UNIVERSITY OF CALIFORNIA
RIVERSIDE

Applying NMR Crystallography to Probe Key Intermediates of the Tryptophan Synthase Mechanism

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

Doctor of Philosophy

in

Chemistry

by

Thomas John Neubauer

December 2014

Dissertation Committee:
Dr. Leonard J. Mueller, Chairperson
Dr. Fancisco Zaera
Dr. Ludwig Bartels

The Dissertation of Thomas John Neubauer is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

I would like to thank my research advisor Professor Leonard Mueller. This research would have been unattainable without his leadership and motivation. I would also like to thank all of my lab mates who made my time during graduate school more enjoyable.

The Following Sections Contain Previously Published Materials.

Chapter II and III
Contains work published in:
J. Am. Chem. Soc., 2011, 133 (1), pp 4–7
DOI: 10.1021/ja106555c
Publication Date (Web): December 10, 2010

Jinfeng Lai †, Dimitri Niks ‡, Yachong Wang †, Tatiana Domratcheva §, Thomas R. M. Barends §, Friedrich Schwarz §,Ryan A. Olsen †, Douglas W. Elliott †, M. Qaiser Fatmi †, Chia-en A. Chang †, Ilme Schlichting §, Michael F. Dunn *‡, andLeonard J. Mueller *†

Chapter IV and IV
Contains work published in:
J. Am. Chem. Soc., 2014, 136 (37), pp 12824–12827
DOI: 10.1021/ja506267d
Publication Date (Web): August 22, 2014

Bethany G. Caulkins †, Baback Bastin †,Chen Yang †, Thomas J. Neubauer †,Robert P. Young †, Eduardo Hilario ‡, Yu-ming M. Huang †, Chia-en A. Chang †, Li Fan ‡, Michael F. Dunn ‡, Michael J. Marsella †, and Leonard J. Mueller *†

ABSTRACT OF THE DISSERTATION

Applying NMR Crystallography to Probe Key Intermediates of the Tryptophan Synthase Mechanism

by

Thomas John Neubauer

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, December 2014
Dr. Leonard Mueller, Chairperson

Protonation and hybridization state are critical phenomena at the chemical-level that are vital in understanding an enzymes mechanism and function. Using high resolution X-ray crystallography unaccompanied is not enough to present a complete account of these details. On the other hand, the chemical shift from solid-state NMR spectroscopy is a remarkably insightful probe of the surrounding electro-chemical environment. The synergistic combination of NMR spectroscopy, X-ray crystallography and ab initio computational chemistry, also referred to as "NMR Crystallography", can be used as a unmatched tool for elucidating high resolution three-dimensional structures of diverse materials. This synergistic approach was used to determine the structures of the indoline-, 2-aminophenol (2AP)- quinonoid, and internal aldimine intermediates in the pyridoxal-5'-phosphate-dependent enzyme tryptophan synthase under conditions of active catalysis. In the presence of side-chain residues fixed at their crystallographically determined coordinates, models of the active site were created and optimized using ab initio computational chemistry from which chemical shifts were calculated. Experimentally measured chemical shifts at particular ^{13}C - and ^{15}N-labeled positions on the

substrate are used to exclusively distinguish and emerge comparable species from the various computational model types; dictated by protonation state of the substrate analogues and nearby catalytic residues. The reduced χ^2 (chi squared) analysis suggests the phenolic / phenolic-acid form to be the predominate protonated species for both the indoline and 2-aminophenol quinonoid intermediates, while the internal aldimine proves to be in the protonated Schiff base form.

TABLE OF CONTENTS

page

LIST OF FIGURES

page

CHAPTER IV

CHAPTER V

CHAPTER VI

LIST OF TABLES

page

CHAPTER I. INTRODUCTION.

Since their inception, both nuclear magnetic resonance (NMR) and *ab initio* calculations have been used to solve important structural and energetic problems in the field of chemistry. As each technique has advanced, there has been an increase in the wealth of information available. The progress in the field of NMR includes more sophisticated spectrometers and more powerful multi-dimensional pulse sequences that give far more detailed information on through-bond and through-space connectivities. Computational chemistry has increased in power due to the continued progression of the modern computer as well as the development of new computational techniques. Alone, each one of these techniques has become an important tool for the chemical research community yet the synergistic combination of these techniques is even more powerful.

NMR, while being one of the most important analytical tools available to chemists, contains a wealth of information beyond the typical spectral interpretation based on empirical rules. The chemical shift observed for a nucleus is a detailed reporter on the local electronic structure. The coupling observed between two nuclei can be used to report either on the distance (in the case of dipolar coupling) or the nature of the bonding (in the case of a J coupling). Yet the detailed analysis of these spectral parameters is hindered by the ability to convert information back to the molecular frame of reference, which often leads to incomplete utilization of the experimental data. The bridge between the spectral parameters of the spin Hamiltonian and the structural parameters of the molecular Hamiltonian is computational chemistry. Rather than simply knowing that the chemical shift reflects a broad chemical environment, one can define the geometry that fragment adopts.

Beyond the chemical shift, NMR is also capable of giving very accurate information on the energetics of processes such as bond rotations. Knowing the energetics of a process from dynamic NMR experiments, we can use *ab initio* theory to tell us details regarding the process. In this way, detailed molecular models can be connected to the energetic components that make up a rotational pathway. Again, the combination of NMR and *ab initio* has the ability to lead to far more information than each individually can contribute.

While the computer and the spectrometer are important physical tools to study problems in chemistry, to understand the link between the information they provide it is essential that we understand the origins of Hartree Fock theory and the chemical shift. This chapter begins with a review of Hartree Fock and *ab initio* computational theory, including the calculation of the chemical shift. The goal is not to be overly pedantic, but to discuss issues important to the practitioner and serve as a guide for others to use when tackling these sorts of problems. There are many potential complications when considering accurate use of *ab initio* theory and the best approach often depends on the type of properties that one is interested in. One of the most important choices that users must make is the nature of the approximation to be employed. Unfortunately, even in the days of gigaflop personal computing, there are still problems that even the largest supercomputer cannot solve. Because of this approximations must often be made and this requires that the user understand the system of study at a level that can dictate the type of calculation to be performed.

The purpose of this section is not the development of new computational procedures but rather the description of variables and pitfalls one runs into during the use of an *ab initio* computational package, focusing on the choice of basis functions and electron correlation. These two items represent the variable one will run into when performing a calculation. For a

more detailed description of *ab initio* theory please refer to such text as [1-4] where much of this section is derived from.

OVERVIEW OF *AB INITIO* THEORY

From quantum mechanics we know the energy of a system is defined by the time dependent Schrödinger equation.

$$H\Psi = E\Psi$$

The Hamiltonian for a molecular system made up of n electron and N nuclei can be partitioned into several terms.

$$H = T_e + T_n + V_{en} + V_{ee} + V_{nn}$$

Here, the wavefunction spans both spin and spatial coordinates of the nuclei and electrons where T_e is the kinetic energy of the electron, T_n is the kinetic energy of the nuclei, V_{en} is the potential energy from electron-nucleus interaction, V_{ee} is the potential energy form electron-electron interaction, and V_{nn} is the potential energy form nucleus-nucleus interaction.

The equation for the kinetic energy of the electron can be written as

$$T_e = -\frac{h^2}{8\pi^2 m}\sum_{i=1}^{n}\nabla_i^2$$

where m is the electron mass, h is Plank's constant, and ∇ is the Laplacian

$$\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}.$$

Likewise, the kinetic energy operator for nuclear coordinates is written as

$$T_n = -\frac{h^2}{8\pi^2}\sum_{v=1}^{N}\frac{1}{M_v}\nabla_v^2$$

where M_v is the mass of nucleus *v*. The potential energy terms in the Hamiltonian are

$$V_{en} = -\sum_{i=1}^{n}\sum_{v=1}^{N} Z_v \frac{e^2}{r_{iv}}$$

where Z_v is atomic charge of nucleus v, e is the electron charge, r_{iv} is the distance from electron i and nucleus v and

$$V_{ee} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \frac{e^2}{r_{ij}}$$

where r_{ij} is the distance from electron i to electron j

$$V_{en} = \frac{1}{2}\sum_{v=1}^{N}\sum_{u\neq v}^{N} Z_v Z_u \frac{e^2}{R_{vu}}$$

This Hamiltonian equation neglects spin interaction and is exactly solvable for a two particle system (i.e. one electron and one nucleus). For a higher order system it is necessary to invoke certain approximations to simplify the equations.

The first is the Born Oppenheimer approximation, where the motions of electrons are separated from the motion of the nuclei due to large difference in the velocity of the electron relative to the nuclei. In this approximation the electron sees some static effective force from the fixed nuclei. On making this assumption we are able to separate the wavefunction into the product of the electronic and nuclear parts

$$\Psi(r,R) = \Psi_R^e(r)\cdot\Psi^n(R)$$

and we concentrate on the electronic wavefunction which satisfies

$$H^e\Psi_R^e(r) = E_R^e\Psi^e(r)$$

with the electronic Hamiltonian made up of the following

$$H^e = T_e + V_{en} + V_{ee}$$

The energy is now an effective energy relative to some nuclear coordinates. Unfortunately this reduced Schrödinger equation is still unsolvable for chemical systems without further approximations. The next approximation is that the electronic wavefunction (made up of n electrons) can be represented in a orbital approximation where

$$\Psi(1,2,3...n) = \Psi_1(1)\Psi_2(2)\Psi_3(3)...\Psi_n(n)$$

In this approximation we assume that the electrons act independently of each other. The spin of the electron must also be taken into account. The one electron wavefunction can be written in the following manner

$$\Psi_i = (x, y, z, \xi) = \phi_i(x, y, z)\eta(\xi)$$

where the wavefunction is broken into a spatial component as well as a spin component. The spin term can have the value of α or β corresponding to the spin up or spin down. In this picture, an orbital is made up of two electrons one with a spin α and one with spin β.

Due to the Pauli principle, the wavefunction is required to be antisymmetric under exchange of electrons.

$$\Psi(1,2,3,...n) = -\Psi(2,1,3,...n)$$

This is not true of a wavefunction represented as a simple product of orbitals. To ensure the antisymmetry, we represent the wavefunction as a Slater determinant built over the orbitals.

$$\Psi(1,2,..n) = \frac{1}{\sqrt{n!}}\begin{vmatrix} \Psi_1(1) & \Psi_1(2) & \cdots & \Psi_1(n) \\ \Psi_2(1) & \Psi_2(2) & \cdots & \Psi_2(n) \\ \vdots & \vdots & \vdots & \vdots \\ \Psi_n(1) & \Psi_n(2) & \cdots & \Psi_n(n) \end{vmatrix}$$

with $\frac{1}{\sqrt{n!}}$ as the normalization factor guaranteeing that

$$\int \Psi^*(1,2,..n)\Psi(1,2,...n)d\tau_1 d\tau_2 d\tau_n = 1$$

with $d\tau = dxdydzd\xi$

We now write the wavefunction for the simplest case of a closed shell system where every orbital is made up of a pair of electrons with different spins.

$$\Psi(1,2,..n) = \frac{1}{\sqrt{2n!}}\begin{vmatrix} \phi_1(1)\alpha(1) & \phi_1(2)\alpha(2) & \cdots & \phi_1(2n)\alpha(2n) \\ \phi_1(1)\beta(1) & \phi_1(2)\beta(2) & & \phi_1(2n)\beta(2n) \\ \phi_2(1)\alpha(1) & \phi_2(2)\alpha(2) & \cdots & \phi_2(2n)\alpha(2n) \\ \phi_2(1)\beta(1) & \phi_2(2)\beta(2) & & \phi_2(2n)\beta(2n) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_n(1)\alpha(1) & \phi_n(2)\alpha(2) & \cdots & \phi_n(2n)\alpha(2n) \\ \phi_n(1)\beta(1) & \phi_n(2)\beta(2) & & \phi_n(2n)\beta(2n) \end{vmatrix}$$

and use the variational principle to assign the optimal molecular orbital. From the variational principle we know that the energy of a system given by an approximate wavefunction is higher than the actual energy. By minimizing the energy

$$E_{approx} = \frac{\int \psi^*_{approx} H \psi_{approx}}{\int \psi^*_{approx} \psi_{approx}}$$

where ψ_{approx} is the approximate wavefunction made by some initial guess, we can in an iterative manner converge at the actual energy due to the fact that

$$E_{approx} \geq E$$

At a simplistic level this is the procedure that is followed in an *ab initio* calculation, where the first step is some guess at an initial wavefunction followed by the evaluation of the energy for

that wavefunction. This procedure is iterated in an attempt to minimize the energy with the actual energy being the lowest possible value of the approximate energy.

In practice this procedure is complicated. To start, the above description is written in terms of the Hartree Fock one electron equations

$$F(1)\phi_i(1) = \varepsilon_i\phi_i(1)$$

where we have now written the individual orbital energies ε_i instead of the overall energy. This is directly analogous to the standard Schrödinger equation with the overall wavefunction replaced by the individual orbital wavefunction and our Hamiltonian is now the Fock operator. The Fock operator for electron 1 is given by the following equation

$$F(1) = h(1) + \sum_{j=1}^{n}\left[2J_j(1) - K_j(1)\right]$$

with

$$h(1) = \frac{h^2}{8\pi^2 m}\nabla_1^2 - \sum_{v=1}^{N}\frac{Z_v e^2}{r_{1v}}$$

$$J_j(1)\phi_i(1) = \left[\int \phi_j(2)\frac{e^2}{r_{12}}\phi_j(2)dv_2\right]\phi_i(1)$$

$$K_j(1)\phi_i(1) = \left[\int \phi_j(2)\frac{e^2}{r_{12}}\phi_i(2)dv_2\right]\phi_i(1)$$

J represents the Coulombic operator (the inter-atomic repulsion) and K represents the spatial exchange operator. It is important to note that each electron in this model moves in an average field created by the other electrons. This is not always a valid approximation and improvement to this simplification include multi-reference wavefunction techniques which can be studied in greater detail in the following reference [5].

The solution of the Hartree Fock equations is not simple and is accomplished using the self consistent field method. In this method one starts with a initial guess of a trial orbital ϕ_i which is used to calculate the exchange and Coulomb operator integrals. Next, the Hartree Fock equations are solved giving rise to a new set of molecular orbitals. This procedure is repeated until a convergence criterion is met.

In practice the molecular orbital is written as a linear combination of pre-defined one electron functions known as basis functions.

$$\phi_i = \sum_{\mu=1}^{N} c_{\mu i} \chi_\mu$$

Where $\chi_1 \cdots \chi_N$ are the basis functions of interest and $c_{\mu i}$ represents the molecular orbital expansion coefficients. The latter are the coefficients that are variationally determined. In the molecular orbital picture this is know as the linear combination of atomic orbitals.

The Hartree Fock procedure as implemented in most all commercially available software packages is somewhat of a black box with the exception of one thing, the choice of a basis function. It is important to take a step back and realize that the basis function is part of *ab initio* procedure that we as user must specify. The size of a basis set, or the number of mathematical functions used to describe the molecular orbital, increases the complexity of the problem. At the same time the larger the basis function, the more accurate of a representation of the energy of the system. Part of the art of computational chemistry is knowing when a certain basis will give acceptable results. Basis sets are typically constructed of a combination of Gaussian functions. This is mainly due to the ease that Gaussian functions can be manipulated mathematically. Gaussian functions alone do not represent electron density accurately. For this reason they are often combined as a sum of multiple Gaussian functions. For instance the early

used STO-3G basis set is described as a Slater Type Orbital (STO) represented by three gaussians (3G) with each Gaussian given a separate variational weighting to facilitate the representation of the electron density. There are a great number of papers and books dealing with the construction of a basis set [6, 7].

The problem is to now find the coefficients c_{ui} for a give basis. Hall and Roothan showed that the coefficients are the solutions of the following

$$\sum_{v=1}^{m}(F_{\mu v}-\varepsilon_i S_{\mu v})c_{vi}$$

with

$$F_{\mu v}=H_{\mu v}+G_{\mu v}$$

$$H_{\mu v}=\int \chi_{\mu}(1)h(1)\chi_{v}(1)dv_1$$

$$G_{\mu v}=\sum_{\lambda p}P_{\lambda p}\left[\langle \mu v|\lambda p\rangle-\frac{1}{2}\langle \mu\lambda|vp\rangle\right]$$

$$S_{\mu v}=\int \chi_{\mu}(1)\chi_{v}(1)dv_1$$

where ε_i represents the energy of orbital *i*, $P_{\lambda p}$ is described as the first order density matrix

$$P_{\lambda p}=2\sum_{i=1}^{n}c_{\lambda i}c_{pi}$$

and $\langle \mu v|\lambda p\rangle$ is the two electron integral

$$\langle \mu v|\lambda p\rangle=\int\int \chi_{\mu}(1)\chi_{v}(1)\frac{e^2}{r_{12}}\chi_{\lambda}(2)\chi_{p}(2)dv_1dv_2$$

In a matrix representation

$$FC=SCE$$

with F being an $m \times m$ matrix with elements of $F_{\mu\nu}$, C an $m \times m$ matrix with elements $c_{\mu i}$, S and $m \times m$ matrix with elements $S_{\mu\nu}$, and E a diagonal matrix with elements ε_i.

Importantly, the solution of the equation gives rise to a set of doubly occupied molecular orbitals, and also a set of orbitals which contain no electrons which are denoted virtual orbitals. It also turns out that the two electron integrals put a large restriction on the size of the calculation that can be accomplished and give rise to the familiar m^4 (basis functions) scaling rules of standard HF calculations.

While the Hartree Fock procedure in many cases gives good results it is important to note that this is not always guaranteed and in other cases there may be different types of calculations that give better results. Often when a HF calculation fails it is due to neglect of electron correlation. The Hartree Fock wavefunctions exclude the population of electrons into low lying excited state orbitals. Most of the treatments of electron correlation come from the use of the virtual orbitals that are left over in a Hartree Fock calculation and ideally would represent the wavefunction as

$$\Psi(1,2,....n) = \sum_I C_I \Psi_I(1,2,....n)$$

Where rather than having one ground wavefunction, there now is a series of wavefunctions where electrons are systematically promoted from their ground state to an excited state. Calculations such as CIS (configuration interaction with single excitations) do just this, by removing an electron from the ground state and placing it in a virtual orbital. Likewise CISD allows for both on electron promotion as well as two electron promotion. Other types of calculations in this family include CCSD (coupled cluster single and double excitation) and BD (Brueckner double excitation) as well as others. These types of procedure come with large

computational overhead. Essentially, every excitation into a virtual state creates a separate wavefunction in the Hartree Fock equation. These types of procedures are typically limited to smaller amounts of basis functions due to this overhead and unfortunately these types of calculations converge quite slowly with the size of their basis functions.

Another popular way to deal with the problem of electron correlation is by the use of Moller-Plesset perturbation theory where excited states are added to the wavefunction as perturbation on the ground state wavefunction. Assuming we have a wavefunction that can be broken into a solvable part and a perturbation to it such that

$$H = H^0 + \lambda V$$

where the perturbation potential is given by

$$V = \sum_{i>j} \frac{1}{r_{ij}} - \frac{1}{2} \sum_i J_i - K_i$$

After expanding out this equation and utilizing perturbation theory we find that the correction to the ground state energy is given by the following

$$E^2 = -\sum_t \frac{\left| \left\langle \Psi^0 \middle| V \middle| \Psi_t \right\rangle \right|^2}{E_t - E_0}$$

Since this equation has truncated the expansion at second order it is commonly abbreviated as the MP2 level. It is possible to also have higher order MP corrections. These techniques are not as computationally extensive compared to the CISD types of electron correlation but are still fairly restrictive in their computational requirements.

The final way to handle electron correlation comes from Density Functional Theory. In density functional theory the energy of a molecule is represented as

$$E = E^T + E^V + E^J + E^{XC}$$

Where E^T represents the kinetic energy term, E^V represents the electron nudei attraction and the nudear repulsion, and E^J represents the electron-electron repulsion. All of these terms are represented in terms of electron density. The final term E^{XC} is distinct in that it represents both the exchange and electron correlation. This includes the effect of the exchange energy from the antisymmetry of the wavefunction as well as dynamic correlation in the motion of the electrons. It is important to note at this point that there are many levels of how these exchange and correlation terms are treated but that they are always treated with some analytical equation and often they contain terms that are optimized for certain standard systems. The choice of what functional to use is important to achieve accurate results. More often utilized approaches are the hybrid density functional which utilize a combination of both the Hartree Fock exchange energy and the density functional exchange correlation energy. One will typically run into these types (B3LYP, BPW91, etc) when reading the literature and one should realize that they are a combination of energies from many different functionals. For instance the exchange correlation energy of the popular B3LYP is actually

$$E_{B3LYP}^{XC} = E_{LDA}^{X} + c_0(E_{HF}^{X} - E_{LDA}^{X}) + c_X(E_{HF}^{X} - E_{B88}^{X}) + E_{VWN3}^{C} + c_C(E_{LYP}^{C} - E_{VWN3}^{C})$$

where $E_{Functional}^{X}$ refers to the exchange energy of that functional and $E_{Functional}^{C}$ refers to correlation energy of the functional and the coefficients c_0, c_X ,and c_C are determined through a comparison of calculated energies to that of experimentally determined values.

Density Functional Theory allows for inclusion of electron correlation without the usual computational expense of the true correlation methods. The drawback being that most of the commonly used functionals are derived from comparison to experimental values so that it is

possible in your calculation the derived coefficients may be incorrect. Despite this shortcoming, it has show to be a very effective computational tool.

Ultimately, the proper choice of level of theory depends on the problem to be solved and in particular the type on the type of bonding involved. In some cases, Hartree Fock theory is too simple of a model to accurately describe a system. In other cases the use of high level electron correlation is unnecessary. Likewise, the choice of basis set requires the discretion on the part of the user. In some cases a small basis set will suffice for accurate results and in others it will not. At the same time, the use of too large of a basis will lead to a substantially longer calculation with no real improvement in accuracy of the results. Carefully weighing these considerations allows one to extend accurate calculations to increasingly complicated (often larger or more interesting) chemical systems.

OTHER COMPUTATION CONSIDERATIONS

In addition to level of theory and basis set, there are more things to consider when performing a calculation, particularly when trying to match calculated properties to some experimental values. An important aspect is the state of the molecule. *Ab initio* theory at its basic level is a gas phase, single molecule calculation. In the real world, we know this approximation to have many shortcomings. If the molecule we are interested in is a solid, we need to consider the molecules that create the framework that surrounds it. If the molecule is a liquid or in a solution it is often important to consider the interactions that are created by the phase surrounding it.

For example, constraints of a solid framework are extremely important to consider when performing an *ab initio* calculation. From the standpoint of a molecule's geometry, the

solid framework is a scaffold that it exists inside of. In many cases the solid state structure is not that which would be calculated for the gas phase. Solid state *ab initio* calculations usually require another set of approximations due to the large size of the problem when considering the effects of multiple molecules. A standard computational program would quickly run out of resources when applying standard techniques to the calculation This problem was made tractable by use of a plane wave basis set which is an oscillatory expansion of the basis functions through space. Rather than having one basis function centered on a molecule, the basis function is extended throughout the solid in a repeating manner. An additional approximation was the use of a pseuodopotential, which freezes the core electrons and only allows the valence orbitals and electrons to be optimized.

As with solid state calculations, it is often important to include the effects that a solvent has on the molecular wavefunction. These techniques range from explicit consideration of solvent molecule to approximations that include the dielectric effect on the molecule. These types of calculations are often important in the accurate determination of such properties as chemical shifts and rotational barriers in the presence of solvents.

For a deeper discussion on other pertinent topics encountered in computational chemistry please refer to [8]. Know we will look at the core of the NMR experiment, the chemical shift. Like the overview of *ab initio* theory, the next section is in part derived from standard references [9, 10].

OVERVIEW OF THE CHEMICAL SHIFT

A spin active nucleus (for the simplest case, spin ½) has an intrinsic magnetic moment that can interact with an external magnetic field with a Hamiltonian

$$H = -\gamma I \cdot B$$

where γ is the gyromagnetic ration of the nuclei and B is the external magnetic field. Transitions between the spin states can be induced by applying a magnetic field perpendicular to the static field at the characteristic frequency of that nucleus. The exact frequency at which the transition occurs in molecules depends on the chemical environment as the electrons around a nucleus have the ability to perturb or "shield" the magnetic field that is observed at the nucleus. This is the most important detail and one that gives rise to the use of NMR as a powerful tool in molecular characterization. The Hamiltonian for the shielding is written as

$$H = -\gamma I_i \cdot \sigma_{ij} \cdot B_j$$

where $\sigma_{ij} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$, the matrix representation of the interaction of the magnetic field with the nuclear spin mediated by the electrons surrounding the nucleus. σ is referred to as the chemical shift. This equation also makes clear the relationship between energy and the chemical shift

$$\sigma_{ij} = (\frac{1}{\gamma\hbar})\frac{\partial^2 E}{\partial B_i \partial I_j}$$

It is often more insightful to partition the chemical shift into two distinct parts

$$\sigma_{ij} = \sigma_{ij}^{\ D} + \sigma_{ij}^{\ P}$$

where $\sigma_{ij}^{\ D}$ is the diamagnetic shielding tensor and $\sigma_{ij}^{\ P}$ is the paramagnetic shielding tensor. The meaning of this separation can be seen from a perturbative theory calculation of the

chemical shift. We first start with the equation for the conjugate momentum of an electron in the presence of a magnetic field

$$p = (i\hbar)\nabla - (e/c)A$$

where ∇ is the differential operator and

$$B = \nabla \times A$$

relating the magnetic field B to the vector potential A. The Hamiltonian for a diatomic molecule with one electron in the presence of a magnetic field is given by the following

$$H - V = (1/2m_e)\left[-i\hbar\nabla - (e/c)A\right]^2$$

In the presence of a nuclear magnetic moment the vector potential becomes the sum of two vector potentials given by

$$A = A^0 + A^n$$

where we now have a term A^0 for the external field and a term A^n for the vector potential from the magnetic moment. Substituting back into the original Hamiltonian we arrive at the following equation

$$H - V = (\Pi^2/2m_e) - (e/2m_e c)\left(\Pi \cdot A^n + A^n \cdot \Pi\right) + (e^2/2m_e c^2)A^{n\,2}$$

where we have defined $\Pi = -i\hbar\nabla - (e/c)A^0$ For perturbation theory we define the first order Hamiltonian

$$H^{(1)} = (e^2/2m_e c^2)(A^0 \cdot A^n + A^n \cdot A^0)$$

We have ignored the second term in the above equation due to the fact that it is independent of the applied field. We now have a term for the first order correction to the energy

$$E^{(1)} = (e^2/2m_e c^2)\langle\psi_0|A^0 \cdot A^n + A^n \cdot A^0|\psi_0\rangle$$

By substituting for both magnetic vector potentials $A^n(r) = (\mu^n \times r^n)/r^{n^3}$ and $A^0 = \frac{1}{2}(B^0 \times r^0)$ we are left with the following

$$E^{(1)} = (e^2/2m_e c^2)\mu_n \langle \psi_0 | r^n \cdot r^0 / r^3{}_n | \psi_0 \rangle B_0$$

which represents the term for the diamagnetic shielding of the nucleus by the electrons. This equation contains terms for both the distance from the nucleus to the electron r^n as well as the distance from the magnetic field to the electron r^0. It is important to note that this term r^0 gives rise to a difficulty in computational chemistry in that the origin (gauge origin) of the magnetic field needs to be chosen. More specifically it needs to be chosen in a way that is invariant to the orientation of the molecule. Many techniques have been developed but by far the most popular and effective is the GIAO approach.

From our Hamiltonian two other terms are also of importance.

$M = (e/m_e c)(i\hbar)(\nabla \cdot A^n)$ and

$N = (e/m_e c)(i\hbar)(\nabla \cdot A^0)$

These two terms give rise to the second order perturbation correction to energy given by the following

$$E^{(2)} = (e^2/2m_e^2 c^2) \times \sum_{q>0} \left[\langle \psi_0 | M | \psi_q \rangle \langle \psi_q | N | \psi_0 \rangle / (E_0 - E_q) \right]$$

After substituting for the vector potential we arrive at

$$E^{(2)} = (e^2/2m^2{}_e c^2) \times \mu_n \sum_{q>0} \left[\frac{\langle \psi_0 | \frac{L_n}{r^3} | \psi_q \rangle \langle \psi_q | L_0 | \psi_0 \rangle}{(E_0 - E_q)} \right] \cdot B$$

where we have defined the new terms $L_0 = (i\hbar)(r^0 \times \nabla)$ and $L_n = (i\hbar)(r^n \times \nabla)$. This term represents the paramagnetic contribution to the chemical shift. It is interesting to note that the diamagnetic term is only dependent on the ground state wavefunction. This makes the accurate calculation of the diamagnetic term much easier. By comparison, we see that the paramagnetic term incorporates both ground and excited state. As we have seen from the *ab initio* description, a standard Hartree Fock procedure does not accurately calculate these types of terms and very often DFT is employed. Also excited states need much larger basis sets to accurately be described.

Another NMR observable that has direct ties to the electronic structure of a molecule is the spin-spin coupling. This term is the through bond coupling that allows the transfer of magnetization from one nucleus to another. Without going into the detail as shown for the chemical shift, we can simply say that the Hamiltonian is given by the following

$$H = \gamma_i \gamma_j \hbar^2 I_i J_{ij} I_j$$

where we now have a coupling term between two magnetic moments. It is common to use the spin-spin coupling (J coupling) as experimental measure of direct bonding. Computationally it has the ability to relate back to geometric effects in the molecule (Karplus equations). Later in this work we will see how the J coupling can be used to create two dimensional NMR experiments that allow structural information to be established.

As with the calculation of the energy of a molecule using *ab initio* techniques, the calculation of the chemical shift requires the user to choose the level of theory as well as the basis set. All of the same considerations apply to the accurate calculation of a chemical shift. The basis set must be of adequate size to accurately represent the electronic structure. The

computational level of theory also must be sufficient to give a detailed representation of the electronic structure that surrounds the nucleus. In many ways the chemical shift requires a more elaborate representation of the electronic structure to accurately represent the chemical shift. It is this connection that we often employ to check models of chemical structure versus experimental NMR data.

REFERENCES

(1) G. Grant, W. Richards, *Computational Chemistry*, Oxford, New York, 1994.

(2) A. Hinchliffe, *Modelling Molecular Structure*, John Wiley and Sons, New York, 1994.

(3) P. W. Atkins, R. S. Freidman, *Molecular Quantum Mechanics*, 3rd ed., Oxford, New York, 1997.

(4) A. Szabo, N. Osterlund, *Modern Quantum Chemistry*, Dover, Mineola, 1996.

(5) M. W. Schmidt, M. S. Gordon, *Annu. Rev. Phys. Chem.* 1998, *49*, 233.

(6) E. R. Davidson, D. Feller, *Chem. Rev.* 1986, *86*, 681.

(7) S. Huzinaga, J. Andzeim, M. Klobukowski, E. Radzio-Andzeim, Y. Sakai, H. Tatewaki, *Gaussian Basis Sets for Molecular Calculations*, Elsevier, Amsterdam, 1984.

(8) J. Foresman, A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, 2nd ed., Gaussian Inc., Pittsburg, 1996.

(9) T. Helgaker, M. Jaszunski, K. Ruud, *Chem. Rev.* 1999, *999*, 293.

(10) D. E. O'Reilley, *Prog. Nucl. Magn Reson. Spectrosc.* 1967, *2*, 1.

CHAPTER II. X-RAY AND NMR CRYSTALLOGRAPHY IN AN ENZYME ACTIVE SITE: THE INDOLINE AND 2AP QUINONOID INTERMEDIATE IN TRYPTOPHAN SYNTHASE.

ABSTRACT

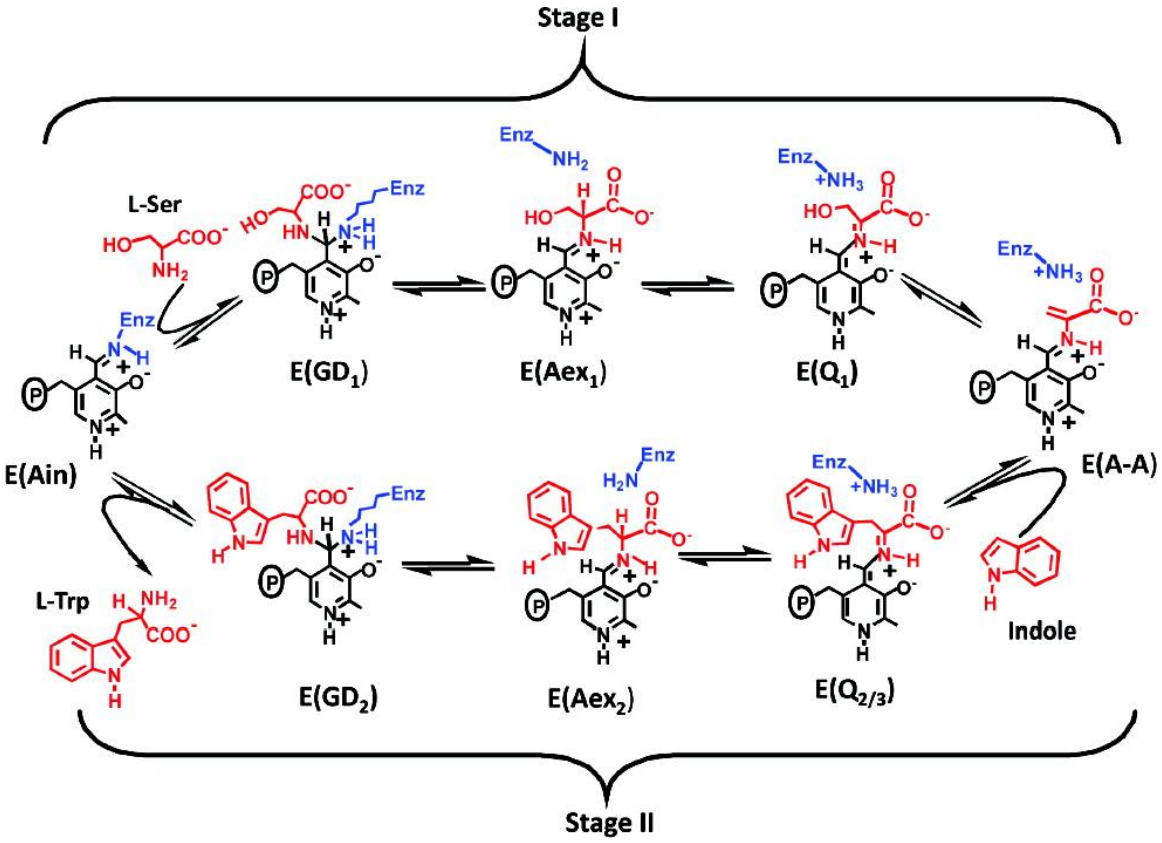
Protonation and hybridization state are critical phenomena at the chemical-level that are vital in understanding an enzymes mechanism and function. Using high resolution X-ray crystallography unaccompanied is not enough to present a complete account of these details. On the other hand, the chemical shift from solid-state NMR spectroscopy is a remarkably insightful probe of the surrounding electro-chemical environment. The synergistic combination of NMR spectroscopy, X-ray crystallography and ab initio computational chemistry, also referred to as "NMR Crystallography", can be used as a unmatched tool for elucidating high resolution three-dimensional structures of diverse materials. This synergistic approach was used to determine the structures of the indoline-, 2-aminophenol (2AP)- quinonoid, and internal aldimine intermediates in the pyridoxal-5'-phosphate-dependent enzyme tryptophan synthase under conditions of active catalysis. In the presence of side-chain residues fixed at their crystallographically determined coordinates, models of the active site were created and optimized using ab initio computational chemistry from which chemical shifts were calculated. Experimentally measured chemical shifts at particular ^{13}C - and ^{15}N-labeled positions on the substrate are used to exclusively distinguish and emerge comparable species from the various computational model types; dictated by protonation state of the substrate analogues and nearby catalytic residues. The reduced χ^2 (chi squared) analysis suggests the phenolic / phenolic-acid form to be the predominate protonated species for both the indoline and 2-aminophenol quinonoid intermediates, while the internal aldimine proves to be in the protonated Schiff base form.

