UNIVERSITY OF CALIFORNIA
RIVERSIDE

Biological Structure and Dynamics at the Interface of Experiment and Theory

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Robert Patrick Young

August 2016

Dissertation Committee
Dr. Leonard J. Mueller, Chairperson
Dr. Chia-en A. Chang
Dr. Francisco Zaera

The Dissertation of Robert Patrick Young is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

I would like to thank my research adviser Len Mueller for his mentorship as well as professional and personal support. I have enjoyed the research I've been a part of here so much that graduating is kind of bittersweet – especially considering all of the exciting possibilities that lay ahead in PLP research. To my lab mates and friends, Beth Caulkins, Ryan Kudla, Tommy Neubauer, Chen Yang, and Lingchao Zhu thank you, I will never forget the times we had together and I hope we always keep in touch. Thank you to Dallas Rabenstein and Cindy Larive who were excellent role models and mentors from the time I first set foot on UCR's campus. Thanks as well to Dan Borchardt who provided so much help in getting many of my experiments going. I would also like to recognize Chia-en Chang, Mike Dunn, Eduardo Hilario, Consuelo Beecher, Chris Roberts, Daryl Bulloch, Derek Langeslay, and Wanli You for all of their assistance in many regards. Thank you to Anna De Angelis and Chris Grant at UCSD as well for their help in the ^{17}O project. Finally, to my family, my parents especially, who have been selfless in their support for me in achieving this goal.

The National Institutes for Health and the National Science Foundation including the NSF Graduate Research Fellowship Program, Award DGE-0813967.

The text of this dissertation, in part or in full, is a reprint of the material as they appear in the following publications:

Chapter 1: Caulkins, B. G.; Yang, C.; Hilario, E.; Fan, L.; Dunn, M. F.; Mueller, L. J. *Biochem. Biophys. Acta.* **2015**, *1854*, 9, 1194-1199.

Chapter 2: Young, R. P.; Caulkins, B. G.; Borchardt, D.; Bulloch, D. N.; Larive, C. K.; Dunn, M. F.; Mueller, L. J. *Angew. Chem. Int. Ed.* **2016**, *55*, 1350-1354.

Chapter 3: Caulkins, B. G.; Bastin, B.; Yang, C.; Neubauer, T. J.; Young, R. P.; Hilario, E.; Huang, Y.-M. M.; Chang, C.-E. A; Fan, L.; Dunn, M. F.; Marsella, M. J.; Mueller, L. J. *J. Am. Chem. Soc.* **2014**, *136*, 12824-12827.

ABSTRACT OF THE DISSERTATION

Biological Structure and Dynamics at the Interface of Experiment and Theory

by

Robert Patrick Young

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, August 2016
Dr. Leonard J. Mueller, Chairperson

The research presented in this thesis reflects advances in the adaptation and application of NMR crystallography – the synergistic combination of X-ray crystallography, NMR spectroscopy, and computational methods – to the active sites of enzymes. The goal is to construct highly detailed and chemically rich structures of the enzyme active site. The direct targets of this work are kinetically competent, quasi-stable intermediates along the reaction pathway of the pyridoxal-5'-phosphate dependent enzyme tryptophan synthase.

NMR crystallography relies on the availability of both X-ray crystal structures and NMR data. This study makes use of multiple NMR restraints, including ^{13}C, ^{15}N, and ^{31}P NMR chemical shifts measured by collaborators in the Mueller group, and adds to these by including ^{17}O as an active site reporter. The acid-base catalysis featured in many enzyme mechanisms relies on oxygen as a key atomic species in amphiprotic functional

groups, making it a potentially important probe nucleus for NMR crystallography in enzyme active sites. Yet oxygen is not considered a standard biological NMR probe given the perceived difficulties related to its quadrupolar nuclear status. In this work we utilize solution state ^{17}O Quadrupole Central Transition NMR to directly measure ^{17}O chemical shifts and probe ionization states of enzyme bound intermediates.

The NMR shifts are combined with X-ray crystallography to build computational models of the active site consistent with both the NMR and X-ray data. This thesis provides an exposition of the synergy of the NMR crystallographic approach applied to important intermediates of tryptophan synthase's catalytic cycle. The results of this approach demonstrate the unprecedented level of structural detail that can be revealed, most remarkably for proton locations, within the active site. This is highlighted in particular by the characterization of proton mediated tautomerization in the "quinonoid/carbanion" intermediate, where first principle calculations of charge allow the mechanistic implications of the observed protonation states to be fully appreciated.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Overview of Dissertation Research

1.1.1 Introduction

The research presented in this thesis reflects advances in the adaptation and application of NMR crystallography – the synergistic combination of X-ray crystallography, NMR spectroscopy, and computational methods – to the active sites of enzymes. The goal is to construct highly detailed and chemically rich structures of the enzyme active site. The direct targets of this work are kinetically competent, quasi-stable intermediates along the reaction pathway of the pyridoxal-5'-phosphate dependent enzyme tryptophan synthase. The results of this thesis demonstrate the unprecedented level of structural detail that can be revealed, most remarkably for proton locations, within the active site. This will be highlighted in particular by the characterization of proton mediated tautomerization in the "quinonoid/carbanion" intermediate, where first principle calculations of charge allow the mechanistic implications of the observed protonation states to be fully appreciated.

NMR crystallography relies on the availability of both X-ray crystal structures and NMR data. This thesis will make use of multiple NMR restraints, including ^{13}C, ^{15}N, and ^{31}P NMR chemical shifts measured by collaborators in the Mueller group, while also adding the additional NMR probe nucleus ^{17}O. The acid-base catalysis featured in many

enzyme mechanisms relies on oxygen as a key atomic species in amphiprotic functional groups, making it a potentially important probe nucleus for NMR crystallography in enzyme active sites. Yet oxygen is not considered a standard biological NMR probe given the perceived difficulties related to its quadrupolar nuclear status.[1] In Chapter 2, the application of ^{17}O Quadrupole Central Transition NMR, which capitalizes on the favorable relaxation properties of its central transition in the limit of slow isotropic motion, to directly measure ^{17}O chemical shifts and probe ionization states of enzyme bound intermediates is demonstrated.[2] Chapter 3 then makes use of this data along with other NMR shifts and the structural framework from X-ray crystallography, to build computational models of the active site consistent with both the NMR and X-ray data. This provides an exposition of the synergy of the NMR crystallographic approach applied to an important intermediate of tryptophan synthase's catalytic cycle, and the result implicates a dynamic equilibrium tautomerization. Finally, Chapter 4 explores the mechanistic implications of the aforementioned exchange using natural population analysis[3] to predict atomic charge. The resulting analysis points to one tautomer as being more catalytically important and also indicates that previously hypothesized protonation states in the active site would likely be detrimental to the desired chemistry.

1.1.2 Description of the Research Problem

Enzymes are biological catalysts and have evolved to perform many complex chemical transformations with exquisite specificity and incredible efficiency. A key to their catalytic prowess lies in the structure of their active sites which preferentially bind

specific substrates in the proper orientation for catalytic residues to carry out the required reaction and is typically assisted by various scales of conformational change in the macromolecule.[4] Complex catalytic mechanisms often take place over the course of several elementary steps.[4] General acid-base catalysis often plays a key role in selecting reaction specificity and enzymes that depend on the small organic cofactor pyridoxal-5'-phosphate, PLP or vitamin-B6, present particularly apt examples.[5-8] PLP is incorporated in many of the enzymes responsible for the metabolism of amine containing biomolecules such as amino acids.[5-8] This single cofactor is utilized to carry out myriad types of different chemical reactions while generally assuming the role as an electrophilic catalyst.[5-8] At the heart of PLP's wide ranging abilities is a complex interplay of proton transfer mediated tautomerization that "tunes" it for the appropriate reaction while minimizing the probability of undesirable side reactions.[8] The unique structure of the active site provides the template for this reaction specificity by positioning key residues to interact with both substrate and cofactor. Understanding how these structural and functional relationships synchronize is made difficult by the large enzyme sizes and rapid reaction rates with natural substrates.[7]

Enzymologists have employed an arsenal of analytical tools such as optical spectroscopy, kinetic isotope effects, site-directed mutagenesis, modified cofactors, substrate and metabolite analogues, and inhibitors in an effort to characterize their catalytic mechanisms.[9-19] X-ray crystallography has proven invaluable in providing the coarse 3-D framework detailing the proximity of active site residues to cofactors and substrates from which much can be inferred.[20, 21] The inability of X-ray crystallography

to locate protons, however, is a drawback with respect to the important role that general acid-base catalysis plays. NMR is unrivalled in its ability to probe such chemical details requiring atomic resolution such as hybridization and ionization states as well as dynamic events associated with conformational or chemical exchange. Solid-state NMR has surmounted the molecular weight issues of the large enzymes;[22, 23] however, the signal overlap from a uniformly enriched protein generally limits active site probes to the substrate and cofactor itself. Interpretation of the NMR chemical shifts thus requires the structural context provided by X-ray crystallography. Ambiguities often remain even by the combination of these highly complementary experimental methods. The continuously advancing approaches offered by computational chemistry provides the vital link by enabling models of the enzyme active site to be constructed in various configurations from which NMR observables can be predicted and compared with experimental data.[22, 24] The synergistic combination of these three powerful techniques is termed NMR crystallography and was originally applied in unison for the refinement of crystal structures of small organic molecules.[25, 26] The research presented herein seeks to advance the adaptation of the NMR crystallographic approach in building the most highly detailed and chemically rich structures of quasi-stable intermediates in the 143 kDa PLP-dependent enzyme tryptophan synthase.

The remainder of this chapter will cover three major topics. First, a brief history of PLP is provided and includes comparisons of several classes of PLP enzymes and the results of recent studies that have more fully characterized tautomerization in PLP, primarily from model compound studies, and have led to the challenge of some

longstanding assumptions with regards to its operation in catalysis. Next, tryptophan synthase, the direct subject of this dissertation is introduced along with the results of a previous NMR crystallography study. Finally, NMR occupies a central theme in this dissertation and a brief summary of its basic aspects are provided.

1.2 Pyridoxal-5'-phosphate and Pyridoxal-5'phosphate Dependent Enzymes

1.2.1 Early History of Prydioxal-5'-phosphate Research

Pyridoxal-5'-phosphate, PLP or vitamin B-6, depicted in Figure 1.1, has been the subject of active research for over 70 years. Isolation and identification of vitamin-B6 as an essential nutrient from metabolic studies predated the determination of its bioactive structural form.[27] Early studies by several groups led to the determination that PLP and the closely related pyridoxamine-5'-phosphate, PMP (Figure 1.1, middle), catalyzed

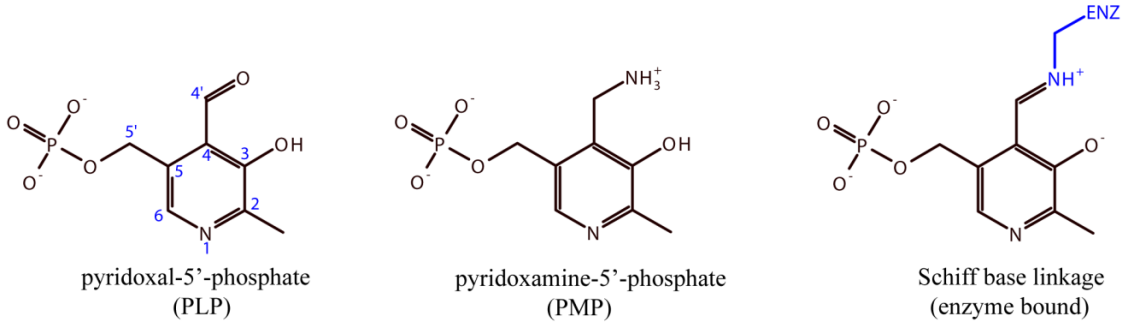


Figure 1.1. Pyridoxal-5'-phosphate, PLP, pyridoxamine-5'-phosphate, PMP, and the canonical Schiff base linkage, internal aldimine, between PLP and the N^{ε} of a sidechain lysine amino group in the enzyme bound state.

reversible non-enzymic reactions that gave products consistent with those observed by Braunstein et al. in transaminase (or aminotransferase) enzymes discovered less than a decade earlier.[28-30] Snell was the first to suggest that PLP and closely related forms may

be cofactors of the aminotransferase enzymes.[30, 31] Paralleling early investigations of PLP alone and in related model compounds, enzymologists were discovering many more PLP-dependent enzymes catalyzing a staggeringly diverse number of reaction types involving amines.[32, 33] Some of the key reactions catalyzed by PLP involving amino acids alone include transamination,[32] α/β/γ elimination or substitution[34], aldol-type reactions, [35] decarboxylation, racemization,[33, 36] and more.[32, 33] PLP enzymes that participate in mechanisms involving radicals[37] and the phosphorylation of glycogen[38], are also known, but given their rather unique nature will not be discussed further. Currently, PLP dependent enzymes have been identified that catalyze over 140 distinct activities as defined by the Enzyme Commission and PLP is currently regarded as the most versatile organic cofactor known.[7, 39] Genetic codes for PLP dependent enzymes have been found in all free-living organisms, in fact, 1.5% of the genome of prokaryotes contain codes for PLP enzymes.[39]

1.2.2 Model Compounds and PLP Enzymes

PLP and derivatives were found to catalyze many of the same reactions observed in enzymes through careful control of typical reaction variables including pH, solvent, temperature, substrate concentration, and even chelation with metal ions.[35, 36, 40] The results of model compound studies demonstrated that PLP forms Schiff base complexes with amines, termed aldimines, at its C4' position eliminating the aldehyde oxygen as water in the process.[32] Similarly, PLP in enzymes is covalently bound to the enzyme via a covalent Schiff-base linkage with the sidechain amine N^{ε} of a lysine residue, termed the internal aldimine, when no substrates are present, see Figure 1.1.[7, 32] The bond to the

enzyme N^{ε} group is replaced by a substrate amino group upon its binding with an enzyme, termed an external aldimine.[6] Figure 1.2 shows examples of the (external) aldimine between PLP and an amino acid, the R representing an amino acid sidechain. The tautomers shown, the two resonance forms of the ketoenamine (Figure 1.2 A and B) and the enolimine tautomer (Figure 1.2, C), are the predominate tautomers involving proton transfer between the Schiff base nitrogen and the nearby phenolic oxygen.[32] Tautomerization plays an important role in PLP chemistry and will be discussed in more detail throughout.

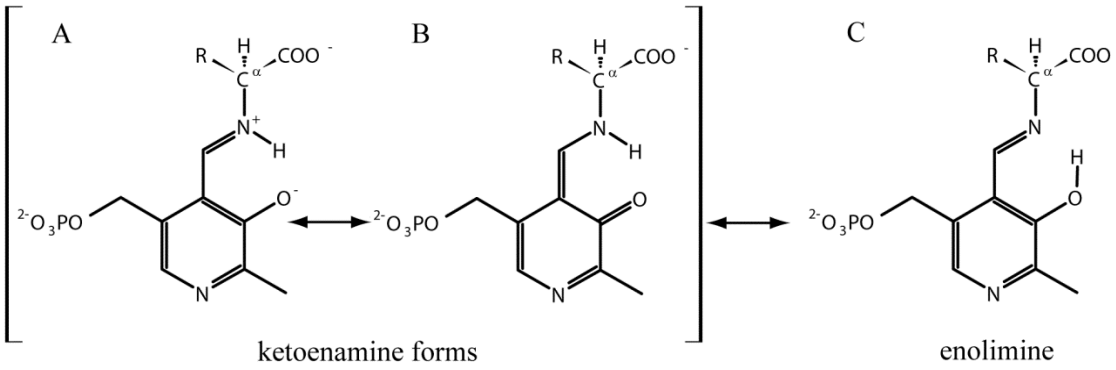


Figure 1.2. External aldimines formed between PLP and amino acid substrates. The ketoenamine (A and B) and enolimine (C) tautomers are the principle forms present in equilibrium.

In both enzymic and non-enzymic reactions of amino acids, the step following the formation of the Schiff base aldimine between PLP and substrate involves the heterolytic cleavage of any one of the bonds to the C^{α} except that of the C^{α}-N.[6, 7, 32, 41-43] It is from this point that PLP reaction types diverge but the bond to be cleaved was hypothesized by Dunathan to be positioned perpendicular to the PLP plane based on stereoelectronic considerations - later supported by first principles calculations and to date no exceptions

have been noted.[44, 45] The fact that enzyme catalyzed reactions proceed at rates considerably faster than those of model compounds and with exquisite specificity compared to non-enzymic reactions is easily attributable to the fact that enzymes have evolved to provide the structural compatibility to preferentially bind their substrates in the proper orientation for the required bond to be cleaved.[32] The heterolytic cleavage of one of the bonds results in a carbanion that can be stabilized through a number of resonance forms, see Figure 1.3.[32] Most famous of these is the quinonoid structure, Figure 1.3B, long thought to be a defining intermediate in PLP catalysis because of its ability to stabilize the resulting negative charge by delocalizing it through its highly conjugated π system further enhanced by the electron withdrawing ability of a

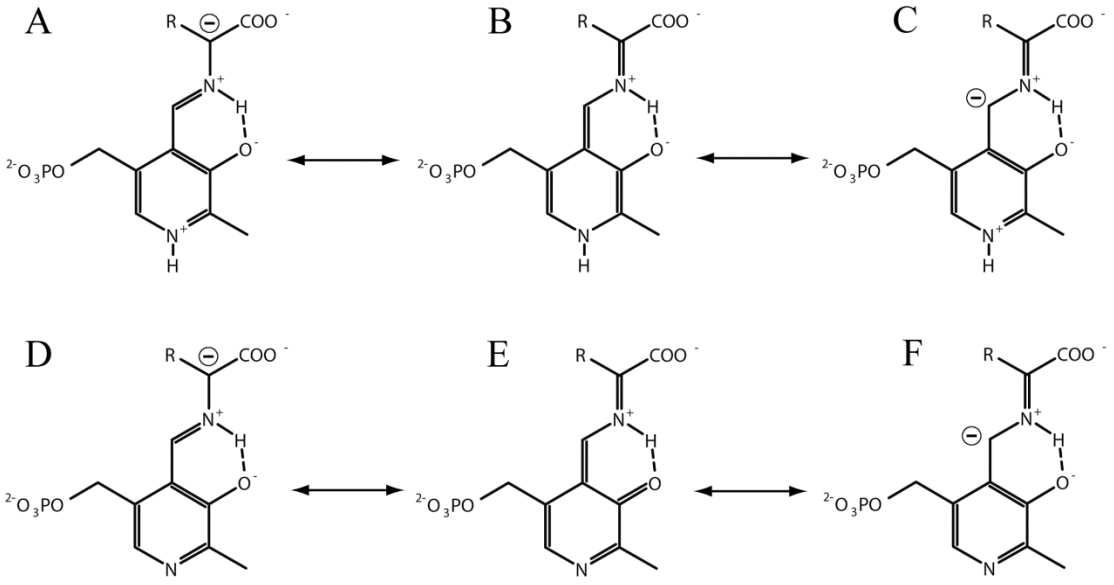


Figure 1.3. Resonance forms of carbanions following the cleavage of one of the bonds to the C^{α}. Structure B is the quinonoid form long thought an obligatory intermediate in PLP catalysis. The bottom row indicates resonance forms without a protonated pyridine ring nitrogen.

protonated pyridine ring nitrogen. [6, 7, 32, 41] Evidence for quinonoid formation was most commonly attributed to the observance of a characteristically narrow UV/Vis absorption maximum in the range 450 nm to 550 nm with a high molar extinction coefficient, ε , relative to other types of intermediates.[9] The bottom row of Figure 1.3, D-F, depicts possible resonance structures in the case of an unprotonated pyridine ring nitrogen. For many amino acid substrates the proton attached to C^{α} is abstracted in the step following formation of the initial external aldimine as depicted in Figure 1.3, but decarboxylases are an obvious exception.[46] As the resonance forms in Figure 1.3 depict, migration of the negative charge to the C4' position may occur, and reprotonation at that site is consistent with the reactions of transaminases in particular.[32, 47] For contrast, racemase enzymes are reprotonated at C^{α} from the face opposite that of the abstracted proton immediately whereas in other enzymes β or γ functional groups are subsequently eliminated in early stages.[41, 48] Many PLP reaction mechanisms involve several steps and intermediates such as carbanions/quinonoids may occur more than once in a given catalytic cycle.[49] It also needs to be pointed out that formation of a distinct quinonoid, or possibly carbanion, is not observed in all PLP dependent enzymes.[50]

A boon to the PLP research community occurred when the first three-dimensional structure of a PLP dependent enzyme, that of aspartate aminotransferase (AAT), was solved by X-ray crystallography in 1980.[20] This was very important as it revealed the identity and placement of the enzyme active site catalytic residues with respect to the cofactor. The second PLP dependent enzyme structure solved was that of tryptophan synthase, the subject of this dissertation, eight years later.[21] The number of X-ray

structures of PLP enzymes has since grown exponentially and includes the structures of a variety of intermediate forms with substrate and metabolite analogues. A search for X-ray structures of enzymes with PLP on the RCSB Protein Data Bank website today will return over 1,200 hits.

The availability of X-ray structures allowed PLP enzymes, previously organized by the type of reaction, to be further organized into families based on common protein structural motifs or fold-types and five are recognized today.[7] The fold-types, I-V, are named by the first enzyme of that type whose three-dimensional structure was determined and they are the Aspartate aminotransferase (fold-type I), Tryptophan synthase (fold-type II), Alanine racemase (fold-type III), D-amino acid aminotransferase (fold-type IV), and Glycogen phosphorylase (fold-type V) families.[7] It must be stressed that, despite the names of the families, all fold-types of PLP enzymes, with the exception of fold-type V, catalyze more than one type of reaction.[5] Also, with few exceptions, PLP enzymes exist as either homodimers or larger multi-subunit proteins with dihedral symmetry and have molecular weights exceeding 90 kDa.[7]

1.2.3 Review of Recent PLP Investigations

To elaborate upon the recurring theme in all PLP reactions, whether in the presence of the enzymes or alone, is that tautomerism and general acid-base chemistry, be it the transfer of protons amongst ionizable sites of the cofactor, substrate, and between either solvent or enzyme active site functional groups plays a critical role in catalysis and in directing reaction specificity. [51-57] Furthermore, studies have found that

the energy differences between many of the possible tautomeric forms are quite similar under conditions of neutral pH and that the pKa values of ionizable sites of PLP and its Schiff base linkage to substrates are interdependent.[51-58] NMR studies of model PLP compounds have provided the most direct experimental evidence of this and further insight has been brought about through the aid of advancing theoretical methods.[59-61] As an example, an NMR study by Sharif and coworkers observed the coupling of functional hydrogen bonds and protonation states in PLP wherein protonation of the pyridine ring nitrogen resulted in an increase of the pKa of the Schiff-base nitrogen and a concomitant decrease in the pKa at the phenolic oxygen site thus promoting transfer to the nearby Schiff-base nitrogen.[52] Quantum chemical calculations by Casasnovas and coworkers predicted large changes in the pKa values at the C4' and C^{α} sites with changing protonation states of the pyridine ring nitrogen, the phenolic oxygen, and the Schiff base nitrogen. These have important implications in reaction specificity as the basicity/acidity of these sites bears a direct influence on the type of overall chemistry carried out. Similar findings have been noted by others.[62, 63]

Studies of PLP catalysis in active enzymes have largely relied on optical spectroscopic methods, especially UV/Vis, which is an indispensable tool in enzymology given its sensitivity and ability to identify even short-lived intermediate states by changes in the absorption spectrum. The drawback to UV/Vis is that it does not report directly on individual protonation states.[64] Advances in X-ray crystallography that have led to an ever increasing number of PLP enzyme structures and their availability has proven critical, as previously mentioned, in providing the ability to locate the relative positions

of catalytic residues and the PLP/substrate complex. Knowing these positions guides investigators performing site-directed mutagenesis which allows testing the importance of individual residues of the enzyme in the active site by measuring the effects on reactions.[12, 14, 65-69] Kinetic isotope effects have also been used extensively and these aid in determining whether catalytic events are rate limiting, especially with regards to proton transfers.[17, 70-73]

As the number of X-ray crystal structures has grown, investigators took note that there were differences in the types of catalytic residues interacting with the key pyridine ring nitrogen. For example, in AAT, the first structure solved, the pyridine ring nitrogen is situated to interact with a carboxylate group of an aspartate sidechain, in TRPS it is a polar hydroxyl sidechain of serine residue, and in alanine racemase, AlaR, it is a basic arginine sidechain, see Figure 1.4 for examples.[20, 74, 75] For just over the past decade the

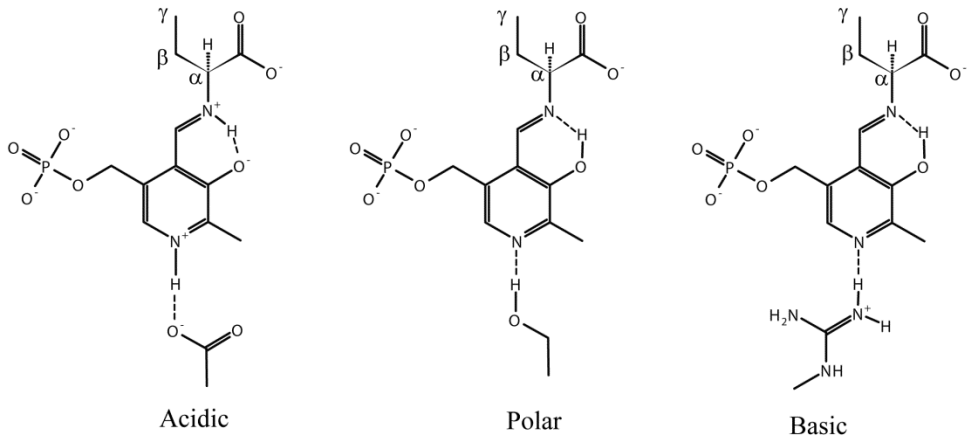


Figure 1.4. Functional enzyme groups observed to interact with the pyridine ring nitrogen of PLP.

assumption that a protonated pyridine ring nitrogen and the ability to form a "proper" quinonoid as a catalytic necessity has been increasingly questioned.[50, 76, 77] As an illustrative example, a recent study by Toney et al. tested the importance of the pyridine ring nitrogen on AAT, AlaR, and O-acetylserine sulfhydrylase (OASS) by supplying the apoenzymes with deaza-PLP.[76] As the name suggests, deaza-PLP lacks the pyridine ring nitrogen which is replaced by a carbon instead. Naturally, deaza- PLP is incapable of forming a quinonoid intermediate. The enzymes studied each differed in the both the type of overall reaction catalyzed as well as in the catalytic residue poised to interact with the PLP pyridine ring nitrogen. In AAT, which interacts with an acidic aspartate residue and catalyzes transamination, introduction of deaza-PLP essentially halted catalytic activity. In AlaR, which interacts with a basic arginine sidechain and catalyzes racemization of L-alanine to D-alanine, an estimated > 700 fold decrease in activity was noted.[76] OASS, which interacts with a polar serine hydroxyl sidechain, catalyzes a β-substitution reaction which involves elimination of the β-acetyl group early on. OASS also catalyzes β-elimination with β-chloroalanine, an analogue of its natural substrate, and the deaza-PLP form saw only a modest decrease in activity.[76] AlaR, when supplied with β-chloroalanine is also capable of performing a β elimination reaction that, like racemization, involves abstraction of the C^{α} proton after external aldimine formation. The deaza-PLP AlaR reaction with β-chloroalanine showed roughly the same modest decrease in activity of the β-elimination reaction as did OASS.[76] The results indicated that a pyridine ring nitrogen and the ability to form a distinct quinonoid intermediate was essential for AAT.[76] The racemization reaction was affected more than the β-elimination

reactions but, unlike AAT, catalytic activity was not halted in AlaR as a result of being supplied deaza-PLP.[76] It was proposed by Cook et al. that OASS proceeds through a concerted-E2 type mechanism that does not require formation of a distinct carbanion/quinonoid species and it was proposed by Toney and coworkers that this alternate pathway was the likely reason deaza-PLP did not affect the activities of OASS and AlaR in the β-elimination reactions.[50, 76]

The delicate balance of ionization states and their effects on reaction specificity that make PLP such a highly "tunable" catalyst has presented quite the challenge to organic chemists and enzymologists alike with regards to determining just how this reactivity is modulated. The examples from more recent studies make it clear that broad generalizations are inadequate and it is imperative that we gain access to the active site with the means to track proton exchange on the atomic level. Much of the groundwork for more detailed study has been set forth by the efforts of biochemists and enzymologists who have identified several means to render many intermediates "quasi-stable" by the use of inhibitors, substrate and metabolite analogues, and allosteric effectors. In addition to affording lifetimes sufficient to acquire X-ray crystal structures, several PLP enzyme intermediates are stable or even active in microcrystals making them amenable for solid state NMR.[22, 75, 78, 79] An initial report of the application of NMR crystallography in the study of catalytically active intermediates of TRPS was published in 2011 and it has continued to be a suitable test-bed for further refinement of the approach.[2, 22, 80-82]

1.3 NMR Crystallography in the Active Site of Tryptophan Synthase

1.3.1 Tryptophan Synthase

TRPS has been the focus of research for over six decades following its discovery and early characterization by Yanofsky.[82] TRPS is a 143 kDa protein consisting of two distinct subunits, α and β, and is arranged as a linear αββα tetramer, see Figure 1.5. The αβ subunits communicate their status through long-range allostery and both α and β subunits assume open (inactive) or closed (active) conformations during its catalytic cycle.[49, 83] Allostery is modulated by substrate binding and the identity of the cation occupying its monovalent cation site.[10, 84] TRPS was the first enzyme identified to sequester a metabolite by shuttling it from the α-site to the β-site via a tunnel, the details of which are still an active area of research.[10, 34, 83, 85-88] Several substrate or metabolite analogues and inhibitors have been identified that attenuate the rates of several key intermediate steps along its β-subunit reaction coordinate allowing for more detailed studies.[11] The *S. typhimurium* TRPS enzyme has favorable crystallization properties allowing for the acquisition of X-ray crystal structures of these quasi-stable intermediates.[11, 75, 83] Microcrystals of TRPS are catalytically active which makes it amenable for solid-state NMR studies.[21, 22, 75, 83]

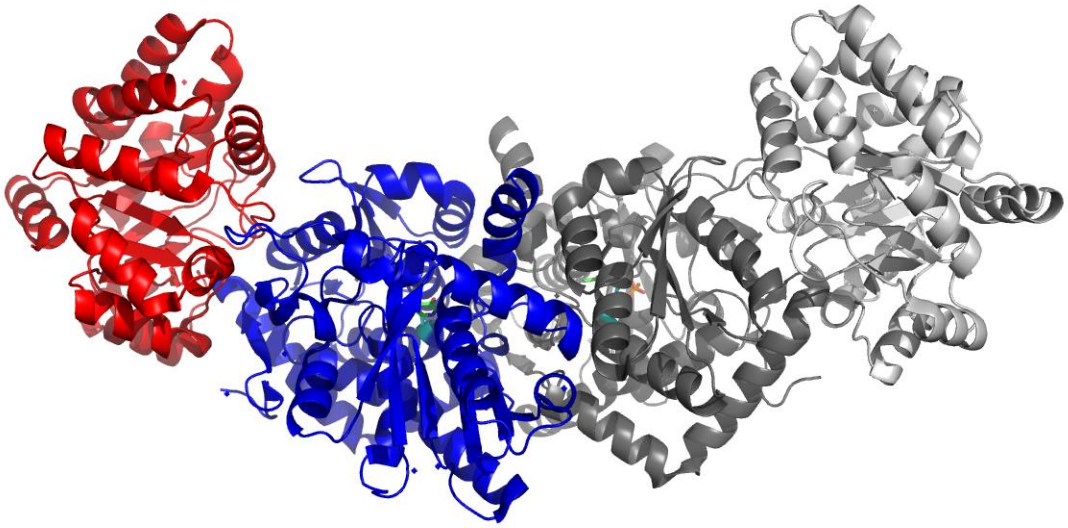


Figure 1.5. Tryptophan synthase full linear αββα tetrameric quaternary structure. For clarity, one α-subunit is colored red and one β-subunit is colored blue.

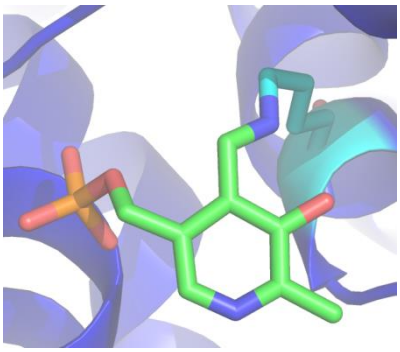


Figure 1.6. Internal aldimine linkage between PLP and βLys87.

1.3.2 Catalytic Mechanism of Tryptophan Synthase

Tryptophan synthase (TRPS) is responsible for catalyzing the final two major steps in the biosynthesis of the amino acid L-tryptophan involving the elimination of the hydroxyl group of L-serine and its replacement with indole. Each αβ subunit pair operates independently.[34] The α-subunit binds indole-glycerol-3-phosphate (IGP) and cleaves it to yield indole and glyceraldehyde-3-phosphate (G3P), see Figure 1.7.[89] The indole is required by the β-subunit at a specific step and is channeled to its active site via a 25Å tunnel.[85] The β-subunit contains the PLP cofactor and the mechanism at the β-site proceeds through several more intermediate steps.

The mechanism for the β-site catalytic cycle is shown in Figure 1.8. The corresponding UV/Vis maximum absorbance wavelengths for intermediates are provided where observed. In the absence of substrates, PLP is covalently bound to the sidechain N^{ε} of βLys87 in the internal aldimine E(Ain). The first step in catalysis involves replacement of the βLys87 bond with the amino group of L-serine and is postulated to proceed through a gem-diamine intermediate, $E(GD_1)$, to give the first external aldimine intermediate $E(Aex_1)$.[86] Abstraction of the proton attached at C^{α} by βLys87 leads to a short-lived first quinonoid intermediate, $E(Q_1)$, during which the β-hydroxyl group of the serine derived substrate is eliminated giving the aminoacrylate intermediate, E(A-A).[86] The E(A-A) form undergoes nucleophilic attack by the C3 of the indole generated at the α-site at the beginning of stage II of the β-subunit cycle leading to a $2^{nd}/3^{rd}$ quinonoid $E(Q_2)$ and $E(Q_3)$.[48] Reprotonation of the C^{α} by βLys87 yields a second external

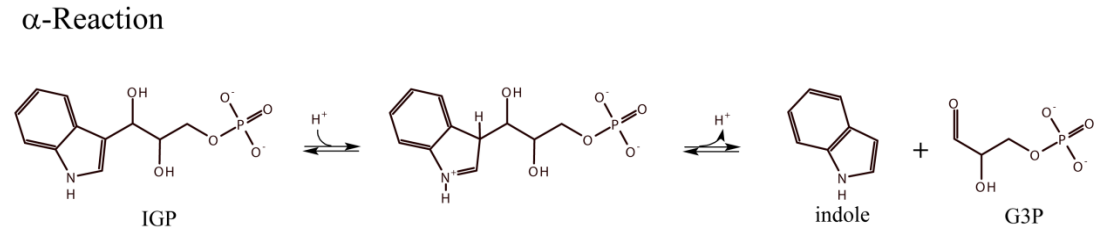


Figure 1.7. The α-site reaction in tryptophan synthase. IGP is cleaved to yield indole and G3P.

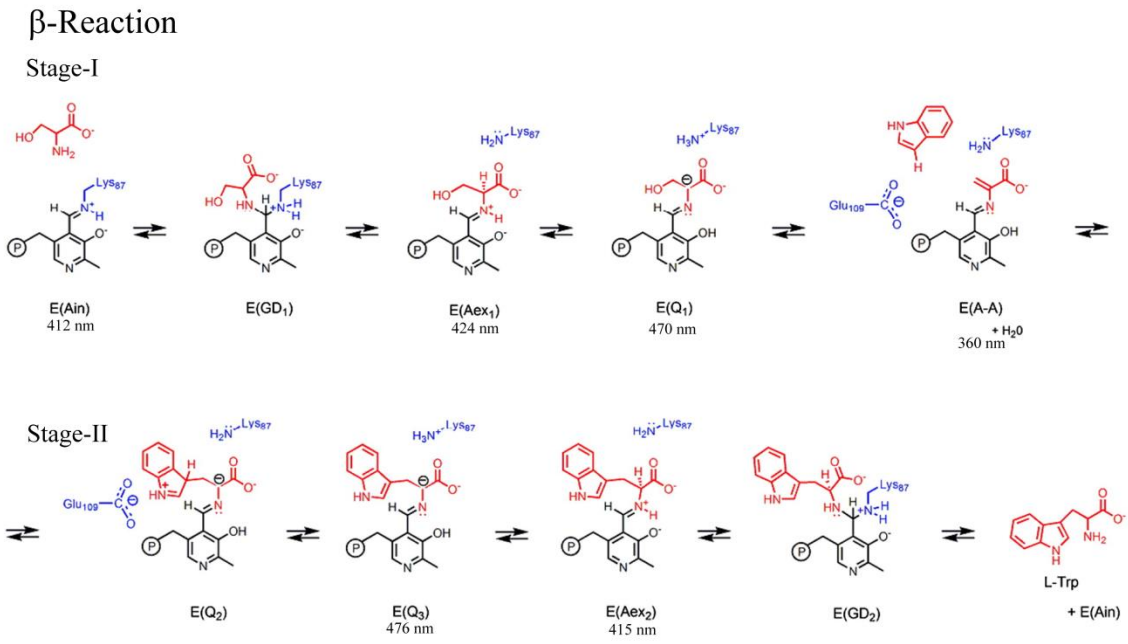


Figure 1.8. The β-site reactions in tryptophan synthase. Pyridoxal-5'-phosphate is depicted in black, enzyme in blue, and substrates and metabolites are indicated in red. Where available, UV/Vis maximum absorption wavelengths are provided. Figure adapted from "Catalytic roles of βLys87 in tryptophan synthase: ^{15}N solid state NMR studies", Vol. 1854 Iss. 9, Bethany G. Caulkins, Chen Yang, Eduardo Hilario, Li Fan, Michael F. Dunn, and Leonard J. Mueller, pp 1194-1199, ©2015, with permission from Elsevier.

aldimine, E(Aex_2), followed by a second gem-diamine intermediate, E(GD_2), that ultimately releases L-tryptophan and the regenerates the E(Ain) between PLP and βLys87.[86] Earlier hypotheses regarding the TRPS catalytic cycle assumed a protonated pyridine ring nitrogen.[86] Confirmation of its unprotonated status as shown resulted from ^{15}N solid state NMR measurements by Mueller group member B. G. Caulkins and first-principles calculations that are the subject of Chapter 3.[81]

1.3.3 Quasi-stable Intermediates

A number of substrate or metabolite analogues of TRPS natural substrates have been identified that render certain intermediate steps quasi-stable. In this research two β-site catalytic intermediates are investigated, the E(Q_3) and the α-aminoacrylate, E(A-A). The E(Q_3) intermediate is an especially desirable target for investigation by NMR crystallography given the importance of quinonoid/carbanion intermediates in PLP enzymology. The E$(Q)_3$ is quasi-stable when supplied with indoline, a close structural analogue of the indole metabolite, and L-serine, the natural β-site substrate, see Figure 1.9, and is referred to as the E$(Q_3)_{indoline}$ intermediate.[11] The lifetime of the intermediate is enhanced by the addition of N-(4'-trifluoro-methoxybenzenesulfonyl)-2-amino-ethyl phosphate (F9), an α- site substrate analogue that activates (closes) the α-subunit, and using Cs^+ as the monovalent cation where its binding aids in closing the β-subunit.[10, 84] The E$(Q_3)_{indoline}$ intermediate does go on to form the product dihydroiso-tryptophan (DIT), a novel amino acid.[11] By supplying the enzyme with 2-aminophenol (2AP) in place of indoline a much longer lived E$(Q_3)_{2AP}$ intermediate results, see Figure 1.9.[11, 83]

The presence of the quinonoid forms is initially confirmed by UV/Vis measurements which show very similar absorption characteristics as observed in the natural enzyme.[22] The X-ray crystal structures of these quinonoids also clearly indicate sp^2 hybridization of the C^{α} as expected for this form.[22, 83] The E(A-A) generated by adding the natural β-substrate L-serine is stable with only the F9 and Cs^+ allosteric effectors added,

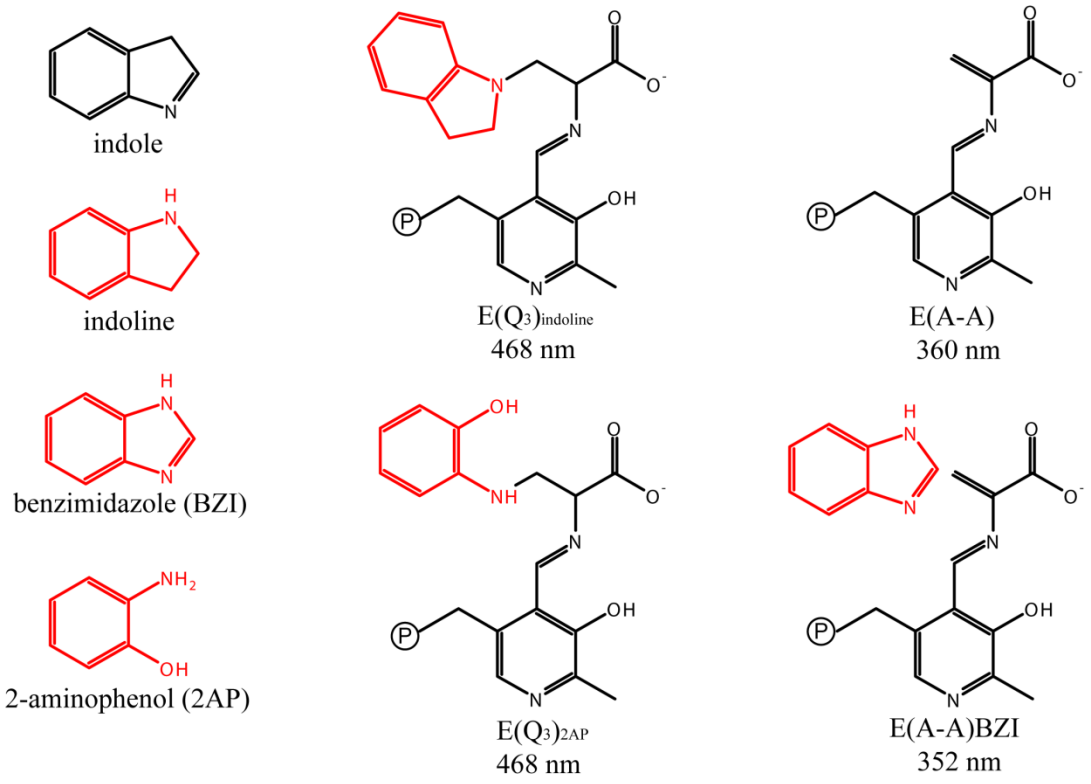


Figure 1.9. Quasi-stable intermediates of the tryptophan synthase β-catalytic cycle studied in this work. Two forms of the $E(Q_3)$ are stabilized by the addition of either 2-aminophenol (2AP) or indoline. The E(A-A) is relatively stable without supplying indole analogues and its lifetime can be extended by adding benzimidazole (BZI) which, as shown, does not bind to the L-serine derived C^{β} position.

however, a much longer lived form can be obtained by adding the non-reactive benzimidazole (BZI) indole analogue to form the E(A-A)BZI complex.[83] As shown in Figure 1.7, BZI occupies a position close to the L-serine derived C^{β} but does not actually bind.[83] Again, the X-ray crystal structures confirm the E(A-A) intermediate and the UV/vis absorption characteristics match those of the intermediate produced with the enzymes natural substrates.[83]

1.3.4 Results from Initial NMR Crystallography Study of $E(Q_3)_{indoline}$

An early application of the NMR crystallography approach involved a study by Lai and coworkers on $E(Q_3)_{indoline}$ in 2011.[22] For context, the canonical form of the quinonoid intermediate is shown in Figure 1.10. The characteristic UV/Vis absorption spectrum including a narrow lineshape, high molar extinction coefficient , and an absorption maximum in the high 400 nm to low 500 nm had for decades been attributed to this canonical form having a protonated pyridine ring and Schiff base nitrogen.[9, 32] Solid state NMR measurements of ^{13}C and ^{15}N enriched L-serine and indoline (enrichment sites identified in Figure 1.8), however, noted a discrepancy at the Schiff-base nitrogen. Its chemical shift was closer to that of deprotonated forms unlike the canonical model.[22] Temperature dependence measurements the Schiff base nitrogen chemical shift indicated that proton exchange involving that site was occurring but the identification of exchange partners could not be determined from this information alone.[22]

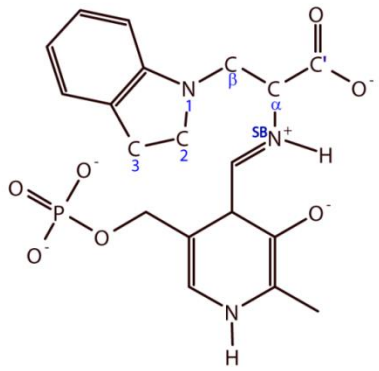


Figure 1.10. Canonical protonated Schiff -base quinonoid. Positions enriched with ^{15}N or ^{13}C on the L-serine derived substrate are marked with Greek letters or SB for Schiff base nitrogen and C' for the carbonyl carbon. Positions enriched on indoline are marked with a blue number.

