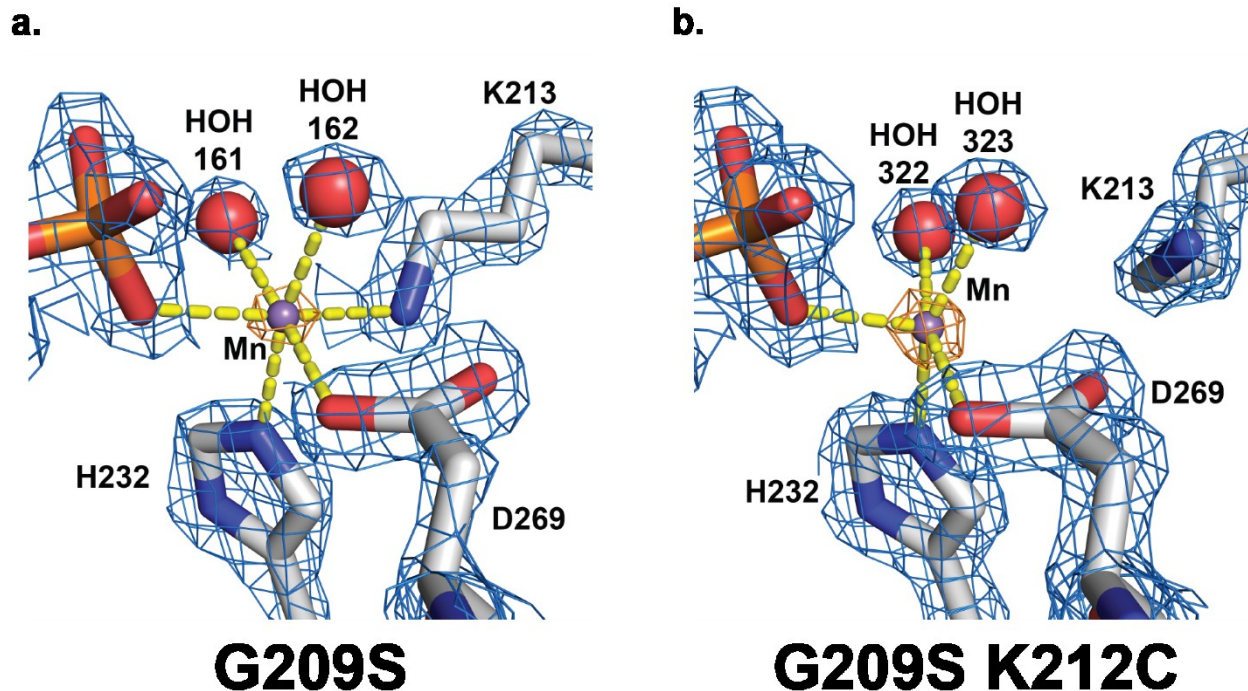
Figure 2.13. Manganese (II) coordination sphere in EcPCK is altered by mutations. Coordination spheres for manganese (II) in EcPCK are shown along with electron density for the ligands (blue, contoured at 2.0 sigma, 2mFo-DFc map) and the manganese ion (orange, contoured at 12.0 sigma, 2mFo-DFc map). a. G209S Mn^{2+} octahedral coordination sphere, identical to previously solved structures of WT EcPCK. The ligands include 3 sidechains, Lys213, His232, and Asp269, along with the γ-phosphate of ATP, and 2 bound water molecules. b. Mn^{2+} coordination sphere in the G209S K212C double mutant shows one major difference from WT behavior: K213 is no longer coordinated to Mn^{2+}, leaving a 5-coordinate sphere in a square pyramidal configuration.
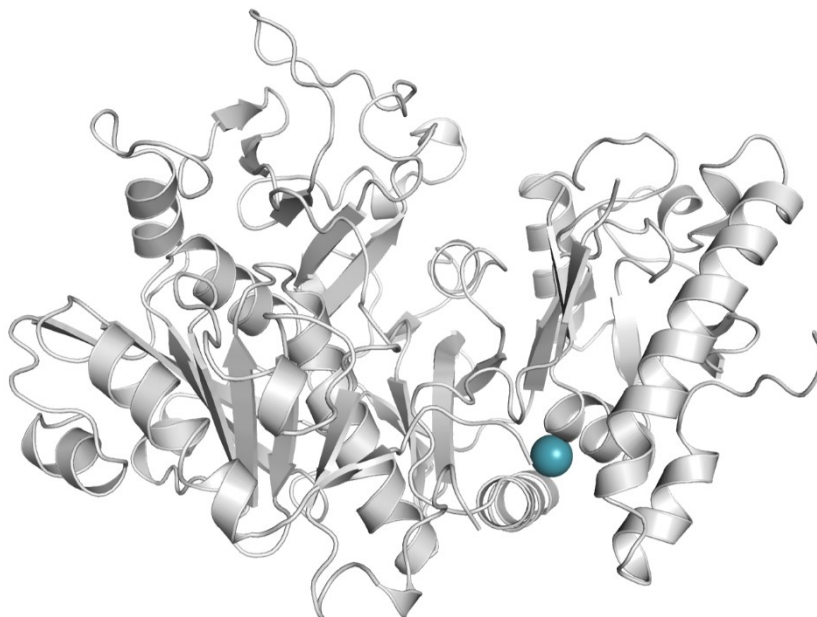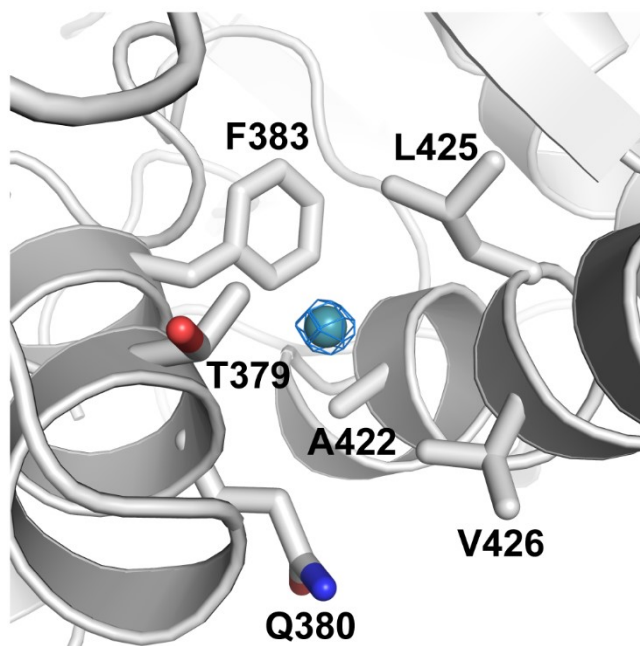
Figure 2.14. Xenon binding pocket in EcPCK. a. Location of the Xe binding pocket relative to the overall structure of EcPCK. Xe is shown as the teal sphere. b. Close-up view of the Xe binding pocket showing the predominantly hydrophobic sidechains within 5 Å of Xe. Blue electron density mesh shown around Xe is contoured to 20 sigma.

2.6.6 Computation prediction of binding in other mutants

All mutants that bound to the nonnative ligands thiosulfate and methansulfonate thus far have involved mutations at the G209 position. The involvement of other residues could also lead to nonnative binding. However, expanding the mutational search space to span all mutations in the three regions of interest would not be feasible experimentally. To increase throughput allowing more comprehensive consideration of mutational space, mutants were screened *in silico* using molecular docking, molecular dynamics (MD) and Molecular Mechanics-Generalized Born Surface Area (MM-GB/SA) methods. Due to the small size of thiosulfate and methansulfonate, using only traditional docking and/or single point MM-GB/SA calculations cannot make reliable predictions. To improve prediction accuracy, we implemented MD-based MM-GB/SA rescoring, which calculates the average binding free energy of each mutant-ligand complex against the entire MD simulation trajectory, thus taking into account receptor flexibility. In addition, we also implemented ligand-receptor interaction energy analysis in order to overcome the difficulty associated with small ligand size. Using such a cross-screening approach, we ranked mutants by their binding scores and interaction energies to the nonnative ligand methanesulfonate (see Table 2.1). One of the top computational hits, the double mutant K212I F216V, was identified and showed crystallographic evidence of methanesulfonate binding and is discussed in the following section. In this case, the binding score alone would not have ranked double mutant K212I F216V high enough to suggest experimental testing. Its calculated average binding free energy is actually slightly less favorable than G209S, but its interaction energy is slightly more favorable than G209S. Therefore, considering both scores together would have ranked this double mutant as one of the top predictions.

System	Average binding free energy (kcal/mol)	Interaction energy (kcal/mol)
G209S	-13.3	-147.6
G209S K212C	-14.9	-133.2
H67Q	-14.1	-132.4
K212V F216S	-13.6	-133.3
*K212I F216V	-12.9	-154.4
H67E G209S	-12.2	-143.1

Table 2.1. Top ranked mutants in terms of methanesulfonate binding. G209S was listed on top as a reference. The other systems were sorted based on their average binding free energies. *K212I F216V is one of three mutants that has both average binding free energy and interaction energy predicted to be comparable with G209S and has crystallographically validated binding. 1 kcal = 4.184 kJ.

2.6.7 K212I F216V double mutations away from the pocket enable methanesulfonate binding

The K212I F216V double mutant predicted to bind methanesulfonate was crystallized under a different condition than the other mutants; use of the previous Bis-Tris condition yielded fine needles that were unsuitable for diffraction studies. This new condition yielded plates rather than rod-like crystals resulting from the other condition. Statistics are presented in Table 2.2. There are two molecules in the ASU as in the wild type, but the space group for this mutant is $P2_1$. The first 6 N-terminal residues, and the last 4 histidine residues of the C-terminal 6xHis-tag were disordered and not modeled.

Although this mutant was crystallized in the presence of ADP, PEP, Mg^{2+}, and Mn^{2+}, none of these ligands or cofactors were present in the electron density. From the difference map, SO_4^{2-} was found to bind to Thr255 in the ATP binding region. This is due to the change in crystallization conditions to one that contains ammonium sulfate as a precipitant. The protein maintained an overall open position that was consistent with the wild type apo form (PDB ID 1OEN). The opening and closing studies with various ligands, sulfonate compounds in particular, are further discussed in Chapter 3.

Refinement was first performed with a protein-only model; no ligands were present to prevent phase bias. Density consistent with a ligand containing four atoms arranged about a fifth central atom was found in the CO_2 binding pocket (Figure 2.15). Two possibilities for the identity of the ligand exist: 1) methanesulfonate, the predicted ligand, and 2) sulfate, from the crystallization condition. Several factors led us to believe that the electron density represented methanesulfonate and not sulfate. First, a sulfate molecule would have tetrahedral symmetry, since all four oxygens are equivalent; there is a clear difference in one of the four substituents attached to the sulfur according to the electron density. Next, two separate refinements were performed with the only difference being the identity of the ligand. The mFo-DFc map of the protein-methanesulfonate complex had fewer difference peaks than that of the protein-sulfate complex. Finally, looking at the B-factors of the group on the sulfonate molecule (the methyl group in the case of methanesulfonate and oxygen in the case of sulfate), the B-factor for the methyl group was more in line with the B-factors for the rest of the ligand as well as atoms in the side chains of the surrounding residues (all ligand atoms had a B-factor of around 30). When inspecting the B-factor of the fourth sulfur on the hypothetical sulfate, the B-factor was 38, significantly higher than adjacent atoms. Because of all these points, we believe that methanesulfonate is the ligand bound, even with the much higher concentration of sulfate present in the condition.

Methanesulfonate bound in the K212I F216V mutant with an occupancy of 0.77, slightly lower than what had been seen for some of the other mutants. Hydrogen bonding is observed for two of the three oxygen atoms on the sulfonate group, O04 to NE of R65 with a distance of 2.85 Å, and O05 to OH of Y207 (2.73 Å) (Figure 2.16). Such a hydrogen bonding pattern is also observed in computationally predicted model, with a distance of 2.92 Å between O04 and NE of R65 and a distance of 2.75 Å between O05 and OH of Y207. O05 also interacts with a water bound by N of R65. O03 is not within hydrogen bonding distance to either the protein itself or any ordered solvent. While ligand binding in the other mutants involved all three oxygen atoms on the sulfonate moiety, only 2/3 of the oxygen atoms are involved here. This might be a reason

why the density is not as clear as in the other cases: the interaction between the ligand and protein is not as strong.
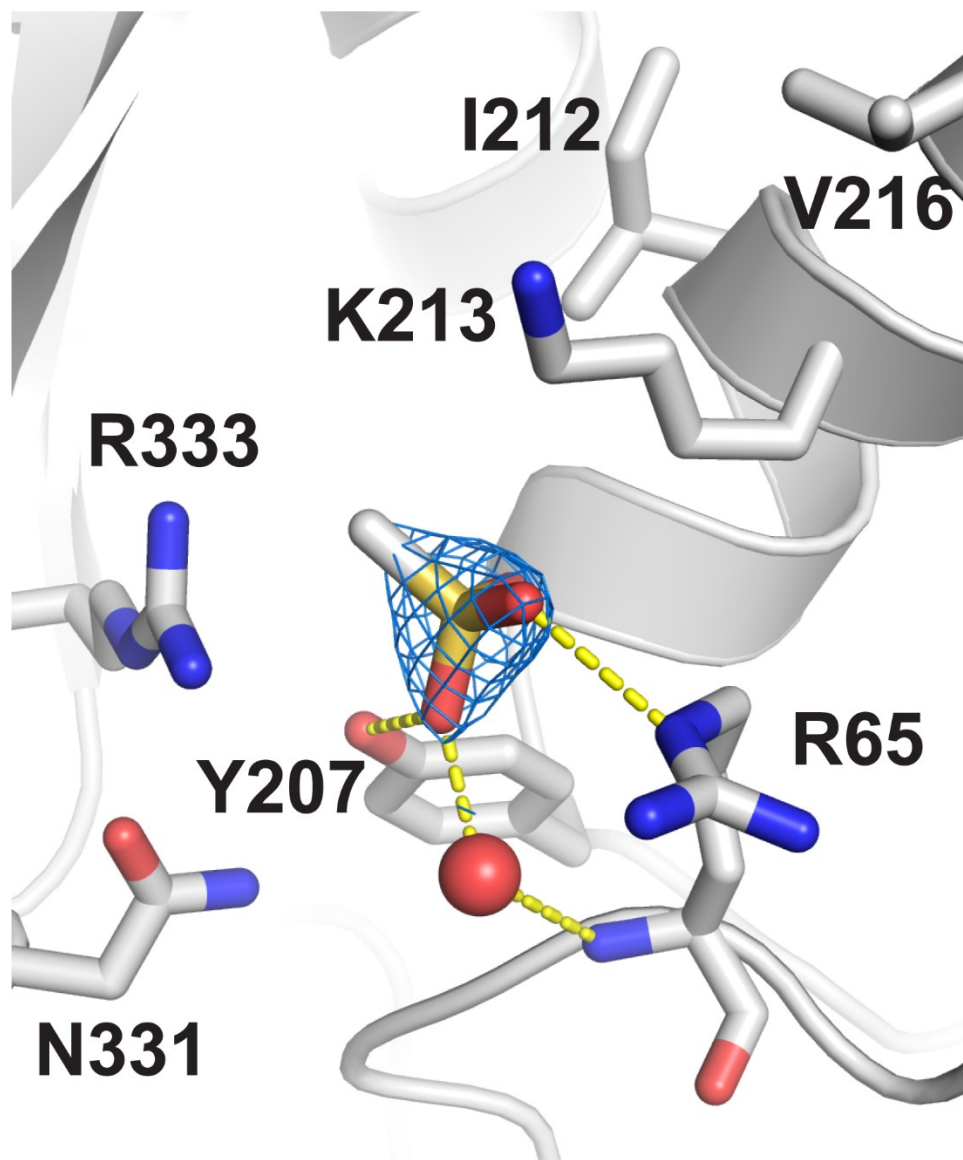
Figure 2.15. Detailed model of the K212I F216V EcPCK CO_2 binding pocket with methanesulfonate bound. Interactions between the bound nonnative ligand methanesulfonate and sidechains are illustrated. The 2mFo-DFc composite omit electron density maps, shown in blue and contoured at 2.0 sigma, surround the nonnative ligand in the pocket. Yellow dashed lines represent hydrogen bonds. Red spheres indicate ordered water molecules.
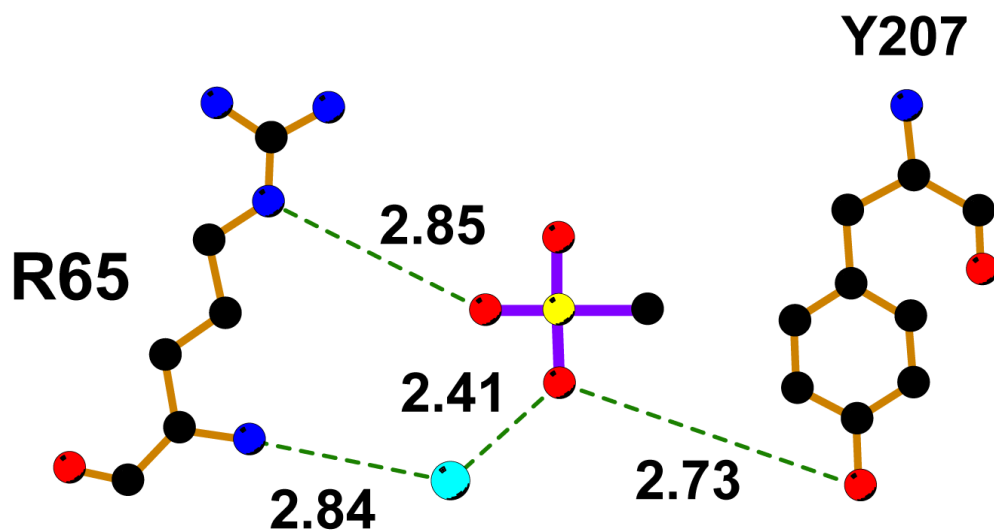
Figure 2.16. 2-D interaction map between methanesulfonate and the K212I F216V CO_2 binding pocket based upon observed distances in the refined model. Green dashed lines represent hydrogen bonding, while blue spheres represent bound water molecules. Distances are reported in Å.