INTRODUCTION

The $\alpha_2\beta_2$ bienzyme complex Tryptophan Synthase catalyzes the final two steps of L-Trp,[1,2] via its pyridoxal-5′-phosphate (PLP, vitamin B_6) cofactor which involves the channeling of indole between the active sites of the α- and β-subunits.[2-5] Long-range allosterics during the catalytic cycle, open enzyme, low activity and closed enzyme, high activity[3-6]control the channeling of indole. In the α-site, 3-indole-D-glycerol-3′-phosphate (IGP) is cleaved to D-glyceraldehyde-3-phosphate and indole. In stage I of the β-reaction (Figure 1.), L-Ser reacts with the internal aldimine [E(Ain)], giving in sequence gem-diamine [E(GD_1)], L-Ser external aldimine [E(A_{ex1})], quinonoid [E(Q_1)], and aminoacrylate Schiff base [E(A-A)] species and a water molecule. In stage II, indole is channeled from the α-site and nucleophilically attacks the E(A-A), giving E($Q_{2/3}$), E(Aex_2), and E(GD_2) intermediates; finally producing, L-Trp. In most PLP enzymes an active-site lysine residue[4,5] is required for proton exchange to C_α of the substrate.

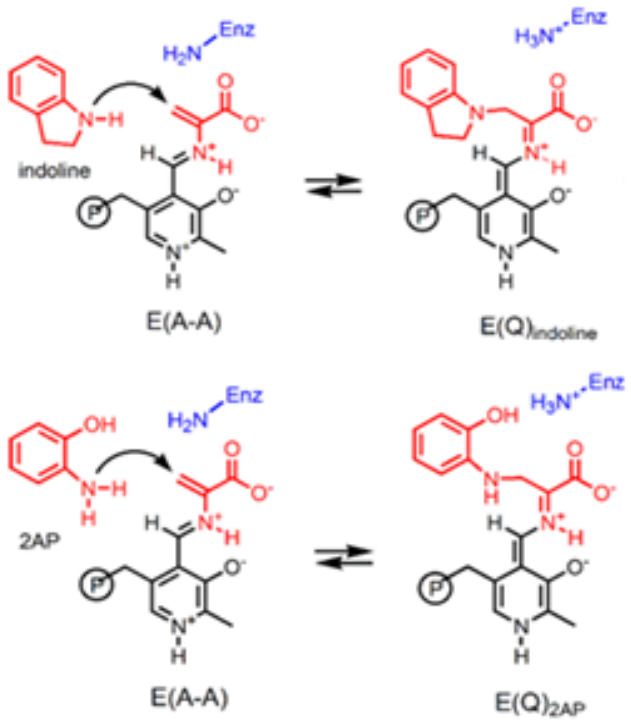
While X-ray crystal structures of open and closed conformations have been reported[6-8] the resulting resolution (nearly atomic) does not establish the protonation states of the substrate or nearby residues. What remain are major holes in answering the substrate transformation during catalysis. Fortunately, solid-state NMR spectroscopy offers critical information at the chemical-level under fundamentally the same conditions used to solve the X-ray crystal structure and can be particularly useful for studying enzyme systems[9-13]. As an example, work by McDowell et al.[13] showed that solid-state ^{13}C NMR spectroscopy could identify the E(A-A) intermediate in the crystalline state for the full tryptophan synthase when labeled L-[3-^{13}C]Ser was supplied as a substrate.

Figure 1. The β-Site Reaction in Tryptophan Synthase



Reprinted with permission from "X-ray and NMR Crystallography in an Enzyme Active Site: The Indoline Quinonoid Intermediate in Tryptophan Synthase. Jinfeng Lai, Dimitri Niks, Yachong Wang, Tatiana Domratcheva, Thomas R. M. Barends, Friedrich Schwarz, Ryan A. Olsen, Douglas W. Elliott, M. Qaiser Fatmi, Chia-en A. Chang, Ilme Schlichting, Michael F. Dunn, and Leonard J. Mueller. Copyright 2011 American Chemical Society"

Figure 2. The Indoline and 2AP Reaction



Adapted with permission from "X-ray and NMR Crystallography in an Enzyme Active Site: The Indoline Quinonoid Intermediate in Tryptophan Synthase. Jinfeng Lai, Dimitri Niks, Yachong Wang, Tatiana Domratcheva, Thomas R. M. Barends, Friedrich Schwarz, Ryan A. Olsen, Douglas W. Elliott, M. Qaiser Fatmi, Chia-en A. Chang, Ilme Schlichting, Michael F. Dunn, and Leonard J. Mueller. Copyright 2011 American Chemical Society"

MATERIALS AND METHODS

X-RAY

The quinonoid species is a catalytic intermediate generated in all PLP-dependent enzymes[14] and is investigated throughout this chapter. When the aminoacrylate intermediate, found in stage II of the β-reaction, is reacted with indoline or 2AP the quasi-stable E(Q)indoline[15,16] and E(Q_{2AP}) respectively, are formed and produce two distinct chemical and structural analogues of the short-lived quinonoid species. (Figure 2.). With the IGP analogue, N-(4'-trifluoromethoxybenzenesulfonyl)-2-aminoethyl phosphate (F9), bound to the α-site, the crystal structure of the Cs^+ form of E(Q)indoline at 1.85 Å resolution was determined and established a closed conformation of the α- and β-subunits.[5,6,8,16] The diffraction data summarized in the supporting information of chapter III showed excellent resolution for electron densities of F9, the indoline, and 2AP quinonoid species.

NMR

Creating an experimental environment where microcrystals of TS remained catalytically active producing independently the $E(Q)_{indoline}$ and $E(Q)_{2AP}$ intermediates and were confirmed using solid-state NMR spectroscopy (Figures 3. and 4.). An abundance of L-serine and indoline/2AP were present in solution which gave way to the overall analysis of the time courses for $E(Q)_{indoline}$ for the mother liquor (Figures 3. and 4.) allowed the identification of unique resonances at 103.6 ppm (red ▲) and 54.1 ppm (red ▼) as the sp^2-hybridized 2-^{13}C (C^{α}) and sp^3-hybridized 3-^{13}C (C^{β}) of $E(Q)_{indoline}$, respectively. Resonances at 61.3 ppm and 57.4 ppm (black ● and ○) were assigned to L-[2,3-^{13}C]serine nonspecifically bound to the microcrystals.

Comparable measurements using [U-^{13}C,^{15}N]serine and indoline (Figures 12 - 16.) yielded the remaining chemical shifts in Table 1. Resonances for $E(Q)_{2AP}$ were determined analogously and tabulated in Table 2. Furthermore the spectroscopic identification of aromatic atom assignments, enriched at C_2, $C_{2'}$, C_3, and N_1 of PLP, were resolved using a protocol outlined in the results and methods section of chapter IV.

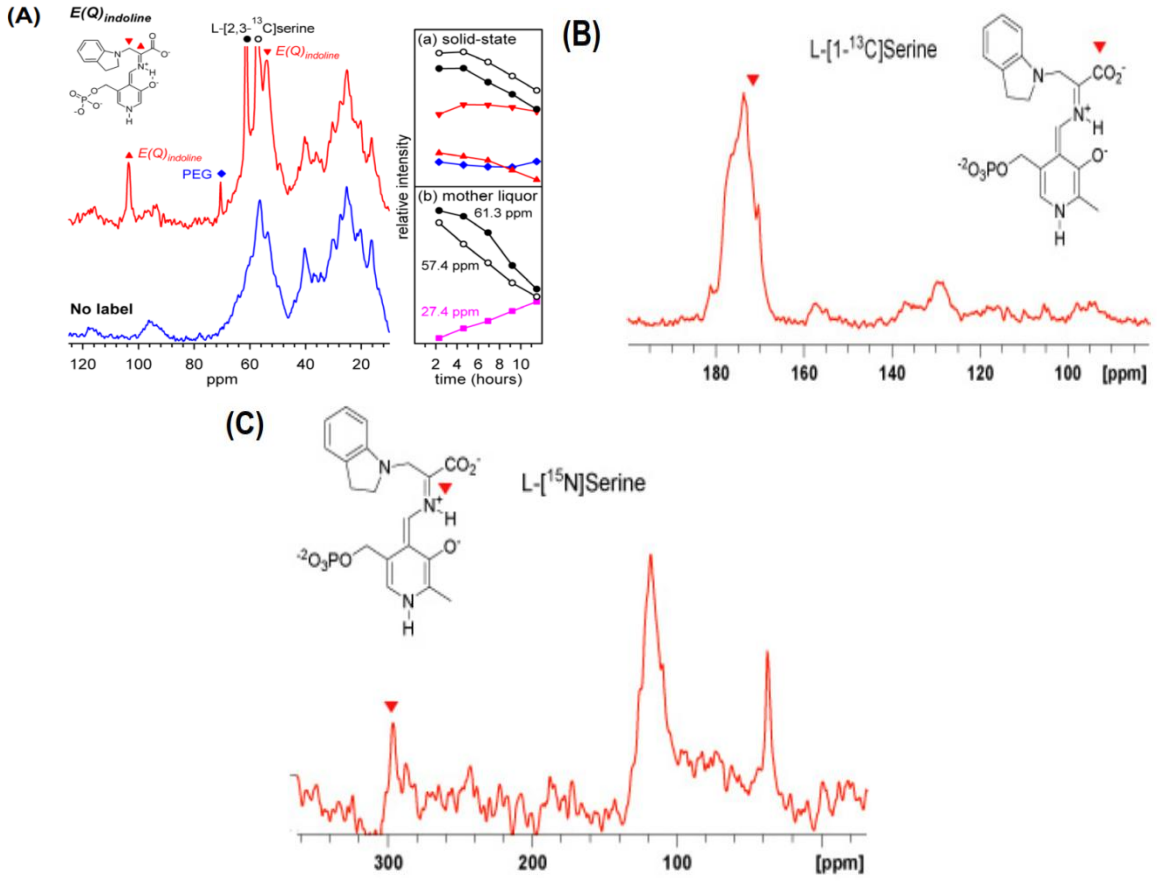


Figure 3. The solid-state NMR spectra (red) of E(Q)Indoline (A) from L-[2,3-^{13}C]Ser and unlabeled L-Ser, (B) from L-[1-^{13}C]Ser, and (C) from L-[^{15}N]Ser.

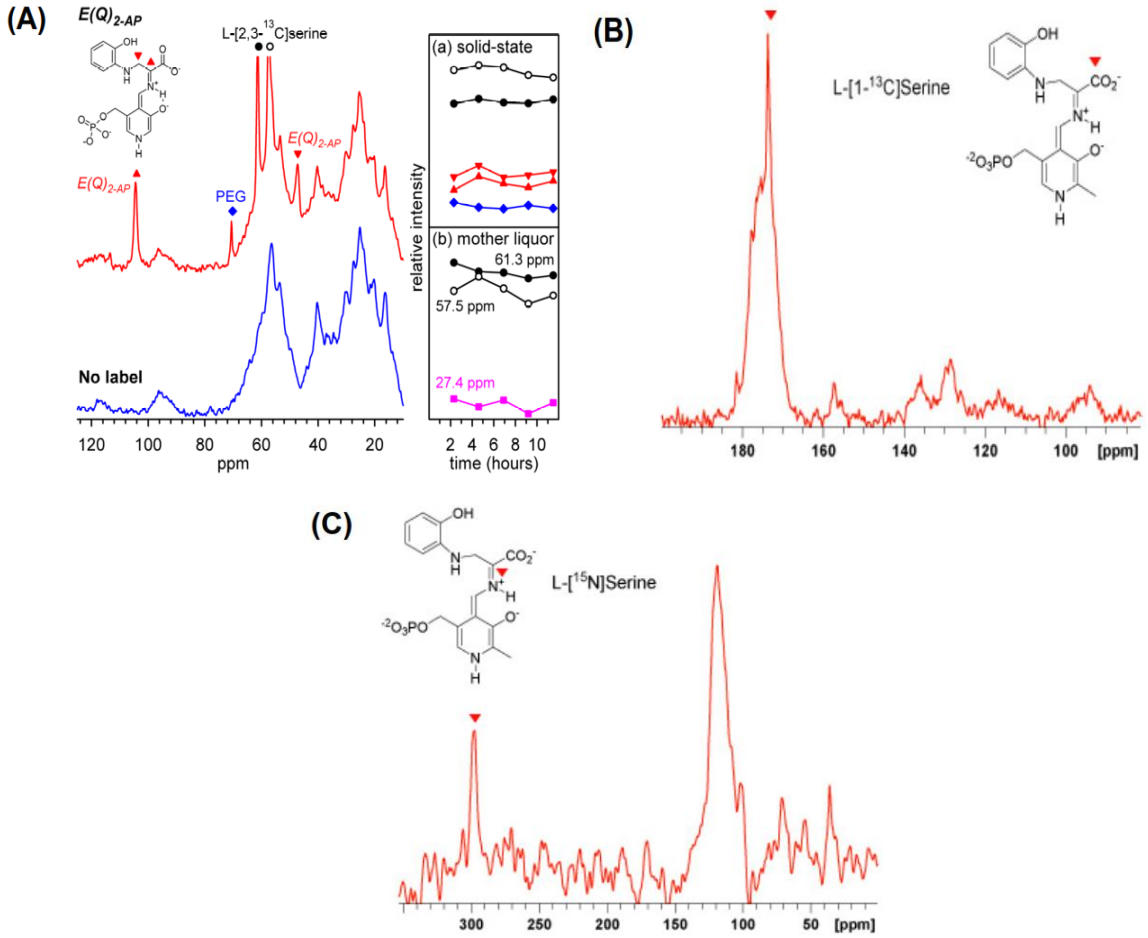


Figure 4. The solid-state NMR spectra (red) of E(Q)2AP (A) from L-[2,3-^{13}C]Ser and unlabeled L-Ser, (B) from L-[1-^{13}C]Ser, and (C) from L-[^{15}N]Ser

Reprinted with permission from "X-ray and NMR Crystallography in an Enzyme Active Site: The Indoline Quinonoid Intermediate in Tryptophan Synthase. Jinfeng Lai, Dimitri Niks, Yachong Wang, Tatiana Domratcheva, Thomas R. M. Barends, Friedrich Schwarz, Ryan A. Olsen, Douglas W. Elliott, M. Qaiser Fatmi, Chia-en A. Chang, Ilme Schlichting, Michael F. Dunn, and Leonard J. Mueller. Copyright 2011 American Chemical Society"

Table 1. Experimental chemical shifts [in ppm] relative to TMS (^{13}C) and $NH_3(l)$ ($^{15}N^a$)]; calculated shifts were produced utilizing linear regression analysis (Chapter VI) for two forms of E(Q)indoline

	Exp.	Calculated Phenolic Form 10_10_00_0	Calculated Phenolic-acid Form 10_10_01_0
$N1_{PLP}$	264.3	256.5	284.3
$C2_{PLP}$	153.6	149.6	151.6
$C3_{PLP}$	145.5	134.2	139.2
$N_{Schiffbase}$	296.5	311.6	299.2
C^{α}	103.6	105.9	97.7
C'	173.0	173.2	161.0
C^{β}	54.1	56.7	54.3
$N_{indoline}$	83.5	92.0	85.2
C2	50.5	51.5	51.8
C3	28.5	35.0	33.7

[a] For comparison, $\delta[NH_3(l)] = \delta[^{15}NH_4Cl(s)] + 39.2$ ppm.

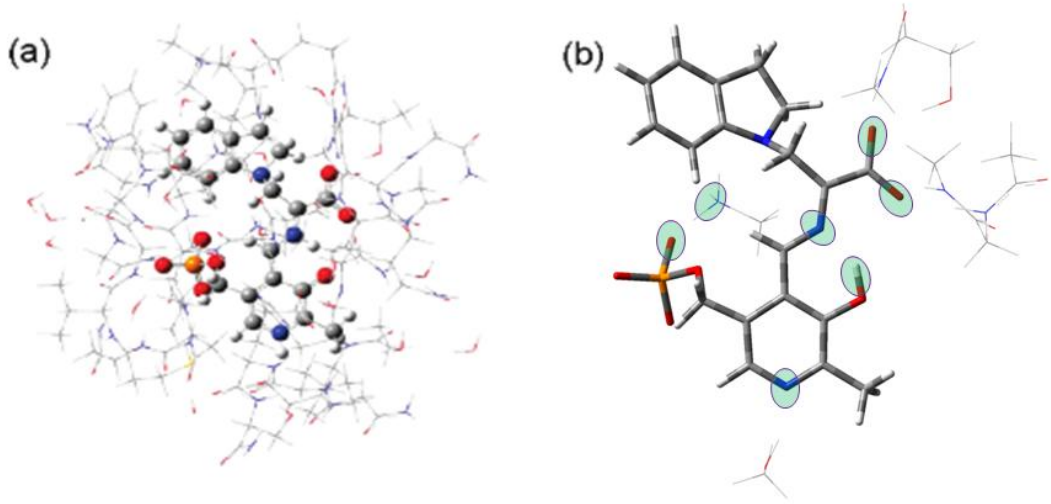


Figure 5. (a) Model of the enzyme active site in the β-subunit of tryptophan synthase, showing (wireframe) the side chains fixed at their crystallographically determined coordinates and treated at the semiempirical PM3 level of theory and (ball and stick) the PLP–substrate complex optimized using DFT at the B3LYP/6-31G(d,p) level. (b) Substructure used for calculating NMR chemical shifts [B3LYP/6-311++G(d,p) DFT]. Green ovals indicate possible sites of protonation on the PLP–substrate complex and the Lys87 side chain. The standard CPK scheme is used to designate the atom colors (H, white; C, gray; N, blue; O, red; P, orange).

Table 2. Experimental chemical shifts [in ppm] relative to TMS (^{13}C) and $NH_3(l)$ ($^{15}N^a$)]; calculated shifts were produced utilizing linear regression analysis (Chapter VI) for two forms of E(Q)2AP

	Exp.	Calculated Phenolic Form 10_10_00_0	Calculated Phenolic-acid Form 10_10_01_0
N_{2AP}	56.0	62.6	59.3
$N1_{PLP}$	265.0	261,7	279.7
$C2_{PLP}$	153.1	150.2	144.6
$C3_{PLP}$	144.6	138.2	138.2
$N_{schiffbase}$	298.6	305.3	327.2
C^{α}	105.1	109.6	104.8
$C^{'}$	173.1	174.3	166.3
C^{β}	47.0	54.1	47.4

[a] For comparison, $\delta[NH_3(l)] = \delta[^{15}NH_4Cl(s)] + 39.2$ ppm.

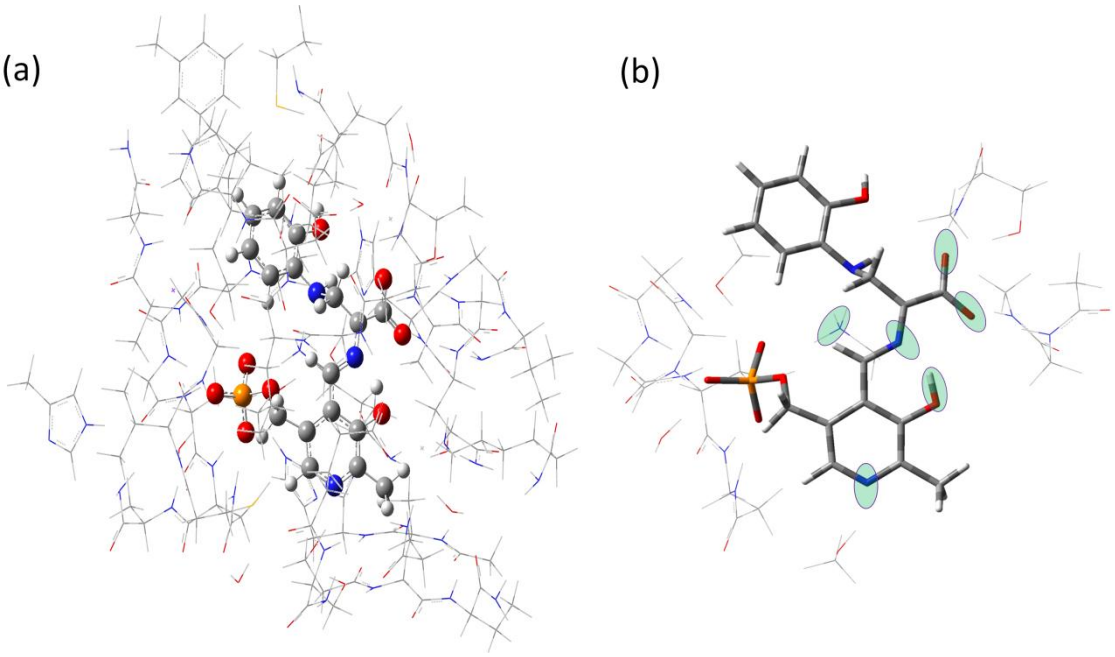


Figure 6. (a) Model of the enzyme active site in the β-subunit of tryptophan synthase, showing (wireframe) the side chains fixed at their crystallographically determined coordinates and treated at the semiempirical PM3 level of theory and (ball and stick) the PLP–substrate complex optimized using DFT at the B3LYP/6-31G(d,p) level. (b) Substructure used for calculating NMR chemical shifts [B3LYP/6-311++G(d,p) DFT]. Green ovals indicate possible sites of protonation on the PLP–substrate complex and the Lys87 side chain. The standard CPK scheme is used to designate the atom colors (H, white; C, gray; N, blue; O, red; P, orange).

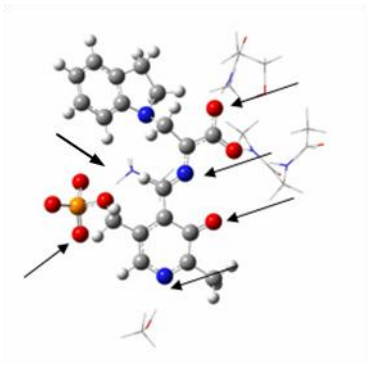
AB INITIO

Calculated chemical shifts of candidate structures were calculated using ab initio electronic structure calculations[17] and cross validated with experimental shifts to deduce the ionization states of the substrate[9] using the following protocol. In order to provide a realistic framework for optimizing the substrate complex (including local charges and hydrogen bonding interactions of nearby residues), a model of the β-subunit enzyme active site (residues fixed at their crystallographic coordinates within 7 Å of the substrate) was constructed (Figures 5- and 6--(a)). Secondly, while employing the ONIOM method[18] [where side chains were treated at the semi empirical level of theory, and the substrate was treated using density functional theory (DFT)] in Gaussian 03,[19] candidate structures for the substrate protonation states were generated then optimized to ground-state geometries. In regards to Ind Qin, the ligand was subject to minimal constraints to combat ring twisting as described in the supplemental information within chapter III. Finally, substrate chemical shifts were calculated using DFT for the excised substructure shown in Figures 4- and 5- (b), which included fragments of residues that were potentially charged or hydrogen-bonded to the substrate. Figure 5 and 6 (b) also shows possible sites of protonation used to generate the candidate structures for the PLP-ligand complex and includes (with respect to the order of the seven digit sequence): the βK87 residue, phosphoryl group, phenolic oxygen, the Schiff base linkage, both carboxylate oxygens, and pyridine nitrogen. Simultaneous protonation was not accounted for and candidate structures with more than a single proton placed in the pocket defined by the pyridoxal oxygen, the Schiff base nitrogen, and the closest carboxylate oxygen were not considered. Further, we also did not allow a doubly protonated carboxylate group and the phosphoryl was also considered dianionic

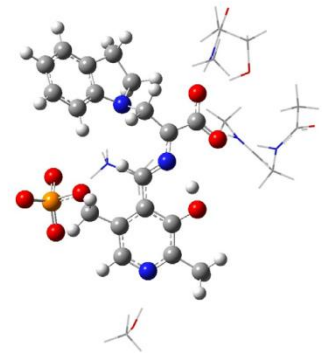
for the 2AP models. This gave 56 possible candidate structures for Ind Qin and 28 total for 2AP, each were geometrically optimized and from which chemical shifts were calculated (Tables 4 and 5 in chapter III) for comparison with the experimental values. This comparison was quantified using the reduced χ^2 statistic (the weighted deviation of the model and experimental shifts), which was calculated using root-mean-square deviation; weights of 4.12 ppm for ^{13}C and 9.75 ppm for ^{15}N. These weights, also known as the nuclide-dependent weighting parameter were determined computationally using linear regression analysis and is the basis of chapter VI and is further defined in chapter III.

Although the experimental results discussed in chapter IV suggest a deprotonated pyridine nitrogen *and* a purely dianionic phosphoryl for these intermediates, we justify for Ind Qin initially, using an extremely large, rigorous, and ultimately unbiased modeling of protonation states as a test of concept. We therefore expect that the reduced chi squared analysis should provide evidence that the experimental environment is captured "computationally" as well. In other words, large reduced χ^2's will be produced form models harboring exotic charge distributions and/or poor structural geometries. The results for Ind Qin reveal that future modeling of intermediates warrants the use of a sleeker, more refined modeling of protonation states as seen in the case of 2AP Qin where we assume appropriately a purely dianionic phosphoryl. This yielded 28 structures for 2AP which were computed and coded analogously. To become familiar with the 7 digit code, Figure 7 demonstrates 24 of the 56 structures considered for Ind Qin, while all 28 structures considered for 2AP can be expected analogously.

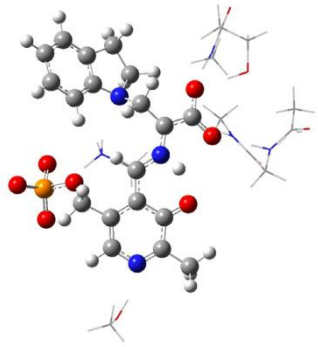
Figure 7. 24 of the 56 total structures of Ind Qin