First principles calculations were performed on models of the enzyme active site in various configurations to predict NMR chemical shifts.[22] The number of possible combinations of protonation states imaginable quickly becomes unmanageable and thus, predicated on experimental UV/Vis data, all models tested included a protonated pyridine ring nitrogen.[22] Establishment of an exchange model consistent with the experimental results predicted tautomeric exchange between the phenolic oxygen, the Schiff base nitrogen, and the carboxylic oxygen nearest the Schiff base nitrogen, see Figure 1.11.[22] Additionally, the protonated carboxylic oxygen, acid-form, was predicted to be the predominant tautomer.[22] NBO calculations of atomic charge indicated that the acid-form built up more negative charge at C^{α} than the canonical protonated Schiff base form indicating that it may have catalytic importance.[22] Efforts had been underway to obtain ^{13}C and ^{15}N enriched PLP but it would be a few years before it was in-hand to test. Probing the ionization state directly using ^{17}O NMR, the topic of Chapter 2, became imperative and was encouraged by a recent demonstration of its feasibility in the solution

state by Zhu and Wu.[1] It would later be determined that the assumption of a protonated pyridine ring nitrogen was not correct and the details of the highly collaborative effort necessary to arrive at this finding are detailed in Chapter 3.

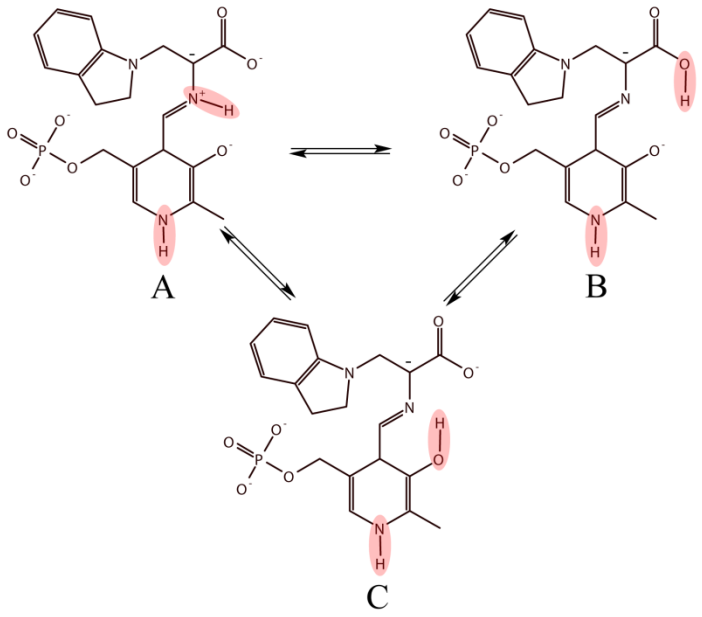


Figure 1.11. Three-site tautomeric equilibrium exchange model from the original $E(Q_3)_{indoline}$ study. (A) Canonical protonated Schiff base quinonoid, (B) acid-form, and (C) phenolic form.

1.4 Brief Overview of NMR Spectroscopy

Nuclear magnetic resonance has emerged as one of the most powerful techniques for probing local chemical structure including local geometry, hybridization, ionization, and protonation states. Additionally, NMR parameters can give details regarding dynamic events such as conformational and chemical exchange, an example of the latter would include determination of proton transfers associated with general acid-base chemistry.

Solid state NMR (SS-NMR) techniques are also capable of determining anisotropic information regarding the local electronic environment. This section provides a brief overview of the NMR phenomenon with the primary purpose of introducing terms and concepts that will be used throughout this dissertation.

NMR active nuclei possess a magnetic dipole moment, **μ**, and therefore an intrinsic magnetic spin angular momentum characterized by the nuclear spin quantum number *I.* The spin-*I* values can assume integer or half-integer values but those with $I = 0$ are not NMR active. The magnitude of the spin angular momentum, a vector quantity, is given by Equation (1).[90]

$$|\boldsymbol{I}| = [\boldsymbol{I} \cdot \boldsymbol{I}]^{1/2} = \hbar[I(I+1)]^{1/2} \quad (1)$$

The magnetic moment is proportional to the magnitude of the spin angular momentum by

$$\mu = \gamma\hbar[I(I+1)]^{1/2} \quad (2)$$

where γ is the isotope specific gyromagnetic ratio most often reported in units of radian per second per Tesla (rad s^{-1} T^{-1}).[91] When an external magnetic field is applied, the degeneracy of magnetic energy in the nuclear ground state is lifted (Zeeman splitting) resulting in $2I + 1$ equally spaced energy levels.[90] By convention in NMR, the static magnetic field is assigned the z-axis and is denoted B_0. The corresponding Hamiltonian for an isolated spin in quantum mechanical operator form is $\widehat{H}_0 = -\gamma B_0 \hat{I}_z$, where $\hat{I}_z$ is the quantum mechanical operator for the z-component of the angular momentum.[92] $\hat{I}_z$ has $2I + 1$ eigenfunctions (and eigenvalues) characterized by the magnetic quantum

numbers, m_I, that assume values from I to $-I$ in integer steps.[92] The eigenvalue equation is given below.

$$\hat{H}_0|I, m_I\rangle = -\gamma B_0 \hat{I}_z|I, m_I\rangle = -m_I \hbar \gamma B_0 |I, m_I\rangle \quad (3)$$

The Zeeman splitting is depicted in Figure 1.10 for spin I = 1/2 (A) and spin I = 5/2 (B) nuclei for the case of a positive γ.[93] The energy differences between the two states for spin-1/2 and between each of the six states for the spin-5/2, Δm = +/-1, due to Zeeman splitting is given by Equation (4).[92] Note that all NMR active nuclei with I > 1/2 are quadrupolar nuclei and are affected by additional terms that will be presented in the next chapter and further discussion will be limited to spin-1/2 nuclei.

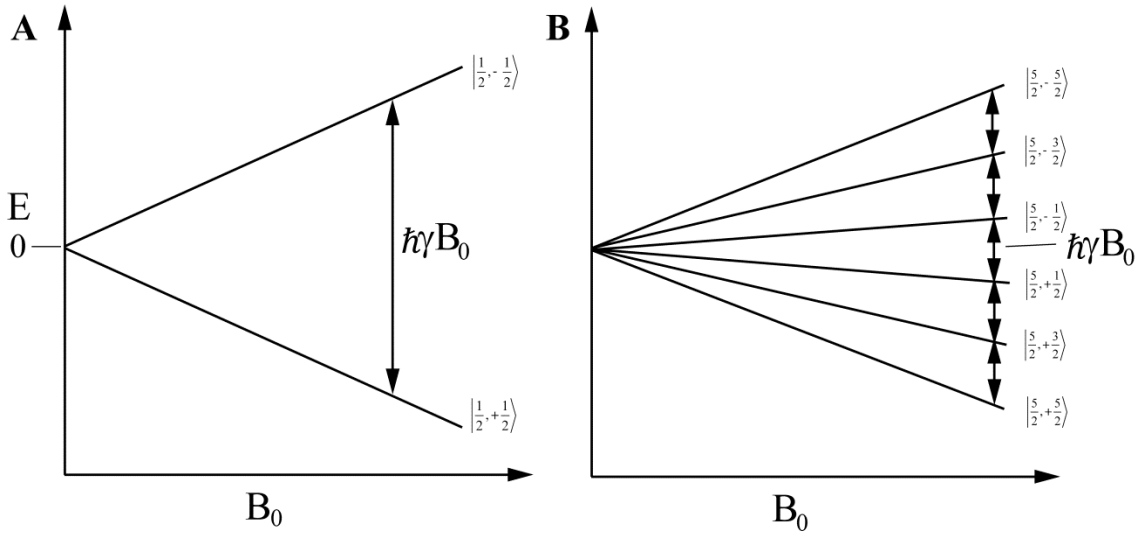


Figure 1.12. Zeeman splitting of the nuclear magnetic energy levels for a spin-1/2 (A) and spin-5/2 (B) nucleus in a static magnetic field for the case of a positive gyromagnetic ratio.

Expressed in terms of photons having frequency ν (Hz) that are resonant with the energy differences we have $\nu_{m\to m'} = h^{-1}\Delta E_{m\to m'} = \gamma B_0/2\pi$ or in angular frequency units (rad/s), $\omega_{m\to m'} = \gamma B_0$.[94] The energy differences separating transitions is quite small, for example, a ^{1}H, γ = 2.675 x 10^8 rad s^{-1} T^{-1} (T = magnetic field strength in Tesla),

$$\Delta E_{m\to m'} = \hbar\gamma B_0 \ (\Delta m = m \to m' = \pm 1) \tag{4}$$

in a B_0 = 21.4 T magnetic field is only 6 x 10^{-25} J which is smaller than the thermal energy product k_bT (T = temperature) where k_b is the Boltzmann constant. Application of a strong, static magnetic field, B_0, causes the magnetic moments of the nuclei to precess about the z-axis of the laboratory frame with a characteristic frequency, $\omega_0 = -\gamma B_0$ (rad/s) or $\nu_0 = -\gamma B_0/2\pi$ (Hz), termed the Larmor frequency.[92] Since there is an energy difference corresponding to the orientation of the magnetic dipole moment in the static field, there is a bias for the low energy alignment resulting in a population difference between m_I = +1/2 and m_I = -1/2, and over time this leads to a net, or bulk magnetization in the sample at thermal equilibrium that is aligned with the static field and will be denoted hereon as M_z^{eq}.[93] For spin-1/2 nuclei, the total number of spins in the low energy level, $N^{+1/2}$, versus those in the higher, $N^{-1/2}$, follow the Boltzmann distribution, Equation (5).

$$\frac{N^-}{N^+} = e^{-\Delta E/k_bT} = e^{-\gamma\hbar B_0/k_bT} \tag{5}$$

Perturbation by a radiofrequency pulse orthogonal to the static field, the x-y plane in the laboratory frame, introduces a new field, B_1. If the B_1 field is resonant with the Larmor frequency the magnetic dipoles will begin to precess about it. A 90° (π/2 rad) pulse equalizes the population difference and generates phase coherence leading to a net

magnetization vector in the transverse plane (x-y), denoted M_{xy}. The components of this transverse magnetization precess about the z-axis following pulse completion. A theoretical bare nucleus would precess in the transverse plane at the Larmor frequency but for nuclei in atoms and molecules the circulation of electrons in the static magnetic field induces local fields which can "shield" the nuclei slightly. The effective field, then, experienced at a given nuclear site is given by Equation (6), where σ is the dimensionless

$$B_{eff} = (1 - \sigma)B_0 \tag{6}$$

shielding constant. The shielding constant is actually a rank two tensor and a dependence upon orientation in the magnetic field is present. In the solution state, however, σ is averaged to a single observable value. As a result of this shielding, the magnetic moments precess at slightly different frequencies, given by Equation (7), depending on their electronic, and hence, chemical environment.

$$v_i(Hz) = \frac{-\gamma B_{eff}}{2\pi} \tag{7}$$

The shielding is relatively small, on the order of parts per million, ppm, but it is detectable. This topic will be expanded upon later.

The precession of magnetic moments in the transverse plane induces a current in a detection coil in the NMR instrument giving rise to the NMR signal. The signal decays over time due to loss of coherence created by differences in local randomly fluctuating magnetic fields at each nucleus for which there are many contributions but ultimately are related to the relative motion of the molecules. The decay constant associated with transverse relaxation is T_2. Additionally, after the pulse the magnetic moments will

reestablish the pre- pulse thermal equilibrium populations with a time constant T_1, termed longitudinal relaxation. Figure 1.13 shows the decay of transverse magnetization, M_{xy}, and the reestablishment of the thermal equilibrium through longitudinal relaxation where M_z^{eq} represents the equilibrium net magnetization aligned with the z-axis.

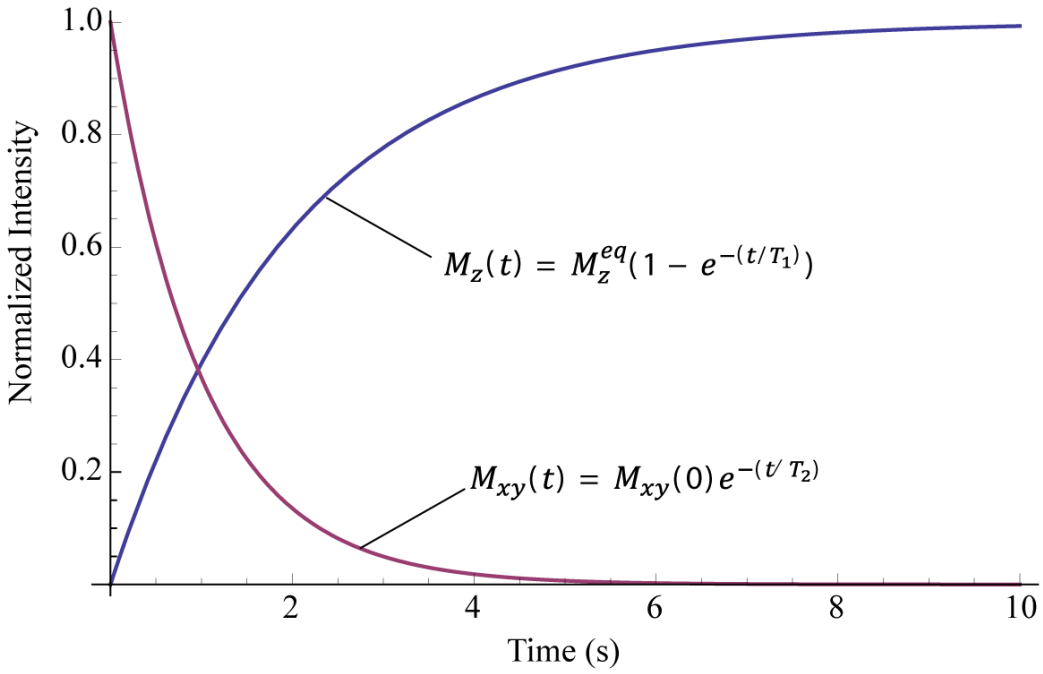


Figure 1.13. Plot of the intensities vs. time of the transverse, M_{xy}, and longitudinal, M_z, magnetization following a 90° pulse.

The resulting NMR signal measured after the pulse is thus a free-induction decay, FID, in the time domain. Fourier transform of the FID gives both real and imaginary spectra as a function of frequency. The real portion is an absorption Lorentzian of the form shown in Equation (8) and depicted graphically in Figure 1.14. Where $R_2 = 1/T_2$ and ν_0 is the offset frequency of the peak (zero for that shown below).

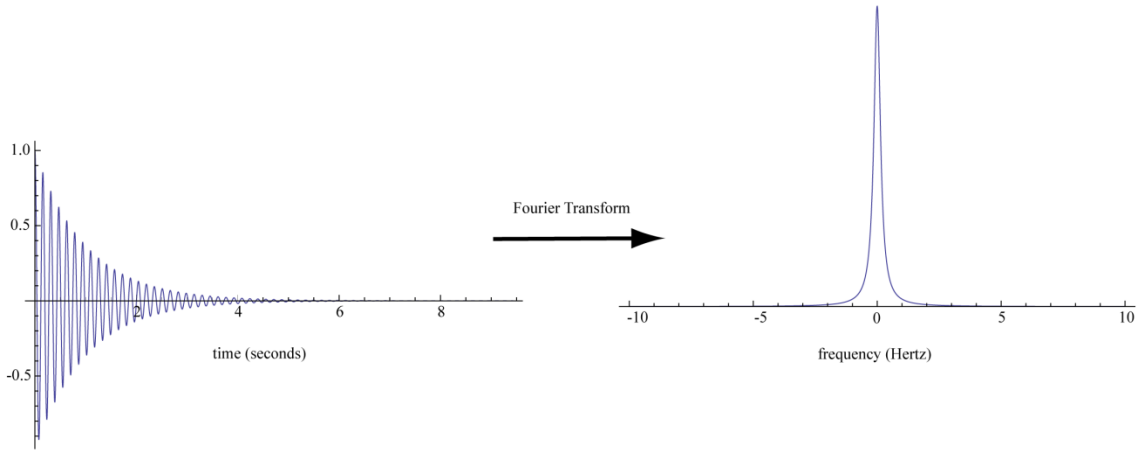


Figure 1.14. Fourier transform of the real component of the FID (time domain) gives the NMR spectrum in the frequency domain in which the signal is an absorptive Lorentzian shown here at zero offset with the carrier frequency.

$$S(\omega) = \frac{R_2}{R_2^2 + 4\pi^2(\nu - \nu_0)^2} \tag{8}$$

There are many different contributions to the transverse relaxation and they will only explicitly be discussed where relevant. The random fluctuations in the local fields, B_{loc}, effecting both transverse and longitudinal relaxation are described by the Fourier transform of an appropriate autocorrelation function that yields the complex spectral density $J(\omega) + iQ(\omega)$. The imaginary component will be discussed in Chapter 2. The real component, $J(\omega)$, gives the density of motions at given frequencies and three are of particular importance, the zero frequency, $J(0)$, the Larmor frequency, $J(\omega_0)$, and twice the Larmor frequency, $J(2\omega_0)$. The reduced forms of the spectral density at these frequencies are given in Equation (9) and are abbreviated J_0, J_1, and J_2, where the subscript indicates the multiple of the Larmor frequency, ω_0.The time constant, τ_c,

$$J_n = \frac{\tau_c}{1 + (n\omega_0\tau_c)^2}, n = 0, 1, or\ 2 \quad (9)$$

is called the rotational correlation time and represents the amount of time it takes an arbitrary point on a molecule to traverse approximately 1 radian. Transverse relaxation, characterized by T_2, for spin-1/2 nuclei is affected by J_0, J_1, and J_2 whereas longitudinal relaxation, characterized by T_1, is dependent on J_1 and J_2 as these alone induce transitions between energy levels. The rotational correlation time has a direct bearing on both the longitudinal and transverse relaxation rates. To indicate this, the longitudinal and transverse relaxation time, T_1 and T_2 respectively, are plotted versus τ_c in Figure 1.13. The relaxation rates are the reciprocal of these characteristic time constants. Motional regimes are defined in NMR as a product of the Larmor frequency (in rad/s) and τ_c, $\omega_0\tau_c$, where the limit of fast isotropic reorientational motion, also known as extreme narrowing, is for $\omega_0\tau_c << 1$, the intermediate motion regime is where the $\omega_0\tau_c \approx 1$, and the limit of slow isotropic motion is defined by $\omega_0\tau_c >> 1$. Naturally, small molecules in non-viscous solvents have very short τ_c values (picosecond) and large molecules, e.g. proteins, have much longer τ_c times (nanosecond).

As indicated in Figure 1.14, the resulting NMR spectrum is a function of frequency. As a means of normalizing measurements made on spectrometers whose field strengths are not identical, the frequency spectrum is converted to a chemical shift spectrum by referencing it to the frequency of a nucleus in an accepted reference

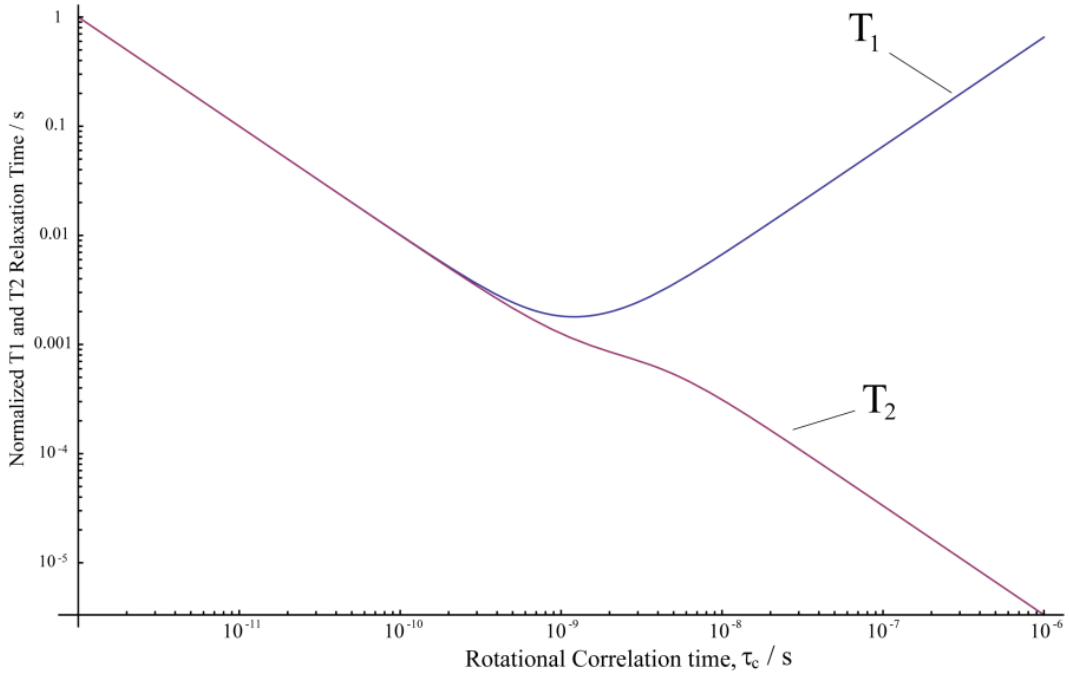


Figure 1.15. Relaxation time constants T_1 and T_2 as a function of the rotational correlation time, τ_c.

compound, for example the methyl protons in tetramethylsilane (TMS) for ^{1}H. As mentioned previously, although the shielding has an orientation dependence, rapid isotropic motion in the solution state averages it to a single observable value. The conversion from frequency to isotropic chemical shift is provided in Equation (11).

$$\delta_{iso}(ppm) = \frac{\nu_i(MHz) - \nu_{ref(}(MHz)}{\nu_{ref(}(MHz)} \times 10^6 \qquad (11)$$

In solid state NMR access to the anisotropic shielding tensor is accessible providing even more information regarding the immediate electronic environment of the nucleus. In practice it is a chemical shift tensor, **δ**, that is experimentally measured. The

shielding and shift are related by: $\boldsymbol{\delta} = \boldsymbol{-\sigma}$. The rank-two chemical shift anisotropy (CSA) tensor is given by Equation (12):

$$\delta_{ij}(ppm) = \begin{bmatrix} \delta_{xx} & \delta_{xy} & \delta_{xz} \\ \delta_{yx} & \delta_{yy} & \delta_{yz} \\ \delta_{zx} & \delta_{zy} & \delta_{zz} \end{bmatrix} \tag{12}$$

Diagonalization of the CSA tensor yields three eigenvalues termed the principle components, δ_{11}, δ_{22}, and δ_{33}, in the CSA principle axis system. These are the most amenable to measurement in the solid state and are ordered as $\delta_{11} \geq \delta_{22} \geq \delta_{33}$ in Mehring notation.[95] The isotropic chemical shift, δ_{iso}, represents 1/3 the Trace of the CSA tensor which is the average of the principal components:

$$\delta_{iso}(ppm) = \frac{1}{3}(\delta_{11} + \delta_{22} + \delta_{33}) \tag{13}$$

Often the isotropic chemical shift value alone is indicative of local chemical environment and sufficient to determine such important details such as hybridization or protonation state. There exist a multitude of approaches to further characterize the local chemical environment available through NMR spectroscopy and, therefore, in the remaining text they will only be introduced where relevant.

1.5 References

1. Zhu, J.F. and G. Wu. Journal of the American Chemical Society, 2011. **133**(4): p. 920-932.

2. Young, R.P., B.G. Caulkins, D. Borchardt, D.N. Bulloch, C.K. Larive, M.F. Dunn, and L.J. Mueller. Angewandte Chemie-International Edition, 2016. **55**(4): p. 1350-1354.

3. Reed, A.E., R.B. Weinstock, and F. Weinhold. Journal of Chemical Physics, 1985. **83**(2): p. 735-746.

4. Hammes, G.G., S.J. Benkovic, and S. Hammes-Schiffer. Biochemistry, 2011. **50**(48): p. 10422-10430.

5. Eliot, A.C. and J.F. Kirsch. Annual Review of Biochemistry, 2004. **73**: p. 383-415.

6. Hayashi, H.. Journal of Biochemistry, 1995. **118**(3): p. 463-473.

7. Jansonius, J.N. Current Opinion in Structural Biology, 1998. **8**(6): p. 759-769.

8. Toney, M.D. Biochimica Et Biophysica Acta, 2011. **1814**(11): p. 1407-18.

9. Metzler, C.M., A.G. Harris, and D.E. Metzler. Biochemistry, 1988. **27**(13): p. 4923-4933.

10. Ngo, H., N. Kimmich, R. Harris, D. Niks, L. Blumenstein, V. Kulik, T.R. Barends, I. Schlichting, and M.F. Dunn. Biochemistry, 2007. **46**(26): p. 7740-7753.

11. Roy, M., S. Keblawi, and M.F. Dunn. Biochemistry, 1988. **27**(18): p. 6698-6704.

12. Ahmed, S.A., H. Kawasaki, R. Bauerle, H. Morita, and E.W. Miles. Biochemical and Biophysical Research Communications, 1988. **151**(2): p. 672-678.

13. Ahmed, S.A., A.M. Kayastha, and E.W. Miles. Faseb Journal, 1991. **5**(5): p. A1149-A1149.

14. Anderson, K.S., A.Y. Kim, J.M. Quillen, E. Sayers, X.J. Yang, and E.W. Miles. Journal of Biological Chemistry, 1995. **270**(50): p. 29936-29944.

15. Hyde, C.C. and E.W. Miles. Bio-Technology, 1990. **8**(1): p. 27-32.

16. Griswold, W.R., A.J. Fisher, and M.D. Toney. Biochemistry, 2011. **50**(26): p. 5918-24.

17. Toney, M.D., J.N. Castro, and T.A. Addington. Journal of the American Chemical Society, 2013. **135**(7): p. 2509-2511.

18. Spies, M.A. and M.D. Toney. Journal of the American Chemical Society, 2007. **129**(35): p. 10678-10685.

19. Zhou, X.Z. and M.D. Toney. Journal of the American Chemical Society, 1998. **120**(50): p. 13282-13283.

20. Ford, G.C., G. Eichele, and J.N. Jansonius. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1980. **77**(5): p. 2559-2563.

21. Hyde, C.C., S.A. Ahmed, E.A. Padlan, E.W. Miles, and D.R. Davies. Journal of Biological Chemistry, 1988. **263**(33): p. 17857-17871.

22. Lai, J.F., D. Niks, Y.C. Wang, T. Domratcheva, T.R.M. Barends, F. Schwarz, R.A. Olsen, D.W. Elliott, M.Q. Fatmi, C.E.A. Chang, I. Schlichting, M.F. Dunn, and L.J. Mueller. Journal of the American Chemical Society, 2011. **133**(1): p. 4-7.

23. Chan-Huot, M., A. Dos, R. Zander, S. Sharif, P.M. Tolstoy, S. Compton, E. Fogle, M.D. Toney, I. Shenderovich, G.S. Denisov, and H.H. Limbach. Journal of the American Chemical Society, 2013. **135**(48): p. 18160-18175.

24. Hartman, J.D., T.J. Neubauer, B.G. Caulkins, L.J. Mueller, and G.J.O. Beran. Journal of Biomolecular Nmr, 2015. **62**(3): p. 327-340.

25. Facelli, J.C. and D.M. Grant. Nature, 1993. **365**(6444): p. 325-327.

26. Harris, R.K., R.E. Wasylishen, and M.J. Duer, eds. 2009, John Wiley & Sons: West Sussex, U.K.

27. Gyorgy, P. Nature, 1934. **133**: p. 498-499.

28. Braunstein, A.E. Nature, 1939. **143**: p. 609-610.

29. Braunstein, A.E. and R.M. Azarkh. Nature, 1939. **144**: p. 669-670.

30. Snell, E.E. Journal of the American Chemical Society, 1945. **67**(2): p. 194-197.

31. Snell, E.E. Journal of Biological Chemistry, 1945. **157**(2): p. 491-505.

32. Metzler, D.E. and P. Christen, eds. *Transaminases*. Biochemistry. 1985, John Wiley & Sons: New York.

33. Olivard, J., D.E. Metzler, and E.E. Snell. Journal of Biological Chemistry, 1952. **199**(2): p. 669-674.

34. Dunn, M.F., D. Niks, H. Ngo, T.R.M. Barends, and I. Schlichting. Trends in Biochemical Sciences, 2008. **33**(6): p. 254-264.

35. Metzler, D.E., J.B. Longenecker, and E.E. Snell. Journal of the American Chemical Society, 1954. **76**(3): p. 639-644.

36. Metzler, D.E. and E.E. Snell. Journal of Biological Chemistry, 1952. **198**(1): p. 353-361.

37. Frey, P.A. and G.H. Reed. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2011. **1814**(11): p. 1548-1557.

38. Palm, D., H.W. Klein, R. Schinzel, M. Buehner, and E.J.M. Helmreich. Biochemistry, 1990. **29**(5): p. 1099-1107.

39. Percudani, R. and A. Peracchi. Embo Reports, 2003. **4**(9): p. 850-854.

40. Metzler, D.E., J.B. Longenecker, and E.E. Snell. Journal of the American Chemical Society, 1953. **75**(11): p. 2786-2787.

41. Toney, M.D.. Archives of Biochemistry and Biophysics, 2005. **433**(1): p. 279-287.

42. Metzler, D.E., M. Ikawa, and E.E. Snell. Journal of the American Chemical Society, 1954. **76**(3): p. 648-652.

43. Ikawa, M. and E.E. Snell. Journal of the American Chemical Society, 1954. **76**(19): p. 4900-4902.

44. Dunathan, H.C.. Proceedings of the National Academy of Sciences of the United States of America, 1966. **55**(4): p. 712-&.

45. Almazov, V.P., Y.V. Morozov, F.A. Savin, and B.S. Sukhareva. International Journal of Quantum Chemistry, 1979. **16**(4): p. 769-775.

46. Toney, M.D. Biochemistry, 2001. **40**(5): p. 1378-1384.

47. Morozov, Y.V., V.P. Almazov, F.A. Savin, and N.P. Bazhulina. Bioorganicheskaya Khimiya, 1982. **8**(8): p. 1119-1132.

48. Barends, T.R.M., T. Domratcheva, V. Kulik, L. Blumenstein, D. Niks, M.F. Dunn, and I. Schlichting. Chembiochem, 2008. **9**(7): p. 1024-1028.

49. Barends, T.R.M., M.F. Dunn, and I. Schlichting. Current Opinion in Chemical Biology, 2008. **12**(5): p. 593-600.

50. Tai, C.H. and P.F. Cook. Accounts of Chemical Research, 2001. **34**(1): p. 49-59.

51. Chan-Huot, M., S. Sharif, P.M. Tolstoy, M.D. Toney, and H.H. Limbach. Biochemistry, 2010. **49**(51): p. 10818-10830.

52. Sharif, S., D. Schagen, M.D. Toney, and H.H. Limbach. Journal of the American Chemical Society, 2007. **129**(14): p. 4440-4455.

53. Sharif, S., G.S. Denisov, M.D. Toney, and H.H. Limbach. Journal of the American Chemical Society, 2006. **128**(10): p. 3375-3387.

54. Schowen, K.B., H.H. Limbach, G.S. Denisov, and R.L. Schowen. Biochimica Et Biophysica Acta-Bioenergetics, 2000. **1458**(1): p. 43-62.

55. Sharif, S., G.S. Denisov, M.D. Toney, and H.H. Limbach. Journal of the American Chemical Society, 2007. **129**(19): p. 6313-6327.

56. Sharif, S., E. Fogle, M.D. Toney, G.S. Denisov, I.G. Shenderovich, G. Buntkowsky, P.M. Tolstoy, M.C. Huot, and H.H. Limbach. Journal of the American Chemical Society, 2007. **129**(31): p. 9558-+.

57. Sharif, S., M.C. Huot, P.M. Tolstoy, M.D. Toney, K.H.M. Jonsson, and H.H. Limbach. Journal of Physical Chemistry B, 2007. **111**(15): p. 3869-3876.

58. Limbach, H.H., M. Chan-Huot, S. Sharif, P.M. Tolstoy, I.G. Shenderovich, G.S. Denisov, and M.D. Toney. Biochimica Et Biophysica Acta, 2011. **1814**(11): p. 1426-37.

59. Casasnovas, R., M. Adrover, J. Ortega-Castro, J. Frau, J. Donoso, and F. Munoz.. Journal of Physical Chemistry B, 2012. **116**(35): p. 10665-10675.

60. Casasnovas, R., J. Ortega-Castro, J. Frau, J. Donoso, and F. Munoz. International Journal of Quantum Chemistry, 2014. **114**(20): p. 1350-1363.

61. Lin, Y.L., J.L. Gao, A. Rubinstein, and D.T. Major. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2011. **1814**(11): p. 1438-1446.

62. Major, D.T. and J.L. Gao. Journal of the American Chemical Society, 2006. **128**(50): p. 16345-16357.

63. Major, D.T., K. Nam, and J.L. Gao. Journal of the American Chemical Society, 2006. **128**(25): p. 8114-8115.

64. Limbach, H.H., M. Chan-Huot, S. Sharif, P.M. Tolstoy, I.G. Shenderovich, G.S. Denisov, and M.D. Toney. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2011. **1814**(11): p. 1426-1437.

65. Birolo, L., E. Sandmeier, P. Christen, and R.A. John. European Journal of Biochemistry, 1995. **232**(3): p. 859-864.

66. Brzovic, P.S., A.M. Kayastha, E.W. Miles, and M.F. Dunn. Biochemistry, 1992. **31**(4): p. 1180-1190.

67. Kayastha, A.M., Y. Sawa, S. Nagata, and E.W. Miles. Journal of Biological Chemistry, 1991. **266**(12): p. 7618-7625.

68. Miles, E.W.. Abstracts of Papers of the American Chemical Society, 1989. **197**: p. 36-Biol.

69. Tai, C.H., W.M. Rabeh, R. Guan, K.D. Schnackerz, and P.F. Cook. Archives of Biochemistry and Biophysics, 2008. **472**(2): p. 115-125.

70. Cash, M.T., E.W. Miles, and R.S. Phillips. Archives of Biochemistry and Biophysics, 2004. **432**(2): p. 233-243.

71. Jhee, K.H., D. Niks, P. McPhie, M.F. Dunn, and E.W. Miles. Biochemistry, 2001. **40**(36): p. 10873-10880.

72. Limbach, H.H., O. Klein, J.M. Lopez del Amo, and J. Elguero. Zeitschrift Fur Physikalische Chemie-International Journal of Research in Physical Chemistry & Chemical Physics, 2004. **218**(1): p. 17-49.

73. Ro, H.S. and E.W. Miles. Journal of Biological Chemistry, 1999. **274**(44): p. 31189-31194.

74. Shaw, J.P., G.A. Petsko, and D. Ringe. Biochemistry, 1997. **36**(6): p. 1329-1342.

75. Ahmed, S.A., C.C. Hyde, G. Thomas, and E.W. Miles. Biochemistry, 1987. **26**(17): p. 5492-5498.

76. Griswold, W.R. and M.D. Toney. Journal of the American Chemical Society, 2011. **133**(37): p. 14823-30.

77. Bach, R.D., C. Canepa, and M.N. Glukhovtsev. Journal of the American Chemical Society, 1999. **121**(28): p. 6542-6555.

78. Ikushiro, H., M.M. Islam, A. Okamoto, J. Hoseki, T. Murakawa, S. Fujii, I. Miyahara, and H. Hayashi. Journal of Biochemistry, 2009. **146**(4): p. 549-562.

79. Phillips, R.S., T.V. Demidkina, and N.G. Faleev. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2003. **1647**(1-2): p. 167-172.

80. Caulkins, B.G., B. Bastin, C. Yang, T.J. Neubauer, R.P. Young, E. Hilario, Y.M.M. Huang, C.E.A. Chang, L. Fan, M.F. Dunn, M.J. Marsella, and L.J. Mueller. Journal of the American Chemical Society, 2014. **136**(37): p. 12824-12827.

81. Caulkins, B.G., C. Yang, E. Hilario, L. Fan, M.F. Dunn, and L.J. Mueller. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2015. **1854**(9): p. 1194-1199.

82. Yanofsky, C. and I.P. Crawford, *Tryptophan Synthase*, in *The Enzymes*, P.D. Boyer, Editor. 1972, Academic Press. p. 1-31.

83. Niks, D., E. Hilario, A. Dierkers, H. Ngo, D. Borchardt, T.J. Neubauer, L. Fan, L.J. Mueller, and M.F. Dunn. Biochemistry, 2013. **52**(37): p. 6396-6411.

84. Woehl, E. and M.F. Dunn. Biochemistry, 1999. **38**(22): p. 7118-7130.

85. Hilario, E., B.G. Caulkins, Y.M.M. Huang, W.L. You, C.E.A. Chang, L.J. Mueller, M.F. Dunn, and L. Fan. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2016. **1864**(3): p. 268-279.

86. Dunn, M.F. Archives of Biochemistry and Biophysics, 2012. **519**(2): p. 154-166.

87. Miles, E.W. Chemical Record, 2001. **1**(2): p. 140-151.

88. Miles, E.W., S. Rhee, and D.R. Davies Journal of Biological Chemistry, 1999. **274**(18): p. 12193-12196.

89. Kulik, V., M. Weyand, R. Seidel, D. Niks, D. Arac, M.F. Dunn, and I. Schlichting. Journal of Molecular Biology, 2002. **324**(4): p. 677-690.

90. Cavanagh, J., W.J. Fairbrother, A.G. Palmer, M. Rance, and N.J. Skelton. 2 ed. 2007, New York: Elsevier.

91. Forsen, S. and B. Lindman, *Ion Binding in Biological Systems Measured by Nuclear Magnetic Resonance Spectroscopy*, in *Methods of Biochemical Analysis*, D. Glick, Editor. 1981, John Wiley & Sons: New York. p. 289-486.

92. Keeler, J., *Understanding NMR Spectroscopy*. 2 ed. 2010, West Sussex, U. K.: John Wiley & Sons.

93. Levitt, M.H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. 2 ed. 2008, West Sussex, U. K.: John Wiley & Sons Ltd.

94. Slichter, C.P., *Principles of Magnetic Resonance*. Solid-State Sciences 1, ed. P. Fulde. Vol. 1. 1980, New York: Springer-Verlag. 397.

95. Mehring, M., *High Resolution NMR Spectroscopy in Solids*. 1st ed. NMR Basic Principles and Progress, ed. P. Diehl, E. Fluck, and R. Kosfield. Vol. 11. 1976, Berlin: Springer-Verlag.

Chapter 2

Solution State ^{17}O Quadrupole Central-Transition NMR Spectroscopy in the Active Site of Tryptophan Synthase

2.1 Introduction

2.1.1 Background

Pyridoxal-5'-phosphate (PLP) plays a critical role as cofactor in a large family of enzymes involved in the metabolism of amino acids and other amine-containing biomolecules. PLP acts as an "electron sink," stabilizing negative charge buildup at multiple steps during the catalytic cycle by means of a highly conjugated π-bond system.[1, 2] A particularly remarkable aspect of PLP is that this single cofactor is capable of participating in a diverse array of chemical transformations, including racemization, transamination, decarboxylation, and β/γ-elimination and substitution.[1, 2] Delineating the factors that fine-tune PLP for a particular reaction remains of great interest.[1-3] Studies of PLP in model compounds have demonstrated the important roles that protonation states and tautomerization play in directing reaction specificity.[4, 5]

NMR is a powerful tool for probing tautomerization and acid-base chemistry at atomic resolution. Tracking the equilibrium involved in tautomerization can be accomplished by interrogating the heteronuclei involved in the proton exchange, yet to date this has largely been limited to ^{15}N chemical shift measurements in the solid state.[3, 6, 7] Oxygen, the other key atomic species often involved in tautomerism, is not yet a standard nuclear probe in biological NMR spectroscopy, despite the potential wealth of chemical insight it can provide.[8-10] Here we report the first application of ^{17}O

quadrupole central transition (QCT) NMR to interrogate kinetically-competent species under conditions of active catalysis, probing two quasi-stable intermediates in the catalytic cycle of the 143 kDa, PLP-dependent enzyme tryptophan synthase (TRPS). This work highlights the promise of ^{17}O QCT NMR to mechanistic enzymology.[11]

^{17}O is the only NMR active isotope of oxygen and possesses a quadrupolar, spin $I = 5/2$ nucleus with a gyromagnetic ratio comparable to ^{15}N and low natural abundance (0.037%).[12] NMR studies of oxygen in biological macromolecules, whether in the solution or solid state, are extraordinarily rare.[13-17] ^{17}O QCT NMR spectroscopy in solution takes advantage of the unique relaxation properties of the ^{17}O central transition; in the limit of slow isotropic motion the linewidth narrows with increasing rotational correlation time, τ_c, an unusual feature compared to spin-1/2 nuclei.[18-21] Applied initially by Lee and Oldfield to carbon monoxide bound to heme-proteins, ^{17}O QCT NMR spectroscopy was recently demonstrated by Zhu and Wu to be applicable even when the quadrupolar coupling constant of the substrate is in the MHz range typical of many organic compounds.[16, 18] A second unusual feature of QCT NMR is that the peak position of the central transition is perturbed by a 2nd order magnetic-field dependent frequency shift; to extract the isotropic chemical shift requires measurements at multiple magnetic fields.[22-24] The field-dependent line shapes also allow for the extraction of quadrupole and chemical shift anisotropy product parameters, P_Q and P_{SA}, that may also be diagnostic of chemical state.[18]

Protonation states and tautomerization in the active site of TRPS have been studied using ^{13}C, ^{15}N, and ^{31}P solid-state NMR (SSNMR) for a number of intermediates

in the catalytic cycle.[6, 7, 25] TRPS catalyzes the final two steps in the biosynthesis of L-tryptophan (L-Trp); a full description of the chemical transformations it catalyzes was presented in Chapter 1. Quasi-stable analogues of the Stage II quinonoid intermediate, $E(Q_3)$, can be formed by supplying the nucleophiles indoline or 2-aminophenol (2AP) in place of the natural substrate indole. These analogues are denoted $E(Q_3)_{indoline}$ and $E(Q_3)_{2AP}$, respectively; the former goes on to form the unnatural amino acid dihydroiso-L-tryptophan (DIT), while the latter does not appear to react further.[26, 27] The α-aminoacrylate intermediate, E(A-A), is relatively stable when no nucleophiles are supplied, and its lifetime can be further enhanced by the addition of the inhibitor benzimidazole (BZI), a non-reactive indole analogue. The structures of these intermediates are shown in Figure 2.1.[26, 27]

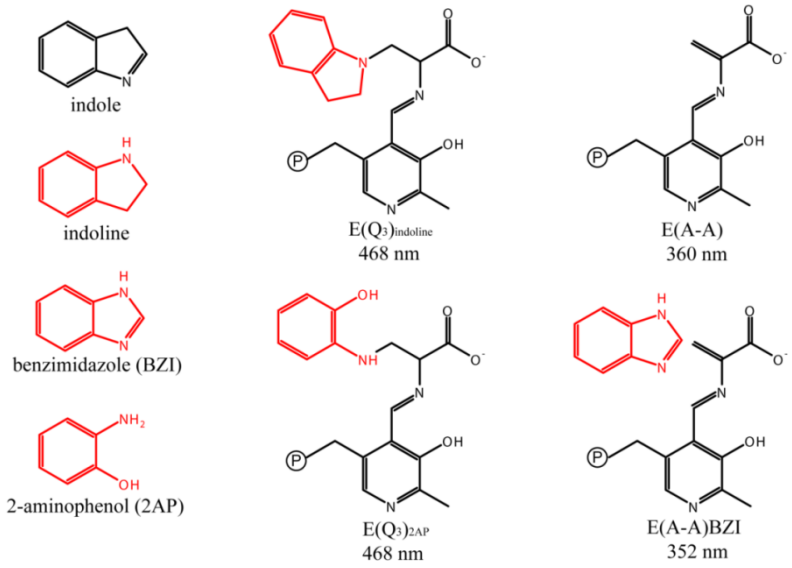


Figure 2.1. Quasi-stable intermediates of the TRPS β-site reaction measured in this study. The $E(Q_3)_{indoline}$ and $E(Q_3)_{2AP}$ are formed by the addition of indoline and 2-aminophenol (2AP), structural analogues of the natural metabolite indole. The E(A-A) forms when no substrate analogues are supplied and the E(A-A)BZI is a more stable form of the E(A-A) and, as shown, the benzimidazole analogue does not covalently bind to the L-serine derived C^{β}.[26, 27]

Initial SSNMR studies of the $E(Q_3)_{indoline}$ intermediate allowed the conclusion to be drawn that an equilibrium involving proton exchange in the pocket adjacent to the Schiff base nitrogen and including the phenolic and nearest carboxylate oxygen was the best description of the chemical state (Figure 2.2) and that an acid-form tautomer was the predominant species at equilibrium.[6] This led to the hypothesis that the acid form of the substrate might play a direct, catalytically-significant role.[6] ^{17}O QCT NMR spectroscopy provides a test of this hypothesis. By supplying the substrate L-serine enriched in ^{17}O at the carboxylic site we are able to explore protonation states in the two quinonoid analogues, $E(Q_3)_{indoline}$ and $E(Q_3)_{2AP}$, and in the aminoacrylate intermediates, E(A-A) and E(A-A)BZI.[11]

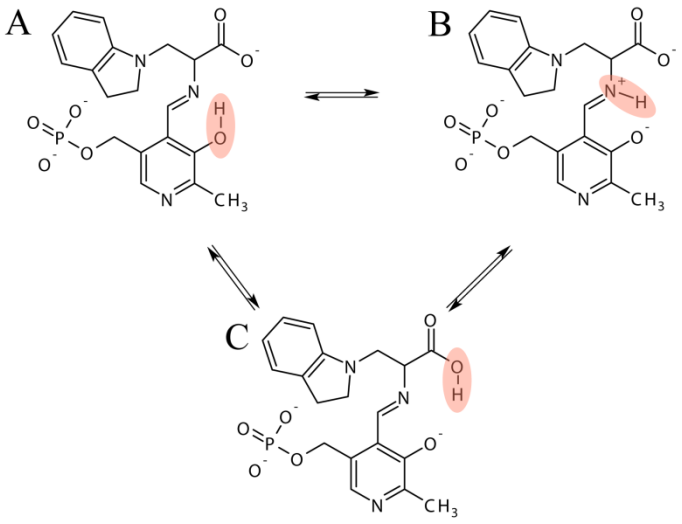


Figure 2.2. Tautomeric equilibrium involving the exchange of a single proton between three ionizable sites of the PLP/substrate complex proposed in the initial E(Q3)indoline investigation. (A) phenolic tautomer, PO, (B) protonated Schiff-base tautomer, SB, and (C).