	EcPCK WT	Y207F	G209N	G209S	G209S K212C	K212I F216V
Resolution range	78.19 - 1.33 (1.38 - 1.33)	47.33 - 1.46 (1.51 - 1.46)	46.27 - 1.44 (1.50 - 1.44)	46.66 - 1.79 (1.85 - 1.79)	47.04 - 1.50 (1.55 - 1.50)	47.3 - 1.55 (1.61 - 1.55)
Space group	P 21 21 21	P 21 21 21	C 1 2 1	C 1 2 1	C 1 2 1	P 1 21 1
Unit cell	94.72 103.61 119.16 90 90 90	94.67 105.86 120.49 90 90 90	125.17 94.36 46.56 90 96.40 90	125.31 93.31 46.53 90 96.23 90	125.71 94.08 46.60 90 96.40 90	60.84 75.61 71.53 90 94.70 90
Total reflections	1645290 (56741)	1610196 (152982)	694878 (65420)	203399 (19707)	306973 (13026)	632781 (32215)
Unique reflections	243952 (14564)	209245 (20200)	92651 (8788)	48917 (4789)	81504 (5722)	85603 (5170)
Multiplicity	6.7 (3.9)	7.7 (7.5)	7.5 (7.4)	4.2 (4.1)	3.8 (2.3)	7.4 (6.2)
Completeness (%)	0.92 (0.55)	0.99 (0.95)	0.95 (0.91)	0.98 (0.96)	0.95 (0.67)	0.91 (0.55)
Mean I/sigma(I)	19.15 (1.18)	11.73 (1.01)	10.55 (1.30)	25.19 (4.68)	12.29 (0.60)	12.75 (1.04)
Wilson B-factor	16.17	20.82	15.66	17.07	17.38	17.8
R-merge	0.05 (0.99)	0.08 (1.71)	0.13 (1.64)	0.04 (0.30)	0.09 (1.51)	0.11 (1.67)
R-meas	0.06 (1.14)	0.09 (1.84)	0.14 (1.76)	0.05 (0.35)	0.10 (1.95)	0.12 (1.82)
CC1/2	0.999 (0.462)	0.999 (0.491)	0.998 (0.629)	0.999 (0.937)	0.998 (0.216)	0.998 (0.333)
CC*	1 (0.795)	1 (0.812)	0.999 (0.879)	1 (0.984)	1 (0.596)	1 (0.707)
Reflections used in refinement	243938 (14562)	208574 (19858)	92630 (8778)	48915 (4788)	81392 (5622)	85582 (5154)
Reflections used for R-free	1999 (119)	1989 (188)1989 (188)	1999 (190)	1949 (191)	1995 (136)	2007 (121)
R-work	13.9% (31.2%)	19.9% (42.5%)	16.4% (33.4%)	16.6% (22.6%)	17.8% (38.8%)	17.4% (34.6%)
R-free	15.3% (31.0%)	22.1% (44.6%)	19.2% (34.3$)	18.5% (25.5%)	20.8% (41.4%)	19.2% (33.0%)

	EcPCK WT	Y207F	G209N	G209S	G209S K212C	K212I F216V
CC(work)	0.969 (0.760)	0.964 (0.691)	0.970 (0.832)	0.962 (0.919)	0.967 (0.520)	0.966 (0.646)
CC(free)	0.964 (0.718)	0.960 (0.634)	0.966 (0.780)	0.957 (0.911)	0.955 (0.461)	0.953 (0.557)
Number of non-hydrogen atoms	8925	9327	4465	4291	4599	4376
macromolecules	8229	8216	4077	4041	4104	4092
ligands	30	39	46	38	39	20
Protein residues	1070	1072	535	529	529	527
RMS(bonds)	0.009	0.005	0.006	0.005	0.008	0.012
RMS(angles)	1.01	0.81	0.91	0.82	0.98	1.13
Ramachandran favored (%)	97	97.09	97.93	98	98	98
Ramachandran allowed (%)	2.4	2.63	1.88	1.9	1.7	1.7
Ramachandran outliers (%)	0.09	0.28	0.19	0.19	0.19	0.19
Rotamer outliers (%)	0.24	0.47	0.24	0.49	0.7	0.7
Clashscore	1.11	2.9	1.24	1.64	3.46	0.99
Average B-factor	20.12	28.29	19.75	20.76	22.48	23.72
macromolecules	19.6	27.14	19.24	20.72	21.58	23.45
ligands	19.89	32.24	17.94	18.09	18.3	33.82
solvent	26.54	36.95	26.09	21.94	30.94	27.15

Table 2.2. Crystallographic data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses

2.7 Discussion

2.7.1 Altering ligand binding through direct and neighboring mutations

From this work, we showed that even mutations consisting of a single amino acid change can have drastic effects on ligand recognition in proteins. Although point mutations are generally thought of as having minimal effect on the protein structure, this work shows that they can lead to nonnative ligand recognition and binding in the mutated region, and furthermore also have long-range effects on nucleotide binding in EcPCK. Specificity thus can depend upon single residue mutations in the active site cleft. This has implications for evolution, medical health, and protein design for synthetic biology.

The first three mutant EcPCK structures presented in this chapter (Y207F, G209N, G209S) are direct mutations, as they interact with the nonnative ligand through the mutated sidechain. These were rationally designed mutants, with Y207F altering the native hydrogen bonding capability conferred through the Tyr207 sidechain, and G209N and S conferring new H-bonding capability by the addition of a new sidechain at a position where one did not exist previously. Alteration of sidechains in close proximity of the ligand-binding pocket should alter activity, and indeed, we see that in these mutants.

One observation arising from this high-resolution dataset is that the mutations are not necessarily directly involved in ligand binding but still have an impact. We refer to these mutations as "neighboring mutations". The α-helix forming part of the CO_2 pocket seems to be especially prone to impact from these neighboring mutations. The first position is K212. While the G209S point mutant was determined by X-ray crystallography to bind only to methanesulfonate, the G209S K212C double point mutant bound exclusively to thiosulfate. The residue 212 sidechain is not involved directly in hydrogen bonding interactions with the ligand in either of these cases; the 212 position is recessed from the CO_2 binding pocket. In the case of this double mutant, mutation of K212 changes the behavior of K213 which is in the CO_2 pocket: the coordination sphere of the Mn^{2+} cation containing K213 is altered, likely changing the local electrostatics of this pocket. This is further discussed in Section 2.7.3.

The second instance where neighboring mutations affect ligand selectivity is the K212I F216V mutant. Neither of these positions have sidechains that protrude into the CO_2 binding pocket, yet this mutant is capable of binding methanesulfonate. The effects of this mutation would have been hard to predict rationally, suggesting that enzyme engineering efforts purely based on a rational design approach may not effectively cover the mutation space to determine effects of mutations; pairing such rational design efforts with a somewhat random element may therefore be more productive. The following sections delve into greater detail of the factors behind the protein-ligand interaction in EcPCK and its mutants.

2.7.2 Hydrogen bonding networks in the *E. coli* PCK CO_2 binding pocket and their influence on ligand selectivity and pose

Thiosulfate was found to bind weakly in the wild type active site with at least two orientations as determined by the shape of the electron density around the ligand. Addition of a hydrogen bond donor at the G209 position locked down the possible orientations of thiosulfate down to one pose, with a significant increase in occupancy compared to the wild type structure. As has been seen before (69, 70), enzyme engineering is capable of increasing ligand affinity, and this example, shown through crystallography, is a clear indicator that we were able to increase ligand affinity.

The number of hydrogen bonds between the ligand and protein do not seem to directly account for the implied ligand affinity (as measured through occupancy): although thiosulfate in WT has 5 H-bonds, the same number as the methanesulfonate ligand bound in G209S, ligand occupancy in WT is lower (87% versus 73%). Investigating this further, we looked at not only the number of hydrogen bonds formed, but also the disorder of the side chains involved in hydrogen bonding. In WT, in addition to H-bonding to the residues involved in native ligand binding, R65 and Y207, two additional residues are involved, K213 and R333. In the case of methanesulfonate binding in the G209S mutant, the only side chains involved are R65 and the newly-mutated G209S, along with water molecules in the Mn^{2+} coordination sphere. The sidechains on K213 and R333 are among the largest and most flexible among the 20 naturally occurring amino acids, with 24 and 33 rotamers respectively, so affinity could be decreased due to the disorder of those sidechains. It turns out that R333 is rather ordered (B-factor of 10-15) relative to the rest of the protein, but K213 is much more disordered (B-factor of 27 at the terminal amine), consistent with the greater restraints on Arg movements from the large planar guanidinium group in Arg. Ser in the G209S mutant is a much shorter sidechain, and R65 is involved in H-bonding of 2/3 of the oxygen atoms of the sulfonate moiety in methanesulfonate, resulting in a well-ordered sidechain for H-bonding. Along with the tightly-bound water molecules in the coordination sphere, the substituents involved in H-bonding of methanesulfonate seem to be much more fixed in position. The relationship between the B-factors of the H-bonding residues and ligand affinity provides a possibility of engineering increased affinity by reducing sidechain disorder in proteins, perhaps by mutagenesis towards shorter sidechains. For example, shortening Glu or Gln sidechains involved in H-bonding to Asp or Asn, respectively, by removing a methylene ($-CH_2-$) group could increase ligand affinity. However, that could lead to a shift in ligand position and possibly alter orientation as well.

2.7.3 Ligand selectivity and orientation based on binding pocket electrostatics

The effects of ligand selectivity upon the change of one residue in the CO_2 binding pocket is intriguing. Experimental evidence has shown that the EcPCK mutants in this study are surprisingly selective in their ligand binding, capable of binding to only one of either thiosulfate or methanesulfonate, but never both. This experimentally validated observation provides a foundation for further investigations of this phenomenon to inform on more general cases of specificity.

Enzymes have evolved to have high ligand selectivity; the primary basis for this evolutionary selection is that binding of the wrong substrate could lead to catalysis of an undesired reaction causing a loss of energy or toxicity. In the context of the cell, that could have catastrophic consequences. With this in mind, there are many mechanisms to enforce this selectivity. First, sterics play a key role in ligand identification. Addition of something as small as a methyl group can have profound effects on ligand affinity. However, that is not the case here. Although there is a size difference between thiosulfate and methanesulfonate, it is not the substantial difference in binding specificity, especially since we know that we are able to accommodate CO_2 and thiosulfate in the WT pocket.

In addition to altering ligand selectivity, the mutations made in these experiments impact ligand orientation and position as well. In some cases, it is apparent why the mutation is leading to these changes. For example, comparing ligand binding in the G209N and G209S mutants, the thiosulfate ligand in the G209N structure is hydrogen bonding to the Asn sidechain, while methanesulfonate in the G209S structure interacts with the Ser sidechain. Asn is bulkier than Ser, so the sidechain is already occupying the space that was occupied by the ligand as seen in WT and the G209S mutant. This in turn pushes the ligand out further to minimize steric repulsions. At this new position, thiosulfate is capable of H-bonding to N331, a residue that was not involved in ligand binding in WT EcPCK. We explored this further by generating the G209S N331Q mutation to see if extending the carboxamide functionality of residue 331 through the introduction of a methylene spacer would act as a compensatory mutation to bring about ligand binding interactions that bear a greater resemblance to the binding mode seen in the G209N mutant as a test to see whether our prediction would hold true. Unfortunately, crystals of the G209S N331Q double mutant did not show binding of thiosulfate or methanesulfonate (data not shown), illustrating that this complex system can behave in nonintuitive ways.

Analyzing the orientation of thiosulfate across the various mutants showed that the thio S of thiosulfate ($\underline{\text{S}}$-SO_3^{2-}) had a tendency to point toward positively charged groups. This effect likely reflects the electrostatic attraction between the weak negative charge of the aforementioned sulfur atom and the positive charge of various sidechains. The partial negative nature of the sulfur can be justified through resonance structures of the thiosulfate anion as there are resonance structures that place the negative charge on the sulfur. This trend can be seen in the crystal structures of the mutants containing thiosulfate: the anions tend to point towards arginine most often and sometimes lysine. This phenomenon is not observed with methanesulfonate as the methyl group of that anion lacks negative charge.

Finally, ligand selectivity in the K212I F216V double mutant is surprising because neither of these mutations were expected to be able to H-bond to the ligand, and indeed that is the case in the crystal structure. From this mutation, we can infer that electrostatic effects of an intermediate distance away from the binding site have a strong role to play in determining ligand selectivity in this system, and perhaps in many others as well. These data thus suggest that further investigation into the roles of electrostatics in protein-ligand interaction will prove to be extremely useful for future protein engineering experiments. This is furthermore one area that could greatly benefit from additional computational input.

2.7.4 Ligand selectivity due to loosely bound water molecules

Since we can largely rule out sterics, another contributor to ligand recognition for thiosulfate versus methanesulfonate would be the effects of ordered solvent in and around the binding site. Upon ligand binding, the displacement of loosely bound water molecules, defined here as water molecules that bear a low enthalpic cost when hydrogen bonding contacts between the water molecule and protein are removed. Replacement of such loosely bound waters results in a favorable change in free energy: the entropy gained by releasing the water molecules, along with the favorable enthalpic change from new interactions between the protein and ligand, outweigh the enthalpic penalty from removing loosely bound water molecules. Notably, the existence of electron density for a bound water molecule defines the localization of bound water on the protein surface but is not in itself an indication of high stability but rather shows a site more favorable than other sites in that same region for a localized water molecule.

In WT EcPCK, thiosulfate binding displaced four loosely bound water molecules, and the space occupied by one of its sulfur atom possibly squeezed the space for another loosely bound water molecule (WAT_A in Figure 2.17). The orientation of thiosulfate allows a potential hydrogen bond between one of its oxygen atoms and the hydroxyl group from Tyr207. In the G209S mutant, however, Tyr207 hydroxyl group points to an opposite direction, leaving a loosely bound water molecule (WAT_B in Figure 2.17) in the pocket. In addition, the Ser209 hydroxyl group not only inhibits binding of the water molecule WAT_A in WT, but also favors the placement of a potential hydrogen bond acceptor. As a result of bound solvent changes upon methanesulfonate binding, the sulfonate group moves closer to Ser209, displacing two loosely bound water molecules and leaving a potential hydrogen bond acceptor, while the hydrophobic methyl moiety displaced WAT_B along with another loosely bound water molecule. The oxygen atoms of methanesulfonate displace two other loosely bound water molecules and form hydrogen bonds with the hydroxyl group of Tyr207 and two water molecules identified in the crystal structure.
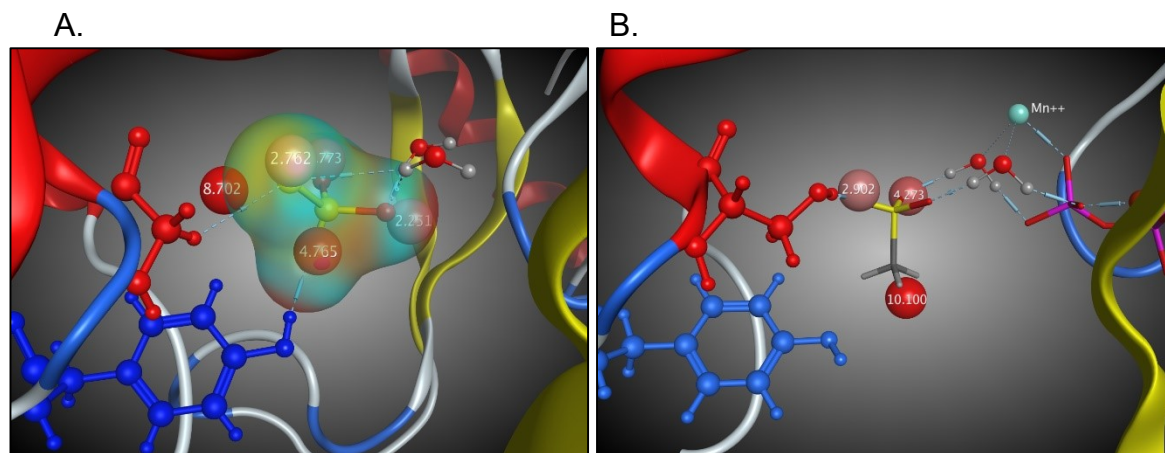
Figure 2.17. Solvent analysis of (A) thiosulfate-WT and (B) methanesulfonate-G209S complexes. Residue Gly209 and Ser209 are shown in red in A and B, respectively, while residue Tyr207 is shown in blue in both figures. The possible predicted loosely bound water sites are shown as red spheres, and the number on the sphere represents the potential free energy gained upon displacement of the labeled water, in kcal/mol. The larger the number, the redder the sphere and the smaller the enthalpic penalty upon displacing the water molecule. The transparent surface surrounding thiosulfate represents the space filling representation of the ligand: clearly one sulfur atom is squeezing the WAT_A (8.7 kcal/mol). The oxygen atom close to Tyr207 displaced another water molecule (4.8 kcal/mol) and forms a hydrogen bond with the hydroxyl group of Tyr207. Upon G209S mutation, the increased size of the sidechain at the residue 209 position eliminated WAT_A, and the hydroxyl group of residue 207 points to an opposite direction, resulting in a very weakly bound WAT_B (10.1 kcal/mol). The methyl group of methanesulfonate displaces WAT_B.