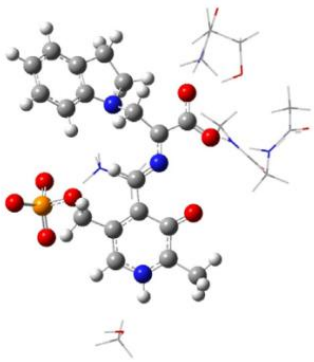
10-00-00-0



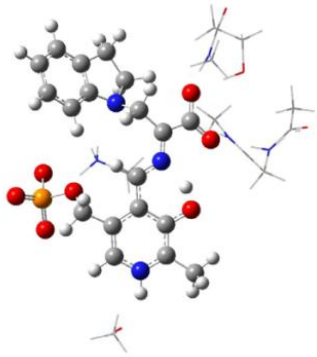
10-10-00-0



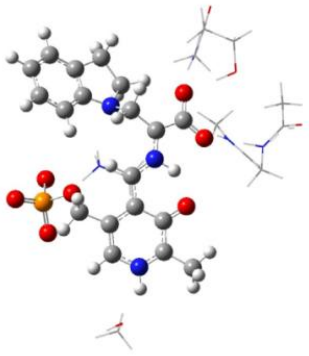
10-01-00-0



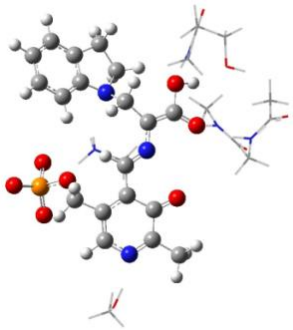
10-00-00-1

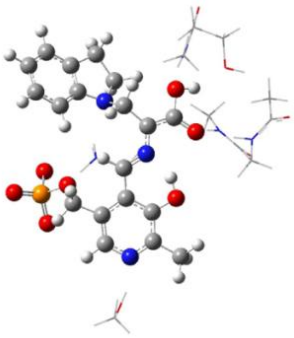


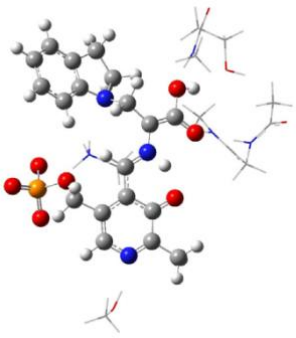
10-10-00-1



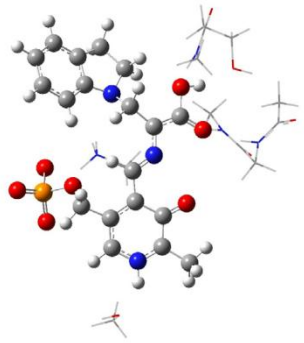
10-01-00-1



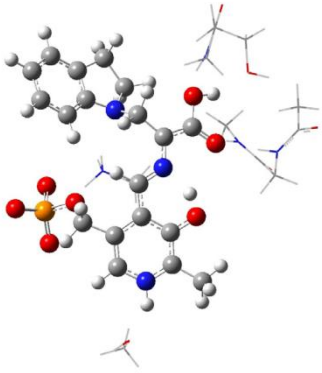




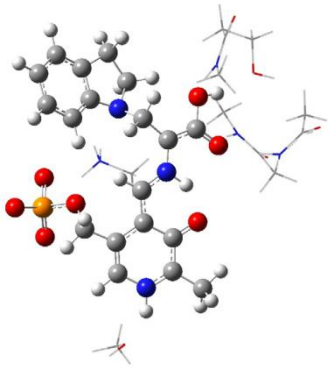
10-00-01-0



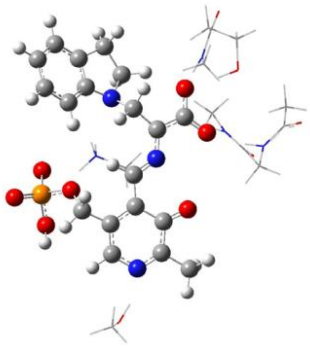
10-10-01-0



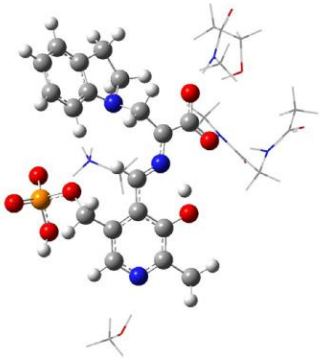
10-01-01-0



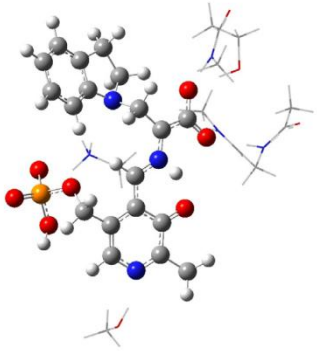
10-00-01-1



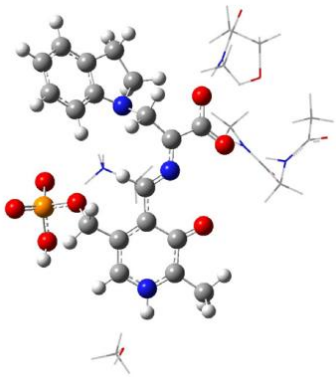
10-10-01-1



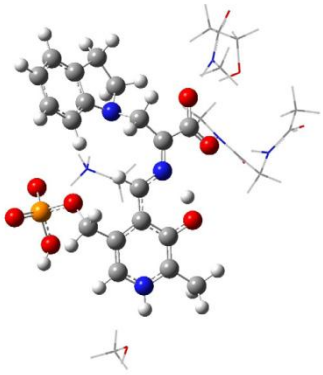
10-01-01-1



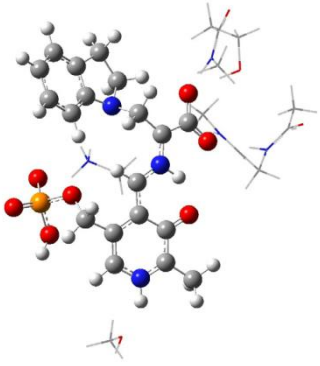
11-00-00-0



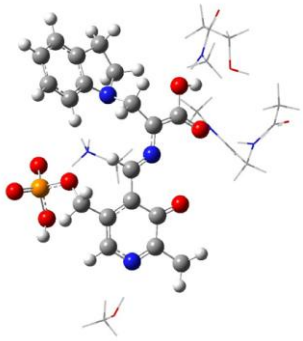
11-10-00-0



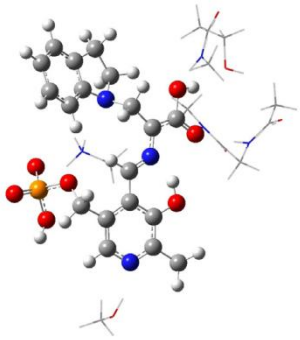
11-01-00-0



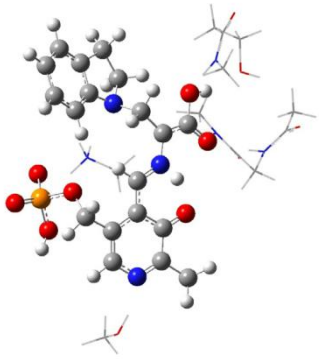
11-00-00-1



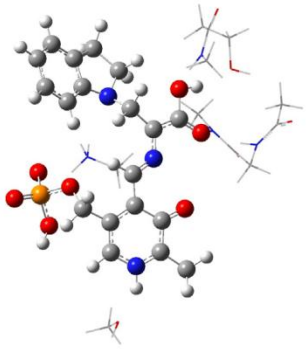
11-10-00-1



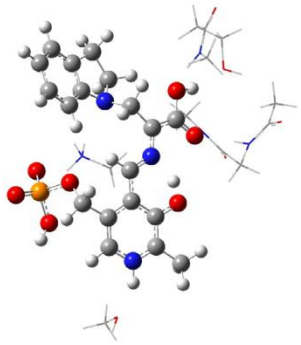
11-01-00-1



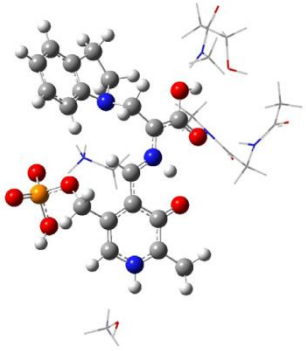
11-00-01-0



11-10-01-0



11-01-01-0

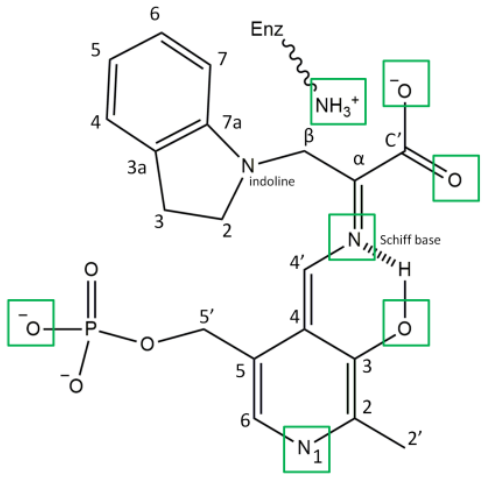


11-00-01-1

11-10-01-1

11-01-01-1

Figure 8. Bar chart representation of reduced χ^2 for Ind Qin



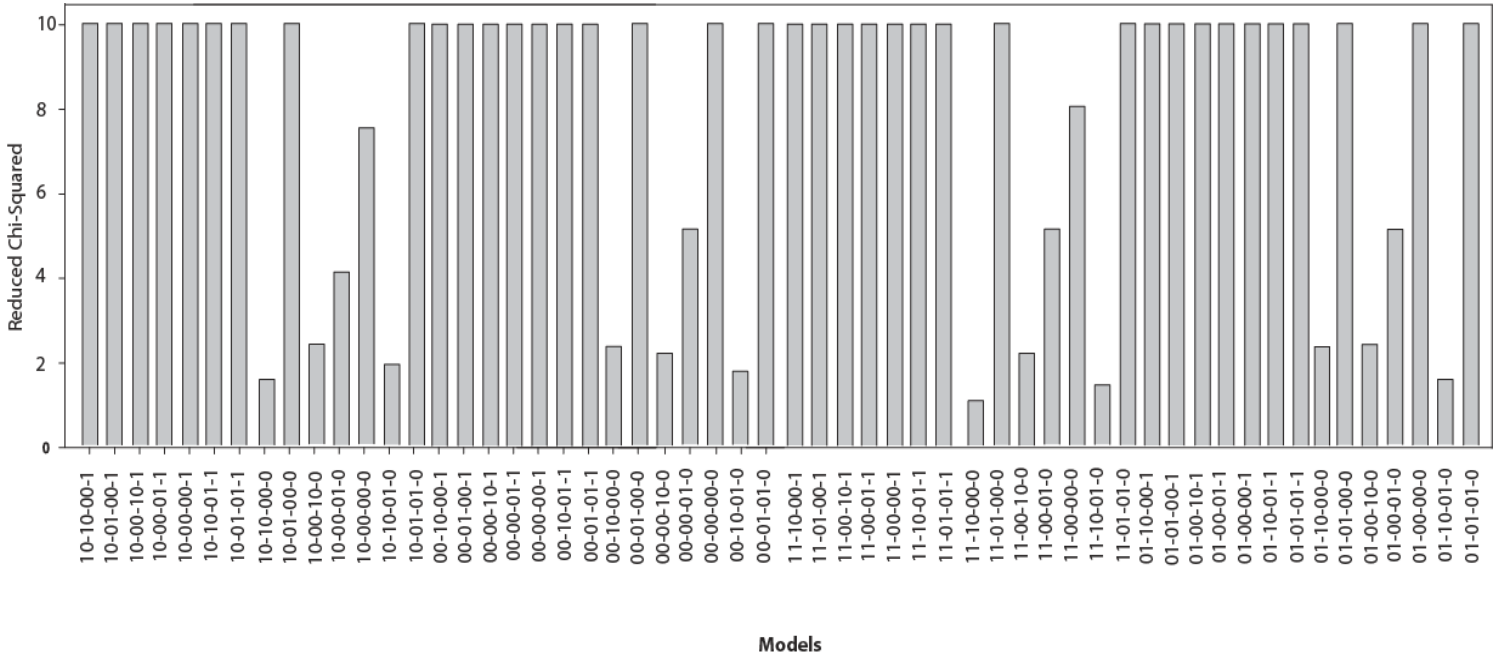
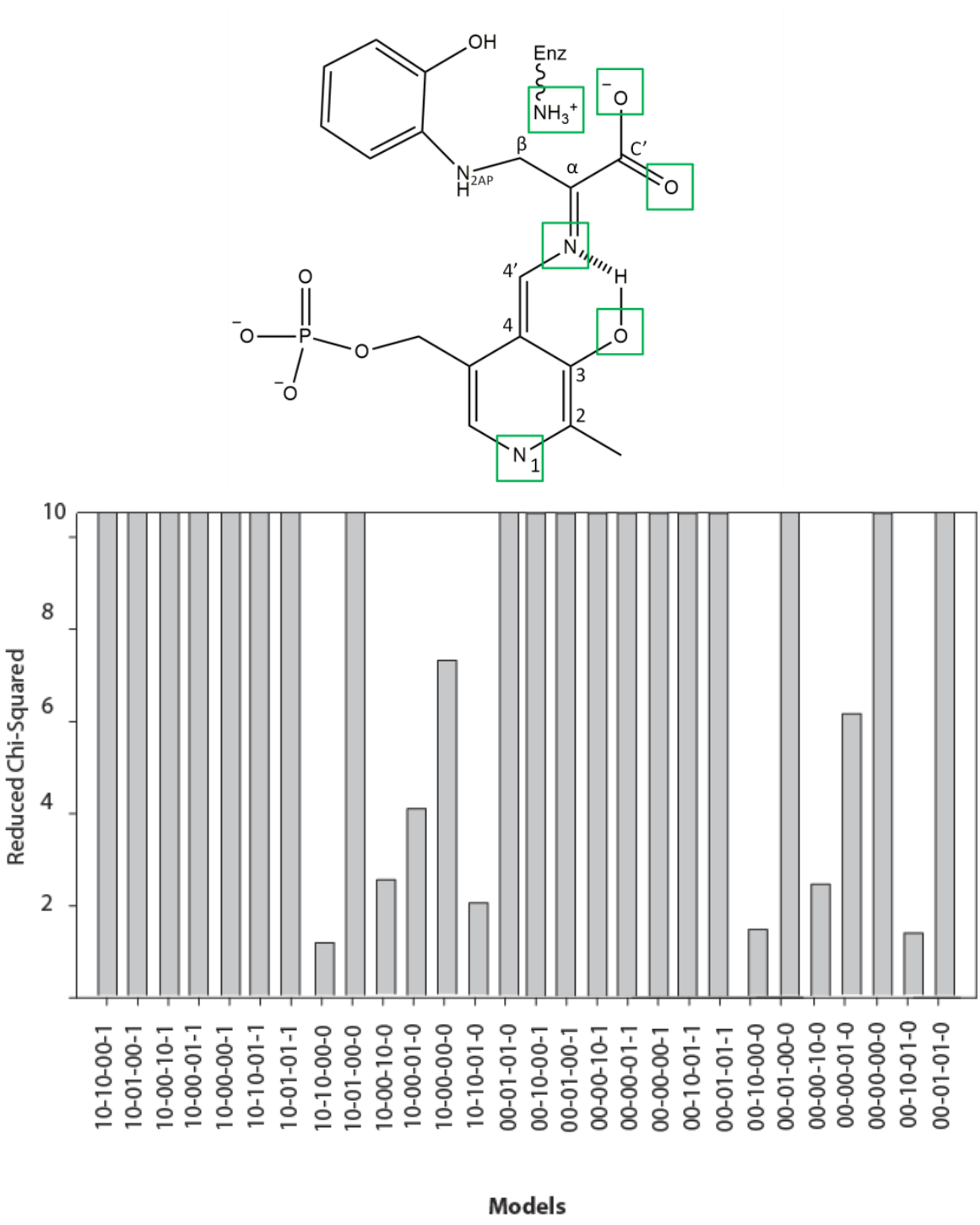


Figure 9. Bar chart representation of reduced χ^2 for 2AP



RESULTS AND DISCUSSION

Of the candidate structures, the phenolic/phenolic-acid forms (model codes: ##-10-00-0 / ##-10-01-0 respectively; where "#" represents 1 or 0) were found to have acceptable agreement between the calculated and experimental ^{13}C and ^{15}N chemical shifts for both quinonoid species. These eight specific phenolic forms of Ind Qin resulted in values producing equivalent patterns of distribution as seen in Figure 8. with the reduced χ^2's in the range of 1.3 - 2.3, where models 10-10-01-0, 01_10_00_0, and 00-10-00-0 (reduced χ^2's of 1.9, 1.9 and 2.3) happens to fall just slightly outside of the statistically accepted 95% confidence interval for 10 experimental data points (0.39 - 1.83)[24]. Similarly the four (4) phenolic forms of the 2AP Qin generated reduced χ^2's ranging from: 1.0 - 2.0, where model 10_10_01_0 (reduced χ^2 of 2.0) extends only 0.06 units outside of the accepted 95% confidence interval for 8 experimental data points (0.34 - 1.94) [24]. Divergences for a majority of the remaining structures in both intermediates takes place at the pyridine and Schiff base nitrogen which collectively have general experimental shifts in the typical ~265 and ~296 ppm range respectively. Established data shows that in inhibited PLP-dependent alanine racemase render a protonated Schiff base nitrogen at 189 ppm,[12] while comparable models convey a neutral Schiff base linkage with methylamine at 313 ppm[20] (both adjusted to δ[NH3(l)]). These results are agree with the calculated chemical shifts of the phenolic models where the Schiff base nitrogen is deprotonated and inhibits a chemical shift attenuated by hydrogen bonding to the phenolic O-H bond.

Although two species of modeling meet the statistical criteria employed, it is important to step back and tally the major findings of the computational results: To distinguish the

difference between the phenolic and phenolic-acid environments it can be thought that the acid portion represents a incredibly strong hydrogen-bond to the residues surrounding the upper carboxylate moiety. Furthermore the overall importance derived is that the phenolic form *still* emerges, whether modeled with a strongly hydrogen bonded upper carboxylate group or positively charged -NH_3 group at the nearby βLys87 residue or maintaining a mono or di-anionic phosphoryl group, and remains the predominate species. Of the 56 unique protonation states tested for Ind Qin, models including nearby βLys87 residue and phosphoryl group do not exhibit a significant role in perturbing the chemical shift on the isotopically labeled positions of the ligand as confirmed from the statistical fitting. Additionally the statistical fitting of the 28 2AP models further support the insignificance and lack of chemical shift perturbation to isotopically labeled atoms of the ligand upon protonation of the βLys87 residue. While these groups protonation states remain non-influencing to the data points extracted from the particular atomic site on the ligand, in every single case including both intermediates the pyridine nitrogen when protonated (namely models: ##_##_##_1) results in reduced χ^2 values that sky rocket ranging from 13 in the lowest and 53 in the highest with an overall average of ≈ 22; suggesting a completely deprotonated pyridine nitrogen, attenuated by hydrogen-bonding to nearby βserine377. This is clearly evident from Figures 8. and 9. where models containing a "1" in the last position of the 7 digit sequence contains a bar that is beyond the chart. Furthermore direct chemical shift evidence from chapter IV shows additional support for an entirely unprotonated pyridine species during all steps of catalysis, including the internal aldimine, for this particular tryptophan synthase system.

These intermediate structures behave in an analogous manner to a sprinter positioned in running blocks, anticipating the sound of the pistol to begin the catalytic race and therefore have counter intuitive implications for the mechanism in tryptophan synthase, which at the next step involves enticing the protonation at the C^{α} site by βK87. The atomic charges calculated from the ab initio electronic structures using natural bond orbital analysis[21] are consistent with the phenolic/phenolic-acid model hindering the next step in catalysis, with C^{α} essentially zero (-0.078 and -0.064 au) in the phenolic form and having only a very slight increase of negative charge (-0.093 and -0.093 au) in the phenolic-acid form for Ind Qin and 2AP respectively. The lack of negative charge at C^{α} plays two roles. First, it attenuates the electric field component along the natural reaction coordinate to slow down the proton donation from N^{ε} of βK87 to the C^{α} site additionally supporting on the other hand that the natural indole-substrate is incredibly short lived in comparison. Second, it increases the energy barrier along the reaction coordinate devoid of charge stabilization.

CONCLUSION

Chemical level phenomena, particularly protonatation and hyrbridation states, are vital for predicting the flow of the mechanism and overall function of the enzyme. Because the chemical shift is an extremely insightful probe to the electro-chemical environment, making use of the unique synergistic combination of X-ray crystallography, solid-state NMR spectroscopy, and ab initio computational theory, termed "NMR-Crystallography", to characterize the enzymatic indoline/2AP quinonoid intermediates in the PLP-dependent tryptophan synthase is key. It was shown that assorted models of charge and protonation states can be singled out by their calculated effect on their unique chemical shifts, ultimately emerging a chemical species most consistent with the experimental results. These support the phenolic/phenolic-acid models and conclusively point towards a more refined understanding of the indoline and 2AP quinonoid protonation states in tryptophan synthase and their responsibility in hindering the next step in catalysis.

REFERENCES

(1) Yanofsky, C.; Crawford, I. P. In The Enzymes; Boyer, P. D., Ed.; Academic Press: New York, 1972; pp 1-31.

(2) Miles, E. W. AdV. Enzymol. Relat. Areas Mol. Biol. 1979, 49, 127-186.

(3) Dunn, M. F.; Aguilar, V.; Brzovic, P.; Drewe, W. F.; Houben, K. F.; Leja, C. A.; Roy, M. Biochemistry 1990, 29, 8598-8607.

(4) Kirschner, K.; Lane, A. N.; Strasser, A. W. M. Biochemistry 1991, 30, 472-478. Pan, P.; Woehl, E.; Dunn, M. F. Trends Biochem. Sci. 1997, 22, 22-27.

(5) Miles, E. W.; Rhee, S.; Davies, D. R. J. Biol. Chem. 1999, 274, 12193- 12196.

(6) Dunn, M. F.; Niks, D.; Ngo, H.; Barends, T. R. M.; Schlichting, I. Trends Biochem. Sci. 2008, 33, 254-264.

(7) Ngo, H.; Kimmich, N.; Harris, R.; Niks, D.; Blumenstein, L.; Kulik, V.; Barends, T. R.; Schlichting, I.; Dunn, M. F. Biochemistry 2007, 46, 7740- 7753.

(8) Rhee, S.; Parris, K. D.; Hyde, C. C.; Ahmed, S. A.; Miles, E. W.; Davies, D. R. Biochemistry 1997, 36, 7664-7680. Sachpatzidis, A.; Dealwis, C.; Lubetsky, J. B.; Liang, P. H.; Anderson, K. S.; Lolis, E. Biochemistry 1999, 38, 12665-12674. Weyand, M.; Schlichting, I.; Marabotti, A.; Mozzarelli, A. J. Biol. Chem. 2002, 277, 10647-10652. Blumenstein, L.; Domratcheva, T.; Niks, D.; Ngo, H.; Seidel, R.; Dunn, M. F.; Schlichting, I. Biochemistry 2007, 46, 14100-14116.

(9) Barends, T. R. M.; Domratcheva, T.; Kulik, V.; Blumenstein, L.; Niks, D.; Dunn, M. F.; Schlichting, I. ChemBioChem 2008, 9, 1024-1028.

(10) Mao, J. H.; Mukherjee, S.; Zhang, Y.; Cao, R.; Sanders, J. M.; Song, Y. C.; Zhang, Y. H.; Meints, G. A.; Gao, Y. G.; Mukkamala, D.; Hudock, M. P.; Oldfield, E. J. Am. Chem. Soc. 2006, 128, 14485-14497.

(11) McDermott, A.; Polenova, T. Curr. Opin. Struct. Biol. 2007, 17, 617-622.

(12) Sharif, S.; Fogle, E.; Toney, M. D.; Denisov, G. S.; Shenderovich, I. G.; Buntkowsky, G.; Tolstoy, P. M.; Huot, M. C.; Limbach, H. H. J. Am. Chem. Soc. 2007, 129, 9558-9559.

(13) Copie, V.; Faraci, W. S.; Walsh, C. T.; Griffin, R. G. Biochemistry 1988, 27, 4966-4970.

(14) McDowell, L. M.; Lee, M. S.; Schaefer, J.; Anderson, K. S. J. Am. Chem. Soc. 1995, 117, 12352-12353.

(15) Dunathan, H. C. Proc. Natl. Acad. Sci. U.S.A. 1966, 55, 712-716.

(16) Roy, M.; Keblawi, S.; Dunn, M. F. Biochemistry 1988, 27, 6698-6704.

(17) Harris, R. M.; Dunn, M. F. Biochemistry 2002, 41, 9982-9990. Harris, R. M.; Ngo, H.; Dunn, M. F. Biochemistry 2005, 44, 16886-16895. Dierkers, A. T.; Niks, D.; Schlichting, I.; Dunn, M. F. Biochemistry 2009, 48, 10997-11010.

(18) Oldfield, E. Philos. Trans. R. Soc., B 2005, 360, 1347-1361.

(19) Vreven, T.; Morokuma, K. J. Comput. Chem. 2000, 21, 1419-1432.

(20) Frisch, M. J.; et al. Gaussian 03, revision B.05; Gaussian, Inc.: Wallingford, CT, 2003.

(21) Sharif, S.; Huot, M. C.; Tolstoy, P. M.; Toney, M. D.; Jonsson, K. H. M.; Limbach, H. H. J. Phys. Chem. B 2007, 111, 3869-3876. Sharif, S.; Schagen, D.; Toney, M. D.; Limbach, H. H. J. Am. Chem. Soc. 2007, 129, 4440-4455.

(22) Foster, J. P.; Weinhold, F. J. Am. Chem. Soc. 1980, 102, 7211-7218. (22) Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Toney, M. D.; Limbach, H.-H. Biochemistry [Online early access]. DOI: 10.1021/bi101061m. Published Online: Nov 10, 2010.

(23) Facelli, J. C.; Grant, D. M. Nature 1993, 365, 325-327. Olsen, R. A.;
Struppe, J.; Elliott, D. W.; Thomas, R. J.; Mueller, L. J. J. Am. Chem. Soc. 2003, 125, 11784-11785. Harris, R. K. Solid State Sci. 2004, 6, 1025- 1037. Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. Phys. Chem. Chem. Phys. 2009, 11, 2610-2621.

(24) Experiments in Physical Chemistry 8th Edition by: Carl Garland, Joseph Nibler, and David Shoemaker; ISBN 978-0072828429

CHAPTER III. SUPPLEMENTAL INFORMATION FOR CHAPTER TWO: X-RAY AND NMR CRYSTALLOGRAPHY IN AN ENZYME ACTIVE SITE: THE INDOLINE AND 2AP QUINONOID INTERMEDIATE IN TRYPTOPHAN SYNTHASE.

ABSTRACT

The synergistic combination of NMR spectroscopy, X-ray crystallography and ab initio computational chemistry, also referred to as "NMR Crystallography", can be used as an unmatched tool for elucidating high resolution three-dimensional structures of diverse materials. This synergistic approach was used to establish the structures of the indoline and 2-aminophenol (2AP) quinonoid intermediates in the pyridoxal-5′-phosphate-dependent enzyme tryptophan synthase. Experimentally measured chemical shifts at particular ^{13}C - and ^{15}N-labeled locations on the ligand were used to exclusively distinguish and emerge a comparable species from the variety of computational model types. The reduced chi squared analysis suggests the phenolic and/or phenolic-acid form to be the predominate protonated species for both indoline and 2-aminophenol quinonoid intermediates.

INTRODUCTION

The purpose of this chapter is to provide supplemental information regarding the techniques discussed in Chapter II.

MATERIALS AND METHODS

The X-ray Crystal Structure of the (F9)E(Q)Indoline Complex.

Purification of TRPS was carried out using a protocol described previously.[1] Using a procedure similar to that described for the formation of an α-site transition-state-like conformation with $E(Q)_{indoline}$[2], the complex of the IGP analogue, N-(4'-trifluoromethoxybenzenesulfonyl)-2- aminoethyl phosphate (F9), with $E(Q)_{indoline}$ was formed. This included a soaking step for the crystals with with F9, L-Ser, CsCl, and indoline. To achieve cryogenic protection an addition of 20% glycerol was used before flash cooling the crystals in liquid nitrogen. With the crystal kept at 100K during data collection, the diffraction data were collected on a beamline PXII (X10SA) of the SLS in Villigen (CH) and processed and refined with XDS[3], CNS[4] and REFMAC5[5] as described previously.[2] Summarized in Table 3. are the unit cell parameters, data collection, and refinement statistics. The electron density maps show indoline reacted with the α-aminoacrylate intermediate at the β-site to give the quinonoid species $E(Q)_{indoline}$ (Figure 10.).

Table 3. Crystal parameters, data collection, and refinement statistics

Table S1: Crystal parameters, data collection, and refinement statistics	
Data collection	**F9-indoline quinonoid structure**
Space group	*C*2
Cell dimensions	
a,*b*,*c* (Å)	184.2, 60.6, 67.4
β (°)	94.62
Resolution (Å)	20-1.85 (1.95-1.85)
R-factor	0.059 (0.358)
I/σ(*I*)	12.1 (3.4)
Completeness (%)	93.9 (86.8)
Redundancy	3.1 (3.1)
Refinement	
Resolution (Å)	19.51-1.85
No. unique reflections	59360
R_{work}/R_{free}	0.2016 (0.2291)
No. atoms	
Protein chain A/B	2001/2963
F9/quinonoid/Cs^+/water	21/28/3/321
B-factors	
Protein chain A/B	41.1/25.8
F9/quinonoid/Cs^+/water	37.7/27.5/43.7/36.1
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.008
Bond angles (°)	1.151
%-age of residues in region of Ramachandran plot	
core	93.7
allowed	6.1
generously allowed	0.2
disallowed	0.0

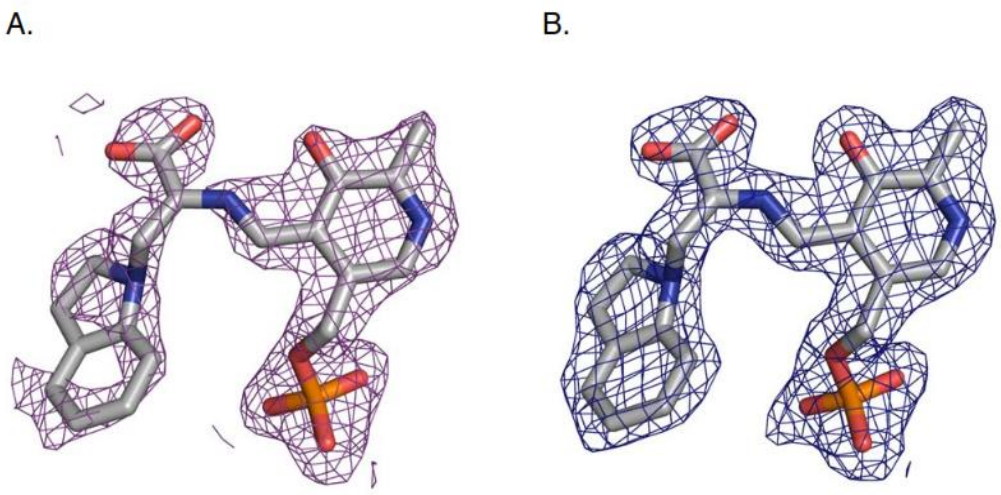


Figure 10. A. $2mF_o$-DF_c electron density map contoured at 1 σ for the indoline quinonoid, after simulated annealing and prior to incorporation of the ligand into the model, overlayed on the final, refined structure. B. Final, refined $2mF_o$-DF_c electron density map (1 σ).

Reprinted with permission from "Supporting Information for X-ray and NMR Crystallography in an Enzyme Active Site: The Indoline Quinonoid Intermediate in Tryptophan Synthase. Jinfeng Lai, Dimitri Niks, Yachong Wang, Tatiana Domratcheva, Thomas R. M. Barends, Friedrich Schwarz, Ryan A. Olsen, Douglas W. Elliott, M. Qaiser Fatmi, Chia-en A. Chang, Ilme Schlichting, Michael F. Dunn, and Leonard J. Mueller. Copyright 2011 American Chemical Society"

UV-Vis

From the same crystal sample, kept at 100 K, a visible light absorption spectrum was collected using an Andor SR-303i-A spectrometer equipped with an Andor DU420A-0E CCD detector (Figure 11.). The peak at 466 nm suggests the indoline quinonoid formed in the crystal[7]; deposited in the Protein data Bank (PDB, www.pdb.org) under accession code 3PR2.

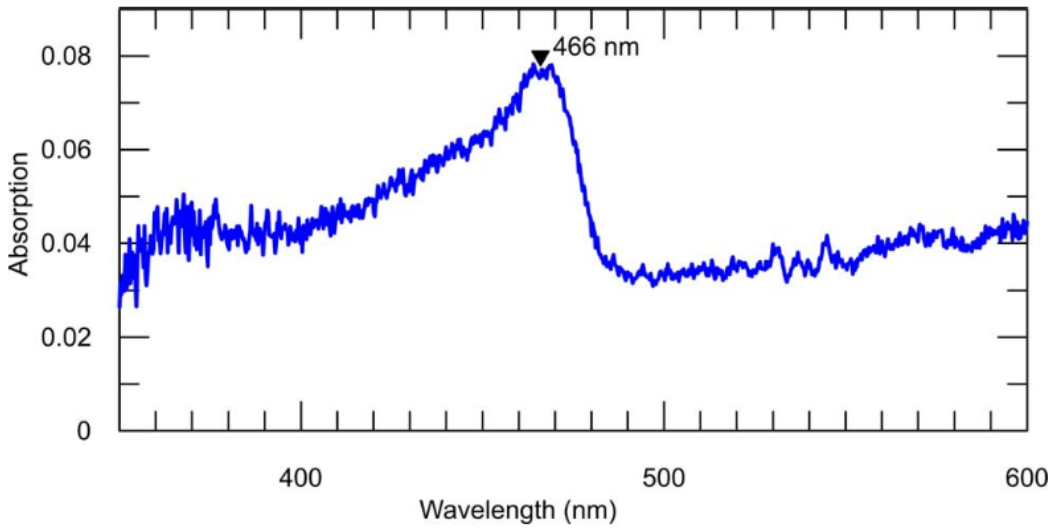


Figure 11. Spectrum collected from the crystal after crystallographic data collection, with the characteristic indoline quinonoid peak at 466nm.

NMR Methods:

Materials

Tryptophan synthase was purified using a protocol described previously.[12] Microcrystals were made by a 1:1 dilution of the enzyme solution with triethanolamine (TEA)-buffer containing 12-14% PEG 8000 and 1.5 mM spermine-HCl.[13] These were then collected and washed with 50 mM TEA, pH 7.8 containing 20% PEG 8000, 1.5 mM Spermine, 3 mM F9, 100 mM CsCl, and 20 mM indoline. Magic-angle-spinning (MAS) rotors contained approximately 16 -

19 mg of protein. A concentrated solution of unlabeled (Sigma) or labeled (Cambridge Isotope Labs) L-Ser in 50 mM TEA, pH 7.8 was added immediately prior to data collection.

NMR Spectroscopy

Magic-angle-spinning (MAS) NMR spectra were obtained on a Bruker DSX 400 Spectrometer (400.4 MHz ^{1}H; 100.7 MHz ^{13}C; 40.6 MHz ^{15}N) using a double resonance 4 mm MAS probe (sample volume ~ 80 l) and a MAS rate of 8 kHz. Sample temperatures were maintained using a Kinetics XRII851 Sample Cooler. 83 kHz ^{1}H and 501kHz ^{13}C/^{15}N $\pi/2$ pulses were used along with either low power (10-20 kHz) or high-power (83 kHz) ^{1}H decoupling during acquisition. 1D ^{13}C/^{15}N spectra were collected beginning with either a single carbon/nitrogen channel pulse and low-power (20 kHz) decoupling (64 scans in 4 min 40 sec, ^{13}C (Figure 16); 256 scans in 17 min 20 sec, ^{15}N) or cross-polarization (CP) from ^{1}H with high-power decoupling (2048 scans in 2 hours 18 min 25 sec (^{13}C) or 4096 scans in 4 hours 36 min 50 sec (15N); CP: 1 ms contact time, 50 kHz 1H spin-lock, 50% - 100% linearly ramped ^{13}C/^{15}N match). These experiments were alternated for at least 24 hrs in order to delineate the free ligand and reaction products in solution and reporting on immobilized (solid) species. In all experiments, 2048 complex time points were sampled with a spectral width of 50 kHz and averaged with a relaxation delay of 4 s. Data sets were zero-filled to 4k complex time points and processed with 50 Hz exponential line broadening for the cross-polarization experiments. No line broadening was used for the one-pulse experiments. ^{13}C spectra were referenced via an external adamantane standard with the downfield shift set to 38.48 ppm from TMS (corresponding to 40.48 ppm from DSS).[14] ^{15}N spectra were referenced indirectly to liquid ammonia (at 0 ppm)

following the protocol outlined in reference (15). This procedure gave a chemical shift of 39.2 ppm for an external U-^{15}N ammonium chloride sample, allowing direct comparison to shifts referenced on that scale (δ(NH3, liq.) = δ ($^{15}NH_4Cl$, solid)+39.2 ppm). ^{15}N shifts referenced to nitromethane are related by δ(CH3NO2, liq.) = δ (15NH4Cl, solid)-341.168 ppm.[16]

In addition to the spectrum presented in the text for the reaction of tryptophan synthase microcrystals with 20 mM L-[2,3-^{13}C]Ser and 20 mM unlabeled indoline, additional spectra were taken with L-[U- ^{13}C]Ser, L-[^{15}N]Ser, and U-[^{13}C,^{15}N]indoline as shown in Figures 12-15. . While the ^{15}N shifts for both Ser (Figure 14.) and indoline (Figure 15.) are easily observed, the carboxylate of Ser (Figure 12.) and C-2 and C-3 for indoline (Figure 13.) are both obscured by low solubility or nonspecific binding of the labeled substrate to the microcrystals. In these cases a strategy was devised in which equilibrium concentrations of the enzyme-bound intermediate were established in the presence of two different concentrations of non-specifically bound substrate, so that difference spectra could be used to observe the resonances from the specifically and non-specifically bound intermediates.

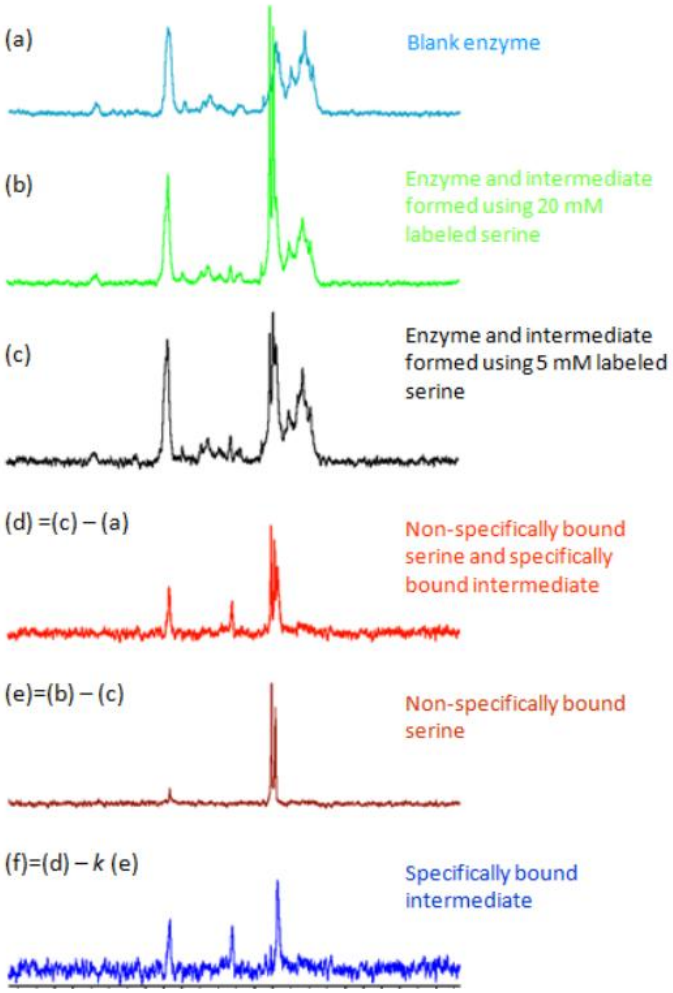


Figure 12. ^{13}C-solid-state NMR cross-polarization magic-angle-spinning (CPMAS) spectra of the reaction of tryptophan synthase microcrystals with 20 mM indoline, 3 mM F9, 100 mM CsCl and (a) no serine, (b) 20 mM L-[U-^{13}C]Ser, and (c) 5 mM L-[U-^{13}C]Ser collected at -10°C and averaged over 4.6 hrs. (d) The difference between the spectra in (c) and (a) gives a spectrum that contains resonances for both the specifically and non-specifically bound serine species. (e) The difference between the spectra in (b) and (c) gives a spectrum that is composed primarily of resonances from non-specifically bound serine. (f) The difference between (d) and (e) (scaled appropriately) gives a spectrum of the specifically-bound intermediate. Data acquired on a 9.4 T Bruker DSX spectrometer equipped with a 4 mm CPMAS probe spinning at a MAS rate of 8 kHz.