2.1.2 Theoretical Aspects of Quadrupole Central Transition NMR

In addition to having a magnetic dipole moment, spin $I > 1/2$ nuclei possess an electric quadrupole moment, eQ, attributed to a non-spherical distribution of nuclear charge.[28, 29] The coupling of the quadrupole moment with the electric field gradients, EFG, at the nucleus generated by surrounding electrons and other nuclei often dominates relaxation processes.[28, 29] Ions, which are primarily subjected to interactions with other ions or nearby dipoles and are often coordinated in highly symmetric arrangements with liganded species which can ameliorate quadrupolar interactions.[30, 31] The quadrupolar interactions are similarly attenuated in highly symmetric covalent bonding situations such as that of inorganic $C^{17}O$ employed by Lee and Oldfield in their early work.[30] Quadrupolar nuclei in organic compounds, on the other hand, are typically exposed to intense EFGs due to asymmetry in their covalent bonding environments.[29, 30] Even with relatively small quadrupole moments, quadrupolar coupling can be on the order of the Larmor frequency, MHz.[18, 30] As a result, NMR studies of quadrupolar nuclei bound to biological macromolecules in solution has been dominated by those involving inorganic ionic species such as Na^+, Ca^{2+}, Mg^{2+}, Al^{3+}, Cl^-, or inorganic molecules like carbon monoxide.[15, 16, 30, 32-39]

^{17}O has a spin, $I = 5/2$, and when placed in a static magnetic field the degeneracy of its magnetic energy eigenstates is lifted by the Zeeman interaction resulting in $2I + 1$ equally spaced energy levels labeled by the angular momentum quantum number m_I, where m assumes values from $+I$ to $-I$ in integer steps. An energy level diagram for ^{17}O

is provided in Figure 2.3. The Hamiltonian for the Zeeman term, H_0, is given by Equation (1), where γ is the nuclide specific gyromagnetic ratio, -3.688×10^{7} rad s^{-1} $T^{-1}$for ^{17}O, B_0

$$\widehat{H}_0|I,m_I\rangle = -\gamma\hbar B_0\hat{I}_z|I,m_I\rangle = -m_I\hbar\gamma B_0|I,m_I\rangle \quad (1)$$

is the static magnetic field strength (T, Tesla), and $\hbar$ (J s) is Plank's constant divided by 2π.[12] The energy spacing between the eigenstates resulting from the Zeeman interaction is given by Equation(2).[30]

$$\Delta E_{m\to m'} = \hbar\gamma B_0\ (\Delta m = m \to m' = \pm1) = \nu_0(s^{-1})\ or\ \omega_0(rad\ s^{-1}) = 2\pi\nu_0 \quad (2)$$

The single quantum coherences (transitions), $\Delta m_I = \pm1$ are grouped as follows, the central transition, CT, corresponds to the Δm_I(-1/2↔+1/2), the first satellite transitions, ST1 correspond to Δm_I(±3/2↔±1/2), and the second satellite transitions, ST2 correspond to Δm_I(±5/2↔±3/2).[20, 24, 40]

2.1.3 First-Order Quadrupolar Interactions

As mentioned previously, quadrupolar nuclei possess an electric quadrupole moment, eQ (C m^2), which couples with the EFGs at the nucleus.[28] The magnitude of the quadrupolar coupling is described by the product of eQ and a 2nd rank EFG tensor.[41] The quadrupole moment for ^{17}O is $-2.558\cdot10^{-30}$ em^2 and the EFG tensor is given by Equation (3) in Cartesian coordinates in an arbitrary molecular axis frame.[12, 42]

$$\boldsymbol{V} = \begin{bmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{yx} & V_{yy} & V_{yz} \\ V_{zx} & V_{zy} & V_{zz} \end{bmatrix},\ V_{\boldsymbol{ii}} = eq_{ii} = e\frac{\partial^2 V(\vec{r})}{\partial i\partial j}\bigg|_{r=0},\ e = unit\ charge \quad (3)$$

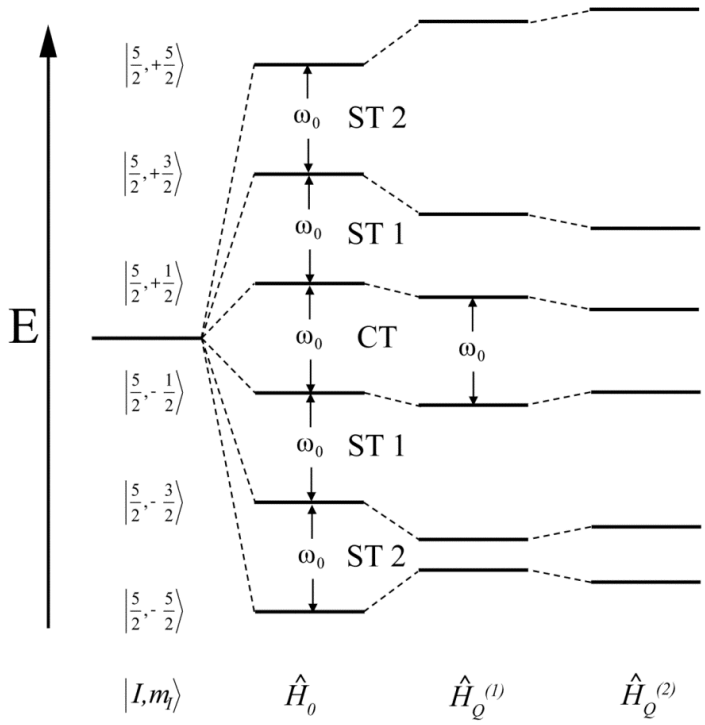


Figure 2.3. Energy level diagram for the spin, I = 5/2, ^{17}O with perturbations from Zeeman splitting, H_0, and the first and second order quadrupolar interactions, $H_Q^{(1)}$ and $H_Q^{(2)}$. The central transition, CT, is unaffected by the $H_Q^{(1)}$, whereas the single quantum coherences separating the satellite transitions, ST 1 and ST 2 change relative to the CT.

The EFG tensor is symmetric, $V_{ij} = V_{ji}$ $(i \neq j)$, and traceless, $V_{xx} + V_{yy} + V_{zz} = 0$. Diagonalization of the EFG tensor yields three eigenvalues, V_{11}, V_{22}, and V_{33}, termed the principal components in the EFG's principal axis system (PAS).[42] By convention, the principal components are ordered $|V_{33}| \geq |V_{22}| \geq |V_{11}|$ and these are also traceless meaning there are only two independent principal components.[30] The EFG tensor is described by its largest principal component, V_{33}, and its asymmetry, η_Q, given by Equation (4).[30]

$$\eta_Q = \frac{V_{22} - V_{11}}{V_{33}}, 0 \leq \eta_Q, \leq 1 \tag{4}$$

The coupling interaction between the EFG and *eQ* is given by the quadrupolar coupling constant, C_Q *(MHz)*, which is the product of the *eQ* and V_{33}, the largest principal component of the EFG and is given as Equation (5).

$$C_Q(MHz) = \frac{e^2 Q V_{33}}{h} \tag{5}$$

The C_Q and η_Q are coupled in equations describing relaxation and for convenience a quadrupole product parameter is defined by Equation (6).[18]

$$P_Q(MHz) = C_Q \sqrt{1 + \frac{\eta_Q^2}{3}} \tag{6}$$

To first order, the quadrupolar interaction perturbs the energy eigenstates further by Equation (7) which gives the eigenvalues, E^Q of the first-order quadrupolar Hamiltonian, $H_Q^{(1)}$.[30]

$$E_{m_I}^{Q(1)} = \frac{P_Q}{8I(2I-1)}(3m_I^2 - I(I+1)) \tag{7}$$

By Equation (7) we see that this shifts m_I levels having the same $|m_I|$ values equally, see Figure 2.2. An important feature is that the energy spacing for the central transition, CT, $|m_I| = 1/2$, is unaffected by the $H_Q^{(1)}$, whereas the energy differences between the satellite transitions ST 1 and ST 2 are no longer degenerate. In the limit of fast isotropic motion, or extreme narrowing, where $\omega_0\tau_c << 1$, the first order quadrupolar interaction is averaged to zero (traceless EFG tensor) and higher order terms are negligible.[43] The degeneracy of the spectral density in extreme narrowing, $J_0 = J_1 = J_2$, results in exponential transverse, $R^{(2)}$, and longitudinal, $R^{(1)}$, relaxation rates given by Equation (8),[18, 20, 21]

$$R^{(2)} = R^{(1)} = \frac{12\pi^2}{125} P_Q^2 \tau_c \qquad (8)$$

where $R^{(2)} = 1/T_2$ and $R^{(1)} = 1/T_1$, T_1 and T_2 being the characteristic time constants of longitudinal and transverse relaxation processes respectively. As Equation (8) indicates, broadening of resonance linewidths, defined by the full width at half-height, $\Delta\nu_{1/2}(Hz) = 1/\pi T_2$, increases with increasing rotational correlation time, τ_c. This is the same type of behavior exhibited by the more popular spin-1/2 nuclei and apparently led many to believe that quadrupolar resonances would be too broad to be detected with increasing molecular size (hence increasing τ_c) in the solution state.[18, 44-46] The theory of quadrupolar relaxation that had been formulated early on in the practice of NMR had predicted non-exponential relaxation behavior outside of extreme narrowing.[20, 40, 47, 48] In the limit of slow isotropic motion, $\omega_0\tau_c >> 1$, relaxation of single quantum transitions would consist of a sum of $I + ½$ decaying exponentials, one of which would narrow with increasing τ_c. The narrowing component corresponds to the central transition in standard pulse NMR.

Beginning in the intermediate motion regime, where $\omega_0\tau_c \approx 1$, the first order quadrupolar interaction is no longer motionally averaged to zero.[20, 43, 49] The real components of the reduced spectral densities are no longer degenerate, that is $J_0 \neq J_1 \neq J_2$, but rather $J_0 \geq J_1 \geq J_2$ and the imaginary component of the spectral density, Q_n, becomes non-negligible leading to a 2nd order quadrupolar term that, in the solution state, results in a dynamic frequency shift of the energy levels and will be described in more detail after discussing the relaxation properties in the intermediate and slow isotropic motional regimes.[22, 50-52] Equations (9) and (10) give the values of the real and imaginary

components of the reduced spectral densities.[18] Figure 2.4 gives the form of the relative magnitudes of each component as a function of $\omega_0\tau_c$.

$$J(n\omega_0) = J_n = \frac{\tau_c}{1 + (n\omega_0\tau_c)^2}, n = 0, 1, or\ 2 \tag{9}$$

$$Q(n\omega_0) = Q_n = n\omega_0\tau_c J_n, n = 0, 1, or\ 2 \tag{10}$$

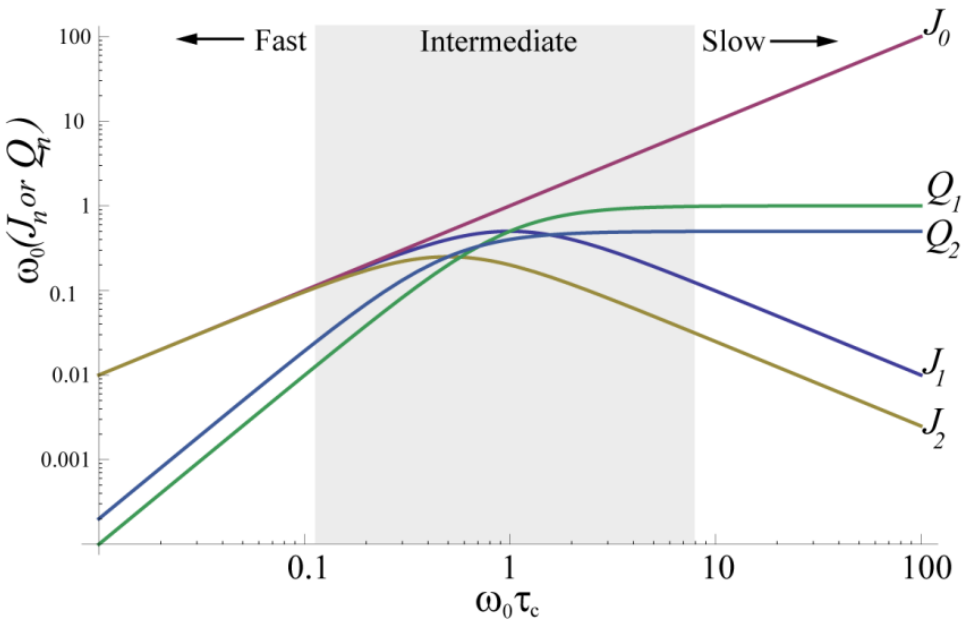


Figure 2.4. Real, J_n, and imaginary, Q_n, components of the reduced spectral density functions plotted as a function of $\omega_0\tau_c$. The limits of fast, intermediate, and slow isotropic reorientational regimes are indicated.

Although no longer purely exponential, longitudinal and transverse relaxation rates in the intermediate motion regime have been approximated as such and are given in Equations (11) and (12) respectively.[18, 49]

$$R^{(1)} = \frac{6\pi^2}{625} P_Q^2 (2J_1 + 8J_2) \tag{11}$$

$$R^{(2)} = \frac{6\pi^2}{625} P_Q^2 (3J_0 + 5J_1 + 2J_2) \tag{12}$$

The transverse relaxation rate, $R^{(2)}$, Equation (12), becomes a poorer approximation at $\omega_0\tau_c > 1.5$, however, the longitudinal relaxation rate, $R^{(1)}$, is a sufficient approximation even beyond the intermediate motion regime.[18, 49]

In the limit of slow isotropic motion, $\omega_0\tau_c >> 1$, $J_0 >> J_1 \approx 4J_2$. Transverse relaxation can no longer be approximated as nearly exponential.[20, 21] As mentioned previously, transverse relaxation of the single quantum coherences is the sum of $I + 1/2$ decaying exponentials, 3 for a spin-5/2 nuclei, and is shown in Equation (13) as a function of time, t, where I, II, and III correspond to the CT, ST 1, and ST 2 respectively, Ω_n is the dynamic frequency shift to be presented shortly, and $A_n^{(2)}$ gives the amplitude.[18]

$$M_{x,y}(t) = \sum_{n=I,II,III} A_n^{(2)} e^{-(i\Omega_n + R_n^{(2)})t} \tag{13}$$

The transverse relaxation rates, $R_n^{(2)}$, of each component are provided in Equations (14) through (16).[18, 21]

$$R_I^{(2)} = \frac{3\pi^2}{1000} P_Q^2 (16J_1 + 56J_2) \tag{14}$$

$$R_{II}^{(2)} = \frac{3\pi^2}{1000} P_Q^2 (6J_0 + 36J_1 + 42J_2) \tag{15}$$

$$R_{III}^{(2)} = \frac{3\pi^2}{1000} P_Q^2 (24J_0 + 48J_1 + 28J_2) \tag{16}$$

The $R_{II}^{(2)}$ and $R_{III}^{(2)}$ rates, corresponding to the ST 1 and ST 2 single quantum coherences, depend on J_0 and would indeed by broadened beyond detectability with a large C_Q.[21]

The $R_I^{(2)}$ for the CT, however, no longer depends on the J_0 and its linewidth, $\Delta\nu_{1/2}$, narrows with τ_c^{-1} as shown in Equation (17).[18, 20, 21, 40, 47-49]

$$\Delta\nu_{\frac{1}{2}}(Hz) = \frac{R_I^{(2)}}{\pi} = \frac{3\pi}{1000} P_Q^2 (16J_1 + 56J_2) \approx 7.2 \times 10^{-3} \left(\frac{P_Q}{\nu_0}\right)^2 \frac{1}{\tau_c} \tag{17}$$

Although the quadrupolar interaction dominates relaxation, as magnetic field strengths increase, relaxation via chemical shift anisotropy, if present, can counter this narrowing.[18] The contribution to relaxation by the shift anisotropy, SA, is given by Equation (18), where P_{SA} is the shift anisotropy product parameter given by Equation (19)

$$R_{SA}^{(2)} = \frac{2\pi^2}{45} (P_{SA}\nu_0)^2 (8J_0 + 6J_1) \tag{18}$$

and its components, $\Delta\delta$, the shift anisotropy, and η_{SA} its asymmetry parameter defined in Equations (20) and (21).

$$P_{SA}(ppm) = \Delta\delta \sqrt{1 + \frac{\eta_{SA}^2}{3}} \tag{19}$$

$$\Delta\delta = \delta_{11} - (\delta_{22} + \delta_{33})/2 \tag{20}$$

$$\eta_{SA} = (\delta_{22} - \delta_{33})/(\delta_{11} - \delta_{iso}) \tag{21}$$

With both quadrupolar coupling and shift anisotropy relaxation contributions included the observed $\Delta\nu_{1/2}(Hz)$ becomes (Equation (22))

$$\Delta\nu_{\frac{1}{2}}(Hz) = 7.2 \times 10^{-3} \left(\frac{P_Q}{\nu_0}\right)^2 \frac{1}{\tau_c} + 1.1(P_{SA}\nu_0)^2 \tau_c. \tag{22}$$

2.1.4 2nd Order Quadrupolar Interaction – Dynamic Frequency Shifts

Although rarely discussed in the NMR of spin-1/2 species, the spectral density function can be written in terms of real and imaginary components.[40, 51] Both components are real-valued. The real component, the Fourier transform of which the reduced spectral density functions, J_n, are derived, is important in transverse and longitudinal relaxation processes. The imaginary component, Q_n, an odd-valued function, has been attributed to induce hyperfine dynamic frequency shifts for certain coherence elements.[51] As seen in Figure 2.4, the magnitudes of Q_n are much smaller than J_n until τ_c is similar to the reciprocal of the Larmor frequency, ω_0^{-1}, where their magnitudes become similar.[51] The dynamic frequency shifts are often much smaller than the broadest resonance, but when one component, in this case the CT, is unaffected by adiabatic contributions to its linewidth and narrows with increasing τ_c, the field-dependent dynamic frequency shift can be significant.[21, 51] Further perturbation of the energy levels by this 2nd order quadrupolar interaction, $H_Q^{(2)}$, is depicted in Figure 2.3 presented earlier. For the CT, the observed chemical shift measured experimentally, δ_{obs}, always appears at a lower value than the isotropic chemical shift, δ_{iso}. In Equation (13), this dynamic frequency shift was denoted Ω_n, and Equations (23) through (25) define its effects on the single quantum coherences for the CT, ST1, and ST2 which are plotted as a function of $\omega_0\tau_c$ in Figure 2.5. A conversion of the dynamic frequency shift to units of parts per million (ppm), $\Delta\delta_d$, is given by Equation (26).[18] Taking note of the form of the imaginary components, Q_n, in Figure 2.5, we see that by $\omega_0\tau_c \gtrapprox 10$ the 2nd order dynamic

$$\Omega_I = \frac{3\pi^2}{1000} P_Q^2(-16Q_1 + 16Q_2) \tag{23}$$

$$\Omega_{II} = \frac{3\pi^2}{1000} P_Q^2(-4Q_1 + 10Q_2) \tag{24}$$

$$\Omega_{III} = \frac{3\pi^2}{1000} P_Q^2(32Q_1 - 8Q_2) \tag{25}$$

$$\Delta\delta_d(ppm) = \frac{\Omega_n}{2\pi\nu_0} \times 10^6 \tag{26}$$

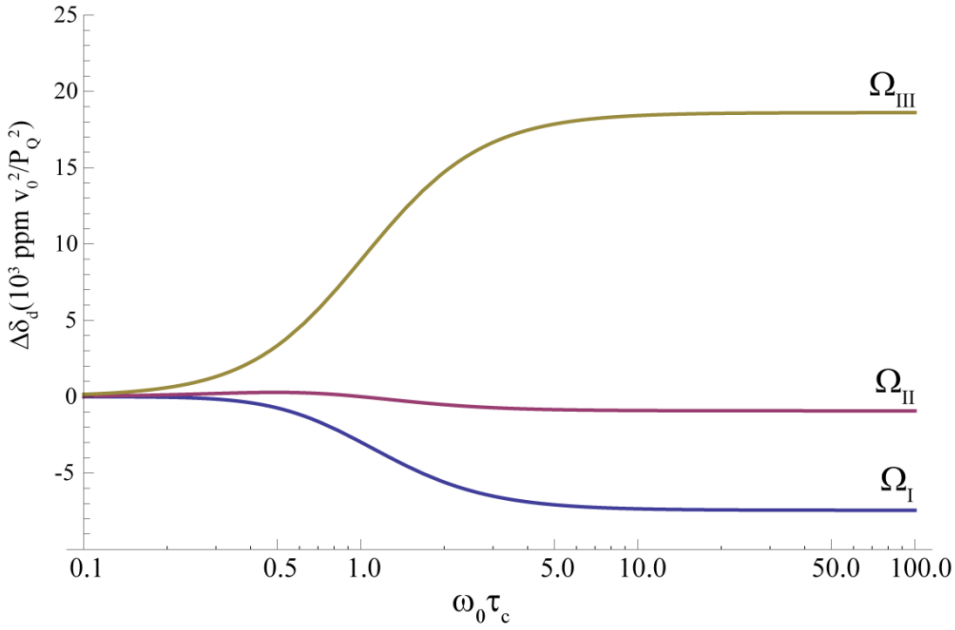


Figure 2.5. Relative effects of the 2nd order quadrupolar interaction for spin I = 5/2 nuclei; dynamic frequency shifts, $\Delta\delta_d$, experienced by the CT, ST 1 and ST 2, labeled Ω_I, Ω_{II}, and Ω_{III} respectively.

frequency shifts essentially reach a limiting value and are constant afterwards.[18] In this regime $\Delta\delta_d$ is linear with the ν_0^{-2}, as shown in Equation (27), and measurement of the δ_{obs}

at multiple field-strengths provides the ability to calculate both the δ_{iso} and the P_Q.[18] It is worth mentioning too that beyond $\omega_0\tau_c \gtrapprox 10$, the constant region of Ω_n, it may be more

$$\Delta\delta_d(ppm) = \delta_{obs} - \delta_{iso} = -6000\left(\frac{P_Q}{\nu_0}\right)^2 \quad (27)$$

appropriate to refer to these as 2nd-order quadrupolar shifts as opposed to dynamic frequency shifts.[24] Since these frequency shifts only apply outside of extreme narrowing, the observed shift from a substrate having a rapid τ_c while free in solution will not change with magnetic field strength, $\delta_{obs} = \delta_{iso}$, whereas the signals corresponding to the bound substrate do.[18] This has at least two distinct advantages. First, discrimination between bound and free resonances is had by multiple field measurements. Second, in the event that the bound and free substrates have similar δ_{iso} shifts the field-dependence of the δ_{obs} for the bound state will not be coincident and, while overlap at some field strength may be inevitable, the δ_{obs} at other field strengths will likely be well resolved from the free signal.[11]

2.1.5 Nutation Rates and Line Intensity of the Central Transition

In the limit of fast isotropic reorientational motion 1st-order quadrupolar interactions are effectively averaged to zero and the energy differences separating the six eigenstates are degenerate and equal to the energy of the Zeeman term, H_0. Outside of this limit only the CT is unperturbed by 1st-order quadrupolar interactions. As a result of non-degenerate energy level spacing, the effects of which increase with increasing C_Q, conventional radiofrequency (rf) pulse powers employed in NMR experiments will be ineffective at exciting the satellite transitions when $\omega_Q >> \omega_{rf}$.[31, 53, 54] The same pulse

that gave effective broadband excitation of all single quantum coherences in the limit of fast motion becomes selective for the CT in the intermediate and slow motion regimes.[21, 31, 53] As a result the CT nutates at a rate three times faster than in the fast motion limit.[21, 31, 53] This means that 90° pulse length at a given power level calibrated on the natural abundance ^{17}O signal from water (in the fast-motion limit) would have to be reduced by three, again at the same power level, to effect a 90° pulse on the CT of resonances outside of extreme narrowing.[53, 54] In this work we take advantage of these differing nutation frequencies depending on $\omega_0\tau_c$ to attenuate intense signals from free substrate and water where overlap obscured bound signals.[11] While the nutation rate for the CT under selective excitation is enhanced by a factor of three, its intensity is reduced by a factor of three for pulse NMR techniques, as opposed to continuous wave.[21, 40, 53] The intensity of the CT is 9/35 of the total intensity contribution from all transitions when $\omega_Q << \omega_{rf}, \omega_0$ hence it is reduced to 3/35 under selective excitation conditions where $\omega_Q >> \omega_{rf}$.[21, 53, 55]

2.1.6 Chemical Shift Range of ^{17}O

The isotropic chemical shift, δ_{iso}, is a supremely sensitive indicator of local electronic environment providing details related to bond order, hybridization, ionization, local geometry, hydrogen bonding, dynamics, and more.[12] The chemical shift range of ^{17}O spans some 2000 ppm, however, in this study the subset of this overall range pertaining to oxygen in the carboxylic site of amino acids is more appropriate.

Solid-state NMR is capable of distinguishing between the different types of oxygen species in carboxylic acids, namely the hydroxyl, COH, and carbonyl, C=O, of protonated acid forms, and the ionized carboxylate, COO^-, forms based on their distinct

δ_{iso} ranges. Wong and Poli recently compiled the results from all reported solid state NMR studies of the amino acids and a summary of these are presented as follows.[10] In acid forms of the carboxylic moiety, the protonated hydroxyl oxygen, COH, δ_{iso} shifts ranged between 167 and 187 ppm with an average of 179 ppm (n = 22). The carbonyl, C=O, δ_{iso} shifts ranged between 303 and 355 ppm with an average of 332 ppm (n=23). The δ_{iso} of carboxylate forms, COO^-, ranged between 250 and 314 ppm with an average of 277 ppm (n = 42). The upfield shift of the COH species reflects an increase in shielding as a result of protonation and, concomitantly, the carbonyl species is deshielded as its bond order increases. The ionized carboxylate forms display average δ_{iso} shifts between those of both acid-form oxygen types. Studies in amino and other carboxylic acids in which both X-ray and neutron diffraction data is available indicate the effects of hydrogen bonding on the oxygen chemical shifts.[56-58] Unsurprisingly, as the distance between an acceptor oxygen and a donor hydrogen becomes smaller, the oxygen chemical shift moves upfield. Simultaneously, as the bond length increases between covalently linked oxygen and hydrogen increases the donor oxygen chemical shift moves downfield.[57] These effects are additive as the number of hydrogen bonds to the oxygen increase and can have quite significant effects on the resulting δ_{iso}.[58]

In aqueous solution, fast exchange between the carboxylic oxygen atoms and water leads to the observance of a single resonance from low to high pH. Gerothanassis et al. measured the chemical shifts of all twenty of the canonical amino acids across a wide range of pH values.[59] At low pH, pH = 0.5, the single fast-exchange δ_{iso} values ranged between 250 and 257 ppm. If one considers that at this pH at least one of the two oxygen

atoms is protonated essentially all the time this would be the simple exchange average of a carbonyl oxygen and hydroxyl oxygen chemical shift. Indeed a 50/50 average of the COH and C=O chemical shifts yields a value of 255 ppm. At mid pH, pH = 6.5, the amino acid is in its zwitterionic state with a negatively ionized carboxylate group. The oxygen δ_{iso} values in this pH range were measured to be between 266 and 274 ppm.[59] In going to high pH conditions, pH = 12.5 ppm, a slight and unexpected shift to upfield values occurred in most amino acids although the and the range observed was 264 to 271 ppm reflecting how small the changes were overall.[59] This was surprising because at high pH the α-amino nitrogen is deprotonated and unionized. The decrease in electronegativity of the nitrogen upon deprotonation should shift the carboxylate oxygen resonances downfield through a decrease in the inductive effect.[59] Some postulates regarding this behavior include the electric field effect of the ammonium ion having a greater role in influencing the shielding at the carboxylic sites than inductive effects.[59] Also, substituent effects by the group attached to the C^{α} have also been considered as oxygen chemical shifts display sensitivities to groups up to four bonds away.[59] The restrictive enzyme active site environment is believed to be more like the solid state and distinct oxygen resonances for both the oxygen atoms in the carboxylic group should be observable unlike bulk solvent.[18]

2.1.7 Quadrupole and Shift Anisotropy Product Parameters

^{17}O QCT NMR has the additional feature that the quadrupole product parameter, P_Q, can be determined from a fit of the δ_{obs} at different fields to Equation (27). The isotropic reorientational correlation time, τ_c, and shift anisotropy product parameter, P_{SA},

can be extracted from lineshape analysis of $\Delta\nu_{1/2}$ at multiple fields, Equation (22). Again, looking to the results from solid-state ^{17}O NMR measurements of amino acids as a guide, the quadrupole coupling constant, C_Q, and asymmetry (or biaxiality) parameter, η_Q, are often reported with the δ_{iso} allowing for the computation of the associated P_Q for comparison with the results obtained from solution state measurements.[10] Using the data compiled by Wong and Poli, the ionized carboxylate oxygen, COO^-, had P_Q values ranging between 6.4 MHz and 8.2 MHz, with an average of 7.6 MHz (n = 40), and for acid-forms the carbonyl oxygen, C=O, P_Q values ranged from 7.9 MHz to 8.7 MHz with an average of 8.4 MHz (n = 23), and hydroxyl oxygen, COH, P_Q values ranged between 7.0 MHz and 7.8 MHz with an average value of 7.5 MHz (n = 22).[10] Naturally, the isotropic chemical shift will often be sufficient to distinguish between the three different forms, however, one can imagine a situation wherein the difference between the average values for carboxylate and acid-form carbonyl P_Q may be helpful in the event of a measured isotropic shift in the low 300 ppm range. It should also be mentioned that the C_Q values, and hence the P_Q, for the hydroxyl oxygen of acid-forms is usually negative, however the sign of the C_Q cannot be determined from NMR experiments.[56, 57] Determining the sign of C_Q requires nuclear quadrupole resonance experiments.[29]

Reports of the principal components of the chemical shift anisotropy tensor for amino acids measured in the solid state are much rarer than the δ_{iso}, C_Q, and η_Q. Data for twenty-five of the carboxylate species oxygen atoms were available and from the principal components provided, the P_{SA} values were calculated from Equations (19) through (21). The values calculated ranged between 262 ppm and 429 ppm with an

average of 342 ppm (n = 25).[10] Data for only four acid-forms were available. The hydroxyl, COH, species P_{SA} ranged between 203 ppm and 238 ppm with an average of 220 ppm (n = 4) and carbonyl type oxygens, C=O, ranged between 474 ppm and 600 ppm with an average of 522 ppm (n = 4).[10] From these limited data it appears that the carboxylate P_{SA} values are intermediate between the hydroxyl types and carbonyl types of the acid-forms much like their isotropic chemical shifts.

2.2 Experimental Section

2.2.1 ^{17}O Enrichment of L-serine

Acid-catalyzed exchange was employed to enrich L-serine at the carboxylic oxygen site(s). 87.6 mg of solid L-serine (99%; Alfa-Aesar) was added to 350 µL of $H_2{}^{17}O$ (90 atom %, Cambridge Isotope Laboratories, Inc.) in a 3 mL conical vial. 35 µL of concentrated ACS Grade HCl(aq) (37%, EMD Millipore) was added to the resulting solution to lower the pH to approximately 2. The vial was capped with a self-sealing Teflon cap, evacuated, and placed under an atmosphere of Ar gas. The vial was placed in an oil bath maintained at a temperature of 55 °C for approximately 8 days after which the solution was neutralized with NaOD (99 atom %, conc. 40% in D_2O, Isotec Inc.), attached to a high vacuum manifold and dried. The duration of the enrichment procedure followed from earlier trials using $H_2{}^{18}O$ (100 atom %, Cambridge Isotope Laboratories, Inc.) which were tracked by electrospray ionization mass spectrometry, ESI-MS, performed by D. N. Bulloch. The fractional abundance of the non-enriched, single-site, and double site enrichment at the carboxylic oxygens in the ^{18}O trials are shown in Figure 2.6 with samples collected at 24, 48, 72, 96, and 120 hours. The distillate from the first

round of drying was collected in a cold trap for later use. The resulting solid was re-dissolved in Millipore-filtered water and dried again using high vacuum distillation to

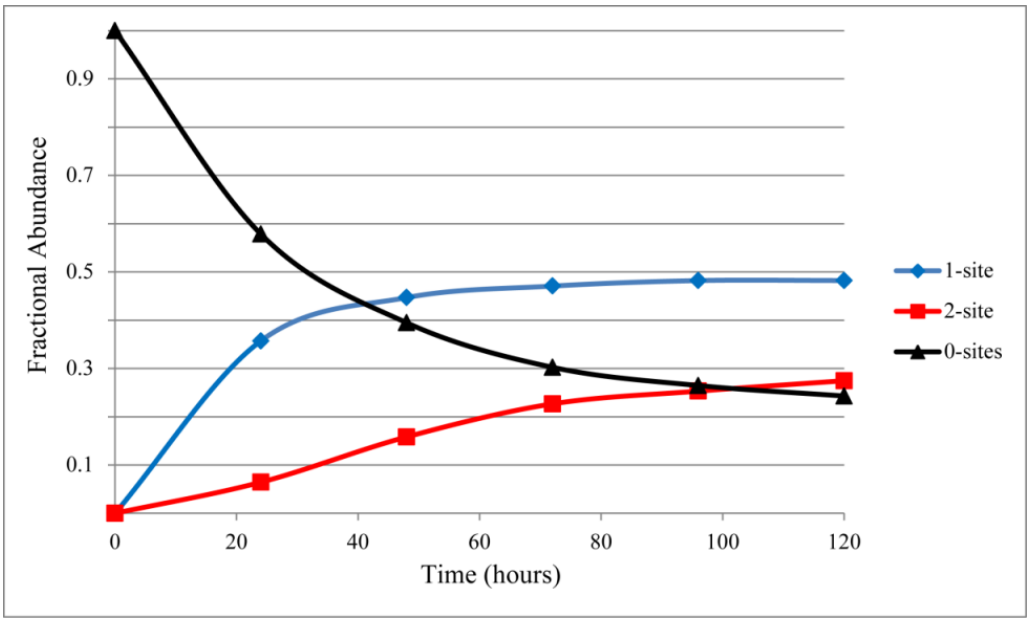


Figure 2.6. Fractional abundance of carboxylic acid enrichment of L-serine trials with $H_2{}^{18}O$ (100 atom %), pH 2, 55°C , monitored by ESI-MS.

remove residual $H_2{}^{17}O$. Incorporation of ^{17}O at the carboxylic oxygen site(s) was confirmed using electrospray ionization mass spectrometry (ESI-MS, performed by D. N. Bulloch) (Micromass Q-TOF Micro) in positive ion mode which indicated enrichment of more than 80%, see Figure 2.7 for the ESI-MS spectrum of the resulting [^{17}O]-L-Ser. Additionally, tandem MS was used to determine the sites of enrichment by D. N. Bulloch. No enrichment of the beta sidechain hydroxyl group was detected indicating that enrichment only occurred at the carboxylic oxygen sites. ^1H NMR spectra of unenriched and enriched ^{17}O L-serine were measured and the latter showed no signs of degradation

or contamination. An ^{17}O-NMR spectrum of the ^{17}O enriched L-serine, 10 mM, is shown below in Figure 2.8. The measurement was performed at 5 °C in 50 mM triethanolamine

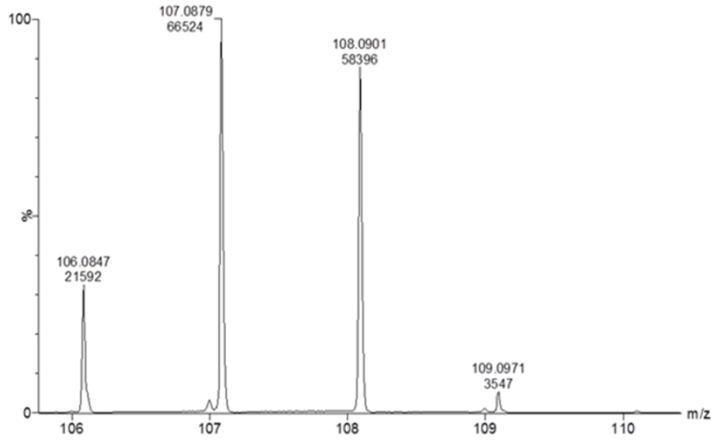


Figure 2.7. ESI-MS mass spectrum of [^{17}O]L-serine following enrichment with $H_2{}^{17}O$ (90 atom %) stirred at 55°C and pH 2 for 8 days. The peaks at 106 m/z, 107 m/z, and 108 m/z correspond to unenriched, single-site, and two-site enrichment respectively.

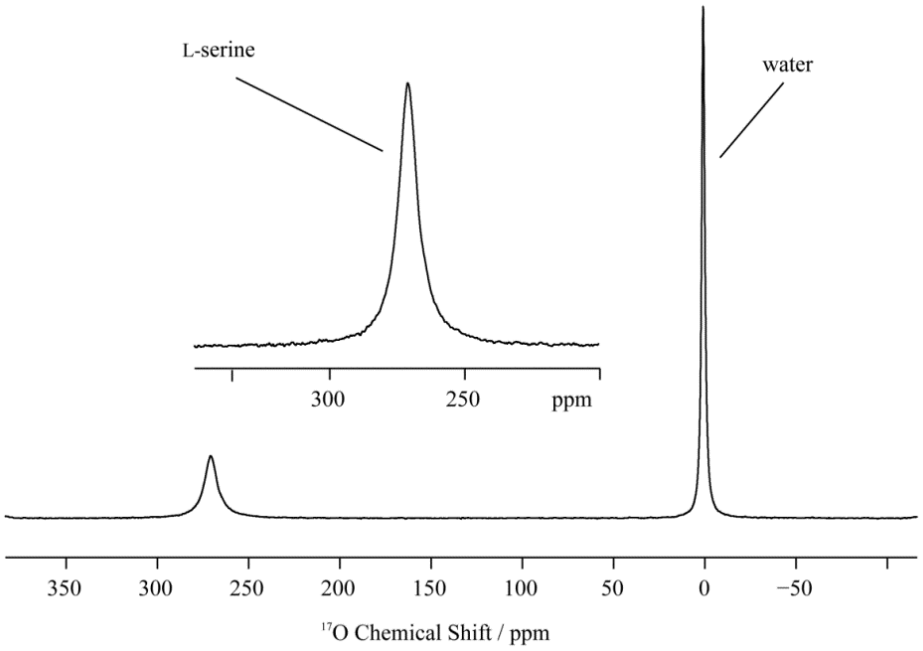


Figure 2.8. ^{17}O-NMR spectrum of L-serine (10 mM) enriched at the carboxylic site (expanded view) measured at 5°C in 50 mM triethanolamine buffer (pH = 7.8), 50 mM CsCl, and 10% D_2O/90% H_2O at a field strength of 14.1 T.

(TEA) buffer, pH 7.8, with 50 mM CsCl and 10% D_2O/90% H_2O in a standard 5 mm glass NMR tube at 14.1 T. The L-serine signal, shown in the expanded view, was observed at 271.2 ppm, as referenced to an external sample of pure H_2O measured at 25 °C. For comparison, Gerothanassis et al. measured the chemical shifts of the amino acids in aqueous solution and obtained values for L-serine of 255.8 ppm at pH 0.5 for the cationic form, 271.8 ppm at pH 6.5 for the zwitterionic form, and 270.1 ppm at pH 12.5 for the anionic form.[59]

2.2.2 Preparation of Enzyme and Enzyme Intermediates

TRPS (*S. typhimirium*) was provided in a 50 mM bicine buffer system, pH 7.8, with 50 mM CsCl (99%, Alfa-Aesar) at concentrations ranging between 9 mg/mL - 39 mg/mL from expression and purifications that were carried out as previously described (B. G. Caulkins).[7] Millipore Amicon Ultra centrifugal concentrators, MWCO = 3kDa or 30kDa (Merck Millipore), were used to exchange the bicine buffer with a 50 mM triethanolamine (99%; Pfaltz & Bauer) buffer, pH 7.8, containing 50 mM CsCl and 1 mM EDTA and to concentrate the protein sufficiently to yield final NMR sample concentrations from 0.8 mM to 1.5 mM active sites (0.4 mM to 0.75 mM TRPS αββα tetramer). TRPS concentrations were determined by measuring UV/vis absorbance values with an HP 8452A Diode Array Spectrophotometer. Total enzyme concentration was determined at 278 nm with a molar extinction coefficient (ε) of 33 $mM^{-1}cm^{-1}$, after which the enzyme was reacted with L-serine and 2-aminophenol (2AP) to yield the $E(Q)_{2AP}$ intermediate, ε = 48 $mM^{-1}cm^{-1}$ at 468 nm, as a check of purity (active/inactive protein).

The purity of all protein samples was > 95 %. The concentration of [^{17}O]-L-serine in all samples ranged between 5 - 15 mM, and the α-site inhibitor N-(4′-trifluoromethoxybenzenesulfonyl)-2-aminoethyl phosphate (F9) was added to all samples at a concentration of 4 mM, and the indole analogues benzimidazole (BZI) (99%; Sigma-Aldrich) and indoline (from indoline·HCl) were added to final concentrations that ranged between 6 – 10 mM by first dissolving solid reagents in ethanol at concentrations of 0.5 – 1 M and then delivering the appropriate amount to enzyme sample to yield the desired final concentration.[60] The addition of 2AP (99%; Acros Organics) was accomplished by first making a saturated solution of 2AP in acetonitrile (HPLC grade, EM Science) from which 2 µL were delivered to all $E(Q_3)_{2AP}$ samples. All intermediates measured at 11.7 T and 14.1 T also contained 10% D_2O (D, 99.9%; Cambridge Isotope Laboratories, Inc.) for lock and 1 mM 2,2-dimethylsilapentane-5-sulfonic acid (DSS) (Stohler Isotope Chemicals) to reference ^{1}H NMR spectra that were also measured as an additional means to track reaction progress.

2.2.3 Experimental ^{17}O Quadrupole Central Transition NMR

^{17}O QCT NMR experiments on TRPS intermediates were conducted at temperatures between 2 °C and 5 °C. All intermediates were measured at 11.7 T and 14.1 T field strengths. 16.4 T measurements were also performed on the $E(Q_3)_{indoline}$, $E(Q_3)_{2AP}$, and E(A-A)BZI intermediates, and the $E(Q_3)_{2AP}$ intermediate was additionally measured at 9.4 T. All spectra were referenced to the natural abundance ^{17}O signal in an external sample of pure water measured at 25°C and assigned the value of 0 ppm. 90° pulse width calibrations were first determined using the natural abundance ^{17}O signal of water and the

values obtained were divided by 3 for optimal excitation of the central transition of TS bound substrate.[18, 53] The 90° pulse width nutation frequencies calibrated for water were 42 kHz, 9 kHz, 30 kHz, and 46 kHz for field strengths of 9.4 T, 11.7 T, 14.1 T, and 16.4 T, respectively. Simultaneous mode quadrature detection was used to acquire ^{17}O spectra at all fields measured. A total (real + imaginary) of 2048 or 4096 points were collected and zero-filled to two times the number of real points, spectral widths ranged from 41 kHz to 100 kHz, relaxation delays were between 5 and 10 ms, and acquisition times were between 20 and 25 ms. A 90°- τ -180°- τ -acquire pulse sequence was used to measure the spectra of all intermediates. A 90°-τ-180°-2τ-180°-2τ-180°-τ -acquire (triple-echo, 3π) was additionally used to measure the spectra of the E(A-A) and E(A-A)BZI intermediates at 14.1T and the E(A-A)BZI intermediate at 16.4 T to improve the observation of bound resonances. This is based on the selective excitation of the central transition in ^{17}O QCT NMR which leads to a nutation rate that is three times faster for protein-bound substrates than that of the broadband excitation felt by free substrate in solution.[18, 53] For all experiments, the τ delay times were between 5 and 50 μs and generally kept as short as the instrumentation timing controller could handle. The pulse diagrams and phase cycling lists are provided in Figure 2.9. All spectra were ultimately processed using Bruker's Topspin 3.2 software. Field specific details including hardware are listed below.

<u>9.4 T (400.37 MHz ^{1}H, 54.28 MHz ^{17}O):</u> Bruker Avance III spectrometer equipped with a Bruker 4 mm MAS broadband tunable probe. Approximately 96 μL of sample was contained in a 4 mm MAS rotor. The sample was not spun during acquisition.

11.7T (500.05 MHz ^{1}H, 67.79 MHz ^{17}O): Varian ^{UNITY}Inova spectrometer equipped with a 5 mm SW PFG probe. Approximately 350 µL of samples were contained in 5 mm Shigemi restricted volume NMR tubes magnetic susceptibility matched to D_2O (Shigemi Inc.).