2.7.5 Energetics of ATP binding versus thiosulfate at the ATP position: implications from structures and mutations

Nucleotide binding in EcPCK is a key event leading to the closure of the enzyme's active site. As such, one would presume that it is a relatively robust feature in the system, and that affinity of ATP would be higher than those of smaller ligands, since more interactions usually result in a higher binding affinity. However, in the experiments, when cocrystallizing WT in the presence of thiosulfate and ATP along with the divalent cations requisite for nucleotide binding, thiosulfate was found to be bound in the ATP binding site. When thiosulfate is bound, the divalent cations in turn are not found in the active site. Normally, several oxygen atoms on the β- and γ-phosphates of ATP complete the octahedral coordination spheres of the Mg^{2+} and Mn^{2+} cations. Since crystallography captures stable, low-energy states in the crystalline form, it must mean that the thiosulfate-bound structure in the presence of free ATP and metals must be at a lower energy state than the ATP-bound structure in the presence of thiosulfate, assuming the crystallographic contacts do not influence the energetics of the system. This can be attributed to the minimization of the Gibbs free energy of the system: although enthalpy from the ATP-protein interactions are lost, the entropy gain from the free nucleotide and metals must be greater than the bond energies. This is most likely attributed to the additional degrees of rotational freely attainable by free ATP in solution, and furthermore consistent with possible strain in the ATP-bound enzyme structure.

As seen by my results, the ATP bound EcPCK show that ATP is present in the syn conformation. This conformation, with the nucleoside oriented so it is over the ribose sugar, is more unstable and higher in energy, negating some of the enthalpic gain from ATP binding. Upon ATP binding, Pi-stacking is also observed between the aromatic system of adenine and the sidechain of R449; the absence of this binding means this favorable interaction is lost as well. Unexpectedly, ATP binding in the presence of thiosulfate was restored for mutants. In all of the mutants described in the results section except for the K212I F216V mutant, $ATP/Mg^{2+}/Mn^{2+}$ binding was observed. Thiosulfate was present in the crystallization conditions at the same concentration as in the WT crystallization condition. The mutations made were not directly in the ATP binding site, since the purpose was to engineer novel binding in the CO_2 pocket. This is another case of the neighboring mutation phenomenon observed in the CO_2 pocket.

3   EcPCK conformational states *in crystallo* and in solution

3.1   Introduction

PCK has been postulated to undergo conformational changes based on crystallographic results. Nucleotide binding triggers the closing of the N- and C-terminal domains, the active cleft running in between these two domains, and the cap that covers the active cleft. The reader is directed to Section 1.1.3 and Figure 1.3 for a more thorough discussion of these structural features.

The dynamic nature of EcPCK during the catalytic cycle along with its robustness and feasible crystallization makes it an ideal prototypic system for studying and understanding protein conformational changes. There are crystal structures of this enzyme in the open and closed domain states (e.g. PDB ID 1OEN for the open state and 2PY7 for the closed PCK-ATP complex), but these models provide an incomplete picture of the entire system under biologically relevant conditions. For example, one issue with the published crystallographic models is that the cap in these structures were often not visible due to its flexibility in the open state, or the electron density was of inadequate quality to support the model. As the cap has been purported to play critical roles in catalysis, understanding this element of the structure is required to understand the enzyme activity and mechanism.

Protein crystallography has been a gold standard for detailed structural analysis. However, a major goal of crystallography and structural biology in general is to understand functional mechanisms, but this goal is often not fully realized. Key issues involve adequately measuring flexibility and defining functionally-important conformational changes. Crystal contacts may not reflect biologically relevant interactions between macromolecules and may impact conformation. Conditions required to maintain a crystal are unlikely to adequately reflect the solvent conditions the protein finds natively. Cryogenic crystallographic data collection at 100 K minimizes radiation damage, but also removes some of the functional dynamics the protein can undergo. Thus, assessing the conformations sampled in cellular environments by crystallography is problematic and typically incomplete at best. Moreover, what is needed to reliably connect structures to function is likely to be a complete experimental description of the functional conformational states, which is typically not feasible by X-ray crystallography alone.

The approach to overcoming these limitations used here is to complement detailed crystallographic structures by studying EcPCK in solution with small angle X-ray scattering (SAXS). We and others (36, 71–79) have been developing small angle X-ray scattering into a powerful complementary technique to overcome the limitations of crystallography in determining the conformations necessary for function. SAXS efficiently provides the spherically averaged pairwise distance between the electrons in a macromolecule. These same pairwise distances may be accurately calculated from an atomic resolution result, thus forming the basis for comparing crystallography and SAXS. Although SAXS is often referred to as a low-

resolution technique, changes on the order of 5 Å are robustly observable and furthermore interpretable by calculating SAXS from modified atomic models. Since SAXS can be conducted in almost any solution condition and in high throughput, the effects of pH, salt or ligand, can be monitored providing a means to determine native state conformations at near atomic resolution in many cases.

Here, we set out to develop a more complete picture of PCK functional conformations and the atomic-level mechanisms used to adopt them by combining crystallography and SAXS on EcPCK. Our study tests and establishes SAXS methods and tools for providing novel insights into overall and high-resolution features of PCK with expected applications to many other systems. To enhance our understanding of PCK conformational changes, we purified and collected SAXS data from and crystallized several single and double mutation variants of EcPCK that we expected to probe functional conformations. We characterized the observed conformations EcPCK undergoes, both in the absence and presence of its ligands and substrates. Based upon these data, a mechanistic model for the PCK opening and closing mechanism is presented. The results provide insights into mechanisms and identify conformations that are likely useful for further engineering of EcPCK for non-native ligand catalysis and for guidance into possible inhibition of human PCK as a means for reducing tumor growth.

3.2   Materials and methods

3.2.1   Mutations to capture new conformational states

Several EcPCK insertion mutants were made to trap the EcPCK domains in intermediate functional states. The loops targeted were the nucleotide binding loop (includes S250), and (the loop that includes F409). One to three amino acid residues from the subset (F, I, L, M) were inserted into either of these two loops by Novici Biotech (Section 2.2.2), and the subsequent protein purified as described in the same section in Chapter 2. Purified protein was then crystallized in the same manner as described in Section 2.2.3.

3.2.2   A quantitative metric for domain motion: principal axes and dot product

A metric to calculate the extent of the openness/closedness of the two domains was created. The metric takes the dot product of two vectors that extend through each domain, lying close to its long axis. For the N-terminal domain, the axis is defined as the principal axis of inertia oriented in the long direction. Code to calculate the principal axes of inertia was modified from existing code (80) but modified to read PDB files using Biopython (81) to allow for better selection of regions of interest. The axis for the N-terminal domain was calculated using residues 10-355. The C-terminal domain's axis was from the center of mass of the domain to the Cα atom of Thr499 to minimize variation in the axis's position since the same method used for the N-terminal domain resulted in variation in the axis even when the C-terminal domains superpose well. This could be due to inclusion of part of the hinge region in the definition of the C-terminal domain. From the two axes, the dot product was calculated using NumPy (82). Histograms were generated using MatPlotLib (83).

3.2.3   Solution state biological small angle X-ray scattering

Purified EcPCK samples were prepared for small angle X-ray scattering experiments. After the size-exclusion chromatography step, samples were concentrated in spin concentrators and the flow-through used as buffers for the SAXS experiments. The buffer used was 20 mM Tris pH 8.0, 200 mM NaCl. Protein was diluted to a final concentration of 5 mg/mL for SAXS experiments. Data were collected in the apo condition (protein only), +ADP (5 mM each of ADP, $MgCl_2$, and $MnCl_2$), and +ATP (5 mM each of ATP, $MgCl_2$, and $MnCl_2$). SAXS data was collected on the SAXS endstation at the SIBYLS Beamline at the Advanced Light Source, Berkeley, California (84). Data was collected on a MarCCD 165 detector or a Dectris Pilatus3 X 2M detector, depending on when data was collected, at 11-12 keV. Exposure schemes varied between the old and new detectors. Integration of data was performed using software provided by the SIBYLS beamline (43). Merging and processing of integrated datasets was done in SCATTER (85).

3.2.4 The conformational space spanned by EcPCK

Models spanning the complete conformational space of EcPCK were generated from four known crystal structures. The Y207K apo and WT+ATP crystal structures provided the basis of the open and closed protein cleft states, while the cap (residues 386-406) was taken from G209N (closed) and Y207F (open). The models were all aligned by superimposition of their C-termini (residues 356-540), and the two caps were placed into the base structures, resulting in a total of 4 models. ATP was removed from the two WT+ATP models for the apo conformational landscape, while ATP/Mg^{2+}/Mn^{2+} was placed into the active cleft of the Y207K model based on the WT+ATP positions. Each one of these models were run through MODELLER (86) to complete flexible loops and the N- and C-termini along with the C-terminal 6xHis-tag that were not observed in the crystallographic electron density. Thus, a total of 8 complete chimeric models served as the reference structures for subsequent analysis.

3.2.5 Generation of theoretical SAXS profiles from atomic models for force plots

The atomic models emerging from crystallographic refinement are often missing residues, especially at the extreme N- and C-terminal regions, and for flexible loops as well. Because SAXS profiles are especially influenced by loops and terminal regions, which are usually located on the periphery of proteins, having a complete model is essential for accurate calculations of theoretical SAXS profiles from atomic models. To overcome this problem, we modeled in missing residues using MODELLER (86).

Theoretical SAXS scattering profiles of complete atomic models were then generated by inputting the aforementioned reference models into FoXS (77) with the appropriate experimental profile (either WT apo or WT+ATP); the best fitting model was selected to obtain a reasonable value for c_1 (the excluded volume parameter) and c_2 (hydration layer parameter) for calculating the remainder of the profiles. Theoretical and experimental profiles were then uploaded to the SAXS similarity server at the SIBYLS Beamline (36) for generation of force plots.

3.2.6 Experimental SAXS Profile Fitting

EcPCK models were fit to experimental profiles using FoXS. The Y207K mutant was fit with an ensemble of 10 models generated by MODELLER from the existing crystal structure; the best fit was obtained by a mixture of two models. WT apo fitting was accomplished through fitting with the G248_L249insIL crystallographic models completed by MODELLER. Finally, the WT+ATP structure was based on the existing WT+ATP crystal structure, but the cap was replaced with the closed cap from Y207F before fitting with FoXS. Plots and residuals were generated with Igor Pro (87).

3.3   Crystallographic conformations of EcPCK

We first characterize observed conformational changes in new and available crystal structures (21, 22, 88). Two multi-residue allosteric phenomena are immediately apparent and have been described previously: the opening and closing of the cleft containing the active site between the N- and C-terminal domains (22) (Figure 3.1a-c), and the cap which closes over the active site cleft (39) (residues 386-406, Figure 3.1d). The cap is required for function (8, 9). We separate our observations of these two phenomena.

We solved two crystal structures where the cap was well resolved and in discrete conformations. In our G209N mutant, the cap region extended away from the active site cleft in a pose we define as "open cap" (OC). Strong peaks in the 2mFo-DFc electron density map contoured to 1.5 rmsd allowed the main chain of the cap to be traced and several sidechains on this segment to be modeled as well. Because of the flexibility of the open cap, electron density for the entire cap sequence was previously not observed. We crystallized and solved the structure of another point mutant, Y207F, where the 2mFo-DFc map with strong signal when contoured to 2.5 rmsd showed the cap tightly packed over the active site cleft in a conformation we call "closed cap" (CC) (Figure 3.1e). An existing deposited structure of EcPCK (PDB ID 1OS1) has a closed cap built into the model. However, that particular deposition's electron density map contoured to 1.0 rmsd did not confidently provide support for the model's cap. Finally, the G248_L249insIL insertion mutant did not have strong electron density for the flexible loop region of the cap, but the β-strands forming the base of the cap were clearly visible and in an intermediate position between the open and closed state. Statistics for the crystallographic datasets and models are presented in Table 3.1 which follows this section.

From my crystallographic results, I assembled a set of contrasting conformations both crystallographically observed and extrapolated from combinations of structures. I fused the open protein and open cap (OPOC), open protein closed cap (OPCC), closed protein open cap (CPOC) and closed protein closed cap (CPCC) conformations.

To characterize the degree to which a particular conformation is open or closed we developed an objective metric. The N- and C-terminal domains of EcPCK were first separated with the N-terminal domain defined by residues 1-355 and the C-terminal domain defined by residues 356-546. These domains hinge about a distal helix located at the bottom of the cleft. The principal axis of inertia was calculated for the N-terminal domain, while for the C-terminal domain, an axis from the center of mass for the domain to the C-alpha carbon of Thr499 was drawn (Figure 3.2a). The dot products of the vectors were calculated where values closer to one indicate the domains are parallel and therefore in a closed conformation. Values closer to zero indicate perpendicular domains and therefore an open structure. A distribution of the values can be seen in Figure 3.2b. The closed states clustered around 0.5 while the open states showed a range of values from 0.144-0.575. Table 3.2 shows the scores of the various mutants analyzed with this method. An apo wild type (WT) open structure (PDB ID 1OEN) scored 0.304 while an ATP bound closed WT structure (PDB ID 1OS1) scored 0.525. Most available crystallographic structures had dot products near 0.52 and were therefore in the closed state reflecting the amenability of this state for crystallization.