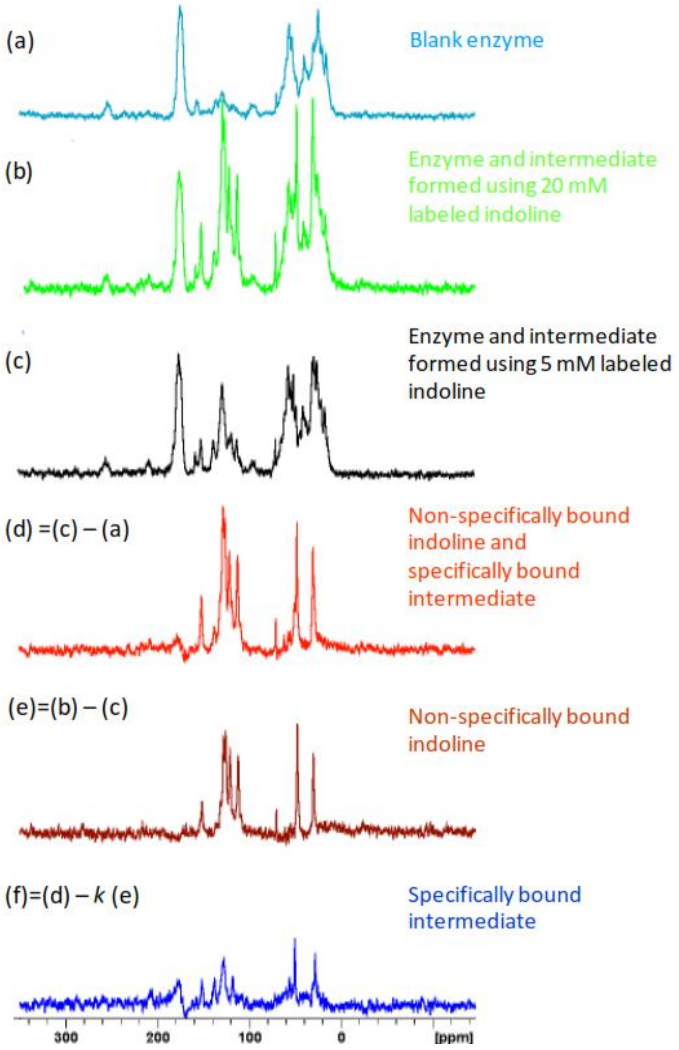


Figure 13. ^{13}C-solid-state NMR cross-polarization magic-angle-spinning (CPMAS) spectra of the reaction of tryptophan synthase microcrystals with 20 mM Ser, 3 mM F9, 100 mM CsCl and (a) no indoline, (b) 20 mM [U-^{13}C,^{15}N]indoline, and (c) 5 mM [U-^{13}C,^{15}N]indoline collected at -10°C and averaged over 4.6 hrs. (d) The difference between the spectra in (c) and (a) gives a spectrum that contains resonances for both the specifically and non-specifically bound indoline species. (e) The difference between the spectra in (b) and (c) gives a spectrum that is composed primarily of resonances from non-specifically bound indoline. (f) The difference between (d) and (e) (scaled appropriately) gives a spectrum of the specifically-bound intermediate. Data acquired on a 9.4 T Bruker DSX spectrometer equipped with a 4 mm CPMAS probe spinning at a MAS rate of 8 kHz.

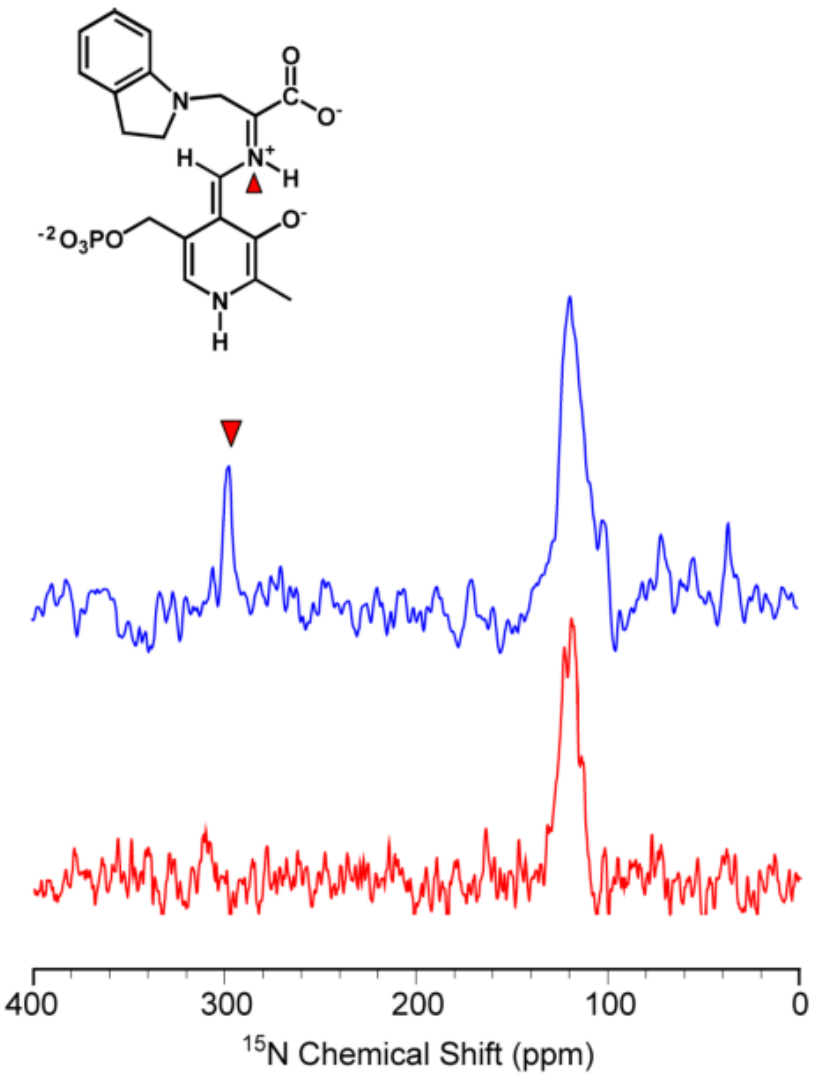


Figure 14. ^{15}N-solid-state NMR cross-polarization magic-angle-spinning (CPMAS) spectra of the reaction of tryptophan synthase microcrystals with 20 mM indoline, 3 mM F9, 100 mM CsCl and no serine (red spectrum) and 20 mM L-[^{15}N]Ser (blue spectrum) collected at -10°C and averaged over 9.2 hrs. Data acquired on a 9.4 T Bruker DSX spectrometer equipped with a 4 mm CPMAS probe spinning at a MAS rate of 8 kHz.

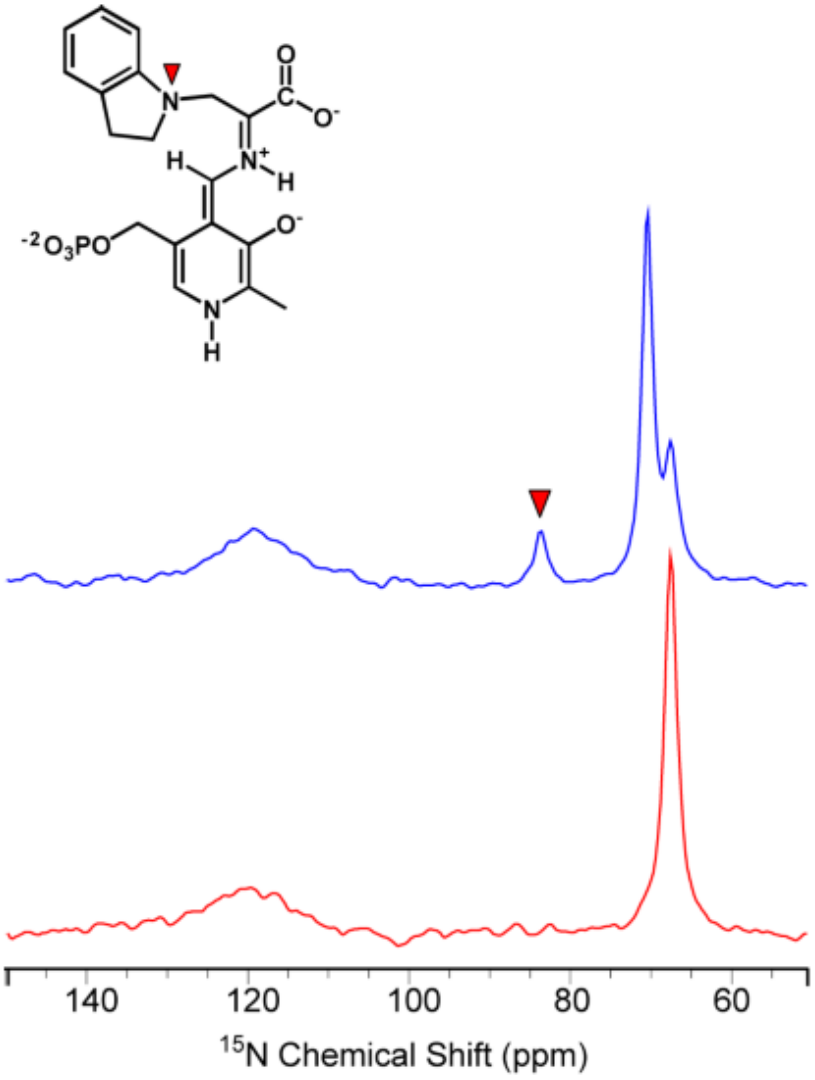


Figure 15. ^{15}N-solid-state NMR CPMAS spectra of the reaction of tryptophan synthase microcrystals with 20 mM L-[U-^{13}C,^{15}N] indoline, 3 mM F9, 100 mM CsCl and no serine (red spectrum) and 20 mM serine (blue spectrum) collected at -10°C and averaged over 9.2 hrs. The resonance at 83.5 ppm (inverted red triangle) is assigned to the indoline quinonoid intermediate, while the peaks at the right arise from non-specifically bound indoline (67.4 ppm) and reaction product (70.2 ppm). Data acquired on a 9.4 T Bruker DSX spectrometer equipped with a 4 mm CPMAS probe spinning at a MAS rate of 8 kHz.

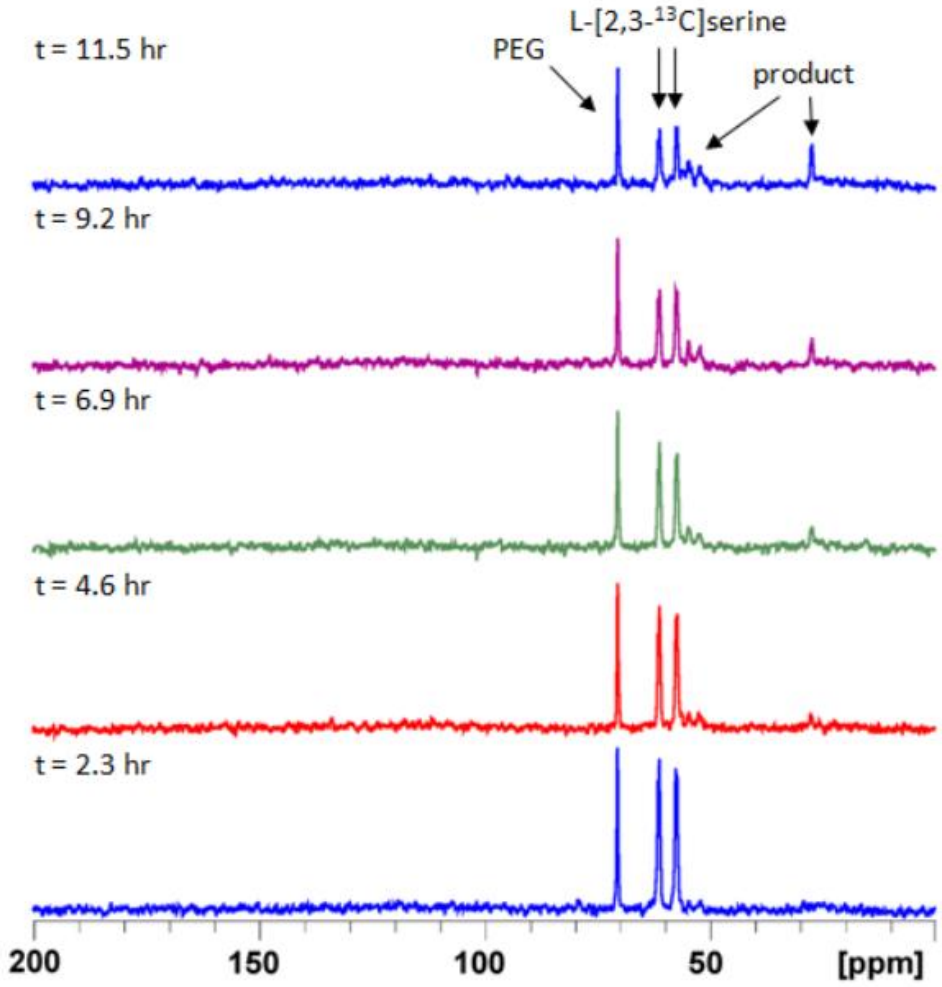
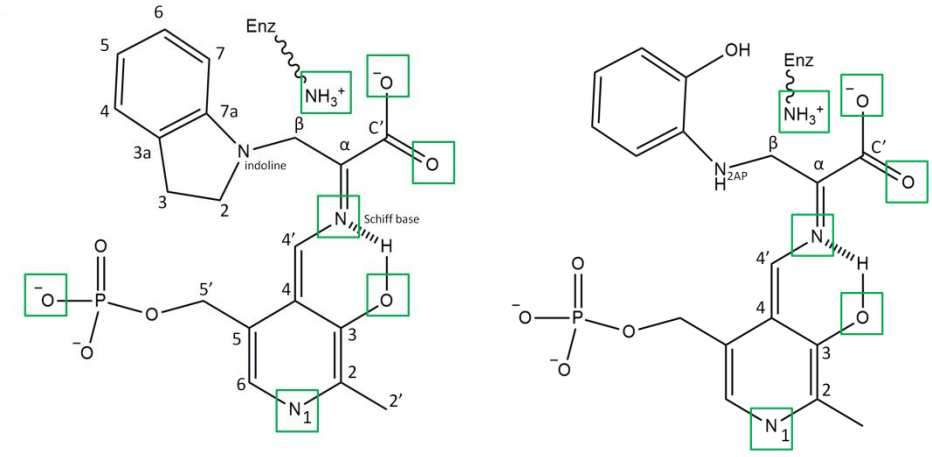


Figure 16. . Direct excitation ^{13}C-solid-state NMR magic-angle-spinning/low power ^{1}H decoupling spectra of the reaction of tryptophan synthase microcrystals with 20 mM indoline, 3 mM F9, 100 mM CsCl and 20 mM L-[2, 3-^{13}C]Ser corresponding to the data shown in Figure 1(b). The time course shows the disappearance of serine and appearance of product signals in the mother liquor. Data collected at -10°C on a 9.4 T Bruker DSX spectrometer equipped with a 4 mm CPMAS probe spinning at a MAS rate of 8 kHz. .

Figure 17. Atom labels and protonation states (boxed) of Ind and 2AP Qin.



Ab initio calculations:

Density functional theory (DFT) using the B3LYP functional within the Gaussian 03[20] software package was used for carrying out calculations. Models of the β-subunit enzyme active site containing residues within 7 Å of the substrate were optimized to their ground state geometries using the ONIOM method[21] (B3LYP/6-31G(d,p):PM3). Bonds that were broken upon this interface were terminated by replacing N-terminal nitrogens with hydrogens and C-terminal carbonyls with carboxamides. In order to provide a framework for optimizing the structure of the ligand, non-hydrogen side chain atoms were fixed at their crystallographically determined coordinates. During geometry optimizations the plane of the indoline ring tilted substantially (>25^{o}) from its orientation in the X-ray crystal structure. Therefore additional constraints were implemented by fixing the dihedral angle formed by each of the indoline benzene carbons with carbons 2, 4, and 6 of the pyridine ring (e.g., dihedral 3a-2-4-6 in Figure 17.).

NMR chemical shifts were calculated[22] at the B3LYP/ 6-311++G(d,p) level. Calculated ^{13}C chemical shifts were referenced via linear regression (CHESHIRENMR.info : slope= -1.03 ,

intercept= 180.62 ppm , carbon nuclide dependent weighting of 4.12) while nitrogens were scaled similarly using the results from chapter VI (CPCM chloroform: slope= -1.05 , intercept= 239.27ppm, nitrogen nuclide dependent weighting of 9.74; measured RMSD of crystal clusters). Protonation states were determined at the following sites: (i) N^{ϵ} of βLys87, (ii) PLP phosphoryl group, (iii) pyridoxyl phenolic oxygen, (iv) Schiff base nitrogen, (v) lower- , (vi) upper-carboxylate oxygen, and (vii) pyridine nitrogen of PLP. The chemical shifts for these structures are summarized in Tables 5 and 7. A shorthand is used to indicate whether a site is protonated ("1" - yes, "0" - no) in the following order: (N^{ϵ} of βLys87)(phosphoryl group[Ind Qin only])-(pyridoxyl phenolic oxygen)(Schiff base nitrogen)-(nearer "lower"-carboxylate oxygen to the Schiff base)(farther "upper"-carboxylate oxygen)-(pyridine nitrogen of PLP). In this nomenclature both species in Figure 17. would correspond to the code sequence: 10-10-00-0.

10-10-00-1	23.6924
10-01-00-1	21.5336
10-00-10-1	12.2411
10-00-01-1	18.2732
10-00-00-1	27.204
10-10-01-1	18.1756
10-01-01-1	25.5024
10-10-00-0	1.55738
10-01-00-0	12.1968
10-00-10-0	2.35536
10-00-01-0	4.09537
10-00-00-0	7.53469
10-10-01-0	1.91299
10-01-01-0	20.8343
00-10-00-1	24.0193
00-01-00-1	20.8945
00-00-10-1	13.0198
00-00-01-1	19.5028
00-00-00-1	31.9572
00-10-01-1	20.9397
00-01-01-1	31.7344
00-10-00-0	2.34947
00-01-00-0	12.1882
00-00-10-0	2.18032
00-00-01-0	5.09829
00-00-00-0	10.5758
00-10-01-0	1.7259
00-01-01-0	24.2471
11-10-00-1	25.9995
11-01-00-1	21.6803
11-00-10-1	12.3601
11-00-01-1	19.1846
11-00-00-1	27.0496
11-10-01-1	13.2425
11-01-01-1	18.1085
11-10-00-0	1.25674
11-01-00-0	10.2418
11-00-10-0	2.36831
11-00-01-0	5.24701
11-00-00-0	8.14591
11-10-01-0	1.49285
11-01-01-0	17.6379
01-10-00-1	30.7798
01-01-00-1	23.7971
01-00-10-1	14.9378
01-00-01-1	22.63
01-00-00-1	53.0121
01-10-01-1	15.6063
01-01-01-1	19.0138
01-10-00-0	1.95767
01-01-00-0	10.3425
01-00-10-0	2.06148
01-00-01-0	5.69102
01-00-00-0	10.0382
01-10-01-0	1.34425
01-01-01-0	16.8655

Table 4. Reduced χ^2 of 56 candidate protonation states of Ind Qin

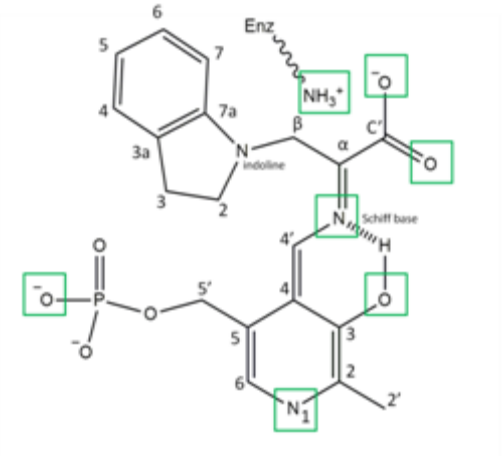


Table 5. Key ab initio chemical shifts for the indoline quinonoid

	$N1_{PLP}$	$C2_{PLP}$	$C3_{PLP}$	$N_{Schiff\ base}$	C^{α}	C'	C^{β}	$N_{indoline}$	C2	C3
10-10-00-1	139.506	148.022	118.222	307.835	123.868	173.241	53.9359	81.7779	51.4755	33.6294
10-01-00-1	164.945	161.389	127.377	206.204	105.638	170.087	55.215	82.252	51.4514	33.725
10-00-10-1	170.207	162.007	131.035	322.571	101.065	167.112	50.9656	88.0412	52.7818	34.3059
10-00-01-1	169.887	160.973	130.317	367.682	101.083	161.897	56.355	97.4762	61.6838	33.1854
10-00-00-1	151.26	161.288	116.132	381.033	111.125	174.927	53.8001	95.29	51.4441	35.33
10-10-01-1	151.387	153.872	120.05	308.798	105.464	162.662	52.8208	83.0018	51.6778	32.6646
10-01-01-1	169.106	165.897	135.193	187.046	93.359	159.539	53.4983	81.6022	51.9927	32.7961
10-10-00-0	256.546	149.576	134.164	311.586	105.851	173.181	56.728	92.0366	51.4947	35.0013
10-01-00-0	273.195	163.973	143.155	197.853	92.853	169.154	58.0806	92.316	51.6274	34.8614
10-00-10-0	273.386	165.652	146.512	294.489	94.7117	164.232	52.412	94.8026	52.9738	35.585
10-00-01-0	271.652	166.066	147.084	330.073	98.0456	158.137	54.2644	95.1324	51.6297	35.4667
10-00-00-0	253.357	164.272	138.248	368.364	103.167	173.174	56.2953	104.678	51.1588	37.5246
10-10-01-0	284.259	151.561	139.213	299.156	97.7265	160.97	54.2735	85.2257	51.8103	33.6669
10-01-01-0	294.83	167.746	148.916	175.158	88.3857	156.741	54.5341	85.3661	51.8576	33.4489
00-10-00-1	141.287	148.647	116.37	312.949	124.053	173.602	53.9606	86.9681	50.2976	33.8753
00-01-00-1	167.563	161.497	126.197	208.101	105.286	170.813	54.4665	88.6801	50.6524	33.7858
00-00-10-1	169.723	162.193	128.665	326.382	102.428	166.908	50.9802	92.2633	52.519	33.7765
00-00-01-1	171.051	162.576	128.628	377.636	104.214	160.144	53.9095	92.5918	51.0431	34.1665
00-00-00-1	146.289	162.193	111.722	389.038	113.024	174.455	53.4783	100.274	50.429	35.2585
00-10-01-1	145.154	151.344	116.511	291.459	104.7	160.972	53.5608	86.1161	51.1365	32.7718
00-01-01-1	161.052	163.233	124.684	176.442	93.3764	157.193	53.057	86.4616	51.3136	32.7338
00-10-00-0	246.874	151.11	131.283	314.018	107.648	173.26	56.3768	95.1506	51.0617	34.4365
00-01-00-0	267.65	165.552	140.021	198.398	94.5262	169.157	56.8993	96.4065	50.5197	34.6013
00-00-10-0	267.489	166.548	142.654	300.189	98.5737	163.545	52.4805	96.1832	51.3239	35.0803
00-00-01-0	265.461	166.572	143.128	342.544	101.785	158.784	55.9368	99.7609	50.8641	35.9636
00-00-00-0	245.721	167.211	130.649	376.043	103.741	172.444	56.3239	109.162	50.8449	37.0188
00-10-01-0	267.398	152.176	132.751	281.081	97.0638	170.025	55.9894	89.0009	50.8139	33.3419
00-01-01-0	278.396	166.726	141.736	158.232	89.229	157.623	54.4871	89.2436	50.9222	33.3621

-Contd. (Table 5. Key ab initio chemical shifts for the indoline quinonoid)

	$N1_{PLP}$	$C2_{PLP}$	$C3_{PLP}$	$N_{Schiff\ base}$	C^{α}	C'	C^{β}	$N_{indoline}$	C2	C3
11-10-00-1	136.106	148.211	120.498	302.505	131.602	171.856	53.7086	79.0647	52.0613	33.2504
11-01-00-1	161.983	160.65	130.55	207.33	111.777	169.029	54.8459	78.2977	51.9815	33.2183
11-00-10-1	167.262	161.398	135.655	328.59	103.991	168.18	50.7136	85.7871	53.2619	33.904
11-00-01-1	168.715	161.396	134.896	378.077	103.179	162.992	56.0681	95.6867	61.0691	32.5743
11-00-00-1	149.947	160.358	118.793	380.758	116.777	174.658	53.5194	91.2422	51.7157	34.4368
11-10-01-1	163.138	150.894	132.028	317.183	111.528	165.196	52.8066	77.3962	52.6099	32.5328
11-01-01-1	177.913	162.659	142.883	205.995	98.3886	161.716	53.0702	76.146	52.7118	32.389
11-10-00-0	255.063	149.318	139.015	312.503	110.617	173.29	56.464	87.9654	51.9901	34.0938
11-01-00-0	270.587	163.043	148.669	204.096	97.1467	169.492	58.1133	87.7025	52.0426	33.8617
11-00-10-0	273.073	164.476	153.205	307.844	95.3614	165.671	52.9116	92.9735	53.4266	34.8091
11-00-01-0	271.204	164.974	153.622	345.016	97.2986	159.636	54.2225	92.4344	51.9358	34.0552
11-00-00-0	254.836	164.007	145.261	376.044	104.677	173.907	56.4934	101.721	51.7058	36.1002
11-10-01-0	284.286	151.205	145.337	308.903	99.4533	162.627	54.0949	82.6302	52.388	32.9563
11-01-01-0	291.915	166.195	155.063	186.403	89.2117	158.554	54.5201	82.5228	52.4709	32.7694
01-10-00-1	126.771	148.431	115.367	308.029	132.923	172.917	53.091	81.9733	50.6672	33.2545
01-01-00-1	155.684	161.023	125.315	209.244	112.287	170.187	53.2229	82.9054	50.6743	33.0844
01-00-10-1	161.559	162.139	129.498	333.378	105.869	168.294	50.3009	89.3014	52.5571	33.4069
01-00-01-1	163.73	162.262	129.525	389.066	106.205	161.56	52.8217	89.2087	51.0306	33.4246
01-00-00-1	96.6336	168.084	103.109	390.086	121.897	175.303	52.5018	94.2876	50.6338	34.2927
01-10-01-1	157.288	152.195	127.059	319.347	112.619	164.888	52.7546	80.4623	51.5459	32.305
01-01-01-1	173.838	163.463	137.728	206.746	99.6656	161.618	52.285	79.7435	51.5884	32.1644
01-10-00-0	246.595	150.792	134.859	314.137	110.469	173.765	55.3013	93.9439	51.023	34.0396
01-01-00-0	266.013	164.333	144.95	204.032	96.8286	170.133	56.087	94.5122	51.1116	33.9363
01-00-10-0	267.894	165.554	147.746	308.529	97.5231	165.17	52.0662	97.4897	52.7285	34.1021
01-00-01-0	267.071	165.918	148.313	350.393	100.433	158.617	55.0024	96.9402	51.1162	34.4036
01-00-00-0	250.823	167.221	139.047	383.402	105.429	174.785	56.5842	98.1971	51.1183	34.9401
01-10-01-0	277.441	152.829	141.609	308.854	101.039	161.806	54.0061	86.7631	51.418	32.9071
01-01-01-0	286.555	167.025	151.659	186.681	91.1627	158.1	53.5848	86.6366	51.4392	32.7659

10-10-00-1	25.1833
10-01-00-1	25.5025
10-00-10-1	14.7766
10-00-00-1	31.0934
10-10-01-1	12.2314
10-01-01-1	21.7742
10-00-01-1	20.6736
00-10-00-1	29.8624
00-01-00-1	26.2415
00-00-10-1	18.114
00-00-00-1	37.6257
00-10-01-1	15.1084
00-01-01-1	21.7911
00-00-01-1	24.0965
10-10-00-0	1.02229
10-01-00-0	13.5521
10-00-10-0	2.22177
10-00-00-0	6.87076
10-10-01-0	2.03989
10-01-01-0	23.1183
10-00-01-0	3.99399
00-10-00-0	1.49721
00-01-00-0	12.7315
00-00-10-0	2.30372
00-00-00-0	35.5612
00-10-01-0	1.47382
00-01-01-0	21.8453
00-00-01-0	6.03768

Table 6. Reduced χ^2 of all 28 candidate protonation states of 2AP Qin

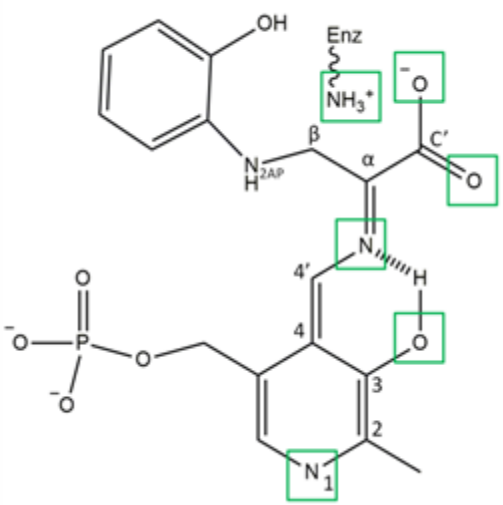


Table 7. Key ab initio chemical shifts for the 2AP Qin

	N_{2AP}	N_{pn}	$C2_{PLP}$	$C3_{PLP}$	$N_{Schiff\ base}$	C^{α}	C'	C^{β}
10-10-00-1	57.3529	146.453	151.471	121.601	291.014	123.926	173.197	49.9267
10-01-00-1	56.7169	165.895	160.834	129.371	211.022	109.265	170.695	51.2907
10-00-10-1	59.6383	171.141	162.092	132.241	328.891	104.591	167.634	47.2757
10-00-00-1	65.9666	152.248	160.147	117.336	374.448	114.613	175.704	50.1954
10-10-01-1	55.6403	174.178	153.472	134.084	309.379	108.65	166.758	47.8446
10-01-01-1	54.7147	184.582	163.339	141.134	205.877	97.9161	163.6	48.5322
10-00-01-1	61.3646	170.538	161.268	131.661	368.357	104.065	163.271	48.7782
00-10-00-1	62.2743	138.023	150.76	119.457	301.71	127.972	173.696	47.6368
00-01-00-1	60.4017	160.486	160.624	127.014	216.629	110.881	171.93	48.6165
00-00-10-1	61.6099	165.798	161.847	129.817	343.361	105.793	166.934	45.8635
00-00-00-1	71.7742	146.237	160.307	113.57	387.039	116.101	175.354	47.8573
00-10-01-1	57.2043	166.581	153.426	131.455	317.35	111.319	166.569	48.0613
00-01-01-1	56.0395	180.28	163.357	139.232	209.244	99.122	164.417	48.0536
00-00-01-1	64.2799	168.337	161.105	130.055	380.189	105.44	162.337	49.3228
10-10-00-0	62.5932	261.66	150.242	138.17	305.348	109.596	174.301	54.0501
10-01-00-0	61.2395	275.007	162.771	146.618	205.069	96.5515	171.163	55.8176
10-00-10-0	63.1072	273.493	164.763	149.34	296.552	98.1529	165.101	49.8711
10-00-00-0	68.3227	253.281	163.458	141.145	361.388	107.781	174.848	53.8087
10-10-01-0	59.387	293.184	151.672	144.18	296.239	99.9536	163.707	50.8592
10-01-01-0	58.6762	297.577	165.606	152.069	184.856	89.6856	160.23	51.221
10-00-01-0	64.6644	271.408	164.29	149.478	329.495	99.8319	160.697	51.4915
00-10-00-0	66.8758	254.262	151.106	136.768	316.183	109.362	174.389	51.585
00-01-00-0	64.5249	270.645	163.673	145.975	207.731	95.9868	173.508	52.9834
00-00-10-0	65.6141	269.323	165.076	148.079	306.504	98.7998	163.899	48.3649
00-00-00-0	76.084	253.274	167.234	145.51	434.76	136.394	180.74	60.2926
00-10-01-0	61.1389	283.1	152.763	142.527	305.74	101.132	163.254	50.2482
00-01-01-0	61.122	292.982	166.258	151.011	187.141	89.9594	161.077	50.289
00-00-01-0	69.947	266.684	165.442	149.158	344.205	101.42	159.267	52.8644

Reduced Chi-Squared Analysis

We quantify the comparison between experimental and calculated chemical shifts for the candidate structures by the reduced χ^2 statistic, the weighted deviation of the model and experimental shifts:

$$\chi^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{calc} - \delta_i^{\exp}\right)^2}{\sigma_i^2}$$

Here N is the number of measured shifts, δ_i^{calc} is the calculated chemical shift (ppm) at site i for the candidate structure, δ_i^{exp} is the experimental chemical shift (ppm) at site i, and σ_i^2 is the nuclide-dependant weighting of the shifts, corresponding to root-mean-squared-deviations (RMSD), σ, of 4.12 ppm for ^{13}C (CHESHIRE) and 9.75 ppm (see chapter VI) for ^{15}N.[24] The reduced chi-squared values for candidate structures are given in Table 6. . Of the candidate structures, ##-10-0#-0 (where # represents 1 or 0) are found to have the closest agreement between the calculated and experimental ^{13}C and ^{15}N chemical shifts for Ind Qin and 2AP.