14.1 T (600.01 MHz ^{1}H, 81.34 MHz ^{17}O): Bruker Avance I spectrometer equipped with a 5 mm BBO Z-grad probe. Approximately 350 µL of samples were contained in 5 mm Shigemi restricted volume NMR tubes magnetic susceptibility matched to D_2O.

16.4 T (699.69 MHz ^{1}H, 94.85 MHz ^{17}O): Bruker Avance I spectrometer equipped with a modified Alderman-Grant coil probe (MAGC) built and tuned to ^{17}O by C. V. Grant at the Center for NMR Spectroscopy and Imaging of Proteins, University of California, San Diego. Approximately 200 µL of samples were contained in 5 mm by 1 cm flat bottom glass NMR tubes.[61]

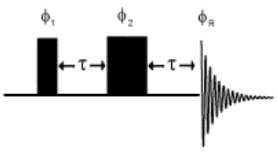
2.3 Results

2.3.1 Experimental Spectra

Stacked QCT NMR spectra for the $E(Q_3)_{indoline}$, $E(Q_3)_{2AP}$, E(A-A), and E(A-A)BZI are shown in Figures 2.10, 2.11, 2.12, and 2.13 respectively. The enzyme-bound ^{17}O QCT signals are marked with an asterisk in all spectra and the dynamic frequency shift of the signals with changing field is evident. The large signals at 0 ppm and in the vicinity of 270 ppm in all spectra are from the ^{17}O signals of water at natural abundance and from free [^{17}O]-L-serine (272 ppm) and pyruvate (268 ppm) respectively, the latter of which is the product of a slow side reaction.[62] With the exception of the $E(Q_3)_{2AP}$ intermediate at all fields measured and the $E(Q_3)_{indoline}$ at 11.7 T, two resonances

pertaining to the distinct carboxylic oxygen atoms bound to the enzyme active site are apparent. In general the bound resonances overlap with each other or with the large signals of the free L-Ser and pyruvate depending on the intermediate and the field at which measured. In particular, both aminoacrylates (Figures 2.12 and 2.13) had enzyme bound resonances at higher fields (14.1T and 16.4T) were very close to the free signals and employing the triple-echo pulse sequence was necessary to even obtain a good value for the δ_{obs}. As an example, Figure 2.14 shows E(A-A)BZI spectra collected at 16.4T with (bottom) and without (top) the triple-echo pulse sequence demonstrating the ability to attenuate free ^{17}O signals by taking advantage of the different nutation rates between free and bound states. In all cases, a robust means to fit the multiple spectra for each intermediate was necessary. Instead of attempting to fit each spectrum individually and combine the results a simultaneous global fit of all of the spectra collected for a given intermediate was accomplished and is discussed in the next section.

Pulse Sequence Diagrams and Phase Cycling Lists:

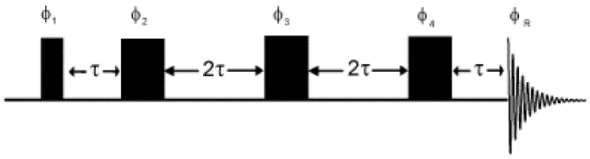


90°- τ -180°- τ –acquire phase cycle list:

φ_1 = 0 1 2 3

φ_2 = 0 0 0 0

φ_R = 0 3 2 1



Triple-echo 90°-τ-180°-2τ -180°-2τ -180°- τ –acquire phase cycle list:

φ_1 = 0 1 2 3

φ_2 = 0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3

φ_3 = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

φ_4 = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

φ_R = 0 3 2 1 2 1 0 3 0 3 2 1 2 1 0 3
2 1 0 3 0 3 2 1 2 1 0 3 0 3 2 1
0 3 2 1 2 1 0 3 0 3 2 1 2 1 0 3
2 1 0 3 0 3 2 1 2 1 0 3 0 3 2 1
2 1 0 3 0 3 2 1 2 1 0 3 0 3 2 1
0 3 2 1 2 1 0 3 0 3 2 1 2 1 0 3
2 1 0 3 0 3 2 1 2 1 0 3 0 3 2 1
0 3 2 1 2 1 0 3 0 3 2 1 2 1 0 3

Notes:

1. Phase tables list (pulse phase in degrees)/90^O
2. Phase lists repeat cyclically from start to match the length of the longest phase cycle

Figure 2.9. Pulse sequences and phase cycling lists for the single echo 90°-τ-180°-τ-acquire and the triple echo, 3π, 90°-τ-180°-2τ--180°-2τ-180°-τ-acquire pulse programs used to acquire ^{17}O-QCT NMR spectra.

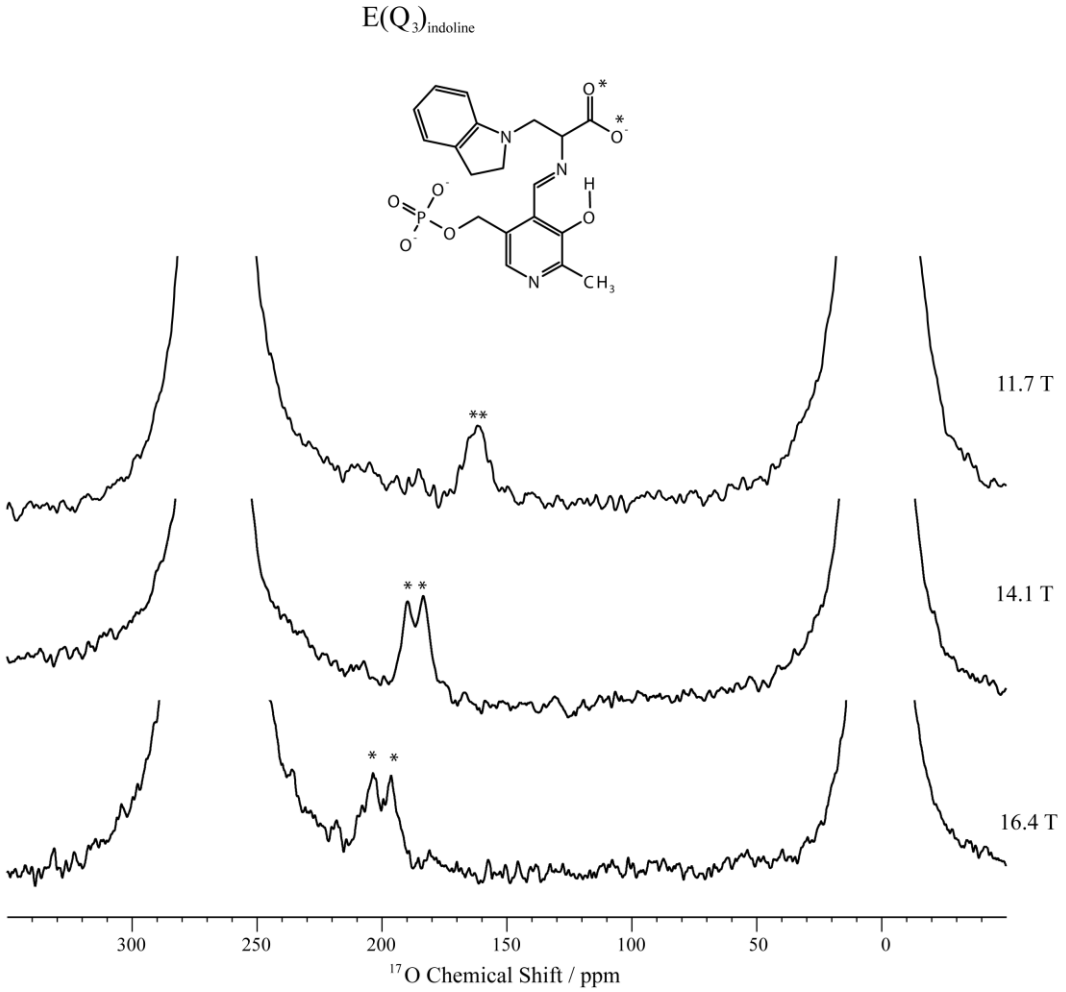


Figure 2.10. ^{17}O QCT NMR spectra of the $E(Q_3)_{indoline}$ intermediate in TRPS at multiple magnetic fields. The enzyme bound intermediates are marked with asterisks. Total number of transients and experiment times: at 11.7 T, 2.4×10^6 transients and 22 hours; at 14.1 T, 2.2×10^6 transients and 20 hours; at 16.4 T, 6.4×10^6 transients and 59 hours. The spectra are displayed with 100 Hz line broadening applied. Experiments were conducted at 3±2°C in pH 7.8 buffered aqueous solution.

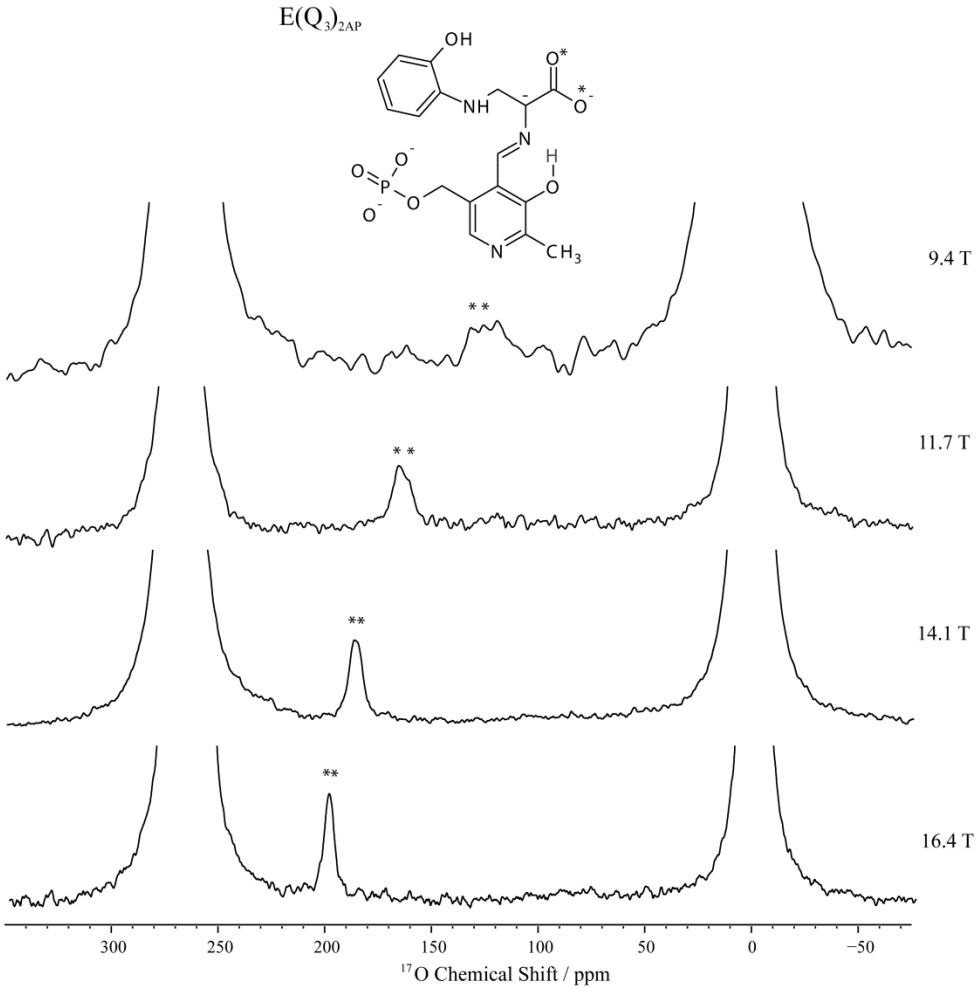


Figure 2.11. ^{17}O QCT NMR spectra of the $E(Q_3)_{2AP}$ intermediate. Enzyme bound signals are marked with asterisks. Total number of transients and experiment times: at 9.4 T, 11.5×10^6 transients, 92 hours; 11.7 T, 2.3×10^6 transients and 17 hours; at 14.1 T, 6.0×10^6 transients and 39 hours; at 16.4 T, 8.9×10^6 transients and 77 hours. The spectra are displayed with 100 Hz line broadening applied, with the exception of that measured at 9.4 T which is displayed with 300 Hz line broadening. Experiments were conducted at 3±2°C in pH 7.8 buffered aqueous solution.

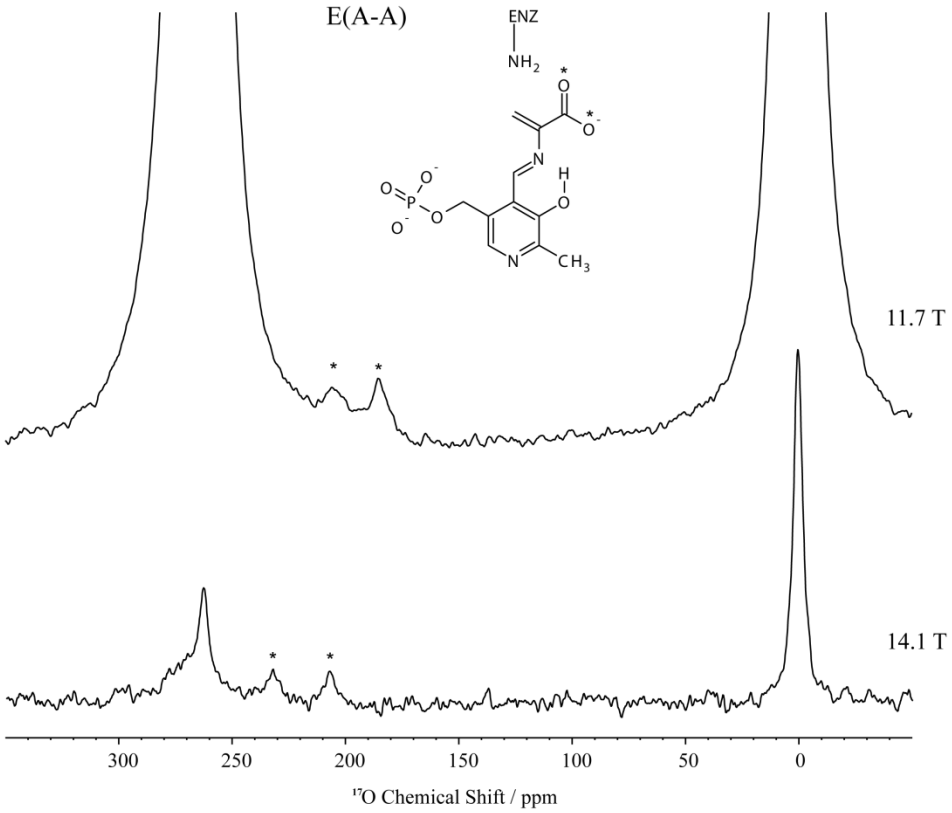


Figure 2.12. ^{17}O QCT NMR spectra of the E(A-A) intermediate at two fields, 11.7 T and 14.1 T as indicated. The spectrum measured at 11.7 T was acquired using the 90°-τ-180°-τ-acquire pulse sequence and consisted of 2.3×10^6 transients for a total acquisition time of 36 hours. The spectrum measure at 14.1 T was acquired using the triple-echo, 3π, 90°-τ-180°-2τ-180°-2τ-180°-τ-acquire pulse sequence to improve resolution of the enzyme bound ^{17}O QCT signals (marked with asterisks) by attenuation of the signals corresponding to free substrate/product and consisted of 0.9×10^6 transients at for a total acquisition time of 12 hours. As an additional note, the sample measured at 14.1 T had a reduced amount of substrate ([^{17}O]-L-serine) added as yet another means to improve resolution of the enzyme bound signals, which in-turn reduced the time available for acquisition. The spectra are displayed with 100 Hz line broadening. Experiments were conducted at 3±2°C in pH 7.8 buffered aqueous solution.

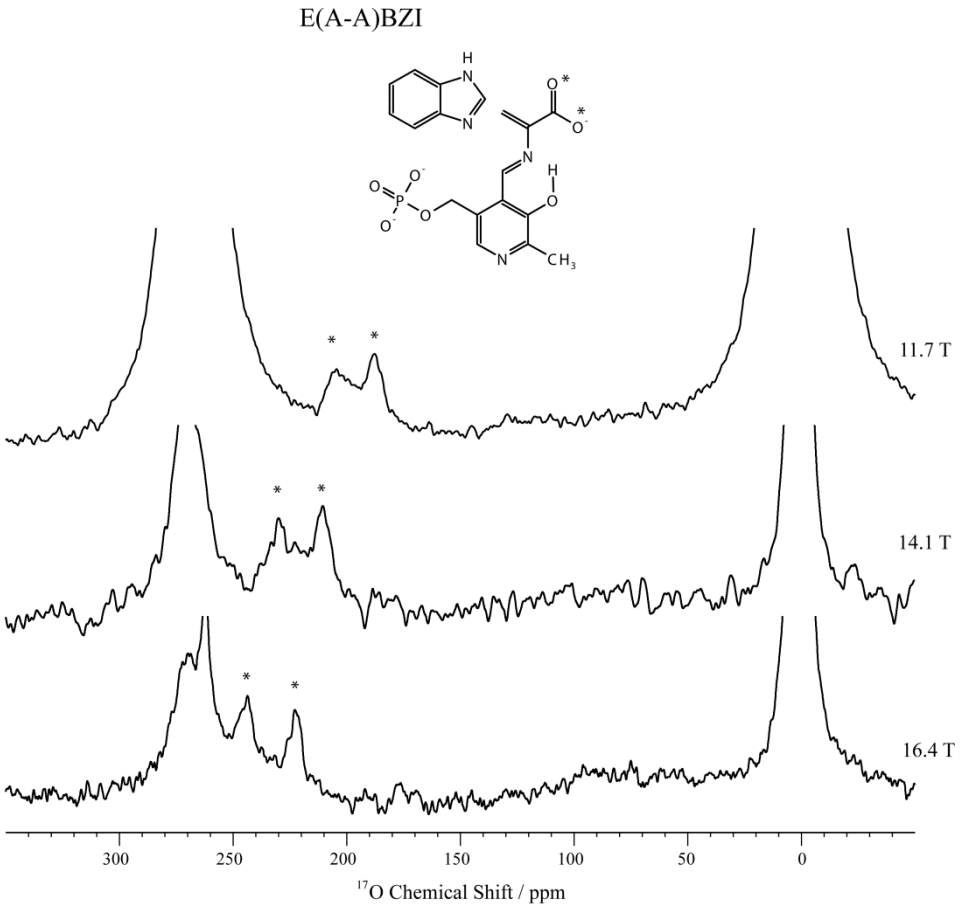


Figure 2.13. ^{17}O QCT NMR spectra of the E(A-A)BZI intermediate in TRPS at multiple fields. The enzyme-bound intermediates are marked with asterisks. Experiments measure at 14.1 T and 16.4 T were acquired using the triple-echo pulse sequence to improve observation of the bound peaks by attenuating free substrate/product signals. Total number of transients and experiment times : at 11.7 T, 7.9×10^6 transients and 72 hours, at 14.1 T, 3.3×10^6 transients and 40 hours, at 16.4 T, 12.9×10^6 transients and 156 hours. The spectra are displayed with 100 Hz line broadening applied. Experiments were conducted at 3±2°C in pH 7.8 buffered aqueous solution.

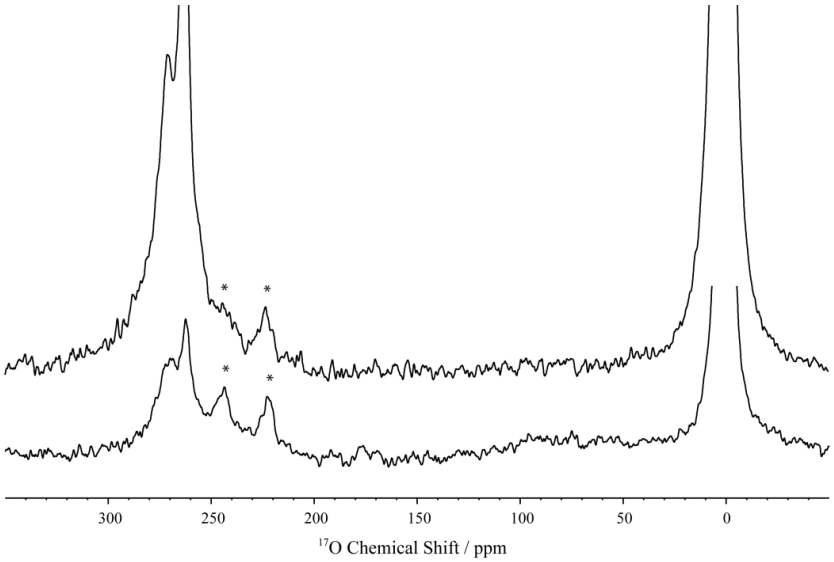


Figure 2.14. Comparison of the E(A-A)BZI spectra acquired with a single spin echo (top) and a triple-echo (bottom). Enzyme-bound signals are marked with asterisks. Both spectra were acquired with 12.9×10^6 transients. Total acquisition times were 121 hours and 156 hours for the single- and triple-echo experiments respectively. The triple-echo pulse sequence was used to attenuate the signals originating from the free substrate/product (and water) to improve observation of the downfield enzyme-bound QCT signal. The triple-echo pulse sequence achieves this by capitalizing on the faster nutation frequency of the central transition (3× faster for spin-5/2) in the limit of slow isotropic motion, effectively selective excitation when the RF field is much smaller than the quadrupolar interaction, versus the broadband excitation of all single quantum coherences in the fast motion limit (extreme narrowing) where the firs-order quadrupolar interaction is effectively averaged.

2.3.2 Global Fitting of Spectral Data

Measurements of the observed chemical shift at multiple fields can be used to calculate the δ_{iso} and P_Q in a straightforward manner given the linear relationship of δ_{obs} with ν_0^{-2} as shown in Equation (27) and presented here again.

$$\Delta\delta_d(ppm) = \delta_{obs} - \delta_{iso} = -6000\left(\frac{P_Q}{\nu_0}\right)^2 \tag{27}$$

The shielding anisotropy product parameter, P_{SA}, and rotational correlation time, τ_c, may be extracted from the field-dependent line shape of the QCT signals according to Equation (22) which derives from the transverse relaxation properties of the CT in the

limit of slow isotropic motion and was presented earlier. Varying extents of signal overlap in all intermediates measured complicated fitting each spectrum individually and, therefore, a more robust means was used to calculate their lineshapes. Taking advantage of the magnetic field dependence ($\nu_0 = -\gamma B_0/2\pi$) of both the δ_{obs} (Equation (27)) and $\Delta\nu_{1/2}$ (Equation (22)), a global, simultaneous fit of all spectra for a given intermediate was accomplished and ν_0 acted as a natural constraint. The simultaneous fitting of the enzyme-bound spectra of a given intermediate at multiple magnetic field strengths to Equations (22) and (27) allows for a more reliable extraction of δ_{iso}, P_Q, P_{SA}, and τ_c. In practice, this gives a Lorentzian line shape for each peak that is described by Equation (28) (written to yield units of ppm in the chemical shift dimension).

$$L\left(I, \nu_0, \delta_{obs}, \Delta\nu_{1/2}\right) = \frac{\nu_0 I(\nu_0)}{\Delta\nu_{1/2}\left(1 + 4\left(\frac{\nu_{0(\delta-\delta_{obs})}}{\Delta\nu_{1/2}}\right)^2\right)} \tag{28}$$

The intensity, $I(\nu_0)$, is not constrained and, therefore, is given as many free parameters as the number of different field strengths measured. As an example, Equation (29) contains a cubic, four free-parameter Equation for the intensity of intermediates measured at four different magnetic field strengths as was accomplished for the $E(Q_3)_{2AP}$.

$$I(\nu_0) = a + b\nu_0 + c\nu_0^2 + d\nu_0^3, \text{(4 different field measurements)} \tag{29}$$

Substitution of Equations (22), (27), and (29) into (28) gives an expression for the field-dependent line shape that were used to fit the spectra and extract δ_{iso}, P_Q, P_{SA}, and τ_c.

To fit the data, the real portion of each spectrum was exported as a JCAMP 6.0 file using Bruker's TopSpin 3.2 program. The spectra used in the fitting procedure were zero-filled only. No other post acquisition processing was performed although spectra

displayed in figures do have line broadening applied, as indicated, for clarity. The JCAMP files were imported into Mathematica. Each spectrum was scaled to the same RMSD noise floor as seen in Figure 2.15. The spectra were then subjected to a global, simultaneous, non-linear least-squares regression to find the best fit parameters satisfying Equation (28). A τ_c of 182 ns was used as a fixed parameter in the final fitting of all spectra. This value was first obtained as a result of the simultaneous fitting of the $E(Q_3)_{2AP}$ spectra at four field strengths and is consistent with the value expected for the τ_c of a 143 kDa protein in the 2 – 5 °C temperature range at which the experiments were conducted. Indeed, calculation of the τ_c using the HYDRONMR program was in excellent agreement with the value determined from the ^{17}O QCT spectral fitting.[63]

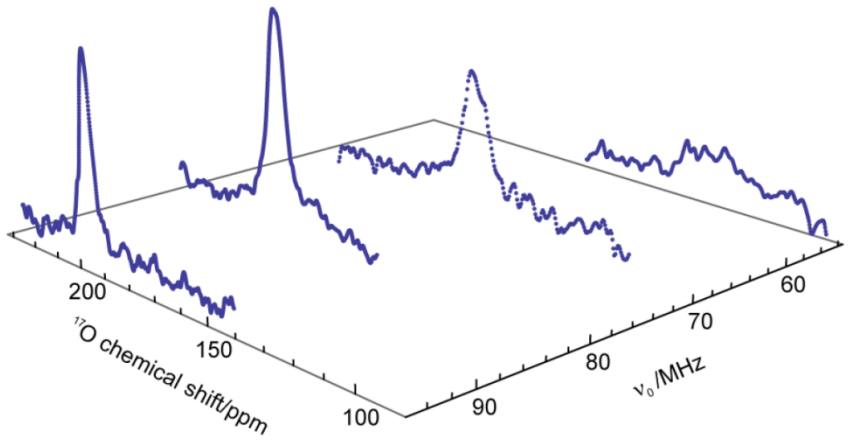


Figure 2.15. $E(Q_3)_{2AP}$ intermediate enzyme-bound region of ^{17}O QCT NMR spectra measured as a function of ν_0. Spectra displayed in this figure were subjected to 100 Hz line broadening.

The $E(Q_3)_{indoline}$ and E(A-A)BZI intermediates were measured at three fields (11.7, 14.1, and 16.4 T). The free substrate/product signals in the E(A-A)BZI intermediate at 16.4 T were fit and subtracted from the spectrum prior to global fitting of

the bound signals. The E(A-A) intermediate was measured at 11.7 and 14.1 T and the δ_{iso} and P_Q were calculated from the δ_{obs} and Equation (27). To improve resolution of the downfield (O1) signal at 14.1 T, a reduced amount of [^{17}O]-L-serine substrate was added and the triple-echo pulse sequence was used, which increases experiment time for a given number of transients and also results in the loss of some S/N for the enzyme bound signals. These factors, combined with the intrinsically shorter lifetime of the E(A-A), did not yield a spectrum of suitable quality to perform a reliable fit. The results for all intermediates are listed in Table 2.1 and Figure 2.16 provides a generic chemical structure of the PLP/substrate complex indicating oxygen atoms O1 and O2. Assignment of O1 and O2 were based on first-principles calculation of chemical shifts and are detailed in Chapter 3. Table 2.2. lists the canonical ranges of the δ_{iso}, P_Q, and P_{SA} for carboxylic acid oxygen types, ionized carboxylate, COO^-, hydroxyl, C-OH, and carbonyl, C=O, from solid state ^{17}O NMR studies discussed in the introduction.[10] Figure 2.17 A-C displays the enzyme bound resonances overlaid with their best-fit lineshapes from the global fitting procedure for the $E(Q_3)_{2AP}$, $E(Q_3)_{indoline}$, and E(A-A)BZI intermediates respectively.

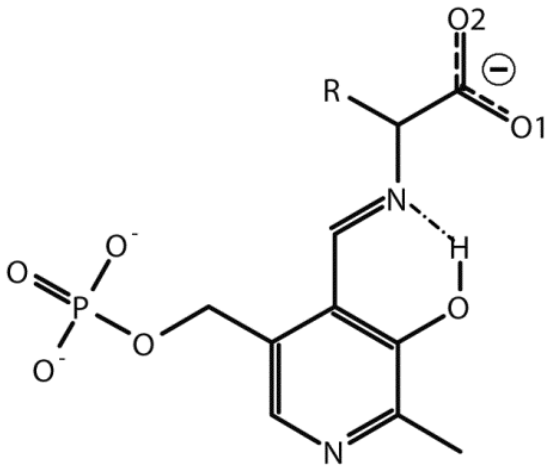


Figure 2.16. Chemical structure of PLP/substrate complex indicating assignment of carboxylic oxygen atoms O1 and O2.

Table 2.1. Isotropic chemical shifts, δ_{iso}, quadrupole, and shift anisotropy product parameters, P_Q and P_{SA} of [^{17}O]-L-serine measured in the enzyme active site of TRPS by ^{17}O QCT NMR.[11]

	δ_{iso}/ppm	P_Q/MHz	P_{SA}/ppm*
$E(Q_3)_{indoline}$			
O1	242.7 ± 0.7	7.66 ± 0.05	226 ± 26
O2	232.7 ± 0.6	7.40 ± 0.05	190 ± 26
$E(Q_3)_{2AP}$			
O1	238.7 ± 0.6	7.69 ± 0.04	196 ± 30
O2	237.2 ± 0.5	7.37 ± 0.04	223 ± 25
E(A-A)BZI			
O1	287 ± 2	7.9 ± 0.1	686 ± 22
O2	258 ± 2	7.2 ± 0.1	705 ± 22
E(A-A)			
O1	292 ± 2	8.2 ± 0.1	-
O2	258 ± 2	7.5 ± 0.1	-

*Extracted using τ_c = 182 ns.

Table 2.2. Canonical ranges for δ_{iso}, P_Q, and P_{SA}, for the carboxylic acid oxygen types, ionized carboxylate, COO^-, and acid-form hydroxyl, C-OH, and carbonyl, C=O, from ^{17}O solid-state NMR measurements of amino acids.[10]

	δ_{iso} / ppm	P_Q / MHz	P_{SA} / ppm
COO^-	250 - 314	6.4 - 8.2	263 - 429
C-OH	167 - 187	7.0 - 7.8	203 - 238
C=O	303 - 355	7.9 - 8.7	474 - 600

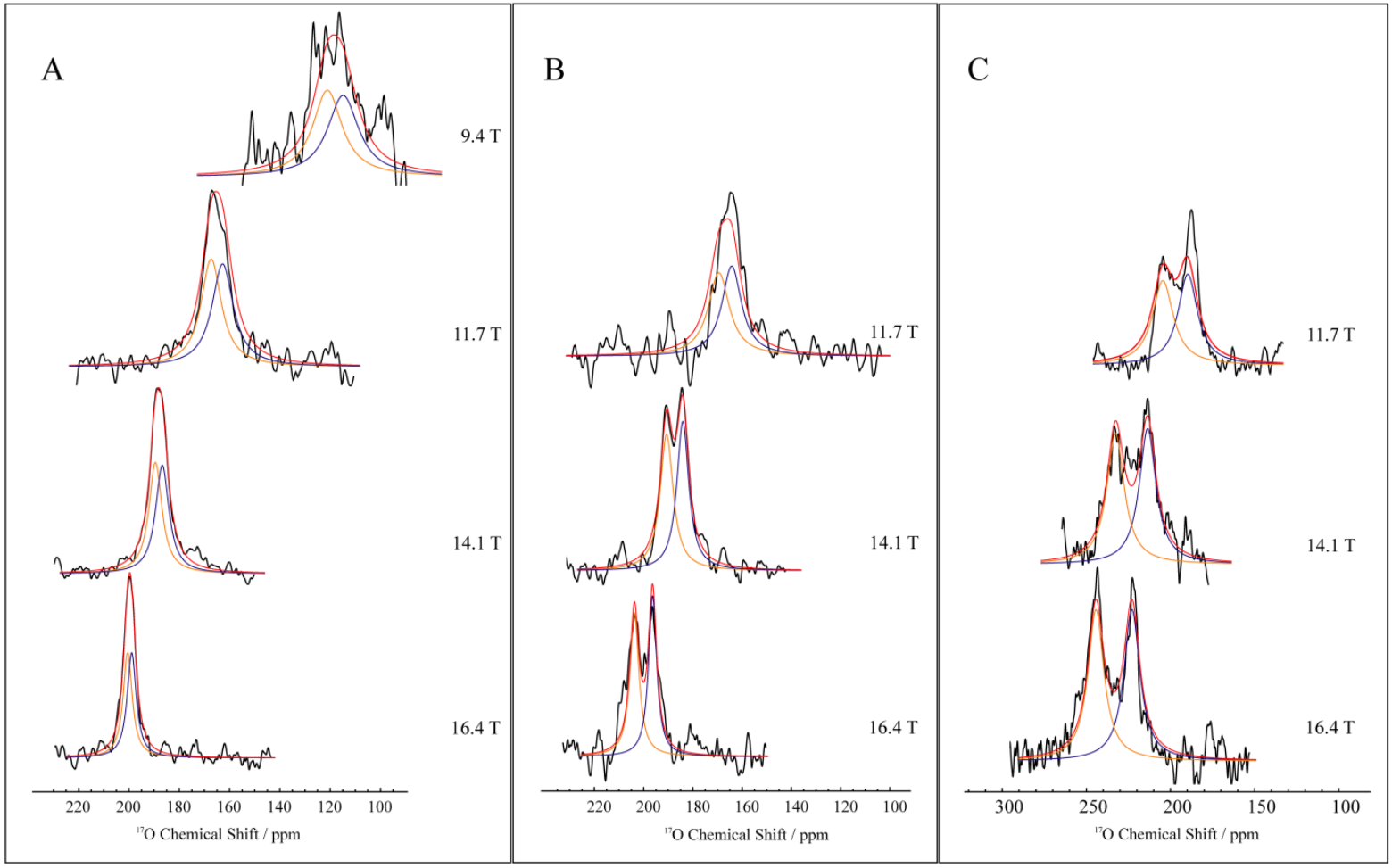


Figure 2.17. Traces of the calculated best-fit component lineshapes for O1 (orange) and O2 (blue), and their sum, O1 + O2 (red) overlaid on experimental ^{17}O QCT NMR spectra at the magnetic fields indicated for the $E(Q_3)_{2AP}$ (A), $E(Q_3)_{indoline}$ (B), and E(A-A)BZI (C) intermediates. 100 Hz line broadening was applied to all spectra.

2.4 Discussion

2.4.1 Quinonoid Intermediates

In comparison with shifts measured for amino acids in the solid state, where carboxylates are observed between 250 ppm – 314 ppm and the carbonyl and hydroxyl oxygens of the acid form range between 303 ppm – 353 ppm and 167 ppm – 187 ppm respectively, the isotropic values of 243 ppm and 233 ppm and 237 ppm and 239 ppm for the $E(Q_3)_{indoline}$ and $E(Q_3)_{2AP}$ intermediates respectively are surprising. In the earlier ^{15}N and ^{13}C NMR crystallography study of the $E(Q_3)_{indoline}$ the combination of experimental input and computational predictions of chemical shift suggested tautomerization in which only the oxygen nearest the Schiff-base nitrogen (O1) became protonated.[6] The fact that both shifts in both quinonoids fall below the canonical range of carboxylates as measured in amino acids would suggest that, at some point, each oxygen must be protonated. Three-site exchange models, however, did not select an equilibrium in which there was a 50/50 average of each oxygen, O1 and O2, spending equal times as hydroxyl and carbonyl in acid forms.[6] The shifts would be expected to diverge if only one oxygen were protonated for a significant amount of time in any equilibrium.[10] The P_Q values measured in the quinonoids do not offer any assistance as they fall within a range common to both carboxylates and hydroxyl species. The P_{SA} values measured do fall below the range of those for carboxylates but do occupy a range consistent with the hydroxyl oxygen of acid forms, although caution must be exercised as the number of experimental values for which the ranges were calculated was relatively small.

2.4.2 Aminoacrylate Intermediates

The E(A-A) intermediate, which precedes E(Q_3) in the catalytic cycle, provides context for interpreting the quinonoid chemical shifts. In the E(A-A) intermediate, the δ_{iso} and P_Q values calculated from the dynamic frequency shifts, Equation (27) were 292 ppm and 258 ppm and 7.9 MHz and 7.2 MHz for O1 and O2 respectively. As mentioned previously, the short life of E(A-A) prevented acquisition of spectra suitable for the global spectral analysis. Fortunately, E(A-A) is stabilized by the addition of benzimidazole, BZI, which extends the life of the intermediate by more than seven-fold allowing for up to a week or longer of data acquisition. The δ_{iso} and P_Q values extracted from the global fitting procedure for the E(A-A)BZI were 287 ppm and 258 ppm and 8.2 MHz and 7.5 MHz for O1 and O2 respectively. Additionally the P_{SA} values were determined from the fitting procedure, using a fixed 182 ns τ_c, to be 686 ppm and 705 ppm for O1 and O2 respectively. The P_{SA} values are slightly higher than the canonical range observed for amino acids in the solid-state. The reason for this is not known and it may be related to the unique enzyme active site environment or perhaps differences in experimental conditions such as magnetic field stability or temperature calibration or perhaps a combination of these factors. The δ_{iso} and P_Q values, which constitute the most reliably extracted parameters, fall solidly within the range of canonical carboxylates.[10]

2.4.3 Conclusions

While E(A-A) and E(A-A)BZI present ^{17}O NMR chemical shifts consistent with carboxylate forms, both $E(Q_3)_{indoline}$ and $E(Q_3)_{2AP}$ show upfield shifted resonances that suggest potential acid form character. But the empirical correlation between shift and

chemical structure presented earlier with regards to amino acids would predict both an upfield and downfield shifted peak for a quinonoid species undergoing fast-exchange between carboxylate and acid forms (assuming, as appears to be the case, that the carboxylate group is not free to rotate). Taking averages of the canonical hydroxyl and carbonyl chemical shifts can produce the fast exchange δ_{iso} values that are observed. An obvious question, however, is how would this be possible with the absence of a suitable hydrogen donating group as the hydrogen bonding partners to O1 and O2 are either backbone amide nitrogens or, for O2, the hydroxyl sidechain of βThr110.[27] An alternate hypothesis is that the increased shieldings in the quinonoid intermediates result from changes in the carboxylate hydrogen bonding interactions that become stronger in going from the E(A-A) to the E(Q_3). The X-ray crystal structures of the quinonoid intermediates (PDBID 3PR2 (indoline) and 4HPJ (2AP)) show no change in the number of hydrogen bond donors compared to the aminoacrylate structures (PDBID 4HN4 (A-A) and 4HPX (BZI)) and only one closer contact, that between βThr110 and O2, in the quinonoids.[6, 27] A third hypothesis is that the increased carboxylate shieldings are due to changes in the electronic structure that are inherent to the quinonoid form, reflecting increased electron density within the π-bonding network, which is formally carrying a negative charge for these non-classical quinonoid intermediates.

Ultimately, interpretation of the observed chemical shift values in the quinonoids would require first principles, quantum calculations of the NMR observables using a cluster model of the enzyme active site. At this point the oxygen chemical shifts are added to the number of critical NMR probes being experimentally measured in the active

site of TRPS in the NMR crystallography approach. The first principles calculations were intently focused on the longer lived $E(Q_3)_{2AP}$ and the process and results are the topic of the next chapter.

2.5 References

1. Toney, M.D. Biochimica Et Biophysica Acta, 2011. **1814**(11): p. 1407-18.

2. Hayashi, H. Journal of Biochemistry, 1995. **118**(3): p. 463-473.

3. Chan-Huot, M., A. Dos, R. Zander, S. Sharif, P.M. Tolstoy, S. Compton, E. Fogle, M.D. Toney, I. Shenderovich, G.S. Denisov, and H.H. Limbach. Journal of the American Chemical Society, 2013. **135**(48): p. 18160-18175.

4. Sharif, S., G.S. Denisov, M.D. Toney, and H.H. LimbachJournal of the American Chemical Society, 2007. **129**(19): p. 6313-6327.

5. Chan-Huot, M., S. Sharif, P.M. Tolstoy, M.D. Toney, and H.H. Limbach. Biochemistry, 2010. **49**(51): p. 10818-10830.

6. Lai, J.F., D. Niks, Y.C. Wang, T. Domratcheva, T.R.M. Barends, F. Schwarz, R.A. Olsen, D.W. Elliott, M.Q. Fatmi, C.E.A. Chang, I. Schlichting, M.F. Dunn, and L.J. Mueller. Journal of the American Chemical Society, 2011. **133**(1): p. 4-7.

7. Caulkins, B.G., B. Bastin, C. Yang, T.J. Neubauer, R.P. Young, E. Hilario, Y.M.M. Huang, C.E.A. Chang, L. Fan, M.F. Dunn, M.J. Marsella, and L.J. Mueller. Journal of the American Chemical Society, 2014. **136**(37): p. 12824-12827.

8. Lemaitre, V., M.E. Smith, and A. Watts. Solid State Nuclear Magnetic Resonance, 2004. **26**(3-4): p. 215-235.

9. Wu, G. Progress in Nuclear Magnetic Resonance Spectroscopy, 2008. **52**(2-3): p. 118-169.

10. Wong, A. and F. Poli. Annual Reports on NMR Spectroscopy, ed. G.A. Webb. Vol. 83. 2014, London: Elsevier.

11. Young, R.P., B.G. Caulkins, D. Borchardt, D.N. Bulloch, C.K. Larive, M.F. Dunn, and L.J. Mueller. Angewandte Chemie-International Edition, 2016. **55**(4): p. 1350-1354.

12. Gerothanassis, I.P.. Progress in Nuclear Magnetic Resonance Spectroscopy, 2010. **56**(2): p. 95-197.

13. Zhu, J.F., E. Ye, V. Terskikh, and G. Wu. Angewandte Chemie-International Edition, 2010. **49**(45): p. 8399-8402.

14. Oldfield, E., H.C. Lee, C. Coretsopoulos, F. Adebodun, K.D. Park, S.T. Yang, J. Chung, and B. Phillips. Journal of the American Chemical Society, 1991. **113**(23): p. 8680-8685.

15. Lee, H.C., K. Cummings, K. Hall, L.P. Hager, and E. Oldfield. Journal of Biological Chemistry, 1988. **263**(31): p. 16118-16124.

16. Lee, H.C. and E. Oldfield. Journal of the American Chemical Society, 1989. **111**(5): p. 1584-1590.

17. Hanashima, S., N. Fujiwara, K. Matsumoto, N. Iwasaki, G.Q. Zheng, H. Torigoe, K. Suzuki, N. Taniguchi, and Y. Yamaguchi. Chemical Communications, 2013. **49**(14): p. 1449-1451.

18. Zhu, J.F. and G. Wu. Journal of the American Chemical Society, 2011. **133**(4): p. 920-932.

19. Bull, T.E., S. Forsen, and D.L. Turner. Journal of Chemical Physics, 1979. **70**(6): p. 3106-3111.

20. Hubbard, P.S.. Journal of Chemical Physics, 1970. **53**(3): p. 985-987.

21. Werbelow, L. and G. Pouzard. Journal of Physical Chemistry, 1981. **85**(25): p. 3887-3891.

22. Werbelow, L.G.. Journal of Chemical Physics, 1979. **70**(12): p. 5381-5383.

23. Butler, A., M.J. Danzitz, and H. Eckert. Journal of the American Chemical Society, 1987. **109**(6): p. 1864-1865.

24. Chung, C.W. and S. Wimperis. Molecular Physics, 1992. **76**(1): p. 47-81.

25. Caulkins, B.G., C. Yang, E. Hilario, L. Fan, M.F. Dunn, and L.J. Mueller. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2015. **1854**(9): p. 1194-1199.

26. Roy, M., S. Keblawi, and M.F. Dunn. Biochemistry, 1988. **27**(18): p. 6698-6704.

27. Niks, D., E. Hilario, A. Dierkers, H. Ngo, D. Borchardt, T.J. Neubauer, L. Fan, L.J. Mueller, and M.F. Dunn. Biochemistry, 2013. **52**(37): p. 6396-6411.

28. Levitt, M.H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. 2 ed. 2008, West Sussex, U. K.: John Wiley & Sons Ltd.

29. Gready, J.E.. Journal of the American Chemical Society, 1981. **103**(13): p. 3682-3691.

30. Forsen, S. and B. Lindman, *Ion Binding in Biological Systems Measured by Nuclear Magnetic Resonance Spectroscopy*, in *Methods of Biochemical Analysis*, D. Glick, Editor. 1981, John Wiley & Sons: New York. p. 289-486.

31. Freude, D., *Quadrupolar Nuclei in Solid-State Nuclear Magnetic Resonance*. Encyclopedia of Magnetic Resonance, ed. R.A. Meyers. 2000, Chichester: John Wiley & Sons.

32. Aramini, J.M., T. Drakenberg, T. Hiraoki, Y. Ke, K. Nitta, and H.J. Vogel. Biochemistry, 1992. **31**(29): p. 6761-6768.

33. Aramini, J.M., M.W. Germann, and H.J. Vogel. Journal of the American Chemical Society, 1993. **115**(21): p. 9750-9753.

34. Aramini, J.M., M.W. Germann, and H.J. Vogel. Journal of Magnetic Resonance, 1997. **129**(2): p. 111-114.

35. Aramini, J.M., T. Hiraoki, M.R. Grace, T.W. Swaddle, E. Chiancone, and H.J. Vogel. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology, 1996. **1293**(1): p. 72-82.