Mutations expanded the range of conformations observed in crystal structures, providing both an extreme conformation and several intermediates as measured by the dot product analysis. One particular mutant, S250_G251insL, was an outlier, and inspection of the model did not show a state that differed from the Y207K mutation. Thus, the Y207K mutation was the most open with a score of 0.265 and adopts a conformation we call "open protein" (OP). Our solved structures of insertion mutants G248_L249insIL and F409_G410insIM were between the open and closed forms (dot product scores 0.376 and 0.508, respectively). The conformation that we refer to as "closed protein" (CP) was set to the WT+ATP complex.
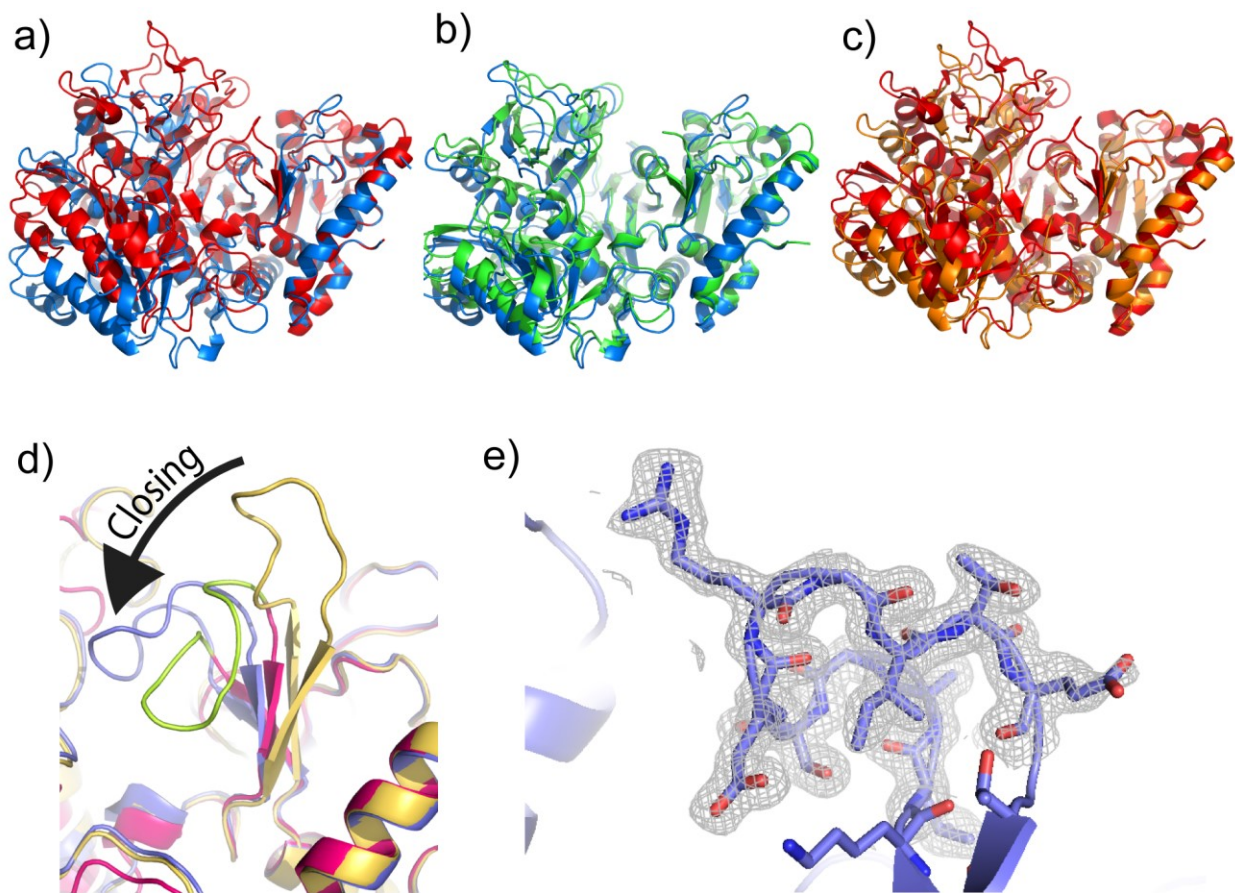
Figure 3.1. Conformational states of EcPCK. a) Overlay of the open (blue, PDB ID 1OEN) and closed (red, PDB ID 1OS1) cleft forms of EcPCK show the available existing structure and serve as a reference point for the two discrete states. The models are aligned with respect to their C-termini, the right half of the structure. b) The crystal structure of the G248_L249insIL insertion mutant (green) exists in a state that is more closed than the previous open structure (blue, same as in panel a) c) The F409_G410insIM insertion mutant (orange) binds nucleotide but exists in a novel closed state relative to the WT structure (red, same as panel a) d) Close-up view of the cap motion that closes over the active cleft. The fully open cap (gold) and fully closed (purple) have been fully observed via X-ray crystallography, and the intermediate state (teal and lime green) have been observed crystallographically as well through the G248-IL insertion mutant, but the lime green section had poor electron density and was therefore modeled in with the MODELLER software. e) Close-up view of the closed cap in Y207F. Electron density (2mFo-DFc) contoured to 1.0 rmsd is shown in grey.
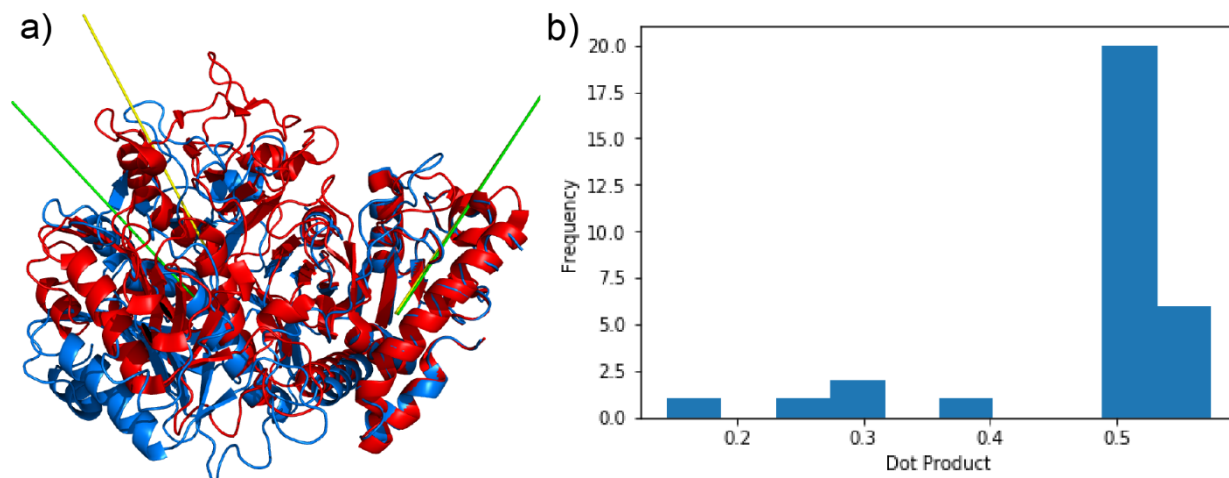
Figure 3.2. Objective metric to determine the degree of EcPCK closure. a) EcPCK in its open (blue, PDB ID 1OEN) and closed (red, PDB ID 1OS1) cleft forms (colored identically as in Figure 3.1) with C-termini (right side) aligned, with principal axis of inertia drawn in for each domain, shown in yellow for the closed form and green for the open form. The dot products of the two vectors were calculated for the open and closed states. b) Calculating the dot product of 37 available EcPCK structures and plotting the results on a histogram shows that we are able to classify the open and closed conformations of EcPCK through the dot product, as well as intermediate states.

	G248_L249insIL	F409_G210insIM
Resolution range	48.8 - 2.02 (2.09 - 2.02)	49.17 - 2.38 (2.46 - 2.38)
Space group	P 21 21 21	P 43 21 2
Unit cell	51.87 77.61 125.50 90 90 90	117.75 117.75 137.41 90 90 90
Total reflections	255713 (15691)	495745 (10954)
Unique reflections	33654 (2993)	38481 (3038)
Multiplicity	7.6 (5.2)	12.9 (3.6)
Completeness (%)	98.77 (88.92)	94.26 (51.05)
Mean I/sigma(I)	14.96 (2.30)	13.28 (0.35)
Wilson B-factor	22.01	49.78
R-merge	0.13 (0.77)	0.21 (3.48)
R-meas	0.14 (0.86)	0.21 (4.04)
R-pim	0.05 (0.35)	0.06 (2.0)
CC1/2	0.997 (0.705)	0.995 (0.0938)
CC*	0.999 (0.909)	0.999 (0.414)
Reflections used in refinement	33652 (2992)	37316 (1976)
Reflections used for R-free	1967 (172)	1939 (108)
R-work	17.7% (27.4%)	20.8% (37.7%)
R-free	20.3% (28.0%)	24.3% (43.4%)
CC(work)	0.960 (0.821)	0.949 (0.308)
CC(free)	0.941 (0.824)	0.949 (0.118)
Number of non-hydrogen atoms	4265	4184
macromolecules	4076	4013
ligands	N/A	29
solvent	189	142
Protein residues	527	524
RMS(bonds)	0.004	0.004
RMS(angles)	0.98	0.79
Ramachandran favored (%)	96.93	92.46
Ramachandran allowed (%)	2.88	5.80
Ramachandran outliers (%)	0.19	1.74
Rotamer outliers (%)	0.47	7.71
Clashscore	0.75	8.22
Average B-factor	25.14	55.99
macromolecules	25.14	56.12
ligands	N/A	60.43
solvent	25.25	51.47

Table 3.1. Crystallographic data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

EcPCK Variant	Dot Product
S250_G251insL	0.144
Y207K	0.265
K212I F216V	0.300
WT PDB ID 1OEN	0.304
G248_L249insIL	0.376
S250T	0.499
F413Y	0.507
F409_G410insIM	0.508
PDB ID 1AYL	0.511
PDB ID 1AQ2	0.512
Y286F	0.515
PDB ID 2OLQ	0.515
S250N	0.518
G209V	0.520
G209S K212V	0.521
T394N	0.521
PDB ID 2PY7	0.522
A392_G393insFL	0.523
PDB ID 2OLR	0.524
PDB ID 2PXZ	0.524
PDB ID 1OS1	0.525
PDB ID 1K3D	0.525
A392_G393insMFL	0.527
PDB ID 1K3D	0.527
G209Q	0.530
G393_T394insFMM	0.532
G393_T394insLL	0.543
G209S K212C	0.546
Y207F	0.549
R65H	0.559
P67E G209S	0.575

Table 3.2. Values of the metric comprising the histogram shown in Figure 3.2b when applied to various PDB models of EcPCK. Values less than 0.45 indicate a structure that has an open active site cleft, while values greater than 0.45 are closed. When more than one chain exists for a particular PDB model, chain A was used for this analysis.

3.4 EcPCK solution state behavior

3.4.1 Mapping out the conformational space spanned by EcPCK in solution

From crystallographic analysis, PCK undergoes two major motions: 1) the opening and closing of the active cleft through motion of the N- and C-terminal domains, and 2) the opening and closing of the cap that covers the active cleft. The range of conformations that are suggested by crystallography are bounded by the extremes of the aforementioned motions. We generated chimeric atomic models as mentioned in Section 3.3: OPOC (open protein cleft, open cap), OPCC (open protein cleft, closed cap), CPOC (closed protein cleft, open cap), and CPCC (closed protein cleft, closed cap) (Figure 3.3a). The SAXS scattering profiles for each of the chimeric models were calculated using the FoXS server and are shown in Figure 3.3b.

Importantly from the calculated SAXS profiles of the four extreme EcPCK states, we see that SAXS is capable of distinguishing these states as they give distinct scattering profiles. It is moreover possible to overlay an experimental SAXS profile onto Figure 3.3b and see which conformational state the experimental profile most resembles. This might not be straightforward if the profiles do not fall on one of these four states or for large datasets.

To provide a comprehensive perspective of large SAXS datasets, we employed a Force Plot method. The SIBYLS Force Plot is an extension of the SAXS Similarity Tool previously described in (36) designed to compare multiple sets of datasets. Instead of visualizing the data on a matrix in the heat map, the force plot represents each SAXS scattering profile as a single circle with its size scaled by the radius of gyration defined by SAXS. The repulsive forces between the profiles are determined by their degree of similarity to one another: similar profiles will be closer together in space on the force plot, and less similar profiles will be further apart. The force of opposition between the nodes is set by the Volatility of Ratio (V_r), described in (36), a similarity metric defined by a normalized comparison of the entire SAXS datasets. Notably, added insight on the four extreme conformational states of EcPCK can also be gleaned from the force plot (Figure 3.4). The quadrilateral spanned by the four landmark models of EcPCK is distinctly trapezoidal, deviating significantly from a square. Since the distances between two nodes reflect how different they are from one another, we obtain objective experimental information on whether the domain or cap motions of EcPCK result in a bigger change in the SAXS profile. The distance between nodes of differing cap states is less than the distance between nodes of differing domain states, showing that the positions of the domains have a larger influence on the SAXS scattering profile, which depends upon the change in scattering of the electron density within each enzyme. This corroborates the fact that the domains comprise of a larger portion of the total atoms in the system. We can also see from the plot that the transition between the CPCC and OPCC states results in a larger SAXS profile alteration than the change between the CPOC and OPOC states.

The force plot provides an effective and comprehensive way to objectively visualize how real experimental data behaves relative to fixed landmarks. Immediately, outliers can be seen, and inferences for the mutants that lie in the space spanned by the chimeric models can be further analyzed. Detailed fits are presented in the following sections.