Figure 18. Bar Graph representation of the reduced χ^2 values for all 56 candidate protonation states of Ind Qin

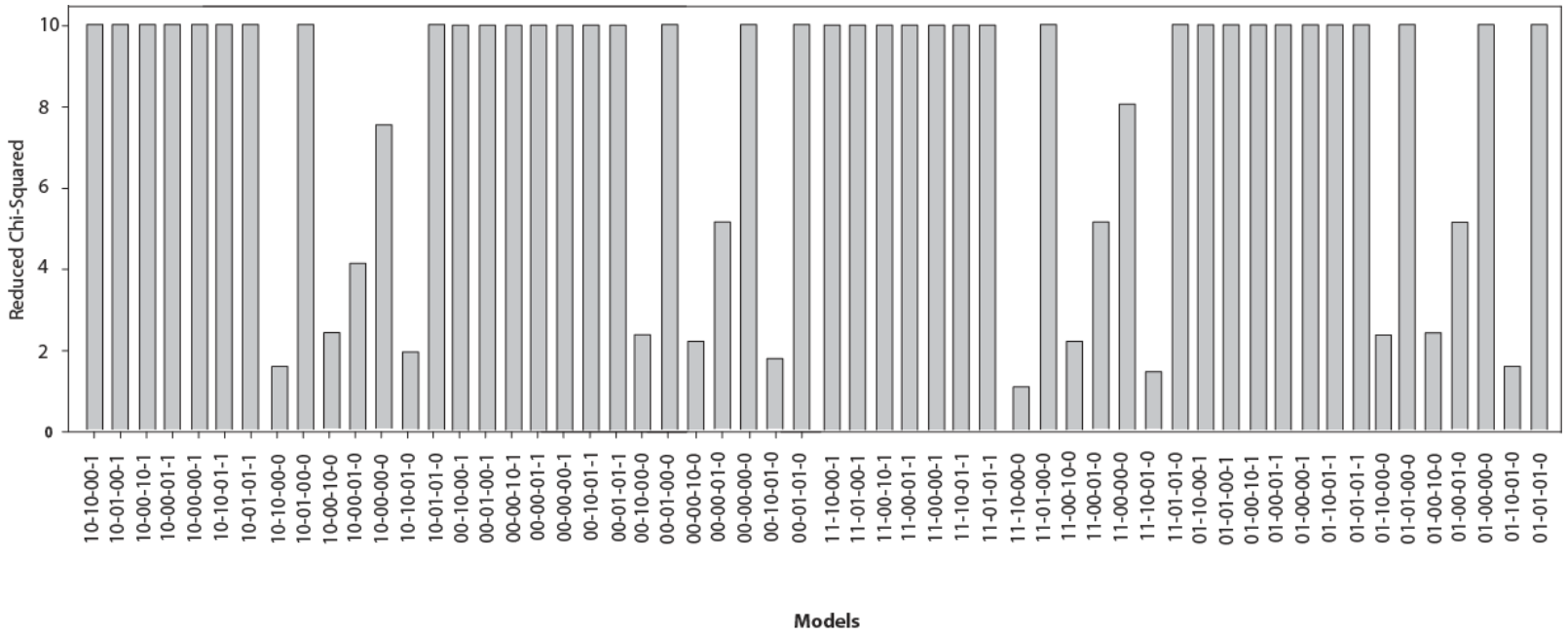
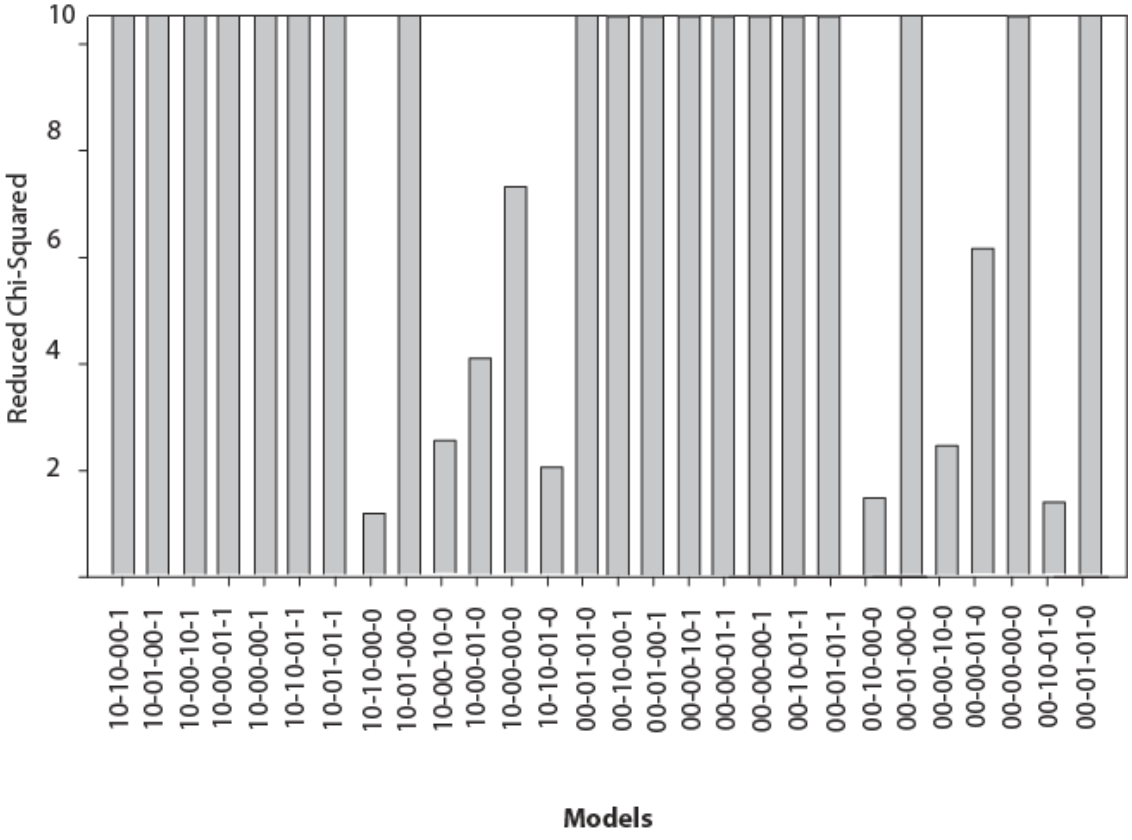


Figure 19. Bar Graph representation of the reduced χ^2 values for all 28 candidate protonation states of 2AP Qin



CONCLUSION

NMR-Crystallography was employed to characterize the enzymatic indoline/2AP quinonoid intermediates in the PLP-dependent tryptophan synthase. It was shown that various models of charge and protonation states can be distinguished by their calculated effect on the observed chemical shifts, ultimately emerging a chemical species most consistent with the experimental results. These support the phenolic/phenolic-acid models.

REFERENCES

(1) Schneider, T. R.; Gerhardt, E.; Lee, M.; Liang, P. H.; Anderson, K. S.; Schlichting, I. Biochemistry-Us 1998, 37, 5394-5406.

(2) Barends, T. R. M.; Domratcheva, T.; Kulik, V.; Blumenstein, L.; Niks, D.; Dunn, M. F.; Schlichting, I. Chembiochem 2008, 9, 1024-1028.

(3) Kabsch, W. Journal of Applied Crystallography 1993, 26, 795-800.

(4) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Acta Crystallographica Section D-Biological Crystallography 1998, 54, 905-921.

(5) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Acta Crystallographica Section D-Biological Crystallography 1997, 53, 240-255.

(6) Ngo, H.; Kimmich, N.; Harris, R.; Niks, D.; Blumenstein, L.; Kulik, V.; Barends, T. R.; Schlichting, I.; Dunn, M. F. Biochemistry-Us 2007, 46, 7740-7753.

(7) Mozzarelli, A.; Peracchi, A.; Rovegno, B.; Dale, G.; Rossi, G. L.; Dunn, M. F. Journal of Biological Chemistry 2000, 275, 6956-6962.

(8) Ngo, H.; Harris, R.; Kimmich, N.; Casino, P.; Niks, D.; Blumenstein, L.; Barends, T. R.; Kulik, V.; Weyand, M.; Schlichting, I.; Dunn, M. F. Biochemistry-Us 2007, 46, 7713-7727.

(9) Dunn, M. F.; Niks, D.; Ngo, H.; Barends, T. R. M.; Schlichting, I. Trends in Biochemical Sciences 2008, 33, 254-264.

(10) Lane, C. F. Synthesis-Stuttgart 1975, 135-146.

(11) Roy, M.; Keblawi, S.; Dunn, M. F. Biochemistry-Us 1988, 27, 6698-6704.

(12) Yang, L. H.; Ahmed, S. A.; Miles, E. W. Protein Expression and Purification 1996, 8, 126-136.

(13) Mozzarelli, A.; Peracchi, A.; Rossi, G. L.; Ahmed, S. A.; Miles, E. W. Journal of Biological Chemistry 1989, 264, 15774-15780; Tian, Y.; Chen, L. L.; Niks, D.; Kaiser, J. M.; Lai, J. F.; Rienstra, C. M.; Dunn, M. F.; Mueller, L. J. Phys Chem Chem Phys 2009, 11, 7078-7086. Morcombe, C. R.; Zilm, K. W. J Magn Reson 2003, 162, 479-486.

(14) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wuthrich, K. Pure and Applied Chemistry 1998, 70, 117-142.

(15) Hayashi, S.; Hayamizu, K. Bulletin of the Chemical Society of Japan 1991, 64, 688-690.

(16) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs, 2005; Vol. 26; Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Journal of Molecular Graphics & Modelling 2006, 25, 247-260.

(17) Song, Y. F.; Gunner, M. R. J Mol Biol 2009, 387, 840-856.

(18) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Journal of Computational Chemistry 2005, 26, 1781-1802.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; J. A. Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; B.05 ed.; Gaussian, Inc.: 2003.

(20) Vreven, T.; Morokuma, K. Journal of Computational Chemistry 2000, 21, 1419-1432.

(21) Oldfield, E. Philosophical Transactions of the Royal Society B-Biological Sciences 2005, 360, 1347-1361; Oldfield, E. Annual Review of Physical Chemistry 2002, 53, 349-378; Oldfield, E. J Biomol Nmr 1995, 5, 217-225; Le, H. B.; Pearson, J. G.; Dedios, A. C.; Oldfield, E. J Am Chem Soc 1995, 117, 3800-3807.

(22) Cai, L.; Fushman, D.; Kosov, D. S. J Biomol Nmr 2008, 41, 77-88.

(23) Pearson, J. G.; Wang, J. F.; Markley, J. L.; Le, H. B.; Oldfield, E. J Am Chem Soc 1995, 117, 8823-8829; Dedios, A. C.; Oldfield, E. J Am Chem Soc 1994, 116, 5307-5314; Szabo, C. M.; Sanders, L. K.; Le, H. C.; Chien, E. Y. T.; Oldfield, E. Febs Lett 2000, 482, 25-30.

CHAPTER IV. PROTONATION STATE OF THE TRYPTOPHAN SYNTHASE INTERNAL ALDIMINE ACTIVE SITE FROM SOLID-STATE NMR SPECTROSCOPY AND AB INITIO COMPUTATIONAL CHEMISTRY.

ABSTRACT

Previously discussed in chapter II, the acid-base chemistry that drives catalysis in pyridoxal-5'-phosphate (PLP) dependent enzymes has been the subject of intense interest and investigation since the initial identification of PLP's role as coenzyme in this extensive class of enzymes. It was first proposed over 50 years ago that the initial step in the catalytic cycle is facilitated by a protonated Schiff base form of the holoenzyme, in which the linking lysine ε-imine nitrogen, which covalently binds the coenzyme, is protonated. Here we provide the first ^{15}N NMR chemical shift measurements of such a Schiff base linkage in the resting holoenzyme form, the internal aldimine state of tryptophan synthase. Double-resonance experiments confirm the assignment of the Schiff base nitrogen, and additional ^{13}C, ^{15}N, and ^{31}P chemical shift measurements of sites on the PLP coenzyme allow a detailed model of coenzyme protonation states to be established. Furthermore computational models of the active site were developed using a synergistic approach in which the structure of the resting holoenzyme were optimized using ab initio computational chemistry in the presence of side-chain residues fixed at their crystallographically determined coordinates. Various models of charge and protonation state for the covalently bound substrate and nearby catalytic residues were uniquely distinguished by their calculated effects on the chemical shifts, ranked by reduced χ^2 analysis, in comparison to the experimentally measured ^{13}C- and ^{15}N-labeled positions on the substrate. The protonated Schiff base (PSB) falls 3.4 units outside of the 95% confidence limit with 4 data points;

superseding however all other models with a minimum difference of reduced $\chi^2 \approx +10$ units to the next nearest model (Figure 24.). The established PSB further supports the importance of the substrate's role in directing the next step of catalysis.

INTRODUCTION

Pyridoxal-5'-Phosphate (PLP), the bioactive form of vitamin B_6, acts as coenzyme in multiple amino acid transformations, including α/β/γ elimination/replacement, racemization, transamination, and decarboxylation.[1,2] At the start of the catalytic cycle, the cofactor is bound through the ε-nitrogen of a lysine side-chain, forming a Schiff base linkage (the internal aldimine state; Figure 20.). It has been proposed that a positively charged, protonated Schiff base (PSB) tautomer activates catalysis at this point by forming a significantly more reactive target for nucleophilic attack than the neutral imine,[3-5] while the protonation states of other sites on the coenzyme - the phenolic oxygen and PLP ring nitrogen in particular - are thought to be critical in establishing the specificity of the reaction pathway.[2,6-8] The experimental determination of protonation states within PLP active sites is typically indirect; the two most common methods of characterization, X-ray crystallography and optical spectroscopy, cannot specifically identify proton locations or report unambiguously on the local chemical environment of individual atoms. Nuclear magnetic resonance (NMR) spectroscopy in tandem with ab initio computational chemistry can provide atomic-resolution characterization, but to date only a handful of NMR and computational studies have been reported for PLP- dependent enzyme active sites.[9-13]

Here we detail the first ^{15}N NMR chemical shift measurements of the Schiff base linkage in the resting holoenzyme state of a PLP-dependent enzyme, the tryptophan synthase internal aldimine complex; these provide direct confirmation of the protonated Schiff base tautomer. Additional ^{13}C, ^{15}N, and ^{31}P chemical shift measurements on the coenzyme allow a chemically-detailed model including all coenzyme protonation states to be established and additionally validated with ab initio calculations.

The tryptophan synthase (TS) $\alpha_2\beta_2$ bienzyme complex relies on PLP to bring together indole and L-serine to form L- tryptophan.[14-17] Figure 20. shows a schematic of the β-subunit active site for the internal aldimine state, E(Ain), with PLP covalently bound to βLys87. In TS (as with all PLP- dependent enzymes), the first step in catalysis is a transimination reaction in which an amino acid substrate makes a nucleophilic attack at C_4' of PLP; this step exchanges the PLP Schiff base linkage to the protein with one to the substrate. The PSB hypothesis posits that the imine nitrogen should be protonated for reactivity toward nucleophiles at the Schiff base carbon, a hypothesis first proposed to explain the remarkable acceleration of enzyme catalyzed reactions compared to PLP catalyzed reactions in solution.[3]

MATERIALS AND METHODS

UV-VIS

UV/vis optical spectroscopy of TS supports the PSB hypothesis. Studies of enzyme internal aldimine states and model compounds with Schiff base linkages to peptides and amino acids in polar, aprotic solvents indicate a conjugation best explained by a protonated imine nitrogen.[4,5,18] The internal aldimine complexes of most PLP enzymes give absorption maxima in the 420 to 430nm range, and the 412nm λ_{max} exhibited by the TS E(Ain), both in solution and in single crystals, correlates well with the postulate of a Schiff base structure conjugated with the PLP π-system.[19] The UV/vis spectrum of TS is independent of pH over the range 5.8 - 10.4,[19-21]

and consequently there is a single ionic form of the E(Ain) β-site and no change in the distribution of tautomeric structures in this pH range.[22] Deprotonation of PLP enzyme Schiff bases causes shifts to shorter wavelength (~ 360nm),[23] while deprotonation of the internal aldimine PLP ring nitrogen has only minor effects on the UV/vis spectrum.[2]

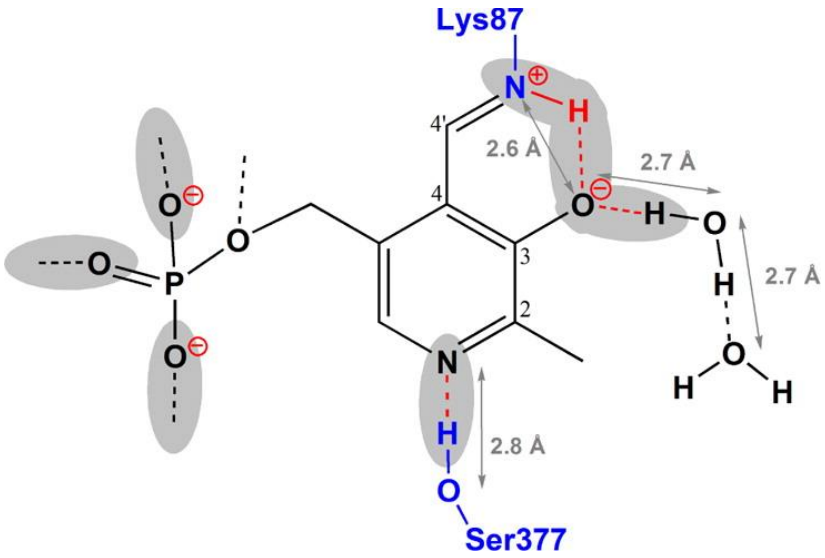


Figure 20. Schematic of PLP covalently bound to βLys87 in the β-subunit active site of tryptophan synthase. Distances and water molecules are from the X-ray crystal structure of the S. typhimurium TS internal aldimine state (PDB code 4HT3). Possible sites of protonation/ionization on the coenzyme and linking imine nitrogen are indicated in gray, along with hydrogen bonds to active-site residues/water. The experimentally determined protonation states from solid-state NMR spectroscopy are highlighted in red. Protein residue fragments are shown in blue and PLP and water in black.

X-RAY

The X-ray crystal structures of TS internal aldimine complexes show the Schiff base nitrogen in close proximity to the 3' oxygen of the PLP ring.[25-30] Key distances from the 1.30 Å resolution X-ray crystal structure of the *S. typhimurium* TS E(Ain) (reported recently by several of us;[30] PDB accession code 4HT3) are shown in Figure 20. The 2.6 Å N-O distance is fully consistent with the presence of a proton in an N-H-O hydrogen bond, but the assignment of donor and acceptor in this bonding system cannot be established directly from the crystal structure. Also apparent, the Schiff base nitrogen is twisted slightly out of the PLP plane, with a C_3-C_4-C_4'-N dihedral angle of 27.2° (consistent with the slight red shift in the UV/vis absorption maximum), and the phenolic oxygen is involved in a hydrogen bonding network with two crystallographically observed water molecules, indicated in the figure.

NMR

While X-ray crystallography and UV/vis spectroscopy suggest a protonated Schiff base form, atomic-resolution probes, such as NMR spectroscopy, have not been applied to directly characterize the linking lysine ε-imine nitrogen in a PLP-dependent enzyme. NMR chemical shifts are particularly sensitive to local chemical structure; protein and model compound Schiff base nitrogen atoms show changes in chemical shifts greater than 100 ppm upon protonation.[31] NMR spectroscopy on PLP model compounds by Limbach and co-workers has demonstrated that the tautomeric form favored can depend upon the substituent on the Schiff base nitrogen, solute-

solvent interactions, the protonation state of the pyridine nitrogen, and hydrogen bonding to the phenolic oxygen.[31-34] For PLP-dependent enzymes, pyridine nitrogen and select carbon atom chemical shifts have been reported for internal aldimines of alanine racemase and aspartate aminotransferase,[11,13] and nitrogen linkages to covalently bound substrates in TS[12] and alanine racemase.[9] Yet no chemical shift measurement for a linking imine nitrogen has been reported for any internal aldimine complex. The challenge to interrogating the Schiff base nitrogen is resolution of this single site within the forest of other peaks. Substrate and coenzyme studies rely on the selective introduction of ^{13}C and ^{15}N-isotopically enriched components to achieve specificity, an approach that is not generally applicable to side-chain sites. Fortunately, the distinct ε-imine chemical shift of a lysine residue covalently bound to the coenzyme can provide resolution of this single site in labeled protein preparations, an approach exploited to study the Schiff base linkage to retinal in bacteriorhodopsin.[35]

AB INITIO

Computational calculations were carried out on the internal aldimine intermediate utilizing the protocol outlined in chapters II and III. Two exceptions to the modeling exist. First, using an ONIOM approach for the NMR chemical shielding calculations was employed to allow a more complete clustering of the ligand while applying a moderate basis set and level of theory (DFT B3LYP/6-31G) around the phosphoryl group and outer regions of the substrate, yielding ~240 atoms total with ~140 in the High Layer; job keyword # ONIOM B3LYP/6-311++G(d,p) : B3LYP/6-31G; DFT in gaussian09. Secondly, because this intermediate does not contain a

carboxylate group and accepted evidence[39] has supported a purely dianionic phosphoryl exceptions to this protocol, regarding all available sites of protonation have been made. This is illustrated in Figure 3. which shows all possibilities for sites of protonation used to generate the candidate structures for the PLP-ligand complex including: the Schiff base linkage, phenolic oxygen, and pyridine nitrogen of PLP eliciting a three digit code I.E. 100 in an analogous fashion revealed in the previous chapters. Interestingly, two specific models (namely 010 and 011), which include a protonated phenolic oxygen, consistently donated it's proton to the local Schiff base nitrogen in all cases where the nitrogen was initially deprotonated during the optimization (Figure 21.), effectively lowering the overall energy by forming a protonated Schiff base species required to converge the system. This phenomena also occurred in cases where the the phenolic proton was deliberately rotated 90°, 180°, and 270° from the plane and *further* while the strict convergence criteria threshold was lifted, using the widely accepted command for difficult convergences cases; SCF=QC (Quadratically Convergent), in Gaussian09.

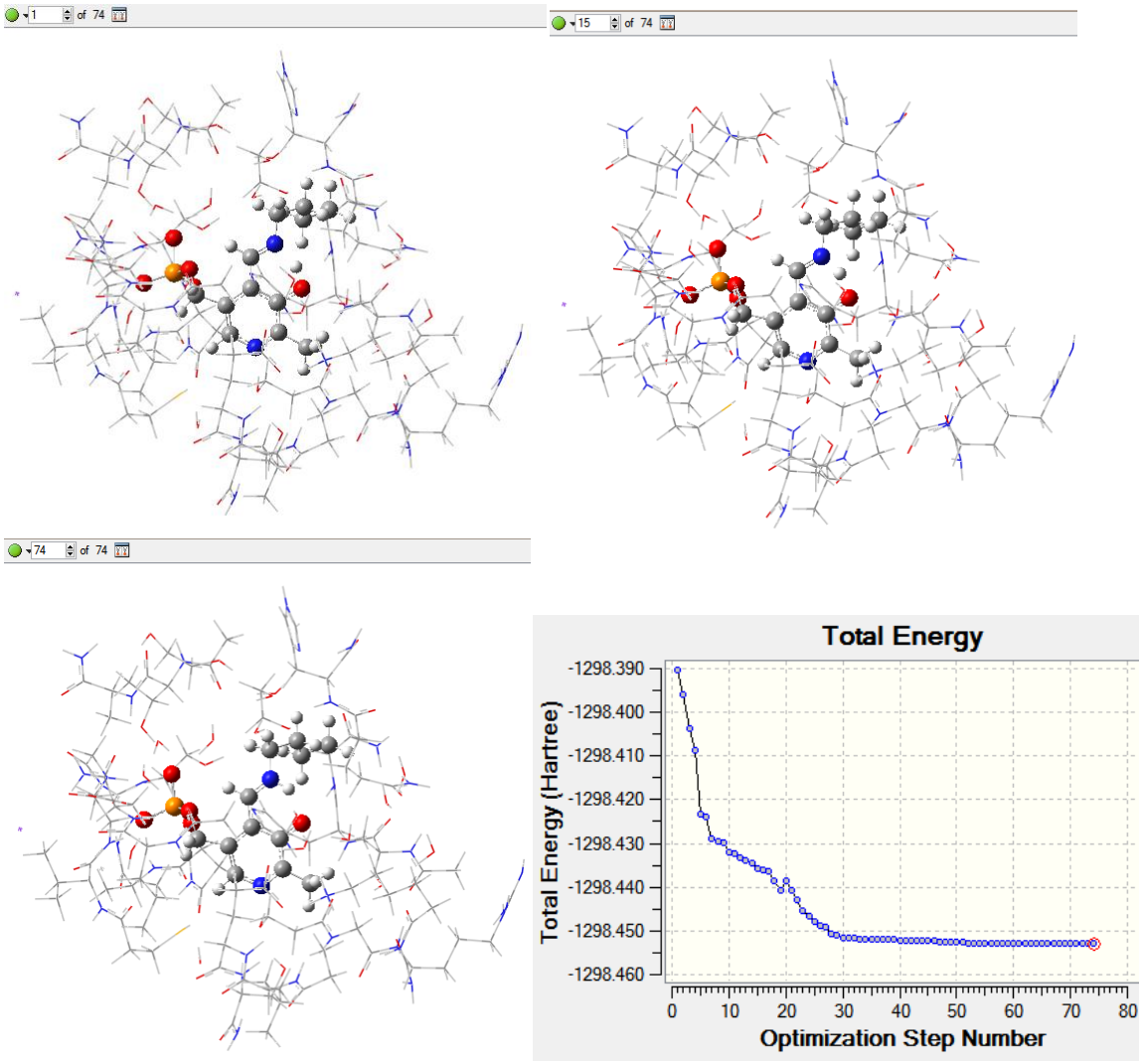
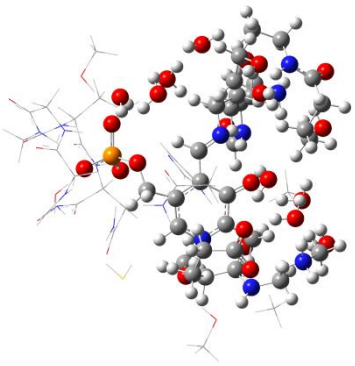


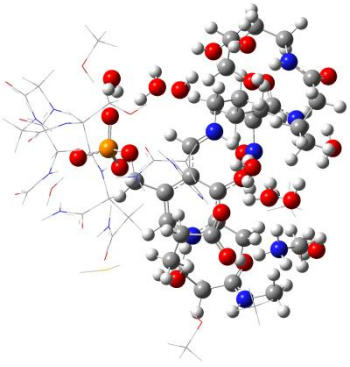
Figure 21. Protonated phenolic oxygen consistently donated it's proton to the local Schiff base nitrogen in all cases where the nitrogen was initially deprotonated during the optimization, effectively lowering the overall energy by forming a protonated Schiff base species required to converge the system.

Overall 6 candidate structures were yielded and the complete set of models are shown in Figure 22. and their respective NMR chemical shifts (Table 8.) were directly compared to experimental values and evaluated using reduced χ^2 statistical analysis where an acceptable range for the 95% confidence interval for 4 data points is 0.101 - 2.295[24].

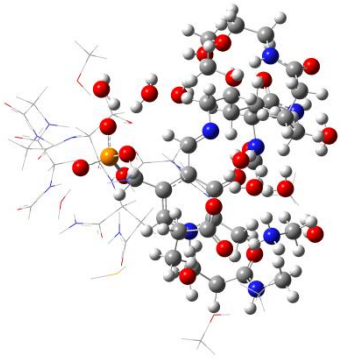
Figure 22. 6 candidate protonation states of the Internal aldimine [### ; 1 = yes, 0 = no; for the presence of a proton at locations: Schiff base, phenolic oxygen, pyridine nitrogen of PLP]



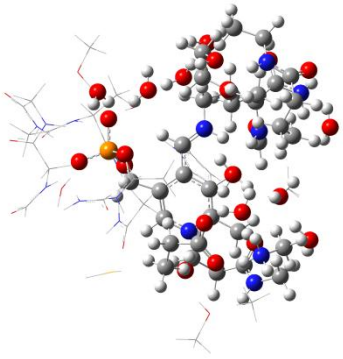
000



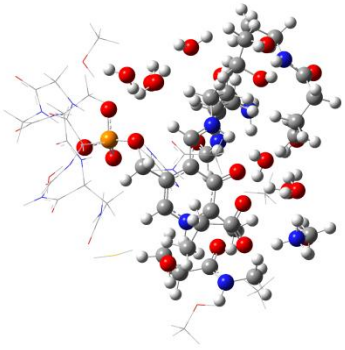
100



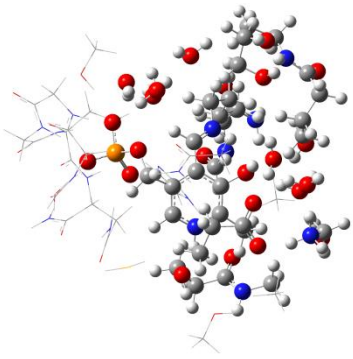
001



110



101



111

Table 8. Experimental and calculated chemical shifts [in ppm] relative to TMS (^{13}C) and $NH_3(l)(^{15}N)^a$] for internal aldimine

	Exp.	Calculated Protonated Schiff Base Form 100
$N_{schiffbase}$	202.3	305.3
$N1_{PLP}$	294.7	261.7
$C2_{PLP}$	158.0	150.2
$C3_{PLP}$	168.7	138.2

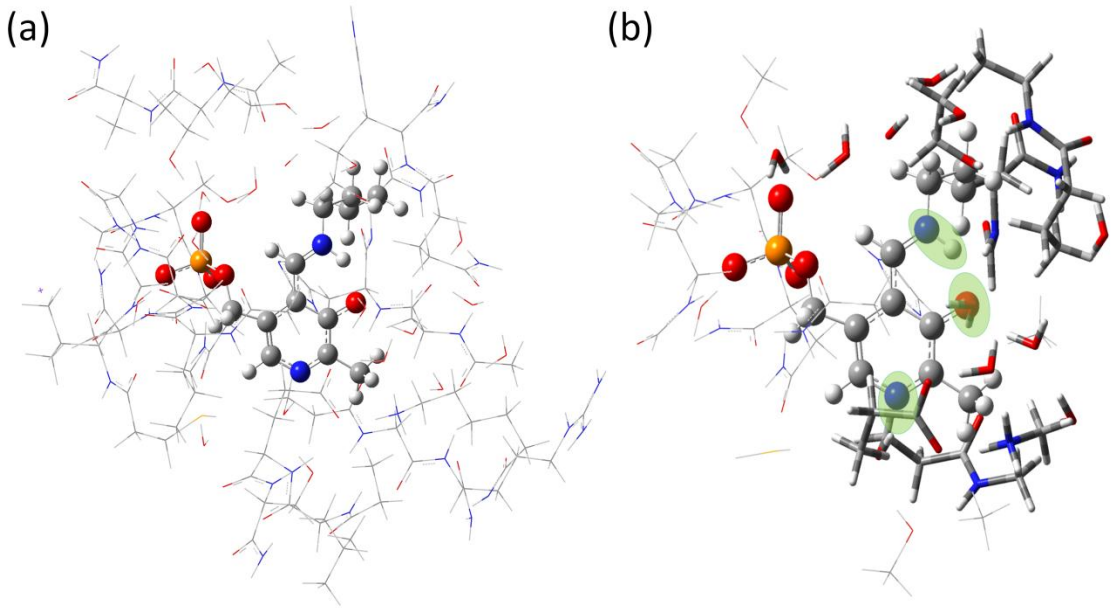
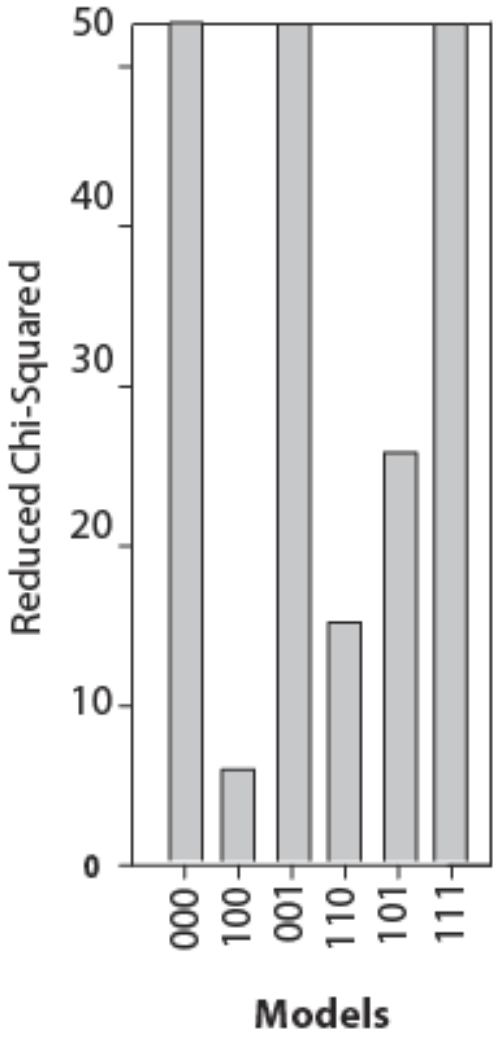
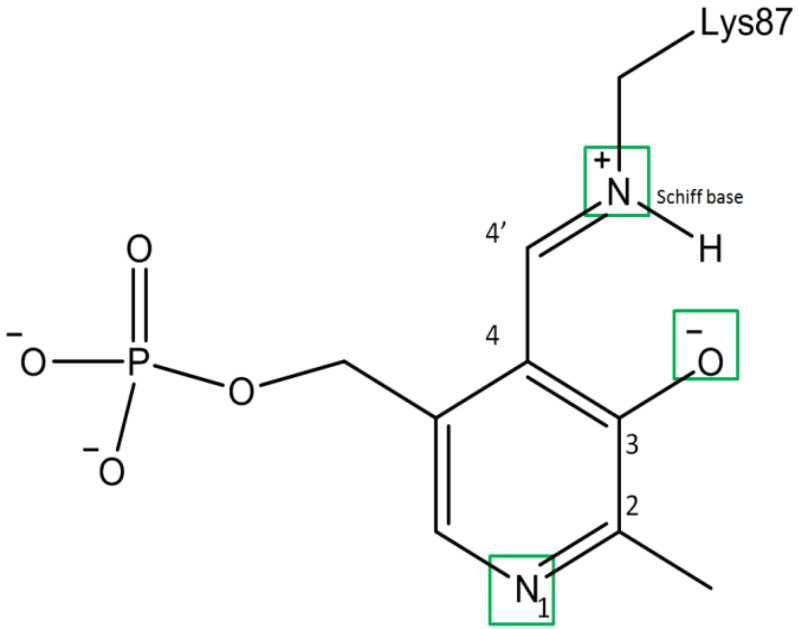


Figure 23. (a) Model of the enzyme active site in the β-subunit of tryptophan synthase, showing (wireframe) the side chains fixed at their crystallographically determined coordinates and treated at the semiempirical PM3 level of theory and (ball and stick) the PLP–substrate complex optimized using DFT at the B3LYP/6-31G(d,p) level. (b) Substructure used for calculating NMR chemical shifts [ONIOM "ball and stick" and "stick" B3LYP/6-311++G(d,p) : "Thin wire frame" (most noticeably around phosphoryl group) B3LYP/6-31G DFT]. Green ovals indicate possible sites of protonation on the PLP–substrate complex. The standard CPK scheme is used to designate the atom colors (H, white; C, gray; N, blue; O, red; P, orange).