36. Aramini, J.M. and H.J. Vogel. Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire, 1998. **76**(2-3): p. 210-222.

37. Vogel, H.J. and W.H. Braunlin. Journal of Magnetic Resonance, 1985. **62**(1): p. 42-53.

38. Bull, T., J.E. Norne, P. Reimarsson, and B. Lindman. Journal of the American Chemical Society, 1978. **100**(15): p. 4643-4647.

39. Delville, A., C. Detellier, and P. Laszlo. Journal of Magnetic Resonance, 1979. **34**(2): p. 301-315.

40. Abragam, A., *Principles of Nuclear Magnetism*. International Series of Monographs on Physics 32. 1961, Oxford: Oxford University Press.

41. Man, P.P., *Quadrupole Couplings in Nuclear Magnetic Resonance*, in *Encyclopedia of Analytical Chemistry*, R.A. Meyers, Editor. 2000, John Wiley & Sons: Chichester. p. 12224-12265.

42. Adiga, S., D. Aebi, and D.L. Bryce. Canadian Journal of Chemistry-Revue Canadienne De Chimie, 2007. **85**(7-8): p. 496-505.

43. Zhu, J.F., E. Ye, V. Terskikh, and G. Wu. Journal of Physical Chemistry Letters, 2011. **2**(9): p. 1020-1023.

44. Zhu, J.F., I.C.M. Kwan, and G. Wu. Journal of the American Chemical Society, 2009. **131**(40): p. 14206-14207.

45. Boykin, D.W., *Oxygen-17 NMR in Organic Chemistry*. 1991: CRC Press.

46. Boykin, D.W., ed. *Oxygen-17 NMR Spectroscopy in Organic Chemistry*. 1991, CRC Press: Boca Raton.

47. Redfield, A.G.. IBM Journal of Research and Development, 1957. **1**(1): p. 19-31.

48. Redfield, A.G., *The Theory of Relaxation Processes*. Advances in Magnetic Resonance, ed. J.S. Waugh. Vol. 1. 1965, New York: Academic Press.

49. Halle, B. and H. Wennerstrom. Journal of Magnetic Resonance, 1981. **44**(1): p. 89-100.

50. Fouques, C.E.M. and L.G. Werbelow. Canadian Journal of Chemistry-Revue Canadienne De Chimie, 1979. **57**(17): p. 2329-2332.

51. Werbelow, L. and R.E. London. Concepts in Magnetic Resonance, 1996. **8**(5): p. 325-338.

52. Werbelow, L.G. and A.G. Marshall. Journal of Magnetic Resonance, 1981. **43**(3): p. 443-448.

53. Man, P.P., J. Klinowski, A. Trokiner, H. Zanni, and P. Papon. Chemical Physics Letters, 1988. **151**(1-2): p. 143-150.

54. Butler, A. and H. Eckert. Journal of the American Chemical Society, 1989. **111**(8): p. 2802-2809.

55. Man, P.P.. Molecular Physics, 1993. **78**(2): p. 307-318.

56. Pike, K.J., V. Lemaitre, A. Kukol, T. Anupold, A. Samoson, A.P. Howes, A. Watts, M.E. Smith, and R. Dupree. Journal of Physical Chemistry B, 2004. **108**(26): p. 9256-9263.

57. Wong, A., K.J. Pike, R. Jenkins, G.J. Clarkson, T. Anupold, A.P. Howes, D.H.G. Crout, A. Samoson, R. Dupree, and M.E. Smith. Journal of Physical Chemistry A, 2006. **110**(5): p. 1824-1835.

58. Dong, S., R. Ida, and G. Wu. Journal of Physical Chemistry A, 2000. **104**(47): p. 11194-11202.

59. Gerothanassis, I.P., R.N. Hunston, and J. Lauterwein. Magnetic Resonance in Chemistry, 1985. **23**(8): p. 659-665.

60. Ngo, H., N. Kimmich, R. Harris, D. Niks, L. Blumenstein, V. Kulik, T.R. Barends, I. Schlichting, and M.F. Dunn. Biochemistry, 2007. **46**(26): p. 7740-7753.

61. Grant, C.V., Y. Yang, M. Glibowicka, C.H. Wu, S.H. Park, C.M. Deber, and S.J. Opella. Journal of Magnetic Resonance, 2009. **201**(1): p. 87-92.

62. Dierkers, A.T., D. Niks, I. Schlichting, and M.F. Dunn. Biochemistry, 2009. **48**(46): p. 10997-11010.

63. de la Torre, J.G., M.L. Huertas, and B. Carrasco. Journal of Magnetic Resonance, 2000. **147**(1): p. 138-146.

Chapter 3

NMR Crystallography in the Active Site of Tryptophan Synthase: Characterization of the 2-Aminophenol Quinonoid Intermediate

3.1 Introduction

3.1.1 Background.

Pyridoxal-5'-phosphate (PLP) enzymes present an excellent example of how complex chemical transformations can take place over the course of many elementary steps and general acid-base catalysis plays a key role in many of the reactions discussed in Chapter 1.[1-5] To date, however, largely indirect methods have been available to enzymologists seeking to characterize enzyme reaction mechanisms.[5-9] Advances in X-ray crystallography have been a boon to the research community as much can be inferred from knowing the relative positions of catalytic residues with respect to cofactors and/or substrates.[10-13] The inability of X-ray diffraction techniques to locate protons, however, often leaves open the question of the protonation states of ionizable species, a serious drawback given the important implications protonation states play in catalysis.[5, 14] Also, the static nature of the crystal structure belies the true dynamic nature of enzymes. NMR spectroscopy is unrivaled in its ability to probe highly localized chemical details such as protonation or ionization states and, moreover, it can be used to characterize dynamical processes such as those associated with chemical or conformational exchange.[5, 15]

Although it is routine to supply ^{13}C and ^{15}N enriched amino acids in protein expression media, primarily for structural calculations on smaller enzymes, < 30 kDa, in the solution state, it is virtually impossible to limit the enrichment only to the active site

residues.[16] In large enzymes, like tryptophan synthase that are generally restricted to solid state NMR methods, the signal overlap from a fully ^{13}C and ^{15}N enriched sample would be difficult, if not impossible, to unravel. As a result**,** NMR measurements of individual nuclei in the active sites of large enzymes are often limited to ^{13}C and ^{15}N enriched cofactors and substrates.[13, 14] The key to interpreting the highly localized information that NMR reporters provide requires detailed information about the surrounding environment - that is the locations of catalytic residues in proximity to the nuclei being probed. It is therefore absolutely essential to have the framework that X-ray crystallography provides. With the information these highly complementary techniques provide, however, NMR data does not always lend itself to straightforward interpretation.[13, 17]

Enzymes create a unique microenvironment and it is not unusual to observe shifts drift from canonical ranges observed for smaller molecules in model compounds.[5] As theoretical methods, such as first-principles and molecular dynamics calculations, continuously advance they are becoming essential components of chemical analysis across all disciplines of chemistry.[13, 18, 19] With input from experimental observations, models of chemical systems can be constructed and systematically varied to test their effects on calculated observables, such as the nuclear magnetic shielding that gives rise to the NMR chemical shift.[13, 19] These can in-turn be compared with experimental observations allowing for vast improvements in the refinement of models by selecting those that are consistent with experimental observations while rejecting those that are not.

Today, the term NMR crystallography has come to refer to the synergistic combination of diffraction, NMR, and theoretical techniques and is being applied to increasingly challenging systems.[13, 18, 20-24] In this chapter, efforts to characterize an important quasi-stable intermediate, the 2-aminophenol (2AP) quinonoid, $E(Q_3)_{2AP}$, in the PLP-dependent enzyme TRPS by adapting the NMR crystallography approach specifically to the enzyme active site are provided. Aiding our efforts, a refined approach to cluster-based calculations of the enzyme active site and linear rescaling parameters to reference calculated magnetic shielding values to experimental chemical shifts specific to the level of theory and basis sets employed in the calculations were presented recently by Hartman and coworkers.[25] The resulting structural refinement of this intermediate by the NMR crystallographic approach has provided unprecedented detail including the identification of competing tautomers existing at equilibrium which have important mechanistic consequences.

As explained in the first chapter, the formation of quinonoid intermediates were long thought to be an essential step in all PLP mechanisms yet evidence has been accumulating that challenges this assumption.[1-4, 26] Recent investigations indicate that the formation of a quinonoid intermediate is important to some types of reactions catalyzed by PLP enzymes, especially in those that involve transamination events, whereas in others its appearance could be detrimental and result in undesirable side reactions.[1, 27, 28] The fact that several substrate and metabolite analogues have been identified in TRPS that result in the accumulation of a quinonoid intermediate, primarily

based on its UV/Vis absorption maximum near 470 nm, has made it an important target for closer study.[29]

Details regarding the catalytic mechanism of TRPS were presented in Chapter 1 and included the introduction of quasi-stable intermediates in the β-site reaction through the use of structural analogues of metabolites and substrates.[13, 29] In Chapter 2, the preparation of such intermediates for solution state ^{17}O QCT NMR was given.[17, 29] The $E(Q_3)$ intermediate can be stabilized with a variety of indole analogues, however, 2AP is particularly desirable because it remains stable for several days.[17, 29] This is important for the measurement of ^{15}N, ^{31}P, and ^{13}C chemical shift anisotropy (CSA) tensors using solid state magic angle spinning NMR (SS-MAS-NMR) which requires longer experiment times compared with isotropic chemical shift measurements. SS-MAS-NMR measurements for ^{15}N, ^{31}P, and ^{13}C nuclei were performed by other Mueller group members, not the author, therefore, only the results of their work will be presented and not the experimental details, some of which have already been published, see references 13, 30 and 31. [13, 30, 31] The chemical structure of the PLP/substrate complex as well as a nearby catalytic residue, βLys87, are shown in Figure 3.1 to highlight the positions enriched in ^{13}C, ^{15}N, and ^{17}O that were measured using various NMR methods.[13, 17, 30, 31] Enrichment for ^{31}P is not required given its high natural abundance (100%).

To recap important features of the $E(Q_3)$ intermediate, a portion of its catalytic cycle is provided in Figure 3.2, which also includes the intermediates preceding and following its formation, namely the aminoacrylate, E(A-A), and the 2nd external aldimine, $E(Aex_2)$.[32] In TRPS, indole generated at the α-subunit active site makes a nucleophilic

attack on the serine derived C^{β} in E(A-A) to yield the electron rich $E(Q_3)$.[31] The transition state leading to the next observable intermediate, $E(Aex_2)$, requires protonation at the serine derived C^{α} carbon, which previous studies indicate is most likely provided by βLYS87 that functions as an acid-base catalyst throughout TRPS's β-site mechanism.[30, 31]

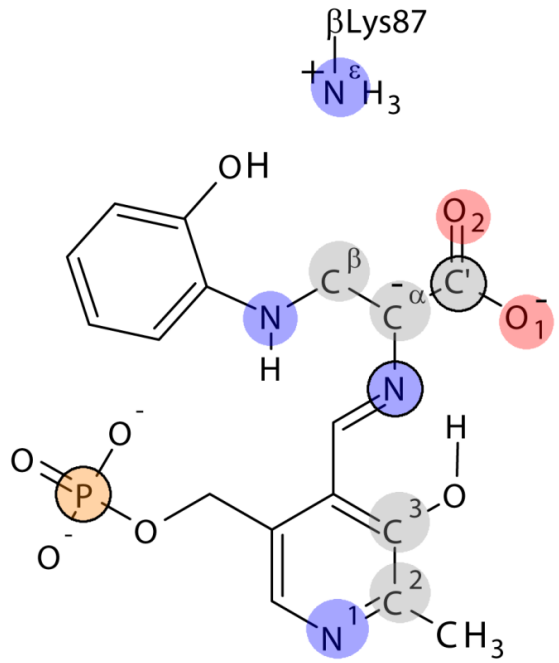


Figure 3.1. NMR sites experimentally measured for the PLP/substrate complex and βlys87 N^{ε} residue in the β-active site of the TRPS $E(Q_3)_{2AP}$ intermediate. Nitrogen, carbon, and oxygen positions were isotopically enriched (^{15}N, ^{13}C, and ^{17}O). CSA tensor measurements were performed on highlighted nuclei circled in black, the L-serine derived carbonyl, C', and Schiff base nitrogen, N, and the PLP phosphoryl, P.

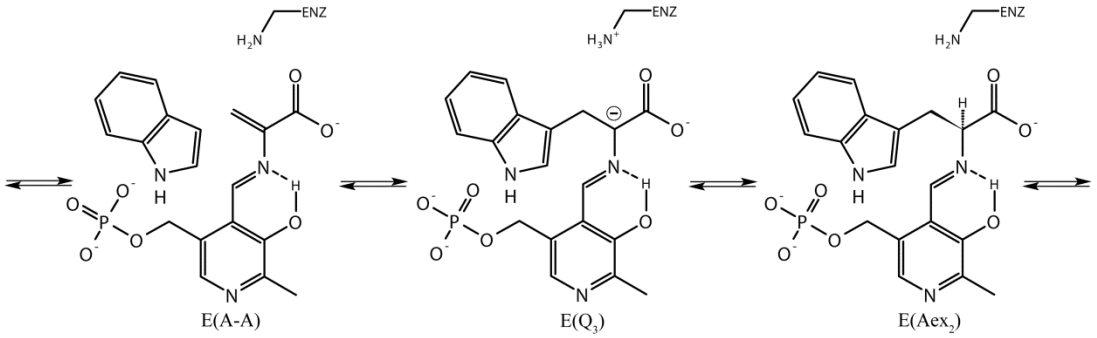


Figure 3.2. Three sequential intermediates along the β-site reaction coordinate, the aminoacrylate, E(A-A), the third quinonoid intermediate with the natural substrate indole, $E(Q_3)$, and the second external aldimine, $E(Aex_2)$.

The remainder of the introduction will focus on four topics. First, the refined approach to cluster based calculations of the active site and linear rescaling for the referencing of chemical shieldings conducted by Hartman and coworkers will be elaborated upon. Next, the conversion of calculated nuclear magnetic shieldings to chemical shifts to facilitate their comparison with corresponding experimental measurements will be presented. Third, the experimental NMR contributions of group members are recognized along with a discussion of important results. Finally, the reduced χ^2 analysis method by which testable models are ranked according to their agreement with experimental results is presented.

3.1.2 Cluster Approach to Modeling the Enzyme Active Site

The ability to calculate NMR chemical shielding using first-principles methods is becoming increasingly more reliable and enhancing our ability to interpret experimentally acquired chemical shifts and other NMR parameters.[25, 33, 34] Despite more efficient algorithms and widespread access to supercomputing resources, calculations on systems the size of biological enzymes remains out of reach. Fortunately, nuclear magnetic

shielding is principally effected most by local interactions with the immediate chemical environment thus allowing for the use of cluster approaches in which the system is truncated to include the components that are of paramount interest, which are those pertaining to experimentally acquired shift data, and enough of the surrounding structure to ensure that all important interactions remain present.[25] The challenge in calculating NMR parameters in large systems lies in determining just how large the cluster needs to be.[25] As the sizes of clusters increase compromises often have to be made in basis set selection, and hence accuracy, to keep the calculations reasonable in terms of computational cost (time).[25]

Hartman and coworkers have performed a systematic approach to find the optimum balance of cluster size and basis set selection in tackling NMR calculations in proteins using the convergence of calculated chemical shielding for specific theoretical models as the primary criterion for performance.[25] Their findings culminated in general recommendations when approaching cluster calculations of chemical shielding in biological systems and those directly applicable to this work are given.[25] First, the cluster needs to extend approximately 7Å from the nuclei of interest. For our study this represents the PLP/substrate complex. Second, it is not necessary to employ a single large basis set for the calculation of all of the atoms in the cluster.[25] A multi-tier approach in which the region of interest, again the PLP/substrate complex, is afforded a large triple-zeta basis set with diffuse and polarization functions on the atoms of interest, e.g. 6-311+G(2d,p).[25] This represents the "high layer". For atoms closely interacting with the region of interest out to 4Å, a triple-zeta basis set without costly diffuse functions, yet

keeping polarization terms, can be assigned, e.g. 6-311G(d,p), representing the "medium" layer. Finally, all remaining atoms, the "low layer", can be assigned an inexpensive double-zeta basis set, e.g. 6-31G.[25] Using these guidelines, NMR calculations on systems having 600 -700 atoms can be routinely accomplished in 1-2 days when run on computer clusters with 32 processors and 48 GB of RAM available in parallel. Also, the mixed basis set approach can be applied to perform a full QM geometry optimization of the cofactor within the active site environment to include intermolecular hydrogen bond partners and charges. In prior studies a combined QM/MM approach (ONIOM) was used in which the PLP/substrate complex atoms were treated at the QM level and all other atoms were computed at the semi-empirical PM3 level of theory.[13] Also, in the past only the PLP/substrate atoms and all hydrogens were permitted to relax their coordinates.[13] In the latest scheme, based on discussions with J. D. Hartman, a better approach is to not only allow the PLP/substrate atoms freedom of movement but also the coordinates of those atoms comprising the "medium layer" (atoms 4Å from the PLP/substrate complex) to capture important intermolecular hydrogen bonding interactions and, as always, all hydrogen atoms.

3.*1.3 NMR Observables: Relationship Between Theory and Experiment*

The output generated from first-principles calculations of NMR observables are absolute magnetic shielding values experienced at the nuclear site, σ (ppm), and is provided as a rank two (3 × 3) shielding tensor, $\boldsymbol{\sigma}$ (tensors are indicated in bold).[35] Experimental NMR measurements measure the isotropic chemical shift, δ_{iso}, and, in the solid state, the principal components of the CSA shift tensor, δ_{11}, δ_{22}, and δ_{33} are

additionally available. The δ_{iso} = (1/3)(δ_{11}+ δ_{22}+δ_{33}) or 1/3Tr[**δ]**. Recall that the isotropic chemical shift is defined as (equation (1))

$$\delta_{iso}(ppm) = \frac{\nu_i(MHz) - \nu_{ref(}(MHz)}{\nu_{ref(}(MHz)} \times 10^6 \quad (1)$$

where $\nu_i = \frac{-\gamma B_0}{2\pi}\left(1 + \delta_{iso,i}\right)$ and $\nu_{ref} = \frac{-\gamma B_0}{2\pi}\left(1 + \delta_{iso,ref}\right)$. The chemical shift, δ_i, and shielding, σ_i, are related by $\delta_i = -\sigma_i$.[36] To compare experimental results with those calculated from first-principles methods the absolute shielding values need to be converted to referenced chemical shifts. Hartman et al. performed calculations on test-sets of crystal structures to determine linear-rescaling parameters that map computed shieldings to referenced chemical shifts specific to the theoretical model and basis sets used.[25, 34] The conversion to isotropic shift values is $\delta_{iso} = A\sigma_{iso} + B$, where A and B are determined from the best fit slope, A, and intercept, B, of computed shieldings versus referenced chemical shifts.[25, 34] Similarly, conversion of rank-two CSA shielding tensors, **σ**, to corresponding shift tensors, **δ**, can be accomplished by $\boldsymbol{\delta} = A\boldsymbol{\sigma} + \mathbb{1}B$, where $\mathbb{1}$ represents the identity matrix. The linear rescaling parameters for the level of theory, DFT with B3LYP, and basis set used to compute NMR shieldings, 6-311+G(2d,p) for ^{13}C, ^{15}N, and ^{17}O are provided in Table 1 along with their corresponding calculated root mean square errors (RMSE) from the fitting procedure. Note, RMSE values are different for δ_{iso} and the individual principal components of the CSA tensor, δ_{11}, δ_{22}, and δ_{33}.[34]

Table 3.1. Linear rescaling parameters mapping calculated chemical shielding values (B3LYP/6-311+G(2d,p) to referenced chemical shift values for the isotopic species indicated. Chemical shift references: $^{15}NH_3$(l),^{13}C TMS, ^{17}O $H_2{}^{17}O$(l), 298K. For ^{13}C see reference 13. ^{15}N and ^{17}O see references 30 and 17 respectively.

Atom	A	B (ppm)	δ_{iso} RMSE (ppm)	CSA RMSE (ppm)
^{15}N	-0.9986	230.57	4.82	13.72
^{13}C	-0.9681	173.63	1.49	4.24
^{17}O	-1.0612	265.79	8.25	NA

Calculating isotropic chemical shifts and principal components from the shielding tensor is straightforward. The shielding tensor can be converted to a shift tensor and 1/3(Tr[$\boldsymbol{\delta}$]) yields the isotropic shift and diagonalization gives the three principal components, in the chemical shift convention based on Mehring notation where $\delta_{11} \geq \delta_{22} \geq \delta_{33}$.[35] To model chemical exchange, e.g. the transfer of protons between ionizable sites, the resulting shielding values from two or more structures are weighted and combined, details are in section *3.3.5*, to yield an exchange averaged result. Taking a weighted average of the isotropic chemical shifts, δ_{iso}, does not present any problems, however, when combining CSA tensors there are a few more steps. First, the rank two CSA tensor calculated is asymmetric and must be symmetrized (before or after referencing) because it contains both symmetric and antisymmetric components the latter of which are not observed experimentally.[37] Extracting the symmetric CSA tensor is accomplished by Equation (2)

$$\boldsymbol{\sigma}_{sym} = \frac{1}{2}\left(\boldsymbol{\sigma}_{asym} + \boldsymbol{\sigma}_{asym}^{T}\right) \tag{2}$$

where $\boldsymbol{\sigma}_{asym}^{T}$ represents the transpose of the asymmetric CSA tensor.[37] Weighted sums of tensors must be computed prior to diagonalization to obtain the eigenvalues representing the principal components of the mixed tensor. This has to do with the fact the principal components relate to the principal axis system in the molecular frame which is not likely to be the same for each individual structure's CSA tensor.

For quadrupolar nuclei, such as ^{17}O, the components of the traceless electric field gradient tensor (EFG) can be calculated in programs such as Gaussian09.[38] Harbison has determined that current basis sets are not equipped to compute accurate EFG parameters because of their inability to capture the distortion of the 1s core orbital in the field of the nuclear quadrupole.[39] As a result, calculations of EFG parameters are systematically overestimated by 10-15%. [39] For purposes of comparison, the quadrupolar coupling constant, C_Q, and asymmetry parameter, η_Q, will be computed from the results of the first-principles calculations and compared to the quadrupole product parameters, P_Q, extracted from the ^{17}O QCT NMR data presented in Chapter 2.[17]

From Gaussian09 output the 2nd rank EFG tensor, $\boldsymbol{V}$, is constructed from the symmetric off-diagonal components provided as XY, XZ, and YZ, and the diagonal elements from the EFG tensor representation 3ii – RR components, where ii = XX, YY, ZZ, see equation (3).[37] In exchange models mixed tensors are weighted and summed prior to diagonalization to yield the EFG principal components$|V_{33}| \geq |V_{22}| \geq |V_{11}|$, in the EFG principle axis system (PAS).[37] V_{33}, the largest tensor component appears in the C_Q, see equation (4).

$$\boldsymbol{V}_{calc} = \begin{bmatrix} 3XX - RR & XY & XZ \\ XY & 3YY - RR & YZ \\ XZ & YZ & 3ZZ - RR \end{bmatrix} \quad (3)$$

$$C_Q(MHz) = \frac{eQV_{33}}{h}, V_{33} = eq_{zz} \quad (4)$$

To convert the Gaussian09 output V_{33} given in atomic units (a.u.) to MHz the conversion factor given by Adiga et al. is employed: C_Q(MHz) = -6.0104 MHz a.u.$^{-1}$(V_{33}(a.u.)).[37] The quadrupole product parameter, P_Q, is given by equation (5) where η_Q is the asymmetry parameter.[40]

$$P_Q(MHz) = \left(\frac{eQV_{33}}{h}\right)\sqrt{1 + \frac{\eta_Q^2}{3}}, \qquad \eta_Q = \frac{V_{22} - V_{11}}{V_{33}}, 0 \leq \eta_Q, \leq 1 \quad (5)$$

3.1.4 Experimental NMR

The NMR crystallographic approach is enhanced by the number of active site probes that can be measured. At 143 kDa, ^{1}H NMR measurements are largely precluded in solution state experiments and are not yet feasible in solid state NMR. Chapter 2 presented the means by which ^{17}O NMR measurements were possible in the solution state given the special properties of quadrupolar nuclei in the limit of slow isotropic motion.[17] Solid-state magic angle spinning NMR, SS-MAS-NMR, provides access to critical ^{15}N, ^{13}C, and ^{31}P data at the active site and can even provide anisotropic data through the determination of chemical shift anisotropy (CSA) tensors which are not available from solution state NMR. This section acknowledges the contributions by group members B. G. Caulkins, R. A. Kudla, J. Lai, T. J. Neubauer, and C. Yang in obtaining ^{15}N, ^{13}C, and ^{31}P chemical shift and CSA tensor data used in conjunction with the X-ray crystallographic data and first-principles calculations used to characterize the $E(Q_3)_{2AP}$

and other TRPS intermediates. Implications of ^{31}P NMR tensor and ^{15}N chemical shift temperature dependence experiments will also be discussed.

Given the low natural abundance of ^{13}C and ^{15}N nuclei, 1.1% and 0.36% respectively, isotopically enriched L-serine, 2-aminophenol, and PLP were required. Uniformly labeled [^{13}C, ^{15}N]-L-serine is commercially available. Group member B. G. Caulkins supplied commercially available [N$^{\varepsilon}$,^{15}N]L-lysine in select protein expressions to measure the chemical shift of the βLys87 residue sidechain epsilon-nitrogen, N$^{\varepsilon}$, responsible for making the Schiff base linkage in the E(Ain) and as a general acid-base catalyst in the remainder of the TRPS catalytic cycle (see reference 31 for details).[31] 2-aminophenol (2AP) was enriched with ^{15}N by Dr. B. Bastin. Synthesis of 2,2',3-^{13}C;^{15}N-PLP was also performed by Dr. B. Bastin and details are provided in reference 30.[30] Figure 1, presented earlier, shows the PLP/substrate complex of the E(Q$_3$)$_{2AP}$ and βlys87 highlighting the ^{13}C, ^{15}N, and ^{17}O enriched positions, ^{17}O was the subject of Chapter 2. ^{31}P does not require enrichment. The highlighted nuclei in Figure 1, that are also outlined in black, the C', Schiff-base N (SB N), and phosphoryl P, represent nuclei for which CSA tensor measurements were also made.

^{13}C and ^{15}N data collected on enriched L-serine and indoline in the E(Q$_3$)$_{indoline}$ intermediate was performed by J. Lai as reported in reference 13. All other ^{13}C and ^{15}N NMR data for the E(Q$_3$)$_{2AP}$, E(Q$_3$)$_{indoline}$, and E(A-A) intermediates appearing herein, see results, was measured by B. G. Caulkins using SS-MAS-NMR methods. In addition to measuring the isotropic chemical shifts of L-serine at C', C$^{\alpha}$, C$^{\beta}$, and N (Schiff base nitrogen, SB N), 2AP at N, βlys87 at N$^{\varepsilon}$, and PLP at N1, C2, C2', and C3, the CSA

tensors were measured for the L-serine derived SB N and C' positions. B. G. Caulkins also performed temperature dependence measurements at the N1 of PLP and the L-serine derived SB N. The N1 position did not show any indication of chemical exchange, however the SB N isotropic chemical shift and linewidth did. Analysis of the temperature dependence of the linewidth and isotropic chemical shift, assuming two-site fast-exchange equilibrium, predicted a protonated SB N as a minor species at approximately 21%. The energy difference between the two states was determined to be approximately 0.7 kcal/mol with an activation free energy, $\Delta G^{\ddagger}$, of approximately 9.7 kcal/mol

Measurement of the PLP ^{31}P isotropic chemical shifts were initially carried out by T. J. Neubauer using SS-MAS-NMR, the results of which pointed to a dianionic phosphoryl group. For further verification, R. A. Kudla performed detailed ^{31}P CSA tensor measurements as they represent the most confident indication of ionization state as the ^{31}P δ_{22} changes from positive to negative values, indicating deshielding, and the δ_{33} component becomes more shielded by > 20 ppm when a monoanionic phosphoryl is deprotonated to yield the dianionoic state.[41, 42] Confirmation of the dianionic state of the phosphoryl group significantly reduced the number of models that had to be constructed and tested by a factor of at least 2.

3.1.5 Reduced Chi Square Analysis: Testing Models vs. Experiment

Reduced χ^2 analysis was used to quantify the agreement between the calculated chemical shifts of the candidate structures to those experimentally measured. The reduced χ^2 statistic provides a numerical value indicating the weighted deviation of the calculated

chemical shifts of the model structure to the experimental chemical shifts as shown equation (6).

$$\chi^2_{CS} = \frac{1}{N}\sum\left(\frac{\left(\delta_{i,calc} - \delta_{i,exp}\right)}{\sigma^{CS}_{i,rmse}}\right)^2 \quad (6)$$

Here N refers to the number of chemical shifts measured (N = 11 for $E(Q_3)_{2AP}$), $\delta_{i,calc}$ (ppm), is the calculated chemical shift at site *i*, $\delta_{i,exp}$ (ppm), is the experimentally measured shift at site *i*, and the $\sigma_{i,rmse}$ value is the root-mean-squared-errors (RMSE) of the particular nuclear species, N, C, or O, at site *i* listed in Table 3.1. Low reduced χ^2 values indicate better agreement between calculated and experimental chemical shifts. To model proton chemical exchange, a two-site model was evaluated according to equation (7).

$$\chi^2_{CS} = \frac{1}{N}\sum\left(\frac{\left((a\delta_{i,calc} - (1-a)\delta_{j,calc}) - \delta_{exp}\right)}{\sigma^{CS}_{rmse}}\right)^2 \quad (7)$$

Here $0 \leq a \leq 1$ and the subscripts *i* and *j* refer to the same nuclear site in different candidate structures. Every pair of candidate structures were evaluated and weighted to give the lowest reduced χ^2 value based on the δ_{iso} chemical shifts (CS). Similarly, a three-site exchange model was also evaluated. Reduced χ^2 ranking of CSA tensor components were calculated in a similar manner for single sites, equation (8), and in two-site exchange model, equation (9), where N = 3, representing the three principal components δ_{11}, δ_{22}, and δ_{33}. Selection of lowest reduced χ^2 in exchange models were based on the isotropic chemical shifts and the resulting contributions were then used to compute and report the reduced χ^2 for the CSA tensors.

$$\chi^2_{CSA} = \frac{1}{N}\sum\left(\frac{\left(\delta_{ii}^{calc} - \delta_{ii}^{exp}\right)}{\sigma_{rmse}^{CSA}}\right)^2 \quad (8)$$

$$\chi^2_{CSA} = \frac{1}{N}\sum\left(\frac{\left(a\delta_{ii,1}^{calc} + (1-a)\delta_{ii,2}^{calc}\right) - \delta_{ii}^{exp}\right)}{\sigma_{rmse}^{CSA}}\right)^2 \quad (9)$$

3.2 Experimental

3.2.1 Enzyme Cuts

Extraction of the enzyme active site cluster for the $E(Q_3)_{2AP}$ proceeded as follows and, unless specifically indicated, was accomplished using the free version of PyMOL.[43] The crystal structure of $E(Q_3)_{2AP}$ of *S. Typhimurium*, PDBID 4HPJ, was obtained from the RCSB Protein Data Bank.[12] Only the highest population conformers were kept in cases where alternates were present. The PLP/substrate complex, comprised of the PLP cofactor, serine derived atoms, and 2AP were selected. In the PDB file, this complex is assigned the residue name 1D0. All atoms within 7 Å of 1D0 were selected and then expanded to include all remaining atoms in their complete amino acid residues. This naturally extended the cluster size beyond 7 Å. All other atoms were deleted. To reduce the total number of atoms while keeping intact continuous segments of backbone and important sidechains at the periphery of the cluster, the structure was further modified. Residues that were not part of continuous backbone segments and having only one atom within 7 Å were deleted (β-site residues Val117, Met134, Glu142, Lys167, Tyr186, Met187, Val201, Ile238, Ser308, and Asn375). Residues with two backbone atoms within 7 Å were converted to alanine or glycine by removing the sidechains lying out of range (β-site residues Leu349, Leu376, and Arg379). Finally, residues having two sidechain

atoms within 7 Å were truncated by removing the outlying backbone atoms (β-site residues Phe280, and His313) thus completing the initial cut.

Hydrogen atoms were then added to the structure using a built-in algorithm in the PyMOL program.[43] Hydrogen atoms added to the C-termini carbonyl carbons of backbone segments were replaced with $–NH_2$ groups to yield amides at these positions. The resulting $[NH_3]^+$ N-termini of backbone segments were replaced with a single hydrogen atom. In all cases, histidine (His) sidechain moieties were only singly protonated (neutral) by the PyMOL program. Referencing a recent molecular dynamics study by Huang et al, 2016, in which pK_a calculations were performed using the MCCE program, and by additional simulations involving MMPBSA, performed by Wanli You, all His residues in the 7 Å cut, βHis86, βHis115, and βHis313 (only the sidechain of βHis313 is present in the structure) were singly protonated on the ring N^{ε} in each.[44, 45] All remaining fragments in the cluster were closely inspected for proper addition of hydrogen atoms and the formal charge for the complete starting structure was determined.

The active site cut was saved as a PDB file and opened in GaussView5 (GV5).[46] The entire structure was inspected closely again and dihedral angles involving hydrogen atoms bonded to oxygen or nitrogen atoms were adjusted to angles suitable for hydrogen bonding with the nearest non-bonded N or O atom if present. The protonation state of the PLP/substrate complex corresponded to that depicted in Figure 1, that is protonated βLys87 N^{ε}, dianionic phosphoryl, deprotonated pyridine ring nitrogen, neutral phenolic oxygen (as opposed to phenolate), deprotonated Schiff-base nitrogen, and an anionic

carboxylate group. Finally, the Cs^+ atom was replaced with a Na^+ atom to maintain the positive charge while obviating the consideration of relativistic effects. This starting structure was then subjected to geometry optimization as described in the next section.

3.2.2 Geometry Optimization

The initial geometry optimization involved freezing all coordinates except the added N atoms and all H atoms. All calculations were performed using the Gaussian09 program and employed the DFT level of theory with the hybrid B3LYP functional.[38, 47-50] The double-zeta 6-31G basis set was assigned to all atoms.[38] After this initial minimization, various protonation states were constructed by adding, removing, and/or moving hydrogen atoms amongst the seven ionizable sites on the PLP/substrate complex and the nearby βLys87 N^{ε}. A binary code scheme with seven places is used to identify the individual protonation state configurations. Figure 3.3 depicts the PLP/substrate and nearby βLys87 residue along with letter codes indicating the place assignments in the code. A value of 1 indicates that the ionizable site contains a proton and is 0 if it does not.

Thirty-two configurations in total were prepared and thirty of these are depicted in Figure 3.4 with their corresponding binary codes. The configurations not shown correspond to the 10-0-00-10-gln and the 10-0-01-00-gln in which the "gln" modifier indicates that, in these structures, the sidechain of βGln114 was positioned to form a hydrogen bond with the phenolic oxygen. This configuration was suggested by molecular dynamics simulations performed by Huang et al. in which the βGln114 sidechain moved from its crystallographic position to form a hydrogen bond with the phenolic oxygen when the SB N or nearest carboxylic group oxygen, O1, were protonated.[44] This is

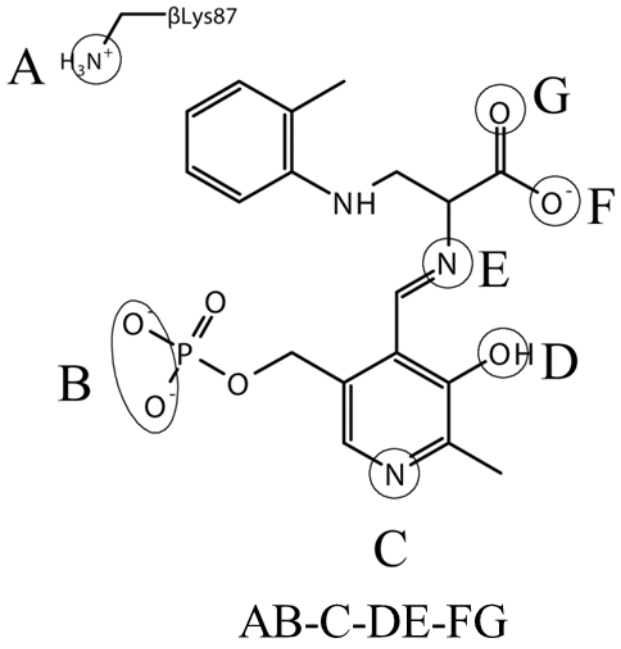


Figure 3.3. Binary code system for identifying protonation state configurations for candidate structures constructed for chemical shift calculations. For each position, A - G, a value of 1 indicates that the site is protonated and a value of 0 indicates that it is unprotonated. For example, the structure depicted above would be described by 10-0-10-00.

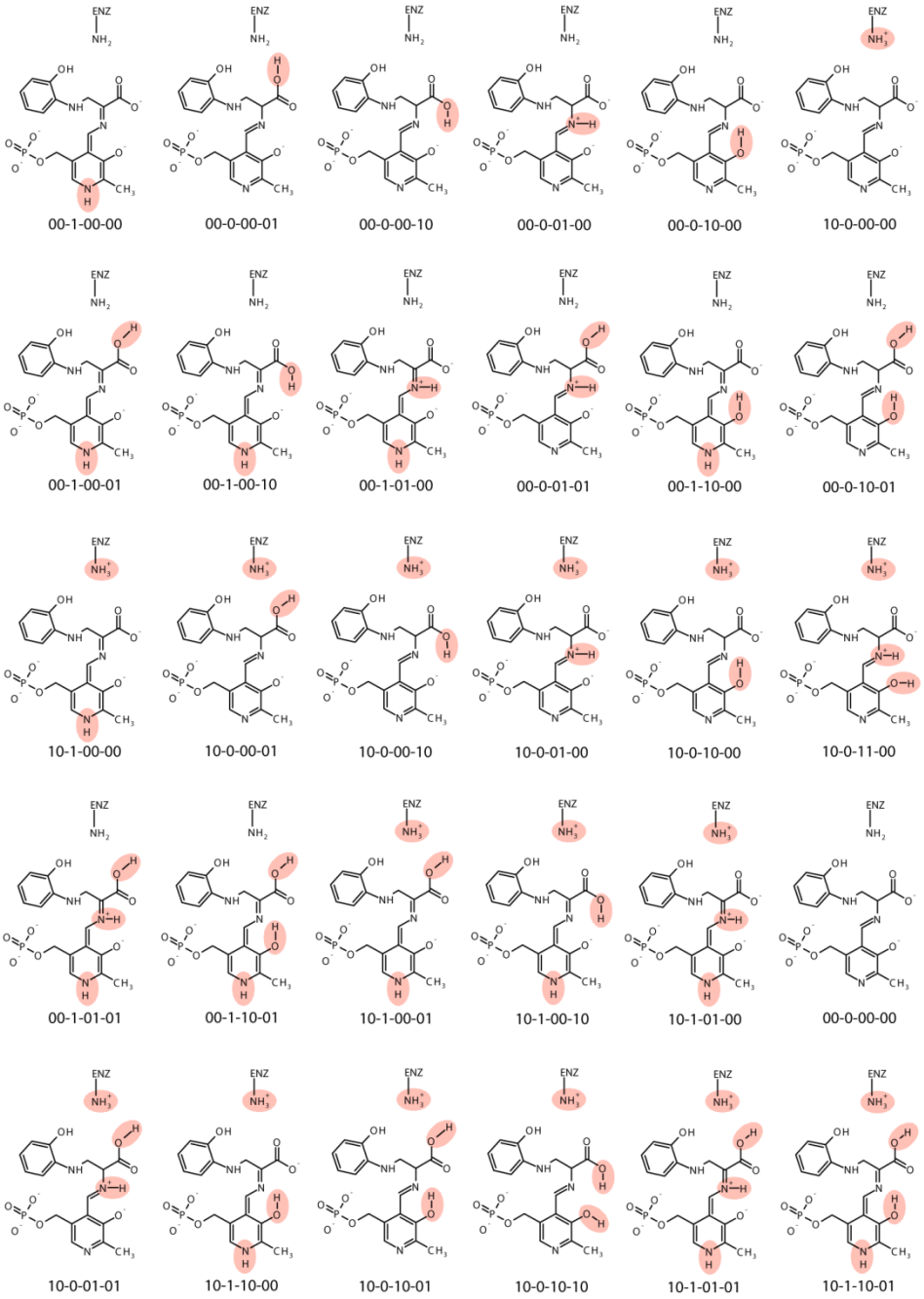


Figure 3.4. Protonation states of 30 of 32 candidate structures constructed for NMR shielding calculations. The structures not shown correspond to the 10-0-01-00-gln and 10-00-10-gln, which differ from the candidate structures having the same binary code in that the sidechain amide group of βGln114 was positioned to form a hydrogen bond with the phenolic oxygen site as observed in molecular dynamics simulations performed by Huang et al.[44]

important to model in these studies as the protonated phenolic oxygen, enolimine configuration, is shown to be more stable when the pyridine nitrogen, PN, is deprotonated in model compound studies.[51] A hydrogen bonding interaction involving a nearby residue could presumably render the phenolic proton more labile thus promoting transfer to the Schiff-base or nearby carboxylic oxygen, O1. Overlays of the 10-0-00-10/10-0-00-10-gln and 10-0-01-00/10-0-01-00-gln structures are shown in Figure 3.5. The absence of structures protonated at the phosphoryl group, position two in the binary code, is due to ^{31}P CSA tensor measurements confirming that was indeed in the dianionic state (measurement by group member R. A. Kudla). The phosphoryl position in the binary code is a vestige of prior work.

Configurations having a protonated pyridine ring nitrogen, N1, and/or a protonated upper carboxylic oxygen atom, O2, which are those with a 1 in the third and/or last position in the seven-place binary code, had to be modified by adjusting the dihedral angles of nearby sidechains to prevent steric clash of hydrogen atoms with nearby enzyme sidechains previously involved in hydrogen bonding to these two atoms. These were the βSer377 sidechain hydroxyl group proton near the N1 pyridine ring nitrogen and the βThr110 sidechain hydroxyl proton near the O2 carboxylic oxygen specifically. Each of the 32 structures was then subjected to further geometry optimization/minimization using a two-tier locally dense basis set assignment in which the atoms of the PLP/substrate complex were calculated with the 6-31G(d,p) basis set and all remaining atoms with the 6-31G basis set. Additionally, all atoms within the PLP/substrate complex and all atoms within 4Å of any atom of the PLP/substrate

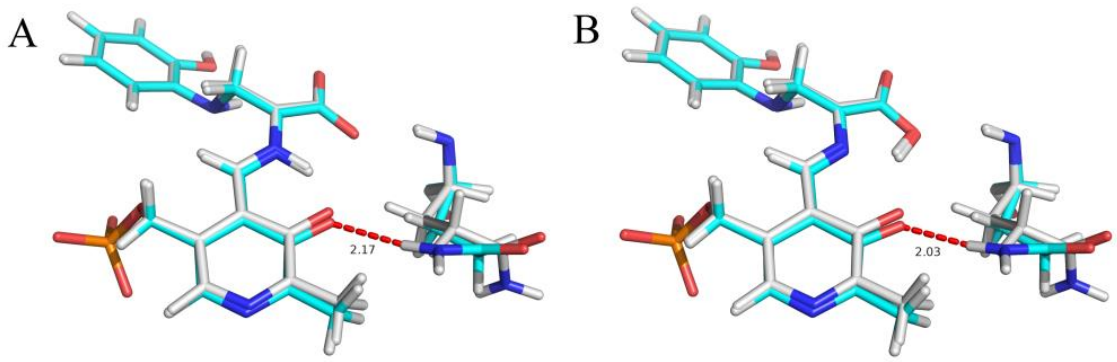


Figure 3.5. Overlays of the 10-0-01-00 (white) and 10-0-01-00-gln (cyan) (A) and 10-0-00-10 (white) and 10-0-00-10-gln (cyan) (B) candidate structures. Hydrogen bond distances of the βGln114 amide sidechain NH to the phenolate oxygen of PLP are indicated by the red dashed lines and the values correspond to the H - - O^- distance in Angstroms (Å).

complex were free to adjust and all remaining coordinates were frozen. Selection of the various layers from the Gaussian input files, coding the coordinates for fixed/free, and generation of the locally dense basis set assignments was accomplished in a custom program written in Mathematica® by Prof. Leonard Mueller. An illustrative example of the enzyme active site cluster that highlights the various layers including mobile and frozen atoms as well as high, medium, and low layers described in the following section is shown in Figure 3.6.