The apo force plot comparison (Figure 3.4) contains the four apo calculated SAXS profiles, along with apo SAXS scattering profiles of selected mutants. The four calculated scattering profiles create conformational landmarks when analyzing experimental data in the force plots. Most of the experimental data lie within a quadrilateral defined by the models, with one outlier, the P67E G209S double mutant. Detailed analysis showed that the radius of gyration (R_g) of this mutant was significantly higher than the others mutants, and fitting the data showed that it exists partially (10%) as a dimer in solution (Figure 3.5) explaining why it deviates from the trapezoid formed by monomeric models. This observation also highlights the strength of SAXS in defining assembly state. Wild type EcPCK and mutant Y207F sit closest to OPCC, a model not observed crystallographically, with some component of similarity with CPCC. A cluster of mutants (G209S K212V, P67Q, R65F K212C, R65H) sits between OPCC and CPOC (models created by combining structures). Experimental profiles that lie between model nodes are intermediate conformations or flexibly sample multiple conformations between the proximal models. Mutant Y207K co-locates near its crystal structure (OPOC) along with K212 F216V. Notably, none of the experimental curves co-located with or are nearest to CPCC, suggesting this conformation is not reached in the absence of nucleotide.
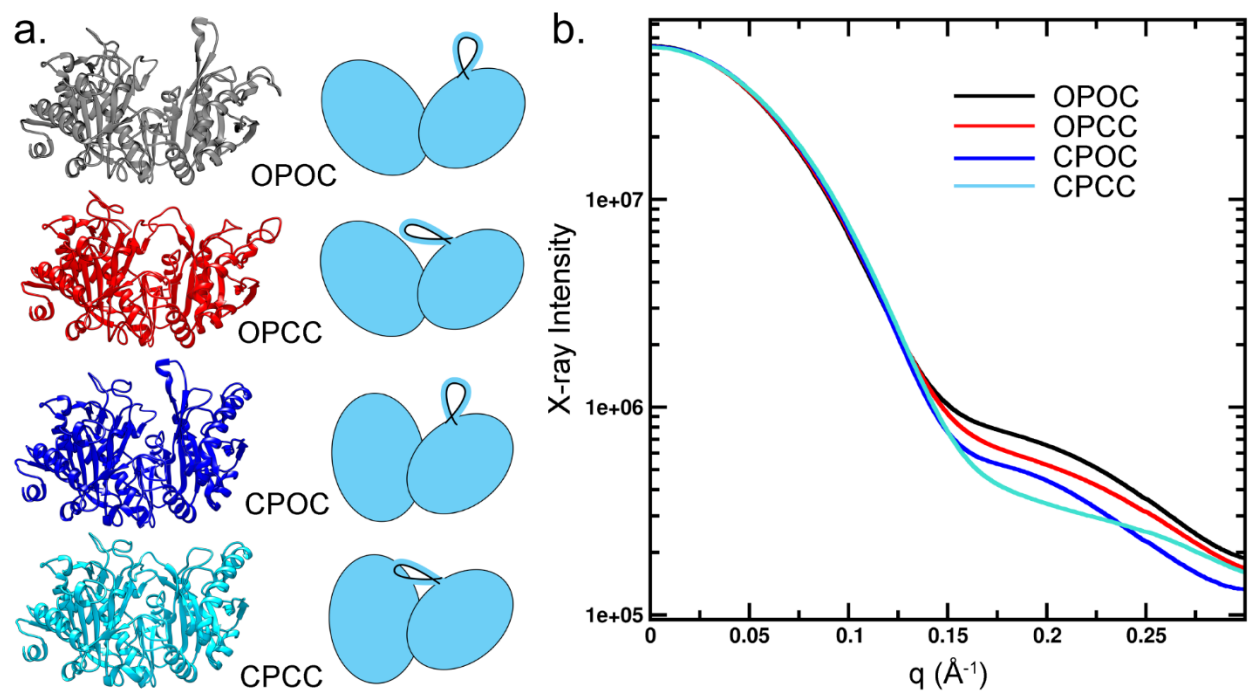
Figure 3.3. The four extreme EcPCK states. a) Structural representation of the four extreme states in cartoon form (left) and as a simplified model where the two domains are represented by the ovals and the cap represented by the protruding loop (right), where OP = open protein, CP = closed protein, OC = open cap and CP = closed cap. b) the calculated SAXS scattering profiles of the four models shown in panel a begin to diverge around $q = 0.14$ Å^{-1}, allowing for the differentiation of the four states through SAXS.
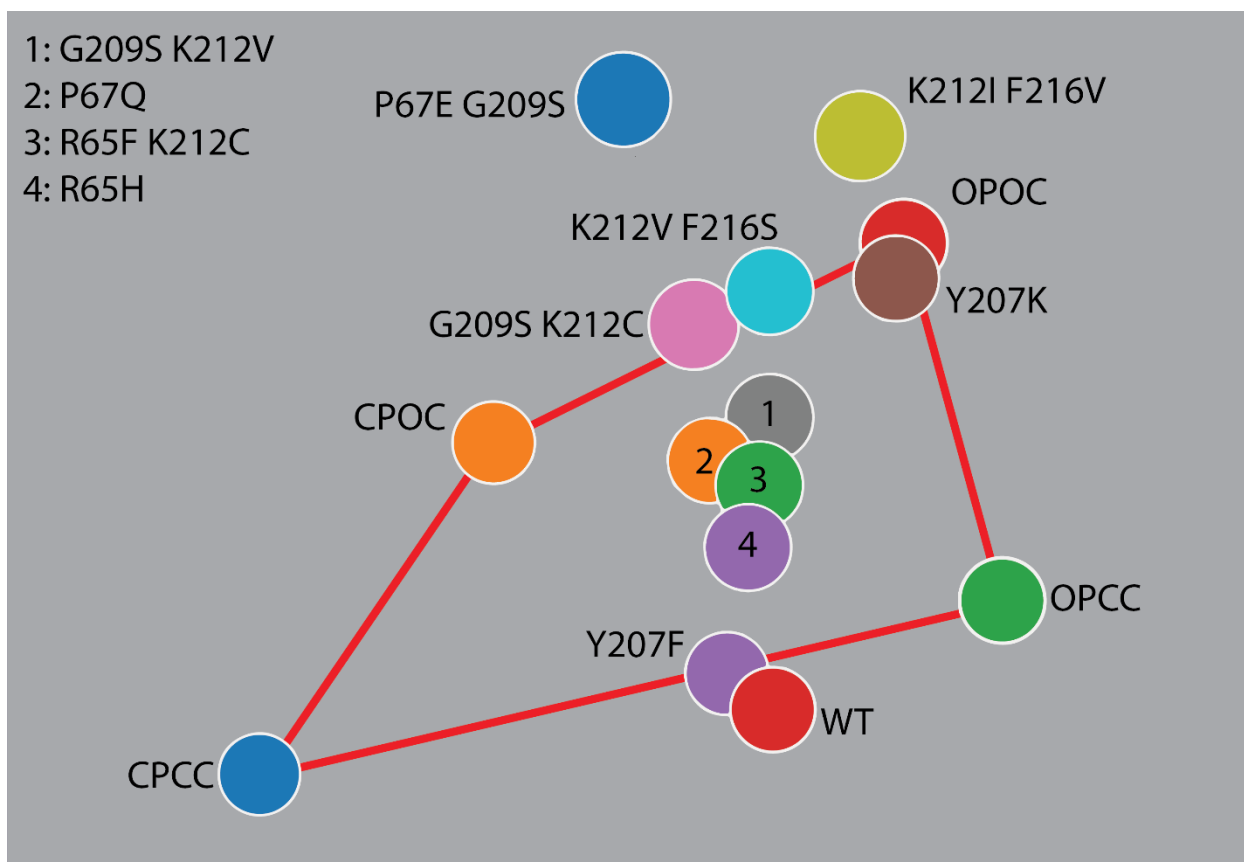
Figure 3.4. Apo force plot of EcPCK variants. Each node represents a SAXS scattering profile, and the distance between nodes represents the comparison of the V_r similarity metric score between the two profiles. Vertices of the quadrilateral map the four theoretical states the WT enzyme can reach: CPCC = closed protein cleft, closed cap, OPOC = open protein cleft, open cap. Edges connecting the vertices are drawn in red. Two profiles are outside the trapezoid: P67E G209S, determined to be partially dimerized in solution, and K212I F216V. Y207K lies extremely close to the CPOC state. The rest of the mutants lie roughly in a vertical line between the OP and CP states, suggesting that in the apo state, the N-terminal and C-terminal domains are in an intermediate position between the fully open and fully closed states as seen in the crystal structures. K212V F216S, G209S K212C, and Y207K mutants exist in solution with an open cap, while Y207F and WT in solution have a closed cap, both of which deduced by the fact that they lie on the edges of the trapezoid. The mutants in the middle, labeled 1-4, are G209S K212V, P67Q, R65F K212C, and R65H, respectively.
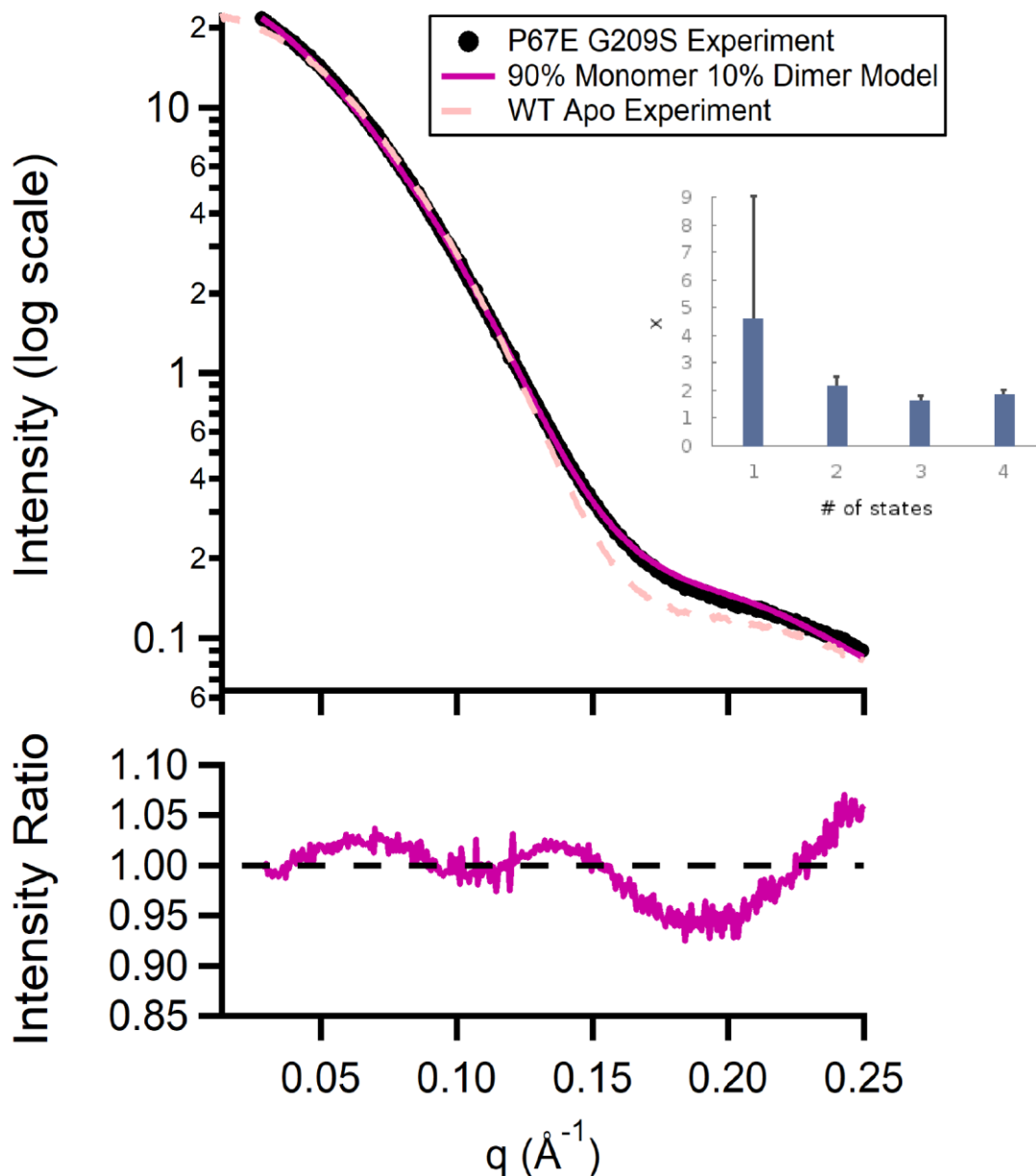
Figure 3.5. Fitting the P67E G209S double mutant profile shows that it is partially dimeric in solution. The radius of gyration of this mutant (low-q region, black) is higher than the WT profile (pink dotted line), suggesting there is a subset that is dimeric, but less than that of a full dimer. The dimer was modeled through crystallographic symmetry: EcPCK crystals in the $P2_12_12_1$ space group have two copies of the protein per asymmetric unit. The fit in the region around q = 0.05 improved drastically upon inclusion of 10% dimer (purple). Inset shows that the best fit is achieved with 2 models; the fit with 3 models is not a statistically significant improvement. Bottom panel shows the ratio between the experimental profile and the model data: a perfect fit between experiment and model would have a ratio of 1.00 throughout.

3.4.2   Y207K point mutant behavior

The Y207K point mutant was crystallized in the presence of ATP/Mg^{2+}/Mn^{2+}, but did not show evidence of binding of any of these compounds in the resulting electron density maps. Coupled with the kinetics results in Section 2.4 showing drastically reduced activity of the Y207K mutant, there is evidence to suggest that ATP binding is significantly hampered by this mutation, though the mechanism for this phenomenon was not yet clear. Since ATP binding has been reported to be the factor in driving conformational change in EcPCK, and this mutant might be incapable of binding ATP, its conformation might predominantly dwell in the apo state. We decided to try and understand this particular mutant first as it might be a degenerate case in the conformations of EcPCK.

From the force plot (Figure 3.4), the Y207K variant of EcPCK was found to be close to the OPOC landmark, suggesting that this mutant in solution is mostly in the OPOC conformation. Fitting the experimental profile with the OPOC model yielded a fit that supports this result (Figure 3.6).

Both the crystallographic and solution state evidence suggests that Y207K exists chiefly in a conformation with the active cleft and the cap open. From the force plot, the solution state behavior of Y207K is quite different than that of WT EcPCK. This can also be seen in the overlay of the scattering profiles in Figure 3.6, where the WT and Y207K profiles deviate past q = 0.13 $Å^{-1}$.
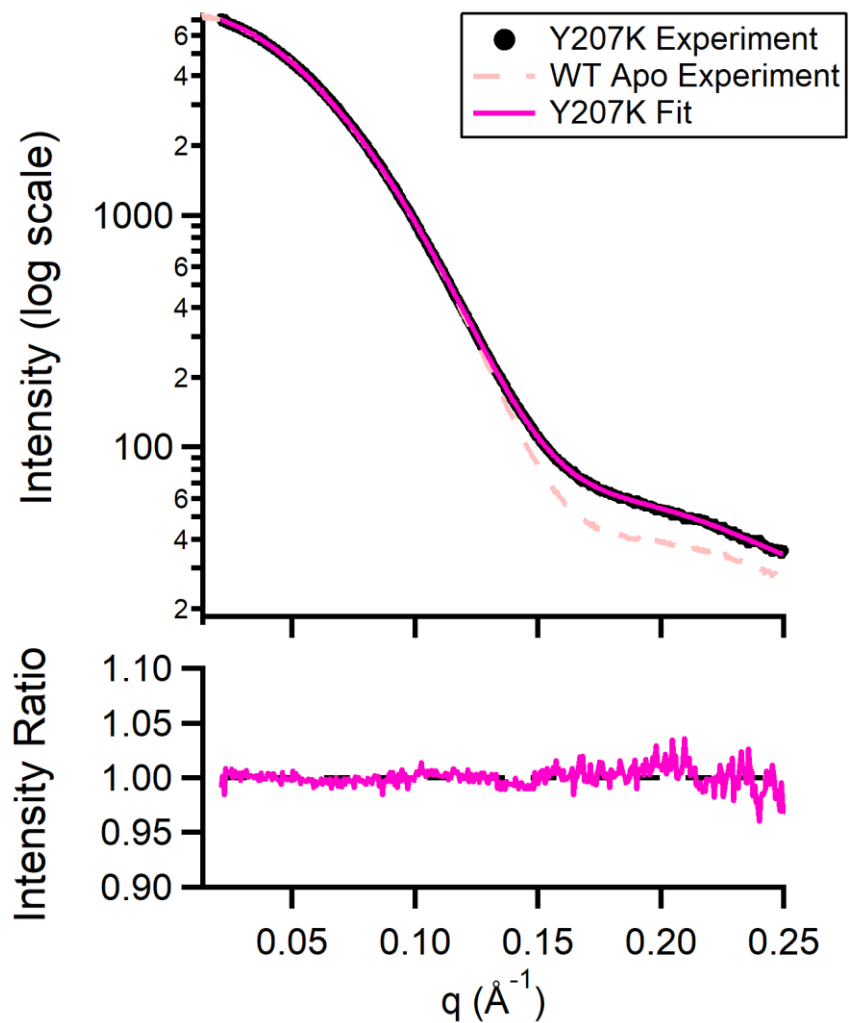
Figure 3.6. The Y207K EcPCK mutant behavior in solution can be explained by its crystal structure. The experimental SAXS profile (black circles) was fit with two models based on the Y207K crystal structure with missing residues completed with MODELLER (magenta solid line). Dotted pink line shows the apo WT EcPCK experimental data for comparison, showing that Y207K behavior is different from the WT. Bottom panel shows the ratio between the experimental profile and the model data: a perfect fit between experiment and model would have a ratio of 1.00 throughout.

3.4.3 Apo WT EcPCK behavior

SAXS scattering profiles were collected on WT EcPCK protein only (the apo state). This section describes the fit of the experimental SAXS profile to atomic models. From the force plot, the state of WT EcPCK in solution seems to fall between the OPCC and CPCC states, which means that the SAXS data indicates that the cap is consistently closed, while the two domains are in a state in between the open and closed crystallographic states.

Testing the fit of the experimental SAXS profile to an existing apo crystal structure, PDB ID 1OEN, resulted in a very poor agreement (Figure 3.7, red line). It is evident that this crystal structure is not describing the behavior of WT EcPCK in solution. We utilized computational techniques to try and change the conformation of EcPCK; tools based on the elastic network model and normal mode analysis were tried, but either did not improve the fit or resulted in biologically infeasible structures (e.g. numerous chain breaks).

Since the force plot suggested an intermediate domain position relative to the open and closed states, and we identified two new intermediate crystallographic states (Section 3.3), those states were used to fit the experimental SAXS profile. An atomic model based on the G248_L249insIL mutant with the closed cap taken from Y207F proved to be the best fit (Figure 3.7, cyan line), suggesting that in solution, the N- and C-terminal domains of apo WT EcPCK adopt the position crystallized in the G248_L249insIL mutant, and with a closed cap.

The solution conformation of the enzyme has a closed cap loop, and a slightly more closed active site cleft than the crystal structure. This is contrary to expectations, since a closed cap would appear to inhibit ingress and egress of substrate. One possible explanation for this behavior would be that the apoenzyme samples various conformations. Because the SAXS data were collected on the second timescale, there would be averaging of the conformational states. Alternatively, substrate has to enter the enzyme from the side of the active site cleft rather than through the top. This is discussed further in Section 3.6.4.
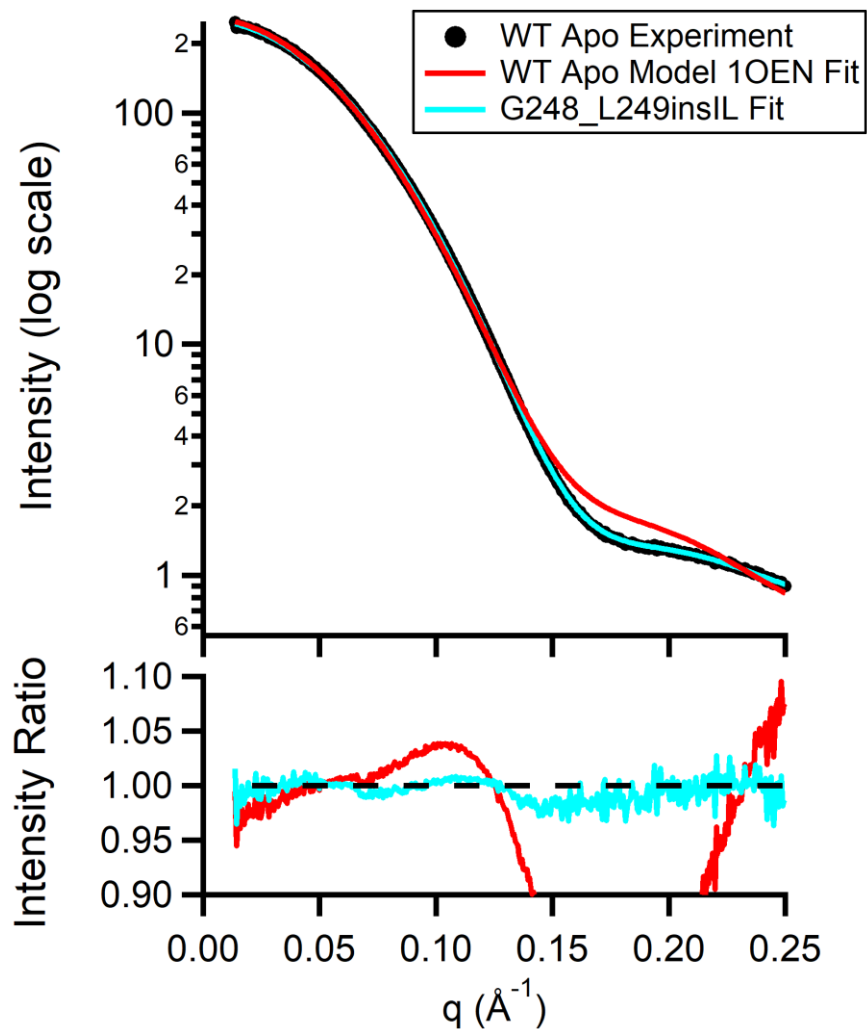
Figure 3.7. Experimental SAXS scattering profile of WT apo EcPCK (black) fit with the crystal structures of WT (red line, PDB ID 1OEN) and G248_L249insIL insertion mutant (blue) shown in Figure 3.1 after modeling in missing residues. The ratio between theoretical and experimental data is shown beneath the scattering profile. The best agreement is between model and experimental data is obtained with the G248_L249insIL insertion mutant model, suggesting that the WT EcPCK cleft and cap position in solution is best represented by the crystal structure G248_L249insIL.

3.5   EcPCK closure upon nucleotide binding

Nucleotide binding plays a large role in the catalytic cycle of PCK. As mentioned in previous sections, this event is purported to drive the conformational change of the enzyme. As such, characterizing the ligand-bound state in solution is important towards understanding the mechanism of this enzyme.

The EcPCK/ATP/Mg^{2+} complex has been crystallized and yielded a closed active cleft conformation, as expected. Density for the cap was fully observed in the open conformation, which was surprising since that contradicts one of its expected roles of excluding bulk solvent from the active site during catalysis. This was a hint that the crystallographic structure might not be accurately capturing a biologically critical state of the system. To better understand the closed conformational state of EcPCK, we performed solution state SAXS experiments.