Figure 24. Bar chart representation of the reduced χ^2's for all 6 candidate structures of Ain





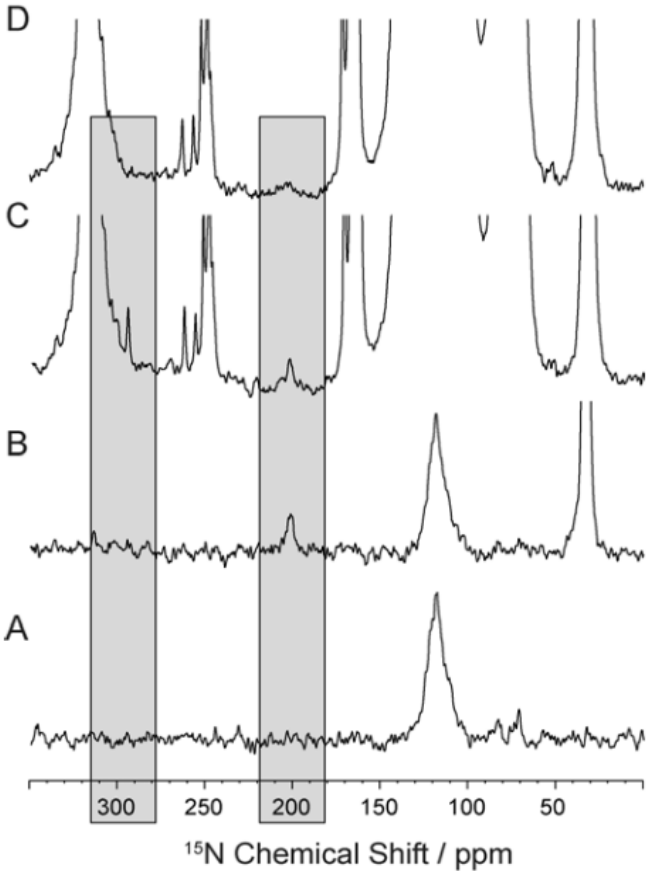
METHODS AND RESULTS

Figures 25. shows ^{15}N solid-state NMR spectra of catalytically active *S. typhimurium* tryptophan synthase microcrystals prepared under the three following conditions: (A) natural abundance isotopomer concentration, (B) selectively ^{15}N- enriched at lysine ε-nitrogen sites (ε-^{15}N-Lys TS), and (2C/2D) uniformly-^{15}N-enriched at protein sites and selectively $^{13}C/^{15}N$ enriched at C_2, C_2', C_3 and N_1 of the PLP coenzyme (U-^{15}N-TS/$_{2,2',3}$-$^{13}C_3$;^{15}N-PLP). The protocols for the synthesis of $_{2,2',3}$-$^{13}C_3$;^{15}N-PLP (adapted from reference [34]) and the preparation of protein samples are given in the supplemental information within chapter V.

At natural abundance, only signal from the large number of protein backbone nitrogens is observed. Upon incorporation of selectively enriched -^{15}N-Lys, two new spectral features are revealed: a large number of mostly overlapped resonances centered at 33 ppm (δ[NH3(l)] scale), and a single resonance at 202.3 ppm. The former correspond to charged ε-amino groups on the labeled lysine residues that have been incorporated into the protein, whereas the latter resonates at the position of a protonated Schiff base[31] and is tentatively assigned to N^{ε} of βLys87, the active-site residue that covalently binds PLP. This assignment is confirmed in two ways. First, the Schiff base linkage to βLys87 is broken by addition of the substrate L-serine, which reacts to form an external aldimine intermediate that subsequently loses water to give an aminoacrylate species;[14] upon addition of 5 μL of 1.2 M serine directly to the microcrystalline sample used to obtain the spectrum in Figure 27., the peak at 202.3 ppm is lost and a new peak at 24.2 ppm appears, suggestive of a neutral amino lysine sidechain for the aminoacrylate intermediate (Figure 25.). Second, rotational-echo double-resonance (REDOR)[36] experiments are used to specifically edit out (dephase) ^{15}N resonances that are dipolar coupled to ^{13}C atoms at

the $_2$, $_{2'}$, and $_3$ positions on the PLP ring. As the dipolar coupling falls off as the inverse cube of the interatomic distance, the $^{15}N\{^{13}C\}$-REDOR editing used here (with 10 ms of dipolar dephasing) is selective for nitrogen atoms within ~3-4 Å of the PLP ^{13}C atoms. The spectra in Figure 25. form a REDOR S_o and S pair: both have a 10 ms echo period on ^{15}N before detection, but differ in the application of dipolar dephasing to ^{13}C for the latter. As expected, there are considerably more peaks in U-^{15}N-TS spectra than in the ε-^{15}N- Lys TS spectrum; there is also remarkable resolution of many individual nitrogen sites. The peak at 202.3 ppm is evident in the REDOR S_o spectrum, but is selectively dephased under dipolar couplings to ^{13}C in S. As this peak arises from a lysine ε-nitrogen, we can conclude that this is the resonance of the Schiff base linkage to PLP and, based on its chemical shift,[31] that it is protonated.

Figure 25. ^{15}N-solid-state NMR cross-polarization magic-angle-spinning (CPMAS) spectra of the tryptophan synthase internal aldimine complex used to assign ^{15}N chemical shifts to the linking lysine ε-imine nitrogen (202.3 ppm) and N1 of PLP (294.7 ppm). Data were acquired on microcrystalline samples of *S. typhimurium* TS prepared with (A) TS at natural abundance isotopomer concentration, (B) ε-^{15}N-Lys TS, and (C/D) U-^{15}N TS/2,2',3-$^{13}C_3$;^{15}N PLP. A and B correspond to direct observation after cross-polarization from ^{1}H to ^{15}N, while C and D form a ^{15}N{^{13}C}-REDOR pair; both have a 10 ms echo period after cross-polarization, but differ in the application of π pulses to ^{13}C (at the quarter and three-quarter mark of each rotor period) in D. Experiments were performed at 9.4 T (400.37 MHz ^{1}H, 100.69 MHz ^{13}C, 40.57 MHz ^{15}N) on a Bruker AVIII spectrometer equipped with an ^{1}H-^{13}C-^{15}N triple resonance 4 mm MAS probe, spinning at a MAS rate of 8 kHz, and with the bearing gas cooled to -15°C, giving an effective sample temperature of -5°C. Cross-polarization was accomplished at a ^{1}H spin-lock field of 45 kHz and ^{15}N spin-lock of 37 kHz (ramped +/- 2 kHz); 85 kHz Spinal64 ^{1}H decoupling was used throughout. 14 μs π pulses were applied to both ^{13}C and ^{15}N during the REDOR echo periods. Each spectrum consists of the sum of 81,920 transients acquired with a relaxation delay of 4s, for a total acquisition time of 3 d 19 h; the REDOR *S* and S_0 spectra were acquired in an interleaved fashion. ^{15}N chemical shifts were referenced indirectly to liq-NH_3 via an external solid-state sample of $^{15}NH_4Cl$, calibrated under MAS conditions as described in the SI. For comparison, $\delta[NH_3(l)] = \delta[^{15}NH_4Cl\ (s)]+39.2$ ppm.



NMR DISCUSSION

Examination of Figure 25. (spectra C and D) show a second peak at 294.7 ppm that is also dephased in the REDOR experiment. This peak is not in spectrum B, so does not arise from an ε- nitrogen label, but as shown in the supporting information of chapter V (Figure 28.) correlates with the introduction of ^{15}N-labeled PLP; this peak is assigned to the PLP pyridine nitrogen (N_1). As expected, strong dipolar coupling of N_1 to its directly bonded neighbor, C_2, leads to efficient REDOR dephasing. This chemical shift of N_1 reports that the pyridine nitrogen is deprotonated.[31] Additional chemical shift measurements (Figures 29. and 30.) for carbons C_2 and C_3, 158.4 ppm and 168.6 ppm (δ[TMS(l)] scale), respectively, and the phosphorus of the PLP phosphoryl, 4.3 ppm (δ[H_3PO_4(85%)] scale), help complete the assignment of ionization states for the coenzyme. For the former, ^{13}C NMR spectroscopy of model Schiff base compounds[37] under conditions in which the ketoenamine form dominates[38] assists in identifying the chemical shifts of C_2 and C_3 as those for PLP with a deprotonated phenolic oxygen. For the latter, the ^{31}P chemical shift of the phosphoryl reports a dianionic group.[39] The experimentally determined protonation states are shown in Figure 20. .

The ^{15}N solid-state NMR chemical shift of the imine nitrogen supports the PSB hypothesis for the internal aldimine state of tryptophan synthase. At the same time, ^{13}C, ^{15}N and ^{31}P chemical shifts on PLP establish that the phosphoryl group, phenolic oxygen, and pyridine ring nitrogen are deprotonated. The pyridine nitrogen in tryptophan synthase interacts with the hydroxyl of βSer377 and, in the absence of an additional proton, would be incapable of assuming the role of hydrogen bond donor. It has been proposed that the protonation state of the pyridine nitrogen plays an important role in steering later reaction specificity for PLP-

dependent enzymes: a protonated pyridine nitrogen enhances electrophilic addition at $C_{4'}$, while a deprotonated pyridine favors reaction at the substrate C^{α}.[2,6-8,40] The deprotonated pyridine nitrogen in TS, assuming it is maintained during the catalytic cycle, is consistent with its catalytic role on the β- replacement pathway that takes serine to tryptophan. Yet protonation of the PLP nitrogen has been shown to promote proton transfer from the phenolic oxygen to the Schiff base nitrogen,[31,41] a defining aspect of the PSB hypothesis. Effecting this transfer for the unprotonated PLP clearly requires a different mode of activation.

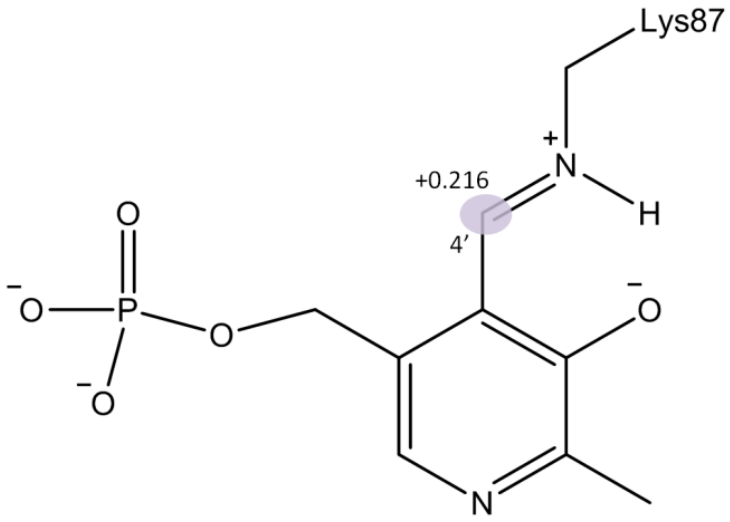
AB INITIO DISCUSSION

Of the candidate structures, PSB form (model code: 100) was found to have the closest agreement (Figure 24.) between the calculated and experimental ^{13}C and ^{15}N chemical shifts with a reduced χ^2 of 5.69; 95% confidence interval 0.101 - 2.295 [24]. A large discrepancy for most of the other structures occurred at the pyridine and Schiff base nitrogen which have experimental shifts of 202.3 and 294.7 ppm respectively. Previous measurements in inhibited PLP-dependent alanine racemase give a protonated Schiff base nitrogen shift of 189 ppm,[12] while model compounds place a neutral Schiff base linkage with methylamine at 313 ppm [20] (both adjusted to δ[NH3(l)]). We note that neither N_1 of PLP nor N^{ε} of βlys87 is at the extreme chemical shift value anticipated for a fully deprotonated or fully protonated nitrogen;[13] this may indicate hydrogen bonding interactions or equilibria between tautomeric forms that, nonetheless, strongly favor the PSB form shown. It is for these reasons that the reduced χ^2 falls just 2.56 units outside of the acceptable 95% confidence interval for 4 data points. Including a

higher enrichment of isotopic labels on the substrate increases the number of data points fitted and ranked by reduced χ^2 analysis thereby cataloging structure(s) under better refinement. The hypothesis of the PSB as a predominate structure is *also* supported in the X-ray analysis where the crystallographic hydrogen bond from the Schiff base nitrogen to the phenolic oxygen distance is 2.57 Å where the computational model is 2.58 Å and is fully consistent with the presence of a proton in an N-H-O hydrogen bond.

The atomic charges calculated from the ab initio electronic structures using natural bond orbital analysis[21] are consistent with fundamental implications for the mechanism in tryptophan synthase, which at the next step involves a transimination reaction. This step requires an amino acid substrate to make a nucleophilic attack at C_4' of PLP exchanging the PLP Schiff base linkage to the protein with one to the substrate[3] and while the NBO calculation predicted C_4' to be +0.216 au in the PSB form it further provides direct evidence and support for the transimination sequence.

Figure 26. NBO charges of PSB



CONCLUSION

The measurement of chemical shifts in the enzyme active site of tryptophan synthase provides an atomic-resolution snapshot of protonation states critical for initiating catalysis in this PLP-dependent enzyme. The ^{15}N chemical shift for the linking lysine ε-imine nitrogen confirms a protonated Schiff base as the majorly contributing tautomer for the internal aldimine state, while ^{13}C, ^{15}N and ^{31}P chemical shifts on PLP report that the phosphoryl group, phenolic oxygen, and pyridine ring nitrogen are deprotonated. These results are consistent both with the PSB hypothesis for PLP activation and tryptophan synthase's enzymatic role as a β elimination/replacement catalyst. Furthermore, the empirical correlation between chemical shift and chemical environment has been used to report on substrate-PLP protonation states in model compounds[12,20,22] and enzyme active sites.[11,12] Here the combination of solid-state NMR spectroscopy with diffraction and computational methods to build chemically detailed structural models with quantitative predictions of chemical shifts has been shown to hold remarkable promise for the structural characterization of enzyme intermediates. Predicated on the accuracy with which chemical shifts can now be calculated,[17] this approach, which is an extension of NMR crystallography[23] to a functioning enzyme system, points to a new hypothesis for the Internal Aldimine intermediate's protonation state in tryptophan synthase and its role in directing the next step in catalysis.

REFERENCES

(1) Metzler, D. E.; Ikawa, M.; Snell, E. E. J Am Chem Soc 1954, 76, 648.

(2) Toney, M. D. Bba-Proteins Proteom 2011, 1814, 1407.

(3) Cordes, E. H.; Jencks, W. P. Biochemistry-Us 1962, 1, 773.

(4) Heinert, D.; Martell, A. E. J Am Chem Soc 1963, 85, 188.

(5) Metzler, D. E. J Am Chem Soc 1957, 79, 485.

(6) Crugeiras, J.; Rios, A.; Riveiros, E.; Richard, J. P. J Am Chem Soc 2011, 133, 3173.

(7) Major, D. T.; Gao, J. L. J Am Chem Soc 2006, 128, 16345.

(8) Major, D. T.; Nam, K.; Gao, J. L. J Am Chem Soc 2006, 128, 8114.

(9) Copie, V.; Faraci, W. S.; Walsh, C. T.; Griffin, R. G. Biochemistry-Us 1988, 27, 4966.

(10) McDowell, L. M.; Lee, M. S.; Schaefer, J.; Anderson, K. S. J Am Chem Soc 1995, 117, 12352.

(11) Sharif, S.; Fogle, E.; Toney, M. D.; Denisov, G. S.; Shenderovich, I. G.; Buntkowsky, G.; Tolstoy, P. M.; Huot, M. C.; Limbach, H. H. J Am Chem Soc 2007, 129, 9558.

(12) Lai, J. F.; Niks, D.; Wang, Y. C.; Domratcheva, T.; Barends, T. R. M.; Schwarz, F.; Olsen, R. A.; Elliott, D. W.; Fatmi, M. Q.; Chang, C. E. A.; Schlichting, I.; Dunn, M. F.; Mueller, L. J. J Am Chem Soc 2011, 133, 4.

(13) Chan-Huot, M.; Dos, A.; Zander, R.; Sharif, S.; Tolstoy, P. M.; Compton, S.; Fogle, E.; Toney, M. D.; Shenderovich, I.; Denisov, G. S.; Limbach, H. H. J Am Chem Soc 2013, 135, 18160.

(14) Dunn, M. F.; Niks, D.; Ngo, H.; Barends, T. R. M.; Schlichting, I. Trends in Biochemical Sciences 2008, 33, 254.

(15) Yanofsky, C.; Crawford, I. P. In The Enzymes; Boyer, P. D., Ed.; Academic Press: New York, 1972, p 1.

(16) Miles, E. W. Advances in Enzymology and Related Areas of Molecular Biology 1979, 49, 127.

(17) Lane, A. N.; Kirschner, K. European Journal of Biochemistry 1983, 129, 571.

(18) Christensen, H. N. J Am Chem Soc 1958, 80, 99.

(19) Peracchi, A.; Bettati, S.; Mozzarelli, A.; Rossi, G. L.; Miles, E. W.; Dunn, M. F. Biochemistry-Us 1996, 35, 1872.

(20) Goldberg, M. E.; Baldwin, R. L. Biochemistry-Us 1967, 6, 2113.

(21) Hur, O.; Niks, D.; Casino, P.; Dunn, M. F. Biochemistry-Us 2002, 41, 9991.

(22) Bazhulina, N. P.; Morozov, Y. V.; Papisova, M. I.; Demidkina, T. V. European Journal of Biochemistry 2000, 267, 1830.

(23) Christen, P.; Metzler, D. E. Transaminases; Wiley: New York, 1985.

(24) Metzler, C. M.; Cahill, A.; Metzler, D. E. J Am Chem Soc 1980, 102, 6075.

(25) Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R. Journal of Biological Chemistry 1988, 263, 17857.

(26) Sachpatzidis, A.; Dealwis, C.; Lubetsky, J. B.; Liang, P. H.; Anderson, K. S.; Lolis, E. Biochemistry-Us 1999, 38, 12665.

(27) Kulik, V.; Weyand, M.; Seidel, R.; Niks, D.; Arac, D.; Dunn, M. F.; Schlichting, I. J Mol Biol 2002, 324, 677.

(28) Weyand, M.; Schlichting, I.; Marabotti, A.; Mozzarelli, A. Journal of Biological Chemistry 2002, 277, 10647.

(29) Ngo, H.; Harris, R.; Kimmich, N.; Casino, P.; Niks, D.; Blumenstein, L.; Barends, T. R.; Kulik, V.; Weyand, M.; Schlichting, I.; Dunn, M. F. Biochemistry-Us 2007, 46, 7713.

(30) Niks, D.; Hilario, E.; Dierkers, A.; Ngo, H.; Borchardt, D.; Neubauer, T. J.; Fan, L.; Mueller, L. J.; Dunn, M. F. Biochemistry-Us 2013, 52, 6396.

(31) Limbach, H. H.; Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Shenderovich, I. G.; Denisov, G. S.; Toney, M. D. Bba-Proteins Proteom 2011, 1814, 1426.

(32) Chan-Huot, M.; Sharif, S.; Tolstoy, P. M.; Toney, M. D.; Limbach, H. H. Biochemistry-Us 2010, 49, 10818.

(33) Sharif, S.; Denisov, G. S.; Toney, M. D.; Limbach, H. H. J Am Chem Soc 2007, 129, 6313.

(34) Sharif, S.; Schagen, D.; Toney, M. D.; Limbach, H. H. J Am Chem Soc 2007, 129, 4440.

(35) Harbison, G. S.; Herzfeld, J.; Griffin, R. G. Biochemistry-Us 1983, 22, 1.

(36) Gullion, T.; Schaefer, J. J Magn Reson 1989, 81, 196.

(37) O'Leary, M. H.; Payne, J. R. The Journal of biological chemistry 1976, 251, 2248.

(38) Sharif, S.; Huot, M. C.; Tolstoy, P. M.; Toney, M. D.; Jonsson, K. H. M.; Limbach, H. H. J Phys Chem B 2007, 111, 3869.

(39) Schnackerz, K. D.; Andi, B.; Cook, P. F. Biochimica Et Biophysica Acta-Proteins and Proteomics 2011, 1814, 1447.

(40) Casasnovas, R.; Adrover, M.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Munoz, F. J Phys Chem B 2012, 116, 10665.

(41) Sharif, S.; Powell, D. R.; Schagen, D.; Steiner, T.; Toney, M. D.; Fogle, E.; Limbach, H. H. Acta Crystallographica Section B-Structural Science 2006, 62, 480.

(42) Experiments in Physical Chemistry 8th Edition by: Carl Garland, Joseph Nibler, and David Shoemaker; ISBN 978-0072828429

CHAPTER V. SUPPLEMENTAL INFORMATION FOR CHAPTER FOUR: PROTONATION STATE OF THE TRYPTOPHAN SYNTHASE INTERNAL ALDIMINE ACTIVE SITE FROM SOLID-STATE NMR SPECTROSCOPY AND AB INITIO COMPUTATIONAL CHEMISTRY.

ABSTRACT

The synergistic combination of NMR spectroscopy, X-ray crystallography and ab initio computational chemistry, also referred to as "NMR Crystallography", can be used as a unmatched tool for elucidating high resolution three-dimensional structures of diverse materials. This synergistic approach was used to determine the structures of the indoline and 2-aminophenol (2AP) quinonoid intermediates in the pyridoxal-5'-phosphate-dependent enzyme tryptophan synthase under conditions of active catalysis. Experimentally measured chemical shifts at particular ^{13}C - and ^{15}N-labeled positions on the substrate were used to exclusively distinguish and emerge a comparable species from the various computational model types. The reduced chi squared analysis establishes the PSB form and supports the importance of the substrate's role in directing the next step of catalysis.

INTRODUCTION

The purpose of this chapter is to provide supplemental information regarding the techniques discussed in Chapter IV.

MATERIALS AND METHODS

X-RAY

The 1.30 Å resolution x-ray crystal structure of (F9) Cs+ *S. typhimurium* TS E(Ain) (PDB accession code 4HT3) complex was determined in an analogous manner as outlined in chapters II and III.

NMR

Solid-state ^{13}C and ^{15}N chemical shifts were indirectly referenced to TMS and liquid-NH3 following the recommendations outlined in Morcombe and Zilm [10] and Markley et al [11]. In brief, the downfield ^{13}C CPMAS peak of solid adamantane was set to 38.48 ppm; this defines a scale on which the corresponding methyl resonance of neat TMS is 0 ppm. The ^{15}N resonance of liquid ammonia was then calculated as the ^{13}C 0 ppm reference frequency multiplied by 10.1329118/25.1450038. On this scale, solid $^{15}NH_4Cl$ was found to resonate at an offset of 39.2 ppm. Solid-state ^{31}P was indirectly referenced to 85% H3PO4 (capillary) following the procedure outlined in Maurer and Kalbitzer [12]. This scale was chosen for direct comparison with the PLP titration results of Schnackerz et al [13-15]. An MAS rotor was first loaded with a 1% solution of DSS in D_2O to determine 0 ppm ^{1}H (DSS). This frequency was multiplied by 0.404806169 to define the ^{31}P 0 ppm point, corresponding to 85% H3PO4 (capillary). For comparison, δ [85% H3PO4 (capillary)]=δ[85% H3PO4 (sphere)] + 0.36 ppm.

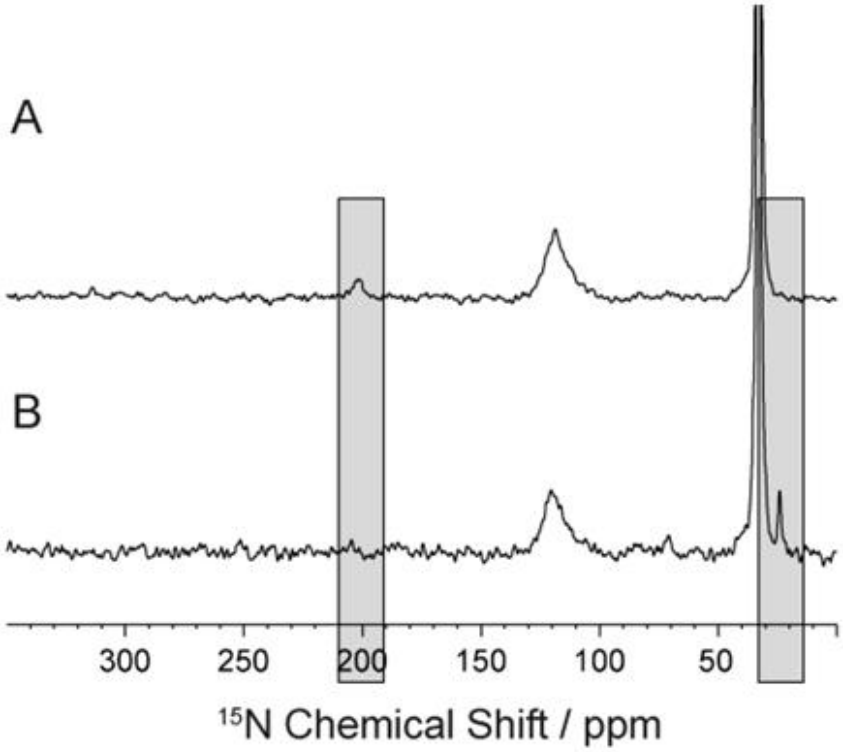


Figure 27. ^{15}N SSNMR of the reaction of tryptophan synthase microcrystals and (A) no serine and (B) 75 mM L-Ser. The conditions in (B) give rise to the aminoacrylate intermediate[3] in which the Schiff base resonance at 202.3 ppm is lost and a new peak at 24.2 ppm appears, suggestive of a neutral amino lysine sidechain for the aminoacrylate intermediate. Experiments were performed at 9.4 T (400.37 MHz ^{1}H, 100.69 MHz ^{13}C, 40.57 MHz ^{15}N) on a Bruker AVIII spectrometer equipped with an ^{1}H-^{13}C-^{15}N triple resonance 4 mm MAS probe, spinning at a MAS rate of 8 kHz, and with the bearing gas cooled to -15°C, giving an effective sample temperature of -5°C. Cross-polarization was accomplished at a ^{1}H spin-lock field of 45 kHz and ^{15}N spin-lock of 37 kHz (ramped +/- 2 kHz); 85 kHz Spinal64 ^{1}H decoupling was used throughout. Each spectrum consists of the sum of 81,920 transients acquired with a relaxation delay of 4s, for a total acquisition time of 3 d 19 h. ^{15}N chemical shifts were referenced indirectly to liq-NH_3 via an external solid-state sample of $^{15}NH_4Cl$ sample, calibrated under MAS conditions as described below.

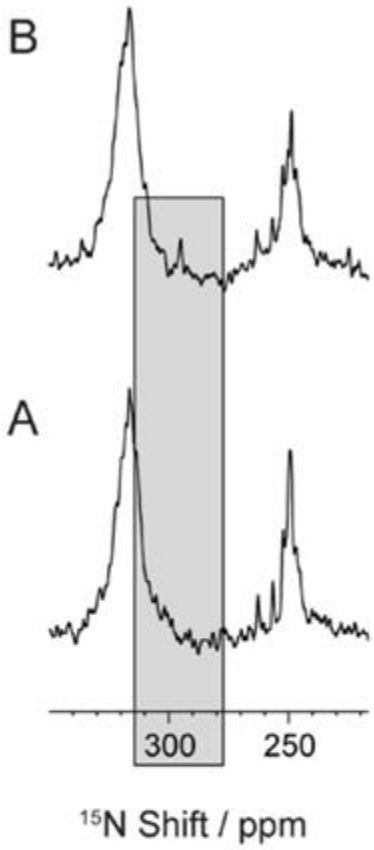


Figure 28. ^{15}N-solid-state NMR cross-polarization magic-angle-spinning (CPMAS) spectra of the tryptophan synthase internal aldimine complex used to assign the ^{15}N chemical shifts for N1 of PLP (294.7 ppm). Data were acquired on microcrystalline samples of *S. typhimurium* TS prepared with (A) U-^{15}N TS/natural abundance PLP and (B) U-^{15}N TS/2,2',3-$^{13}C_3$;^{15}N PLP. Experiments were performed at 9.4 T (400.37 MHz ^{1}H, 100.69 MHz ^{13}C, 40.57 MHz ^{15}N) on a Bruker AVIII spectrometer equipped with an ^{1}H-^{13}C-^{15}N triple resonance 4 mm MAS probe, spinning at a MAS rate of 8 kHz, and with the bearing gas cooled to -15°C, giving an effective sample temperature of -5°C. Cross-polarization was accomplished at a ^{1}H spin-lock field of 45 kHz and ^{15}N spin-lock of 37 kHz (ramped +/- 2 kHz); 85 kHz Spinal64 ^{1}H decoupling was used throughout. Each spectrum consists of the sum of 81,920 transients acquired with a relaxation delay of 4s, for a total acquisition time of 3 d 19 h. ^{15}N chemical shifts were referenced indirectly to liq-NH_3 via an external solid-state sample of $^{15}NH_4Cl$ sample, calibrated under MAS conditions as described below.

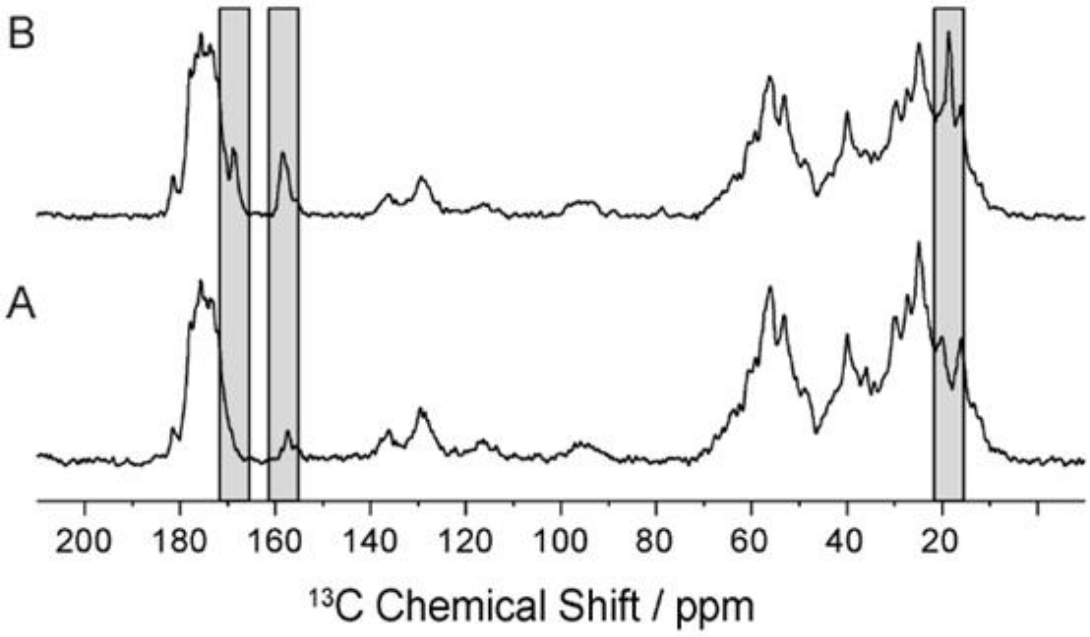


Figure 29. ^{13}C-solid-state NMR cross-polarization magic-angle-spinning (CPMAS) spectra of the tryptophan synthase internal aldimine complex used to assign the ^{13}C chemical shifts for C2 and C3 of PLP. Data were acquired on microcrystalline samples of *S. typhimurium* TS prepared with (A) natural abundance TS/natural abundance PLP and (B) natural abundance TS/2,2',3-$^{13}C_3$;^{15}N PLP. Experiments were performed at 9.4 T (400.37 MHz ^{1}H, 100.69 MHz ^{13}C, 40.57 MHz ^{15}N) on a Bruker AVIII spectrometer equipped with an ^{1}H-^{13}C-^{15}N triple resonance 4 mm MAS probe, spinning at a MAS rate of 8 kHz, and with the bearing gas cooled to -15°C, giving an effective sample temperature of -5°C. Cross-polarization was accomplished at a ^{1}H spin-lock field of 45 kHz and ^{13}C spin-lock of 53 kHz (ramped +/- 2 kHz); 85 kHz Spinal64 ^{1}H decoupling was used throughout. Each spectrum consists of the sum of 16,384 transients acquired with a relaxation delay of 4s, for a total acquisition time of 18.3 h. ^{13}C chemical shifts were referenced indirectly to neat TMS via an external solid-state sample of adamantane, calibrated under MAS conditions as described below.