3.2.3 NMR Calculations

NMR shielding calculations were performed on the optimized structures using the Gaussian09 program and the gauge-independent atomic orbital (GIAO) method at the DFT level of theory with the B3LYP hybrid functional.[52, 53] A three-tier locally dense basis set assignment was employed.[25] The high layer, consisting of the PLP/substrate complex, was assigned the 6-311+G(2d,p) basis set, the medium layer, comprised of any

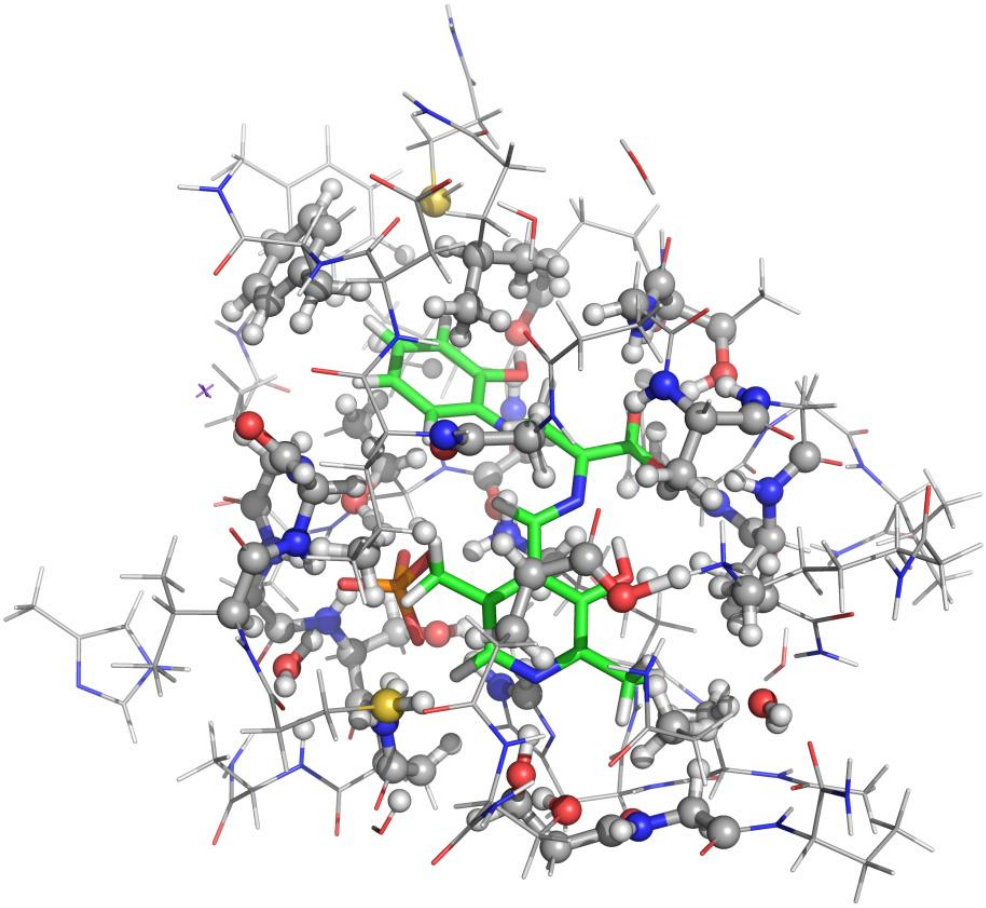


Figure 3.6. 7Å active site cluster extracted from the $E(Q_3)_{2AP}$ X-ray crystal structure, PDBID 4HPJ. The high layer, PLP/substrate complex is depicted in green (carbon) licorice form, the medium layer is represented in gray (carbon) ball and stick representation, and the low layer is depicted in gray (carbon) line representation.

atom within 4Å of any PLP/substrate atom, were assigned the 6-311G(d,p) basis set, and all remaining atoms (low layer) assigned the 6-31G basis set. See Figure 3.6, again, for a visual depiction of the three tier basis set assignment scheme.

3.2.4 First Principles Calculations of Other TRPS Intermediates

First principles calculations were completed on a limited number of protonation state configurations for the $E(Q_3)_{indoline}$ and E(A-A) for comparison with the calculations of the $E(Q_3)_{2AP}$. The procedures used were identical to those listed for the $E(Q_3)_{2AP}$. Differences in the structures of these intermediates as a result of different indole analogues led to slight differences in the residues deleted or modified.

3.2.5 Reduced χ^2 Analysis of Isotropic Chemical Shift

Reduced χ^2 calculations were performed using Mathematica®. A reduced χ^2value was calculated for each of the 32 structures with respect to agreement of experimental δ_{iso} shifts, 11 in total, and CSA tensors for the C' and SB N atoms. Two-site exchange models were calculated by taking every pair of single site structures and finding the best weighted combination that produced the lowest reduced χ^2 value based on the 11 δ_{iso} shifts. The C' and SB N CSA tensors for each pair were mixed accordingly and their corresponding reduced χ^2 values reported with each two-site combination. Similarly, a three-site exchange model was also evaluated.

3.3 Results

3.3.1 Single Site Models

Table 3.2 lists the calculated δ_{iso} shifts from all 32 configurations modeled along with their corresponding reduced χ^2 values. The reduced χ^2 values are also shown graphically in as a column chart in Figure 3.7. Only two configurations had reduced χ^2 values less than 24. The lowest value, 3.33, corresponded to the 10-0-10-00 structure,

which will be referred to as PO, representing a protonated phenolic oxygen and βLys87 N^{ε}. The next lowest χ^2 value, 8.76, pertained to the 00-0-10-00 configuration which differs from the PO only in the protonation state of βLys87 N^{ε}. The remaining configurations had reduced χ^2 values ranging from 24.85 to 101.55. The reduced χ^2 values for the C' and the SB N CSA tensors of the best single structure based on the 11 isotropic shifts, the PO, were 1.47 and 7.91 respectively.

Table 3.2. Calculated isotropic chemical shifts, δ_{iso}, in parts per million, ppm, ranked by reduced χ^2 value for all 32 candidate structures.

Structure	N1	C2	C3	SB N	C^{α}	C'	C^{β}	O1	O2	N^{ε}	2AP N	Red. χ^2
10-0-10-00	257.0	141.0	150.1	320.5	107.8	172.8	47.3	238.9	233.6	26.9	54.1	3.33
00-0-10-00	253.0	136.0	150.7	325.1	109.5	172.4	46.5	235.7	227.9	12.7	53.2	8.76
10-0-10-01	282.4	144.7	151.3	310.7	100.6	162.7	45.5	214.1	128.6	26.8	51.6	24.85
10-0-00-00	251.8	144.8	163.4	369.3	104.1	172.4	46.6	227.2	216.2	26.9	60.0	25.32
00-0-10-01	280.2	144.2	152.5	312.1	102.1	161.7	48.5	207.6	129.9	19.1	44.5	25.98
10-0-10-10	278.5	140.8	146.0	329.7	104.8	165.4	43.3	110.0	200.4	27.6	51.3	34.17
00-0-00-00	256.1	144.8	165.8	376.3	115.9	172.3	45.8	237.6	224.9	21.8	56.8	35.83
10-0-11-00	273.0	139.2	142.8	212.9	106.9	169.4	44.9	232.8	252.9	27.6	49.0	36.14
10-0-00-01	267.7	149.9	164.6	334.3	102.7	160.5	46.1	190.0	125.6	28.3	54.4	38.64
10-0-00-10	270.5	149.2	164.9	302.1	98.3	164.2	44.2	110.0	167.6	27.4	53.8	40.75
00-0-00-01	267.2	147.9	165.0	340.8	104.1	159.3	47.5	185.1	127.0	18.0	55.5	42.27
00-0-00-10	270.8	147.6	165.3	306.6	100.2	163.1	43.8	106.9	161.2	15.4	53.6	44.29
10-0-00-10-gln	275.9	153.0	165.5	304.0	98.9	164.2	42.7	108.8	164.6	28.1	54.3	45.04
00-0-01-00	271.5	145.1	164.4	204.7	94.4	170.1	46.1	225.1	218.6	13.3	51.4	47.26
10-1-10-01	184.0	134.2	151.0	320.8	107.7	164.4	44.3	239.0	135.9	26.7	50.6	48.42
10-0-01-00	272.0	149.4	164.1	202.2	93.3	170.1	47.4	229.5	224.6	27.8	52.2	49.54
10-0-01-00-gln	275.7	149.8	164.8	201.2	93.6	169.9	46.1	227.3	220.0	27.6	53.1	51.13
00-1-10-01	181.4	132.1	151.7	320.7	109.8	164.2	47.3	232.0	137.2	29.8	43.0	52.46
10-1-00-10	188.1	139.3	162.5	323.0	102.6	166.0	42.8	115.5	188.4	27.4	52.6	54.74
10-1-01-00	182.7	136.9	161.7	209.2	103.1	169.8	45.7	243.3	245.6	27.1	50.3	62.34
00-1-00-10	180.9	137.1	163.6	330.2	105.9	165.0	44.6	111.6	179.4	14.0	48.3	66.58
10-1-00-01	183.5	140.1	162.5	365.8	104.9	162.4	43.1	214.1	130.2	26.8	54.7	68.09
00-1-01-00	176.1	133.0	162.1	211.6	106.1	170.1	46.1	240.1	242.0	16.0	47.3	69.06
10-1-10-00	159.8	127.1	147.7	321.8	122.8	172.4	45.4	260.0	259.0	26.4	51.8	71.02
00-1-00-01	181.8	137.0	162.7	367.0	109.6	162.0	46.9	199.2	131.7	26.7	49.6	73.03
10-1-00-00	166.9	129.5	161.4	384.6	115.2	173.8	43.9	248.8	237.9	26.2	58.1	81.49
00-1-10-00	152.5	124.0	149.2	318.1	124.8	173.0	45.9	256.1	256.3	14.7	48.6	84.42
00-1-01-01	193.3	141.3	164.9	199.2	96.4	159.7	46.3	214.3	127.5	17.4	45.3	92.29
10-1-01-01	196.2	143.3	164.6	193.7	93.6	161.3	44.3	222.4	127.3	27.5	48.9	94.18
00-1-00-00	160.7	127.5	160.4	387.6	120.6	173.4	44.1	248.3	236.3	29.0	55.5	95.57
00-0-01-01	288.2	152.4	166.8	186.6	90.3	159.4	45.6	199.2	122.3	14.9	50.4	100.02
10-0-01-01	290.4	152.5	166.3	184.4	89.5	159.6	45.2	205.1	121.5	27.5	50.4	101.55

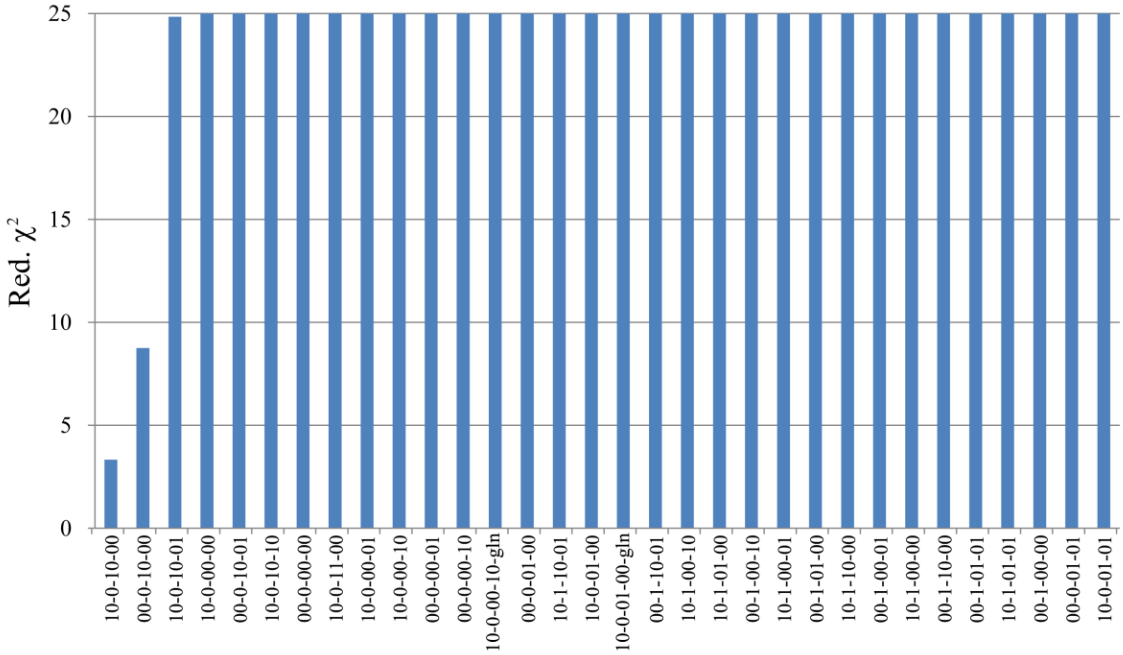


Figure 3.7. Column chart indicating the reduced χ^2 values for all 32 candidate structures.

3.3.2 Single-site Carbonyl (C') CSA Tensors

The reduced χ^2 values for the C' CSA tensors computed for all 32 structures are shown in Figure 3.8A and a plot of the tensor principal components δ_{11}, δ_{22}, and δ_{33} with an overlay of the δ_{iso} is provided in Figure 3.8B. The C' reduced χ^2 values make a large jump (8 to 60) in going from the 14th to the 15th structures. Not unsurprisingly, all of the structures having large reduced χ^2 values are acid-forms. The plot of the C' CSA tensor principal component, ordered according to reduced χ^2 value, indicates why this is the case as the δ_{11} and δ_{22} components shift in opposite directions upon protonation of one of the carboxylic oxygen atoms while the δ_{33} remains relatively stable.[54, 55] The cancelling effects of the shifts in the δ_{11} and δ_{22} components upon protonation reveal why the δ_{iso} is relatively insensitive and, therefore, a poor indicator of protonation state alone.[54, 55]

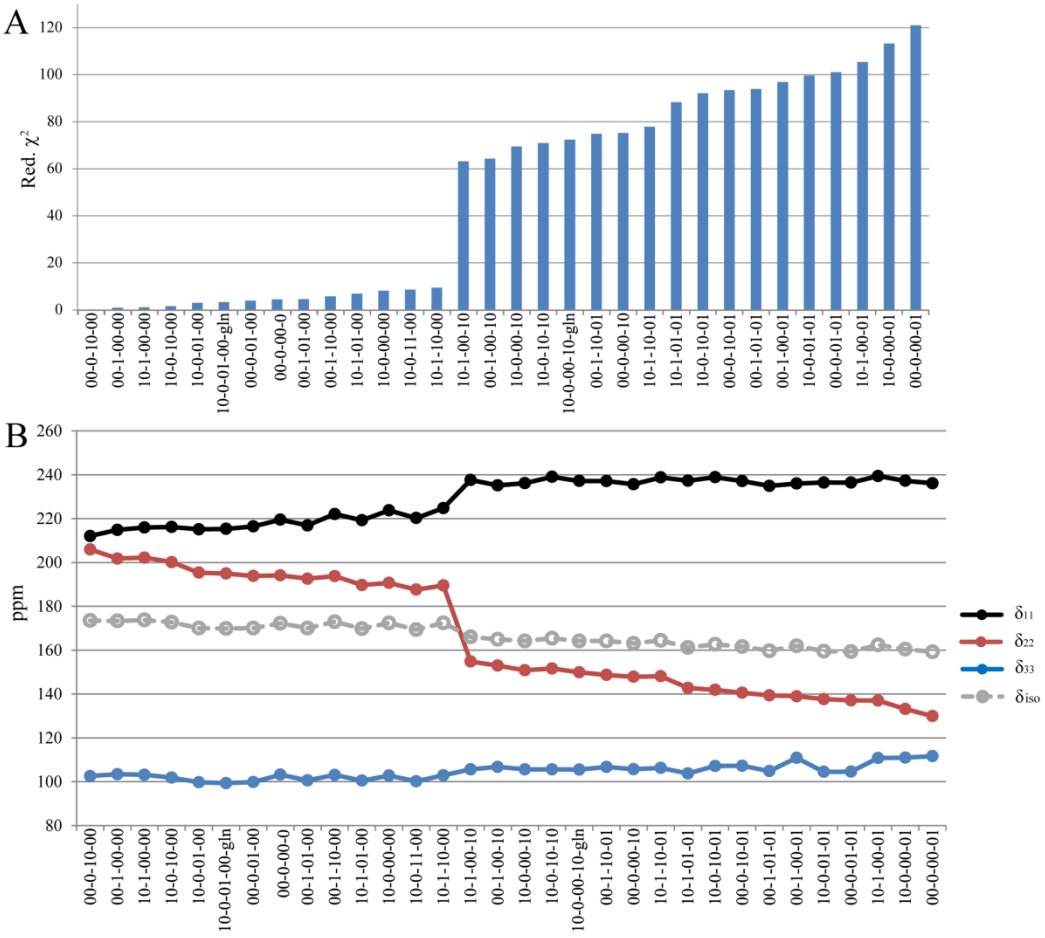


Figure 3.8. Column chart (A) carbonyl, C', CSA reduced χ^2 values. (B) Plot of the principle tensor components and isotropic chemical shift of the C' CSA as indicated.

3.3.3 Single-site Schiff Base Nitrogen (SB N) CSA Tensors

The SB N CSA tensor reduced χ^2 values are shown in Figure 3.9A and their principal tensor components are plotted in Figure 3.9B. The reduced χ^2 values for the SB N CSA tensors make a similar large jump (14.7 to 43.2) as observed in the C' tensors from the 14th to the 15th structures and another large increase (44.7 to 73.6) from the 16th to the 17th structure. The oscillatory nature of the principal components indicates that not

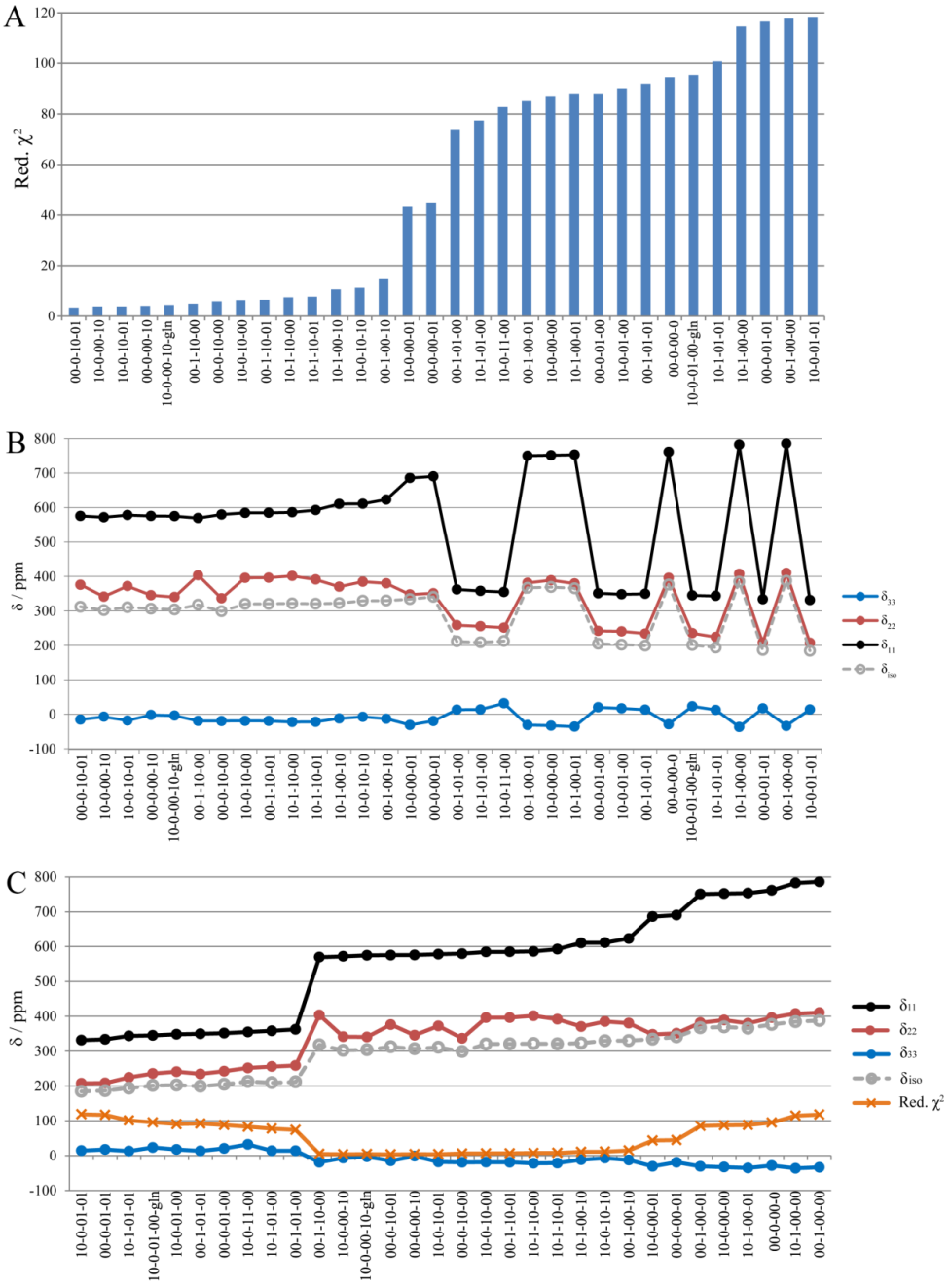


Figure 3.9. Column chart (A) of the Schiff base nitrogen CSA tensor reduced χ^2 values, (B) a plot of the principal tensor components for each structure ranked by lowest reduced χ^2 values, and (C) a plot of the principal tensor components and isotropic chemical shift for each candidate structure ranked by lowest δ_{11} component.

just one feature is responsible for producing very poor reduced χ^2 values unlike the C' tensor where it was easily attributable to the protonation of the carboxylic group.

The principal tensor components are replotted in Figure 3.9C from lowest to highest largest principal component, δ_{11}, with the δ_{iso} and reduced χ^2 values overlaid to make patterns in their change more readily interpretable. What we see in Figure 3.9C is that all protonated SB N structures, 1st through 10th, have lower δ_{11} and δ_{22} components compared with all others and noticeably poor reduced χ^2 values. This provides an indication that a protonated SB N is not the major species. The 25th to the 32nd structures represent the second set in which the reduced χ^2 values increase markedly. Common to these (25th -32nd) structures is that neither the phenolic oxygen nor the SB N is protonated. In looking at the structures that had relatively low reduced χ^2 values, the 11th to 24th, the first thing that stands out is that all structures with a protonated phenolic oxygen are represented here with the exception of the 10-0-11-0, in which both the phenolic oxygen and SB N are protonated. This is unsurprising because clearly the largest effect on the SB N tensor components is its protonation state. The second thing that stands out is that for any structure in which the PO is not protonated the carboxylic oxygen nearest the SB N, O1, is protonated. If we again shift our attention to the 25th to 32nd structures, which like the protonated SB N forms had very poor reduced χ^2 values, we note that the carboxylic acid group is either unprotonated or protonated at O2. It is suspected that hydrogen bonding between O1 and the SB N or between the phenolic oxygen and the SB N in the structures from the 11th to 24th, as "looking the same" from the SB N point of view. In the 25th through 32nd structures, the SB N is not only

deprotonated but also incapable of forming any hydrogen bonding interactions and this is reflected most noticeably in its δ_{11} principal component which increases sharply. The δ_{33} component also decreases which probably explains the slower increase in δ_{iso} with changes in hydrogen bonding interactions compared with direct protonation and deprotonation of the SB N. Finally, unlike the C', the SB N is reporting directly on its own protonation state and we see that δ_{11} and δ_{22} components tensor components shift in the same direction changes in protonation state. The δ_{33} is the least affected but still makes noticeable decreasing steps in the three groupings based on reduced χ^2 value, 1 through 10, 11 through 24, and 25 through 32. Overall, the δ_{iso} shifts noticeably to higher values when the SB N is deprotonated (+100 ppm and greater) indicating that it is an excellent reporter of protonation state.

3.3.4 Two-site Exchange Models

The 2-site exchange model, in which every possible pair of single structures were evaluated and weighted to give the lowest reduced χ^2 value based on the δ_{iso} shifts improved the values considerably as may be expected. In analyzing the resulting combinations, only pairs of structures containing the same number of protons were subsequently considered and this was the only restriction applied. The best 2-site combination had a reduced χ^2 value of 0.39 and represents an 80%:20% equilibrium of the 10-0-10-00 and 10-0-01-00 respectively. The major species corresponds to the PO, which ranked as the best single configuration, and the minor species, 10-0-01-00, represents transfer of a proton from the phenolic oxygen to the SB N, hereafter referred to as the SB tautomer. An essentially identical equilibrium was selected with the 10-0-01-0-

gln configuration, in which the βGln114 was moved from its crystallographic position to form a hydrogen bond with the phenolic oxygen, indicating that this configuration does not alter the predicted chemical shifts to any noticeable extent. The corresponding mixed CSA tensors for the C' and SB N in the best 2-site exchange model had reduced χ^2 values of 1.64 and 0.67 respectively. Table 3.3 provides the isotropic chemical shift values and corresponding reduced χ^2 values for the PO, SB, their 2-site exchange equilibrium, and the experimental values. Tables 3.4 and 3.5 provide the CSA tensor results for the SB N and C' positions respectively. The SB structure alone had reduced χ^2 values of 49.5, 2.78, and 85.9 for the δ_{iso} shifts, C'CSA, and SB N CSA tensor respectively. For comparison, Figure 3.10 provides a column chart with the χ^2 values for the isotropic chemical shifts, C', and SB N CSA tensors for the 10 best 2-site combinations calculated and are ranked in ascending order based on their δ_{iso} shifts reduced χ^2 value. The 80%:20% PO:SB was the only exchange combination in which all three reduced χ^2 values are less than two. The best single configuration, 100% PO, ranks ninth and is the last to have a reduced χ^2 value based on δ_{iso} shifts less than four.

3.3.5 Three-site Exchange Model

Finally, a three-site exchange model was also evaluated applying the same restrictions regarding zero net change in protons between structures. The resulting equilibrium weights essentially selected the best two-site model, 80:20 PO:SB, with negligible contributions from any others, on the order of 10^{-4} to 10^{-7} %, with only one exception. This exception corresponded to the PO:SB at roughly the same 80:20 PO:SB, with negligible contributions from any others, on the order of 10^{-4} to 10^{-7} %, with only

Table 3.3. Calculated δ_{iso} chemical shifts (ppm) for each candidate structure comprising the best-fit 2-site exchange, PO and SB, their calculated 80:20 PO:SB equilbrium δ_{iso}, and the experimentally measured δ_{iso}(ppm). The reduded χ^2 values for each individual structure and the 2-site equilibrium are also provided.

Atom	PO	SB	2-Site	Exp.
N1	257.0	272.0	260.0	262.0
C2	141.0	149.4	142.7	144.6
C3	150.1	164.1	152.9	153.1
SB N	320.5	202.2	297.2	298.6
C^{α}	107.8	93.3	105.0	105.1
C'	172.8	170.1	172.2	173.1
C^{β}	47.3	47.4	47.3	47.0
O1	239	230	237	239
O2	234	225	232	237
N^{ε}	26.9	27.8	27.1	31.5
2AP N	54.1	52.2	53.7	56.0
χ^2 CS	3.33	49.5	0.37	-

Table 3.4. Calculated and experimental Schiff base nitrogen (SB N) δ_{ii} chemical shifts (ppm) for the prinicpal components of the chemical shift anisotropy (CSA) tensor.

Component	PO	SB	2-Site	Exp.
δ_{33}	-19	17	-12	-6
δ_{22}	396	241	383	368
δ_{11}	585	348	521	533
SB N χ^2	7.91	85.9	0.63	-

Table 3.5. Calculated and experimental carbonyl, C', principal component, δ_{ii} (ppm), chemical shifts for the CSA tensor. Reduced χ^2 values are listed for the calculated values.

Component	PO	SB	2-Site	Exp.
δ_{33}	102	100	101	105
δ_{22}	200	195	199	203
δ_{11}	216	215	216	211
C' χ^2	1.47	2.78	1.65	-

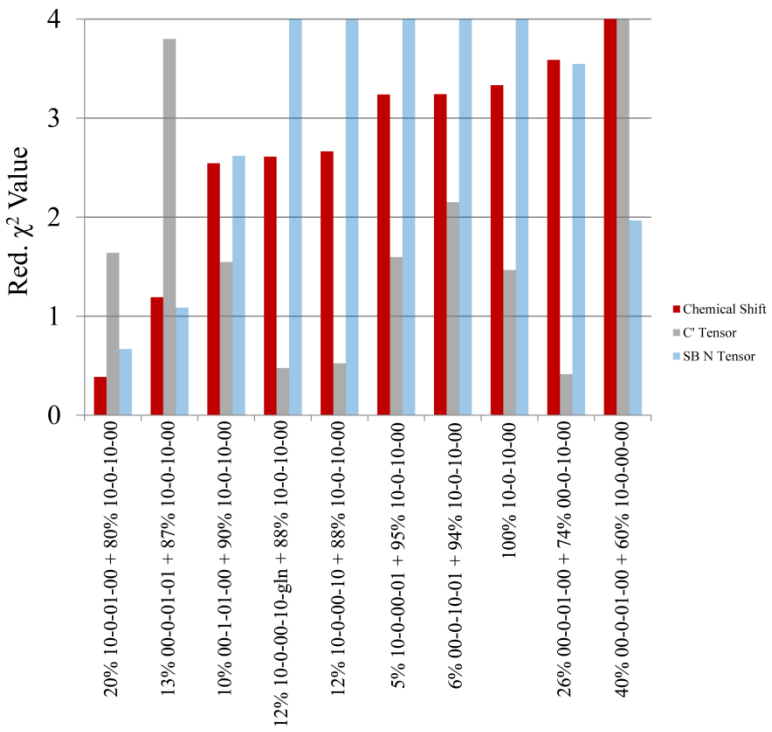


Figure 3.10. Column chart providing the 10 2-site exchange equilbrium models with the lowest calculated reduced χ^2 values based on the 11 δ_{iso} chemical shifts (red). The reduced χ^2 values of the corresponding carbonyl, C' (dark grey), and Schiff base nitrogen, SB N (light blue), CSA tensors are also provided for comparison.

one exception. This exception corresponded to the PO:SB at roughly the same 80:20 equilibrium plus a 0.5 % contribution from the 10-0-00-10-gln configuration, which represents protonation of the carboxylic oxygen nearest the SB N, O1, and a hydrogen bond from the βGln114 sidechain to the phenolic oxygen. This structure will be referred to as the AF structure. The three forms PO, SB, and AF are shown in Figure 3.11. The reduced χ^2 for the isotropic chemical shifts was higher than the best two-site model by 0.01 and those for the corresponding CSA tensors changed by -0.2 for the C' and +0.02 for the SB N.

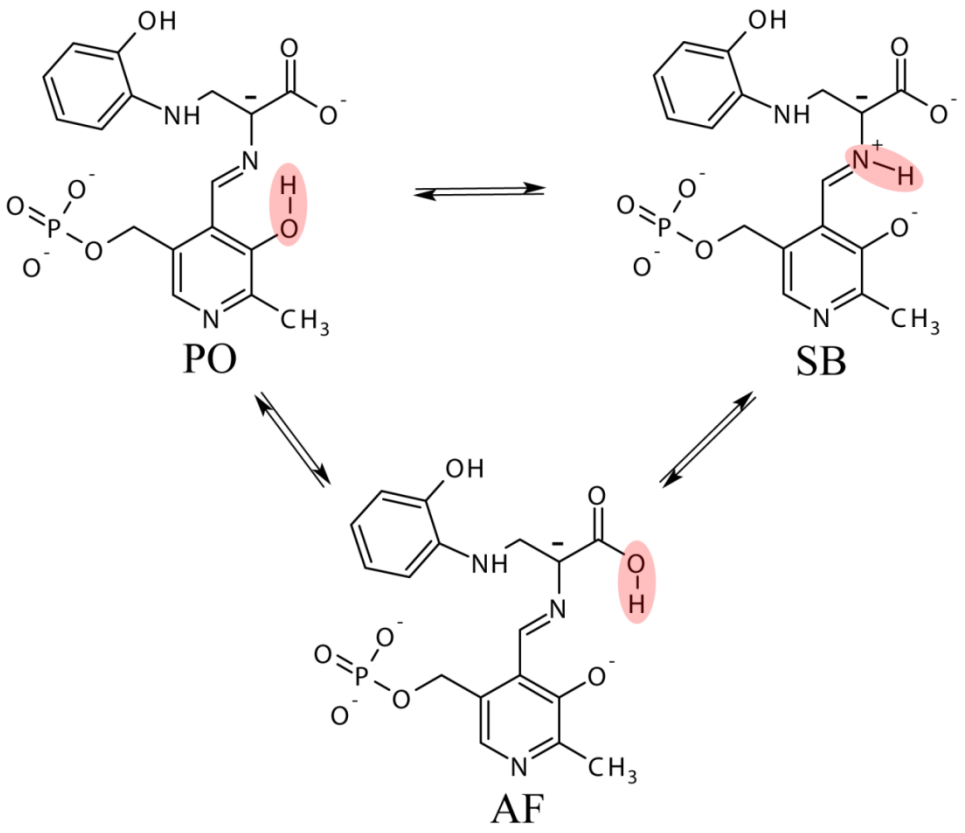


Figure 3.11. Depiction of 3-site tautomeric exchange in which a single proton is shuttled between the phenolic oxygen, PO, Schiff base nitrogen, SB, or carboxylic oxygen nearest the SB N, AF. The best two-site equilibrium model selected an exchange between PO and SB with populations of 80% and 20% respectively. In a three-site model a contribution of less than 1% was attributed to the AF form.

3.4 Discussion

3.4.1 Best-fit Results

The PO tautomer ranked as the best single configuration and, from model compound studies, would be the presumed major tautomeric species in cases where the pyridine ring nitrogen, N1, is deprotonated.[51, 56] The largest discrepancy, > 20 ppm, between calculated and experimental chemical shifts at the SB N in the PO tautomer is resolved in the two-site exchange model that selects the SB tautomer as the minor species

at 20% as the best combination. The reduced χ^2 of this two-site model with respect to the isotropic chemical shifts, at 0.39, is considerably better than that of the single PO configuration. The differences between the experimental and calculated shifts for the PO:SB equilibrium are less than one RMSE for all but the PLP pyridine ring C2 carbon yet it is still well within two RMSE. The calculated βLys87 N^{ε} is also consistent with experimental results that indicate a protonated species even though predicted values tended to be slightly lower than the 31.5 ppm experimental value, predictions for the deprotonated form were considerably lower.[30] Fittingly, a vast improvement in the reduced χ^2 value for the SB N CSA tensor is also noted at 0.67 versus 7.91 in the PO configuration and 90.2 in the SB configuration. What is particularly remarkable with regards to the SB N CSA tensor is how even single structures with very poor reduced χ^2 values alone, can be combined to produce an excellent match with the experimental results. The C' CSA reduced χ^2 did increase slightly from 1.47 to 1.64 but, as seen in Figure 3.10, several combinations produced relatively low reduced χ^2 values with two exceptions, the second and tenth structures. In structures where the C' tensors reduced χ^2 values were considerably lower than the best fit 2-site exchange model, 3rd, 4th, and 9th, the corresponding SB N CSA tensor and/or the isotropic chemical shift reduced χ^2 values were relatively high indicating a poor overall fit.

The ability to experimentally measure the site specific atomic probes within the enzyme active site by NMR, build testable models of the enzyme active site based on the X-ray crystal structure, and to reliably calculate chemical shifts have afforded us the ability to characterize the $E(Q_3)_{2AP}$ intermediate with unprecedented detail. Establishment

of the tautomeric equilibrium involving dynamic proton exchange, including populations, provides a better foundation with which to understand the relationships between the active site structure and function and begin to explore their mechanistic implications. This will begin with a discussion of the N1 and SB N chemical shifts observed here with respect to model PLP compound studies and predictions. Second, the sensitivity of the carbon chemical shifts is discussed. The third topic focuses on the observed oxygen chemical shifts and how an important question presented in the previous chapter can now be viewed with far greater acuity in light of the first principles calculations. Then an initial exploration of the mechanistic implications of the observed tautomeric equilibrium will be discussed relative to the predictions of chemical shifts for other configurations of the $E(Q_3)_{2AP}$ intermediate modeled. Finally, initial calculations from other intermediates will then be presented for comparison.

3.4.2 Schiff Base Nitrogen Chemical Shift

Studies on model PLP compounds in the solid state identified coupling between intra- and intermolecular hydrogen bonds involving PLP and interacting species.[51, 56] The predicted 2-site equilibrium in this study agrees well with those findings with respect to the PO tautomer being predominant in cases where the pyridine nitrogen, N1, is not protonated.[51] In cases where N1 is protonated, the proton previously on the phenolic oxygen migrates to the SB N.[51, 56] At 298 ppm, the chemical shift of the SB N is closer to that of a deprotonated species, however, its chemical shift and linewidth demonstrated a temperature dependence that, upon analysis presented earlier, predicted the protonated SB N population at 21%. (All SB N NMR measurements were made by B. G. Caulkins).

The SB N CSA tensor improvement in the two-site exchange model selecting 20% SB was in excellent agreement of the experimental results.

3.4.3 Pyridine Ring Nitrogen Chemical Shift

The isotropic chemical shift value of the N1 nitrogen measured by B. G. Caulkins was 262 ppm. Similar measurements in related PLP enzymes and in model PLP compounds generally observe pyridine ring nitrogen chemical shifts typically around 320 ppm when deprotonated and close to 200 ppm when protonated.[51, 56] Based upon the measured chemical shift alone, 262 ppm, one might assume proton exchange was likely and this would perhaps fit the narrative that protonation at N1 was promoting the tautomeric exchange between the phenolic oxygen and the SB N. B. G. Caulkins performed temperature dependence measurements for the N1 nitrogen and the chemical shift was insensitive to temperature change and its linewidth remained relatively narrow, two indicators that argue against exchange. The calculated chemical shifts in the models tested indeed indicate that the N1 can experience shielding within the active site sufficient to lower its chemical shift value to 262 ppm or even lower. This leads to questions regarding what factors promote the transfer of the proton between the phenolic oxygen and Schiff-base nitrogen in TRPS if the pyridine nitrogen in fact remains unprotonated throughout the mechanism and what catalytic purpose does protonation of the SB N serve.

Clues to the first question can be related to studies of model compounds and the active site structures of other PLP enzymes such as alanine racemase, AlaR, aspartate aminotransferase, AAT, and O-acetylserine sulfhydrylase, OASS.[56] Studies of model

compounds by Chan-Huot et al. aimed at determining the factors for protonating, and thus activating, the SB N in internal aldimines also analyzed X-ray structures of PLP enzymes from different families.[56] They proposed that protonation of N1 or very strong hydrogen bond interactions occurring at N1 with nearby enzyme residue was one way to promote this tautomeric shift.[56] AAT's N1 is protonated through its interaction with an Asp sidechain carboxylate.[56] In the TRPS $E(Q_3)_{2AP}$ intermediate crystal structure N1 is 2.62 Å from the sidechain hydroxyl group of βSer377.[12] In the models tested in this study, this distance was maintained within +/- 0.02 Å. In AlaR, however, there is not a strong hydrogen bond interaction apparent in the crystal structure given the near 3.0Å distance between N1 and the nearest sidechain nitrogen of a basic Arg residue.[56] Several crystallographic waters are near the phenolic oxygen in AlaR, however, and Chan-Huot et al. hypothesized that hydrogen bonding directly to the phenolic oxygen may be sufficient to transfer the proton to the SB N. In other enzymes, such as O-acetylserine sulfhydrylase where, like TRPS, N1 interacts with an active site serine sidechain hydroxyl group, but an Asn residue sidechain is within hydrogen bonding distance to the phenolic oxygen.[56] The open form TRPS internal aldimine crystal structure contains many structural waters in the vicinity of the phenolic oxygen but the closed structure $E(Q_3)_{2AP}$ does not.[12] Furthermore, no obvious catalytic residues capable of hydrogen bonding to the phenolic oxygen are evident in the X-ray crystal structure.[12] Molecular dynamics simulations conducted by Huang et al., however, did observe the βGln114 sidechain move from its crystallographic position to engage in hydrogen bonding with the with the phenolic oxygen when it was in the phenolate form.[44] Models incorporating the

βGln114 positioned in the hydrogen bonding interaction were tested for the SB and AF. As mentioned in the results section, the best two-site models selecting equilibrium between the SB and PO were essentially identical and included the SB tautomer with and without the βGln114 positioned for hydrogen bonding. The results indicate that the positioning of the βGln114 sidechain is a possibility for promoting the proton transfer leading to the observed exchange from the phenolic oxygen to the SB N if the hydrogen bonding interaction at N1 is not sufficiently strong.

Addressing the question regarding the protonation state of the pyridine ring nitrogen, N1, was important given the assumption that its protonation state had been previously assumed from its UV/Vis absorption maximum at 468 nm.[57, 58] The results of this study demonstrate that protonation of N1 is not required to explain the experimentally observed chemical shifts and two possible modes affecting transfer of the proton from the phenolic oxygen to the SB N are present in relation to investigations by others.[44, 56] The mechanistic implications of the protonation state at the N1 position are discussed later relative to intuitive assumptions regarding magnetic shielding and the next catalytic step in the sequence. A deeper look at the mechanistic implications by other theoretical methods is the subject of the next chapter. At this point, it would be more appropriate to refer to the TRPS quinonoid intermediates as carbanions, however to avoid complications by switching the name, quinonoid will be retained.

3.4.4 Carbon Chemical Shifts

Naturally, the atoms directly involved in proton exchange are by far the most sensitive to the protonation state. In the results section the ostensible insensitivity of the

C' δ_{iso} to the protonation of the carboxylic oxygen atoms was related to the cancelling effects of shifts in the δ_{11} and δ_{22} principal components which were attainable by measuring the CSA tensor in the solid-state (measurements by group member B. G. Caulkins).[54, 55] The ^{13}C δ_{iso} shifts of the serine derived C^{β} and the pyridine ring C2 were not very sensitive reporters of the tautomeric equilibrium occurring in the $E(Q_3)_{2AP}$ intermediate.[59] The ^{13}C chemical shifts of the pyridine ring C3 and the serine derived C^{α} do stand out among carbon atoms as the differences in their chemical shifts were predicted to span 14 ppm between the SB and PO for both. The chemical shift differences predicted at the PLP C3 carbon between protonated phenolic and deprotonated phenolate are also consistent with model compound studies by O'Leary and indicate that, within the enzyme active site, C3 may serve well as a proxy for the phenolic oxygen protonation state.[60] This is important because enrichment of the phenolic site with ^{17}O is a difficult organic synthesis problem. The C^{α} chemical shift also displayed a 14 ppm difference in predicted chemical shifts between the PO and SB protonation state, however, given the commercial availability of ^{15}N and ^{13}C enriched amino acids, both the SB N and C^{α} will generally be amenable to direct measurement in PLP enzymes.

3.4.5 Oxygen Chemical Shifts

The variety of configurations tested also provides valuable insight regarding the observed isotropic ^{17}O chemical shifts.[17] In the last chapter three hypotheses were proposed regarding the upfield shifts of both oxygen atoms of the substrate derived carboxylate.[17] The first was that a fast exchange equilibrium occurred in which each oxygen atom carried a proton for some amount of time. The second hypothesis suggested

close hydrogen bonding contacts with nearby enzyme residues were responsible for the upfield shifts. Finally, the third proposed that the extra shielding was the result of the increased electron density in the quinonoid/carbanion intermediates. The predicted chemical shifts for the oxygen atoms of the carboxylate are in good agreement with the PO configuration alone as well as the suggested equilibrium of 20% SB and 80% PO.[17] The predicted chemical shifts for the oxygen atoms of acid-form configurations, those having a 1 in the penultimate or last digit in the binary code scheme, both shift far upfield. The protonated oxygen chemical shifts are predicted to be lower than the same for amino or carboxylic acids in the solid state.[61] More unusual are the shifts predicted for the unprotonated oxygen which moves to values well below that for an ionized carboxylate.[61] In carboxylic acids the oxygen of acid-forms that is not bonded to the hydrogen shifts downfield of the ionized carboxylate form.[61] The agreement between the 2-site exchange model and the experimental shifts strongly suggest that increased shielding concomitant with an increase in electron density in the quinonoid/carbanion species as being the major contributor to the upfield shifts of the oxygen sites. Essentially, the inclusion of an acid-form tautomer, first suggested in the $E(Q_3)_{indoline}$ study, is not necessary to explain the observed shifts, however, its presence cannot be ruled out entirely.[13, 17]

As previously discussed, only one 3-site exchange model had a contribution greater than 10^{-4}% of a third tautomer before χ^2 values exceeded 3 and it pertained to the 10-0-10-00-gln configuration at 0.5% with essentially the same 20%:80% SB:PO.[17] In the fast-exchange limit, an equilibrating mixture of the acid (AF), SB, and PO forms has

an average chemical shift, $\langle\delta\rangle = p_{AF}\delta_{AF} + p_{SB}\delta_{SB} + p_{PO}\delta_{PO}$, where p_i and δ_i are the corresponding populations and chemical shifts of the tautomers.[17] Making use of the first principles calculations, the average chemical shift of the exchanging tautomer system can be predicted and quantitatively compared to the experimental shift through the root-mean-squared deviation (RMSD),

$$RMSD = \sqrt{\left(\langle\delta\rangle - \delta_{exp}\right)^2} = \sqrt{(p_{AF}\delta_{AF} + p_{SB}\delta_{SB} + (1 - p_{AF} - p_{SB})\delta_{PO} - \delta_{exp})^2}.$$

RMSD's for O1 and O2 are plotted as a function of p_{AF} and p_{SB} in Figure 3.12 below and allow an upper limit of approximately 13% on the AF population before the RMSD of O1 is twice that of the linear rescaling test set (test set RMSD=8.25 ppm for oxygen).[17]

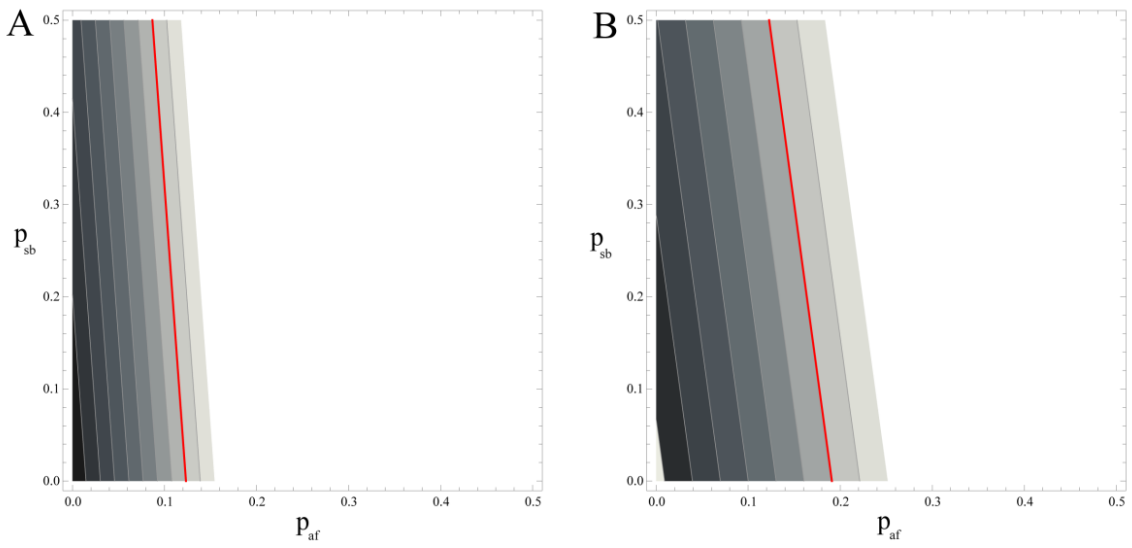


Figure 3.12. RMSE surfaces for (A) O1 and (B) O2 as a function of the AF and SB populations. The red line indicates an RMSE of 16 ppm, twice the RMSE of the linear rescaling test set. Contours are shown from 2-20 ppm with a spacing of 2 ppm.