SAXS data of WT EcPCK were collected in the presence of ATP, along with Mg^{2+}, and Mn^{2+}, the cofactors involved in ATP binding to capture EcPCK in its closed conformation. Because SAXS is a solution-based technique, we can sample a much wider range of conditions compared to X-ray crystallography. As long as we can prepare a buffer with the ligands, we are able to collect information on the buffer-subtracted protein data.

To understand the ATP-bound holoenzyme of EcPCK, care must be taken when comparing it to the apoenzyme data. Because SAXS measures the electron pair distance distribution function (the P(r) function) in a given sample, it is sensitive not only to just conformational changes, but also to ligand binding. Although the ligand is present in the buffer and thus the electron pair distance information within the ligand molecule is subtracted out from the sample, ligand binding introduces additional terms since there are now interactions between the protein and the ligand. Furthermore, with a conformational change, subtracting the contribution of the protein-only term obtained from the apoenzyme will not be appropriate because the new conformation will also alter the P(r) function.

Because of this, we cannot simply place the EcPCK+ATP experimental SAXS data into the previous force plot. A new force plot with new landmarks must be generated: the landmarks will have the same EcPCK conformation, but with ATP/Mg^{2+}/Mn^{2+} modeled into the active site as well. This proved to be more challenging than generating the landmarks for the apo force plot, since the ligands can easily be removed from the PDB file of an atomic model. We placed ATP and associated cofactors by aligning the C-terminal domains of an apo structure with an ATP-bound structure, then placed the ligands into the apo PDB files at the same position as the superposed ATP-bound structure. This method worked because positioning of these ligands in the active site does not have to be exact; the P(r) function is more sensitive to the absence of these electrons rather than their exact placement.

The ATP force plot of the four EcPCK states (Figure 3.8) looks markedly different than that of the apo state. While the landmarks of the apo force plot form a trapezoid, this one has an elongated triangular structure with CPCC at the far end, and the other three states on the other end. This suggests that it would be challenging to differentiate the CPOC, OPCC, and OPOC states by SAXS. This is likely an artifact due to deficiencies in modeling the hydration layer in the active cleft, and will be further elaborated in Section 3.6.1. Even so, placing the WT+ATP experimental data on the +ATP force plot shows that it is unambiguously near the CPCC node.

This finding suggests that the domain positions are correctly captured by crystallography, and the cap is closed over the active site, which is in line with what is intuitively expected for the behavior of the cap.

To verify that this is the correct state, an atomic model of the CPCC state with ATP bound was fit to the WT+ATP experimental profile. This fit is shown in Figure 3.9 as the green curve, providing additional evidence that the ATP-bound form of EcPCK exists in solution in the CPCC conformation.
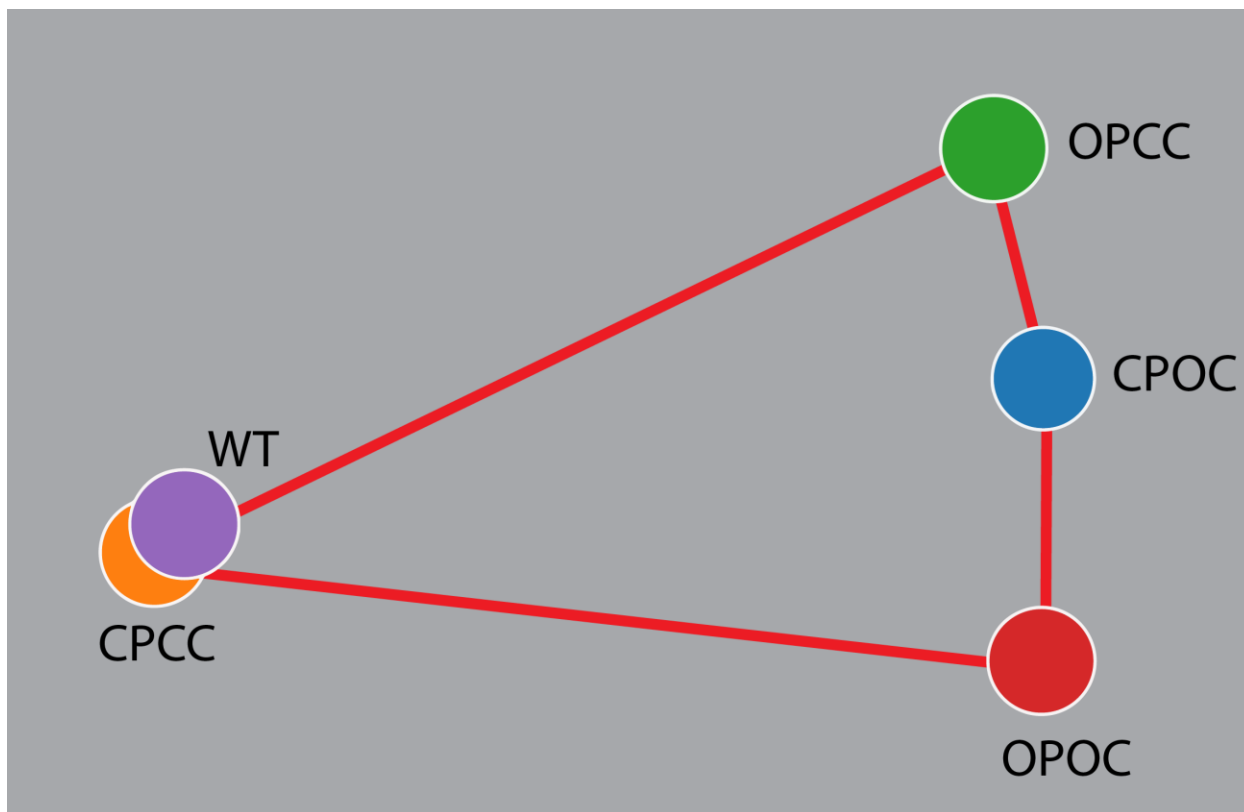
Figure 3.8. The ATP-bound force plot of EcPCK shows a markedly different landscape than the apo force plot. While the four landmarks in the apo force plot (Figure 3.4) are spread out in space, everything except for the CPCC state are similar (right cluster). WT+ATP EcPCK experimental SAXS data shows the greatest agreement with the CPCC state establishing this as the solution state.
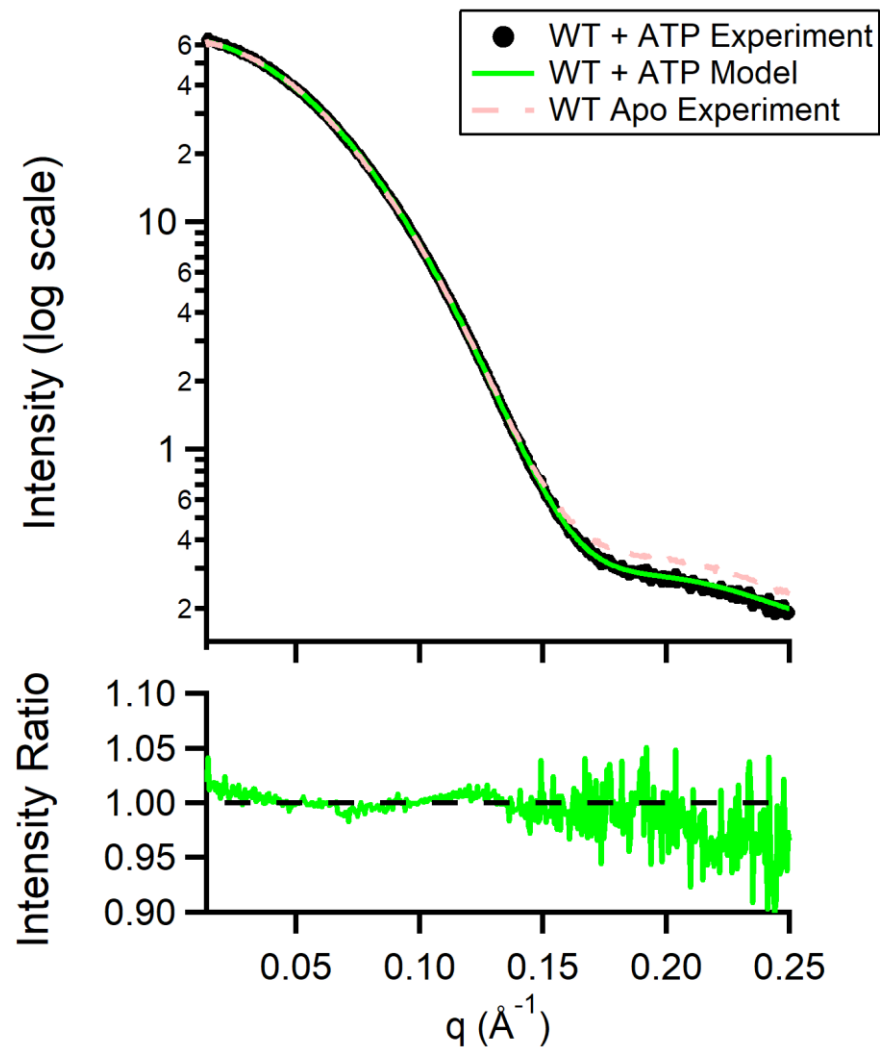
Figure 3.9. The ATP/Mg^{2+}/Mn^{2+} EcPCK complex shows good agreement with the CPCC model state. Experimental SAXS profile (solid black circles) is fit with a CPCC model of EcPCK with ATP bound (solid green line). Apo WT SAXS profile (pink dotted line) is shown for comparison.

3.6 Discussion

3.6.1 Sensitivity of SAXS to small scale changes: moving away from "blobology"

In this work, we have seen real life examples where small changes in the model can result in distinguishable fits. This is in part due to improvements to the precision with which SAXS measurements can be made, with the introduction of high quality detectors along with progress in synchrotron radiation. From the fits of the solution scattering data of the WT apo state versus the Y207K mutant, it can be seen that SAXS can differentiate the position of the cap and the separation between the two lobes: an incorrect position for these two components results in a suboptimal fit. One can see another example of this by fixing the position of the entire model except for the cap; generating the theoretical SAXS profiles with a varying cap results in drastically different SAXS scattering profiles (Figure 3.3b). In the force plots, the same model was used for the OPOC/OPCC and CPOC/CPCC states; only the cap was changed, yet the scattering profile varies drastically.

During the process of fitting the WT+ATP model to the experimental data, it was seen that the theoretical SAXS profiles of the PDB models with and without ATP/Mg^{2+}/Mn^{2+} showed a substantial difference. Combined with the results with the different cap and relative domain conformations, the data shows that movements of small loops, relatively small shifts in domains, or binding of small molecules can be observed by SAXS. Although SAXS is usually treated as a low-resolution technique, mostly stemming from the bead models generated for *ab initio* shape determination, direct comparison of experimental solution scattering profiles under various conditions amongst themselves and especially to existing crystallographic models can provide insight to biological function at much smaller length scales than is traditionally expected from SAXS. From the crystal structures of the open and closed caps shown in Figure 3.1d, the β-strands forming the base of the cap start from the same position and at the end are separated by 9 Å, and the maximum distance between the two cap states are 15Å. From averaging these distances, an 8-10 Å estimated resolution would be needed to distinguish the cap positions. Together, these results suggest that the development of tools that combine crystallographic and solution scattering information has far-reaching impacts on drawing physiologically relevant conclusions from macromolecular structures, since both high-resolution and solution state information are elucidated.

While SAXS can determine small-scale changes, certain experimental conditions make interpretation more challenging, especially when ligands are involved in scattering. From this work, the four extreme conformational states can be differentiated as seen in the apo force plot (Figure 3.4), but upon addition of ATP into the active cleft, the differentiability of these four states decreases on the force plot (Figure 3.8). The SAXS profiles of these states are theoretical and calculated from chimeric atomic models, suggesting differentiability could be reduced by the way the theoretical scattering profiles are calculated. While the atomic positions of the protein's atoms and the ligands are clear, there is a hydration layer around the protein; the partially ordered solvent in this region is different than the bulk solvent. The hydration layer and its contribution to scattering when analyzing high resolution features is an area of continued research. The active site cleft with ATP and the divalent cations bound pose a challenge to existing models because they diverge from the chemical environments that are routinely

encountered with the twenty amino acids at the surface of proteins. In fact, this is a problem even for high-resolution X-ray crystal structures. Examination of crystallographic data and the refined structural models uncovers a gap in between the level of error in the data and the error in our best refined X-ray crystallographic models, which has been termed the R-factor gap (89). Examination of difference maps indicates that this R-factor gap is partly inside the model in flexible regions and partly in the solvent: current X-ray crystal structures do not accurate model the data for flexible regions and for bound solvent. The results presented here suggest that the combined SAXS and crystallographic data from this work can serve to help improve hydration layer models for structural biology including high resolution crystal structures.

3.6.2 Insights from solution EcPCK conformational states

Solution state data deviate from crystallographic results

Although EcPCK and other members of the PCK superfamily have been crystallized extensively in the past, the atomic models were not always able to fit the solution state data without modification. This suggests that the solution state of EcPCK, and by logical extension likely the majority of proteins, differ from the crystallographic structures. Reasons for these discrepancies include crystallographic packing of protein molecules, and differing conditions between SAXS data collection and crystal growth. One clear example of where crystallographic packing could influence the protein conformation is the EcPCK cap.

In the G209N mutant where the open cap is fully observed, there are interactions between the cap and residues 75-82 on a second EcPCK molecule that is present due to crystallographic symmetry. As for the second point, protonation states of some sidechains vary based on the pH environment. The SAXS data were collected at pH 8.0, while the conditions for crystal growth were at pH 5.78. Histidine would have different protonation states between these two conditions. Since electrostatics have effects on conformation (e.g. the Y207K mutation in Section 3.4.2), the crystal structure, with its different charge state, would differ from the solution state information. Crystallization of the protein under more biologically-relevant conditions would generate data that would likely better reflect *in vivo* conditions, but proteins dictate the conditions under which crystals grow, not the crystallographer. As such, since SAXS can be accurately calculated from the crystal structure, SAXS is an excellent solution-base complement to verify the extent to which crystallography results reflect native solution behavior and suggest alternate conformations and distributions of conformations.

Induced fit vs. conformational selection

There are two prevailing models on how enzymes undergo conformational change to bind to substrate: induced fit and conformational selection (90). To briefly summarize these two models, the induced fit model states that substrate binding induces a conformational change of the enzyme to correctly position residues and other structural elements into catalytically-active positions, while the conformational selection model states that an enzyme naturally undergoes fluctuations in its conformations ranging from its apo state to the state capable of ligand binding, and addition of ligand shifts the population of conformations towards the latter state. Previous

research on PCK has provided support for the induced fit model (33): the crystal structures shown in this work capture both the encounter complex (the enzyme-substrate complex before the enzyme is reshaped to be catalytically active) as well as the active enzyme-substrate complex. However, these conclusions were drawn from crystallographic structures which only provide a static snapshot of the conformation under specific conditions. Our solution state data showed that apo EcPCK adopted a conformation that had not been seen in any one crystal structure: the domains were in an intermediate state between the fully open and fully closed forms, and the cap was in a closed state. This is contrary to the induced fit model, since cap closure should be induced by nucleotide binding but we observed the closed cap in the apo condition. There is no SAXS evidence to support the open cap conformation in the WT enzyme, but observation of the open cap in Y207K shows that we are able to observe this state. Our data provides support for some degree of sampling of EcPCK states, suggesting that neither of the induced fit or conformational sampling models adequately describe the interaction of EcPCK with its nucleotide substrate.

Lid closing agreement with previous results

Evidence for cap closing in the presence of ATP but without OAA or PEP was observed computationally by Drummond et al. (91) Their MD simulations revealed a partially closed conformation of the cap that does not quite close to the same extent to the conformation in the *A. succiniciproducens* structure (PDB ID 1YTM). The closed cap conformation we observed in the Y207F mutant overlays with the cap in 1YTM; however, the substrates in these structures are different. While 1YTM was complexed to oxalate, Mg^{2+}, and Mn^{2+}, chain B of our Y207F EcPCK mutant was only complexed with thiosulfate. Since we observed cap closure with only thiosulfate bound at the ATP site, the interactions at the α-phosphate position of ATP are sufficient for cap closure.

Similarity of SAXS scattering profiles between ATP-bound and apo EcPCK states

The scattering profiles of the EcPCK apo state and the ATP-bound state are similar, as can be seen in Figure 3.9. This is not surprising, given that the solution conformation of apo EcPCK as determined by SAXS is between the OPCC and CPCC state, and the ATP-bound state is CPCC. The SAXS data show that we are only capturing the CPCC state of the PCK-ATP complex, and not the encounter state that was previously mentioned.