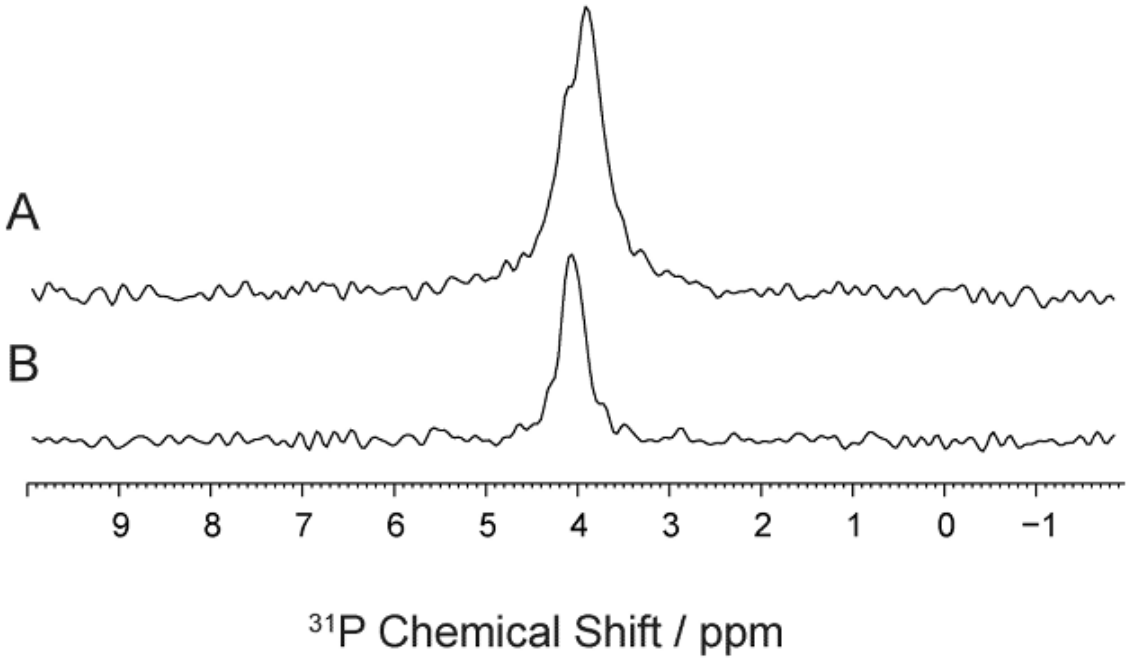


Figure 30. ^{31}P SSNMR spectra of TS internal aldimine prepared (A) with and (B) without the α-site binding ligand N-(4'-trifluoromethoxybenzenesulfonyl)-2-aminoethyl phosphate (F9). Experiments were performed at 14.1 T (600.01 MHz ^{1}H, 242.89 MHz ^{31}P) on a Bruker AVANCE spectrometer equipped with an ^{1}H-X double resonance 4 mm MAS probe, spinning at a MAS rate of 10 kHz, and with the bearing gas cooled to -20 °C, giving an effective sample temperature of -15 °C. Cross-polarization was accomplished at a ^{1}H spin-lock field of 45 kHz and ^{31}P spin-lock of 55 kHz (ramped +/- 5 kHz); 85 kHz Spinal64 ^{1}H decoupling[5] was used throughout. Each spectrum consists of the sum of 8,192 transients acquired with a relaxation delay of 4 s, for a total acquisition time of 9.1 h. ^{31}P chemical shifts were referenced indirectly to 85% H_3PO_4 (capillary) as described below.

Preparation of ε-^{15}N-Lys TS and U-^{15}N TS:

Tryptophan synthase was expressed and purified as previously described[1,2] with the following modifications. Bacteria were grown in LB containing tryptone (10 g/L), NaCl (10 g/L), yeast extract (5 g/L), and ampicillin (50 mg/L) until the late log phase (6 hours, 37 °C) when cells were harvested and transferred to minimal media. Expression was induced through addition of isopropyl β-D-1- thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM in minimal media enriched with 2 g L-1 15 NH4Cl (U-^{15}N labeled sample only, Cambridge Isotope Laboratories) for 12 hours at 25 °C. Minimal media for unlabeled and ε -^{15}N-Lys TS samples were supplemented with 40 mL unlabeled 10x BioExpress Cell Growth Media (CIL), while U-^{15}N TS samples were supplemented with 40 mL 10x U-^{15}N BioExpress Cell Growth Media (CIL). ε -^{15}N-Lys TS was prepared by addition of 250 mg L- ε -^{15}N-Lys hydrochloride (CIL) to unlabeled minimal media. Microcrystals were prepared by diluting enzyme solution 1:1 with 50 mM Cs-bicine buffer, pH = 7.8 containing 14% PEG-8000 and 3.0 mM spermine as described previously.[1] Microcrystals were collected and washed with 50 mM Cs-bicine, pH 7.8 containing 8% PEG-8000, 1.5 mM spermine, and 3 mM N-(4'-trifluoromethoxybenzenesulfonyl)-2-aminoethyl phosphate (F9; a high affinity alpha site ligand). Magic-angle-spinning rotors were packed at 10,000 rpm, and each rotor contained approximately 25-30 mg of protein.

Preparation of U-^{15}N TS/2,2',3-^{13}C3;^{15}N-PLP:

Tryptophan synthase $\alpha_2\beta_2$ subunits were dissociated and reconstituted with labeled PLP as previously described4 with attention called to the following details. Addition of potassium thiocyanate (KSCN) to a final concentration of 1 M to the enzyme solution was followed by a five minute wait at room temperature until reaction initiation. Hydroxylamine was then added to a

final concentration of 10 mM, followed by five more minutes at room temperature, completing oxime formation. Dissociation was completed with dialysis against Cs-bicine, pH 7.8, containing 1 M KSCN for four hours at 4 °C. This was followed by two more dialyses against plain 50 mM Cs-bicine, pH 7.8, at 4 °C. UV-vis tests on the apoenzyme showed no activity in the presence of substrates. The holoenzyme was resolved by addition of 2,2',3-^{13}C3;^{15}N-PLP to the dialyzed protein so the final PLP concentration was three times greater than the protein concentration. This solution was incubated in a water bath at 35 °C for 10 minutes. Reassociation of the subunits was completed with addition of L-serine to a final concentration of 15 mM and allowed to sit for 10 more minutes in the warm water bath. The enzyme solution was cooled in an ice bath for 30 minutes, then dialyzed twice against 50 mM Cs-bicine, pH 7.8, at 4 °C. UV/vis activity tests performed on the reconstituted holoenzyme showed fully active enzyme with 70 % recovery of the initial concentration, and incorporation of the 2,2',3-^{13}C3;^{15}N-PLP into the enzyme was verified with both ^{13}C and ^{15}N SSNMR.

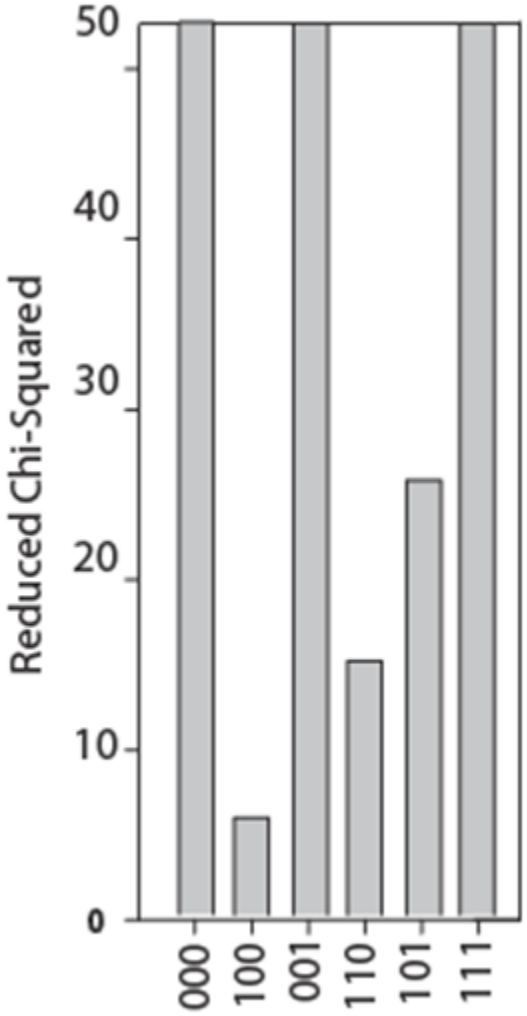
AB INITIO

Computational calculations were carried out on the internal aldimine intermediate utilizing the protocol outlined in chapters II and III. Two exceptions to the modeling exist. First, using an ONIOM approach for the NMR chemical shielding calculations was employed to allow a more complete clustering of the ligand while applying a moderate basis set and level of theory (DFT B3LYP/6-31G) around the phosphoryl group and outer regions of the substrate, yielding ~240 atoms total with ~140 in the High Layer; job keyword # ONIOM B3LYP/6-311++G(d,p) : B3LYP/6-31G; DFT in gaussian09. Secondly, because this intermediate does not contain a carboxylate group and accepted evidence[39] has supported a purely dianionic phosphoryl

exceptions to this protocol, regarding all available sites of protonation have been made. This is illustrated in Figure 22. which shows all possibilities for sites of protonation used to generate the candidate structures for the PLP-ligand complex including: the Schiff base linkage, phenolic oxygen, and pyridine nitrogen of PLP eliciting a three digit code I.E. 100 in an analogous fashion revealed in the previous chapters. Interestingly, two specific models (namely 010 and 011), which include a protonated phenolic oxygen, consistently donated it's proton to the local Schiff base nitrogen in all cases where the nitrogen was initially deprotonated during the optimization (Figure 21.), effectively lowering the overall energy by forming a protonated Schiff base species required to converge the system. This phenomena also occurred in cases where the the phenolic proton was deliberately rotated 90°, 180°, and 270° from the plane and *further* while the strict convergence criteria threshold was lifted, using the widely accepted command for difficult convergences cases; SCF=QC (Quadratically Convergent), in Gaussian09.

Table 9. Chemical shifts of all 6 models of Ain

Model	NSchiff Base	N1	C2	C2
100	{164.031,	320.69,	157.801,	166.661}
001	{341.061,	208.585,	150.162,	156.89}
101	{180.69,	206.686,	146.66,	157.335}
110	{193.727,	346.346,	153.496,	145.666}
111	{282.156,	201.74,	160.239,	103.108}
000	{339.189,	282.527,	148.737,	161.876}



Model	Reduced χ^2
100	5.69513
001	73.2637
101	25.4374
110	15.3346
111	103.029
000	51.7429

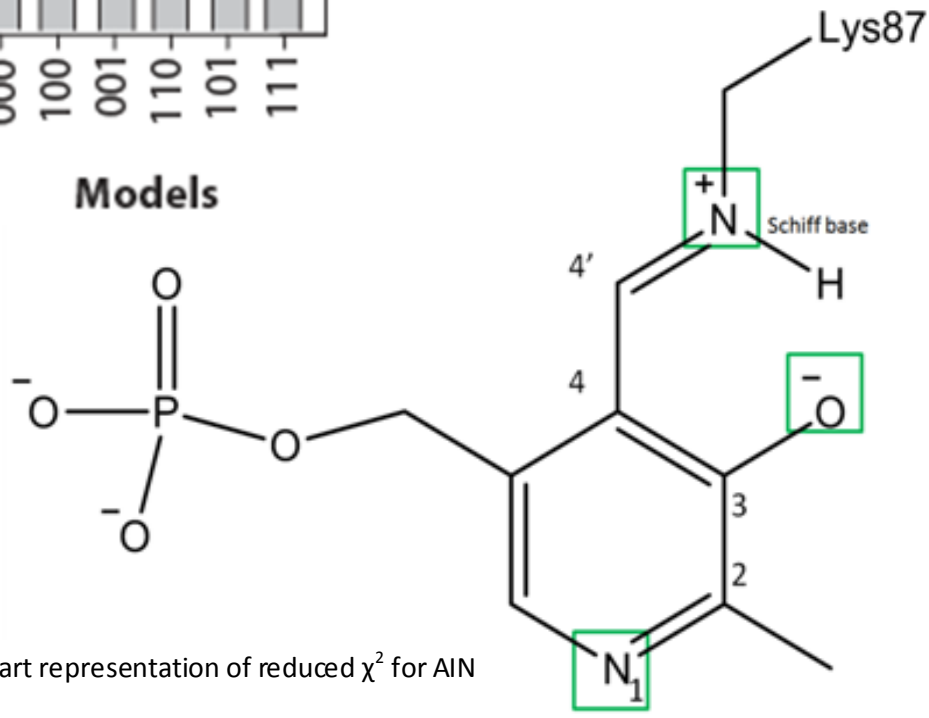


Figure 31. Bar chart representation of reduced χ^2 for AIN

CONCLUSION

The measurement of chemical shifts in the enzyme active site of tryptophan synthase provides an atomic-resolution snapshot of protonation states critical for initiating catalysis in this PLP-dependent enzyme. The ^{15}N chemical shift for the linking lysine ε-imine nitrogen confirms a protonated Schiff base for the internal aldimine state, while ^{13}C, ^{15}N and ^{31}P chemical shifts on PLP report that the phosphoryl group, phenolic oxygen, and pyridine ring nitrogen are deprotonated. These results are consistent both with the PSB hypothesis for PLP activation and tryptophan synthase's enzymatic role as a β elimination/replacement catalyst.

REFERENCES

(1) Tian, Y.; Chen, L.; Niks, D.; Kaiser, J. M.; Lai, J.; Rienstra, C. M.; Dunn, M. F.; Mueller, L. J. Phys Chem Chem Phys 2009, 11, 7078.

(2) Yang, L.; Ahmed, S. A.; Miles, E. W. Protein Expr Purif 1996, 8, 126.

(3) Dunn, M. F.; Niks, D.; Ngo, H.; Barends, T. R. M.; Schlichting, I. Trends in Biochemical Sciences 2008, 33, 254.

(4) Miles, E. W.; Moriguchi, M. J Biol Chem 1977, 252, 6594.

(5) Lee, S. J.; Kim, E.; Seo, M. L.; Do, Y.; Lee, Y. A.; Lee, S. S.; Jung, J. H.; Kogiso, M.; Shimizu, T. Tetrahedron 2008, 64, 1301.

(6) Dean, A.; Ferlin, M. G.; Brun, P.; Castagliuolo, I.; Badocco, D.; Pastore, P.; Venzo, A.; Bombi, G. G.; Di Marco, V. B. Dalton T 2008, 1689.

(7) Sharif, S.; Schagen, D.; Toney, M. D.; Limbach, H. H. J Am Chem Soc 2007, 129, 4440.

(8) Chan-Huot, M.; Niether, C.; Sharif, S.; Tolstoy, P. M.; Toney, M. D.; Limbach, H. H. J Mol Struct 2010, 976, 282.

(9) V. L. Florentiev, V. I. I., M. Ya. Karpeisky In Methods in Enzymology; Donald B. McCormick, L. D.

W., Ed.; Elsevier: 1970; Vol. 18, Part A, p 567

(10) Morcombe, C. R.; Zilm, K. W. J Magn Reson 2003, 162, 479.

(11) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wuthrich, K. Pure and Applied Chemistry 1998, 70, 117.

(12) Maurer, T.; Kalbitzer, H. R. Journal of Magnetic Resonance Series B 1996, 113, 177.

(13) Schnackerz, K. D.; Andi, B.; Cook, P. F. Biochimica Et Biophysica Acta-Proteins and Proteomics 2011, 1814, 1447.

(14) Schnackerz, K. D.; Keller, J.; Phillips, R. S.; Toney, M. D. Biochimica Et Biophysica Acta-Proteins and Proteomics 2006, 1764, 230.

(15) Schnackerz, K. D.; Mozzarelli, A. Journal of Biological Chemistry 1998, 273, 33247.

(16) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.;

Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs, 2005; Vol. 26.

(17) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Journal of Molecular Graphics & Modelling 2006, 25, 247.

(18) Song, Y. F.; Gunner, M. R. J Mol Biol 2009, 387, 840.

(19) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Journal of Computational Chemistry 2005, 26, 1781.

CHAPTER VI. CONVERSION OF CALCULATED ISOTROPIC SHIELDING TO CHEMICAL SHIFT USING LINEAR REGRESSION ANALYSIS.

ABSTRACT

Correlations between experimentally determined ^{15}N chemical shifts and Gauge-Independent Atomic Orbitals (GIAO)-calculated isotropic shielding constants, δ_{EXPT} = [(intercept) $-\sigma_{CALCD}$]/(-slope), are reported that were obtained from libraries[1-5] of nitrogen-containing heterocycles. These groups of compounds are categorized by their experimentally determined physical state (solid or liquid) (pyridine/Schiff-base-like or pyrrolidine/pyrimidine-like) structures, respectively. For the 27 liquid compounds (test set) a conductor-like polarizable continuum model (CPCM)[6,7], an implicit solvent technique, was used during geometry optimizations *and* GIAO-NMR chemical shielding calculations while employing density functional theory (DFT), utilizing the B3LYP hybrid functional with the notoriously accepted basis sets 6-31G(d,p) and 6-311G++(d,p), respectively. Calculations for the 16 solid state crystal compounds (probe set) were carried out analogously, however making use of a cluster-like approach in exchange for a CPCM. Absolute nitrogen shieldings (unreferenced SCF GIAO isotropic magnetic shielding tensor value in ppm) were then tabulated and plotted against established isotropic chemical shift values, referenced to external liquid ammonia. The performance of theoretical NMR calculations and the resulting eligibility for routine practical use were assessed from convergence of R^2 to 1 and slope to -1.0, indicating a decrease in systematic error and ultimately a well-performing method. An improvement of the original linear regression of Dokalik et al. was established by re-modeling the computational technique employed, notably a CPCM. This resulted in two distinct data sets for which respective experimental results yielded R^2s = 0.9967 and 0.9957 and slope= -1.05 and 1.01 for chloroform and DMSO respectively. This re-modeling

resulted in an overall decrease in systematic error and further pushed the original linear regression results from Dokalik et al. into the well-performing method regime and improved the accuracy and precision for conversion of calculated isotropic shielding to chemical shift. Most noticeably this improved linear regression test set revealed that the computational modeling resulting from the chloroform CPCM captures, more effectively, the experimental environment ultimately producing a genuine foundation from which RMSDs of a broader range of diverse model systems can be compared. Further, a cluster-like approach for solid state structures (probe set) showed an indiscriminate distribution of points localized about the improved linear regression of the Chloroform-CPCM test set data. The difference from data points of the independently well-performing probe set, measured to the line of this test set, yielded an overall RMSD of 9.738. Correspondingly this value is considerably lower than the RMSD (11.943) calculated when measured to the line formed by the improved, well-performing DMSO test set. Finally, as a proof of concept, a simplified gas-phase-like probe set (extracted directly from the previously optimized clustered models) measured to the line produced by the CPCM-Chloroform test set resulted in an overall RMSD of 19.668. This provides direct evidence that as the computational modeling of the chemical environment diverges from and loses resemblance to the experimental environment, the RMSDs of the probe set disbursed about the well defined test set increase drastically. It is this phenomena that motivates the entire investigation established in this study, and equally important, the ability to quantify the effects of environmental changes between modeling systems.

INTRODUCTION

Over the last 30 years, ab initio electronic structure theory and computational chemistry have advanced to the point where it is now possible to routinely model individual molecules (either in the gas or condensed phases) and predict molecular properties. This has added to the arsenal of techniques available to help identify unknown compounds. In particular, the ability to generate ab initio chemical shifts with high accuracy *and* precision has made it possible to screen and rank competing structural and conformational models for consistency with experimental shifts from nuclear magnetic resonance spectroscopy. In favorable cases, this allows a single structural model to be delineated. This general approach of screening structural models against NMR parameters has been essential in the development of nuclear magnetic resonance (NMR) crystallography which seeks to define the atomic-resolution, three-dimensional structure of crystalline solids using a synergistic combination of X-ray diffraction (single crystal or powder), first-principles computational chemistry, and solid-state NMR spectroscopy. There are many approaches to NMR crystallography (due perhaps to the diversity of NMR methods and parameters that can be measured), and no single approach has yet reached consensus; in most implementations, however, NMR shifts are used to test the consistency of the final structural models with experimental observables.

For solid-state systems, two different approaches have developed. The first uses a cluster model in which first, second, or higher shells of neighboring molecules are included about a central molecule of interest. The goal is to provide a representative chemical environment for the central molecule, enabling an accurate modeling of its properties using localized, Gaussian bases. Related fragment-based approaches look for computationally-

efficient means to calculate the energies and properties. A second approach uses a plane-wave basis and periodic boundary conditions on the unit cell to capture the physics of the condensed state. The practical choice of methodology often comes down to the size of the unit cell. For unit cells with several hundred atoms, as one might find in organic and inorganic molecular crystals, plane wave methods are appropriate. For extended unit cells with thousands of atoms and where only a smaller region may be of interest – an enzyme active site in a crystalline protein, for example – cluster methods are favored. Indeed, recently, we have exploited cluster-based ab initio calculations as part of the application of NMR crystallography to the enzyme active site in tryptophan synthase.

To make use of ab initio shieldings to screen competing models in a quantitative manner, it is necessary to have an a priori estimate of the expected agreement between predicted and experimental shifts. This is typically quantified as root-mean-square-deviations (RMSD) of ab initio absolute shieldings from experimental chemical shifts, measured relative to a reference compound such as TMS or DSS. Ideally such correlations would have a slope of -1 and an intercept that corresponds to the absolute shielding of the reference compound. In practice, slopes typically deviate from -1, indicating a systematic error due to approximations in the level of theory employed and/or basis set limitations. Nonetheless, as long as the correlation is known, computational shieldings can be accurately converted into predicted chemical shifts. Considerable attention has been given to the first-principles calculation of ^{1}H and ^{13}C chemical shifts and correlations have been established for various combinations of theory and basis set to experimental values.

As the experimental NMR chemical shifts are solvent dependent, the influence of intermolecular interactions between solute and solvent cannot be ignored. This raises a second important issue (distinct from the systematic errors discussed above): how well does a given model of the molecular system capture the essential physics. Isolated-molecule (gas phase) calculations may not precisely or accurately predict chemical shifts in condensed phases. For example, Dokalik et. al. found distinct correlations when comparing gas-phase, ab initio ^{15}N chemical shieldings for a test set of molecules to experimental shifts for molecules in solutions of DMSO and chloroform. This is undesirable, as it restricts the application of shielding/shift conversions to molecules dissolved in solvents that have been well-characterized and limits the generality with which such conversions can be extended to other condensed phase systems, such as cluster models of crystalline solids. Implicit and explicit solvent models can help rectify this situation, potentially allowing a single correlation curve between ab initio shieldings and experimental shifts in different solvents.

When considering computational methods one must find a reasonable tradeoff between computational cost and acceptable accuracy. Because DFT with B3LYP alongside a moderate basis set is unable to accurately predict absolute nitrogen chemical shieldings (typical MAD +-15ppm), we employed the well developed and widely accepted approach of Forsyth and Sebag[8] for the conversion of chemical shieldings using linear regression analyasis from a test set of molecules. This method paves the way for a more robust conversion of shielding to shift for libraries of nitrogen containing compounds and species alike in condensed phases. Specifically, we address and further develop the computational methodology of the linear correlations calculated by Dokalik et al. to illustrate that a crystal-clustered approach for solid state

structures are applicable to solution state data when employing an implicit solvent model environment.

MATERIALS AND METHODS

SOLUTION-STATE STRUCTURES

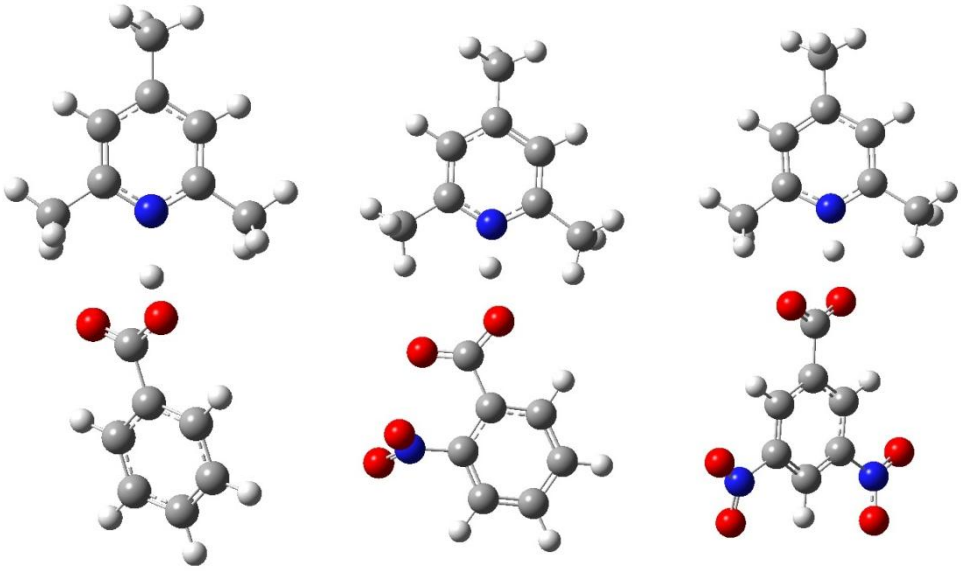
All quantum chemical calculations were performed with the Gaussian09 Revision D.01 program package[11]. Initial solution-state structures were built using gaussview 5.0.9 and geometry optimized using density functional theory utilizing the B3LYP hybrid density functional with the 6-31G(d,p) basis set incorporating an implicit solvent model, a Capacitor-like Polarizable Continum Model (CPCM) with radii=uaks. Geometrically converged structures were then subjected to GIAO NMR chemical shift calculations in an implicit CPCM solvent (DMSO or chloroform) under the same method and level of theory using the 6-311G++(d,p) basis set. Absolute nitrogen shieldings (unreferenced Gaussian SCF GIAO isotropic magnetic shielding tensor values in ppm) were then tabulated and plotted against experimental isotropic chemical shift values; referenced to neat external liquid ammonia. Of the 29 total nitrogen containing compounds taken from dokalik et al, 2 molecules were not used from this data set, namely pyrazole and imidazole for 2 specific reasons: 1. The complete experiment was not carried out by dokalik et al; 2. The inherent and statistical ambiguity of assigning and fitting nitrogen data to a coalescing species. We believe that subjectively averaging the calculated nitrogen values effectively reduces, unproportionately, the number of data points in each data set to represent equivalent nitrogens compromises an almost vigorous approach because it requires prior knowledge of the tautomeric distribution and further is statistically problematic when comparing the effects of two similar models containing a different number of total data points.

SOLID-STATE STRUCTURES

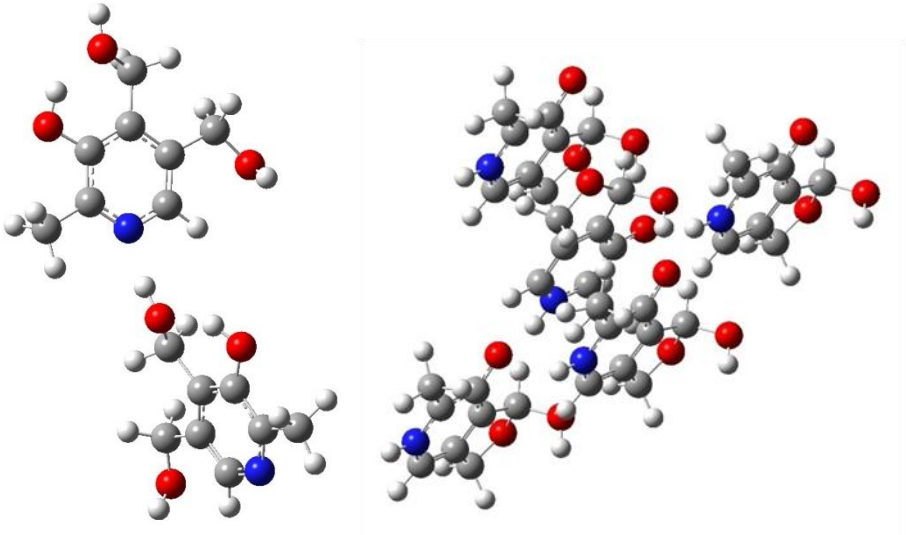
The experimental results from the compounds used in this study were selected from various authors, Liquids data: Dokalik et al. and Solids: Limbach, McDermott et al. and are referenced in Tables 10. and 11. respectfully. All structures were assembled in Gauss View 5.0.9 while solid state structural coordinates were originally derived from the CCSD[12] software package: conquest and mercury viewer and downloaded from the CSD and used at the X-ray resolution provided. Initial solid-state crystal structures were then taken from the Crystal Structure Database (CSD), regenerated in Gauss View 5.0.9 and cut out from a large three dimensional repetition of the unit cell. This insured a proper addressing of the electronic chemical environment including all major hydrogen bond and van der Waals' contributions to the entire molecule of choice, and more specifically, the experimentally measured nitrogen(s) of interest. The refined cluster models were chosen to be of sufficient size to provide a representative chemical and electronic environment for the central molecule as shown in Figure 32. . Heavy (non-hydrogen) atoms were fixed at their crystallographically and well-determined coordinates, thereby maintaining the overall motif of the crystal packing environment, while hydrogen atoms were geometry-optimized within this framework. Hydrogen geometry optimizations and GIAO NMR chemical shifts were performed using the same method and level of theory for liquids mentioned previously, although no CPCM was employed.