The quadrupolar coupling constants are not directly measurable in solution state ^{17}O QCT NMR experiments, however, a range of their values can be calculated based on the

limitations to the magnitude of the asymmetry parameter, η_Q (see equations (4) and (5) presented earlier). If η_Q is equal to zero, then the C_Q is equal to the P_Q. If the η_Q is at its maximum value of 1, then C_Q is equal to $P_Q/1.154$. The range of possible C_Q values for both oxygen atoms in the $E(Q_3)_{2AP}$ intermediate are provided in Table 3.6 along with the original calculated C_Q values, which are known to be systematically over estimated from 10% -15%, and C_Q values adjusted to compensate for the overage by reducing the values by 10%.[39] The adjusted C_Q values calculated from the 2-site exchange model fall within the range of those possible as calculated from ^{17}O experimental P_Q in reference.[17] The P_Q values between experimental and calculated results was not performed because it is unknown how the individual principal components are affected as opposed to the value of C_Q, which has been examined by Harbison.[39]

Table 3.6. Experimental C_Q range, in MHz, Calculated from P_Q.[17] Corresponding C_Q calculated, C_Q Calc. (MHz), and adjusted by reduction of 10% to account for systematic errors in calculations, C_Q Calc.,Adj.[39]

	Exp. C_Q Range	C_Q Calc.	C_Q Calc., Adj.
O1	6.66 to 7.69	7.83	7.12
O2	6.38 to 7.37	7.03	6.39

3.4.6 Mechanistic Implications from Nuclear Magnetic Shielding.

The catalytic intermediate following the $E(Q_3)$ quinonoid/carbanion intermediate is a 2nd external aldimine, $E(Aex_2)$ and results from reprotonation of the C^α carbon.[31] As mentioned previously, it had long been thought that the formation of a true quinonoid, that is a protonated N1 and, likely, a protonated SB N, were necessary to stabilize the negative charge at the C^α by maximizing the electron withdrawing ability of the pyridine ring.[62] Supporting this position was the UV/Vis absorption maximum at 468 nm which

was consistent with that of known quinonoids.[57] The first principles calculations indicate that a protonated N1 does not contribute to the observed experimental NMR chemical shifts and presumably to the UV/Vis spectrum. Having calculated models protonated at N1 provides the ability to compare their predictions to the best-fit exchange model.

Table 3.7. Calculated δ_{iso} (ppm) shifts for select nuclei of six candidate structures of the $E(Q_3)_{2AP}$ intermediate. The PNPO, PNSB, and PNAF differ from PO, SB, and AF by having a protonated pyridine ring nitrogen, N1.

Structure	ID	N1	SB N	C^{α}	O1	O2
10-0-10-00	PO	257.0	320.5	107.8	238.9	233.6
10-0-01-00	SB	272.0	202.2	93.3	229.5	224.6
10-0-00-10	AF	270.5	302.1	98.3	110.0	167.6
10-1-10-00	PNPO	159.8	321.8	122.8	260.0	259.0
10-1-01-00	PNSB	182.7	209.2	103.1	243.3	245.6
10-1-00-10	PNAF	188.1	323.0	102.6	115.5	188.4

The chemical shifts for the SB, PO, and acid-form AF, and their counterparts having a protonated N1, PNSB, PNPO, and PNAF are given in Table 3.7. If we consider an increase in magnetic shielding, leading to a decrease in chemical shift, as an indicator of enhanced electron density in the vicinity of a particular nuclear species we see some clear trends emerge. First, in comparing the chemical shifts of N1with a transfer of a proton from the SB N in the SB configuration to the phenolic oxygen in the PO configuration we see a corresponding increase in shielding (decrease in chemical shift) by 15 ppm. In the corresponding PNSB and PNPO configurations, the N1 chemical shift decreases by 22.9 ppm for the same proton transfer. This indicates that tautomerization occurring between the SB and PO modulates electron transfer to the pyridine ring. This is also reflected in the chemical shifts of the carboxylic oxygen atoms and the C^{α}. The

oxygen atoms are deshielded by 9.4 ppm and 9.0 ppm for O1 and O2 respectively and the C^{α} by 14.5 ppm in going from the SB to the PO tautomer. In the corresponding PNSB and PNPO structures the same proton transfer deshields the oxygen atoms by 16.7 ppm and 13.4 ppm for O1 and O2 respectively and by 19.7 ppm at C^{α}. These results indicate that protonation at N1 also modulates electron transfer to the pyridine ring and perhaps more strongly than a shift of a proton to the phenolic oxygen alone as evidenced by the fact that the O1 and O2 are more deshielded in PNPO and PNSB than in the SB and PO tautomers. The trend at C^{α} is that it experiences the most shielding when the SB N is protonated and in order of most to least shielding rank as SB > PNSB > PO > PNPO. The acid forms, AF and PNAF do not alter the shielding at C^{α} much from SB and PNSB respectively. The trends observed here are in line with the conclusions of a theoretical study regarding the relative acidities of the C^{α} and C4' carbons by Casasnovas et al.[27, 28] Their conclusion was that the protonation state of N1 was a primary determinant of which atom would be more acidic and that the protonation states at the phenolic oxygen and SB N modulated the relative acidity between the C^{α} and C4'.[27] From the results herein, it appears that the SB would prime C^{α} for reprotonation whereas the PNPO would delocalize the electron density away from the imine moiety to the pyridine ring, possibly favoring a reaction at C4'as in aminotransferases. The C^{α} chemical shifts are much closer between the PNSB and PO configurations and both are at values in between those of the PNPO and SB. In the original $E(Q_3)_{indoline}$ study by Lai et al. and in several theoretical studies of PLP model compounds, such as those by Casasnovas et al. various methods were used to calculate the charges at C^{α} and C4' including natural bond orbital analysis

and CHelpG, however, none have been done in the presence of the full complement of active site residues.[13, 28] Using the cluster based approach as accomplished for the calculation of NMR parameters a study was initiated to test the effects of catalytic residues on the computed charges with the hope that it would provide further insight into how reaction specificity is modulated in PLP enzymes and is the topic of Chapter 4.

In the ^{17}O QCT NMR study, two E(Q_3) and two E(A-A) intermediates, brought about using different indole analogues were measured to test the generality of the observations.[17] The similarities between the E$(Q_3)_{indoline}$ and E$(Q_3)_{2AP}$ quinonoids were consistent as were those between the aminoacrylates, E(A-A) and E(A-A)(BZI).[17] Although complete first-principles studies for the other TRPS intermediates are only in their preliminary stages, select structures corresponding to the PO, SB, AF, PNPO and PNSB were calculated for the E$(Q_3)_{indoline}$ and the results are listed in Table 3.8. The E(A-A) intermediate results are provided in Table 3.9 and correspond to the PO, SB, PNPO, PNSB, and an acid-form in which the O2 oxygen is protonated, U-AF, because an acid-form protonated at O1 was unstable. The results of the E(A-A) are described after first making a comparison of the quinonoid/carbanion species.

3.4.7 Comparisons of NMR Shielding with Select E$(Q_3)_{indoline}$ Structures

The E$(Q_3)_{indoline}$ shows the same trends as observed for the E$(Q_3)_{2AP}$. Both oxygen chemical shifts in the AF shift markedly upfield as was observed in the E$(Q_3)_{2AP}$ indicating this is a general feature of carbanion intermediates (here still referred to as quinonoids) and their corresponding additional electron density. The experimental chemical shift values for the E$(Q_3)_{indoline}$ oxygen atoms were 243 ppm and 233 ppm, and

the differences are reflected in the all of the corresponding structures which are greater than those calculated for the $E(Q_3)_{2AP}$. The same trends regarding the effects of tautomeric exchange between PO and SB (and PNPO and PNSB) was observed in the calculated structures and, like the $E(Q_3)_{2AP}$, the SB structure sees the most shielding at C^{α}.

3.4.8 Comparisons of NMR Shielding with Select E(A-A) Structures

The E(A-A) intermediate structures calculated corresponded to the PO, SB, PNPO, and PNSB. Placement of a proton on O1 was unstable and would migrate to the SB N. A form in which O2 was protonated was stable, likely because there was no suitable proton acceptor in the immediate vicinity, this form is called U-AF to differentiate it from the AF calculated for the quinonoids. Transfer of a proton from O1 to the SB N was not observed in the quinonoid forms and perhaps these were more stable resulting from the extra electron density in the structure as evidenced by the low ^{17}O δ_{iso} values. Also notable with the U-AF structure is a divergence in chemical shifts when one of the carboxylic oxygen atoms in protonated. O2, protonated in U-AF has a predicted chemical shift of 166 ppm and O1's shift is predicted at 336 ppm. Both of these shifts are completely consistent with those observed for carboxylic acids measured in the solid state.[61] Additionally, protonation at N1 has a negligible effect on the shielding at the carboxylic oxygen sites (PNSB vs SB and PNPO vs PO). The largest shielding is observed for the oxygen atoms in the SB form for which the shielding at C^{α} is also the greatest. The shielding at N1 between SB and PO shows negligible change as does the

Table 3.8. Calculated δ_{iso} chemical shifts (ppm) of select structures of the $E(Q_3)_{indoline}$ intermediate. Experimental values are given in the last row.

Structure	Code	N1	C2	C3	SB N	C^{α}	C'	C^{β}	O1	O2	N^{ε}
10-0-10-00	PO	261.9	142.2	150.1	325.5	107.3	171.6	51.0	254.1	227.4	23.7
10-0-01-00	SB	276.4	152.8	165.3	198.9	92.3	168.5	51.1	241.7	215.7	24.2
10-1-10-00	PNPO	151.6	127.5	147.7	326.2	123.7	171.6	49.6	277.8	255.8	24.8
10-1-01-00	PNSB	173.7	138.8	161.8	209.9	104.0	168.7	49.8	261.9	241.8	25.2
10-0-00-10	AF	273.9	155.8	166.3	298.6	100.4	164.1	48.4	113.0	160.4	25.4
	Experimental	265.0	145.4	154.1	296.0	103.5	173.0	54.1	243.0	233.0	31.5

Table 3.9. Calculated δ_{iso} chemical shifts (ppm) of select structures of the E(A-A) intermediate. Experimental values are given in the last row.

Structure	Code	N1	C2	C3	SB N	C^{α}	C'	C^{β}	O1	O2	N^{ε}
00-0-10-00	PO	303.4	145.8	154.1	306.0	142.3	169.7	119.6	298.2	262.3	24.0
00-1-10-00	SB	305.2	156.2	171.7	158.1	130.7	167.3	118.4	289.4	260.8	23.6
00-0-01-00	PNPO	204.5	138.5	155.8	317.2	139.1	168.9	124.3	295.0	263.1	22.7
00-1-01-00	PNSB	204.5	138.5	155.8	317.2	139.1	168.9	124.3	295.0	263.1	22.7
00-0-00-01	U-AF	290.9	156.5	166.8	296.2	144.7	170.7	119.9	336.3	166.0	22.7
	Experimental	298.0	151.5	158.2	292.5	146.0	NA	118.9	292.0	258.0	24.2

shielding at N1 between the PNSB and PNPO configurations. The muted effects of proton transfers in the E(A-A) compared with the quinonoids likely results from the lack of excess electron density to shift between pyridine and imine moieties.

3.5 References

1. Toney, M.D. Biochimica Et Biophysica Acta, 2011. **1814**(11): p. 1407-18.

2. Hayashi, H. Journal of Biochemistry, 1995. **118**(3): p. 463-473.

3. Jansonius, J.N. Current Opinion in Structural Biology, 1998. **8**(6): p. 759-769.

4. Eliot, A.C. and J.F. Kirsch. Annual Review of Biochemistry, 2004. **73**: p. 383-415.

5. Limbach, H.H., M. Chan-Huot, S. Sharif, P.M. Tolstoy, I.G. Shenderovich, G.S. Denisov, and M.D. Toney. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2011. **1814**(11): p. 1426-1437.

6. Metzler, C.M., A. Cahill, and D.E. Metzler. Journal of the American Chemical Society, 1980. **102**(19): p. 6075-6082.

7. Christensen, H.N. Journal of the American Chemical Society, 1958. **80**(1): p. 99-105.

8. Metzler, D.E. Journal of the American Chemical Society, 1957. **79**(2): p. 485-490.

9. Robitaille, P.M., R.D. Scott, J.Y. Wang, and D.E. Metzler. Journal of the American Chemical Society, 1989. **111**(8): p. 3034-3040.

10. Hyde, C.C., S.A. Ahmed, E.A. Padlan, E.W. Miles, and D.R. Davies. Journal of Biological Chemistry, 1988. **263**(33): p. 17857-17871.

11. Ford, G.C., G. Eichele, and J.N. Jansonius. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1980. **77**(5): p. 2559-2563.

12. Niks, D., E. Hilario, A. Dierkers, H. Ngo, D. Borchardt, T.J. Neubauer, L. Fan, L.J. Mueller, and M.F. Dunn. Biochemistry, 2013. **52**(37): p. 6396-6411.

13. Lai, J.F., D. Niks, Y.C. Wang, T. Domratcheva, T.R.M. Barends, F. Schwarz, R.A. Olsen, D.W. Elliott, M.Q. Fatmi, C.E.A. Chang, I. Schlichting, M.F. Dunn, and L.J. Mueller. Journal of the American Chemical Society, 2011. **133**(1): p. 4-7.

14. Sharif, S., E. Fogle, M.D. Toney, G.S. Denisov, I.G. Shenderovich, G. Buntkowsky, P.M. Tolstoy, M.C. Huot, and H.H. Limbach. Journal of the American Chemical Society, 2007. **129**(31): p. 9558.

15. Tycko, R., ed. *Nuclear Magnetic Resonance Probes of Molecular Dynamics*. 1994, Kluwer Academic Publishers: Dordrecht, The Netherlands. 550.

16. Cavanagh, J., W.J. Fairbrother, A.G. Palmer, M. Rance, and N.J. Skelton, *Protein NMR Spectroscopy, Principles and Practice*. 2 ed. 2007, New York: Elsevier.

17. Young, R.P., B.G. Caulkins, D. Borchardt, D.N. Bulloch, C.K. Larive, M.F. Dunn, and L.J. Mueller. Angewandte Chemie-International Edition, 2016. **55**(4): p. 1350-1354.

18. Macholl, S., D. Tietze, and G. Buntkowsky. Crystengcomm, 2013. **15**(43): p. 8627-8638.

19. Mueller, L.J. and M.F. Dunn. Accounts of Chemical Research, 2013. **46**(9): p. 2008-2017.

20. Harris, R.K., R.E. Wasylishen, and M.J. Duer, eds. *NMR Crystallography*. 2009, John Wiley & Sons: West Sussex, U.K.

21. Facelli, J.C. and D.M. Grant. Nature, 1993. **365**(6444): p. 325-327.

22. Rajeswaran, M., T.N. Blanton, N. Zumbulyadis, D.J. Giesen, C. Conesa-Moratilla, S.T. Misture, P.W. Stephens, and A. Huq. Journal of the American Chemical Society, 2002. **124**(48): p. 14450-14459.

23. Olsen, R.A., J. Struppe, D.W. Elliott, R.J. Thomas, and L.J. Mueller. Journal of the American Chemical Society, 2003. **125**(39): p. 11784-11785.

24. Gupta, R., G.J. Hou, R. Renirie, R. Wever, and T. Polenova. Journal of the American Chemical Society, 2015. **137**(16): p. 5618-5628.

25. Hartman, J.D., T.J. Neubauer, B.G. Caulkins, L.J. Mueller, and G.J.O. Beran. Journal of Biomolecular Nmr, 2015. **62**(3): p. 327-340.

26. Bach, R.D., C. Canepa, and M.N. Glukhovtsev. Journal of the American Chemical Society, 1999. **121**(28): p. 6542-6555.

27. Casasnovas, R., M. Adrover, J. Ortega-Castro, J. Frau, J. Donoso, and F. Munoz. Journal of Physical Chemistry B, 2012. **116**(35): p. 10665-10675.

28. Casasnovas, R., A. Salva, J. Frau, J. Donoso, and F. Munoz. Chemical Physics, 2009. **355**(2-3): p. 149-156.

29. Roy, M., S. Keblawi, and M.F. Dunn. Biochemistry, 1988. **27**(18): p. 6698-6704.

30. Caulkins, B.G., B. Bastin, C. Yang, T.J. Neubauer, R.P. Young, E. Hilario, Y.M.M. Huang, C.E.A. Chang, L. Fan, M.F. Dunn, M.J. Marsella, and L.J. Mueller. Journal of the American Chemical Society, 2014. **136**(37): p. 12824-12827.

31. Caulkins, B.G., C. Yang, E. Hilario, L. Fan, M.F. Dunn, and L.J. Mueller. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2015. **1854**(9): p. 1194-1199.

32. Dunn, M.F., D. Niks, H. Ngo, T.R.M. Barends, and I. Schlichting. Trends in Biochemical Sciences, 2008. **33**(6): p. 254-264.

33. Hartman, J.D. and G.J.O. Beran. Journal of Chemical Theory and Computation, 2014. **10**(11): p. 4862-4872.

34. Hartman, J.D., S. Monaco, B. Schatschneider, and G.J.O. Beran. Journal of Chemical Physics, 2015. **143**(10).

35. Saito, H., I. Ando, and A. Ramamoorthy. Progress in Nuclear Magnetic Resonance Spectroscopy, 2010. **57**(2): p. 181-228.

36. Levitt, M.H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. 2 ed. 2008, West Sussex, U. K.: John Wiley & Sons Ltd.

37. Adiga, S., D. Aebi, and D.L. Bryce. Canadian Journal of Chemistry-Revue Canadienne De Chimie, 2007. **85**(7-8): p. 496-505.

38. Frisch, M.J., G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N.J. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.

39. Harbison, G.S. Journal of Magnetic Resonance, 2015. **257**: p. 24-31.

40. Zhu, J.F. and G. Wu. Journal of the American Chemical Society, 2011. **133**(4): p. 920-932.

41. Challoner, R., C.A. Mcdowell, W. Stirtan, and S.G. Withers. Biophysical Journal, 1993. **64**(2): p. 484-491.

42. Taguchi, J.E., S.J. Heyes, D. Barford, L.N. Johnson, and C.M. Dobson. Biophysical Journal, 1993. **64(**2): p. 492-501.

43. Schrodinger, LLC, *The Pymol Molecular Graphics System, Version 1.8*. 2015.

44. Huang, Y.M.M., W.L. You, B.G. Caulkins, M.F. Dunn, L.J. Mueller, and C.E.A. Chang. Protein Science, 2016. **25**(1): p. 166-183.

45. Song, Y.F., J.J. Mao, and M.R. Gunner. Journal of Computational Chemistry, 2009. **30**(14): p. 2231-2247.

46. Dennington, R., T. Kieth, and J. Millam, *Gauss View, Version 5*. 2009, Semichem Inc.: Shawnee Mission, KS.

47. Becke, A.D. Journal of Chemical Physics, 1993. **98**(2): p. 1372-1377.

48. Lee, C.T., W.T. Yang, and R.G. Parr. Physical Review B, 1988. **37**(2): p. 785-789.

49. Vosko, S.H., L. Wilk, and M. Nusair. Canadian Journal of Physics, 1980. **58**(8): p. 1200-1211.

50. Stephens, P.J., F.J. Devlin, C.F. Chabalowski, and M.J. Frisch. Journal of Physical Chemistry, 1994. **98**(45): p. 11623-11627.

51. Sharif, S., D. Schagen, M.D. Toney, and H.H. Limbach. Journal of the American Chemical Society, 2007. **129**(14): p. 4440-4455.

52. Ditchfield, R.. Molecular Physics, 1974. **27**(4): p. 789-807.

53. Wolinski, K., J.F. Hinton, and P. Pulay. Journal of the American Chemical Society, 1990. **112**(23): p. 8251-8260.

54. Gu, Z.T. and A. McDermott. Journal of the American Chemical Society, 1993. **115**(10): p. 4282-4285.

55. Gu, Z.T., R. Zambrano, and A. McDermott. Journal of the American Chemical Society, 1994. **116**(14): p. 6368-6372.

56. Chan-Huot, M., A. Dos, R. Zander, S. Sharif, P.M. Tolstoy, S. Compton, E. Fogle, M.D. Toney, I. Shenderovich, G.S. Denisov, and H.H. Limbach. Journal of the American Chemical Society, 2013. **135**(48): p. 18160-18175.

57. Metzler, C.M., A.G. Harris, and D.E. Metzler. Biochemistry, 1988. **27**(13): p. 4923-4933.

58. Barends, T.R.M., T. Domratcheva, V. Kulik, L. Blumenstein, D. Niks, M.F. Dunn, and I. Schlichting. Chembiochem, 2008. **9**(7): p. 1024-1028.

59. London, R.E. Journal of Magnetic Resonance, 1980. **38**(1): p. 173-177.

60. Oleary, M.H. and J.R. Payne. Journal of Biological Chemistry, 1976. **251**(8): p. 2248-2254.

61. Wong, A. and F. Poli, *Solid-State Oxygen-17 NMR Studies of Biomolecules*. Annual Reports on NMR Spectroscopy, ed. G.A. Webb. Vol. 83. 2014, London: Elsevier.

62. Richard, J.P., T.L. Amyes, J. Crugeiras, and A. Rios. Current Opinion in Chemical Biology, 2009. **13**(4): p. 475-483.

Chapter 4

Explorations of Chemical Mechanism in the 2AP-Quinonoid of Tryptophan Synthase by NPA Charge Calculations

4.1 Introduction

4.1.1 Background

Application of the NMR crystallography approach to the study of the 2AP-quinonoid intermediate, $E(Q_3)_{2AP}$, in tryptophan synthase, TRPS, has identified tautomeric exchange between enolimine (PO) and ketoenamine (SB) forms of the pyridoxal-5'-phosphate (PLP)/substrate complex brought about by transfer of a proton from PLP's phenolic oxygen to the nearby Schiff base nitrogen (Figure 4.1). The equilibrium populations determined by reduced χ^2 square fitting of computationally predicted shifts to those experimentally obtained were 80% PO to 20% SB. The prediction of 20% SB tautomer is further supported by ^{15}N chemical shift and linewidth temperature dependence measurements made by group member B. G. Caulkins.

Another critical feature determined by the study outlined in Chapter 3 was that the pyridine ring nitrogen is not protonated. As stressed in Chapters 1 and 3, formation of a quinonoid brought about by protonation of the pyridine ring nitrogen was long thought a general feature of PLP mechanisms.[1-3] More recent studies highlighted earlier have begun to challenge this assumption.[4-8]

The availability of X-ray structures provided the early clues to the deprotonated pyridine nitrogen by indicating that enzyme functional groups interacting with PLP's

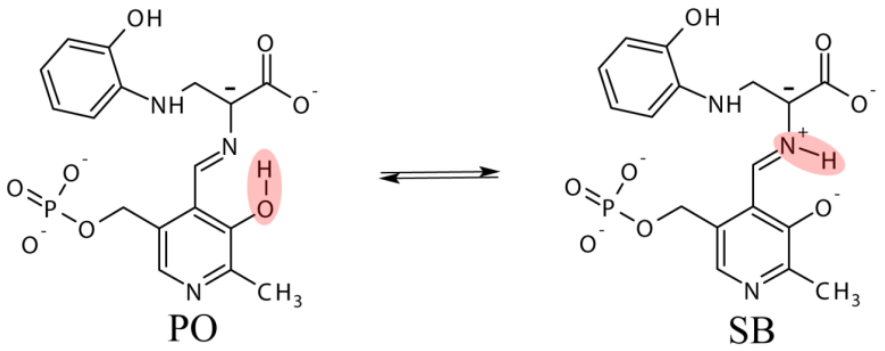


Figure 4.1. Equilibrium tautomers identified through NMR crystallography in the tryptophan synthase (TRPS) β-site $E(Q_3)_{2AP}$ intermediate. The enolimine tautomer, PO, is the major species at 80%, and the ketoenamine, SB, the minor at 20%.

pyridine nitrogen were not invariably acidic groups.[9, 10] For example, enzymes such as TRPS and O-acetylserine sulfhydrylase (OASS) have a serine hydroxyl group and alanine racemase (AlaR) has a basic arginine guanidinium group positioned near the pyridine nitrogen.[9, 10] Adding to the growing evidence that a protonated pyridine nitrogen is not a required element for catalysis have been PLP enzymes that do not exhibit the characteristic UV/Vis absorption maxima ascribed to quinonoids, but still perform heterolytic bond cleavage of one of the bonds to the Cα in their mechanisms.[10] Modified cofactors and site-directed mutagenesis studies have shed even more light on the issue indicating that formation of a quinonoid species is absolutely critical in some PLP enzymes yet in others is detrimental.[10] TRPS presents an interesting case as it does exhibit a UV/Vis absorption maximum consistent with those identified as quinonoid forms yet, in the most highly detailed study of any enzymatic PLP intermediate, the combination of NMR and first-principles calculations see no evidence of a protonated pyridine.[11]

In an effort to determine what role the protonation states of the PLP/substrate complex play in directing the subsequent catalytic event for E(Q_3), natural population analysis (NPA) atomic charge calculations were performed on the experimentally observed tautomers as well as forms having a protonated pyridine ring nitrogen and protonated carboxylic oxygen. By employing a similar approach used to calculate NMR shifts outlined in Chapter 3, the NPA charges are computed with the PLP/substrate complex confined within the local structure the active site out to 7Å which maintains important intermolecular interactions.[12] A comparison of NPA charge calculations performed on the PLP/substrate extracted from the active site environment and solvated in implicit water as well as structures allowed to freely optimize under the implicit solvent conditions reveal that it is indeed a combination of the electronic environment of the active site as well as the geometry of the PLP/substrate complex optimized within its framework that effect the overall atomic charges.

There has been a long-standing interest in the PLP literature on the important roles acid-base catalysis and proton mediated tautomerism play in directing reaction specificity in PLP mechanisms.[13-16] The ability of PLP to participate in such a wide variety of different chemical transformations is no doubt attributable to its highly "tunable" nature.[17] Elucidating the finer details of how the unique environment of an enzyme's active site adjust its catalytic power according to the desired overall chemistry when undesirable side reactions are not that distant, from an energetic standpoint, is an important aim for enzymologists.[18] One means by which investigators have sought insight regarding the roles of various protonation states on reaction specificity in PLP has

been through theoretical partial atomic charge calculations.[6, 7, 19] In the context of the $E(Q_3)_{2AP}$ intermediate in TRPS β-site catalytic mechanism, to proceed to the next identifiable intermediate, the second external aldimine, $E(Aex_2)$, the Cα requires protonation by βLys87, see Figure 4.2.[16, 20] Chemical intuition would predict that the Cα is a better target for reprotonation by the charged amino group of βLys87 when the

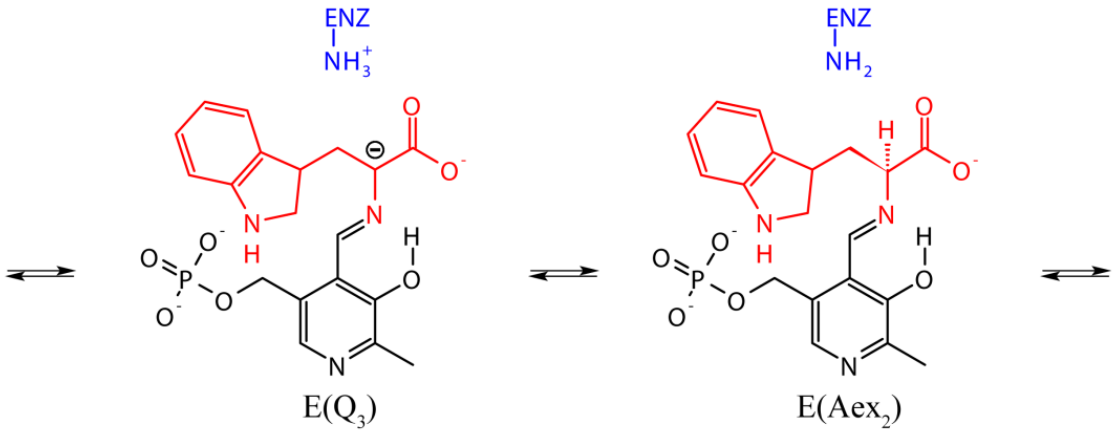


Figure 4.2. Sequential intermediates in the TRPS β-site reaction mechanism. Protonation of the third quinonoid/carbanion, $E(Q_3)$, at the Cα position by βLys87 (blue) yields the second external aldimine intermediate $E(Aex_2)$.

localization of negative charge is maximal at its site.[6, 7, 19, 20] In the X-ray crystal structure of $E(Q_3)_{2AP}$ (PDBID 4HPJ), the βLys87 N^{ε} is positioned almost equidistant from both the Cα and C4' carbons as shown in Figure 4.3.[21] In the study of the $E(Q_3)_{indoline}$ conducted by Lai and coworkers, NBO charges were computed in competing resonance structures identified using the NMR crystallography approach that included a protonated Schiff base, protonated phenolic oxygen, and acid-form tautomer (carboxylic oxygen nearest the Schiff base nitrogen protonated).[19] The NBO calculations predicted much more negative charge buildup at Cα in the latter species which led to the prediction that

the acid-form was the catalytically active form.[19] In all three structures the pyridine ring nitrogen was assumed to be protonated, a feature recently discovered to be incorrect (Chapter 3).[16] Casasnovas et al. have performed comprehensive theoretical atomic charge studies aimed at determining the effects of protonation states on charge distribution between the pyridine and imine moieties on model compounds as well as their effects on C4' and Cα acidity, as these two sites are particularly relevant in PLP mechanisms.[6, 7] A discussion of their findings follows.

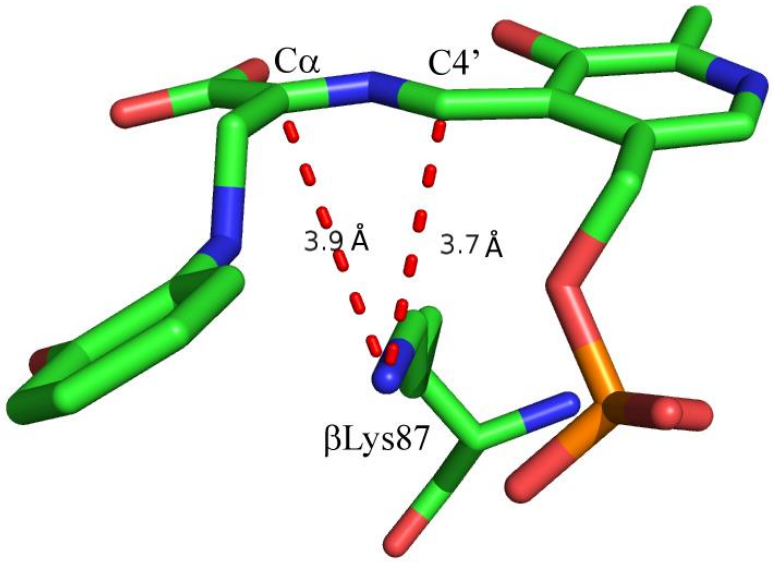


Figure 4.3. Position of βLys87 with respect to the Cα and C4' carbons in the X-ray crystal structure of the $E(Q_3)_{2AP}$ intermediate (PDBID 4HPJ).[21]

4.1.2 Relevant Atomic Charge Studies of PLP Tautomerism

Casasnovas et al. studied the effects of protonation states on a truncated model of PLP, 3-hydroxy-pyridine, 3HP, complexed with alanine via a Schiff-base linkage at 3HP's C4'. The quinonoid/carbanion form of the intermediate was also constructed by

removing the proton attached at the alanine Cα position. The protonation states they tested included forms similar to the PO and SB shown in Figure 4.1 as well as their protonated pyridine forms. Structures with and without acidic or basic groups positioned to interact with the pyridine nitrogen were also constructed. The structures were solvated in implicit (CPCM)[22, 23] water and geometry optimized prior to atomic charge calculations using the CHelpG scheme.[24] The resulting calculations placed a negative charge on C4' in all configurations tested. The Cα was predicted to have a negative charge in only the ketoenamine forms lacking a protonated pyridine nitrogen.[7] This would be the equivalent of the SB structure. The Cα was never predicted to have a more negative charge than at C4' and the authors concluded that the C4' was always nucleophilic and thus always a target for reprotonation.[7] The greatest difference in charge between the C4' and Cα was in the structure that would correspond to the PO with a protonated pyridine ring nitrogen. Another observance obtained by computing the delocalization of charge indicated that in all structures, except the one corresponding to the SB, the negative charge of the carbanion was always delocalized on the pyridine ring and always to a greater extent when the pyridine ring nitrogen was protonated. In the SB structure there was virtually no observed net delocalization of the negative charge and it was concluded that the protonated Schiff base nitrogen was the factor responsible for preventing further delocalization of charge at Cα thus increasing its probability for reprotonation.[7]

In a follow-up study focused on delineating the roles of tautomerism with respect to carbon acidity at the C4' and Cα positions, Casasnovas et al. combined theoretical

atomic charge calculations with both theoretical and experimental pKa measurements of PLP analogues.[6] Most relevant to this specific study were the NBO atomic charge (synonymous with NPA) calculations and corresponding determination of delocalized charge between the imine and pyridine moieties in models of glycine bound to PLP.[6] In all cases, the negative charge of carbanions, generated by removing a proton from Cα, was delocalized onto the pyridine ring with the most delocalization noted for a form that would correspond to our PO with a protonated pyridine nitrogen as noted in their previous study.[6] A form having the protonation states of the PO was second in terms of delocalizing charge to the pyridine. A protonated SB form with protonated pyridine was third followed by a form similar to the SB depicted in Figure 4.1 experiencing the least delocalization.[6] The more negative charge delocalized onto the pyridine ring the more favorable C4' became for reprotonation. The protonated Schiff base/deprotonated pyridine nitrogen, showing the least delocalization towards the pyridine would hence be the most amenable to reprotonation at Cα.[6] With respect to their findings and our determination that the prevailing equilibrium tautomers in the $E(Q_3)_{2AP}$ are the PO and SB forms, (both without a protonated pyridine nitrogen) the SB form should be more conducive for protonation at Cα. What we would like to determine are the effects the enzyme active site has on the charges computed because, to our knowledge, a cluster based calculation of atomic charges on a PLP/substrate quinonoid intermediate in the presence of the full complement of active site residues has not been accomplished.

4.1.3 NPA Atomic Charge Calculations

The concept of atoms in molecules having a net partial charge has been a useful construct to aid in the interpretation of the theoretically calculated electronic structure and corresponds to an intuitive chemical interpretation. There are many approaches for calculating partial atomic charges, which are not experimentally observable quantities. Given the large size of the active site cluster model in this study, > 600 atoms, the natural population analysis, NPA, method of Reed, Weinstock, and Weinhold was deemed the most appropriate. The NPA method is considered superior to the more widely used Mulliken and Löwdin population analyses.[25] The NPA approach is based on the construction of an orthonormal set of natural atomic orbitals, NAOs, in an arbitrary atomic orbital basis set that involves a diagonalization procedure on the one-electron density matrix followed by the removal of interatomic overlap.[25] The first step leads to pre-NAOs that are divided by occupancy into a minimal set for those with non-zero occupancy and a Rydberg set for those with zero occupancy.[25] The second step is symmetric orthogonalization procedure that is weighted by occupancy to preserve the highly populated atom-centered orbitals that leads to the final NAOs.[25] The NAO's of the minimal set form the natural minimal basis, NMB, and those of the Rydberg set form the natural Rydberg basis, NRB. The NMB preserves as much of the atomic character as possible and any remaining electron density is assigned as occupancy to the NRB. The occupancy of the NRB is interpreted as the change that occurs when the atom is placed in the molecular environment.[25] The sum of all the orbital contributions to a specific atom in the molecule gives its atomic charge.[26] NPA is less sensitive to the quality of the basis

sets employed and numerically very stable.[25-28] Criticisms of the approach are that it tends to produce results that accentuate bond polarity by producing larger negative/positive values at atoms in bonds.[29] For the purposes of this study, we are more interested in the relative changes in computed charges rather than their absolute values.

4.2 Experimental

4.2.1General Features

All calculations performed in this work were conducted using Gaussian09[30] and the NBO 6.0 program.[31] All calculations were performed at the DFT level of theory using the B3LYP hybrid functional.[32-35] Calculations were performed on UC Riverside's Bioinformatics Biocluster using between 6 and 32 processors running in parallel with 1.5 GB of ram per processor assigned.

4.2.2 Active-site NPA Charge Calculations

The NPA calculations were performed on the same structures optimized for the NMR shielding calculations and were detailed in Chapter 3 Sections 3.2.1 to 3.2.3. Six candidate structures were selected for comparison. These structures are the PO and SB, identified earlier as the two tautomers observed at equilibrium in $E(Q_3)_{2AP}$, an acid-form, AF, and the corresponding protonated pyridine ring nitrogen structures denoted PNPO, PNSB, and PNAF and are shown in Figure 4.4. The formal charges are highlighted in each structure. The surrounding enzyme residues comprising the 7Å active site cut have a net charge of zero and, therefore, the formal charge on each structure is equal to the net charge on the PLP/substrate complex.

Calculations on the active site cluster were performed using three different multi-tier basis set schemes. For a description of the high/medium/low layers see Chapter 3 Section 3.2.3 and Figure 3.6. The first basis set combination (high/medium/low)

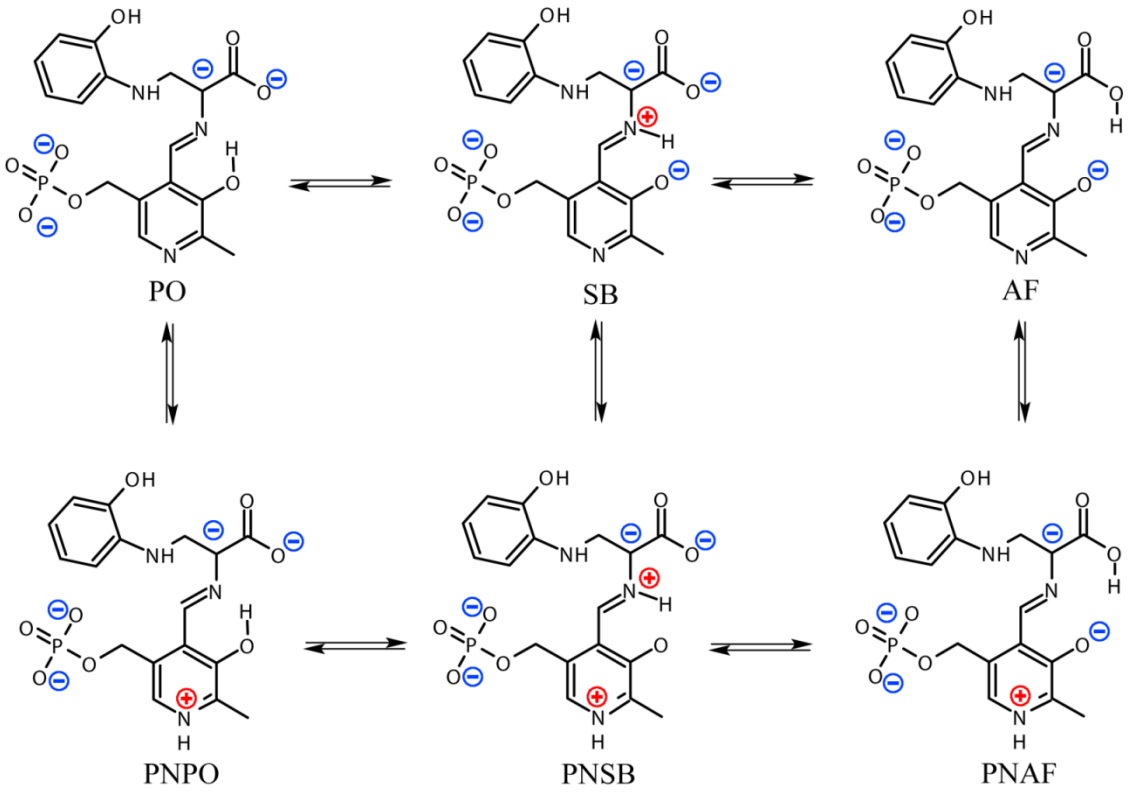


Figure 4.4. Candidate structures for NPA charge calculations. The PO and SB represent the enolimine/ketoenamine tautomers observed at equilibrium in the NMR crystallography study.

consisted of the same combination used for the NMR calculations, 6-311+G(2d,p)/6-311G(d,p)/6-31G. It was noted that while the overall charges on the entire cluster summed to the correct overall formal charge, the PLP/substrate complex did not sum to its own formal charge. To test whether the use of diffuse or polarization functions had any effect on the total charge calculated for the PLP/substrate complex two more combinations were used. The second combination used was thus 6-311G(2d,p)/6-

311G(d,p)/6-31G which eliminates the diffuse function on the high layer. The third basis set combination employed was 6-311G/6-311G/6-31G which, in addition to having no diffuse functions, also lacks polarization functions on both the high and medium layers.

4.2.3 NPA Charges of Extracted PLP/Substrate Complex in Implicit Solvent

To determine the extent to which intermolecular interactions from the surrounding catalytic residues of the active site influenced the charges computed for the PLP/substrate complex, the surrounding enzyme residues were deleted. NPA charge calculations were performed on the PLP/substrate complexes in the geometries determined from their optimization in the presence of the active site residues (see Chapter 3 experimental for details). The charges were calculated in implicit water, dielectric ≈ 80, dichloromethane (DCM), dielectric ≈ 9, and in the gas phase to test whether or not different dielectric mediums would affect the charges computed. The CPCM implicit solvation model was used for both water and DCM. All calculations were performed using B3LYP/6-311+G(2d,p).[22, 23]

4.2.4 NPA Charges of Freely Optimized Extracted PLP/Substrate Complexes

Finally, to test for the effects of geometry on the charges calculated, each of the extracted PLP/substrate complexes was allowed to freely relax in CPCM implicit water and DCM. Optimizations were performed with B3LYP/6-31+G(d,p). Using modified redundant coordinates, dihedral angles for the phosphate group and 2AP sidechain, C5-C5'-O5'-P, Cα-Cβ-N2*-C2*, and Cβ-N2*-C2*-C1* (*asterisks indicate 2AP atoms) as indicated in Figure 4.5, had to be frozen because initial attempts at optimization led to wild fluctuations in the positions of these groups. It is clear that the enzyme can enforce

geometries that would be unstable outside of the active site. NPA charges were calculated using B3LYP/6-311+G(2d,p).

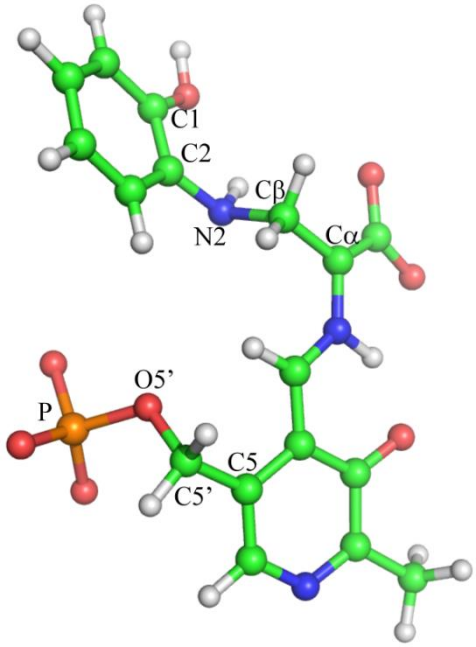


Figure 4.5. $E(Q_3)_{2Ap}$ extracted cofactor/substrate indicating dihedral angles frozen in CPCM implicit solvent geometry optimizations – PLP: C5'C5'-O5'-P, substrate derived: Cα-Cβ-N2-C2 and Cβ-N2-C2-C1.