Since ATP binding in solution induces the CPCC conformation, it would not be possible to experimentally capture the other three conformations of EcPCK with ATP bound using static SAXS methods. However, time-resolved experiments, either with SAXS or other biophysical techniques, could aid in deciphering these phenomena. With SAXS, one can perform a time-resolved experiment by controlling the addition of substrate to the system. One possibility would be to do stopped-flow SAXS experiments, mixing EcPCK with ATP, to capture the conformational change as it is occurring. The other option would be to use a substrate that is present in solution but unable to be used by the enzyme until an external trigger is applied, for example, caged ATP that is photoreleased by ultraviolet light. The timescales for those experiments, on the order of 10 ms, might be able to capture this domain motion of EcPCK, but the technical sophistication of SAXS measurements with short exposure times is not trivial,

84

especially when probing small motions such as the ones described in this work. The faster the timescale, the less signal in each SAXS measurement for the same enzyme concentration, so doing these experiments is outside the scope of this dissertation, but my data suggest this as an approach to examine this process. Domain motions are generally on the μs to ms timescale (92). Another possible way to probe for this behavior would be a time-resolved FRET experiment, although the effects of chromophore labeling on the dynamics of smaller features, such as the EcPCK cap, could be significant.

3.6.3   Additional states of EcPCK captured by this work

Crystallography of several mutants, including the G248_L249insIL and F409_G410insIM mutants, revealed structures that were open to an intermediate degree relative to the open and closed forms of WT EcPCK, providing greater detail to the transformation that occurs upon closing of the active cleft upon nucleotide binding. Furthermore, crystallizing the Y207F mutant in the presence of thiosulfate revealed a clear picture of the closed cap.

Our work also adds evidence to the fact that in solution, proteins may have deviations from their crystallographic structures. While the differences between EcPCK crystal structures and the SAXS data mostly involve the cap, a 20 residue region that may seem insignificant relative to the whole 540 amino acid chain length of EcPCK, the importance of the cap in the role of catalysis cannot be overstated. While it is known that the protein is open in its apo state, finding the cap in a closed conformation in solution was unexpected. We had expected the cap to undergo a large degree of motion while EcPCK was in solution in its apo state, because it often appeared disordered in the crystallographic electron density, and also because an open cap should facilitate substrate ingress and egress from the active site more easily. Perhaps in its open state, EcPCK's cap is not a significant hindrance to ligand exchange as substrate can enter from the side of the active cleft. In this case, the cap opening process may act in expulsion of product after catalysis is complete. Further experiments could distinguish these possibilities identified in part by this work and shed light on the unique role of the cap.

3.6.4   The role of the Y207K mutation

From the scattering data, it is evident that the Y207K point mutation dramatically alters EcPCK from its wild type solution behavior. Looking more closely into why Y207K mutants are incapable of binding ATP can provide insight on the mechanistic underpinnings of EcPCK. Tyr207, the natural residue at this position, contains a largely hydrophobic sidechain with a polar hydroxyl group, while the Lys sidechain is of the same length but positively charged under biologically-relevant pH. In the closed cap structure of Y207F, it can be seen that R396 of the cap, another positively charged residue, is in close proximity to the residue 207 position. With the Y207K mutation, there would be repulsion of the two positively charged sidechains, leading to an inability for the cap to close. Cap closing might be a required motion for ATP binding in EcPCK, providing yet another reason to study loop dynamics in solution. Potential ways to accomplish this would be to perform time-resolved SAXS experiments, or SAXS experiments

with caged ATP substrates. If SAXS does not provide the timescales needed to resolve the conformational change, fluorescence experiments such as FRET could provide insight.

3.6.5 Conformational change and enzyme engineering

At this point, it is clear that defining conformational changes of EcPCK, or any other enzyme, is important for linking structure to function. One method to carry out enzyme engineering experiments is to make mutations to an existing enzyme. As one mutation is usually not enough to define functionally-relevant conformational space, large libraries of mutants are often generated, many of which are nonfunctional. Screening large libraries of mutants can be time-consuming or even infeasible. So combining mutant libraries with a high-throughput technique such as SAXS, as developed here, is an efficient way to eliminate mutants that are improperly folded and also, from what we have learned in this work, unable to undergo a requisite conformational change. The combined results presented here test SAXS in the analysis of functional conformational states. The results show that SAXS can be a powerful tool in structural biology with extended capabilities for examining conformational states in high-throughput expected to be beneficial for the burgeoning field of enzyme engineering field.

4 Conclusions and Future Prospects

4.1 EcPCK is a prototypic model system to examine ligand specificity and conformational states linked to chemistry

This work on EcPCK reveals that it is a robustly complex system to study ligand binding specificity: its discrete binding sites, which bind the three substrates and two divalent cation cofactors needed in the reaction, are coupled to motions the enzyme undergoes to achieve catalysis. As such, EcPCK is a prototypic model system to examine ligand specificity and conformational states linked to chemistry. Furthermore the role PCK plays in central metabolism along with its newfound function in cancer biology cements its biological importance, underscoring the importance of the research performed in this field to define specificity determinants for PCK and to more generally understand the role of residue interactions in ligand specificity.

I have expanded the knowledge of this system by providing experimental results and insights two aspects: the effects of mutations in the CO_2 binding pocket, and the functionally-relevant conformations of this system in solution including the impact of ligand binding. The CO_2 binding pocket region of the active site cleft had not been well-characterized prior to the work reported here. Furthermore, because conformation is so important for chemical activity and function, I probed EcPCK behavior in solution to analyze the position of its domains and loops under biologically-relevant conditions and to couple this conformational data with the atomic models generated from crystallographic data.

4.2 Crystallography reveals nonnative ligand binding modes in EcPCK

CO_2 in the atmosphere is the primary carbon source for life on Earth. In fact, CO_2 is not only a key molecule in cellular function, but it is also of great environmental significance and an important chemical pathway for control by synthetic biology and design efforts. Environmentally, CO_2 is the most significant and long-lived greenhouse gas in the atmosphere. In terms of ligand binding and chemistry, as an enzyme that unusually directly reacts with CO_2, EcPCK contains a binding pocket for this molecule. Moreover understanding and harnessing the specificity and reactivity of this pocket by replacing CO_2 binding with another molecule would lead to novel carbon-carbon bond formation. A first step toward this goal of understanding and controlling PCK is engineering novel ligand binding into this CO_2 pocket.

By systematically examining the effects of mutations in this area, I showed that certain residues were more critical to function than others, as seen by the loss of activity for some mutants while others retained native activity. The identity of the sidechains around the active sitealso played a role in determining how much activity is altered, and could be largely rationalized in terms of the active cleft sidechain chemistry, protein conformation and bound solvent.

Certain mutations led to nonnative ligand binding. The G209 position seems to play a large role in this phenomenon, where G209N and G209S point mutations led to nonnative ligand binding. This can be rationalized through the addition of a hydrogen bond donor at this position

where there was previously no sidechain. Aside from mutations at this position that led to nonnative ligand binding through new direct interactions with the ligands, neighboring mutations near the CO_2 binding pocket led to novel ligand binding characteristics as well, even though the mutations did not directly interact with the ligand. This shows that the pocket is sensitive to changes in electrostatic environment because of these mutations. Finally, ligand selectivity in these mutant enzymes is partly rationalized through the displacement of loosely bound water molecules identified through computational studies.

This work furthers our understanding of protein-ligand interactions in the CO_2 pocket of EcPCK, could serve to provide guidance for PEPCK inhibitor design for biology and cancer medicine, and provides that first step forward toward new synthetic chemistry via redesigned PCK.

4.3 Solution state information is complementary to crystallographic conformations and reveals missing conformational states

The detailed atomic resolution models provided by x-ray crystallography play an integral part in our understanding of proteins. Yet, due to the inherent nature of crystal packing and growth, crystallographic models may have deviations from biologically-active conformations and are unlikely to define all biologically-relevant conformational states due to the importance of flexibility in function. In this work, I explored the conformational space spanned by EcPCK states through mutating active site loops and performing comprehensive solution small angle x-ray scattering (SAXS) experiments to characterize key functional states of EcPCK and examine the structural similarities for ligand-binding mutants in solution.

The Y207K point mutant did not bind ATP when cocrystallized in its presence, hinting that it might not be able to undergo the necessary conformational change. I collected SAXS data on this mutant, and it was distinct from the WT enzyme. The Y207K SAXS scattering profile matched its crystal structure. This is not the case with the WT SAXS data, since the Y207K and WT crystal structures superpose with minimal differences yet the SAXS profiles do not overlap well.

In part I was able to examine comparative conformational state similarity through the use of the SIBYLS force plot developed by Greg Hura (93) and further tested and validated by my work in this system. From the calculated SAXS profiles of the four extreme EcPCK conformational states for the open and closed protein and cap, I showed that SAXS is capable of distinguishing these states as they give distinct scattering profiles. With SAXS force plots, the WT apo state in solution was found to have its N- and C-terminal domains in an intermediate state between the fully open and fully closed states, and with a closed cap. The partially closed nature of the apo WT solution state disagrees with previous results on the induced fit model of PCK. Furthermore, the Y207K variant was found to be close to the OPOC landmark, suggesting that this mutant in solution is mostly in the fully open conformation explaining its inability to bind ATP and linking ATP binding to PCK conformational closing. Results such as these push SAXS analysis beyond comparison of shapes toward measurement of structural states associated with ligand specificity and enzyme chemistry.

Importantly, the ATP-bound state of WT EcPCK was characterized both by crystallography and SAXS as well. Although the crystal structure of the complex had an open cap, the solution state data indicated that the cap was closed. Crystallographic packing of the loop against another copy enzyme in the adjacent unit cell is the likely reason for this discrepancy indicating that SAXS can be more accurate in defining functional conformations than crystallography, despite the inherent advantages in precision from the X-ray diffraction experiment where the signal is immensely amplified by the crystal lattice repeats.

Both the insertion mutants and the solution state conformational data provide added insights into EcPCK behavior that were not captured by the WT crystal structures alone. These new data from combined x-ray methods of scattering to measure solution conformation and diffraction to examine atomic detail provide a better understanding of EcPCK ligand specificity and enzyme activities. I hope that these may furthermore add insights informing our knowledge of the connections of protein conformation to chemistry protein conformation in general.

4.4 Future prospects

4.4.1 Analyzing networks of changes triggered by ligand binding including loop movements

One feature observed in the WT-thiosulfate crystallographic complex of EcPCK was that the overall conformation of the enzyme was closed. Previous studies have shown that nucleotide binding is the triggering event for closing the enzyme, but there was no nucleotide bound in my crystal structure. Instead, thiosulfate was bound at the α-phosphate position of where ATP would have been. This suggests that binding of an oxyanion at this position may be sufficient to trigger closing. Furthermore, this insight shows that SAXS is capable of providing relatively high resolution information on functionally-important conformational changes. Comparison of the ATP-bound and thiosulfate-bound closed structures of WT EcPCK might reveal key residues and possibly a network of residues that shift upon binding, showing a mechanistic trigger in the closing of this enzyme.

4.4.2 Inhibitor development leveraging CO_2 pocket binding data

When cocrystallized with thiosulfate, the WT enzyme was found to bind thiosulfate at the ATP and CO_2 binding sites. Due to the conservation of key catalytic residues across the PCK superfamily, this knowledge could potentially be used for developing inhibitors for PCK. An inhibitor could be designed by starting with thiosulfate and building out from there to increase the number of atoms and interactions it makes with the protein to increase specificity. Fragment screening by crystallography and then growth of experimentally-observed bound fragments is a successful strategy for inhibitor design (94–96).

89

4.4.3 Mechanistic insights from time-resolved experiments

Experiments with high temporal resolution could be designed to study the closing of EcPCK upon ligand binding. This would aid in capturing the ligand-bound open protein state, which would provide more information on whether the apo conformation of EcPCK in solution is capable of binding ligand. Also, kinetics data on the timescales of the various motions of EcPCK can be gained as well. As the timescales involved are far too long to be computationally simulated as yet, we will have to rely on experiments to get this information. In the future when computational methods mature, EcPCK may serve as a test system to validate computational molecular dynamics algorithms.

4.4.4 CO_2 sequestration using EcPCK

As the most significant and long-lived greenhouse gas in the atmosphere, CO_2 has been linked to rising global temperatures (97). Reducing its concentration through carbon sequestration has the potential to mitigate or even reverse its effects, if we have sufficient chemical knowledge to make this practical and efficient. Since EcPCK uses CO_2 in its reaction, it is a logical candidate to be considered for carbon sequestration projects. The K_m for CO_2 binding by PCK can be as low as 20 µM (98). Notably, this is comparable to the K_m for CO_2 binding by RuBisCO from C3 and CAM plants (12–26 µM) or from C4 plants (28–63 µM) (98, 99). This K_m for CO_2 binding means that the PCK reaction may be reversible in vivo, that PCK might act as a carboxylase in plants, and that PCK could in principle be redesigned to use CO_2 as one for synthetic biology and environmental remediation purposes (see Figure 1.2). PCK could in this sense be a major component in synthetic biology designs for carbon capture. For example, if one overexpresses PCK while simultaneously upregulating pathways that supply phosphoenolpyruvate, it would lead to the generation of oxaloacetate which feeds into many metabolic pathways, most notably the citric acid cycle. Logically, one could control the balance between carboxylation and decarboxylation by means of the ATP:ADP ratio plus the relative concentrations of OAA, PEP, and CO_2. Downstream reactions could use products arising from these reactions, such as acetyl-CoA, and generate long fatty acid chains for various applications that would help fund the carbon capture process.

In fact, the binding of CO_2 by EcPCK and other enzymes has been studied (91, 100, 101) for carbon sequestration purposes. However, our new structures and kinetics information on the mutation of the EcPCK CO_2 binding site provide added data and insight that could prove to be useful for improving such synthetic biology efforts including computational algorithms for understanding CO_2 binding. Indeed, taken together the results and methods presented here provide a roadmap for applying crystallography and high-throughput SAXS together to advance PCK biological understanding and protein engineering by providing the means to experimentally define any missing conformational states.

4.4.5   Toward addressing a grand challenge for structural biology

A major challenge for crystallography and structural biology in general is to understand functional mechanisms, but this challenge is often not fully met. The results presented in this dissertation pertain to understanding the ligand-binding interactions in the PCK active site cleft, how bound water molecules may influence ligand specificity, and how conformations may be linked to binding. My results may ultimately translate into approaches for inhibitor and drug development. Yet, overall these results may help link structure to chemistry by providing an exemplary approach and methods for the comprehensive experimental assessment of enzyme conformational states. The grand challenge of connecting structure to activity has proven difficult to address even for the well-funded Protein Structure Initiative (102). My results on PCK show that at least part of the answer to this challenge lies in the need to more completely map functional conformational states. Specifically, I show that a complete experimental description of the functional conformational states is typically not feasible by X-ray crystallography alone and that high-throughput SAXS has sufficient resolution to identify many such states. Thus, a more complete understanding of PCK and how ligand binding and conformational states are linked to enzyme chemistry comes from the integrated X-ray methods described here. These methods and insights may therefore help address the grand challenge of better linking structure to biological activity. Time will tell.