Figure 32. Probe set of all 16 solid state crystal cluster-like structures

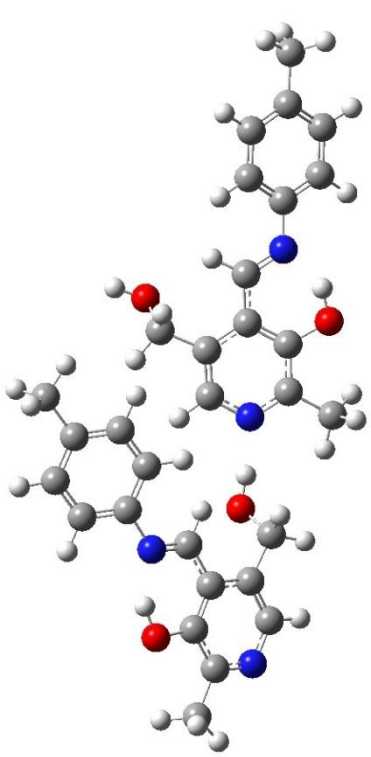
Limbach compounds:



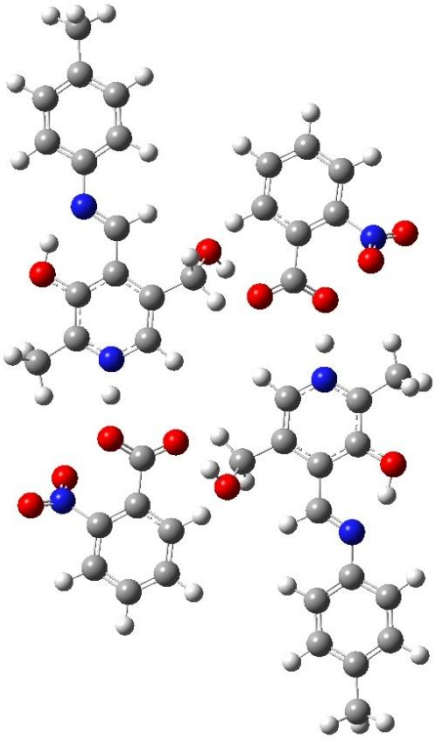
GODNAO GODNES GODNIW



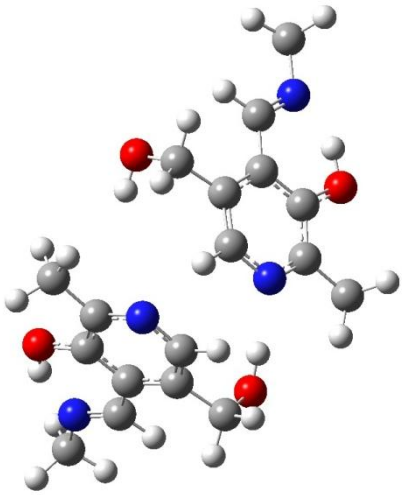
BITZAF BIHKEI01



GEHHAD

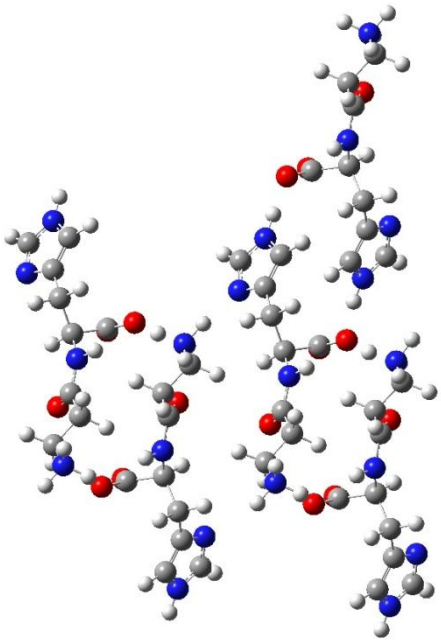


GEHHEH

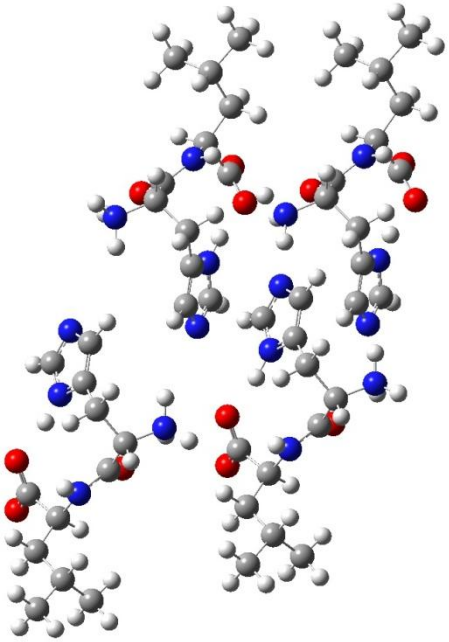


GEHHIL

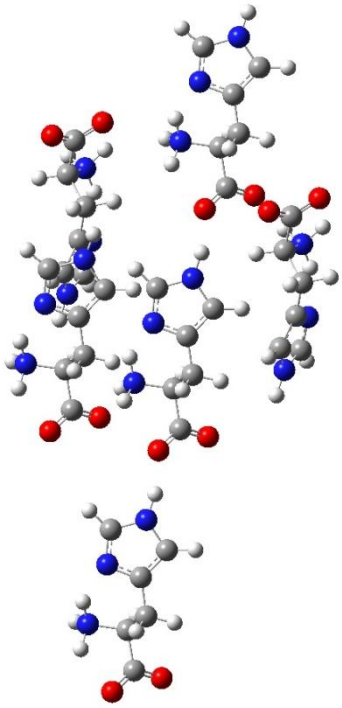
McDermott Model Compounds:



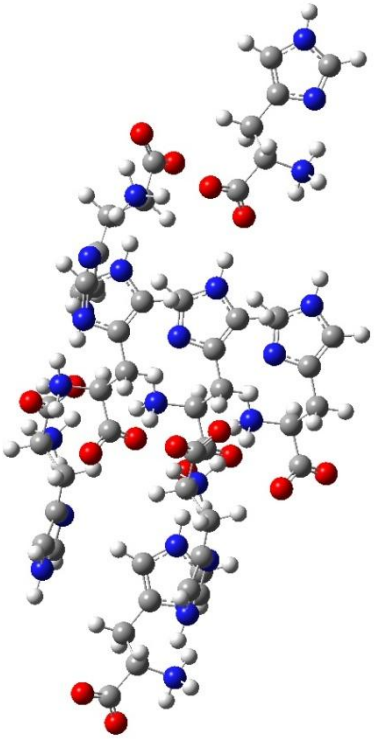
BALHIS01



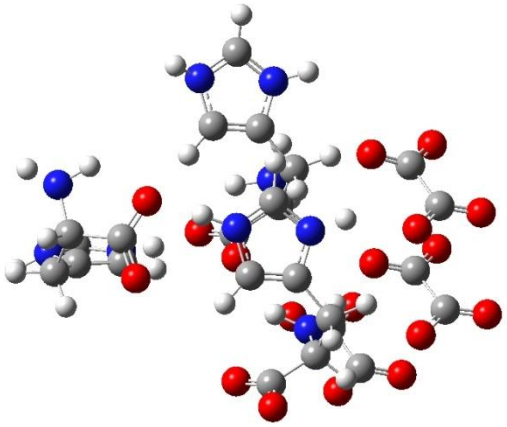
JUKMOR



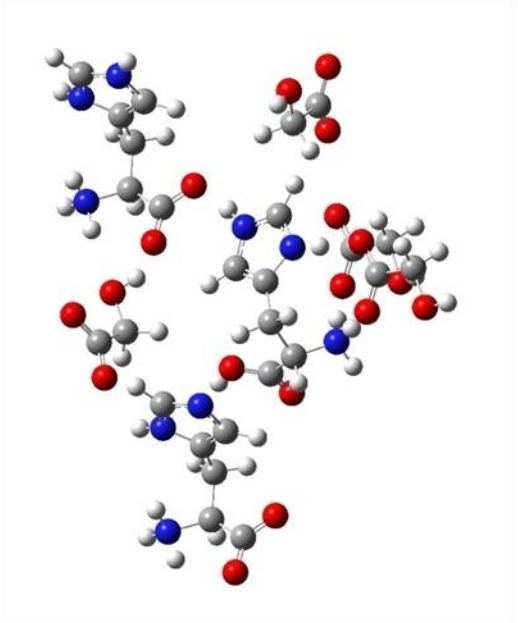
LHISTD02



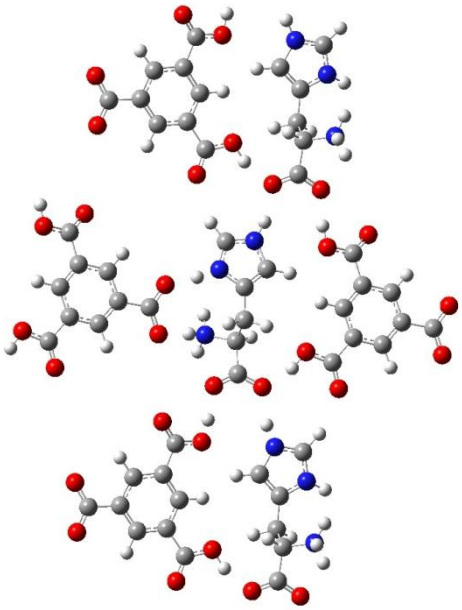
LHISTD13



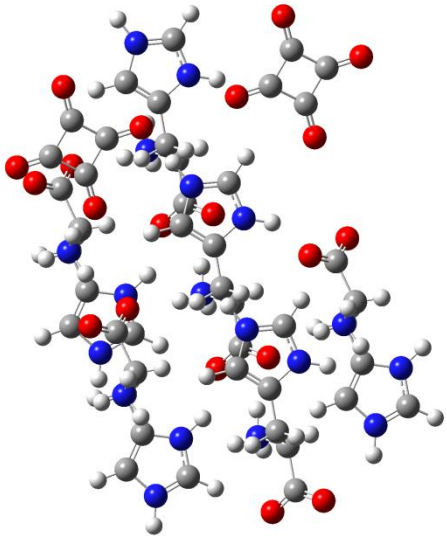
RARXOX



TEJWAG



LHISTM



TIWXEC

RESULTS AND DISCUSSION

From Tables 10. and 11. the following three sections summarize the conclusions about the predictive accuracy of the linear regression produced by the GIAO calculations:

1. Improving the computational modeling of Dokalik et al.

Dokalik's modeling included one set of calculations that resulted in a single set of shielding values to which shifts from two separate experimental data sets were fit. In other words a computational gas-phase like environment, fit to two separate experimental results using a polar (DMSO) and non-polar (chloroform) solvent, yielding two lines seen in Figure 33. a titled, "Gas-Phase Modeling of Liquids Data". For this reason a re-modeling of each experimental condition utilizing a CPCM was employed generating two distinct computational data sets for which their two respective experimental data sets would be compared as see in Figure 33. b titled, "CPCM Modeling of Liquids Data". The linear regression reveals (R^2s = 0.9967 and 0.9957 and slope= -1.05 and 1.01 for chloroform and DMSO respectively) that this re-modeling is a well performing method and improves the accuracy and precision for conversion of calculated isotropic shielding to chemical shift. In order to justify whether a model is considered "well performing"[9] linear fits produced by a particular test set must achieve an R^2 no smaller than 0.995 and a slope that does not deviate from -1 by more than 0.05. The 27 nitrogen containing molecules experimentally measured and calculated by Dokalik et al yield slopes of -1.10 and -1.11 for chloroform and DMSO respectively, indicating that the gas-phase-like approach, although a major improvement to the 30 structures originally fit by Dokalik et. al., is still not a well performing method. It should be noted that although the RMSDs between Dokaliks data and the CPCMs are essentially the same, falling within 0.387 and 0.206 of each other for

chloroform and DMSO respectively, this is not indicative of a poorly performing method; in fact a well performing method provides further certainty of the RMSDs produced. Most noticeably, the linear regression data reveals that the computational modeling utilizing a CPCM for chloroform specifically, does a better job of effectively capturing the experimental environment providing a genuine foundation to compare RMSDs of a broader range of diverse model systems.

Figure 33a.

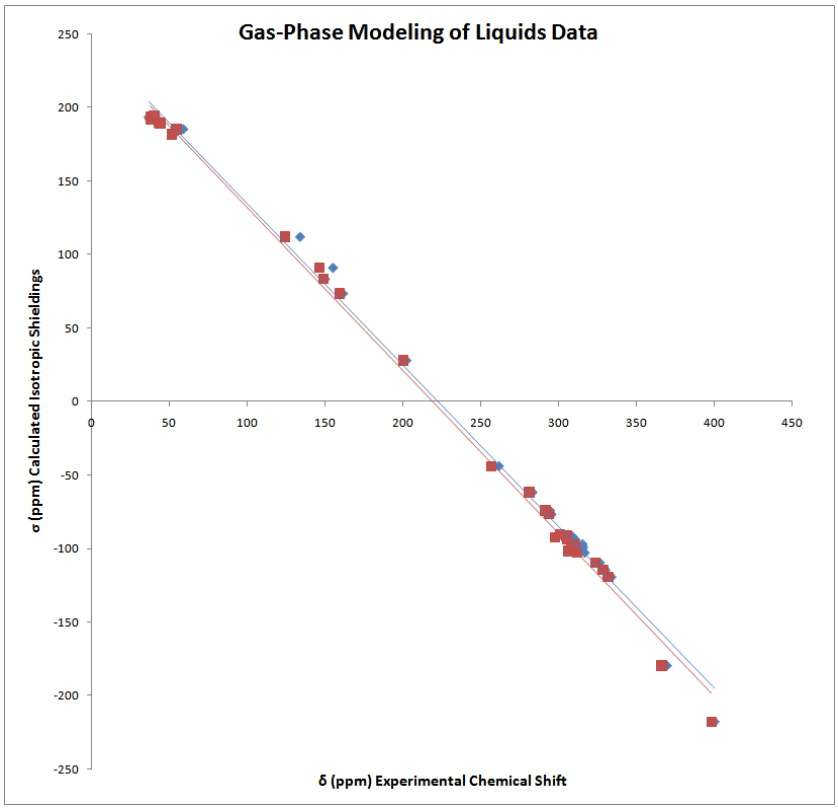
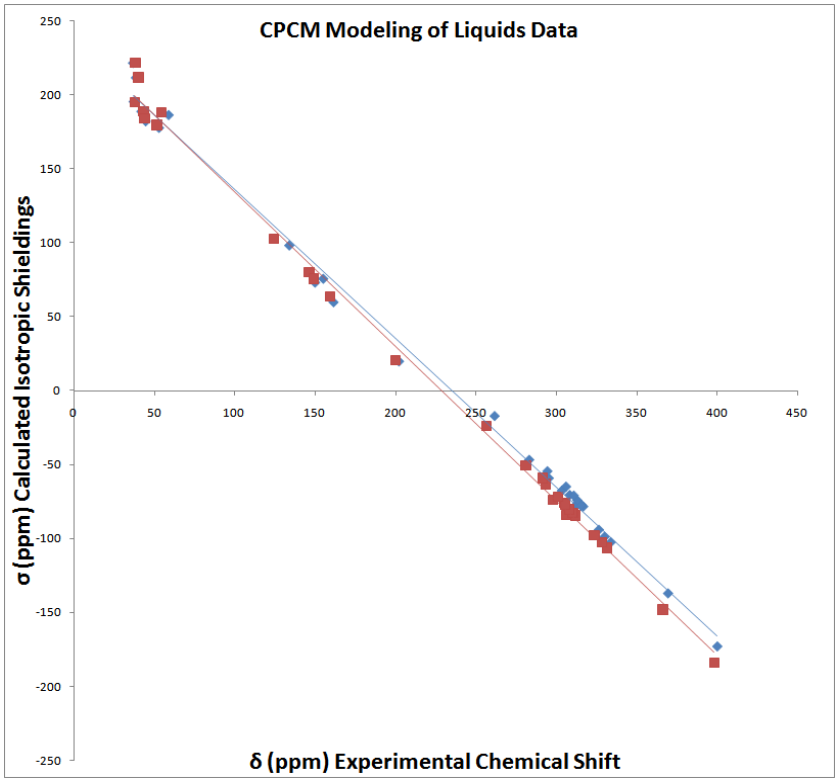


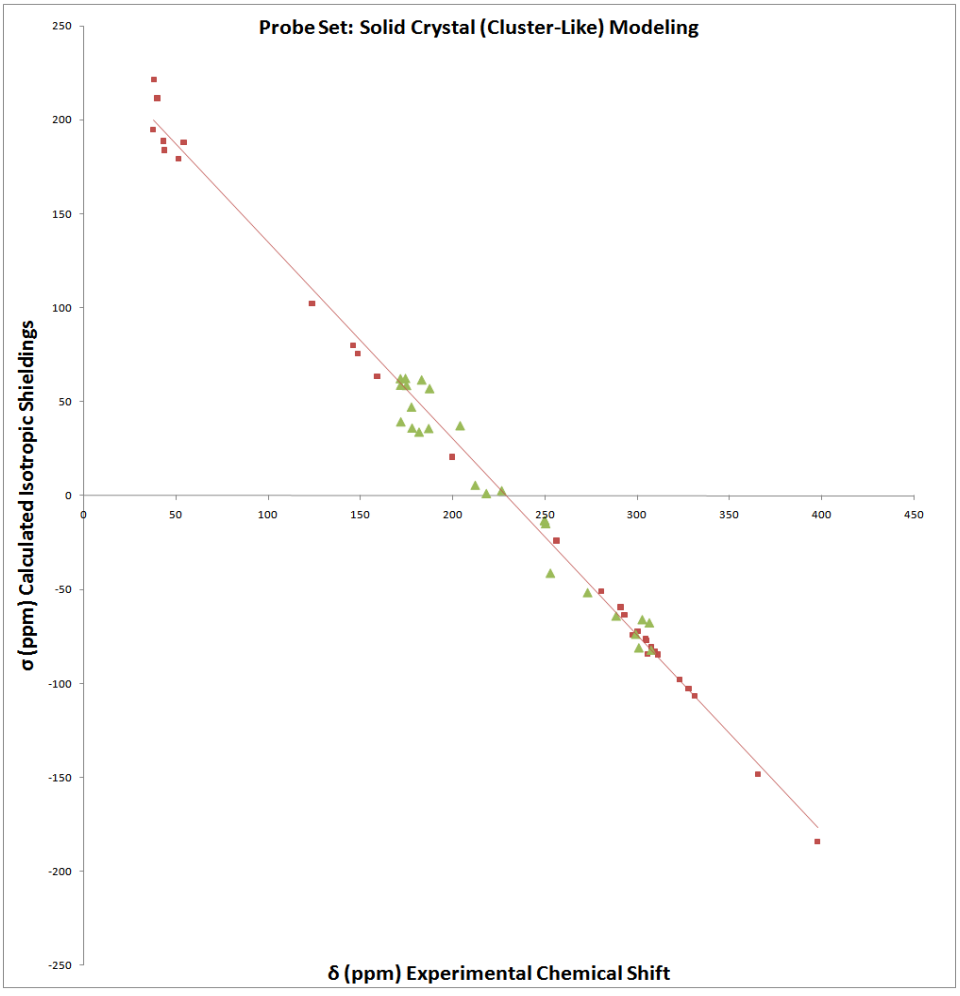
Figure 33b.



2. Solid Crystal (Cluster-like model probe set) plotted about linear regression formed by improved CHLOROFORM-CPCM test set.

Plotting all 26 nitrogen data points produced from a probe set of 16 individual solid crystal cluster-like models (Limbach and McDermott) shows an indiscriminate distribution of points localized about the improved linear regression of Chloroform-CPCM test set data. These disbursed probe molecules, when measured to the line formed by the improved Chloroform test set, resulted in an overall RMSD of 9.738. This RMSD value is considerably lower than the RMSD produced when measuring to the line formed by the improved DMSO test set, 11.943.

Figure 34.



3. Solid Crystal (Gas-phase-like model probe set) plotted about linear regression formed by improved CHLOROFORM-CPCM test set.

As a test of concept it can be imagined that by directing a model towards a gas phase like environment would result in RMSDs that are vastly larger than that of more genuine models making use of a cluster approach. Therefore we tested this idea by removing all previously geometry optimized clusters surrounding the central molecule containing the nitrogen(s) of interest as seen in Figure 32. This insured that the difference in the NMR shieldings illustrate the direct effect of the absence of the clusters and do not depend on atomic geometries between these methods: clustered and gas-phase. Also, and not as trivial, all protons found between close neighbors of the pyridine-like nitrogen and carboxylate oxygen were included in the NMR calculations, namely GODNAO, GODNES, GODNIW, and GEHHEH. This inclusion permits *only* influences from clustering and intentionally disregards the argument of whether a cluster (regardless of tautomeric proportion) is the donor or acceptor; addressing the fundamental issue of whether there exists a loosely covalently-bound or tightly hydrogen-bonded proton between two clusters. In fact the results show that all 26 nitrogen data points from this simplified probe set including: 16 non-clustered (gas-phase-like) models measured to the line produced by the CPCM-Chloroform test set resulted in an overall RMSD of 19.668. This provides direct, inarguable evidence that as the computational modeling of the chemical environment diverges from and loses resemblance to the experimental environment, the RMSDs of the probe set disbursed about the well defined test set increase drastically. It is this phenomina that motivates the entire investigation established here and more importantly our ability to quantify without bias, the effects of environmental changes between modeling systems.

Figure 35.

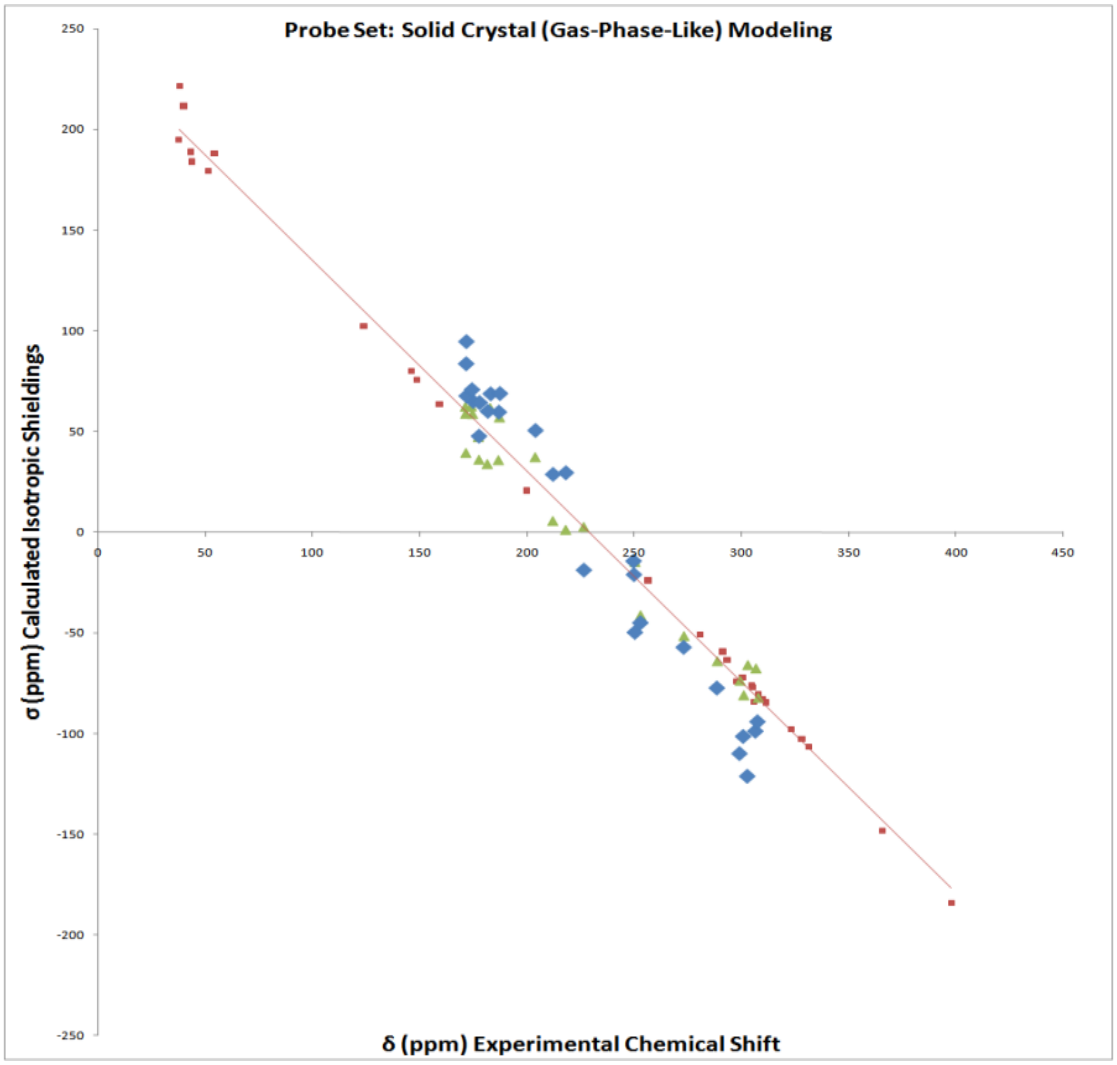

Probe Set: Solid Crystal (Gas-Phase-Like) Modeling
σ (ppm) Calculated Isotropic Shieldings
δ (ppm) Experimental Chemical Shift
250
200
150
100
50
0
-50
-100
-150
-200
-250
0
50
100
150
200
250
300
350
400
450

Figure 36. Generating input files for NMR calculation from geometry optimized clusters by deleting clusters surrounding molecule containing nitrogen(s) of choice.

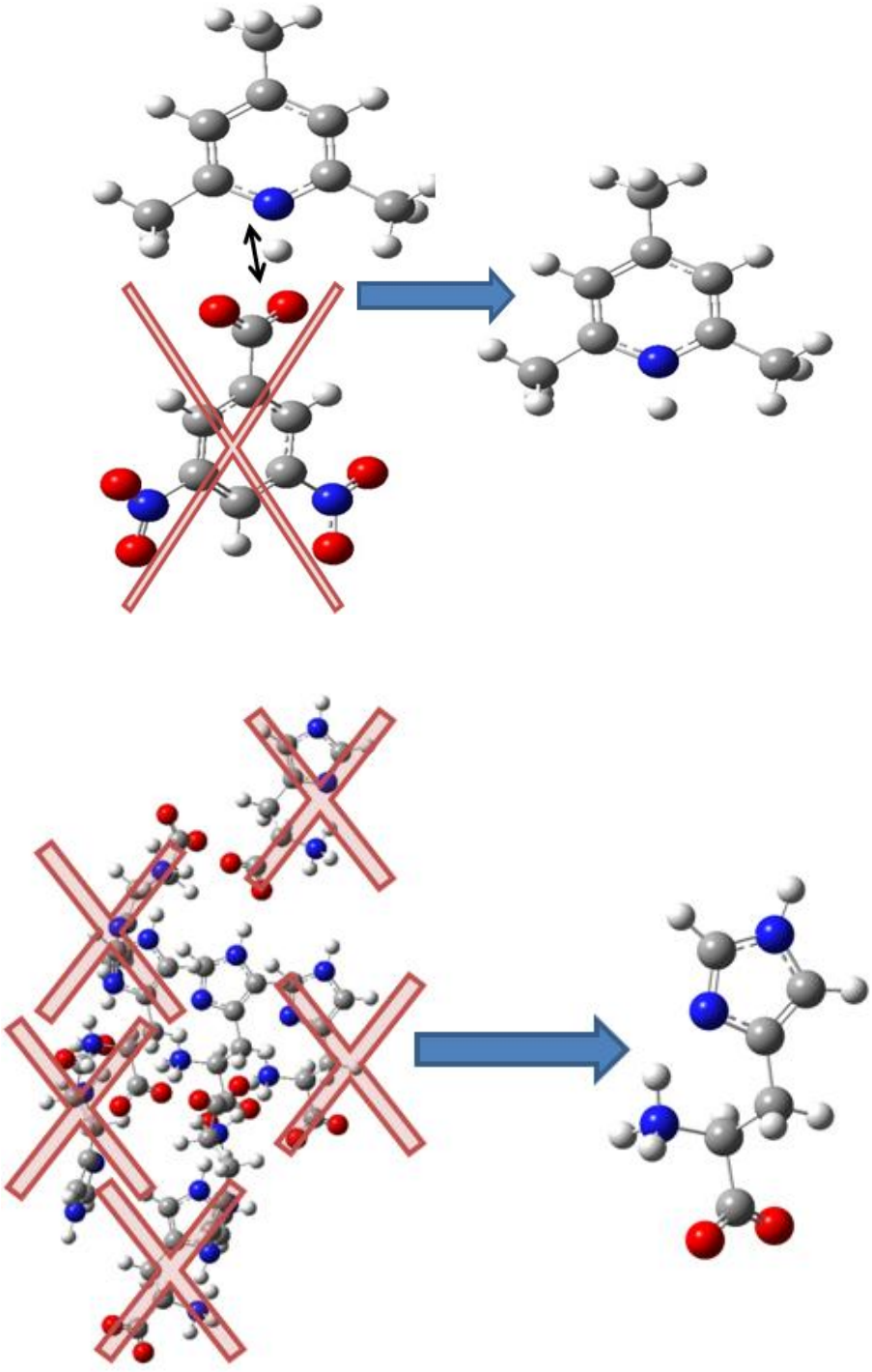


Table 10. Liquid state experimental NMR shift and calculated shielding values of Nitrogen containing compounds in solvents Chloroform and DMSO.

Structures		Chloroform		DMSO	
		δ_{EXPT}	σ_{CALCD}	δ_{EXPT}	σ_{CALCD}
Pyrrolidine	N	37.9	197.0	37.3	195.3
N-Methylpyrrolidine	N	43.3	188.8	42.1	188.6
Piperidine	N	38.2	221.7	36.9	221.4
N-Methylpiperidine	N	40.1	211.7	38.7	211.5
Aniline	N	54.4	188.2	59.0	186.2
N-Methylaniline	N	51.5	179.4	53.0	177.6
N,N-Dimethylaniline	N	43.9	183.9	44.8	182.0
Pyrrole	N-1	146.2	80.3	155.1	75.5
N-Methylpyrrole	N-1	148.8	75.6	150.0	72.9
N-Methylpyrazole	N-1	199.9	20.7	202.2	19.6
	N-2	304.9	-75.8	308.3	-70.7
N-Methylimidazole	N-1	159.3	63.5	161.5	59.6
	N-3	256.5	-23.8	261.6	-17.3
Pyridine	N	311.6	-84.4	316.7	-78.5
2-Picoline	N	310.1	-82.9	315.7	-77.5
3-Picoline	N	305.9	-83.0	311.6	-78.1
4-Picoline	N	297.6	-73.9	303.4	-67.8
2,6-Lutidine	N	309.5	-82.9	315.3	-78.0
Pyridazine	N	397.9	-183.8	400.0	-172.1
Pyrimidine	N	293.3	-63.1	295.3	-59.3
Pyrazine	N	331.5	-106.3	333.8	-102.7
Indole	N	124.1	102.5	134.0	98.0
Quinoline	N	308.0	-80.3	313.0	-74.3
Isoquinoline	N	305.4	-76.8	310.8	-71.1
Phthalazine	N	365.9	-147.9	369.4	-137.2
Quinazoline	N1	280.8	-50.5	283.1	-46.9
	N3	291.3	-59.2	294.4	-54.7
Quinoxaline	N	328.2	-102.6	329.9	-99.0
Acridine	N	300.6	-71.9	306.0	-65.1
Phenazine	N	323.3	-97.8	326.4	-94.2

[a] Reference[1], Relative to external ammonia.

Table 11. Experimental solid state chemical shift δ and calculated shielding σ values for nitrogen containing model compounds.

CSD Code		δ_{EXPT}	σ_{CALCD} Clustered	σ_{CALCD} Gas-phase-like
GODNAO[a]	N	273.2	-51.5532	-57.1987
GODNIW[a]	N	218.2	1.1457	29.3343
GODNES[a]	N	212.2	5.5327	28.541
BIHKEI01	N	204.0	37.2025	50.2964
BITZAF[b]	N	288.7	-64.0261	-77.3718
GEHHAD[b]	N_{pn}	301.0	-80.8598	-101.37
	N_{sb}	299.3	-73.6455	-109.866
GEHHEH[b]	N_{pn}	226.6	2.65235	-18.9162
	N_{sb}	306.7	-67.5847	-98.7922
GEHHIL[b]	N_{pn}	307.7	-82.1843	-94.0678
	N_{sb}	302.9	-65.8879	-121.111
BALHIS01[c]	N_{ε}-H	171.7	58.7415	94.3316
	N_{δ}...NH_3	250.4	-14.8218	-49.7833
JUKMOR[c]	N_{δ}-H	177.6	47.1139	47.4811
	N_{ε}...NH_3	253	-41.2095	-45.0552
LHISTD02[c]	N_{ε}-H	171.8	39.2826	67.4083
	N_{δ}...NH_3	250	-13.2667	-21.1517
LHISTD13[c]	N_{ε}-H	171.6	62.1978	83.3649
	N_{δ}...NH_3	249.8	-13.1501	-14.4577
RARXOX[c,d]	N_{ε}-H	187.5	56.9255	68.6178
	N_{δ}-H	177.6	35.9379	64.1571
TEJWAG[c]	N_{ε}-H	183.1	61.5432	68.4725
LHISTM[d]	N_{δ}-H	186.9	35.6998	59.4184
	N_{ε}-H	174.9	58.6285	64.7375
TIWXEC[d]	N_{ε}-H	181.7	33.7849	59.8173
	N_{δ}-H	174.3	62.3503	70.5128

a Reference[2], b Reference[3], c Reference[4], d Reference[5]

Relative to external ammonia; NH3 @ -50C (NH4Cl is 35.9ppm downfield from NH3 at 0ppm)

Table 12. Linear regression results

Data Set	Number of Structures in Library	Data Points	$\delta_{EXPT} = [(\text{intercept}) - \sigma_{CALCD}]/(-\text{slope})$		RMSD	R^2	Figure
			Intercept	Slope			
Chloroform[a]	27	30	243.03	-1.10	6.31	0.9971	1a.
DMSO[a]	27	30	244.19	-1.11	7.58	0.9959	1a.
$CPCM_{Chloroform}$	27	30	239.27	-1.05	6.69	0.9967	1b.
$CPCM_{DMSO}$	27	30	236.61	-1.01	7.78	0.9957	1b.
Crystal Cluster Model	16	26			9.74[b]		2
Crystal Gas-Phase-Like Model	16	26			19.67[b]		3

[a] Reference[1], [b] RMSD calculated from the deviation of the 26 data points from the $Crystal_{Cluster\ Model}$ to the line fit from $CPCM_{Chloroform}$, NA= not applicable

CONCLUSION

The performance of theoretical NMR calculations and the resulting eligibility for routine practical use were assessed from convergence of R^2 to 1 and slope to -1.0, and an improvement of the original linear regression of Dokalik et al. was established by re-modeling the computational technique employed, notably a CPCM. This re-modeling resulted in an overall decrease in systematic error and further pushed the original linear regression results from Dokalik et al. into the well-performing method regime and improved the accuracy and precision for conversion of calculated isotropic shielding to chemical shift. Most noticeably this improved linear regression test set revealed that the computational modeling resulting from the chloroform CPCM captures, more effectively, the experimental environment ultimately producing a genuine foundation from which RMSDs of a broader range of diverse model systems can be compared. Further, a cluster-like approach for solid state structures (probe set) showed an indiscriminate distribution of points localized about the improved linear regression of the Chloroform-CPCM test set data. The difference from data points of the independently well-performing probe set, measured to the line of this test set, yielded an overall RMSD of 9.738. Finally, as a proof of concept, a simplified gas-phase-like probe set measured to the line produced by the CPCM-Chloroform test set resulted in an overall RMSD of 19.668. This provides direct evidence that as the computational modeling of the chemical environment diverges from and loses resemblance to the experimental environment, the RMSDs of the probe set disbursed about the well defined test set increase drastically. It is this phenomena that motivates the entire investigation established in this study, and equally important, the ability to quantify the effects of environmental changes between modeling systems.

REFERENCES

(1) Dokalik, A.; Kalchhauser, H.; Mikenda, W.; Schweng, G. Magnetic Resonance in Chemistry 1999, 37, 895.

(2) Lorente, P.; Shenderovich, I. G.; Golubev, N. S.; Denisov, G. S.; Buntkowsky, G.; Limbach, H.-H. Magnetic Resonance in Chemistry 2001, 39, S18.

(3) Sharif, S.; Schagen, D.; Toney, M. D.; Limbach, H. H. Journal of the American Chemical Society 2007, 129, 4440.

(4) Song, X.-j.; Rienstra, C. M.; McDermott, A. E. Magnetic Resonance in Chemistry 2001, 39, S30.

(5) Wei, Y.; de Dios, A. C.; McDermott, A. E. Journal of the American Chemical Society 1999, 121, 10389.

(6) Barone, V.; Cossi, M. The Journal of Physical Chemistry A 1998, 102, 1995.

(7) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Journal of computational chemistry 2003, 24, 669.

(8) Forsyth, D. A.; Sebag, A. B. Journal of the American Chemical Society 1997, 119, 9483.

(9) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Chem Rev 2012, 112, 1839.

(11) Gaussian 09, Revision D.01, M. J. Frisch, 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. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. 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, Inc., Wallingford CT, 2009.

(12) http://www.ccdc.cam.ac.uk/pages/Home.aspx