4.2.5 Calculating Charge Delocalization

To quantify the delocalization of charge with various protonation state configurations a method based on that of Casasnovas et al. was employed.[7] Figure 4.6 shows the atoms comprising the imine moiety (in blue) and the pyridine moiety (in red). The imine moiety includes the C4' and attached hydrogen. As mentioned previously, in cluster models of the enzyme active site, although the total charge summed to the total formal charge for the entire cluster, the charges of individual fragments did not. The PLP/substrate complex always summed to less than its formal charge. As a result, a modified scheme was employed to quantify the delocalization for these cases – which

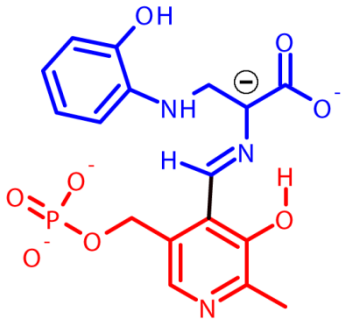


Figure 4.6. Division of $E(Q_3)_{2AP}$ PLP/substrate complex into pyridine (red) and imine (blue) moieties.

works equally well for the extracted fragments whose formal and total charges were always equal as expected. Formal charges for the imine and pyridine were calculated by summing the formal charges as shown in Figure 4.4 assuming no delocalization as depicted. For this method the formal charge of the imine moiety is denoted *FI*, the formal charge of the pyridine is denoted *FP*, and the total formal charge is denoted *FT*. The actual charges were calculated by summing all atoms in each moiety from the NPA charge output are denoted *AI*, *AP*, and their total *AT*. The expected charges, *EI*, *EP* are calculated according to Equation (1).

$$EI = \frac{FI}{FT} AI \text{ and } EP = \frac{FP}{FT} AP \quad (1)$$

The delocalization is then determined by subtracting the expected charge from the actual, to give DI and DP denoting delocalized charge for the imine and pyridine respectively.

$$DI = AI - EI \text{ and } DP = AP - EP \quad (2)$$

By this scheme, if one moiety has more negative charge than expected it will have a negative delocalization value. Naturally the charge gained by one should equal that lost by the other whose positive value indicates the transfer of negative charge away from it.

4.3 Results

4.3.1 Active Site NPA Charges Including Basis Set Effects

The formal, actual, expected, and delocalized , *FX*, *AX*, *EX*, and *DX* (where *X = I*, *P*, or *T* for imine, pyridine, or total respectively) values for the charges computed using the largest basis set on the high layer are provided in Table 4.1. The basis set combinations are denoted Full for the largest basis set (6-311+G(2d,p)) on the high layer, ND for the basis set lacking diffuse functions, and ND NP for the basis set combination lacking both diffuse and polarization functions. Calculation of atomic charge in the clusters for AF and PNAF were not calculated (NC) for the ND and ND NP basis set combinations. The NPA charges at the C4' and Cα positions are of particular interest for determining whether one of those sites would be expected to make a more attractive target for reprotonation by the βLys87 which is in a position almost equidistant from both of those sites in the X-ray crystal structure, see Figure 4.3, and is positively charged in $E(Q_3)_{2AP}$.[16] Reprotonation at C4' is a requirement in transaminases and other related PLP enzymes that proceed through a ketamine/pyridoxamine (PMP) intermediate during at least one step.[13] The atomic charges for Cα and C4' as a function of basis set combination for the PO, SB, PNSB, and PNPO candidate structures are given in Table 4.2 and are plotted in Figure 4.7. The variation in charge at the C4' was particularly insensitive to the basis set used, however, there was a dependence noted at Cα. The

Table 4.1. Formal charges for the imine and pyridine moieties (all numbers in units of elementary charge, e), FI, FP, and their total FT. AI, AP, and AT were the sums of the charges resulting from the NPA calculations for each moiety and the total respectively. The expected charges, EI and EP, were calculated from the fraction of the total formal charge each moiety was expected to possess multiplied by the actual charge calculated. DI and DP represent delocalization of negative charge by subtracting the expected charge from the actual charge.

	PO	SB	AF	PNAF	PNSB	PNPO
FI	-2	-1	-1	-1	-1	-2
FP	-2	-3	-3	-2	-2	-1
FT	-4	-4	-4	-3	-3	-3
AI	-1.50	-0.91	-0.92	-0.84	-0.80	-1.34
AP	-1.93	-2.52	-2.60	-1.88	-1.89	-1.34
AT	-3.43	-3.43	-3.52	-2.72	-2.68	-2.68
FI/FT	0.50	0.25	0.25	0.33	0.33	0.67
FP/FT	0.50	0.75	0.75	0.67	0.67	0.33
EI	-1.71	-0.86	-0.88	-0.91	-0.89	-1.79
EP	-1.71	-2.57	-2.64	-1.81	-1.79	-0.89
DI	0.22	-0.05	-0.04	0.07	0.10	0.45
DP	-0.22	0.05	0.04	-0.07	-0.10	-0.45

Table 4.2. Partial atomic charges calculated at Cα and C4’ with different basis sets as indicated. All values are in elementary units of charge, e.

Basis	Atom	PO	SB	PNSB	PNPO
Full	Cα	-0.082	-0.116	-0.045	0.012
	C4'	-0.078	0.015	-0.032	-0.103
ND	Cα	-0.061	-0.099	-0.029	0.030
	C4'	-0.074	0.014	-0.030	-0.097
ND NP	Cα	-0.047	-0.075	-0.013	0.035
	C4'	-0.084	0.009	-0.030	-0.108

reason for this is not known, perhaps this is a result of the Cα being in closer proximity to atoms, the carboxylic oxygens, that are directly involved in intermolecular interactions with enzyme residues. The C4’ is relatively isolated in this regard. The decision was made to compute NPA charges for the remaining two structures, AF and PNAF, and all extracted PLP/substrate complexes with the 6-311+G(2d,p) basis set.

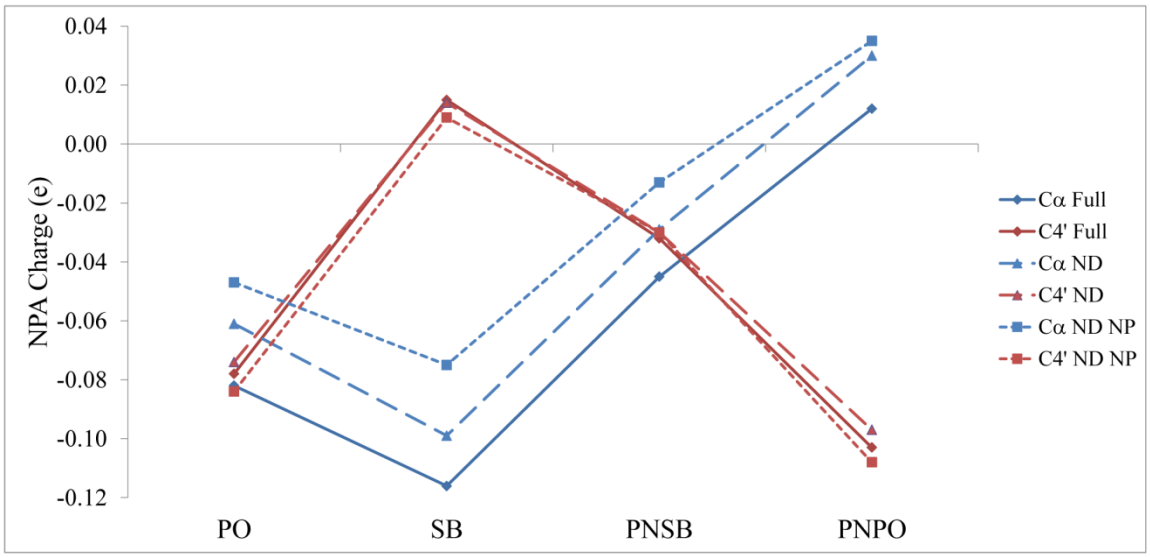


Figure 4.7. NPA charges calculated at Cα and C4' of $E(Q_3)_{2AP}$ in the 7Å active site cluster using three different multi-tier basis set schemes. Full corresponds to 6-311+G(2d,p) for the high layer, ND refers to no diffuse functions, 6-311G(2d,p), and ND NP removes both diffuse and polarization functions from the high and medium layers, 6-311G (H/M).

The fractions of the total formal charge for the PLP/substrate complex computed in the active site for the three basis set combinations used are provided in Table 4.3. The effects of the basis sets on the fractional charge of the PLP/substrate complex were surprising. It was expected that the inclusion of diffuse functions would see the lowest fraction of the formal charge remaining on the PLP/substrate complex in the active site. This was not the case although the values obtained were similar. The charges computed with basis sets lacking either diffuse or both diffuse and polarization functions were identical and lower than the charges calculated with the full basis set by 4% in all structures. A trend is observed in which the forms lacking a protonated pyridine ring nitrogen contain 3% less formal charge than their counterparts having a protonated pyridine ring nitrogen. The PNAF maintained the highest fraction of formal charge in the

full basis set calculations. The trends observed in the atomic charges and their mechanistic interpretation follows in the discussion section.

Table 4.3. Fraction of total formal charge maintained on the PLP/substrate complex when calculated in the presence of the enzyme active-site residues with basis sets that included diffuse and polarization functions, Full, no diffuse functions, ND, and both no diffuse or polarization functions, ND NP.

Basis	PO	SB	AF	PNAF	PNSB	PNPO
Full	0.86	0.86	0.88	0.91	0.89	0.89
ND	0.82	0.82	NC	NC	0.85	0.85
ND NP	0.82	0.82	NC	NC	0.85	0.85

4.3.2 Extracted PLP/Substrate Charge Calculations in Implicit Solvents

Charges computed at C4' and Cα for the extracted PLP/substrate complexes in the active site optimized geometry (ASG) of PO, SB, PNSB, and PNPO were similar for the CPCM implicit solvents of water and DCM. This trend also held for the charges computed on the freely optimized (OPT) extracted PLP/substrate complex in the implicit CPCM solvation. The gas phase charges computed for C4' and Cα in the ASG followed the same general trend as those computed in implicit solvent and were closer to those of the OPT forms, however, the values of C4' in each structure were shifted higher and charge for the Cα in the SB was shifted noticeably lower. The charges computed for PO, SB, PNSB, and PNPO in the ASG in implicit CPCM solvation and in the gas phase as well as the OPT charges in the implicit solvents are displayed in Figure 4.8 and tabulated in Table 4.4. A gas phase optimization was not performed. The nearly identical charges computed in both high and low dielectric mediums relative to each other and the difference of these to the vacuum/gas phase indicates that it is advisable to employ some dielectric medium greater than 1 although more work would have to be accomplished

determine if there were in fact a minimum threshold. All other charge calculations on the extracted PLP/substrate complexes, ASG or OPT, were carried out only in CPCM water.

Table 4.4. NPA charges, in elementary units, e, calculated at Cα and C4' outside of the active site with the geometry optimized, ASG, in CPCM water, CPCM dichloromethane (DCM), or gas phase, and in freely optimized geometries in CPCM water or DCM.

CPCM	Geometry	Atom	PO	SB	PNSB	PNPO
Water	ASG	Cα	-0.058	-0.086	0.016	0.065
		C4'	-0.103	-0.018	-0.051	-0.097
DCM	ASG	Cα	-0.057	-0.087	0.024	0.074
		C4'	-0.101	-0.016	-0.050	-0.091
Water	OPT	Cα	-0.011	-0.051	0.057	0.113
		C4'	-0.138	-0.053	-0.084	-0.110
DCM	OPT	Cα	-0.011	-0.053	0.063	0.114
		C4'	-0.135	-0.050	-0.084	-0.108
None (gas ph.)	ASG	Cα	-0.018	-0.072	0.065	0.114
		C4'	-0.074	-0.007	-0.041	-0.056

It is clear in Figure 4.8 that the geometry influences the computed NPA charges. We observe increases in the charge computed at the Cα and decreases in the charges computed at C4' for all structures in moving from the ASG forms to the OPT geometries. It is well known that geometry has an influence atomic charge calculations but what is important is that the geometries are sufficiently different when optimized within and outside of the active site to result in these marked shifts in the charges. In the discussion section direct comparisons will be made between charges computed in the presence of the active site residues to the extracted complex values having the ASG or OPT geometry.

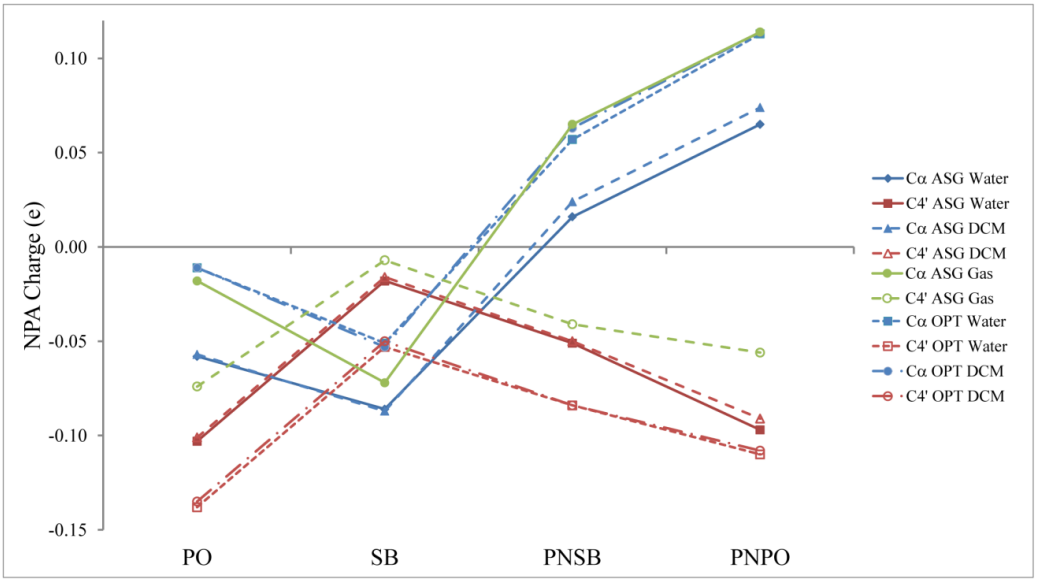


Figure 4.8. Variation in partial atomic charges at Cα and C4' calculated in CPCM implicit solvents, water and dichloromethane (DCM) in the active site geometry (ASG) and freely optimized geometries (OPT). Gas phase charges are given for ASG only.

4.4 Discussion

4.4.1 Influence of the Active Site Residues and Geometry on Atomic Charge

The results of the charges calculated at the Cα and C4' positions of the PLP/substrate complex in the three distinct environments and/or geometry, ENZ, ASG, and OPT are shown in Figure 4.9. It is clear that in addition to enforcing a particular geometry, the intermolecular interactions or effects of the surrounding active site do influence the atomic charges at Cα and C4'. If we consider the PO and the PNSB structures in the ENZ forms we see that the charges predicted for Cα and C4' are both negative but also very similar to each other. The transfer of a proton from the phenolic oxygen to the Schiff base nitrogen (PO to SB) results in the Cα becoming more negative and the C4' assumes a positive value. This is the type of difference and separation in

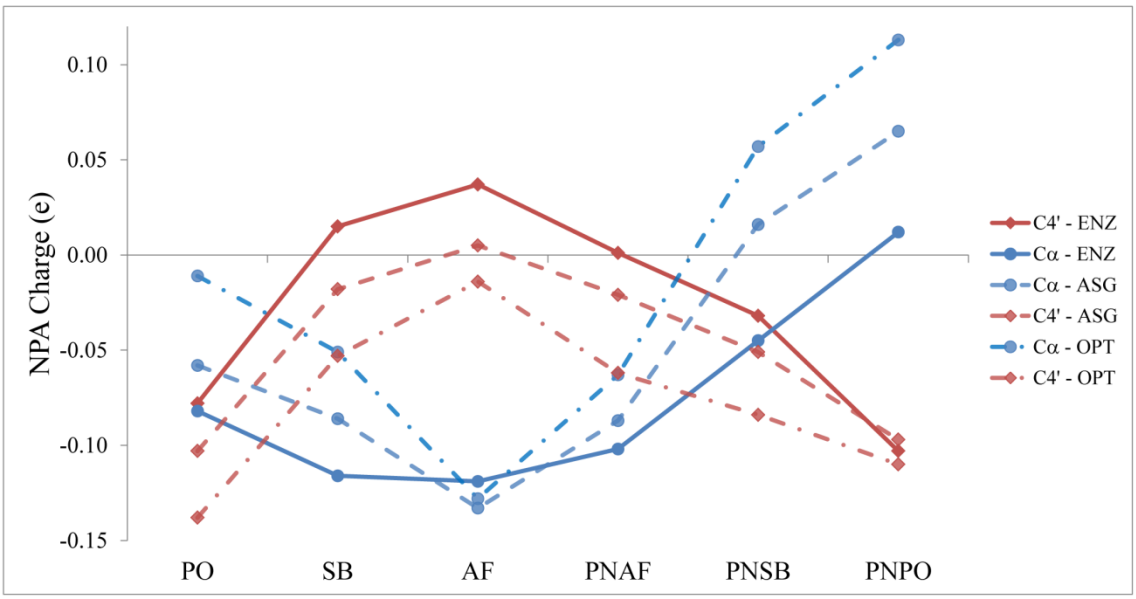


Figure 4.9. NPA charges calculated for Cα and C4' of the PLP/substrate complex in the enzyme active site, ENZ, outside of the enzyme but with the active site optimized geometry in implicit water, ASG, and outside the enzyme active site optimized in implicit water, OPT.

charge we would expect if the Cα were to be the target for reprotonation as it is in TRPS. Transfer of the proton from Schiff base nitrogen to the nearby carboxylic oxygen (SB to AF) sees a slight enhancement in the positive charge at C4' and almost no change at Cα from the SB. The effect of protonating the pyridine ring nitrogen in the AF, AF to PNAF, sees the charge at C4' drop from positive to almost neutral as the Cα makes a shift to a more positive/less negative value. A transfer of the proton back to the Schiff base nitrogen with a protonated pyridine (PNAF to PNSB) further increases the charge (towards positive) at Cα and decreases the charge at C4' leaving them both slightly negative and, like in the PO, almost equal. Finally, a transfer of the proton from the Schiff base nitrogen in the PNSB to the phenolic oxygen leading to the PNPO structure

sees a dramatic increase in in the charge at Cα leading to a positive value and a sharp decrease in the charge at C4' giving it its most negative value amongst the ENZ structures tested. The charges at C4' and Cα in the PNPO indicate that the C4' is a more attractive target for reprotonation in this case versus the opposite situation encountered in the SB and AF structures. The effects of the protonation state of the pyridine ring nitrogen are quite clear as it increases negative charge at C4' in all protonated species versus their corresponding deprotonated species with a concomitant increase (less negative/more positive) in the partial charge predicted at Cα. The effects of protonation state on charge delocalization between the imine and pyridine moieties will be discussed later.

Shifts in the charges calculated for C4' and Cα are observed in all of the ASG forms relative to the ENZ forms. The C4' values are lower (more negative) than computed in ENZ forms except for the PNPO where it is almost identical. At Cα the shifts are to more positive values than those computed in the active site except for the AF alone where it is slightly lower. Allowing the extracted PLP/substrate complex to optimize its geometry in the implicit CPCM water (OPT structures) results in yet another shift of the charges predicted at Cα and C4' in the same directions as observed for the ASG structures. The aforementioned shifts, again, indicate that both the electronic environment (active site versus implicit solvent) and the geometry (active site versus freely optimized) effect the values of the NPA charges computed.

In an effort to quantify the extent to which each variable, environment or geometry, contributes to the change in the calculated charges, the percent changes were

calculated for the charges computed for the ENZ to ASG and ASG to OPT with respect to the overall change (ENZ to OPT) and are tabulated in Table 4.5. The percent changes in moving from the ENZ to the ASG forms were between 34-60%. Naturally, the remainder was attributable to the changes in geometry in going from ASG to OPT structures. It is clear from this that changes from ENZ to ASG and from ASG to OPT are both significant. Most changes in partial charge at Cα and C4' for each of the candidate structures were stepwise and in the same direction, see Figure 4.11 panels B and D, except for the Cα in the AF and the C4' in the PNPO structures which explains their unusual percent changes in Table 4.5. The structures of the ENZ/ASG forms were

Table 4.5. NPA charges at Cα and C4' for all structures tested in the three different environment/geometry combinations, ENZ, ASG and OPT and the absolute percent change in NPA charge moving from ENZ to ASG and ASG to OPT.

	ENZ	ASG	OPT	Range (max -min)	|%Δ| ENZ to ASG	|%Δ| ASG to OPT
Cα PO	-0.082	-0.058	-0.011	0.071	34	66
C4' PO	-0.078	-0.103	-0.138	0.060	42	58
Cα SB	-0.116	-0.086	-0.051	0.065	46	54
C4' SB	0.015	-0.018	-0.053	0.068	49	51
Cα AF	-0.119	-0.133	-0.125	0.014	100	57
C4' AF	0.037	0.005	-0.023	0.060	53	47
Cα PNAF	-0.102	-0.087	-0.063	0.039	38	62
C4' PNAF	0.001	-0.021	-0.062	0.063	35	65
Cα PNSB	-0.045	0.016	0.057	0.102	60	40
C4' PNSB	-0.032	-0.051	-0.084	0.052	37	63
Cα PNPO	0.012	0.065	0.113	0.101	52	48
C4' PNPO	-0.103	-0.097	-0.110	0.013	46	100

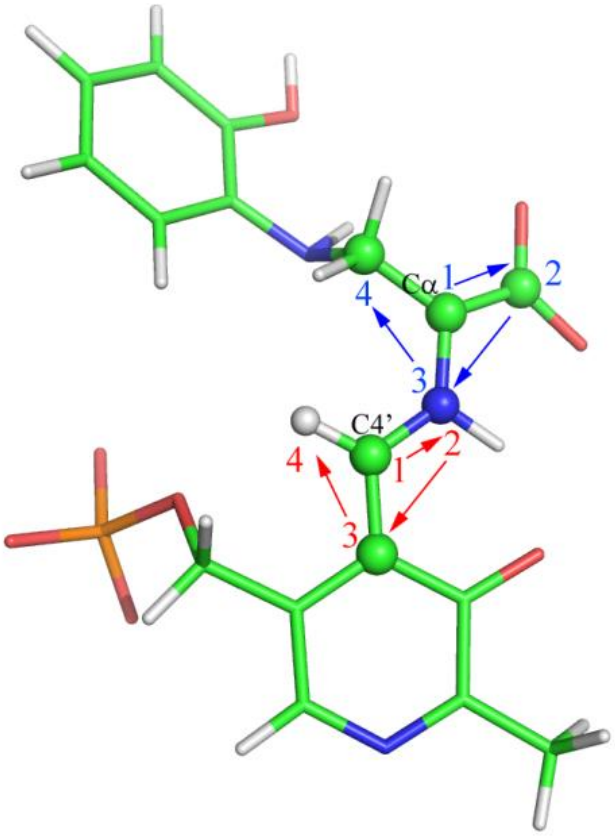


Figure 4.10. Measurement of improper dihedral angles at Cα and C4'. The order at Cα is highlighted by the blue numbers/arrows. The order at C4' is indicated by the red numbers/arrows.

overlaid with the OPT forms and, in particular, the geometry at the Cα site especially stood out. As a representative example, Figures 4.12 shows overlays of ASG and OPT forms of the SB candidate structure highlighting this observation. Improper dihedrals angles were calculated at C4' and Cα to measure their deviation from a planar sp^2 form as depicted in Figure 4.10. Additionally, NPA charges and improper dihedrals were measured for intermediate structures (Int.) in the optimization process. The most

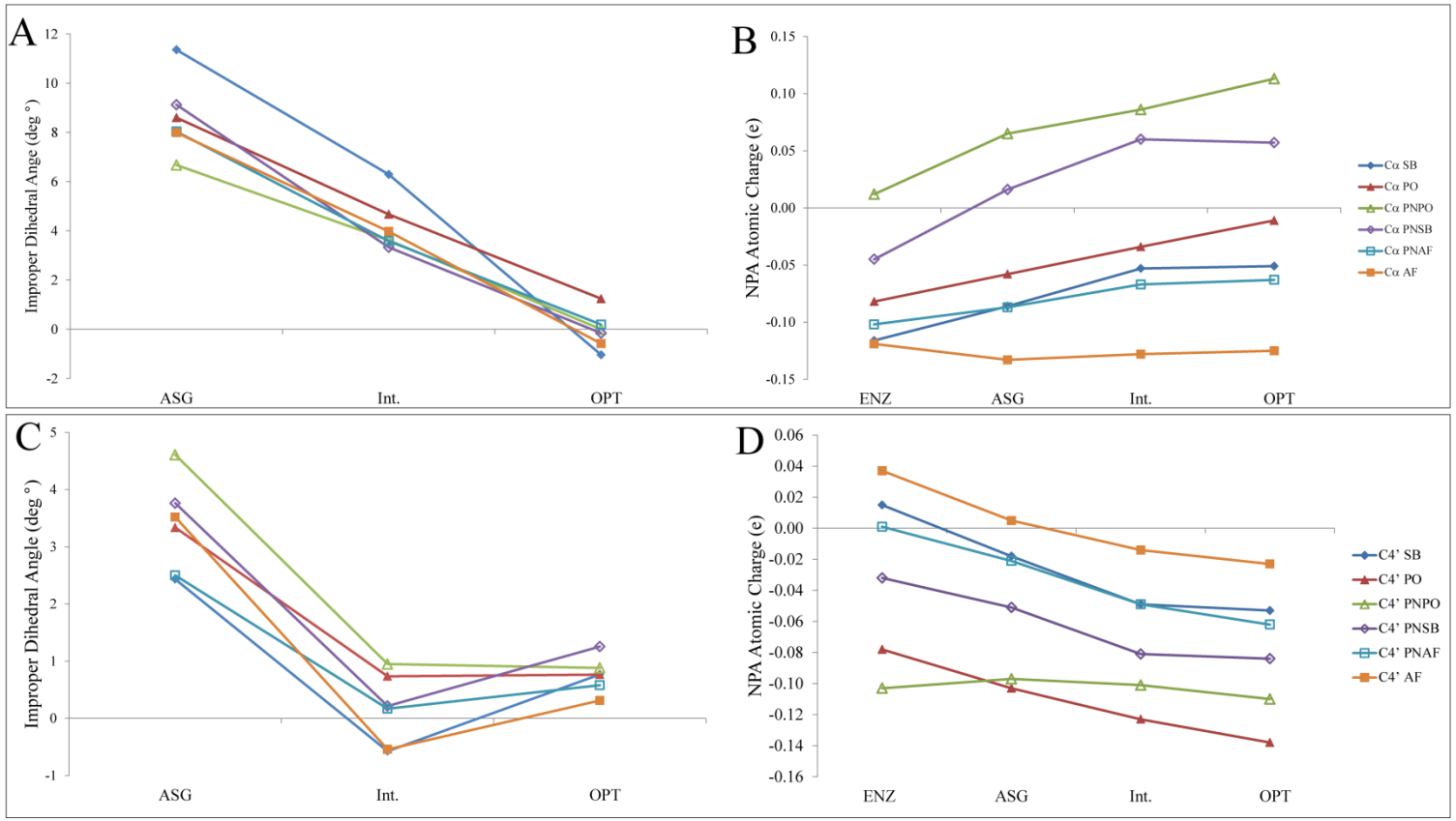


Figure 4.11. Variation in improper dihedral angle and corresponding NPA charge at the Cα, A and B, and C4', C and D, positions. Intermediate structures from the optimization were included in the plots of the improper dihedral and NPA charge. The charges computed for the PLP/substrate complex within the active-site cluster (ENZ) were added for comparison.

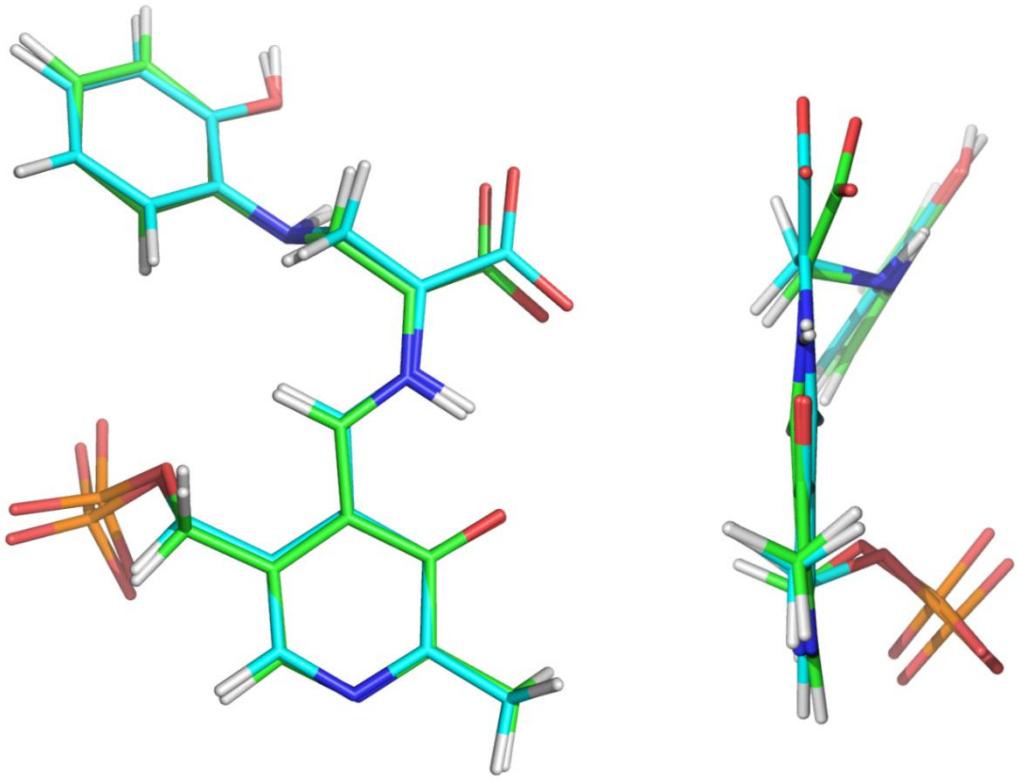


Figure 4.12. SB structures optimized within the active site cluster, ENZ/ASG, (green) and optimized in CPCM water, OPT, (cyan).

pronounced deviation from planarity is for the Cα carbons in the active site geometry (6°-12°) and in particular for the SB structure. The values for the C4' deviated about half as much (2°-5°)as those observed for Cα and the greatest deviation observed at Cα corresponded to the smallest deviation at C4' and vice versa indicating that at the extremes, at least, these are perhaps correlated. As a result of the optimizations, both the Cα and C4' became much more planar than the corresponding active site geometry. The improper dihedral measured at Cα in the X-ray crystal structure was approximately 2° to put the changes of the ENZ forms geometry optimized in the active site in perspective. In

the corresponding plots of the charges at these sites (the charges computed for ENZ structures were added to the figure for comparison) we see that a move towards planarity shifts the charge higher (towards more positive) for the Cα and lower for C4' which is unsurprising. It is also observed that, in general, the charge in the ENZ structures represent the lowest (most negative) for the Cα and highest (most positive) for the C4'. Two exceptions are the Cα in the AF and C4' in the PNPO. A discussion of these results is expanded upon with a focus on the mechanistic implications following the results of the charge delocalization calculations.

4.4.2 Protonation States and Charge Delocalization

The electron withdrawing properties of PLP's pyridine ring, enhanced when the pyridine ring nitrogen is protonated, and its ability to delocalize negative charge via its extended π network through the Schiff base linkage with amino acid substrates are widely known and appreciated. As others have pointed out, however, it is likely that its power as an electrophile is modulated through interactions with the protein as well as the tautomerism occurring between the phenolic oxygen and Schiff base nitrogen.[6, 17] Essentially, it may not be conducive to the mechanism for PLP to act in its maximal capacity as an electron sink.[17] The NPA charge study of the $E(Q_3)_{2AP}$ provides the opportunity to try and elucidate what effects the protonation states have on charge delocalization both within and outside the active site environment. The details regarding how delocalization is quantified were presented in the experimental section.

Figure 4.13 presents a plot of delocalized charge versus structure for the charges calculated in the presence of the active site (ENZ), from the extracted PLP/substrate

complex in CPCM water with the same geometry as the active site (ASG), and structures allowed to optimize in the presence of the CPCM water (OPT). The values are also provided in Table 4.6. The charges are divided between the pyridine and imine moieties as depicted in Figure 4.6. Overall the same general trend is observed between the ENZ, ASG, and OPT forms but with clear shifts for each structure computed in the different

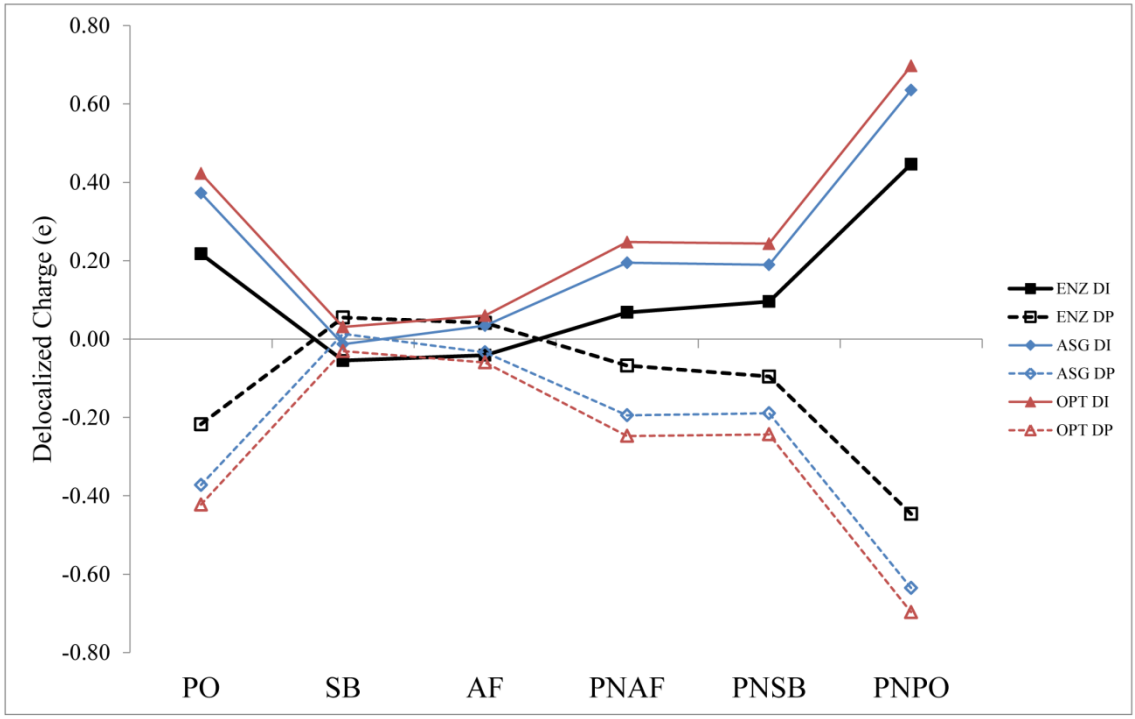


Figure 4.13. Delocalized charge between imine and pyridine moieties in the three combinations of environment and geometry.

environments/geometries. The behavior of the delocalization is especially noteworthy at the SB and AF structures. In the OPT forms the delocalized negative charge is always towards the pyridine ring and the least delocalization in charge is at the SB structure. In the ASG SB structure delocalization of negative charge switches towards the imine moiety slightly but switches back towards the pyridine in the AF structure. For the ENZ,

the delocalization of negative charge shifts more pronouncedly towards the imine in the SB structure than the ASG and the remains towards the imine in the AF as well before crossing back over in the PNAF where the delocalization of the negative charge is again towards the pyridine.

The electron withdrawing properties of the pyridine ring, when protonated, are clearly evident as they double the amount of delocalized charge onto it in moving from the PO to PNPO and SB to PNSB in the ENZ, ASG, and OPT whereas there is more variability in going from the AF to the PNAF but is approximately double or more. Also, the phenolic oxygen plays a significant role in delocalizing negative charge towards the pyridine ring as well with a higher ratio, 4.5, in going from SB to PO and PNSB to PNPO. Just as Casasnovas and coworkers observed, the most potent combination in delocalizing negative charge onto the pyridine ring is a protonated pyridine and phenolic oxygen.[6, 7] Maintaining a localized charge near Cα is aided by a protonated Schiff base nitrogen and deprotonated pyridine ring nitrogen as they found as well.[6, 7]

Table 4.6. Delocalized charge per environment/geometry combination in units of elementary charge, e.

Geometry	Moiety	PO	SB	AF	PNAF	PNSB	PNPO
ENZ	DI	0.22	-0.05	-0.04	0.07	0.10	0.45
	DP	-0.22	0.05	0.04	-0.07	-0.10	-0.45
ASG	DI	0.37	-0.01	0.03	0.19	0.19	0.64
	DP	-0.37	0.01	-0.03	-0.19	-0.19	-0.64
OPT	DI	0.42	0.03	0.06	0.25	0.24	0.70
	DP	-0.42	-0.03	-0.06	-0.25	-0.24	-0.70

4.4.3 Mechanistic Implications of Tautomerism in $E(Q_3)_{2AP}$

Based upon the relative values of partial charges calculated in this study with respect to the experimentally determined PO-SB tautomerism we would certainly conclude that the SB has the catalytic importance of localizing negative charge at Cα. This was the conclusion by Casasnovas and coworkers as well as the same general trends were observed, in fact the values for the OPT structures are essentially identical.[6, 7] What is significant about the findings presented here are the shifts in not only the delocalization between imine and pyridine but the corresponding values of the charges at Cα and C4' when these are calculated in the presence of the enzyme active site. Unlike the OPT forms, the ENZ forms in the SB and AF were capable of delocalizing negative charge on/to the imine. Another important difference noted was the fact that the relative charges at Cα and C4' were about equal in the PO (and in the PNSB) in the active site unlike the predictions for their values in the OPT form especially. It appears as though the protonated phenolic oxygen/unprotonated pyridine ring nitrogen combination in the PO of the active site is sufficient to "stabilize" the carbanion without making C4' a favorable target for protonation as in the OPT form. Stability is not conducive to reactivity and it is hypothesized that the proton transfer to the Schiff base nitrogen helps actually destabilize the quinonoid by concentrating the negative charge at Cα and we note that in the ENZ form a noticeable shift from sp^2 planarity occurs. This is not to make such a bold claim that the SB is the "catalytically active" tautomer in the $E(Q_3)$ but the SB may certainly be setting the stage.

The AF forms were also tested for their effects in light of the original predictions proposed for the $E(Q_3)_{indoline}$ intermediate and the inability to eliminate the possibility of an AF form in the $E(Q_3)_{2AP}$. The AF in the ENZ only enhances the difference in charges between the Cα and C4', a condition which is also observed for the ASG and OPT AF forms as well. Had the pyridine ring nitrogen been protonated, the PNAF would be predicted to be more catalytically important for localizing negative charge at Cα than the PNSB as Lai and coworkers concluded.[19] The PNPO would be the least favorable protonation state combination for protonation at Cα as Casasnovas and coworkers had previously found as well.

4.4.4 Conclusions

Our results in every environment/geometry tested would predict the same tautomers and pyridine protonation states as being more or less favorable for directing reprotonation either between the C4' or Cα in a relative sense to those tested in the study by Casasnovas et al.[6, 7] What this study has demonstrated, however, is that the enzyme active site is capable of modulating the electrophilic capability of the pyridine ring to the extent that the direction of delocalization can shift to the imine instead of the pyridine moiety. The observance of near identical values for partial charge at both Cα and C4' in PO in the ENZ form, as opposed to their divergence favoring negative charge buildup at C4' in the ASG and even more so in the OPT, demonstrates how the active site environment can stabilize the additional negative charge without favoring the reprotonation at the wrong carbon for the desired catalytic outcome.

4.5 References

1. Hayashi, H. Journal of Biochemistry, 1995. **118**(3): p. 463-473.

2. Schlenk, F. and E.E. Snell. Journal of Biological Chemistry, 1945. **157**(1): p. 425-426.

3. Jansonius, J.N. Current Opinion in Structural Biology, 1998. **8**(6): p. 759-769.

4. Toney, M.D. Biochemistry, 2001. **40**(5): p. 1378-1384.

5. Bach, R.D., C. Canepa, and M.N. Glukhovtsev. Journal of the American Chemical Society, 1999. **121**(28): p. 6542-6555.

6. Casasnovas, R., M. Adrover, J. Ortega-Castro, J. Frau, J. Donoso, and F. Munoz. Journal of Physical Chemistry B, 2012. **116**(35): p. 10665-10675.

7. Casasnovas, R., A. Salva, J. Frau, J. Donoso, and F. Munoz. Chemical Physics, 2009. **355**(2-3): p. 149-156.

8. Richard, J.P., T.L. Amyes, J. Crugeiras, and A. Rios. Current Opinion in Chemical Biology, 2009. **13**(4): p. 475-483.

9. Hyde, C.C., S.A. Ahmed, E.A. Padlan, E.W. Miles, and D.R. Davies. Journal of Biological Chemistry, 1988. **263**(33): p. 17857-17871.

10. Griswold, W.R. and M.D. Toney. Journal of the American Chemical Society, 2011. **133**(37): p. 14823-30.

11. Dunn, M.F. Archives of Biochemistry and Biophysics, 2012. **519**(2): p. 154-166.

12. Hartman, J.D., T.J. Neubauer, B.G. Caulkins, L.J. Mueller, and G.J.O. Beran. Journal of Biomolecular NMR, 2015. **62**(3): p. 327-340.

13. Metzler, D.E. and P. Christen, eds. *Transaminases*. Biochemistry. 1985, John Wiley & Sons: New York.

14. Metzler, D.E., M. Ikawa, and E.E. Snell. Journal of the American Chemical Society, 1954. **76**(3): p. 648-652.

15. Caulkins, B.G., B. Bastin, C. Yang, T.J. Neubauer, R.P. Young, E. Hilario, Y.M.M. Huang, C.E.A. Chang, L. Fan, M.F. Dunn, M.J. Marsella, and L.J. Mueller. Journal of the American Chemical Society, 2014. **136**(37): p. 12824-12827.

16. Caulkins, B.G., C. Yang, E. Hilario, L. Fan, M.F. Dunn, and L.J. Mueller. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2015. **1854**(9): p. 1194-1199.

17. Toney, M.D. Biochimica Et Biophysica Acta, 2011. **1814**(11): p. 1407-18.

18. Limbach, H.H., M. Chan-Huot, S. Sharif, P.M. Tolstoy, I.G. Shenderovich, G.S. Denisov, and M.D. Toney. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2011. **1814**(11): p. 1426-1437.

19. Lai, J.F., D. Niks, Y.C. Wang, T. Domratcheva, T.R.M. Barends, F. Schwarz, R.A. Olsen, D.W. Elliott, M.Q. Fatmi, C.E.A. Chang, I. Schlichting, M.F. Dunn, and L.J. Mueller. Journal of the American Chemical Society, 2011. **133**(1): p. 4-7.

20. Barends, T.R.M., T. Domratcheva, V. Kulik, L. Blumenstein, D. Niks, M.F. Dunn, and I. Schlichting. Chembiochem, 2008. **9**(7): p. 1024-1028.

21. Niks, D., E. Hilario, A. Dierkers, H. Ngo, D. Borchardt, T.J. Neubauer, L. Fan, L.J. Mueller, and M.F. Dunn. Biochemistry, 2013. **52**(37): p. 6396-6411.

22. Mennucci, B., M. Cossi, and J. Tomasi. Journal of Chemical Physics, 1995. **102**(17): p. 6837-6845.

23. Cossi, M., N. Rega, G. Scalmani, and V. Barone. Journal of Computational Chemistry, 2003. **24**(6): p. 669-681.

24. Breneman, C.M. and K.B. Wiberg. Journal of Computational Chemistry, 1990. **11**(3): p. 361-373.

25. Reed, A.E., R.B. Weinstock, and F. Weinhold. Journal of Chemical Physics, 1985. **83**(2): p. 735-746.

26. Jensen, F., *Introduction to Computational Chemistry*. 2 ed. 2007, West Sussex, England: John Wiley & Sons.

27. Bultinck, P., W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, and J.P. Tollenaere. Journal of Physical Chemistry A, 2002. **106**(34): p. 7887-7894.

28. Bultinck, P., W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, C. Van Alsenoy, and J.P. Tollenaere. Journal of Physical Chemistry A, 2002. **106**(34): p. 7895-7901.

29. Guerra, C.F., J.W. Handgraaf, E.J. Baerends, and F.M. Bickelhaupt. Journal of Computational Chemistry, 2004. **25**(2): p. 189-210.

30. Frisch, M.J., G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N.J. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.

31. Glendening, E.D., C.R. Landis, and F. Weinhold. Journal of Computational Chemistry, 2013. **34**(16): p. 1429-1437.

32. Becke, A.D.. Journal of Chemical Physics, 1993. **98**(2): p. 1372-1377.

33. Lee, C.T., W.T. Yang, and R.G. Parr. Physical Review B, 1988. **37**(2): p. 785-789.

34. Vosko, S.H., L. Wilk, and M. Nusair. Canadian Journal of Physics, 1980. **58**(8): p. 1200-1211.

35. Stephens, P.J., F.J. Devlin, C.F. Chabalowski, and M.J. Frisch. Journal of Physical Chemistry, 1994. **98**(45): p. 11623-11627.