Bibliography

1.   Utter, M. F., and Kurahashi, K. (1954) Purification of Oxalacetic Carboxylase from Chicken Liver. *J. Biol. Chem.* **207**, 787–802

2.   Utter, M. F., Kurahashi, K., and Rose, I. A. (1954) Some properties of oxalacetic carboxylase. *J. Biol. Chem.* **207**, 803–819

3.   Utter, M. F., and Kurahashi, K. (1954) Mechanism of action of oxalacetic carboxylase. *J. Biol. Chem.* **207**, 821–841

4.   Goldie, A. H., and Sanwal, B. D. (1980) Allosteric control by calcium and mechanism of desensitization of phosphoenolpyruvate carboxykinase of Escherichia coli. *J. Biol. Chem.* **255**, 1399–1405

5.   Delbaere, L. T. J., Vandonselaar, M., Glaeske, D., Jabs, C., and Goldie, H. (1991) Crystallization of the Calcium-Activated Phosphoenolpyruvate Carboxykinase from Escherichia-Coli-K12. *J. Mol. Biol.* **219**, 593–594

6.   Sudom, A., Walters, R., Pastushok, L., Goldie, D., Prasad, L., Delbaere, L. T. J., and Goldie, H. (2003) Mechanisms of activation of phosphoenolpyruvate carboxykinase from Escherichia coli by Ca2+ and of desensitization by trypsin. *J. Bacteriol.* **185**, 4233–42

7.   Cotelesage, J. J. H., Puttick, J., Goldie, H., Rajabi, B., Novakovski, B., and Delbaere, L. T. J. (2007) How does an enzyme recognize CO2? *Int. J. Biochem. Cell Biol.* **39**, 1204–10

8.   Johnson, T. A., and Holyoak, T. (2010) Increasing the conformational entropy of the omega-loop lid domain in phosphoenolpyruvate carboxykinase impairs catalysis and decreases catalytic fidelity. *Biochemistry*. **49**, 5176–5187

9.   Johnson, T. A., and Holyoak, T. (2012) The Ω-loop lid domain of phosphoenolpyruvate carboxykinase is essential for catalytic function. *Biochemistry*. **51**, 9547–59

10.   Hundal, R. S., Krssak, M., Dufour, S., Laurent, D., Lebon, V., Chandramouli, V., Inzucchi, S. E., Schumann, W. C., Petersen, K. F., Landau, B. R., and Shulman, G. I. (2000) Mechanism by which metformin reduces glucose production in type 2 diabetes. *Diabetes*. **49**, 2063–2069

11.   Fell, D. (1997) *Understanding the control of metabolism*, 10.1016/S0307-4412(97)87557-7

12.   Burgess, S. C., He, T., Yan, Z., Lindner, J., Sherry, A. D., Malloy, C. R., Browning, J. D., and Magnuson, M. A. (2007) Cytosolic Phosphoenolpyruvate Carboxykinase Does Not Solely Control the Rate of Hepatic Gluconeogenesis in the Intact Mouse Liver. *Cell Metab.* **5**, 313–320

13.   Stark, R., and Kibbey, R. G. (2014) The mitochondrial isoform of phosphoenolpyruvate carboxykinase (PEPCK-M) and glucose homeostasis: Has it been overlooked? *Biochim.*

Biophys. Acta - Gen. Subj. **1840**, 1313–1330

14. Méndez-Lucas, A., Hyroššová, P., Novellasdemunt, L., Viñals, F., and Perales, J. C. (2014) Mitochondrial PEPCK is a Pro-Survival, ER-Stress Response Gene Involved in Tumor Cell Adaptation to Nutrient Availability. *J. Biol. Chem.* **289**, 22090–22102

15. Leithner, K., Hrzenjak, A., Trötzmüller, M., Moustafa, T., Köfeler, H. C., Wohlkoenig, C., Stacher, E., Lindenmann, J., Harris, L., Olschewski, A., and Olschewski, H. (2015) PCK2 activation mediates an adaptive response to glucose depletion in lung cancer. *Oncogene.* **34**, 1044–1050

16. Vincent, E. E., Sergushichev, A., Griss, T., Gingras, M. C., Samborska, B., Ntimbane, T., Coelho, P. P., Blagih, J., Raissi, T. C., Choinière, L., Bridon, G., Loginicheva, E., Flynn, B. R., Thomas, E. C., Tavaré, J. M., Avizonis, D., Pause, A., Elder, D. J. E., Artyomov, M. N., and Jones, R. G. (2015) Mitochondrial Phosphoenolpyruvate Carboxykinase Regulates Metabolic Adaptation and Enables Glucose-Independent Tumor Growth. *Mol. Cell.* **60**, 195–207

17. Montal, E. D., Dewi, R., Bhalla, K., Ou, L., Hwang, B. J., Ropell, A. E., Gordon, C., Liu, W. J., DeBerardinis, R. J., Sudderth, J., Twaddel, W., Boros, L. G., Shroyer, K. R., Duraisamy, S., Drapkin, R., Powers, R. S., Rohde, J. M., Boxer, M. B., Wong, K. K., and Girnun, G. D. (2015) PEPCK Coordinates the Regulation of Central Carbon Metabolism to Promote Cancer Cell Growth. *Mol. Cell.* **60**, 571–583

18. DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G., and Thompson, C. B. (2008) The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metab.* **7**, 11–20

19. Balsa-Martinez, E., and Puigserver, P. (2015) Cancer Cells Hijack Gluconeogenic Enzymes to Fuel Cell Growth. *Mol. Cell.* **60**, 509–511

20. Nelson, D. L., and Cox, M. M. (2005) *Lehninger Principles of Biochemistry*, 4/e, W. H. Freeman and Company, New York, NY

21. Matte, A., Goldie, H., Sweet, R. M., and Delbaere, L. T. (1996) Crystal structure of Escherichia coli phosphoenolpyruvate carboxykinase: a new structural family with the P-loop nucleoside triphosphate hydrolase fold. *J. Mol. Biol.* **256**, 126–143

22. Tari, L., Matte, A., and Pugazhenthi, U. (1996) Snapshot of an enzyme reaction intermediate in the structure of the ATP–Mg2+–oxalate ternary complex of Escherichia coli PEP carboxykinase. *Nat. Struct. Biol.* **3**, 355–363

23. Sudom, A. M., Prasad, L., Goldie, H., and Delbaere, L. T. (2001) The phosphoryl-transfer mechanism of Escherichia coli phosphoenolpyruvate carboxykinase from the use of AlF(3). *J. Mol. Biol.* **314**, 83–92

24. Sugahara, M., Ohshima, N., Ukita, Y., Sugahara, M., and Kunishima, N. (2005) Structure of ATP-dependent phosphoenolpyruvate carboxykinase from Thermus thermophilus HB8

showing the structural basis of induced fit and thermostability. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 1500–1507

25. Leduc, Y. A., Prasad, L., Laivenieks, M., Zeikus, J. G., and Delbaere, L. T. J. (2005) Structure of PEP carboxykinase from the succinate-producing Actinobacillus succinogenes: A new conserved active-site motif. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 903–912

26. Cotelesage, J. J. H., Prasad, L., Zeikus, J. G., Laivenieks, M., and Delbaere, L. T. J. (2005) Crystal structure of Anaerobiospirillum succiniciproducens PEP carboxykinase reveals an important active site loop. *Int. J. Biochem. Cell Biol.* **37**, 1829–1837

27. Aich, S., Prasad, L., and Delbaere, L. T. J. (2008) Structure of a GTP-dependent Bacterial PEP-carboxykinase from Corynebacterium glutamicum. *Int. J. Biochem. Cell Biol.* **40**, 1597–1603

28. Trapani, S., Linss, J., Goldenberg, S., Fischer, H., Craievich, A. F., and Oliva, G. (2001) Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from Trypanosoma cruzi at 2 Å resolution. *J. Mol. Biol.* **313**, 1059–1072

29. Machová, I., Snášel, J., Dostál, J., Brynda, J., Fanfrlík, J., Singh, M., Tarábek, J., Vaněk, O., Bednárová, L., and Pichová, I. (2015) Structural and Functional Studies of Phosphoenolpyruvate Carboxykinase from Mycobacterium tuberculosis. *PLoS One*. **10**, e0120682

30. Carlson, G. M., and Holyoak, T. (2009) Structural insights into the mechanism of phosphoenolpyruvate carboxykinase catalysis. *J. Biol. Chem.* **284**, 27037–27041

31. Balan, M. D., McLeod, M. J., Lotosky, W. R., Ghaly, M., and Holyoak, T. (2015) Inhibition and Allosteric Regulation of Monomeric Phosphoenolpyruvate Carboxykinase by 3-Mercaptopicolinic Acid. *Biochemistry*. **54**, 5878–5887

32. Sullivan, S. M., and Holyoak, T. (2007) Structures of rat cytosolic PEPCK: Insight into the mechanism of phosphorylation and decarboxylation of oxaloacetic acid. *Biochemistry*. **46**, 10078–10088

33. Sullivan, S. M., and Holyoak, T. (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13829–13834

34. Stiffin, R. M., Sullivan, S. M., Carlson, G. M., and Holyoak, T. (2008) Differential inhibition of cytosolic PEPCK by substrate analogues. Kinetic and structural characterization of inhibitor recognition. *Biochemistry*. **47**, 2099–2109

35. Hanson, R. W. (2009) Thematic minireview series: A perspective on the biology of phosphoenolpyruvate carboxykinase 55 years after its discovery. *J. Biol. Chem.* **284**, 27021–27023

36. Hura, G. L., Budworth, H., Dyer, K. N., Rambo, R. P., Hammel, M., McMurray, C. T., and Tainer, J. A. (2013) Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods*. **10**, 453–4

37. Matte, A., Tari, L. W., and Delbaere, L. T. (1998) How do kinases transfer phosphoryl groups? *Structure*. **6**, 413–419

38. Delbaere, L. T. J., Sudom, A. M., Prasad, L., Leduc, Y., and Goldie, H. (2004) Structure/function studies of phosphoryl transfer by phosphoenolpyruvate carboxykinase. *Biochim. Biophys. Acta*. **1697**, 271–8

39. Holyoak, T., Sullivan, S. M., and Nowak, T. (2006) Structural insights into the mechanism of PEPCK catalysis. *Biochemistry*. **45**, 8254–63

40. Fessner, W. D. (1998) Enzyme mediated C-C bond formation. *Curr. Opin. Chem. Biol.* **2**, 85–97

41. Bugg, T. D. H. (2012) Enzymatic Carbon-Carbon Bond Formation. in *Introduction to Enzyme and Coenzyme Chemistry*, pp. 148–180, John Wiley & Sons, Ltd, Chichester, UK, 10.1002/9781118348970.ch7

42. Jeong, J. Y., Yim, H. S., Ryu, J. Y., Lee, H. S., Lee, J. H., Seen, D. S., and Kang, S. G. (2012) One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics Studies. *Appl. Environ. Microbiol.* **78**, 5440–5443

43. Classen, S., Hura, G. L., Holton, J. M., Rambo, R. P., Rodic, I., McGuire, P. J., Dyer, K., Hammel, M., Meigs, G., Frankel, K. A., and Tainer, J. A. (2013) Implementation and performance of SIBYLS: A dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **46**, 1–13

44. Kabsch, W. (2010) Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132

45. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674

46. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221

47. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501

48. Schrödinger LLC The PyMOL Molecular Graphics System, Version 1.8

49. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **51**, 2778–2786

50. Chemical Computing Group Inc. Molecular Operating Environment (MOE), 2013.08 Ed.

51. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **65**, 712–725

52. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174

53. Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280

54. Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **16**, 1357–1377

55. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, Jr., J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. (2004) Gaussian 03, revision C.02. Gaussian, Inc., Wallingford, CT.

56. Li, P., and Merz, K. M. (2016) MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **56**, 599–604

57. Kleywegt, G. J. (2006) Crystallographic refinement of ligand complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 94–100

58. Liebeschuetz, J., Hennemann, J., Olsson, T., and Groom, C. R. (2012) The good, the bad and the twisted: A survey of ligand geometry in protein crystal structures. *J. Comput. Aided. Mol. Des.* **26**, 169–183

59. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) Electronic ligand builder and optimization workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 1074–1080

60. Debreczeni, J. É., and Emsley, P. (2012) Handling ligands with Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 425–430

61. Emsley, P. (2017) Tools for ligand validation in Coot. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 203–210

62. Tickle, I. J. (2012) Statistical quality indicators for electron-density maps. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 454–467

63. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. Sect. A*. **42**, 140–149

64. Hodel, A., Kim, S. -H, and Brünger, A. T. (1992) Model bias in macromolecular crystal structures. *Acta Crystallogr. Sect. A*. **48**, 851–858

65. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Adams, P. D., Read, R. J., Zwart, P. H., and Hung, L. W. (2008) Iterative-build OMIT maps: Map improvement by iterative model building and refinement without model bias. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 515–524

66. Perry, J. J. P., Shin, D. S., Getzoff, E. D., and Tainer, J. A. (2010) The structural biochemistry of the superoxide dismutases. *Biochim. Biophys. Acta - Proteins Proteomics*. **1804**, 245–262

67. Borgstahl, G. E. O., Parge, H. E., Hickey, M. J., Beyer, W. F., Hallewell, R. A., and Tainer, J. A. (1992) The structure of human mitochondrial manganese superoxide dismutase reveals a novel tetrameric interface of two 4-helix bundles. *Cell*. **71**, 107–118

68. Williams, R. S., Moncalian, G., Williams, J. S., Yamada, Y., Limbo, O., Shin, D. S., Groocock, L. M., Cahill, D., Hitomi, C., Guenther, G., Moiani, D., Carney, J. P., Russell, P., and Tainer, J. A. (2008) Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair. *Cell*. **135**, 97–109

69. Posner, B. A., Li, L., Bethell, R., Tsuji, T., and Benkovic, S. J. (1996) Engineering specificity for folate into dihydrofolate reductase from Escherichia coli. *Biochemistry*. **35**, 1653–1663

70. Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. a, Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design. *Nature*. **453**, 190–195

71. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. a (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285

72. Pelikan, M., Hura, G., and Hammel, M. (2009) Structure and flexibility within proteins as

identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **28**, 174–189

73. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., Tsutakawa, S. E., Jenney, F. E., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S.-J., Scott, J. W., Dillard, B. D., Adams, M. W. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods*. **6**, 606–12

74. Hammel, M. (2012) Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *Eur. Biophys. J.* **41**, 789–799

75. Dainese, E., Sabatucci, A., and Cozzani, I. (2005) Small Angle X-Ray Scattering: A Powerful Tool to Analyze Protein Conformation in Solution. *Curr. Org. Chem.* **9**, 1781–1800

76. Caracciolo, G., Amiconi, G., Bencivenni, L., Boumis, G., Caminiti, R., Finocchiaro, E., Maras, B., Paolinelli, C., and Congiu Castellano, A. (2001) Conformational study of proteins by SAXS and EDXD: The case of trypsin and trypsinogen. *Eur. Biophys. J.* **30**, 163–170

77. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974

78. Doniach, S., and Lipfert, J. (2009) *Use of small angle X-ray scattering (SAXS) to characterize conformational states of functional RNAs.*, 1st Ed., Elsevier Inc., 10.1016/S0076-6879(09)69011-X

79. Mertens, H. D. T., and Svergun, D. I. (2010) Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172**, 128–141

80. Poulain, P. Principal Axes. https://github.com/pierrepo/principal_axes

81. Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and De Hoon, M. J. L. (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. **25**, 1422–1423

82. Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011) The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30

83. Hunter, J. D. (2007) Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 99–104

84. Dyer, K. N., Hammel, M., Rambo, R. P., Tsutakawa, S. E., Rodic, I., Classen, S., Tainer, J. A., and Hura, G. L. (2014) High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach. in *Structural Genomics: General Applications* (Chen, Y. W. ed), pp. 245–258, Humana Press, Totowa, NJ, 10.1007/978-1-62703-691-7_18

85.   Rambo, R. P. (2015) Scatter

86.   Webb, B., and Sali, A. (2014) Comparative Protein Structure Modeling Using
       MODELLER. in *Current Protocols in Bioinformatics*, p. 5.6.1-5.6.32, John Wiley &
       Sons, Inc., Hoboken, NJ, USA, **2014**, 5.6.1-5.6.32

87.   Igor Pro, version 6.37. WaveMetrics, Inc.

88.   Matte, A., Tari, L. W., Goldie, H., and Delbaere, L. T. J. (1997) Structure and Mechanism
       of Phosphoenolpyruvate Carboxykinase. *J. Biol. Chem.* **272**, 8105–8108

89.   Holton, J. M., Classen, S., Frankel, K. A., and Tainer, J. A. (2014) The R-factor gap in
       macromolecular crystallography: an untapped potential for insights on accurate structures.
       FEBS J. **281**, 4046–4060

90.   Csermely, P., Palotai, R., and Nussinov, R. (2010) Induced fit, conformational selection
       and independent dynamic segments: An extended view of binding events. *Trends
       Biochem. Sci.* **35**, 539–546

91.   Drummond, M. L., Wilson, A. K., and Cundari, T. R. (2012) Nature of protein-CO2
       interactions as elucidated via molecular dynamics. *J. Phys. Chem. B.* **116**, 11578–11593

92.   Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., and Kern, D. (2007)
       A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature.* **450**,
       913–916

93.   Lai, Y., Hura, G. L., Dyer, K. N., Tang, H. Y. H., Tainer, J. A., and Yeates, T. O. (2016)
       Designing and defining dynamic protein cage nanoassemblies in solution. *Sci. Adv.* **2**,
       e1501855

94.   Winter, A., Higueruelo, A. P., Marsh, M., Sigurdardottir, A., Pitt, W. R., and Blundell, T.
       L. (2012) Biophysical and computational fragment-based approaches to targeting protein–
       protein interactions: applications in structure-guided drug discovery. *Q. Rev. Biophys.* **4**,
       1–44

95.   Murray, C. W., and Blundell, T. L. (2010) Structural biology in fragment-based drug
       design. *Curr. Opin. Struct. Biol.* **20**, 497–507

96.   Blundell, T. L., Sibanda, B. L., Montalvao, R. W., Brewerton, S., Chelliah, V., Worth, C.
       L., Harmer, N. J., Davies, O., and Burke, D. (2006) Structural biology and bioinformatics
       in drug design: opportunities and challenges for target identification and lead discovery.
       Philos. Trans. R. Soc. B Biol. Sci. **361**, 413–423

97.   Intergovernmental Panel on Climate Change (2014) *Climate Change 2014 Synthesis
       Report*, 10.1017/CBO9781107415324

98.   Yeoh, H.-H., Badger, M. R., and Watson, L. (1981) Variations in Kinetic Properties of
       Ribulose-1,5-bisphosphate Carboxylases among Plants. *Plant Physiol.* **67**, 1151–1155

99.   Yeoh, H.-H., Badger, M. R., and Watson, L. (1980) Variations in Km(CO2) of Ribulose-1,5-bisphosphate Carboxylase among Grasses. *Plant Physiol.* **66**, 1110–1112

100. Drummond, M. L., Cundari, T. R., and Wilson, A. K. (2012) Protein-based carbon capture: progress and potential. *Greenh. Gases Sci. Technol.* **2**, 223–238

101. Cundari, T. R., Wilson, A. K., Drummond, M. L., Gonzalez, H. E., Jorgensen, K. R., Payne, S., Braunfeld, J., De Jesus, M., and Johnson, V. M. (2009) CO 2 -Formatics: How Do Proteins Bind Carbon Dioxide? *J. Chem. Inf. Model.* **49**, 2111–2115

102. Harrison, S. C. (2007) Comments on the NIGMS PSI. *Structure*. **15**, 1344–1346